On Invariant Coordinate System (ICS) Functionals

Pauliina Ilmonen

School of Health Sciences, FI-33014 University of Tampere, Finland, and Aalto University School of Economics, FI-00076 Aalto, Finland.

Hannu Oja

School of Health Sciences, FI-33014 University of Tampere, Finland.

Robert Serfling

Department of Mathematics, University of Texas at Dallas, Richardson, Texas 75080-3021, USA.

Summary. Equivariance and invariance issues often arise in multivariate statistical analysis. Statistical procedures have to be modified sometimes to obtain an affine equivariant or invariant version. This is often done by preprocessing the data, e.g., by standardizing the multivariate data or by transforming the data to an invariant coordinate system. In this article standardization of multivariate distributions, and characteristics of invariant coordinate system (ICS) functionals and statistics, are examined. Also, invariances up to some groups of transformations are discussed. Constructions of ICS functionals are addressed. In particular the construction based on the use of two scatter matrix functionals presented by Tyler et al. (2009), and direct definitions based on the approach presented by Chaudhuri and Sengupta (1993) are examined. Diverse applications of ICS functionals are discussed.

Keywords: Invariant coordinate system functionals, Multivariate analysis, Nonparametric methods.

1. Introduction

Equivariance and invariance issues often arise in multivariate statistical analysis. Statistical procedures have to be modified sometimes to obtain an affine equivariant or invariant version. This can be done by preprocessing the data, e.g., by standardizing the multivariate data or by transforming the data to an invariant coordinate system. Some traditionally used approaches are not fully affine invariant. For example the data transformation used in principal component analysis is invariant only under orthogonal transformations.

Location and scatter functionals and statistics are used to describe and standardize the multivariate distribution or data. Let $\mathbb{X} = \{\boldsymbol{X}_1, ..., \boldsymbol{X}_n\}$ be a random sample of size n from a d-variate distribution $F_{\boldsymbol{X}}$. Let $\mathcal{A}$ be the set of all nonsingular $d \times d$ matrices and $\mathcal{S}$ the set of all positive definite symmetric $d \times d$ matrices.

Definition 1.1. *A location functional $\boldsymbol{T}(F) \in \mathbb{R}^d$ is a vector-valued functional which is affine equivariant, that is,*

$$\boldsymbol{T}(F_{\boldsymbol{AX}+\boldsymbol{b}}) = \boldsymbol{AT}(F_{\boldsymbol{X}}) + \boldsymbol{b}$$

for all $\boldsymbol{A} \in \mathcal{A}$ and $\boldsymbol{b} \in \mathbb{R}^d$.

Definition 1.2. *A scatter functional or a covariance functional* $\boldsymbol{S}(F) \in \mathcal{S}$ *is a matrix valued functional which is affine equivariant in the sense that*

$$\boldsymbol{S}(F_{\boldsymbol{AX}+\boldsymbol{b}}) = \boldsymbol{A}\boldsymbol{S}(F_{\boldsymbol{X}})\boldsymbol{A}'$$

for all $\boldsymbol{A} \in \mathcal{A}$ *and all* $\boldsymbol{b} \in \mathbb{R}^d$*. If* $\boldsymbol{S}(F) \in \mathcal{S}$ *satisfies a weaker condition*

$$\boldsymbol{S}(F_{\boldsymbol{AX}+\boldsymbol{b}}) = k\boldsymbol{A}\boldsymbol{S}(F_{\boldsymbol{X}})\boldsymbol{A}'$$

where $k = k(\boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{X}})$*, then* $\boldsymbol{S}(F)$ *is called a weak covariance functional (Serfling (2010)).*

The corresponding sample statistics are obtained if the functionals are applied to the empirical cumulative distributions F_n based on $\mathbb{X}$. We then write $\boldsymbol{T}(F_n)$ and $\boldsymbol{S}(F_n)$, or, $\boldsymbol{T}(\mathbb{X})$ and $\boldsymbol{S}(\mathbb{X})$, respectively, and $\boldsymbol{A}\mathbb{X} + \boldsymbol{b} = \{\boldsymbol{A}\boldsymbol{X}_1 + \boldsymbol{b}, ..., \boldsymbol{A}\boldsymbol{X}_n + \boldsymbol{b}\}$. The location and scatter sample statistics then also satisfy

$$\boldsymbol{T}(\boldsymbol{A}\mathbb{X} + \boldsymbol{b}) = \boldsymbol{A}\boldsymbol{T}(\mathbb{X}) + \boldsymbol{b} \quad \text{and} \quad \boldsymbol{S}(\boldsymbol{A}\mathbb{X} + \boldsymbol{b}) = \boldsymbol{A}\boldsymbol{S}(\mathbb{X})\boldsymbol{A}'$$

for all $\boldsymbol{A} \in \mathcal{A}$ and all $\boldsymbol{b} \in \mathbb{R}^d$.

The first examples of location and scatter functionals are the mean vector and the regular covariance matrix:

$$\boldsymbol{T}_1(F_{\boldsymbol{X}}) = Mean(\boldsymbol{X}) \quad \text{and} \quad \boldsymbol{S}_1(F_{\boldsymbol{X}}) = Cov(\boldsymbol{X})$$

Location and scatter functionals can be based on the third and fourth moments, respectively, as well. A location functional based on third moments is

$$\boldsymbol{T}_2(F_{\boldsymbol{X}}) = \frac{1}{d} E\left((\boldsymbol{X} - E(\boldsymbol{X}))' \boldsymbol{S}_1(F_{\boldsymbol{X}})^{-1} (\boldsymbol{X} - E(\boldsymbol{X}))\boldsymbol{X}\right)$$

and a scatter matrix functional based on fourth moments is

$$\boldsymbol{S}_2(F_{\boldsymbol{X}}) = \frac{1}{d+2} E\left((\boldsymbol{X} - E(\boldsymbol{X}))(\boldsymbol{X} - E(\boldsymbol{X}))' \boldsymbol{S}_1(F_{\boldsymbol{X}})^{-1} (\boldsymbol{X} - E(\boldsymbol{X}))(\boldsymbol{X} - E(\boldsymbol{X}))'\right).$$

M-functionals of location and scatter, introduced by Maronna (1976), are commonly used in many statistical analyses. They are defined as solutions of the two equations

$$\boldsymbol{T}(F_{\boldsymbol{X}}) = E(w_1(r))^{-1} E(w_1(r)\boldsymbol{X})$$

and

$$\boldsymbol{S}(F_{\boldsymbol{X}}) = E(w_2(r)(\boldsymbol{X} - \boldsymbol{T}(F_{\boldsymbol{X}}))(\boldsymbol{X} - \boldsymbol{T}(F_{\boldsymbol{X}}))'),$$

where $w_1(r)$ and $w_2(r)$ are nonnegative continuous functions of the Mahalanobis distance $r = ||\boldsymbol{S}(F_{\boldsymbol{X}})^{-1/2}(\boldsymbol{X} - \boldsymbol{T}(F_{\boldsymbol{X}}))||$. (The $|| \cdot ||$ here denotes the l_2 norm of $\cdot$, and $\boldsymbol{S}(F_{\boldsymbol{X}})^{-1/2}$ is the symmetric version (see Section 2).) The mean vector and the regular covariance matrix are M-functionals with $w_1(r) = w_2(r) = 1$. For example Hettmansperger-Randles functionals (Hettmansperger and Randles (2002)) are obtained with choices $w_1(r) = \frac{1}{r}$ and $w_2(r) = \frac{p}{r^2}$.

There are several other location and scatter functionals, even families of them, having different desirable properties (robustness, efficiency, limiting multivariate normality, fast computations etc). See for example Lopuhaä (1989); Maronna et al. (2006); Davies (1987); Kent and Tyler (1996).

A scatter matrix $\mathbf{\Sigma} \in \mathcal{S}$ can be decomposed into two parts by $\mathbf{\Sigma} = \sigma^2 \mathbf{\Sigma}_0$ where $\sigma^2 = \sigma^2(\mathbf{\Sigma})$ is scalar-valued, and $\mathbf{\Sigma}_0 = \sigma^{-2}\mathbf{\Sigma}$ is a shape matrix. The scale function $\sigma^2(\mathbf{\Sigma})$ satisfies $\sigma^2(\boldsymbol{I}_d) = 1$ and $\sigma^2(c\mathbf{\Sigma}) = c\sigma^2(\mathbf{\Sigma})$. In the literature, choices of the scale function $\sigma^2(\mathbf{\Sigma})$ are $\mathbf{\Sigma}_{11}$, $tr(\mathbf{\Sigma})/d$, $d/tr(\mathbf{\Sigma}^{-1})$, and $det(\mathbf{\Sigma})^{1/d}$, for example. See Paindaveine (2008); Frahm (2009) and references therein. If $\boldsymbol{S}(F)$ is a scatter functional or a weak covariance functional, then one can define the corresponding *scale and shape functionals* as $\sigma^2(\boldsymbol{S}(F))$ and $\boldsymbol{S}_0(F) = [1/\sigma^2(\boldsymbol{S}(F))]\boldsymbol{S}(F)$, respectively. Note that the shape functional $\boldsymbol{S}_0(F)$ is naturally also a weak covariance functional.

By definitions, it is required that multivariate location and scatter statistics are affine equivariant. Multivariate testing and estimation procedures in general are often hoped to be affine invariant and affine equivariant, respectively. (However, some of the traditional methods do not fulfill this desired property.) Invariance and maximal invariance (Eaton (1989); Lehmann and Romano (2005)) are defined as follows.

Definition 1.3. *(i) A statistic $Q(\mathbb{X})$ is invariant under a group $\mathcal{G}$ of transformations if*

$$Q(g\mathbb{X}) = Q(\mathbb{X}), \quad \textit{for all } g \in \mathcal{G}.$$

(ii) A statistic $Q(\mathbb{X})$ is maximal invariant under $\mathcal{G}$ if it is invariant under $\mathcal{G}$ and if

$$Q(\mathbb{Y}) = Q(\mathbb{X}) \quad \Rightarrow \quad \mathbb{Y} = g\mathbb{X}, \quad \textit{for some } g \in \mathcal{G}.$$

If $Q_0(\mathbb{X})$ is maximal invariant, then any other invariant statistic $Q(\mathbb{X})$ is a function of $Q_0(\mathbb{X})$. This means that $Q_0(\mathbb{X})$ is the most informative one among the set of invariant statistics and is therefore preferred in further analysis of data. Another extreme case among invariant statistics is a constant variable $Q(\mathbb{X}) = c$ that does not carry any information at all.

