On Strong Invariant Coordinate System (SICS) Functionals

Robert Serfling[1]

University of Texas at Dallas

first draft July, 2009
updated June, 2010
updated March, 2011
currently in further development

[1]Department of Mathematical Sciences, University of Texas at Dallas, Richardson, Texas 75080-3021, USA. Email: `serfling@utdallas.edu`. Website: `www.utdallas.edu/~serfling`.

Abstract

In modern multivariate statistical analysis, affine invariance or equivariance for statistical procedures are properties of paramount interest and importance. A statistical procedure lacking such a property can sometimes acquire it if carried out not on the original data but rather on suitably transformed data, in some cases accompanied by a retransformation back to the original coordinate system. Three types of relevant transformation, weak covariance (WC), transformation-retransformation (TR), and strong invariant coordinate system (SICS), are treated in Serfling (2010), where the the WC and TR types are seen to be essentially equivalent, and the SICS type is introduced as a very productive special case of the TR type. It also is seen that any TR type can serve to convert the spatial multivariate quantile function into an affine equivariant version. However, the single TR functional that has been used in practice for this purpose (Chakraborty and Chaudhuri, 1996, and Chakraborty, Chaudhuri, and Oja, 1998), has proved difficult to understand theoretically. In fact, this particular TR functional happens to be a particular case of SICS type and is better understood by consideringing this aspect. The present paper studies SICS functionals with some generality, developing their basic properties, constructing particular families of them, illustrating them explicitly for the bivariate case, and examining their roles in several application contexts.

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

Key words and phrases: Invariant coordinate system functionals; Multivariate analysis; Nonparametric methods.

1 Introduction

A leading focus of modern exploratory data analysis and data mining is the search for important features and artifacts of a multivariate data set. Here one must insist that any such discoveries be invariant under affine transformation of the data, at least up to a homogeneous scale change and a translation. Otherwise, what one perceives is at least in part merely an artifact of the particular coordinate system being used.

That is, one should confine attention to artifacts which, when expressed as a statistic computed from the data, are functions of a suitable "maximal invariant" statistic. With reference to multivariate data in $\mathbb{R}^d$ and the group $\mathcal{G}$ of *affine transformations* $\boldsymbol{Ax} + \boldsymbol{b}$ with nonsingular $d \times d$ $\boldsymbol{A}$ and d-vector $\boldsymbol{b}$, a maximal invariant statistic is any labeling of the orbits of $\mathcal{G}$ (see Lehmann and Romano, 2005, §6.2). It turns out that in this setting a maximal invariant is obtained by applying a suitable type of matrix-valued transformation $\boldsymbol{M}(\mathbb{X}_n)$ to the data $\mathbb{X}_n$, producing $\boldsymbol{M}(\mathbb{X}_n)\mathbb{X}_n$ as the desired maximal invariant.

In particular, for this purpose, $\boldsymbol{M}(\cdot)$ must be a *strong invariant coordinate system* (SICS) functional as defined in Serfling (2010), tightening the notion of invariant coordinate system (ICS) functional defined and treated in detail in Tyler, Critchley, Dümbgen, and Oja (2009). Following Serfling (2010, Definition 6.2 and Lemma 6.1), we state

Definition 1.1 A *strong invariant coordinate system (SICS) functional* is a $d \times d$ matrix-valued functional $\boldsymbol{M}(\cdot)$ satisfying, for $\boldsymbol{Y} = \boldsymbol{AX} + \boldsymbol{b}$ with any nonsingular $d \times d$ matrix $\boldsymbol{A}$ and d-vector $\boldsymbol{b}$,

$$\boldsymbol{M}(F_{\boldsymbol{Y}}) = k_1 \boldsymbol{M}(F_{\boldsymbol{X}})\boldsymbol{A}^{-1}, \tag{1}$$

with $k_1 = k_1(\boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}})$ a scalar constant.

Lemma 1.1 *For a SICS functional $\boldsymbol{M}(F)$ and $\boldsymbol{Y} = \boldsymbol{AX} + \boldsymbol{b}$ with any nonsingular $d \times d$ matrix $\boldsymbol{A}$ and d-vector $\boldsymbol{b}$,*

$$\boldsymbol{M}(F_{\boldsymbol{Y}})\boldsymbol{Y} = k_1 \boldsymbol{M}(F_{\boldsymbol{X}})\boldsymbol{X} + \boldsymbol{c}, \tag{2}$$

with $k_1 = k_1(\boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}})$ a scalar constant and $\boldsymbol{c} = \boldsymbol{c}(\boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}}) = k_1 \boldsymbol{M}(F_{\boldsymbol{X}})\boldsymbol{A}^{-1}\boldsymbol{b}$ a vector constant.

The property given by (2) makes SICS functionals very useful in a wide range of practical applications. Thus the $\boldsymbol{M}(\cdot)$-transformed coordinates $\boldsymbol{M}(F_{\boldsymbol{Y}})\boldsymbol{Y}$ constructed after affine transformation agree with the original $\boldsymbol{M}(\cdot)$-transformed coordinates $\boldsymbol{M}(F_{\boldsymbol{X}})\boldsymbol{X}$, except for *a homogeneous scale change* and *a translation*. A similar statement applies to the sample versions $M(\mathbb{X}_n)\mathbb{X}_n$ and $M(\mathbb{Y}_n)\mathbb{Y}_n$. The use of $M(\mathbb{X}_n)\mathbb{X}_n$ instead of $\mathbb{X}_n$ means that structure and patterns which are mere artifacts of the choice of coordinate system, which should be disregarded no matter how striking, do indeed become eliminated. Statistical procedures which are not affine invariant functions of data $\mathbb{X}_n$ become affine invariant in the above sense if redefined as functions of $M(\mathbb{X}_n)\mathbb{X}_n$, i.e., if carried out not on $\mathbb{X}_n$ but rather on $\mathbb{X}_n^* = M(\mathbb{X}_n)\mathbb{X}_n$. Likewise, statistical procedures which are not affine equivariant functions of data $\mathbb{X}_n$ become affine equivariant if first carried out on $M(\mathbb{X}_n)\mathbb{X}_n$ and then retransformed back to $\mathbb{X}_n$ via $M(\mathbb{X}_n)^{-1}$. In particular, as seen later, the coordinatewise multivariate median can be made affine equivariant, and projection pursuit outlyingness functions based on only finitely many projections can be made affine invariant.

