Equivariance and Invariance Properties of Multivariate Quantile and Related Functions, and the Role of Standardization

Robert Serfling[1]

University of Texas at Dallas

Final Preprint Version for *Journal of Nonparametric Statistics*
November 2009

[1]Department of Mathematical Sciences, University of Texas at Dallas, Richardson, Texas 75083-0688, USA. Email: serfling@utdallas.edu. Website: www.utdallas.edu/~serfling. Support by NSF Grants DMS-0103698 and DMS-0805786 and NSA Grant H98230-08-1-0106 is gratefully acknowledged.

Abstract

Equivariance and invariance issues arise as a fundamental but often problematic aspect of multivariate statistical analysis. For multivariate quantile and related functions, we provide coherent definitions of these properties. For standardization of multivariate data to produce equivariance or invariance of procedures, three important types of matrix-valued functional are studied: "weak covariance" (or "shape"), "transformation-retransformation" (TR), and "strong invariant coordinate system" (SICS). Clarification of TR affine equivariant versions of the sample spatial quantile function is obtained. It is seen that geometric artifacts of SICS-standardized data are invariant under affine transformation of the original data followed by standardization using the same SICS functional, subject only to translation and homogeneous scale change. Some applications of SICS standardization are described.

AMS 2000 Subject Classification: Primary 62G99 Secondary 62H99.

Key words and phrases: Multivariate quantile functions; Multivariate outlyingness functions; Affine equivariance and invariance; Covariance functionals; Transformation-retransformation functionals; Invariant coordinate system functionals; Nonparametric methods.

1 Introduction

Equivariance and invariance issues arise as a fundamental but often problematic aspect of multivariate statistical analysis. In particular, it is highly important that multivariate quantile functions and closely associated depth, outlyingness, and rank functions all satisfy appropriate equivariance and invariance criteria. For these, however, the limited existing definitions are rather *ad hoc*. Here we provide coherent formal definitions that unify and extend previous notions. As a second, closely related thrust, we treat the role of suitable standardization of multivariate data in order to produce desired equivariance or invariance of statistical procedures. Three relevant types of matrix-valued functional are studied in detail: "*weak covariance*" (WC), or "shape", "*transformation-retransformation*" (TR), and "*strong invariant coordinate system*" (SICS). Certain clarifications of the well-established TR approach to construct affine equivariant versions of the sample spatial quantile function are obtained. For SICS-standardized data, it is seen that geometric artifacts and patterns are invariant under affine transformation of the original data followed by restandardization using the same SICS transformation, subject only to translation and homogeneous scale change. Applications of SICS standardization are discussed, including construction of new highly robust, computationally simple, affine invariant outlyingness functions.

To proceed to a more detailed description of these results, let us consider as a benchmark the *standardization* of points $\boldsymbol{x}$ in $\mathbb{R}^d$ that was introduced by Mahalanobis (1936),

$$\boldsymbol{\Sigma}(F)^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu}(F)), \qquad (1)$$

relative to a distribution F on $\mathbb{R}^d$, with $\boldsymbol{\mu}(F)$ and $\boldsymbol{\Sigma}(F)$ location and covariance functionals, respectively, and with $\boldsymbol{C}^{1/2}$ denoting any square root of a matrix $\boldsymbol{C}$. In current practice, robust choices of $\boldsymbol{\mu}(F)$ and $\boldsymbol{\Sigma}(F)$ are used. The related *Mahalanobis distance*

$$\|\boldsymbol{\Sigma}(F)^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu}(F))\|,$$

with $\|\cdot\|$ the usual Euclidean distance, has become a fundamental tool widely used for many purposes in multivariate statistics and data mining. A leading application is to consider it as an *outlyingness function*, say $O_{\rm MD}(\boldsymbol{x}, F)$, measuring the outlyingness of a point $\boldsymbol{x}$ relative to the distribution F. It is immediate that $O_{\rm MD}(\boldsymbol{x}, F)$ is *affine invariant*: for $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$ and $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ with any nonsingular $d \times d$ $\boldsymbol{A}$ and any $\boldsymbol{b}$, we have $O_{\rm MD}(\boldsymbol{y}, F_{\boldsymbol{Y}}) = O_{\rm MD}(\boldsymbol{x}, F_{\boldsymbol{X}})$. While this approach has mathematical tractability and strong intuitive appeal, it also has a limitation: the contours of equal outlyingness are necessarily *ellipsoidal*, regardless of whether F is elliptically symmetric.

Here we consider outlyingness measures more broadly, within a formulation of four closely associated entities: multivariate *quantile*, *depth*, *outlyingness*, and *centered rank* functions. We exhibit a general "paradigm" that expresses the *equivalence* of these four functions. In this setting, we study issues of their *equivariance and invariance*. In particular, *quantile* functions should be *affinely equivariant*, and *outlyingness* functions *affinely invariant*. That is, the quantile representation of a point after affine transformation should agree with its original quantile representation similarly transformed, and its outlyingness measure should remain unchanged. Central to our development is a general definition of *affine equivariance for multivariate quantile functions*, which then yields corresponding versions for the above-mentioned related functions. This is nontrivial since the affine equivariance property for a quantile function must necessarily include a suitable re-indexing that enables the associated outlyingness function simultaneously to be affine invariant.

There arises the question of whether and how a statistical procedure can be modified, if necessary, to obtain an affinely equivariant or invariant version. In this connection, we study the role of *standardization of multivariate data* for construction of multivariate quantile and related functions that indeed possess desired equivariance or invariance properties, focusing on the above-mentioned WC, TR, and SICS functionals. In particular, the well-known scatter functional of Tyler (1987) and its symmetrized version given by Dümbgen (1988) are TR functionals. It is seen that WC functionals admit square-root like factorizations in terms of TR functionals, and that SICS functionals form a very important and productive special case of TR functional. It also is shown that *any TR functional* suffices to create affine equivariant modifications of the sample spatial quantile function (defined in Section 2).

This provides a broader view of a particular TR transformation used for constructing affine equivariant versions of the sample spatial quantiles (Chakraborty and Chaudhuri, 1996, Chakraborty, Chaudhuri, and Oja, 1998, and Chakraborty, 2001). It is seen that what is actually estimated by the so-constructed "TR sample spatial quantile function" is not the population spatial quantile function but rather a different, although closely related, "Malahanobis spatial quantile function" which is associated with the particular choice of TR functional. In comparison with the *Mahalanobis distance* quantile function, such *Mahalanobis spatial* quantile functions yield more flexible outlyingness contours and are worth studying in their own right. These results strengthen the motivation for the TR approach as a practical method, widening the scope and technique of its application.

Tyler, Critchley, Dümbgen, and Oja (2009) define "invariant coordinate system" (ICS) functionals and study them in depth. Here we introduce "strong invariant coordinate system" (SICS) functionals, which represent a special case of both ICS and TR functionals and play an important practical role: the geometric structure of SICS-standardized multivariate data is affine invariant in the sense that *the SICS-standardized transformed data agrees with the SICS-standardized original data*, subject only to *a translation* and *a homogeneous scale change*. That is, standardization by a SICS functional eliminates structure and pattern which are merely artifacts of the particular coordinate system and thus are spurious despite any striking appearance. The SICS approach may be used to correct the lack of full affine equivariance or invariance of some existing multivariate statistical procedures already in the literature. The above-mentioned TR transformation of Chakraborty and Chaudhuri (1996) is a SICS functional. We provide other examples of SICS functionals and also mention well-known TR functionals which are not SICS. Finally, we indicate applications of SICS standardization, including development of new multivariate outlyingness functions that are *highly robust*, *affine invariant*, and *computationally simple*, by adapting projection pursuit using just a (relatively) few projections of the multivariate points onto lines.

Our treatment is organized as follows. Multivariate depth, outlyingness, quantile, and centered rank functions are defined precisely in Section 2, within a coherent framework. Their equivariance and invariance properties are formulated and exemplified in Section 3. WC and TR functionals are treated in Section 4, discussing their interconnections and some important TR functionals. The role of TR functionals in defining "Mahalanobis spatial quantile functions" is examined in Section 5. In Section 6 we treat SICS functionals, their connections with TR functionals, and aspects of their application.

2 Multivariate Depth, Outlyingness, Quantile, and Rank functions

In order to formulate precise notions of equivariance and invariance for multivariate depth, outlyingness, quantile, and rank functions, these functions must themselves be defined in a precise and coherent manner. Here we provide relevant definitions and several examples, drawing upon and elaborating a previous treatment in Serfling (2006).

2.1 Definitions

2.1.1 Quantile Functions in $\mathbb{R}$ and $\mathbb{R}^d$

For a univariate distribution F, the quantile function is $F^{-1}(p) = \inf\{x : F(x) \geq p\}$, $0 < p < 1$. As a preliminary to extending to the multivariate case, where a natural linear order is lacking, we orient to a "center" by re-indexing to the open interval $(-1, 1)$ via $u = 2p - 1$ and representing the quantile function as $Q(u, F) = F^{-1}\left(\frac{1+u}{2}\right)$, $-1 < u < 1$. Each point $x \in \mathbb{R}$ has a quantile representation $x = Q(u, F)$ for some choice of u. The *median* is $Q(0, F)$. For $u \neq 0$, the index u indicates through its sign *the direction of* x *from the median* and through its magnitude $|u|$ a measure of the *outlyingness of* x *from the median*. For $|u| = c \in (0, 1)$, the "contour" $\left\{F^{-1}\left(\frac{1-c}{2}\right), F^{-1}\left(\frac{1+c}{2}\right)\right\}$ demarks the upper and lower tails of equal probability weight $\frac{1-c}{2}$. Then $|u| = c$ is also the *probability weight* of the enclosed "central region".

To extend to a parallel formulation for a distribution F on $\mathbb{R}^d$, an associated *quantile function* is indexed by $\boldsymbol{u}$ in the unit ball $\mathbb{B}^{d-1}(\mathbf{0})$ in $\mathbb{R}^d$, attaches to each point $\boldsymbol{x}$ a *quantile representation* $\boldsymbol{Q}(\boldsymbol{u}, F)$, and generates *nested* contours $\{\boldsymbol{Q}(\boldsymbol{u}, F) : \|\boldsymbol{u}\| = c\}$, $0 \leq c < 1$. For $\boldsymbol{u} = \mathbf{0}$, the most central point $\boldsymbol{Q}(\mathbf{0}, F)$ is interpreted as a *d-dimensional median* $\boldsymbol{M}_F$. For $\boldsymbol{u} \neq \mathbf{0}$, the index $\boldsymbol{u}$ represents *direction* in some sense, for example, direction to $\boldsymbol{Q}(\boldsymbol{u}, F)$ from $\boldsymbol{M}_F$, or *expected* direction to $\boldsymbol{Q}(\boldsymbol{u}, F)$ from random $\boldsymbol{X} \sim F$. The magnitude $\|\boldsymbol{u}\|$ represents an *outlyingness parameter*, higher values corresponding to more extreme points. The contours for $\|\boldsymbol{u}\| = c$ thus represent equivalence classes of points of equal outlyingness. (But in general c need not be the enclosed probability weight.)

2.1.2 Centered rank, outlyingness, and depth functions

Three functions closely related to the quantile function $\boldsymbol{Q}(\boldsymbol{u}, F)$, have special meanings and distinct roles.

Centered rank function. The quantile function $\boldsymbol{Q}(\boldsymbol{u}, F)$, $\boldsymbol{u} \in \mathbb{B}^{d-1}(\mathbf{0})$, has an *inverse*, given at each point $\boldsymbol{x} \in \mathbb{R}^d$ by the point $\boldsymbol{u}$ in $\mathbb{B}^{d-1}(\mathbf{0})$ for which $\boldsymbol{x}$ has $\boldsymbol{Q}(\boldsymbol{u}, F)$, as its quantile representation, i.e., by the solution $\boldsymbol{u}$ of the equation

$$\boldsymbol{x} = \boldsymbol{Q}(\boldsymbol{u}, F).$$

(Under the typical condition that the nested contours of $\boldsymbol{Q}(\cdot, F)$ do not intersect for different $\boldsymbol{u}$, the solution $\boldsymbol{u}$ is unique. In the univariate case, it suffices that F is strictly increasing.) The solutions $\boldsymbol{u}$ taken over different $\boldsymbol{x}$ define a *centered rank function* $\boldsymbol{R}(\boldsymbol{x}, F)$, $\boldsymbol{x} \in \mathbb{R}^d$, which takes values in $\mathbb{B}^{d-1}(\mathbf{0})$, with the origin assigned as rank of the multivariate median $\boldsymbol{Q}(\mathbf{0}, F)$. Thus $\boldsymbol{R}(\boldsymbol{x}, F) = \mathbf{0}$ for $\boldsymbol{x} = \boldsymbol{Q}(\mathbf{0}, F)$, and $\boldsymbol{R}(\boldsymbol{x}, F)$ gives a "directional rank" in $\mathbb{B}^{d-1}(\mathbf{0})$ for other $\boldsymbol{x}$. In the univariate case, we have $R(x, F) = 2F(x) - 1$, with its *sign* giving "direction" (from the median $F^{-1}(1/2)$), and its *magnitude* providing the "rank" of x. For testing $H_0 : \boldsymbol{M}_F = \boldsymbol{\theta}_0$, the sample version of $\boldsymbol{R}(\boldsymbol{\theta}_0, F)$ provides a natural *test statistic*, a multivariate version of the *univariate sign test*.