In this paper we mainly consider invariance and equivariance under the groups of transformations $\mathcal{G}_0 = \{g : g\mathbb{X} = \boldsymbol{A}\mathbb{X}, \quad \boldsymbol{A} \in \mathcal{A}\}$ and $\mathcal{G}_1 = \{g : g\mathbb{X} = \boldsymbol{A}\mathbb{X} + \boldsymbol{b}, \quad \boldsymbol{A} \in \mathcal{A}, \boldsymbol{b} \in \mathbb{R}^d\}$. Then the procedures based on invariant statistics do not depend on the chosen coordinate system for the observation vectors. It turns out that in the first setting a maximal invariant is obtained by applying a suitable type of matrix-valued transformation $\boldsymbol{G}(\mathbb{X}) \in \mathcal{A}$ to the data $\mathbb{X}$, producing $\boldsymbol{G}(\mathbb{X})\mathbb{X}$ as the desired maximal invariant for the first group of transformations. Then a noninvariant statistic $\Theta(\mathbb{X})$ can be made into an invariant one by taking $\Theta(\boldsymbol{G}(\mathbb{X})\mathbb{X})$. (Choosing $Q(\mathbb{X}) = \boldsymbol{G}(\mathbb{X})\mathbb{X}$ here to be just invariant, instead of maximal invariant, would also naturally make $\Theta(Q(\mathbb{X}))$ invariant.) In the second setting a maximal invariant of the type $\boldsymbol{G}(\mathbb{X})(\mathbb{X} - \boldsymbol{T}(\mathbb{X}))$ can be found. In particular, we will explore when $\boldsymbol{G}(\mathbb{X})$ may be given by $\boldsymbol{S}(\mathbb{X})^{-1/2}$ for some scatter matrix $\boldsymbol{S}(\mathbb{X})$. Note that, since transformations $g\mathbb{X} = \boldsymbol{A}\mathbb{X} + \boldsymbol{b}, \quad \boldsymbol{A} \in \mathcal{A}, \boldsymbol{b} \in \mathbb{R}^d$ are called affine transformations, invariance under $\mathcal{G}_1$ or $\mathcal{G}_0$ is called *affine invariance.*

The paper is organized as follows. Standardization of multivariate data is discussed in Section 2. The definition for a matrix $\mathbf{\Sigma}^{-1/2}$ and different ways to choose it uniquely are examined in detail. Definitions of invariant coordinate system (ICS) functionals and statistics are given in Section 3. Also, invariances up to some groups of transformations are discussed. Section 4 addresses construction of ICS functionals. In particular the construction based on the use of two scatter matrix functionals presented by Tyler et al. (2009), and direct definitions based on the approach presented by Chaudhuri and Sengupta (1993) are examined. Several applications of ICS functionals are discussed in Section 5.

2. Standardization of data

Let $\mathbb{X} = \{\boldsymbol{X}_1, ..., \boldsymbol{X}_n\}$ be a random sample of size n from a d-variate distribution $F_{\boldsymbol{X}}$. Location and scatter functionals are often used to center and standardize the distribution. For the standardization we need the following definition of matrix $\boldsymbol{\Sigma}^{-1/2}$.

Definition 2.1. *Let $\boldsymbol{\Sigma} \in \mathcal{S}$. Then $\boldsymbol{\Sigma}^{-1/2}$ denotes any matrix $\boldsymbol{G}$ which satisfies*

$$\boldsymbol{G}\boldsymbol{\Sigma}\boldsymbol{G}' = \boldsymbol{I}_d.$$

Note that $\boldsymbol{\Sigma}^{-1/2}$ is defined only up to an orthogonal transformation: if $\boldsymbol{G}\boldsymbol{\Sigma}\boldsymbol{G}' = \boldsymbol{I}_d$, then also $(\boldsymbol{V}\boldsymbol{G})\boldsymbol{\Sigma}(\boldsymbol{V}\boldsymbol{G})' = \boldsymbol{I}_d$ for all orthogonal matrices $\boldsymbol{V}$. Fix

$$\boldsymbol{\Sigma} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}',$$

such that $\boldsymbol{U}$ is a fixed unique orthogonal matrix and $\boldsymbol{\Lambda}$ is a diagonal matrix with diagonal elements in a decreasing order of magnitude. The eigenvalue and eigenvector composition $\boldsymbol{\Sigma} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}'$ always exists and it can be chosen uniquely. Later in this section when we use the notation $\boldsymbol{\Sigma} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}'$, we assume that the eigenvalue matrix $\boldsymbol{\Lambda}$ and the orthogonal eigenvector matrix $\boldsymbol{U}$ are chosen uniquely. Now one can choose $\boldsymbol{\Sigma}^{-1/2}$ in a unique way by requiring, for example, that

(a) $\boldsymbol{\Sigma}^{-1/2}$ is unique lower diagonal (e.g. the inverse of the lower diagonal matrix in the Cholesky decomposition of $\boldsymbol{\Sigma}$),
(b) $\boldsymbol{\Sigma}^{-1/2}$ is unique upper diagonal (e.g. formed by permuting the rows of the inverse of the lower diagonal matrix in the Cholesky decomposition of $\boldsymbol{\Sigma}$),
(c) $\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{U}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}'$ (symmetric version),
(d) $\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}'$ (rows are rescaled eigenvectors), or
(e) $\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}'$ where $\boldsymbol{V}$ is the VARIMAX rotation (factor analysis)

In the above developments $\boldsymbol{\Lambda}^{-1/2}$ is the unique diagonal matrix satisfying $\boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^{-1/2} = \boldsymbol{I}_d$. The use of the above mentioned matrices $\boldsymbol{\Sigma}^{-1/2}$, especially the ones defined in points (a) and (c), in standardizing data, were examined for example in Li and Zhang (1998).

The following result is utilized from a well known result for positive definite matrices, see for example Horn and Johnson (1985), and is therefore presented just as a remark (and not as a theorem.) However, the proof for this particular case (real positive definite matrices) is presented to make the result more clear.

Remark 2.1. *In general, for all $\boldsymbol{\Sigma} \in \mathcal{S}$, $\boldsymbol{G}$ can be regarded as $\boldsymbol{\Sigma}^{-1/2}$ if and only if*

$$\boldsymbol{G} \in \{\boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}' \ : \ \boldsymbol{V} \text{ orthogonal and } \boldsymbol{\Sigma} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}'\},$$

and if $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$ can be regarded as $\boldsymbol{\Sigma}^{-1/2}$, then $\boldsymbol{G}_1 = \boldsymbol{V}\boldsymbol{G}_2$ for some orthogonal $\boldsymbol{V}$.

PROOF. Assume that $\boldsymbol{\Sigma} \in \mathcal{S}$ and let $\boldsymbol{\Sigma} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}'$. Define $\boldsymbol{G}_0 = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}'$. Now clearly $\boldsymbol{V}\boldsymbol{G}_0\boldsymbol{\Sigma}\boldsymbol{G}_0'\boldsymbol{V}' = \boldsymbol{I}_d$ for any orthogonal matrix $\boldsymbol{V}$. Assume that $\boldsymbol{G}\boldsymbol{\Sigma}\boldsymbol{G}' = \boldsymbol{I}_d$. Now $\boldsymbol{\Sigma} = \boldsymbol{G}^{-1}(\boldsymbol{G}')^{-1} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}'$ and it follows that $\boldsymbol{G}^{-1} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}'\boldsymbol{G}'$. Thus

$$\boldsymbol{G} = (\boldsymbol{G}')^{-1}\boldsymbol{U}\boldsymbol{\Lambda}^{-1}\boldsymbol{U}' = (\boldsymbol{G}')^{-1}\boldsymbol{U}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}'.$$

Clearly

$$(\boldsymbol{G}')^{-1}\boldsymbol{U}\boldsymbol{\Lambda}^{-1/2}((\boldsymbol{G}')^{-1}\boldsymbol{U}\boldsymbol{\Lambda}^{-1/2})' = I_d$$

i.e. the matrix $(\boldsymbol{G}')^{-1}\boldsymbol{U}\boldsymbol{\Lambda}^{-1/2}$ is orthogonal. Thus $\boldsymbol{G} = \boldsymbol{\Sigma}^{-1/2}$ if and only if

$$\boldsymbol{G} \in \{\boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}' \ : \ \boldsymbol{V} \text{ orthogonal and } \boldsymbol{\Sigma} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}'\}.$$

Assume then that $\boldsymbol{G}_1\boldsymbol{\Sigma}\boldsymbol{G}_1' = \boldsymbol{I}_d$ and $\boldsymbol{G}_2\boldsymbol{\Sigma}\boldsymbol{G}_2' = \boldsymbol{I}_d$. Now $\boldsymbol{G}_1 = \boldsymbol{V}_1\boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}'$ and $\boldsymbol{G}_2 = \boldsymbol{V}_2\boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}'$ for some orthogonal matrices $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$. Thus $\boldsymbol{G}_1 = \boldsymbol{V}_1\boldsymbol{V}_2^{-1}\boldsymbol{G}_2$. Since $\boldsymbol{V}_1\boldsymbol{V}_2^{-1}$ is clearly orthogonal, $\boldsymbol{G}_1 = \boldsymbol{V}\boldsymbol{G}_2$ for some orthogonal $\boldsymbol{V}$.

For practical calculations one must of course have a rule for the choice of $\boldsymbol{V}$ in $\boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}'$.

For a scatter or a weak covariance functional $\boldsymbol{S}(F) \in \mathcal{S}$, define the corresponding unique eigenvector and eigenvalue functionals implicitly by

$$\boldsymbol{S}(F) = \boldsymbol{U}(F)\boldsymbol{\Lambda}(F)\boldsymbol{U}(F)'.$$

Definition 2.2. *Let the functional $\boldsymbol{S}(F) \in \mathcal{S}$. Then $\boldsymbol{S}^{-1/2}(F)$ denotes any functional $\boldsymbol{G}(F)$ which satisfies*

$$\boldsymbol{G}(F)\boldsymbol{S}(F)\boldsymbol{G}(F)' = \boldsymbol{I}_d.$$

The next Lemma follows directly from Definition 2.2 and Remark 2.1.

Lemma 2.1. *For any functional $\boldsymbol{S}(F) \in \mathcal{S}$*

$$\boldsymbol{S}^{-1/2}(F) = \boldsymbol{V}(F)\boldsymbol{\Lambda}(F)^{-1/2}\boldsymbol{U}(F)',$$

where $\boldsymbol{V}(F)$ is an orthogonal matrix functional.

To fix the functional $\boldsymbol{S}^{-1/2}(F)$ uniquely we thus have to fix the functional $\boldsymbol{V}(F)$. Possible choices of $\boldsymbol{V}(F)$ are $\boldsymbol{U}(F)$ or $\boldsymbol{I}_d$ or a matrix $\boldsymbol{V}(F)$ that makes $\boldsymbol{S}^{-1/2}(F)$ unique upper or lower triangular, for example. However, none of the choices discussed so far guarantees the invariance of $\boldsymbol{S}^{-1/2}(\mathbb{X})\mathbb{X}$.

Theorem 2.1. *Let $\boldsymbol{A} \in \mathcal{A}$ and $\boldsymbol{b} \in \mathbb{R}^d$.*

- *If $\boldsymbol{S}(F)$ is a scatter functional then $\boldsymbol{S}^{-1/2}(\boldsymbol{A}\mathbb{X} + \boldsymbol{b}) = \boldsymbol{U}\boldsymbol{S}^{-1/2}(\mathbb{X})\boldsymbol{A}^{-1}$ for some orthogonal $\boldsymbol{U} = \boldsymbol{U}(\mathbb{X}, \boldsymbol{A})$.*
- *If $\boldsymbol{S}(F)$ is a weak covariance functional then $\boldsymbol{S}^{-1/2}(\boldsymbol{A}\mathbb{X} + \boldsymbol{b}) = k\boldsymbol{U}\boldsymbol{S}^{-1/2}(\mathbb{X})\boldsymbol{A}^{-1}$ for some orthogonal $\boldsymbol{U} = \boldsymbol{U}(\mathbb{X}, \boldsymbol{A})$ and some constant $k = k(\mathbb{X}, \boldsymbol{A})$.*

PROOF. Let $\boldsymbol{S}(F)$ be a scatter functional. Clearly

$$\boldsymbol{S}^{-1/2}(\mathbb{X})\boldsymbol{A}^{-1}(\boldsymbol{S}(\boldsymbol{A}\mathbb{X}+\boldsymbol{b}))(\boldsymbol{S}^{-1/2}(\mathbb{X})\boldsymbol{A}^{-1})'$$

$$= \boldsymbol{S}^{-1/2}(\mathbb{X})\boldsymbol{A}^{-1}\boldsymbol{A}\boldsymbol{S}(\mathbb{X})\boldsymbol{A}'(\boldsymbol{A}')^{-1}(\boldsymbol{S}^{-1/2}(\mathbb{X}))' = I_d.$$

Thus it follows from Remark 2.1, that $\boldsymbol{S}^{-1/2}(\boldsymbol{A}\mathbb{X}+\boldsymbol{b}) = \boldsymbol{U}\boldsymbol{S}^{-1/2}(\mathbb{X})\boldsymbol{A}^{-1}$ for some orthogonal $\boldsymbol{U} = \boldsymbol{U}(\mathbb{X}, \boldsymbol{A})$. The second point of the Theorem may be proven similarly.