Instances of such uses of SICS functionals have been defined *ad hoc* in particular nonparametric inference settings in Chaudhuri and Sengupta (1993), Chakraborty and Chaudhuri (1996), and Chakraborty, Chaudhuri, and Oja (1998), and in a certain semiparametric inference setting in Ilmonen, Nevalainen, and Oja (2010). Here we augment the introduction of the general concept of SICS functionals in Serfling (2010) by a detailed study having complete generality in the nonparametric setting. Key structural properties of SICS functionals are presented in Section **2**. Approaches toward their construction are discussed in Section **3**. Explicit illustration of them is carried out for the bivariate case in Section **4**. Their diverse roles in applications are illustrated briefly in Section **5**.

For later reference, we make precise the interconnections between SICS and other key matrix-valued functionals. As indicated above and discussed in Serfling (2010), SICS functionals are a productive special case of *invariant coordinate system* (ICS) functionals. The latter are defined in Tyler *et al.* (2009) as 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. Also, as seen in Serfling (2010), SICS functionals are closely related to transformation-retransformation (TR) functionals and weak covariance (WC) functionals. That is, any SICS functional $\boldsymbol{M}(F)$ is a *transformation-retransformation* (TR) functional, i.e., one satisfying

$$\boldsymbol{A}'\boldsymbol{M}(F_{\boldsymbol{Y}})'\boldsymbol{M}(F_{\boldsymbol{Y}})\boldsymbol{A} = k_2\boldsymbol{M}(F_{\boldsymbol{X}})'\boldsymbol{M}(F_{\boldsymbol{X}}) \tag{3}$$

for $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$, with $k_2 = k_2(\boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}})$ a positive scalar function of $\boldsymbol{A}$, $\boldsymbol{b}$, and $F_{\boldsymbol{X}}$. (In turn, any TR functional $\boldsymbol{M}(F)$) induces a corresponding *weak covariance* (WC) functional $\boldsymbol{C}(F) = (\boldsymbol{M}(F)'\boldsymbol{M}(F))^{-1}$, i.e., a symmetric positive definite $d \times d$ matrix-valued functional $\boldsymbol{C}(F)$ satisfying

$$\boldsymbol{C}(F_{\boldsymbol{Y}}) = k_3\boldsymbol{A}\boldsymbol{C}(F_{\boldsymbol{X}})\boldsymbol{A}', \tag{4}$$

with $k_3 = k_3(\boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}})$ a positive scalar function of $\boldsymbol{A}$, $\boldsymbol{b}$, and $F_{\boldsymbol{X}}$.)

2 Structural properties of SICS functionals

2.1 On formulation of SICS functionals

To understand the nature of SICS functionals, it is helpful to rewrite (1) as

$$\boldsymbol{M}(F_{\boldsymbol{Y}})^{-1} = k_1^{-1}\boldsymbol{A}\boldsymbol{M}(F_{\boldsymbol{X}})^{-1}. \tag{5}$$

This shows that the *inverse* of a SICS functional is *translation invariant* and *multiplicatively affine equivariant*, subject to a *homogeneous scale change*. These familiar properties make the representation (5) perhaps more straightforward to understand than (1) and motivate the following constructive definition.

Definition 2.1 A nonsingular $d \times d$ matrix-valued functional $\boldsymbol{W}(\cdot)$ is a *pre-SICS* functional if it satisfies, for $\boldsymbol{Y} = \boldsymbol{AX} + \boldsymbol{b}$ with any nonsingular $d \times d$ matrix $\boldsymbol{A}$ and d-vector $\boldsymbol{b}$,

$$\boldsymbol{W}(F_{\boldsymbol{Y}}) = k_0 \boldsymbol{A} \boldsymbol{W}(F_{\boldsymbol{X}}), \tag{6}$$

with $k_0 = k_0(\boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}})$ a scalar constant. □

(Note that $\boldsymbol{W}(F) = (0)_{d\times d}$, where $(c)_{d\times d}$ denotes the $d \times d$ matrix with all elements equal to c for any real c, is also a solution of (6) but not a nonsingular one.) Thus there are two alternative approaches toward formulation of SICS functionals:

1. Seek a SICS functional $\boldsymbol{M}(F)$ satisfying (1), or
2. Seek a pre-SICS functional $\boldsymbol{W}(F)$ satisfying (6) and then take $\boldsymbol{M}(F) = \boldsymbol{W}(F)^{-1}$.

Both approaches are illustrated in Section 3.

Remark 2.1 For the second approach, we might start by constructing a *sample* version, i.e., a (nonsingular) solution to the sample version of (6),

$$\boldsymbol{W}(\mathbb{Y}_n) = k_{0s} \boldsymbol{A} \boldsymbol{W}(\mathbb{X}_n), \tag{7}$$

with $k_{0s} = k_{0s}(\boldsymbol{A}, \boldsymbol{b}, \mathbb{X}_n)$ a scalar constant, and obtain the corresponding sample SICS functional by

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

Then we may pass to population versions via

$$\boldsymbol{M}(F) = E\{\boldsymbol{W}(\mathbb{X}_{n_0})^{-1}\} \tag{8}$$

$$\boldsymbol{W}(F) = \boldsymbol{M}(F)^{-1} = (E\{\boldsymbol{W}(\mathbb{X}_{n_0})^{-1}\})^{-1}, \tag{9}$$

for some suitable conceptual sample size n_0.

