WORKING PAPERS

N° 1496

February 2024

"An invitation to intrinsic compositional data analysis using projective geometry and Hilbert's metric"

Olivier Faugeras

Toulouse School of Economics

AN INVITATION TO INTRINSIC COMPOSITIONAL DATA ANALYSIS USING PROJECTIVE GEOMETRY AND HILBERT'S METRIC

BY OLIVIER P. FAUGERAS[1,a],

[1] *Toulouse School of Economics, University of Toulouse Capitole, France* , [a]*olivier.faugeras@tse-fr.eu*

We propose to study Compositional Data (CoDa) from the projective geometry viewpoint. Indeed, CoDa, as equivalence classes of proportional vectors, corresponds to projective points in a projective space, and thus can be studied using the tools, language and framework of projective geometry. Combined with the partial order structure induced by the non-negativity of CoDa, the projective viewpoint highlights the inherent geometrical and structural properties CoDa, irrespective of a particular representation and an arbitrarily chosen coordinate system. This intrinsic approach helps to clarify the relationships and offers much needed geometric insight between other competing coordinate-based approaches, such as Aitchison's log-ratio, Watson's spherical, or plain affine representations in the simplex. Our first objective is thus to give a tutorial on projective geometry geared towards compositional data analysis.

In addition, owing to the projective or ordering structures of CoDa, the positive CoDa space can be endowed with an intrinsic metric, Hilbert's projective metric, which is independent of any metrization of any ambient space. Such non-smooth metric is well-suited with the principles of compositional analysis (in particular, subcompositional coherence) and is compatible with both Aitchison's vector space geometry in log coordinates and the straight affine geometry of the simplex. In view of statistical applications, a smooth and strictly convex approximation of Hilbert's metric is constructed and is shown to share most properties of the original metric.

Our second objective is then to establish the firsts steps of such an intrinsic statistical analysis of CoDa, based on Hilbert's metric and the projective viewpoint. To that regards, we show how Hilbert metric and its smooth surrogate allow to build extrinsic and intrinsic measures of location and spread, in particular Fréchet means and variance, how to construct analogues of the Gaussian/Laplace distribution and how to perform nonparametric regression. All three examples are supplemented by numerical simulations. We close by drawing some perspectives for further research, inviting for new directions for CoDa analysis based on the intrinsic projective viewpoint.

CONTENTS

MSC2020 subject classifications: Primary 62H05; secondary 51A05, 62G08.

Keywords and phrases: Compositional Data, Projective geometry, Hilbert's projective metric, Intrinsic statistical analysis, Fréchet mean, Nonparametric regression, Gaussian distribution.

1. Introduction.

1.1. *Aitchison's classical approach to Compositional Data.* Compositional data (CoDa) analysis deals with statistical analysis of multivariate data $\mathbf{x} = (x_0, x_1, \ldots, x_d) \in \mathbb{R}^{d+1}$ where each x_i describe the amount of the ith component of a composition. It is understood that the magnitude of any component does not have any significance in itself, but only in its proportion relative to other components. Thus, CoDa is traditionally represented mathematically by requiring that $\mathbf{x}$ be i) non-negative, $\mathbf{x} \geq 0$ (as each component represent a physical amount) and ii) that $\mathbf{x}$ be subject to a constraint on the sum of its components $\sum_{i=0}^{d} x_i = \kappa$, where the sum κ is irrelevant for the analysis, so that only relative amounts matters. For $\kappa = 1$, this amounts to consider $\mathbf{x}$ as an element of the unit simplex

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

Aitchison's geometry, pioneered by [2], [3], only considers positive CoDa elements, i.e. $\mathbf{x}$ in the positive simplex

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

so that they can be studied through a variety of log-ratio transforms. These include:

- The additive log-ratio transform $\mathrm{alr} : \mathbb{R}_{++}^{d+1} \to \mathbb{R}^d$, defined by
$$\mathrm{alr}(\mathbf{x}) = (\ln(x_1/x_0), \ldots, \ln(x_d/x_0)) =: \mathrm{alr}_0(\mathbf{x}),$$
where the latter notation is used when one wants to specify w.r.t. which base coordinate x_0 the log-ratio is computed.
- The centered log-ratio transform $\mathrm{clr} : \mathbb{R}_{++}^{d+1} \to \mathbb{R}^{d+1}$, defined by
$$\mathrm{clr}(\mathbf{x}) = (\ln(x_0/g(\mathbf{x})), \ldots, \ln(x_d/g(\mathbf{x}))),$$
where $g(\mathbf{x}) = (x_0 x_1 \ldots x_d)^{1/(d+1))}$ is the geometric mean of $\mathbf{x}$.
- The isometric log-ratio transform $\mathrm{ilr} : \mathbb{R}_{++}^{d+1} \to \mathbb{R}^d$, which is constructed by taking an orthonormal basis $(\mathbf{1}/||\mathbf{1}||, \mathbf{v}_1, \ldots, \mathbf{v}_d)$ of $\mathbb{R}^{d+1}$, where $\mathbf{1} = (1, \ldots, 1)$, and projecting the $\mathrm{clr}(\mathbf{x})$ transformed vector onto the subspace spanned by $(\mathbf{v}_1, \ldots, \mathbf{v}_d)$. Setting $V = (\mathbf{v}_1 \ldots \mathbf{v}_d) \in \mathbb{R}^{(d+1)\times d}$ the so-called contrast matrix, ilr writes $\mathrm{ilr}_V(\mathbf{x}) = V^T \mathbf{ln}\, \boldsymbol{x} = V^T \mathrm{clr}(\mathbf{x})$, with $\mathbf{x}$ viewed as a column vector. The choice of V is left to the user and can be made using "balances" or "pivots" and a sequential binary partition methodology, which is deemed useful for interpretation.

These transforms allow to reduce the non-vectorial CoDa into log-coordinates representatives sitting in the usual Euclidean vector space $(\mathbb{R}^d, +, ., \langle .|. \rangle)$.[1] By pulling back this structure to the original positive simplex Δ^d_{++}, positive CoDa are endowed with a global Euclidean space structure $(\Delta^d_{++}, \oplus, \odot, \langle .|. \rangle_A)$, where perturbation $\oplus$ and powering $\odot$ are the vector space operations, and $\langle .|. \rangle_A$ is Aitchison's scalar product, see Sections 2.4 and 3 for more details. In addition to [3], book references include [44], [61], [46], [25].

1.2. *Motivation.* In spite of its successes, Aitchison's log-ratio Euclidean approach has some drawbacks. First, it can be confusing for the practitioner, which transform and framework is best suited for the data and analysis at hand. alr is sensitive to which part is taken in the denominator and is thus not permutation invariant. clr maps Δ^d_{++} to a hyperplane of $\mathbb{R}^{d+1}$, and leads to singular variance matrices. ilr is a family of transforms, which depends on an arbitrary choice of an orthonormal basis. Selection of a basis following the "balance tree" methodology, with its specific terminology and methods, is somehow complicated for the newcomer in the field.

Second, there exists alternatives approaches to CoDa: [60] and [61] mention four distinct frameworks (rplus, aplus, rcomp, acomp) to analyse compositional data. In particular, the raw or "do nothing"" approach, which simply considers CoDa points as simple vectors of $\mathbb{R}^{d+1}$ measuring absolute magnitudes is sometimes advocated for e.g. in archaeometry or geology (see [8], [7], [9], [10]). Notably, [59] points that, for some datasets, the empirical mean based on Aitchison's distance may lie outside the bulk of the dataset, and thus may be inappropriate. Another popular approach is based on spherical coordinates [63] via square root transforms, and more generally powering and Box-Cox transforms [57], [59]. These coordinate representations allow to deal with CoDa having zeros in their components, which is sometimes crucial in some applications. This has lead to debate on what is the "right" approach to compositional data analysis, with contradicting views on the matter, see e.g. [51], and more confusion for an outsider applied data scientist.

In both Aitchison's approach and its alternatives, the analysis is *extrinsic* as it is based on a special choice of representation and coordinates: the analysis hinges on the fact that the sample space can be embedded into an auxiliary ambient space (usually $\mathbb{R}^d$ or $\mathbb{R}^{d+1}$), endowed with a particular metric (usually the Euclidean distance).[2] For example, in the simplex representation, Aitchison's distance is the Euclidean distance of the clr or ilr coordinates (see e.g. [25] Chapter 3). Yet, for an appropriate statistical analysis it is essential to consider the inherent geometrical properties of the sample space of observations. This calls for an *intrinsic* approach to CoDa analysis based on the underlying structural properties of the observation space, irrespective of a particular representation and an arbitrarily chosen coordinate system.

Due to the scale-invariance requirement, it is now increasingly recognised ([4], [46], [25]) that the sample space of CoDa is made of equivalence classes of proportional vectors. Such equivalence classes are known in the geometry literature as *projective points* of the projective space $\mathbb{P}^d = \mathbb{P}(\mathbb{R}^{d+1})$, see the forthcoming Section 2 and Appendix A. Projective points $[\mathbf{x}]$ can be thought of, geometrically, as lines through the origin, or, algebraically, as one-dimensional sub-spaces of $\mathbb{R}^{d+1}$. Combined with the non-negativity/positivity requirement, this calls for a redefinition of CoDa points $[\mathbf{x}]_+$ as projective points in the non-negative orthant cone $\mathbb{R}^{d+1}_+$. Thus, CoDa are defined as elements of a new space $\mathbb{P}^d_+$ (to be defined precisely thereafter), represented geometrically as rays (i.e. directed half-lines) through the origin in the non-negative orthant cone. Hence, the simplex, resp. the non-negative sphere, is just one possible model/representation of these non-negative projective points.

[1] or $\mathbb{R}^{d+1}$ for the clr coordinates.

[2] with the chordal or spherical distance in the spherical representation.

1.3. *Purpose.* It is surprising that projective geometry has been largely ignored in the CoDa literature. A possible explanation is that projective geometry is not as well known/taught as classical vector geometry in the applied mathematics and statistics curriculum. Another possible reason is the multiplicity of approaches, models and representations used in projective geometry, some of which are sketched in Appendix A. We believe that such a projective geometry framework gives a natural setting[3] for studying CoDa as equivalence classes, as it leads to an intrinsic, coordinate-free approach. This often gives a simpler description of CoDa constructs and geometric insight, which is often obscured in algebraic approaches based on coordinates. Moreover, it is possible to define *intrinsically* a notion of distance, called Hilbert's projective distance, which will be shown to be well-suited for CoDa and its statistical analysis.

The goal of this paper is to bring this intrinsic projective geometry viewpoint into the statistical analysis of CoDa. Specifically our objectives are two-folds: i) we aim at bridging the gap between abstract projective geometry concepts and applied statistical analysis of CoDa. Therefore, some parts of the paper are expository, intended to explain the basics of projective geometry and Hilbert's metric to a non-specialist audience. We intent to show how it gives a simpler and advantageous description of CoDa and its structure. To ease understanding, we link those concepts to the ones familiar with CoDa, in particular with Aitchison's log-ratio analysis. ii) we also aim at establishing the first steps of such an intrinsic statistical analysis approach of positive CoDa, based on Hilbert's metric and the projective viewpoint. In particular, we show how statistical descriptive statistics; like mean, variance; distributions, like the Laplace/Gaussian; and estimation techniques like nonparametric regression can be built from such an intrinsic projective viewpoint with Hilbert's metric.

1.4. *Outline.* The detailed outline of the paper is as follows: In Section 2, we describe the projective and order/convex cone structure of CoDa. This leads to a redefinition of CoDa as "non-negative equivalence classes", i.e. as projective points in the non-negative orthant cone. We contrast such an unnormalised view of CoDa, as a full equivalence class, with classical normalised approaches, where a CoDa point is seen through its vector/affine coordinates in a special reference frame. The intrinsic projective approach allows to give a simple description of compositional maps on CoDa spaces, obtained from quotienting non-negative and non singular linear maps on vector spaces. In the case of positive CoDa elements, we show how the vector space structure and linear maps can be defined directly in an intrinsic manner, without reference to special vector coordinates or choice of a vector basis.

Section 3 deepens the intrinsic study of Aitchison's log (ratio) geometry on positive CoDa from the projective viewpoint, as initiated in the previous section. It shows how the $\mathbf{ln}$ map transforms isomorphically equivalence classes of positive CoDa vectors $[\mathbf{x}]_+$ into other equivalence classes $[\mathbf{ln}\,\boldsymbol{x}]_\sim$, which can be interpreted as parallel lines. This allows to give a much needed geometric description and insight of how the alr, clr, ilr coordinates are obtained.

In Section 4, we show how the positive CoDa space can be endowed with Hilbert's projective metric, based on the cross-ratio, in such way that it is compatible with this ordering and projective structure. We explain how such a metric can also be constructed from an order theory point of view, as pioneered by [12], which gives a simple explicit formula for Hilbert's distance. The important point is the intrinsic character of such a metric, as it only depends on the order structure of the CoDa space, and not on a metrization (e.g. in Euclidean distance) of an auxiliary ambient space. In Section 5 we study properties of this metric. It is shown that Hilbert's metric is compatible with Aitchison's vector space structure, is scale

[3] Information geometry is also another interesting viewpoint, see [24].

and permutation invariant, and is subcompositionally coherent, thus satisfy all principles of CoDa analysis. We derive alternative expressions w.r.t to some (pseudo)-metrics in real vector spaces, establishing some isometric embeddings/isomorphisms between the positive CoDa space with Hilbert's metric and real normed vector spaces, endowed with suitable metrics. This also shed a new light on the $\ln$ transforms of Aitchison's classical log-coordinates approach.

Being endowed with such a metric structure, one can propose statistical applications based on comparisons of CoDa points w.r.t. Hilbert metric. As those applications are often formulated as optimization problems, we first study the directional differentiability of Hilbert metric for CoDa and propose a smooth proxy whose properties are investigated. At last, Statistical applications per se are eventually studied in Section 7. We explain the difference between intrinsic and extrinsic descriptive statistics and propose to define as intrinsic measures of location and scatter Fréchet means/median and variance w.r.t. Hilbert's metric. We also introduce a Hilbert metric analogue of the Gaussian/Laplace distribution. In addition, Nadaraya-Watson type estimators based on Hilbert metric are also introduced to perform nonparametric regression with CoDa covariate and/or output variables. These applications are illustrated by numerical applications. We conclude in Section 8 and draw some perspectives for further research. Proofs are relegated to Appendix B.

1.5. *Author's contributions and related works.* In view of our first objective to provide a tutorial on metric projective geometry applied to CoDa, the manuscript gathers and synthesises from the projective CoDa viewpoint numerous results of various authors. In addition, the manuscript provides original contributions, which are summarised below:

- the formal redefinition of CoDa points (Definition 2.1) as non-negative projective points, combining the order and projective structures;
- the notion of compositional morphisms and compositional groups (Lemma 2.2);
- the intrinsic definition of linear compositional mappings on $(\mathbb{P}^d_{++}, \oplus, \odot)$ (Definition 2.3);
- the formal definition of the vector space $(\mathbb{R}^{d+1}/\sim, \tilde{+}, \tilde{\cdot})$ and the isomorphism with $(\mathbb{P}^d_{++}, \oplus, \odot)$ (Proposition 3.2);
- the geometrical interpretation of the alr, clr, ilr transforms (Sections 3.2, 3.3, 3.4);
- the normed vector space properties of Hilbert's metric (Propositions 5.1 and 5.2);
- the directional differentiability of Hilbert's metric w.r.t. to the m and e linear structures (Proposition 6.1 and 6.2);
- the properties of the smooth approximate Hilbert distance (Proposition 6.5);
- the definition and existence of the intrinsic Fréchet mean/median/variance (Definition 7.1, Lemma 7.2, Theorem 7.3;
- the use of Hilbert's smooth metric as a practical surrogate to compute the mean (as in Example 2);
- the generalised Hilbert Gaussian/Laplace distributions (Definition 7.4);
- the use of Hilbert's smooth metric in nonparametric Coda Regression (Example 3 in Section 7.3).

The application of Hilbert's cross-ratio distance for clustering on the simplex was proposed by [40]. Towards the end of completing our manuscript, we became aware of [41]. [41] also offers a nice overview of Hilbert's metric. Their objective is different (the study of non-linear embeddings), while we focus more on the projective/CoDa aspects. Also, we had the same idea as the authors of [41] of approximating Hilbert's distance with the well-known log-sum-exp trick, and so we were quite disappointed to not be able to include Definition 6.3, which we had developed independently, as an original contribution. We acknowledge their priority and improved our work using theirs (in particular, regarding strict convexity of the smooth Hilbert metric).

Notations. The following notations will be introduced throughout the text and are collected here for convenience of the reader.

- (Column) vectors in $\mathbb{R}^{d+1}$ are written in bold letters, $\mathbf{x} = (x_0, x_1, \dots, x_d) \in \mathbb{R}^{d+1}$, and operations on vector are interpreted component wise, i.e. $\mathbf{ln}\, \boldsymbol{x} = (\ln x_0, \dots, \ln x_d)$.
- $||.|| = ||.||_2$ denotes the usual norm (Euclidean/L^2), $||\mathbf{x}||_1 = \sum_{i=0}^{d} |x_i|$ the L^1 norm, and $||\mathbf{x}||_\infty = \max_i |x_i|$ the L^∞ norm.
- $\mathbb{R}_+ = \{x \in \mathbb{R}, x \geq 0\}$ stands for the non-negative part of $\mathbb{R}$, $\mathbb{R}_{++} = \{x \in \mathbb{R}, x > 0\}$ for the positive part.
- Δ_+^d the d-dimensional (unit or probability) simplex of $\mathbb{R}^{d+1}$, $\Delta_{++}^d = \mathring{\Delta}_+^d$ the positive simplex.
- $\mathbb{P}^d$ real projective space of dimension d induced by $\mathbb{R}^{d+1}$, $\mathbb{P}_+^d$ non-negative unnormalised CoDa vectors, $\mathbb{P}_{++}^d$ positive unnormalised CoDa vectors.
- $[\mathbf{x}] \in \mathbb{P}^d$, $[\mathbf{x}]_+ \in \mathbb{P}_+^d$ are projective, resp. CoDa, equivalence classes of $\mathbf{x}$.
- $[\mathbf{x}]_1 = \mathcal{C}(\mathbf{x})$ rescaled-to-unit-sum representative of $[\mathbf{x}]_+$.
- $\delta(.,.)$ Hilbert's projective metric, $d_H(.,.)$ Birkhoff's version of Hilbert metric.

2. The intrinsic projective nature of CoDa. In this section, we argue that CoDa should be considered as elements of a projective space, i.e. as an equivalence class, in order to encode the scale invariance of CoDa. This "unnormalised" intrinsic view is in contrast with the definitions traditionally encountered in the literature: in these "normalised" views, CoDa are given particular, extrinsic, coordinates representations as vectors in a subset of an Euclidean space.

2.1. *The ordering and projective structure of CoDa .* As explained in the introduction, a CoDa element, construed for the moment as a vector $\mathbf{x} = (x_0, \dots, x_d) \in \mathbb{R}_+^{d+1} \setminus \{\mathbf{0}\}$, has two main features: i) it has non-negative coordinates, i.e.

$$\boldsymbol{x} \geq \mathbf{0} \qquad \Longleftrightarrow \qquad x_i \geq 0, \quad i = 0, \dots, d, \tag{3}$$

and ii) $\mathbf{x}$ only carries relative information on the composition: if one defines the (collinearity/scaling invariance) equivalence relation $\equiv$ as

$$\mathbf{x} \equiv \mathbf{y} \Longleftrightarrow \exists \lambda \in \mathbb{R}^* \text{ s.t. } \mathbf{x} = \lambda \mathbf{y}, \tag{4}$$

two equivalent vectors, $\mathbf{x} \equiv \mathbf{y}$, carries the same compositional information and are thus identified. There are thus two fundamental structures underlying each CoDa element:

i) the partial order structure $\leq$ on $\mathbb{R}^{d+1}$, induced by the positive orthant convex cone $\mathbb{R}_+^{d+1}$, viz.

$$\boldsymbol{x} \leq \boldsymbol{y} \Longleftrightarrow \mathbf{y} - \mathbf{x} \in \mathbb{R}_+^{d+1},$$

in the sense that (3) only allows non-negative components;

ii) the real projective space structure of $\mathbb{P}^d$, obtained by quotienting $\mathbb{R}^{d+1} \setminus \{\mathbf{0}\}$ by the equivalence relation $\equiv$, i.e.

$$\mathbb{P}^d := {(\mathbb{R}^{d+1} \setminus \{\mathbf{0}\})}\big/_{\equiv} = \frac{\mathbb{R}^{d+1} \setminus \{\mathbf{0}\}}{\mathbb{R} \setminus \{0\}}.$$

Indeed, by denoting $[\mathbf{x}]$, the equivalence class of $\mathbf{x}$ for (4), viz.

$$[\mathbf{x}] := \{\lambda \mathbf{x}, \lambda \in \mathbb{R}^*\} \tag{5}$$

it is readily seen that condition (4) mandates CoDa elements to be regarded as equivalence classes (restricted to the positive orthant by condition (3)). From the geometrical viewpoint, $[\mathbf{x}]$ is a line in $\mathbb{R}^{d+1}$ through the origin $\mathbf{0}$, viz. the undirected direction span$(\mathbf{x})$. From the projective geometry viewpoint, $[\mathbf{x}]$ is a projective point and the set $\{(\lambda x_0, \ldots, \lambda x_d), \lambda \in \mathbb{R}^*\}$ are its homogeneous coordinates. See Appendix A for a description of projective spaces and their main representations.

Combining both structures, we thus define the set of CoDa elements as the "non-negative part"[4] of the projective space $\mathbb{P}^d$, hence the formal definition:

DEFINITION 2.1. A CoDa point is an equivalence class $[\mathbf{x}]_+$ of the quotient space,

$$\mathbb{P}^d_+ := \frac{\mathbb{R}^{d+1}_+ \setminus \{\mathbf{0}\}}{\mathbb{R}_{++}},$$

that is to say, a CoDa point is defined as the set

$$[\mathbf{x}]_+ := \{\lambda \mathbf{x}, \lambda > 0, \}, \tag{6}$$

for some $\mathbf{x} \geq \mathbf{0}$. $\mathbf{x} \in \mathbb{R}^{d+1}$ is a called a representative[5] of $[\mathbf{x}]_+ \in \mathbb{P}^d_+$.

REMARK 1. $[\mathbf{x}]_+$ is also the equivalence class for the positive scaling relation $\equiv_+$, defined on $\mathbb{R}^{d+1}$ by

$$\mathbf{x} \equiv_+ \mathbf{y} \Longleftrightarrow \exists \lambda > 0 \text{ s.t. } \mathbf{x} = \lambda \mathbf{y}.$$

So that one can also define $\mathbb{P}^d_+$ as the punctured non-negative orthant $\mathbb{R}^{d+1}_+ \setminus \{\mathbf{0}\}$, quotiented by such positive scaling relation, viz.

$$\mathbb{P}^d_+ = \frac{\mathbb{R}^{d+1}_+ \setminus \{\mathbf{0}\}}{\equiv_+} = \{[\mathbf{x}]_+ : \mathbf{x} \in \mathbb{R}^{d+1}_+ \setminus \{\mathbf{0}\}\}.$$

Geometrically speaking, a CoDa point $[\mathbf{x}]_+$ is thus interpreted as a non-negative direction, viz. a ray in the positive orthant cone, emanating through the origin, or equivalently, as the intersection of a line through the origin with the positive orthant $\mathbb{R}^{d+1}_+$. The set of *positive* CoDa points (i.e. with positive components) is defined similarly as

$$\mathbb{P}^d_{++} := \frac{\mathbb{R}^{d+1}_{++} \setminus \{\mathbf{0}\}}{\mathbb{R}_{++}} = \{[\mathbf{x}]_+ : \mathbf{x} \in \mathbb{R}^{d+1}_{++} \setminus \{\mathbf{0}\}\}.$$

2.2. *Affine representations in $\mathbb{R}^{d+1}$ of projective/CoDa points by normalised vectors.* We have thus considered the space of CoDa elements as the set of equivalence classes $[\mathbf{x}]_+$ in the non-negative orthant, i.e. a CoDa element is in fact the *whole* set $[\mathbf{x}]_+$ of *unnormalised* vectors $\mathbf{x}$ satisfying (6). It is essential, in order to avoid confusion, to clearly distinguish between a CoDa element $[\mathbf{x}]_+ \in \mathbb{P}^d_+$, thought as an equivalence class, and its (many) *vector representative* $\mathbf{x} \in \mathbb{R}^{d+1}_+$: a CoDa element has *homogeneous coordinates* $[x_0 : x_1 : \ldots : x_d]_+$, whereas (one of) its vector representative has (classical) vector coordinates $(x_0, \ldots, x_d)$, representing the absolute/raw size of the components of the composition. The (surjective) quotient map,

$$\begin{aligned} [.]_+ : \mathbb{R}^{d+1}_+ \setminus \{\mathbf{0}\} &\to \mathbb{P}^d_+ \\ \mathbf{x} &\mapsto [\mathbf{x}]_+, \end{aligned}$$

[4] or, following Appendix A, as the projectivization of the non-negative orthant cone.

