WORKING PAPERS

N° 1580

October 2024

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

Olivier Faugeras

Toulouse School of Economics

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, France.

13 August 2024

Abstract

The presence of zeroes in Compositional Data (CoDa) is a thorny issue for Aitchison's classical log-ratio analysis. Building upon our previous geometric approach (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 will be studied in the separate companion paper Faugeras (2024a).

Keywords: Compositional data, Affine geometry, Barycentric coordinates, Barycentric divergence, Fréchet mean, Gaussian distribution, Barycentric variation matrix.

MSC2020: 62R20, 62H20.

*olivier.faugeras@tse-fr.eu

Contents

1 Introduction

1.1 A primer on Compositional data (CoDa)

Compositional data (CoDa) analysis deals with the statistical analysis of non-negative multivariate data $\mathbf{a} = (a_0, a_1, \ldots, a_d) \in \mathbb{R}^{d+1}$, where each $a_i \geq 0$ describe the amount of the ith component of a composition. It is understood that the raw magnitude a_i of any component does not have any significance in itself, but only in its proportion relative to other components. 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 traditional approach (Aitchison (1986)) normalizes the raw composition vector $\mathbf{a}$ by its sum, i.e. expresses the data in percentages, an operation called closure in the CoDa literature,

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

where $||.||_1$ stands for the ℓ_1 norm, and the equality follows since $\boldsymbol{a} \geq \mathbf{0}$. This leads to the consideration of *normalised* (i.e. after rescaling to unit sum) CoDa element as a vector

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

constrained to take its values in the d dimensional unit simplex,

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

Due to these unit-sum and non-negativity constraints, CoDa are no longer genuine vectors and thus can not be directly tackled with classical statistical multivariate analysis. As a consequence, it is well known (see, e.g., Pearson (1897), Chayes (1971)) that the naive covariance matrix Σ of CoDa components, i.e.

$$\Sigma := \big(\mathrm{cov}(x_i, x_j)\big) \in \mathbb{R}^{(d+1)^2} \tag{3}$$

is non-informative, exhibiting spurious correlation and being subcompositionally incoherent. Therefore, these simplex-normalised compositional data points are studied through a variety of log-ratio transforms, (alr, clr, ilr), which mandates to restrict attention to the positive simplex $\Delta_{++}^d = \{\boldsymbol{x} > \mathbf{0}, \boldsymbol{x} \in \Delta_+^d\}$. These transforms turn the positive simplex into an Euclidean vector space in log-transformed coordinates, on which the classical techniques apply. For recent accounts on this classical vector space approach, see, e.g., Greenacre (2018), Vera Pawlowsky-Glahn, Juan José Egozcue, and Tolosana-Delgado (2015), Boogaart and Tolosana-Delgado (2013).

In particular, the CoDa literature has looked for a CoDa analogue of the classical variance matrix of Euclidean vectors in order to study the codependence of components. The main approaches promoted are those based on the variance matrix of log-ratio transformed (alr,clr,ilr) data: Aitchison (1986) defines the variation matrix (also called the log-ratio variance matrix) $T := \left(t_{ij}\right) \in \mathbb{R}^{(d+1)^2}$, by

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

Similarly, the clr-variance matrix $\Sigma^{\text{clr}} \in \mathbb{R}^{(d+1)^2}$, resp. the ilr-variance matrix $\Sigma^{\text{clr}} \in \mathbb{R}^{d^2}$, is the variance-covariance matrix of the clr-transformed dataset, resp., ilr-transformed data, see e.g. Aitchison (1986), Greenacre (2018)) or Boogaart and Tolosana-Delgado (2013).

1.2 Motivation

In spite of its many successes, the main challenge of the classical log-ratio transformed vector space approach is its *intrinsic* inability to handle CoDa with some zeroes components: the logarithm is (negative) infinite for zero. Thus, log-ratios of CoDa components are undefined whenever the numerator or denominator of the ratio is zero. Such an issue can be severe and fundamental to some fields of applications, like chemometrics and bioinformatics, where compositional data sets (e.g. microbiome data) typically exhibit a high proportion of zeros. This has motivated a growing literature (and debate) on effective methods for replacing zeroes, see e.g. Lubbe, Filzmoser, and Templ (2021), Martín-Fernández et al. (2015) and the references therein for a review of some of (the more and more involved) imputation methods.

Such zeroes issue is particularly prevalent for the log-ratio variation matrices (4) and its relatives based on ilr and clr: due to the large (negative) values of the logarithm function when the argument is small, large relative error are likely to occur in the log-ratio variance of the small/replaced components, thus distorting any multivariate analysis based on such log-ratio variances. Any log-ratio based analysis risks to become particularly sensitive to the imputation method used to replace the zeroes (see Greenacre (2021)).

This motivates the search of possible alternative representations of CoDa on the full simplex, in order to have a direct and unified way to deal with CoDa, with or without zeroes.

1.3 Aims and scope

In Faugeras (2023), we proposed a geometric view of CoDa as an element of the projectivization $\mathbb{P}^d_+$ of the non-negative orthant cone $\mathbb{R}^{d+1}_+$: CoDa are defined as projective points in the space $\mathbb{P}^d_+$ of equivalence classes $[\mathbf{x}]_+$ of non-negative vectors $\boldsymbol{x} \geq \mathbf{0}$, $\mathbf{x} \in \mathbb{R}^{d+1}$, where $[\mathbf{x}]_+$ is the equivalence class for the (positive) scaling relation,

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

The components of $\mathbf{x}$ are homogeneous coordinates of $[\mathbf{x}]_+$, and are noted $[x_0 : x_1 : \ldots : x_d]$. In turn, these equivalence classes admits several representatives, which corresponds to several geometric models. In particular, the simplex Δ^d_+ is one particular affine model of $\mathbb{P}^d_+$.

This change of perspective allows to study CoDa using the tools and framework of projective and/or affine geometry. Indeed, a CoDa element $\mathbf{x} \in \Delta_+^d$ of the simplex should not be envisioned as a *vector* embedded in a $d+1$ dimensional Euclidean space $\mathbb{R}^{d+1}$, with $(x_0, x_1, \ldots, x_d)$ as *Cartesian* coordinates, but as an *affine representation* of a *projective point* in the $d-$dimensional projective space $\mathbb{P}_+^d$, with $[x_0 : x_1 : \ldots : x_d]$ as *homogeneous* coordinates. Thus, a CoDa projective point $[\mathbf{x}]_+$ identifies with an *affine point* $\mathbf{x} \in \Delta_+^d$, and $(x_0, x_1, \ldots, x_d)$ are then its *barycentric* (and not Cartesian) coordinates. As a consequence, the displacement vector from point $[\mathbf{y}]_+$ to $[\mathbf{x}]_+$ 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 the Cartesian coordinates of simplex representatives, (as would be the case for $\mathbf{x}, \mathbf{y}$ vectors), but is given by a more complicated formula (see the forthcoming Lemma 2.4). Consequently, the naive covariance matrix Σ of (3) based on the expected scalar product $\langle . | . \rangle$ of the displacement vectors from the data $\mathbf{x}, \mathbf{y} \in \Delta_+^d$ to their respective mean $E\mathbf{x}, E\mathbf{y} \in \Delta_+^d$, needs to account for the fact that $\mathbf{x}, \mathbf{y} \in \Delta_+^d$ are affine points in (homogeneous) barycentric coordinates.

In the projective viewpoint, the notion of displacement vector between two equivalence classes $[\mathbf{x}]_+$, $[\mathbf{y}]_+$ does not exist. What is meaningful is to consider the pair $([\mathbf{x}]_+, [\mathbf{y}]_+)$ as a projective line passing between these two points, which corresponds to the vector plane $\text{span}(\mathbf{x}, \mathbf{y})$ in the ambient space $\mathbb{R}^{d+1}$. The (average) relative orientation between the planes $\text{span}(\mathbf{x}, E\mathbf{x})$ and $\text{span}(\mathbf{y}, E\mathbf{y})$ can serve as a basis upon which one can define a notion of covariance and correlation between points. Grassmann's exterior (wedge) product $\wedge$ is the key fundamental algebraic tool which allows to synthetically construct lines from pairs of points and to decompose the orientations of a pair of planes into components. These result in bivectors, $\mathbf{x} \wedge (E\mathbf{x})$ and $\mathbf{y} \wedge (E\mathbf{y})$, which interprets geometrically as oriented parallelograms. Their components and scalar product serve as analogues upon which one can construct notions of distance, and covariance matrix for CoDa.

