WORKING PAPERS

N° 1320

March 2022

"Detecting outliers in compositional data using Invariant Coordinate Selection"

Anne Ruiz-Gazen, Christine Thomas-Agnan, Thibault Laurent and Camille Mondon

Toulouse School of Economics

Detecting outliers in compositional data using Invariant Coordinate Selection

Anne Ruiz-Gazen, Christine Thomas-Agnan, Thibault Laurent,
and Camille Mondon

Abstract Invariant Coordinate Selection (ICS) is a multivariate statistical method introduced by Tyler et al. (2009) and based on the simultaneous diagonalization of two scatter matrices. A model based approach of ICS, called Invariant Coordinate Analysis, has already been adapted for compositional data in Muehlmann et al. (2021). In a model free context, ICS is also helpful at identifying outliers (Nordhausen and Ruiz-Gazen, 2022). We propose to develop a version of ICS for outlier detection in compositional data. This version is first introduced in coordinate space for a specific choice of ilr coordinate system associated to a contrast matrix and follows the outlier detection procedure proposed by Archimbaud et al. (2018a). We then show that the procedure is independent of the choice of contrast matrix and can be defined directly in the simplex. To do so, we first establish some properties of the set of matrices satisfying the zero-sum property and introduce a simplex definition of the Mahalanobis distance and the one-step M-estimators class of scatter matrices. We also need to define the family of elliptical distributions in the simplex. We then show how to interpret the results directly in the simplex using two artificial datasets and a real dataset of market shares in the automobile industry.
Keywords: compositional data, invariant coordinate selection, outlier detection,

Anne Ruiz-Gazen
Toulouse School of Economics, University of Toulouse 1 Capitole, e-mail: anne.ruiz-gazen@tse-fr.eu

Christine Thomas-Agnan
Toulouse School of Economics, University of Toulouse 1 Capitole, e-mail: christine.thomas@tse-fr.eu

Thibault Laurent
Toulouse School of Economics, CNRS e-mail: thibault.laurent@tse-fr.eu

Camille Mondon
ENS, e-mail: camille.mondon@ens.fr

1 Introduction

Compositional data are by nature multivariate. Indeed, vectors with positive components are considered as compositional data when the interest lies in the relative information between their components: this last fact implies that they can be represented by a unique element in a simplex by dividing the components by their sum. Classical statistical techniques need to be adapted to deal with these constraints (positivity, sum equal to one). A common approach consists in transforming the data using the centered log-ratio (clr) or the isometric log-ratio (ilr) transformations (see Egozcue et al. (2011)), and apply standard techniques in this coordinate space. Filzmoser et al. (2012)) propose to use the ilr transformation and detect outliers with the usual or the robust version of the Mahalanobis distance. Because of the affine invariance property of the Mahalanobis distance, the authors notice that the identified outliers do not depend on the choice of the ilr transformation. Moreoever, they propose some graphical tools in the coordinate space based on robust Principal Component Analysis (PCA) and biplots representation in order to interpret the outliers. Their interpretation is only done in coordinate space. In the present work, we consider adapting the Invariant Coordinate Selection (ICS) technique for outlier detection to compositional data. ICS is a multivariate statistics method based on the joint diagonalization of two scatter matrices and aimed at detecting interesting features in multivariate data sets such as outliers or clusters (see, e.g., Tyler et al. (2009) and Archimbaud et al. (2018a)). Compared to the Mahalanobis distance criterion, ICS includes a dimension reduction step. Compared to PCA, the components of ICS are invariant under affine transformations. We first propose to introduce ICS in coordinate space using an ilr transformation. Following Archimbaud et al. (2018a), we focus on the case of a small proportion of outliers and use the invariant components associated with the largest eigenvalues of the joint diagonalization of two particular scatter matrices. As with the Mahalanobis distance, the identification of outliers with ICS does not depend on the choice of the ilr (see also Muehlmann et al. (2021)). In order to go beyond the coordinate space and interpret the outliers in the simplex, we introduce new algebra tools and define eigen-elements of endomorphisms of the simplex. We also introduce a class of one-step M-scatter estimators in the simplex. Thanks to these tools we are able to write a reconstruction formula of the data in the simplex that decompose the data in a proper way for outlier identification and interpretation using ternary diagrams. In Section 2, we recall some facts about the ICS method and its application to outlier detection. Section 3 is a reminder about compositional data analysis. In Section 4, we develop some tools necessary for Section 5. First comes some properties of the algebra of $D \times D$ matrices with the zero-sum property: in particular, their rank, their inverses, their eigen-elements. Then Section 4.2 defines one-step M-scatter functionals for simplex-valued random variables together with an adapted version of Mahalanobis distance. Finally, Section 4.3 introduces the family of elliptical distributions in the simplex. Section 5 first introduces ICS in coordinate space, then reformulates ICS directly in the simplex. In subsection 5.3, we present a formula for reconstructing the data from ICS in coordinate space and in the simplex. Section 6 is dedicated to three applications, with two toy datasets (with small and large dimensions) and a real marketing application from the automobile industry.

2 Reminder about ICS and outlier detection

Invariant Coordinate Selection is a multivariate statistical method based on the simultaneous diagonalization of two scatter matrices. As detailed in Nordhausen and Ruiz-Gazen (2022), the method belongs to a large family of multivariate statistics methods and is useful in particular for outlier detection as described below.

2.1 Scatter matrices

The family of scatter matrices generalizes the notion of covariance matrix (see Tyler et al., 2009; Nordhausen and Tyler, 2015, among others) and it has the following functional definition. For a p-dimensional vector $\mathbf{X}$ with distribution function $F_{\mathbf{X}}$, a functional $\mathbf{S}(F_{\mathbf{X}})$, also denoted by $\mathbf{S}(\mathbf{X})$, is called a scatter functional if it is a $p \times p$ symmetric positive definite and affine equivariant matrix. Note that in Nordhausen and Tyler (2015), the definition is less stringent than that in Tyler et al. (2009), and assumes that a scatter matrix is only semipositive definite. We recall that an affine equivariant matrix $\mathbf{S}(\mathbf{X})$ is such that

$$\mathbf{S}(\mathbf{AX} + \mathbf{b}) = \mathbf{AS}(\mathbf{X})\mathbf{A}^T,$$

where ^{T} denotes the transpose operator, $\mathbf{A}$ is a full rank $p \times p$ matrix and $\mathbf{b}$ a p-vector.

For a p-variate dataset $\mathbf{X}_n = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^T$, the empirical version $\mathbf{S}(\mathbf{X}_n)$ of a scatter functional is the scatter functional $\mathbf{S}(F_n)$ where F_n is the empirical distribution. A scatter matrix estimator is a $p \times p$ symmetric positive definite and affine equivariant matrix such that

$$\mathbf{S}(\mathbf{X}_n\mathbf{A} + \mathbf{1}_n\mathbf{b}^T) = \mathbf{A}^T\mathbf{S}(\mathbf{X}_n)\mathbf{A},$$

where $\mathbf{A}$ is a full rank $p \times p$ matrix, $\mathbf{b}$ a p-vector and $\mathbf{1}_n$ an n-dimensional vector of ones.

There exist many scatter matrices as detailed for example in Tyler et al. (2009). The most well-known scatter matrix is the covariance matrix. As many other scatter matrices, the covariance involves the mean which is an affine equivariant location estimator. We recall that an affine equivariant location estimator $\mathbf{T}$ is such that:

$$\mathbf{T}(\mathbf{AX} + \mathbf{b}) = \mathbf{AT}(\mathbf{X}) + \mathbf{b},$$

for the functional version and

$$\mathbf{T}(\mathbf{X}_n\mathbf{A}^T + \mathbf{1}_n\mathbf{b}^T) = \mathbf{A}\mathbf{T}(\mathbf{X}_n) + \mathbf{b},$$

for the empirical version where $\mathbf{A}$ is a full rank $p \times p$ matrix and $\mathbf{b}$ a p-vector.

A general class of scatter matrices is the class of one-step M-estimators with a functional defined by:

$$\mathbf{COV}_w(\mathbf{X}) = \mathbf{E}\left[w(M^2(\mathbf{X}))(\mathbf{X} - \mathbf{E}(\mathbf{X}))(\mathbf{X} - \mathbf{E}(\mathbf{X}))^T\right],$$

where w is a non-negative and continuous weight function and

$$M^2(\mathbf{X}) = (\mathbf{X} - \mathbf{E}(\mathbf{X}))^T \mathbf{COV}(\mathbf{X})^{-1} (\mathbf{X} - \mathbf{E}(\mathbf{X})) \tag{1}$$

is the square Mahalanobis distance with $\mathbf{E}(\mathbf{X})$ the expectation of $\mathbf{X}$ and $\mathbf{COV}(\mathbf{X})$ its covariance matrix. The sample version of one-step M-estimators is:

$$\mathbf{COV}_w(\mathbf{X}_n) = \frac{1}{n}\sum_{i=1}^{n} w(M^2(\mathbf{x}_i))(\mathbf{x}_i - \bar{\mathbf{x}}_n)(\mathbf{x}_i - \bar{\mathbf{x}}_n)^T,$$

where $\bar{\mathbf{x}}_n = 1/n \sum_{i=1}^{n} \mathbf{x}_i$ is the empirical mean and

$$M^2(\mathbf{x}_i) = (\mathbf{x}_i - \bar{\mathbf{x}}_n)^T \mathbf{COV}(\mathbf{X}_n)^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}_n)$$

is the empirical version of the square Mahalanobis distance.

Note that the covariance matrix $\mathbf{COV}$ is obtained with $w(d) = 1$, while the fourth-moment based estimator $\mathbf{COV}_4$ is obtained with $w(d) = d/(p+2)$ which is widely used in the blind source separation literature (see, e.g., Theis and Inouye, 2006; Nordhausen and Virta, 2019) but also in the context of outlier detection (see Archimbaud et al., 2018a).

For elliptical distributions with second moments, scatter functionals are all proportional to the covariance matrix (see, e.g., Bilodeau and Brenner, 2008). We recall that an elliptical distribution is obtained as an affine transformation of a spherical distribution which is a distribution invariant by orthogonal transformation. Multivariate normal and Student distribution belong to this family of distributions.