[5] $\mathbb{P}^d_+$ has dimension d, hence the notation. Note, however, that the representatives $\mathbf{x}$ lie in $\mathbb{R}^{d+1}_+$, a space of dimension $d+1$.

associates to an absolute vector $\mathbf{x}$ its compositional/homogeneous part $[\mathbf{x}]_+$. In particular, as will be clear below, the representation of $[\mathbf{x}]_+$ by a vector $\mathbf{x}$ constrained in the simplex, is merely one, among many possible representations.

Indeed, equivalence classes, whether they are projective points $[\mathbf{x}]$ of (5), or CoDa point $[\mathbf{x}]_+$ of (6), being scalar multiples of a vector representative, admits several distinguished representations as a single point/vector in the Euclidean space $\mathbb{R}^{d+1}$. We will call these standardised representations *normalised* projective (resp. CoDa) points.

i) For projective points $[\mathbf{x}]$ which are not orthogonal to $\mathbf{1} = (1, \dots, 1)$, i.e. such that there exists a vector representative $\mathbf{x}$ with $\sum_{i=0}^d x_i \neq 0$, one can standardise the latter by the sum of its components: define the rescaled-to-unit-sum representative $[\mathbf{x}]_1 \in \mathbb{R}^{d+1}$ of $[\mathbf{x}] \in \mathbb{P}^d$ by

$$[\mathbf{x}]_1 := \frac{\mathbf{x}}{\langle \mathbf{1} | \mathbf{x} \rangle} = \frac{\mathbf{x}}{\sum_{i=0}^d x_i}. \tag{7}$$

Correspondingly, for CoDa elements $[\mathbf{x}]_+$, the conditions $\mathbf{x} \neq \mathbf{0}$ and $\boldsymbol{x} \geq \mathbf{0}$ guarantee that $\sum_{i=0}^d x_i \neq 0$, so that one can always represent a CoDa point by such a rescaled-sum representative. Note that since $\boldsymbol{x} \geq \mathbf{0}$, (7) also writes as

$$[\mathbf{x}]_1 = \frac{\mathbf{x}}{||\mathbf{x}||_1} = \mathcal{C}(\mathbf{x}),$$

where $||.||_1$ is the L^1-norm, and $\mathcal{C}$ is the closure operation of the CoDa literature. This yields an interpretation of the simplex representation (7) as the directional part in the polar decomposition $\mathbf{x} \mapsto (||\mathbf{x}||_1, [\mathbf{x}]_1)$ w.r.t. the L^1 norm of a vector $\mathbf{x}$ into a magnitude $||\mathbf{x}||_1$ and a direction/composition $[\mathbf{x}]_1$.[6]

Geometrically, let $\mathcal{H}_{\mathtt{sum}} = \{\mathbf{x} \in \mathbb{R}^{d+1} : \sum_{i=0}^d x_i = 1\}$ be the affine hyperplane in the Euclidean space $\mathbb{R}^{d+1}$, with normal vector $\mathbf{1} = (1, \dots, 1)$ and passing through, say, $(1, 0, \dots, 0)$. Then, a projective point $[\mathbf{x}] \in \mathbb{P}^d$, (resp. a CoDa point $[\mathbf{x}]_+ \in \mathbb{P}^d_+$), is represented in the Euclidean space $\mathbb{R}^{d+1}$ by the corresponding point on $\mathcal{H}_{\mathtt{sum}}$, intersected by the line $[\mathbf{x}]$, (resp. by the ray $[\mathbf{x}]_+$, see Figure 1. In other words, for CoDa $[\mathbf{x}]_+ \in \mathbb{P}^{\mathbf{d}}_+$, $[\mathbf{x}]_1$ corresponds to the traditional representation of CoDa as a vector element constrained in the simplex Δ^d_+ of (1), and (7) is the radial projection on it.

ii) The classical approach in projective geometry is to standardise by a ratio w.r.t. a component. For $\mathbf{x} = (x_0, \dots, x_d)$, if $x_0 \neq 0$, one can represent the projective point $[\mathbf{x}] \in \mathbb{P}^d$ (resp. $[\mathbf{x}]_+ \in \mathbb{P}^d_+$) by the point

$$\pi_0(\mathbf{x}) := \mathbf{x}/x_0 = (1, x_1/x_0, \dots, x_d/x_0) \tag{8}$$

of the Euclidean space $\mathbb{R}^{d+1}$. Geometrically, the Euclidean representative $\mathbf{x}/x_0$ is located on the affine hyperplane

$$\mathcal{H}_0 := \{\mathbf{x} \in \mathbb{R}^{d+1} : x_0 = 1\} \tag{9}$$

and corresponds to the point of intersection of the line $[\mathbf{x}]$ (resp. of the ray $[\mathbf{x}]_+$) with $\mathcal{H}_0$, see Figure 2. Forgetting the first constant 1 coordinate, one can identify $\pi_0(\mathbf{x})$ with a point of $\mathbb{R}^d$. If $x_0 = 0$, then one can no longer divide by x_0. Nonetheless, the projective point $[\mathbf{x}]$, (resp. $[\mathbf{x}]_+$) can be interpreted as a *point at infinity* in the direction of the vector $(0, x_1, \dots, x_d)$. This leads to the synthetic view of projective geometry, where $\mathbb{P}^d$ is envisioned as $\mathbb{R}^d$, "completed" with a set of points "at infinity", see Appendix A.

[6] This shows the ubiquity of CoDa for (non-negative) multivariate vectors.

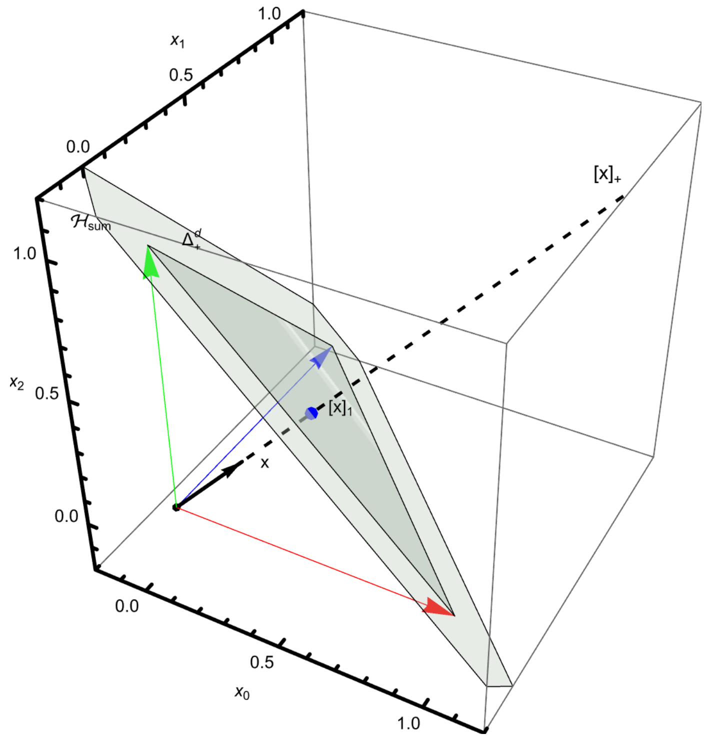

FIG 1. *Rescaled-to-sum representation of $\mathbb{P}_+^d$ in $\mathbb{R}^{d+1}$. Given an unnormalised vector $\mathbf{x} \in \mathbb{R}^{d+1}$ (black arrow), the CoDa $[\mathbf{x}]_+$ (black dashed ray) intersects the (light blue) hyperplane $\mathcal{H}_{sum}$ at a (blue) point $[\mathbf{x}]_1$, located on the simplex Δ_+^d (light blue triangle). Red, blue, green arrows: coordinate unit vectors of $\mathbb{R}^{d+1}$, (here, $\mathbb{R}^3$).*

Correspondingly, for CoDa elements $[\mathbf{x}]_+$, i.e. with the added constraint $\mathbf{x} \geq 0$, the CoDa elements s.t. $x_0 \neq 0$ can be represented by a point $\pi_0(\mathbf{x}) \in \mathcal{H}_0 \cap \mathbb{R}_+^{d+1}$. $\pi_0(\mathbf{x})$ somehow corresponds to the alr_0 coordinate, but without the log part of the transform. See also Section 3 for a more elaborate study of Aitchison's log-ratio transforms from the projective viewpoint.

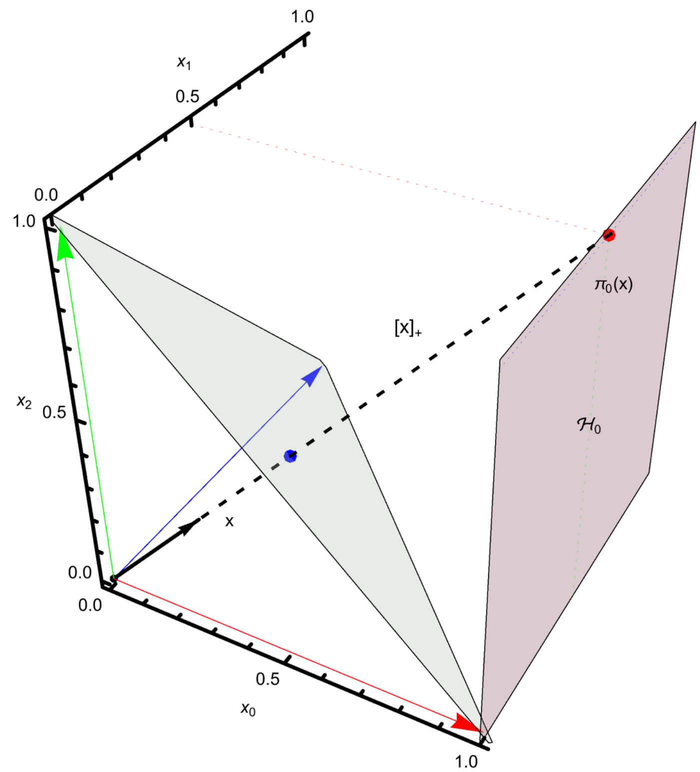

FIG 2. *Ratio representation of $\mathbb{P}_+^d$ in $\mathbb{R}^{d+1}$. The CoDa point $[\mathbf{x}]_+$ (black dashed half-line) is represented by the (red) point $\pi_0(\mathbf{x})$ on the affine hyperplane $\mathcal{H}_0 = \{\mathbf{x} : x_0 = 1\}$ (light red). The previous normalised representative $[\mathbf{x}]_1 \in \Delta_+^d$ on the simplex is shown as the blue point.*

iii) Other models of projective points and their CoDa analogues can be envisioned. Appendix A gives a short review of the most frequent occurring representations of projective points encountered in the literature. This review gives insight on the geometric nature of projective points and adds perspective on the recurring debate (see e.g. [51]) on the pros and cons of the possible approaches for dealing with CoDa: from the projective viewpoint, all these approaches are based on extrinsic representations, which are particular models of the same space $\mathbb{P}_+^d$, endowed with these two fundamental structures of a projective space and a partial order induced by a convex cone.

Figure 3 illustrates and summarizes the distinction between the unnormalised and normalised view of CoDa based on the two main representations i) and ii): CoDa elements $[\mathbf{x}]_+ \in \mathbb{P}_+^d$, seen as equivalence classes, possibly obtained by quotienting a raw/absolute vector $\mathbf{x} \in \mathbb{R}_+^{d+1}$, can be identified with a single vector/affine point $[\mathbf{x}]_1$ or $\pi_0(\mathbf{x})$. The maps $[\mathbf{x}]_+ \to [\mathbf{x}]_1$ and $[\mathbf{x}]_+ \to \pi_0(\mathbf{x})$ are bijective on the appropriate[7] domains.

[7] For π_0, one must restrict $[\mathbf{x}]_+$ to be s.t. $x_0 \neq 0$.

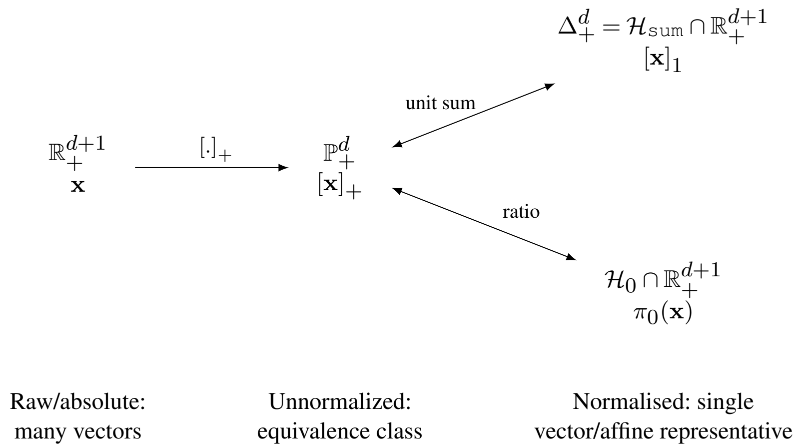

FIG 3. *Unnormalised (Projective) and Normalised representation (rescaled to unit sum and ratio) of CoDa.*

Let us make a few comments about the affine representations i) and ii) above. First, they both represent CoDa elements $[\mathbf{x}]_+$ by a normalised vector whose coordinates are ratios of the coordinates of vector representative $\mathbf{x}$ of $[\mathbf{x}]_+$. In particular, there is no need to take the $\log$ of these ratio coordinates to achieve such a representation of the whole $[\mathbf{x}]_+$ by a single normalised vector. The $\ln$ transform is only mandatory if one wants to give a global vector space structure to CoDa, as will be explained in Section 3, at the price of necessarily restricting the analysis to *positive* CoDa elements in $\mathbb{P}_{++}^d$. The use of log-free affine representations to deal with CoDa elements with possible zeros in their components, will be explored in a separate paper.

Second, it is clear that these affine representations are not canonical, in the sense that they depend on the choice of a particular hyperplane ($\mathcal{H}_0$ or $\mathcal{H}_{\texttt{sum}}$). Surely, the choice of $H_{\texttt{sum}}$ may appear more convenient, as it is more symmetric, and a single chart $[.]_1$ works for all CoDa elements. Yet, it remains a convention, viz. a choice among all possible hyperplanes. From the theoretical standpoint, one can consider the CoDa space as a manifold with its atlas composed of all its affine charts $\pi_0, \pi_1, \ldots, \pi_d$, where π_i is the ratio transform (8), but w.r.t. base component x_i.

In addition, such extrinsic view of CoDa, by identifying $[\mathbf{x}]_+$ with a vector/affine representative in the ambient space $\mathbb{R}^{d+1}$, construe CoDa as a $d+1$ dimensional object, whereas the true dimension of $\mathbb{P}_+^d$ is d. Surely, both $[\mathbf{x}]_1$ and $\pi_0(\mathbf{x})$ live on hyperplanes of dimension d, so in particular, if one forgets the first 1 coordinate in $\pi_0(\mathbf{x})$, one can simply identify such ratio representation with a vector of $\mathbb{R}^d$. But, for $[\mathbf{x}]_1$, the criticism is valid, even though $[\mathbf{x}]_1$ can in turn be identified with a d-dimensional coordinate representation by a change of basis to an orthonormal basis having $\mathbf{1}$ as element, and dropping the corresponding constant coordinate, as in the ilr transform.

Consequently, it appears that CoDa analysis based on the normalised vector/affine representatives i) and ii) or their log-transformed analogues do not seem ideal. It would thus be desirable to have a more intrinsic approach based on direct analysis at the level of the equivalence class $[\mathbf{x}]_+$.

2.3. *Compositional morphisms.* The intrinsic projective approach allows to give a simple description of the analogue of projective maps[8] for CoDa in $\mathbb{P}_+^d$, resp. $\mathbb{P}_{++}^d$, which we call

[8]also known as homographies.

compositional morphisms. The importance of such transformations stems from the fact that the main CoDa operations, like taking a subcomposition, amalgamation, permutation, can be described by such compositional morphisms. See Section 5.2 for more details.

From the discussion in Appendix A, a bijective linear endomorphism (viz. an automorphism) $A \in \mathbb{GL}(\mathbb{R}^{d+1})$ maps one-dimensional subspaces into one-dimensional subspaces and thus induces, by projectivization, a map, denoted[9] by $[A] : \mathbb{P}^d \to \mathbb{P}^d$, defined by setting

$$[A]([\mathbf{x}]) = [A(\mathbf{x})],$$

where the right-hand side is the equivalence class (5) of $A(\mathbf{x})$. Equivalently, $[A]$ can be viewed as the equivalence class $[A] = \{\lambda A, \lambda \in \mathbb{R}^*\}$ of non-zero scalar multiples of A. In other words, the quotient map $\mathbf{x} \to [\mathbf{x}]$, can be extended to a (covariant) functor, sending invertible linear maps $A : \mathbb{R}^{d+1} \to \mathbb{R}^{d+1}$ into projective maps $[A] : \mathbb{P}^d \to \mathbb{P}^d$, in a way compatible with the composition of maps. That is to say, $[AB] = [A][B]$, which can be equivalently illustrated by the following commutative diagram:

$$\begin{tikzcd}
\mathbf{x} \arrow[r, "A"] \arrow[d, "{[.]}"] & A\mathbf{x} \arrow[r, "B"] \arrow[d, "{[.]}"] & BA\mathbf{x} \arrow[d, "{[.]}"] \\
{[\mathbf{x}]} \arrow[r, "{[A]}"] & {[A\mathbf{x}]} \arrow[r, "{[B]}"] & {[BA\mathbf{x}]}
\end{tikzcd}$$

FIG 4. *The quotient map as a functor from linear spaces to projective spaces.*

Taking into account the non-negatitivity, resp. positivity of $\mathbb{P}^d_+$, resp. $\mathbb{P}^d_{++}$, one readily sees that the linear map A must also send the non-negative cone $\mathbb{R}^{d+1}_+$, resp. positive cone $\mathbb{R}^{d+1}_{++}$, into itself, in order to induce a mapping $[A]_+$ of $\mathbb{P}^d_+$, resp. $\mathbb{P}^d_{++}$. This translates in matrix terms, identifying A with its matrix $(a_{ij})_{0\le i,j\le d}$, by requiring that $a_{ij} \ge 0$, resp. $a_{ij} \ge 0$ and $\sum_j a_{ij} > 0$, $0 \le i,j \le d$, (the latter condition means that A has no rows full of zeros), as one readily checks that

$$A(\mathbb{R}^d_+) \subset \mathbb{R}^d_+ \Leftrightarrow a_{ij} \ge 0,$$

$$A(\mathbb{R}^d_{++}) \subset \mathbb{R}^d_{++} \Leftrightarrow a_{ij} \ge 0, \text{ and } \sum_{j=0}^{d} a_{ij} > 0.$$

As for projective spaces, one can thus extend the CoDa quotient map $\mathbf{x} \to [\mathbf{x}]_+$ in a functorial way, sending non-negative, resp. positive, bijective map to their positive scalar multiples:

$$A \to [A]_+ := \{\lambda A, \lambda > 0\}.$$

This leads to the definition of the natural analogue of the projective group for CoDa elements: set

$$\mathbb{PGL}^d_+ := \{[A]_+, A \in \mathbb{GL}(\mathbb{R}^{d+1}), A \ge \mathbf{0}\},$$

$$\mathbb{PGL}^d_{++} := \{[A]_+, A \in \mathbb{GL}(\mathbb{R}^{d+1}), A \ge \mathbf{0}, A\mathbf{1} > \mathbf{0}\}.$$

Then,

LEMMA 2.2. *$\mathbb{PGL}^d_+$, resp. $\mathbb{PGL}^d_{++}$, is a group for the composition, acting on $\mathbb{P}^d_+$, resp. $\mathbb{P}^d_{++}$, which we call the CoDa projective group, resp. positive CoDa projective group.*

[9] $[A]$ was denoted by $\mathbb{P}(A)$ in the Appendix.

For a general linear map $A : \mathbb{R}^{d+1} \to \mathbb{R}^{k+1}$, one can obtain similar results, under the proviso on $\ker(A)$ mentioned in Appendix A. This leads to a general definition of CoDa maps as non-negative (resp. positive) equivalence classes $[A]_+ : \mathbb{P}^d_+ \to \mathbb{P}^k_+$, via projectivization of the (restriction to $\mathbb{R}^{d+1} \setminus \ker(A)$) of non-negative (resp. positive) invertible linear operators. We omit the details.

2.4. *Intrinsic vector space structure of positive CoDa points $\mathbb{P}^d_{++}$ at the level of equivalence classes.* In the traditional simplex approach, positive CoDa elements, seen as elements of the positive simplex Δ^d_{++}, can be endowed with a global vector space structure: for $\mathbf{x}, \mathbf{y} \in \Delta^d_{++}$, [45] define the vector addition $\oplus$ (perturbation) and scalar multiplication $\odot$ (powering) as (see p.12):

$$\mathbf{x} \oplus \mathbf{y} := \mathcal{C}(\boldsymbol{x} \times \boldsymbol{y}) = [\boldsymbol{x} \times \boldsymbol{y}]_1, \tag{10}$$

$$\lambda \odot \mathbf{x} := \mathcal{C}(\boldsymbol{x}^\lambda) = [\boldsymbol{x}^\lambda]_1, \tag{11}$$

where $\boldsymbol{x} \times \boldsymbol{y} = (x_0 y_0, \dots, x_d y_d)$ is the coordinate-wise multiplication. These operations are based on the isomorphism of the $\exp : (\mathbb{R}, +, .) \to (\mathbb{R}_{++}, \times, \hat{})$:

$$e^{a+b} = e^a \times e^b, \quad e^{\lambda a} = (e^a)^\lambda.$$

These operations are thus the pull-back of the usual vector space operations $+, .$, for elements of Δ^d_{++} expressed in log-coordinates: for $\mathbf{x}, \mathbf{y} \in \Delta^d_{++}$,

$$\mathbf{x} \oplus \mathbf{y} := [\mathbf{exp}(\mathbf{ln}\, \boldsymbol{x} + \mathbf{ln}\, \boldsymbol{y})]_1,$$

$$\lambda \odot \mathbf{x} := [\mathbf{exp}(\boldsymbol{\lambda}.\, \mathbf{ln}\, \boldsymbol{x})]_1.$$

This definition is extrinsic, as it is based on the simplex representation $[.]_1$. Note that the rescaling/closure operation (7) is needed after each operation to get back to an element of Δ^d_{++}.

It is noteworthy to remark that this vector space structure of positive CoDa elements does not depend on the representation $[.]_1$ chosen. Indeed, from the projective viewpoint, the operations $\oplus, \odot$ can be directly expressed as operations on equivalence classes: for any representative $\boldsymbol{x} > \mathbf{0}$ and $\boldsymbol{y} > \mathbf{0}$ of $[\mathbf{x}]_+, [\mathbf{y}]_+ \in \mathbb{P}^d_{++}$ and $\lambda \in \mathbb{R}$, define,

$$[\mathbf{x}]_+ \oplus [\mathbf{y}]_+ := [\boldsymbol{x} \times \boldsymbol{y}]_+, \tag{12}$$

$$\lambda \odot [\mathbf{x}]_+ := [\mathbf{x}^\lambda]_+. \tag{13}$$

It is readily checked that these operations are well defined and do not depend on the choice of the representative $\mathbf{x}$ in the equivalence class $[\mathbf{x}]_+$: for $\nu, \mu > 0$, $\mu\mathbf{x} \in [\mathbf{x}]_+$, $\nu\mathbf{y} \in [\mathbf{y}]_+$, and

$$(\mu\mathbf{x}) \times (\nu\mathbf{y}) = \mu\nu.(\boldsymbol{x} \times \boldsymbol{y}) \in [\boldsymbol{x} \times \boldsymbol{y}]_+,$$

$$(\mu\mathbf{x})^\lambda = \mu^\lambda.\mathbf{x}^\lambda \in [\mathbf{x}^\lambda]_+.$$

So in particular, using the ratio standardisation (8), one also has that, for $x_0, y_0 \neq 0$,

$$[\mathbf{x}]_+ \oplus [\mathbf{y}]_+ := [\pi_0(\mathbf{x}) \times \pi_0(\mathbf{y})]_+,$$

$$\lambda \odot [\mathbf{x}]_+ := [\pi_0(\mathbf{x})^\lambda]_+$$

In other words, the ratio representative (8) of $[\mathbf{x}]_+ \oplus [\mathbf{y}]_+$ is simply

$$\pi_0([\mathbf{x}]_+ \oplus [\mathbf{y}]_+) = \pi_0(\mathbf{x}) \times \pi_0(\mathbf{y}),$$

and similarly for $\odot$. This confirms the intrinsic character of the $\oplus, \odot$ operations, as one could have defined them using ratio representatives instead of simplex ones (provided $x_0, y_0 \neq 0$ of course).

