# Log-Free Divergence and Covariance matrix for Compositional Data I: The Affine/Barycentric Approach

Olivier P. Faugeras

# LOG-FREE DIVERGENCE AND COVARIANCE MATRIX FOR COMPOSITIONAL DATA I: THE AFFINE/BARYCENTRIC APPROACH

OLIVIER P. FAUGERAS
TOULOUSE SCHOOL OF ECONOMICS, UNIVERSITÉ TOULOUSE 1 CAPITOLE,
INSTITUT DE MATHÉMATIQUES DE TOULOUSE

ABSTRACT. The presence of zeroes in Compositional Data (CoDa) is a thorny issue for Aitchison's classical log-ratio analysis. Building upon the geometric approach by (Faugeras 2023), we study the full CoDa simplex from the perspective of affine geometry. This view allows to regard CoDa as points (and not vectors), naturally expressed in barycentric coordinates. A decomposition formula for the displacement vector of two CoDa points yields a novel family of barycentric dissimilarity measures. In turn, these barycentric divergences allow to define i) Fréchet means and their variants, ii) isotropic and anisotropic analogues of the Gaussian distribution, and importantly iii) variance and covariance matrices. All together, the new tools introduced in this paper provide a log-free, direct and unified way to deal with the whole CoDa space, exploiting the linear affine structure of CoDa, and effectively handling zeroes. A strikingly related approach based on the projective viewpoint and the exterior product is studied in the separate companion paper (Faugeras 2025a).

## CONTENTS



2020 *Mathematics Subject Classification.* 62R20, 62H20.
*Key words and phrases.* compositional data, affine geometry, barycentric coordinates, barycentric divergence, Fréchet mean, Gaussian distribution, barycentric variation matrix.

## 1. Introduction

Compositional data (CoDa) analysis provides a framework for the statistical treatment of simplex-valued data, offering insights into proportional relationships. Traditional CoDa methods rely on log-ratio transformations to address inherent constraints, yet these approaches falter when zero values are present. This paper introduces a novel affine geometric perspective on CoDa, leveraging barycentric coordinates to define divergence measures and variance matrices applicable to the entire simplex, including compositions with zeros. By reframing CoDa as affine points rather than Euclidean vectors, we overcome limitations of classical techniques, providing unified tools for statistical analysis.

1.1. **A primer on Compositional data (CoDa).** Compositional data (CoDa) consist of non-negative multivariate data $\mathbf{a} = (a_0, a_1, \ldots, a_d) \in \mathbb{R}^{d+1}$, where each $a_i \geq 0$ represents the quantity of the $i$-th component in a composition. It is understood that the raw magnitude $a_i$ is meaningless in isolation; only its proportion relative to other components matter. Composition of soil in geology, elements in a mixture in chemistry, sources of calories in nutrition, or vote shares in an election are examples of CoDa.

The standard approach, pioneered by J. Aitchison 1986, normalizes the raw composition vector $\mathbf{a}$ by its sum, i.e. expresses the data in percentages, via the

closure operation:

$$\mathcal{C}(\mathbf{a}) := \frac{\mathbf{a}}{\sum_0^d a_i} = \frac{\mathbf{a}}{||\mathbf{a}||_1}, \tag{1}$$

where $||.||_1$ stands for the $\ell_1$-norm. This leads to the consideration of *normalised* CoDa element as a vector

$$\mathbf{x} = (x_0, \ldots, x_d) = \mathcal{C}(\mathbf{a}),$$

confined to the $d$-dimensional unit simplex,

$$\Delta_+^d := \{\mathbf{x} = (x_0, \ldots, x_d) \in \mathbb{R}^{d+1} : x_i \geq 0, \sum_{i=0}^{d} x_i = 1\}. \tag{2}$$

The unit-sum and non-negativity constraints preclude the direct application of conventional multivariate statistics: the naive covariance matrix,

$$\Sigma := \big(\mathrm{cov}(x_i, x_j)\big) \in \mathbb{R}^{(d+1)\times(d+1)}, \tag{3}$$

produces spurious correlations and lacks subcompositional coherence (Pearson 1897, Chayes 1971). To circumvent this, log-ratio transformations—additive (alr), centered (clr), and isometric (ilr)—map the *positive* simplex $\Delta_{++}^d = \{\mathbf{x} > 0, \mathbf{x} \in \Delta_+^d\}$ into a Euclidean space, enabling standard analyses (Greenacre 2018, Vera Pawlowsky-Glahn, Juan José Egozcue, and Tolosana-Delgado 2015, Van Den Boogaart and Tolosana-Delgado 2013).

A central challenge in CoDa analysis is quantifying component interdependence. The main approaches are based on the variance matrix of log-ratio transformed data. J. Aitchison 1986 proposed the variation matrix $T = (t_{ij}) \in \mathbb{R}^{(d+1)\times(d+1)}$, defined as:

$$t_{ij} := \mathrm{Var}\left(\ln \frac{x_i}{x_j}\right), \tag{4}$$

while the clr- and ilr-variance matrices ($\Sigma^{\mathrm{clr}}$ and $\Sigma^{\mathrm{ilr}}$) derive from the variance matrix of corresponding log-transformed data (J. Aitchison 1986, Greenacre 2018) or Van Den Boogaart and Tolosana-Delgado 2013).

1.2. **Motivation.** Despite its successes, the log-ratio approach fails when CoDa include zero components, as logarithmic transformations become undefined. This limitation is critical in fields like chemometrics and bioinformatics, where datasets (e.g., microbiome profiles) frequently contain numerous zeros. Efforts to mitigate this through zero-imputation methods (e.g., Lubbe, Filzmoser, and Templ 2021, J.-A. Martín-Fernández et al. 2015, Palarea-Albaladejo and J. A. Martín-Fernández 2015) introduce sensitivity to replacement techniques, often distorting log-ratio variance estimates due to the logarithm's behavior near zero (Greenacre 2021). Quoting Greenacre 2021 p. 295, "Zeros in compositional data are the Achilles heel of the logratio approach". These challenges underscore the need for alternative representations that accommodate zeros without imputation.

1.3. **Aims, scope and contributions.** This paper redefines CoDa within an affine geometric framework, building on the projective viewpoint of Faugeras 2023, where compositions were treated as projective points, i.e. as equivalence classes $[\mathbf{x}]_+$ for the (positive) scaling relation,

$$\mathbf{z} \in [\mathbf{x}]_+ \Leftrightarrow \mathbf{z} = \lambda\mathbf{x}, \text{ for some } \lambda > 0$$

of non-negative vectors $\boldsymbol{x} \geq \mathbf{0}$, $\mathbf{x} \in \mathbb{R}^{d+1} \setminus \{\mathbf{0}\}$. The set of these equivalence classes $\mathbb{P}^d_+ := \{[\mathbf{x}]_+, \mathbf{x} \in \mathbb{R}^{d+1}_+ \setminus \{\mathbf{0}\}\}$, is the CoDa space. The components of the vector representative $\mathbf{x}$ of $[\mathbf{x}]_+$, are its homogeneous coordinates, denoted by $[x_0 : x_1 : \ldots : x_d]$. In turn, these equivalence classes admits several representations within different geometric models.

Here, we focus on the affine model of $\mathbb{P}^d_+$ provided by the simplex $\Delta^d_+$, interpreting/identifying a CoDa projective point $[\mathbf{x}]_+$ as an *affine point* with barycentric coordinates $(x_0, x_1, \ldots, x_d)$ rather than a *Euclidean vector* with Cartesian coordinates. As a consequence, the displacement vector from CoDa point $\mathbf{y}$ to $\mathbf{x}$ in the simplex does not write plainly as the vector

$$\mathbf{x} - \mathbf{y} = (x_0 - y_0, \ldots, x_d - y_d)$$

obtained by the difference of their coordinates, (as would be the case for $\mathbf{x}, \mathbf{y}$ vectors), but is given by a more complicated formula (see the forthcoming Lemma A.4). Such a formula for the displacement vector in barycentric coordinates is key to build notions of proximity–barycentric divergence–between CoDa affine points and related barycentric variance matrices.

This shift of perspective eliminates reliance on log-ratios, enabling analysis across the full simplex, including zero-containing compositions. Our primary contributions are twofold:

- **Barycentric Divergences:** We introduce a family of $\alpha$-barycentric divergence measures, applicable to all CoDa, and explore their geometric properties, extending them to infinite-dimensional positive measures.
- **Barycentric Variance-Covariance matrices:** We define covariance and variance matrices based on displacement vectors in barycentric coordinates, offering a log-free alternative to traditional log-ratio variance matrices for assessing component proportionality.

1.4. **Outline.** The paper outlines these concepts using affine geometry, avoiding the complexity of projective methods detailed in a companion study (Faugeras 2025a). We begin in Section 2 by elaborating on the affine viewpoint of CoDa and the formula of the displacement vector in barycentric coordinates. Subsequent sections develop the mathematical foundations (Section 3 and 6), derive statistical applications (e.g., Fréchet means (Section 4, with additional study and simulations in Appendix B), Gaussian distributions (Section 5)), and validate the approach through simulations (Section 7).

Appendix A gives a primer on the basics of affine spaces and barycentric coordinates, outlining the differences with vector spaces and Cartesian coordinates. In particular, we give the key basic formula (Lemma A.4) for the displacement vector between two points expressed in barycentric coordinates. Readers unfamiliar with affine geometry are encouraged to review Appendix A prior to proceeding, while those with prior knowledge may proceed directly, consulting Lemma A.4 as needed.

## 2. CoDa as an affine point in barycentric coordinates

In this section, we elaborate on the simplex representation of CoDa as affine points expressed in barycentric coordinates. In particular, we show how the amalgamation, subcomposition and partition operations on CoDa correspond to affine

operations on barycenters. More importantly, we show that the displacement between two CoDa elements can be decomposed in terms of displacements of the different pairs of basis frame parts.

Indeed, in view of Appendix A's discussion of affine geometry and the homogeneous character of barycentric coordinates, we elucidate why CoDa elements within the simplex $\Delta_+^d$ are to be regarded as affine points in barycentric coordinates rather than Euclidean vectors: identifying each part $i$ of a composition with an affine point $A_i$, it is reasonable to assume that the frame $\mathcal{F} = \{A_i, i = 0, \ldots, d\}$ is made of affinely independent points, since parts relate to different entities and are then not related to each others. Consequently, a (simplex representative of) CoDa $\mathbf{x} \in \Delta_+^d$ corresponds to the affine point

$$\mathbf{x} = \sum_{i=0}^{d} x_i A_i, \quad \sum_{i=0}^{d} x_i = 1, \quad x_i \geq 0, \tag{5}$$

in normalized barycentric coordinates, as in (28). For example, for the classical foraminiferal compositional dataset (J. Aitchison 1986, Dataset 34 in Appendix D, see also Section 7), the points $A_0, \ldots, A_3$ would correspond to the four different fossil species (*Neogloboquadrina atlantica*, *Neogloboquadrina pachyderma*, *Globorotalia obesa* and *Globigerinoides triloba*), and $x_0, \ldots x_3$ are the corresponding species proportions[1].

2.1. **CoDa operations as affine operations on points.** We now show how the Amalgamation, Subcomposition and Partition operations on CoDa of J. Aitchison 1986 have a simple geometric description in terms of barycenters and affine combinations of points. Recall that given a CoDa $\mathbf{x} = (x_0, \ldots, x_d) \in \Delta_+^d$, an amalgamation of order 1 is a mapping

$$\Delta_+^d \ni \mathbf{x} \mapsto \mathbf{t} \in \Delta_+^1,$$

obtained when the parts of a $(d+1)-$ composition are separated into two mutually exclusive and exhaustive subsets, and the composition within each subset are added together. This results in a $2-$parts composition in $\Delta_+^1$. For example, $\mathbf{x} = (x_0, x_1, x_2, x_3) \in \Delta_+^3$ can be amalgamated into $\mathbf{t} = (t_0, t_1)$ with $t_0 = x_0 + x_1$, $t_1 = x_2 + x_3$. A subcomposition

$$\Delta_+^d \ni \mathbf{x} \mapsto \mathbf{c} \in \Delta_+^k$$

is obtained by selecting, say, $k+1$ parts of a composition, with $1 \leq k < d$, and closing the selected subvector to obtain a subcomposition in $\Delta_+^k$. Finally, a partition of order one is the separation of a $(d+1)-$parts composition into two disjoint and exhaustive subsets, and recording the amalgamation and subcomposition of each subsets. For example, the order one partition

$$(x_0, \ldots, x_k | x_{k+1}, \ldots, x_d)$$

cuts the $(d+1)-$parts at position $0 \leq k < d$ and yields an amalgamation vector $\mathbf{t} = (t_0, t_1)$, with $t_0 = x_0 + \ldots + x_k$, $t_1 = x_{k+1} + \ldots + x_d$, together with the two

[1]One could also consider CoDa as a (discrete) measure on a set of categories, with the affine points $A_i$ corresponding to a Dirac measure on each category. See e.g. Faugeras 2026 for approaches to CoDa from this probabilistic standpoint.

vectors of subcompositions

$$\mathbf{c}_0 = \mathcal{C}(x_0, \ldots, x_k) = \frac{(x_0, \ldots, x_k)}{t_0}, \quad \mathbf{c}_1 = \mathcal{C}(x_{k+1}, \ldots, x_d) = \frac{(x_{k+1}, \ldots, x_d)}{t_1}.$$

By Property 2.10 and 2.11 of J. Aitchison 1986, this results in a bijective transformation

$$\Delta_+^d \ni \mathbf{x} \mapsto (\mathbf{t}, \mathbf{c}_0, \mathbf{c}_1) \in \Delta_+^1 \times \Delta_+^k \times \Delta_+^{d-k-1}.$$

Identifying a CoDa element $\mathbf{x} = (x_0, \ldots, x_d) \in \Delta_+^d$ with the affine point $P = \sum_{i=0}^d x_i A_i$ expressed as a barycenter of the base parts-points $A_i$, the point $P$ can be decomposed as a sum of two points

$$\begin{aligned} P &= \left(\sum_{i=0}^k x_i A_i\right) + \left(\sum_{i=k+1}^d x_i A_i\right) \\ &:= C_0 + C_1. \end{aligned}$$

$C_0$ and $C_1$ write, in *normalized* barycentric coordinates, as

$$\begin{aligned} C_0 &= \sum_{i=0}^k \frac{x_i}{\sum_{j=0}^k x_j} A_i = \sum_{i=0}^k \frac{x_i}{t_0} A_i = \sum_{i=0}^k c_{0i} A_i, \\ C_1 &= \sum_{i=k+1}^d \frac{x_i}{\sum_{j=k+1}^d x_j} A_i = \sum_{i=k+1}^d \frac{x_i}{t_1} A_i = \sum_{i=k+1}^d c_{1i} A_i, \end{aligned}$$

where $(c_{0i})$, resp. $(c_{1i})$, are the components of $\mathbf{c}_0$, resp. $\mathbf{c}_1$. The weighted point $(1, P)$ writes, also in *normalized* barycentric coordinates, as

$$P = t_0 C_0 + t_1 C_1, \quad t_0 + t_1 = 1. \tag{6}$$

In other words, the composition $P$ can be partitioned into two subcompositions $C_0$ and $C_1$, whose normalized barycentric coordinates $\mathbf{c}_0, \mathbf{c}_1$ corresponds to the subcomposition operation of J. Aitchison 1986. In addition, the original composition point $P$ writes as the barycenter of these two subcompositions points, with barycentric coordinates $\mathbf{t} = (t_0, t_1)$ w.r.t. $C_0, C_1$, corresponding to the amalgamation operation. The equality (6) is the statement of Properties 2.10 and 2.11 of the partitioning operation of J. Aitchison 1986 in affine geometric language and corresponds to the well-known property of associativity/reduction of the barycenter, that is to say that a barycenter can be computed from sub-barycenters.

These considerations, although elementary, shed a geometric light on the basic operations on compositions and thus vindicate the affine viewpoint espoused in this paper. In particular, it clarifies the role of the total/amalgamation/closure in the treatment of CoDa with a total.

2.2. **Displacement vectors of CoDa points.** We now turn our focus on the key formula for the displacement vector of CoDa points. Indeed, for CoDa points $\mathbf{x}, \mathbf{y} \in \Delta_+^d$, the unit sum normalization $||\mathbf{x}||_1 = ||\mathbf{y}||_1 = 1$ entails a simplification in formula (29) in Lemma A.4, as

$$\overrightarrow{\mathbf{xy}} = \sum_{i<j} \det \begin{vmatrix} x_i & y_i \\ x_j & y_j \end{vmatrix} \overrightarrow{A_i A_j} \tag{7}$$

and makes it clear why the displacement between two CoDa points should not write as a difference (25) of their coordinates in a frame of reference (see Appendix A.4).

Formula (7) is the key ingredient for the affine viewpoint of this paper: it gives the decomposition of the displacement vector from $\mathbf{x}$ to $\mathbf{y}$ in terms of the $d(d+1)/2$ displacements $\overrightarrow{A_iA_j}$ of two different points $A_i, A_j$, $i \neq j$, of the affine frame $\mathcal{F}$.

**Remark 1** (Displacement coordinates as weighted ratios)**.** Formula (7) can be written in terms of (weighted) ratios. Indeed, let

$$v_{ij} := \det \begin{vmatrix} x_i & y_i \\ x_j & y_j \end{vmatrix}, \quad 0 \leq i \neq j \leq d, \tag{8}$$

be the determinantal coefficient of the component of the displacement in the $\overrightarrow{A_iA_j}$ direction in formula (7). It writes, for $x_i \neq 0$, $y_i \neq 0$, as

$$v_{ij} = x_iy_j - y_ix_j = x_iy_i \left( \frac{y_j}{y_i} - \frac{x_j}{x_i} \right). \tag{9}$$

Let $\mathbb{U}_i := \{[\mathbf{x}]_+ \in \mathbb{P}_+^d : x_i \neq 0\}$ be the subset of the CoDa space with non-null (hence positive) $i$-th coordinate. Then, following Faugeras 2023, $\mathbb{U}_i$ can be identified with the non-negative part of the affine hyperplane $\{\mathbf{x} : x_i = 1\}$ of $\mathbb{R}^{d+1}$: a projective CoDa point $[\mathbf{x}]_+ \in \mathbb{U}_i$ with homogeneous coordinates

$$[x_0 : \ldots : x_i : \ldots : x_d]_+ = \left[ \frac{x_0}{x_i} : \ldots : 1 : \ldots : \frac{x_d}{x_i} \right]_+$$

can in turn be identified, after dropping the constant 1 at the $i$th position, with an affine point $X_{/i} \in \mathbb{R}^d$ with inhomogeneous coordinates

$$X_{/i} := \left( \frac{x_0}{x_i}, \ldots, \frac{x_{i-1}}{x_i}, \frac{x_{i+1}}{x_i}, \ldots, \frac{x_d}{x_i} \right).$$

Thus, if both $[\mathbf{x}]_+$ and $[\mathbf{y}]_+$ belong to $\mathbb{U}_i$, equation (9) interprets as a weighted difference $y_j/y_i - x_j/x_i$ of the $j$th inhomogeneous coordinate of the points $X_{/i}$ and $Y_{/i}$, in the affine patch corresponding to $\mathbb{U}_i$, with a weight $x_iy_i$ given by the product of their $i$th (simplex) coordinate. If one sets conventionally $0/0 := 1$, then formula (9) becomes true for all (non-negative) Coda elements $[\mathbf{x}]_+, [\mathbf{y}]_+ \in \mathbb{P}_+^d$. In other words, the displacement for CoDa elements decomposes as a weighted difference, not of their coordinate components $(x_j)$, but of their ratios $(x_j/x_i)_{j \neq i}$. The weight $x_iy_i$ translates the relative importance of part $i$ of $\mathbf{x}$ and $\mathbf{y}$. This description gives a geometrical interpretation of the components of the displacement vector in terms of their inhomogeneous coordinates in each affine hyperplanes $\{\mathbf{x} : x_i = 1\}$.

## 3. Barycentric divergence on the CoDa space

In this section, we introduce our first major contribution, barycentric divergences, and study its theoretical properties.

### 3.1. Motivation and definition.

Lemma A.4, and its specialization via formula (7) to CoDa represented on the simplex $\Delta_+^d$, by giving the decomposition of the displacement vector of CoDa $\mathbf{x}$ to $\mathbf{y}$ in terms of the $d(d+1)/2$ displacements $\overrightarrow{A_iA_j}$ of two different points $A_i, A_j$, provides a natural way to measure the distance or proximity between two CoDa elements.

Indeed, each part $i = 0, \ldots, d$ of a composition identifies with a point $A_i$ in a $d-$dimensional affine space. However, parts of a composition are just different entities with no proper geometric properties. Hence, although the displacements

$\overrightarrow{A_iA_j}$, $i<j$ are dependent from the affine geometric viewpoint, it makes sense, from the compositional viewpoint, to consider each pair $(i,j)$ of parts, identified with the displacements $\overrightarrow{A_iA_j}$, $i \neq j$ in formula (7), *as if* they were "orthogonal". In fact, the principle of subcompositional coherence in CoDa is based on the idea that, quoting Vera Pawlowsky-Glahn, Juan José Egozcue, and Tolosana-Delgado 2015 p. 16, "subcompositions should behave like orthogonal projections in real analysis". This principle somehow motivates (heuristically) the idea that two different elementary displacements $\overrightarrow{A_iA_j}$, and $\overrightarrow{A_kA_l}$, for $(i,j) \neq (k,l)$, should be thought as "orthogonal" and that each corresponding determinantal coordinate coefficient in (7) measures "orthogonal" characteristics of a pair of CoDa, from which one can build a measure of distance or proximity between CoDa points. (A more principled mathematical justification will be given in Faugeras 2025a, based on exterior products.) Taking a measure of the magnitude of the displacement vector $\overrightarrow{\mathbf{xy}}$, via the choice of a norm $||.||$, e.g., an $\ell_\alpha$ norm, leads to the following definition:

**Definition 3.1.** Let $[\mathbf{x}]_+, [\mathbf{y}]_+ \in \mathbb{P}_+^d$ be two CoDa elements, and $\alpha$ be a real number. Then, the $\alpha$-barycentric divergence is defined, for $1 \leq \alpha < \infty$, as

$$d_\alpha([\mathbf{x}]_+, [\mathbf{y}]_+) := \frac{\left(\sum_{i<j} \left|\det \begin{vmatrix} x_i & y_i \\ x_j & y_j \end{vmatrix}\right|^\alpha\right)^{1/\alpha}}{||\mathbf{x}||_1||\mathbf{y}||_1}, \tag{10}$$

and, for $\alpha = \infty$, as

$$d_\alpha([\mathbf{x}]_+, [\mathbf{y}]_+) := \frac{\max_{i<j} \left|\det \begin{vmatrix} x_i & y_i \\ x_j & y_j \end{vmatrix}\right|}{||\mathbf{x}||_1||\mathbf{y}||_1}.$$

Both expressions reduces to the numerator in case of simplex representatives $\mathbf{x}, \mathbf{y} \in \Delta_+^d$.

The following theorem studies properties of such divergences, and justifies the heuristic motivation which had lead to their definition.

**Theorem 3.2.** *i)* $d_\alpha$ *is* ***well defined*** *on* $\mathbb{P}_+^d$, $d_\alpha : \mathbb{P}_+^d \times \mathbb{P}_+^d \to \mathbb{R}_+$, *i.e. does not depend on the representatives* $\mathbf{x}, \mathbf{y}$ *of* $[\mathbf{x}]_+, [\mathbf{y}]_+$, *viz.*

$$d_\alpha([\mathbf{x}]_+, [\mathbf{y}]_+) = d_\alpha([\lambda\mathbf{x}]_+, [\mu\mathbf{y}]_+), \quad \lambda, \mu > 0.$$

*ii)* ***Symmetry****:* $d_\alpha([\mathbf{x}]_+, [\mathbf{y}]_+) = d_\alpha([\mathbf{y}]_+, [\mathbf{x}]_+)$.

