On Invariant Within Equivalence Coordinate System (IWECS) Transformations

Robert Serfling

March, 2015

Abstract In exploratory data analysis and data mining in the very common setting of a data set $\mathbb{X}$ of vectors from $\mathbb{R}^d$, the search for important features and artifacts of a geometrical nature is a leading focus. Here one must insist that such discoveries be invariant under selected changes of coordinates, at least within some specified equivalence relation on geometric structures. Otherwise, interesting findings could be merely artifacts of the coordinate system. To avoid such pitfalls, it is desirable to transform the data $\mathbb{X}$ to an associated data cloud $\mathbb{X}^*$ whose geometric structure may be viewed as intrinsic to the given data $\mathbb{X}$ but also invariant in the desired sense. General treatments of such "invariant coordinate system" transformations have been developed from various perspectives. As a timely step, here we formulate a more structured and unifying framework for the relevant concepts. With this in hand, we develop results that clarify the roles of so-called transformation-retransformation transformations. We illustrate by treating invariance properties of some outlyingness functions. Finally, we examine productive connections with maximal invariants.

1 Introduction

In exploratory data analysis and data mining in the very common setting of a data set $\mathbb{X}$ of vectors from $\mathbb{R}^d$, a leading focus is the search for important features and artifacts of a geometrical nature. Here one should insist that such discoveries be invariant under selected changes of coordinates, or at least be invariant under such changes up to a particular equivalence relation on geometric structures. Otherwise, what appears to be interesting geometric structure could be nothing but an artifact of the particular coordinate system adopted. To avoid such a pitfall, the data $\mathbb{X}$ can be

---

Robert Serfling
Department of Mathematical Sciences, University of Texas at Dallas, Richardson, Texas 75083 USA, e-mail: `serfling@utdallas.edu`, website: `http://www.utdallas.edu/~serfling`

transformed to an associated new data cloud $\mathbb{X}^*$ having geometric structure that is intrinsically related to $\mathbb{X}$ but also invariant in the desired sense. General treatments of such "invariant coordinate system" transformations have been developed from various perspectives. As a timely next step, here we introduce a more structured and unifying framework for the relevant concepts, develop results that clarify the use of transformation-retransformation transformations, illustrate by treating invariance of some popular outlyingness functions, and productively examine connections with maximal invariants.

The topic of transformation to an "*invariant coordinate system (ICS)*" is treated broadly in the seminal paper of Tyler, Critchley, Dümbgen, and Oja (2009), and further general treatments are provided in Ilmonen, Nevalainen, and Oja (2010), Serfling (2010), and Ilmonen, Oja, and Serfling (2012). See also Nordhausen (2008) for useful results. Collectively, these sources treat two quite different approaches toward construction of "ICS" transformations and discuss a diversity of interesting practical applications.

However, the various treatments to date are not completely coherent and precise with respect to what is actually meant by "ICS". Indeed, for many of the examples and applications, the desired invariance is achieved only within some equivalence relation defined on the geometric structures of data sets. For example, for a data set $\mathbb{X}$, when we seek to identify its geometric structure that is invariant under affine transformation, it might be the case for the given application that differences due to homogeneous scale changes, coordinatewise sign changes, and translations may be ignored. That is, for any affine transformation of the given data cloud $\mathbb{X}$ to $\mathbb{Y}$, we might require only that the corresponding invariant coordinate systems $\mathbb{X}^*$ and $\mathbb{Y}^*$ agree only within such a specified equivalence. To accommodate a variety of such practical applications, the notions and terminology of "ICS" have evolved very productively but in somewhat loose fashion.

It is now timely and useful to have a more structured conceptual framework that draws together the various "ICS" results and adds perspective. For this purpose, we introduce and study in **Section 2** a precise notion of "*invariant within equivalence coordinate system (IWECS)*" transformation: $\mathbf{M}(\mathbb{X})$ such that the transformed data $\mathbf{M}(\mathbb{X})\mathbb{X}$ is invariant under transformation of $\mathbb{X}$ relative to a transformation group $\mathscr{G}$, subject to equivalence relative to another transformation group $\mathscr{F}$. That is, for $g \in \mathscr{G}$, $\mathbf{M}(\mathbb{X})\mathbb{X}$ and $\mathbf{M}(g\mathbb{X})g\mathbb{X}$ need not be equal but must fall in the same equivalence class. Specifically, $\mathbf{M}(\mathbb{X})\mathbb{X}$ is to be $\mathscr{G}$-invariant within $\mathscr{F}$-equivalence.

It is seen in Serfling (2010) that the ICS transformations of practical interest fall within the class of transformation-retransformation (TR) transformations, which are essentially inverse square roots of covariance matrices. The chief purpose of TR transformations is standardization of data, so that estimators, test statistics, and other sample statistics become affine invariant or equivariant when defined on the standardized data. However, in some such cases a strong type of TR transformation is needed, namely an ICS transformation. Also, it is of interest to know when a TR transformation may directly play the role of an ICS transformation. Here we note that, to serve additionally as an ICS transformation, a TR matrix must be rather atypical, since the "usual" ICS transformations cannot be symmetric or triangular

(Serfling, 2010). In **Section 3** we provide further clarifications on TR versus ICS transformations, as follows. Theorem 2 provides the narrowest equivalence (i.e., the smallest $\mathscr{F}$) for which TR transformations can serve as IWECS transformations relative to linear or affine invariance. Theorem 3 exhibits a key special class of TR transformations for which the corresponding IWECS are affine invariant relative to the smallest possible nontrivial choice of $\mathscr{F}$. As illustrations of the application of these theorems, we treat affine invariant TR versions of the spatial outlyingness function and of the projection outlyingness function when the number of projections used is finite.