It is tempting to define the population version $\boldsymbol{W}(F)$, alternatively, as simply $E\{\boldsymbol{W}(\mathbb{X}_{n_0})\}$ for some suitable conceptual sample size n_0 and then take $\boldsymbol{M}(F) = \boldsymbol{W}(F)^{-1}$. However, this is defeated by two considerations. First, the constant k_{0s} may depend upon $\mathbb{X}_n$, in which case taking expectations across (7) does not lead to a solution of (6). Also, even for $k_{0s} = k_{0s}(\boldsymbol{A}, \boldsymbol{b})$ *not* depending on $\mathbb{X}_n$, as in fact holds for the examples in Section 3, it can happen, and in fact does happen for these same examples, that

$$E\{\boldsymbol{W}(\mathbb{X}_{n_0})\} = (0)_{d\times d}. \tag{10}$$

Hence the population functional $E\{\boldsymbol{W}(\mathbb{X}_{n_0})\}$ provides only a noninformative "parameter" and yields only the trivial *singular* solution of (6). Nevertheless, sample pre-SICS functionals $\boldsymbol{W}(\mathbb{X}_n)$ constructed in practice are indeed *nonsingular* and perform quite effectively.

The indirect approach using (8) and (9) requires only that $\boldsymbol{M}(F)$ as defined by (8) be finite. (This is an analogue of defining the parameters $\theta = E(1/W)$ and $\eta = 1/\theta$ for a univariate random variable W having mean 0.) □

2.2 Properties of pre-SICS and SICS functionals

Remark 2.2 A SICS functional $\boldsymbol{M}(F)$ is *neither symmetric nor triangular.* As discussed in Serfling (2010), a TR functional is some choice of square root of the inverse of the associated WC functional $\boldsymbol{C}(F) = (\boldsymbol{M}(F)'\boldsymbol{M}(F))^{-1}$. Typical choices of such square roots are either symmetric or triangular. However, when the TR functional $\boldsymbol{M}(F)$ is symmetric or triangular and also SICS, then $\boldsymbol{M}(F)\boldsymbol{A}^{-1}$ must also be symmetric or triangular for arbitrary $\boldsymbol{A}$. It is easy to find counterexamples to this possibility. □

Remark 2.3 *Using two SICS functionals successively.* Let $\boldsymbol{M}_1(F)$ and $\boldsymbol{M}_2(F)$ be two SICS functionals. If $\mathbb{X}_n$ is transformed to $\mathbb{X}_n^* = \boldsymbol{M}_1(\mathbb{X}_n)\mathbb{X}_n$, and $\mathbb{X}_n^*$ to $\mathbb{X}_n^{**} = \boldsymbol{M}_2(\mathbb{X}_n^*)\mathbb{X}_n^*$, then we have

$$\begin{aligned} \boldsymbol{M}_2(\mathbb{X}_n^*)\mathbb{X}_n^* &= \boldsymbol{M}_2(\boldsymbol{M}_1(\mathbb{X}_n)\mathbb{X}_n)\boldsymbol{M}_1(\mathbb{X}_n)\mathbb{X}_n \\ &= k_2(\boldsymbol{M}_1, \mathbf{0}, \mathbb{X}_n)\boldsymbol{M}_2(\mathbb{X}_n)\boldsymbol{M}_1(\mathbb{X}_n)^{-1}\boldsymbol{M}_1(\mathbb{X}_n)\mathbb{X}_n \\ &= k_2(\boldsymbol{M}_1, \mathbf{0}, \mathbb{X}_n)\boldsymbol{M}_2(\mathbb{X}_n)\mathbb{X}_n. \end{aligned}$$

That is, carrying out a series of successive SICS standardizations is equivalent to just using the most recent one in the first place. □

In the following result, we use the notation $\boldsymbol{D}_c$ for a diagonal matrix with each diagonal element equal to c.

Lemma 2.1 *(i) If $\boldsymbol{W}(F)$ and $\boldsymbol{M}(F)$ are pre-SICS and SICS, respectively, then so are $c\boldsymbol{W}(F)$ and $c\boldsymbol{M}(F)$, respectively, for any constant c. The same holds for sample versions.*

(ii) If $\boldsymbol{X} \overset{d}{=} c\boldsymbol{Z}$ for some constant c, then

$$\boldsymbol{M}(F_{\boldsymbol{X}}) = k_1(\boldsymbol{D}_c, \mathbf{0}, F_{\boldsymbol{Z}})\, c^{-1}\boldsymbol{M}(F_{\boldsymbol{Z}}) \tag{11}$$

and

$$\boldsymbol{W}(\mathbb{X}_n) \overset{d}{=} k_{0s}(\boldsymbol{D}_c, \mathbf{0}, \mathbb{Z}_n)\, c\,\boldsymbol{W}(\mathbb{Z}_n). \tag{12}$$

(iii) More generally, if $\boldsymbol{X}$ and $\boldsymbol{Z}$ are affinely equivalent in distribution, i.e., $\boldsymbol{X} \overset{d}{=} \boldsymbol{A}\boldsymbol{Z}+\boldsymbol{b}$, then $\boldsymbol{M}(F_{\boldsymbol{X}})$ and $\boldsymbol{M}(F_{\boldsymbol{Z}})$ are proportional, and likewise for $\boldsymbol{W}(\mathbb{X}_n)$ and $\boldsymbol{W}(\mathbb{Z}_n)$.

Statements (ii) and (iii) give conditions under which the determination of a pre-SICS or SICS functional for a given distribution can be reduced to finding it for an associated standard distribution, within a constant of proportionality. In many applications, such constants drop out.

Proof. (i) This is immediate from the definitions.

(ii) If $\boldsymbol{X} \overset{d}{=} c\boldsymbol{Z} = \boldsymbol{D}_c\boldsymbol{Z}$, then $\boldsymbol{M}(F_{\boldsymbol{X}}) = \boldsymbol{M}(F_{c\boldsymbol{Z}}) = \boldsymbol{M}(F_{\boldsymbol{D}_c\boldsymbol{Z}}) = k_1(\boldsymbol{D}_c, \mathbf{0}, F_{c\boldsymbol{Z}})\boldsymbol{M}(F_{\boldsymbol{Z}})\boldsymbol{D}_c^{-1}$ $= k_1(\boldsymbol{D}_c, \mathbf{0}, F_{c\boldsymbol{Z}})\, c^{-1}\boldsymbol{M}(F_{\boldsymbol{Z}})$, establishing (11). Similar steps using $\boldsymbol{W}(\mathbb{X}_n) \overset{d}{=} \boldsymbol{W}(c\mathbb{Z}_n)$ lead to (12).