*iii)* ***Permutation invariance****: let* $\boldsymbol{\sigma} = (\sigma_0, \sigma_1, \ldots, \sigma_d)$ *be a permutation of* $\{0, 1, \ldots, d\}$ *and write* $\mathbf{x}_\sigma$ *for the vector* $(x_{\sigma_0}, \ldots, x_{\sigma_d})$. *Then,*

$$d_\alpha([\mathbf{x}_{\boldsymbol{\sigma}}]_+, [\mathbf{y}_{\boldsymbol{\sigma}}]_+) = d_\alpha([\mathbf{x}]_+, [\mathbf{y}]_+).$$

*iv)* ***Boundedness****:* $0 \leq d_\alpha([\mathbf{x}]_+, [\mathbf{y}]_+) \leq 1$.

*v)* ***Positive-definiteness****:* $d_\alpha([\mathbf{x}]_+, [\mathbf{y}]_+) \geq 0$ *and* $d_\alpha([\mathbf{x}]_+, [\mathbf{y}]_+) = 0 \Leftrightarrow [\mathbf{x}]_+ = [\mathbf{y}]_+$

*vi)* ***Zeroes subcompositional coherence****: if* $[\mathbf{x}]_+, [\mathbf{y}]_+ \in \mathbb{P}_+^d$ *are seen as subcompositions of larger compositions* $[\tilde{\mathbf{x}}]_+ := [\mathbf{x} : \mathbf{0}]_+ \in \mathbb{P}_+^{d+k}$ *and* $[\tilde{\mathbf{y}}]_+ := [\mathbf{y} : \mathbf{0}]_+ \in \mathbb{P}_+^{d+k}$, *where* $\mathbf{0} \in \mathbb{R}^k$, *for some integer* $k > 0$, *and where* $\mathbf{x} : \mathbf{0}$ *stands for the concatenation of* $\mathbf{x} \in \mathbb{R}_+^{d+1}$ *and* $\mathbf{0}$, *then*

$$d_\alpha([\tilde{\mathbf{x}}]_+, [\tilde{\mathbf{y}}]_+) = d_\alpha([\mathbf{x}]_+, [\mathbf{y}]_+).$$

*vii)* ***Weighted subcompositional dominance****: let $I \subset \{0, 1, \dots, d\}$ be a subset of indices of cardinality at least 2, and for $\mathbf{x} \in \mathbb{R}^{d+1}$, set $\mathbf{x}_I := (x_i)_{i \in I}$ be the corresponding*[2] *sub-vector of $\mathbf{x}$ made of the parts in $I$. If $[\mathbf{x}]_+, [\mathbf{y}]_+ \in \mathbb{P}^d_+$, then*

$$||\mathbf{x}_I||_1 \, ||\mathbf{y}_I||_1 d_\alpha([\mathbf{x}_I]_+, [\mathbf{y}_I]_+) \leq ||\mathbf{x}||_1 \, ||\mathbf{y}||_1 d_\alpha([\mathbf{x}]_+, [\mathbf{y}]_+).$$

*viii)* ***Amalgamation dominance for*** $\alpha = 1$*: let $I$, $\mathbf{x}_I$ as in vii), and set $J := \{0, \dots, d\} \setminus I$ be the complementary index set of $I$. For $\mathbf{x} \in \mathbb{R}^{d+1}$, set $\mathscr{A}_I(\mathbf{x}) = (\sum_{i \in I} x_i, \mathbf{x}_J) \in \mathbb{R}^{\mathrm{card}\,(J)+1}$ the vector with components in $I$ amalgamated. Then, if $[\mathbf{x}]_+, [\mathbf{y}]_+ \in \mathbb{P}^d_+$, one has that*

$$d_1([\mathscr{A}_I(\mathbf{x})]_+, [\mathscr{A}_I(\mathbf{y})]_+) \leq d_1([\mathbf{x}]_+, [\mathbf{y}]_+).$$

*Proof.* i) Since $[\mathbf{x}]_+, [\mathbf{y}]_+ \in \mathbb{P}^d_+$, $\mathbf{x}, \mathbf{y} \neq \mathbf{0}$. Thus, $||\mathbf{x}||_1, ||\mathbf{y}||_1 \neq 0$. Scale invariance $\mathbf{x} \leftarrow \lambda \mathbf{x}$, $\mathbf{y} \leftarrow \mu \mathbf{y}$, in the r.h.s. of (10) follows from multilinearity of the determinant.

ii) and iii) follows easily from the definition.

iv) For $1 \leq \alpha < \infty$, Minkowski's inequality yields

$$\begin{aligned}
\left( \sum_{i<j} |x_i y_j - x_j y_i|^\alpha \right)^{1/\alpha} &\leq \left( \sum_{i<j} |x_i|^\alpha |y_j|^\alpha \right)^{1/\alpha} + \left( \sum_{i<j} |y_i|^\alpha |x_j|^\alpha \right)^{1/\alpha} \\
&\leq \left( \sum_{i,j} |y_i|^\alpha |x_j|^\alpha \right)^{1/\alpha} \\
&= \left( \sum_i |y_i|^\alpha \right)^{1/\alpha} \left( \sum_j |x_j|^\alpha \right)^{1/\alpha} = ||\mathbf{x}||_\alpha ||\mathbf{y}||_\alpha \\
&\leq ||\mathbf{x}||_1 ||\mathbf{y}||_1,
\end{aligned}$$

where the last line follows from the $L_p$ inequality, and which gives the result.

For $\alpha = \infty$, the results follows from

$$|x_i y_j - x_j y_i| \leq \max(x_i y_j, x_j y_i) \leq ||\mathbf{x}||_1 ||\mathbf{y}||_1.$$

v) Assume

$$d_\alpha([\mathbf{x}]_+, [\mathbf{y}]_+) = 0 \Leftrightarrow x_i y_j = x_j y_i \text{ for all } i \neq j \in \{0, 1, \dots, d\}. \tag{11}$$

Set $I = \{i : x_i \neq 0\}$. Since $\mathbf{x} \neq \mathbf{0}$, $I \neq \varnothing$. By permutation invariance iii), one can assume w.l.o.g. that $x_0 \neq 0$. Then, (11) with $i = 0$ yields

$$y_j = x_j \left( \frac{y_0}{x_0} \right), \quad \forall j \neq 0.$$

Set $\lambda := y_0/x_0$. $\lambda \neq 0$ because, if $\lambda = 0$, then, $y_0 = 0$ and the above equation yields $y_j = 0$, $\forall j \neq 0$. This would lead to a contradiction, since $\mathbf{y} \neq \mathbf{0}$. One has thus $\mathbf{y} = \lambda \mathbf{x}$, with $\lambda > 0$, viz. $[\mathbf{x}]_+ = [\mathbf{y}]_+$.

The converse direction is obvious from the anti-symmetry property of the determinant.

[2] written e.g. in the usual order on the integer. By permutation invariance, it does not matter in which order one writes the components of $\mathbf{x}_I$ and $\mathbf{y}_I$, as long the same order is used for both vectors. See also Remark 2.

vi) Both the numerator and the denominator in (10) remain the same if some zeroes are added to the components of $\mathbf{x}$ and $\mathbf{y}$.

vii) One has, for $1 \leq \alpha < \infty$,

$$||\mathbf{x}_I||_1||\mathbf{y}_I||_1 d_\alpha([\mathbf{x}_I]_+,[\mathbf{y}_I]_+) = \left(\sum_{\substack{i<j\\ i,j\in I}} \left|\det\begin{vmatrix} x_i & y_i\\ x_j & y_j\end{vmatrix}\right|^\alpha\right)^{1/\alpha}$$
$$\leq \left(\sum_{i<j} \left|\det\begin{vmatrix} x_i & y_i\\ x_j & y_j\end{vmatrix}\right|^\alpha\right)^{1/\alpha} = ||\mathbf{x}||_1||\mathbf{y}||_1 d_\alpha([\mathbf{x}]_+,[\mathbf{y}]_+),$$

since $I \subset \{0,\ldots,d\}$, and similarly for $\alpha = \infty$.

viii) Assume w.l.o.g that $\mathbf{x},\mathbf{y} \in \Delta_+^d$ are the simplex representatives of $[\mathbf{x}]_+, [\mathbf{y}]_+ \in \mathbb{P}_+^d$, so that $||\mathbf{x}||_1 = ||\mathbf{y}||_1 = 1$. One has $\mathscr{A}_I(\mathbf{x}) = (||\mathbf{x}_I||_1, \mathbf{x}_J)$, so that

$$||\mathscr{A}_I(\mathbf{x})||_1 = ||\mathbf{x}_I||_1 + \sum_{j\in J}|x_j| = ||\mathbf{x}_I||_1 + ||\mathbf{x}_J||_1 = ||\mathbf{x}||_1 = 1.$$

(Amalgamation does not the change the total mass). Thus, by distinguishing between the first and remaining components of $\mathscr{A}_I(\mathbf{x})$, $\mathscr{A}_I(\mathbf{y})$, one has that

$$d_1[(\mathscr{A}_I(\mathbf{x})]_+,[\mathscr{A}_I(\mathbf{y})]_+) = \sum_{j\in J}\left|\det\begin{vmatrix} \sum_{i\in I}x_i & \sum_{i\in I}y_i\\ x_j & y_j\end{vmatrix}\right| + \sum_{\substack{i<j\\ i,j\in J}}\left|\det\begin{vmatrix} x_i & y_i\\ x_j & y_j\end{vmatrix}\right|$$

$$= \sum_{j\in J}\left|\sum_{i\in I}\det\begin{vmatrix} x_i & y_i\\ x_j & y_j\end{vmatrix}\right| + \sum_{\substack{i<j\\ i,j\in J}}\left|\det\begin{vmatrix} x_i & y_i\\ x_j & y_j\end{vmatrix}\right| \tag{12}$$

$$\leq \sum_{j\in J}\sum_{i\in I}\left|\det\begin{vmatrix} x_i & y_i\\ x_j & y_j\end{vmatrix}\right| + \sum_{\substack{i<j\\ i,j\in J}}\left|\det\begin{vmatrix} x_i & y_i\\ x_j & y_j\end{vmatrix}\right| \tag{13}$$

$$\leq \sum_{i<j}\left|\det\begin{vmatrix} x_i & y_i\\ x_j & y_j\end{vmatrix}\right| = d_1([\mathbf{x}]_+,[\mathbf{y}]_+), \tag{14}$$

where (12) follows by multi-linearity of the determinant, (13) by the triangle inequality, and (14) by the inclusion principle.

□

**Remark 2.** Let $k$, $1 \leq k \leq d$, be an integer. Any injection $\iota : \{0,\ldots,k\} \to \{0,\ldots,d\}$ between indices induces an injective linear map between vectors, $\iota* : \mathbb{R}^{k+1} \ni \mathbf{x} \to \iota * \mathbf{x} \in \mathbb{R}^{d+1}$, defined coordinate-wise by

$$(\iota * \mathbf{x})_j = \begin{cases} x_i & \text{if } j = \iota(i) \text{ for some } i \in \{0,\ldots,k\}\\ 0 & \text{otherwise}\end{cases}$$

for $j \in \{0,\ldots,d\}$. Then, properties iii) ($k = d$) and vi) ($k < d$) of Theorem 3.2 can be combined in the single property:

$$d_\alpha([\iota * \mathbf{x}]_+,[\iota * \mathbf{y}]_+) = d_\alpha([\mathbf{x}]_+,[\mathbf{y}]_+)$$

for all $[\mathbf{x}]_+, [\mathbf{y}]_+ \in \mathbb{P}_+^k$ and all injections $\iota : \{0,\ldots,k\} \to \{0,\ldots,d\}$.

3.2. **Discussion.** Properties vi)–viii) relate to the "Principles of CoDa Analysis" of the Log-Ratio School, particularly subcompositional and amalgamation coherence. These principles have sparked significant debate within the compositional data (CoDa) community (e.g., Scealy and Welsh 2014). Recognizing that discussions of CoDa principles can be nuanced, especially in the context of CoDa with zeros, we have strived to present an objective analysis of how the proposed barycentric divergences align or not with these principles. The discussion below, supported by relevant context, allows readers to assess the relevance of these properties for their applications.

3.2.1. *Perturbation invariance for CoDa with zeros.* The Log-Ratio School (John Aitchison 1992) often requires that a measure of compositional difference must satisfy some other properties, in particular compatibility with Aitchison's vector space operations $\oplus, \odot$. For CoDa with zeros, the perturbation operation $\oplus$ is no longer adequate as it is not simply transitive[3]: given $\mathbf{x}, \mathbf{y} \in \Delta_+^d$, since $0.\lambda = 0$ for any $\lambda \in \mathbb{R}$, i) there is no $\mathbf{q} \in \Delta_+^d$ s.t. $\mathbf{y} = \mathbf{x} \oplus \mathbf{q}$ if $y_i > 0$ and $x_i = 0$, for some part $i$, and ii) there is not a unique $\mathbf{q} \in \Delta_+^d$ s.t. $\mathbf{y} = \mathbf{x} \oplus \mathbf{q}$ if $x_i > 0$ and $y_i = 0$, for some part $i$. In other words, for CoDa with zeros, the operation $\oplus$ can not always encode the change between two compositions in an unambiguously way[4]. This suggests the inadequacy, for CoDa with zeros, of distances based on, quoting, "perturbation as the measure of change between two compositions", i.e. inter-compositional ratios $y_i/x_i$, as was argued in John Aitchison 1992. In particular, it makes no sense to require perturbation invariance in a dissimilarity measure for CoDa with zeros.

3.2.2. *Zeroes compositional coherence.* Property vi) (zeroes compositional coherence) of Theorem 3.2 means that the barycentric divergences behave logically with respect to components $x_i, y_i$ that are both absent (have zero weight) as such elements can be ignored. It is a standard requirement in analysis and geometry (e.g. Minkowski's $\ell^p$ norms and distances are compatible with each other as the dimension increase).

In contrast, log coordinates are infinite for CoDa with zeros and Aitchison's distance $d_A$ becomes undefined. There is thus a conceptual difficulty for handling zeroes in the Log-Ratio School, as noted by e.g. Scealy and Welsh 2014: a vector $\mathbf{x} \neq 0$ containing (some) zeroes is a well-defined vector, thus leading by projectivization to a well-defined CoDa projective point $[\mathbf{x}]_+$, but is not a valid basis for a composition in Aitchison's geometry. For example, J. Park, Yoon, et al. 2022 construct, for some $\mathbf{x} \in \Delta_+^d$ on the boundary of the simplex (i.e. with a zero), two sequences of CoDa points $\mathbf{x}_n, \mathbf{y}_n \in \Delta_{++}^d$ in the interior of the simplex (thus zero-free so that $d_A(\mathbf{x}_n, \mathbf{y}_n) < \infty$), s.t. $\mathbf{x}_n \to \mathbf{x}$, $\mathbf{y}_n \to \mathbf{x}$, yet with $d_A(\mathbf{x}_n, \mathbf{y}_n) \to \infty$, as $n \to \infty$.

The trick to avoid zeroes in the Log-Ratio School is to resort to sub-setting (i.e. taking a positive sub-vector). However, the subsetting operation $\pi_I : \mathbb{R}^{d+1} \ni \mathbf{x} \mapsto \mathbf{x}_I \in \mathbb{R}^{\operatorname{card}(I)}$ is not an endomorphism, as the codomain $R^{\operatorname{card}(I)}$ of $\pi_I$ is not the same as the domain $\mathbb{R}^{d+1}$, and is thus not a projection in the sense of linear

[3]Axiom T1 in John Aitchison 1992.

[4]In contrast, in affine geometry, any two points, regardless of the presence of zeros in their coordinates, can always be fully related uniquely by the vector made of their difference. See Appendix A.

algebra[5]. The point here is that subcompositional coherence is often motivated in the Log-Ratio School via the idea that "subcompositions [...] should behave like orthogonal projections in real analysis." (Vera Pawlowsky-Glahn, Juan José Egozcue, and Tolosana-Delgado 2015 p. 16). This results in logical discrepancies between projected elements (thus with zeros for components in the kernel) and between subsetted elements (Scealy and Welsh 2014).

Ad-hoc zero replacements methods with some small $\epsilon > 0$, though popular in the Log-Ratio School, introduce instabilities and distortions of the data: as $|\ln \epsilon| \to \infty$ and the derivative of the ln is unbounded when $\epsilon \to 0$, zero-replaced components are likely to oversize non-zero ones and small variation near zeroes to produce large fluctuations in log-coordinates. Such issues compound when the number of zeros increases, e.g. for high-dimensional microbiome data (J. Park, Ahn, and C. Park 2023). Empirical studies (Nearing et al. 2022, Greenacre 2021) have confirmed those issues by showing that data analysis become inconsistent depending on how the zeros are replaced. These zero-replacement methods are often motivated for data with rounded zeroes. However, the distinction between essential and rounded zeroes is not a mathematical one: a zero is a zero, mathematically, and log ratios are undefined with zeros.

Consequently, it is becoming increasingly recognized (Scealy and Welsh 2014, Greenacre 2021, J. Park, Yoon, et al. 2022, J. Park, Ahn, and C. Park 2023, etc.) that an adequate treatment of zeros require new approaches departing from Aitchison's framework.

3.2.3. *Subcompositional dominance.* Property vii) of Theorem 3.2 is a weak form of subcompositional dominance: barycentric divergences are subcompositionally dominant, when they are properly weighted by the corresponding product of the total weight (amalgamation total) of the (sub)-barycenters. Such property can be thought as the global analogue of formula (9) of Remark 1, where the displacement vector writes as a weighted sum of difference of (subcompositionally coherent) ratios.

Property vii) reflect the fact that the operations of Subcomposition and Amalgamation work in conjunction from the affine geometry perspective, as was discussed in Section 2.1: per (5), CoDa simplex representative $\mathbf{x}, \mathbf{y} \in \Delta_+^d$ are the affine points $\mathbf{x} = \sum_{i=0}^d x_i A_i$, $\mathbf{y} = \sum_{i=0}^d y_i A_i$ in normalized barycentric coordinates. $\mathbf{x}$ and $\mathbf{y}$ then split into two sub-barycenters in $I$ and its complement $J := \{0, \ldots, d\} \setminus I$, i.e.

$$\mathbf{x} = \left(\sum_{i\in I} x_i A_i\right) + \left(\sum_{i\in J} x_i A_i\right)$$
$$\mathbf{y} = \left(\sum_{i\in I} y_i A_i\right) + \left(\sum_{i\in J} y_i A_i\right).$$

If $||\mathbf{x}_I||_1 \neq 0$, and $||\mathbf{y}_I||_1 \neq 0$, the $I$-sub-barycenters $\sum_{i\in I} x_i A_i$, $\sum_{i\in I} y_i A_i$ write as

$$\sum_{i\in I} x_i A_i = ||\mathbf{x}_I||_1 \sum_{i\in I} \frac{x_i}{||\mathbf{x}_I||_1} A_i$$
$$\sum_{i\in I} y_i A_i = ||\mathbf{y}_I||_1 \sum_{i\in I} \frac{y_i}{||\mathbf{y}_I||_1} A_i, \tag{15}$$

[5]It is not idempotent as it can not be composed with itself.

and thus have total mass (which corresponds to amalgamation) $||\mathbf{x}_I||_1$, $||\mathbf{y}_I||_1$ and write in *normalised* barycentric coordinates as $\mathbf{x}_I/||\mathbf{x}_I||_1$, $\mathbf{y}_I/||\mathbf{y}_I||_1$ (which corresponds to subcompositions of the Log-Ratio School).

Therefore, from the affine geometry perspective, CoDa are *weighted* points, with weight the amalgamation total, and composition the *normalized* barycentric coordinate. Thus, evaluating displacements and divergences between sub-barycenters must take into account both their normalized coordinates (the subcomposition) and their weights (their amalgamation total) in order to maintain coherence in equations (15), i.e. between sub-setting/projection on the l.h.s. and amalgamation+subcomposition on the r.h.s. In contrast, the literature often considers amalgamation dominance and subcomposition dominance as opposing properties, see e.g. Scealy and Welsh 2014.

3.2.4. *On Subcompositional coherence.* Even for data without zeroes, it has been argued in the literature that subcompositional coherence may not be a desirable property (Scealy and Welsh 2014, J. Park, Ahn, and C. Park 2023, Gajer and Ravel 2025). The verbal definition of Subcompositional coherence as "agreement between statements on a subcomposition whether they are obtained from analyzing the full composition or just the subcomposition" is vague (Scealy and Welsh 2014), who further argue that "it is not satisfied by [...] log-ratio methods". A formalization of Subcompositional coherence is given in Gajer and Ravel 2025, who show that clr is not subcompositionally coherent (while alr is) in their sense. J. Park, Ahn, and C. Park 2023 construct a (toy) example in the context of microbiome data, where the presence of a disease is activated by the deficiency (i.e. small value) of some taxa. They show that when passing to the subcomposition of the incriminated taxa, the "relative abundance to the total is lost when taking a subcomposition" and may yield uninformative data regarding the absence or presence of the disease. This motivates their subsequent use of amalgamation coherence instead as a guiding principle, as advocated by Greenacre 2020, Scealy and Welsh 2014.

Property viii) of Theorem 3.2 shows that the 1-barycentric divergence is amalgamation dominant. See also Section 7, for an empirical study of amalgamation based on barycentric variance matrices.

3.3. **The case $\alpha = 2$.** One has thus obtained a family of symmetric, permutation-invariant, bounded divergences on the full Coda space (i.e. also for CoDa with zeroes). Among all possible divergences, noticeable cases occur for $\alpha = 1, 2, \infty$. Indeed, for $\alpha = 2$, the divergence write as follows:

**Lemma 3.3.** *The $2-$barycentric divergence writes as*

$$d_2([\mathbf{x}]_+, [\mathbf{y}]_+) = \frac{\sqrt{||\mathbf{x}||_2^2||\mathbf{y}||_2^2 - \langle \mathbf{x}|\mathbf{y}\rangle^2}}{||\mathbf{x}||_1||\mathbf{y}||_1} \tag{16}$$

$$= \frac{||\mathbf{x}||_2||\mathbf{y}||_2}{||\mathbf{x}||_1||\mathbf{y}||_1} \sin\theta_{\mathbf{xy}}, \tag{17}$$

*where $\theta_{\mathbf{xy}} \in [0, \pi/2]$ is the acute angle between the rays $[\mathbf{x}]_+$ and $[\mathbf{y}]_+$.*

*Proof.* By symmetry,

$$\begin{aligned}\sum_{i<j}(x_iy_j - x_jy_i)^2 &= \frac{1}{2}\sum_{i,j}(x_iy_j - x_jy_i)^2\\ &= \frac{1}{2}\sum_{i,j}\left(x_i^2y_j^2 + x_j^2y_i^2 - 2x_iy_ix_jy_j\right)\\ &= \frac{1}{2}\left(\sum_i x_i^2\sum_j y_j^2 + \sum_i x_i^2\sum_j y_j^2 - 2\sum_i x_iy_i\sum_j x_jy_j\right)\\ &= ||\mathbf{x}||_2^2||\mathbf{y}||_2^2 - \langle\mathbf{x}|\mathbf{y}\rangle^2,\end{aligned}$$

which gives (16). Together with $0 \leq \langle\mathbf{x}|\mathbf{y}\rangle = ||\mathbf{x}||_2||\mathbf{y}||_2\cos\theta_{\mathbf{xy}}$, with $\theta_{\mathbf{xy}} \in [0, \pi/2]$, it yields formula (17). □

**Remark 3.**

i) Since $||\mathbf{x}||_2 \leq ||\mathbf{x}||_1$, formula (17) implies that

$$0 \leq d_2([\mathbf{x}]_+, [\mathbf{y}]_+) \leq 1,$$

which gives, for the case of the 2-divergence, another proof of the upper bound in Theorem 3.2 iv).

ii) Formula (17) involves the sine of the (acute) angle between rays $[\mathbf{x}]_+$ and $[\mathbf{y}]_+$. Since $||\mathbf{x}||_2 \leq ||\mathbf{x}||_1 \leq \sqrt{d+1}||\mathbf{x}||_1$, one has that

$$\frac{\sin\theta_{\mathbf{xy}}}{d+1} \leq d_2([\mathbf{x}]_+, [\mathbf{y}]_+) \leq \sin\theta_{\mathbf{xy}},$$

thus $d_2$ is Lipshitz-equivalent to the sine distance on rays.