The construction of TR matrices that possess the structural properties requisite to be ICS (or IWECS) is somewhat challenging. Relative to the linear and affine transformation groups, connections between a useful strong special case of ICS and IWECS transformations and the relevant maximal invariant statistics are examined in **Section 4.1**. Thus maximal invariant statistics can play a role in constructing ICS and IWECS transformations in this special case. We provide background references on two distinctive approaches that have been developed along these lines. Further exploiting connections with maximal invariant statistics, in **Section 4.2** we revisit classical treatments (Lehmann, 1959) of maximal invariants relative to these groups and "discover" a competitive third approach, one offering greater simplicity and less computational burden.

The present paper treats only the case of data from a Euclidean space, as does all of the literature to date except for extension to complex-valued data (Ilmonen, 2013). However, the concepts we present in fact can have very general extension and potentially have application in quite diverse contexts. We provide brief discussion in **Section 5**.

As the literature we cite in this paper amply portrays, there has been a prominent guiding influence in developing, studying, and applying ICS transformations. The contributions of the present paper are dedicated as a tribute to Hannu Oja and his leadership.

2 A General Framework for Formulation of Invariant Within Equivalence Coordinate Systems in $\mathbb{R}^d$

2.1 General Framework

Here we draw together and extend recent general treatments of invariant coordinate system (ICS) transformations (Nordhausen, 2008, Tyler, Critchley, Dümbgen, and Oja, 2009, Ilmonen, Nevalainen, and Oja, 2010, Serfling, 2010, and Ilmonen, Oja, and Serfling, 2012). A general framework for describing the inherent geometrical structure of a data set in $\mathbb{R}^d$ via *invariant within equivalence coordinate system representations* is defined as follows.

Definition 1. An *invariant within equivalence coordinate system (IWECS)* in $\mathbb{R}^d$ consists of three components $(\mathscr{G}, \mathbf{M}(\cdot), \mathscr{E})$,

1. a group $\mathscr{G}$ of transformations g on data sets $\mathbb{X}$ of observations from $\mathbb{R}^d$,
2. a data-based $d \times d$ matrix transformation $\mathbf{M}(\mathbb{X})$ taking $\mathbb{X}$ to $\mathbf{M}(\mathbb{X})\mathbb{X}$, and
3. an equivalence relation $\mathscr{E}$ on the transformed data sets $\mathbf{M}(\mathbb{X})\mathbb{X}$,

such that $\mathbf{M}(g\mathbb{X})g\mathbb{X}$, $g \in \mathscr{G}$, all lie in the same equivalence class relative to $\mathscr{E}$, i.e., are invariant relative to $\mathscr{G}$ within $\mathscr{E}$-equivalence. □

Thus the $\mathscr{E}$-orbit to which $\mathbf{M}(\mathbb{X})\mathbb{X}$ belongs is invariant under transformation of $\mathbb{X}$ by $g \in \mathscr{G}$. We call the matrix $\mathbf{M}(\cdot)$ an *invariant within equivalence coordinate system (IWECS) transformation*, and resulting the transformed data $\mathbf{M}(\mathbb{X})\mathbb{X}$ is the desired *invariant within equivalence coordinate system*.

Whereas $\mathscr{G}$ concerns transformations on initial data sets $\mathbb{X}$ in $\mathbb{R}^d$ and represents a criterion for invariance, the equivalence relation $\mathscr{E}$ concerns the transformed data sets $\mathbf{M}(\mathbb{X})\mathbb{X}$ and represents a criterion for equivalent geometric structure. There are many possibilities for $(\mathscr{G}, \mathscr{E})$. In the example of Section 1, $\mathscr{G}$ consists of affine transformations and $\mathscr{E}$ represents data sets as having equivalent geometric structure if they differ only with respect to homogeneous scale change, coordinatewise sign changes, and translation.

For given $(\mathscr{G}, \mathscr{E})$, the challenge is to find a suitable $\mathbf{M}(\cdot)$ satisfying Definition 1. In the sequel, we consider the special case that $\mathscr{E}$ corresponds to invariance under a group of transformations $\mathscr{F}$ and denote the above framework by $(\mathscr{G}, \mathbf{M}(\cdot), \mathscr{F})$. Then a key criterion for finding a solution is provided by the following result, which follows immediately from Definition 1.

Theorem 1. *For given $(\mathscr{G}, \mathscr{F})$, a suitable IWECS transformation is given by any $\mathbf{M}(\cdot)$ such that, for any $g \in \mathscr{G}$, there exists $f_0 = f_0(g, \mathbb{X}) \in \mathscr{F}$ for which*

$$\mathbf{M}(g\mathbb{X})g\mathbb{X} = f_0\,\mathbf{M}(\mathbb{X})\mathbb{X}. \tag{1}$$

The following result provides a useful sufficient condition for (1) in the form of a structural requirement on the matrix $\mathbf{M}(\cdot)$ that in practice serves essentially as the definition of an IWECS transformation. The proof is immediate.