(iii) We use similar arguments as for (ii). □

Lemma 2.2 *Let $\boldsymbol{\theta}(\mathbb{X}_n)$ be a translation invariant d-vector. If $\boldsymbol{W}(\mathbb{X}_n)$ is pre-SICS with corresponding SICS functional $\boldsymbol{M}(\mathbb{X}_n) = \boldsymbol{W}(\mathbb{X}_n)^{-1}$, and with proportionality constants $k_{0s} = k_{0s}(\boldsymbol{A}, \boldsymbol{b})$ and $k_1 = k_1(\boldsymbol{A}, \boldsymbol{b})$ not depending on $\mathbb{X}_n$, then also pre-SICS is*

$$\widetilde{\boldsymbol{W}}(\mathbb{X}_n) = \boldsymbol{W}(\widetilde{\mathbb{X}}_n),$$

where $\widetilde{\mathbb{X}}_n = \{\widetilde{\boldsymbol{X}}_i,\ 1 \leq i \leq n\}$, with

$$\widetilde{\boldsymbol{X}}_i = \|\boldsymbol{M}(\mathbb{X}_n)(\boldsymbol{X}_i - \theta(\mathbb{X}_n))\|^{\alpha} (\boldsymbol{X}_i - \boldsymbol{\theta}(\mathbb{X}_n)),\ 1 \leq i \leq n,$$

for any constant α, $-\infty < \alpha < \infty$.

Proof. Let $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$. Then $\boldsymbol{M}(\mathbb{Y}_n) = k_1(\boldsymbol{A}, \boldsymbol{b})\boldsymbol{M}(\mathbb{X}_n)\boldsymbol{A}^{-1}$ and it is easily seen that $\widetilde{\mathbb{Y}}_n = k_1(\boldsymbol{A}, \boldsymbol{b})^{\alpha}\boldsymbol{A}\widetilde{\mathbb{X}}_n$. Then

$$\begin{aligned}
\widetilde{\boldsymbol{W}}(\mathbb{Y}_n) &= \boldsymbol{W}(\widetilde{\mathbb{Y}}_n) \\
&= \boldsymbol{W}(k_1(\boldsymbol{A}, \boldsymbol{b})^{\alpha}\boldsymbol{A}\widetilde{\mathbb{X}}_n) \\
&= k_{0s}(k_1(\boldsymbol{A}, \boldsymbol{b})^{\alpha}\boldsymbol{A}, \boldsymbol{0})\boldsymbol{A}\boldsymbol{W}(\widetilde{\mathbb{X}}_n) \\
&= k_{0s}(k_1(\boldsymbol{A}, \boldsymbol{b})^{\alpha}\boldsymbol{A}, \boldsymbol{0})\boldsymbol{A}\widetilde{\boldsymbol{W}}(\mathbb{X}_n),
\end{aligned}$$

so that $\widetilde{\boldsymbol{W}}(\mathbb{X}_n)$ is pre-SICS with constant $\widetilde{k}_{0s}(\boldsymbol{A}, \boldsymbol{b}) = k_{0s}(k_1(\boldsymbol{A}, \boldsymbol{b})^{\alpha}\boldsymbol{A}, \boldsymbol{0})$ □

3 Construction of pre-SICS and SICS functionals

Let $\mathbb{X}_n = \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\}$ be i.i.d. observations in $\mathbb{R}^d$ with distribution $F_{\boldsymbol{X}}$. Our chief interest is in construction of pre-SICS or SICS functionals by general approaches available in *completely nonparametric* (i.e., no restrictions on F) settings. This is pursued in Sections 3.1–3.4 below. Also, in Section 3.5 we briefly indicate approaches which set substantial restrictions on F and thus are *semiparametric*.

3.1 The Chaudhuri and Sengupta (1993) example

Evidently, the first example of SICS functional in the literature was introduced by Chaudhuri and Sengupta (1993) as a pre-SICS functional in the context of testing $H : \boldsymbol{\theta} = \boldsymbol{0}$ versus $H : \boldsymbol{\theta} \neq \boldsymbol{0}$ in the location model $F_{\boldsymbol{X}} = F_0(\boldsymbol{x} - \boldsymbol{\theta})$. Since $\boldsymbol{A}\boldsymbol{\theta} = \boldsymbol{0}$ if and only if $\boldsymbol{\theta} = \boldsymbol{0}$, Chaudhuri and Sengupta suggest using a test function ϕ satisfying $\phi(\boldsymbol{A}\boldsymbol{x}) = \phi(\boldsymbol{x})$, all nonsingular $\boldsymbol{A}$, thus making the same decision before and after any nonsingular transformation of the coordinate system. This motivates choosing the test procedure to be some function of a maximal invariant statistic relative to the group of nonsingular transformations $\boldsymbol{A}$.

Based on $\mathbb{X}_n$, and for each fixed choice of d distinct indices $\mathbb{J} = \{i_1, \ldots, i_d\}$ from $\{1, \ldots, n\}$, Chaudhuri and Sengupta define the matrix

$$\boldsymbol{W}_{\mathbb{J}}(\mathbb{X}_n) = [\boldsymbol{X}_{i_1}, \ldots, \boldsymbol{X}_{i_d}]_{d \times d} \tag{13}$$

and show that for $F_{\boldsymbol{X}}$ absolutely continuous the transformed observations (i.e, "data-driven coordinates")

$$\mathbb{Y}_n^{(\mathbb{J})} = \boldsymbol{W}_{\mathbb{J}}(\mathbb{X}_n)^{-1}\mathbb{X}_n$$

form a maximal invariant statistic with respect to the nonsingular transformations $\boldsymbol{A}$.