The dissymmetry between the $||.||_2$ and $||.||_1$ norms in the fraction in (17) suggests that we (should) eliminate this ratio-of-norms coefficient by replacing in the definition (17) the denominator $||\mathbf{x}||_1||\mathbf{y}||_1$ by $||\mathbf{x}||_2||\mathbf{y}||_2$. This amounts to normalizing the CoDa elements $[\mathbf{x}]_+$, $[\mathbf{y}]_+$ by the $||.||_2$ norm instead of the $||.||_1$ norm, i.e. to replace the closure operation (1) which radially projects the ray $[\mathbf{x}]_+$ on the simplex by a radial projection on the unit sphere. Such a step will be performed in Faugeras 2025a, and justified mathematically from the projective viewpoint. In addition, such a change will lead to improved properties.

iii) Let $A = \begin{pmatrix}\mathbf{x}\,\mathbf{y}\end{pmatrix} \in \mathbb{R}^{(d+1)\times 2}$ be the matrix with columns $\mathbf{x}, \mathbf{y}$. Then, formula (16) writes

$$d_2([\mathbf{x}]_+, [\mathbf{y}]_+) = \frac{\sqrt{\det(A^TA)}}{||\mathbf{x}||_1||\mathbf{y}||_1}$$

Hence, the numerator is the square root of the determinant of the Gram matrix,

$$A^TA = \begin{pmatrix}\langle\mathbf{x}|\mathbf{x}\rangle & \langle\mathbf{x}|\mathbf{y}\rangle\\ \langle\mathbf{x}|\mathbf{y}\rangle & \langle\mathbf{y}|\mathbf{y}\rangle\end{pmatrix},$$

and thus interprets geometrically to the area of the parallelogram spanned by $\{\mathbf{x}, \mathbf{y}\}$. Thus, for simplex representatives $\mathbf{x}, \mathbf{y} \in \Delta_+^d$, $d_2([\mathbf{x}]_+, [\mathbf{y}]_+)$ interprets geometrically has twice the area of the triangle $\boldsymbol{Oxy}$, with $\mathbf{O}$ the origin of the ambient vector space $\mathbb{R}^{d+1}$. In addition, this form allows to re-derive some of the properties of Theorem 3.2, and more importantly, suggests the more abstract approach of Faugeras 2025a.

3.4. **Infinite dimensional version.** As a CoDa element is simply a discrete probability distribution on a finite number of locations whose locations are forgotten (see, e.g., Faugeras 2026), Definition 3.1 can be generalized to general probability measures and even to $\sigma-$finite positive measures $P, Q$ on some measurable space $(\Omega, \mathcal{A})$, with $0 < P(\Omega), Q(\Omega) < \infty$. (Infinite dimensional versions of CoDa vector spaces are called Bayes space in the CoDa literature, see, e.g., J. J. Egozcue, Díaz-Barrero, and V. Pawlowsky-Glahn 2006). Let $\mu$ be a measure dominating $P$ and $Q$, (e.g., $\mu = (P+Q)/2$). By Radon-Nikodym's theorem, $P$, $Q$ have densities $f = \frac{dQ}{d\mu}$, $g = \frac{dQ}{d\mu}$. Let $L_\alpha(\Omega, \mathcal{A}, \mu)$ be the Lebesgue space of $(\Omega, \mathcal{A})$ measurable functions $h$ s.t. $\int |h|^\alpha d\mu < \infty$. One can then define in such general case the following symmetric divergence:

**Definition 3.4.** Let $\alpha \geq 1$. If $P$, resp. $Q$, with densities $f$, resp. $g$ are such that $f, g \in L_\alpha(\Omega, \mathcal{A}, \mu)$, then the finite symmetric divergence,

$$D_\alpha(P, Q) := \frac{\left(\iint |f(x)g(y) - f(y)g(x)|^\alpha \mu(dx) \times \mu(dy)\right)^{1/\alpha}}{P(\Omega)Q(\Omega)}$$

which reduces to the numerator in case $P$ and $Q$ are probability measures, is well defined.

*Proof.* From the inequality, $|a-b|^\alpha \leq 2^{\alpha-1}(|a|^\alpha + |b|^\alpha)$, (which itself follows from convexity of $x \mapsto |x|^\alpha$ for $\alpha \geq 1$), one has that

$$|f(x)g(y) - f(y)g(x)|^\alpha \leq 2^{\alpha-1}(|f(x)|^\alpha |g(y)|^\alpha + |f(y)|^\alpha |g(x)|^\alpha).$$

Hence, if $||f||_\alpha, ||g||_\alpha < \infty$, then the numerator is finite. □

To our knowledge, such divergence has not been introduced before in the probabilistic literature. It allows to compare measures with different total masses[6] and possibly different supports. In contrast, for the infinite dimensional Aitchison distance, the measures must have the same null sets, i.e. positive densities, while information divergences (ratio-)based on the density of $P$ w.r.t. $Q$, like the Kullback-Leibler divergence (relative entropy), require absolute continuity of $P$ w.r.t. $Q$ to be finite.

## 4. Fréchet means based on the barycentric divergences

4.1. **Definitions and basic properties.** Having a notion of divergence between CoDa points, one can now define notions of center and measures of dispersion of a cluster of points, following the metric approach of Fréchet 1948 (See also Faugeras 2023 Section 7, and M. T. Tsagris, Preston, and Wood 2011). Indeed, given a CoDa sample $[\mathbf{x}^1]_+, \ldots, [\mathbf{x}^n]_+ \in \mathbb{P}_+^d$, for any Coda point $[\mathbf{m}]_+ \in \mathbb{P}_+^d$, the following functional

$$\mathcal{F}_{\alpha,\beta}([\mathbf{m}]_+) := \sum_{i=1}^n d_\alpha^\beta([\mathbf{m}]_+, [\mathbf{x}^i]_+), \tag{18}$$

with $1 \leq \alpha \leq \infty$, $\beta > 0$, gives a measure of the outlyingness[7] (i.e. is a depth function) of the point $[\mathbf{m}]_+$ w.r.t. the data points $[\mathbf{x}^1]_+, \ldots, [\mathbf{x}^n]_+$. Minimizing

[6] More precisely, it allows to compare the compositional part of measures, in case they have unequal total mass.

[7] or, following Huygens' terminology in mechanics of solids, is the inertia of the data set relative to the point $[\mathbf{m}]_+$.

such functional over the whole CoDa space thus gives a notion of central point, and the value of $\mathcal{F}_{\alpha,\beta}$ at a minimum gives a measure of dispersion of the cloud of points. We thus introduce the following definition:

**Definition 4.1.** An empirical $(\alpha, \beta)$-barycentric Fréchet mean is defined as a minimizer of (18). It is simply called an $\alpha$-barycentric Fréchet mean for $\beta = \alpha$.

Among possible choices for $\beta$, one can consider three interesting cases, i) $\beta = 2$, ii) $\beta = 1$ and iii) $\beta = \alpha$. Case i) gives the well-known Fréchet mean, but for the different $\alpha$-divergences of equation (10). The notion of Fréchet mean is a natural extension of the standard mean in (semi-)metric spaces. In particular, the case $\alpha = \beta = 2$ leads to a quadratic program, and the resolution of a linear system. Case ii) gives a notion of spatial median, again for the different $\alpha$-divergences considered, which is a more robust version of center than the Fréchet mean, but is usually more computationally difficult. Case iii) is a sort of generalized Fréchet mean, which gives the median for the 1-divergence, the Fréchet mean for the $2-$divergence. By removing the power of the outer bracket in the determinantal formula (10), it somehow appears as a natural choice.

The following theorem easily ensues.

**Theorem 4.2.** *i) For $1 \leq \alpha \leq \infty$, $\beta > 0$, a $(\alpha, \beta)$-barycentric Fréchet mean always exists.*

*ii) For $\beta \geq \alpha \geq 1$ and $\alpha = \infty, \beta \geq 1$, stationary points of (18) are $(\alpha, \beta)$-barycentric Fréchet means.*

*iii) For $\infty > \beta \geq \alpha > 1$, a $(\alpha, \beta)$-barycentric Fréchet mean is unique.*

*Proof.* i) Restricting to simplex representatives in $\Delta_+^d$, the functional writes (18) as a sum of (absolute values of) powers on a compact convex space, hence is continuous, and Weierstrass theorem ensues that minimizers always exists.

ii) The functional (18) is convex (for $\beta \geq \alpha$ and $\alpha = \infty, \beta \geq 1$), so that local minima are global ones.

iii) For $\infty > \beta \geq \alpha > 1$, (18) is strictly convex.

□

It is worth noting some advantages of such means/medians: compared to the classical Aitchison's mean, which is based on log-ratios and results in the geometric mean, the proposed approach enables handling situations where zeroes are present in the composition. Compared to the Fréchet means/medians based on Hilbert' projective metric, introduced in Faugeras 2023, characterizing and computing the barycentric means/medians is a much easier convex problem. In particular, the case $\alpha = \beta = 2$ stands out as particularly appealing since unicity is guaranteed and the problem reduces to a simple basic quadratic program.

4.2. **Discussion of variants and extensions.** One can consider useful variants and extensions of the Fréchet mean based on barycentric divergences, which we now discuss.

i) The barycentric means/medians are based on the minimization of the Fréchet functional (18) over the whole CoDa space $\mathbb{P}_+^d$, (equivalently, $\Delta_+^d$). One can also consider several versions of so-called medoids (see e.g. Kaufman and Peter J. Rousseeuw 1990) where the minimization of (18 is restricted to

the finite set of data points $[\mathbf{x}^1]_+, \ldots, [\mathbf{x}^n]_+$. This allows to obtain a center which is always a member of the data set. This can be useful in cases where the data set has a special geometric structure (e.g. sits on a line, a curve, or more generally a manifold), and one wants to ensure that the central point is representative of the data structure. It is also beneficial in terms of interpretability of the center. See Examples B.4 and B.5.

In addition, the computation reduces to a discrete optimization problem, i.e. computing all pairwise divergences between points and identification of the minimal one. This requires at most $O(n^2)$ distance evaluations, and there exists some algorithms (D. Wang and Eppstein 2006, Baharav and Tse 2019) which allows to reduce the number of distance evaluations to an almost linear time. This can be crucial for (moderately) large datasets, where computation time is the main bottleneck of the method.

ii) The Fréchet mean/median/medoid look for a single central point. It can be generalized to $k$-mean/median/medoid clustering (see e.g. Everitt et al. 2011, Simovici 2021) which ask for the location of $k$ cluster centers and a partition the $n$ observations into $k$ cluster sets, $\mathcal{W} = \{S_1, S_2, ..., S_k\}$, so as to minimize the sum of $\beta$-powers of the $\alpha$-divergences from each sample point to its nearest cluster: the objective is to find

$$\arg\min_{\mathcal{W}} \sum_{j=1}^{k} \sum_{[\mathbf{x}]_+ \in S_j} d_\alpha^\beta([\mathbf{x}]_+, [\mathbf{m}^i]_+)$$

where $[\mathbf{m}^i]_+$ is itself the Fréchet mean/median

$$[\mathbf{m}^i]_+ := \arg\min_{[\mathbf{m}]_+ \in \mathbb{P}_+^d} \sum_{[\mathbf{x}]_+ \in S_i} d_\alpha^\beta([\mathbf{x}]_+, [\mathbf{m}]_+).$$

Several variants/algorithms can be envisioned. In particular, the forthcoming Definition 5.1 of Generalized Gaussian distribution based on the barycentric divergence (10) suggests to investigate Expectation-Minimization clustering algorithms based on a model of mixture of such barycentric Gaussian distributions. We leave this suggestion for further research.

iii) The measure of the Fréchet functional (18) at the minimum gives a measure of the global variability of the cloud of CoDa points. An alternative robust measure of dispersion of a cloud of CoDa points can be obtained by replacing the sum in (18), calculated at the mean/median point $[\mathbf{m}]_+$, by the median. In particular, taking $\alpha = \beta = 1$ gives the Mean Absolute Deviation, defined as

$$\text{MAD} = \text{Median } \left(d_1([\mathbf{x}^i]_+, \text{Med}([\mathbf{x}^1]_+, \ldots, [\mathbf{x}^n]_+))\right),$$

where $\text{Med}([\mathbf{x}^1]_+, \ldots, [\mathbf{x}^n]_+)$ is a Fréchet median of the data points based on the $1-$barycentric divergence ($\beta = \alpha = 1$). See, e.g., Peter J Rousseeuw and Croux 1993, Gauss 1816.

iv) A further generalization is provided by minimizing a weighted version of the Fréchet functional (18). This corresponds to considering weighted versions of the barycentric divergences, see Definition 5.2 in Section 5. In particular, if the weight depends on some covariate, this allows to perform Fréchet regression of CoDa, see Section 7.

Further study and numerical illustrations of the various Fréchet means based on barycentric divergences, for varying values of $\alpha$ and $\beta$, are provided in Appendix B.

## 5. Barycentric generalized Laplace-Gaussian distribution

5.1. **Isotropic generalized Laplace-Gaussian distributions.** As in Faugeras 2023, one can define a family of generalized Gaussian distributions, based on the family of divergences (10).

**Definition 5.1.** A random $[\mathbf{X}]_+ \in \mathbb{P}_+^d$ follows a Generalized Barycentric Gaussian distribution with parameters $([\mathbf{m}]_+, \sigma, \alpha) \in \mathbb{P}_+^d \times \mathbb{R}_{++} \times [1, \infty]$, if its distribution admits a density w.r.t. to the uniform measure $\nu$ on $\mathbb{P}_+^d$ given by

$$f_\alpha([\mathbf{x}]_+; [\mathbf{m}]_+, \sigma) := Z_\alpha^{-1}([\mathbf{m}]_+, \sigma) \exp\left(-\left(\frac{d_\alpha([\mathbf{x}]_+, [\mathbf{m}]_+)}{\sigma}\right)^\alpha\right) \tag{19}$$

for $1 \leq \alpha < \infty$, and by

$$f_\infty([\mathbf{x}]_+; [\mathbf{m}]_+, \sigma) := Z_\infty^{-1}([\mathbf{m}]_+, \sigma) \exp\left(-\frac{d_\infty([\mathbf{x}]_+, [\mathbf{m}]_+)}{\sigma}\right), \quad \alpha = \infty,$$

where $Z_\alpha([\mathbf{m}]_+, \sigma)$ is a normalizing constant.

Definition 5.1 gives an analogue of the Gaussian and Laplace distributions when $\alpha = 2$ and $\alpha = 1$, respectively, with $[\mathbf{m}]_+$ a mean parameter and $\sigma$ a dispersion parameter. Figure 1 shows density ternary plots (i.e. for $d = 2$) of such distributions with a centered and non-centered mean parameter, in the noticeable cases $\alpha = 1, 2, \infty$. The level sets of the density shows the geometry of the balls for the corresponding divergence. In particular, $\alpha = 2$ give the usual Euclidean distance geometry, but truncated on the (full) simplex. The cases $\alpha = 1$ and $\alpha = \infty$ give polygonal balls. It is interesting to note that the level sets of the centered distributions (i.e., for $\mathbf{m} = \mathbf{1}$) have the same (truncated) hexagonal shape for $\alpha = 1$ and $\alpha = \infty$, which is also similar to the balls in Hilbert projective distance, see Faugeras 2023. The level sets of the non-centered distribution for $\alpha = \infty$ (lower right panel) appears, due to the truncation, as (part of) lozenges.

5.2. **Anisotropic generalized Laplace-Gaussian distributions.** As a further generalization, one can consider weighted versions of the barycentric divergences (10) and corresponding Gaussian-type distributions of Definition 5.1. Indeed, in classical multivariate analysis, when measuring distance between vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{d+1}$ made of heterogeneous components, it is common to standardize the variables by their standard deviation, in order to balance out the contributions of each variable. This corresponds to measuring the distance between two sample elements $\mathbf{x}, \mathbf{y}$ with the Standardized Euclidean distance

$$d_{\mathbf{S}}(\mathbf{x}, \mathbf{y}) := \sqrt{\sum_{i=0}^{d} \left(\frac{x_i}{s_i} - \frac{y_i}{s_i}\right)^2},$$

where $\mathbf{s} = (s_0, \ldots, s_d)$, with $s_i$ the standard deviation of the $i$th variable. A generalization of this principle, taking into account the correlation of the data, leads to the definition of the well-known Mahalanobis distance,

$$d_\Sigma(\mathbf{x}, \mathbf{y}) := \sqrt{(\mathbf{x} - \mathbf{y})^T \Sigma^{-1} (\mathbf{x} - \mathbf{y})},$$

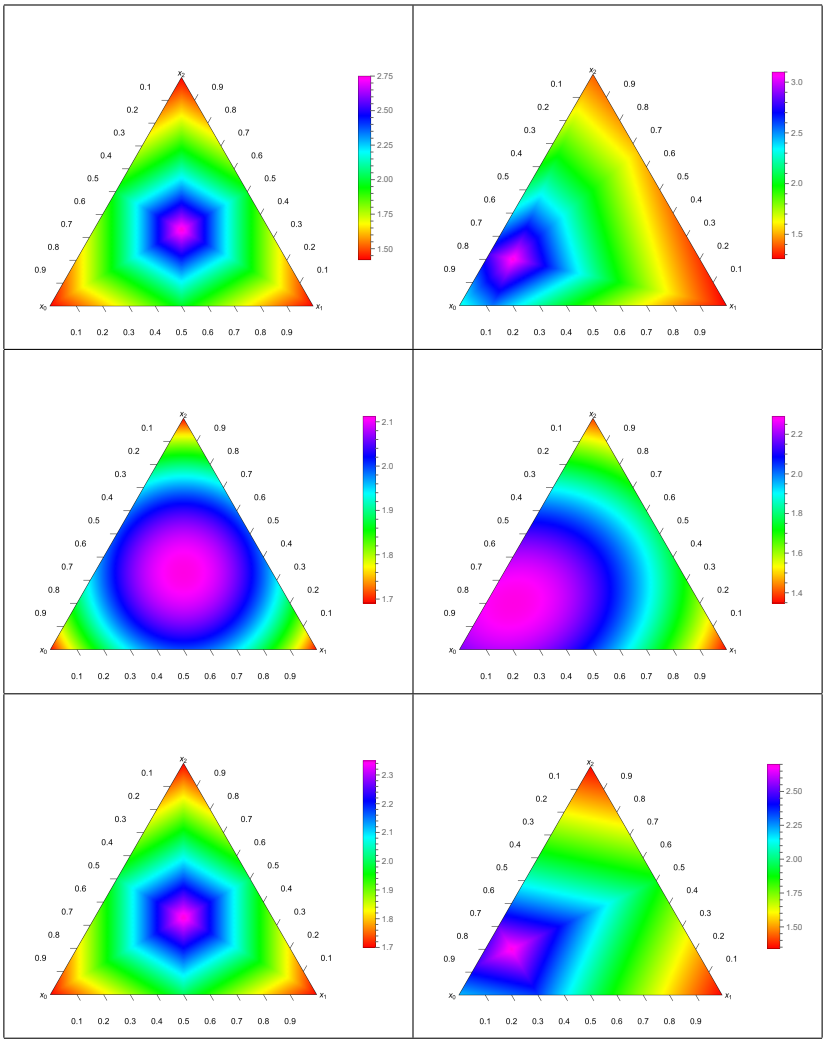

FIGURE 1. Generalised Barycentric Gaussian distributions with $\alpha$-divergence. Left column: centered distribution with $[\mathbf{m}]_+ = [1 : 1 : 1]_+$. Right column: a non-centered distribution with $\mathbf{m} = (0.7, 0.1, 0.2)$. $\alpha = 1$ (up), $\alpha = 2$ (center), $\alpha = \infty$ (down). $\sigma = 1$.

where $\Sigma$ is the covariance matrix.

Here, we can apply this idea to CoDa, as follows:

**Definition 5.2.** Let $W \in \mathbb{R}^{(d+1)\times(d+1)}$ be a symmetric matrix, with positive components $w_{ij} > 0$, and null diagonal. Let $[\mathbf{x}]_+, [\mathbf{y}]_+ \in \mathbb{P}_+^d$ be two CoDa elements. Then, the $W$-weighted barycentric $\alpha$-divergence is defined, for $1 \leq \alpha < \infty$, as

$$d_{\alpha,W}([\mathbf{x}]_+, [\mathbf{y}]_+) := \frac{\left(\sum_{i<j} w_{ij}^{-1} \left|\det \begin{vmatrix} x_i & y_i \\ x_j & y_j \end{vmatrix}\right|^\alpha\right)^{1/\alpha}}{||\mathbf{x}||_1 ||\mathbf{y}||_1} \tag{20}$$

and, for $\alpha = \infty$, as

$$d_{\infty,W}([\mathbf{x}]_+, [\mathbf{y}]_+) := \frac{\max_{i<j}\left(w_{ij}^{-1}\left|\det\begin{vmatrix} x_i & y_i \\ x_j & y_j \end{vmatrix}\right|\right)}{||\mathbf{x}||_1||\mathbf{y}||_1}.$$

It is easily seen that the $W$-weighted barycentric $\alpha$-divergence satisfy all properties of Theorem 3.2, with the exception that the upper bound 1 has to be replaced by the minimal weight.

The corresponding Generalized Laplace-Gaussian distribution is defined analogously to (19), now with the added parameter matrix $W$. Without loss of generality, one can constrain $W$ further by requiring that $\sum_{i<j} w_{ij} = 1$. This allows to interpret the parameters as follows: $W$ controls the shape of the balls in weighted $\alpha$-divergence, while the $\sigma$ parameter measures their overall size. Thus, we define

$$\mathcal{W}_0 = \{W \in \mathbb{R}^{(d+1)\times(d+1)} : w_{ij} = w_{ji} > 0, i \neq j; w_{ii} = 0; \sum_{i<j} w_{ij} = 1\}$$

as the resulting set of constrained symmetric weight matrices with zero diagonal and positive weights.

**Definition 5.3.** A random $[\mathbf{X}]_+ \in \mathbb{P}_+^d$ is said to follow a Generalized Weighted Barycentric Gaussian distribution with parameters $([\mathbf{m}]_+, W, \sigma, \alpha) \in \mathbb{P}_+^d \times \mathcal{W}_0 \times \mathbb{R}_{++} \times [1, \infty]$, if its distribution admits a density w.r.t. to the uniform measure $\nu$ on $\mathbb{P}_+^d$ given by

$$f_\alpha([\mathbf{x}]_+; [\mathbf{m}]_+, W, \sigma) := Z_\alpha^{-1}([\mathbf{m}]_+, W, \sigma) \exp\left(-\left(\frac{d_{\alpha,W}([\mathbf{x}]_+, [\mathbf{m}]_+)}{\sigma}\right)^\alpha\right) \tag{21}$$

for $1 \leq \alpha < \infty$, and, for $\alpha = \infty$, by

$$f_\infty([\mathbf{x}]_+; [\mathbf{m}]_+, W, \sigma) := Z_\infty^{-1}([\mathbf{m}]_+, W, \sigma) \exp\left(-\frac{d_\infty([\mathbf{x}]_+, [\mathbf{m}]_+)}{\sigma}\right),$$

where $Z_\alpha([\mathbf{m}]_+, W, \sigma)$ is a normalizing constant.