Summarizing, the $\oplus$ and $\odot$ operations, originally defined in the literature for unit-sum simplex representatives, can be lifted to the full equivalence class (6), and are intrinsic to $\mathbb{P}^d_{++}$. Here, no closure operation (7) is needed at such equivalence class level. A more detailed study of $(\mathbb{P}^d_{++}, \oplus, \odot)$ will be given in Section 3

2.5. *Linear compositional mappings for positive CoDa in* $(\mathbb{P}^d_{++}, \oplus, \odot)$. In the literature, linear applications on the positive CoDa space $(\mathbb{P}^d_{++}, \oplus, \odot)$ are defined extrinsically, via vector log coordinates of simplex representatives: given a contrast matrix $V \in \mathbb{R}^{(d+1)\times d}$, and $A : \mathbb{R}^d \to \mathbb{R}^d$ an endomorphism of $\mathbb{R}^d$, the action of A on $\mathbf{x} \in \Delta^d_{++}$ is given by defining an endomorphism $\psi_A : \Delta^d_{++} \to \Delta^d_{++}$ as

$$\psi_A(\mathbf{x}) := \mathrm{ilr}_V^{-1}(A\,\mathrm{ilr}_V(\mathbf{x})) = \mathcal{C}(\mathbf{exp}(VAV^T\mathbf{ln}\,\boldsymbol{x})).$$

Setting $A^\Delta := VAV^T$, the action of A on on $\mathbf{x} \in \Delta^d_{++}$ is often noted using the $\boxed{\cdot}$ symbol, as $A^\Delta \boxed{\cdot}\, \mathbf{x} := \psi_A(\mathbf{x})$. It is then shown that $A^\Delta \boxed{\cdot}\, \mathbf{x}$ does not depend on the (sub)-basis matrix V and can be written as $A^\Delta \boxed{\cdot}\, \mathbf{x} = \mathrm{clr}^{-1}(A^\Delta \mathrm{clr}(\mathbf{x}))$. The $(d+1)\times(d+1)$ matrix A^Δ is called the matrix of the endomorphism ψ_A, and it is shown that A^Δ satisfy the "zero-sum property", viz. $A^\Delta\mathbf{1} = (A^\Delta)^T\mathbf{1} = \mathbf{0}$. For a general linear map $A : \mathbb{R}^\ell \to \mathbb{R}^d$, the associated map $\psi_A : \Delta^\ell_{++} \to \Delta^d_{++}$ is defined similarly as $\psi_A(\mathbf{x}) = \mathrm{ilr}^{-1}_{V_d}(A\,\mathrm{ilr}_{V_\ell}(\mathbf{x}))$, where V_d, resp. V_ℓ are contrast matrices of $\mathbb{R}^d$, resp. $\mathbb{R}^\ell$.

As in Section 2.4, we show below that linear applications on $(\mathbb{P}^d_{++}, \oplus, \odot)$ can be defined intrinsically, directly at the level of equivalence classes, without reference to special coordinates or basis. Given a linear application $A : \mathbb{R}^{d+1} \to \mathbb{R}^{d+1}$, simply set $A^+ : \mathbb{P}^d_{++} \to \mathbb{P}^d_{++}$ as

$$A^+([\mathbf{x}]_+) := [\mathbf{exp}(A\,\mathbf{ln}\,\boldsymbol{x})]_+. \tag{14}$$

In other words, one follows the path illustrated in Figure 5, where ι denotes the selection operation of a vector representative (a reciprocal of the quotient map $\mathbf{x} \to [\mathbf{x}]_+$).

$$\begin{tikzcd}
{[\mathbf{x}]_+ \in \mathbb{P}^d_{++}} \arrow[r, "\iota"] & \mathbf{x} \in \mathbb{R}^{d+1}_{++} \arrow[r, "\mathbf{ln}"] & \mathbf{ln}\,\boldsymbol{x} \in \mathbb{R}^{d+1} \arrow[d, bend left, "A"] \\
A^+([\mathbf{x}]_+) & \mathbf{exp}(A\,\mathbf{ln}\,\boldsymbol{x}) \arrow[l, "{[.]_+}"'] & A\,\mathbf{ln}\,\boldsymbol{x} \arrow[l, "\boldsymbol{exp}"']
\end{tikzcd}$$

FIG 5. *Linear compositional applications on* $(\mathbb{P}^d_{++}, \oplus, \odot)$.

In order to be well-defined, A^+ must be independent of the selection operation ι: if $\mathbf{y} = \lambda\mathbf{x}$, $\lambda > 0$, is another vector representative of $[\mathbf{x}]_+$, then by definition (14) and linearity of A,

$$\begin{aligned}
A^+([\mathbf{y}]_+) &= [\mathbf{exp}(A\,\mathbf{ln}\,\boldsymbol{y})]_+ = [\mathbf{exp}(A(\mathbf{ln}\,\boldsymbol{x} + (\ln\lambda)\mathbf{1}))]_+ \\
&= [\mathbf{exp}(A(\mathbf{ln}\,\boldsymbol{x}) + (\ln\lambda)A(\mathbf{1}))]_+ \\
&= [\mathbf{e}^{A(\mathbf{ln}\,\boldsymbol{x})} \times \mathbf{e}^{(\ln\lambda)A(\mathbf{1})}]_+
\end{aligned}$$

The latter is equal to $[\mathbf{e}^{A(\mathbf{ln}\,\boldsymbol{x})}]_+$ if and only if the component-wise product $\mathbf{e}^{A(\mathbf{ln}\,\boldsymbol{x})} \times \mathbf{e}^{(\ln\lambda)A(\mathbf{1})}$ of two vectors reduces to a product of a vector by a scalar, for all $\lambda > 0$. This is the case iff the vector $\mathbf{e}^{(\ln\lambda)A(\mathbf{1}))}$ has identical coordinates, viz. is a scalar multiple of $\mathbf{1}$, which happens iff $\mathbf{1}$ is an eigenvector of A. One must then restrict the set of possible $A : \mathbb{R}^{d+1} \to \mathbb{R}^{d+1}$ to those s.t. $A\mathbf{1} = \alpha\mathbf{1}$, for some $\alpha \in \mathbb{R}$. The latter property interprets as a constant row sum property on the matrix representation of A. It is also easy to check that A^+ is indeed linear on $(\mathbb{P}^d_{++}, \oplus, \odot)$.

Next, we enquire which endomorphisms $A : \mathbb{R}^{d+1} \to \mathbb{R}^{d+1}$ s.t. $A\mathbf{1} = \alpha\mathbf{1}$, for some $\alpha \in \mathbb{R}$, give rise to the same $A^+ : \mathbb{P}^d_{++} \to \mathbb{P}^d_{++}$. Decompose $\mathbb{R}^{d+1}$ into the direct sum $\mathbb{R}^{d+1} = \text{span}(\mathbf{1}) + \mathcal{V}$, so that any $\mathbf{z} \in \mathbb{R}^{d+1}$ writes $\mathbf{z} = \zeta_0\mathbf{1} + \mathbf{v}$, with unique $\zeta_0 \in \mathbb{R}, \mathbf{v} \in \mathcal{V}$. Thus,

$$A\mathbf{z} = \zeta_0 A\mathbf{1} + A(\mathbf{v}) = \zeta_0\alpha\mathbf{1} + A(\mathbf{v})$$

In turn, $A(\mathbf{v})$ writes $A(\mathbf{v}) = a_0\mathbf{1} + \mathbf{v}'$, for some unique $a_0 \in \mathbb{R}, \mathbf{v}' \in \mathcal{V}$. Thus,

$$\mathbf{e}^{A\mathbf{z}} = \mathbf{e}^{(\zeta_0\alpha + a_0)\mathbf{1} + \mathbf{v}'} = \mathbf{e}^{(\zeta_0\alpha + a_0)\mathbf{1}} \times \mathbf{e}^{\mathbf{v}'} = e^{(\zeta_0\alpha + a_0)}.\mathbf{e}^{\mathbf{v}'},$$

where the last operation . is multiplication of the vector $\mathbf{e}^{\mathbf{v}'}$ by a scalar. By quotienting, $[\mathbf{e}^{A\mathbf{z}}]_+ = [\mathbf{e}^{\mathbf{v}'}]_+$ and it is seen that the part $a_0\mathbf{1}$ of $A(\mathbf{v})$ in span$(\mathbf{1})$ does not influence the value of $[\mathbf{e}^{A\mathbf{z}}]_+$, i.e. leads to the same $A^+([\mathbf{x}]_+)$, for $\mathbf{z} = \mathbf{ln}\,\boldsymbol{x}$. In other words, only endomorphisms such that $\mathcal{V}$ is stable by A, i.e. s.t. $A(\mathcal{V}) \subset \mathcal{V}$, will lead to a differing $A^+ : \mathbb{P}^d_{++} \to \mathbb{P}^d_{++}$. In projective language, any linear $A : \mathbb{R}^{d+1} \to \mathbb{R}^{d+1}$ which maps the projective point $[\mathbf{1}]$ and the projective hyperplane $[\mathcal{V}]$ to themselves, gives rise to a unique linear map on $\mathbb{P}^d_{++}$. In matrix terms, if $(\mathbf{1}, \mathbf{v}_1, \ldots, \mathbf{v}_d)$ is a basis of $\mathbb{R}^{d+1} = \text{span}(\mathbf{1}) + \mathcal{V}$, the matrix of A writes in such a basis

$$A = \begin{pmatrix} \alpha & \mathbf{0}^T \\ \mathbf{0} & \tilde{A} \end{pmatrix},$$

where $\mathbf{0} = (0, \ldots, 0) \in \mathbb{R}^d$ and $\tilde{A} : \mathbb{R}^d \to \mathbb{R}^d$. Any two such matrices $A \in \mathbb{R}^{(d+1)\times(d+1)}$, i.e. with the same $\tilde{A}$, but differing only by α, gives rise to the same linear compositional mapping on $\mathbb{P}^d_{++}$. One recovers the result that any endomorphism (noted here $\tilde{A}$) of $\mathbb{R}^d$ induces a unique linear compositional mapping A^+ on $\mathbb{P}^d_{++}$. One thus obtains the following definition:

DEFINITION 2.3. Any linear application $A : \mathbb{R}^{d+1} \to \mathbb{R}^{d+1}$, s.t. $\mathbf{1}$ is an eigenvector of A, induces a linear application A^+ on $(\mathbb{P}^d_{++}, \oplus, \odot)$. Conversely, the set of linear applications on $(\mathbb{P}^d_{++}, \oplus, \odot)$ is isomorphic to the set of linear applications $\tilde{A} : \mathbb{R}^d \to \mathbb{R}^d$.

REMARK 2. i) There is a slight difference in our approach to defining A^+, as it is based on a $A : \mathbb{R}^{d+1} \to \mathbb{R}^{d+1}$, compared to the classical CoDa literature where A^Δ is directly defined from an application $A : \mathbb{R}^d \to \mathbb{R}^d$. Confusingly, the literature often notes A^* for A and A for our A^Δ. Note also that in the equivalence class viewpoint, there is no need to introduce the extra cumbersome notation $\boxed{\cdot}$.

ii) We used a different notation to distinguish between A^+ and the compositional morphism $[A]_+$ introduced in Section 2.3: $[A]_+$ is defined on the larger $\mathbb{P}^d_+$ space whereas A^+ only acts on the positive $\mathbb{P}^d_{++}$, A^+ is linear whereas $[A]_+$ is not, and the conditions on A to construct A^+ and $[A]_+$ are different.

iii) Similar constructions at the equivalence class level can be obtained for a linear map A between spaces of different dimension. We omit the details.

3. Log coordinates on $\mathbb{P}^d_{++}$ at the level of equivalence classes and connections with Aitchison's geometry. On the simplex representation Δ^d_{++} of $\mathbb{P}^d_{++}$, Aitchison's log-ratio analysis is based on the set of log-coordinates introduced in Section 1. In this section, we deepen the intrinsic approach, begun in Sections 2.4 and 2.5, of studying $\mathbb{P}^d_{++}$ directly at the level of equivalence classes.

3.1. *Isomorphism of vector spaces.* The $\mathbf{ln}$ map acts on equivalence classes of positive CoDa vectors $[\mathbf{x}]_+ \in \mathbb{P}^d_{++}$ as follows: for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{d+1}_{++}$,

$$
\begin{aligned}
[\mathbf{x}]_+ = [\mathbf{y}]_+ &\Leftrightarrow \exists \lambda > 0 \text{ s.t. } \mathbf{x} = \lambda \mathbf{y} \\
&\Leftrightarrow \exists \lambda > 0 \text{ s.t. } \mathbf{ln}\mathbf{x} = \mathbf{ln}\mathbf{y} + \ln \lambda \mathbf{1},
\end{aligned}
$$

where $\mathbf{1} = (1, \dots, 1) \in \mathbb{R}^{d+1}$. This suggest to consider the following equivalence relation $\sim$ on $\mathbb{R}^{d+1}$: for $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^{d+1}$,

$$
\mathbf{z}_1 \sim \mathbf{z}_2 \Leftrightarrow \exists \mu \in \mathbb{R} \text{ s.t. } \mathbf{z}_1 = \mathbf{z}_2 + \mu \mathbf{1},
$$

which means that two vectors of $\mathbb{R}^{d+1}$ whose coordinates only differ by a common constant are identified. Set $\mathbb{R}^{d+1}/\sim$ the corresponding quotient space, and denote by

$$
[\mathbf{z}]_\sim = \{\mathbf{z} + \mu \mathbf{1}, \mu \in \mathbb{R}\} \tag{15}
$$

the equivalence classes of $\mathbb{R}^{d+1}/\sim$. In other words, the $\mathbf{ln}$ map can be lifted to an operation on equivalence classes, transforming positive CoDa elements $[\mathbf{x}]_+ \in \mathbb{P}^d_{++}$, viewed geometrically as rays in $\mathbb{R}^{d+1}_{++}$, to elements $[\mathbf{ln}\, \boldsymbol{x}]_\sim$ of the quotient space $\mathbb{R}^{d+1}/\sim$, interpreted geometrically as parallel lines in $\mathbb{R}^{d+1}$ along the direction $\mathbf{1}$,

$$
\mathbf{ln}([\mathbf{x}]_+) = [\mathbf{ln}\, \boldsymbol{x}]_\sim, \quad \mathbf{exp}([\mathbf{z}]_\sim) = [\boldsymbol{e}^{\boldsymbol{z}}]_+. \tag{16}
$$

In addition, this $\mathbf{ln}$ transform is bijective, as shown in the next lemma:

LEMMA 3.1. $\mathbf{ln} : \mathbb{P}^d_{++} \to \mathbb{R}^{d+1}/\sim$ *is bijective.*

Moreover, since each representative $\mathbf{z} + \mu \mathbf{1}$ of the line $[\mathbf{z}]_\sim$ is a vector of the ambient vector space $(\mathbb{R}^{d+1}, +, .)$, the equivalence classes of $\mathbb{R}^{d+1}/\sim$ inherit from the latter space its vector space structure by pulling back its vector space operations. Indeed, let us define the vector space operations $\tilde{+}, \tilde{\cdot}$ on the quotient space $\mathbb{R}^{d+1}/\sim$, by

$$
\begin{aligned}
&[\mathbf{z}_1]_\sim \tilde{+} [\mathbf{z}_2]_\sim := [\mathbf{z}_1 + \mathbf{z}_2]_\sim, \quad \mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^{d+1} \\
&\lambda \tilde{\cdot} [\mathbf{z}]_\sim := [\lambda \mathbf{z}]_\sim, \quad \mathbf{z} \in \mathbb{R}^{d+1}, \lambda \in \mathbb{R}.
\end{aligned} \tag{17}
$$

(One readily checks that these operations are well defined, i.e. do not depend on the vector representative). In turn, $(\mathbb{R}^{d+1}/\sim, \tilde{+}, \tilde{\cdot})$ can be pull-backed to $\mathbb{P}^d_{++}$ by the $\mathbf{exp}$ map to obtain $(\mathbb{P}^d_{++}, \oplus, \odot)$: in view of (16),

$$
\begin{aligned}
\mathbf{exp}([\mathbf{z}_1]_\sim \tilde{+} [\mathbf{z}_2]_\sim) &= \mathbf{exp}([\mathbf{z}_1 + \mathbf{z}_2]_\sim) \\
&= [e^{\mathbf{z}_1 + \mathbf{z}_2}]_+ = [e^{\mathbf{z}_1} \times e^{\mathbf{z}_2}]_+ \\
&= [e^{\mathbf{z}_1}]_+ \oplus [e^{\mathbf{z}_2}]_+ \\
\mathbf{exp}(\lambda \tilde{\cdot} [\mathbf{z}]_\sim) &= \mathbf{exp}([\lambda \mathbf{z}]_\sim) = [e^{\lambda \mathbf{z}}]_+ = [(e^{\mathbf{z}})^\lambda]_+ \\
&= \lambda \odot [e^{\mathbf{z}}]_+
\end{aligned}
$$

In other words, one has obtained the following proposition:

PROPOSITION 3.2. *The* $\mathbf{ln}$ *and* $\mathbf{exp}$ *maps are isomorphisms of the vector spaces* $(\mathbb{P}^d_{++}, \oplus, \odot)$ *and* $(\mathbb{R}^{d+1}/\sim, \tilde{+}, \tilde{\cdot})$*, as illustrated in the commutative diagram of Figure 6.*

This description of Aitchison's operations $\oplus, \odot$ at the level of equivalence classes thus gives new insight on the intrinsic algebraic structure of $\mathbb{P}^d_{++}$.

$$
\begin{tikzcd}
(\mathbb{R}^{d+1}_{++}, \times, \hat{}) \arrow[d, "{[.]_+}"] \arrow[r, bend left, "\mathbf{ln}"] & (\mathbb{R}^{d+1}, +, .) \arrow[l, "\mathbf{exp}"] \arrow[d, "{[.]_\sim}"] \\
(\mathbb{P}^d_{++}, \oplus, \odot) \arrow[r, bend right, "\mathbf{ln}"'] & (\mathbb{R}^{d+1}/\sim, \tilde{+}, \tilde{\cdot}) \arrow[l, "\mathbf{exp}"]
\end{tikzcd}
$$

FIG 6. *Isomorphism of vector spaces in log-coordinates.*

3.2. *Normalised log representatives: alr coordinates.* From such an approach in terms of equivalence classes and with Figure 6 in mind, one can obtain a geometrical description of the alr, clr and ilr vector coordinates mentionned in Section 1. Indeed, we can identify the quotient space $\mathbb{R}^{d+1}/\sim$, i.e. the stack of lines parallel to $\mathbf{1}$, with a subset of points in $\mathbb{R}^{d+1}$ by taking as vector representatives $\mathbf{z} \in \mathbb{R}^{d+1}$ of $[\mathbf{z}]_\sim$, the intersection of the line $[\mathbf{z}]_\sim$ with any hyperplane of $\mathbb{R}^{d+1}$ not parallel to $\mathbf{1}$.

For example, one can consider representatives $\mathbf{z}$ with $z_0 = 0$. Thus, a CoDa point $[\mathbf{x}]_+$, identified by Lemma 3.1 via the $\mathbf{ln}$ transform to $[\mathbf{z}]_\sim := [\mathbf{ln}\, \boldsymbol{x}]_\sim$, is now identified, by choosing $\mu = -\ln x_0$ in (15), with the vector

$$
\boldsymbol{\zeta} := \begin{pmatrix} 0 \\ \ln x_1 - \ln x_0 \\ \vdots \\ \ln x_d - \ln x_0 \end{pmatrix} = \begin{pmatrix} 0 \\ \mathrm{alr}_0(\mathbf{x}) \end{pmatrix} \in \mathbb{R}^{d+1}.
$$

In turn, by dropping the first null coordinate, the latter can be identified with the element $\mathrm{alr}_0(\mathbf{x}) = (\ln(x_1/x_0), \ldots, \ln(x_d/x_0))$ of $\mathbb{R}^d$, as described in Section 1. See Figure 7.

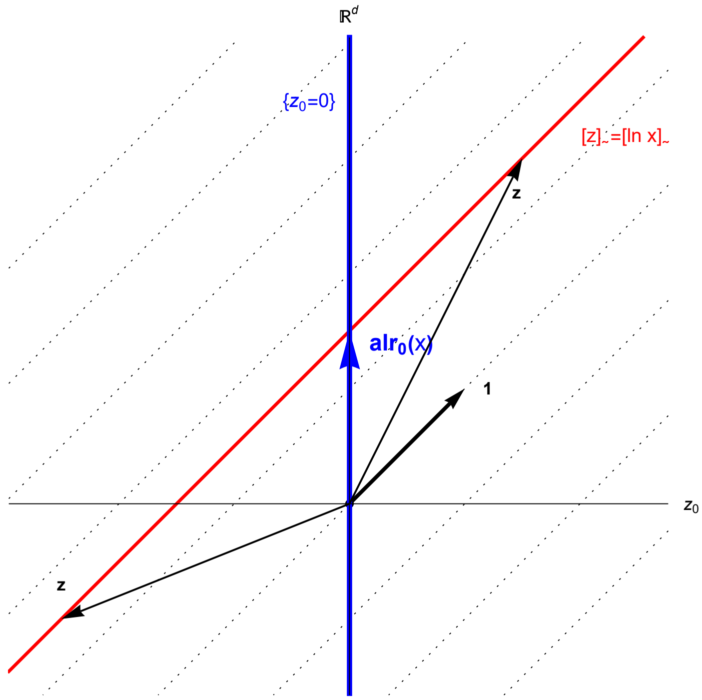

FIG 7. *$alr_0(\mathbf{x})$ as a vector representative of $[\ln \boldsymbol{x}]_\sim$. Equivalence classes $[\mathbf{z}]_\sim$ are shown as dotted lines parallel to the vector $\mathbf{1}$ (thick black arrow) in $\mathbb{R}^{d+1}$. For a given $[\mathbf{z}]_\sim = [\ln \boldsymbol{x}]_\sim$ (red line) with two vector representatives $\mathbf{z}$ (thin black arrows), the $alr_0(\mathbf{x})$ (blue arrow) vector representative is located on the intersection of $[\mathbf{z}]_\sim$ with the vector hyperplane $\{\mathbf{z} : z_0 = 0\}$ (blue vertical line).*

3.3. *Normalised log representatives: clr coordinates.* One can also associate to each log transformed equivalence class $[\ln \boldsymbol{x}]_\sim$ of positive CoDa element $[\mathbf{x}]_+$ a unique normalised/distinguished vector representative in $\mathbb{R}^{d+1}$ as follows: decomposes any vector representative $\mathbf{z} \in \mathbb{R}^{d+1}$ of the line $[\mathbf{z}]_\sim := [\ln \boldsymbol{x}]_\sim$ into its component parallel to $\mathbf{1}$ and its component $\mathbf{z}^\perp \in \mathbb{R}^{d+1}$ orthogonal to $\mathbf{1}$: one sets

$$\mathbf{z} = \mathbf{z}^\perp + \mu \mathbf{1}, \text{ s.t. } \langle \mathbf{z}^\perp | \mathbf{1} \rangle = 0.$$

In other words, one defines the operation $^\perp : \mathbb{R}^{d+1} \to \mathbb{R}^{d+1}$ as

$$\mathbf{z}^\perp := \mathbf{z} - \text{proj}_{\mathbf{1}}(\mathbf{z}) = \mathbf{z} - \frac{\langle \mathbf{z} | \mathbf{1} \rangle}{d+1} \mathbf{1},$$

where $\text{proj}_{\mathbf{1}}$ is the orthogonal projection on $\mathbf{1}$. It is easily seen that the mapping $[\mathbf{z}]_\sim \to \mathbf{z}^\perp$ is well defined: if $\mathbf{z}_1 = \mathbf{z} + \mu \mathbf{1}$ is another representative of $[\mathbf{z}]_\sim$, then

$$\mathbf{z}_1^\perp = \mathbf{z} + \mu \mathbf{1} - \frac{\langle \mathbf{z}_1 | \mathbf{1} \rangle}{d+1} \mathbf{1} = \mathbf{z} + \mu \mathbf{1} - \frac{\langle \mathbf{z} | \mathbf{1} \rangle}{d+1} \mathbf{1} - \mu \mathbf{1} = \mathbf{z}^\perp.$$

By definition, $\mathbf{z}^\perp$ lies in the "null-sum" vector subspace

$$\mathcal{H}_{\mathbf{1}} := \{\mathbf{z} \in \mathbb{R}^{d+1} : \sum_{i=0}^{d} z_i = 0\},$$

(of dimension d), which is the vector hyperplane parallel to the affine hyperplane $\mathcal{H}_{\texttt{sum}}$ of the affine representation i) of Section 2.2. See Figure 8.