Corollary 1. *For given $(\mathscr{G}, \mathscr{F})$, a suitable IWECS transformation is given by any $\mathbf{M}(\cdot)$ such that, for any $g \in \mathscr{G}$, there exists $f_0 = f_0(g, \mathbb{X}) \in \mathscr{F}$ for which*

$$\mathbf{M}(g\mathbb{X}) = f_0\,\mathbf{M}(\mathbb{X})g^{-1}. \tag{2}$$

With $\mathscr{A}$ the set of all nonsingular $d \times d$ matrices, important choices for $\mathscr{G}$ are

$\mathscr{G}_0 = \{g : g\mathbb{X} = \mathbf{A}\mathbb{X},\ \mathbf{A} \in \mathscr{A}\}$ (nonsingular linear transformation),
$\mathscr{G}_1 = \{g : g\mathbb{X} = \mathbf{A}\mathbb{X} + \mathbf{b},\ \mathbf{A} \in \mathscr{A},\ \mathbf{b} \in \mathbb{R}^d\}$ (affine transformation).

For $\mathscr{F}$, key choices are

$\mathscr{D}_0 = \{f : f\mathbb{Y} = c\mathbb{Y},\ c > 0\}$ (homogeneous rescaling),

$\mathscr{D} = \{f : f\mathbb{Y} = \mathrm{diag}(c_1,\ldots,c_d)\mathbb{Y},\ c_i > 0,\ i = 1,\ldots,d\}$ (heterogeneous rescaling),

$\mathscr{J} = \{f : f\mathbb{Y} = \mathrm{diag}(c_1,\ldots,c_d)\mathbb{Y},\ c_i = \pm 1,\ i = 1,\ldots,d\}$ (heterogeneous sign changing),

$\mathscr{P} = \{f : f\mathbb{Y} = \mathbf{P}\mathbb{Y},\ \mathbf{P}$ is a permutation matrix$\}$ (permutation),

$\mathscr{U} = \{f : f\mathbb{Y} = \mathbf{U}\mathbb{Y},\ \mathbf{U}$ is an orthogonal matrix$\}$ (rotation and/or reflection).

The above $\mathscr{G}$ and $\mathscr{F}$ arise quite naturally in nonparametric multivariate inference, as discussed in Ilmonen, Oja, and Serfling (2012), where $\mathscr{G}_1$ is especially emphasized. Their equation (4) corresponds to our equation (2) specialized to $\mathscr{G}_1$ and in that form is given as their definition of what here we call an IWECS transformation. Let us also note that certain combinations of the above choices of $\mathscr{F}$ are of special interest, for example:

$\mathscr{F}_0 = \{f : f\mathbb{Y} = c\,\mathbb{Y} + \mathbf{b},\ c > 0,\ \mathbf{b} \in \mathbb{R}^d\}$
(translation, homogeneous rescaling),

$\mathscr{F}_1 = \{f : f\mathbb{Y} = c\,\mathrm{diag}(c_1,\ldots,c_d)\,\mathbb{Y} + \mathbf{b},\ c > 0,\ c_i = \pm 1,\ i = 1,\ldots,d,\ \mathbf{b} \in \mathbb{R}^d\}$
(translation, homogeneous rescaling, heterogeneous sign changing),

$\mathscr{F}_2 = \{f : f\mathbb{Y} = \mathrm{diag}(\pm c_1,\ldots,\pm c_d)\,\mathbb{Y} + \mathbf{b},\ c_i > 0,\ i = 1,\ldots,d,\ \mathbf{b} \in \mathbb{R}^d\}$
(translation, heterogeneous rescaling, heterogeneous sign changing),

$\mathscr{F}_3 = \{f : f\mathbb{Y} = \mathbf{U}(c\,\mathbb{Y} + \mathbf{b}),\ c > 0,\ \mathbf{U}$ orthogonal, $\mathbf{b} \in \mathbb{R}^d\}$
(translation, homogeneous rescaling, rotation, reflection),

$\mathscr{F}_4 = \{f : f\mathbb{Y} = \mathbf{U}(\mathrm{diag}(c_1,\ldots,c_d)\,\mathbb{Y} + \mathbf{b}),\ c_i > 0,\ i = 1,\ldots,d,\ \mathbf{U}$ orthogonal, $\mathbf{b} \in \mathbb{R}^d\}$
(translation, heterogeneous rescaling, rotation, reflection).

In particular, the example discussed in Section 1 concerns $\mathscr{G}_1$ and $\mathscr{F}_1$.

The property that a transformation is IWECS with respect to $(\mathscr{G},\mathscr{F})$ becomes weaker if $\mathscr{F}$ acquires additional transformations. In this respect, let us note that $\mathscr{F}_0 \subset \mathscr{F}_1 \subset \mathscr{F}_2$ and $\mathscr{F}_0 \subset \mathscr{F}_3 \subset \mathscr{F}_4$ so that here the strongest case corresponds to $\mathscr{F} = \mathscr{F}_0$. Of course, still stronger is the ideal case that the equivalence relation $\mathscr{E}$ (i.e., the group $\mathscr{F}$) may be omitted and the invariance relative to $\mathscr{G}$ is strict: the transformed sets $\mathbf{M}(g\mathbb{X})g\mathbb{X}$, $g \in \mathscr{G}$, are identical without qualification and $\mathbf{M}(\cdot)$ is a purely ICS transformation. Generally, however, this aspiration is too stringent and must be relaxed, adopting an equivalence criterion that is as narrow as possible.

3 TR Matrices as IWECS Transformations

In seeking IWECS transformations relative to the popular affine group $\mathscr{G}_1$, one may inquire whether widely used standardizing transformations such as the inverse square roots of scatter matrices suffice for this purpose. That is, more precisely, may a transformation-retransformation (TR) matrix serve as an IWECS matrix? As shown in Serfling (2010), the answer is negative except for some very special cases that exclude popular ones. Hence it becomes of interest to explore what "minimal" $\mathscr{F}$ suffices for an arbitrary TR transformation to serve as an IWECS transformation relative to $(\mathscr{G}_1, \mathscr{F})$. In Section 3.1 we review the definition of TR matrices, and in Sections 3.2 and 3.3 we develop explicit answers to this question.