We observe here that, for each $\mathbb{J}$, the matrix $\boldsymbol{W}_{\mathbb{J}}(\mathbb{X}_n)$ is a pre-SICS transformation in the sense of (7), in this case with $k_{0s} \equiv 1$. This fact is given as a step in the proof by Chaudhuri and Sengupta of maximal invariance of $\mathbb{Y}_n^{(\mathbb{J})}$ (with $k_{0s} \equiv 1$ playing an essential role), but they do not otherwise comment on the SICS property.

Let $\mathcal{C}(d, n)$ denote the class of all sets of d distinct integers from $\{1, \ldots, n\}$. It follows that the statistic

$$\xi_n = \{\mathbb{Y}_n^{(\mathbb{J})}, \mathbb{J} \subset \mathcal{C}(d, n)\}$$

is also maximal invariant and has the further desirable property of being invariant over permutations of the indices of the observations, i.e., ξ_n is symmetric in the observations, although this latter is obtained at the cost of considerable extra computation. Chaudhuri and Sengupta develop a family of multivariate sign tests based on ξ_n.

Remark 3.1 The SICS transformation solution to the problem of constructing affine invariant location tests is more sweeping than necessary for some purposes. Namely, if $\boldsymbol{M}(\mathbb{X}_n)$ is *any* SICS transformation relative to the group of nonsingular transformations $\boldsymbol{A}$, and $\phi(x_1, \ldots, x_n)$ is *any* test function, then the test function $\phi(\boldsymbol{M}(\mathbb{X}_n)x_1, \ldots, \boldsymbol{M}(\mathbb{X}_n)x_n)$ is affine invariant. This is quite encompassing. In some situations, however, one may have reason to impose certain requirements that somewhat narrow the class of test functions ϕ under consideration. In such a case, one might be able to obtain affine invariance of $\phi(\boldsymbol{M}(\mathbb{X}_n)x_1, \ldots, \boldsymbol{M}(\mathbb{X}_n)x_n)$ for $\boldsymbol{M}(\mathbb{X}_n)$ not necessarily as strong as SICS. For example, Randles (2000) achieves some favorable distribution-free properties by restricting to test functions ϕ which are orthogonally invariant functions of their arguments and are applied to the data through their *signs*. For such ϕ, it follows that $\phi(\boldsymbol{M}(\mathbb{X}_n)x_1, \ldots, \boldsymbol{M}(\mathbb{X}_n)x_n)$ is affine invariant if merely $\boldsymbol{M}(\mathbb{X}_n)$ is a TR functional (see also Serfling, 2010, for related discussion in connection with the spatial quantile function). □

3.2 The Chakraborty and Chaudhuri (1996) example

In the setting of *estimating* location rather than testing a specified value, Chakraborty and Chaudhuri (1996) introduce a variation of the Chaudhuri and Sengupta (1993) pre-SICS transformation that eliminates any specifed value, namely

$$\boldsymbol{W}_{\mathbb{J}}(\mathbb{X}_n) = \left[(\boldsymbol{X}_{i_1} - \boldsymbol{X}_{i_{d+1}}), \ldots, (\boldsymbol{X}_{i_d} - \boldsymbol{X}_{i_{d+1}})\right]_{d \times d}, \tag{14}$$

with the index set $\mathbb{J} = \{i_1, \ldots, i_{d+1}\}$ in $\mathcal{C}(d+1, n)$, thus using $d+1$ sample observations, and show that it is a maximal invariant with respect to invertible affine transformations $\boldsymbol{Ax} + \boldsymbol{b}$. They then show that computing the coordinate-wise median on the observations $\{Y_i^{(\mathbb{J})}, i \notin \mathbb{J}\}$ $= \boldsymbol{W}_{\mathbb{J}}(\mathbb{X}_n)^{-1}X_i, i \notin \mathbb{J}\}$ and retransforming that result back to the original coordinates via the inverse $\boldsymbol{W}_{\mathbb{J}}(\mathbb{X}_n)$ yields a fully affine equivariant version of the sample coordinatewise median, the "transformation-retransformation (TR)" coordinatewise median. The key step in the proof is an application of the pre-SICS property (7), which is immediately satisfied by this $\boldsymbol{W}_{\mathbb{J}}(\mathbb{X}_n)$ (with $k_{0s} \equiv 1$, although this is not essential to the proof).

Chakraborty and Chaudhuri select $\mathbb{J}$ to make the matrix $\boldsymbol{W}_{\mathbb{J}}' \widehat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{W}_{\mathbb{J}}$ 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 $\widehat{\boldsymbol{\Sigma}}^{-1/2} \boldsymbol{W}_{\mathbb{J}}$ as orthonormal as possible, where $\widehat{\boldsymbol{\Sigma}}$ is a consistent (at least proportionally) estimator of the population scatter matrix. (A highly robust, computationally efficient in low dimension, example is the FAST-MCD sample scatter matrix.) While only $d+1$ observations are used in defining $\boldsymbol{W}_{\mathbb{J}}(\mathbb{X}_n)$, all of the data now have been "looked at" in the process of choosing $\mathbb{J}$. However, we note that the computational burden in this method includes more than the first step of getting $\widehat{\Sigma}$. The continuing steps to find the "optimal" set $\mathbb{J}$ by checking all combinations are of order $O(n^{d+1})$ and become prohibitive very quickly as d increases.

Remark 3.2 (i) The WC functional $\boldsymbol{C}_{\mathbb{J}}(\mathbb{X}_n)$ corresponding to $\boldsymbol{W}_{\mathbb{J}}(\mathbb{X}_n)$ is $\boldsymbol{W}_{\mathbb{J}}(\mathbb{X}_n)\boldsymbol{W}_{\mathbb{J}}(\mathbb{X}_n)'$, with expected value $E\{\boldsymbol{C}_{\mathbb{J}}(\mathbb{X}_n)\} = 2d\boldsymbol{\Sigma}(F_{\boldsymbol{X}})$.