Corollary 2.1. *Let $\boldsymbol{S}(F)$ be a scatter or a weak covariance functional, and $\boldsymbol{T}(F)$ a location functional. Then*

- $\boldsymbol{S}^{-1/2}(\mathbb{X})\mathbb{X}$ *is not necessarily invariant under group of transformations* $\{g : g\mathbb{X} = \boldsymbol{A}\mathbb{X}, \boldsymbol{A} \in \mathcal{A}\}$*, and*
- $\boldsymbol{S}^{-1/2}(\mathbb{X})(\mathbb{X} - \boldsymbol{T}(\mathbb{X}))$ *is not necessarily invariant under group of transformations* $\{g : g\mathbb{X} = \boldsymbol{A}\mathbb{X} + \boldsymbol{b}, \boldsymbol{A} \in \mathcal{A},\ \boldsymbol{b} \in \mathbb{R}^d\}$*.*

PROOF. Let $\boldsymbol{S}(F)$ be a scatter functional. Let $\boldsymbol{S}(\mathbb{X}) = \boldsymbol{U}(\mathbb{X})\boldsymbol{\Lambda}(\mathbb{X})\boldsymbol{U}(\mathbb{X})'$, where the eigenvalue matrix $\boldsymbol{\Lambda}(\mathbb{X})$ has distinct diagonal elements and where $\boldsymbol{U}(\mathbb{X})$ is the unique orthogonal eigenvector matrix. Choose $\boldsymbol{S}^{-1/2}(\mathbb{X}) = \boldsymbol{U}(\mathbb{X})\boldsymbol{\Lambda}(\mathbb{X})^{-1/2}\boldsymbol{U}(\mathbb{X})'$ (unique symmetric version). Assume that $\boldsymbol{V}$ is some orthogonal matrix. Now the unique symmetric version of $\boldsymbol{S}^{-1/2}(\boldsymbol{V}\mathbb{X})$ is $\boldsymbol{V}\boldsymbol{U}(\mathbb{X})\boldsymbol{\Lambda}(\mathbb{X})^{-1/2}\boldsymbol{U}(\mathbb{X})'\boldsymbol{V}'$. It now follows that

$$\boldsymbol{S}^{-1/2}(\mathbb{X})\mathbb{X} = \boldsymbol{U}(\mathbb{X})\boldsymbol{\Lambda}(\mathbb{X})^{-1/2}\boldsymbol{U}(\mathbb{X})'\mathbb{X}$$

and

$$\boldsymbol{S}^{-1/2}(\boldsymbol{V}\mathbb{X})\boldsymbol{V}\mathbb{X}$$

$$= \boldsymbol{V}\boldsymbol{U}(\mathbb{X})\boldsymbol{\Lambda}(\mathbb{X})^{-1/2}\boldsymbol{U}(\mathbb{X})'\boldsymbol{V}'\boldsymbol{V}\mathbb{X} = \boldsymbol{V}\boldsymbol{U}(\mathbb{X})\boldsymbol{\Lambda}(\mathbb{X})^{-1/2}\boldsymbol{U}(\mathbb{X})'\mathbb{X}.$$

Thus

$$\boldsymbol{S}^{-1/2}(\boldsymbol{V}\mathbb{X})\boldsymbol{V}\mathbb{X} = \boldsymbol{V}\boldsymbol{S}^{-1/2}(\mathbb{X})\mathbb{X}.$$

The sets $\boldsymbol{S}^{-1/2}(\mathbb{X})\mathbb{X}$ and $\boldsymbol{V}\boldsymbol{S}^{-1/2}(\mathbb{X})\mathbb{X}$ in general are not the same sets unless $\boldsymbol{V} = \boldsymbol{I}_d$ or $\mathbb{X}$ is very specific. Similar examination can be done for all unique choices $\boldsymbol{S}^{-1/2}(\mathbb{X})$ that were presented in page 4.

Note that, for example, the transformation $\boldsymbol{U}(\mathbb{X})'\mathbb{X}$ used in principal component analysis is not affine invariant. Thus the results of principal component analysis may vary depending on the used coordinate system. For other examples and discussion, see Serfling (2010).

Remark 2.2. *We will show later that, with a suitable choice of $\boldsymbol{V}(F)$, the functional $\boldsymbol{S}^{-1/2}(F)$ can be made affine equivariant.*

Remark 2.3. *Serfling (2010) called a matrix valued functional $\boldsymbol{M}(F)$ a transformation-retransformation functional if $\boldsymbol{S}(F) = (\boldsymbol{M}(F)'\boldsymbol{M}(F))^{-1}$ is a weak covariance functional. Clearly then $\boldsymbol{M}(F) = \boldsymbol{S}^{-1/2}(F)$. If $\boldsymbol{M}(F)$ is a transformation-retransformation functional then so is $\boldsymbol{V}(F)\boldsymbol{M}(F)$ where $\boldsymbol{V}(F)$ is any orthogonal matrix valued functional.*

3. Invariant coordinate system (ICS)

Ideally (in order to obtain maximal invariance of $\boldsymbol{G}(\mathbb{X})(\mathbb{X}-\boldsymbol{T}(\mathbb{X}))$ under affine transformations, see Theorem 3.1), we wish to find a transformation matrix functional $\boldsymbol{G}(F)$ or statistic $\boldsymbol{G}(\mathbb{X})$ that satisfies the conditions stated in the following definition.

Definition 3.1. *(i) An invariant coordinate system (ICS) functional (under affine transformations) is a matrix-valued functional* $\boldsymbol{G}(F)\in\mathcal{A}$ *satisfying*

$$\boldsymbol{G}(F_{\boldsymbol{AX}+\boldsymbol{b}})=\boldsymbol{G}(F_{\boldsymbol{X}})\boldsymbol{A}^{-1},\quad \textit{for all } \boldsymbol{A}\in\mathcal{A} \textit{ and } \boldsymbol{b}\in\mathbb{R}^d. \tag{1}$$

(ii) An invariant coordinate system (ICS) statistic (under affine transformations) is a $d\times d$ *matrix-valued sample statistic* $\boldsymbol{G}(\mathbb{X})$ *satisfying*

$$\boldsymbol{G}(\boldsymbol{A}\mathbb{X}+\boldsymbol{b})=\boldsymbol{G}(\mathbb{X})\boldsymbol{A}^{-1},\quad \textit{for all } \boldsymbol{A}\in\mathcal{A} \textit{ and } \boldsymbol{b}\in\mathbb{R}^d. \tag{2}$$

For the ICS statistic $\boldsymbol{G}(\mathbb{X})$ we then have the following result.

Theorem 3.1. *(i) If* $\boldsymbol{G}(\mathbb{X})\in\mathcal{A}$ *satisfies* $\boldsymbol{G}(\boldsymbol{A}\mathbb{X})=\boldsymbol{G}(\mathbb{X})\boldsymbol{A}^{-1}$ *for all* $\boldsymbol{A}\in\mathcal{A}$*, then* $\boldsymbol{G}(\mathbb{X})\mathbb{X}$ *is maximal invariant under the transformations in* $\{g: g\mathbb{X}=\boldsymbol{A}\mathbb{X}\}$.
(ii) If $\boldsymbol{G}(\mathbb{X})\in\mathcal{A}$ *satisfies* $\boldsymbol{G}(\boldsymbol{A}\mathbb{X}+\boldsymbol{b})=\boldsymbol{G}(\mathbb{X})\boldsymbol{A}^{-1}$ *for all* $\boldsymbol{A}\in\mathcal{A}$ *and* $\boldsymbol{b}\in\mathbb{R}^d$ *and* $\boldsymbol{T}(\mathbb{X})$ *is a location statistic, then* $\boldsymbol{G}(\mathbb{X})(\mathbb{X}-\boldsymbol{T}(\mathbb{X}))$ *is maximal invariant under the transformations in* $\{g: g\mathbb{X}=\boldsymbol{A}\mathbb{X}+\boldsymbol{b}\}$.

Proof. (i) Assume that $\boldsymbol{G}(\boldsymbol{A}\mathbb{X})=\boldsymbol{G}(\mathbb{X})\boldsymbol{A}^{-1}$ for all $\boldsymbol{A}\in\mathcal{A}$. Then $\boldsymbol{G}(\boldsymbol{A}\mathbb{X})\boldsymbol{A}\mathbb{X}=\boldsymbol{G}(\mathbb{X})\boldsymbol{A}^{-1}\boldsymbol{A}\mathbb{X}=\boldsymbol{G}(\mathbb{X})\mathbb{X}$ for all $\boldsymbol{A}$. Thus $\boldsymbol{G}(\mathbb{X})\mathbb{X}$ is invariant under the transformations in $\{g: g\mathbb{X}=\boldsymbol{A}\mathbb{X}\}$. If $\boldsymbol{G}(\mathbb{X})\mathbb{X}=\boldsymbol{G}(\mathbb{Y})\mathbb{Y}$ then $\mathbb{Y}=\boldsymbol{A}\mathbb{X}$ where $\boldsymbol{A}=\boldsymbol{G}(\mathbb{Y})^{-1}\boldsymbol{G}(\mathbb{X})$. Thus $\boldsymbol{G}(\mathbb{X})\mathbb{X}$ is a maximal invariant statistic. (ii) Assume that $\boldsymbol{T}(\mathbb{X})$ is a location statistic and that $\boldsymbol{G}(\boldsymbol{A}\mathbb{X}+\boldsymbol{b})=\boldsymbol{G}(\mathbb{X})\boldsymbol{A}^{-1}$ for all $\boldsymbol{A}\in\mathcal{A}$ and $\boldsymbol{b}\in\mathbb{R}^d$. Then $\boldsymbol{G}(\boldsymbol{A}\mathbb{X}+\boldsymbol{b})(\boldsymbol{A}\mathbb{X}+\boldsymbol{b}-\boldsymbol{T}(\boldsymbol{A}\mathbb{X}+\boldsymbol{b}))=\boldsymbol{G}(\mathbb{X})\boldsymbol{A}^{-1}(\boldsymbol{A}\mathbb{X}+\boldsymbol{b}-\boldsymbol{A}\boldsymbol{T}(\mathbb{X})-\boldsymbol{b})=\boldsymbol{G}(\mathbb{X})(\mathbb{X}-\boldsymbol{T}(\mathbb{X}))$ for all $\boldsymbol{A}$ and $\boldsymbol{b}$. Thus $\boldsymbol{G}(\mathbb{X})(\mathbb{X}-\boldsymbol{T}(\mathbb{X}))$ is invariant under the transformations in $\{g: g\mathbb{X}=\boldsymbol{A}\mathbb{X}+\boldsymbol{b}\}$. If $G(\mathbb{X})(\mathbb{X}-\boldsymbol{T}(\mathbb{X}))=\boldsymbol{G}(\mathbb{Y})(\mathbb{Y}-\boldsymbol{T}(\mathbb{Y}))$ then $\mathbb{Y}=\boldsymbol{A}\mathbb{X}+\boldsymbol{b}$ where $\boldsymbol{A}=\boldsymbol{G}(\mathbb{Y})^{-1}\boldsymbol{G}(\mathbb{X})$ and $\boldsymbol{b}=\boldsymbol{T}(\mathbb{Y})-G(\mathbb{Y})^{-1}\boldsymbol{G}(\mathbb{X})\boldsymbol{T}(\mathbb{X})$. Thus $\boldsymbol{G}(\mathbb{X})(\mathbb{X}-\boldsymbol{T}(\mathbb{X}))$ is maximal invariant.