For a positive CoDa $[\mathbf{x}]_+ \in \mathbb{P}^d_{++}$, $(\mathbf{ln}\,\boldsymbol{x})^\perp$ corresponds to the clr coordinates. Indeed, in coordinates, $\mathbf{z}^\perp$ writes in matrix terms as

$$\mathbf{z}^\perp = (I_{d+1} - \frac{\mathbf{1}\mathbf{1}^T}{d+1})\mathbf{z} = \begin{pmatrix} \frac{d}{d+1} & \frac{-1}{d+1} & \cdots & \frac{-1}{d+1} \\ \frac{-1}{d+1} & \frac{d}{d+1} & \cdots & \frac{-1}{d+1} \\ \vdots & \ddots & \ddots & \vdots \\ \frac{-1}{d+1} & \cdots & \frac{-1}{d+1} & \frac{d}{d+1} \end{pmatrix} \mathbf{z},$$

where I_{d+1} is the identity matrix of $\mathbb{R}^{d+1}$. For $\mathbf{z} = \mathbf{ln}\,\boldsymbol{x}$, one has

$$\begin{aligned}(\mathbf{ln}\,\boldsymbol{x})^\perp &= \mathbf{ln}\,\boldsymbol{x} - \left(\frac{1}{d+1}\sum_{j=0}^{d} \ln x_j\right)\mathbf{1} \\ &= \mathbf{ln}\left(\frac{\mathbf{x}}{g(\mathbf{x})}\right) = \mathrm{clr}(\mathbf{x}),\end{aligned}$$

where $g(\mathbf{x})$ is the geometric mean of $\mathbf{x}$. One thus re-obtains the algebraic expression of clr of Section 1, which is often given in the literature without geometric insight.

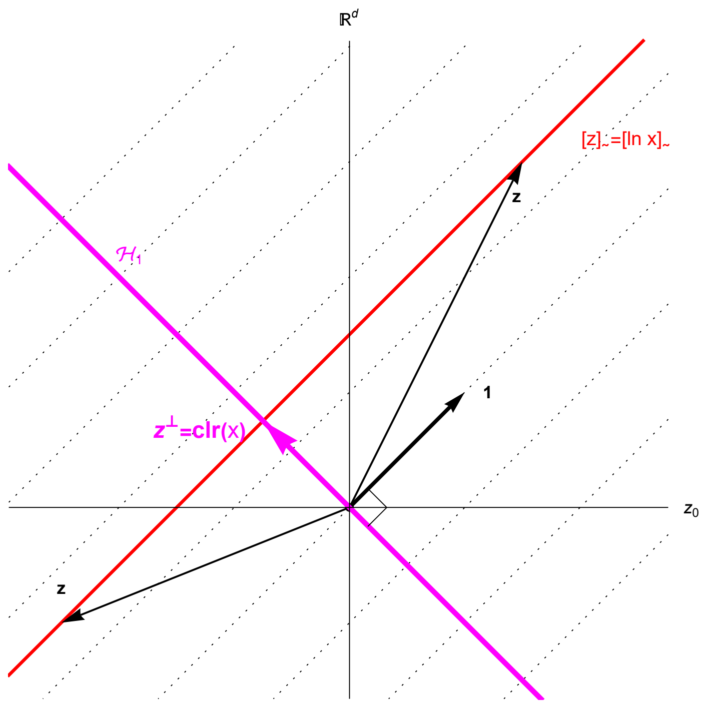

FIG 8. *clr$(\mathbf{x})$ as a vector representative of $[\mathbf{ln}\,\boldsymbol{x}]_\sim$. For a given $[\mathbf{z}]_\sim = [\mathbf{ln}\,\boldsymbol{x}]_\sim$ (red line) with two vector representatives $\mathbf{z}$ (thin black arrows), the $\mathbf{z}^\perp = clr(\mathbf{x})$ (pink arrow) vector representative is located on the intersection of $[\mathbf{z}]_\sim$ with the null-sum vector hyperplane $\mathcal{H}_\mathbf{1}$ (pink diagonal line).*

REMARK 3. Since $\mathcal{H}_\mathbf{1}$ is a vector subspace, the representatives $\mathbf{z}^\perp$ inherits its vector space structure. This gives a more concrete, yet equivalent, way to define the $\tilde{+}, \tilde{\cdot}$ operations, via these orthogonal representatives, as

$$\begin{aligned}&[\mathbf{z}_1]_\sim \,\tilde{+}\, [\mathbf{z}_2]_\sim := [\mathbf{z}_1^\perp + \mathbf{z}_2^\perp]_\sim, \quad \mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^{d+1} \\ &\lambda \,\tilde{\cdot}\, [\mathbf{z}]_\sim := [\lambda \mathbf{z}^\perp]_\sim, \quad \mathbf{z} \in \mathbb{R}^{d+1}, \lambda \in \mathbb{R}.\end{aligned}$$

By pulling back to $\mathbb{P}^d_{++}$ by the $\mathbf{exp}$ map, one obtains the original definition of Aitchison's operations in terms of clr coordinates. Here, one sees that the clr coordinates are again derived vector coordinates from the underlying equivalence classes $[.]_\sim$, and $[.]_+$.

3.4. *Normalised log representatives: ilr coordinates.* A third way to associate to a unique normalised/distinguished vector representative in $\mathbb{R}^d$ to a positive CoDa element $[\mathbf{x}]_+$ is by coordonatization in an orthonormal basis. Indeed, the previous section has shown that to $[\mathbf{z}]_\sim = [\mathbf{ln}\,\boldsymbol{x}]_\sim$ is associated a unique vector representative $\mathbf{z}^\perp$ of $\mathbb{R}^{d+1}$. By definition $\mathbf{z}^\perp \in \mathcal{H}_{\mathbf{1}}$, the null-sum vector hyperplane (3.3), which is of dimension d. Hence, by changing the canonical basis $(\mathbf{e}_0, \mathbf{e}_1, \ldots, \mathbf{e}_d)$ of $\mathbb{R}^{d+1}$ into an orthonormal basis $(\mathbf{1}/||\mathbf{1}||, \mathbf{v}_1, \ldots, \mathbf{v}_d)$, one can express $\mathbf{z}^\perp$ as

$$\mathbf{z}^\perp = \langle \mathbf{z}^\perp | \mathbf{1} \rangle \frac{\mathbf{1}}{||\mathbf{1}||^2} + \sum_{j=1}^{d} \langle \mathbf{z}^\perp | \mathbf{v}_j \rangle \mathbf{v}_j = \sum_{j=1}^{d} \langle \mathbf{z}^\perp | \mathbf{v}_j \rangle \mathbf{v}_j$$

and thus identify $\mathbf{z}^\perp$ with its coordinate vector representative $(\langle \mathbf{z}^\perp | \mathbf{v}_1 \rangle, \ldots, \langle \mathbf{z}^\perp | \mathbf{v}_j \rangle)$ in the orthonormal basis $(\mathbf{v}_1, \ldots, \mathbf{v}_d)$ of $\mathbf{R}^d$.

Matricially, this vector representative in $\mathbb{R}^d$ is expressed as follows: set $\mathbf{V} = (\mathbf{v}_1 \ldots \mathbf{v}_d) \in \mathbb{R}^{(d+1)\times d}$ the sub-matrix of column orthonormal vectors, then the orthonormal coordinate representation of $\mathbf{z}^\perp$ in $\mathbb{R}^d$ writes $V^T \mathbf{z}^\perp$. In terms of the original Coda element $[\mathbf{x}]_+$, one can thus define the ilr transform, directly at the level of the equivalence class $[\mathbf{x}]_+$ as

$$\mathrm{ilr}_V([\mathbf{x}]_+) = V^T \mathbf{ln}\,\boldsymbol{x} \in \mathbb{R}^d.$$

Again, as for alr and clr, the algebraic expression of the ilr coordinates is obtained from a geometric reasoning. Note again that there is no need for closure, i.e. to go through a simplex representation. One can now complete Figure 6 by Figure 9, to give a full picture of the geometric underpinnings and relations between the different Aitchison's log-ratio transforms.

$$\begin{tikzcd}
\mathbf{x}\in\mathbb{R}^{d+1}_{++} \arrow[r, bend left, "\mathbf{ln}"] \arrow[d, "{[.]_+}"] & \mathbf{ln}\,\boldsymbol{x}\in\mathbb{R}^{d+1} \arrow[l, "\mathbf{exp}"] \arrow[d, "{[.]_\sim}"] & (\mathbf{ln}\,\boldsymbol{x})^\perp \in \mathcal{H}_1 \subset \mathbb{R}^{d+1} \\
{[\mathbf{x}]_+ \in \mathbb{P}^d_{++}} \arrow[r, bend right, "\mathbf{ln}"'] & {[\mathbf{ln}\,\boldsymbol{x}]_\sim \in \mathbb{R}^{d+1}/\sim} \arrow[l, "\mathbf{exp}"'] \arrow[ur, "\mathrm{clr}"] \arrow[d, "\mathrm{alr}_0"] \arrow[dr, "\mathrm{ilr}_V"] & \\
 & \mathbf{ln}(\mathbf{x}/x_0)\in\mathbb{R}^d & V^T\mathbf{ln}\,\boldsymbol{x}\in\mathbb{R}^d
\end{tikzcd}$$

FIG 9. *Aitchison's log-ratio alr, clr and ilr transforms from the equivalence classes viewpoint.*

4. Hilbert's projective metric on $\mathbb{P}^d_{++}$.

4.1. *Desiderata for the definition of a metric on CoDa.* Having identified in Section 2 the two fundamental, projective and order, structures of CoDa, we can now enquire for an adequate metric to compare CoDa points. In view of the preceding discussion, it is thus required that a sensible metric on $\mathbb{P}^d_+$ should

i) be compatible with these partial ordering and projective structures;

ii) be intrinsic, in the sense that it does not depend on a particular representation of the equivalence class, nor on a particular metricization of an ambient space.[10]

Equivalently, this amounts to building a distance[11] function d on the non-negative orthant cone, $d : \mathbb{R}_+^{d+1} \times \mathbb{R}_+^{d+1} \to [0, \infty]$, which is scale-invariant, viz.

$$d(\lambda \mathbf{x}, \mu \mathbf{y}) = d(\mathbf{x}, \mathbf{y}), \quad \lambda, \mu > 0. \tag{18}$$

In addition to these structural conditions (and in particular scale invariance), [3] states two other invariance conditions that any statistical method applied to compositions should satisfy: iii) permutation invariance, and iv) subcompositional coherence. In short, permutation invariance means that the analysis should give equivalent results when the ordering of the parts in the composition is changed and subcompositional coherence that the distance between two full compositions must be greater than, or equal to, the distance between them when considering any subcomposition.

It will be shown in this section and the next, that, for positive CoDa elements in $\mathbb{P}^d_{++}$, Hilbert's projective metric appears to be particularly well-suited. Actually there are two versions of Hilbert's metric: the original [29] metric defined by cross-ratio on the relative interior of a bounded convex set and [12]'s version as a distance between pairs of rays in a cone. These projective metrics have proven useful in numerous fields, such as Perron-Froebenius theorems, matrix scaling/DAD/IPFP/Sinkhorn algorithm, etc.. See [43] and [33] for an overview of Hilbert geometry. Article reviews are [34], [17], [32]. The application of Hilbert's cross-ratio distance for clustering on the simplex was proposed by [40] (see also [38]). We first recall Hilbert's version in the next subsection.

4.2. *Hilbert's original definition of a projective metric on a bounded convex set.* In a letter to Klein, [29] noted that a formula of Klein, based on the cross-ratio of projective geometry, provides a metric on any bounded convex domain. The definition is as follows. Let Ω be a bounded, open, convex subset of a real finite dimensional affine normed space. Given $\mathbf{x} \neq \mathbf{y} \in \Omega$, set the cross-ratio

$$R(\boldsymbol{x'}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y'}) := \frac{||\mathbf{x}' - \mathbf{y}||}{||\mathbf{x}' - \mathbf{x}||} \frac{||\mathbf{y}' - \mathbf{x}||}{||\mathbf{y}' - \mathbf{y}||},$$

where $\mathbf{x}', \mathbf{y}'$ are the intersection of the line $\{\lambda \mathbf{x} + (1-\lambda)\mathbf{y}, \lambda \in \mathbb{R}\}$ with Ω, and are ordered as in Figure 10 below.

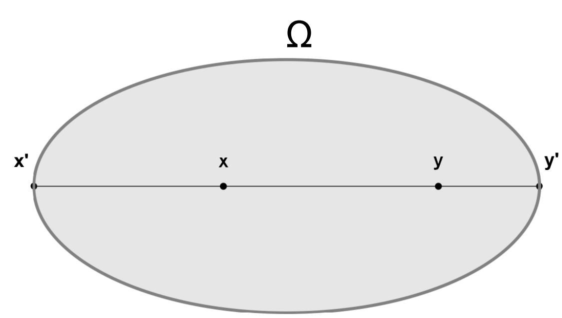

FIG 10. *Hilbert's projective metric for a bounded convex set*

[10] Aitchison's distance depends on the Euclidean metric of the embedding space $\mathbb{R}^{d+1}$.

[11] At this preliminary stage, we allow for the search of an infinite metric on $\mathbb{P}^d_+$. The proposed Hilbert metric will only be finite on $\mathbb{P}^d_{++}$. See also Remark 4.

Hilbert's metric δ is defined as the log of the cross ratio: for $\mathbf{x} \neq \mathbf{y}$, set

$$\delta(\boldsymbol{x}, \boldsymbol{y}) := \ln R(\boldsymbol{x'}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{y'}) = \ln \frac{||\mathbf{x}' - \mathbf{y}||}{||\mathbf{x}' - \mathbf{x}||} \frac{||\mathbf{y}' - \mathbf{x}||}{||\mathbf{y}' - \mathbf{y}||},$$

(sometimes a factor $1/2$ is included). Set also $\delta(\mathbf{x}, \mathbf{x}) = 0$. Then, one can show (see e.g. [43]) that

i) δ defines a finite metric on the relative interior of Ω, and the latter space equipped with this metric becomes a complete metric space.
ii) δ is invariant under projective transformations.
iii) the Hilbert metric is projective in the sense that the straight lines are geodesics: $\delta(\mathbf{x}, \mathbf{y}) = \delta(\mathbf{x}, \mathbf{z}) + \delta(\mathbf{z}, \mathbf{y})$ whenever $\mathbf{z}$ is on the segment $\{\lambda\mathbf{x} + (1-\lambda)\mathbf{y}, \lambda \in [0,1]\}$.

In view of (7) in Section 2.2, the simplex representative $[\mathbf{x}]_1$ of positive CoDa elements $[\mathbf{x}]_+$ lie in the open, bounded, convex set Δ^d_{++}, so that Hilbert's metric can be applied to $\Omega = \Delta^d_{++}$. This leads to a definition of Hilbert distance δ for positive CoDa elements $[\mathbf{x}]_+, [\mathbf{y}]_+ \in \mathbb{P}^d_{++}$ as

$$\delta([\mathbf{x}]_+, [\mathbf{y}]_+) = \delta([\mathbf{x}]_1, [\mathbf{y}]_1).$$

The invariance of the cross-ratio under projective transformations, i.e. that $R(\mathbf{x}', \mathbf{x}, \mathbf{y}, \mathbf{y}') = R(T(\mathbf{x}'), T(\mathbf{x}), T(\mathbf{y}), T(\mathbf{y}))$, where $T(\mathbf{x}'), T(\mathbf{x}), T(\mathbf{y}), T(\mathbf{y})$ are the intersections of a hyperplane with the rays through $\mathbf{x}', \mathbf{x}, \mathbf{y}$ and $\mathbf{y}'$, respectively, mean that δ for positive Coda does not depend on this simplex representation[12] and translates in the sought for scale invariance property (18). Moreover, since δ is based on the ratios of distances $||\mathbf{x}' - \mathbf{y}||/||\mathbf{x}' - \mathbf{x}||$ and $||\mathbf{y}' - \mathbf{x}||/||\mathbf{y}' - \mathbf{y}||$ of aligned triples of points, it only depends on the relative positions of the four points $\mathbf{x}', \mathbf{x}, \mathbf{y}, \mathbf{y}'$, and not on the norm $||.||$ used. Indeed, setting $\mathbf{x}' - \mathbf{x} = t(\mathbf{x}' - \mathbf{y})$ and $\mathbf{y}' - \mathbf{y} = s(\mathbf{y}' - \mathbf{x})$, with $0 < s, t < 1$, then $R(\mathbf{x}', \mathbf{x}, \mathbf{y}, \mathbf{y}') = 1/(st) > 1$, which only depends on the scalar s, t defining $\mathbf{x}', \mathbf{y}'$ and not on the norm used to measure the distances.

4.3. *Birkhoff's ordered version of Hilbert's metric on a cone.* Birkhoff's version of Hilbert's projective metric ([12]) is based on the partial order induced by a cone of a, possibly infinite dimensional, real Banach space V, see [34], [17]. For our purposes, we simply to need the case $V = \mathbb{R}^{d+1}$. Let K a closed pointed solid convex cone with non-empty interior, i.e. s.t.

i) $K \subset V$, K is closed, the interior $\mathring{K}$ is not empty;
ii) $K + K \subset K$;
iii) $\lambda K \subset K$ for all $\lambda \geq 0$;
iv) $K \cap -K = \{0\}$.

Then, K induces on V an (Archimedean) partial order $\leq$,

$$\boldsymbol{x} \leq \boldsymbol{y} \Longleftrightarrow \mathbf{y} - \mathbf{x} \in K$$

A way to quantify this order is via the following pair of asymmetric quantities, defined for $\mathbf{x}, \mathbf{y} \in K^* := K \setminus \{0\}$:

$$M(\mathbf{x}, \mathbf{y}) := \inf\{\lambda \geq 0 : \mathbf{x} \leq \lambda \mathbf{y}\},$$
$$m(\mathbf{x}, \mathbf{y}) := \sup\{\mu \geq 0 : \mu \boldsymbol{y} \leq \boldsymbol{x}\}, \tag{19}$$

[12]This will be also clear using Birkhoff's order approach to the definition of Hilbert's metric. See next subsection.

with $m(\mathbf{x},\mathbf{y}) = \infty$, $M(\mathbf{x},\mathbf{y}) = \infty$ if the corresponding set is empty. These quantities have a geometric interpretation: given $\mathbf{x} \in K^*$ chosen as a reference point, $M(\mathbf{x},\mathbf{y})$ tells how much one has to expand the length of $\mathbf{y}$ (keeping its direction fixed) to make it larger than $\mathbf{x}$, and conversely, $m(\mathbf{x},\mathbf{y})$ tells how much one has to shrink $\mathbf{y}$ to make it lower than $\mathbf{x}$. It is noteworthy that $m(\mathbf{x},\mathbf{y})$, $M(\mathbf{x},\mathbf{y})$ are finite whenever $\mathbf{x},\mathbf{y} \in \mathring{K}$, even though $\mathbf{x},\mathbf{y}$ may not be comparable, see Example 1.

EXAMPLE 1. Take $K = \mathbb{R}_2^+$, $\mathbf{x} = (1/3,2/3)$, $\mathbf{y} = (1/2,1/2)$. $\mathbf{x},\mathbf{y}$ are not comparable: $\mathbf{y}$ is larger than $\mathbf{x}$ w.r.t. to the first coordinate, and is lower w.r.t. to the second coordinate. One has to enlarge $\mathbf{y}$ by at least $4/3$ to have $4/3\boldsymbol{y} \geq \boldsymbol{x}$, and shrink it by at least $2/3$ to have $2/3\boldsymbol{y} \leq \boldsymbol{x}$, viz. $M(\mathbf{x},\mathbf{y}) = 4/3$, $m(\mathbf{x},\mathbf{y}) = 2/3$, see Figure 11.

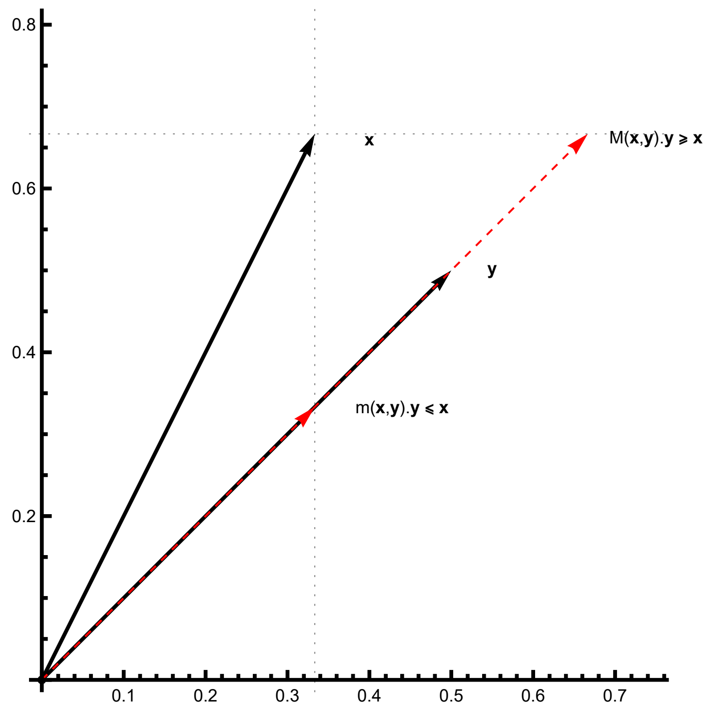

FIG 11. *Quantization of the orthant order via the M and m functionals of Example 1.* $\mathbf{x} = (1/3,2/3)$, $\mathbf{y} = (1/2,1/2)$ *(black arrows) give* $M(\mathbf{x},\mathbf{y}) = 4/3$, $m(\mathbf{x},\mathbf{y}) = 2/3$.

In particular, for $K = \mathbb{R}_+^{d+1}$, one has the following explicit formulas for m and M,

$$M(\mathbf{x},\mathbf{y}) := \max_i \frac{x_i}{y_i}, \quad m(\mathbf{x},\mathbf{y}) := \min_i \frac{x_i}{y_i}, \quad \mathbf{x},\mathbf{y} > \mathbf{0}, \tag{20}$$

which justify their notation. Birkhoff's version of Hilbert's projective metric d_H is then defined on K^* as follows:

DEFINITION 4.1. Hilbert's projective metric d_H on K^* is defined by $d(\mathbf{x},\mathbf{0}) = 0$, $d(\mathbf{0},\mathbf{y}) = 0$ and

$$d_H(\mathbf{x},\mathbf{y}) = \ln \frac{M(\mathbf{x},\mathbf{y})}{m(\mathbf{x},\mathbf{y})}, \quad \mathbf{x},\mathbf{y} \in K^*. \tag{21}$$

For $K = \mathbb{R}_+^{d+1}$ and $\mathbf{x}, \mathbf{y} > \mathbf{0}$, since

$$\frac{1}{\min_j\{x_j/y_j\})} = \max_j\{(x_j/y_j)^{-1}\} = \max_j\{y_j/x_j\},$$

one has

$$d_H(\mathbf{x},\mathbf{y}) = \ln\frac{\max_i x_i/y_i}{\min_j x_j/y_j} = \ln\max_{i,j}\frac{x_i y_j}{x_j y_i}.$$

The connection between the two versions of Hilbert's projective metric is explained in [32]. For the convenience of the reader, we reproduce their argument: let $\mathbf{x},\mathbf{y} \in \mathring{K}$ and write $m = m(\mathbf{x},\mathbf{y})$ and $M = M(\mathbf{x},\mathbf{y})$. Replacing $\mathbf{x}$ by $\lambda\mathbf{x}$ for a suitable $\lambda > 0$, if necessary, will insure that the line through $\mathbf{x}$ and $\mathbf{y}$ leaves K at two points, $\mathbf{a}$ and $\mathbf{b}$, in the two-dimensional subspace spanned by $\mathbf{x}$ and $\mathbf{y}$, see Figure 12. As explained in Example 1, the point $\mathbf{x} - m\mathbf{y}$ is obtained by moving from $\mathbf{x}$ in the $-\mathbf{y}$ direction until the non-negativity constraint is violated. By similar triangles, we see that $m = \frac{\overline{\mathbf{ax}}}{\overline{\mathbf{ay}}}$ and that $M = \frac{\overline{\mathbf{xb}}}{\overline{\mathbf{yb}}}$. Therefore,

$$d_H(\mathbf{x},\mathbf{y}) = \ln\frac{M(\mathbf{x},\mathbf{y})}{m(\mathbf{x},\mathbf{y})} = \ln\frac{\overline{\mathbf{xb}}.\overline{\mathbf{ay}}}{\overline{\mathbf{ax}}.\overline{\mathbf{yb}}} = \ln R(\mathbf{a},\mathbf{x},\mathbf{y},\mathbf{b}) = \delta(\mathbf{x},\mathbf{y}),$$

which establishes the connection with the original definition of Hilbert's metric via cross-ratio. For another proof, see [34] Theorem 2.2.

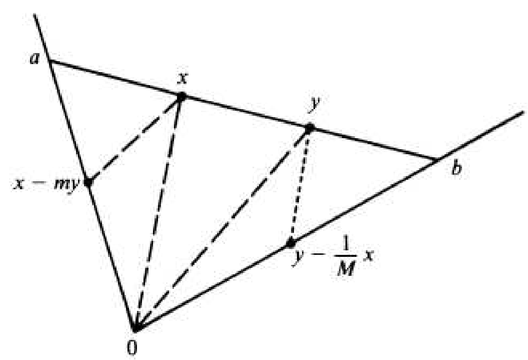

FIG 12. *Birkhoff's ordered version of Hilbert's metric as a projective metric based on the cross-ratio.*

Hilbert's projective metric turns $(\mathring{K}, d_H)$ into a pseudo-metric space, with finite d_H, where $\mathring{K}$ is the interior of K. More precisely, one has the following properties, which follows from those of m, M. (See e.g. [17], [32] or [34]).