Figure 2 illustrates Definition 5.3 for the case $\alpha = 2$. The left column shows a centered Generalized Weighted Barycentric Gaussian distribution, i.e. with mean parameter $[\mathbf{m}]_+ = [1 : 1 : 1]_+$. The upper, resp., lower, left panels have shape parameter

$$W = \begin{pmatrix} 0 & 0.8 & 0.1 \\ 0.8 & 0 & 0.1 \\ 0.1 & 0.1 & 0 \end{pmatrix}, \quad \text{resp. } W = \begin{pmatrix} 0 & 0.1 & 0.8 \\ 0.1 & 0 & 0.1 \\ 0.8 & 0.1 & 0 \end{pmatrix}.$$

Compared to Figure 1, one has elongated the balls in barycentric 2-divergence in the direction having the higher weight 0.8, that is the contour plots obtained are ellipses stretched out in the $\overrightarrow{A_0A_1}$ direction (up) and $\overrightarrow{A_0A_2}$ direction (down panel). The right column shows the same distributions as the left column, but with a non-centered mean, viz. $\mathbf{m} = (0.7, 0.1, 0.2)$, so one can also compare with the isotropic distributions of Figure 1.

We provide in the Supplementary Material (Faugeras 2025b), Section 1, some supplementary simulations of these generalized weighted barycentric Gaussian distribution, in the cases $\alpha = 1, \infty$, for illustration and comparison purposes.

We have thus obtained an analogue of the multivariate Laplace-Gaussian distribution and its generalizations on the whole CoDa space $\mathbb{P}_+^d$: $\alpha$ sets the general form

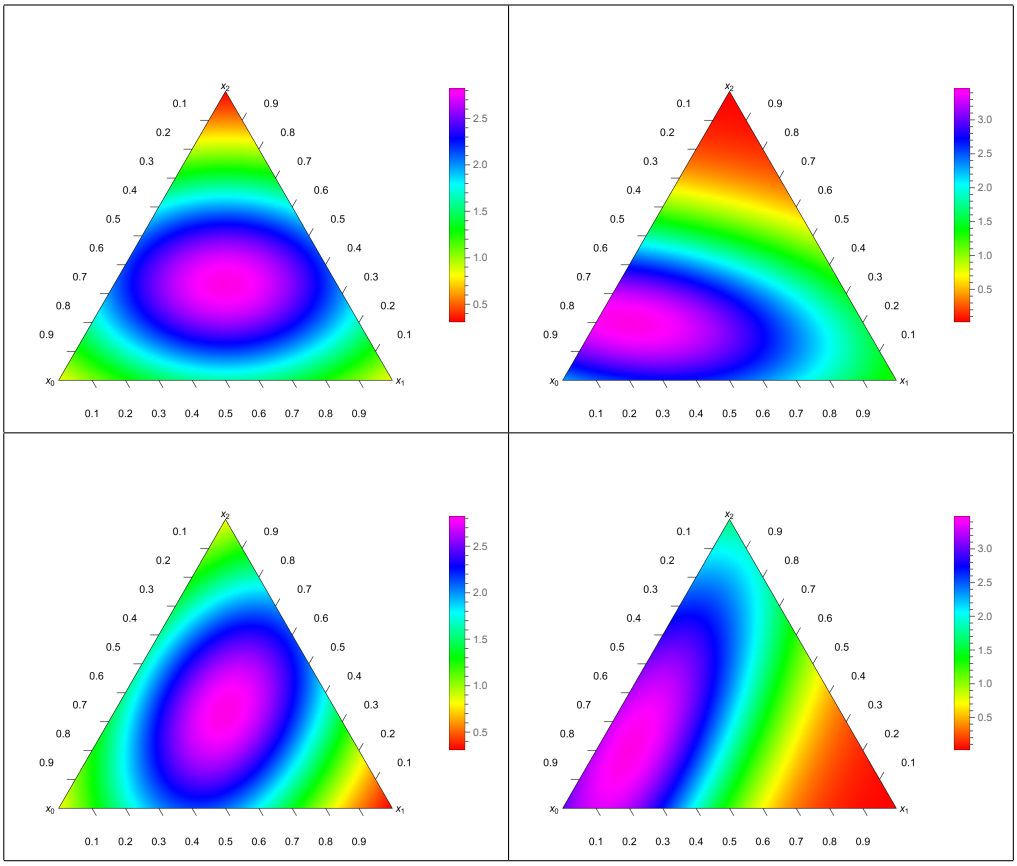

FIGURE 2. Generalised Weighted Barycentric Gaussian distributions with 2-divergence. $\sigma = 1$. Left column: centered distribution with $\mathbf{m} = (1 : 1 : 1)$. Right column: a non-centered distribution with $\mathbf{m} = (0.7, 0.1, 0.2)$. $(w_{01}, w_{02}, w_{12}) = (0.8, 0.1, 0.1)$ (up), $(w_{01}, w_{02}, w_{12}) = (0.1, 0.8, 0.1)$ (down).

of the balls in $\alpha$-divergence, $[\mathbf{m}]_+$ is the location parameter, and $(W, \sigma)$ the dispersion ones. Such distributions can thus be made to accommodate for a large variety of shapes of the data points and should prove useful for modeling and estimation purposes. For example, one could build nonparametric (density or regression) estimators with a kernel based on such distributions. We leave this investigation, together with inference issues of the parameters of the distribution, for further research.

**Remark 4** (Weighted Hilbert projective distance)**.** We remark that the idea of weighting the components entering in the formula of the barycentric $\alpha$-divergence can also be applied to Hilbert's projective metric (see Faugeras 2023). We can thus generalize the latter and define the weighted Hilbert projective metric as follows:

**Definition 5.4.** Let $W \in \mathcal{W}_0$. The weighted Hilbert projective metric on $\mathbb{P}^d_{++}$, with weight matrix $W$, is defined as

$$d_{H,W}([\mathbf{x}]_+, [\mathbf{y}]_+) := \max_{0\leq i<j\leq d} w_{ij} \left| \ln \frac{x_i}{x_j} - \ln \frac{y_i}{y_j} \right|.$$

In turn, the corresponding Gaussian type distribution is defined analogously to Definition 5.3 and Definition 7.4 in Faugeras 2023, with $d_{H,W}$ replacing $d_{\alpha,W}$ in (21). It also gives risen to an anisotropic Generalized Gaussian type distribution, this time based on the weighted version of Hilbert's projective metric. The Supplementary Material (Faugeras 2025b), Section 1, provides some simulations for illustration purposes.

## 6. Variance and covariance matrices

We now turn our attention to our second major contribution, the definition of a notion of covariance matrix for CoDa, based on the barycentric/affine viewpoint.

6.1. **Definitions.** The basic idea is to construct a covariance matrix based on averaged scalar product of displacement vectors, as in the Euclidean vector case, but now with taking into account the affine nature of the data points expressed in barycentric coordinates.

More precisely, let $[\mathbf{x}]_+, [\mathbf{y}]_+ \in \mathbb{P}^d_+$ be a pair of random CoDa (projective) points. Assume one has some corresponding (deterministic) mean points

$$[\boldsymbol{\mu}^{\mathbf{x}}]_+ = [\mu^{\mathbf{x}}_0 : \ldots : \mu^{\mathbf{x}}_d]_+, \quad \text{resp. } [\boldsymbol{\mu}_{\mathbf{y}}]_+ = [\mu^{\mathbf{y}}_0 : \ldots : \mu^{\mathbf{y}}_d]_+.$$

A priori, one could consider a variety of mean points, such as the arithmetic mean (centroid), the geometric (Aitchison) mean, the Fréchet-Hilbert mean (Faugeras 2023), the $(\alpha, \beta)$-barycentric Fréchet mean of Definition 4.1, etc. It will turn out that most interesting properties are obtained using the centroid means. We thus only consider these thereafter and set

$$[\boldsymbol{\mu}^{\mathbf{x}}]_+ = [E\mathbf{x}]_+, \quad [\boldsymbol{\mu}^{\mathbf{y}}]_+ = [E\mathbf{y}]_+.$$

From the discussion of Section 2, one can regard these four projective points as affine points, expressed in barycentric coordinates w.r.t. the frame $\mathcal{F} = \{A_0, \ldots, A_d\}$. Thus, formula (29) applied to $M \equiv [\boldsymbol{\mu}^{\mathbf{x}}]_+$ and $N \equiv [\mathbf{x}]_+$, resp. $M \equiv [\boldsymbol{\mu}^{\mathbf{y}}]_+$ and $N \equiv [\mathbf{y}]_+$, allows to compute the displacement vectors to the CoDa points from the corresponding means as,

$$\mathbf{v}_{\mathbf{x}} := \overrightarrow{[\boldsymbol{\mu}^{\mathbf{x}}]_+ [\mathbf{x}]_+} = \frac{\sum_{i<j} \det \begin{vmatrix} \mu^{\mathbf{x}}_i & x_i \\ \mu^{\mathbf{x}}_j & x_j \end{vmatrix} \overrightarrow{A_iA_j}}{||\boldsymbol{\mu}^{\mathbf{x}}||_1 ||\mathbf{x}||_1},$$

$$\mathbf{v}_{\mathbf{y}} := \overrightarrow{[\boldsymbol{\mu}^{\mathbf{y}}]_+ [\mathbf{y}]_+} = \frac{\sum_{i<j} \det \begin{vmatrix} \mu^{\mathbf{y}}_i & y_i \\ \mu^{\mathbf{y}}_j & y_j \end{vmatrix} \overrightarrow{A_iA_j}}{||\boldsymbol{\mu}^{\mathbf{y}}||_1 ||\mathbf{y}||_1}.$$

A term-by-term product of the $\overrightarrow{A_iA_j}$ component of $\mathbf{v}_{\mathbf{x}}$ and $\mathbf{v}_{\mathbf{y}}$, viz.

$$\frac{\det \begin{vmatrix} \mu^{\mathbf{x}}_i & x_i \\ \mu^{\mathbf{x}}_j & x_j \end{vmatrix} \times \det \begin{vmatrix} \mu^{\mathbf{y}}_i & y_i \\ \mu^{\mathbf{y}}_j & y_j \end{vmatrix}}{||\boldsymbol{\mu}^{\mathbf{x}}||_1 ||\mathbf{x}||_1 ||\boldsymbol{\mu}^{\mathbf{y}}||_1 ||\mathbf{y}||_1}$$

gives a measure of the (random) covariation of the displacement vectors $\mathbf{v}_\mathbf{x}$ and $\mathbf{v}_\mathbf{y}$ in the same direction $\overrightarrow{A_iA_j}$. Taking expectation, this gives an analogue of the covariance between two Coda points $[\mathbf{x}]_+$ and $[\mathbf{y}]_+$ as the average of the displacement vectors of the Coda point $[\mathbf{x}]_+$ and $[\mathbf{y}]_+$ from their respective mean, in the common direction from part $i$ to $j$. This leads to the following definition of the (cross-)covariance matrix of the random CoDa points $[\mathbf{x}]_+$ and $[\mathbf{y}]_+$:

**Definition 6.1** (Covariance matrix for a pair of CoDa)**.** Let $[\mathbf{x}]_+, [\mathbf{y}]_+ \in \mathbb{P}_+^d$ be random CoDa points, with corresponding mean point $[\boldsymbol{\mu}^\mathbf{x}]_+, [\boldsymbol{\mu}^\mathbf{y}]_+ \in \mathbb{P}_+^d$. The barycentric covariance matrix of $[\mathbf{x}]_+$ and $[\mathbf{y}]_+$, w.r.t. $[\boldsymbol{\mu}^\mathbf{x}]_+, [\boldsymbol{\mu}^\mathbf{y}]_+$, is defined as the following symmetric matrix (with null diagonal) of size $d+1$

$$\mathrm{Cov}([\mathbf{x}]_+, [\mathbf{y}]_+) := \left(\mathrm{Cov}([\mathbf{x}]_+, [\mathbf{y}]_+)_{i,j}\right) \in \mathbb{R}^{(d+1)\times(d+1)},$$

where the $(i,j)$ component is set as

$$\mathrm{Cov}([\mathbf{x}]_+, [\mathbf{y}]_+)_{i,j} := E\left(\frac{\det\begin{vmatrix}\mu_i^\mathbf{x} & x_i\\ \mu_j^\mathbf{x} & x_j\end{vmatrix} \times \det\begin{vmatrix}\mu_i^\mathbf{y} & y_i\\ \mu_j^\mathbf{y} & y_j\end{vmatrix}}{||\boldsymbol{\mu}^\mathbf{x}||_1||\mathbf{x}||_1||\boldsymbol{\mu}^\mathbf{y}||_1||\mathbf{y}||_1}\right). \tag{22}$$

If all four representatives are normalized to sit on the simplex, viz. $\mathbf{x}, \mathbf{y}, \boldsymbol{\mu}^\mathbf{x}, \boldsymbol{\mu}^\mathbf{y} \in \Delta_+^d$, the previous expression (22) reduces to its numerator.

Taking $[\mathbf{y}]_+ = [\mathbf{x}]_+$ and $[\boldsymbol{\mu}^\mathbf{x}]_+ = [\boldsymbol{\mu}^\mathbf{y}]_+$ in the previous definition leads to the definition of the analogue of a variance matrix[8] for a random CoDa $[\mathbf{x}]_+$:

**Definition 6.2** (Variance matrix for CoDa)**.** The barycentric variance matrix of $[\mathbf{x}]_+$ w.r.t. the deterministic mean point $[\boldsymbol{\mu}^\mathbf{x}]_+$ is defined as the following symmetric matrix (with null diagonal)

$$\begin{aligned}\mathrm{Var}([\mathbf{x}]_+) &:= \mathrm{Cov}([\mathbf{x}]_+, [\mathbf{x}]_+) \in \mathbb{R}^{(d+1)\times(d+1)}\\ &= \left(E\left(\frac{\det^2\begin{vmatrix}\mu_i^\mathbf{x} & x_i\\ \mu_j^\mathbf{x} & x_j\end{vmatrix}}{||\boldsymbol{\mu}^\mathbf{x}||_1^2||\mathbf{x}||_1^2}\right)\right)_{\substack{i=0,\cdots,d\\ j=0,\cdots,d}}.\end{aligned} \tag{23}$$

In case both representatives of $[\mathbf{x}]_+$, $[\boldsymbol{\mu}^\mathbf{x}]_+$ are chosen on the simplex, viz. $\mathbf{x}, \boldsymbol{\mu}^\mathbf{x} \in \Delta_+^d$, the $(i,j)$ component of $\mathrm{Var}([\mathbf{x}]_+)$, simplifies as

$$\mathrm{Var}([\mathbf{x}]_+)_{i,j} = E\left(\det^2\begin{vmatrix}\mu_i^\mathbf{x} & x_i\\ \mu_j^\mathbf{x} & x_j\end{vmatrix}\right).$$

Since the squared 2−barycentric divergence $d_2^2([\mathbf{x}]_+, [\boldsymbol{\mu}^\mathbf{x}]_+)$ is a separable function of its $i<j$ components, the expected divergence between $[\mathbf{x}]_+$ and its mean decomposes as a sum of the expected pairwise barycentric divergence, along the $i<j$ components,

$$Ed_2^2([\mathbf{x}]_+, [\boldsymbol{\mu}^\mathbf{x}]_+) = \sum_{i<j} E\left(\frac{\det^2\begin{vmatrix}\mu_i^\mathbf{x} & x_i\\ \mu_j^\mathbf{x} & x_j\end{vmatrix}}{||\boldsymbol{\mu}^\mathbf{x}||_1^2||\mathbf{x}||_1^2}\right) = \sum_{i<j}\mathrm{Var}([\mathbf{x}]_+)_{i,j}$$

[8] or auto-covariance matrix

Therefore, it is natural to define the total variance, which quantifies the total variability in a compositional data set, as the sum of the variance components of the variance matrix.

**Definition 6.3** (Total Variance for CoDa)**.** The total variance of $[\mathbf{x}]_+$ w.r.t. the deterministic mean point $[\boldsymbol{\mu}^{\mathbf{x}}]_+$ is the scalar

$$\mathrm{TVar}([\mathbf{x}]_+) = \sum_{i<j} \mathrm{Var}([\mathbf{x}]_+)_{i,j}.$$

By dividing the Variance matrix components $\mathrm{Var}([\mathbf{x}]_+)_{i,j}$ by the Total Variance $\mathrm{TVar}([\mathbf{x}]_+)$, one obtains a normalised Variance matrix, (called the contained variance in Greenacre 2021), which allows to quantify the importance of each variance component to the total.

Eventually, a measure of correlation is obtained by combining Definitions 6.1 and 6.2:

**Definition 6.4** (Correlation matrix for CoDa)**.** The barycentric correlation matrix of $[\mathbf{x}]_+$ and $[\mathbf{y}]_+$, w.r.t. $[\boldsymbol{\mu}^{\mathbf{x}}]_+, [\boldsymbol{\mu}^{\mathbf{y}}]_+$, is defined as

$$\rho([\mathbf{x}]_+, [\mathbf{y}]_+) := \left(\rho([\mathbf{x}]_+, [\mathbf{y}]_+)_{i,j}\right) \in \mathbb{R}^{(d+1)\times(d+1)},$$

with, for $i \neq j$,

$$\begin{aligned}\rho([\mathbf{x}]_+, [\mathbf{y}]_+)_{i,j} &:= \frac{\mathrm{Cov}([\mathbf{x}]_+, [\mathbf{y}]_+)_{i,j}}{\sqrt{\mathrm{Var}([\mathbf{x}]_+)_{i,j}\mathrm{Var}([\mathbf{y}]_+)_{i,j}}}, \\ &= \frac{E\left(\frac{\det\begin{vmatrix}\mu_i^{\mathbf{x}} & x_i\\ \mu_j^{\mathbf{x}} & x_j\end{vmatrix}\times\det\begin{vmatrix}\mu_i^{\mathbf{y}} & y_i\\ \mu_j^{\mathbf{y}} & y_j\end{vmatrix}}{||\boldsymbol{\mu}^{\mathbf{x}}||_1||\mathbf{x}||_1||\boldsymbol{\mu}^{\mathbf{y}}||_1||\mathbf{y}||_1}\right)}{\sqrt{E\left(\frac{\det^2\begin{vmatrix}\mu_i^{\mathbf{x}} & x_i\\ \mu_j^{\mathbf{x}} & x_j\end{vmatrix}}{||\boldsymbol{\mu}^{\mathbf{x}}||_1^2||\mathbf{x}||_1^2}\right)}\sqrt{E\left(\frac{\det^2\begin{vmatrix}\mu_i^{\mathbf{y}} & y_i\\ \mu_j^{\mathbf{y}} & y_j\end{vmatrix}}{||\boldsymbol{\mu}^{\mathbf{x}}||_1^2||\mathbf{x}||_1^2}\right)}}\end{aligned}$$

and $\rho_{i,i} = 0$.

In definition 6.4, the normalisation by the $||.||_1$ norm enters both in the expectation of the numerator and in the expectation of denominator. This suggests the following modified version of correlation, directly as a ratio of determinants:

**Definition 6.5** (Modified correlation matrix for CoDa)**.** The modified barycentric correlation matrix of $[\mathbf{x}]_+$ and $[\mathbf{y}]_+$, w.r.t. $[\boldsymbol{\mu}^{\mathbf{x}}]_+, [\boldsymbol{\mu}^{\mathbf{y}}]_+$, is defined as

$$r([\mathbf{x}]_+, [\mathbf{y}]_+)_{i,j} := \frac{E\left(\det\begin{vmatrix}\mu_i^{\mathbf{x}} & x_i\\ \mu_j^{\mathbf{x}} & x_j\end{vmatrix}\times\det\begin{vmatrix}\mu_i^{\mathbf{y}} & y_i\\ \mu_j^{\mathbf{y}} & y_j\end{vmatrix}\right)}{\sqrt{E\left(\det^2\begin{vmatrix}\mu_i^{\mathbf{x}} & x_i\\ \mu_j^{\mathbf{x}} & x_j\end{vmatrix}\right)}\sqrt{E\left(\det^2\begin{vmatrix}\mu_i^{\mathbf{y}} & y_i\\ \mu_j^{\mathbf{y}} & y_j\end{vmatrix}\right)}}, \quad i \neq j,$$

and $r([\mathbf{x}]_+, [\mathbf{y}]_+)_{i,i} = 0$.

**Remark 5.** i) By definition of $\mathbb{P}_+^d$, if $[\mathbf{x}]_+ \in \mathbb{P}_+^d$, then $\mathbf{x} \neq \mathbf{0}$ and $||\mathbf{x}||_1 \neq 0$. Hence, the ratios in Definitions 6.1 and 6.2 are well-defined. By linearity of $||.||_1$ and multilinearity of the determinant, (22) is invariant by positive

rescaling $\mathbf{x} \leftarrow \alpha\mathbf{x}$, $\mathbf{y} \leftarrow \beta\mathbf{y}$, $\boldsymbol{\mu}^{\mathbf{x}} \leftarrow \gamma\boldsymbol{\mu}^{\mathbf{x}}$, $\boldsymbol{\mu}^{\mathbf{y}} \leftarrow \delta\boldsymbol{\mu}^{\mathbf{y}}$, with $\alpha, \beta, \gamma, \delta > 0$[9]. Hence, Definitions 6.1 and 6.2 are well-defined on $\mathbb{P}_+^d$.

Definition 6.4, is undefined when $\mathrm{Var}([\mathbf{x}]_+)_{i,j} = 0$ or $\mathrm{Var}([\mathbf{y}]_+)_{i,j} = 0$ (and the discussion is similar for Definition 6.5). The situation here is analogous to Euclidean vectors, where Pearson's classical correlation coefficient is undefined for a degenerate (Dirac) random variable. We thus set the corresponding coefficient equal to 0 in such a case.

ii) When all four points $[\mathbf{x}]_+$, $[\mathbf{y}]_+$, $[\boldsymbol{\mu}^{\mathbf{x}}]_+$, $[\boldsymbol{\mu}^{\mathbf{y}}]_+$, are simplex normalized, viz. $\mathbf{x}, \mathbf{y}, \boldsymbol{\mu}^{\mathbf{x}}, \boldsymbol{\mu}^{\mathbf{y}} \in \Delta_+^d$, Definition 6.5 coincides with Definition 6.4. The difference lies in the fact that Definition 6.5 is scale-invariant only for non-random rescaling, whereas Definition 6.4 is scale-invariant for deterministic and random rescaling alike.

6.2. **Discussion and properties of barycentric variation matrices.** The component $i, j$ of the Variance matrix (23) measures the proportionality of parts $i, j$, as shown in the next Proposition.