The closely related projective and affine viewpoints on CoDa thus naturally suggest two related approaches to defining distance/divergence and covariance/correlation on CoDa. As readers may not be familiar with projective geometry, we separate the projective and affine approaches in two different papers, which can be read independently. The present article only uses notions of affine geometry and is thus conceptually simpler than Faugeras (2024a), which is based on the wedge product. The purpose of this paper is thus to define notions of divergence and variance matrix on the whole CoDa space, based on the affine viewpoint, using barycentric coordinates.

1.4 Outline

In Section 2, we give 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 2.4) for the displacement vector between two points expressed in barycentric coordinates, w.r.t. a frame of affinely independent points. Section 3 elaborates 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.

This decomposition allows to introduce in Section 4 the first main object of the paper, which is a family of divergences on the whole CoDa space (thus allowing for zeroes): the $\alpha-$barycentric divergences. We study their properties and generalize to the infinite dimensional case, viz. for general positive measures. The case $\alpha = 2$ has a nice geometric interpretation in terms of areas of triangles on the simplex.

Section 5 gives a first statistical application by defining various notions of empirical Fréchet means/medians for a sample cloud of CoDa points, based on these barycentric divergences. Simulations illustrate and allow to compare the variants obtained. The case of the Fréchet mean for the $\alpha = 2$ barycentric divergence appears particularly interesting, and reduces to the centroid (arithmetic mean) for $2-$ parts compositions.

Section 6 defines corresponding isotropic Laplace-Gaussian distributions based on the barycentric divergences. Introducing weighted versions of the barycentric divergences allows to generalize further and to define anisotropic generalized Laplace-Gaussian distributions. This gives analogues of the multivariate Laplace-Gaussian distribution on the whole CoDa space, with parameters for the location, the overall dispersion and the relative directions of variation.

Eventually, we introduce in Section 7 the second major object of the paper: a notion of variation matrix for CoDa based on an averaged "scalar product" of displacement vectors, as in the Euclidean vector case, but now with the key basic formula taking into account the affine nature of the Coda points expressed in barycentric coordinates. More precisely, we define affine notions of covariance and correlation matrices for a pair of random CoDa, and of variance matrix for the study of the intra-dispersion of a single random CoDa among its parts. The barycentric variance matrix is a log-free analogue of the classical log-variation matrices based on the variance of log-ratio transformed CoDa of the literature, in its ability to measure proportionality of CoDa components. Its properties are studied and illustrated by basic simulations. We conclude in Section 8.

2 A primer on affine geometry and barycentric coordinates

We briefly recall the basics of affine geometry and of barycentric coordinate systems. These are the prerequisites to understand CoDa as affine points.

2.1 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.[1] 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 2.1 (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*

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

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 $\mathcal{A}$. 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

[1] Yet, see the forthcoming Section 2.2

the vector basis $(\mathbf{e}_1, \dots, \mathbf{e}_d)$, viz.

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

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

$$(\beta_1 - \alpha_1, \dots, \beta_d - \alpha_d). \tag{5}$$

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).

2.2 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 2.2 (Affine independence). *A set $\{A_1, \dots, A_N\}$ of N points in a d-dimensional affine space, $d \geq 2$, is said to be affine independent if the $N-1$ vectors $\overrightarrow{A_1A_k}$, $k = 2, \dots, N$, are linearly independent*[2]*. A simplex of affine independent 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 2.3 (Barycentric coordinates). *Let $\{A_0, \dots, A_d\}$ be $d+1$ affinely independent points in a d-dimensional affine space $\mathcal{A}$. Let P be a given point. There are scalars $p_0, \dots, p_d$, with $\sum_{i=0}^{d} p_i \neq 0$, such that, for all points Q,*

$$(\sum_{i=0}^{d} p_i)\overrightarrow{QP} = \sum_{i=0}^{d} p_i \overrightarrow{QA_i}. \tag{6}$$

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

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

Proof. Since $\mathcal{A}$ is d-dimensional and $\{A_0, \ldots, A_d\}$ are affinely independent, there exist unique scalars α_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 (6) 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 (6) as

$$P = Q + \overrightarrow{QP} = Q + \sum_{i=0}^{d} \frac{p_i}{\sum_{j=0}^{d} p_j} \overrightarrow{QA_i}, \tag{7}$$

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

$$P = \sum_{i=0}^{d} p_i A_i, \quad \text{with } \sum_{i=0}^{d} p_i = 1, \tag{8}$$

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 (λ_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 (8). 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.

2.3 Formula of the displacement vector from barycentric coordinates of points

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

Lemma 2.4. *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

$$\mathbf{v} = \frac{\sum_{i<j} det \begin{vmatrix} m_i & n_i \\ m_j & n_j \end{vmatrix} \overrightarrow{A_i A_j}}{\sum_i m_i \sum_j n_j}. \tag{9}$$

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

$$
\begin{aligned}
\mathbf{v} = \overrightarrow{MN} = -M + N &= -\frac{\sum_i m_i A_i}{\sum_i m_i} + \frac{\sum_j n_j A_j}{\sum_j n_j} \\
&= \frac{\sum_i \sum_j n_j m_i (A_j - A_i)}{\sum_i m_i \sum_j n_j} \\
&= \frac{\sum_{i<j} n_j m_i (A_j - A_i) + \sum_{i>j} n_j m_i (A_j - A_i)}{\sum_i m_i \sum_j n_j} \\
&= \frac{\sum_{i<j} n_j m_i (A_j - A_i) + \sum_{i<j} n_i m_j (A_i - A_j)}{\sum_i m_i \sum_j n_j} \qquad (10) \\
&= \frac{\sum_{i<j} (m_i n_j - n_i m_j)(-A_i + A_j)}{\sum_i m_i \sum_j n_j} \\
&= \frac{\sum_{i<j} \det \begin{vmatrix} m_i & n_i \\ m_j & n_j \end{vmatrix} \overrightarrow{A_i A_j}}{\sum_i m_i \sum_j n_j}.
\end{aligned}
$$

where (10) 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.

3 CoDa as an affine point in barycentric coordinates

The quick reminder on affine geometry of the previous Section 2, especially the homogeneous character of barycentric coordinates, justify the claims of Section 1.2 and make it clear why CoDa elements, identified as elements of the simplex Δ_+^d, are to be seen as affine *points in barycentric coordinates*, and not as vectors: identifying each component i of a composition with an affine point A_i, it is reasonable to assume that $\mathcal{F} = \{A_i, i = 0, \ldots, d\}$ are affinely independent, since components relate to different entities and are thus not related to each others. Then, a (simplex representative of) CoDa $\mathbf{x} \in \Delta_+^d$ corresponds to the affine point

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

in normalized barycentric coordinates, as in (8).

3.1 CoDa operations as affine operations on points

In particular, the Amalgamation, Subcomposition and Partition operations on CoDa of 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 Δ_+^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 $k+1$ parts of a composition and closing the selected subvector to obtain a subcomposition in Δ_+^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

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

cuts the $(d+1)-$parts at position $0 \le k \le d$ 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 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 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

$$P = t_0 C_0 + t_1 C_1, \quad t_0 + t_1 = 1. \tag{11}$$

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 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 (11) is the statement of Properties 2.10 and 2.11 of the partitioning operation of 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.

3.2 Displacement vectors of CoDa points

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 (9) in Lemma 2.4, as

$$\overrightarrow{\mathbf{xy}} = \sum_{i<j} \det \begin{vmatrix} x_i & y_i \\ x_j & y_j \end{vmatrix} \overrightarrow{A_i A_j} \tag{12}$$

and makes it clear why the displacement between two CoDa points should not write as a difference (5) of their coordinates in a frame of reference (see the forthcoming Remark 2). Formula (12) 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 $\overrightarrow{A_i A_j}$ of two different points A_i, A_j, $i \neq j$, of the affine frame $\mathcal{F}$.