2.2 ICS principle

Let $\mathbf{S}_1(\mathbf{X})$ and $\mathbf{S}_2(\mathbf{X})$ be two scatter functionals. ICS consists of the simultaneous diagonalization of $\mathbf{S}_1(\mathbf{X})$ and $\mathbf{S}_2(\mathbf{X})$. If the random vector $\mathbf{X}$ follows an elliptical distribution, the two scatter matrices will be proportional and the result will be useless. However, as mentioned in Tyler et al. (2009), comparing two different scatter functionals may help to reveal interesting departures from an elliptical distribution. This is the case in particular for anomaly detection. The method searches for a $p \times p$ matrix $\mathbf{H}(\mathbf{X})$ and a diagonal matrix $\mathbf{\Lambda}(\mathbf{X})$ such that:

$$\mathbf{H}(\mathbf{X})^T\mathbf{S}_1(\mathbf{X})\mathbf{H}(\mathbf{X}) = \mathbf{I}_p \quad \text{and} \quad \mathbf{H}(\mathbf{X})^T\mathbf{S}_2(\mathbf{X})\mathbf{H}(\mathbf{X}) = \mathbf{\Lambda}(\mathbf{X}), \tag{2}$$

where $\mathbf{I}_p$ denotes the $p \times p$ identity matrix. The matrix $\mathbf{\Lambda}(\mathbf{X})$ contains the eigenvalues of $\mathbf{S}_1(\mathbf{X})^{-1}\mathbf{S}_2(\mathbf{X})$ in decreasing order, while the columns of the matrix $\mathbf{H}(\mathbf{X}) = (\mathbf{h}_1, \ldots, \mathbf{h}_p)$ contain the corresponding eigenvectors such that:

$$\mathbf{S}_2(\mathbf{X})\mathbf{H}(\mathbf{X}) = \mathbf{S}_1(\mathbf{X})\mathbf{H}(\mathbf{X})\mathbf{\Lambda}(\mathbf{X})$$

$$\text{or else} \quad \mathbf{S}_1(\mathbf{X})^{-1}\mathbf{S}_2(\mathbf{X})\mathbf{H}(\mathbf{X}) = \mathbf{H}(\mathbf{X})\mathbf{\Lambda}(\mathbf{X}).$$

These eigenvalues and eigenvectors can also be derived through the spectral decomposition of the following symmetric matrix:

$$\mathbf{S}_1(\mathbf{X})^{-1/2}\mathbf{S}_2(\mathbf{X})\mathbf{S}_1(\mathbf{X})^{-1/2} = \mathbf{U}(\mathbf{X})\mathbf{\Lambda}(\mathbf{X})\mathbf{U}(\mathbf{X})^T. \tag{3}$$

with $\mathbf{U}(\mathbf{X})$ a $p \times p$ orthogonal matrix and the same eigenvalues in the diagonal matrix $\mathbf{\Lambda}(\mathbf{X})$. We have:

$$\mathbf{H}(\mathbf{X}) = \mathbf{S}_1(\mathbf{X})^{-1/2}\mathbf{U}(\mathbf{X}).$$

We also have:

$$\mathbf{H}(\mathbf{X})\mathbf{H}(\mathbf{X})^T = \mathbf{S}_1(\mathbf{X})^{-1} \text{ and } \mathbf{H}(\mathbf{X})^{-1} = \mathbf{U}(\mathbf{X})^T\mathbf{S}_1(\mathbf{X})^{1/2}.$$

Tyler et al. (2009) give an interesting interpretation of the eigenvalues $\lambda_1, \ldots, \lambda_p$ in terms of kurtosis. Using the optimality property of eigen-elements, we have that $\mathbf{h}_1$ maximizes the ratio:

$$\frac{\mathbf{h}^T\mathbf{S}_2(\mathbf{X})\mathbf{h}}{\mathbf{h}^T\mathbf{S}_1(\mathbf{X})\mathbf{h}}$$

over all possible vectors $\mathbf{h}$ in $\mathbb{R}^p$ and that λ_1 is equal to the maximum. This ratio of two scale measures can be viewed as a generalized measure of kurtosis and λ_1 can thus be interpreted as a maximum kurtosis. The other eigenvalues and eigenvectors can be defined is a similar way by maximizing the same ratio over vectors $\mathbf{h}$ that verify additional orthogonality conditions (see Tyler et al. (2009) for details).

Using any affine equivariant location estimator $\mathbf{T}(\mathbf{X})$, the ICS scores $\mathbf{Z} = (z_1, \ldots, z_p)^T$ are defined by:

$$\mathbf{Z} = \mathbf{H}(\mathbf{X})^T(\mathbf{X} - \mathbf{T}(\mathbf{X})), \tag{4}$$

or equivalently by $z_k =< \mathbf{h}_k, \mathbf{X} - \mathbf{T}(\mathbf{X}) >$ where $< .,. >$ denotes the usual scalar product. The scores define the affine invariant coordinates. The square Euclidian norm of these coordinates is given by:

$$\begin{aligned}\mathbf{Z}^T\mathbf{Z} &= (\mathbf{X} - \mathbf{T}(\mathbf{X}))^T\mathbf{H}(\mathbf{X})\mathbf{H}(\mathbf{X})^T(\mathbf{X} - \mathbf{T}(\mathbf{X})) \\ &= (\mathbf{X} - \mathbf{T}(\mathbf{X}))^T\mathbf{S}_1(\mathbf{X})^{-1}(\mathbf{X} - \mathbf{T}(\mathbf{X}))\end{aligned}$$

The last expression is a generalization of the Mahalanobis distance (1) of $\mathbf{X}$ with the location parameter $\mathbf{T}(\mathbf{X})$ (instead of $\mathbf{E}(\mathbf{X})$) and with respect to the scatter matrix

$\mathbf{S}_1(\mathbf{X})$ (instead of $\mathbf{COV}(\mathbf{X})$). In the special case where $\mathbf{T}(\mathbf{X}) = \mathbf{E}(\mathbf{X})$ and $\mathbf{S}_1(\mathbf{X}) = \mathbf{COV}(\mathbf{X})$, we have:

$$\mathbf{Z}^T\mathbf{Z} = \sum_{k=1}^{p} z_k^2 = M^2(\mathbf{X}). \tag{5}$$

The empirical version of ICS consists of the joint diagonalization of a scatter pair of estimators $\mathbf{S}_1(\mathbf{X}_n)$ and $\mathbf{S}_2(\mathbf{X}_n)$. For a $p \times p$ matrix $\mathbf{H}(\mathbf{X}_n)$ and a diagonal matrix $\mathbf{\Lambda}(\mathbf{X}_n)$ we have:

$$\mathbf{H}(\mathbf{X}_n)^T\mathbf{S}_1(\mathbf{X}_n)\mathbf{H}(\mathbf{X}_n) = \mathbf{I}_p \quad \text{and} \quad \mathbf{H}(\mathbf{X}_n)^T\mathbf{S}_2(\mathbf{X}_n)\mathbf{H}(\mathbf{X}_n) = \mathbf{\Lambda}(\mathbf{X}_n).$$

Using any affine equivariant location estimator $\mathbf{T}(\mathbf{X}_n)$, the ICS scores are given by:

$$\mathbf{Z}_n = (\mathbf{z}_1, \ldots, \mathbf{z}_n)^T = (\mathbf{X}_n - \mathbf{1}_n\mathbf{T}(\mathbf{X}_n)^T)\mathbf{H}(\mathbf{X}_n) \tag{6}$$

and are affine invariant. Similar to (5), if $T(\mathbf{X}_n) = \bar{\mathbf{x}}_n$ and $\mathbf{S}_1(\mathbf{X}_n) = \mathbf{COV}(\mathbf{X}_n)$, we have:

$$M^2(\mathbf{x}_i) = \mathbf{z}_i^T\mathbf{z}_i.$$

2.3 ICS for outlier detection

As already stated in Tyler et al. (2009), one possible application of ICS is outlier detection. The Mahalanobis distance is a well-known tool to detect outliers (see Rousseeuw and Van Zomeren (1990)) but it does not offer the possibility of dimension reduction. ICS gives the possibility of selecting components that are helpful in detecting anomalies (see Archimbaud et al. (2018a) for details). In the case of a small proportion of outliers, the theoretical properties of ICS (see Archimbaud et al. (2018a) for details) lead us to only focus on the invariant components associated with the largest kurtosis and thus the largest eigenvalues. In this context, Archimbaud et al. (2018a) show that the scatter pair $\mathbf{S}_1(\mathbf{X}) = \mathbf{COV}(\mathbf{X})$ and $\mathbf{S}_2(\mathbf{X}) = \mathbf{COV}_4(\mathbf{X})$ is not only simple and fast to compute but also effective in detecting outliers when compared to other pairs that involve robust scatter estimators. Archimbaud et al. (2018a) propose different automatic procedures for invariant components selection based on hypothesis testing. Details can be found in Archimbaud et al. (2018a) but in short the idea is to test sequentially the normality of each of the invariant components using some classical tests like the D'Agostino test. After selecting k invariant components among p, the last step of the procedure is the outlier identification. Let us consider the empirical version of ICS. For each observation $i = 1, \ldots, n$, the square "ICS distance" is the square Euclidian norm in the invariant coordinate system accounting for the k first coordinates:

$$(\text{ICS distance})_{i,k}^2 = \sum_{j=1}^{k} \left(z_i^j\right)^2, \tag{7}$$

where z_i^j denotes the jth coordinate of the score $\mathbf{z}_i$. In Archimbaud et al. (2018a), an observation is flagged as an outlier when its ICS distance using k components is larger than a cutoff based on Monte Carlo simulations from the standard Gaussian distribution. Given a data dimension, a scatter pair and a number k of selected components, many Gaussian samples are generated and the ICS distances are computed. A cutoff is derived for a fixed level γ as the $1-\gamma$ quantile of these distances. The whole ICS procedure for outlier detection is available in the `R` package `ICSOutlier` described in Archimbaud et al. (2018b) and used in Section 6 below.

3 Reminder about compositional data analysis

A D-composition $\mathbf{u}$ is a vector of D parts (or shares) of some whole which carries relative information. There exists a unique representation of this vector in the unit simplex space

$$\mathbf{S}^D = \left\{ \mathbf{u} = (u_1, \ldots, u_D)^T : u_m > 0, m = 1, ..., D; \sum_{m=1}^{D} u_m = 1 \right\}.$$

For any vector $\mathbf{w} \in \mathbb{R}^{+D}$, its representer in the simplex is obtained by the closure operation

$$C(\mathbf{w}) = \left(\frac{w_1}{\sum_{m=1}^{D} w_m}, \cdots, \frac{w_D}{\sum_{m=1}^{D} w_m} \right).$$

The following operations endow the unit simplex with a vector space structure

1. $\oplus$ is the perturbation operation, corresponding to the addition in $\mathbb{R}^D$:

$$\text{For} \quad \mathbf{u}, \mathbf{v} \in \mathbf{S}^D, \mathbf{u} \oplus \mathbf{v} = C(u_1 v_1, \ldots, u_D v_D),$$

2. $\odot$ is the power operation, corresponding to the scalar multiplication in $\mathbb{R}^D$:

$$\text{For} \quad \lambda \in \mathbb{R}, \mathbf{u} \in \mathbf{S}^D \quad \lambda \odot \mathbf{u} = C(u_1^\lambda, \ldots, u_D^\lambda).$$

Compositional data analysis uses log-ratio transformations such as the centered log-ratio (clr) and the isometric log-ratio (ilr) transformations. The clr transformation of a vector $\mathbf{u} \in \mathbf{S}^D$ is defined by

$$\text{clr}(\mathbf{u}) = \mathbf{G}_D \ln \mathbf{u}$$

where $\mathbf{G}_D = \mathbf{I}_D - \frac{1}{D} \mathbf{1}_D \mathbf{1}_D{}^T$, $\mathbf{I}_D$ is a $D \times D$ identity matrix, $\mathbf{1}_D$ is the D-vector of ones and where the logarithm of $\mathbf{u} \in \mathbf{S}^D$ is understood componentwise.