**Proposition 6.6.** *Assume w.l.o.g. that $\mathbf{x} \in \Delta_+^d$ is the simplex normalized representative. Var$([\mathbf{x}]_+)_{i,j} = 0$ if and only if $x_i$ and $x_j$ are proportional or one of them is zero.*

*Proof.* $\Leftarrow$: If $\mathbf{x} \in \Delta_+^d$, $\sum_{i=0}^d x_i = 1$ implies $1 = \sum_{i=0}^d Ex_i = \sum_{i=0}^d \mu_i^{\mathbf{x}}$, that is to say, $\boldsymbol{\mu}^{\mathbf{x}} \in \Delta_+^d$ and $\boldsymbol{\mu}_i^{\mathbf{x}} = Ex_i$. Therefore, if, say, $x_j = \lambda x_i$ a.s., for some $\lambda > 0$, then, $\boldsymbol{\mu}_j^{\mathbf{x}} = \lambda\boldsymbol{\mu}_i^{\mathbf{x}}$ and $\det \begin{vmatrix} \mu_i^{\mathbf{x}} & x_i \\ \mu_j^{\mathbf{x}} & x_j \end{vmatrix} = \det \begin{vmatrix} \mu_i^{\mathbf{x}} & x_i \\ \lambda\mu_i^{\mathbf{x}} & \lambda x_i \end{vmatrix} = 0$ a.s. (two proportional rows). Thus, $\mathrm{Var}([\mathbf{x}]_+)_{i,j} = 0$.

On the other hand, if, say, $x_i = 0$ a.s. then $\mu_i^{\mathbf{x}} = 0$ and $\mathrm{Var}([\mathbf{x}]_+)_{i,j} = 0$ also.

$\Rightarrow$: Notice that one always has that

$$E\det \begin{vmatrix} \mu_i^{\mathbf{x}} & x_i \\ \mu_j^{\mathbf{x}} & x_j \end{vmatrix} = E(x_j\mu_i^{\mathbf{x}} - x_i\mu_j^{\mathbf{x}}) = \mu_j^{\mathbf{x}}\mu_i^{\mathbf{x}} - \mu_i^{\mathbf{x}}\mu_j^{\mathbf{x}} = 0.$$

If $\mathrm{Var}([\mathbf{x}]_+)_{i,j} = 0$, Tchebychev's inequality entails that, for all $t > 0$,

$$P\left(\left|\det \begin{vmatrix} \mu_i^{\mathbf{x}} & x_i \\ \mu_j^{\mathbf{x}} & x_j \end{vmatrix}\right| > t\right) = 0.$$

Thus, $x_i\mu_j^{\mathbf{x}} - x_j\mu_i^{\mathbf{x}} = 0$ a.s. Since $x_i \geq 0$, $\mu_i^{\mathbf{x}} = 0$ iff $x_i = 0$ a.s. Therefore, $x_i\mu_j^{\mathbf{x}} - x_j\mu_i^{\mathbf{x}} = 0$ a.s. entails $x_i$ and $x_j$ are proportional or one of them is zero.

□

This result is intuitively clear since $\mathrm{Var}([\mathbf{x}]_+)_{i,j}$ measures the quadratic displacement variation of $[\mathbf{x}]_+$ around its mean $[\boldsymbol{\mu}^{\mathbf{x}}]_+$, along the $i, j$ parts: if both components $x_i$ and $x_j$ are proportional, there is no variation in the $\overrightarrow{A_iA_j}$ direction (and similarly if one of them is always zero). One thus obtains for the variance matrix an object similar to the log variation matrix of J. Aitchison 1986, (or of its variants to be found in Lovell et al. 2015, Erb and Notredame 2016, Filzmoser and Hron 2009, Juan José Egozcue and Vera Pawlowsky-Glahn 2023), in its ability to measure the proportionality of parts.

[9]Note that $\alpha, \beta$ may be random.

**Remark 6.** i) Although named barycentric variance matrix, (23) is not a classical variance-covariance matrix (in particular, it has null diagonal), which can be somehow confusing for newcomers to CoDa. This feature reflects the affine nature of CoDa: from the affine geometry perspective where CoDa are points (and not vectors), it takes two points to make a direction. Hence, the variation can only be measured along the vectors $\overrightarrow{A_iA_j} = A_j - A_i$ corresponding to pairs of parts. The latter is the null vector if $i = j$. This explains why diagonal elements of the barycentric variance matrix is zero. This null-diagonal feature is common to all measures of proportionality afore-mentioned (and in particular for J. Aitchison 1986's log variation matrix): it reflects that only the *relative* variation of one part $i$ w.r.t. another $j$ is sensible in CoDa analysis, see e.g. J. Aitchison 1986. We nevertheless chose this term as the barycentric variance matrix derives from the barycentric covariance matrix.

ii) One could also define the partial barycentric variance contribution of part $i$ ($\mathrm{PVar}_i$) as

$$\mathrm{PVar}_i([\mathbf{x}]_+) := \sum_{j=0}^{d} \mathrm{Var}([\mathbf{x}]_+)_{i,j}$$

which measures the total contribution of part $i$ in the total variance, as one has

$$\mathrm{TVar}([\mathbf{x}]_+) = \frac{1}{2}\sum_{i=0}^{d} \mathrm{PVar}_i([\mathbf{x}]_+).$$

This allows to have the above decomposition of the total barycentric variance in terms of the $d+1$ components themselves[10], instead of the $d(d+1)/2$ terms of Definition 6.3.

It is worth stressing some of advantageous features of the barycentric variances. First, being log-free, the proposed matrix (23) is defined on the whole CoDa space $\mathbb{P}_+^d$, and is now able to process CoDa with zeroes. This is in contrast to all of the above-mentioned measures, which fail to be defined whenever some zeroes are existent in a component. Second, even if the data has no zeroes, log transformations will turn parts with small values into large values, resulting in large variations in the log-ratio variance. The relative error in the small components are likely to be high and to distort any multivariate analysis based on such log-ratio variances. This issue occurs in particular with imputation methods for CoDa with zeroes, see e.g. Greenacre 2021 Section 7, and our discussion in Section 3.2. To the contrary, the proposed variation matrix (23) does not alter the scale of the parts by a nonlinear transformation. At last, it has been argued that when two parts are not exactly proportional, the log-ratio variance of J. Aitchison 1986 has no intrinsic scale and so is hard to interpret. This is especially relevant for determining a cut-off for selecting variables, and motivated Lovell et al. 2015 to propose their scale-free variant $\phi$ which puts the log-ratio variance in relation to the size of the single variances involved. Here, since the Total Variance of Definition 6.3 interprets as the average displacement of $[\mathbf{x}]_+$ w.r.t. to its centroid mean, it makes sense to

[10]This was a question asked by a referee, to obtain an expression of the total variance as a sum over the components, so that it mimics that of the clr variance matrix.

scale the components $\mathrm{Var}([\mathbf{x}]_+)_{i,j}$ of the Variance matrix (23) by the (scalar) total variance, i.e. to set the Normalised Variance matrix

$$\mathrm{NVar}([\mathbf{x}]_+) := \frac{\mathrm{Var}([\mathbf{x}]_+)}{\mathrm{TVar}([\mathbf{x}]_+)}, \tag{24}$$

as a scale-free version of the Variance Matrix, with components between 0 and 1.

Let us illustrate these points with basic examples.

**Example 6.1** (Variance matrix with two identical components)**.** We take the same distribution as in Example B.4, Figure 11, but with $n = 10000$ sample points. The empirical variance matrix is

$$\mathrm{Var}([\mathbf{x}]_+) = \begin{pmatrix} 0. & 0.0132371 & 0.0132371 \\ 0.0132371 & 0. & 0. \\ 0.0132371 & 0. & 0. \end{pmatrix}.$$

This empirically confirms the result of Proposition 6.6: since the data (blue points) sit on the straight line $x_1 = x_2$ in the triangle, there is no variation in the direction $A_1A_2$. The empirical variance $\mathrm{var}([\mathbf{x}]_+)_{1,2}$ computed on the data is exactly zero as $x_1$ and $x_2$ carry the same proportional information.

**Example 6.2** (Variance matrix with uniform Dirichlet distribution)**.** Let $\mathbf{x} = (x_0, x_1, x_2)$ be distributed according to the Dirichlet$(1, 1, 1)$ distribution, which corresponds to the uniform distribution on the simplex $\Delta^2_+$. The empirical arithmetic mean computed on 10000 i.i.d. replications is given by

$$\boldsymbol{\mu}^{\mathbf{x}} = (0.331853, 0.335558, 0.33259).$$

The empirical barycentric variance matrix and normalized version are

$$\mathrm{Var}([\mathbf{x}]_+) = \begin{pmatrix} 0. & 0.0186154 & 0.0184052 \\ 0.0186154 & 0. & 0.0187276 \\ 0.0184052 & 0.0187276 & 0. \end{pmatrix},$$

$$\mathrm{NVar}([\mathbf{x}]_+) = \begin{pmatrix} 0. & 0.333919 & 0.330149 \\ 0.333919 & 0. & 0.335932 \\ 0.330149 & 0.335932 & 0. \end{pmatrix}.$$

One obtains approximately the same values in all directions for the empirical mean and variance matrix, as expected with such an isotropic distribution.

If the composition above was in fact a four-parts composition with, say, null second component, i.e. if one adds a column of zeros at the second component, so that $x_1 = 0$, $(x_0, x_2, x_3) \sim$ Dirichlet$(1, 1, 1)$, then the new normalized barycentric variance matrix is

$$\mathrm{NVar}([\mathbf{x}]_+) = \begin{pmatrix} 0. & 0. & 0.333919 & 0.330149 \\ 0. & 0. & 0. & 0. \\ 0.333919 & 0. & 0. & 0.335932 \\ 0.330149 & 0. & 0.335932 & 0. \end{pmatrix}.$$

In other words the new normalized barycentric variance matrix is unchanged, except for an additional row and column of zeroes at position 2 corresponding to the null component $x_1 = 0$: there is zero variation in the directions $A_1A_j$, $j = 0, 2, 3$, in agreement with Proposition 6.6.

6.3. **Discussion and properties of barycentric correlation matrices.** The barycentric covariance and correlation matrices of Definitions 6.1, 6.4 and 6.5 allow to measure the joint variation of a pair of random CoDa elements w.r.t. to their respective arithmetic mean, in a direction $A_iA_j$ corresponding to the parts $i, j$, as shown in the next Proposition:

**Theorem 6.7** (Properties of Covariance, Correlation)**.** *i) Boundedness:*

$$\begin{aligned} &Cov^2([\mathbf{x}]_+, [\mathbf{y}]_+)_{i,j} \leq Var([\mathbf{x}]_+)_{i,j}\, Var([\mathbf{y}]_+)_{i,j}, \\ &-1 \leq \rho([\mathbf{x}]_+, [\mathbf{y}]_+)_{i,j} \leq 1, \\ &-1 \leq r([\mathbf{x}]_+, [\mathbf{y}]_+)_{i,j} \leq 1. \end{aligned}$$

*ii) Zero covariance when pair of simplex representatives are independent: Assume $\mathbf{x}, \mathbf{y} \in \Delta_+^d$ are simplex representatives of $[\mathbf{x}]_+$, $[\mathbf{y}]_+$. If the pair $(x_i, x_j)$ is independent of the pair $(y_i, y_j)$, then*

$$Cov([\mathbf{x}]_+, [\mathbf{y}]_+)_{i,j} = \rho([\mathbf{x}]_+, [\mathbf{y}]_+)_{i,j} = r([\mathbf{x}]_+, [\mathbf{y}]_+)_{i,j} = 0.$$

*iii) Zero modified covariance when independence of pairs of raw amounts: Assume the compositional data is obtained by closure of the raw amounts, viz. $\mathbf{x} = \mathcal{C}(\mathbf{a})$, $\mathbf{y} = \mathcal{C}(\mathbf{b})$. If the pair $(a_i, a_j)$ is independent of the pair $(b_i, b_j)$, and the means representatives are chosen to be the expectations of the raw amounts, i.e.*

$$\boldsymbol{\mu}^{\mathbf{x}} = E\mathbf{a}, \quad \boldsymbol{\mu}^{\mathbf{y}} = E\mathbf{b},$$

*then*

$$r([\mathbf{x}]_+, [\mathbf{y}]_+)_{i,j} = 0.$$

*Note that $(a_i, a_j) \perp (b_i, b_j)$ does not imply $(x_i, x_j)$ independent of $(y_i, y_j)$.*

*Proof.* Let us denote $\mathcal{X}_{ij} := \begin{vmatrix} \mu_i^{\mathbf{x}} & x_i \\ \mu_j^{\mathbf{x}} & x_j \end{vmatrix}$, $\mathcal{Y}_{ij} := \begin{vmatrix} \mu_i^{\mathbf{y}} & y_i \\ \mu_j^{\mathbf{y}} & y_j \end{vmatrix}$.

i) By Cauchy-Schwarz,

$$\left(E\left(\frac{\det\mathcal{X}_{ij}}{||\boldsymbol{\mu}^{\mathbf{x}}||_1||\mathbf{x}||_1} \times \frac{\det\mathcal{Y}_{ij}}{||\boldsymbol{\mu}^{\mathbf{y}}||_1||\mathbf{y}||_1}\right)\right)^2 \leq E\left(\frac{\det^2\mathcal{X}_{ij}}{||\boldsymbol{\mu}^{\mathbf{x}}||_1^2||\mathbf{x}||_1^2}\right) E\left(\frac{\det^2\mathcal{Y}_{ij}}{||\boldsymbol{\mu}^{\mathbf{y}}||_1^2||\mathbf{y}||_1^2}\right),$$

which is

$$\mathrm{Cov}^2([\mathbf{x}]_+, [\mathbf{y}]_+)_{i,j} \leq \mathrm{Var}([\mathbf{x}]_+)_{i,j}\mathrm{Var}([\mathbf{y}]_+)_{i,j},$$

and yields the result for $\rho$. The proof for $r$ is similar.

ii) By independence, $E(\det\mathcal{X}_{ij} \times \det\mathcal{Y}_{ij}) = E(\det\mathcal{X}_{ij}) \times E(\det\mathcal{Y}_{ij})$. Since $[\boldsymbol{\mu}^{\mathbf{x}}]_+$ and $[\boldsymbol{\mu}^{\mathbf{y}}]_+$ are the centroid means, $Ex_i = \lambda\mu_i^{\mathbf{X}}$, $Ex_j = \lambda\mu_j^{\mathbf{X}}$, for some $\lambda > 0$. Therefore, $E(\det\mathcal{X}_{ij}) = \lambda(\mu_j^{\mathbf{X}}\mu_i^{\mathbf{X}} - \mu_i^{\mathbf{X}}\mu_j^{\mathbf{X}}) = 0$, and similarly, $E(\det\mathcal{Y}_{ij}) = 0$. Thus, $\mathrm{Cov}([\mathbf{x}]_+, [\mathbf{y}]_+)_{i,j} = 0$.

iii) By choosing as representatives of

$$[\mathbf{x}] = [\mathbf{a}]_+, \quad [\mathbf{y}]_+ = [\mathbf{b}]_+,$$

directly the raw amounts $\mathbf{a}, \mathbf{b}$, independence of $(a_i, a_j)$ with $(b_i, b_j)$ imply independence of

$$\det\mathcal{X}_{ij} = \det\begin{vmatrix} a_i & Ea_i \\ a_j & Ea_j \end{vmatrix}, \quad \text{with} \quad \det\mathcal{Y}_{ij} = \det\begin{vmatrix} b_i & Eb_i \\ b_j & Eb_j \end{vmatrix}.$$

The rest of the proof is the same as ii).

□

Property iii) means that if the raw/absolute amounts are available, the modified correlation matrix allows to detect independence of pairs of components.

## 7. Applications to real data

7.1. **Foraminiferal data: amalgamation of CoDa with zeroes using the barycentric variance matrix.** The **foraminiferal** dataset is a classical palaeoecological compositional dataset studied by J. Aitchison 1986 (Dataset 34 in Appendix D), and is available in the R package **CoDA.base** Comas-Cufí 2023. This dataset contains the proportions of four different fossil species (*Neogloboquadrina atlantica*, *Neogloboquadrina pachyderma*, *Globorotalia obesa* and *Globigerinoides triloba*) measured across 30 depth levels within core samples. Figure 3 show the ternary plots for all combinations of three species out of four.

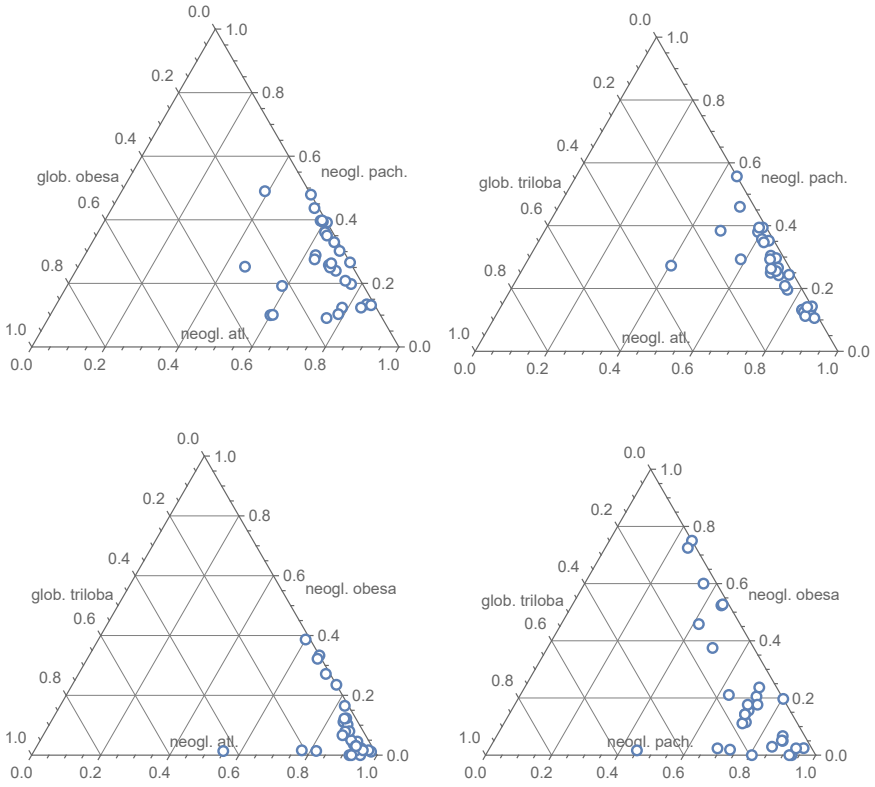


Figure 3. Ternary plots of the foraminiferal dataset.

The dataset, positioned near the sample space boundary, contains (a small number of) zeros—specifically, three in part 3 (Globorotalia obesa) and two in part 4 (Globigerinoides triloba), with no simultaneous occurrences across these species. This pattern enables zero-value mitigation by amalgamating components 3 and 4.

| | neogl. atl. | neogl. pach. | glob. obesa | glob. triloba |
|---|---|---|---|---|
| neogl_atl | 0.0000000 | 0.0084447 | 0.0033752 | 0.0015425 |
| neogl_pach | 0.0084447 | 0.0000000 | 0.0007140 | 0.0002204 |
| glob_obesa | 0.0033752 | 0.0007140 | 0.0000000 | 0.0000444 |
| glob_triloba | 0.0015425 | 0.0002204 | 0.0000444 | 0.0000000 |

TABLE 1. Barycentric variance matrix for foraminiferal data: species 3 (*Globorotalia obesa*) and 4 (*Globigerinoides triloba*) have the smallest barycentric variance component value 0.0000444 and thus are the most similar. Total barycentric variance 0.01434119.

More broadly, J. Aitchison 1986 p. 267 advocated amalgamation of similar components as a general strategy, posing: "Is there a feasible amalgamation which still allows investigation of the main issues and at the same time removes all zero components?". He noted that "If, however, we know that species 3 and 4 are similar, at any rate more related than any other pair, then perhaps little will be lost if we decide to amalgamate species 3 and 4."

This raises a critical question: how can we quantify the similarity between components in a compositional dataset, so that we can identify such most similar pair of components? When zeros are present in the dataset, traditional log-ratio transformations and the corresponding log-variation matrices (4) become undefined for the components of interest. The solution found by Aitchison was to use the *external covariate* depth and to check, *a posteriori*, via a significance test within a linear *parametric* compositional regression model with amalgamated species 3 and 4 as reference, whether there is any dependence of composition on depth, see J. Aitchison 1986 p. 267.

To the contrary, the barycentric variance matrix (23) of Definition 6.2 enables such a quantification of similarity between components (with or without zeroes), and allows for a simple, *direct*, *model-free*, *internal* identification of the most similar pair. The results for the foraminiferal dataset are given in Tables 1 and 2, which present the Barycentric variation matrix and its normalized version (24), i.e. divided by the (scalar) total variance Tvar = 0.01434119 (see Definition 6.3). The barycentric variance component between species 3 and 4 indeed has the lowest value 0.0000444 among all pairs, which is one order of magnitude below its nearest contender 0.0002204 (between species 2 (*Neogloboquadrina pachyderma*) and 4 (*Globigerinoides triloba*)). In more readable relative terms (Table 2), the barycentric variance component between species 3 and 4 accounts for roughly 0.31 % of the total, compared to 1.53 % between species 2 and 4 (the second lowest variance component), and 58.88 % between species 1 and 2 (the highest variance component).

This example thus shows that the barycentric variance matrix thus provides a framework for quantifying the extent to which pairs of components exhibit similar compositional variation. By leveraging this matrix, one can systematically evaluate which pairs of parts are more similar and, consequently, make informed decisions about amalgamations. It is interesting to note that our proposed approach is used in this example from the perspective of classical log-ratio CoDa analysis, (with the goal of removing zeroes by amalgamation), and thus can be used in complement,

| | neogl. atl. | neogl. pach. | glob. obesa | glob. triloba |
|---|---|---|---|---|
| neogl_atl | 0.0000000 | 0.5888452 | 0.2353532 | 0.1075558 |
| neogl_pach | 0.5888452 | 0.0000000 | 0.0497844 | 0.0153683 |
| glob_obesa | 0.2353532 | 0.0497844 | 0.0000000 | 0.0030932 |
| glob_triloba | 0.1075558 | 0.0153683 | 0.0030932 | 0.0000000 |

TABLE 2. Normalised Barycentric variance matrix for foraminiferal data (in percentages of the total barycentric variance): species 3 (*Globorotalia obesa*) and 4 (*Globigerinoides triloba*) only account for 0.309 % of the total barycentric variance.

and not in opposition, to more familiar methods based on log-ratio vector space representations of non-zero CoDa.

7.2. **Clustering of compositional data and variables using barycentric divergence and variance: FA composition of marine copepods.** Admittedly, it was mainly the low dimensionality and low number of zeros in the foraminiferal dataset that enabled J. Aitchison 1986 to manually determine which parts to amalgamate. For datasets having dozens or hundreds of parts and/or numerous zeros (e.g. microbiome data), such an approach by direct inspection is likely to become impractical.

In contrast, the identification and selection of relevant variables by computation of the barycentric variation matrix can be performed in these more difficult datasets. Indeed, one needs $O(d^2)$ product operations to compute the barycentric divergence, so computation of the barycentric variance matrix scales relatively well with the number of parts. This allows for an automated selection of the most important components using a hierarchical clustering procedure.

We illustrate this possibility on the data set **FA compositions of marine copepods** studied by Greenacre 2021. The data set records the composition of 40 fatty acids in a sample of 42 copepods in the Arctic, see Table 2 of Greenacre 2021, and is also available online. We performed a hierarchical clustering procedure with complete linkage on the computed barycentric variance matrix. The results are displayed on Figure 4, which shows the dendogram, ranking the parts by decreasing degree of similarity.

The dendogram shows a large number of similar FA components (the large vertical leaf in the central part, stretching from 16:3(n-4) to 18:2(n-6)), while dissimilar FA components (those who are merged later in the tree) are those listed on the top and bottom of the tree (from 16:00 to 18:4(n-3) and from 18-:1(n-9) to 16:1(n-7)). Interestingly, these most dissimilar components correspond pretty much to those in the 13-part subcomposition selected by Greenacre 2021, see the list displayed on his Table 2.