Remark 1 (Displacement coordinates as weighted ratios)**.** *Let*

$$v_{ij} := det \begin{vmatrix} x_i & y_i \\ x_j & y_j \end{vmatrix}, \quad 0 \leq i \neq j \leq d, \tag{13}$$

be the determinantal coefficient of the component of the displacement in the $\overrightarrow{A_i A_j}$ *direction in formula (12). It writes, for* $x_i \neq 0$, $y_i \neq 0$, *as*

$$v_{ij} = x_i y_j - y_i x_j = x_i y_i \left(\frac{y_j}{y_i} - \frac{x_j}{x_i} \right). \tag{14}$$

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*

$$[x_0 : \ldots : x_i : \ldots : x_d]_+ = [x_0/x_i : \ldots : 1 : \ldots : x_d/x_i]_+$$

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*

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

Thus, if both $[\mathbf{x}]_+$ *and* $[\mathbf{y}]_+$ *belong to* $\mathbb{U}_i$*, equation (14 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* $x_i y_i$ *given by the product of their* i*th (simplex) coordinate. In other words, the displacement for positive CoDa elements decomposes as a weighted difference, not of their coordinate components* (x_j)*, but of their ratios. The weight* $x_i y_i$ *translates the relative importance of components* i *of* $\mathbf{x}$ *and* $\mathbf{y}$*.*

If one sets conventionally $0/0 := 1$*, then formula (14) becomes true for all (non-negative) Coda elements* $[\mathbf{x}]_+, [\mathbf{y}]_+ \in \mathbb{P}^d_+$*. This interpretation of coordinate displacements as a weighted sum of components ratios is important from the subcompositional coherence point of view: only ratios of CoDa components are subcompositionally coherent, see, e.g., Greenacre (2021)*

Remark 2 (Displacement vector in the reference frame $(A_0; \overrightarrow{A_0A_1}, \ldots, \overrightarrow{A_0A_d})$)**.** *Formulas (12) and (9) expresses the displacement vectors in terms of the* $d(d+1)/2$ *vectors* $\overrightarrow{A_iA_j}$*. 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* $\mathbf{e}_i = \overrightarrow{A_0A_i}$*,* $i = 1, \ldots, d$*. Then, either by setting* $Q = A_0$ *in (6) and computing the difference of points or decomposing* $\overrightarrow{A_iA_j} = \overrightarrow{A_iA_0} + \overrightarrow{A_0A_j}$ *in (12), one can write the displacement vector of* $\overrightarrow{\mathbf{xy}}$*, for simplex-normalized* $\mathbf{x}, \mathbf{y} \in \Delta^d_+$*, w.r.t. the reference frame* $(A_0; \overrightarrow{A_0A_1}, \ldots, \overrightarrow{A_0A_d})$*, as*

$$\overrightarrow{\mathbf{xy}} = \sum_{j=1}^{d} \left(\sum_{0 \le i \ne j \le d} v_{ij} \right) \mathbf{e}_j = \sum_{j=1}^{d} (y_j - x_j)\mathbf{e}_j. \tag{15}$$

This gives the usual formula (5) of the displacement vector between two points expressed in $d+1$ *barycentric coordinates in terms of the difference of* d *independent coordinates. (Note the absence of the zero coordinates* x_0 *and* y_0*). However, the formula is not symmetric w.r.t coordinate components as one component serves as origin and thus play a distinguished role.*

Note that if one directly applies formula (6), then, obviously, for any point Q*,*

$$\overrightarrow{\mathbf{xy}} = \overrightarrow{Q\mathbf{y}} - \overrightarrow{Q\mathbf{x}} = \sum_{i=0}^{d} (y_i - x_i)\, \overrightarrow{QA_i}.$$

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. In addition, it introduces the extrinsic element Q, whose choice is arbitrary. This gives another reason why analysis of CoDa elements based on the usual coordinate difference is clearly wrong.

4 Barycentric divergence on the CoDa space

4.1 Motivation and definition

Lemma 2.4, and its specialization via formula (12) to CoDa represented on the simplex Δ^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 component $i = 0, \ldots, d$ of a composition identifies with a point A_i in a $d-$dimensional affine space. However, components 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 components, identified with the displacements $\overrightarrow{A_iA_j}$, $i \neq j$ in formula (12), *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 (12) 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 (2024a), 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 ℓ_α norm, leads to the following definition:

Definition 4.1. *Let $[\mathbf{x}]_+, [\mathbf{y}]_+ \in \mathbb{P}^d_+$ be two CoDa elements, and α be a real number. Then, the α-barycentric divergence is defined, for $1 \leq \alpha < \infty$, as*

$$d_\alpha([\mathbf{x}]_+, [\mathbf{y}]_+) := \frac{\left(\sum_{i<j} \left| det \begin{vmatrix} x_i & y_i \\ x_j & y_j \end{vmatrix} \right|^\alpha \right)^{1/\alpha}}{||\mathbf{x}||_1 ||\mathbf{y}||_1}, \tag{16}$$

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 4.2. *i)* d_α *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* $(x_{\sigma_0}, \ldots, x_{\sigma_1}, x_{\sigma_d})$. *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*

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

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 (16) 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

$$d_\alpha([\mathbf{x}]_+, [\mathbf{y}]_+) = 0 \Leftrightarrow x_i y_j = x_j y_i \text{ for all } i \neq j \in \{0, 1, \ldots, d\}. \tag{17}$$

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, (17) 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.

vi) Both numerator and denominator in (16) remain the same if some zeroes are added to the components of $\mathbf{x}$ and $\mathbf{y}$.

□

4.2 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 4.3. *The 2−barycentric divergence writes as*

$$d_2([\mathbf{x}]_+, [\mathbf{y}]_+) = \frac{\sqrt{||\mathbf{x}||_2^2 ||\mathbf{y}||_2^2 - \langle \mathbf{x}|\mathbf{y} \rangle^2}}{||\mathbf{x}||_1 ||\mathbf{y}||_1} \tag{18}$$

$$= \frac{||\mathbf{x}||_2 ||\mathbf{y}||_2}{||\mathbf{x}||_1 ||\mathbf{y}||_1} \sin \theta_{\mathbf{xy}}, \tag{19}$$

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 (18). Together with $0 \le \langle\mathbf{x}|\mathbf{y}\rangle = ||\mathbf{x}||_2||\mathbf{y}||_2\cos\theta_{\mathbf{xy}}$, with $\theta_{\mathbf{xy}} \in [0, \pi/2]$, it yields formula (19). □

Remark 3. *i) Since* $||\mathbf{x}||_2 \le ||\mathbf{x}||_1$*, formula (19) implies that*

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

which gives, for the case of the 2-divergence, another proof of the upper bound in Theorem 4.2 iv).

ii) Formula (19) involves the the sine of the (acute) angle between rays $[\mathbf{x}]_+$ *and* $[\mathbf{y}]_+$*. Since* $||\mathbf{x}||_2 \le ||\mathbf{x}||_1 \le \sqrt{d+1}||\mathbf{x}||_2$*, one has that*

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

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

The dissymmetry between the $||.||_2$ *and* $||.||_1$ *norms in the fraction in (19) suggests that we (should) eliminate this ratio-of-norms coefficient by replacing in the definition (19) 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 (2024a), 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 (18) 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 4.2, and more importantly, suggests the more abstract approach of Faugeras (2024a).

4.3 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 (2024b)), Definition 4.1 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 μ be a measure dominating P and Q, (e.g., $\mu = (P+Q)/2$). By Radon-Nikodym's theorem, P, Q have densities $f = \frac{dQ}{d\mu}$, $g = \frac{dQ}{s\mu}$. One can then define in such general case the following symmetric divergence:

Definition 4.4. *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 and possibly disjoint supports[3].

5 Fréchet means based on the barycentric divergences

5.1 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

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

approach of Fréchet (1948) (See also Faugeras (2023) Section 7). Indeed, given a CoDa sample $[\mathbf{x}^1]_+, \dots, [\mathbf{x}^n]_+ \in \mathbb{P}^d_+$, for any Coda point $[\mathbf{m}]_+ \in \mathbb{P}^d_+$, the following functional

$$\mathcal{F}_{\alpha,\beta}([\mathbf{m}]_+) := \sum_{i=1}^{n} d_\alpha^\beta([\mathbf{m}]_+, [\mathbf{x}^i]_+), \tag{20}$$

with $1 \leq \alpha \leq \infty$, $\beta > 0$, gives a measure of the outlyingness[4] (i.e. is a depth function) of the point $[\mathbf{m}]_+$ w.r.t. the data points $[\mathbf{x}^1]_+, \dots, [\mathbf{x}^n]_+$. Minimizing 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 5.1. *An empirical (α, β)-barycentric Fréchet mean is defined as a minimizer of (20). It is simply called an α-barycentric Fréchet mean for $\beta = \alpha$.*

Among possible choices for β, one can consider three interesting cases, i) $\beta = 2$, ii) $\beta = 1$ and iii) $\beta = \alpha$. Case i) gives the the well-known Fréchet mean, but for the different α-divergences of equation (16). 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 α-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 (16), it somehow appears as a natural choice.