(ii) The same $\boldsymbol{W}_{\mathbb{J}}(\mathbb{X}_n)$ is used by Chakraborty, Chaudhuri, and Oja (1998) to develop an affine equivariant (TR) modification of the sample spatial median and an affine invariant multivariate spatial sign test. They also introduce a strategy to reduce the amount of computation by stopping the search over all subsets $\mathbb{J}$ of size $d+1$ from $\{1, \ldots, n\}$ as soon as the ratio between the geometric mean and harmonic mean of the eigenvalues of $\boldsymbol{W}_{\mathbb{J}}' \widehat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{W}_{\mathbb{J}}$ becomes less than $1 + \varepsilon$. Chakraborty (2001) extends this TR approach to the entire spatial quantile function. (As shown in Serfling, 2010, one may also obtain such affine equivariance and invariance properties for sample spatial statistics using any TR transformation.) □

3.3 Generalization: a family of pre-SICS functionals

Here we discuss in detail the family of pre-SICS functionals introduced in Serfling (2010) and containing (14) as a special case. Also, in Section 3.4 we introduce a related but different family.

Suppose now that a subsample $\mathbb{X}_N^* = \{\boldsymbol{X}_1^*, \ldots, \boldsymbol{X}_N^*\}$. of size $N \leq n$ is selected by some method. by For example, $\mathbb{X}_N^*$ may be the whole sample ($N = n$), or the subset of observations indexed by $\mathbb{J}$ in the Chakraborty and Chaudhuri (1996) approach ($N = d+1$), or the subset of observations selected in the FAST-MCD algorithm (e.g., $N = \lfloor (n+d+1)/2 \rfloor$). In Remark 3.4, we discuss a an approach using "spatial trimming", with any arbitrary choice of N, which may be computed easily in any dimension.

Next let us choose an integer $m \leq N/(d+1)$, for example $m = \lfloor N/(d+1) \rfloor$, and form $d+1$ means each using m distinct observations from $\mathbb{X}_N^*$:

$$\overline{\boldsymbol{X}}_i^* = m^{-1} \sum_{j=1}^{m} \boldsymbol{X}_{(i-1)m+j}^*, \ 1 \leq i \leq d+1.$$

Finally, we form

$$\boldsymbol{W}_{(m,N)}(\mathbb{X}_n) = \Big[(\overline{\boldsymbol{X}}_1^* - \overline{\boldsymbol{X}}_{d+1}^*), \ldots, (\overline{\boldsymbol{X}}_d^* - \overline{\boldsymbol{X}}_{d+1}^*)\Big]_{d\times d}. \tag{15}$$

It is immediate that $\boldsymbol{W}_{(m,N)}(\mathbb{X}_n)$ satisfies (7) with $k_{0s} \equiv 1$. It also uses nearly all if not all of the N selected observations. With suitable choice of $\mathbb{J}$ it is invariant over permutations. Also, it is potentially computationally attractive, the main computational burden being the steps to derive $\mathbb{X}_N^*$ from $\mathbb{X}_n$ and the steps to compute the $d+1$ means. Further, with suitable choice of $\mathbb{J}$ it can be quite robust.

The case $(m, N) = (1, d+1)$, i.e., $\boldsymbol{W}_{(1,d+1)}(\mathbb{X}_n)$, is just (14). The case of chief interest to us here, however, is $(m, N) = (\lfloor N/(d+1)\rfloor, N)$, which uses as many observations as possible in each mean. In any case, for all choices of (m, N), the properties of $\boldsymbol{W}_{(m,N)}(\mathbb{X}_n)$ are very similar. For example, $E\{\boldsymbol{W}_{(m,N)}(\mathbb{X}_n)\} = (0)_{d\times d}$.

An important structural feature of $\boldsymbol{W}_{(m,N)}(\mathbb{X}_n)$ may be stated as follows. Let $\boldsymbol{Y}$ denote a random variable equal in distribution to the mean $\overline{\boldsymbol{X}}^{(m)}$ of m observations on $F_{\boldsymbol{X}}$, i.e., $F_{\boldsymbol{Y}} = F_{\overline{\boldsymbol{X}}^{(m)}}$. Let $\mathbb{Y}_{d+1}$ be a sample of size $d+1$ from $F_{\boldsymbol{Y}}$. Then we have

$$\boldsymbol{W}_{(m,N)}(\mathbb{X}_n) \stackrel{d}{=} \boldsymbol{W}_{(1,d+1)}(\mathbb{Y}_{d+1}). \tag{16}$$

That is, we may behave as if the "data" is given by the reduced data

$$\mathbb{Y}_{d+1} = \{\boldsymbol{Y}_i = \overline{\boldsymbol{X}}_i^{(m)},\ 1 \le i \le d+1\}.$$

Remark 3.3 (i) An important role in some applications is played by the special case that $\boldsymbol{X}$ is a standard d-variate normal random variable, say $\boldsymbol{Z}$, i.e., $F_{\boldsymbol{Z}} = N(\boldsymbol{0}, \boldsymbol{I}_d)$. Then $\boldsymbol{Y}$ has distribution $N(\boldsymbol{0}, m^{-1}\boldsymbol{I}_d)$, i.e., $\boldsymbol{Y} \stackrel{d}{=} m^{-1/2}\boldsymbol{Z}$, and $F_{\boldsymbol{Y}} = F_{m^{-1/2}\boldsymbol{Z}}$. We now apply Lemma 2.1(ii) and obtain in this normal case

$$\boldsymbol{W}_{(m,N)}(\mathbb{X}_n) \stackrel{d}{=} \boldsymbol{W}_{(1,d+1)}(\mathbb{Y}_{d+1}) \stackrel{d}{=} m^{-1/2}\boldsymbol{W}_{(1,d+1)}(\mathbb{Z}_{d+1}). \tag{17}$$