In practical problems full invariance as in Definition 3.1 is not always needed. Weaker concepts of invariant coordinate system are obtained if we only require invariance up to some groups of transformations. For example for the tests based on marginal signs and ranks it is sufficient to have ICS functionals up to permutation, and (heterogenous) sign changes and scales. Let $\mathcal{C}$ be some subgroup of nonsingular $d\times d$ matrices. Then we define:

Definition 3.2. *(i) A* $d\times d$ *matrix-valued functional* $\boldsymbol{G}(F)$ *is an invariant coordinate system (ICS) functional up to a group of transformations* $\mathcal{C}$ *if, for any* $\boldsymbol{A}\in\mathcal{A}$ *and any* $\boldsymbol{b}\in\mathbb{R}^d$*,*

$$\boldsymbol{G}(F_{\boldsymbol{AX}+\boldsymbol{b}})=\boldsymbol{C}\boldsymbol{G}(F_{\boldsymbol{X}})\boldsymbol{A}^{-1}. \tag{3}$$

for some $\boldsymbol{C}\in\mathcal{C}$.
(ii) A $d\times d$ *matrix-valued sample statistic* $\boldsymbol{G}(\mathbb{X})$ *is an invariant coordinate system statistic up to a group of transformations* $\mathcal{C}$ *if, for any* $\boldsymbol{A}\in\mathcal{A}$ *and any* $\boldsymbol{b}\in\mathbb{R}^d$*,*

$$\boldsymbol{G}(\boldsymbol{A}\mathbb{X}+\boldsymbol{b})=\boldsymbol{C}\boldsymbol{G}(\mathbb{X})\boldsymbol{A}^{-1}. \tag{4}$$

for some $\boldsymbol{C}\in\mathcal{C}$.

Theorem 3.2. *Let $Q(\mathbb{X})$ be invariant under transformations in $\mathcal{C}$, that is, $Q(\boldsymbol{C}\mathbb{X}) = Q(\mathbb{X})$, for all $\boldsymbol{C} \in \mathcal{C}$. If $\boldsymbol{G}(\mathbb{X})$ is an ICS statistic up to $\mathcal{C}$, then $Q(\boldsymbol{G}(\mathbb{X})\mathbb{X})$ is affine invariant, that is, $Q(\boldsymbol{G}(\boldsymbol{A}\mathbb{X})\boldsymbol{A}\mathbb{X}) = Q(\boldsymbol{G}(\mathbb{X})\mathbb{X})$ for all $\boldsymbol{A} \in \mathcal{A}$.*

Proof. Assume that $Q(\mathbb{X})$ is invariant under transformations in $\mathcal{C}$ and that $\boldsymbol{G}(\mathbb{X})$ is an ICS statistic up to $\mathcal{C}$. Let $\boldsymbol{A} \in \mathcal{A}$. Now, since $\boldsymbol{G}(\mathbb{X})$ is an ICS statistic up to $\mathcal{C}$,

$$Q(\boldsymbol{G}(\boldsymbol{A}\mathbb{X})\boldsymbol{A}\mathbb{X}) = Q(\boldsymbol{C}\boldsymbol{G}(\mathbb{X})\boldsymbol{A}^{-1}\boldsymbol{A}\mathbb{X})$$

for some $\boldsymbol{C} \in \mathcal{C}$ and since $Q(\mathbb{X})$ is invariant under transformations in $\mathcal{C}$,

$$Q(\boldsymbol{C}\boldsymbol{G}(\mathbb{X})\mathbb{X}) = Q(\boldsymbol{G}(\mathbb{X})\mathbb{X}).$$

What are then the interesting groups of transformations in practice? In the literature of multivariate nonparametric statistics one often has invariance under the above mentioned $\mathcal{G}_0 = \{g \ : \ g\mathbb{X} = \boldsymbol{A}, \ \boldsymbol{A} \in \mathcal{A}\}$ (nonsingular matrix multiplication) and $\mathcal{G}_1 = \{g \ : \ g\mathbb{X} = \boldsymbol{A}\mathbb{X} + \boldsymbol{b}, \quad \boldsymbol{A} \in \mathcal{A}, \boldsymbol{b} \in \mathbb{R}^d\}$ (affine transformation), but also under the following groups of transformations.

(a) $\mathcal{D}_0 = \{c\boldsymbol{I}_d \ : \ c > 0\}$ (homogeneous rescaling),
(b) $\mathcal{D} = \{\text{diag}(c_1, ..., c_d) \ : \ c_i > 0, \ i = 1, ..., d\}$ (heterogeneous rescaling),
(c) $\mathcal{J} = \{\text{diag}(c_1, ..., c_d) \ : \ c_i = \pm 1, \ i = 1, ..., d\}$ (heterogeneous sign changing),
(d) $\mathcal{P} = \{\boldsymbol{P} \ : \ \boldsymbol{P}$ is a permutation matrix $\}$ (permutation), and
(e) $\mathcal{U} = \{\boldsymbol{U} \ : \ \boldsymbol{U}$ is orthogonal$\}$ (rotation and reflection).

Serfling (2010) called $\boldsymbol{G}$ a *strong ICS functional* if it is an ICS functional up to $\mathcal{D}_0$. Using these definitions we can now also say that, if $\boldsymbol{S}$ is a scatter matrix functional, then $\boldsymbol{S}^{-1/2}$ is an ICS functional up to $\mathcal{U}$. Similarly, if $\boldsymbol{S}$ is a weak covariance functional, then $\boldsymbol{S}^{-1/2}$ is an ICS functional up to $\mathcal{U}_c = \{c\boldsymbol{U} \ : \ c > 0$ and $\boldsymbol{U} \in \mathcal{U}\}$. However, for many applications, these $\mathcal{U}$ and $\mathcal{U}_c$ are too broad. In the independent component analysis ICS functionals can be used to recover independent components. See Section 5.5. For that application, for example, an ICS functional up to $\mathcal{U}$ or $\mathcal{U}_c$ does not work.

ICS functionals are often used to preprocess the data to obtain affine invariant or equivariant statistical procedures. Theorem 3.2 then shows what is needed for full invariance. For the tests based on spatial signs and ranks for example we need ICS functionals only up to transformations in $\mathcal{U}$ and as already mentioned, for the tests based on marginal signs and ranks it is sufficient to have ICS functionals up to a group of transformations

$$\mathcal{C}_{PDJ} = \{\boldsymbol{PDJ} \ : \ \boldsymbol{P} \in \mathcal{P}, \ \boldsymbol{D} \in \mathcal{D}, \text{ and } \boldsymbol{J} \in \mathcal{J} \ \}.$$

These will be discussed in more detail in Section 5.

4. Construction of ICS functionals

Construction of ICS functionals is addressed in this section. An approach based on the use of two scatter matrix functionals presented by Tyler et al. (2009) is examined in Section 4.1 and a direct approach based on the construction presented by Chaudhuri and Sengupta

(1993) is examined in Section 4.2. Both approaches are nonparametric in the sense that parametric model assumptions are not needed. The approach presented in Section 4.1 is based on constructing functionals that can easily be applied to empirical distributions. Using functionals enables to derive asymptotical results straightforwardly. The approach presented in Section 4.2 is a very practical data driven approach.

4.1. The use of two scatter functionals

In this subsection we first construct ICS functionals; the ICS statistics are then obtained if these functionals are applied to empirical distributions. ICS functionals can be based on any two different scatter or weak covariance functionals: Let $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ be two weak covariance functionals, and consider the set of distributions

$$\mathcal{F} = \{F \ : \ \boldsymbol{S}_1^{-1}(F)\boldsymbol{S}_2(F) \text{ has distinct eigenvalues}\}.$$

Set $\mathcal{F}$ is clearly a nonempty set. If for example $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are the regular covariance matrix and the scatter matrix based on fourth moments, and if $\boldsymbol{Z}$ has mutually independent component with distinct fourth moments and $\boldsymbol{X} = \boldsymbol{A}\boldsymbol{Z} + \boldsymbol{b}$, then $F_{\boldsymbol{X}} \in \mathcal{F}$. The assumption that $\boldsymbol{S}_1^{-1}(F)\boldsymbol{S}_2(F)$ has distinct eigenvalues, is further needed to ensure that the eigenvectors of $\boldsymbol{S}_1^{-1}(F)\boldsymbol{S}_2(F)$ are unique up to the signs, scales, and order. Some discussion about this restriction can be found from page 10. In this model of distributions $\mathcal{F}$ one can define ICS functionals in the following ways.

(a) Find a transformation matrix functional $\boldsymbol{G}$ and diagonal matrix valued functional $\boldsymbol{L}$ as a solution of eigenvector and eigenvalue problems

$$\boldsymbol{S}_1^{-1}\boldsymbol{S}_2\boldsymbol{G}' = \boldsymbol{G}'\boldsymbol{L}.$$

As the lengths, signs, and order of the eigenvectors are not fixed, any solution $\boldsymbol{G}$ is an ICS functional in $\mathcal{F}$ up to permutation, heterogenous rescaling, and heterogenous sign changes, i.e. $\boldsymbol{G}$ is an ICS functional up to a group of transformations $\mathcal{C}_{PDJ}$ (page 8). See Tyler et al. (2009).

(b) Find a transformation matrix functional $\boldsymbol{G}$ and diagonal matrix valued functional $\boldsymbol{L}$ which solve the above eigenvector and eigenvalue problem and satisfy

$$\boldsymbol{G}\boldsymbol{S}_1\boldsymbol{G}' = \boldsymbol{I}_d \ \text{ and } \ \boldsymbol{G}\boldsymbol{S}_2\boldsymbol{G}' = \boldsymbol{L}$$

where the eigenvalues in $\boldsymbol{L}$ are now in a decreasing order. Note that the assumptions $\boldsymbol{G}\boldsymbol{S}_1\boldsymbol{G}' = \boldsymbol{I}_d$ and $\boldsymbol{G}\boldsymbol{S}_2\boldsymbol{G}' = \boldsymbol{L}$ do not further restrict the underlying distribution. These are just used to choose $\boldsymbol{G}$ to be a certain version of $\boldsymbol{S}_1^{-1/2}$, to be specific, the assumptions $\boldsymbol{G}\boldsymbol{S}_1\boldsymbol{G}' = \boldsymbol{I}_d$ and $\boldsymbol{G}\boldsymbol{S}_2\boldsymbol{G}' = \boldsymbol{L}$ here are used to to fix the scales and order of the eigenvectors (column vectors of $\boldsymbol{G}'$). With these assumptions any solution $\boldsymbol{G}$ is an ICS functional up to a group of transformations

$$\{\boldsymbol{D}_0\boldsymbol{J} \ : \ \boldsymbol{D}_0 \in \mathcal{D}_0, \quad \boldsymbol{J} \in \mathcal{J}\}$$

Note that, if $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are scatter functionals, then $\boldsymbol{G}$ is an ICS functional up to sign changes ($\mathcal{J}$, page 8.) only.