The following theorem easily ensues.

Theorem 5.2. *i) For $1 \leq \alpha \leq \infty$, $\beta > 0$, a (α, β)-barycentric Fréchet mean always exists.*

ii) For $\beta \geq \alpha \geq 1$ and $\alpha = \infty, \beta \geq 1$, stationary points of (20) are (α, β)-barycentric Fréchet means.

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

Proof. i) Restricting to simplex representatives in Δ^d_+, the functional writes (20) 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 (20) is convex (for $\beta \geq \alpha$ and $\alpha = \infty, \beta \geq 1$), so that local minima are global ones.

iii) For $\infty > \beta \geq \alpha > 1$, (20) is strictly convex.

□

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

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.

Remark 4. *i) The barycentric means/medians are based on the minimization of the Fréchet functional (20) over the whole CoDa space $\mathbb{P}_+^d$, (equivalently, Δ_+^d). One can also consider several versions of so-called medoids (see e.g. Kaufman and Peter J. Rousseeuw (1990)) where the minimization of (20 is restricted to the finite set of data points $[\mathbf{x}^1]_+, \dots, [\mathbf{x}^n]_+$. 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 4 and 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 (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 β-powers of the α-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 6.1 of Generalized Gaussian distribution based on the barycentric divergence (16) 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 (20) 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 (20), calculated at the mean/median point* $[\mathbf{m}]_+$*, by the median. In particular, taking* $\alpha = \beta = 1$ *gives the Mean Absolute Deviation, defined as*

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

where $Med([\mathbf{x}^1]_+, \ldots, [\mathbf{x}^n]_+)$ *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).*

5.2 Numerical experiments and comparison with the centroid

We illustrate and investigate the different kind of means based on α-barycentric divergence. In the following, we always choose $\beta = \alpha$ in Definition 5.1, except for $\alpha = \infty$, for which we set $\beta = 1$.

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 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 λ *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* ∞*-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 5.3. *For $d = 1$, the empirical Fréchet mean based on the $2-$divergence is the arithmetic mean.*

Proof. Assume $\mathbf{x}^1, \dots, \mathbf{x}^n$ is a simplex-normalized sample, so that $\mathbf{x}^i = (x_0^i, x_1^i) \in \Delta_+^1$, $i = 1, \dots, n$. 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 2, where the data is made of only two data points.

Example 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 1, 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.*

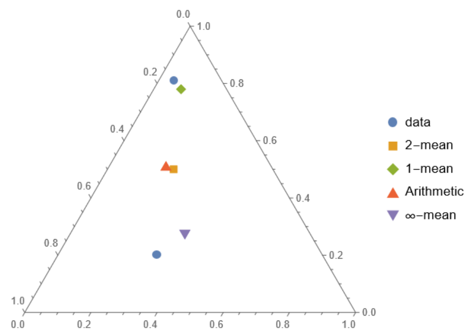

Figure 1: Comparison of the means for a toy example of 2 data points. Sample points (blue), Arithmetic mean (red upward triangle), Fréchet means for the α-divergence: $\alpha = 2$ (orange square), $\alpha = 1$ (green lozenge), $\alpha = \infty$ (downward violet triangle).

A more comprehensive picture of the influence of the α parameter on the location of the $\alpha-$barycentric mean is given in Example 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 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 2. 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* ℓ_1*-norm in the denominator of the* 2*-divergence by the* ℓ_2 *norm. Also shown on the Figure are the Fréchet mean based on the* 1*-divergence (green lozenge), the Fréchet mean for the* ∞*-divergence (violet downward triangle), and several barycentric Fréchet mean for the* α*-divergence (orange squares), for* α *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".*

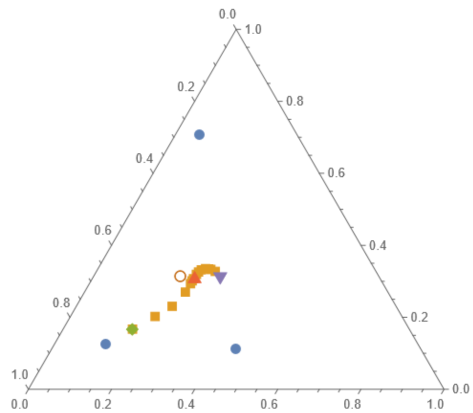

Figure 2: Comparison of the means for a toy example of 3 data points. Sample points (blue), Fréchet mean for the 2 divergence (red upward triangle), arithmetic mean (red circle), Fréchet mean for the 1-divergence (green lozenge), Fréchet mean for the ∞-divergence (violet downward triangle). Fréchet mean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

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 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 3 (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 3. It is noteworthy to remark that four 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. ∞-barycentric Fréchet mean (violet downward triangle). As the sample size grows, we notice empirically that the 1 and ∞ barycentric Fréchet mean seem to become indistinguishable, on this example, see Table 2.*

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
∞− 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
∞− barycentric	0.310345	0.344828	0.344828

Table 1: Comparison of the different kind of means of Example 4.

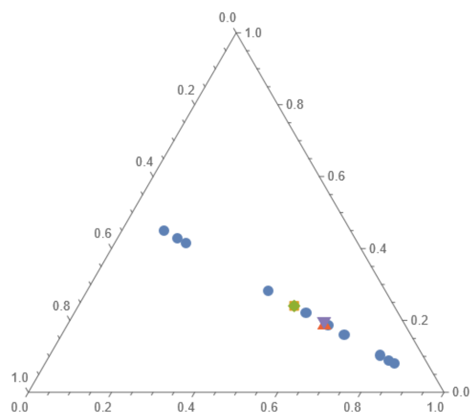

Figure 3: Comparison of the means for Example 4 of 10 data points on a line. Sample points (blue), Fréchet mean for the 2 divergence (orange square), arithmetic mean (green lozenge), Fréchet mean for the 1-divergence ((red upward triangle)), Fréchet mean for the ∞-divergence (violet downward triangle).

The previous example showed barycentric Fréchet means aligned with the line where the data sits. However, the picture in Figure 3 is somehow misleading and is due to the symmetry in the data of Example 4. In general, this is not the case, as illustrated on the following example.

Example 5. *As in Example 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 4. In this example, only the arithmetic mean and the ∞-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.*

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 2: Comparison of the different kind of means of Example 5.

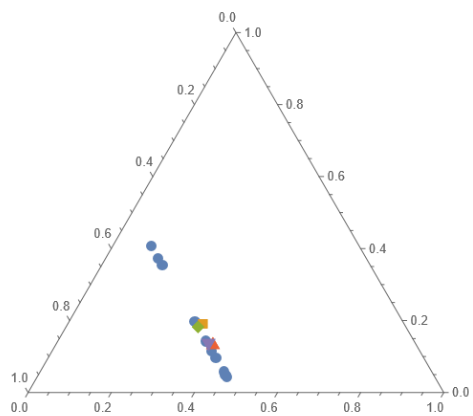

Figure 4: Comparison of the means for Example 5 of 10 data points on a line. Sample points (blue), Fréchet mean for the 2 divergence (orange square), arithmetic mean (green lozenge), Fréchet mean for the 1-divergence ((red upward triangle)), Fréchet mean for the ∞-divergence (violet downward triangle).

6 Generalized Laplace-Gaussian distribution based on the determinantal barycentric divergence

6.1 Isotropic Generalized Laplace-Gaussian distributions

As in Faugeras (2023), one can define a family of generalized Gaussian distributions, based on the family of divergences (16).

Definition 6.1. *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 distri-*

bution admits a density w.r.t. to the uniform measure ν *on* $\mathbb{P}^d_+$ *given by*

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

for $1 \le \alpha < \infty$*, 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 6.1 gives an analogue of the Gaussian, resp. Laplace, distribution when $\alpha = 2$, resp., $\alpha = 1$, with $[\mathbf{m}]_+$ a mean parameter and σ a dispersion parameter. Figure 5 show 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.