(ii) In applications involving the $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ model for $\boldsymbol{X}$ with unknown parameters $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the *distribution theory* pertinent to the use of $\boldsymbol{W}_{(m,N)}(\mathbb{X}_n)^{-1}$ as a standardizing transformation for the data $\mathbb{X}_n$ can be reduced without loss of generality to the case that $\boldsymbol{X}$ is *standard* d-variate normal. Then by (17) the distribution theory for $\boldsymbol{W}_{(m,N)}(\mathbb{X}_n)$ is given by that for $m^{-1/2}\boldsymbol{W}_{(1,d+1)}(\mathbb{Z}_{d+1})$, which can be determined as a characteristic of the known population $N(\boldsymbol{0}, \boldsymbol{I}_d)$. For example, we thus have $\boldsymbol{M}_{(m,N)}(F_{\boldsymbol{X}}) = E\{\boldsymbol{W}_{(m,N)}(\mathbb{X}_n)^{-1}\} = m^{-1/2}E\{\boldsymbol{W}_{(1,d+1)}(\mathbb{Z}_{d+1})^{-1}\} = m^{-1/2}\boldsymbol{M}_{(1,d+1)}(F_{\boldsymbol{Z}})$. This motivates special investigation of $\boldsymbol{M}_{(1,d+1)}(F_{\boldsymbol{Z}})$. □

Remark 3.4 On choice of N and $\mathbb{J}$. Augmenting the approaches for $\mathbb{J}$ discussed above, we present a method that both is robust and entails low computational burden regardless of the dimension. For this, we apply the "spatial trimming" method of Mazumder (2010) and Mazumder and Serfling (2010), summarized as follows:

1. Evaluate the sample Dümbgen-Tyler TR matrix $\boldsymbol{M}_0(\mathbb{X}_n)$. This is easily computed in any dimension.

2. Compute the corresponding TR spatial outlyingness for the sample observations $\boldsymbol{X}_i$, $1 \leq i \leq n$:

$$O_{\mathrm{S}}^{(\mathrm{TR})}(\boldsymbol{X}_i, \mathbb{X}_n) = \left\| n^{-1} \sum_{j=1}^{n} \boldsymbol{S}\left(\boldsymbol{M}_0(\mathbb{X}_n)(\boldsymbol{X}_i - \boldsymbol{X}_j)\right) \right\|, \ 1 \leq i \leq n.$$

   This also is easily computed in any dimension.

3. For a specified integer $N \leq n$, let $\mathbb{J} = \{j_1, j_2, \ldots, j_N\}$ denote the indices of the $\boldsymbol{X}_i$ possessing the N smallest values of $O_{\mathrm{S}}^{(\mathrm{TR})}(\boldsymbol{X}_i, \mathbb{X}_n)$.

The set $\mathbb{X}_N^* = \{\boldsymbol{X}_1^*, \ldots, \boldsymbol{X}_N^*\} = \{\boldsymbol{X}_{j_1}, \ldots, \boldsymbol{X}_{j_N}\}$ of the N "innermost observations" is thus robust, is symmetric in the data set $\mathbb{X}_n$, and is obtained with computational ease in any dimension. □

3.4 Another family of SICS functionals

Let $\boldsymbol{M}(\mathbb{X}_n)$ be a TR functional as per (3) with scalar constant $k_2(\boldsymbol{A}, \boldsymbol{b}, \mathbb{X}_n)$. Then, with $\boldsymbol{S}(\cdot)$ the usual sign function on $\mathbb{R}^d$,

$$\widetilde{\boldsymbol{W}}_{(m,N)}(\mathbb{X}_n) = \boldsymbol{M}(\mathbb{X}_n)^{-1} \left[\boldsymbol{S}\left(\boldsymbol{M}(\mathbb{X}_n)(\overline{\boldsymbol{X}}_1^* - \overline{\boldsymbol{X}}_{d+1}^*)\right), \ldots, \boldsymbol{S}\left(\boldsymbol{M}(\mathbb{X}_n)(\overline{\boldsymbol{X}}_d^* - \overline{\boldsymbol{X}}_{d+1}^*)\right)\right]_{d\times d}$$

is a pre-SICS functional with scalar constant $k_{0s}(\boldsymbol{A}, \boldsymbol{b}, \mathbb{X}_n) = k_2(\boldsymbol{A}, \boldsymbol{b}, \mathbb{X}_n)^{1/2}$. To see this, let $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$ and without loss of generality let $\boldsymbol{b} = \boldsymbol{0}$. Then

$$\begin{aligned}
\widetilde{\boldsymbol{W}}_{(m,N)}(\mathbb{Y}_n) &= \boldsymbol{M}(\mathbb{Y}_n)^{-1} \left[\boldsymbol{S}\left(\boldsymbol{M}(\mathbb{Y}_n)(\overline{\boldsymbol{Y}}_1^* - \overline{\boldsymbol{Y}}_{d+1}^*)\right), \ldots\right]_{d\times d} \\
&= \left[\frac{\boldsymbol{A}(\overline{\boldsymbol{X}}_1^* - \overline{\boldsymbol{X}}_{d+1}^*)}{\left((\overline{\boldsymbol{X}}_1^* - \overline{\boldsymbol{X}}_{d+1}^*)'\boldsymbol{A}'(k_2\boldsymbol{A}(\boldsymbol{M}(\mathbb{X}_n)'\boldsymbol{M}(\mathbb{X}_n))^{-1}\boldsymbol{A}')^{-1}\boldsymbol{A}(\overline{\boldsymbol{X}}_1^* - \overline{\boldsymbol{X}}_{d+1}^*)\right)^{1/2}}, \ldots\right]_{d\times d} \\
&= k_2^{1/2}\boldsymbol{A}\,\boldsymbol{M}(\mathbb{X}_n) \left[\boldsymbol{S}\left(\boldsymbol{M}(\mathbb{X}_n)(\overline{\boldsymbol{X}}_1^* - \overline{\boldsymbol{X}}_{d+1}^*)\right), \ldots\right]_{d\times d} \\
&= k_2^{1/2}\boldsymbol{A}\,\widetilde{\boldsymbol{W}}_{(m,N)}(\mathbb{X}_n).
\end{aligned}$$

A robust and computationally easy implementation of this construction is as follows. Choose N and obtain a set of inner observations via spatial trimming as in Remark 3.4. For this set of observations indexed by $\mathbb{J}$, obtain the usual covariance matrix, say $\boldsymbol{\Sigma}_{\mathbb{J}}(\mathbb{X}_n)$. Then take $\boldsymbol{M}(\mathbb{X}_n) = \boldsymbol{\Sigma}_{\mathbb{J}}(\mathbb{X}_n)^{-1/2}$ or any other TR factorization of $\boldsymbol{\Sigma}_{\mathbb{J}}(\mathbb{X}_n)$.