Outlyingness function. The magnitude $\|\boldsymbol{R}(\boldsymbol{x}, F)\|$ of the centered rank function defines an *outlyingness function* $O(\boldsymbol{x}, F)$, $\boldsymbol{x} \in \mathbb{R}^d$, giving a *center-inward ordering* of points $\boldsymbol{x}$ in $\mathbb{R}^d$. Higher values represent greater "outlyingness".

Depth function. A corresponding *depth function* $D(\boldsymbol{x}, F) = 1 - O(\boldsymbol{x}, F)$ provides a *center-outward ordering* of points $\boldsymbol{x}$ in $\mathbb{R}^d$, higher depth corresponding to higher "centrality". Multivariate data points may be ordered by their sample depths.

2.1.3 Equivalence: a "D-O-Q-R paradigm"

Under suitable regularity conditions on F, *the associated depth, outlyingness, quantile, and rank functions are equivalent*:

- $\boldsymbol{Q}(\boldsymbol{u}, F)$ and $\boldsymbol{R}(\boldsymbol{x}, F)$ are equivalent (inversely).
- $D(\boldsymbol{x}, F)$ and $O(\boldsymbol{x}, F)$ are equivalent (inversely).
- These are linked by

  a) $O(\boldsymbol{x}, F) = \|\boldsymbol{R}(\boldsymbol{x}, F)\|$ $(= \|\boldsymbol{u}\|)$,

  b) $D(\boldsymbol{x}, F)$ induces a corresponding $\boldsymbol{Q}(\boldsymbol{u}, F)$. (See Example 2.1 below.)

We emphasize that the inverse of $\boldsymbol{Q}(\boldsymbol{u}, F)$ is not F but rather the associated $\boldsymbol{R}(\boldsymbol{x}, F)$, which in the univariate case conveniently happens to be equivalent to F. Each of D, O, $\boldsymbol{Q}$, and $\boldsymbol{R}$ thus generates the others, and technically they all may be used interchangeably, but each provides a different conceptual and intuitive standpoint. As detailed in Remark 2.1, the "$\boldsymbol{Q}$" generated by "D" can differ in its indexing from that "$\boldsymbol{Q}$" which generates "D", although the contours will still be the same.

2.2 Examples

In view of the paradigm of Section 2.1.3, one may introduce any particular example by specifying a particular depth function, outlyingness function, quantile function, or rank function, as may be convenient.

Example 2.1 *Depth-induced quantile functions.* For $D(\boldsymbol{x}, F)$ possessing nested contours enclosing the "median" $\boldsymbol{M}_F$ and bounding "central regions" of form $\{\boldsymbol{x} : D(\boldsymbol{x}, F) \geq \alpha\}$, $\alpha > 0$, the depth contours induce $\boldsymbol{Q}(\boldsymbol{u}, F)$, $\boldsymbol{u} \in \mathbb{B}^{d-1}(\mathbf{0})$, with each $\boldsymbol{x} \in \mathbb{R}^d$ given a quantile representation, as follows. For $\boldsymbol{x} = \boldsymbol{M}_F$, denote it by $\boldsymbol{Q}(\mathbf{0}, F)$. For $\boldsymbol{x} \neq \boldsymbol{M}_F$, denote it by $\boldsymbol{Q}(\boldsymbol{u}, F)$ with $\boldsymbol{u} = p\boldsymbol{v}$, where p is the probability weight of the central region with $\boldsymbol{x}$ on its boundary and $\boldsymbol{v}$ is the unit vector toward $\boldsymbol{x}$ from $\boldsymbol{M}_F$. In this case, $\boldsymbol{u} = \boldsymbol{R}(\boldsymbol{x}, F)$ indicates direction toward $\boldsymbol{x} = \boldsymbol{Q}(\boldsymbol{u}, F)$ from $\boldsymbol{M}_F$, and the outlyingness parameter $\|\boldsymbol{u}\| = \|\boldsymbol{R}(\boldsymbol{x}, F)\|$ is the probability weight of the central region with $\boldsymbol{Q}(\boldsymbol{u}, F)$ on its boundary.

All of the various depth functions considered in Liu, Parelius, and Singh (1999) and in Zuo and Serfling (2000), for example, induce associated outlyingness, quantile, and rank functions. In particular, included here is the *halfspace depth* studied in detail by Donoho and Gasko (1992) and, recently, with new perspectives by Kong and Mizera (2008) and Hallin, Paindaveine, and Šiman (2009). □

Example 2.2 *The spatial quantile function.* For univariate Z with $E|Z| < \infty$, the pth quantile for $0 < p < 1$ may be characterized as any value θ minimizing

$$E\{|Z - \theta| + (2p - 1)(Z - \theta)\} \tag{2}$$

(Ferguson, 1967, p. 51). In the median-centered notation via $u = 2p - 1$, and defining $\Phi(u, t) = |t| + ut$, $-1 < u < 1$, we may equivalently obtain θ by minimizing

$$E\{\Phi(u, Z - \theta) - \Phi(u, Z)\}, \tag{3}$$

where subtraction of $\Phi(u, Z)$ eliminates the need of a moment assumption on Z. As a multivariate extension, d-dimensional "spatial" or "geometric" quantiles were introduced by Dudley and Koltchinskii (1992) and Chaudhuri (1996). Following the latter, we extend the index set to the open unit ball $\mathbb{B}^{d-1}(\mathbf{0})$ and minimize a generalized form of (3). Specifically, for random vector $\boldsymbol{X}$ having cdf F on $\mathbb{R}^d$, and for $\boldsymbol{u}$ in $\mathbb{B}^{d-1}(\mathbf{0})$, the *$\boldsymbol{u}$th spatial quantile* $\boldsymbol{Q}_{\mathrm{S}}(\boldsymbol{u}, F)$ is given by $\boldsymbol{\theta}$ minimizing

$$E\{\Phi(\boldsymbol{u}, \boldsymbol{X} - \boldsymbol{\theta}) - \Phi(\boldsymbol{u}, \boldsymbol{X})\}, \tag{4}$$

where $\Phi(\boldsymbol{u}, \boldsymbol{t}) = \|\boldsymbol{t}\| + \boldsymbol{u}'\boldsymbol{t}$. In particular, $\boldsymbol{Q}_{\mathrm{S}}(\mathbf{0}, F)$ is the well-known *spatial median.*

Equivalently, in terms of the *spatial sign function* (or *unit vector function*),

$$\boldsymbol{S}(\boldsymbol{x}) = \begin{cases} \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}, & \boldsymbol{x} \in \mathbb{R}^d, \boldsymbol{x} \neq \mathbf{0}, \\ \mathbf{0}, & \boldsymbol{x} = \mathbf{0}, \end{cases}$$

the quantile $\boldsymbol{Q}_{\mathrm{S}}(\boldsymbol{u}, F)$ may be represented as the solution $\boldsymbol{x}$ of

$$E\{\boldsymbol{S}(\boldsymbol{x}-\boldsymbol{X})\} = \boldsymbol{u}. \tag{5}$$

By (5), $\boldsymbol{Q}_{\mathrm{S}}(\boldsymbol{u}, F)$ is obtained by inverting the map $\boldsymbol{x} \mapsto E\{\boldsymbol{S}(\boldsymbol{x}-\boldsymbol{X})\}$, from which it is seen that spatial quantiles are a special case of the "M-quantiles" introduced by Breckling and Chambers (1988) and also treated by Koltchinskii (1997) and Breckling, Kokic and Lübke (2001). Solving (5) for $\boldsymbol{u}$ yields the *spatial centered rank function*,

$$\boldsymbol{R}_{\mathrm{S}}(\boldsymbol{x}, F) = E\{\boldsymbol{S}(\boldsymbol{x}-\boldsymbol{X})\}, \tag{6}$$

the *expected direction* to $\boldsymbol{x}$ from a random point $\boldsymbol{X} \sim F$. (Compare with the interpretation of the *depth-induced* $\boldsymbol{R}(\boldsymbol{x}, F)$ as *direction from the median*, as seen in Example 2.1.) Möttönen and Oja (1995) apply the function $\boldsymbol{R}_{\mathrm{S}}(\boldsymbol{x}, F)$ in hypothesis testing, and Vardi and Zhang (2000) treat the related *spatial depth function*, $D_{\mathrm{S}}(\boldsymbol{x}, F) = 1 - \|\boldsymbol{R}_{\mathrm{S}}(\boldsymbol{x}, F)\|$, whose inverse $O_{\mathrm{S}}(\boldsymbol{x}, F) = \|\boldsymbol{R}_{\mathrm{S}}(\boldsymbol{x}, F)\|$ $(= \|\boldsymbol{u}\|$ relative to equation (5)) may be interpreted as a *spatial outlyingness function*. Koltchinskii (1994a,b, 1997) develops a Bahadur-Kiefer represention for the sample spatial quantile function and derives other theoretical properties of $\boldsymbol{Q}_{\mathrm{S}}(\boldsymbol{u}, F)$. Serfling (2004) treats further properties of $\boldsymbol{Q}_{\mathrm{S}}(\boldsymbol{u}, F)$ and introduces related nonparametric multivariate descriptive measures for location, spread, skewness, and kurtosis. Extension to *spatial U-quantiles* is provided in Zhou and Serfling (2008). □

Remark 2.1 *Equivalent depth-based expression for the spatial quantile function.* As seen in Example 2.2, $\boldsymbol{Q}_{\mathrm{S}}(\cdot, F_{\boldsymbol{X}})$ generates an associated depth function via

$$\boldsymbol{Q}_{\mathrm{S}}(\cdot, F_{\boldsymbol{X}}) \longrightarrow \boldsymbol{R}_{\mathrm{S}}(\boldsymbol{x}, F_{\boldsymbol{X}}) \longrightarrow O_{\mathrm{S}}(\boldsymbol{x}, F_{\boldsymbol{X}}) \longrightarrow D_{\mathrm{S}}(\boldsymbol{x}, F_{\boldsymbol{X}}) = 1 - \|E\boldsymbol{S}(\boldsymbol{x}-\boldsymbol{X})\|.$$

On the other hand, as seen in Example 2.1, $D_{\mathrm{S}}(\cdot, F_{\boldsymbol{X}})$ itself generates an associated quantile function, say $\widetilde{\boldsymbol{Q}}_{\mathrm{S}}(\cdot, F_{\boldsymbol{X}})$, where $\widetilde{\boldsymbol{Q}}_{\mathrm{S}}(\widetilde{\boldsymbol{u}}, F_{\boldsymbol{X}})$ is the quantile representation of a point $\boldsymbol{x}$ lying on the boundary of some depth-based "central region" $\{\boldsymbol{x} : D_{\mathrm{S}}(\boldsymbol{x}, F_{\boldsymbol{X}}) \geq \alpha\}$ and in the direction $\boldsymbol{u}$ from the median $\boldsymbol{M}_{\boldsymbol{X}} = \boldsymbol{Q}_{\mathrm{S}}(\boldsymbol{0}, F_{\boldsymbol{X}}) = \widetilde{\boldsymbol{Q}}_{\mathrm{S}}(\boldsymbol{0}, F_{\boldsymbol{X}})$. Here $\widetilde{\boldsymbol{u}} = p\boldsymbol{v}$, where p is the probability weight of the central region with $\boldsymbol{x}$ on its boundary and $\boldsymbol{v}$ is the unit vector toward $\boldsymbol{x}$ from $\boldsymbol{M}_{\boldsymbol{X}}$. For this version of the spatial quantile function, the outlyingness parameter $\|\widetilde{\boldsymbol{u}}\|$ is the probability weight of the central region with $\boldsymbol{x} = \widetilde{\boldsymbol{Q}}_{\mathrm{S}}(\widetilde{\boldsymbol{u}}, F_{\boldsymbol{X}})$ on its boundary. While $\boldsymbol{Q}_{\mathrm{S}}(\cdot, F_{\boldsymbol{X}})$ and $\widetilde{\boldsymbol{Q}}_{\mathrm{S}}(\cdot, F_{\boldsymbol{X}})$ generate the same contours, but indexed differently, the correspondence $\boldsymbol{Q}_{\mathrm{S}}(\boldsymbol{u}, F_{\boldsymbol{X}}) = \widetilde{\boldsymbol{Q}}_{\mathrm{S}}(\widetilde{\boldsymbol{u}}, F_{\boldsymbol{X}})$ is nontrivial to characterize explicitly. □

Example 2.3 *"Mahalanobis distance" outlyingness function.* Substitution of multivariate location and spread measures $\boldsymbol{\mu}(F)$ and $\boldsymbol{\Sigma}(F)$ in the classical *univariate scaled deviation outlyingness* $O(x, F) = |(x - \mu(F))/\sigma(F)|$ of Mosteller and Tukey (1977) yields the very popular *Mahalanobis distance outlyingness* $O_{\mathrm{MD}}(\boldsymbol{x}, F) = \|\boldsymbol{\Sigma}(F)^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu}(F))\|$ discussed in Section 1. The outlyingness contours are *ellipsoidal*, whatever the shape of F. □

3 Equivariance of Multivariate Quantile Functions

Quantile functions on $\mathbb{R}^d$ are desirably *equivariant*, and *outlyingness functions* should be *invariant*. That is, the new quantile representation of a point $\boldsymbol{x}$ after affine transformation should agree with the original representation similarly transformed, and its outlyingness measure should remain unchanged. In this section we formalize these requirements in suitable technical definitions, and in subsequent sections we see how to produce desired equivariance or invariance through *standardization using suitable types of transformation.*

Definition 3.1 An $\mathbb{R}^d$-valued quantile function $\boldsymbol{Q}(\boldsymbol{u}, F)$, $\boldsymbol{u} \in \mathbb{B}^{d-1}(\boldsymbol{0})$, is *affine equivariant* if, for $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$ with any nonsingular $d \times d$ $\boldsymbol{A}$ and any $\boldsymbol{b}$,

$$\boldsymbol{Q}(\boldsymbol{v}, F_{\boldsymbol{Y}}) = \boldsymbol{A}\,\boldsymbol{Q}(\boldsymbol{u}, F_{\boldsymbol{X}}) + \boldsymbol{b},\ \ \boldsymbol{u} \in \mathbb{B}^{d-1}(\boldsymbol{0}), \tag{7}$$

with a $\mathbb{B}^{d-1}(\boldsymbol{0})$-valued re-indexing $\boldsymbol{v} = \boldsymbol{v}(\boldsymbol{u}, \boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}})$ which satisfies

$$\|\boldsymbol{v}(\boldsymbol{u}, \boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}})\| = \|\boldsymbol{u}\|,\ \ \boldsymbol{u} \in \mathbb{B}^{d-1}(\boldsymbol{0}). \tag{8}$$

□

For the *median* $\boldsymbol{Q}(\boldsymbol{0}, F_{\boldsymbol{X}})$, the equivariance property may be stated simply $\boldsymbol{Q}(\boldsymbol{0}, F_{\boldsymbol{Y}}) = \boldsymbol{A}\,\boldsymbol{Q}(\boldsymbol{0}, F_{\boldsymbol{X}}) + \boldsymbol{b}$. Note that condition (8) builds *outlyingness invariance* into the definition of affine equivariance of $\boldsymbol{Q}(\cdot, F)$. This is made more precise in Remark 3.1 below.