PROPOSITION 4.2. i) $d_H(\mathbf{x},\mathbf{y}) \geq 0$, $d_H(\mathbf{x},\mathbf{y}) < \infty$ *for all* $\mathbf{x},\mathbf{y} \in \mathring{K}$.
ii) *On* $\mathring{K}$*,* $d_H(\mathbf{x},\mathbf{y}) = 0$ *if and only if* $\mathbf{x} = \lambda\mathbf{y}$*, for some* $\lambda > 0$.
iii) *Symmetry:* $d_H(\mathbf{x},\mathbf{y}) = d_H(\mathbf{y},\mathbf{x})$.
iv) *Triangle:* $d_H(\mathbf{x},\mathbf{z}) \leq d_H(\mathbf{x},\mathbf{y}) + d_H(\mathbf{y},\mathbf{z})$
v) *Scale invariance:* $d_H(\lambda\mathbf{x},\mu\mathbf{y}) = d_H(\mathbf{x},\mathbf{y})$*, for all* $\lambda,\mu > 0$, $\mathbf{x},\mathbf{y} \in K$.

4.4. *Hilbert metric for positive CoDa vectors.* Properties ii) and v) of Proposition 4.2 are precisely the requirements of independence of the representative and of scale invariance (18), we asked for constructing a suitable metric for unnormalised CoDa vectors, i.e. for equivalence classes (6). Birkhoff's approach to Hilbert metric also makes immediately clear that d_H is intrinsic on rays of $\mathbb{R}^d_+$: no metricization of $\mathbb{R}^d$ was used in the definition of d_H, only the (partial) order structure of the cone $\mathbb{R}^d_+$. One therefore defines Hilbert's metric on the space $\mathbb{P}^d_{++}$ of positive CoDa vectors as follows:

DEFINITION 4.3. Let $K = \mathbb{R}^{d+1}_+$ be the positive orthant cone in definition (21). For $[\mathbf{x}]_+, [\mathbf{y}]_+, \in \mathbb{P}^d_{++}$, let

$$d_H([\mathbf{x}]_+, [\mathbf{y}]_+) := d_H(\mathbf{x}, \mathbf{y}) \tag{22}$$

$$= \ln\left(\max_{i,j} \frac{x_i y_j}{x_j y_i}\right) \tag{23}$$

$$= \max_{i,j} \ln\left(\frac{x_i y_j}{x_j y_i}\right) \tag{24}$$

where $\mathbf{x}, \mathbf{y}$ are any representatives in $\mathbb{R}^{d+1}_{++}$ of $[\mathbf{x}]_+, [\mathbf{y}]_+$. Then, d_H is finite and well defined on $\mathbb{P}^d_{++}$. $(\mathbb{P}^d_{++}, d_H)$ is a complete metric space.

REMARK 4 (On distances on $\mathbb{P}^d_+$). Hilbert's distance $d_H([\mathbf{x}]_+, [\mathbf{y}]_+)$ become infinite whenever $[\mathbf{x}]_+$ or $[\mathbf{y}]_+$ has some zero components. By considering $\phi(d_H([\mathbf{x}]_+, [\mathbf{y}]_+))$, where ϕ is a bounded, sub-additive, monotonically increasing function such that $\phi(0) = 0$, (e.g. $\phi(x) = 1/(1+x)$), one obtains an equivalent bounded metric on the whole $\mathbb{P}^d_+$. This new metric $\phi \circ d_H$ inherits the nice properties of d_H of Section 5, when restricted to elements of $\mathbb{P}^d_{++}$, and attains the upper bound 1 whenever $[\mathbf{x}]_+$ or $[\mathbf{y}]_+$ has some zero components.

In order to have a more meaningful comparison of CoDa with zeros, another possibility is to consider extrinsic metrics defined on the whole $\mathbb{P}^d_+$, as suggested by the different representations listed in Appendix A. In particular, the representation of $\mathbb{P}^d_+$ as the "non-negative sphere" suggest to consider the spherical/angular distance or the chordal (i.e. Hellinger) distance, while log-free affine representations suggest to use the Euclidean distance on these affine subsets. The latter approach will be pursued elsewhere.

5. Properties of Hilbert metric.

5.1. *Compatibility with the vector space structure $\oplus, \odot$ and permutation invariance.* Hilbert's metric is well suited for positive CoDa vectors, as it is translation and permutation invariant, as shown in the next Proposition.

PROPOSITION 5.1. i) ***Compatibility with the vector space structure*** $\oplus, \odot$***:*** *one has, for* $[\mathbf{x}]_+, [\mathbf{y}]_+, [\mathbf{p}]_+ \in \mathbb{P}^d_{++}$,

$$d_H([\mathbf{x}]_+ \oplus [\mathbf{p}]_+, [\mathbf{y}]_+ \oplus [\mathbf{p}]_+) = d_H([\mathbf{x}]_+, [\mathbf{y}]_+)$$

and, for $\lambda \in \mathbb{R}$,

$$d_H(\lambda \odot [\mathbf{x}]_+, \lambda \odot [\mathbf{y}]_+) = |\lambda| d_H([\mathbf{x}]_+, [\mathbf{y}]_+)$$

ii) ***Permutation invariance****: Let* $\mathbf{s} = (s(0), \ldots, s(d))$ *be a permutation of* $\{0, 1, \ldots, d\}$ *and write* $\mathbf{x_s}$ *for the vector* $(x_{s(0)}, \ldots, x_{s(d)})$*. Then,*

$$d_H([\mathbf{x_s}]_+, [\mathbf{y_s}]_+) = d_H([\mathbf{x}]_+, [\mathbf{y}]_+)$$

In fact, taking into account the vector space structure of $(\mathbb{P}^d_{++}, \oplus, \odot)$, Hilbert's distance is induced by a norm: setting the Hilbert norm as

$$||[\mathbf{x}]_+||_H := d_H([\mathbf{x}]_+, [\mathbf{1}]_+) = \max_{i,j} \ln\left(\frac{x_i}{x_j}\right), \tag{25}$$

then, the positive Coda space endowed with Hilbert metric becomes a normed vector space:

PROPOSITION 5.2. *$(\mathbb{P}^d_{++}, \oplus, \odot, ||.||_H)$ is a normed vector space, with*

$$d_H([\mathbf{x}]_+, [\mathbf{y}]_+) = ||[\mathbf{x}]_+ \ominus [\mathbf{y}]_+||_H.$$

Since every finite dimensional normed vector space is complete, this gives another proof of the completeness of $(\mathbb{P}^d_{++}, d_H)$.

5.2. *The Lipshitz property in Hilbert metric and subcompositional coherence.* As announced in Section 2.3, the unnormalised approach to CoDa vectors as equivalence class allows to express the basic operations on CoDa, like extracting a subcomposition and amalgamation, as non-negative linear transformations on the vector representatives. For example, selection of the first k components, $[x_0 : \ldots : x_{k-1}]_+ \in \mathbb{P}^{k-1}_+$, $1 \le k \le d+1$, of a CoDa $[\mathbf{x}]_+ = [x_0 : \ldots : x_d] \in \mathbb{P}^d_+$, is simply obtained by matrix multiplication of the column vector representative $\mathbf{x} \in \mathbb{R}^{d+1}_+$, with the matrix $A = \begin{pmatrix} I_k & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{k \times (d+1)}$, (or any positive scalar multiple of A), where $\mathbf{0}$ is the null matrix of size $k \times (d+1-k)$, and $\mathbf{I_k}$ the identity matrix of size k. The resulting vector $\mathbf{y} = A\mathbf{x} = (x_0, \ldots, x_{k-1})$ is an unnormalised representative of the $k-$subcomposition $[\mathbf{y}]_+ \in \mathbb{P}^{k-1}_+$, and the latter is simply obtained by taking the quotient map of such representative. In the language of compositional morphisms of Section 2.3,

$$[\mathbf{y}]_+ = [A]_+([\mathbf{x}]_+) = [A\mathbf{x}]_+.$$

Similarly, the operations of amalgamation of several components into one, as e.g. in $[\mathbf{x}]_+ = [x_0 : \ldots : x_d] \in \mathbb{P}^d_+$ transformed into $[x_0 + x_1 : x_2 : \ldots : x_d]_+ \in \mathbb{P}^{d-1}_+$, and permutation of the components can also both be conveniently described by a compositional morphism, viz. positive scalar multiples of a positive linear transformation, as the resulting matrix A is made of 1 and 0.

This is in contrast with the classical simplex approach. where a subcomposition or an amalgamation have to be rescaled to unity by dividing with the remaining mass. This results in operations expressed as fractional linear transformations, i.e. as ratio of linear transformations. Thus, by considering the full equivalence class as our objects of study, these transformations are simpler to express.

The fact that these transformations are positive linear implies that subcompositional coherence is automatically satisfied with Hilbert's metric. Indeed, Hilbert's metric has the remarkable general property that it transforms an homogeneous, monotone mapping into a Lipschitz's one, with Lipshitz constant 1, see e.g. [33] Corollary 2.1.4. Let $K \subset E, L \subset F$ two closed cones of some vector spaces E, F and $T : K \to L$ be a transformation which is r-homogeneous, viz.

$$T(\lambda \mathbf{x}) = \lambda^r T(\mathbf{x}), \quad \lambda > 0, \mathbf{x} \in K$$

and monotone, viz.

$$\mathbf{x} \le_K \mathbf{y} \Rightarrow T(\mathbf{x}) \le_L T(\mathbf{y}), \quad \mathbf{x}, \mathbf{y} \in K.$$

Then,

$$d_H(T(\mathbf{x}), T(\mathbf{y})) \le r d_H(\mathbf{x}, \mathbf{y})$$

where the Hilbert distances are w.r.t. to the corresponding cones K, L.

In our particular case, $K = \mathbb{R}^d_+$, $L = \mathbb{R}^{k-1}_+$, and as the transformation A expressing the subcomposition operation is linear, it is homogeneous of degree one, so that A is $1-$Lipschitz[13] w.r.t to the Hilbert metric of the corresponding spaces,

$$d_H([A\mathbf{x}]_+, [A\mathbf{y}]_+) \le d_H([\mathbf{x}]_+, [\mathbf{y}]_+). \tag{26}$$

[13] 1-Lipshitz maps between metric spaces are also called non-expansive maps or metric maps in the literature.

The latter is the expression of subcompositional coherence. One has thus he important property:

PROPOSITION 5.3. *The Hilbert metric is subcompositionally coherent, i.e. it satisfies (26) for A expressing the subcomposition operation, and in general for any positive linear transformation.*

REMARK 5.

i) In its rescaled-to-unit-sum representation $[\mathbf{x}]_1$, a CoDa vector can be identified with (the weights) of a probability measure. Then, taking a subcomposition van be expressed as transforming such discrete probability measure by a Markov kernel. In the field of information geometry, information monotonicity is the generalisation of subcompositional coherence and also expresses a contracting property of Markov transformed probability measures, i.e. that the statistical information of a sub-model can only decrease, see [6], [62].

ii) Theorem 2.9 in [32] shows that Hilbert's metric is essentially the only metric on the positive cone which i) is invariant on rays, ii) is projective in the sense that the straight line between two points is a geodesic, viz. a shortest possible path between two points, and iii) contracts under positive linear transformations, i.e. which satisfy (18) and (26). This is in contrast with Aitchison's distance where geodesics are curved.

5.3. *Isometric embeddings and isometries for Hilbert distance and their relation to Aitchison's log-geometry..* In this subsection, we show how Hilbert's metric can be expressed w.r.t to some (pseudo)-metrics in Euclidean spaces. This is helpful in getting a better grasp of the geometric picture of Hilbert's distance of CoDa elements. In addition, it establishes some metric connections with the log-coordinate classical approach of Aitchison, thereby completing the study of Section 3.

5.3.1. *Isometric embedding in $(\mathbb{R}^{d(d+1)/2}, ||.||_\infty)$.* Hilbert's distance writes, for any representatives $\mathbf{x}, \mathbf{y}$,

$$\begin{aligned} d_H([\mathbf{x}]_+, [\mathbf{y}]_+) &= \ln\left(\max_{0\le i,j\le d} \frac{x_i y_j}{y_i x_j}\right) = \max_{0\le i,j\le d}\left(\ln\frac{x_i}{x_j} - \ln\frac{y_i}{y_j}\right) \\ &= \max_{0\le i<j\le d}\left|\ln\frac{x_i}{x_j} - \ln\frac{y_i}{y_j}\right|. \end{aligned}$$

Since there are $d(d+1)/2$ distinct pairs (i,j) with $0 \le i < j \le d$, this suggests to define $\Psi : \mathbb{P}^d_{++} \mapsto \mathbb{R}^{d(d+1)/2}$ by

$$\Psi_{ij}([\mathbf{x}]_+) = \ln\frac{x_i}{x_j}, \quad \text{for } 0 \le i < j \le d. \tag{27}$$

Note that Ψ is well-defined on $\mathbb{P}^d_{++}$ since (27) is scale-invariant. Hilbert's distance now writes,

$$d_H([\mathbf{x}]_+, [\mathbf{y}]_+) = ||\Psi([\mathbf{x}]_+) - \Psi([\mathbf{y}]_+)||_\infty \tag{28}$$

One has thus proved the following:

PROPOSITION 5.4. *$(\mathbb{P}^d_{++}, d_H)$ is isometrically embedded into $(\mathbb{R}^{d(d+1)/2}, ||.||_\infty)$ by the distance-preserving map Ψ, viz.*

$$(\mathbb{P}^d_{++}, d_H) \hookrightarrow (\mathbb{R}^{d(d+1)/2}, ||.||_\infty).$$

REMARK 6. i) One can also directly check that Ψ is injective:

$$\begin{aligned}\Psi([\mathbf{x}]_+) = \Psi([\mathbf{y}]_+) &\Longleftrightarrow \ln(x_i/x_j) = \ln(y_i/y_j), \quad 0 \le i < j \le d \\ &\Longleftrightarrow x_i/y_i = x_j/y_j = \lambda, \quad 0 \le i < j \le d \\ &\Longleftrightarrow \mathbf{x} = \lambda \mathbf{y} \Longleftrightarrow [\mathbf{x}]_+ = [\mathbf{y}]_+\end{aligned}$$

for some $\lambda > 0$.

ii) The Ψ coordinates (27) can be arranged into an anti-symmetric matrix $\Psi([\mathbf{x}]_+) := \big(\Psi_{ij}([\mathbf{x}]_+)\big) \in \mathbb{R}^{(d+1)\times(d+1)}$. This bears some analogy with the Plücker coordinates in projective geometry, where projective subspaces can also be represented by anti symmetric matrices (Plücker matrices). Hilbert's metric then writes as a distance between matrices,

$$d_H([\mathbf{x}]_+, [\mathbf{y}]_+) = ||\Psi([\mathbf{x}]_+) - \Psi([\mathbf{y}]_+)||_{\max},$$

where the distance is the (element-wise) $\infty-$matrix norm $||.||_{\max}$, viz. $||A||_{\max} := \max_{i,j} |a_{ij}|$. Notice that such matrix norm is not sub-multiplicative. This suggests the possibility to use other matrix norms and more generally to view CoDa as matrix-valued elements.

iii) Ψ can be interpreted as a generalization of the additive log-ratio alr_0 transform introduced in Section 1. Ψ partially remedies some drawbacks of the alr_0 transform alluded earlier: whereas alr_0 is not symmetrical, as the first coordinate x_0 is used as reference for ratio standardization (see (8)) and thus plays a distinguished role, Ψ is symmetrical as it considers simultaneously all alr_i transforms, $i = 0, \dots, d$.

Proposition 5.4 shows that, with appropriate distances, Ψ establishes an isometric embedding of $\mathbb{P}^d_{++}$ into $\mathbb{R}^{d(d+1)/2}$. However, Ψ is clearly not surjective, as these spaces do not have the same dimensions. This implies in particular, that these Ψ coordinates are not independent, but are related by a system of equations, as is the case with Plücker coordinates of projective points. (For example, when $d = 2$, one can directly check that $\Psi_{12} = -\Psi_{01} + \Psi_{02}$.) In other words, the image of Ψ is a subset of $\mathbb{R}^{d(d+1)/2}$. Of course, ψ is a (bijective) isometry onto its image $\psi(\mathbb{P}^d_{++})$. Regarding alr and isometries, see also the forthcoming Corollary 5.6 and Remark 7.

5.3.2. *Isometry with $\mathbb{R}^d$.* For $\mathbf{z} \in \mathbb{R}^{d+1}$, define

$$||\mathbf{z}||_{\mathrm{MM}} = \max_{0\le i\le d} z_i - \min_{0\le i\le d} z_i. \tag{29}$$

Indeed, it is readily checked that $\mathbf{z} \mapsto ||\mathbf{z}||_{\mathrm{MM}}$ is i) non-negative, ii) sub-additive, iii) absolute homogeneous, but iv) $||\mathbf{z}||_{\mathrm{MM}}$=0 only implies $\mathbf{z} = \lambda\mathbf{1}$, for some $\lambda \in \mathbb{R}$, equivalently $[\mathbf{z}]_\sim = [\mathbf{0}]_\sim$, where $[.]_\sim$ is the equivalence class (15), defined in Section 3. Thus, $||.||_{\mathrm{MM}}$ is only a pseudo-norm on $\mathbb{R}^{d+1}$, but a genuine norm on $\mathbb{R}^{d+1}/\sim$. Combining this variation norm with the operations $\tilde{+}, \tilde{\cdot}$ of (17), one obtains that $(\mathbb{R}^{d+1}/\sim, \tilde{+}, \tilde{\cdot}, ||.||_{\mathrm{MM}})$ is a normed vector space.

On the other hand, for two positive CoDa $[\mathbf{x}]_+, [\mathbf{y}]_+ \in \mathbb{P}^d_{++}$, Hilbert's metric also writes

$$\begin{aligned}d_H([\mathbf{x}]_+, [\mathbf{y}]_+) = d_H(\mathbf{x}, \mathbf{y}) &= \ln \max_{i,j} \frac{x_i y_j}{y_i x_j} = \max_{i,j} \ln \frac{x_i}{y_i}\frac{y_j}{x_j} \\ &= \max_i \max_j \ln \frac{x_i}{y_i} - \ln \frac{x_j}{y_j} \\ &= \max_i \ln \frac{x_i}{y_i} + \max_j(-\ln \frac{x_j}{y_j})\end{aligned}$$

$$
\begin{aligned}
&= \max_i \ln \frac{x_i}{y_i} - \min_i \ln \frac{x_i}{y_i} \\
&= \max_i (\ln x_i - \ln y_i) - \min_i (\ln x_i - \ln y_i) \\
&= ||\mathbf{ln}\, \boldsymbol{x} - \mathbf{ln}\, \boldsymbol{y}||_{\text{MM}} \\
&= ||[\mathbf{ln}\, \boldsymbol{x}]_\sim \,\tilde{-}\, [\mathbf{ln}\, \boldsymbol{y}]_\sim||_{\text{MM}} \qquad (30)
\end{aligned}
$$

where in (30), $[\mathbf{ln}\, \boldsymbol{x}]_\sim \,\tilde{-}\, [\mathbf{ln}\, \boldsymbol{y}]_\sim$ stands for $[\mathbf{ln}\, \boldsymbol{x} - \mathbf{ln}\, \boldsymbol{y}]_\sim$, in agreement with the definition of the $\tilde{+}, \tilde{\cdot}$ operations of Section 3. Combined with Lemma 3.1, (30) thus establishes that the $\mathbf{ln}$ map is an isometry[14] between the normed[15] vector spaces $(\mathbb{P}^d_{++}, \oplus, \odot, ||.||_H)$ and $(\mathbb{R}^{d+1}/\sim, \tilde{+}, \tilde{\cdot}, ||.||_{\text{MM}})$. We have thus established the following:

PROPOSITION 5.5. *$(\mathbb{P}^d_{++}, \oplus, \odot, ||.||_H)$ and $(\mathbb{R}^{d+1}/\sim, \tilde{+}, \tilde{\cdot}, ||.||_{MM})$ are isometrically isomorphic.*

At last, and as explained in Section 3, we can identify the quotient space $\mathbb{R}^{d+1}/\sim$ with $\mathbb{R}^d$, by taking a vector representatives $\mathbf{z} \in \mathbb{R}^{d+1}$ located on any vector hyperplane of $\mathbb{R}^{d+1}$ not parallel to $\mathbf{1}$. It is instructive to see what happens when one chooses the alr coordinates: for $\mathbf{z}$ with $z_0 = 0$, a CoDa point $[\mathbf{x}]_+$, identified by Proposition 5.5 via the $\mathbf{ln}$ transform to $[\mathbf{z}]_\sim := [\mathbf{ln}\, \boldsymbol{x}]_\sim$, is now identified with the vector

$$\boldsymbol{\zeta} := \begin{pmatrix} 0 \\ \text{alr}_0(\mathbf{x}) \end{pmatrix} \in \mathbb{R}^{d+1},$$

where $(0, \text{alr}_0(\mathbf{x}))$ stands for the concatenation of 0 with alr_0. (30) then writes

$$d_H([\mathbf{x}]_+, [\mathbf{y}]_+) = ||(0, \text{alr}_0(\mathbf{x})) - (0, \text{alr}_0(\mathbf{y}))||_{\text{MM}}.$$

The above can be written as a distance between elements $\text{alr}_0(\mathbf{x})$ and $\text{alr}_0(\mathbf{y})$ of $\mathbb{R}^d$ by introducing on $\mathbb{R}^d$ the modified variation norm $||.||_{\text{MM}_0}$ (compare with (29) as,

$$||\mathbf{z}||_{\text{MM}_0} := \max(0, z_1, \ldots, z_d) - \min(0, z_1, \ldots, z_d), \quad \mathbf{z} \in \mathbb{R}^d.$$

It is readily checked that $||.||_{\text{MM}_0}$ is a norm (and not only a pseudo-norm) on $\mathbb{R}^d$, since $||\mathbf{z}||_{\text{MM}_0} = 0$ implies

$$\begin{pmatrix} 0 \\ \mathbf{z} \end{pmatrix} = \mu \begin{pmatrix} 1 \\ \mathbf{1} \end{pmatrix} \Rightarrow \mu = 0 \Rightarrow \mathbf{z} = \mathbf{0}.$$

Eventually, one has that

$$d_H([\mathbf{x}]_+, [\mathbf{y}]_+) = ||\text{alr}_0(\mathbf{x}) - \text{alr}_0(\mathbf{y})||_{\text{MM}_0},$$

having thus proved the following:

COROLLARY 5.6. *alr is a bijective isometry between $(\mathbb{P}^d_{++}, d_H)$ and $(\mathbb{R}^d, ||.||_{MM_0})$.*

The following figure illustrates the results and constructions established in this subsection.

[14] A bijective distance-preserving map.

[15] The norm corresponding to d_H on $\mathbb{P}^d_{++}$ is obviously $||[\mathbf{x}]_+||_H := d_H([\mathbf{x}]_+, [\mathbf{1}]_+)$.

$$
\begin{tikzcd}[column sep=large, row sep=large]
{[\mathbf{x}]_+ \in (\mathbb{P}^d_{++}, d_H)} \arrow[r, bend left, "\mathbf{ln}"] \arrow[d, bend left, "\mathcal{C}=[.]_1"] & {[\ln \boldsymbol{x}]_\sim \in (\mathbb{R}^{d+1}/\sim, ||.||_{\mathrm{MM}})} \arrow[l, "\mathbf{exp}"] \arrow[d, bend left, "\zeta"] \\
{\mathbf{x} \in (\Delta^d_{++}, d_H)} \arrow[u, bend left, "{[.]_+}"] \arrow[r, bend right, "\mathrm{alr}"'] & {\mathrm{alr}_0(\mathbf{x}) \in (\mathbb{R}^d, ||.||_{\mathrm{MM}_0})} \arrow[l] \arrow[u, bend left]
\end{tikzcd}
$$

FIG 13. *Bijective isometry of the positive CoDa space* $(\mathbb{P}^d_{++}, d_H)$ *with* $(\mathbb{R}^d, ||.||_{MM_0})$

REMARK 7. We have thus shown that alr is indeed a (bijective) isometry, when $\mathbb{R}^d$ is endowed with the correct metric. This is a noticeable feature, which goes against the prevailing conceptions regarding alr in the CoDa literature. Indeed, [45] state that "the essential problem with alr coordinates is the non-isometric character of this transformation", when the Euclidean distance $||.||_2$ is used in the alr coordinates domain $\mathbb{R}^d$. To that regard, the standard Euclidean distance used in the classical approach in the CoDa literature does not seem to be well suited.

5.4. *About geodesics in the CoDa space with Hilbert distance.*

We discuss the remarkable fact that straight line segments, in both the affine and Aitchison's log coordinates, are geodesics for the Hilbert metric, i.e. length minimising curves. This gives some insight with the manifold approach of Information geometry and vindicate the usefulness of using Hilbert metric to measure distances on the space of (discrete) distributions.