For a vector $\mathbf{u}$ in the orthogonal space $\mathbf{1}_D^\perp$ (orthogonality with respect to the usual scalar product of $\mathbb{R}^D$), the inverse clr transformation is defined by

$$\text{clr}^{-1}(\mathbf{u}) = C(\exp(\mathbf{u})).$$

The simplex $\mathcal{S}^D$ of dimension $D-1$ can be equipped with the Aitchison scalar product

$$< \mathbf{u}, \mathbf{v} >_A = < \text{clr}(\mathbf{u}), \text{clr}(\mathbf{v}) >,$$

where the right hand side scalar product is the usual scalar product in $\mathbb{R}^D$.

For any given orthonormal basis $(\mathbf{e}_1, \cdots, \mathbf{e}_{D-1})$ of $\mathbf{S}^D$, orthonormality being understood with respect to the Aitchison scalar product here, one can define a so-called contrast matrix $\mathbf{V}$ of dimension $D \times (D-1)$ (e.g. Pawlowsky-Glahn et al., 2015) given by $\mathbf{V} = \text{clr}(\mathbf{e}_1, \cdots, \mathbf{e}_{D-1})$, where clr is understood columnwise. To each such matrix $\mathbf{V}$ is associated an isometric log-ratio transformation by:

$$\text{ilr}_V(\mathbf{u}) = \mathbf{V}^T \ln(\mathbf{u}).$$

The inverse transformation is given by:

$$\mathbf{u} = \text{ilr}_V^{-1}(\mathbf{u}^*) = \mathcal{C}(\exp(\mathbf{V}\mathbf{u}^*)).$$

The link between the ilr and clr transformations is $\text{clr}(\mathbf{u}) = \mathbf{V}\text{ilr}_V(\mathbf{u})$.

4 Multivariate tools for compositional data

4.1 Algebra of endomorphisms of the simplex and eigendecomposition

Let $\mathcal{A}$ be the set of $D \times D$ matrices such that $\mathbf{A}\mathbf{1}_D = \mathbf{0}_D$ and $\mathbf{A}^T\mathbf{1}_D = \mathbf{0}_D$ where $\mathbf{0}_D$ denotes the D-dimensional column vector of zeros: this condition is called the zero-sum property. Egozcue et al. (2011) define endomorphisms of the simplex using the ilr transformation and prove that their corresponding matrix belongs to $\mathcal{A}$. The linearity here refers to the vector space structure of the simplex based on the perturbation and powering operations. Let us introduce an equivalent formulation based on the clr transformation: for $\mathbf{u} \in \mathbf{S}^D$ and $\mathbf{A} \in \mathcal{A}$, endomorphisms of the simplex are defined by maps $\mathbf{u} \mapsto \mathbf{A} \boxdot \mathbf{u} := \text{clr}^{-1}(\mathbf{A}\text{clr}(u))$.

The composition of endomorphisms corresponds to the ordinary matrix product since it is clear that $\mathbf{A} \boxdot (\mathbf{B} \boxdot \mathbf{u}) = \mathbf{AB} \boxdot \mathbf{u}$ and therefore $\mathcal{A}$ is an algebra with neutral element $\mathbf{G}_D$. We are now going to extend the definition of the ilr transformation to matrices of $\mathcal{A}$.

Theorem 1 *Let* $\mathbf{V}$ *be a* $D \times (D-1)$ *contrast matrix and let* $\mathbf{P}_V$ *be the* $D \times D$ *block matrix* $[\mathbf{V}\ \frac{1}{\sqrt{D}}\mathbf{1}_D]$. *For a* $D \times D$ *matrix* $\mathbf{A} \in \mathcal{A}$, *the* $(D-1) \times (D-1)$ *matrix* $\mathbf{A}^* := ilr_V(\mathbf{A}) = \mathbf{V}^T\mathbf{A}\mathbf{V}$ *is such that* $\mathbf{A} = \mathbf{V}\mathbf{A}^*\mathbf{V}^T = \mathbf{P}_V \begin{pmatrix} \mathbf{A}^* & 0 \\ 0 & 0 \end{pmatrix} \mathbf{P}_V^T$ *and satisfies the following properties*

1. *the rank of* $\mathbf{A}$ *is equal to the rank of* $ilr_V(\mathbf{A})$
2. *if* $ilr_V(\mathbf{A})$ *is invertible then* $\mathbf{A}$ *is invertible in* $\mathcal{A}$ *and we have the following expressions for its Moore-Penrose pseudo-inverse*

$$\mathbf{A}^{-1} = (\mathbf{A} + \frac{1}{D}\mathbf{1}_D\mathbf{1}_D^T)^{-1} - \frac{1}{D}\mathbf{1}_D\mathbf{1}_D^T = \mathbf{V}(\mathbf{V}^T\mathbf{A}\mathbf{V})^{-1}\mathbf{V}^T = \mathbf{P}_V \begin{pmatrix} \mathbf{A}^{*-1} & 0 \\ 0 & 0 \end{pmatrix} \mathbf{P}_V^T.$$

3. *$ilr_V(\mathbf{AB}) = ilr_V(\mathbf{A})ilr_V(\mathbf{B})$. If $\mathbf{A}$ is invertible, then $ilr_V(\mathbf{A}^{-1}) = (ilr_V(\mathbf{A}))^{-1}$. If $(ilr_V(\mathbf{A}))^{1/2}$ exists then $ilr_V(\mathbf{A}^{1/2}) = (ilr_V(\mathbf{A}))^{1/2}$.*

The matrix $\mathrm{ilr}_V(\mathbf{A})$ is simply the matrix corresponding to $\mathbf{A}$ in coordinate space when the coordinates are defined by ilr_V. We also extend the definition of the clr transformations to matrices.

Theorem 2 *For a $D \times D$ matrix $\mathbf{B}$, let us define its clr transformation by*

$$clr(\mathbf{B}) = \mathbf{G}_D\mathbf{B}\mathbf{G}_D. \tag{8}$$

We then have the following properties

1. *if $\mathbf{A} \in \mathcal{A}$, then $clr(\mathbf{A}) = \mathbf{A}$*
2. *if $\mathbf{B} \notin \mathcal{A}$, then $clr(\mathbf{B}) \in \mathcal{A}$ and for any $\mathbf{x} \in \mathcal{S}^{\mathcal{D}}$*

$$\mathbf{B} \boxdot \mathbf{x} := clr^{-1}(clr(\mathbf{B})clr(x)) = clr(\mathbf{B}) \boxdot \mathbf{x} \tag{9}$$

3. *if $\mathbf{B} \notin \mathcal{A}$, then the unique element $\mathbf{A} \in \mathcal{A}$ such that $ilr_V(\mathbf{A}) = ilr_V(\mathbf{B})$ is $\mathbf{A} = clr(\mathbf{B})$.*
4. *for any contrast matrix $\mathbf{V}$ and any $\mathbf{A} \in \mathcal{A}$ we have $clr(\mathbf{A}) = \mathbf{V}ilr_V(\mathbf{A})\mathbf{V}^T$*

Note that the matrix product $\boxdot$ can be defined even when the matrix $\mathbf{B}$ does not belong to $\mathcal{A}$ but in that case it is not linear. Note also that the ilr and clr transformations preserve symmetry.

The next proposition links the eigen-elements of $\mathbf{A}$ to those of $\mathrm{ilr}(\mathbf{A})$. Let us first define the notion of $\mathcal{A}$-diagonalizable for a matrix of $\mathcal{A}$.

Definition 1 A matrix $\mathbf{A} \in \mathcal{A}$ is said $\mathcal{A}$-diagonalizable if there exists a basis $\mathbf{e}_1, \ldots, \mathbf{e}_{D-1}$ of $\mathcal{S}^{\mathcal{D}}$ and $D-1$ reals λ_j $(j = 1, \ldots D-1)$ such that

$$\mathbf{A} \boxdot \mathbf{e}_j = \lambda_j \odot \mathbf{e}_j \quad \forall j = 1, \ldots D-1 \tag{10}$$

We will say that $\mathbf{e}_j$ is an $\mathcal{A}$-eigenvector of $\mathbf{A}$. It is clear that $\mathrm{clr}(\mathbf{e}_j)$ is then an eigenvector of $\mathrm{clr}(\mathbf{A}) = \mathbf{A}$, and that for any contrast matrix $\mathbf{V}$, $\mathrm{ilr}_V(\mathbf{e}_j)$ is an eigenvector of $\mathrm{ilr}_V(\mathbf{A})$. Note that $\mathbf{1}_D$ is an eigenvector of $\mathcal{A}$ associated to the eigenvalue 0. It is natural to say that a matrix $\mathbf{A} \in \mathcal{A}$ is diagonal in a given basis $\mathbf{e}_1, \ldots, \mathbf{e}_{D-1}$ of $\mathcal{S}^{\mathcal{D}}$ if equation (10) is satisfied for these vectors.

Theorem 3 *Let $\mathbf{V}$ be a $D \times (D-1)$ contrast matrix. For a $D \times D$ matrix $\mathbf{A} \in \mathcal{A}$, we have the following properties:*

1. *if $\mathbf{e}_j^* \in \mathbb{R}^{D-1}$ is an eigenvector of $ilr_V(\mathbf{A})$, then $\mathbf{e}_j = ilr^{-1}(\mathbf{e}_j^*) \in \mathcal{S}^{\mathcal{D}}$ is an $\mathcal{A}$-eigenvector of $\mathbf{A}$ and $\mathbf{w}_j = clr(\mathbf{e}_j) \in \mathbb{R}^D$ an eigenvector of $\mathbf{A}$.*
2. *the set of eigenvalues of $\mathbf{A}$ contain the eigenvalue 0. The other $D-1$ eigenvalues of $\mathbf{A}$ coincide with the eigenvalues of $ilr_V(\mathbf{A})$ for any contrast matrix $\mathbf{V}$.*

3. $ilr_V(\mathbf{A})$ is diagonalizable if and only if $\mathbf{A}$ is diagonalizable, and if and only if $\mathbf{A}$ is $\mathcal{A}$-diagonalizable.