Denote the family of *contours* of a quantile function $\boldsymbol{Q}(\cdot, F)$ by

$$\widetilde{\boldsymbol{Q}}(c, F) = \{\boldsymbol{Q}(\boldsymbol{u}, F) : \|\boldsymbol{u}\| = c\},\ \ 0 < c < 1.$$

Then, if $\boldsymbol{Q}(\cdot, F)$ is affine equivariant, equivalently so are the contours: for $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$, we have

$$\widetilde{\boldsymbol{Q}}(c, F_{\boldsymbol{Y}}) = \boldsymbol{A}\,\widetilde{\boldsymbol{Q}}(c, F_{\boldsymbol{X}}) + \boldsymbol{b},\ \ 0 < c < 1.$$

Here the mapping $\boldsymbol{u} \mapsto \boldsymbol{v}(\boldsymbol{u}, \boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}})$ is left implicit.

Example 3.1 *The univariate case.* For the univariate quantile function in median-centered notation, $Q(u, F) = F^{-1}(\frac{1+u}{2})$, $-1 < u < 1$, as discussed in Section 2.1 above, the usual translation and scale equivariance takes the form

$$Q(\mathrm{sgn}(a)u, F_{aX+b}) = a\,Q(u, F_X) + b,\ \ -1 < u < 1,$$

for all $a, b \in \mathbb{R}$, which is (7) with $v(u, a, b, F_X) = \mathrm{sgn}(a)u$, satisfying (8). (Equivalently, we could choose $v(u, a, b, F_X) = -\mathrm{sgn}(a)u$.) □

Remark 3.1 The equivariance of $\boldsymbol{Q}(\boldsymbol{u}, F)$ yields corresponding equivariance and invariance properties for the related D, O, and R functions, using their definitions. By the definition of $\boldsymbol{R}(\boldsymbol{x}, F)$, (7) immediately yields *equivariance of the centered rank function* in the following sense:

$$\boldsymbol{R}(\boldsymbol{y}, F_{\boldsymbol{Y}}) = \boldsymbol{v}\left(\boldsymbol{R}(\boldsymbol{x}, F_{\boldsymbol{X}}), \boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}}\right). \tag{9}$$

In turn, the relation $O(\boldsymbol{x}, F) = \|\boldsymbol{R}(\boldsymbol{x}, F)\|$ yields through (8) and (9) *invariance of the outlyingness function*, $O(\boldsymbol{y}, F_{\boldsymbol{Y}}) = O(\boldsymbol{x}, F_{\boldsymbol{X}})$, and likewise of the *depth function*, $D(\boldsymbol{y}, F_{\boldsymbol{Y}}) = D(\boldsymbol{x}, F_{\boldsymbol{X}})$. (These latter also follow from the definition of equivariance of $\boldsymbol{Q}(\cdot, F)$.) □

Example 3.2 $\boldsymbol{v}(\boldsymbol{u}, \boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}})$ *for depth-induced quantile functions.* Suppose that a quantile function $\boldsymbol{Q}(\boldsymbol{u}, F)$ is constructed as in Example 2.1 and satisfies (7). Then (in obvious notation) $\boldsymbol{M_Y} = \boldsymbol{A}\,\boldsymbol{M_X} + \boldsymbol{b}$, from which it follows that the unnormalized direction vector toward $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ from $\boldsymbol{M_Y}$ is given by $\boldsymbol{y} - \boldsymbol{M_Y} = \boldsymbol{A}(\boldsymbol{x} - \boldsymbol{M_X})$. Therefore, for some constant c_0, we have $\boldsymbol{R}(\boldsymbol{y}, F_{\boldsymbol{Y}}) = c_0\boldsymbol{A}\boldsymbol{R}(\boldsymbol{x}, F_{\boldsymbol{X}})$, or, equivalently, $\boldsymbol{v}(\boldsymbol{u}, \boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}}) = c_0\boldsymbol{A}\boldsymbol{u}$. Then the requirement (8) is satisfied if and only if $|c_0| = \|\boldsymbol{u}\|/\|\boldsymbol{A}\boldsymbol{u}\|$, yielding

$$\boldsymbol{v}(\boldsymbol{u}, \boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}}) = \pm\frac{\|\boldsymbol{u}\|}{\|\boldsymbol{A}\boldsymbol{u}\|}\boldsymbol{A}\boldsymbol{u}, \tag{10}$$

for either choice of sign. In the univariate case, (10) reduces to $\pm\text{sgn}(a)u$, in agreement with Example 3.1. Note that for $\boldsymbol{A}$ orthogonal (10) becomes simply

$$\boldsymbol{v}(\boldsymbol{u}, \boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}}) = \pm\boldsymbol{A}\boldsymbol{u}. \tag{11}$$

Many typical depth functions, for example the *halfspace depth*, the *simplicial depth*, the *Mahalanobis distance depth*, and the *projection depth*, are *affine invariant* and so induce corresponding related O, Q, and R functions which are *affine invariant/equivariant*. □

Example 3.3 $\boldsymbol{v}(\boldsymbol{u}, \boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}})$ *for the spatial quantile function.* A well-known limitation of the spatial quantile function is its *orthogonal*, rather than fully affine, equivariance. Indeed, as pointed out by Chaudhuri (1996, Fact 2.2.1), for $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$ with any *orthogonal* $d \times d$ $\boldsymbol{A}$ and any $\boldsymbol{b}$,

$$\boldsymbol{Q}_{\text{S}}(\boldsymbol{A}\boldsymbol{u}, F_{\boldsymbol{Y}}) = \boldsymbol{A}\,\boldsymbol{Q}_{\text{S}}(\boldsymbol{u}, F_{\boldsymbol{X}}) + \boldsymbol{b},\ \boldsymbol{u} \in \mathbb{B}^{d-1}, \tag{12}$$

which corresponds to Definition 3.1 restricted to *orthogonal* $\boldsymbol{A}$ and thus with $\boldsymbol{v}(\boldsymbol{u}, \boldsymbol{A}, F_{\boldsymbol{X}})$ $= \boldsymbol{A}\boldsymbol{u}$. Consequently, a point $\boldsymbol{x}$ labeled a (spatial) "outlier" or "nonoutlier" would have the same classification after *orthogonal* transformation to a new coordinate system but not necessarily after transformation by *heterogeneous scale changes*. □

4 Weak Covariance Functionals and Transformation-Retransformation Functionals

In this section we examine two types of functional that arise in standardizing multivariate data or in transforming it to satisfy some purpose. Application to define "Mahalanobis spatial quantile functions" is carried out in Section 5. A third type of standardizing functional is treated in Section 6.

4.1 Weak Covariance (WC) Functionals

A symmetric positive definite $d \times d$ matrix-valued functional $\boldsymbol{C}(F)$ defined on distributions F on $\mathbb{R}^d$ is called a *covariance functional* if it satisfies *covariance equivariance*, $\boldsymbol{C}(F_{\boldsymbol{A}\boldsymbol{X}+\boldsymbol{b}})$ $= \boldsymbol{A}\boldsymbol{C}(F_{\boldsymbol{X}})\boldsymbol{A}'$ for all vectors $\boldsymbol{b}$ and all nonsingular $d \times d$ A (Rousseeuw and Leroy, 1987, and Lopuhaä and Rousseeuw, 1991). A weak version that suffices for most purposes, yet offers useful additional flexibility, is the following

Definition 4.1 A symmetric positive definite $d \times d$ matrix-valued functional $\boldsymbol{C}(F)$ is called a *weak covariance (WC) functional* if, for $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$ with any nonsingular $\boldsymbol{A}$ and any $\boldsymbol{b}$,

$$\boldsymbol{C}(F_{\boldsymbol{Y}}) = k_1 \boldsymbol{A}\boldsymbol{C}(F_{\boldsymbol{X}})\boldsymbol{A}', \tag{13}$$

with $k_1 = k_1(\boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}})$ a positive scalar function of $\boldsymbol{A}$, $\boldsymbol{b}$, and $F_{\boldsymbol{X}}$. □

The sample version for a data set $\mathbb{X}_n = \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\}$ in $\mathbb{R}^d$ may be expressed, with $\mathbb{Y}_n = \boldsymbol{A}\mathbb{X}_n + \boldsymbol{b}$ and $k_1 = k_1(\boldsymbol{A}, \boldsymbol{b}, \mathbb{X}_n)$, as

$$\boldsymbol{C}_n(\mathbb{Y}_n) = k_1 \boldsymbol{A}\boldsymbol{C}_n(\mathbb{X}_n)\boldsymbol{A}'. \tag{14}$$

WC functionals are known in the literature as "scatter functionals" (e.g., Dümbgen and Tyler, 2005) and "shape functionals" (e.g., Paindaveine, 2008, and Tyler, Critchley, Dümbgen, and Oja, 2009), among other labels. If $\boldsymbol{C}(F)$ is a WC functional, then so is any multiple, and sometimes in practice a canonical representative is chosen by imposing some normalizing condition such as $\det(\boldsymbol{C}(F)) = 1$ or $\mathrm{trace}(\boldsymbol{C}(F)) = d$. The determinant-based normalization is shown in Paindaveine (2008) to possess uniquely certain favorable decision-theoretic properties in LAN elliptical and semiparametric elliptical families. Many statistical procedures such as principal components analysis, canonical correlations, linear discriminant analysis, and multiple and partial correlations involve covariance matrices only up to a scale factor, in which case a covariance functional may be replaced by a weak version. In these types of application, $\boldsymbol{C}(F)$ is one of the target parameters. Here, however, we examine the role of weak covariance functionals in *standardization* with the aim of obtaining affine equivariance of certain procedures. In such a case, $\boldsymbol{C}(F)$ is rather a nuisance parameter.

4.2 Transformation-Retransformation (TR) Functionals

Beginning with Chaudhuri and Sengupta (1993), and continuing with Chakraborty and Chaudhuri (1996), Chakraborty, Chaudhuri, and Oja (1998), and Chakraborty (2001), a particular "data-driven" transformation of the data was introduced and applied as a means of modifying certain multivariate sign tests, multivariate sample medians, and multivariate sample quantiles into versions achieving full affine invariance or equivariance. This data-based nonsingular $d \times d$ matrix transformation $\boldsymbol{M}_0(\mathbb{X}_n)$ is discussed in detail in Section 4.3 below. The method consists of transforming the data to $\boldsymbol{M}_0(\mathbb{X}_n)\mathbb{X}_n$, then carrying out the procedure (for some sort of location inference) on the transformed data $\boldsymbol{M}_0(\mathbb{X}_n)\mathbb{X}_n$, and finally retransforming that result back to the original coordinate system via $\boldsymbol{M}(\mathbb{X}_n)^{-1}$. In such applications of the approach, the desired equivariance or invariance is established by separate arguments.

A special property of $\boldsymbol{M}_0(\mathbb{X}_n)$ established and discussed in Chaudhuri and Sengupta (1993) is that the transformed data $\boldsymbol{M}_0(\mathbb{X}_n)\mathbb{X}_n$ is invariant under affine transformations $\boldsymbol{A}\boldsymbol{X} \mapsto \boldsymbol{Y}$ and indeed is a maximal invariant of the data. However, this has not been used or needed in the lines of application of $\boldsymbol{M}_0(\mathbb{X}_n)$ to produce affinely invariant or equivariant sample procedures in location inference. Nevertheless, in Section 6 we return to this property of $\boldsymbol{M}_0(\mathbb{X}_n)$ and discuss its importance from the perspective of SICS transformations.

The above-described method using $\boldsymbol{M}_0(\mathbb{X}_n)$ to produce equivariance or invariance in location problems is called the "transformation-retransformation" (TR) approach. Randles (2000) introduces a general view of the TR approach by formulating a necessary (although not sufficient) affine equivariance requirement for any matrix transformation $\boldsymbol{M}(\mathbb{X}_n)$ *potentially* to produce an affine equivariant or invariant procedure via the TR scheme. We adopt this notion here, introducing the following definition as a population analogue of the sample version given by Randles (2000).