5.4.1. e *and* m *straight line segments of Information Geometry.*

A CoDa element $[\mathbf{p}]_+$ can be thought of as the distribution of a discrete random variable, and thus CoDa can also be studied from the point of view of Information Geometry ([24]). Information Geometry ([5], [6], [24]) views the space of probability measures as a manifold, and defines special systems of coordinates to describe it. This manifold is then endowed with a Riemannian metric and a dually flat structure. We briefly explain the basics below and connects with the projective approach.

Let $\mathbf{X}$ be a discrete random variable over, say, $\{0, 1, \dots, d\}$. Its distribution is given by

$$p(x) := P(X = x) = \sum_{i=0}^{d} p_i \delta_i(x), \tag{31}$$

where $\delta_i(x) = \mathbb{1}_{i=x}$ is the Kronecker function, and $\mathbf{p} = (p_0, \dots, p_d) \in \Delta^d_+$ is a constrained vector of the probability simplex. (31) expresses the distribution of X as a mixture of the Dirac distributions δ_i, so $\mathbf{p}$ are called the $m-$coordinates of the distribution (m for mixture). Thus, $m-$coordinates in Information Geometry corresponds to the affine representations in the simplex (7).

The discrete distribution can also be expressed as an exponential family: since $1 = \sum_{i=0}^{d} \delta_i(x)$,

$$\ln p(\mathbf{x}) = \sum_{i=0}^{d} (\ln p_i)\, \delta_i(x) = \sum_{i=1}^{d} \ln(p_i/p_0)\, \delta_i(x) + \ln p_0.$$

Hence, $p(x)$ writes as an exponential family as

$$p(x) = p(x;\boldsymbol{\theta}) := \exp\Big(\sum_{i=1}^{d} \theta_i \delta_i(x) - Z(\theta)\Big),$$

where $\theta_i = \ln(p_i/p_0)$, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$, and

$$Z(\boldsymbol{\theta}) = -\ln p_0 = \ln\left(1 + \sum_{i=1}^{d} e^{\theta_i}\right)$$

is the cumulant generating function. Thus, the $\boldsymbol{\theta}$ coordinates, viz. the natural parameter in the exponential family, is another system of coordinates, called the $e-$coordinates in Information Geometry (e for exponential). $e-$ coordinates correspond to the alr transform of the CoDa literature,

$$\boldsymbol{\theta} = \ln \mathbf{p}/p_0 = \mathrm{alr}_0(\mathbf{p}).$$

(Note that one can dispense with introducing the cumulant function, by mixing m and e coordinates, and writing $p(x;\boldsymbol{\theta}) := p_0 \exp\left(\sum_{i=1}^{d} \theta_i \delta_i(x)\right)$.)

Each coordinate system, $\mathbf{p}$ and $\boldsymbol{\theta}$, defines a linear structure on the space of discrete probability measures, inducing two notions of straight line segments (called geodesics in Information Geometry). Linearity is obtained by simply declaring that the corresponding coordinate system is affine, and the segments are obtained by linear interpolation of the coordinates. Thus, the mixture coordinate system defines a $m-$straight line segment $\gamma^m : [0,1] \to \Delta^d_+$ connecting two distributions $\mathbf{p}$ and $\mathbf{q}$ by taking the linear interpolation of the two distributions,

$$\gamma^m(t) := (1-t)\mathbf{p} + t\mathbf{q}.$$

Geometrically, γ^m is a straight line segment in the simplex Δ^d_+. Similarly, the exponential coordinate system defines a $e-$straight line segment $\gamma^e : [0,1] \to \Delta^d_+$ connecting two distributions $p(x;\boldsymbol{\theta_p})$ and $p(x;\boldsymbol{\theta_q})$ by linearly interpolating the natural parameter $\boldsymbol{\theta}$ as

$$\boldsymbol{\theta}(t) := (1-t)\boldsymbol{\theta_p} + t\boldsymbol{\theta_q}.$$

Thus, by taking the logarithm, it corresponds to a linear interpolation of the two distributions in the logarithmic (alr) scale: such e-straight segment writes

$$\gamma^e(t) := p(x, \boldsymbol{\theta}(t)),$$

or, more explicitly,

$$\mathrm{alr}_0(\gamma^e(t)) = (1-t)\mathrm{alr}_0(\mathbf{p}) + t\,\mathrm{alr}_0(\mathbf{q})$$

Hence, in the CoDa framework, this $e-$segment corresponds to a vector segment in Aitchison's geometry, i.e. w.r.t. $\oplus$, $\odot$ operations,

$$[\gamma^e(t)]_+ = (1-t) \odot [\mathbf{p}]_+ \oplus t \odot [\mathbf{q}]_+.$$

From the projective viewpoint, it can also be described as

$$[\ln \gamma^e(t)]_\sim = (1-t) \,\tilde{\cdot}\, [\mathbf{ln}\,\boldsymbol{p}]_\sim \,\tilde{+}\, t \,\tilde{\cdot}\, [\mathbf{ln}\,\boldsymbol{q}]_\sim.$$

Geometrically, the latter expresses the $e-$segment as a linear interpolation of the lines $[\mathbf{ln}\,\boldsymbol{p}]_\sim$ and $[\mathbf{ln}\,\boldsymbol{q}]_\sim$, as in Figure 7.

REMARK 8. In addition, Information Geometry shows that these two coordinate systems are dual, in the sense that the negative entropy $\phi(\mathbf{p}) := \sum_{i=0}^{d} p_i \ln p_i$ is a convex function of $\mathbf{p}$, and that $\boldsymbol{\theta}$ is the gradient of ϕ, $\theta = \nabla\phi(\mathbf{p})$: $\boldsymbol{\theta}$ and $\mathbf{p}$ are dual in the sense of the classical Fenchel-Legendre transform of convex analysis. The Bregman divergence based on ϕ is the Kullback-Leibler divergence and can be written in terms of ϕ, its convex conjugate ϕ^* (which is the cumulant generating function) and a mix of the e and m coordinates. This gives, among other, a Riemannian metric (the Fisher information matrix), and generalized Pythagorean and projection theorems. More generally, a divergence function endows a Riemannian metric and a pair of dually coupled affine connections on the space of probability measures. See e.g. [5] for more details.)

5.4.2. *Information geometry geodesics w.r.t. Hilbert distance..* Now, let us study these $e-$ and $m-$segments from the (Hilbert) metric geometry viewpoint (see e.g. [15]). Recall that a geodesic path γ in a metric space (M, d) connecting $x, y \in M$ is a (continuous) map $\gamma : [0,1] \to M$, such that

$$d(\gamma(s), \gamma(t)) = |s-t| d(x,y) \quad \text{for all } s, t \in I.$$

(sometimes the multiplicative constant $d(x,y)$ on the right hand side is normalised to 1). A metric space is said to be a geodesic space if for each $x, y \in M$, there exists a geodesic path $\gamma : [0,1] \to M$ joining x and y, i.e. $\gamma(0) = x$ and $\gamma(1) = y$.

Regarding the CoDa space $(\mathbb{P}^d_+, d_H)$ as a metric space, it is a geodesic space. This is obvious when one regards $(\mathbb{P}^d_+, \oplus, \odot, ||.||_H)$ as a normed vector space, i.e. one considers $e-$segments. However, the geodesics are not unique. In particular, $m-$segments, i.e. straight-line segments in the simplex representation Δ^d_{++} are also geodesics for Hilbert metric. This was Hilbert's original motivation for the definition of his metric. This is the content of the next Proposition.

PROPOSITION 5.7. *The $m-$ and $e-$ segments are geodesics for the Hilbert metric.*

6. Differentiability properties and smooth approximations of Hilbert distance. In view of statistical applications, which are often formulated as optimisation problems, we study in this section the differentiability properties of Hilbert projective distance. This leads us to introduce a smoothed approximation of Hilbert's distance which is shown to satisfy many (but not all) properties of the original Hilbert's distance.

6.1. *Differentiability properties of Hilbert distance.*

6.1.1. *Directional differentiability.* For a function $f : \mathbb{R}^{d+1} \to \mathbb{R}$, recall that f is said to admit a directional derivative of f at $\mathbf{x}$ in the direction $\mathbf{v} \in \mathbb{R}^{d+1}$ if the limit

$$f'(\mathbf{x}, \mathbf{v}) := \lim_{\alpha \downarrow 0} \frac{1}{\alpha}[f(\mathbf{x} + \alpha\mathbf{v}) - f(\mathbf{x})]$$

exists, and f is said to be directionally differentiable at $\mathbf{x}$ if it admits a directional derivative at $\mathbf{x}$ for every direction $\mathbf{v}$. The following Proposition shows that Hilbert's metric, considered as a function on $(\mathbb{R}^{d+1}_{++}, +, .)$ is directionally differentiable.

PROPOSITION 6.1. *For some fixed $\mathbf{y} \in \mathbb{R}^{d+1}_{++}$, consider the function $f : \mathbb{R}^{d+1}_{++} \ni \mathbf{x} \mapsto d_H(\mathbf{x}, \mathbf{y})$. Then, for all $\mathbf{x} \in \mathbb{R}^{d+1}_{++}$, f is directionally differentiable with directional derivative*

$$f'(\mathbf{x}, \mathbf{v}) = \max_{i \in U(\mathbf{x})} v_i/x_i - \min_{i \in L(\mathbf{x})} v_i/x_i, \quad \mathbf{v} \in \mathbb{R}^{d+1}, \tag{32}$$

where $U(\mathbf{x}) = \{i = 0, \ldots, d : \ln(x_i/y_i) = \max_j \ln(x_j/y_j)\}$ and $L(\mathbf{x}) = \{i = 0, \ldots, d : \ln(x_i/y_i) = \min_j \ln(x_j/y_j)\}$.

REMARK 9. The sets $U(\mathbf{x})$ and $L(\mathbf{x})$ of (32) remain invariant by rescaling $\mathbf{x} \leftarrow \lambda\mathbf{x}$, $\mathbf{y} \leftarrow \mu\mathbf{y}$, for $\lambda, \mu > 0$. However,

$$f'(\lambda x, \mathbf{v}) = \frac{1}{\lambda} f'(\mathbf{x}, \mathbf{v}).$$

Thus, the value of the directional derivative f' depends on the choice of the representative $\mathbf{x} \in \mathbb{R}^{d+1}_{++}$. Note, however, that $f'(\lambda\mathbf{x}, \lambda v) = f'(\mathbf{x}, \mathbf{v})$, so that f' is scale invariant, when both the point $\mathbf{x}$ and direction $\mathbf{v}$ are rescaled.

6.1.2. *Log-directional differentiability.* In view of the vector space structure of $(\mathbb{P}^d_{++}, \oplus, \odot)$ (see Proposition 5.2), and/or the linear structure of $e-$segments of Section 5.4, it also makes sense to define directional differentiability of a mapping $f : \mathbb{P}^d_{++} \to \mathbb{R}$ w.r.t. the $\oplus, \odot$ operations, viz. w.r.t. log coordinates: following [44] Chapter 13, let us say that f admits a log-directional derivative at $[\mathbf{x}]_+$ in the direction $[\mathbf{v}]_+$, if the limit

$$f'^{\oplus}([\mathbf{x}]_+, [\mathbf{v}]_+) := \lim_{\alpha \downarrow 0} \frac{f([\mathbf{x}]_+ \oplus \alpha \odot [\mathbf{v}]_+) - f([\mathbf{x}]_+)}{\alpha}$$

exists, and f is said to be log-directionally differentiable at $[\mathbf{x}]_+$ if it admits a log-directional derivative for every direction $[\mathbf{v}]_+$.

One can then gives the log version of Proposition 6.1:

PROPOSITION 6.2. *For some fixed $[\mathbf{y}]_+ \in \mathbb{P}^d_{++}$, consider the function $f : \mathbb{P}^d_{++} \ni [\mathbf{x}]_+ \mapsto d_H([\mathbf{x}]_+, [\mathbf{y}]_+)$. Then, for all $[\mathbf{x}]_+ \in \mathbb{P}^d_{++}$, f is log-differentiable with log-directional derivative*

$$f'^{\oplus}([\mathbf{x}]_+, [\mathbf{v}]_+) = \max_{i \in U([\mathbf{x}]_+)} \ln v_i - \min_{i \in L([\mathbf{x}]_+)} \ln v_i, \quad \mathbf{v} \in \mathbb{R}^{d+1}, \tag{33}$$

where $U([\mathbf{x}]_+) = \{i = 0, \ldots, d : \ln(x_i/y_i) = \max_j \ln(x_j/y_j)\}$ and $L([\mathbf{x}]_+) = \{i = 0, \ldots, d : \ln(x_i/y_i) = \min_j \ln(x_j/y_j)\}$.

REMARK 10 (Bouligand differentiability). In fact, Hilbert's distance satisfy a stronger form of directional differentiability called Bouligand differentiability, which was introduced by Robinson in [49] for locally Lipschitz continuous functions, and studied in [52]. In short, a function is B-differentiable if it is directionally differentiable and if in addition, the directional derivative is a first-order approximation of f. Compared to the usual (Fréchet) derivative, one does not require that the derivative be linear in the direction, simply positively homogeneous. The first-order approximation condition in the definition, is the key to have an operative concept (in particular the chain rule) for optimisation purposes. The concept of B-differentiability avoid using set-valued analysis (unlike other form of generalized derivatives) and apply to the large class of non-smooth functions, in particular piece-wise differentiable functions (see e.g. Proposition 4.1.3 in [52]). Since Hilbert distance is the composition of the $\max$ operation and smooth functions, it is piece-wise differentiable hence B-differentiable.

6.2. *Smooth approximations of Hilbert's metric.*

6.2.1. *Definition.* The non-differentiability of Hilbert's metric comes from the $\max$ operation in the expressions (23) or (24) defining it. This suggests to use a smooth approximation of the maximum in order to obtain a differentiable approximate proxy of the Hilbert metric. In particular, one can use the log-sum-exp function (see e.g. [13]), defined for $\mathbf{x} = (x_1, \ldots, x_m) \in \mathbb{R}^m$ as

$$\mathrm{lse}(\mathbf{x}) = \ln\left(\sum_{i=1}^{m} e^{x_i}\right)$$

or its rescaled version $\mathrm{lse}_c(\mathbf{x}) := c^{-1}\mathrm{lse}(c\mathbf{x})$, where $c > 0$ controls the degree of approximation of $\max$ by lse_c, see Lemma 6.4 below. In view of (24), the smooth approximate version of Hilbert's metric is then defined as follows:

DEFINITION 6.3 (Smooth Hilbert distance). For $[\mathbf{x}]_+, [\mathbf{y}]_+, \in \mathbb{P}^d_{++}$ and $c > 0$, the smooth Hilbert approximate distance is defined as

$$d_{H,c}([\mathbf{x}]_+, [\mathbf{y}]_+) := \mathrm{lse}_c\left(\ldots, \ln\left(\frac{x_i y_j}{x_j y_i}\right), \ldots\right), \quad 0 \le i, j \le d, \tag{34}$$

$$= \frac{1}{c}\ln\left(\sum_{0\le i,j\le d}\left(\frac{x_i y_j}{x_j y_i}\right)^c\right)$$

where $\mathbf{x}, \mathbf{y}$ are any representatives in $\mathbb{R}^{d+1}_{++}$ of $[\mathbf{x}]_+, [\mathbf{y}]_+$.

It is readily checked that $d_{H,c}$ is well defined, irrespective of the chosen representatives. Note that we could have considered several variants, as explained in the following remark.

REMARK 11 (Variants). i) One could have used equation (23) instead of (24) as the formula of Hilbert's metric and approximate the $\max$ inside the logarithm. However, from an approximation viewpoint it is better to first take the logarithms and then approximate the $\max$, as the reverse would propagate the approximation error through a function with unbounded derivative. Also, we checked that the structural properties of the resulting Hilbert's metric proxy would be less satisfying (for example we do not have scalar equivariance as for (34), see Proposition 6.5 below).

ii) For given $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{d+1}_{++}$, set $R(i,j) := \frac{x_i y_j}{x_j y_i}$ the odds-ratio, so that Hilbert's distance writes $d_H([\mathbf{x}]_+, [\mathbf{y}]_+) = \max_{i,j} \ln R(i,j)$. Since $R(i,i) = 1$ and $\ln R(i,j) = -\ln R(j,i)$, it is clear that one can restrict the $\max$ operation in the definition of Hilbert's metric to distinct pairs $i \ne j$, i.e. $d_H([\mathbf{x}]_+, [\mathbf{y}]_+) = \max_{i\ne j} \ln R(i,j)$. In turn, applying the log-sum-exp trick to distinct pairs yields a variant $\widetilde{d_{H,c}}$ of Definition (34), which writes as

$$\widetilde{d_{H,c}}([\mathbf{x}]_+, [\mathbf{y}]_+) := \frac{1}{c}\ln\left(\sum_{i<j}\left(R(i,j)^c + \frac{1}{R(i,j)^c}\right)\right).$$

Similar reasoning shows that the proposed smooth proxy of Hilbert's metric (34) can also be written as,

$$d_{H,c}([\mathbf{x}]_+, [\mathbf{y}]_+) := \frac{1}{c}\ln\left(d + 1 + \sum_{i<j}\left(R(i,j)^c + \frac{1}{R(i,j)^c}\right)\right)$$

$$= \frac{1}{c}\ln\left(d + 1 + 2\sum_{i<j}\cosh(c\ln R(i,j))\right).$$

Note that adding a constant in the sum of exponentials yields a strictly convex approximation of the $\max$, as shown in Appendix A of [39], where the authors consider, instead of lse, $\mathrm{lse}^+(x_1, \ldots, x_m) := \log(1 + \sum_{i=1}^m e_i^x)$ as a strictly convex smooth approximation of the $\max$.

6.2.2. *Properties of the smoothed Hilbert metric.* We collect in the following lemma some elementary properties of the rescaled lse function, which will be useful for studying the properties of $d_{H,c}$:

LEMMA 6.4. i) *Approximation inequality: for* $\mathbf{x} = (x_1, \ldots, x_m)$,

$$\max(x_i) \leq lse_c(\mathbf{x}) \leq \frac{\ln m}{c} + \max(x_i)$$

ii) *lse_c is convex and strictly increasing w.r.t. to each coordinate.*
iii) *Sub-additivity: $lse_c(\mathbf{x} + \mathbf{y}) \leq lse_c(\mathbf{x}) + lse_c(\mathbf{y})$*
iv) *Positive scalar equivariance: $lse_c(\lambda \mathbf{x}) = \lambda \times lse_{\lambda c}(\mathbf{x})$, for $\lambda > 0$.*

In turn, the smooth Hilbert approximate distance has the following properties:

PROPOSITION 6.5. i) *Smooth approximation of Hilbert's metric from above:*

$$d_H([\mathbf{x}]_+, [\mathbf{y}]_+) \leq d_{H,c}([\mathbf{x}]_+, [\mathbf{y}]_+) \leq d_H([\mathbf{x}]_+, [\mathbf{y}]_+) + \frac{2\ln(d+1)}{c}$$

ii) *Distance-like properties:*
  a) *Non-negativity and "almost separability":*

$$d_{H,c}([\mathbf{x}]_+, [\mathbf{y}]_+) \geq d_{H,c}([\mathbf{x}]_+, [\mathbf{x}]_+) = \frac{2\ln(d+1)}{c} \geq 0$$

  b) *Symmetry: $d_{H,c}([\mathbf{x}]_+, [\mathbf{y}]_+) = d_{H,c}([\mathbf{y}]_+, [\mathbf{x}]_+)$.*
  c) *Triangular inequality:*

$$d_{H,c}([\mathbf{x}]_+, [\mathbf{z}]_+) \leq d_{H,c}([\mathbf{x}]_+, [\mathbf{y}]_+) + d_{H,c}([\mathbf{y}]_+, [\mathbf{z}]_+)$$

iii) *Partial compatibility with the vector space structure $\oplus, \odot$:*
  a) *for $[\mathbf{x}]_+, [\mathbf{y}]_+, [\mathbf{p}]_+ \in \mathbb{P}^d_{++}$,*

$$d_{H,c}([\mathbf{x}]_+ \oplus [\mathbf{p}]_+, [\mathbf{y}]_+ \oplus [\mathbf{p}]_+) = d_{H,c}([\mathbf{x}]_+, [\mathbf{y}]_+)$$

  b) *Almost scalar equivariance for $\lambda \in \mathbb{R}^*$,*

$$d_{H,c}(\lambda \odot [\mathbf{x}]_+, \lambda \odot [\mathbf{y}]_+) = |\lambda| d_{H,c|\lambda|}([\mathbf{x}]_+, [\mathbf{y}]_+)$$

  c) *Strict convexity w.r.t $\oplus, \odot$: for fixed $[\mathbf{y}]_+$, the mapping $[\mathbf{x}]_+ \to d_{H,c}([\mathbf{x}]_+, [\mathbf{y}]_+)$ is strictly convex w.r.t. the vector operations $\oplus, \odot$.*

iv) *Permutation invariance:*
  If $\mathbf{s}$ is a permutation of $\{0, 1, \ldots, d\}$, then

$$d_{H,c}([\mathbf{x_s}]_+, [\mathbf{y_s}]_+) = d_{H,c}([\mathbf{x}]_+, [\mathbf{y}]_+).$$

7. Statistical applications: Fréchet-Hilbert means, Gaussian-Hilbert distributions and nonparametric regression. The previous sections have explained how the positive CoDa space $\mathbb{P}^d_{++}$, equipped with Hilbert's projective metric d_H, is now a metric space. This section eventually proposes statistical applications based on such comparisons of distances of CoDa points.

7.1. *Intrinsic measures of location and spread based on Hilbert distance.* We begin our statistical applications by considering Hilbert's metric analogues of the basic descriptive statistics like the mean, median and variance. Following [47], we first explain the different notions of extrinsic and intrinsic means which can be defined on a general topological space. Specialising to the CoDa space, Aitchison's mean fits into this framework as an extrinsic mean, and other such means could be considered. Switching to the intrinsic viewpoint, one can also define a Fréchet mean, median and variance based on Hilbert distance. Existence and unicity are discussed.

7.1.1. *Extrinsic means on a general topological space.* Let $(\Omega, \mathcal{A}, P)$ be a probability space. Consider a random element $X : \Omega \to \mathcal{X}$ on some topological space $\mathcal{X}$. The first basic descriptive statistic of the distribution $P^{\mathbf{X}}$ of X is provided by a notion of center/mean. For a general space $\mathcal{X}$, the mean EX is not defined since $\mathcal{X}$ does not carry the structure of a linear space. (In particular, there is no notion of arithmetic averaging so that $\int_{\mathcal{X}} xP^X(dx)$ does not even make sense.)

When there is an embedding (i.e. an injective map) $\phi : \mathcal{X} \hookrightarrow \mathbb{R}^m$ into an ambient Euclidean space $\mathbb{R}^m$, $\mathcal{X}$ can be identified with a subset $\phi(\mathcal{X})$ of $\mathbb{R}^m$. The latter space both has a linear structure (i.e. is a vector space) and a metric structure (is endowed with the Euclidean distance). Thus, one can define the *ambient mean* (w.r.t. ϕ), as

$$\mathbf{a}(X) := E(\phi(X)) \in \mathbb{R}^m,$$

and the latter expectation can be defined,

i) either algebraically, as the vector integral

$$\mathbf{a}(X) = \int_{\mathbb{R}^m} \mathbf{y} P^{\mathbf{Y}}(d\mathbf{y}),$$

where $P^{\mathbf{Y}}$ is the multivariate distribution of $\mathbf{Y} = \phi(X) \in \mathbb{R}^m$;

ii) or metrically, as the minimizer of the sum of squares distances

$$\mathbf{a}(X) = \arg\min_{\mathbf{z} \in \mathbb{R}^m} \int_{\mathbb{R}^m} d^2(\mathbf{y}, \mathbf{z}) P^{\mathbf{Y}} d\mathbf{y}, \tag{35}$$

where d is a distance on the ambient space $\mathbb{R}^m$. When d is the usual Euclidean distance, both definitions coincide.

For a sample $X_1, \ldots, X_n \in \mathcal{X}$, the empirical ambient mean $\overline{\mathbf{a}}_n := \overline{\mathbf{a}}_n(X)$ is obtained as above by replacing $P^{\mathbf{Y}}$ by the corresponding empirical measure. Hence, i) yields a definition of the ambient mean via the vector space structure of $\mathbb{R}^m$, as the arithmetic average $\overline{\mathbf{a}}_n = n^{-1} \sum_{i=1}^n \phi(X_i)$, or equivalently as the unique point such that the residuals $(\phi(X_1) - \overline{\mathbf{a}}_n) + \ldots + (\phi(X_n) - \overline{\mathbf{a}}_n)$ sum to zero, while ii) is based solely on the metric space structure of $\mathbb{R}^m$, as $\overline{\mathbf{a}}_n = \arg\min_{\mathbf{z} \in \mathbb{R}^m} \sum_{i=1}^n d^2(\phi(X_i), \mathbf{z})$.