All symmetric matrices in $\mathcal{A}$ are $\mathcal{A}$-diagonalizable. Note that the vectors $\mathbf{e}_j = \text{clr}^{-1}(\mathbf{e}_j^*)$ are independent of the contrast matrix $\mathbf{V}$. Let $\mathbf{A}$ be a symmetric matrix of $\mathcal{A}$. Since the vector $\mathbf{1}_D$ is an eigenvector of $\mathbf{A}$, $\mathbf{A}$ cannot be diagonal in the canonical basis of $\mathbb{R}^D$ but it can be diagonal in a basis obtained by completing $\mathbf{w}_D = \frac{1}{D}\mathbf{1}_D$ with $D-1$ orthogonal eigenvectors in $\mathbf{1}_D^{\perp}$, say $\mathbf{w}_1, \ldots \mathbf{w}_{D-1}$. Then $\mathbf{e}_j = \text{clr}^{-1}(\mathbf{w}_j) \in \mathcal{S}^{\mathcal{D}}$ $(j = 1, \ldots D-1)$, is an orthonormal basis of $\mathcal{S}^{\mathcal{D}}$ for the Aitchison metric since $< \mathbf{e}_i, \mathbf{e}_j >_A = < \mathbf{w}_i, \mathbf{w}_j >_E = \delta_{ij}$ where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise, and these vectors are $\mathcal{A}$-eigenvectors of $\mathbf{A}$. If $\mathbf{W} = [\mathbf{w}_1 \ldots \mathbf{w}_{D-1}]$ is the corresponding contrast matrix, then $\text{ilr}_W(\mathbf{A})_{ij} = \mathbf{w}_i^T \mathbf{A} \mathbf{w}_j = \lambda_j \mathbf{w}_i^T \mathbf{w}_j = \lambda_i \delta_{ij}$ which shows that $\text{ilr}_{\mathbf{W}}(\mathbf{A})$ is the $(D-1) \times (D-1)$ diagonal matrix $\mathbf{\Lambda}$ with the λ_i as diagonal elements. Then using Theorem 1, we can write that $\mathbf{A} = \mathbf{P}_W \begin{pmatrix} \mathbf{\Lambda} & 0 \\ 0 & 0 \end{pmatrix} \mathbf{P}_W^T$ showing that $\mathbf{A}$ is similar to the diagonal matrix $\begin{pmatrix} \mathbf{\Lambda} & \mathbf{0} \\ \mathbf{0} & 0 \end{pmatrix}$. This last result gives us the general form of diagonal matrices of $\mathcal{A}$ with the corresponding spectral representation $\mathbf{A} = \sum_{i=1}^{D-1} \lambda_i \mathbf{w}_i \mathbf{w}_i^T$.

4.2 One-step M-scatter functionals of a compositional random vector

For a simplex valued random vector $\mathbf{X}$ (see Pawlowsky-Glahn et al. (2015)) , let us recall the following definition of expectation

$$\mathbf{E}^{\oplus}\mathbf{X} := \text{clr}^{-1}(\mathbf{E}\text{clr}\mathbf{X})$$

and the following definition of the (clr-)covariance matrix $\mathbf{COV}^{\oplus}\mathbf{X}$ (see Aitchison (1982)) given by the $D \times D$ matrix

$$\mathbf{COV}^{\oplus}\mathbf{X} := \mathbf{COV}(\text{clr}\mathbf{X}).$$

Using the same principles, let us now introduce a simplex adapted definition of the square Mahalanobis distance as being the square Mahalanobis distance in the usual sense of the clr coordinates of $\mathbf{X}$

$$M^2(\mathbf{X}) = (\text{clr}\mathbf{X} - \mathbf{E}^{\oplus}\mathbf{X})^T (\mathbf{COV}^{\oplus}\mathbf{X})^{-1} (\text{clr}\mathbf{X} - \mathbf{E}^{\oplus}\mathbf{X}).$$

In the same line, let us define the following one-step M-scatter matrix of a simplex valued random vector as the corresponding scatter of its clr coordinates

$$\mathbf{COV}_w^{\oplus}\mathbf{X} := \mathbf{COV}_w(\text{clr}\mathbf{X}) = \mathbf{E}[w(M(\mathbf{X}))(\text{clr}\mathbf{X} - \mathbf{E}\text{clr}\mathbf{X})(\text{clr}\mathbf{X} - \mathbf{E}\text{clr}\mathbf{X})^T]$$

For $w(d) = d/(D+2)$, we get the fourth-moment based scatter matrix $\mathbf{COV}_4^{\oplus}\mathbf{X}$:

$$\mathbf{COV}_4^{\oplus}\mathbf{X} := \mathbf{COV}_4(\text{clr}\mathbf{X}) = \frac{1}{D+2}\mathbf{E}[M^2(\mathbf{X})(\text{clr}\mathbf{X} - \mathbf{E}\text{clr}\mathbf{X})(\text{clr}\mathbf{X} - \mathbf{E}\text{clr}\mathbf{X})^T]$$

All these characteristics can also be expressed using the ilr coordinates associated to any contrast matrix $\mathbf{V}$ by the following formulas

$$\mathbf{E}^{\oplus}\mathbf{X} = \text{ilr}_V^{-1}(\mathbf{E}\text{ilr}_V(\mathbf{X}))$$

$$\mathbf{COV}^{\oplus}\mathbf{X} = \mathbf{COV}(\text{clr}\mathbf{X}) = \mathbf{COV}(\mathbf{V}\text{ilr}_V(\mathbf{X}))$$

and thus

$$\mathbf{COV}^{\oplus}\mathbf{X} = \mathbf{V}\mathbf{COV}(\text{ilr}_V(\mathbf{X}))\mathbf{V}^T = \text{ilr}_V^{-1}(\mathbf{COV}(\text{ilr}_V(\mathbf{X}))),$$

$$M^2(\mathbf{X}) = M^2(\text{ilr}_V(\mathbf{X})),$$

and

$$\mathbf{COV}_w^{\oplus}\mathbf{X} = \mathbf{V}\mathbf{COV}_w(\text{ilr}_V(\mathbf{X}))\mathbf{V}^T = \text{ilr}_V^{-1}(\mathbf{COV}_w(\text{ilr}_V(\mathbf{X})))$$

Note that the scatter functionals $\mathbf{COV}_w^{\oplus}\mathbf{X}$ belong to the algebra $\mathcal{A}$ and thus we also have:

$$\mathbf{COV}_w^{\oplus}\mathbf{X} = \text{clr}^{-1}(\mathbf{COV}_w(\text{clr}\mathbf{X})).$$

Given a sample of size n, the empirical versions of the previous scatter matrices can be derived easily.

4.3 Elliptical distribution in the simplex

Mateu-Figueras et al. (2021) review some distributions in the simplex including the multivariate Student distribution. We define a new family of elliptical distributions in the simplex. A random vector $\mathbf{X}$ with values in $\mathcal{S}^{\mathcal{D}}$ is said to follow an elliptical distribution if any of its ilr coordinates follows an elliptical distribution in $\mathbb{R}^{D-1}$. This definition makes sense due to the following theorem.

Theorem 4 *Given two contrast matrices* $\mathbf{V}_1$ *and* $\mathbf{V}_2$*, if* $\mathbf{X}_1^* = ilr_{V_1}(\mathbf{X})$ *follows an elliptical distribution with parameters* $(\mu_1^*, \mathbf{\Sigma}_1^*)$*, then* $\mathbf{X}_2^* = ilr_{\mathbf{V}_2}(\mathbf{X})$ *follows an elliptical distribution with parameters* (μ_2^*, Σ_2^*) *with*

$$\begin{aligned}
\mathbf{V}_2\mu_2^* &= \mathbf{V}_1\mu_1^* \\
\mathbf{V}_2\mathbf{\Sigma}_2^*\mathbf{V}_2^T &= \mathbf{V}_1\mathbf{\Sigma}_1^*\mathbf{V}_1^T, \\
\mathbf{V}_2\mathbf{\Sigma}_2^{*-1}\mathbf{V}_2^T &= \mathbf{V}_1\mathbf{\Sigma}_1^{*-1}\mathbf{V}_1^T
\end{aligned}$$

From this theorem, we can say that $\mu = \mathbf{V}\mu^*$ is an invariant which characterizes the location parameter in the simplex of the elliptical distribution, $\Sigma = \mathbf{V}\Sigma^*\mathbf{V}^T$ is an invariant that characterizes the scatter matrix in the simplex and $\mathbf{Q} = \mathbf{V}\Sigma^{*-1}\mathbf{V}^T$ is an invariant that characterizes the precision matrix of this distribution in the simplex.

As in Comas-Cufí et al. (2016), we can extend this definition to a mixture of elliptical distributions.

5 ICS for compositional data

5.1 ICS in coordinate space

With the definitions introduced in Section 4.2, we can now define ICS for a compositional random vector $\mathbf{X}$. For a given choice of contrast matrix $\mathbf{V}$, let $\mathbf{X}^* = \mathrm{ilr}_V(\mathbf{X})$. In the ilr coordinate space, ICS consists of the joint diagonalization of two scatter matrices $\mathbf{S}_1(\mathbf{X}^*)$ and $\mathbf{S}_2(\mathbf{X}^*)$. Let us focus on $\mathbf{S}_1(\mathbf{X}^*) = \mathbf{COV}(\mathbf{X}^*)$ and $\mathbf{S}_2(\mathbf{X}^*) = \mathbf{COV}_4(\mathbf{X}^*)$. From equation (3) in Subsection 2.2, we can derive the affine invariant coordinates by diagonalizing the $(D-1)\times(D-1)$ symmetric matrix

$$\mathbf{L}^* = \mathbf{COV}(\mathbf{X}^*)^{-1/2}\mathbf{COV}_4(\mathbf{X}^*)\mathbf{COV}(\mathbf{X}^*)^{-1/2}. \tag{11}$$

Let $\lambda_1 \geq \ldots \geq \lambda_{D-1}$ be the eigenvalues of $\mathbf{L}^*$ in descending order, and $\mathbf{\Lambda}$ be the $(D-1)\times(D-1)$ diagonal matrix with the vector of eigenvalues on its diagonal. Let $\mathbf{u}_k^*$, k ranging from 1 to $D-1$, be the $D-1$ corresponding eigenvectors of $\mathbf{L}^*$ and $\mathbf{U}^* = [\mathbf{u}_1^* \ldots \mathbf{u}_{D-1}^*]$ be the matrix whose columns are these eigenvectors. By cnstruction, the matrix $\mathbf{U}^*$ is orthogonal (with respect to the usual scalar product in $\mathbb{R}^{D-1}$). We have for all $k = 1, \ldots, D-1$:

$$\mathbf{L}^*\mathbf{u}_k^* = \lambda_k \mathbf{u}_k^*.$$

Let us denote by $\mathbf{h}_k^*$, $k = 1, \ldots, D-1$ the column vectors of $\mathbf{H}^* = \mathbf{COV}(\mathbf{X}^*)^{-1/2}\mathbf{U}^*$. We have

$$\mathbf{H}^{*T}\mathbf{COV}(\mathbf{X}^*)\mathbf{H}^* = \mathbf{I}_{D-1} \tag{12}$$

$$\mathbf{H}^{*T}\mathbf{COV}_4(\mathbf{X}^*)\mathbf{H}^* = \mathbf{\Lambda}. \tag{13}$$

Equations (12) and (13) correspond to the joint diagonalization of $\mathbf{COV}(\mathbf{X}^*)$ and $\mathbf{COV}_4(\mathbf{X}^*)$. As for equation (2), we also have:

$$\mathbf{COV}_4(\mathbf{X}^*)\mathbf{H}(\mathbf{X}^*) = \mathbf{COV}(\mathbf{X}^*)\mathbf{H}(\mathbf{X}^*)\mathbf{\Lambda}(\mathbf{X}) \tag{14}$$

The scores or invariant coordinates of $\mathbf{X}^*$ are given by:

$$\mathbf{Z}^* = \mathbf{H}^{*T}(\mathbf{X}^* - \mathbf{E}\mathbf{X}^*) \tag{15}$$

or equivalently by $z_k^* =< \mathbf{h}_k^*, \mathbf{X}^* - \mathbf{E}\mathbf{X}^* >$, $k = 1, \ldots, D-1$.