Definition 4.2 A positive definite $d \times d$ matrix-valued functional $\boldsymbol{M}(F)$ (not necessarily symmetric) is called a *transformation-retransformation (TR) functional* if, for $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$ with any nonsingular $\boldsymbol{A}$ and any $\boldsymbol{b}$,

$$\boldsymbol{A}'\boldsymbol{M}(F_{\boldsymbol{Y}})'\boldsymbol{M}(F_{\boldsymbol{Y}})\boldsymbol{A} = k_2\boldsymbol{M}(F_{\boldsymbol{X}})'\boldsymbol{M}(F_{\boldsymbol{X}}), \tag{15}$$

with $k_2 = k_2(\boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}})$ a positive scalar function of $\boldsymbol{A}$, $\boldsymbol{b}$, and $F_{\boldsymbol{X}}$. □

Every TR functional $\boldsymbol{M}(F)$ induces a corresponding WC functional, via

$$\boldsymbol{C}(F) = (\boldsymbol{M}(F)'\boldsymbol{M}(F))^{-1}. \tag{16}$$

Conversely, every WC functional has factorizations of form (16), and each such $\boldsymbol{M}(F)$ is necessarily a TR functional. In particular, one may choose $\boldsymbol{M}(F)$ to be the symmetric square root of $\boldsymbol{C}(F)^{-1}$ or the unique upper triangular matrix in the Cholesky factorization with "1" in the uppermost diagonal cell. Thus the choice of $\boldsymbol{M}(F)$ in (16) is not unique. Also, besides these structurally differing cases, for any solution $\boldsymbol{M}(F)$ we have that additional solutions are given by $c\,\boldsymbol{O}\,\boldsymbol{M}(F)$ for any constant c and orthogonal matrix $\boldsymbol{O}$. Other solutions, quite different from these structurally, will be seen in Section 6.

The following result makes precise the connection between WC and TR functionals.

Lemma 4.1 *We have*

(i) Given $\boldsymbol{M}(F)$ satisfying (15), the functional $\boldsymbol{C}(F)$ defined by (16) satisfies (13).

(ii) Given $\boldsymbol{C}(F)$ satisfying (13), any functional $\boldsymbol{M}(F)$ satisfying (16) satisfies (15).

Proof. Let $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$, for any nonsingular $\boldsymbol{A}$ and any $\boldsymbol{b}$.

(i) Given $\boldsymbol{M}(F)$ satisfying (15), the functional $\boldsymbol{C}(F) = (\boldsymbol{M}(F)'\boldsymbol{M}(F))^{-1}$ satisfies

$$\begin{aligned} C(F_{\boldsymbol{Y}}) &= (\boldsymbol{M}(F_{\boldsymbol{Y}})'\boldsymbol{M}(F_{\boldsymbol{Y}}))^{-1} \\ &= (k_2(\boldsymbol{A}')^{-1}\boldsymbol{M}(F_{\boldsymbol{X}})'\boldsymbol{M}(F_{\boldsymbol{X}})\boldsymbol{A}^{-1})^{-1} \\ &= k_2^{-1}\boldsymbol{A}C(F_{\boldsymbol{X}})\boldsymbol{A}', \end{aligned}$$

fulfilling (13) with $k_1 = k_2^{-1}$.

(ii) Given $\boldsymbol{C}(F)$ satisfying (13) and $\boldsymbol{M}(F)$ satisfying (16), we have

$$\begin{aligned} \boldsymbol{A}'\boldsymbol{M}(F_{\boldsymbol{Y}})'\boldsymbol{M}(F_{\boldsymbol{Y}})\boldsymbol{A} &= \boldsymbol{A}'C(F_{\boldsymbol{Y}})^{-1}\boldsymbol{A} \\ &= \boldsymbol{A}'(k_1\boldsymbol{A}C(F_{\boldsymbol{X}})\boldsymbol{A}')^{-1}\boldsymbol{A} \\ &= = k_1^{-1}\boldsymbol{M}(F_{\boldsymbol{X}})'\boldsymbol{M}(F_{\boldsymbol{X}}), \end{aligned}$$

fulfilling (15) with $k_2 = k_1^{-1}$. □

For obvious reasons, we call (16) a *TR factorization* of $\boldsymbol{C}(F)$. A given WC functional has many TR factorizations, not only due to variation in multiplicative constants, but also with $\boldsymbol{M}(F)$ quite different structurally.

4.3 Some Particular TR (or WC) Functionals

By Lemma 4.1, WC and TR functionals may be discussed interchangeably. For practical applications, we desire that sample versions be both *robust* and *computationally efficient*. The selection of a such functional involves a *trade-off* between these criteria. Below we consider several cases from the spectrum of choices.

The Chaudhuri and Sengupta (1993) "Data-Driven" TR functional. Let $\mathbb{X}_n = \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\}$ (column vectors) be i.i.d. observations in $\mathbb{R}^d$ with continuous distribution F. Let $\alpha = \{i_0, i_1, \ldots, i_d\}$ be distinct indices in $\{1, \ldots, n\}$ with $i_0 < i_1 < \cdots < i_d$. Denote by $\boldsymbol{X}(\alpha, i_0)$ the $d \times d$ matrix with columns $(\boldsymbol{X}_{i_k} - \boldsymbol{X}_{i_0})$, $k = 1, \ldots, d$. The data-based matrix-valued functional $\boldsymbol{M}_0(\mathbb{X}_n) = \boldsymbol{X}(\alpha, i_0)^{-1}$, which exists with probability 1, is readily seen to be a sample TR functional, as follows. For $\mathbb{Y}_n = \boldsymbol{A}\mathbb{X}_n + \boldsymbol{b}$, it is quickly checked that

$$\boldsymbol{M}_0(\mathbb{Y}_n) = \boldsymbol{M}_0(\mathbb{X}_n)\boldsymbol{A}^{-1}, \tag{17}$$

which immediately yields $\boldsymbol{A}'\boldsymbol{M}_0(\mathbb{Y}_n)'\boldsymbol{M}_0(\mathbb{Y}_n)\boldsymbol{A} = \boldsymbol{M}_0(\mathbb{X}_n)'\boldsymbol{M}_0(\mathbb{X}_n)$, the sample version of (15) with $k_2 = 1$. The relation (17) has independent interest and indicates a significant further property of $\boldsymbol{M}_0(\cdot)$ that we discuss in Section 6.

A population analogue of $\boldsymbol{M}_0(\mathbb{X}_n)$ has not been given in the literature and indeed is not straightforward (see Serfling, 2009, for discussion). However, for the corresponding sample WC functional $\boldsymbol{C}_0(\mathbb{X}_n) = (\boldsymbol{M}_0(\mathbb{X}_n)'\boldsymbol{M}_0(\mathbb{X}_n))^{-1} = \boldsymbol{X}(\alpha, i_0)\boldsymbol{X}(\alpha, i_0)'$, the expected value defines a population analogue,

$$\begin{aligned} \boldsymbol{C}_0(F) &= E\left\{[(\boldsymbol{X}_1 - \boldsymbol{X}_{d+1})\cdots(\boldsymbol{X}_d - \boldsymbol{X}_{d+1})]\,[(\boldsymbol{X}_1 - \boldsymbol{X}_{d+1})\cdots(\boldsymbol{X}_d - \boldsymbol{X}_{d+1})]'\right\} \\ &= E\left\{[(\boldsymbol{X}_i - \boldsymbol{X}_{d+1})(\boldsymbol{X}_j - \boldsymbol{X}_{d+1})']_{d\times d}\right\}, \end{aligned} \tag{18}$$

which is an analogue of the usual covariance matrix with $\boldsymbol{X}_{d+1}$ playing the role of the sample mean. Clearly, $\boldsymbol{C}_0(\mathbb{X}_n)$ estimates $\boldsymbol{C}_0(F)$.

The transformation $\boldsymbol{M}_0(\mathbb{X}_n)$ as formulated above depends on the choice of the set α and thus is not invariant over permutations of the observations. However, the set α may be selected to meet some optimality criteria. Detailed considerations are found in Chakraborty and Chaudhuri (1998, 1999), Chakraborty, Chaudhuri, and Oja (1998), and Chakraborty (2001), with orientation to attaining optimal efficiency in the case of elliptically symmetric distributions. A particular recommendation is to find the choice of α making the matrix $\boldsymbol{X}(\alpha, i_0)'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X}(\alpha, i_0)$ become as close as possible to a matrix of form $\lambda \boldsymbol{I}_d$, i.e., so as to make the coordinate system represented by $\widehat{\boldsymbol{\Sigma}}^{-1/2}\boldsymbol{X}(\alpha, i_0)$ as orthonormal as possible, where $\widehat{\boldsymbol{\Sigma}}$ is a consistent and equivariant (WC) estimator of the population scatter matrix. This optimization step is combinatorial and entails extensive computation beyond that required for computation of $\widehat{\boldsymbol{\Sigma}}$, for which efficient algorithms are available as noted below. For treatment of robustness of the multivariate medians based on $\boldsymbol{M}_0(\mathbb{X}_n)$ and using a high breakdown estimator $\widehat{\boldsymbol{\Sigma}}$, see Chakraborty and Chaudhuri (1999).

The Tyler (1987) scatter estimator. A particular matrix-valued estimator of scatter was introduced by Tyler (1987) and shown to be "most robust" by virtue of minimizing over a class of M-estimators of scatter the maximum asymptotic variance over a class of elliptically symmetric distributions. Adrover (1998) establishes that the maximum bias is minimized as well. With respect to a specified location functional $\theta(F)$, the Tyler scatter functional is defined as $\boldsymbol{V}(F) = (\boldsymbol{M}_s(F)'\boldsymbol{M}_s(F))^{-1}$, with $\boldsymbol{M}_s(F)$ the unique symmetric square root of $\boldsymbol{V}^{-1}$ obtained through the M-estimation equation

$$E\left\{\left(\frac{\boldsymbol{M}_s(F_{\boldsymbol{X}})(\boldsymbol{X}-\boldsymbol{\theta}(F_{\boldsymbol{X}}))}{\|\boldsymbol{M}_s(F_{\boldsymbol{X}})(\boldsymbol{X}-\boldsymbol{\theta}(F_{\boldsymbol{X}}))\|}\right)\left(\frac{\boldsymbol{M}_s(F_{\boldsymbol{X}})(\boldsymbol{X}-\boldsymbol{\theta}(F_{\boldsymbol{X}}))}{\|\boldsymbol{M}_s(F_{\boldsymbol{X}})(\boldsymbol{X}-\boldsymbol{\theta}(F_{\boldsymbol{X}}))\|}\right)'\right\} = d^{-1}\boldsymbol{I}_d. \tag{19}$$

Tyler establishes that the matrix $\boldsymbol{V}^{-1}$ so obtained from any solution of (19) is unique up to a constant of proportionality. For (19) expressed in terms of the variable $\boldsymbol{Y}=\boldsymbol{A}\boldsymbol{X}+\boldsymbol{b}$, both $\boldsymbol{M}_s(F_{\boldsymbol{Y}})$ and $\boldsymbol{M}_s(F_{\boldsymbol{X}})\boldsymbol{A}^{-1}$ are solutions, in which case we have $\boldsymbol{M}_s(F_{\boldsymbol{Y}})'\boldsymbol{M}_s(F_{\boldsymbol{Y}}) = c(\boldsymbol{M}_s(F_{\boldsymbol{X}})\boldsymbol{A}^{-1})'\boldsymbol{M}_s(F_{\boldsymbol{X}})\boldsymbol{A}^{-1}$ for some constant c, by which it is seen that $\boldsymbol{M}_s(F)$ is a TR functional, satisfying (15) with $k_2(\boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}}) \equiv d^2$. It follows by Lemma 4.1(i) that $\boldsymbol{V}(F)$ is a WC functional. We note in passing, for later reference, that we do *not* have for $\boldsymbol{M}_s(F)$ the analogue of (17), because $\boldsymbol{M}_s(F_{\boldsymbol{X}})\boldsymbol{A}^{-1}$ is not in general symmetric. (In fact, counterexamples are easily found.)

In terms of the sign function, (15) may be expressed

$$E\left\{[\boldsymbol{S}(\boldsymbol{M}_s(F_{\boldsymbol{X}})(\boldsymbol{X}-\boldsymbol{\theta}(F_{\boldsymbol{X}})))][\boldsymbol{S}(\boldsymbol{M}_s(F_{\boldsymbol{X}})(\boldsymbol{X}-\boldsymbol{\theta}(F_{\boldsymbol{X}})))]'\right\} = d^{-1}\boldsymbol{I}_d, \tag{20}$$

which states that the *sign covariance matrix* of the $\boldsymbol{M}_s(F)$-transformed (centered) observations is $d^{-1}\boldsymbol{I}_d$, reflecting uncorrelated components.

In the treatment of Tyler (1987), the parameter $\boldsymbol{\theta}(F)$ is assumed known and thus a constant in the sample version of (19). In this case, if $n > d(d-1)$, the sample solution $\boldsymbol{M}_s(\mathbb{X}_n)$ is *unique* up to multiplicative constants and the n transformed data points lie on axes corresponding to unit vectors approximately "equally spaced". An iterative algorithm using Cholesky factorizations to compute $\boldsymbol{M}_s(\mathbb{X}_n)$ quickly in any practical dimension is given in Tyler (1987).