If ϕ is bijective, then $\mathbf{a}(X)$ can be pulled back to $\mathcal{X}$, so that one defines the *extrinsic mean* of X on $\mathcal{X}$ based on ϕ, as

$$E^{\phi}(X) := \phi^{-1}(\mathbf{a}(X)) = \phi^{-1}(E\,\phi(X)). \tag{36}$$

If ϕ is not surjective, it may happen that $\mathbf{a}(X) \notin \phi(S)$, so that one can not directly pull-back the ambient mean to $\mathcal{X}$. However, in such a case, The Euclidean structure of the ambient space allows to orthogonally project the ambient mean $\mathbf{a}(X)$ to $\phi(\mathcal{X})$: Define $\boldsymbol{\mu}(X) \in \phi(\mathcal{X}) \subset \mathbb{R}^d$ as

$$\boldsymbol{\mu}(X) = \arg\min_{\mathbf{z} \in \phi(\mathcal{X})} ||\mathbf{z} - \mathbf{a}(X)||. \tag{37}$$

Provided that $\boldsymbol{\mu}(X)$ exists and is unique, its pull-back to the original space $\mathcal{X}$ is

$$E^{\phi}(X) := \phi^{-1}(\boldsymbol{\mu}(X)) = \phi^{-1}(\arg\min_{\mathbf{z} \in \phi(\mathcal{X})} ||\mathbf{z} - E\,\phi(X)||),$$

In general, $\boldsymbol{\mu}(X)$ (a fortiori $E^{\phi}(X)$) may not exists or may be not be single-valued. Figure 14 illustrates the conceptual procedure.

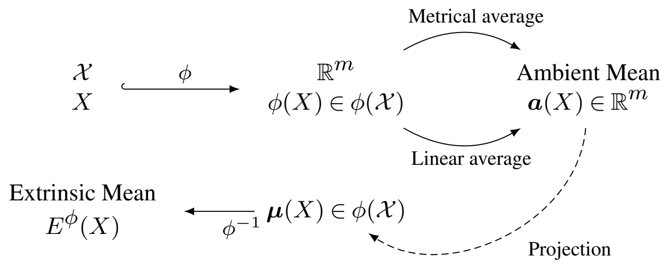

FIG 14. *Extrinsic mean on a topological space $\mathcal{X}$ based on an embedding $\phi : \mathcal{X} \hookrightarrow \mathbb{R}^m$.*

7.1.2. *Aitchison's mean as an extrinsic mean.* For positive Coda in $\mathbb{P}^d_{++}$, Aitchison's approach follows such an extrinsic definition of the mean, using clr as embedding map ϕ: for $\mathbf{X} \in \Delta^d_{++}$, a positive CoDa random element represented in the simplex, since $\mathcal{H}_{\mathbf{1}} \subset \mathbb{R}^{d+1}$ (as defined in (3.3) is a vector subspace, $E(\text{clr}(X)) \in \mathcal{H}_{\mathbf{1}}$. Moreover, since clr is an isomorphism between Δ^d_{++} and $\mathcal{H}_{\mathbf{1}}$, Aitchison's simplex mean can then be defined via (36) as

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

It is then shown that $E^{\oplus}(X)$ is equal to $\text{ilr}_V^{-1}(E(\text{ilr}_V(X)))$, for any contrast matrix V. The empirical version is simply the arithmetic average $n^{-1} \odot (\mathbf{x}_1 \oplus \ldots \oplus \mathbf{x}_n)$, which corresponds to the closed geometric mean. Aitchison's CoDa mean is thus a quasi-arithmetic or generalised mean (see [14]). The extrinsic character of Aitchison's mean comes from the special choice of log-coordinates and the use of the Euclidean distance to compute the mean in the log-coordinates spaces.

The use of such mean has been criticized in the literature [59], [50]: geodesics corresponding to Aitchison's distance are curved, which may not be suitable for some datasets where Aitchison's mean may lie outside the data cloud. Moreover, the $\oplus$ operation corresponds more to a change of unit (see e.g. [61]) than, say mixing cocktails into a single one ([50]), and thus may not be appropriate when one wants to aggregate compositions to obtain an "average" composition. Eventually, Remark 7 hinted that the Euclidean distance might not be an "adequate" metric for CoDa.

Let us just mention, that, in view of the embeddings studied in section 5, one can consider other kinds of extrinsic means than Aitchison's mean. In particular, one can take the embedding map $\Psi : \mathbb{P}^d_{++} \hookrightarrow \mathbb{R}^{d(d+1)/2}$ defined in (27). This time, obtaining an extrinsic mean $E^{\Psi}([\mathbf{X}]_+)$ on $\mathbb{P}^d_{++}$ as in Figure 14, based on Ψ, possibly requires the additional step (37) of projecting the ambient mean $\mathbf{a}([\mathbf{X}]_+)$ to $\boldsymbol{\mu}([\mathbf{X}]_+) \in \Psi(\mathbb{P}^d_{++})$. We shall not pursue further, as we will focus on intrinsic means, to be defined thereafter.

7.1.3. *Intrinsic Fréchet mean based on Hilbert distance.* The above approach for defining a generalised notion of mean on a general topological space $\mathcal{X}$ is extrinsic in the sense that it is depends on the auxiliary ambient space, through the embedding map ϕ and the distance d chosen. When $(\mathcal{X}, d_{\mathcal{X}})$ is a metric space with intrinsic distance $d_{\mathcal{X}}$, equation (35)'s metrical characterisation of the mean as a solution of a least-squares problem suggest a more intrinsic definition of the mean. Following [26], the Fréchet intrinsic mean on the metric space $(\mathcal{X}, d_{\mathcal{X}})$ is directly defined as

$$M^F(X) := \arg \inf_{m \in \mathcal{X}} \int_{\mathcal{X}} d^2_{\mathcal{X}}(m, x) P^X(dx).$$

In general, $M^F(X)$, if it exists, is a set, and the question of uniqueness is often non-trivial (see e.g. [18], [1]). The intrinsic Fréchet variance is then the corresponding average square distance to the Fréchet mean, viz.

$$Var^F(X) := \int_{\mathcal{X}} d^2_{\mathcal{X}}(M^F(X), x) P^X(dx) = E\,[d^2_{\mathcal{X}}(M^F(X), X)].$$

These formulas are thus the natural generalisation of the mean and variance of Euclidean vectors, with the Euclidean distance replaced by the distance $d_{\mathcal{X}}$, of $\mathcal{X}$.

Applied to the CoDa space P^d_{++}, and in view of the intrinsic character of Hilbert's projective metric of Sections 4-6, these considerations suggest to define measures of location and scatter for CoDa as follows:

DEFINITION 7.1 (Intrinsic Fréchet-Hilbert mean and variance). For $[\mathbf{X}]_+ \in \mathbb{P}^d_{++}$ a random element with distribution μ, set the Fréchet energy as

$$\mathcal{F}_\mu([\mathbf{m}]_+) := E\left(d^2_H([\mathbf{X}]_+, [\mathbf{m}]_+)\right).$$

Then, the intrinsic Fréchet-Hilbert mean of $[\mathbf{X}]_+$ is defined as

$$[\mathbf{m}^H]_+ := \arg \inf_{[\mathbf{m}]_+ \in \mathbb{P}^d_{++}} \mathcal{F}_\mu([\mathbf{m}]_+), \tag{38}$$

and the Fréchet-Hilbert variance as

$$\mathrm{Var}^H([\mathbf{X}]_+) := \mathcal{F}_\mu([\mathbf{m}^H]_+)),$$

provided $[\mathbf{m}^H]_+$ exists and is unique.

The empirical version, for a sample $[\mathbf{X}_1]_+, \ldots, [\mathbf{X}_n]_+$, writes

$$[\overline{\mathbf{m}}^H_n]_+ := \arg \inf_{[\mathbf{m}]_+ \in \mathbb{P}^d_{++}} \sum_{i=1}^n d^2_H([\mathbf{X}_i]_+, [\mathbf{m}]_+),$$

with corresponding empirical variance

$$\overline{\mathrm{Var}}^H_n := \frac{1}{n} \sum_{i=1}^n d^2_H([\mathbf{X}_i]_+, [\overline{\mathbf{m}}^H_n]_+).$$

In analogy to the Euclidean case, one can also define a metric/geometric/Fermat-Weber-Torricelli median, by setting

$$\mathcal{L}_\mu([\mathbf{m}]_+) := E\left(d_H([\mathbf{X}]_+, [\mathbf{m}]_+)\right),$$

and defining the intrinsic Fréchet-Hilbert median as

$$[\mathtt{Med}^H]_+ := \arg \inf_{[\mathbf{m}]_+ \in \mathbb{P}^d_{++}} \mathcal{L}_\mu([\mathbf{m}]_+),$$

and similarly for the empirical case. The corresponding median absolute deviation in Hilbert distance is $E(d_H([\mathbf{X}]_+, [\mathtt{Med}^H]_+))$.

7.1.4. *Existence and Characterisation of Fréchet-Hilbert means and medians.* We say that the distribution μ of $[\mathbf{X}]_+$ has finite second Hilbert moment, if $\mathcal{F}_\mu([\mathbf{m}]_+) < \infty$ for some $[\mathbf{m}]_+ \in \mathbb{P}^d_{++}$. We collect in the following lemma some elementary property of the $\mathcal{F}_\mu$, $\mathcal{L}_\mu$ functionals.

LEMMA 7.2. i) *$\mathcal{L}_\mu$ is $1-$Lipschitz continuous, $\mathcal{F}_\mu$ is Lipschitz on each compact set of $\mathbb{P}^d_{++}$.*

ii) *If μ finite second Hilbert moment, then* $\mathcal{L}_\mu([\mathbf{m}]_+), \mathcal{F}_\mu(\mathbf{m}]_+) < \infty$ *for all* $[\mathbf{m}]_+ \in \mathbb{P}^d_{++}$.
iii) $\mathcal{L}_\mu, \mathcal{F}_\mu$ *are coercive, i.e.* $\mathcal{L}_\mu([\mathbf{m}]_+), |\mathcal{F}_\mu([\mathbf{m}]_+) \to \infty$, *as* $[\mathbf{m}]_+$ *converges to the boundary of* $\mathbb{P}^d_{++}$, *i.e. converges to a point with some zero components.*
iv) $\mathcal{L}_\mu, \mathcal{F}_\mu$ *are convex w.r.t. the vector space operations* $\oplus, \odot$ *of* $\mathbb{P}^d_{++}$.

Existence of the intrinsic Fréchet-Hilbert mean and median then easily ensues:

THEOREM 7.3. *Let the distribution μ of $[\mathbf{X}]_+$ has finite second, resp. first, Hilbert moment. Then, the intrinsic Fréchet-Hilbert mean $[\mathbf{m}^H]_+$, resp. Fréchet-Hilbert median $[\mathtt{Med}^H]_+$, of $[\mathbf{X}]_+$ exist.*

In particular, the empirical Fréchet-Hilbert mean and median always exist. Unlike the Euclidean case, where the squared distance is a smooth, strictly convex function of the two points, the squared Hilbert metric is a non-smooth, convex (w.r.t. $\oplus, \odot$), but not strictly convex function. This implies that unicity is not guaranteed. This issue is also related to the non-unicity of the geodesics w.r.t. Hilbert metric, as exemplified by Proposition 5.7.

From convexity, a local minimum of $\mathcal{F}_\mu$, resp. $\mathcal{L}_\mu$, is a global minimum. From directional differentiability of Hilbert distance (Propositions 6.1 and 6.2), a necessary and sufficient condition for a local minimum of a convex function on a convex set is given by the classical condition of non-negativity of the directional derivative in all directions emanating from the minimum. This gives the following variational inequality characterisation of a Fréchet-Hilbert mean $[\mathbf{m}^H]_+$: $[\mathbf{m}^H]_+$ is a Fréchet-Hilbert mean, if and only if, for all $[\mathbf{m}]_+ \in \mathbb{P}^d_{++}$,

$$\mathcal{F}'^{\oplus}_\mu([\mathbf{m}^H]_+; [\mathbf{m}]_+ - [\mathbf{m}^H]_+) \geq 0,$$

where $F'^{\oplus}_\mu([\mathbf{x}]_+; [\mathbf{v}]_+)$ is the log-directional derivative of the Fréchet functional in the direction $[\mathbf{v}]_+$ at $[\mathbf{x}]_+$. From the additive form of $\mathcal{F}_\mu$ and Proposition 6.2, one could give a more explicit form, but the resulting expression is not particularly tractable, and is thus omitted.

7.1.5. *Smooth Fréchet-Hilbert Mean: a practical surrogate.* Hopefully, replacing d_H in the Fréchet functional of Definition 7.1 with the smooth approximate Hilbert distance $d_{H,c}$ of Definition 6.3 allows to alleviate these issues of non-unicity of the Fréchet-Hilbert mean and medians and non-differentiability of the objective function. Indeed, from Proposition 6.5 iii) (c), it follows that $d_{H,c}$ is strictly convex w.r.t. the vector space operations $\oplus, \odot$. This guarantees the unicity of the Fréchet-Hilbert mean and median based on the smooth Hilbert distance $d_{H,c}$. In addition, the objective is now a convex (w.r.t. $\oplus, \odot$) and differentiable function, and can be computed practically using classical methods of smooth and convex optimisation.

Interestingly, numerical experiments reported in Example 2 below indicates the good behaviour of the smoothed Fréchet Hilbert mean: for toy data located on a line in the simplex, the smoothed Fréchet-Hilbert mean is visually on the line, and thus manages to give an average compatible with the intrinsic geometric structure of the data. See also the extension to nonparametric regression in Section 7.3.

EXAMPLE 2. Figure 15 illustrates the dissimilarity between the different kind of empirical means on a toy example dataset. The raw data consists of 10 points on the (projective) line $x_2 = x_1 + x_0$, with x_0, x_1 i.i.d. uniform on $[0,1]$. Aitchison's mean visually lie outside the line where sits the data. Hilbert's Fréchet mean, computed by a general numerical local minimizer, give a point close, but still not on the line. In addition, the position depends on the starting point in the local minimizer procedure, which illustrates the fact that the Fréchet-Hilbert mean is in general not unique (in the example depicted here, the starting point is $(0.5, 0.5, 0.5)$). The smooth Fréchet-Hilbert mean, with $c = 100$, is visually on the line. In

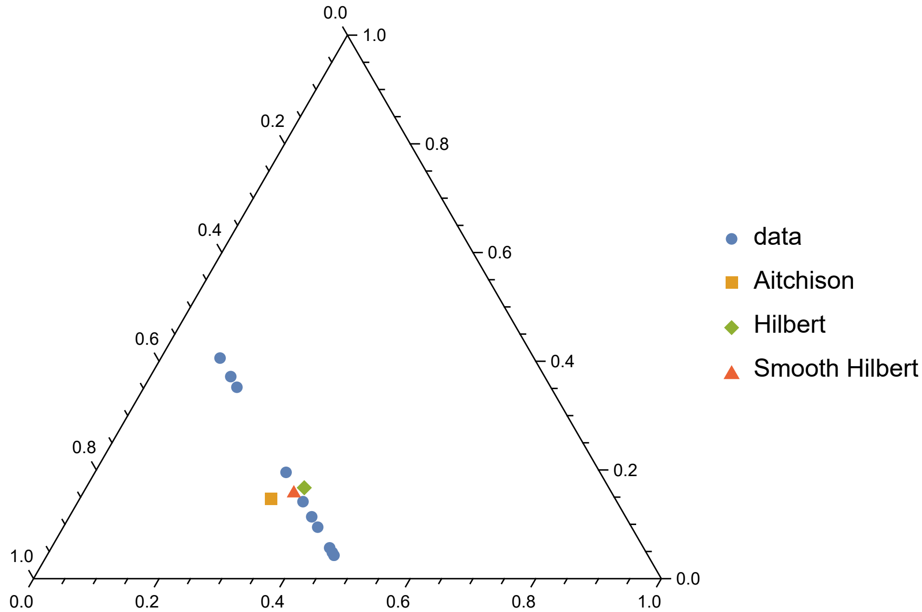

FIG 15. *Comparison of empirical means for* 10 *data points (blue points) on a line. Aitchison's geometric mean (orange square), Hilbert's Fréchet mean (green diamond), Smooth Hilbert's mean (red triangle).*

Aitchison	Hilbert	Smooth Hilbert
0.0555	0.0171	0.0057

TABLE 1

(Euclidan) Distance of the different means to the data line: the smooth Hilbert mean is the closest.

addition to this qualitative assessment, Table 1, computes the distance of the different means to the projective line where the data sits: the smooth Hilbert mean is the closest and thus best captures the intrinsic geometric feature of the data.

7.2. *Hilbert Gaussian/Laplace distribution via Least Squares of Hilbert distances.* Defining an analogue of the Gaussian distribution for CoDa is important to provide a statistical foundation for inference based on Fréchet Mean with respect to Hilbert's projective metric. In this section, we define such an analogue on the positive CoDa space $(\mathbb{P}^d_{++}, d_H)$, based on the statistical characterisation of the Gaussian distribution.

7.2.1. *Gauss' statistical characterisation of the Normal distribution.* Indeed, recall that Gauss' original characterisation of the Gaussian distribution is that, in the observation-with-error linear regression model

$$\mathbf{x} = \boldsymbol{m} + \boldsymbol{\epsilon},$$

where $\boldsymbol{m} \in \mathbb{R}^d$ is the value to be estimated and $\mathbf{x} \in \mathbb{R}^d$ the observation, the distribution of measurement errors $\boldsymbol{\epsilon}$ is Gaussian if and only if Maximum Likelihood estimation of $\boldsymbol{m}$ is equivalent to the method of Least-Squares. In other words, the empirical mean $n^{-1}\sum_{i=1}^n \mathbf{x}_i$, obtained by minimizing the sum of squares

$$\sum_{i=1}^n d(\mathbf{x}_i, \boldsymbol{m})^2,$$

where $(\mathbf{x}_1,\ldots,\mathbf{x}_n)$ is a sequence of i.i.d. observations and d the usual Euclidean distance, if, and only if $\mathbf{x}$ follows a $\mathcal{N}(\boldsymbol{m},\sigma^2 I_d)$ distribution with density

$$\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{d^2(\mathbf{x}-\boldsymbol{m})}{2\sigma^2}\right).$$

Similarly, the Laplace distribution (m,σ) has density

$$Z(\sigma)\exp\left(-\frac{d(\mathbf{x}-\boldsymbol{m})}{\sigma}\right),$$

where $Z(\sigma)$ is a normalising factor, is associated to the fact that the maximum likelihood estimate of $\boldsymbol{m}$ is the empirical median obtained by minimizing the sum of (non-squared) distances $\sum_{i=1}^n d(\mathbf{x}_i,\boldsymbol{m})$.

7.2.2. *Generalised Hilbert Gaussian distribution.* This statistical characterisation of the Gaussian/Laplace distribution suggest to define the following analogue on the positive CoDa space $(\mathbb{P}^d_{++},d_H)$: consider the sub-density,

$$g_\alpha([\mathbf{x}]_+;[\boldsymbol{m}]_+,\sigma):=\exp\left(-\left(\frac{d_H([\mathbf{x}]_+,[\boldsymbol{m}]_+)}{\sigma}\right)^\alpha\right)$$

parametrized by $([\boldsymbol{m}]_+,\sigma)\in\mathbb{P}^d_{++}\times\mathbb{R}_{++}$, where $\alpha\in\mathbb{R}_{++}$. Set the normalizing factor,

$$Z_\alpha([\mathbf{m}]_+,\sigma):=\int_{\mathbb{P}^d_+} g_\alpha([\mathbf{x}]_+;[\mathbf{m}]_+,\sigma)\nu(d[\mathbf{x}]_+)$$

where ν denotes the uniform measure on $\mathbb{P}^d_{++}$ (see below). Since $d_H\geq 0$, $Z_\alpha([\mathbf{m}]_+,\sigma)$ is finite, and

$$f_\alpha([\mathbf{x}]_+;[\mathbf{m}]_+,\sigma):=(Z_\alpha([\mathbf{m}]_+,\sigma))^{-1}g_\alpha([\mathbf{x}]_+;\boldsymbol{m},\sigma) \tag{39}$$

is a well-defined probability density on $(\mathbb{P}^d_{++},d_H)$, akin to the Gaussian density for $\alpha=2$, and to the Laplace density for $\alpha=1$ (and the uniform density for $\alpha=0$). One has thus:

DEFINITION 7.4 (Generalised Hilbert Gaussian distribution). Let $([\boldsymbol{m}]_+,\sigma,\alpha)\in\mathbb{P}^d_{++}\times\mathbb{R}_{++}\times\mathbb{R}_+$. Then, $[\mathbf{X}]_+\in\mathbb{P}^d_{++}$ follows a Generalised Hilbert Gaussian distribution with parameters $[\boldsymbol{m}]_+,\sigma,\alpha$ if its density w.r.t. the uniform measure ν on $\mathbb{P}^d_+$ writes as (39). Let us call it the Hilbert Gaussian distribution for $\alpha=2$, and the Hilbert Laplace distribution for $\alpha=1$.

Figures 16 and 17 gives density plots of a standard Gaussian, resp. Laplace distribution, for $d=2$, i.e., with $[\mathbf{m}]_+=[1:1:1]_+$ and $\sigma=1$. The hexagonal shape of the density levels corresponds to the balls in Hilbert metric. When the distribution is no longer centered around the neutral point of the simplex, the shape is distorted accordingly. As an illustration, Figure 18 shows a Hilbert Gaussian distribution with $[\mathbf{m}]_+=[0.7:0.1:0.2]_+$ and $\sigma=1$.

For concrete computations, one can take the simplex representation 7 and compute probabilities by transfer. In particular, having w.l.o.g. $x_0=1-\sum_{i=1}^d x_i$ as fixed component, the $d-$ dimensional uniform measure ν on the simplex Δ^d_{++} writes

$$\nu(dx_1,\ldots,dx_d)=\frac{1}{d!}\prod_{i=1}^d \mathbb{1}_{x_i>0}\mathbb{1}_{\sum_{i=1}^d x_i<1}\,dx_1\ldots dx_d,$$

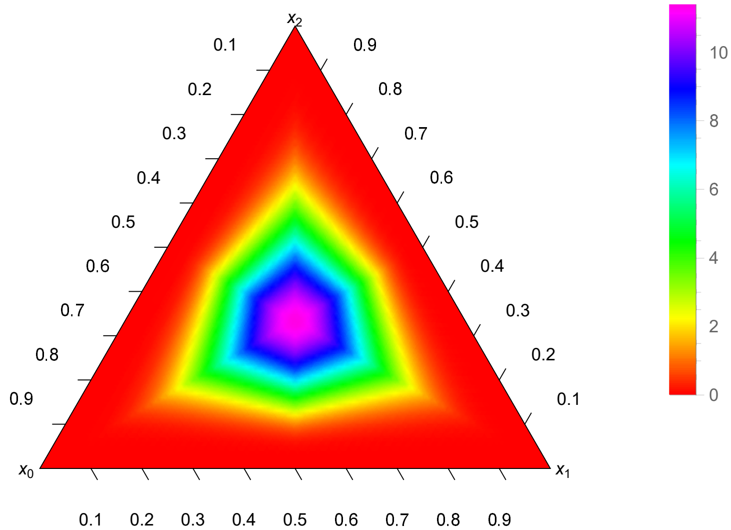

FIG 16. *Standard Gaussian-Hilbert distribution.*

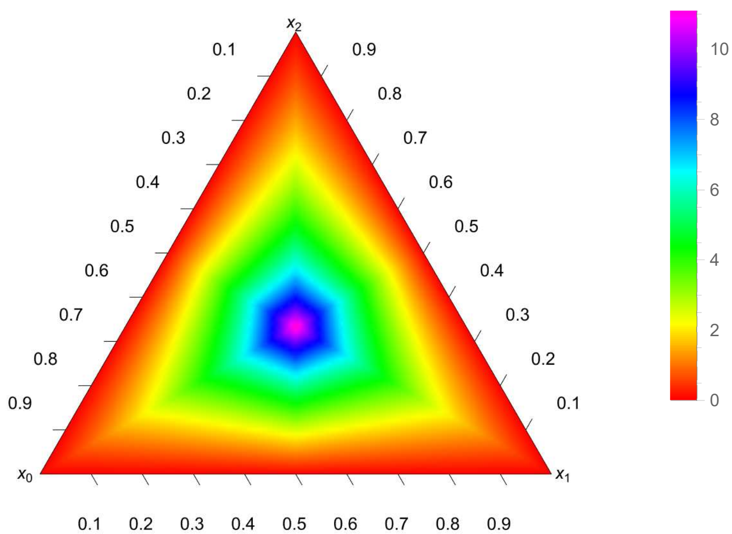

FIG 17. *Standard Laplace-Hilbert distribution.*

so that the Hilbert Gaussian/Laplace distribution of a random CoDa $[\mathbf{X}]_+ \in \mathbb{P}^d_{++}$ writes, for any Borel set A of $\mathbb{P}^d_{++}$,

$$\mathbb{P}([\mathbf{X}]_+ \in A) = \int_{[A]_1} \frac{g_\alpha([\mathbf{x}]_1; [\boldsymbol{m}]_1, \sigma)}{Z_\alpha([\boldsymbol{m}]_+, \sigma)} \frac{1}{d!} \prod_{i=1}^{d} \mathbb{1}_{x_i > 0} \mathbb{1}_{\sum_{i=1}^d x_i < 1} \, dx_1 \dots dx_d,$$

where $[A]_1$ stands for the simplex transformed image of A. Here, one sets $x_0 = 1 - \sum_{i=1}^d x_i$ in the expression of $[\mathbf{x}]_1$, so that the resulting integral is a d-fold classical Lebesgue integral.