5.2 ICS in the simplex

Let us now use Section 4 to obtain a formulation of the previous results back in the simplex. This presentation of ICS involves elements (scatter matrices, eigenvalues and eigenvectors) which are independent of the particular choice of contrast matrix, thus justifying this approach. Let us denote

$$\mathbf{L} = (\mathbf{COV}^{\oplus}\mathbf{X})^{-1/2}\mathbf{COV}_4^{\oplus}\mathbf{X}(\mathbf{COV}^{\oplus}\mathbf{X})^{-1/2}. \tag{16}$$

By Theorem 1, we have that

$$\mathrm{ilr}_{\mathbf{V}}(\mathbf{L}) = \mathbf{L}^* \tag{17}$$

and by Theorem 3, we have that, for $k = 1, \ldots, D$,

$$\mathbf{L} \boxdot \mathbf{u}_k = \lambda_k \odot \mathbf{u}_k$$

where $\mathbf{u}_k = \mathrm{ilr}^{-1}(\mathbf{u}_k^*)$ for $k = 1, \ldots, D-1$, and $\mathbf{u}_D = \mathbf{1}_D/\sqrt{D}$ corresponding to $\lambda_D = 0$. We have $< \mathbf{u}_k, \mathbf{u}_l >_A = \delta_{kl}$, for $k, l = 1, \ldots, D$. The vectors $\mathbf{u}_k$ are the $\mathcal{A}$-eigenvectors of $\mathbf{L}$. We can write the following spectral representation of $\mathbf{L}$:

$$\mathbf{L} = \sum_{k=1}^{D-1} \lambda_k \mathrm{clr}(\mathbf{u}_k)\mathrm{clr}(\mathbf{u}_k)^T$$

If we denote by $\mathbf{h}_k = \mathrm{ilr}^{-1}(\mathbf{h}_k^*) = (\mathbf{COV}^{\oplus}\mathbf{X})^{-1/2} \boxdot \mathbf{u}_k$, $k = 1, \ldots, D$, we get:

$$\mathbf{COV}_4^{\oplus}\mathbf{X} \boxdot \mathbf{h}_k = \lambda_k \odot \mathbf{COV}^{\oplus}\mathbf{X} \boxdot \mathbf{h}_k$$

and

$$(\mathbf{COV}^{\oplus}\mathbf{X})^{-1}\mathbf{COV}_4^{\oplus}\mathbf{X} \boxdot \mathbf{h}_k = \lambda_k \odot \mathbf{h}_k.$$

The scores $\mathbf{Z}^* = (z_1^*, \ldots, z_{D-1})$ defined by (15) do not depend on the constrast matrix as already mentioned in Muehlmann et al. (2021), and are given by

$$z_k^* =< \mathbf{h}_k^*, \mathbf{X}^* - \mathbf{E}\mathbf{X}^* > = < \mathbf{h}_k, \mathbf{X} \ominus \mathbf{E}^{\oplus}\mathbf{X} >_A . \tag{18}$$

This equation shows that the scores can be used for outlier detection independently of the contrast matrix.

5.3 Reconstruction formula

From (15), it is easy to derive the reconstruction formula in coordinate space:

$$\mathbf{X}^* = \mathbf{E}\mathbf{X}^* + (\mathbf{H}^{*T})^{-1}\mathbf{Z}^* \tag{19}$$

Let $\mathbf{a}_k^*$ denote the column vectors of the matrix $(\mathbf{H}^{*T})^{-1} = \mathbf{COV}(\mathbf{X}^*)^{1/2}\mathbf{U}^*$ for $k = 1, \ldots, D-1$. Let us define the scalar product with respect to the metric $\mathbf{COV}(\mathbf{X}^*)^{-1}$ by:

$$< \mathbf{u}^*, \mathbf{v}^* >_{\mathbf{COV}(\mathbf{X}^*)^{-1}} = \mathbf{u}^{*T}\,\mathbf{COV}(\mathbf{X}^*)^{-1}\mathbf{v}^*.$$

Equation (12) shows that the vectors $\mathbf{a}_k^*$, $k = 1, \ldots, D-1$ are orthonormal in the sense of this scalar product since the equation can be rewritten

$$(\mathbf{H}^*)^{-1}\mathbf{COV}(\mathbf{X}^*)^{-1}(\mathbf{H}^{*T})^{-1} = \mathbf{I}_{D-1}. \tag{20}$$

This orthogonality implies that the reconstruction formula can also be obtained by

$$\mathbf{X}^* - \mathbf{EX}^* = \sum_{k=1}^{D-1} < \mathbf{a}_k^*, \mathbf{X}^* - \mathbf{EX}^* >_{\mathbf{COV}(\mathbf{X}^*)^{-1}} \mathbf{a}_k^*. \tag{21}$$

Denoting the coordinates of $\mathbf{Z}^*$ by $(z_1^*, \ldots, z_{D-1}^*)$, the coordinates of the $(D-1)$ vector:

$$(\mathbf{H}^*)^{-1}\mathbf{COV}(\mathbf{X}^*)^{-1}(\mathbf{X}^* - \mathbf{EX}^*)$$

are $< \mathbf{a}_k^*, \mathbf{X}^* - \mathbf{EX}^* >_{\mathbf{COV}(\mathbf{X}^*)^{-1}}$.

Using (19), this vector can be written:

$$(\mathbf{H}^*)^{-1}\mathbf{COV}(\mathbf{X}^*)^{-1}(\mathbf{X}^* - \mathbf{EX}^*) = (\mathbf{H}^*)^{-1}\mathbf{COV}(\mathbf{X}^*)^{-1}(\mathbf{H}^{*T})^{-1}\mathbf{Z}^* \tag{22}$$

Using (22) and (20), we get:

$$(\mathbf{H}^*)^{-1}\mathbf{COV}(\mathbf{X}^*)^{-1}(\mathbf{X}^* - \mathbf{EX}^*) = \mathbf{Z}^* \tag{23}$$

and thus

$$< \mathbf{a}_k^*, \mathbf{X}^* - \mathbf{EX}^* >_{\mathbf{COV}(\mathbf{X}^*)^{-1}} = z_k^*.$$

Combining with (21), we get the final reconstruction formula in coordinate space

$$\mathbf{X}^* = \mathbf{EX}^* + \sum_{k=1}^{D-1} z_k^* \mathbf{a}_k^*. \tag{24}$$

Applying $\mathrm{ilr}_{\mathbf{V}}^{-1}$ to equation (24) we get the following simplex version of the reconstruction formula

$$\mathbf{X} = \mathbf{E}^{\oplus}\mathbf{X} \bigoplus_{k=1}^{D-1} z_k^* \odot \mathbf{a}_k \tag{25}$$

where

$$\mathbf{a}_k = \mathrm{ilr}_{\mathbf{V}}^{-1}(\mathbf{a}_k^*) = (\mathbf{COV}^{\oplus}\mathbf{X})^{1/2} \boxdot \mathbf{u}_k. \tag{26}$$

The vectors $\mathbf{a}_k$ are related to the $\mathcal{A}$-eigenvectors of $\mathbf{L}$ by (26). They generate simplex-lines called ICS-axes which are the sets of vectors $\alpha \odot \mathbf{a}_1$, for $\alpha \in \mathbb{R}$. In the next section, we use the empirical version of the reconstruction formula (25) in order to plot the projection of the data on the vector $\mathbf{a}_1$ in some ternary diagrams in situations where the number of selected invariant components is one.

We can also write (25) in terms of the vectors $\mathbf{h}_k$:

$$\mathbf{X} = \mathbf{E}^{\oplus}\mathbf{X} \bigoplus_{k=1}^{D-1} < \mathbf{h}_k, \mathbf{X} \ominus \mathbf{E}^{\oplus}\mathbf{X} >_A \odot \mathbf{COV}^{\oplus}\mathbf{X} \boxdot \mathbf{h}_k . \tag{27}$$

6 Examples of application

We first consider two artificial data sets following a mixture of two normal distributions with 10% of observations that differ from the 90% constituting the main bulk of the data. The dimension is $D = 3$ for the first example and $D = 20$ for the second one. The contrasts matrices we use for the ilr transformations in this section are triangular Helmert matrices corresponding to the original ilr transformation defined by Egozcue et al. (2003).

6.1 Toy examples

For the first example, the contrast matrix is given by $\mathbf{V}^T = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{2}\sqrt{\frac{2}{3}} & -\frac{1}{2}\sqrt{\frac{2}{3}} & \sqrt{\frac{2}{3}} \end{pmatrix}$.
In this toy example, $n = 100$ observations are generated in the ilr space with $D - 1 = 2$ dimensions from a mixture of two Gaussian distributions. The mean and the covariance matrix of the first 90% of the observations (sample 1) are respectively

$$\mu_1^* = (0, 0)^T \text{ and } \mathbf{\Sigma}_1^* = 0.02\mathbf{I}_2 + 0.02\mathbf{1}_2\mathbf{1}_2^T,$$

while the mean vector and the covariance matrix of the remaining 10% (sample 2) are

$$\mu_2^* = \left(\frac{2}{\sqrt{2}} \log 2, \frac{-1}{\sqrt{6}} \log 2\right)^T \text{ and } \Sigma_2^* = 0.05\mathbf{I}_2.$$

Figure 1 on the left (resp. in the middle) shows the data set in the simplex $\mathcal{S}^3$ (resp. in the ilr space). The points in cyan (resp. magenta) belong to sample 1 (resp. sample 2) and we can see that component x_2 has higher values in sample 2 than in sample 1, to the detriment of x_1 and x_3. We perform the ICS method in the ilr space using the `ICSOutlier` package (Archimbaud et al., 2018a). The eigenvalues are 1.57 and 0.81 and the D'Agostino test for normality leads to the selection of a single invariant component. Note that this test is based on the ICS scores and thus does not depend on the ilr transformation (see Archimbaud et al. (2018a) for more details). Figure 1 on the right reports the ICS distances as in Equation (7) for each observation. The horizontal line represents a cut-off value based on Monte Carlo simulations and a 90% quantile. The choice of the quantile order can be done with respect to the expected percentage of outliers in the data. The ICS distances and the cutoff are also independent of the ilr transformation since they depend on the ICS scores only. Figure 1 on the right allows us to identify outliers represented by filled circles. On this example, all 10 observations from sample 2 are identified as outliers whereas only 1 out of the 90 observations from sample 1 is incorrectly identified (at the limit of the cutoff value).