All of the above discussion applies unchanged to the solution $\boldsymbol{M}_t(F)$ of (15) given by the upper triangular square root of $\boldsymbol{V}^{-1}$. The sample version $\boldsymbol{M}_t(\mathbb{X}_n)$ of this TR functional is computed by the same algorithm and is used by Randles (2000) in designing a new multivariate sign test of location. In such location testing problems, the parameter $\boldsymbol{\theta}(F)$ is specified by the null hypothesis.

For inference situations when $\boldsymbol{\theta}(F)$ is not known or specified, a symmetrized version of $\boldsymbol{M}_s(F)$ is given by Dümbgen (1998): $\boldsymbol{M}_{s1}(F) = \boldsymbol{M}_s(F \ominus F)$, where $F \ominus F$ denotes the distribution of the difference $\boldsymbol{X}_1 - \boldsymbol{X}_2$ of two independent observations on F.

Robustness of $\boldsymbol{M}_s(\mathbb{X}_n)$ and $\boldsymbol{M}_{s1}(\mathbb{X}_n)$ is of interest more generally than under an elliptical symmetry assumption. Breakdown points of these estimators are treated in Dümbgen and Tyler (2005), with the following results. For $\boldsymbol{M}_s(\mathbb{X}_n)$ the finite sample breakdown point (BP) is $(\lceil n/d \rceil - 1)/n \sim 1/d$, this limit being the maximum possible for multivariate M-estimates of scatter. For $\boldsymbol{M}_{s1}(\mathbb{X}_n)$, a BP of $1 - \sqrt{1 - 1/d}$, taking values between $1/(2d)$ and $1/d$, is obtained for a contamination model. Dümbgen and Tyler (2005) also discuss in detail the nature of the contamination that can cause breakdown of these estimators. Convenient R-packages (e.g., ICSNP) are available for computation of these estimators.

High Breakdown Point Scatter Estimators. Of course, estimators with much higher breakdown points are often desired. For example, the Minimum Covariance Determinant (MCD) estimator of Rousseeuw (1985), attains a BP of $\lfloor (n - d + 1)/2 \rfloor /n \sim 1/2$. However, to gain such a high BP while retaining full affine covariance equivariance, computational efficiency is sacrificed. On the other hand, an efficient algorithm *Fast-MCD* of Rousseeuw and Van Driessen (1999) *approximates* the MCD and is implemented in the R packages *MASS*, *rrcov*, and *robustbase*, for example, as well as in other software packages. Other well-known covariance functionals also attain the preceding high BP, again at the expense of computational complexity. See Maronna, Martin, and Yohai, (2006) for general discussion.

5 "Mahalanobis Spatial" Quantile Functions

5.1 Formulation and Key Property

As seen from the discussion of the *spatial quantile function* in Example 2.2, the so-called *spatial median* $\boldsymbol{Q}_{\mathrm{S}}(\boldsymbol{0}, F)$ is the minimizer $\boldsymbol{\theta}$ of *expected Euclidean distance* $E\|\boldsymbol{X} - \boldsymbol{\theta}\|$. It has a long history and literature (reviewed nicely in Small, 1990). Isogai (1985) and Rao (1988), however, suggest minimizing instead the *expected Mahalanobis distance*,

$$E\|\boldsymbol{\Sigma}(F)^{-1/2}(\boldsymbol{X} - \boldsymbol{\theta})\|, \tag{21}$$

with $\boldsymbol{\Sigma}(F)$ the usual covariance matrix of F and $\boldsymbol{\Sigma}(F)^{-1/2}$ a square root in the standard sense, producing a fully affine equivariant multivariate median, in comparison with the (only) orthogonally equivariant spatial median. Even so, it is the latter that has received preference (e.g., Chakraborty, Chaudhuri, and Oja, 1998), primarily on the grounds that the coordinate system resulting from standardization as in (21) lacks a "simple and natural geometric interpretation". On the other hand, standardization is a fundamental step in statistical analysis, motivated by the principle that deviations should be interpreted in a relative sense. Thus (21) indeed has the same supporting arguments that have made the Mahalanobis distance itself a basic tool.

In this spirit, augmenting the above-discussed "Mahalanobis spatial median", we may define an analogue of the entire spatial quantile function through *standardization*, and for this purpose let us consider *any TR functional*. For $\boldsymbol{X}$ having cdf F on $\mathbb{R}^d$, and for a given TR functional $\boldsymbol{M}(\cdot)$, the corresponding *Mahalanobis spatial quantile function* $\boldsymbol{Q}_{\mathrm{MS}}(\boldsymbol{u}, F_{\boldsymbol{X}})$, $\boldsymbol{u} \in \mathbb{B}^{d-1}(\boldsymbol{0})$, is defined at $\boldsymbol{u}$ as the vector $\boldsymbol{\theta}$ minimizing

$$E\left\{\Phi\left(\boldsymbol{u}, \boldsymbol{M}(F_{\boldsymbol{X}})(\boldsymbol{X}-\boldsymbol{\theta})\right) - \Phi\left(\boldsymbol{u}, \boldsymbol{M}(F_{\boldsymbol{X}})\boldsymbol{X}\right)\right\}, \tag{22}$$

analogous to the definition of spatial quantile but using a *standardized* deviation. For $\boldsymbol{u} = \boldsymbol{0}$, we have the *Mahalanobis spatial median*, which minimizes

$$E\left\{\|\boldsymbol{M}(F_{\boldsymbol{X}})(\boldsymbol{X}-\boldsymbol{\theta})\| - \|\boldsymbol{M}(F_{\boldsymbol{X}})\boldsymbol{X}\|\right\},$$

as discussed above with $\boldsymbol{M}(\cdot) = \boldsymbol{\Sigma}(\cdot)^{-1/2}$, but now using a standard equivalent formulation not requiring $\boldsymbol{X}$ to have finite mean. Unlike the *Mahalanobis distance quantile function* $\boldsymbol{Q}_{\mathrm{MD}}(\boldsymbol{u}, F)$ associated with the Mahalanobis distance outlyingness function of Example 2.3, the quantile function $\boldsymbol{Q}_{\mathrm{MS}}(\boldsymbol{u}, F)$ does not necessarily have elliptical contours. This is a well-known attractive feature of the spatial quantile function $\boldsymbol{Q}_{\mathrm{S}}(\boldsymbol{u}, F)$ and it immediately carries over to $\boldsymbol{Q}_{\mathrm{MS}}(\boldsymbol{u}, F)$ via the "TR representation" (27) treated in Section 5.2. Therefore, $\boldsymbol{Q}_{\mathrm{MS}}(\boldsymbol{u}, F)$ like $\boldsymbol{Q}_{\mathrm{S}}(\boldsymbol{u}, F)$ follows the shape of F, ellipsoidal or not.

We establish in Theorem 5.1 below that, for any choice of TR functional $\boldsymbol{M}(\cdot)$, the corresponding quantile function $\boldsymbol{Q}_{\mathrm{MS}}(\boldsymbol{u}, F)$ is *affine equivariant*. The proof will utilize the following lemma, which gives an important property of a key matrix that arises in using TR functionals.

Lemma 5.1 *For any TR functional $\boldsymbol{M}(F)$ satisfying (15), and for $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$ with any nonsingular $d \times d$ $\boldsymbol{A}$ and any $\boldsymbol{b}$, the matrix*

$$\widetilde{\boldsymbol{A}}(\boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}}) = k_2(\boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}})^{1/2}(\boldsymbol{M}(F_{\boldsymbol{Y}})')^{-1}(\boldsymbol{A}')^{-1}\boldsymbol{M}(F_{\boldsymbol{X}})' \tag{23}$$

is orthogonal.

Proof. With $\widetilde{\boldsymbol{A}} = \widetilde{\boldsymbol{A}}(\boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}})$, we have,

$$\begin{aligned}
\widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{A}}' &= k_2(\boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}})(\boldsymbol{M}(F_{\boldsymbol{Y}})')^{-1}(\boldsymbol{A}')^{-1}\boldsymbol{M}(F_{\boldsymbol{X}})'\boldsymbol{M}(F_{\boldsymbol{X}})\boldsymbol{A}^{-1}\boldsymbol{M}(F_{\boldsymbol{Y}})^{-1} \\
&= k_2(\boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}})(\boldsymbol{M}(F_{\boldsymbol{Y}})')^{-1}(\boldsymbol{A}')^{-1}\boldsymbol{C}(F_{\boldsymbol{X}})^{-1}\boldsymbol{A}^{-1}\boldsymbol{M}(F_{\boldsymbol{Y}})^{-1} \\
&= (\boldsymbol{M}(F_{\boldsymbol{Y}})')^{-1}\boldsymbol{C}(F_{\boldsymbol{Y}})^{-1}\boldsymbol{M}(F_{\boldsymbol{Y}})^{-1} \\
&= \boldsymbol{I}_d.
\end{aligned}$$

□

For the Tyler (1987) TR functional discussed in Section 4.3, the sample version of Lemma 5.1 is noted and utilized by Randles (2000).

Theorem 5.1 *Let $\boldsymbol{M}(\cdot)$ be a TR functional and $\boldsymbol{Q}_{MS}(\boldsymbol{u}, F)$ the corresponding Mahalanobis spatial quantile function. For $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$ with any nonsingular $d \times d$ $\boldsymbol{A}$ and any $\boldsymbol{b}$,*

$$\boldsymbol{Q}_{\mathrm{MS}}(\widetilde{\boldsymbol{A}}\boldsymbol{u}, F_{\boldsymbol{Y}}) = \boldsymbol{A}\,\boldsymbol{Q}_{\mathrm{MS}}(\boldsymbol{u}, F_{\boldsymbol{X}}) + \boldsymbol{b}, \tag{24}$$

with $\widetilde{\boldsymbol{A}} = \widetilde{\boldsymbol{A}}(\boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}})$ as given in (23) of Lemma 5.1 (and hence orthogonal).

That is, $\boldsymbol{Q}_{\mathrm{MS}}(\boldsymbol{u}, F)$ satisfies Definition 3.1 with $\boldsymbol{v}(\boldsymbol{u}, \boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}}) = \widetilde{\boldsymbol{A}}(\boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}})\,\boldsymbol{u}$. As seen in Remark 3.1, this equivariance yields corresponding equivariance and invariance properties for the related functions $\boldsymbol{R}_{\mathrm{MS}}(\boldsymbol{x}, F)$, $O_{\mathrm{MS}}(\boldsymbol{x}, F)$, and $D_{\mathrm{MS}}(\boldsymbol{x}, F)$. In particular, (9) yields

$$\boldsymbol{R}_{\mathrm{MS}}(\boldsymbol{y}, F_{\boldsymbol{Y}}) = \widetilde{\boldsymbol{A}}(\boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}})\,\boldsymbol{R}_{\mathrm{MS}}(\boldsymbol{x}, F_{\boldsymbol{X}}). \tag{25}$$

Restricting $\boldsymbol{A}$ to be proportional to an *orthogonal* matrix, it is seen that $\boldsymbol{I}_d$ is a TR functional. Also, with $\boldsymbol{M}(F) \equiv \boldsymbol{I}_d$ the Mahalanobis spatial quantile function is just the spatial quantile function $\boldsymbol{Q}_{\mathrm{S}}(\boldsymbol{u}, F)$. Then (24) yields the *orthogonal equivariance* of $\boldsymbol{Q}_{\mathrm{S}}(\boldsymbol{u}, F)$, as stated earlier in (12).

Proof of Theorem 5.1. Let $\boldsymbol{\theta}$ minimize (22) and put $\boldsymbol{\eta} = \boldsymbol{A}\boldsymbol{\theta} + \boldsymbol{b}$ and $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$, with $\boldsymbol{A}$ nonsingular. It suffices for (24) to show that $\boldsymbol{\eta}$ minimizes

$$E\left\{\Phi\left(\widetilde{\boldsymbol{A}}\boldsymbol{u}, \boldsymbol{M}(F_{\boldsymbol{Y}})(\boldsymbol{Y} - \boldsymbol{\eta})\right) - \Phi\left(\widetilde{\boldsymbol{A}}\boldsymbol{u}, \boldsymbol{M}(F_{\boldsymbol{Y}})\boldsymbol{Y}\right)\right\}. \tag{26}$$

Using $\boldsymbol{Y} - \boldsymbol{\eta} = \boldsymbol{A}(\boldsymbol{X} - \boldsymbol{\theta})$, the TR equivariance condition (15), and nonsingularity of $\boldsymbol{A}$, it is quickly checked that

$$\|\boldsymbol{M}(F_{\boldsymbol{Y}})(\boldsymbol{Y} - \boldsymbol{\eta})\| = k_2(\boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}})^{1/2}\,\|\boldsymbol{M}(F_{\boldsymbol{X}})(\boldsymbol{X} - \boldsymbol{\theta})\|.$$

Also, again applying (15), we obtain

$$(\widetilde{\boldsymbol{A}}\boldsymbol{u})'\boldsymbol{M}(F_{\boldsymbol{Y}})(\boldsymbol{Y} - \boldsymbol{\eta}) = k_2(\boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}})^{1/2}\,\boldsymbol{u}'\boldsymbol{M}(F_{\boldsymbol{X}})(\boldsymbol{X} - \boldsymbol{\theta}).$$

Hence

$$\Phi\left(\widetilde{\boldsymbol{A}}\boldsymbol{u}, \boldsymbol{M}(F_{\boldsymbol{Y}})(\boldsymbol{Y} - \boldsymbol{\eta})\right) = k_2(\boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}})^{1/2}\,\Phi\left(\boldsymbol{u}, \boldsymbol{M}(F_{\boldsymbol{X}})(\boldsymbol{X} - \boldsymbol{\theta})\right).$$

Consequently, with $\boldsymbol{\eta} = \boldsymbol{A}\boldsymbol{\theta} + \boldsymbol{b}$ the first term in (26) is equal to $k_2(\boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}})^{-1/2}$ times the first term in (22), and so both quantities are simultaneously minimized by $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$, respectively. (The other terms in (26) and (22) do not depend on by $\boldsymbol{\theta}$ or $\boldsymbol{\eta}$.) This completes the proof. □

For any choice of TR functional $\boldsymbol{M}(\cdot)$, the corresponding Mahalanobis spatial quantile function $\boldsymbol{Q}_{\mathrm{MS}}(\cdot, F)$ enjoys an *explicit representation* in terms of $\boldsymbol{Q}_{\mathrm{S}}(\cdot, F)$. Also, the *sample version* of $\boldsymbol{Q}_{\mathrm{MS}}(\cdot, F)$ corresponding to a particular TR functional $\boldsymbol{M}_0(\cdot)$ has already been introduced in the literature and applied to obtain an affine equivariant version of the sample spatial quantile function. These topics we discuss in the next section.