3.1 Definition of TR Matrices

A *transformation-retransformation (TR)* matrix is a positive definite $d \times d$ matrix $\mathbf{M}(\mathbb{X})$ (not necessarily symmetric) such that, for $\mathbb{Y} = \mathbf{A}\mathbb{X} + \mathbf{b}$ with any nonsingular $\mathbf{A}$ and any $\mathbf{b}$,

$$\mathbf{A}^\top \mathbf{M}(\mathbb{Y})^\top \mathbf{M}(\mathbb{Y}) \mathbf{A} = k_2 \mathbf{M}(\mathbb{X})^\top \mathbf{M}(\mathbb{X}),$$

with $k_2 = k_2(\mathbf{A}, \mathbf{b}, \mathbb{X})$ a positive scalar function of $\mathbf{A}$, $\mathbf{b}$, and $\mathbb{X}$. Such TR matrices are equivalently given by factorizations of weak covariance (WC) matrices, i.e., via

$$\mathbf{C}(\mathbb{X}) = (\mathbf{M}(\mathbb{X})^\top \mathbf{M}(\mathbb{X}))^{-1},$$

where the symmetric positive definite $d \times d$ WC matrix $\mathbf{C}(\mathbb{X})$ satisfies

$$\mathbf{C}(\mathbb{Y}) = k_1 \mathbf{A} \mathbf{C}(\mathbb{X}) \mathbf{A}^\top,$$

with $k_1 = k_1(\mathbf{A}, \mathbf{b}, \mathbb{X})$ a positive scalar function of $\mathbf{A}$, $\mathbf{b}$, and $\mathbb{X}$. For $k_1 = 1$, $\mathbf{C}(\mathbb{X})$ is a strict “covariance” matrix. Typical standardizations of data $\mathbb{X}$ for various purposes are given by $\mathbf{M}(\mathbb{X})\mathbb{X}$. See Serfling (2010) and Ilmonen, Oja, and Serfling (2012) for detailed discussion and examples.

3.2 TR Matrices as IWECS Transformations

We now explore whether such an $\mathbf{M}(\cdot)$ can be IWECS. In particular, Serfling (2010) shows, in different notation, that any IWECS transformation relative to $(\mathscr{G}_1, \mathscr{F}_1)$ is TR, but not conversely, one counter-example being the popular Tyler (1987) TR matrix. But is there a broader $\mathscr{F}$ for which *any TR matrix* is in fact IWECS? The following result answers this in the affirmative, for both $\mathscr{G}_1$ and $\mathscr{G}_0$, with $\mathscr{F} = \mathscr{F}_3$.

Theorem 2. *Every TR matrix is IWECS relative to* $(\mathscr{G}_1, \mathscr{F}_3)$ *and also to* $(\mathscr{G}_0, \mathscr{F}_3)$.

Proof. (i) Let us first consider $(\mathscr{G}_1, \mathscr{F}_3)$. Let $g \in \mathscr{G}_1$ be given by $g\mathbf{x} = \mathbf{A}\mathbf{x} + \mathbf{b}$ for some nonsingular $\mathbf{A}$ and any $\mathbf{b}$. It is shown in Serfling (2010), Lemma 5.1, that, for any TR matrix $\mathbf{M}(\cdot)$, and for $\mathbb{Y} = \mathbf{A}\mathbb{X} + \mathbf{b}$ with any nonsingular $\mathbf{A}$ and any $\mathbf{b}$, the matrix

$$\mathbf{U}_0 = \mathbf{U}_0(\mathbf{A}, \mathbf{b}, \mathbb{X}) = k_2(\mathbf{A}, \mathbf{b}, \mathbb{X})^{1/2} (\mathbf{M}(\mathbb{Y})^\top)^{-1} (\mathbf{A}^\top)^{-1} \mathbf{M}(\mathbb{X})^\top$$

is orthogonal. Then we readily obtain

$$\mathbf{M}(\mathbb{Y}) = k_2^{1/2} \mathbf{U}_0 \mathbf{M}(\mathbb{X}) \mathbf{A}^{-1} \quad (3)$$

and in turn

$$\mathbf{M}(\mathbb{Y})\mathbb{Y} = k_2^{1/2}\mathbf{U}_0[\mathbf{M}(\mathbb{X})\mathbb{X} + \mathbf{M}(\mathbb{X})\mathbf{A}^{-1}\mathbf{b}] = f_0\mathbf{M}(\mathbb{X})\mathbb{X}, \tag{4}$$

where $f_0 = f_0(\mathbf{A}, \mathbf{b}, \mathbb{X})$ represents translation of $\mathbf{M}(\mathbb{X})\mathbb{X}$ by the constant $\mathbf{M}(\mathbb{X})\mathbf{A}^{-1}\mathbf{b}$, followed by homogeneous scale change by $k_2(\mathbf{A}, \mathbf{b}, \mathbb{X})$ and then rotation/reflection by orthogonal $\mathbf{U}_0$. Thus $f_0 \in \mathscr{F}_3$ and equation (1) in Theorem 1 is satisfied for the given $g \in \mathscr{G}_1$.

(ii) For $(\mathscr{G}_0, \mathscr{F}_3)$, the proof is similar. □

Theorem 2 shows explicitly the precise strengths and limitations of TR matrices as IWECS transformations. We can apply this result through various straightforward corollaries, for example the following.