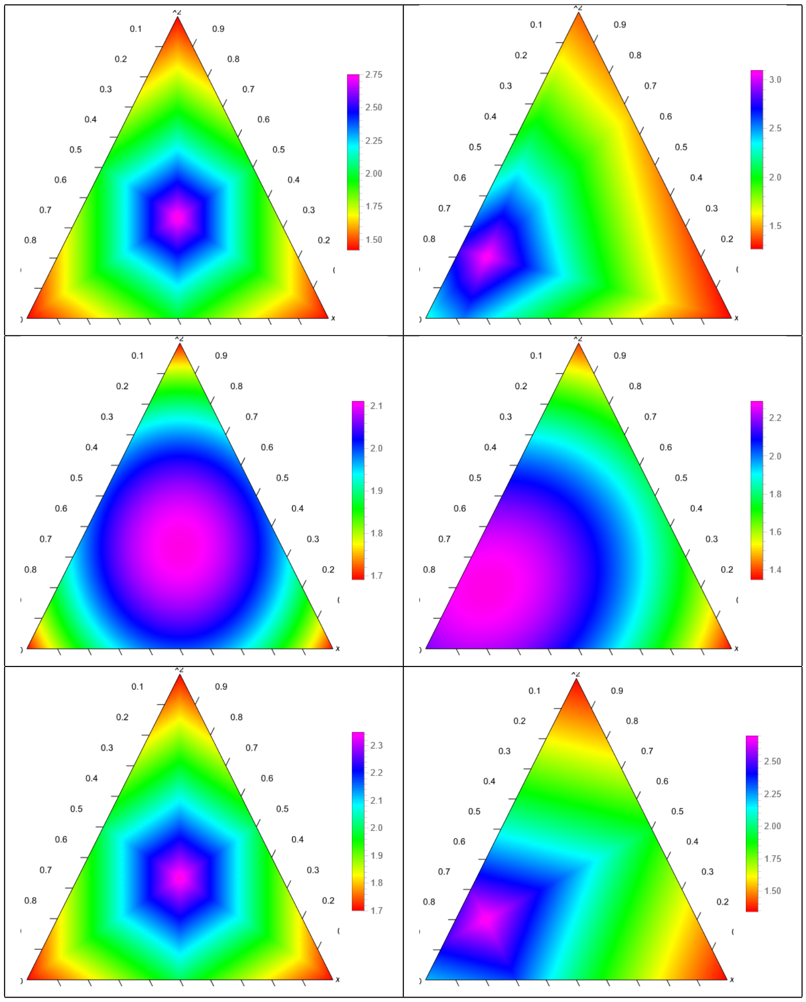

Figure 5: Generalised Barycentric Gaussian distributions with α-divergence. Left column: centered distribution with $[\mathbf{m}]_+ = [1:1:1]_+$. Right column: a non-centered distribution with $\mathbf{m} = (0.7, 0.1, 0.2)$. $\alpha = 1$ (up), $\alpha = 2$ (center), $\alpha = \infty$ (down). $\sigma = 1$.

6.2 Anisotropic Generalized Laplace-Gaussian distributions

As a further generalization, one can consider weighted versions of the barycentric divergences (16) and corresponding Gaussian-type distributions of Definition 6.1. 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 ith variable. A generalization of this principle, taking into account the correlation of the data, leads to the definition of the well-known Mahalanobis distance,

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

where Σ is the covariance matrix.

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

Definition 6.2. *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 α-divergence is defined, for $1 \leq \alpha < \infty$, as*

$$d_{\alpha,W}([\mathbf{x}]_+, [\mathbf{y}]_+) := \frac{\left(\sum_{i<j} w_{ij}^{-1} \left| det \begin{vmatrix} x_i & y_i \\ x_j & y_j \end{vmatrix} \right|^{\alpha} \right)^{1/\alpha}}{||\mathbf{x}||_1 ||\mathbf{y}||_1} \tag{22}$$

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 α-divergence satisfy all properties of Theorem 4.2, 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 (21), 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 α-divergence, while the σ 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 6.3. *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* ν *on* $\mathbb{P}_+^d$ *given by*

$$f_\alpha([\mathbf{x}]_+; [\mathbf{m}]_+, W, \sigma) := Z_\alpha^{-1}([\mathbf{m}]_+, W, \sigma) \exp\left(- \left(\frac{d_{\alpha,W}([\mathbf{x}]_+, [\mathbf{m}]_+)}{\sigma} \right)^\alpha \right) \quad (23)$$

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 6 illustrates Definition 6.3 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 5, 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 5.

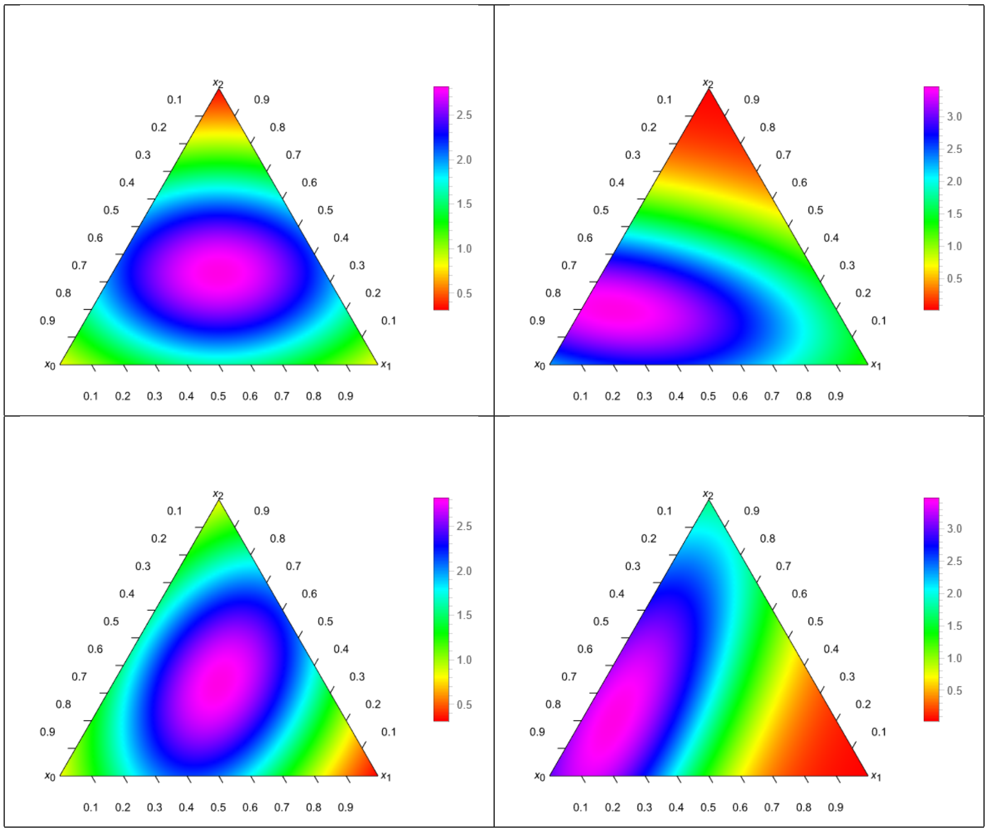

Figure 6: Generalised Weighted Barycentric Gaussian distributions with 2-divergence. $\sigma = 1$. Left column: centered distribution with $\mathbf{m} = (1 : 1 : 1)$. Right column: a non-centered distribution with $\mathbf{m} = (0.7, 0.1, 0.2)$. $(w_{01}, w_{02}, w_{12}) = (0.8, 0.1, 0.1)$ (up), $(w_{01}, w_{02}, w_{12}) = (0.1, 0.8, 0.1)$ (down).

We provide in Appendix 9 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 $\mathbb{P}_+^d$: α sets the general form of the balls in α-divergence, $[\mathbf{m}]_+$ is the location parameter, and (W, σ) 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 for further research.