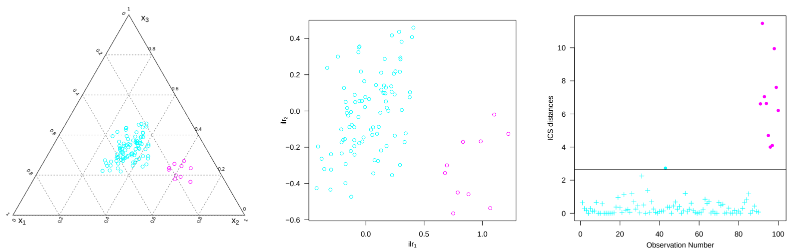

Fig. 1 First toy example: data in the simplex (left), data in the ilr space (middle), identification of the outlying observations using ICS (right).

The two vectors generating the ICS-axes (dashed lines on Figure 2) are equal to $a_1^* = (0.31, -0.1)$ and $a_2^* = (0.12, 0.22)$ in the ilr space and $a_1 = (0.27, 0.43, 0.30)$ and $a_2 = (0.28, 0.33, 0.39)$ in the simplex space. To better understand the role of the ICS components and how they discriminate the observations, we represent on Figure 2, the projections of the observations on the first ICS axis (left plots) and the second ICS axis (right plots) in the ilr space (top plots) and in the simplex space (bottom plots). The first ICS axis allows to discriminate the observations with a high value of x_2 relatively to the other shares, and results in a good discrimination of the two groups. On the contrary, the second axis which seems to separate observations with high values of x_1 against observations with high values of x_3, does not allow to discriminate the two groups.

Finally, using the cutoff value, we represent in grey on Figure 3 the zones or areas of the ilr space (left plot) and of the simplex (right plot) where the observations are considered as outliers. It confirms that the observations with a large or a small value of x_2 relatively to the other shares are in the outlying zone.

For the second toy example, we generate a higher dimensional data with $D = 20$, using two multivariate Gaussian distributions. The first sample is of size $n_1 = 90$ with

$$\mu_1^* = (0, 0, \ldots, 0)^T \quad \Sigma_1^* = 0.02\mathbf{I}_{D-1} + 0.02\mathbf{1}_{D-1}\mathbf{1}_{D-1}^T,$$

and the second sample is of size $n_2 = 10$ with

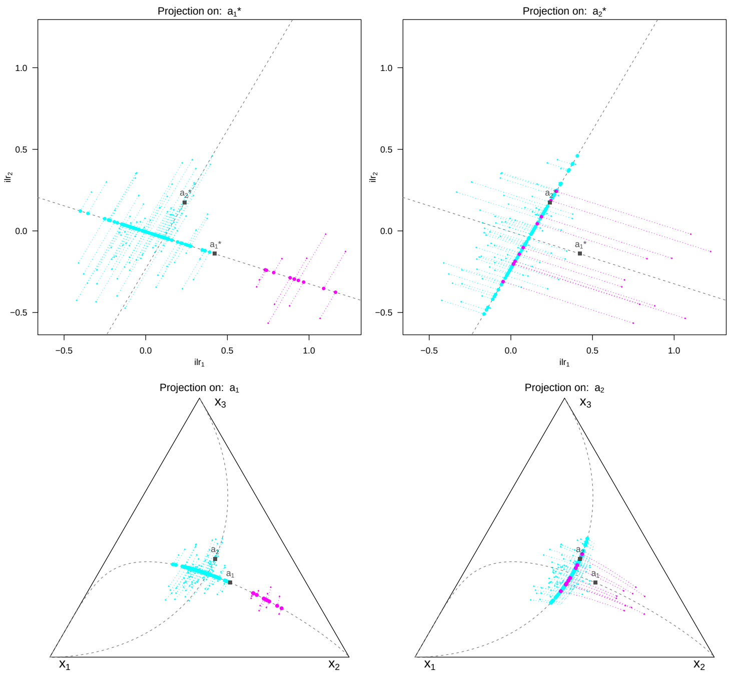

Fig. 2 First toy example: plot of the ICS-axes and projections of the data on the ICS axes (ICS 1 on the left and ICS 2 on the right) in the ilr space (top plots) and in the simplex (bottom plots).

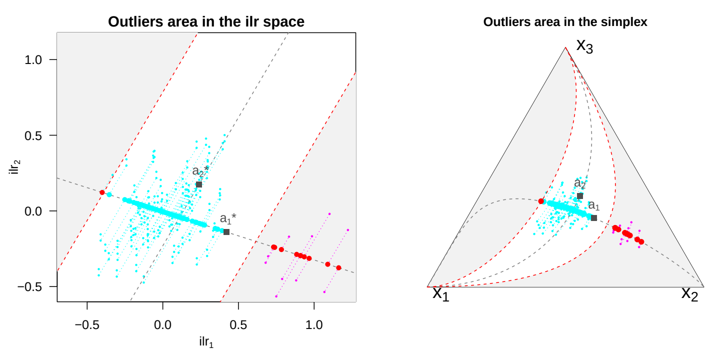

Fig. 3 First toy example: outliers zones in grey in the ilr space (left) and in the simplex space (right).

$$\mu_2^* = \left(\frac{2}{\sqrt{2}}\log 2, \frac{-1}{\sqrt{6}}\log 2, 0, \ldots, 0\right)^T \quad \Sigma_2^* = \begin{pmatrix} 0.05\mathbf{I}_2 & \mathbf{0} \\ \mathbf{0} & 0.02\mathbf{I}_{D-1} + 0.02\mathbf{1}_{D-3}\mathbf{1}_{D-3}^T \end{pmatrix}.$$

When $D > 3$, several options can be used for representing compositional data. One possibility is to plot ternary diagrams using sub-compositions as described in van den Boogaart and Tolosana-Delgado (2008). For instance, one can choose to plot a ternary diagram with x_1, x_2 and the sum of the remaining parts $x_3 + \ldots + x_D$. Another possibility is to replace the sum of the remaining parts by their geometric mean. If $D > 3$ is not too large, these sub-ternary diagrams can be gathered in a square matrix of dimension $D(D-1)/2$.

In order to identify the outliers, we implement the ICS method using `ICSOutlier` in coordinate space. The procedure selects only the first invariant component. The left plot of Figure 4 displays the ICS distances and the cutoff value as an horizontal line to identify outliers. This plot is the same for all ilr transformations. 9 observations out of 10 are detected as outliers in sample 2 while none of the observations from sample 1 are identified as outliers. The plot on the right represents several sub-ternary diagrams, but not all of them because of the large dimension $D = 20$. The selected ternary diagrams plot two parts among x_1 to x_5 against the geometric mean of the rest. However, the diagrams that are not shown are very similar to the ones that focus on x_3, x_4 and x_5 (see the rows and columns 3, 4 and 5 on the matrix plot). Observations with the cross (resp. circle) symbol belong to sample 1 (resp. sample 2). The sub-ternary diagrams confirm that x_1 and x_2 are the composition parts playing a role in explaining the outlyingness of the red points. In fact, the sample 1 observations are clearly visible and separated from the other group when considering the x_1 and x_2 components and the geometric mean of the other parts. On the contrary when looking at the ternary diagrams that do not take x_1 and x_2 separately from the other parts, the outliers are not distinct from the other observations.

We represent on Figure 5 the sub-ternary diagram $(x_1, x_2, *)$ (where $*$ represents the geometric mean of the rest), with small circles in cyan (resp.magenta) for sample 1 (resp. sample 2). The vector $\mathbf{a}_1$ is plotted together with the ICS axis represented by a dashed line. We see that the data projected on the first ICS axis are clearly discriminated by high values of x_2 relatively to x_1.

6.2 Market shares example

This market share dataset has been simulated from a model fitted on the real European cars market in 2015 and is available in Barreiro et al. (2022). The plot on the top of Figure 6 represents the shares in the French automobile market of 5 segments ($D = 5$), from January 2003 to August 2015, denoted by A, B, C, D and E (European cars market segments, from the cheapest cars to the most powerful and luxury ones). We perform the ICS method in the ilr space and represent on the bottom of Figure 6 the ICS distances for each observation. The normality test of the ICS procedure reveals that only the first component is important for outliers identification. The

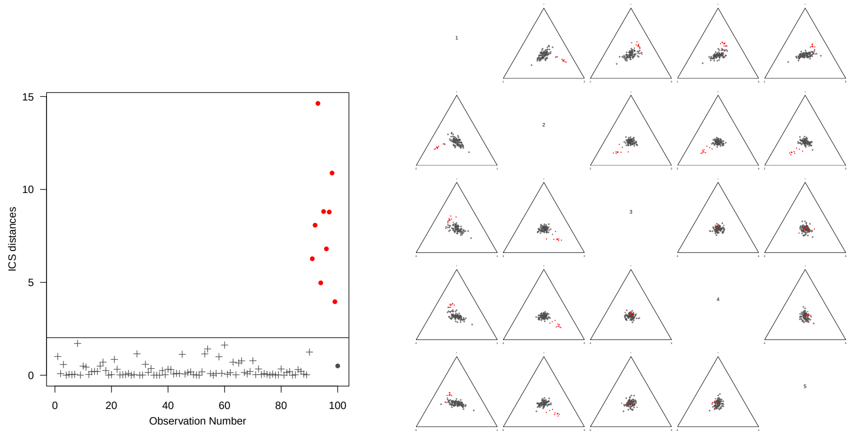

Fig. 4 Second toy example: ICS distances (left), sub-ternary diagrams of the first five composition parts, (right) with circles for sample 2 and red for detected outliers.

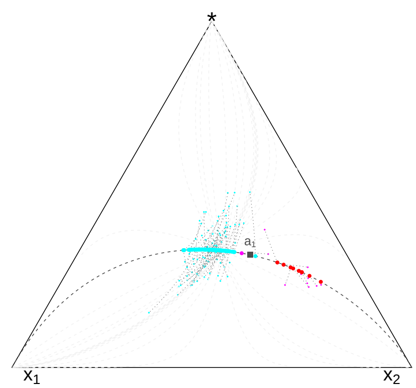

Fig. 5 Second toy example: plot of the ICS-axis $\mathbf{a}_1$ and projections of the data on this axis in the ternary diagram $(x_1, x_2, *)$.

cut-off value is based on the quantile of order 97.5%. All the identified outliers are concentrated in a time interval between September 2008 and May 2009. During this period, the global automobile market was undergoing a crisis with worldwide sales significantly down and political solutions have been provided such as the scrapping bonus at the end of 2008.