We emphasize once more that our proposed method enables the direct computation of a similarity measure between components, applicable to CoDa both with and without zeros. This is in contrast to Greenacre 2021, where he obtained his list of 13 FAs by excluding *a priori* all FAs that had an average occurrence below 0.01, in his initial analysis based on log-ratio distances and variances.

To analyze the complete 40-parts dataset, Greenacre 2021 compare in his Section 7 several zero replacement methods, and conclude that "since the substituted

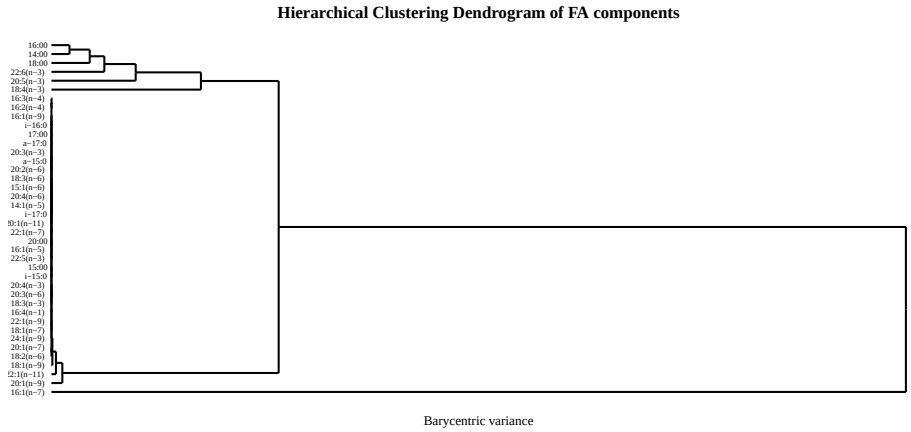


FIGURE 4. Dendogram of CoDa parts using hierarchical clustering with complete linkage based on the 2-barycentric variance matrix.

values will be small numbers, engendering large negative or positive logratios", "the method of zero substitution can have a strong effect on the structure of the compositional data set and influence its subsequent analysis". An additional warning against zero-replacement methods is his observation that "some of the methods broke down", which prompted the exclusion of the parts with too many zeroes (!).

Turning to the analysis of CoDa points, one can perform hierarchical clustering, this time of the data points, based on the full 40-part composition. This clustering is based on the matrix of 2-barycentric divergences between the 42 data points. The barycentric divergence is particularly useful as it measures the proximity of CoDa, even when zeros are present. The resulting dendogram for CoDa points is displayed on Figure 5.

7.3. **Fréchet local linear regression of CoDa using barycentric divergence.** Returning to the foraminiferal dataset, further statistical analysis can be obtained by performing, e.g., a nonparametric regression of the composition on the depth covariate. If one does not amalgamate species 3 and 4, the data points contain components with zeros, and thus the CoDa log-ratio regression techniques based on classical multivariate analysis can not be applied.

In contrast, an estimate of the conditional expectation of the full compositional response with respect to the depth covariate can be obtained by solving a weighted variant of the barycentric divergence-based Fréchet mean, as defined in Definition 4.1. The description is as follows: consider a random pair $(X, [\mathbf{Y}]_+) \in \mathbb{R} \times \mathbb{P}_+^3$ of the 4-parts foraminiferal CoDa response variable $[\mathbf{Y}]_+$), with depth covariate $X$. The goal of Fréchet regression is to estimate the conditional Fréchet mean

$$[\mathbf{m}]_+(x) := \arg \min_{[\mathbf{m}]_+ \in \mathbb{P}_+^3} E\left\{ d_2^2([\mathbf{Y}]_+, [\mathbf{m}]_+) | X = x \right\}$$

where $d_2$ is the $2-$barycentric divergence (10) of Definition 3.1 and Equation (16). The conditional Fréchet mean is the metric generalization of the classical regression function/conditional mean $m(x) = E[Y|X = x]$ to a non-Euclidean response variate $Y$ taking its values in a space equipped with divergence $d_2$. Given the sample

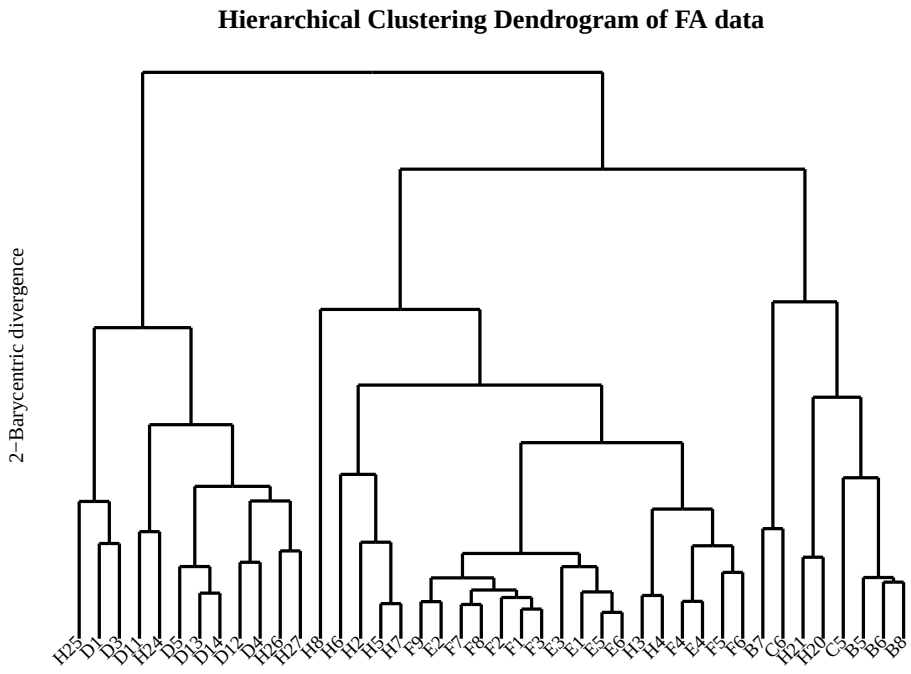


FIGURE 5. Dendogram of data points using hierarchical clustering with complete linkage based on the 2-barycentric divergence for the FA data.

$(x_i, [\mathbf{y}_i]_+)$, $i = 1, \ldots, 30$, an empirical estimate $[\hat{\mathbf{m}}]_+(x)$ of $[\mathbf{m}]_+(x)$ is obtained by solving

$$[\hat{\mathbf{m}}]_+(x) := \arg \min_{[\mathbf{m}]_+ \in \mathbb{P}_+^3} \frac{1}{n} \sum_{i=1}^{n} K_h\left(x_i - x\right) d_2^2([\mathbf{y}_i]_+, [\mathbf{m}]_+),$$

where $n = 30$ is the sample size, $K$ is a smoothing kernel such as the Epanechnikov kernel or Gaussian Kernel and $h$ is a bandwidth, with $K_h(.) = h^{-1}K(./h)$. $[\hat{\mathbf{m}}]_+$ is the metric generalization of the classical Nadaraya-Watson estimate to a Non-Euclidean response variable, see e.g. Hein 2009, and a weighted version of the empirical Fréchet mean defined via (18). See also M. Tsagris, Alenazi, and Stewart 2023 for kernel regression with the $\alpha-$transformation.

Going beyond Nadaraya-Watson smoothers, Petersen and Müller 2019 show how to generalize nonparametric local linear regression methods (Fan and Gijbels 1996, Loader 1999) for response data in a metric space with Euclidean predictors. This local Fréchet regression proves to be superior to Nadaraya-Watson smoothing, especially near the boundaries. It is obtained by solving

$$[\hat{\mathbf{m}}]_+(x) := \arg \min_{[\mathbf{m}]_+ \in \mathbb{P}_+^3} \frac{1}{n} \sum_{i=1}^{n} s(x_i, x, h) d_2^2([\mathbf{y}_i]_+, [\mathbf{m}]_+),$$

where $s(x_i, x, h)$ are local linear weight, replacing the Nadaraya-Watson weights $K_h\left(x_i - x\right)$.

We implemented such a local linear Fréchet estimate based on the 2−barycentric divergence for the foraminiferal dataset, with Epanechnikov kernel and a Nearest-Neighbor type bandwidth. The latter corresponds to setting $h$ as the $k-th$ smallest (usual) distance between the fitting point $x$ and the data points $x_i$, see Loader 1999. The results are displayed in Figure 6. The figure shows the influence of the depth

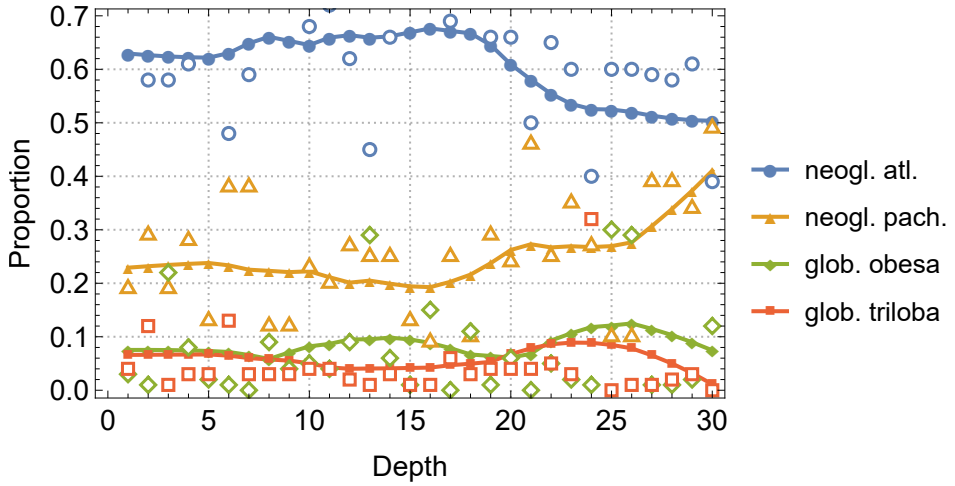


FIGURE 6. Local linear Fréchet regression of the foraminiferal CoDa on the depth covariate. Shallow markers (circle, triangle, diamond, square): data points for the four components. Solid lines with filled markers: Fréchet regression estimate. Epanechnikov kernel, 10-points Nearest Neighbor bandwidth.

covariate on the full-composition: the composition remains roughly stable from 0 to 15 meters, with species 1(*Neogloboquadrina atlantica*) as dominating species. From 15 to 30 meters, one observes a gradual shift, with the emergence of species 2(*Neogloboquadrina pachyderma*) and corresponding decrease of species 1. Species 3 remain at comparatively low levels, with moderate surge around 13 and 25 meters. Species 4(*Globigerinoides triloba*) remains low at all depth levels, and the curve suggest that observation 24 is an outlier.

7.4. **Further statistical Fréchet regression analysis.** To conclude this section, we briefly discuss how more advanced statistical analyses can be conducted within the Fréchet regression framework. We also provide references to relevant literature for further exploration.

Assessment of the model performance can be obtained using either an in-sample performance measure, by a Fréchet analogue of the coefficient of determination $R^2$, or an out-of-sample performance, by splitting the sample into training and testing sets, estimating the regression sample on the training set, and evaluating its predictive performance on the testing set, see e.g. Ghosal, Meiring, and Petersen 2023.

When the number of predictor variables increase, the local polynomial estimator is known to suffer from the curse of dimensionality. A way to alleviate this effect

is to use semi-parametric techniques like single-index models, where the regression function is assumed to depend only on a scalar projection of the multivariate predictor. Such single-index models have recently been extended to Fréchet regression by Ghosal, Meiring, and Petersen 2023 and Bhattacharjee and Müller 2023. An additional interest of the single-index model also owes to its interpretability as the weights of the single index quantify the contribution each covariate/predictor component on the output. In particular, Bhattacharjee and Müller 2023 provides asymptotic normality results on the estimator of the index, which allows to perform significance tests and selection of the variables.

Another prominent machine learning method for non-parametric regression is random forest (Breiman 2001, Genuer et al. 2020). It has excellent predictive performance, requires only a few parameters to tune, is able to learn complex relationships between input and output and can handle high-dimensional data efficiently. Capitaine et al. 2024, Qiu, Yu, and Zhu 2024 generalized random forests to the Fréchet case where inputs and outputs lie in metric spaces. By endowing the CoDa space with the present barycentric divergence or the log-free distance of Faugeras 2025a, one can apply this state-of-the-art machine learning method to CoDa with zeros. Additionally, Fréchet random forests outputs out-of-bag error, which provides a direct estimation of the prediction error. Similarly, the out-of-bag samples can be used to build variable importance scores which allow to effectively assess the importance of each variable in the learning task, and thus to perform variable selection.

Other methods to tame the curse of dimensionality are based on dimension reduction (see e.g. Ghojogh et al. 2023 for an overview). In particular, Zhang, Xue, and Li 2022 Ying and Yu 2020, Dong and Wu 2022 and Weng, Ke, and P. Wang 2024 extends sufficiency dimension reduction techniques like sliced inverse regression to Fréchet regression. At last, let us mention that deep Neural networks, combined with dimension reduction (Isomap), have also been extended to Fréchet regression in Iao, Zhou, and Müller 2024.

## 8. Conclusion

This study reinterprets compositional data (CoDa) through an affine geometric lens, leveraging barycentric coordinates to address limitations of classical log-ratio methods, particularly the handling of zero values. Building on the foundational affine geometry outlined in Appendix A–notably the displacement vector formula (Lemma A.4), which gives a decomposition in terms of displacements of the different pairs of basis frame parts–we established in Section 2 that CoDa within the simplex $\Delta_+^d$ are naturally represented as affine points, with operations like amalgamation, subcomposition, and partition corresponding to barycentric manipulations. This perspective, though seemingly overlooked in CoDa literature despite its affinity with ternary plots, offers a unified framework for analysis irrespective of zero components. Our primary contributions are twofold. First, Section 3 introduced $\alpha-$barycentric divergences, applicable across the entire CoDa space, with properties extensible to infinite-dimensional measures and a geometric interpretation for $\alpha = 2$. These divergences enabled to define some first statistical tools. In Section 4, we defined empirical Fréchet means and medians for CoDa sample clouds, minimizing the expected $\alpha$-barycentric divergence to a fixed point. The $\alpha = 2$ case,

reducing to the centroid for two-part compositions, offers a robust measure of central tendency, as demonstrated through simulations comparing variants (Appendix B). Section 5 extended this framework by defining isotropic Laplace-Gaussian distributions based on barycentric divergences, characterized by a location parameter and uniform dispersion. Weighted versions introduced anisotropic generalizations, incorporating directional variation parameters akin to multivariate Gaussian distributions in Euclidean spaces. These distributions provide a probabilistic model for CoDa with zeros, as illustrated in simulations.

Section 6 presented our second major contribution, the barycentric variance matrix, a log-free analogue to log-ratio variance matrices, quantifying component proportionality via averaged scalar products of displacement vectors in barycentric coordinates. Its practical utility was showcased in Section 7 through real datasets, enabling hierarchical clustering for component selection and local Fréchet regression for trend and outlier detection. The potential integration with advanced techniques—such as random forests or single-index models—highlights its adaptability to modern statistical challenges.

Reflecting on J.A. Dieudonné's quip, "Who ever uses barycentric coordinates?" (quoted by Pedoe 1970), this work asserts CoDa as a compelling application domain. By providing mathematical grounding and novel tools, we aim to catalyze further statistical exploration of CoDa from an affine viewpoint, free from positivity constraints. A companion paper (Faugeras 2025a) explores a related projective approach based on Grassmann's algebra and outer product, suggesting convergence of geometric perspectives that may enrich future research.

## Appendix A. A primer on affine geometry and barycentric coordinates

A.1. **Affine spaces.** Informally, an affine space is a vector space without a fixed choice of origin. It describes the geometry of points and free vectors in space, distinguishing between the two types of objects. As a consequence of the lack of origin, points in affine space cannot be (linearly) added together by plainly adding their coordinates as is the case for vectors, since the notion of linear combination of points is frame dependent, see, e.g., Gallier 2011 Chapter 2.1.[11] However, a vector $\mathbf{v}$ may be added to a point $P$ by placing the initial point of the vector at $P$ and then transporting $P$ to the terminal point. The operation thus described $P \to P + \mathbf{v}$ is the translation of $P$ along $\mathbf{v}$.

This suggests to develop affine geometry over linear algebra: an affine space is a set of points equipped with a set of transformations (that is bijective mappings), the translations, which form a vector space, and such that for any given ordered pair of points there is a unique translation sending the first point to the second; the composition of two translations is their sum in the vector space of the translations.

Weyl's axiomatization of an affine space formalizes these considerations:

**Definition A.1** (Affine space)**.** An affine space is a pair $(\mathcal{A}, \overrightarrow{\mathcal{A}})$ consisting of a nonempty set $\mathcal{A}$ (whose elements are called points) and a real vector space $\overrightarrow{\mathcal{A}}$ (the space of vectors) such that there is a mapping

$$\mathcal{A} \times \mathcal{A} \to \overrightarrow{\mathcal{A}}$$

[11] Yet, see the forthcoming Section A.2

denoted by

$$(P, Q) \in \mathcal{A} \times \mathcal{A} \mapsto \overrightarrow{PQ} \in \overrightarrow{\mathcal{A}}$$

satisfying the following axioms:

i) for any $P, Q, R \in \mathcal{A}$, we have $\overrightarrow{PR} = \overrightarrow{PQ} + \overrightarrow{QR}$;
ii) for any $P \in \mathcal{A}$ and for any $\mathbf{x} \in \overrightarrow{\mathcal{A}}$ there is one and only one $Q \in \mathcal{A}$ such that $\mathbf{x} = \overrightarrow{PQ}$.

$\mathcal{A}$ is often said to be the affine space associated to $\overrightarrow{\mathcal{A}}$, or conversely that $\overrightarrow{\mathcal{A}}$ is the associated vector space for the affine space $A$. It is also convenient to write $Q = P + \mathbf{v}$ or $\mathbf{v} = Q - P$, instead of $\mathbf{v} = \overrightarrow{PQ}$. The dimension of $\mathcal{A}$ is defined as that of $\overrightarrow{\mathcal{A}}$. When $\mathcal{A} = \overrightarrow{\mathcal{A}}$ and $\overrightarrow{PQ} = Q - P$, the vector space $\overrightarrow{\mathcal{A}}$ itself is regarded as an affine space.

A point $O \in \mathcal{A}$ (called the origin) and a vector basis $(\mathbf{e}_1, \ldots, \mathbf{e}_d)$ of (a finite $d-$dimensional) $\overrightarrow{\mathcal{A}}$ together are called a frame of reference in the affine space $\mathcal{A}$. The affine coordinates of a point $P \in \mathcal{A}$ in the frame of reference $(O; \mathbf{e}_1, \ldots, \mathbf{e}_d)$ are defined as the (vector) coordinates $(\alpha_1, \ldots, \alpha_d)$ of the vector $\mathbf{x} := \overrightarrow{OP}$ in the vector basis $(\mathbf{e}_1, \ldots, \mathbf{e}_d)$, viz.

$$\mathbf{x} := \overrightarrow{OP} = \sum_{i=1}^{d} \alpha_i \mathbf{e}_i.$$

If relative to the frame of reference $(O; \mathbf{e}_1, \ldots, \mathbf{e}_d)$, the point $P$ has coordinates $(\alpha_1, \ldots, \alpha_d)$, while the point $Q$ has coordinates $(\beta_1, \ldots, \beta_d)$, then the vector $\overrightarrow{PQ}$ has, with respect to the basis $(\mathbf{e}_1, \ldots, \mathbf{e}_d)$, coordinates

$$(\beta_1 - \alpha_1, \ldots, \beta_d - \alpha_d). \tag{25}$$

Further details on affine transformations, affine subspaces, etc..., can be found on any textbook on affine geometry, see, e.g., Gallier 2011, Shafarevich and Remizov 2013.

A.2. **Barycentric coordinates.** Instead of locating a point with respect to a frame made of a point and a vector basis, one can locate points in a reference system made solely of points. Barycentric coordinates, introduced by Möbius 1827, specify the location of a point w.r.t. a simplex (of $d+1$ points in a $d-$dimensional affine space). Barycentric calculus interprets as a method of treating geometry by considering a point as the center of gravity of certain other points to which weights are ascribed. It is particularly useful to describe triangle centers (the centroid, orthocenter, incenter, circumcenter, etc.), which enjoy simple barycentric coordinate representations with respect to the vertices of their reference triangle. Barycentric coordinates are used, e.g., in geometric modeling, in computer graphics, in geophysics, or in the finite element method for interpolation on polygons.

Barycentric coordinates are defined w.r.t. to a simplex of affine independent points. Hence, we recall the notion of affine independence:

**Definition A.2** (Affine independence)**.** A set $\{A_1, \ldots, A_N\}$ of $N$ points in a $d$-dimensional affine space, $d \geq 2$, is said to be affine independent if the $N-1$ vectors $\overrightarrow{A_1 A_k}$ , $k = 2, \ldots, N$, are linearly independent[12]. A simplex of affine independent

[12]Thus, necessarily, $N - 1 \leq d$.

points, i.e. a set of $d+1$ affinely independent points in a $d$-dimensional affine space, is simply called an affine frame.

Barycentric coordinates are then defined as follows:

**Definition A.3** (Barycentric coordinates)**.** Let $\{A_0, \ldots, A_d\}$ be $d+1$ affinely independent points in a $d$-dimensional affine space $\mathcal{A}$. Let $P$ be a given point. There are scalars $p_0, \ldots, p_d$, with $\sum_{i=0}^{d} p_i \neq 0$, such that, for all points $Q$,

$$(\sum_{i=0}^{d} p_i)\overrightarrow{QP} = \sum_{i=0}^{d} p_i \overrightarrow{QA_i}. \tag{26}$$

The elements of a $(d+1)$ tuple $(p_0, p_1, \ldots, p_d)$ that satisfies this equation are called barycentric coordinates of $P$ with respect to $\{A_0, \ldots, A_d\}$.

*Proof.* Since $\mathcal{A}$ is $d$-dimensional and $\{A_0, \ldots, A_d\}$ are affinely independent, there exist unique scalars $\alpha_i$, $i = 1, \ldots, d$, s.t. $P$ writes w.r.t. to the frame of reference $(A_0; \overrightarrow{A_0A_1}, \ldots, \overrightarrow{A_0A_d})$ as

$$P = A_0 + \sum_{i=1}^{d} \alpha_i \overrightarrow{A_0A_i}.$$

Thus, for all $Q \in \mathcal{A}$,

$$\begin{aligned}
\overrightarrow{QP} &= \overrightarrow{QA_0} + \overrightarrow{A_0P} \\
&= \overrightarrow{QA_0} + \sum_{i=1}^{d} \alpha_i (\overrightarrow{A_0Q} + \overrightarrow{QA_i}) \\
&= (1 - \sum_{i=1}^{d} \alpha_i)\overrightarrow{QA_0} + \sum_{i=1}^{d} \alpha_i \overrightarrow{QA_i}
\end{aligned}$$

Thus, equation (26) is satisfied with $p_0 = 1 - \sum_i a_i$, $p_i = \alpha_i$, $i = 1, \ldots, d$, and $\sum_{i=0}^{d} p_i = 1 \neq 0$. □

Conversely, a family of scalars $(p_0, \ldots, p_d)$ s.t. $\sum_{i=0}^{d} p_i \neq 0$ define a unique point $P$ via the vector $\overrightarrow{QP}$ of (26) as

$$P = Q + \overrightarrow{QP} = Q + \sum_{i=0}^{d} \frac{p_i}{\sum_{j=0}^{d} p_j} \overrightarrow{QA_i}, \tag{27}$$

where $Q$ can be chosen arbitrarily, (see, e.g., Gallier 2011 Lemma 2.1 (1)). Barycentric coordinates are then homogeneous: scaling each coordinate $p_i \leftarrow \lambda p_i$ by a common factor $\lambda \neq 0$ defines the same point $P$ in (27). Thus, in barycentric coordinates only ratios of coordinates are relevant. Therefore, in analogy with homogeneous coordinates of projective geometry, barycentric coordinates of the point $P$ will be denoted by $(p_0 : p_1 : \ldots : p_d)$. The affine independence of the affine frame insures that the barycentric coordinate representation of a point with respect to the affine frame is unique, up to scaling. Imposing the condition $\sum_{i=0}^{d} p_i = 1$ yields unicity of the $(p_0, \ldots, p_d)$ and result in *normalised barycentric coordinates*, which are sometimes given as the definition of barycentric coordinates. In this case, the point $P$

of (27) is simply written as a combination of the points in the frame,

$$P = \sum_{i=0}^{d} p_i A_i, \quad \text{with } \sum_{i=0}^{d} p_i = 1, \tag{28}$$

and is called the barycenter (with weight 1) of the weighted points $(p_i, A_i)$, $i = 0, \ldots, d$.