Remark 5 (Weighted Hilbert projective distance)**.** *We remark that the idea of weighting the components entering in the formula of the barycentric α-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 6.4. *Let $W \in \mathcal{W}_0$. The weighted Hilbert projective metric on $\mathbb{P}^d_{++}$, with weight matrix W, is defined as*

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

In turn, the corresponding Gaussian type distribution is defined analogously to Definition 6.3 and Definition 7.4 in Faugeras (2023), with $d_{H,W}$ replacing $d_{\alpha,W}$ in (23). It also gives risen to an anisotropic Generalised Gaussian type distribution, this time based on the weighted version of Hilbert's projective metric. Appendix 9 provides some simulations for illustration purposes.

7 Variance and covariance matrices

7.1 Definitions

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

$$[\boldsymbol{\mu}^{\mathbf{x}}]_+ = [\mu_0^{\mathbf{x}} : \ldots : \mu_d^{\mathbf{x}}]_+, \quad \text{resp. } [\boldsymbol{\mu}_{\mathbf{y}}]_+ = [\mu_0^{\mathbf{y}} : \ldots : \mu_d^{\mathbf{y}}]_+.$$

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 (α, β)-barycentric Fréchet mean of Definition 5.1, 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 3, 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 (9) 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,

$$\mathbf{v_x} := \overrightarrow{[\boldsymbol{\mu}^{\mathbf{x}}]_+ \, [\mathbf{x}]_+} = \frac{\sum_{i<j} \det \begin{vmatrix} \mu_i^{\mathbf{x}} & x_i \\ \mu_j^{\mathbf{x}} & x_j \end{vmatrix} \overrightarrow{A_i A_j}}{||\boldsymbol{\mu}^{\mathbf{x}}||_1 ||\mathbf{x}||_1},$$

$$\mathbf{v_y} := \overrightarrow{[\boldsymbol{\mu}^{\mathbf{y}}]_+ \, [\mathbf{y}]_+} = \frac{\sum_{i<j} \det \begin{vmatrix} \mu_i^{\mathbf{y}} & y_i \\ \mu_j^{\mathbf{y}} & y_j \end{vmatrix} \overrightarrow{A_i A_j}}{||\boldsymbol{\mu}^{\mathbf{y}}||_1 ||\mathbf{y}||_1}.$$

A term-by-term product of the $\overrightarrow{A_i A_j}$ component of $\mathbf{v_x}$ and $\mathbf{v_y}$, viz.

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

gives a measure of the (random) covariation of the displacement vectors $\mathbf{v_x}$ and $\mathbf{v_y}$ in the same direction $\overrightarrow{A_i A_j}$. 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 component i to j. This leads to the following definition of the covariance matrix of the random CoDa points $[\mathbf{x}]_+$ and $[\mathbf{y}]_+$:

Definition 7.1 (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 $d+1$*

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

where the (i,j) component is set as

$$Cov([\mathbf{x}]_+, [\mathbf{y}]_+)_{i,j} := E\left(\frac{det \begin{vmatrix} \mu_i^{\mathbf{x}} & x_i \\ \mu_j^{\mathbf{x}} & x_j \end{vmatrix} \times det \begin{vmatrix} \mu_i^{\mathbf{y}} & y_i \\ \mu_j^{\mathbf{y}} & y_j \end{vmatrix}}{||\boldsymbol{\mu}^{\mathbf{x}}||_1 ||\mathbf{x}||_1 ||\boldsymbol{\mu}^{\mathbf{y}}||_1 ||\mathbf{y}||_1}\right). \tag{24}$$

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 (24) 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 for a random CoDa $[\mathbf{x}]_+$:

Definition 7.2 (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)*

$$
\begin{aligned}
Var([\mathbf{x}]_+) &:= Cov([\mathbf{x}]_+, [\mathbf{x}]_+) \in \mathbb{R}^{(d+1)\times(d+1)} \\
&= \left(E\left(\frac{det^2 \begin{vmatrix} \mu_i^{\mathbf{x}} & x_i \\ \mu_j^{\mathbf{x}} & x_j \end{vmatrix}}{||\boldsymbol{\mu}^{\mathbf{x}}||_1^2 ||\mathbf{x}||_1^2} \right) \right)_{\substack{i=0,\cdots,d \\ j=0,\cdots,d}} .
\end{aligned} \tag{25}
$$

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* $Var([\mathbf{x}]_+)$, *simplifies as*

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

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,

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

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

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

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 7.1 and 7.2:

Definition 7.4 (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 $i \neq j$,

$$\rho([\mathbf{x}]_+, [\mathbf{y}]_+)_{i,j} := \frac{Cov([\mathbf{x}]_+, [\mathbf{y}]_+)_{i,j}}{\sqrt{Var([\mathbf{x}]_+)_{i,j}\, Var([\mathbf{y}]_+)_{i,j}}},$$

$$= \frac{E\left(\frac{det\begin{vmatrix}\mu_i^{\mathbf{x}} & x_i\\ \mu_j^{\mathbf{x}} & x_j\end{vmatrix}\times det\begin{vmatrix}\mu_i^{\mathbf{y}} & y_i\\ \mu_j^{\mathbf{y}} & y_j\end{vmatrix}}{||\boldsymbol{\mu}^{\mathbf{x}}||_1||\mathbf{x}||_1||\boldsymbol{\mu}^{\mathbf{y}}||_1||\mathbf{y}||_1}\right)}{\sqrt{E\left(\frac{det^2\begin{vmatrix}\mu_i^{\mathbf{x}} & x_i\\ \mu_j^{\mathbf{x}} & x_j\end{vmatrix}}{||\boldsymbol{\mu}^{\mathbf{x}}||_1^2||\mathbf{x}||_1^2}\right)}\sqrt{E\left(\frac{det^2\begin{vmatrix}\mu_i^{\mathbf{y}} & y_i\\ \mu_j^{\mathbf{y}} & y_j\end{vmatrix}}{||\boldsymbol{\mu}^{\mathbf{x}}||_1^2||\mathbf{x}||_1^2}\right)}}$$

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

In definition 7.4, 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 7.5 (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 6. *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 7.1 and 7.2 are well-defined. By linearity of $||.||_1$ and multilinearity of the determinant, (24) 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 $\alpha, \beta, \gamma, \delta > 0$[5]. Hence, Definitions 7.1 and 7.2 are well-defined on $\mathbb{P}_+^d$.*

Definition 7.4, is undefined when $Var([\mathbf{x}]_+)_{i,j} = 0$ or $Var([\mathbf{y}]_+)_{i,j} = 0$ (and the discussion is similar for Definition 7.5). 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 7.5 coincides with Definition 7.4. The difference lies in the fact that Definition 7.5 is scale-invariant only for non-random rescaling, whereas Definition 7.4 is scale-invariant for deterministic and random rescaling alike.

[5]Note that α, β may be random.

7.2 Discussion and properties of barycentric variation matrices

The component i, j of the Variance matrix (25) measures the proportionality of parts i, j, as shown in the next Proposition.

Proposition 7.6. *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 components: if both components x_i and x_j are proportional, there is no variation in the $(A_i A_j)$ 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 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.

It is worth stressing some of its advantageous features. First, being log-free, the proposed variation matrix (25) 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 abovementioned 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. To the contrary, the proposed variation matrix (25) 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 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 ϕ which puts the log-ratio variance in relation to the size of the single variances involved. Here, since the Total Variance of Definition 7.3 interprets as the average displacement of $[\mathbf{x}]_+$ w.r.t. to its centroid mean, it makes sense to scale the components $\mathrm{Var}([\mathbf{x}]_+)_{i,j}$ of the Variance matrix (25) by the (scalar) total variance, i.e. to set the Normalised Variance matrix

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

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 (Variance matrix with two identical components)**.** *We take the same distribution as in Example 4, Figure 3, but with $n = 10000$ sample points. The empirical variance matrix is*

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

This empirically confirms the result of Proposition 7.6: since the data (blue points) sits on the straight line $x_1 = x_2$ in the triangle, there is no variation in the direction A_1A_2. The empirical variance $var([\mathbf{x}]_+)_{1,2}$ computed on the data is exactly zero as x_1 and x_2 carry the same proportional information.