(c) Let $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ be any two different location functionals. Find a transformation matrix functional $\boldsymbol{G}$ and diagonal matrix valued functional $\boldsymbol{L}$ which solve the above eigenvector and eigenvalue problem and satisfy

$$\boldsymbol{G}\boldsymbol{S}_1\boldsymbol{G}' = \boldsymbol{I}_d, \quad \boldsymbol{G}\boldsymbol{S}_2\boldsymbol{G}' = \boldsymbol{L}, \quad \text{and} \quad \boldsymbol{G}(\boldsymbol{T}_1 - \boldsymbol{T}_2) \geq 0$$

where the eigenvalues in $\boldsymbol{L}$ are in a decreasing order. The condition $\boldsymbol{G}(\boldsymbol{T}_1 - \boldsymbol{T}_2) \geq 0$ (where $\geq$ is meant componentwise) is used to fix the signs of the eigenvectors (column vectors of $\boldsymbol{G}'$) and it does not bring any further restrictions on the underlying distribution. If $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are scatter functionals and if

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

then the functional $\boldsymbol{G}$ is an ICS functional in $\mathcal{F}_t$. See Ilmonen et al. (2010) and Nordhausen et al. (2010).

The assumption that $\boldsymbol{S}_1^{-1}(F)\boldsymbol{S}_2(F)$ has distinct eigenvalues is needed to ensure that G can be chosen to be a unique ICS functional (or an ICS functional up to above mentioned groups of transformations). If one or some of the eigenvalues of $\boldsymbol{S}_1^{-1}(F)\boldsymbol{S}_2(F)$ are multiple, then the eigenvectors corresponding to those eigenvalues are not unique and the matrix $\boldsymbol{G}$ is not uniquely defined. How restrictive is the assumption? If $\boldsymbol{X} = \boldsymbol{A}\boldsymbol{Z} + \boldsymbol{b}$ for some $\boldsymbol{A} \in \mathcal{A}$ and $\boldsymbol{b} \in \mathbb{R}^d$ where $\boldsymbol{J}\boldsymbol{P}\boldsymbol{Z} \sim \boldsymbol{Z}$ for all $\boldsymbol{J} \in \mathcal{J}$ and $\boldsymbol{P} \in \mathcal{P}$ then $\boldsymbol{S}_1(F_{\boldsymbol{X}})$ and $\boldsymbol{S}_2(F_{\boldsymbol{X}})$ are proportional and $F_{\boldsymbol{X}} \notin \mathcal{F}$. This means that the ICS functional based on two scatter matrices is not uniquely defined for example for the distributions in the elliptic model, or for the distributions where the components of $\boldsymbol{Z}$ are i.i.d. To obtain pure ICS functional, a further assumption $\boldsymbol{G}(F)(\boldsymbol{T}_1(F) - \boldsymbol{T}_2(F)) > 0$ is also needed. This assumption rules out all symmetric distributions, but in most application ICS up to sign changes is enough and the assumption can be dropped.

Remark 4.1. *Note that, if $\mathbb{X}$ is a random sample from a continuous d-variate distribution, then $F_n \in \mathcal{F}_t$ with probability one and $\boldsymbol{G}(\mathbb{X}) = \boldsymbol{G}(F_n)$ is an ICS statistic.*

Sample statistics $\boldsymbol{G}(\mathbb{X})$ and $\boldsymbol{L}(\mathbb{X})$ as defined in point c above are thus affine equivariant and invariant in the sense that

$$\boldsymbol{G}(\boldsymbol{A}\mathbb{X} + \boldsymbol{b}) = \boldsymbol{G}(\mathbb{X})\boldsymbol{A}^{-1} \quad \text{and} \quad \boldsymbol{L}(\boldsymbol{A}\mathbb{X} + \boldsymbol{b}) = \boldsymbol{L}(\mathbb{X})$$

for all $\boldsymbol{A} \in \mathcal{A}$ and $\boldsymbol{b} \in \mathbb{R}^d$. For the asymptotics, it is therefore not a restriction to assume that $\mathbb{X}$ is a random sample from a distribution F with $\boldsymbol{S}_1(F) = \boldsymbol{I}_d$ and $\boldsymbol{S}_2(F) = \boldsymbol{\Lambda}$ where the diagonal elements of $\boldsymbol{\Lambda}$ are $\lambda_1 \geq ... \geq \lambda_d > 0$. Ilmonen et al. (2010) then proved that if all the diagonal elements of $\boldsymbol{\Lambda}$ are distinct and if

$$\sqrt{n}(\boldsymbol{S}_1(\mathbb{X}) - \boldsymbol{I}_d) = O_p(1) \quad \text{and} \quad \sqrt{n}(\boldsymbol{S}_2(\mathbb{X}) - \boldsymbol{\Lambda}) = O_p(1)$$

then

$$\begin{aligned}
\sqrt{n}(\boldsymbol{G}(\mathbb{X})_{ii} - 1) &= -\frac{1}{2}\sqrt{n}(\boldsymbol{S}_1(\mathbb{X})_{ii} - 1) + o_p(1), \\
(\lambda_i - \lambda_j)\sqrt{n}\boldsymbol{G}(\mathbb{X})_{ij} &= \sqrt{n}\boldsymbol{S}_2(\mathbb{X})_{ij} - \lambda_i\sqrt{n}\boldsymbol{S}_1(\mathbb{X})_{ij} + o_p(1), \ i \neq j, \quad \text{and} \\
\sqrt{n}(\boldsymbol{L}(\mathbb{X})_{ii} - \lambda_i) &= \sqrt{n}(\boldsymbol{S}_2(\mathbb{X})_{ii} - \lambda_i) - \lambda_i\sqrt{n}(\boldsymbol{S}_1(\mathbb{X})_{ii} - 1) + o_p(1).
\end{aligned}$$

With a tiny modification in the proof of the above results in Ilmonen et al. (2010) one can show that the three equations above are in fact true if λ_i is distinct from all the other eigenvalues λ_j, $j \neq i$. The limiting joint distributions of the sample eigenvectors and sample eigenvalues for a subset with distinct population eigenvalues can then be derived from the limiting distributions of $\boldsymbol{S}_1(\mathbb{X})$ and $\boldsymbol{S}_2(\mathbb{X})$.

4.2. *Direct definition of $\boldsymbol{G}(\mathbb{X})$*

4.2.1. *The Chaudhuri and Sengupta (1993) example*

Evidently, the first example of ICS functional in the literature was introduced by Chaudhuri and Sengupta (1993) in the context of testing $H : \boldsymbol{\theta} = \boldsymbol{0}$ versus $H : \boldsymbol{\theta} \neq \boldsymbol{0}$ in the location model $F_{\boldsymbol{X}} = F(\boldsymbol{x} - \boldsymbol{\theta})$. Since $\boldsymbol{A\theta} = \boldsymbol{0}$ if and only if $\boldsymbol{\theta} = \boldsymbol{0}$, Chaudhuri and Sengupta (1993) suggest using a test function Q satisfying $Q(\boldsymbol{A}\mathbb{X}) = Q(\mathbb{X})$ for all nonsingular $\boldsymbol{A} \in \mathcal{A}$, 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 $\{g : g\mathbb{X} = \boldsymbol{A}\mathbb{X},\ \boldsymbol{A} \in \mathcal{A}\}$.

Let $\mathbb{X}$ be a random sample of size n from an absolutely continuous distribution. For each fixed choice of d distinct indices $\mathbb{J} = \{i_1, \ldots, i_d\}$ from $\{1, \ldots, n\}$, the matrix $(\boldsymbol{X}_{i_1}, \ldots, \boldsymbol{X}_{i_d})$ is then invertible with probability one, and, for all possible choices of $\mathbb{J}$, its inverse

$$\boldsymbol{G}_{\mathbb{J}}(\mathbb{X}) = (\boldsymbol{X}_{i_1}, \ldots, \boldsymbol{X}_{i_d})^{-1} \tag{5}$$

is an ICS statistic in the sense that $\boldsymbol{G}_{\mathbb{J}}(\boldsymbol{A}\mathbb{X}) = \boldsymbol{G}_{\mathbb{J}}(\mathbb{X})\boldsymbol{A}^{-1}$. Chaudhuri and Sengupta (1993) then show that the transformed observations (i.e, “data-driven coordinates”)

$$\boldsymbol{G}_{\mathbb{J}}(\mathbb{X})\mathbb{X}$$

form a maximal invariant statistic with respect to the nonsingular transformations $\{g : g\mathbb{X} = \boldsymbol{A}\mathbb{X},\ \boldsymbol{A} \in \mathcal{A}\}$. The multivariate signs of transformed observations $\boldsymbol{G}_{\mathbb{J}}(\mathbb{X})\mathbb{X}$ are then used as basic building blocks for various affine invariant versions of multivariate sign tests.

Let $\mathcal{C}(d, n)$ denote the class of all sets of d distinct integers from $\{1, ..., n\}$. Again in practice, a general rule for the choice of $\mathbb{J}$ is needed. The question then remains on how to choose $\mathbb{J} \in \mathcal{C}(d, n)$ for example in such a way that $\boldsymbol{G}_{\mathbb{J}}(\mathbb{X})$ converges to some population quantity.

4.2.2. *The Chakraborty and Chaudhuri (1996) example*

In the setting of *estimating* location rather than testing a specified value, Chakraborty and Chaudhuri (1996) introduced a variant of the Chaudhuri and Sengupta (1993) transformation, namely

$$\boldsymbol{G}_{\mathbb{J}}(\mathbb{X}) = \big((\boldsymbol{X}_{i_1} - \boldsymbol{X}_{i_{d+1}}), \ldots, (\boldsymbol{X}_{i_d} - \boldsymbol{X}_{i_{d+1}})\big)^{-1}, \tag{6}$$

with the index set $\mathbb{J} = \{i_1, \ldots, i_{d+1}\}$ in $\mathcal{C}(d+1, n)$, thus using $d+1$ sample observations. Then $\boldsymbol{G}_{\mathbb{J}}(\mathbb{X})$ is an ICS statistic ($\boldsymbol{G}_{\mathbb{J}}(\boldsymbol{A}\mathbb{X} + \boldsymbol{b}) = \boldsymbol{G}_{\mathbb{J}}(\mathbb{X})\boldsymbol{A}^{-1}$) and

$$\boldsymbol{G}_{\mathbb{J}}(\mathbb{X})(\mathbb{X} - \boldsymbol{X}_{i_{d+1}})$$

is a maximal invariant statistic with respect to the nonsingular transformations $\{g : g\mathbb{X} = \boldsymbol{A}\mathbb{X} + \boldsymbol{b},\ \boldsymbol{A} \in \mathcal{A}, \boldsymbol{b} \in \mathbb{R}^d\}$. Chakraborty and Chaudhuri (1996) then show that computing the coordinate-wise median on the observations $\{\boldsymbol{G}_{\mathbb{J}}(\mathbb{X})\boldsymbol{X}_i, i \notin \mathbb{J}\}$ and retransforming that result back to the original coordinates via the inverse $\boldsymbol{G}_{\mathbb{J}}(\mathbb{X})^{-1}$ yields a fully affine equivariant version of the sample coordinatewise median, the "transformation-retransformation (TR)" coordinatewise median.