5.2 TR Representation for $\boldsymbol{Q}_{\mathbf{MS}}(\cdot, F)$ in Terms of $\boldsymbol{Q}_{\mathbf{S}}(\cdot, F)$

For a given TR functional $\boldsymbol{M}(\cdot)$, let us consider the corresponding $\boldsymbol{Q}_{\mathrm{MS}}(\cdot, F)$. Note that (22) is simply (4) with $\boldsymbol{X}$ and $\boldsymbol{\theta}$ replaced by $\boldsymbol{M}(F_{\boldsymbol{X}})\boldsymbol{X}$ and $\boldsymbol{M}(F_{\boldsymbol{X}})\boldsymbol{\theta}$, respectively. Therefore, if $\boldsymbol{\theta}$ minimizes (22), then $\boldsymbol{M}(F_{\boldsymbol{X}})\boldsymbol{\theta}$ minimizes (4) with $\boldsymbol{X}$ replaced by $\boldsymbol{M}(F_{\boldsymbol{X}})\boldsymbol{X}$. That is, $\boldsymbol{M}(F_{\boldsymbol{X}})\boldsymbol{Q}_{\mathrm{MS}}(\boldsymbol{u}, F_{\boldsymbol{X}}) = \boldsymbol{Q}_{\mathrm{S}}(\boldsymbol{u}, F_{\boldsymbol{M}(F_{\boldsymbol{X}})\boldsymbol{X}})$. This relationship may be characterized as a *TR representation* for the Mahalanobis spatial quantile function in terms of the spatial quantile function:

$$\boldsymbol{Q}_{\mathrm{MS}}(\boldsymbol{u}, F_{\boldsymbol{X}}) = \boldsymbol{M}(F_{\boldsymbol{X}})^{-1}\, \boldsymbol{Q}_{\mathrm{S}}(\boldsymbol{u}, F_{\boldsymbol{M}(F_{\boldsymbol{X}})\boldsymbol{X}}). \tag{27}$$

For helpful perspective, note that (27) asserts

$$(\textit{Mahalanobis } \boldsymbol{u}\text{th quantile of } \boldsymbol{X}) \;=\; \boldsymbol{M}(F_{\boldsymbol{X}})^{-1} \times (\textit{spatial } \boldsymbol{u}\text{th quantile of } \boldsymbol{M}(F_{\boldsymbol{X}})\boldsymbol{X}),$$

extending the univariate relationship "(uth quantile of X) $= \sigma \times$ (uth quantile of $\sigma^{-1}X$)".

The TR representation (27) for the Mahalanobis spatial quantile function readily yields similar representatons for the Mahalanobis spatial centered rank, depth, and outlyingness functions, as follows. The definition of $\boldsymbol{R}_{\mathrm{MS}}(\boldsymbol{x}, F_{\boldsymbol{X}})$ via $\boldsymbol{x} = \boldsymbol{Q}_{\mathrm{MS}}(\boldsymbol{R}_M(\boldsymbol{x}, F_{\boldsymbol{X}}), F_{\boldsymbol{X}})$, along with (5), yields

$$\boldsymbol{R}_{\mathrm{MS}}(\boldsymbol{x}, F_{\boldsymbol{X}}) = \boldsymbol{R}_{\mathrm{S}}\left(\boldsymbol{M}(F_{\boldsymbol{X}})\boldsymbol{x}, F_{\boldsymbol{M}(F_{\boldsymbol{X}})\boldsymbol{X}}\right) = E\left\{\boldsymbol{S}\left(\boldsymbol{M}(F_{\boldsymbol{X}})(\boldsymbol{x} - \boldsymbol{X})\right)\right\}, \tag{28}$$

which in turn immediately yields

$$O_{\mathrm{MS}}(\boldsymbol{x}, F_{\boldsymbol{X}}) = O_{\mathrm{S}}\left(\boldsymbol{M}(F_{\boldsymbol{X}})\boldsymbol{x}, F_{\boldsymbol{M}(F_{\boldsymbol{X}})\boldsymbol{X}}\right) = \left\|E\left\{\boldsymbol{S}\left(\boldsymbol{M}(F_{\boldsymbol{X}})(\boldsymbol{x} - \boldsymbol{X})\right)\right\}\right\|. \tag{29}$$

For depth, we use $D_{\mathrm{MS}}(\boldsymbol{x}, F_{\boldsymbol{X}}) = 1 - O_{\mathrm{MS}}(\boldsymbol{x}, F_{\boldsymbol{X}})$.

5.3 TR Sample Spatial Quantile Function

Chakraborty (2001), as proposed in Chaudhuri (1996), develops a *TR sample spatial quantile function* using the particular sample TR functional $\boldsymbol{M}_0(\mathbb{X}_n)$ discussed above in Section 4.3 and already applied in Chakraborty and Chaudhuri (1996) for the sample coordinatewise median and in Chakraborty, Chaudhuri, and Oja (1998) for the sample spatial median. This is given by

$$\boldsymbol{Q}_{\mathrm{S}}^{(\mathrm{TR})}(\boldsymbol{u}, \mathbb{X}_n) = \boldsymbol{M}_0(\mathbb{X}_n)^{-1}\, \boldsymbol{Q}_{\mathrm{S}}(\boldsymbol{\nu}(\boldsymbol{u}, \mathbb{X}_n), \boldsymbol{M}_0(\mathbb{X}_n)\mathbb{X}_n), \tag{30}$$

where $\boldsymbol{\nu}(\boldsymbol{u}, \mathbb{X}_n) = (\|\boldsymbol{u}\|/\|\boldsymbol{M}_0(\mathbb{X}_n)\boldsymbol{u}\|)\boldsymbol{M}_0(\mathbb{X}_n)\boldsymbol{u}$. The re-indexing has population analogue $\boldsymbol{\nu}(\boldsymbol{u}, F_{\boldsymbol{X}}) = \|\boldsymbol{u}\|/\|\boldsymbol{M}_0(F_{\boldsymbol{X}})\boldsymbol{u}\|\boldsymbol{M}_0(F_{\boldsymbol{X}})\boldsymbol{u}$, and we note that $\|\boldsymbol{\nu}(\boldsymbol{u}, \mathbb{X}_n)\| = \|\boldsymbol{\nu}(\boldsymbol{u}, F_{\boldsymbol{X}})\| = \|\boldsymbol{u}\|$, indicating an *outlyingness invariant* re-indexing. However, in view of the TR representation (27) and the result given by Theorem 5.1, it is clear that that no re-indexing is required for construction of a proper "TR sample spatial quantile function". The work is done completely by the TR steps. That is, we could just as well define an affine equivariant TR sample spatial quantile function simply by

$$\widetilde{\boldsymbol{Q}}_{\mathrm{S}}^{(\mathrm{TR})}(\boldsymbol{u}, \mathbb{X}_n) = \boldsymbol{M}_0(\mathbb{X}_n)^{-1}\, \boldsymbol{Q}_{\mathrm{S}}(\boldsymbol{u}, \boldsymbol{M}_0(\mathbb{X}_n)\mathbb{X}_n). \tag{31}$$

This also clarifies that what are estimated by $\boldsymbol{Q}_{\mathrm{S}}^{(\mathrm{TR})}(\boldsymbol{u}, \mathbb{X}_n)$ and $\widetilde{\boldsymbol{Q}}_{\mathrm{S}}^{(\mathrm{TR})}(\boldsymbol{u}, \mathbb{X}_n)$, respectively, are not $\boldsymbol{Q}_{\mathrm{S}}(\boldsymbol{\nu}(\boldsymbol{u}, F_{\boldsymbol{X}}), F_{\boldsymbol{X}})$ and $\boldsymbol{Q}_{\mathrm{S}}(\boldsymbol{u}, F_{\boldsymbol{X}})$, but rather $\boldsymbol{Q}_{\mathrm{MS}}(\boldsymbol{\nu}(\boldsymbol{u}, F_{\boldsymbol{X}}), F_{\boldsymbol{X}})$ and $\boldsymbol{Q}_{\mathrm{MS}}(\boldsymbol{u}, F_{\boldsymbol{X}})$. Thus, in particular, the medians $\boldsymbol{Q}_{\mathrm{S}}(\boldsymbol{0}, F_{\boldsymbol{X}})$ and $\boldsymbol{Q}_{\mathrm{MS}}(\boldsymbol{0}, F_{\boldsymbol{X}})$ are *different* location parameters of $F_{\boldsymbol{X}}$ (and only the latter is fully affine equivariant).

On this basis, we call the version given by (31) the *sample Mahalanobis spatial quantile function* and denote it by $\boldsymbol{Q}_{\mathrm{MS}}(\boldsymbol{u}, \mathbb{X}_n)$. That is, for *any TR functional* $\boldsymbol{M}(\cdot)$, we may consider the sample version of the TR representation (27) for $\boldsymbol{Q}_{\mathrm{MS}}(\boldsymbol{u}, F_{\boldsymbol{X}})$ to be given by

$$\boldsymbol{Q}_{\mathrm{MS}}(\boldsymbol{u}, \mathbb{X}_n) = \boldsymbol{M}(\mathbb{X}_n)^{-1}\, \boldsymbol{Q}_{\mathrm{S}}(\boldsymbol{u}, \boldsymbol{M}(\mathbb{X}_n)\mathbb{X}_n). \tag{32}$$

Once $\boldsymbol{M}(\mathbb{X}_n)$ is computed for a data set (a computation which may or not be intensive, depending on the choice), the computation of the $\boldsymbol{u}$th sample Mahalanobis spatial quantile $\boldsymbol{Q}_{\mathrm{MS}}(\boldsymbol{u}, \mathbb{X}_n)$ may be carried out in straightforward fashion by solving

$$-\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{S}\left(\boldsymbol{M}(\mathbb{X}_n)^{-1/2}(\boldsymbol{x} - \boldsymbol{X}_i)\right) = \boldsymbol{u}. \tag{33}$$

See Chaudhuri (1996) for discussion with respect to the usual spatial quantiles and Chakraborty (2001) for discussion with respect to the TR spatial quantiles defined by (30).

Since *any TR functional* suffices for formulation of an affine equivariant "TR" sample spatial quantile function, alternatives to the Chaudhuri and Sengupta matrix $\boldsymbol{M}_0(\mathbb{X}_n)$ for this purpose are the Tyler and Dümbgen matrices $\boldsymbol{M}_s(\mathbb{X}_n)$ and $\boldsymbol{M}_{s1}(\mathbb{X}_n)$, for example. It is also of interest to explore other possibilities.

6 Invariant Coordinate System (ICS) Functionals

In nonparametric settings, especially other than elliptically symmetric families, different types of multivariate scatter estimators may measure different aspects of the underlying distribution. In this spirit, Tyler, Critchley, Dümbgen, and Oja (2009) (TCDO) present a general approach based on the use of two different matrix-valued scatter measures together to discover interesting features. Specifically, the eigenvalue-eigenvector decomposition of one such measure relative to the other yields a transformation that takes the given data into a new *invariant coordinate system* (ICS) within which perceived structures in the data remain invariant under (further) affine transformation of the data. This broadly generalizes methods of "generalized principal components analysis" for creating diagnostic plots and yields useful results related to the area of independent components analysis.

ICS transformations are of general interest in their own right and can be constructed by other methods as well. Here we focus on an approach that produces a special class of ICS transformations that are especially productive in applications involving outlier detection, among others. We also examine interconnections among ICS, WC, and TR functionals.

6.1 Definitions and Lemmas

We start with the definition given by TCDO.

Definition 6.1 An *invariant coordinate system (ICS) functional* is a positive definite $d \times d$ matrix-valued functional $\boldsymbol{D}(F)$ such that the $\boldsymbol{D}(\cdot)$-transformed variable $\boldsymbol{D}(F_{\boldsymbol{X}})\boldsymbol{X}$ remains unaltered after affine transformation to $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$, for any nonsingular $\boldsymbol{A}$ and any $\boldsymbol{b}$, followed by $\boldsymbol{D}(\cdot)$-transformation to $\boldsymbol{D}(F_{\boldsymbol{Y}})\boldsymbol{Y}$, *except for coordinatewise scale changes*, *coordinatewise sign changes*, and *translations*. □

The appeal of such a transformation is that geometric structures or patterns identified in a $\boldsymbol{D}(\cdot)$-transformed data set $D(\mathbb{X}_n)\mathbb{X}_n$ remain unaltered after affine transformation to $\mathbb{Y}_n = \boldsymbol{A}\mathbb{X}_n + \boldsymbol{b}$ followed by $\boldsymbol{D}(\cdot)$-transformation to $D(\mathbb{Y}_n)\mathbb{Y}_n$, subject, however, to coordinatewise scale changes, coordinatewise sign changes, and translations.