Example 7 (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 Δ^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

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

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

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

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

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 7.6.*

7.3 Discussion and properties of barycentric correlation matrices

The barycentric covariance and correlation matrices of Definitions 7.1, 7.4 and 7.5 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 7.7 (Properties of Covariance, Correlation)**.** *i) Boundedness:*

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

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 $\mathcal{X}_{ij} := \begin{vmatrix} \mu_i^{\mathbf{x}} & x_i \\ \mu_j^{\mathbf{x}} & x_j \end{vmatrix}$, $\mathcal{Y}_{ij} := \begin{vmatrix} \mu_i^{\mathbf{y}} & y_i \\ \mu_j^{\mathbf{y}} & xy_j \end{vmatrix}$.

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

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

and yields the result for ρ. 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.

8 Conclusion

"Who ever uses barycentric coordinates ?" once asked the famous french mathematician J.A. Dieudonné[6]. Well, one would be tempted to reply that one goal of this paper was to demonstrate that CoDa is a natural domain of application of barycentric coordinates. Thinking geometrically about Coda as points in an affine space in barycentric coordinates gives a direct and unified way to deal with CoDa, with or without zeroes alike.

One key element of our approach is a decomposition formula for the displacement between two CoDa elements in terms of the displacements of the different pairs of the basis frame parts. This allows to define a family of barycentric divergences on the whole CoDa simplex space. In turn, these novel barycentric

[6]quoted by Pedoe (1970).

divergences enables to build essential statistical constructs like Fréchet means and their variants, Gaussian-type distributions accounting for anisotropic dispersion, and eventually log-free variance-covariance matrices set up as averaged "scalar product" of the displacement between Coda and their mean points.

In retrospect, such an affine viewpoint for the simplex seems natural, (after all, a ternary plot is really a plot in barycentric coordinates and we have shown that classical subcompositional operations are related to barycenters), but appears neglected in the CoDa literature. By providing for the necessary mathematical background on affine geometry and barycenters, our ambition is to supply essential tools for further statistical analysis of CoDa from such an affine viewpoint, free of the positivity constraint induced by the log transforms in classical log-ratio analysis. It is remarkable that the projective viewpoint aided with the exterior product gives strikingly similar divergences and covariances matrices concepts. Such a related approach will be treated in a separate companion paper (Faugeras (2024a)).

9 Appendix: Supplementary simulations

9.1 Anisotropic Generalised Barycentric Gaussian distributions with $\alpha = 1, \infty$.

For illustration and comparison purposes, we present in Figures 7 and 8 a sample of density plots of the weighted barycentric Gaussian distributions (see Definition 6.3), based on the W-weighted barycentric α-divergence, for $\alpha = 1, \infty$ and varying shape W and location $[\mathbf{m}]_+$ parameters.

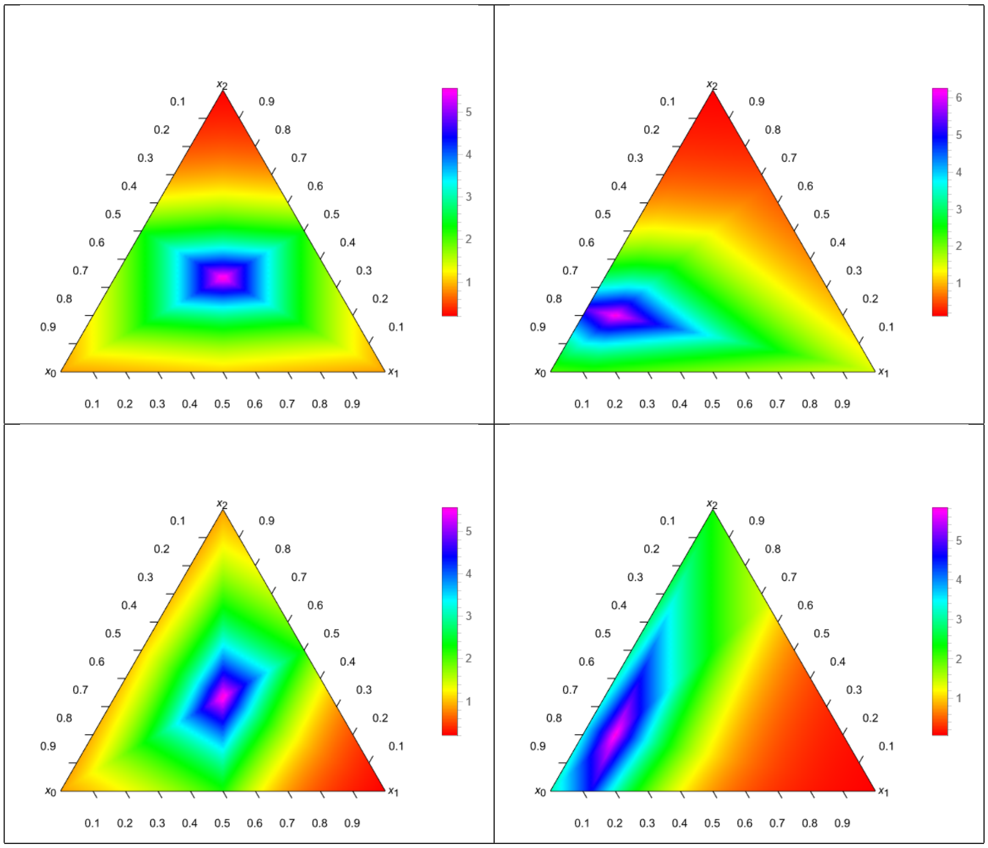

Figure 7: Generalised Weighted Barycentric Gaussian distributions with $\alpha = 1$-divergence. Left column: centered distribution with $[\mathbf{m}]_+ = [1 : 1 : 1]_+$. Right column: a non-centered distribution with $\mathbf{m} = (0.7, 0.1, 0.2)$. $(w_{01}, w_{02}, w_{12}) = (0.8, 0.1, 0.1)$ (up), $(w_{01}, w_{02}, w_{12}) = (0.1, 0.8, 0.1)$ (down). $\sigma = 2$.

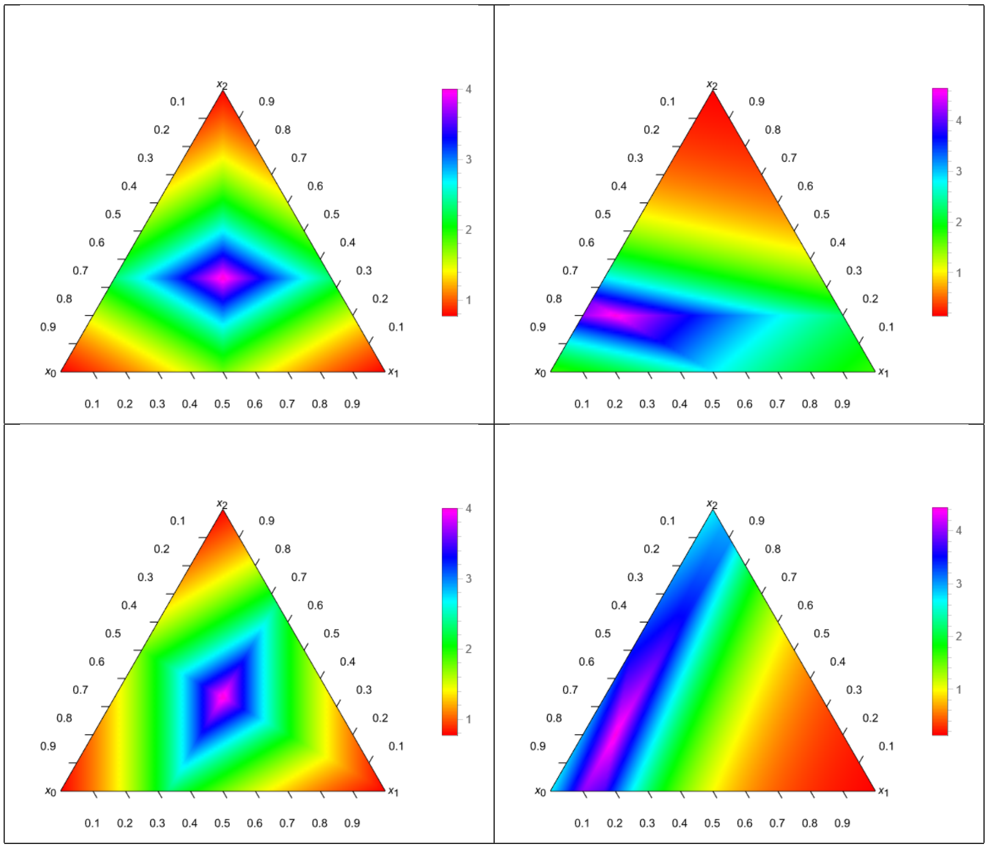

Figure 8: Generalised Weighted Barycentric Gaussian distributions with $\alpha = \infty$-divergence. Left column: centered distribution with $[\mathbf{m}]_+ = [1 : 1 : 1]_+$. Right column: a non-centered distribution with $\mathbf{m} = (0.7, 0.1, 0.2)$. $(w_{01}, w_{02}, w_{12}) = (0.8, 0.1, 0.1)$ (up), $(w_{01}, w_{02}, w_{12}) = (0.1, 0.8, 0.1)$ (down). $\sigma = 2$.