In Figure 7 we represent the matrix of sub-ternary diagrams with detected outliers in red. It seems that among all ternary diagrams, the ones including segment A are the best possible in order to identify the outliers. More precisely, the sub-ternary

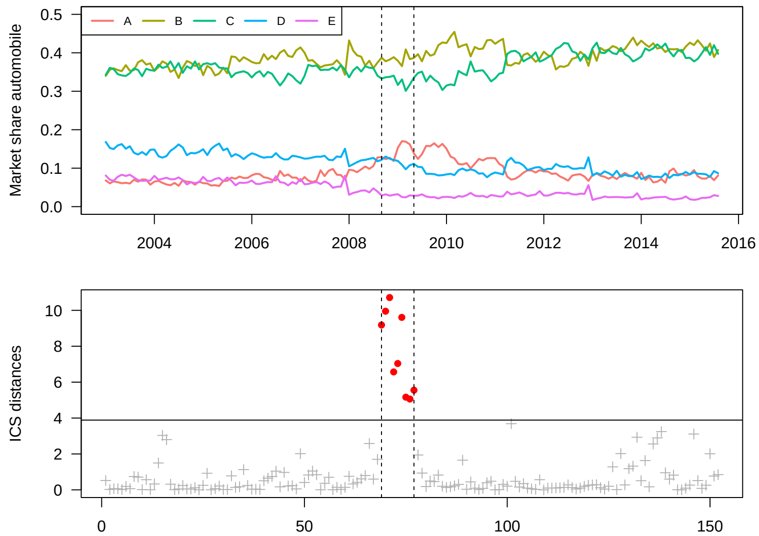

Fig. 6 French Market automobile shares example: from January 2003 to August 2015 in 5 segments (top) and identification of the outlying observations using ICS distances (bottom). The dotted vertical lines represent the period in which outliers were identified (September 2008 to May 2009)

diagram that includes segments A, D, and the others separates the most the two groups. Thus, we plot on Figure 8 the data in the sub-ternary diagram $(A, D, *)$ where $*$ represents now the sum of the other components. We also represent the vector $\mathbf{a}_1$, the ICS axis, and the projections of the data on this axis.

The time points which are detected as outlying correspond to observations with high values of segment A, compared to more normal values of D and low values of $B + C + E$. This interpretation is confirmed when looking at the top plots of Figure 6.

7 Conclusion

The present contribution extends ICS for outlier detection to the context of compositional data. As for standard data, ICS with the scatter pair **COV** and $\mathbf{COV}_4$ is a powerful tool to detect a small proportion of outliers. The definition of ICS in coordinate space is straightforward and the identification of outliers does not depend on the choice of the isometric log-ratio transformation. The definition of ICS in the simplex is more challenging and some algebra tools have been introduced to tackle the problem. Using a reconstruction formula, ICS axes can be plotted on ternary diagrams that help interpreting the outliers. Further interpretation tools are

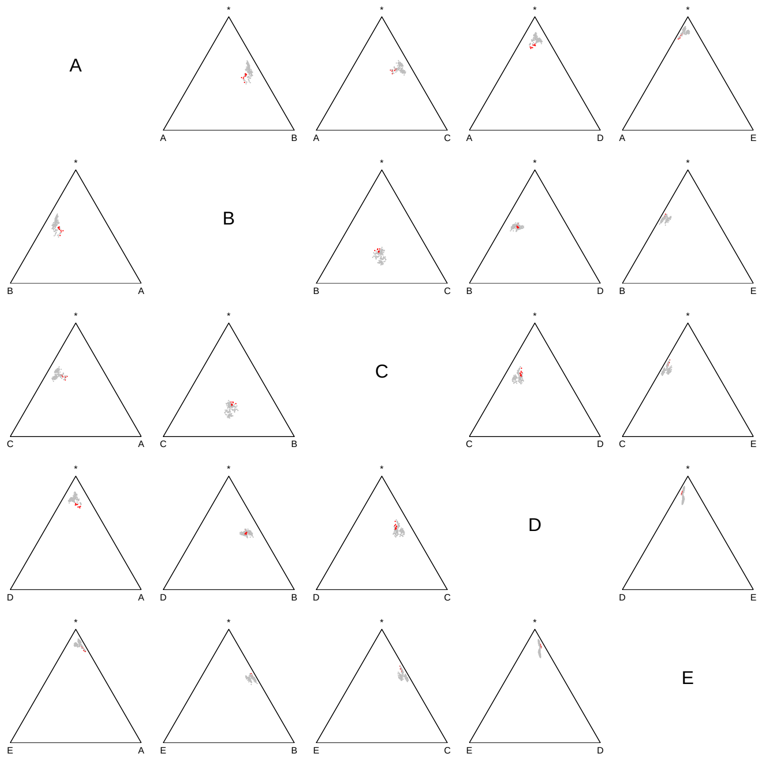

Fig. 7 French Market automobile shares example: outliers identification on the matrix sub-ternary diagram.

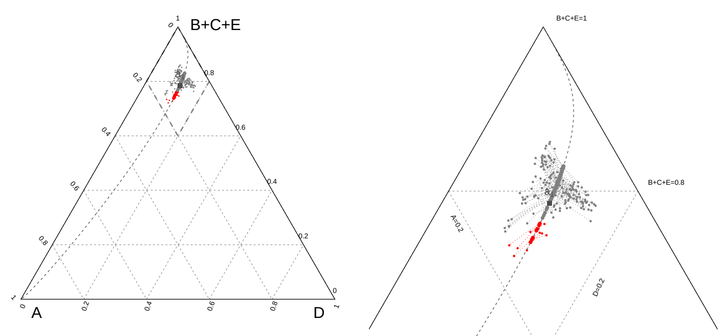

Fig. 8 French Market automobile shares example: projection of the data on the first ICS axis $\mathbf{a}_1$ in the sub-ternary diagram defined by A, D, and the amalgamation of others components (left). Zoom on the interesting part of the ternary diagram (right).

work in progress. Among the perspectives, we can mention the extension of ICS to compositional functional data (see Rieser and Filzmoser (2022)).

Acknowledgements The authors would like to thank Pr. Dr. David Tyler for all his work in multivariate data analysis, which is an immense source of inspiration. We acknowledge funding from the French National Research Agency (ANR) under the Investments for the Future (Investissements d'Avenir) program, grant ANR-17-EURE-0010.

8 Appendix

Proof of Theorem 1

1. Let $\mathbf{P}_V$ be the $D \times D$ bloc matrix $[\mathbf{V}\ \frac{1}{\sqrt{D}}\mathbf{1}_D]$. Then $\mathbf{P}_V^T\mathbf{P}_V = \mathbf{I}_D$ and $\mathbf{P}_V\mathbf{P}_V^T = \mathbf{V}\mathbf{V}^T + \frac{1}{D}\mathbf{1}_D\mathbf{1}_D^T = \mathbf{I}_D$ therefore $\mathbf{P}_V$ is invertible and its inverse is equal to $\mathbf{P}_V^T$. If $\mathbf{A} = \mathbf{V}A^*\mathbf{V}^T$ for a $(D-1)\times(D-1)$ matrix $\mathbf{A}^*$, then $\mathbf{A} = \mathbf{P}_V \begin{pmatrix} \mathbf{A}^* & \mathbf{0} \\ \mathbf{0} & 0 \end{pmatrix} \mathbf{P}_V^T = \mathbf{P}_V \begin{pmatrix} \mathbf{A}^* & \mathbf{0} \\ \mathbf{0} & 0 \end{pmatrix} \mathbf{P}_V^{-1}$ therefore $\mathbf{A}$ is similar to $\mathbf{A}^*$ and their rank is equal.
2. if $\mathbf{A}^*$ is invertible, by the previous property, $\mathbf{A} = \mathbf{V}\mathbf{A}^*\mathbf{V}^T$ is also invertible. Then, let us first prove that $(\mathbf{A} + \frac{1}{D}\mathbf{1}_D\mathbf{1}_D^T)$ is invertible. We can write

$$\mathbf{P}_V \begin{pmatrix} \mathbf{A}^* & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix} \mathbf{P}_V^T = \mathbf{A} + \frac{1}{D}\mathbf{1}_D\mathbf{1}_D^T.$$

The rank of the central matrix is D therefore $\mathbf{A} + \frac{1}{D}\mathbf{1}_D\mathbf{1}_D^T$ is invertible and its inverse is given by

$$\left(\mathbf{P}_V \begin{pmatrix} \mathbf{A}^* & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix} \mathbf{P}_V^T\right)^{-1} = \left(\mathbf{P}_V \begin{pmatrix} \mathbf{A}^* & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix} \mathbf{P}_V^{-1}\right)^{-1} = \mathbf{P}_V \begin{pmatrix} \mathbf{A}^{*-1} & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix} \mathbf{P}_V^T.$$

Then let us check that the inverse of $\mathbf{A}$ in $\mathcal{A}$ is given by $\mathbf{P}_V \begin{pmatrix} \mathbf{A}^{*-1} & \mathbf{0} \\ \mathbf{0} & 0 \end{pmatrix} \mathbf{P}_V^T$. Indeed $\mathbf{P}_V \begin{pmatrix} \mathbf{A}^{*-1} & \mathbf{0} \\ \mathbf{0} & 0 \end{pmatrix} \mathbf{P}_V^T\mathbf{A} = \mathbf{P}_V \begin{pmatrix} \mathbf{A}^{*-1} & \mathbf{0} \\ \mathbf{0} & 0 \end{pmatrix} \mathbf{P}_V^T\mathbf{P}_V \begin{pmatrix} \mathbf{A}^* & \mathbf{0} \\ \mathbf{0} & 0 \end{pmatrix} \mathbf{P}_V^T = \mathbf{V}\mathbf{V}^T = \mathbf{G}_D$. Same for the other direction. Since $\mathbf{P}_V \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix} \mathbf{P}_V^T = \frac{1}{D}\mathbf{1}_D\mathbf{1}_D^T$, we have

$$\mathbf{A}^{-1} = \mathbf{P}_V \begin{pmatrix} \mathbf{A}^{*-1} & \mathbf{0} \\ \mathbf{0} & 0 \end{pmatrix} \mathbf{P}_V^T = \mathbf{P}_V \begin{pmatrix} \mathbf{A}^{*-1} & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix} \mathbf{P}_V^T - \mathbf{P}_V \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix} \mathbf{P}_V^T$$

and thus $\mathbf{A}^{-1} = (\mathbf{A} + \frac{1}{D}\mathbf{1}_D\mathbf{1}_D^T)^{-1} - \frac{1}{D}\mathbf{1}_D\mathbf{1}_D^T$. An alternative formula is

$$\mathbf{A}^{-1} = \mathbf{V}(\mathbf{V}^T\mathbf{A}\mathbf{V})^{-1}\mathbf{V}^T$$

3. $\mathrm{ilr}_V(\mathbf{A})\mathrm{ilr}_V(\mathbf{B}) = \mathbf{V}^T\mathbf{A}\mathbf{V}\mathbf{V}^T\mathbf{B}\mathbf{V} = \mathbf{V}^T\mathbf{A}\mathbf{B}\mathbf{V} = \mathrm{ilr}_V(\mathbf{A}\mathbf{B})$. If $\mathbf{A}$ is invertible, then $\mathrm{ilr}_V(\mathbf{A}^{-1}) = \mathbf{V}^T\mathbf{V}(\mathbf{V}^T\mathbf{A}\mathbf{V})^{-1}\mathbf{V}^T\mathbf{V} = (\mathbf{V}^T\mathbf{A}\mathbf{V})^{-1} = (\mathrm{ilr}_V(\mathbf{A}))^{-1}$. If $(\mathrm{ilr}_V(\mathbf{A}))^{1/2}$ exists, let us define $\mathbf{A}^{1/2} = \mathrm{ilr}^{-1}\left((\mathrm{ilr}_V(\mathbf{A}))^{1/2}\right) = \mathbf{V}(\mathrm{ilr}_V(\mathbf{A}))^{1/2}\mathbf{V}^T$. We have $\mathbf{A}^{1/2}\mathbf{A}^{1/2} = \mathbf{V}(\mathrm{ilr}_V(\mathbf{A}))^{1/2}\mathbf{V}^T\mathbf{V}(\mathrm{ilr}_V(\mathbf{A}))^{1/2}\mathbf{V}^T = \mathbf{V}\mathrm{ilr}_V(\mathbf{A})\mathbf{V}^T = \mathbf{A}$.