If $\sum_{i=0}^{d} p_i = 0$, then, by Gallier 2011 Lemma 2.1 (2),

$$\sum_{i=0}^{d} p_i \overrightarrow{QA_i}$$

is independent of $Q$ and thus defines a unique vector. Combining both cases allows to give a meaning to general linear combination of points

$$\sum_{i \in I} \lambda_i P_i, \quad \lambda_i \in \mathbb{R},$$

where $(P_i)_{i \in I}$ is a family of points and $I$ an index set: it will yield

i) either a point, if $\sum_{i \in I} \lambda_i \neq 0$, defined as the barycenter of the weighted points $(\lambda_i, P_i)$. If $\sum_{i \in I} \lambda_i \neq 1$, writing $P = \sum_{i \in I} \lambda_i P_i$ expresses $P$ in *homogeneous* barycentric coordinates w.r.t $(P_i)_{i \in I}$ and thus corresponds to the barycenter

$$P = \sum_{i \in I} \frac{\lambda_i}{\sum_{j \in I} \lambda_j} P_i$$

in *normalised* homogeneous coordinates (28). In other words, the point $P$ is given the weight $\sum_i \lambda_i$.

ii) or a vector, if $\sum_{i \in I} \lambda_i = 0$, defined as $\sum_{i=0}^{d} \lambda_i \overrightarrow{QP_i}$ with $Q$ chosen arbitrarily. In particular, the difference $P_1 - P_2$ of two points gives a vector, hereby justifying the notation of a vector in affine space as a difference of two points.

A.3. **Formula of the displacement vector from barycentric coordinates of points.** Let $\mathcal{F} = \{A_0, \ldots, A_d\}$ be $d+1$ affine independent points in the affine space $\mathbb{R}^d$. Let the points $M$, resp. $N$, with barycentric coordinates $(m_0 : \ldots : m_d)$, resp. $(n_0 : \ldots : n_d)$, w.r.t. $\mathcal{F}$. Then, the following key (elementary) lemma gives the displacement vector $\mathbf{v} = \overrightarrow{MN}$ from $M$ to $N$:

**Lemma A.4.** *For $M$ and $N$ defined by barycentric coordinates as*

$$M = \frac{\sum_i m_i A_i}{\sum_i m_i}, \quad N = \frac{\sum_j n_j A_j}{\sum_j n_j},$$

*the displacement vector $\mathbf{v} = \overrightarrow{MN}$ from $M$ to $N$ writes*

$$\mathbf{v} = \frac{\sum_{i<j} det \begin{vmatrix} m_i & n_i \\ m_j & n_j \end{vmatrix} \overrightarrow{A_i A_j}}{\sum_i m_i \sum_j n_j}. \tag{29}$$

*Proof.* One has, by definition of $M$ and $N$,

$$\begin{aligned}
\mathbf{v} = \overrightarrow{MN} &= -M + N = -\frac{\sum_i m_i A_i}{\sum_i m_i} + \frac{\sum_j n_j A_j}{\sum_j n_j} \\
&= \frac{\sum_i \sum_j n_j m_i (A_j - A_i)}{\sum_i m_i \sum_j n_j} \\
&= \frac{\sum_{i<j} n_j m_i (A_j - A_i) + \sum_{i>j} n_j m_i (A_j - A_i)}{\sum_i m_i \sum_j n_j} \\
&= \frac{\sum_{i<j} n_j m_i (A_j - A_i) + \sum_{i<j} n_i m_j (A_i - A_j)}{\sum_i m_i \sum_j n_j} \qquad (30) \\
&= \frac{\sum_{i<j} (m_i n_j - n_i m_j)(-A_i + A_j)}{\sum_i m_i \sum_j n_j} \\
&= \frac{\sum_{i<j} \det \begin{vmatrix} m_i & n_i \\ m_j & n_j \end{vmatrix} \overrightarrow{A_i A_j}}{\sum_i m_i \sum_j n_j}.
\end{aligned}$$

where (30) follows by exchanging the role of $i$ and $j$. □

Notice that $\mathbf{v}$ is homogeneous, i.e. is invariant w.r.t. rescalings $\mathbf{m} \leftarrow \lambda \mathbf{m}$ and $\mathbf{n} \leftarrow \mu \mathbf{n}$, $\lambda, \mu \neq 0$, of the barycentric coordinates of $M$ and $N$.

A.4. **Comparison with the formula of the displacement vector in a reference frame.** The subsequent discussion elucidates the distinction, noted in the introduction, between Cartesian and barycentric coordinates in calculating displacement vectors.

Assume $M$ and $N$ are expressed in normalized barycentric coordinates, i.e. $\sum_i m_i = \sum_i n_i = 1$, and set also

$$v_{ij} := m_i n_j - m_j n_i,$$

so that (29) reduces to the numerator, viz.

$$\overrightarrow{MN} = \sum_{i=0}^{d} \sum_{j=i+1}^{d} v_{ij} \overrightarrow{A_i A_j}. \qquad (31)$$

Such formula (31) expresses the displacement vector $\overrightarrow{MN}$ in terms of $d(d+1)/2$ components $v_{ij}$ along the coordinate axis vectors $\overrightarrow{A_i A_j}$, $i < j$. It is more common in affine geometry to write the displacement vector w.r.t. to a reference frame $(O; \mathbf{e}_1, \ldots, \mathbf{e}_d)$. W.lo.g. we take the affine frame with $A_0$ as origin and $\mathbf{e}_i = \overrightarrow{A_0 A_i}$, $i = 1, \ldots, d$. Indeed, by distinguishing $i = 0$ from $i \neq 0$ and decomposing $\overrightarrow{A_i A_j} =$

$\overrightarrow{A_iA_0} + \overrightarrow{A_0A_j} = \mathbf{e}_j - \mathbf{e}_i$ in (31), one can write

$$\begin{aligned}
\overrightarrow{MN} &= \sum_{j=1}^{d} v_{0j}\overrightarrow{A_0A_j} + \sum_{i=1}^{d}\sum_{j=i+1}^{d} v_{ij}\overrightarrow{A_iA_j} \\
&= \sum_{j=1}^{d} v_{0j}\mathbf{e}_j + \sum_{i=1}^{d}\sum_{j>i} v_{ij}\mathbf{e}_j + \sum_{i=1}^{d}\sum_{j>i}(-v_{ij})\mathbf{e}_i \qquad (32)\\
&= \sum_{j=1}^{d} v_{0j}\mathbf{e}_j + \sum_{i=1}^{d}\sum_{j>i} v_{ij}\mathbf{e}_j + \sum_{j=1}^{d} + \sum_{i>j}(-v_{ji})\mathbf{e}_j \\
&= \sum_{j=1}^{d} v_{0j}\mathbf{e}_j + \sum_{i=1}^{d}\sum_{j>i} v_{ij}\mathbf{e}_j + \sum_{j=1}^{d} + \sum_{i>j} v_{ij}\mathbf{e}_j \\
&= \sum_{j=1}^{d}(m_0n_j - n_0m_j)\mathbf{e}_j + \sum_{1\le i\ne j}^{d} v_{ij}\mathbf{e}_j \qquad (33)
\end{aligned}$$

where we exchanged $i$ and $j$ in the third sum of (32), used antisymmetry of $v_{ij}$, and the second sum in (33) denotes a double sum. By use of normalized barycentric coordinates,

$$m_0 = 1 - \sum_{i=1}^{d} m_i,$$
$$n_0 = 1 - \sum_{i=1}^{d} n_i.$$

Plugging these into the first sum of (33) yields

$$\begin{aligned}
\overrightarrow{MN} &= \sum_{j=1}^{d}(n_j - m_j)\mathbf{e}_j + \sum_{j=1}^{d}\sum_{i=1}^{d}(-n_jm_i + m_jn_i)\,\mathbf{e}_j + \sum_{1\le i\ne j}^{d} v_{ij}\mathbf{e}_j \\
&= \sum_{j=1}^{d}(n_j - m_j)\mathbf{e}_j + \sum_{i,j=1}^{d}(-v_{ij})\,\mathbf{e}_j + \sum_{1\le i\ne j}^{d} v_{ij}\mathbf{e}_j \\
&= \sum_{j=1}^{d}(n_j - m_j)\mathbf{e}_j, \qquad (34)
\end{aligned}$$

since $v_{ii} = 0$ by antisymmetry. Alternatively to these tedious but elementary[13] algebraic manipulations, note that one could have simply put $Q = A_0$ in (26) to obtain directly (34).

Formula (34) expresses the displacement vector between two points in $d+1$ barycentric coordinates as a difference $n_j - m_j$ of $d$ *independent* coordinates. Thus, one has a formula similar to the usual difference formula (25) of two vectors in Cartesian coordinates, with the notable difference that the zero barycentric coordinates $m_0$ and $n_0$ are absent. However, the formula is not symmetric w.r.t coordinate components as one component serves as origin and thus play a distinguished role.

[13]They are somehow illuminating for understanding the difference between Cartesian and Barycentric coordinates

Note that if one directly applies formula (26), then, obviously, for any point $Q$,

$$\overrightarrow{MN} = \overrightarrow{QN} - \overrightarrow{QM} = \sum_{j=0}^{d} (n_j - m_j)\,\overrightarrow{QA_j}, \tag{35}$$

so that the displacement vector is expressed as the difference of, this time, all $d+1$ barycentric coordinates along the vectors $(\overrightarrow{QA_0}, \overrightarrow{QA_1}, \ldots, \overrightarrow{QA_d})$. However, $(\overrightarrow{QA_0}, \overrightarrow{QA_1}, \ldots, \overrightarrow{QA_d})$ is not a vector basis, since it now contains $d+1$ vectors in a $d-$dimensional space: this explains why analysis of CoDa as plain vectors of $\mathbb{R}^{d+1}$, as was sometimes the usage in the early days of CoDa analysis prior to J. Aitchison 1986, is clearly ill-conceived. In addition, it introduces the extrinsic element $Q$, whose choice is arbitrary.

The lack of symmetry of (34) and the above criticisms of (35) explains why we base our further considerations on formula (29) of the key Lemma A.4.

A.5. **Illustrations.** The following simple examples illustrate the aforementioned concepts and Lemma A.4.

**Example A.1** ($d = 1$)**.** For $d = 1$, the affine space $\mathcal{A}$ is the line $\mathbb{R}$, and two distinct points $A_0, A_1$ on the line are affinely independent. Set $M, N \in \mathbb{R}$ with respective normalized barycentric coordinates w.r.t. $(A_0, A_1)$ as $M = (0.5 : 0.5)$, $N = (0.8 : 0.2)$, viz.

$$M = 0.5A_0 + 0.5A_1, \quad N = 0.8A_0 + 0.2A_1,$$

see Figure 7. Vectorially, this means, by taking $Q = A_0$ in (6), that

$$\overrightarrow{A_0M} = 0.5\overrightarrow{A_0A_0} + 0.5\overrightarrow{A_0A_1} = 0.5\overrightarrow{A_0A_1},$$
$$\overrightarrow{A_0N} = 0.8\overrightarrow{A_0A_0} + 0.2\overrightarrow{A_0A_1} = 0.2\overrightarrow{A_0A_1},$$

since $\overrightarrow{A_0A_0} = \overrightarrow{0}$ is the null vector. In other words, the position vectors $\overrightarrow{A_0M}$ and $\overrightarrow{A_0N}$ have respective (Cartesian) coordinates 0.5 and 0.2 in the frame of reference $(A_0, \overrightarrow{A_0A_1})$. The displacement vector

$$\overrightarrow{MN} = \overrightarrow{A_0N} - \overrightarrow{A_0M} = -0.3\overrightarrow{A_0A_1}$$

writes as the difference of the Cartesian coordinates $0.2 - 0.5 = -0.3$, and obviously coincides with formula (29) of Lemma A.4,

$$\overrightarrow{MN} = \det \begin{vmatrix} 0.5 & 0.8 \\ 0.5 & 0.2 \end{vmatrix} \overrightarrow{A_0A_1} = -0.3\overrightarrow{A_0A_1},$$

which is computed directly from the barycentric coordinates of the initial points.

Here, for $d = 1$, formula (29) give the same number of components in the decomposition of the displacement vector along the $d(d+1)/2$ axis pairs $\overrightarrow{A_iA_j}$, $i < j$, as the classical decomposition of the displacement vector along the $d$ vector basis $\overrightarrow{A_0A_i}$: one single component along $\overrightarrow{A_0A_1}$.

**Example A.2** ($d = 2$)**.** For $d = 2$, the affine space $\mathcal{A}$ is the plane $\mathbb{R}^2$, and any three distinct points $A_0, A_1, A_2$ not on the same line are affinely independent. Take, e.g., $A_0, A_1, A_2$ forming an equilateral triangle. Let $M, N \in \mathbb{R}^2$ with respective

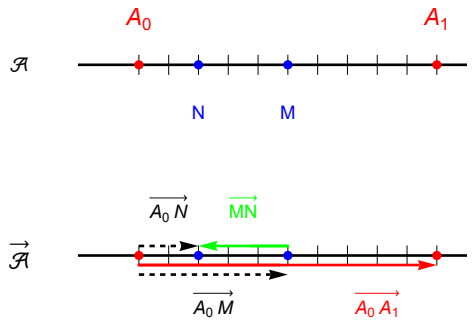


FIGURE 7. Illustration of barycentric coordinates and displacement vectors for $d = 1$. Up: two (blue) points $M, N$ with barycentric coordinates $M = (0.5 : 0.5)$, $N = (0.8 : 0.2)$ w.r.t. the (red) points $A_0, A_1$, on the affine space $\mathcal{A}$. Down: corresponding position vectors $\overrightarrow{A_0M}$, $\overrightarrow{A_0N}$ (dashed black arrows) w.r.t. to the vector basis $\overrightarrow{A_0A_1}$ (red arrow), with displacement vector $\overrightarrow{MN}$ (green arrow), in the associated vector space $\vec{\mathcal{A}}$.

normalized barycentric coordinates w.r.t. $(A_0, A_1, A_2)$ as $M = (1/3 : 1/3 : 1/3)$, $N = (0.8 : 0.2 : 0)$, viz.

$$M = \frac{1}{3}A_0 + \frac{1}{3}A_1 + \frac{1}{3}A_2, \quad N = \frac{4}{5}A_0 + \frac{1}{5}A_1 + 0A_2,$$

so that $M$ is on the center of the triangle and $N$ is on the edge $A_0A_1$, see Figure 8. In the frame of reference $(A_0; \overrightarrow{A_0A_1}, \overrightarrow{A_0A_2})$, the position vectors corresponding to $M$ and $N$ write

$$\overrightarrow{A_0M} = \frac{1}{3}\overrightarrow{A_0A_1} + \frac{1}{3}\overrightarrow{A_0A_2}$$
$$\overrightarrow{A_0N} = \frac{1}{5}\overrightarrow{A_0A_1}$$

i.e. have respective Cartesian coordinates $(1/5, 0)$ and $(1/3, 1/3)$ w.r.t. to the vector basis $(\overrightarrow{A_0A_1}, \overrightarrow{A_0A_2})$. The displacement vector writes classically as the difference of these Cartesian coordinates of position vectors,

$$\overrightarrow{MN} = -\frac{2}{15}\overrightarrow{A_0A_1} - \frac{1}{3}\overrightarrow{A_0A_2}. \tag{36}$$

On the other hand, formula (29) of Lemma A.4 writes

$$\begin{aligned}\overrightarrow{MN} &= \det\begin{vmatrix}1/3 & 4/5\\ 1/3 & 1/5\end{vmatrix}\overrightarrow{A_0A_1} + \det\begin{vmatrix}1/3 & 4/5\\ 1/3 & 0\end{vmatrix}\overrightarrow{A_0A_2} + \det\begin{vmatrix}1/3 & 1/5\\ 1/3 & 0\end{vmatrix}\overrightarrow{A_1A_2}\\ &= -\frac{1}{5}\overrightarrow{A_0A_1} - \frac{4}{15}\overrightarrow{A_0A_2} - \frac{1}{15}\overrightarrow{A_1A_2}. \end{aligned} \tag{37}$$

Since in any plane triangle one has $\overrightarrow{A_0A_1} + \overrightarrow{A_1A_2} + \overrightarrow{A_2A_0} = \overrightarrow{0}$, i.e.

$$\overrightarrow{A_1A_2} = -\overrightarrow{A_0A_1} + \overrightarrow{A_0A_2},$$

it is readily checked that (37) coincides with (36).

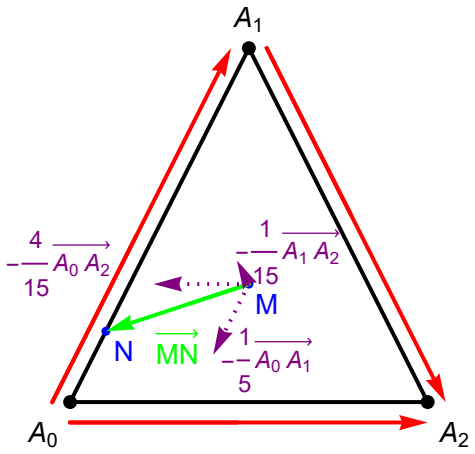


FIGURE 8. Illustration of the displacement vector formula (29) of Lemma A.4 for $d = 2$. Two (blue) points $M, N$ with barycentric coordinates $M = (1/3 : 1/3 : 1/3)$, $N = (4/5 : 1/5, 0)$ w.r.t. the affine frame $A_0, A_1, A_2$. The displacement vector $\overrightarrow{MN}$ (green arrow) decomposes along the coordinate axis $(\overrightarrow{A_0A_1}, \overrightarrow{A_0A_2}, \overrightarrow{A_1A_2})$ (red arrows) as the sum of $-\frac{1}{5}\overrightarrow{A_0A_1}$, $-\frac{4}{15}\overrightarrow{A_0A_2}$, $-\frac{1}{15}\overrightarrow{A_1A_2}$ (purple dotted arrows).

Compared to Example A.1, here, for $d = 2$, one has one extra axis component, $\overrightarrow{A_1A_2}$, in the decomposition of the displacement vector in the $d(d+1)/2 = 3$ coordinate axis $\overrightarrow{A_iA_j}$, $i < j$, made of ordered pairs of distinct points $(A_i, A_j)$, provided by Lemma A.4, compared to the classical decomposition of the displacement vector on the $d = 2$ vectors basis $(\overrightarrow{A_0A_1}, \overrightarrow{A_0A_2})$.

## APPENDIX B. NUMERICAL EXPERIMENTS ON BARYCENTRIC FRÉCHET MEANS

For $\alpha = 2$, the barycentric Fréchet mean sometimes corresponds to the arithmetic mean, i.e. the centroid defined as the arithmetic average of the data points on the

simplex, viz.

$$[\overline{\mathbf{x}}]_+ := \left[\frac{1}{n}\sum_{i=1}^{n}\mathcal{C}(\mathbf{x}^i)\right]_+,$$

where $\mathcal{C}$ is the closure operation of (1). Also, the barycentric median ($\alpha = 1$) may not be unique. The following toy example illustrates these points.

**Example B.1** (Toy example: centroid/barycenter of the triangle)**.** A simple calculation shows that the 2-barycentric Fréchet mean $(m_0, m_1, m_2) \in \Delta_d^+$ of the vertices of the basic triangle $A_1A_2A_3$, with $A_1 = (1, 0, 0)$, $A_2 = (0, 1, 0)$, $A_3 = (0, 0, 1)$ minimizes the Lagrangian

$$L(\mathbf{m}) := 2(m_0^1 + m_1^2 + m_2^2) + \lambda(m_0 + m_1 + m_2 - 1),$$

where $\lambda$ is the Lagrange multiplier, and is thus easily seen to be equal to $\mathbf{m} = (1/3, 1/3, 1/3)$, i.e. the barycenter or centroid of the triangle.

On the other hand, the Fréchet functional for the 1-divergence is constant and equal to 2, hence any point in the triangle is a 1-barycentric Fréchet mean. For the $\infty$-divergence, the Fréchet functional writes, for normalized $\mathbf{m} \in \Delta_d^+$, as

$$\max(m_0, m_1) + \max(m_0, m_2) + \max(m_1, m_2),$$

and is also minimal for the barycenter.

Note that this toy example, how trite as it may appear, illustrates a case which can not be dealt with Aitchison's log-ratio approaches, since the data contains some zero components.

When CoDa is one-dimensional (i.e., has two components), one can prove that the empirical Fréchet mean based on the $2-$barycentric divergence coincides with the arithmetic mean, as shown in the following Proposition.

**Proposition B.1.** *For $d = 1$, the empirical Fréchet mean based on the $2-$divergence is the arithmetic mean.*

*Proof.* Assume $\mathbf{x}^1, \ldots, \mathbf{x}^n$ is a simplex-normalized sample, so that $\mathbf{x}^i = (x_0^i, x_1^i) \in \Delta_+^1$, $i = 1, \ldots, n$. The normalized Fréchet mean $\mathbf{m} \in \Delta_+^1$ minimizes

$$\begin{aligned} F(\mathbf{m}) &= \sum_{i=1}^{n}(x_0^i m_1 - x_1^i m_0)^2 = \sum_{i=1}^{n}(x_0^i - x_0^i m_0 - x_1^i m_0)^2 \\ &= \sum_{i=1}^{n}(x_0^i - m_0)^2, \end{aligned}$$

since $m_0 + m_1 = 1$ and $x_0^i + x_1^i = 1$. The latter is obviously minimized by taking the arithmetic mean of the first coordinate $m_0 = \frac{1}{n}\sum_{i=1}^{n} x_0^i$, which yields $m_1 = \frac{1}{n}\sum_{i=1}^{n} x_1^i$. □

However, in general, the Fréchet mean based on the $2-$barycentric divergence is different from the arithmetic (linear) mean. We illustrate this fact with the simple Example B.2, where the data is made of only two data points.

**Example B.2** (Toy counter-example: two data points)**.** One considers two data points $(0.05, 0.85, 0.15)$, $(0.3, 0.2, 0.5)$, located on the left of the triangle, as depicted by the blue points on Figure 9, together with the barycentric Fréchet means for $\alpha = 2$ (orange square), $\alpha = 1$ (green lozenge), $\alpha = \infty$ (downward violet triangle)

and the arithmetic mean (red upward triangle). Here, only the 2-barycentric mean remain close, yet distinct, from the arithmetic mean, which is the mid-point between the to data points. All Fréchet means appears somehow skewed towards the right of the triangle, away from the segment line where the data sits.