How then to choose $\mathbb{J}$? Write $\boldsymbol{W}_{\mathbb{J}}(\mathbb{X}) = \boldsymbol{G}_{\mathbb{J}}(\mathbb{X})^{-1}$ and let $\boldsymbol{S}$ be a scatter functional. Chakraborty and Chaudhuri (1996) select $\mathbb{J}$ to make the matrix

$$\boldsymbol{W}_{\mathbb{J}}(\mathbb{X})'\boldsymbol{S}(\mathbb{X})^{-1}\boldsymbol{W}_{\mathbb{J}}(\mathbb{X}) \quad \text{or equivalently} \quad \boldsymbol{G}_{\mathbb{J}}(\mathbb{X})\boldsymbol{S}(\mathbb{X})\boldsymbol{G}_{\mathbb{J}}(\mathbb{X})'$$

become as close as possible to a matrix of form $\lambda\boldsymbol{I}_d$, i.e., $\boldsymbol{G}_{\mathbb{J}}(\mathbb{X})$ is made to be close to a version of $\boldsymbol{S}(\mathbb{X})^{-1/2}$ up to a constant. (The criterion for the closeness could be e.g. the variance of the eigenvalues of $\boldsymbol{G}_{\mathbb{J}}(\mathbb{X})\boldsymbol{S}(\mathbb{X})\boldsymbol{G}_{\mathbb{J}}(\mathbb{X})'$.) While only $d+1$ observations are used in defining $\boldsymbol{G}_{\mathbb{J}}(\mathbb{X})$, 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 $\boldsymbol{S}(\mathbb{X})$. 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.

Note that, for all choices of $\mathbb{J}$, $\boldsymbol{S}_{\mathbb{J}}(\mathbb{X}) = \boldsymbol{W}_{\mathbb{J}}(\mathbb{X})\boldsymbol{W}_{\mathbb{J}}(\mathbb{X})'$ is a scatter statistic with expected value $2dCov(\boldsymbol{X})$. The distribution of $\boldsymbol{S}_{\mathbb{J}}(\mathbb{X})$ is however the same for all $\mathbb{J}$ and all $n > d+1$ so that it is not consistent to any population value. Next we consider the behavior of $\boldsymbol{S}_{\hat{\mathbb{J}}}(\mathbb{X})$ when $\hat{\mathbb{J}}$ is chosen so that $\boldsymbol{S}_{\hat{\mathbb{J}}}(\mathbb{X})$ is as close as possible to the value of an auxiliary scatter matrix $\boldsymbol{S}(\mathbb{X})$. If $\mathbb{X}$ is a random sample from F then the finite-sample and limiting distributions of $\boldsymbol{S}_{\hat{\mathbb{J}}}(\mathbb{X})$ depend on both F and the choice of $\boldsymbol{S}(\mathbb{X})$. We next show that, if $\boldsymbol{S}(\mathbb{X})$ is consistent, then $\boldsymbol{S}_{\hat{\mathbb{J}}}(\mathbb{X})$ is consistent to the same population value. (For a similar result but with a different selection criterion, see Corollary 3.8 in Chakraborty (2001).)

Theorem 4.1. *Assume that* $\boldsymbol{S}(\mathbb{X}) \to_P \boldsymbol{S}(F)$ *and write* $\lambda_{\mathbb{J},1} \geq \dots \geq \lambda_{\mathbb{J},d}$ *for the eigenvalues of* $\boldsymbol{G}_{\mathbb{J}}(\mathbb{X})\boldsymbol{S}(\mathbb{X})\boldsymbol{G}'_{\mathbb{J}}(\mathbb{X})$. *If* $\hat{\mathbb{J}} = \mathbb{J}(\mathbb{X})$ *is the value* $\mathbb{J}$ *that minimizes* $\sum_{i=1}^d (\lambda_{\mathbb{J},i} - 1)^2$ *then also* $\boldsymbol{S}_{\hat{\mathbb{J}}}(\mathbb{X}) \to_P \boldsymbol{S}(F)$.

Proof. Write shortly $\boldsymbol{S} = \boldsymbol{S}(F)$, $\hat{\boldsymbol{S}} = \boldsymbol{S}(\mathbb{X})$, and $\boldsymbol{G}_{\mathbb{J}} = \boldsymbol{G}_{\mathbb{J}}(\mathbb{X})$. First note that $||\boldsymbol{G}_{\mathbb{J}}\hat{\boldsymbol{S}}\boldsymbol{G}'_{\mathbb{J}} - \boldsymbol{I}_d||^2 = \sum_{i=1}^d (\lambda_{\mathbb{J},i} - 1)^2$. Fix $\epsilon > 0$, and find $M > 0$ such that, if

$$\mathfrak{J} = \{\mathbb{J} \ : \ ||\boldsymbol{G}_{\mathbb{J}}|| < M \text{ and } ||\boldsymbol{G}_{\mathbb{J}}\boldsymbol{S}\boldsymbol{G}'_{\mathbb{J}} - \boldsymbol{I}_d|| < \epsilon/2\},$$

then $P(\mathfrak{J} \neq \emptyset) \to 1$. Let

$$\tilde{\mathbb{J}} = \begin{cases} \{1, ..., d+1\}, & \text{if } \mathfrak{J} = \emptyset \\ \mathbb{J}_0, & \text{if } \mathfrak{J} \neq \emptyset \text{ and } \mathbb{J}_0 \in \mathfrak{J} \text{ randomly chosen .} \end{cases}$$

As

$$||\boldsymbol{G}_{\mathbb{J}}\hat{\boldsymbol{S}}\boldsymbol{G}'_{\mathbb{J}} - \boldsymbol{I}_d|| \leq ||\boldsymbol{G}_{\mathbb{J}}\boldsymbol{S}\boldsymbol{G}'_{\mathbb{J}} - \boldsymbol{I}_d|| + ||\boldsymbol{G}_{\mathbb{J}}(\hat{\boldsymbol{S}} - \boldsymbol{S})\boldsymbol{G}_{\mathbb{J}}||,$$

it holds that

$$\begin{aligned} P\left(||\boldsymbol{G}_{\hat{\mathbb{J}}}\hat{\boldsymbol{S}}\boldsymbol{G}'_{\hat{\mathbb{J}}} - \boldsymbol{I}_d|| > \epsilon\right) &\leq P\left(||\boldsymbol{G}_{\tilde{\mathbb{J}}}\hat{\boldsymbol{S}}\boldsymbol{G}'_{\tilde{\mathbb{J}}} - \boldsymbol{I}_d|| > \epsilon\right) \\ &\leq P\left(||\boldsymbol{G}_{\tilde{\mathbb{J}}}\boldsymbol{S}\boldsymbol{G}'_{\tilde{\mathbb{J}}} - \boldsymbol{I}_d|| > \frac{\epsilon}{2}\right) + P\left(||\boldsymbol{G}_{\tilde{\mathbb{J}}}(\hat{\boldsymbol{S}} - \boldsymbol{S})\boldsymbol{G}'_{\tilde{\mathbb{J}}}|| > \frac{\epsilon}{2}\right) \\ &\to 0. \end{aligned}$$

This holds as

$$P\left(||\boldsymbol{G}_{\tilde{\mathbb{J}}}\boldsymbol{S}\boldsymbol{G}'_{\tilde{\mathbb{J}}} - \boldsymbol{I}_d|| > \frac{\epsilon}{2}\right) \leq P(\mathfrak{J} = \emptyset) \to 0$$

and

$$\begin{aligned} P\left(||\boldsymbol{G}_{\tilde{\mathbb{J}}}(\hat{\boldsymbol{S}} - \boldsymbol{S})\boldsymbol{G}'_{\tilde{\mathbb{J}}}|| > \frac{\epsilon}{2}\right) &\leq P\left(||\boldsymbol{G}_{\tilde{\mathbb{J}}}||^2||\hat{\boldsymbol{S}} - \boldsymbol{S}|| > \frac{\epsilon}{2}\right) \\ &\leq P(\mathfrak{J} = \emptyset) + P\left(M^2||\hat{\boldsymbol{S}} - \boldsymbol{S}|| > \frac{\epsilon}{2}\right) \\ &\to 0. \end{aligned}$$

As a similar construction can be made for all $\epsilon > 0$, it follows that $\boldsymbol{G}_{\hat{\mathbb{J}}}\hat{\boldsymbol{S}}\boldsymbol{G}'_{\hat{\mathbb{J}}} \to_P \boldsymbol{I}_d$ and therefore $\boldsymbol{S}_{\hat{\mathbb{J}}} - \hat{\boldsymbol{S}} \to_P \boldsymbol{0}$ and, finally $\boldsymbol{S}_{\hat{\mathbb{J}}} \to_P \boldsymbol{S}$.

Remark 4.2. The same $\boldsymbol{W}_{\mathbb{J}}(\mathbb{X}_n)$ is used by Chakraborty et al. (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 $\boldsymbol{W}'_{\mathbb{J}}\boldsymbol{S}(\mathbb{X})^{-1}\boldsymbol{W}_{\mathbb{J}}$ 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, see Section 5.3.

5. Selected applications of ICS functionals

Here we augment the applications mentioned above with several further contexts.

5.1. Multivariate one-sample location problem

Assume that $\mathbb{X}$ is a random sample from a symmetric distribution with the symmetry center $\boldsymbol{\mu}$ (i.e. $\boldsymbol{X} - \boldsymbol{\mu} \stackrel{d}{=} \boldsymbol{\mu} - \boldsymbol{X}$). We wish to test the null hypothesis $H_0: \ \boldsymbol{\mu} = \boldsymbol{0}$. It is then often natural to restrict the family of possible test statistics $Q(\mathbb{X})$ to those which are invariant under the group of affine transformations

$$\mathcal{G} = \{g \ : \ g\mathbb{X} = \boldsymbol{A}\mathbb{X} \ \text{ for some nonsingular } d \times d \text{ matrix } \boldsymbol{A}\}$$

as the null hypothesis is invariant under $\mathcal{G}$. The test statistic $Q(\mathbb{X})$ is invariant under $\mathcal{G}$ if $Q(g\mathbb{X}) = Q(\mathbb{X})$ for all $g \in \mathcal{G}$.

(a) If the data is expressed as a matrix by $\boldsymbol{W} = (\boldsymbol{X}_1, ..., \boldsymbol{X}_n)$, then $\boldsymbol{W}'(\boldsymbol{W}\boldsymbol{W}')^{-1}\boldsymbol{W}$ is a maximal invariant and Hotelling's T^2 test statistic

$$Q^2(\mathbb{X}) = \boldsymbol{1}'_n\boldsymbol{W}'(\boldsymbol{W}\boldsymbol{W}')^{-1}\boldsymbol{W}\boldsymbol{1}_n$$

is invariant under $\mathcal{G}$. The companion location estimate, the sample mean $\hat{\boldsymbol{\mu}}(\mathbb{X}) = \bar{\boldsymbol{X}}$ is affine equivariant, that is, $\hat{\boldsymbol{\mu}}(\boldsymbol{A}\mathbb{X} + \boldsymbol{b}) = \boldsymbol{A}\hat{\boldsymbol{\mu}}(\mathbb{X}) + \boldsymbol{b}$ for all $\boldsymbol{A}$ and $\boldsymbol{b}$.