Corollary 2. *If an $\mathbb{R}^m$-valued statistic $\mathbf{Q}(\mathbb{X})$ is invariant with respect to $\mathscr{F}_3$, then its evaluation at a TR-based IWECS $\mathbf{M}(\mathbb{X})\mathbb{X}$ relative to either $(\mathscr{G}_1, \mathscr{F}_3)$ or $(\mathscr{G}_0, \mathscr{F}_3)$ is invariant with respect to $\mathscr{G}_1$ or $\mathscr{G}_0$, respectively.*

Example 1. Invariance of spatial outlyingness function. The *spatial outlyingness function* (Serfling, 2010) is defined as

$$O_S(\mathbf{x}, \mathbb{X}) = \|\mathbf{R}_S(\mathbf{x}, \mathbb{X})\|, \ \mathbf{x} \in \mathbb{R}^d,$$

where $\mathbf{R}_S(\mathbf{x}, \mathbb{X})$ is the spatial centered rank function (Oja, 2010) in $\mathbb{R}^d$ given by

$$\mathbf{R}_S(\mathbf{x}, \mathbb{X}) = n^{-1}\sum_{i=1}^{n}\mathbf{S}(\mathbf{x} - \mathbf{X}_i), \ \mathbf{x} \in \mathbb{R}^d,$$

with $\mathbf{S}(\mathbf{y})$ the *spatial sign function* (or *unit vector function*) in $\mathbb{R}^d$ given by

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

It is readily checked that $\mathbf{R}_S(\mathbf{x}, \mathbb{X})$ is *translation and homogeneous scale invariant* and *orthogonally equivariant*:

$$\begin{aligned} \mathbf{R}_S(\mathbf{x} - \mathbf{b}, \mathbb{X} - \mathbf{b}) &= \mathbf{R}_S(\mathbf{x}, \mathbb{X}), \\ \mathbf{R}_S(c\mathbf{x}, c\mathbb{X}) &= \mathbf{R}_S(\mathbf{x}, \mathbb{X}), \\ \mathbf{R}_S(\mathbf{U}\mathbf{x}, \mathbf{U}\mathbb{X}) &= \mathbf{U}\mathbf{R}_S(\mathbf{x}, \mathbb{X}). \end{aligned}$$

Then $O_S(\mathbf{x}, \mathbb{X})$ is *translation, homogeneous scale, and orthogonally invariant, i.e., invariant with respect to the group $\mathscr{F}_3$*, but is not affine invariant. However, fully affine invariant versions are immediately obtained via Corollary 2: For for any TR matrix $\mathbf{M}(\cdot)$, the so-called *TR spatial outlyingness function* corresponding to $\mathbf{M}(\cdot)$,

$$O_S^{(\mathrm{TR})}(\mathbf{x}, \mathbb{X}) = O_S(\mathbf{M}(\mathbb{X})\mathbf{x}, \mathbf{M}(\mathbb{X})\mathbb{X}), \ \mathbf{x} \in \mathbb{R}^d,$$

is *affine invariant*. □

3.3 Special Types of TR Matrix for IWECS Transformations

It would be desirable to have the result of Theorem 2 for a smaller choice of $\mathscr{F}$ than $\mathscr{F}_3$. In this vein, we ask: *What additional property is required of a TR matrix* $\mathbf{M}(\cdot)$ *in order for it to serve as an IWECS transformation relative to* $\mathscr{G}_1$ *for a choice of* $\mathscr{F}$ *narrower than* $\mathscr{F}_3$*?*

A clue is given by equation (3) in the proof of Theorem 2. If we simply require that $\mathbf{M}(\cdot)$ satisfy this equation *without* the factor $\mathbf{U}_0$, then equation (4) would hold without the presence of $\mathbf{U}_0$, yielding the following very strong conclusion.

Theorem 3. *Let* $\mathbf{M}(\cdot)$ *be a TR matrix such that, for* $\mathbb{Y} = \mathbf{A}\mathbb{X} + \mathbf{b}$ *as above,*

$$\mathbf{M}(\mathbb{Y}) = k_2^{1/2}\mathbf{M}(\mathbb{X})\mathbf{A}^{-1}, \tag{5}$$

with $k_2 = k_2(\mathbf{A}, \mathbf{b}, \mathbb{X})$ *as in the definition of the given TR matrix. Then* $\mathbf{M}(\cdot)$ *is IWECS relative to* $(\mathscr{G}_1, \mathscr{F}_0)$ *and to* $(\mathscr{G}_0, \mathscr{F}_0)$*.*

An analogue of Corollary 2 is

Corollary 3. *If an* $\mathbb{R}^m$*-valued statistic* $\mathbf{Q}(\mathbb{X})$ *is invariant with respect to* $\mathscr{F}_0$*, then its evaluation at a TR-based IWECS* $\mathbf{M}(\mathbb{X})\mathbb{X}$ *for* $\mathbf{M}(\cdot)$ *satisfying (5) is, relative to either* $(\mathscr{G}_1, \mathscr{F}_0)$ *or* $(\mathscr{G}_0, \mathscr{F}_0)$*, invariant with respect to* $\mathscr{G}_1$ *or* $\mathscr{G}_0$*, respectively.*

In comparison with Corollary 2, Corollary 3 requires more of the TR matrix but yields a stronger conclusion by allowing $\mathscr{F}_0$ instead of $\mathscr{F}_3$.