9.2 Anisotropic Generalised Hilbert-Gaussian distributions

For illustration purposes, we present in Figure 9 density plots of the weighted Hilbert Gaussian distributions, based on the (square of) the $W-$weighted Hilbert Projective metric, defined in Remark 5.

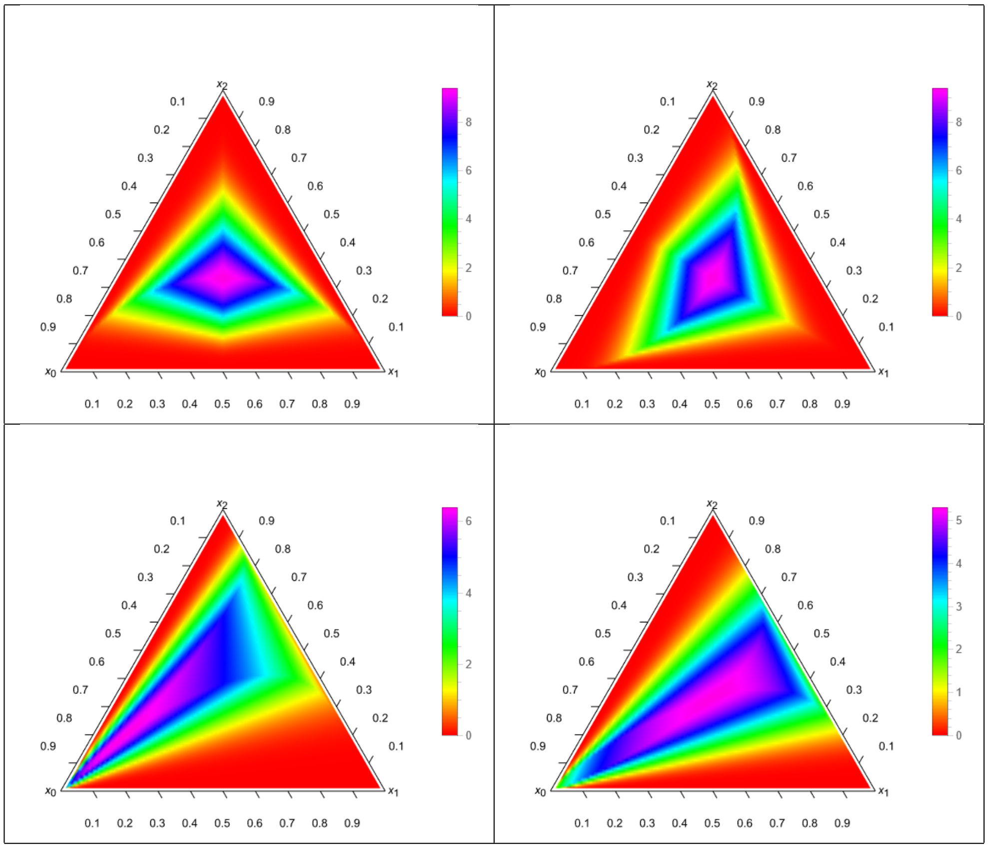

Figure 9: Generalised Weighted Hilbert-Gaussian distributions based on the W-weighted Hilbert projective metric. $[\mathbf{m}]_+ = [1:1:1]_+$, $(w_{01}, w_{02}, w_{12}) = (0.8, 0.1, 0.1)$ (upper left), $[\mathbf{m}]_+ = [1:1:1]_+$, $(w_{01}, w_{02}, w_{12}) = (0.1, 0.8, 0.1)$ (upper right), $\mathbf{m} = (0.7, 0.1, 0.2)$, $(w_{01}, w_{02}, w_{12}) = (0.4, 0.5, 0.1)$ (lower left), $[\mathbf{m}]_+ = [1:1:1]_+$, $(w_{01}, w_{02}, w_{12}) = (0.4, 0.5, 0.1)$ (lower right). $\sigma = 100$. $\alpha = 2$.

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. Chapman & Hall, London, pp. xvi+416.

ISBN: 0-412-28060-4. DOI: 10.1007/978-94-009-4109-0. URL: https://doi.org/10.1007/978-94-009-4109-0.

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*. Curran Associates Inc.: Red Hook, NY, USA.

Boogaart, K. Gerald van den and Raimon Tolosana-Delgado (2013). *Analyzing compositional data with R*. Use R! Springer, Heidelberg, pp. xvi+258. ISBN: 978-3-642-36808-0. DOI: 10.1007/978-3-642-36809-7. URL: https://doi.org/10.1007/978-3-642-36809-7.

Chayes, Felix (1971). *Ratio correlation: a manual for students of petrology and geochemistry*. University of Chicago Press.

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. John Wiley & Sons, Ltd., Chichester, pp. xvi+330. ISBN: 978-0-470-74991-3. DOI: 10.1002/9780470977811. URL: https://doi.org/10.1002/9780470977811.

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://www.tse-fr.eu/fr/publications/invitation-intrinsic-compositional-data-analysis-using-projective-geometry-and-hilberts-metric.

Faugeras, Olivier P. (July 2024a). “Log-Free Distance and Covariance matrix for Compositional Data II: The projective/exterior product approach”. Manuscript in preparation.

Faugeras, Olivier P. (Jan. 2024b). “The Stick-Breaking and Ordering Representation of Compositional Data: Copulas and Regression models”. TSE Working Paper, n° 24-1500. 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 (1948). “Les éléments aléatoires de nature quelconque dans un espace distancié”. In: *Annales de l’institut Henri Poincaré*. Vol. 10. 4, pp. 215–310.

Gallier, Jean (2011). *Geometric methods and applications*. Second. Vol. 38. Texts in Applied Mathematics. For computer science and engineering. Springer, New York, pp. xxviii+680. ISBN: 978-1-4419-9960-3. DOI: 10.1007/978-1-4419-9961-0. URL: https://doi.org/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.

Greenacre, Michael (2018). *Compositional data analysis in practice*. CRC press.

Greenacre, Michael (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.

Kaufman, Leonard and Peter J. Rousseeuw (1990). "Partitioning Around Medoids (Program PAM)". In: *Finding Groups in Data*. John Wiley & Sons, Ltd. Chap. 2, pp. 68–125. ISBN: 9780470316801. DOI: https://doi.org/10.1002/9780470316801.ch2. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470316801.ch2. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470316801.ch2.

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 zeros in compositional data sets". In: *Statistical Modelling* 15(2), pp. 134–158. DOI: 10.1177/1471082X14535524. eprint: https://doi.org/10.1177/1471082X14535524. URL: https://doi.org/10.1177/1471082X14535524.

Möbius, August Ferdinand (1827). *Der barycentrische Calcul, ein Hülfsmittel zur analytischen Behandlung der Geometrie (etc.)* Barth.

Pawlowsky-Glahn, Vera, Juan José Egozcue, and Raimon Tolosana-Delgado (2015). *Modeling and analysis of compositional data*. Statistics in Practice. John Wiley & Sons, Ltd., Chichester, 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.

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.

Shafarevich, Igor R. and Alexey O. Remizov (2013). *Linear algebra and geometry*. Translated from the 2009 Russian original by David Kramer and

Lena Nekludova. Springer, Heidelberg, pp. xxii+526. ISBN: 978-3-642-30993-9. DOI: 10.1007/978-3-642-30994-6. URL: https://doi.org/10.1007/978-3-642-30994-6.

Simovici, Dan A (2021). *CLUSTERING: Theoretical and Practical Aspects*. World Scientific.

Wang, David and Joseph Eppstein (2006). "Fast approximation of centrality". In: *Graph algorithms and applications* 5(5), p. 39.