A particular approach for construction of ICS functionals is given by TCDO. For any two WC functionals $\boldsymbol{V}_1(F)$ and $\boldsymbol{V}_2(F)$ with the eigenvalues of $\boldsymbol{V}_1(F)^{-1}\boldsymbol{V}_2(F)$ all distinct, the matrix $\boldsymbol{D}(F)$ of corresponding eigenvectors is an ICS functional in the above sense. Also, the method is extended to allow multiplicities among the eigenvalues. Various choices of pairs $(\boldsymbol{V}_1(\cdot), \boldsymbol{V}_2(\cdot))$ are studied.

In some applications, however, for example those requring *outlyingness invariance*, the transformed data $D(\mathbb{Y}_n)\mathbb{Y}_n$ should agree with $D(\mathbb{X}_n)\mathbb{X}_n$ except only for *homogeneous* scale changes and *homogeneous* sign changes. Further, for some applications, it is desirable that an ICS functional also play the role played by a TR functional, but an ICS functional is not in general TR. To address these issues, we introduce two successively stronger versions of ICS functional and examine their properties. The first version requires homogeneity of *scale changes* and possesses a structure defined by (34) in Lemma 6.1 below. The second possesses the same structure but imposes in addition homogeneity of *sign changes* and is formulated in Definition 6.2 below.

Lemma 6.1 *Let $\boldsymbol{D}(F)$ be a positive definite $d \times d$ matrix-valued functional satisfying, for $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$ with any nonsingular $\boldsymbol{A}$ and any $\boldsymbol{b}$,*

$$\boldsymbol{D}(F_{\boldsymbol{Y}}) = k_3 \boldsymbol{J}\boldsymbol{D}(F_{\boldsymbol{X}})\boldsymbol{A}^{-1}, \tag{34}$$

with $k_3 = k_3(\boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}})$ a positive scalar function and $\boldsymbol{J} = \boldsymbol{J}(\boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}})$ a sign change matrix (diagonal with ± 1 diagonal values). Then $\boldsymbol{D}(F)$ is both ICS and TR (and thus $(\boldsymbol{D}(F)'\boldsymbol{D}(F))^{-1}$ is WC).

Proof. For $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$ with any nonsingular $\boldsymbol{A}$ and any $\boldsymbol{b}$, we have

$$\boldsymbol{D}(F_{\boldsymbol{Y}})\boldsymbol{Y} = k_3\boldsymbol{J}\boldsymbol{D}(F_{\boldsymbol{X}})\boldsymbol{A}^{-1}(\boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}) = k_3\boldsymbol{J}\boldsymbol{D}(F_{\boldsymbol{X}})\boldsymbol{X} + k_3\boldsymbol{J}\boldsymbol{D}(F_{\boldsymbol{X}})\boldsymbol{A}^{-1}\boldsymbol{b},$$

so that $\boldsymbol{D}(F_{\boldsymbol{Y}})\boldsymbol{Y}$ agrees with $\boldsymbol{D}(F_{\boldsymbol{X}})\boldsymbol{X}$ up to a homogenous scale change, coordinatewise sign changes, and a translation. Thus $\boldsymbol{D}(F_{\boldsymbol{X}})$ is ICS. Also, we have

$$\begin{aligned}
\boldsymbol{A}'\boldsymbol{D}(F_{\boldsymbol{Y}})'\boldsymbol{D}(F_{\boldsymbol{Y}})\boldsymbol{A} &= \boldsymbol{A}'(k_3\boldsymbol{J}\boldsymbol{D}(F_{\boldsymbol{X}})\boldsymbol{A}^{-1})'(k_3\boldsymbol{J}\boldsymbol{D}(F_{\boldsymbol{X}})\boldsymbol{A}^{-1})\boldsymbol{A} \\
&= k_3^2\boldsymbol{A}'(\boldsymbol{A}')^{-1}\boldsymbol{D}(F_{\boldsymbol{X}})'\boldsymbol{J}'\boldsymbol{J}\boldsymbol{D}(F_{\boldsymbol{X}}) \\
&= k_3^2\boldsymbol{D}(F_{\boldsymbol{X}})'\boldsymbol{D}(F_{\boldsymbol{X}}),
\end{aligned}$$

fulfilling (15) with $k_2 = k_3^2$. □

Definition 6.2 A *strong ICS (SICS) functional* $\boldsymbol{D}(\cdot)$ is one satisfying (34) with $\boldsymbol{J} = \boldsymbol{I}_d$, thus satisfying

$$\boldsymbol{D}(F_{\boldsymbol{Y}}) = k_3 \boldsymbol{D}(F_{\boldsymbol{X}}) \boldsymbol{A}^{-1}, \tag{35}$$

with $k_3 = k_3(\boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}})$ a scalar constant. □

The key property making SICS functionals very useful in practice is given in the following result, whose proof is immediate following the lines used in proving Lemma 6.1.

Lemma 6.2 *A SICS functional* $\boldsymbol{D}(F)$ *satisfies*

$$\boldsymbol{D}(F_{\boldsymbol{Y}})\boldsymbol{Y} = k_3 \boldsymbol{D}(F_{\boldsymbol{X}})\boldsymbol{X} + \boldsymbol{c}, \tag{36}$$

with $k_3 = k_3(\boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}})$ *a scalar constant and* $\boldsymbol{c} = c_3(\boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}}) = k_3 \boldsymbol{D}(F_{\boldsymbol{X}})\boldsymbol{A}^{-1}\boldsymbol{b}$ *a vector constant.*

For a SICS functional, $\boldsymbol{D}(F_{\boldsymbol{Y}})\boldsymbol{Y}$ agrees with $\boldsymbol{D}(F_{\boldsymbol{X}})\boldsymbol{X}$ subject only to a *homogenous scale change and a translation*. A similar statement applies to the sample versions $D(\mathbb{Y}_n)\mathbb{Y}_n$ and $D(\mathbb{X}_n)\mathbb{X}_n$. Conversion of the "data" $\mathbb{X}_n$ to $D(\mathbb{X}_n)\mathbb{X}_n$ eliminates structure and patterns which are mere artifacts of the choice of coordinate system and should be disregarded as spurious no matter how striking.

Remark 6.1 *A SICS functional is neither symmetric nor triangular.* If a SICS functional $\boldsymbol{D}(F)$ is symmetric (resp., triangular), then $\boldsymbol{D}(F_{\boldsymbol{Y}})$ in (35) is symmetric (resp., triangular), and hence so is $\boldsymbol{D}(F_{\boldsymbol{X}})\boldsymbol{A}^{-1}$ for *arbitrary* nonsingular $\boldsymbol{A}$. Counterexamples are easily found. Consequently, the Tyler and Dümbgen TR functionals $\boldsymbol{M}_s(\cdot)$ and $\boldsymbol{M}_{s1}(\cdot)$ considered earlier cannot be SICS. However, the Chaudhuri and Sengupta functional $\boldsymbol{M}_0(\cdot)$ is indeed SICS, satisfying (35) with $k_3 = 1$, as seen earlier in (17). □

Although $\boldsymbol{M}_0(\cdot)$ is the only explicit example of SICS functional in the literature to date, this is mainly because heretofore SICS functionals have not been identified or targeted for theoretical investigation or practical application. Example 6.1 formulates a useful class of SICS functionals within which $\boldsymbol{M}_0(\cdot)$ is just a special case.

Example 6.1 *A family of SICS functionals.* Let $\mathbb{Z}_N = \{\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_N\}$ be a subset of $\mathbb{X}_n$ of size N obtained through some permutation-invariant procedure. Next, for $m = \lfloor N/(d+1) \rfloor$, form $d+1$ means $\overline{\boldsymbol{Z}}_1, \ldots, \overline{\boldsymbol{Z}}_{d+1}$ based, respectively, on consecutive blocks of size m from $\mathbb{Z}_N$. Finally, defining the matrix

$$\boldsymbol{W}(\mathbb{X}_n) = \left[(\overline{\boldsymbol{Z}}_1 - \overline{\boldsymbol{Z}}_{d+1}), \ldots, (\overline{\boldsymbol{Z}}_d - \overline{\boldsymbol{Z}}_{d+1})\right]_{d \times d},$$

it is readily seen that

$$\boldsymbol{D}(\mathbb{X}_n) = \boldsymbol{W}(\mathbb{X}_n)^{-1}$$

is a SICS functional.

The Chaudhuri and Sengupta (1993) TR functional $\boldsymbol{M}_0(\mathbb{X}_n)$ is the case that each mean is based on just one observation from a selected subset of size $N = d + 1$. As discussed in Section 4.3, this special subset is derived by extensive computation beyond that for the robust scatter matrix $\widehat{\boldsymbol{\Sigma}}$.

Alternatively, let $\mathbb{Z}_N$ be the set of observations selected and used in computing $\widehat{\boldsymbol{\Sigma}}$ (e.g., the Minimum Covariance Determinant scatter matrix based on $\mathbb{X}_n$) with, say, $N \approx 0.75n$. This uses all of the data in selecting $\mathbb{Z}_N$ and all of those observations in defining $\boldsymbol{W}(\mathbb{X}_n)$. As a result, there is little computation beyond that for $\widehat{\boldsymbol{\Sigma}}$. Such an approach is followed in Mazumder and Serfling (2009). $\square$

Remark 6.2 *Population versions of sample SICS functionals.* The population version of a SICS functional may be defined directly, as in Definition 6.2, and then the sample analogue version is immediate. However, the reverse direction is not straightforward. For example, starting with $\boldsymbol{W}(\mathbb{X}_n)$ in the above example, one might consider $E\{\boldsymbol{W}(\mathbb{X}_n)\}^{-1}$ as a natural population analogue of $\boldsymbol{D}(\mathbb{X}_n) = \boldsymbol{W}(\mathbb{X}_n)^{-1}$. But $E\{\boldsymbol{W}(\mathbb{X}_n)\}$ is a matrix of zeros. This very interesting issue is examined in detail in a general treatment of SICS functionals in Serfling (2009). $\square$

Remark 6.3 *Applications of SICS functionals.* Of course, any application of TR functionals can served by a SICS functional. However, some applications need the extra property provided by SICS transformations. In particular, some existing procedures which lack full affine equivariance or invariance can acquire it if carried out on SICS-transformed data. For example, *principal components analysis* can be made affine invariant by performing it on the data $\boldsymbol{D}(\mathbb{X}_n)\mathbb{X}_n$ instead of $\mathbb{X}_n$.

Another context is *fast dispersion matrix estimates.* For example, a method based on pairwise robust covariance estimation was proposed by Gnanadesikan and Kettenring (1972) and later modified by Maronna and Zamar (2002). See Maronna, Martin, and Yohai (2006) for general discussion of what is now called the "orthogonalized Gnanadesikan-Kettenring estimate" (OGK). The main drawback is its lack of affine equivariance. However, OGK is covariance equivariant under homogeneous scale change and translation. Consequently, if preceded by a SICS transformation of the data, it becomes fully affine covariance equivariant.

Full development of the above modifications of existing procedures is beyond the scope of the present paper. However, below we provide a general result on construction of affine invariant functionals using SICS-transformation and illustrate briefly in the context of *robust outlyingness functions.* $\square$

6.2 A Result on Constructing Affine Invariance

Here we provide a result showing how to construct suitable affine invariant functionals from more basic functionals, through the use of a SICS functional.

Theorem 6.1 *Let $T(\boldsymbol{x}, F)$ be a real-valued functional of d-vector $\boldsymbol{x}$ and distribution F on $\mathbb{R}^d$ that is invariant under homogeneous scale change and translation applied to $\boldsymbol{x}$, in the* *sense that*

$$T(a\boldsymbol{x}+\boldsymbol{b}, F_{a\boldsymbol{X}+\boldsymbol{b}}) = T(\boldsymbol{x}, F_{\boldsymbol{X}}) \tag{37}$$

for any scalar a and any vector $\boldsymbol{b}$*. Let* $\boldsymbol{D}(F)$ *be a SICS functional. Then the functional*

$$T(\boldsymbol{D}(F_{\boldsymbol{X}})\boldsymbol{x}, F_{\boldsymbol{D}(F_{\boldsymbol{X}})\boldsymbol{X}})$$

is affine invariant.