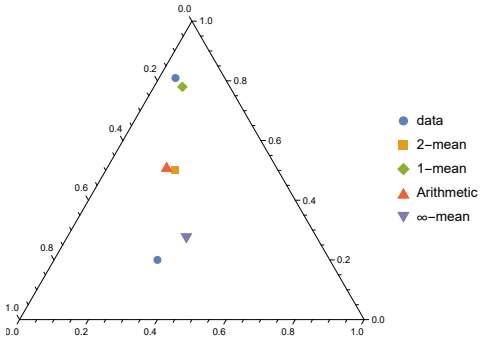


FIGURE 9. Comparison of the means for a toy example of 2 data points. Sample points (blue), Arithmetic mean (red upward triangle), Fréchet means for the $\alpha$-divergence: $\alpha = 2$ (orange square), $\alpha = 1$ (green lozenge), $\alpha = \infty$ (downward violet triangle).

A more comprehensive picture of the influence of the $\alpha$ parameter on the location of the $\alpha-$barycentric mean is given in Example B.3, still very basic with only three data points. In general, when the number of points is larger, the different $\alpha-$means have a tendency to be less spread apart.

**Example B.3** (A counter example with three points)**.** One consider the means of the three (blue) data points $(1/8, 1/8, 3/4)$, $(1/17, 12/17, 4/17)$, and $(4/9, 1/9, 4/9)$, see Figure 10. The 2-Fréchet mean is, approximately, $(0.160, 0.462, 0.378)$ and is depicted by the red upward triangle, while the arithmetic mean (red circle) is approximately equals to $(0.140, 0.467, 0.393)$, and are thus clearly different. Note that the means remain different if one replaces the normalization by $\ell_1$-norm in the denominator of the 2-divergence by the $\ell_2$ norm. Also shown on the Figure are the Fréchet mean based on the 1-divergence (green lozenge), the Fréchet mean for the $\infty$-divergence (violet downward triangle), and several barycentric Fréchet mean for the $\alpha$-divergence (orange squares), for $\alpha$ varying from 1.1 to 10. For $\alpha = 1$, the Fréchet mean corresponds statistically to the "median" (the Fermat-Weber-Torricelli problem), albeit with a different notion of "distance".

When the data have a special geometric structure, in particular when it is on a straight (Euclidean) lines, the different Fréchet means exhibit intriguing properties, as illustrated in the next two examples.

**Example B.4.** Let $a_0, a_1$ be independent, uniformly distributed r.v. on $[0, 1]$ and set $a_2 = a_1$, so that the raw amounts $a_1$ and $a_2$ are co-monotonic. CoDa is obtained by closure, i.e. $\mathbf{x} = \mathcal{C}(\mathbf{a})$. A sample of $n = 10$ i.i.d. replications of $\mathbf{x}$ is shown on Figure 11 (blue points) and sits on the straight line $x_1 = x_2$ in the triangle. The Fréchet Means for $\alpha = 2$, $\alpha = 1$, $\alpha = \infty$ , together with the arithmetic mean, are computed and displayed on Figure 11. It is noteworthy to remark that four

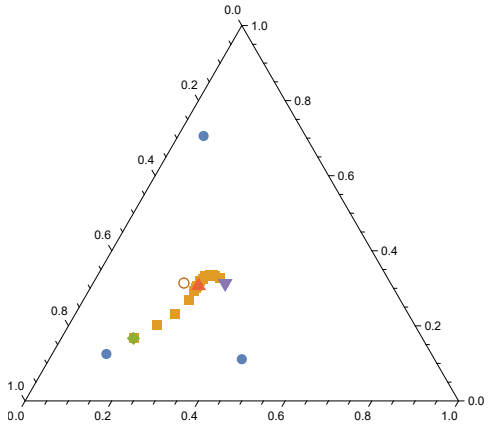


FIGURE 10. Comparison of the means for a toy example of 3 data points. Sample points (blue), Fréchet mean for the 2 divergence (red upward triangle), arithmetic mean (red circle), Fréchet mean for the 1-divergence (green lozenge), Fréchet mean for the $\infty$-divergence (violet downward triangle). Fréchet mean for the $\alpha$-divergence (orange squares), for $\alpha$ varying from 1.1 to 10.

| Mean | $x_0$ | $x_1$ | $x_2$ |
|---|---|---|---|
| $n = 10$ | | | |
| Arithmetic | 0.519382 | 0.240309 | 0.240309 |
| 2− barycentric | 0.519382 | 0.240309 | 0.240309 |
| 1− barycentric | 0.616009 | 0.191995 | 0.191995 |
| $\infty$− barycentric | 0.615063 | 0.192469 | 0.192469 |
| $n = 50$ | | | |
| Arithmetic | 0.335202 | 0.332399 | 0.332399 |
| 2− barycentric | 0.335202 | 0.332399 | 0.332399 |
| 1− barycentric | 0.310344 | 0.344828 | 0.344828 |
| $\infty$− barycentric | 0.310345 | 0.344828 | 0.344828 |

TABLE 3. Comparison of the different kind of means of Example B.4.

means considered respect the geometry of the data, in the sense that they all lie on the line $x_1 = x_2$. On this example, the 2-barycentric Fréchet mean (orange square) and the arithmetic mean (green lozenge) coincide, and are distinct from the 1-barycentric Fréchet mean (red upward triangle), resp. $\infty$-barycentric Fréchet mean (violet downward triangle). As the sample size grows, we notice empirically that the 1 and $\infty$ barycentric Fréchet mean seem to become indistinguishable, on this example, see Table 3.

The previous example showed barycentric Fréchet means aligned with the line where the data sits. However, the picture in Figure 11 is somehow misleading and is due to the symmetry in the data of Example B.4. In general, this is not the case, as illustrated on the following example.

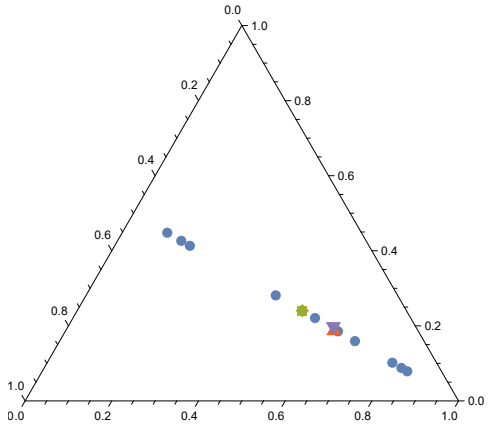


FIGURE 11. Comparison of the means for Example B.4 of 10 data points on a line. Sample points (blue), Fréchet mean for the 2 divergence (orange square), arithmetic mean (green lozenge), Fréchet mean for the 1-divergence ((red upward triangle)), Fréchet mean for the $\infty$-divergence (violet downward triangle).

**Example B.5.** As in Example B.4, we sample $a_0, a_1$ as independent, uniformly distributed r.v. on $[0,1]$, but we now set $a_2 = a_0 + a_1$, and eventually, $\mathbf{x} = \mathcal{C}(\mathbf{a})$. The data (blue points) sits on the line $x_2 = 1/2$ displayed on Figure 12. In this example, only the arithmetic mean and the $\infty$-barycentric mean sit on the line $x_2 = x_0 + x_1 = 0.5$. In addition, the $2-$barycentric Fréchet mean now clearly differs from the arithmetic mean. All mean remain close to each others, yet different, and seem to converge to the same point as the sample size increases.

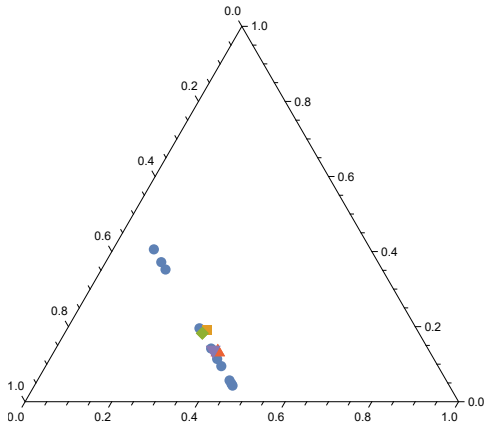


FIGURE 12. Comparison of the means for Example B.5 of 10 data points on a line. Sample points (blue), Fréchet mean for the 2 divergence (orange square), arithmetic mean (green lozenge), Fréchet mean for the 1-divergence ((red upward triangle)), Fréchet mean for the $\infty$-divergence (violet downward triangle).

| Mean | $x_0$ | $x_1$ | $x_2$ |
|---|---|---|---|
| $n = 10$ | | | |
| Arithmetic | 0.317668 | 0.182332 | 0.5 |
| $2-$ barycentric | 0.324976 | 0.191089 | 0.483935 |
| $1-$ barycentric | 0.375724 | 0.137818 | 0.486458 |
| $\infty-$ barycentric | 0.368017 | 0.131983 | 0.5 |
| $n = 50$ | | | |
| Arithmetic | 0.236695 | 0.263305 | 0.5 |
| $2-$ barycentric | 0.241752 | 0.268185 | 0.490062 |
| $1-$ barycentric | 0.238849 | 0.266468 | 0.494683 |
| $\infty-$ barycentric | 0.237808 | 0.262192 | 0.5 |

TABLE 4. Comparison of the different kind of means of Example B.5.

## ACKNOWLEDGEMENTS

Olivier P. Faugeras acknowledges funding from the French National Research Agency (ANR) under the Investments for the Future (Investissements d'Avenir) program, grant ANR-17-EURE-0010.

## REFERENCES

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. Monographs on Statistics and Applied Probability. London: Chapman & Hall, pp. xvi+416. ISBN: 0-412-28060-4. DOI: 10.1007/978-94-009-4109-0.

Aitchison, John (1992). "On Criteria For Measures of Compositional Difference". In: *Math. Geol.* 24.4, pp. 365–379. ISSN: 0882-8121. DOI: 10.1007/BF00891269. URL: https://doi.org/10.1007/BF00891269.

Baharav, Tavor Z. and David Tse (2019). "Ultra Fast Medoid Identification Via Correlated Sequential Halving". In: *Proceedings of the 33rd International Conference on Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc., p. 10.

Bhattacharjee, Satarupa and Hans-Georg Müller (2023). *Single Index Fréchet Regression*. arXiv: 2108.05437 [stat.ME]. URL: https://arxiv.org/abs/2108.05437.

Breiman, Leo (2001). "Random Forests". In: *Machine learning* 45, pp. 5–32.

Capitaine, Louis et al. (2024). "Fréchet Random Forests for Metric Space Valued Regression with Non Euclidean Predictors". In: *J. Mach. Learn. Res.* 25, Paper No. 355, 41. ISSN: 1532-4435.

Chayes, Felix (1971). *Ratio Correlation: a Manual for Students of Petrology and Geochemistry*. Chicago: University of Chicago Press.

Comas-Cufí, Marc (2023). *coda.base: A Basic Set of Functions for Compositional Data Analysis*. R package version 0.5.5. URL: https://CRAN.R-project.org/package=coda.base.

Dong, Yushen and Yichao Wu (2022). "Fréchet Kernel Sliced Inverse Regression". In: *Journal of Multivariate Analysis* 191, p. 105032. ISSN: 0047-259X. DOI: https://doi.org/10.1016/j.jmva.2022.105032. URL: https://www.sciencedirect.com/science/article/pii/S0047259X22000495.

Egozcue, J. J., J. L. Díaz-Barrero, and V. Pawlowsky-Glahn (2006). "Hilbert Space of Probability Density Functions Based on Aitchison Geometry". In: *Acta Math. Sin. (Engl. Ser.)* 22.4, pp. 1175–1182. ISSN: 1439-8516. DOI: 10.1007/s10114-005-0678-2. URL: https://doi.org/10.1007/s10114-005-0678-2.

Egozcue, Juan José and Vera Pawlowsky-Glahn (2023). "Subcompositional Coherence and a Novel Proportionality Index of Parts". In: *SORT* 47.2, pp. 229–244. ISSN: 1696-2281. DOI: 10.1504/ijor.2023.131493. URL: https://doi.org/10.1504/ijor.2023.131493.

Erb, Ionas and Cedric Notredame (2016). "How Should We Measure Proportionality on Relative Gene Expression Data?" In: *Theory in Biosciences* 135, pp. 21–36.

Everitt, Brian S. et al. (2011). *Cluster Analysis*. Fifth. Wiley Series in Probability and Statistics. Chichester: John Wiley & Sons, Ltd., pp. xvi+330. ISBN: 978-0-470-74991-3. DOI: 10.1002/9780470977811.

Fan, J. and I. Gijbels (1996). *Local Polynomial Modelling and its Applications*. Vol. 66. Monographs on Statistics and Applied Probability. London: Chapman & Hall, pp. xvi+341. ISBN: 0-412-98321-4.

Faugeras, Olivier P. (Dec. 2023). "An Invitation to Intrinsic Compositional Data Analysis Using Projective Geometry and Hilbert's Metric". TSE Working Paper, no. 23-1496. URL: https://hal.science/hal-04344403v2.

– (May 2025a). "Log-Free Distance and Covariance Matrix for Compositional Data II: The Projective/Exterior Product Approach". In: *Austrian Journal of Statistics* 54, pp. 121–159. URL: https://ajs.or.at/index.php/ajs/article/view/2105.

– (Feb. 2025b). "Log-Free Divergence and Covariance Matrix for Compositional Data I: The Affine/Barycentric Approach". Supplementary material.

– (Jan. 2026). "The Stick-Breaking and Ordering Representation of Compositional Data: Copulas and Regression models". In: *Austrian Journal of Statistics* 55, pp. 1–31. URL: https://hal.science/hal-04409705.

Filzmoser, Peter and Karel Hron (2009). "Correlation Analysis for Compositional Data". In: *Mathematical Geosciences* 41, pp. 905–919.

Fréchet, Maurice (Apr. 1948). "Les Éléments Aléatoires de Nature Quelconque Dans un Espace Distancié". In: *Annales de l'institut Henri Poincaré*. Vol. 10, pp. 215–310.

Gajer, Pawel and Jacques Ravel (2025). *A New Approach to Compositional Data Analysis using $L^\infty$-normalization with Applications to Vaginal Microbiome*. arXiv: 2503.21543 [stat.CO]. URL: https://arxiv.org/abs/2503.21543.

Gallier, Jean (2011). *Geometric Methods and Applications*. Second. Vol. 38. Texts in Applied Mathematics. New York: Springer, pp. xxviii+680. ISBN: 978-1-4419-9960-3. DOI: 10.1007/978-1-4419-9961-0.

Gauss, Carl Friedrich (1816). "Bestimmung der Genauigkeit der Beobachtungen". In: *Zeitschrift für Astronomie und Verwandte Wissenschaften* 1, pp. 187–197.

Genuer, Robin et al. (2020). *Random Forests*. Cham, Switzerland: Springer.

Ghojogh, Benyamin et al. (2023). *Elements of Dimensionality Reduction and Manifold Learning*. Cham, Switzerland: Springer.

Ghosal, Aritra, Wendy Meiring, and Alexander Petersen (2023). "Fréchet Single Index Models For Object Response Regression". In: *Electronic Journal of Statistics* 17.1, pp. 1074–1112. DOI: 10.1214/23-EJS2120. URL: https://doi.org/10.1214/23-EJS2120.

Greenacre, Michael (2018). *Compositional Data Analysis in Practice*. New York: CRC press.

– (2020). "Amalgamations are valid in compositional data analysis, can be used in agglomerative clustering, and their logratios have an inverse transformation". In: *Applied Computing and Geosciences* 5, p. 100017. ISSN: 2590-1974. DOI: https://doi.org/10.1016/j.acags.2019.100017. URL: https://www.sciencedirect.com/science/article/pii/S2590197419300175.

– (2021). "Compositional Data Analysis". In: *Annu. Rev. Stat. Appl.* 8, pp. 271–299. ISSN: 2326-8298. DOI: 10.1146/annurev-statistics-042720-124436. URL: https://doi.org/10.1146/annurev-statistics-042720-124436.

Hein, Matthias (2009). "Robust Nonparametric Regression with Metric-Space Valued Output". In: *Advances in Neural Information Processing Systems*. Ed. by Y. Bengio et al. Vol. 22. New York: Curran Associates, Inc.

Iao, Su I, Yidong Zhou, and Hans-Georg Müller (2024). *Deep Fréchet Regression*. arXiv: 2407.21407 [stat.ME]. URL: https://arxiv.org/abs/2407.21407.

Kaufman, Leonard and Peter J. Rousseeuw (1990). "Partitioning Around Medoids (Program PAM)". In: *Finding Groups in Data*. Hoboken, New Jersey: John Wiley & Sons, Ltd, pp. 68–125. ISBN: 9780470316801. DOI: https://doi.org/10.1002/9780470316801.ch2. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470316801.ch2.

Loader, Clive (1999). *Local Regression and Likelihood*. Statistics and Computing. New York: Springer-Verlag, pp. xiv+290. ISBN: 0-387-98775-4.

Lovell, David et al. (2015). "Proportionality: a Valid Alternative to Correlation for Relative Data". In: *PLoS computational biology* 11.3, e1004075.

Lubbe, Sugnet, Peter Filzmoser, and Matthias Templ (2021). "Comparison of Zero Replacement Strategies For Compositional Data With Large Numbers of Zeros". In: *Chemometrics and Intelligent Laboratory Systems* 210, p. 104248. ISSN: 0169-7439. DOI: https://doi.org/10.1016/j.chemolab.2021.104248. URL: https://www.sciencedirect.com/science/article/pii/S0169743921000162.

Martín-Fernández, Josep-Antoni et al. (2015). "Bayesian-Multiplicative Treatment of Count Eeros in Compositional Data Sets". In: *Statistical Modelling* 15.2, pp. 134–158. DOI: 10.1177/1471082X14535524.

Möbius, August Ferdinand (1827). *Der Barycentrische Calcul, ein Hülfsmittel zur Analytischen Behandlung der Geometrie (etc.)* Leipzig: Barth.

Nearing, Jacob T et al. (2022). "Microbiome Differential Abundance Methods Produce Different Results Across 38 Datasets". In: *Nature communications* 13.1, p. 342.

Palarea-Albaladejo, Javier and Josep Antoni Martín-Fernández (2015). "zCompositions — R Package for Multivariate Imputation of Left-Censored Data Under a Compositional Approach". In: *Chemometrics and Intelligent Laboratory Systems* 143, pp. 85–96. ISSN: 0169-7439. DOI: https://doi.org/10.1016/j.chemolab.2015.02.019. URL: https://www.sciencedirect.com/science/article/pii/S0169743915000490.

Park, Junyoung, Jeongyoun Ahn, and Cheolwoo Park (July 2023). "Kernel Sufficient Dimension Reduction and Variable Selection for Compositional Data via Amalgamation". In: *Proceedings of the 40th International Conference on Machine Learning*. Ed. by Andreas Krause et al. Vol. 202. Proceedings of Machine

Learning Research. PMLR, pp. 27034–27047. URL: https://proceedings.mlr.press/v202/park23a.html.

Park, Junyoung, Changwon Yoon, et al. (July 2022). "Kernel Methods for Radial Transformed Compositional Data with Many Zeros". In: *Proceedings of the 39th International Conference on Machine Learning*. Ed. by Kamalika Chaudhuri et al. Vol. 162. Proceedings of Machine Learning Research. PMLR, pp. 17458–17472. URL: https://proceedings.mlr.press/v162/park22d.html.

Pawlowsky-Glahn, Vera, Juan José Egozcue, and Raimon Tolosana-Delgado (2015). *Modeling and Analysis of Compositional Data*. Statistics in Practice. Chichester: John Wiley & Sons, Ltd., pp. xx+247. ISBN: 978-1-118-44306-4.

Pearson, Karl (1897). "Mathematical Contributions to the Theory of Evolution.—on a Form of Spurious Correlation Which May Arise When Indices Are Used in the Measurement of Organs". In: *Proceedings of the Royal Society of London* 60.359-367, pp. 489–498.

Pedoe, Daniel (1970). "Thinking Geometrically". In: *Amer. Math. Monthly* 77.7, pp. 711–721. ISSN: 0002-9890. DOI: 10.2307/2316201. URL: https://doi.org/10.2307/2316201.

Petersen, Alexander and Hans-Georg Müller (2019). "Fréchet Regression For Random Objects With Euclidean Predictors". In: *Ann. Statist.* 47.2, pp. 691–719. ISSN: 0090-5364. DOI: 10.1214/17-AOS1624. URL: https://doi.org/10.1214/17-AOS1624.

Qiu, Rui, Zhou Yu, and Ruoqing Zhu (2024). "Random Forest Weighted Local Fréchet Regression with Random Objects". In: *Journal of Machine Learning Research* 25.107, pp. 1–69. URL: http://jmlr.org/papers/v25/23-0811.html.

Rousseeuw, Peter J and Christophe Croux (1993). "Alternatives to the Median Absolute Deviation". In: *Journal of the American Statistical association* 88.424, pp. 1273–1283.

Scealy, J. L. and A. H. Welsh (2014). "Colours and Cocktails: Compositional Data Analysis 2013 Lancaster Lecture". In: *Aust. N. Z. J. Stat.* 56.2, pp. 145–169. ISSN: 1369-1473. DOI: 10.1111/anzs.12073. URL: https://doi.org/10.1111/anzs.12073.

Shafarevich, Igor R. and Alexey O. Remizov (2013). *Linear Algebra and Geometry*. Translated from the 2009 Russian original by David Kramer and Lena Nekludova. Heidelberg: Springer, pp. xxii+526. ISBN: 978-3-642-30993-9. DOI: 10.1007/978-3-642-30994-6.

Simovici, Dan A (2021). *Clustering: Theoretical and Practical Aspects*. Singapore: World Scientific. DOI: 10.1142/12394. eprint: https://www.worldscientific.com/doi/pdf/10.1142/12394.

Tsagris, Michail, Abdulaziz Alenazi, and Connie Stewart (2023). "Flexible Non-Parametric Regression Models for Compositional Response Data With Zeros". In: *Stat. Comput.* 33.5, Paper No. 106, 17. ISSN: 0960-3174. DOI: 10.1007/s11222-023-10277-5. URL: https://doi.org/10.1007/s11222-023-10277-5.

Tsagris, Michail T, Simon Preston, and Andrew TA Wood (2011). "A Data-Based Power Transformation for Compositional Data". In: *Proceedings of the 4th Compositional Data Analysis Workshop*. Girona, Spain.

Van Den Boogaart, K. Gerald and Raimon Tolosana-Delgado (2013). *Analyzing Compositional Data With R*. Use R! Heidelberg: Springer, pp. xvi+258. ISBN: 978-3-642-36808-0. DOI: 10.1007/978-3-642-36809-7.

Wang, David and Joseph Eppstein (2006). "Fast Approximation of Centrality". In: *Graph algorithms and applications* 5.5, p. 39.

Weng, Jiaying, Chenlu Ke, and Pei Wang (Jan. 2024). "Sparse Fréchet Sufficient Dimension Reduction via Nonconvex Optimization". In: *Conference on Parsimony and Learning*. Ed. by Yuejie Chi et al. Vol. 234. Proceedings of Machine Learning Research. USA: PMLR, pp. 39–53. URL: https://proceedings.mlr.press/v234/weng24a.html.

Ying, Chao and Zhou Yu (2020). *Fréchet Sufficient Dimension Reduction for Random Objects*. arXiv: 2007.00292 [math.ST]. URL: https://arxiv.org/abs/2007.00292.

Zhang, Qi, Lingzhou Xue, and Bing Li (2022). *Dimension Reduction for Fréchet Regression*. arXiv: 2110.00467 [stat.ME]. URL: https://arxiv.org/abs/2110.00467.

Toulouse School of Economics, Université Toulouse 1 Capitole, Institut de Mathématiques de Toulouse, 1 Esplanade de l'université, 31080 Toulouse Cedex 06, France

*Email address*: olivier.faugeras@tse-fr.eu

*URL*: https://www.tse-fr.eu/fr/people/olivier-faugeras