Example 2. Invariance of projection outlyingness with finitely many projections. With ν the median and η the MAD (median absolute deviation from the median), the well-known *projection outlyingness function* given by

$$O_P(\mathbf{x}, \mathbb{X}) = \sup_{\|\mathbf{u}\| = 1} \left| \frac{\mathbf{u}^\top \mathbf{x} - \nu(\mathbf{u}^\top \mathbb{X})}{\eta(\mathbf{u}^\top \mathbb{X})} \right|, \ \mathbf{x} \in \mathbb{R}^d, \tag{6}$$

represents the worst case scaled deviation outlyingness of projections of $\mathbf{x}$ onto lines. It is affine invariant, highly masking robust (Dang and Serfling, 2010), and does not impose ellipsoidal contours as does the very popular Mahalanobis distance outlyingness function, which also is affine invariant. However, $O_P(\mathbf{x}, \mathbb{X})$ is highly computational, and to overcome this burden Serfling and Mazumder (2013) develop and study a modified version entailing only finitely many selected projections, say $\Delta = \{\mathbf{u}_1, \ldots, \mathbf{u}_K\}$, i.e.,

$$O_P^{(\Delta)}(\mathbf{x}, \mathbb{X}) = \sup_{\mathbf{u} \in \Delta} \left| \frac{\mathbf{u}^\top \mathbf{x} - \nu(\mathbf{u}^\top \mathbb{X})}{\eta(\mathbf{u}^\top \mathbb{X})} \right|, \ \mathbf{x} \in \mathbb{R}^d. \tag{7}$$

However, $O_P^{(\Delta)}(\mathbf{x}, \mathbb{X})$ with finite Δ is no longer affine invariant. Nor is it orthogonally invariant, so Corollary 2 is inapplicable and thus an arbitrary TR version does not achieve affine invariance. On the other hand, simply using invariance of $O_P^{(\Delta)}(\mathbf{x}, \mathbb{X})$ with respect to $\mathscr{F}_0$, it follows by Corollary 3 that any TR version

$$O_P^{(\Delta,TR)}(\mathbf{x}, \mathbb{X}) = O_P^{(\Delta)}(\mathbf{M}(\mathbb{X})\mathbf{x}, \mathbf{M}(\mathbb{X})\mathbb{X}),\ \mathbf{x} \in \mathbb{R}^d,$$

with $\mathbf{M}(\cdot)$ satisfying (5) is indeed affine invariant. Of course, standardizing by $\mathbf{M}(\cdot)$ introduces a further computational issue, and Serfling and Mazumder (2013) also develop computationally attractive choices of $\mathbf{M}(\cdot)$ satisfying (5). □

Remark 1. A TR matrix satisfying the special condition (5) is distinguished as a "strong invariant coordinate system" (SICS) transformation in Serfling (2010) and Ilmonen, Oja, and Serfling (2012), where also other results like Theorem 3 are seen corresponding to replacement of $\mathbf{U}_0$ in (3) by possibilities other than simply the identity matrix and hence corresponding to $\mathscr{F}$ larger than $\mathscr{F}_0$. □

Remark 2. Let us compare condition (5) with the somewhat similar condition given by (2), which in the present setting would be expressed as

$$\mathbf{M}(\mathbb{Y}) = f_0 \mathbf{M}(\mathbb{X}) g^{-1} \tag{8}$$

for any given $g \in \mathscr{G}$ for some related $f_0 \in \mathscr{F}_0$. For the case $\mathscr{G} = \mathscr{G}_1$, let g be given by $g\mathbf{x} = \mathbf{A}\mathbf{x} + \mathbf{b}$ for some nonsingular $\mathbf{A}$ and any $\mathbf{b}$. Now it is readily checked that the transformation g^{-1} consists of translation by $-\mathbf{b}$ followed by application of $\mathbf{A}^{-1}$, or equivalently application of $\mathbf{A}^{-1}$ followed by translation by $-\mathbf{A}^{-1}\mathbf{b}$. *Thus (5) is equivalent to (2) with suitable choice of* $f_0 \in \mathscr{F}_0$. The argument for $\mathscr{G} = \mathscr{G}_0$ is similar. In dealing with $\mathscr{G} = \mathscr{G}_1$, the use of (5) is more direct and convenient. □

4 Some Connections With Maximal Invariants

A natural "invariance principle" is that artifacts of the data $\mathbb{X}$ which are invariant without qualification relative to a group $\mathscr{G}$ of transformations should be functions of a suitable "maximal invariant" statistic that constitutes a labeling of the orbits of $\mathscr{G}$. See Lehmann and Romano (2005), §6.2, for elaboration. For a data set $\mathbb{X}$ = $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ of observations in $\mathbb{R}^d$, a maximal invariant is obtained via some suitable matrix-valued transformation $\mathbf{B}(\mathbb{X})$ applied to $\mathbb{X}$, producing $\mathbf{B}(\mathbb{X})\mathbb{X}$ as the desired maximal invariant.

In this case, an ICS $\mathbf{M}(\mathbb{X})\mathbb{X}$ relative to $\mathscr{G}$ should be expressible as a function of $\mathbf{B}(\mathbb{X})\mathbb{X}$. However, this need not be true for an IWECS, of course. In Section 4.1, relative to $\mathscr{G}_0$ and $\mathscr{G}_1$, we exhibit connections between ICS transformations and the pertinent maximal invariant transformations. These connections are exploited in Section 4.2 to "discover" from some classical results a new approach toward construction of ICS and IWECS transformations. In Section 5 the connections are extended to the case of an arbitrary $\mathscr{G}$.