3.5 Construction of SICS functionals in a semiparametric setting

Tyler, Critchley, Dümbgen, and Oja (2009) and Ilmonen, Nevalainen, and Oja (2010) construct ICS functionals and certain cases of SICS functionals by methods based on two WC functionals $\boldsymbol{S}_1(F)$ and $\boldsymbol{S}_2(F)$ and (in some cases) also two location functionals let $\boldsymbol{T}_1(F)$ and $\boldsymbol{T}_2(F)$. Restrictions on F are imposed, thus making the setting *semiparametric*. For example, let a matrix-valued functional $\boldsymbol{M}(F)$ and a diagonal matrix-valued functional $\boldsymbol{L}(F)$ be defined as solutions of the eigenvector and eigenvalue problem

$$\boldsymbol{S}_1(F)^{-1}\boldsymbol{S}_2(F)\boldsymbol{M}(F)' = \boldsymbol{M}(F)'\boldsymbol{L}(F),$$

subject to

$$\begin{aligned} \boldsymbol{M}(F)\boldsymbol{S}_1(F)\boldsymbol{M}(F) &= \boldsymbol{I}_d \\ \boldsymbol{M}(F)\boldsymbol{S}_2(F)\boldsymbol{M}(F)' &= \boldsymbol{L}(F) \\ \boldsymbol{M}(F)(\boldsymbol{T}_1(F) - \boldsymbol{T}_2(F)) &\geq 0, \end{aligned}$$

where the eigenvalues in $\boldsymbol{L}(F)$ are in decreasing order. If F belongs to

$$\mathcal{F} = \{F : \ \boldsymbol{L}(F) \text{ has distinct diagonal elements and } \boldsymbol{M}(F)(\boldsymbol{T}_1(F) - \boldsymbol{T}_2(F)) > 0\},$$

then the functional $\boldsymbol{M}(F)$ is a uniquely defined SICS functional. However, this particular $\mathcal{F}$. excludes elliptical F and distributions with i.i.d. components, for example. For details, see Ilmonen, Nevalainen, and Oja (2010), who point out that on the other hand this $\mathcal{F}$ does include with probability 1 the sample distribution function $\widehat{F}_n$ based on a random sample $\mathbb{X}_n$ from any continuous d-variate distribution F, in which case the sample version $\boldsymbol{M}(\mathbb{X}_n)$ is then SICS and has asymptotic behavior derivable from the joint asymptotic behavior of $\boldsymbol{S}_1(\widehat{F}_n)$, $\boldsymbol{S}_2(\widehat{F}_n)$, $\boldsymbol{T}_1(\widehat{F}_n)$, and $\boldsymbol{T}_2(\widehat{F}_n)$.

4 Explicit illustration of SICS functionals in the bivariate case

5 Selected applications of SICS functionals

Here we augment the applications in Sections 3.1 and 3.2 with several further contexts.

5.1 Projection pursuit scaled deviation vectors

The projection pursuit approach toward formulation of multivariate outlyingness (or depth) functions is well-established and uses the supremum of the univariate scaled deviation outlyingness

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

taken over all univariate projections, for location and spread measures $\mu(F)$ and $\sigma(F)$ (e.g., see Zuo, 2003). More generally, for any set Δ of unit vectors $\boldsymbol{u}$ in $\mathbb{R}^d$, we may define the corresponding projection pursuit outlyingness of a point $\boldsymbol{x}$ in $\mathbb{R}^d$ by

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

For Δ the set of *all* projections, this is the above projection outlyingness and is affine invariant. However, for Δ *finite*, not even orthogonal invariance holds. On the other hand, after first standardizing the data with a SICS transformation $\boldsymbol{M}(F)$, the modified outlyingness function defined by

$$\widetilde{O}_\Delta(\boldsymbol{x}, F) = \sup_{\boldsymbol{u} \in \Delta} O(\boldsymbol{u}'\boldsymbol{M}(F_{\boldsymbol{X}})\boldsymbol{x}, F_{\boldsymbol{u}'\boldsymbol{M}(F_{\boldsymbol{X}})\boldsymbol{X}})$$

is affine invariant for any choice of finite Δ. See Serfling (2010) for detailed discussion.

Acknowledgements

The author gratefully acknowledges very helpful comments by G. L. Thompson and support under National Science Foundation Grants DMS-0805786 and DMS-0940165 and National Security Agency Grant H98230-08-1-0106.

References

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

[2] 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.

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

[4] 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.

[5] Ilmonen, P., Nevalainen, J., and Oja, H. (2010). Characteristics of multivariate distributions and the invariant coordinate system. Preprint.

[6] Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*, 3rd edition. Springer.

[7] Mazumder, S. (2010). Affine Invariant, Robust, and Computationally Easy Multivariate Outlier Identification and Related Methods. Ph.D. dissertation, University of Texas at Dallas, May 2010.

[8] Mazumder, S. and Serfling, R. (2010). Spatial trimming, with applications to robustify sample spatial quantile and outlyingness functions, and to construct a new robust scatter estimator. Submitted.

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

[10] Serfling, R. (2010). Equivariance and invariance properties of multivariate quantile and related functions, and the role of standardization. *Journal of Nonparametric Statistics* **22** 915–936.

[11] 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.

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