(b) The multivariate one-sample location tests $Q(\mathbb{X})$ based on marginal signs and signed-ranks, see Puri and Sen (1971), are invariant under the group of permutations, heterogenous sign changes and heterogeneous rescalings, that is, under

$$\mathcal{C} = \{\boldsymbol{PDJ} \ : \ \boldsymbol{P} \in \mathcal{P},\ \boldsymbol{D} \in \mathcal{D}, \text{ and } \boldsymbol{J} \in \mathcal{J}\ \}.$$

(Even more is true: The tests are also invariant under strictly increasing transformations $g : \mathbb{R}^+ \to \mathbb{R}^+$ to the absolute values of the marginal variables. However, we are here interested only in affine transformations.) The statistic $Q(\boldsymbol{S}^{-1/2}(\mathbb{X})\mathbb{X})$ is not affine invariant but if $\boldsymbol{G}(\mathbb{X})$ and $\boldsymbol{L}(\mathbb{X})$ give any solution to the eigenvector and eigenvalue problem

$$\boldsymbol{S}_1^{-1}(\mathbb{X})\boldsymbol{S}_2(\mathbb{X})\boldsymbol{G}(\mathbb{X})' = \boldsymbol{G}(\mathbb{X})'\boldsymbol{L}(\mathbb{X}),$$

then $Q(\boldsymbol{G}(\mathbb{X})\mathbb{X})$ is affine invariant. (In fact it is sufficient here that S_1 and S_2 are scatter matrices with respect to the origin.) Naturally also $Q(\boldsymbol{G}_{\mathbb{J}}(\mathbb{X})\mathbb{X})$, $\mathbb{J} \in \mathcal{C}(d,n)$, is affine invariant for any choice of $\mathbb{J}$. The companion location estimates $\hat{\boldsymbol{\mu}}(\mathbb{X})$, the vector of componentwise medians and the vector of componentwise Hodges-Lehmann estimates, can be made affine equivariant using a transformation and retransformation: $\tilde{\boldsymbol{\mu}}(\mathbb{X}) = \boldsymbol{G}(\mathbb{X})^{-1}\hat{\boldsymbol{\mu}}(\boldsymbol{G}(\mathbb{X})\mathbb{X})$, see Nordhausen et al. (2006).

(c) The one-sample spatial sign and signed-rank test statistics $Q(\mathbb{X})$, see Oja (2010), are invariant under the group of transformations in $\{d\boldsymbol{U} \ : \ d \neq 0,\ \boldsymbol{U} \in \mathcal{U}\}$. Therefore $Q(\boldsymbol{S}^{-1/2}(\mathbb{X})\mathbb{X})$ is affine invariant for any weak covariance functional $\boldsymbol{S}$ and for any version of $\boldsymbol{S}^{-1/2}$. The corresponding location estimates $\hat{\boldsymbol{\mu}}(\mathbb{X})$, the spatial medians and the spatial Hodges-Lehmann estimate, can be made affine equivariant by $\tilde{\boldsymbol{\mu}}(\mathbb{X}) = \boldsymbol{S}(\mathbb{X})^{1/2}\hat{\boldsymbol{\mu}}(\boldsymbol{S}^{-1/2}(\mathbb{X})\mathbb{X})$.

5.2. *Multivariate two-sample location problem*

Assume that $\mathbb{X}_1$ and $\mathbb{X}_2$ are independent random samples of size n_1 and n_2 from distributions $F(\cdot)$ and $F(\cdot - \boldsymbol{\Delta})$, respectively. Write $\mathbb{X} = \mathbb{X}_1 \cup \mathbb{X}_2$ and $n = n_1 + n_2$. We wish to test the null hypothesis $H_0 : \ \boldsymbol{\Delta} = \boldsymbol{0}$. The test statistic $Q(\mathbb{X}_1, \mathbb{X}_2)$ is often required to be invariant under the group of affine transformations

$$\mathcal{G} = \{g \ : \ g\mathbb{X} = \boldsymbol{A}\mathbb{X} + \boldsymbol{b} \ \text{ for some nonsingular } d \times d \text{ matrix } \boldsymbol{A} \text{ and } d\text{-vector } \boldsymbol{b}\}.$$

(a) Assume that $\boldsymbol{X}_1, ..., \boldsymbol{X}_{n_1}$ is the first sample and $\boldsymbol{X}_{n_1+1}, ..., \boldsymbol{X}_{n_1+n_2}$ the second sample. Write $\bar{\boldsymbol{X}}$ for the sample mean vector, $\boldsymbol{S}$ for the sample covariance matrix, both calculated from $\mathbb{X}$, $\boldsymbol{Z}_i = \boldsymbol{S}^{-1/2}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})$, $i = 1, ..., n$, and $\boldsymbol{W} = (\boldsymbol{W}_1, \boldsymbol{W}_2) = (\boldsymbol{Z}_1, ..., \boldsymbol{Z}_n)$. Then $\boldsymbol{W}'\boldsymbol{W}$ is a maximal invariant. Hotelling's two-sample T^2 test statistic

$$\frac{1}{n_1}\boldsymbol{1}_{n_1}'\boldsymbol{W}_1'\boldsymbol{W}_1\boldsymbol{1}_{n_1} + \frac{1}{n_2}\boldsymbol{1}_{n_2}'\boldsymbol{W}_2'\boldsymbol{W}_2\boldsymbol{1}_{n_2}$$

is affine invariant and the corresponding estimate (difference of sample means) is affine equivariant.

(b) The multivariate two-sample location tests $Q(\mathbb{X}_1, \mathbb{X}_2)$ based on marginal signs and ranks (Puri and Sen (1971)) are invariant under

$$\{g \ : \ g\mathbb{X} = \boldsymbol{PDJ}\mathbb{X} + \boldsymbol{b} \ \text{ where } \ \boldsymbol{P} \in \mathcal{P},\ \boldsymbol{D} \in \mathcal{D},\ \boldsymbol{J} \in \mathcal{J},\ \boldsymbol{b} \in \mathbb{R}^d\}.$$

Let again $\boldsymbol{G}(\mathbb{X})$ and $\boldsymbol{L}(\mathbb{X})$ give any solution to the eigenvector and eigenvalue problem

$$\boldsymbol{S}_1^{-1}(\mathbb{X})\boldsymbol{S}_2(\mathbb{X})\boldsymbol{G}(\mathbb{X})' = \boldsymbol{G}(\mathbb{X})'\boldsymbol{L}(\mathbb{X}).$$

Then $Q(\boldsymbol{G}(\mathbb{X})\mathbb{X}_1, \boldsymbol{G}(\mathbb{X})\mathbb{X}_2)$ is affine invariant. Naturally also $Q(\boldsymbol{G}_{\mathbb{J}}(\mathbb{X})\mathbb{X}_1, \boldsymbol{G}_{\mathbb{J}}(\mathbb{X})\mathbb{X}_2)$, $\mathbb{J} \in \mathcal{C}(d+1, n)$, is affine invariant for any choice of $\mathbb{J}$. The difference of the marginal medians and the two-sample Hodges-Lehmann estimate can be made affine equivariant using again the transformation and retransformation technique, see Nordhausen et al. (2006).

(c) The two-sample spatial sign and rank tests $Q(\mathbb{X}_1, \mathbb{X}_2)$ (Oja (2010)) are invariant under

$$\{g \ : \ g\mathbb{X} = d\boldsymbol{U}\mathbb{X} + \boldsymbol{b}, \quad d \neq 0, \ \boldsymbol{U} \in \mathcal{U}, \ \boldsymbol{b} \in \mathbb{R}^d\}.$$

Then $Q(\boldsymbol{S}^{-1}(\mathbb{X})\mathbb{X}_1, \boldsymbol{S}^{-1/2}(\mathbb{X})\mathbb{X}_2)$ is invariant for any choices of weak covariance matrix $\boldsymbol{S}$ and for any version of $\boldsymbol{S}^{-1/2}$. Again, the corresponding estimates $\hat{\boldsymbol{\Delta}}(\mathbb{X}_1, \mathbb{X}_2)$ can be made affine equivariant by $\boldsymbol{S}^{1/2}(\mathbb{X})\hat{\boldsymbol{\Delta}}(\boldsymbol{S}^{-1/2}(\mathbb{X})\mathbb{X}_1, \boldsymbol{S}^{-1/2}(\mathbb{X})\mathbb{X}_2)$.

5.3. Multivariate quantile estimation

For a distribution F on $\mathbb{R}^d$, an associated *quantile function* attaches to each point $\boldsymbol{x}$ a *quantile representation* $\boldsymbol{Q}(\boldsymbol{u}, F)$, indexed by $\boldsymbol{u}$ in the unit ball $\mathbb{B}^d(\boldsymbol{0})$ in $\mathbb{R}^d$. For $\boldsymbol{u} = \boldsymbol{0}$, the most central point $\boldsymbol{Q}(\boldsymbol{0}, F)$ is interpreted as a *d-dimensional median* $\boldsymbol{M}_F$. For $\boldsymbol{u} \neq \boldsymbol{0}$, 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.

Quantile functions on $\mathbb{R}^d$ are desirably *equivariant*, and the associated *outlyingness functions invariant*, under the group of transformations

$$\mathcal{G} = \{g \ : \ g\mathbb{X} = \boldsymbol{A}\mathbb{X} + \boldsymbol{b} \ \text{ for some nonsingular } d \times d \text{ matrix } \boldsymbol{A} \text{ and } d\text{-vector } \boldsymbol{b}\}.$$

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. This is captured in the following definition.

Definition 5.1. *An $\mathbb{R}^d$-valued quantile function $\boldsymbol{Q}(\boldsymbol{u}, F)$, $\boldsymbol{u} \in \mathbb{B}^d(\boldsymbol{0})$, is affine equivariant if, for $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$ with any nonsingular $d \times d$ $\boldsymbol{A}$ and any $\boldsymbol{b}$,*

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

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

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

For the *median* $\boldsymbol{Q}(\boldsymbol{0}, F_{\boldsymbol{X}})$, the equivariance property may be stated simply $\boldsymbol{Q}(\boldsymbol{0}, F_{\boldsymbol{Y}}) = \boldsymbol{A}\,\boldsymbol{Q}(\boldsymbol{0}, F_{\boldsymbol{X}}) + \boldsymbol{b}$. Condition (8) builds *outlyingness invariance* into the definition of affine equivariance of $\boldsymbol{Q}(\cdot, F)$.

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

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

These contours represent equivalence classes of points of equal outlyingness. (But in general c need not be the enclosed probability weight.) Typically, and desirably, they are *nested*. If $\boldsymbol{Q}(\cdot, F)$ is affine equivariant, then 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.

A number of different multivariate quantile functions have been formulated. Some are affine equivariant in the above sense, and some are not. For example, the well-known spatial quantile function is only orthogonally equivariant. However, computation of the spatial quantile function on the data after standardization by a TR transformation, followed by retransformation back to the original coordinates, yields a fully affine equivariant version. See Serfling (2010) for elaboration.

5.4. *Multivariate skewness and kurtosis*

Some other descriptive statistics for multivariate data, like skewness and kurtosis statistics, are desired to be invariant under the group of transformations

$$\mathcal{G} = \{g\ :\ g\mathbb{X} = \boldsymbol{A}\mathbb{X} + \boldsymbol{b}\ \text{ for some nonsingular } d \times d \text{ matrix } \boldsymbol{A} \text{ and } d\text{-vector } \boldsymbol{b}\}.$$

Let again $\bar{\boldsymbol{X}}$ be the sample mean vector and $\boldsymbol{S}$ the sample covariance matrix. Then write $\boldsymbol{Z}_i = \boldsymbol{S}^{-1/2}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})$, $i = 1, ..., n$, and $\boldsymbol{W} = (\boldsymbol{Z}_1, ..., \boldsymbol{Z}_n)$. As mentioned before, $\boldsymbol{W}'\boldsymbol{W}$ is a maximal invariant. The Mardia (1970) skewness and kurtosis statistics

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\boldsymbol{W}'\boldsymbol{W})_{ij}^3 \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{W}'\boldsymbol{W})_{ii}^2$$

are then naturally invariant under $\mathcal{G}$.