4.1 Connections in the Case of Groups $\mathscr{G}_0$ and $\mathscr{G}_1$

With reference to the groups $\mathscr{G}_0$ and $\mathscr{G}_1$, the following results of Ilmonen, Oja, and Serfling (2012), Theorem 3.1, connect maximal invariants with TR matrices $\mathbf{M}(\cdot)$ satisfying a strong special case of equation (5), namely

$$\mathbf{M}(\mathbb{Y}) = \mathbf{M}(\mathbb{X})\mathbf{A}^{-1}, \tag{9}$$

where $\mathbb{Y}$ denotes $\mathbf{A}\mathbb{X}$ in the case of $\mathscr{G}_0$ and $\mathbf{A}\mathbb{X}+\mathbf{b}$ in the case of $\mathscr{G}_1$. They show under (9) that

(i) $\mathbf{M}(\mathbb{X})\mathbb{X}$ is a maximal invariant under $\mathscr{G}_0$,

(ii) $\mathbf{M}(\mathbb{X})(\mathbb{X}-T(\mathbb{X}))$ is a maximal invariant under $\mathscr{G}_1$, for any location statistic $T(\mathbb{X})$.

Of course, in view of (9), $\mathbf{M}(\cdot)$ is an IWECS transformation. Also, note that under (9) we have $\mathbf{M}(\mathbb{X}) = \mathbf{M}(\mathbb{X}-T(\mathbb{X}))$, and thus the maximal invariant in (ii) may also be written as $\mathbf{M}(\mathbb{X}-T(\mathbb{X}))(\mathbb{X}-T(\mathbb{X}))$.

In case (i), $\mathbf{M}(\cdot)$ is a very strong special case of IWECS transformation, namely a pure ICS transformation without qualification by an equivalence relation, for we have $\mathbf{M}(g\mathbb{X})g\mathbb{X} = \mathbf{M}(\mathbb{X})\mathbb{X}$, $g \in \mathscr{G}_0$. Note that under merely (5) instead of the strengthening to (9), we have by Theorem 3 that $\mathbf{M}(\cdot)$ is an IWECS transformation relative to $(\mathscr{G}_1, \mathscr{F}_0)$, a slightly weaker conclusion although still quite strong, and the IWECS $\mathbf{M}(\mathbb{X})\mathbb{X}$ is no longer a maximal invariant.

In case (ii), and even under merely (5), we have that $\mathbf{M}(\cdot)$ is IWECS relative to $(\mathscr{G}_1, \mathscr{F}_0)$, as per Theorem 3. However, the IWECS $\mathbf{M}(\mathbb{X})\mathbb{X}$ is not a maximal invariant. Consequently, under (9), $\mathbf{M}(\cdot)$ is closely associated with both obtaining an IWECS and obtaining a maximal invariant, although neither solution directly yields the other. Since typical TR matrices do not satisfy (9), special types are required.

Particular constructions of $\mathbf{M}(\cdot)$ satisfying (9) with reference to $\mathscr{G}_0$ and $\mathscr{G}_1$ have been developed and applied to obtain affine invariant multivariate sign and angle tests and affine equivariant multivariate coordinate-wise and spatial medians, in a series of papers by Chaudhuri and Sengupta (1993), Chakraborty and Chaudhuri (1996), and Chakraborty, Chaudhuri, and Oja (1998). Further approaches are treated in Serfling (2010), Ilmonen, Nevalainen, and Oja (2010), and Ilmonen, Oja, and Serfling (2012), covering a range of applications and exploring the formal properties of these transformations. Treatments are carried out in the setting of complex valued independent component analysis by Ilmonen (2013) and in the setting of supervised invariant coordinate selection by Liski, Nordhausen and Oja (2014).

4.2 Some Pertinent Classical Results

Maximal invariant statistics relative to $\mathscr{G}_0$ and $\mathscr{G}_1$ have been treated in detail as early as Lehmann (1959), and those results are pertinent here. In particular, for $\mathbb{X}$ a $d \times n$ matrix of n column d-vectors, and relative to the group $\mathscr{G}_0$, Lehmann (1959) derives the maximal invariant

$$\mathbb{P} = \mathbb{X}^\top(\mathbb{X}\mathbb{X}^\top)^{-1}\mathbb{X},$$

which corresponds to $\mathbf{M}_0(\mathbb{X})\mathbb{X}$ with

$$\mathbf{M}_0(\mathbb{X}) = \mathbb{X}^\top(\mathbb{X}\mathbb{X}^\top)^{-1}.$$

We readily find that $\mathbf{M}_0(\cdot)$ satisfies (9) and hence is both an ICS and a maximal invariant transformation.

Three noteworthy aspects of the Lehmann maximal invariant $\mathbb{P}$ are as follows:

1. $\mathbf{M}_0(\mathbb{X})$ is more directly computed than existing ICS matrices relative to $\mathscr{G}_1$.
2. $\mathbb{P}$ has interesting geometric interpretations as discussed in Lehmann (1959).
3. $\mathbb{P}$ is $n \times n$ rather than $d \times n$ as would be $\mathbf{M}_0(\mathbb{X})\mathbb{X}$ were $\mathbf{M}_0(\cdot)$ a $d \times d$ TR matrix satisfying (9). However, as easily seen, assuming the rows of $\mathbf{M}_0(\mathbb{X})$ are linearly independent as should hold with probability 1, any d rows of $\mathbf{M}_0(\mathbb{X})$ form a $d \times d$ TR matrix $\mathbf{M}_1(\cdot)$, say, also satisfying (9) and thus yield what we might call a *minimal dimension* maximal invariant $\mathbb{P}_0 = \mathbf{M}_1(\mathbb{X})\mathbb{X}$, say.