Proof of Theorem 2

1. 1 is a clear consequence of the fact that $\mathbf{G}_D$ is the neutral element of $\mathcal{A}$.
2. It is clear that $\mathrm{clr}(\mathbf{B})\mathrm{clr}(\mathbf{x}) \in \mathbf{1}^\perp$, hence by definition $\mathrm{clr}^{-1}(\mathrm{clr}(\mathbf{B})\mathrm{clr}(x)) = \mathrm{clr}(\mathbf{B}) \boxdot \mathbf{x}$.
3. If $\mathbf{V}^T\mathbf{B}\mathbf{V} = \mathbf{V}^T\mathbf{A}\mathbf{V}$, then multiplying on the left by $\mathbf{V}$ and on the right by $\mathbf{V}^T$ and using the fact that $\mathbf{V}\mathbf{V}^T = \mathbf{G}_D$, we get $\mathbf{G}_D\mathbf{B}\mathbf{G}_D = \mathbf{G}_D\mathbf{A}\mathbf{G}_D$ and hence $\mathrm{clr}(\mathbf{A}) = \mathrm{clr}(\mathbf{B})$. Then if $\mathbf{A} \in \mathcal{A}$, then $\mathrm{clr}(\mathbf{A}) = \mathbf{A} = \mathrm{clr}(\mathbf{B})$.
4. by Theorem 1, we have $\mathbf{A} = \mathbf{V}\mathrm{ilr}_{\mathbf{V}}(\mathbf{A})\mathbf{V}^T$ and $\mathbf{A} = \mathrm{clr}(\mathbf{A})$ by 1.

Proof of Theorem 3

$\mathbf{A}^*$ is diagonalizable if there exists a basis $\mathbf{v}_1^*, \dots \mathbf{v}_{D-1}^*$ of $\mathbb{R}^{D-1}$ and $D-1$ real values λ_j such that $\mathbf{A}^*\mathbf{v}_j^* = \lambda_j\mathbf{v}_j^*$. Then let $\mathbf{e}_j = \mathrm{ilr}^{-1}(\mathbf{v}_j^*)$, we get by applying ilr^{-1}: $\mathbf{A} \boxdot \mathbf{e}_j = \mathrm{ilr}^{-1}(\lambda_j\mathbf{v}_j^*) = \lambda_j \odot \mathrm{ilr}^{-1}(\mathbf{v}_j^*) = \lambda_j \odot \mathbf{e}_j$ so that $\mathbf{e}_j$ is an $\mathcal{A}$-eigenvector of $\mathbf{A}$. Now applying the clr transformation, we also get that if $\mathbf{w}_j := \mathrm{clr}(\mathbf{e}_j)$, then $\mathbf{A}\mathrm{clr}(\mathbf{e}_j) = \lambda_j\mathrm{clr}(\mathbf{e}_j)$ so that $\mathbf{A}\mathbf{w}_j = \lambda_j\mathbf{w}_j$ showing that $\mathbf{w}_j$ is an eigenvector of $\mathbf{A}$. $\mathbf{1}_D/\sqrt{D}$ is an eigenvector of $\mathbf{A}$ associated to the eigenvalue 0 when $\mathbf{A} \in \mathcal{A}$ and this completes the basis in $\mathbb{R}^D$ since the vectors $\mathbf{w}_j$ belong to $\mathbf{1}_D^\perp$, $j = 1, \dots D-1$.

Proof of Theorem 4

The density of the elliptical distribution of $\mathbf{x}_1^* = \mathrm{ilr}_{V_1}(\mathbf{x})$ is a function of $R = (\mathrm{ilr}_{V_1}(\mathbf{x}) - \mu_1^*)^T\mathbf{\Sigma}_1^{*-1}(\mathrm{ilr}_{V_1}(\mathbf{x}) - \mu_1^*)$. Since $\mathrm{ilr}_{V_1}(\mathbf{x}) = \mathbf{V}_1^T\mathrm{clr}(\mathbf{x})$, an alternative formulation for R is

$$R = (\mathrm{clr}(\mathbf{x}) - \mathrm{clr}(\mu))^T\mathbf{V}_1^T\mathbf{\Sigma}_1^{*-1}\mathbf{V}_1(\mathrm{clr}(\mathbf{x}) - \mathrm{clr}(\mu))$$

Now if we let $\mu_2^* = \mathbf{V}_2^T\mathbf{V}_1\mu_1^*$, we have $\mathbf{V}_2\mu_2^* = \mathbf{V}_1\mu_1^*$. Similarly let $\mathbf{\Sigma}_2^* = \mathbf{V}_2^T\mathbf{V}_1\mathbf{\Sigma}_1^*\mathbf{V}_1^T\mathbf{V}_2$, we have $\mathbf{V}_2\mathbf{\Sigma}_2^*\mathbf{V}_2^T = \mathbf{V}_1\mathbf{\Sigma}_1^*\mathbf{V}_1^T$. Therefore substituting this expression in R, we see that R is invariant to the specification of the contrast matrix and going backwards we can rewrite $R = (\mathrm{ilr}_{V_2}(\mathbf{x}) - \mu_2^*)^T\mathbf{\Sigma}_2^{*-1}(\mathrm{ilr}_{V_2}(\mathbf{x}) - \mu_2^*)$ which shows that $\mathrm{ilr}_{V_1}(\mathbf{x})$ follows an elliptical distribution with parameters μ_2^* and $\mathbf{\Sigma}_2^*$. Now using the properties of contrast matrices $\mathbf{V}\mathbf{V}^T = \mathbf{G}_D$ and $\mathbf{V}^T\mathbf{V} = \mathbf{I}_{D-1}$, we have

$$(\mathbf{V}_2^T\mathbf{V}_1\mathbf{\Sigma}_1^*\mathbf{V}_1^T\mathbf{V}_2)(\mathbf{V}_2^T\mathbf{V}_1\mathbf{\Sigma}_1^{*-1}\mathbf{V}_1^T\mathbf{V}_2) = \mathbf{I}_{D-1}$$

which proves the last part of the theorem.

References

Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(2):139–160.

Archimbaud, A., Nordhausen, K., and Ruiz-Gazen, A. (2018a). ICS for multivariate outlier detection with application to quality control. *Computational Statistics & Data Analysis*, 128:184–199.

Archimbaud, A., Nordhausen, K., and Ruiz-Gazen, A. (2018b). Unsupervized outlier detection with ICSOutlier. *The R Journal*, 10(1):234–250.

Barreiro, I. R., Laurent, T., and Thomas-Agnan, C. (2022). *Regression models involving compositional variables*. R package, https://github.com/tibo31/codareg.

Bilodeau, M. and Brenner, D. (2008). *Theory of Multivariate Statistics*. Springer, New York.

Comas-Cufí, M., Martín-Fernández, J. A., and Mateu-Figueras, G. (2016). Log-ratio methods in mixture models for compositional data sets. *Sort*, 1:349–374.

Egozcue, J., Pawlowsky-Glahn, V., Mateu-Figueras, G., and Barceló-Vidal, C. (2003). Isometric logratio transformations for compositional data analysis. *Mathematical Geology*, 35(3):279–300.

Egozcue, J. J., Barceló-Vidal, C., Martín-Fernández, J. A., Jarauta-Bragulat, E., Díaz-Barrero, J. L., and Mateu-Figueras, G. (2011). *Elements of Simplicial Linear Algebra and Geometry*, chapter 11, pages 139–157. John Wiley & Sons, Ltd.

Filzmoser, P., Hron, K., and Reimann, C. (2012). Interpretation of multivariate outliers for compositional data. *Computers & Geosciences*, 39:77–85.

Mateu-Figueras, G., Monti, G. S., and Egozcue, J. (2021). Distributions on the simplex revisited. In *Advances in Compositional Data Analysis*, pages 61–82. Springer.

Muehlmann, C., Fačevicová, K., Gardlo, A., Janečková, H., and Nordhausen, K. (2021). *Independent Component Analysis for Compositional Data*, pages 525–545. Springer International Publishing.

Nordhausen, K. and Ruiz-Gazen, A. (2022). On the usage of joint diagonalization in multivariate statistics. *Journal of Multivariate Analysis*, 188:104844.

Nordhausen, K. and Tyler, D. E. (2015). A cautionary note on robust covariance plug-in methods. *Biometrika*, 102(3):573–588.

Nordhausen, K. and Virta, J. (2019). An overview of properties and extensions of FOBI. *Knowledge-Based Systems*, 173:113–116.

Pawlowsky-Glahn, V., Egozcue, J. J., and Tolosana-Delgado, R. (2015). *Modelling and Analysis of Compositional Data*. John Wiley & Sons, Ltd.

Rieser, C. and Filzmoser, P. (2022). Outlier detection for pandemic-related data using compositional functional data analysis. In *Pandemics: Insurance and Social Protection*, pages 251–266. Springer, Cham.

Rousseeuw, P. J. and Van Zomeren, B. C. (1990). Unmasking multivariate outliers and leverage points. *Journal of the American Statistical association*, 85(411):633–639.

Theis, F. J. and Inouye, Y. (2006). On the use of joint diagonalization in blind signal processing. In *IEEE International Symposium on Circuits and Systems*, pages 3589–3593. IEEE.

Tyler, D. E., Critchley, F., Dümbgen, L., and Oja, H. (2009). Invariant co-ordinate selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(3):549–592.

van den Boogaart, K. G. and Tolosana-Delgado, R. (2008). "compositions": A unified R package to analyze compositional data. *Computers & Geosciences*, 34(4):320–338.