Consider next the approach based on a simultaneous use of two location functionals $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ and two scatter functionals $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$. Find a transformation matrix functional $\boldsymbol{G}$ and diagonal matrix valued functional $\boldsymbol{L}$ which solve the above eigenvector and eigenvalue problem and satisfy

$$\boldsymbol{G}\boldsymbol{S}_1\boldsymbol{G} = \boldsymbol{I}_d, \quad \boldsymbol{G}\boldsymbol{S}_2\boldsymbol{G}' = \boldsymbol{L}, \quad \text{and} \quad \boldsymbol{G}(\boldsymbol{T}_1 - \boldsymbol{T}_2) \geq 0$$

where the eigenvalues in $\boldsymbol{L}$ are in a decreasing order. As mentioned before, $\mathbb{Z} = \boldsymbol{G}(\mathbb{X})(\mathbb{X} - \boldsymbol{T}_1(\mathbb{X}))$ is then a maximal invariant. The skewness and kurtosis can be now defined as $\boldsymbol{T}_2(\mathbb{Z})$ and $\boldsymbol{S}_2(\mathbb{Z})$ or

$$||\boldsymbol{T}_2(\mathbb{Z})||^2 \quad \text{and} \quad ||\boldsymbol{S}_2(\mathbb{Z}) - \boldsymbol{I}_d||^2$$

See Kankainen et al. (2007); Nordhausen et al. (2010); Ilmonen et al. (2010).

More generally, if $\mathbb{X}$ is a random sample from $F(\boldsymbol{A} \cdot + \boldsymbol{b})$ with unknown $\boldsymbol{A}$ and $\boldsymbol{b}$ then the model checking should be based on an invariant sample statistic. Nordhausen et al. (2010) used $\boldsymbol{T}_2(\mathbb{Z})$ and $\boldsymbol{S}_2(\mathbb{Z})$ above to distinguish between a wide range of models.

5.5. Independent component analysis

One important and timely example of the use of the ICS functionals is the independent component analysis (ICA). The field of applications of ICA is wide and constantly expanding, varying from biomedical image data applications to signal processing.

In the *independent component (IC) model* it is assumed that the d-variate vector $\boldsymbol{X}$ can be written as

$$\boldsymbol{X} = \boldsymbol{\Omega}\boldsymbol{Z},$$

for some full-rank $d \times d$ *mixing matrix* $\boldsymbol{\Omega}$ and for some d-vector $\boldsymbol{Z}$ with independent components. In the *independent component analysis (ICA)* the aim is to find an estimate for an *unmixing matrix* $\boldsymbol{\Gamma}$ such that $\boldsymbol{\Gamma}\boldsymbol{X}$ has independent components. Naturally $\boldsymbol{\Gamma} = \boldsymbol{\Omega}^{-1}$ is one possible unmixing matrix. The IC model can be formulated in several ways: If the independent components are permuted or multiplied by nonzero scalars they still remain independent. Then the ICA problem reduces to estimating an unmixing matrix $\boldsymbol{\Omega}^{-1}$ only up to the order, signs and scales of the row vectors. In order to be able to identify a mixing matrix one has to assume that at most one of the components of $\boldsymbol{Z}$ is normally distributed. Excellent overviews of independent component analysis are given in Hyvärinen et al. (2001) and Cichocki and Amari (2006).

Definition 5.2. *A scatter functional $\boldsymbol{S}$ is said to possess the independence property if $\boldsymbol{S}(F_{\boldsymbol{X}})$ is a diagonal matrix for all $\boldsymbol{X}$ with independent components.*

Independence property, which most scatter matrices do not enjoy, is essential in the independent component analysis. The regular covariance matrix is a scatter matrix with the independence property. Another scatter matrix with the independence property is the matrix based on fourth moments. In general M functionals (for example Hettmansperger-Randles scatter functional) or other scatter matrix functional families do not posses the independence property. However, for any scatter matrix $\boldsymbol{S}(F_{\boldsymbol{X}})$, its symmetrized version $\boldsymbol{S}_{sym}(F_{\boldsymbol{X}}) = \boldsymbol{S}(F_{\boldsymbol{X}_1 - \boldsymbol{X}_2})$, where $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent copies of $\boldsymbol{X}$, has the independence property (Oja et al. (2006); Tyler et al. (2009)).

Let $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ be two scatter functional with the independence property. We then find a transformation matrix functional $\boldsymbol{G}$ and diagonal matrix valued functional $\boldsymbol{L}$ as a solution of

$$\boldsymbol{S}_1^{-1}\boldsymbol{S}_2\boldsymbol{G}' = \boldsymbol{G}'\boldsymbol{L}.$$

Any solution $\boldsymbol{G}(F_{\boldsymbol{X}})$ is then an unmixing matrix in the IC model and $\boldsymbol{G}(\mathbb{X})$ may be used as a estimate of it. If $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are the moment based functionals given above, then the functional $\boldsymbol{G}$ is the well-known fourth order blind identification (FOBI) functional (Cardoso (1989)).

5.6. Sliced inverse regression

High dimensional data analysis has grown to be an extremely important topic in the field of statistics. Dimension reduction plays a key role in high dimensional data analysis. One then wishes to reduce the dimension of a d-variate random vector $\boldsymbol{X}$ with cumulative distribution function $F_{\boldsymbol{X}}$ using a $k \times d$ transformation matrix $\boldsymbol{B}$ such that $\boldsymbol{X} \to \boldsymbol{Z} = \boldsymbol{B}\boldsymbol{X}$. Moreover, this transformation should be done without losing any information.

If the goal is to use the reduced number of variables in the new coordinate system to predict the value of a known response variable $\boldsymbol{Y}$, then the joint distribution of $\boldsymbol{X}$ and $\boldsymbol{Y}$ should play a role in dimension reduction of $\boldsymbol{X}$. This is nicely conducted in sliced inverse regression (SIR) (Li (1991)). Sliced inverse regression can be seen as an ICS functional application based on two scatter matrices

$$\boldsymbol{S}_1(F_{\boldsymbol{X}}) = Cov(\boldsymbol{X}) \quad \text{and} \quad \boldsymbol{S}_2(F_{\boldsymbol{X}}) = Cov\left(E(\boldsymbol{X}|Y)\right).$$

Note that the second scatter matrix $\boldsymbol{S}_2$ is now supervised in the sense that it depends on the joint distribution of $\boldsymbol{X}$ and $\boldsymbol{Y}$. One again solves the eigenvector and eigenvalue problem

$$\boldsymbol{S}_1^{-1}\boldsymbol{S}_2\boldsymbol{G}' = \boldsymbol{G}'\boldsymbol{L}.$$

and the data can be transformed to a subspace corresponding to nonzero eigenvalues.

SIR can be seen as an example of supervised invariant coordinate selection. For a comprehensive discussion on SIR and supervised invariant coordinate selection, see Liski et al. (2011).

5.7. *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 transformation $\boldsymbol{M}(F)$ that is ICS up to $\mathcal{D}_0$ (see page 8), then 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 authors wish to thank the three Reviewers for their very insightful comments and suggestions, that helped to improve the paper.

The two first authors are grateful of the support from the Academy of Finland

The third author gratefully acknowledges very helpful comments by G. L. Thompson and support under National Science Foundation Grants DMS-0805786 and DMS-1106691 and National Security Agency Grant H98230-08-1-0106.

References

Cardoso, J.-F. (1989). Source separation using higher moments. In *Proceedings of IEEE international conference on acoustics, speech and signal processing*, pp. 2109–2112.

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

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

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

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

Cichocki, A. and S. Amari (2006). *Adaptive Blind Signal and Image Processing*. Chichester: John Wiley & Sons.

Davies, P. L. (1987). Asymptotic behavior of S-estimates of multivariate location parameters and dispersion matrices. *The Annals of Statistics 15*, 1269–1292.

Eaton, M. L. (1989). *Group Invariance Applications in Statistics*. Institute of Mathematical Statistics and the American Statistical Association.

Frahm, G. (2009). Asymptotic distributions of robust shape matrices and scales. *Journal of Multivariate Analysis 100*, 1329–1337.

Hettmansperger, T. P. and R. H. Randles (2002). A practical affine equivariant multivariate median. *Biometrika 89*, 851–860.

Horn, R. A. and C. R. Johnson (1985). *Matrix Analysis*. Cambridge University Press.

Hyvärinen, A., J. Karhunen, and E. Oja (2001). *Independent Component Analysis*. New York: John Wiley & Sons.

Ilmonen, P., J. Nevalainen, and H. Oja (2010). Characteristics of multivariate distributions and the invariant coordinate system. *Statistics and Probability Letters 80(23–24)*, 1844–1853.

Kankainen, A., S. Taskinen, and H. Oja (2007). Tests of multinormality based on location vectors and scatter matrices. *Statistical Methods & Applications 16*, 357–379.

Kent, J. T. and D. E. Tyler (1996). Constrained M-estimation of multivariate location and scatter. *The Annals of Statistics 24*, 1346–1370.

Lehmann, E. L. and J. P. Romano (2005). *Testing Statistical Hypotheses* (Third ed.). Springer.

Li, G. and J. Zhang (1998). Sphering and its properties. *Sankhya The Indian Journal of Statistics 60*, 119–133.

Li, K.-C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association 86*, 316–327.

Liski, E., K. Nordhausen, and H. Oja (2011). Supervised invariant coordinate selection. *submitted*.

Lopuhaä, H. P. (1989). On the relation between S-estimators and M-estimators of multivariate location and covariance. *The Annals of Statistics 17*, 1662–1683.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika 57*, 519–530.

Maronna, R. A. (1976). Robust M-estimators of multivariate location and scatter. *The Annals of Statistics 4*, 51–67.

Maronna, R. A., D. R. Martin, and V. J. Yohai (2006). *Robust Statistics: Theory and Methods*. Chichester: John Wiley & Sons.

Nordhausen, K., H. Oja, and E. Ollila (2010). Multivariate models and the first four moments. In D. Hunter, J. Rosenberger, and D. Richards (Eds.), *Festschrift for Thomas P. Hettmansperger*, Volume in press.

Nordhausen, K., H. Oja, and D. E. Tyler (2006). On the efficiency of invariant multivariate sign and rank tests. In E. P. Liski, J. Isotalo, J. Niemel, S. Puntanen, and G. P. H. Styan (Eds.), *Festschrift for Tarmo Pukkila on his 60th Birthday*, pp. 217–231. Tampere: University of Tampere.

Oja, H. (2010). *Multivariate Nonparametric Methods With R*. New York, USA: Springer-Verlag.

Oja, H., S. Sirkiä, and J. Eriksson (2006). Scatter matrices and independent component analysis. *Austrian Journal of Statistics 35*, 175–189.

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

Puri, M. L. and P. K. Sen (1971). *Nonparametric Methods in Multivariate Analysis*. New York, USA: John Wiley & Sons.

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

Tyler, D. E., F. Critchley, L. Dümbgen, and H. Oja (2009). Invariant co-ordinate selection. *Journal of the Royal Statistical Society, Series B 71*, 549–592.

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