Remark 3. Note that $\mathbb{P}_0$ (a) serves as an ICS relative to $\mathscr{G}_0$, (b) serves as a maximal invariant relative to $\mathscr{G}_1$ via

$$\mathbf{M}_1(\mathbb{X})(\mathbb{X} - T(\mathbb{X})) = \mathbf{M}_1(\mathbb{X} - T(\mathbb{X}))(\mathbb{X} - T(\mathbb{X})),$$

and (c) serves as an IWECS relative to $(\mathscr{G}_1, \mathscr{F}_0)$. The reduction of the "full" maximal invariant $\mathbb{P}$ to the minimal dimension version $\mathbb{P}_0$ gives up some data, but only what is redundant of that which is retained, as far as a labeling of orbits is concerned. It should be noted that the computational burden posed by $\mathbb{P}$ and $\mathbb{P}_0$ is relatively light. Full investigation of $\mathbb{P}$ and $\mathbb{P}_0$ is deferred to a future study. □

5 Extensions for General $\mathbb{X}$ and General $\mathscr{G}$

In the present paper we have focused on $\mathscr{X} = \mathbb{R}^d$ and $\mathscr{G} = \mathscr{G}_0$ and $\mathscr{G}_1$. However, Definition 1 can immediately be formulated more generally, allowing the data $\mathbb{X}$ to be observations from any space $\mathscr{X}$ and taking $\mathbf{M}$ to be a data based operator on elements of $\mathscr{X}$.

Also, the connections (i) and (ii) of Section 4.1 regarding maximal invariance under (9) have a completely general extension, corresponding to a tightening of (2) in the same way that (9) tightens (5), as follows.

Theorem 4. *For any group $\mathscr{G}$ of transformations on data sets $\mathbb{X}$ from any space $\mathscr{X}$, let $\mathbf{M}(\cdot)$ be such that $\mathbf{M}(\mathbb{X})$ itself belongs to $\mathscr{G}$ for any data set $\mathbb{X}$ and suppose that*

$$\mathbf{M}(g\mathbb{X}) = \mathbf{M}(\mathbb{X})g^{-1}. \tag{10}$$

Then $\mathbf{M}(\mathbb{X})\mathbb{X}$ is both an ICS and a maximal invariant with respect to $\mathscr{G}$.

Proof. Invariance of $\mathbf{M}(\mathbb{X})\mathbb{X}$ follows immediately from (10). Now suppose that $\mathbf{M}(\mathbb{X})\mathbb{X} = \mathbf{M}(\mathbb{X}^*)\mathbb{X}^*$ for two data sets $\mathbb{X}$ and $\mathbb{X}^*$. Then $\mathbb{X}^* = [\mathbf{M}(\mathbb{X}^*)^{-1}\mathbf{M}(\mathbb{X})]\mathbb{X} = g^*\mathbb{X}$, where $g^* = \mathbf{M}(\mathbb{X}^*)^{-1}\mathbf{M}(\mathbb{X}) \in \mathscr{G}$. Hence $\mathbb{X}$ and $\mathbb{X}^*$ lie in the same orbit of $\mathscr{G}$, establishing maximality. □

Statistical inference procedures which are invariant or equivariant with respect to some group $\mathscr{G}$ can be obtained by evaluating suitable preliminary versions at some appropriate functions either of an IWECS or of a maximal invariant, whichever is more convenient. In light of Theorem 4, these constructions and studies may be explored in greater generality than for $\mathscr{X} = \mathbb{R}^d$. For example, a potential application of the IWECS framework arises in the study of similarity between time series with invariance to (various combinations of) the distortions of warping, uniform scaling, offset, amplitude scaling, phase, occlusions, uncertainty and wandering baseline. This and other applications are being pursued in separate investigations.

Acknowledgements The author thanks G. L. Thompson for encouragement and two anonymous reviewers for comments and suggestions helpful to improving the paper. Also, support by NSF Grant DMS-1106691 is greatly appreciated.

References

[1] 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.
[2] Chakraborty, B., Chaudhuri, P. and Oja, H. (1998). Operating transformation retransformation on spatial median and angle test. *Statistica Sinica* **8** 767–784.
[3] 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.
[4] Dang, X. and Serfling, R. (2010). Nonparametric depth-based multivariate outlier identifiers, and masking robustness properties. *Journal of Statistical Planning and Inference* **140** 198–213.
[5] Ilmonen, P. (2013). On asymptotic properties of the scatter matrix based estimates for complex valued independent component analysis. *Statistics and Probability Letters* **83** 1219–1226.
[6] Ilmonen, P., Nevalainen, J., and Oja, H. (2010). Characteristics of multivariate distributions and the invariant coordinate system. *Statistics and Probability Letters* **80** 1844–1853.
[7] Ilmonen, P., Oja, H., and Serfling, R. (2012). On invariant coordinate system (ICS) functionals. *International Statistical Review* **80** 93–110.
[8] Lehmann, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.
[9] Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*, 3rd Edition. Springer, New York.

[10] Liski, E., Nordhausen, K. and Oja, H. (2014). On supervised coordinate selection. *Statistics* **48** 711–731.

[11] Nordhausen, K. (2008). On Invariant Coordinate Selection and Nonparametric Analysis of Multivariate Data. Ph.D. Dissertation, University of Tampere, Finland.

[12] Oja, H. (2010). *Multivariate Nonparametric Methods with R: An approach based on spatial signs and ranks*. Springer.

[13] 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.

[14] Serfling, R. and Mazumder, S. (2013). Computationally easy outlier detection via projection pursuit with finitely many directions. *Journal of Nonparametric Statistics* **25** 447–461.

[15] 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.