Proof. For $\boldsymbol{Y} = \boldsymbol{AX}+\boldsymbol{b}$ and $\boldsymbol{y} = \boldsymbol{Ax}+\boldsymbol{b}$ with any nonsingular $\boldsymbol{A}$ and any $\boldsymbol{b}$, and using (35), we have

$$\begin{aligned} T(\boldsymbol{D}(F_{\boldsymbol{Y}})\boldsymbol{y}, F_{\boldsymbol{D}(F_{\boldsymbol{Y}})\boldsymbol{Y}}) &= T(\boldsymbol{D}(F_{\boldsymbol{Y}})\boldsymbol{Ax}, F_{\boldsymbol{D}(F_{\boldsymbol{Y}})\boldsymbol{AX}}) \\ &= T(k_3\boldsymbol{D}(F_{\boldsymbol{X}})\boldsymbol{x}, F_{k_3\boldsymbol{D}(F_{\boldsymbol{X}})\boldsymbol{X}}) \\ &= T(\boldsymbol{D}(F_{\boldsymbol{X}})\boldsymbol{x}, F_{\boldsymbol{D}(F_{\boldsymbol{X}})\boldsymbol{X}}). \end{aligned}$$

□

Example 6.2 *Projection pursuit outlyingness functions using finitely many directions.* The projection pursuit approach toward formulation of multivariate outlyingness functions is well-established and uses the supremum of univariate scaled deviation outlyingness

$$O(x, F) = \left| \frac{x-\mu(F)}{\sigma(F)} \right|,$$

taken over all univariate projections of a data point $\boldsymbol{x}$ in $\mathbb{R}^d$. (These deviations were used in a different way in Example 2.3 to define Mahalanobis distance outlyingness.) More generally, for any set Δ of unit vectors $\boldsymbol{u}$ in $\mathbb{R}^d$, we may define outlyingness of a point $\boldsymbol{x}$ in $\mathbb{R}^d$ as some given function of the values of $O(\boldsymbol{u}'\boldsymbol{x}, F_{\boldsymbol{u}'\boldsymbol{X}})$ over the projections onto $\boldsymbol{u} \in \Delta$, i.e., as a function of the quantities

$$\{O(\boldsymbol{u}'\boldsymbol{x}, F_{\boldsymbol{u}'\boldsymbol{X}}), \boldsymbol{u} \in \Delta\}. \tag{38}$$

One possibility, for example, is to take

$$O_\Delta^{(\mathrm{sup})}(\boldsymbol{x}, F) = \sup_{\boldsymbol{u} \in \Delta} O(\boldsymbol{u}'\boldsymbol{x}, F_{\boldsymbol{u}'\boldsymbol{X}}).$$

For Δ the set of *all* projections, this is the usual *projection outlyingness* (e.g., see Zuo, 2003) and is affine invariant. However, for Δ *finite*, not even orthogonal invariance holds. Another possibility with finite Δ is to take $O_\Delta(\boldsymbol{x}, F)$ to be a quadratic form $O_\Delta^{(\mathrm{qf})}(\boldsymbol{x}, F)$ in the quantities in (38), but again even orthogonal invariance fails.

On the other hand, Peña and Prieto (2001) introduce an affine invariant method of using $O_\Delta^{(\mathrm{sup})}(\boldsymbol{x}, F)$ with a finite set Δ consisting of $2d$ *data-driven* directions. These are selected using univariate measures of kurtosis over candidate directions, choosing d with local extremes of high kurtosis and d with local extremes of low kurtosis. Ultimately, in their complex algorithm, the "outliers" are selected using Mahalanobis distances and thus the method has elliptical outlyingness contours. Filzmoser, Maronna, and Werner (2008) incorporate this approach into a more elaborate one, using also a principal components step, that achieves certain improvements in performance for detection of location outliers, especially in high dimension, but gives up equivariance (a SICS standardization might regain this, however). See also Maronna, Martin, and Yohai (2006) for general discussion.

The use of finite Δ with a fixed set of deterministic directions has strong appeal on computational grounds, and it is desirable that the directions be approximately uniformly scattered on the d-dimensional unit sphere. (Fang and Wang, 1994, provide convenient numerical algorithms for this purpose.) In this spirit, Pan, Fung, and Fang (2000) develop an approach using a quadratic form $O_\Delta^{(\mathrm{qf})}(\boldsymbol{x}, \mathbb{X}_n, F)$ based on the differences

$$\{O(\boldsymbol{u}'\boldsymbol{x}, \boldsymbol{u}'\mathbb{X}_n) - O(\boldsymbol{u}'\boldsymbol{x}, F_{\boldsymbol{u}'\boldsymbol{X}}), \boldsymbol{u} \in \Delta\}.$$

These involve the unknown F and thus a bootstrap step is incorporated. Also the method is not affine invariant (although a SICS transformation could correct for this).

A more straightforward approach is to use the quantities (38), but as modified by a SICS transformation, and employing a univariate outlyingness function $O(x, F)$ satisfying the assumptions of Theorem 6.1. With any SICS functional $\boldsymbol{D}(\cdot)$, this results in an affine invariant outlyingness function $O_\Delta(\boldsymbol{x}, F)$ defined in terms of the quantities

$$\{O(\boldsymbol{u}'\boldsymbol{D}(F_{\boldsymbol{X}})\boldsymbol{x}, F_{\boldsymbol{u}'\boldsymbol{D}(F_{\boldsymbol{X}})\boldsymbol{X}}),\ \boldsymbol{u} \in \Delta\}$$

(We note that the set Δ does *not* become transformed when transforming observations, since then the desired uniform scattering over the unit sphere would become modified). In particular, the above scaled deviation $O(x, F)$, i.e., taking

$$T(x, F) = \left|\frac{x - \mu(F)}{\sigma(F)}\right|,$$

immediately meets the required conditions. This approach is implemented in Mazumder and Serfling (2009), using sample outlyingness functions $O_\Delta^{(\mathrm{sup})}(\boldsymbol{x}, \mathbb{X}_n)$ and $O_\Delta^{(\mathrm{qf})}(\boldsymbol{x}, \mathbb{X}_n)$ based on the quantities

$$\{O(\boldsymbol{u}'\boldsymbol{D}(\mathbb{X}_n)\boldsymbol{x}, \boldsymbol{u}'\boldsymbol{D}(\mathbb{X}_n)\mathbb{X}_n),\ \boldsymbol{u} \in \Delta\}$$

(instead of those in (38)) and exploring several SICS transformations $\boldsymbol{D}(\cdot)$. The resulting outlier detection methods are robust, computationally attractive, and affine invariant, and their masking and swamping performance is extremely competitive in both low and high dimension.

Acknowledgments

The author greatly appreciates several stimulating discussions with Marc Hallin and Davy Paindaveine, useful input from Hannu Oja, and constructive comments by three anonymous reviewers. This has led to substantial improvements in the manuscript. The author also thanks Satyaki Mazumder for finding errors in earlier drafts. The support of National Science Foundation Grants DMS-0103698 and DMS-0805786 and National Security Agency Grant H98230-08-1-0106 is gratefully acknowledged.

References

[1] Adrover, J. G. (1998). Minimax bias-robust estimation of the dispersion matrix of a multivariate distribution. *Annals of Statistics* **26** 2301–2320.

[2] Breckling, J. and Chambers, R. (1988). M-quantiles. *Biometrika* **75** 761–771.

[3] Breckling, J., Kokic, P. and Lübke, O. (2001). A note on multivariate M-quantiles. *Statistics and Probability Letters* **55** 39–44.

[4] Chakraborty, B. (2001). On affine equivariant multivariate quantiles. *Annals of the Institute of Statistical Mathematics* **53** 380–403.

[5] Chakraborty, B., and Chaudhuri, P. (1996). On a transformation and re-transformation technique for constructing an affine equivariant multivariate median. *Proceedings of the American Mathematical Society* **124** 2539–2547.

[6] Chakraborty, B., and Chaudhuri, P. (1998). On an adaptive transformation–retransformation estimate of multivariate location. *Journal of the Royal Statistical Society, Series B* **60** 145–157.

[7] Chakraborty, B., and Chaudhuri, P. (1999). A note on the robustness of multivariate medians. *Statistics and Probability Letters* **45** 269–276.

[8] Chakraborty, B., Chaudhuri, P., and Oja, H. (1998). Operating transformation and retransformation on spatial median and angle test. *Statistica Sinica* **8** 767–784.

[9] Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *Journal of the American Statistical Association* **91** 862–872.

[10] 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.

[11] Donoho, D. L. and Gasko, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Annals of Statistics* **20** 1803–1827.

[12] Dudley, R. M. and Koltchinskii, V. I. (1992). The spatial quantiles. Unpublished manuscript.

[13] Dümbgen, L. (1998). On Tyler's M-functional of scatter in high dimension. *Annals of the Institute of Statistical Mathematics* **50** 471–491.

[14] Dümbgen, L. and Tyler, D. E. (2005). On the breakdown properties of some multivariate M-functionals. *Scandinavian Journal of Statistics* **32** 247–264.

[15] Fang, K. T, and Wang, Y. (1994). *Number Theoretic Methods in Statistics*. Chapman and Hall, London.

[16] Ferguson, T. S. (1967). *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press, New York.

[17] Filzmoser, P., Maronna, R., and Werner, M. (2008). Outlier identification in high dimensions. *Computational Statistics & Data Analysis* **52** 1694–1711.

[18] Gnanadesikan, R. and Kettenring, J. R. (1972). Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics* **28** 81–124.

[19] Hallin, M., Paindaveine, D., and Šiman, M. (2009). Multivariate quantiles and multiple-output regression quantiles: from L_1 optimization to halfspace depth (with discussion). *Annals of Statistics*, in press.

[20] Isogai, T. (1985). Some extension of Haldane's multivariate median and its application. *Annals of Institute of Statistical Mathematics* **37** 289–301.

[21] Koltchinskii, V. (1994a). Bahadur-Kiefer approximation for spatial quantiles. In: Hoffman-Jørgensen, J., Kuelbs, J., Marcus, M. (Eds.), *Probability in Banach Spaces 9*, Birkhaüser, Basel, 401–415.

[22] Koltchinskii, V. (1994b). Nonlinear transformations of empirical processes: functional inverses and Bahadur-Kiefer representations. In: *Proceedings of 6th International Vilnius Conference on Probability and Mathematical Statistics*, VSP-TEV, Netherlands, 423–445.

[23] Koltchinskii, V. (1997). *M*-estimation, convexity and quantiles. *Annals of Statistics* **25** 435–477.

[24] Kong, L., and Mizera, I. (2008). Quantile tomography: Using quantiles with multivariate data. Preprint.

[25] Liu, R. Y., Parelius, J. M. and Singh, K. (1999). Multivariate analysis by data depth: Descriptive statistics, graphics and inference (with discussion). *Annals of Statistics* **27** 783–858.

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

[27] Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Science of India* **12** 49–55.

[28] Maronna, R. A., Martin, R. D., and Yohai, V. J. (2006). *Robust Statistics: Theory and Methods*. Wiley, Chichester, England.

[29] Maronna, R. A. and Zamar, R. H. (2002). Robust estimation of location and dispersion for high-dimensional data sets. *Technometrics* **44** 307–317.

[30] Mazumder, S. and Serfling, R. (2009). Robust multivariate outlyingness functions based on Mahalanobis standardization and projected scaled deviations. In preparation.

[31] Mosteller, C. F. and Tukey, J. W. (1977). *Data Analysis and Regression*. Addison-Wesley, Reading, Mass.

[32] Möttönen, J. and Oja, H. (1995). Multivariate spatial sign and rank methods. *Journal of Nonparametric Statistics* **5** 201–213.

[33] Paindaveine, D. (2008). A canonical definition of shape. *Statistics and Probability Letters* **78** 2240–2247.

[34] Pan, J.-X., Fung, W.-K., and Fang, K.-T. (2000). Multiple outlier detection in multivariate data using projection pursuit techniques. *Journal of Statistical Planning and Inference* **83** 153–167.

[35] Peña, D. and Prieto, F. J. (2001). Robust covariance matrix estimation and multivariate outlier rejection. *Technometrics* **43** 286–310.

[36] Randles, R. H. (2000). A simpler, affine-invariant, multivariate, distribution-free sign test. *Journal of the American Statistical Assocation* **95** 1263–1268.

[37] Rao, C. R. (1988). Methodology based on the L_1 norm in statistical inference. *Sankhyā Series A* **50** 289–313.

[38] Rousseeuw P. (1985). Multivariate estimation with high breakdown point. In: W. Grossmann, G. Pflug, and W. Wertz (Eds.), *Mathematical Statistics and Applications, Vol. B*, Reidel Publishing, Dordrecht, 283–297.

[39] Rousseeuw, P. and Van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics* **41** 212–223.

[40] Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression and Outlier Detection.* John Wiley & Sons, New York.

[41] Serfling, R. (2004). Nonparametric multivariate descriptive measures based on spatial quantiles. *Journal of Statistical Planning and Inference* **123** 259–278.

[42] Serfling, R. (2006). Depth functions in nonparametric multivariate analysis. In *Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications* (R. Y. Liu, R. Serfling, D. L. Souvaine, eds.), pp. 1–16. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Volume 72, American Mathematical Society, Providence, Rhode Island.

[43] Serfling, R. (2009). On strong invariant coordinate system (SICS) functionals. In preparation.

[44] Small, C. G. (1990). A survey of multidimensional medians. *International Statistical Institute Review* **58** 263–277.

[45] Tyler, D. E. (1987). A distribution-free M-estimator of multivariate scatter. *Annals of Statistics* **15** 234–251.

[46] Tyler, D. E., Critchley, F., Dümbgen, L. and Oja, H. (2009). Invariant co-ordinate selection. *Journal of the Royal Statistical Society*, *Series B* **71** 1–27.

[47] Vardi, Y. and Zhang, C.-H. (2000). The multivariate L_1-median and associated data depth. *Proceedings National Academy of Science USA* **97** 1423–1426.

[48] Zhou, W. and Serfling, R. (2008). Multivariate spatial U-quantiles: a Bahadur-Kiefer representation, a Theil-Sen estimator for multiple regression, and a robust dispersion estimator. *Journal of Statistical Planning and Inference* **138** 1660–1678.

[49] Zuo, Y. (2003). Projection-based depth functions and associated medians. *Annals of Statistics* **31** 1460–1490.

[50] Zuo, Y. and Serfling, R. (2000). General notions of statistical depth function. *Annals of Statistics* **28** 461–482.