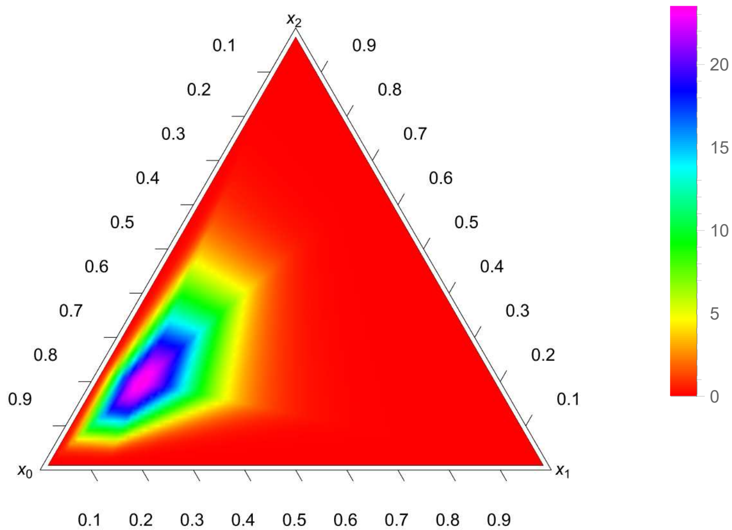

FIG 18. *Gaussian-Hilbert distribution with* $[\mathbf{m}]_+ = [0.7 : 0.1 : 0.2]_+$ *and* $\sigma = 1$.

REMARK 12 (Haar measure on $\mathbb{P}^d_{++}$). The analogue of the uniform/Lebesque measure on $\mathbb{P}^d_{++}$ could have been defined intrinsically as a Haar measure. Let $G = (\mathbb{R}^{d+1}_{++}, \times)$ be the multiplicative Abelian group of positive vectors and $H = \{h_\lambda := \lambda\mathbf{1}, \lambda \in \mathbb{R}\}$ be its subgroup of positive vectors along the ray $\mathbf{1}$. Then, $\mathbb{P}^d_{++} = G/H$ and by Haar's Theorem there is, up to a positive multiplicative constant, a unique countably additive non-trivial measure ν on the Borel subsets of (P^d_{++}, d_H) s.t. for all Borel set B of $\mathbb{R}^d_{++}$ and $h_\lambda \in H$, $\nu(h_\lambda B) = \nu(B)$. However, we believe that measures on $\mathbb{P}^d_{++}$ are easier to grasp by working extrinsically with representatives $[\mathbf{x}]_1 \in \Delta^d_{++}$.

7.3. *Nonparametric CoDa regression estimator based on Hilbert distance.*

7.3.1. *Basic principle.* One can generalise the empirical intrinsic Fréchet-Hilbert mean to a regression framework into the estimation of the conditional mean. Recall that for an Euclidean random vector $(\mathbf{X}, \mathbf{Y}) \in \mathbb{R}^p \times \mathbb{R}^q$, the Nadaraya-Watson estimate of the regression function $r(\mathbf{x}) := E(\mathbf{Y}|\mathbf{X} = \mathbf{x})$,

$$\hat{r}(\mathbf{x}) = \frac{\sum_{i=1}^n \mathbf{Y}_i K\left(\frac{||\mathbf{X}_i - \mathbf{x}||}{h}\right)}{\sum_{i=1}^n K\left(\frac{||\mathbf{X}_i - \mathbf{x}||}{h}\right)},$$

where K is a univariate kernel and $h > 0$ a bandwidth, is the solution of the local constant weighted least square,

$$\hat{r}(\mathbf{x}) = \arg\inf_a \sum_{i=1}^n ||\mathbf{Y}_i - \mathbf{a}||^2 K\left(\frac{||\mathbf{X}_i - \mathbf{x}||}{h}\right),$$

where $||.||$ are the Euclidean norms on the corresponding spaces. Replacing the (square) Euclidean distance for $\mathbf{X}$ or $\mathbf{Y}$ by Hilbert's distance allows to define a nonparametric estimator of the regression function, with $\mathbf{X}$ and/or $\mathbf{Y}$ of the CoDa type. In particular,

- For an Euclidean covariate $\mathbf{X} \in \mathbb{R}^p$ and a CoDa response $[\mathbf{Y}]_+ \in \mathbb{P}^d_{++}$: The Nadraya-Watson type estimator $[\hat{\mathbf{r}}(\mathbf{x})]_+$ of the regression function of $[\mathbf{Y}]_+$ given $\mathbf{X} = \mathbf{x}$ is a minimizer of

$$\sum_{i=1}^{n} K_\Sigma(\mathbf{x}_i - \mathbf{x}) d_H^2([\mathbf{y}_i]_+, [\hat{\mathbf{r}}]_+)$$

over $[\mathbf{r}]_+ \in \mathbb{P}^d_{++}$, with K_Σ multivariate kernel with bandwidth matrix Σ.
- For CoDa Covariate $[\mathbf{X}]_+ \in \mathbb{P}^d_{++}$ and $[\mathbf{Y}]_+ \in \mathbb{P}^d_{++}$ The Nadaraya-Watson type estimator minimizes

$$\sum_{i=1}^{n} K(d_H([\mathbf{x}_i]_+, [\mathbf{x}]_+)/\sigma) d_H^2([\mathbf{y}_i]_+, [\hat{\mathbf{r}}]_+)$$

over $[\mathbf{r}]_+ \in \mathbb{P}^d_{++}$, this time with a univariate kernel K and bandwidth σ.

In view of the normed vector space structure (Proposition 5.2), more sophisticated Hilbert metric based local linear estimates could be constructed by fitting a local linear model instead of a locally constant one. Also, nearest-neighbour based on (smoothed) Hilbert distance can be adapted. See e.g. [58], [36].

7.3.2. *Numerical illustrations.* The following example illustrates the feasibility of the proposed approach to nonparametric CoDa regression based on the smoothed version of Hilbert's distance.

EXAMPLE 3. Figure 19 shows the Nadaraya-Watson estimate of the regression function of a CoDa response variable $[\mathbf{Y}]_+ = [Y_0 : Y_1 : Y_2]_+$, with a univariate regressor x. The data points (blue circles) are generated according to the model

$$\begin{aligned} Y_0 &= \cos(x\pi/2) + 0.2 + \epsilon_0 \\ Y_1 &= \sin(x\pi/2) + 0.2 + \epsilon_1 \\ Y_2 &= \sin(x\pi) + 0.2 + \epsilon_2, \end{aligned}$$

with $\epsilon_0, \epsilon_1, \epsilon_2$ i.i.d. uniform on $[-0.2, 0.2]$, and x regularly spaced between $[0.01, 0.99]$ with increments of 0.05 (197 data points). The smoothed Fréchet-Hilbert conditional mean estimator (green diamonds, bandwidth $\sigma = 0.1$.) is able to track the nonlinear regression curve (orange triangles). Notice, however that the estimated curve misses the beginning and end of the regression curve. This boundary effect, which is typical of a local constant estimator, could be, in principle, be dealt with more sophisticated local linear regression techniques.

8. Conclusion and perspectives. We have thus tried to give a guided tour on the projective geometry viewpoint for CoDa analysis. The main message is that the projective viewpoint gives a natural setting for an intrinsic study of CoDa as equivalence classes, giving much needed geometric insight on the structure, representations and properties of CoDa. As CoDa is studied irrespective of the subjective choice of a particular representation and coordinate systems, our claim is that the projective approach gives a unified view and more general perspective on CoDa. It thus appears at least complementary to the more down-to-earth coordinate-based approaches, which are seen, from the projective viewpoint, as special representations/geometries of the same object. Our hope is that this unified view should foster some advances on the methodological side, and, maybe, some reconciling in the debates over the competing coordinate-based approaches. To that respect, Hilbert's projective metric, with its intrinsic character and its compatibility with both Aitchison's log ratio vector space and

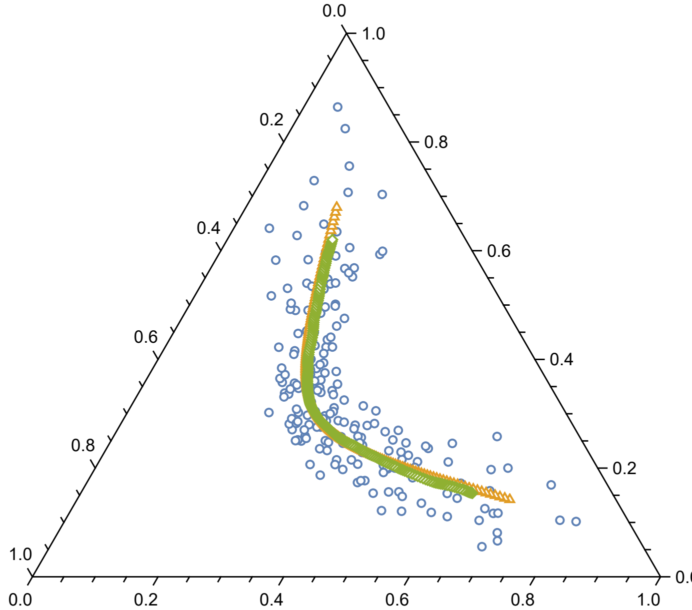

FIG 19. *Nadaraya-Watson nonparametric regression estimator based on smoothed Hilbert metric for a CoDa response with a real regressor.*

the affine simplex structure, seems a candidate to challenge the view espoused in [42] that "it is not possible to come up with a compelling choice for either method based on purely a priori or theoretical grounds, and that a more pragmatic approach is to make a data-dependent choice of metric." In addition, and as was shown in the numerical simulations, its smooth and strictly convex approximate seems to offer a practical surrogate for paving the way to an intrinsic statistical analysis of CoDa.

For length reasons, we have barely scratched the surface of statistical applications based on such an intrinsic projective geometry and Hilbert's metric. Our objective was to quickly validates the feasibility of an intrinsic projective statistical analysis based on Hilbert's metric. Much more needs to be done to establish a full framework. So let us close the article by briefly indicating some directions of further research based on the intrinsic projective viewpoint. The statistical analysis of Fréchet mean and median deserves a deeper study. In particular, non-asymptotic confidence intervals for Fréchet means could be constructed, adapting methods of e.g. [28] [31]. Efficient computations of the Fréchet mean or its smooth surrogate ([23]) is another topic of further research. Another obvious candidate for an interesting alternative to the log-sum-exp smooth approximation of Hilbert's metric would be the Moreau-Yosida regularisation based on infimal convolution (see e.g. [13]). This would, in principle, yield an efficient computation method of a smooth Fréchet-Hilbert mean based on proximal minimisation algorithms. Nonparametric density estimation can be obtained by using a kernel based on measuring distances with Hilbert metric, for example the Hilbert-Gaussian distribution we introduced. Distance covariance (see e.g. [55], [56], [21], [35]) i.e. covariance measures and dependence coefficients based on certain expectations of pairwise distance suggest to study the analogue based on Hilbert's metric. More generally, the study of dependence (inter and intra) deserves a thorough separate study, some of which is current work in progress.

APPENDIX A: PROJECTIVE SPACES AND THEIR (USUAL) REPRESENTATIONS

There are several equivalent ways to define a projective space. The axiomatic approach defines it as an abstract structure verifying certain axioms (in particular incidence axioms). A more intuitive approach is by means of concrete models, using concepts from linear algebra and Euclidean geometry. Recommended references are [48], [37], [53], [11], [16]. See also [54] for two-sided oriented projective geometry, and its applications to computer vision.

A.1. The vector space model of a projective space. In the vector space model of a projective space, a projective space is viewed, via the operation of projectivization $\mathbb{P}(.)$, of a given vector space.

DEFINITION A.1 (Projectivization of a vector space). Let $\mathbb{E}$ be a vector space. The projective space $\mathbb{P}(\mathbb{E})$ induced by $\mathbb{E}$ is the set of of one-dimensional sub-spaces of $\mathbb{E}$,

$$\mathbb{P}(\mathbb{E}) = \{\text{span}(\mathbf{x}), \mathbf{x} \in \mathbb{E}, \mathbf{x} \neq \mathbf{0}, \},$$

with $\text{span}(\mathbf{x}) = \{\lambda\mathbf{x}, \lambda \in \mathbb{R}\}$. Elements of $\mathbb{P}(\mathbb{E})$ are called (projective) points and will be denoted by $[\mathbf{x}]$. $\mathbf{x} \in \mathbb{E}$, $\mathbf{x} \neq \mathbf{0}$ is called a a representative of $[\mathbf{x}]$.

Geometrically, a *point* in $\mathbb{P}(\mathbb{E})$ is a *line* in $\mathbb{E}$ passing through the origin, i.e. an unoriented direction. Similarly, a projective line corresponds to a two dimensional plane in $\mathbb{E}$, and similarly for higher dimensional objects. If $\mathbb{E}$ is finite dimensional, the (projective) dimension of $\mathbb{P}(\mathbb{E})$ is $\dim \mathbb{E} - 1$. We will usually consider $\mathbb{E} = \mathbb{R}^{d+1}$ the Euclidean vector space of dimension $d+1$, and denote simply by $\mathbb{P}^d$ the projective space (of dimension d) induced by $\mathbb{R}^{d+1}$.

- Projective subspaces: The projectivization operation allows to naturally consider projective subspaces from vector subspaces: If $\mathbb{E}' \subset \mathbb{E}$ is a vector subspace of $\mathbb{E}$, then $\mathbb{P}(\mathbb{E}') \subset \mathbb{P}(\mathbb{E})$, since every line $\text{span}(\mathbf{x})$ contained in $\mathbb{E}'$ is also contained in $\mathbb{E}$.
- Projective mappings: Let $\mathbb{L}$ and $\mathbb{M}$ be two linear subspaces and $f : \mathbb{L} \to \mathbb{M}$ a linear mapping. If $\ker f = \{\mathbf{0}\}$ then f maps any straight line from $\mathbb{L}$ into a uniquely determined straight line in $\mathbb{M}$ and hence induces the mapping $\mathbb{P}(f) : \mathbb{P}(\mathbb{L}) \to \mathbb{P}(\mathbb{M})$, called the projectivization of f. In particular, if f is an isomorphism, then $\mathbb{P}(f)$ is called a projective isomorphism. When $\ker f \neq \{\mathbf{0}\}$, straight lines contained in $\ker f$, that is, consisting of the projective subspace $\mathbb{P}(\ker f) \subset \mathbb{P}(\mathbb{L})$, are mapped into zero, which does not determine any point in $\mathbb{P}(\mathbb{M})$. Therefore, the projectivization $\mathbb{P}(f)$ is determined only on the complement $\mathbb{P}(\mathbb{L}) \setminus \mathbb{P}(\ker f)$.
- Projective group: Let $\mathbb{L} = \mathbb{M}$ and f, g be bijective. Then, i) $\mathbb{P}(id_{\mathbb{L}}) = id_{\mathbb{P}(\mathbb{L})}$, ii) $\mathbb{P}(fg) = \mathbb{P}(f)\mathbb{P}(g)$ iii) $\mathbb{P}(f^{-1}) = \mathbb{P}(f)^{-1}$. $\mathbb{P}(f)$ runs through the group of mappings of $\mathbb{P}(\mathbb{L})$ into itself, which is called the projective group of the space $\mathbb{P}(\mathbb{L})$ and is denoted by $\mathbb{PGL}(\mathbb{L})$. Every mapping $\mathbb{P}(f)$ maps the projective subspaces of $\mathbb{P}(\mathbb{L})$ into projective subspaces, preserving dimension and all incidence relations.

A.2. Analytical model: homogeneous coordinates. The analytical model represents points of $\mathbb{P}^d$ by their homogeneous coordinates. Let $(\mathbf{e}_0, \mathbf{e}_1, \ldots, \mathbf{e}_d)$ be the canonical basis of $\mathbb{R}^{d+1}$. A point M of $\mathbb{P}^d$ is represented by any non-null vector $\mathbf{x} \in \mathbb{R}^{d+1}$, with coordinates $(x_0, x_1, \ldots, x_d)$ s.t. the vector line $\text{span}(\mathbf{x})$ corresponds to the projective point M. The coordinates $(x_0, x_1, \ldots, x_d)$ of the vector $\mathbf{x}$ in the basis $(\mathbf{e}_0, \mathbf{e}_1, \ldots, \mathbf{e}_d)$ of $\mathbb{E}$ are called the homogeneous coordinates of the point M. Since any collinear vector $\mathbf{y} = \lambda\mathbf{x}$, with $\lambda \neq 0$, spans the same one-dimensional subspace, $\text{span}(\mathbf{y}) = \text{span}(\mathbf{x})$, thus corresponds to the same projective point M, it is readily seen that homogeneous coordinates are not unique but consists of multiples $(\lambda x_0, \lambda x_1, \ldots, \lambda x_d)$, for $\lambda \in \mathbb{R}^*$ and fixed $(x_0, x_1, \ldots, x_d) \in \mathbb{R}^{d+1} \setminus \mathbf{0}$. We will denote by

$$[x_0 : x_1 : \ldots : x_d]$$

these homogeneous coordinates of the projective point M, with scalar multiples identified. Hence, each point $M \in \mathbb{P}^d$ has infinitely many sets of homogeneous coordinates.

If a bijective linear mapping $f : \mathbb{E} \to \mathbb{E}$ is represented in terms of coordinates of the matrix A, then $\mathbb{P}(f)$ in appropriate homogeneous coordinates is represented by the same matrix A or any matrix λA proportional to it.

A.3. Algebraic model. The algebraic model builds the projective space $\mathbb{P}^d(\mathbb{E})$ as a quotient space, as described in Section 2. For $\mathbf{x}, \mathbf{y} \in \mathbb{E}$ a general vector space, define the equivalence relation $\equiv$ as in (4), and the equivalence class $[\mathbf{x}]$ of $\mathbf{x}$ for (4) as in (5), then, the projective space $\mathbb{P}^d(\mathbb{E})$, obtained by quotienting $\mathbb{E} \setminus \{\mathbf{0}\}$ by the equivalence relation $\equiv$, i.e. as

$$\mathbb{P}^d := (\mathbb{E} \setminus \{\mathbf{0}\})/_{\equiv} = \frac{\mathbb{E} \setminus \{\mathbf{0}\}}{\mathbb{R} \setminus \{0\}}.$$

A.4. Spherical model. The spherical model of $\mathbb{P}^d$ consists of in identifying a projective point with (a pair of) points on the surface of the sphere. Depending on the norm chosen, several "spherical" models arise.

A.4.1. *Unit Euclidean sphere: direction cosines.* Since the intersection of a line passing through the origin and the unit sphere is a pair of diametrically opposed points, the radial projection

$$\mathcal{S} : \mathbf{x} \mapsto \frac{\mathbf{x}}{||\mathbf{x}||_2}, \quad \mathbf{x} \neq \mathbf{0},$$

sends a projective point $[\mathbf{x}]$ to points on the unit sphere

$$\mathbb{S}_2(\mathbb{E}) := \{\mathbf{x} \in \mathbb{E} : ||\mathbf{x}||_2 = 1\},$$

with opposite points identified. Projective lines are then represented by the great circles. For CoDa points $[\mathbf{x}]_+$, the non-negativity constraint $\mathbf{x} \geq \mathbf{0}$, ensures that the CoDa projective point $[\mathbf{x}]_+$ is identified with just one point on the unit sphere. The components of $\mathcal{S}(\mathbf{x})$ are the direction cosines ([63]). One can thus parametrize a CoDa point by the vector of corresponding angles,

$$\boldsymbol{\alpha} = \arccos \mathcal{S}(\mathbf{x}) \in [0, \pi/2]^{d+1},$$

which satisfy the constraint $\sum_{i=0}^{d} \cos^2 \alpha_i = 1$.

Regarding the metric structure, such a spherical representation of the projective space suggests a natural way to measures distance between points, resp. CoDa, as angles between lines, resp. non-negative half-lines, (see e.g. [11] p. 171 Definition 8.6.3): As $\frac{\langle \mathbf{x}|\mathbf{y}\rangle}{||\mathbf{x}||||\mathbf{y}||}$ does not depend on the representatives, one can define, after normalisation[16] by $2/\pi$, the angular/spherical/elliptic distance for CoDa as

$$d_S([\mathbf{x}]_+, [\mathbf{y}]_+) = \frac{2}{\pi} \mathbf{arccos} \frac{\langle \mathbf{x}|\mathbf{y}\rangle}{||\mathbf{x}||||\mathbf{y}||} \tag{40}$$

It satisfy the following properties:

[16] This normalisation is optional.

i) boundedness: $0 \le d_S([\mathbf{x}]_+, [\mathbf{y}]_+) \le 1$
ii) $d_S([\mathbf{x}]_+, [\mathbf{y}]_+) = 0$ iff $[\mathbf{x}]_+ = [\mathbf{y}]_+$. Note that $d_S([\mathbf{x}]_+, [\mathbf{y}]_+) = 1$ iff $\mathbf{x} \perp \mathbf{y}$ (which can only happen if both elements are on the "edges", i.e. have some null components).
iii) symmetry is obvious.
iv) hereditary property: if $\mathbb{F} \subset \mathbb{E}$ is a subspace containing $[\mathbf{x}]_+$ and $[\mathbf{y}]_+$, the angles are the same whether measured in $\mathbb{F}$ or in $\mathbb{E}$. The latter property is a version of subcompositional coherence.
v) Invariance w.r.t rotation: (see [11] Proposition 8.6.6). This property is the analogue of invariance by translation of Proposition 5.1 i), but with vector addition replaced by rotations. This imply in particular invariance w.r.t. permutation of the labelling of the components of CoDa.
vi) triangle inequality. (see [11] Theorem 18.4.2 and Section 19.1).

In addition, this distance is defined for all points of $\mathbb{P}^d_+$, and not only for positive CoDa points of $\mathbb{P}^d_{++}$: it can handle CoDa with zeros.

A.4.2. *Triangular/Simplex representation.* Similarly, by taking the L_1 norm, the radial projection[17]

$$\mathcal{C} : \mathbf{x} \mapsto \frac{\mathbf{x}}{||\mathbf{x}||_1}$$

maps the projective point $[\mathbf{x}]$ to (a pair of identified) points on the unit diamond-shaped "sphere" w.r.t. L_1 norm, $\mathbb{S}_1(E) := \{\mathbf{x} \in \mathbb{E} : ||\mathbf{x}||_1 = 1\}$.

For CoDa points, the non-negativity constraint $\mathbf{x} \ge \mathbf{0}$, ensures that the CoDa projective point $[\mathbf{x}]_+$ is identified with just one point on $\mathbb{S}_1(E)$. This corresponds to standardizing a vector $\mathbf{x} \in \mathbb{R}^{d+1}_+ \setminus \{\mathbf{0}\}$ representing the ray $[\mathbf{x}]_+$ by the sum of its components. This is the simplex representation of a CoDa point, which was presented in Section 2.2 from the affine viewpoint, as a projection on the affine hyperplane $\mathcal{H}_{\texttt{sum}}$ (see below).

A.4.3. *Unit Euclidean sphere: square root.* By combining the simplex representation with the square root, one obtains another way to map a CoDa point to the unit sphere w.r.t. the L_2 norm,

$$\mathcal{R} : \mathbf{x} \mapsto \sqrt{\frac{\mathbf{x}}{||\mathbf{x}||_1}} = \sqrt{\mathcal{C}(\mathbf{x})}$$

This is the square-root representation of [63].

This representation leads to an interesting connection with information geometry: the Euclidean distance between two square-root transformed points $\mathcal{R}(\mathbf{x})$ and $\mathcal{R}(\mathbf{y})$ interprets probabilistically as is the Hellinger distance between the corresponding probability vectors $\mathcal{C}(\mathbf{x})$ and $\mathcal{C}(\mathbf{y})$. Geometrically, the latter is the chordal distance between $\mathcal{R}(\mathbf{x})$ and $\mathcal{R}(\mathbf{y})$.

A.5. Affine/Ratio representation. The affine representation was presented in Section 2.2, taking as affine subspace either $\mathcal{H}_{\texttt{sum}}$ or $\mathcal{H}_0$ of $\mathbb{R}^{d+1}$. For the latter, $\pi_0(\mathbf{x})$ can be further identified with a point in $\mathbb{R}^d$, by dropping the first constant coordinate.

More generally, for each $i = 0, \dots, d$, let

$$\mathbb{U}_i = \{[\mathbf{x}] \in \mathbb{P}^d, \text{ with } x_i \neq 0\}$$

[17] i.e., the closure operation in the CoDa literature, for $\boldsymbol{x} > \mathbf{0}$.

and $\pi_i : \mathbb{U}_i \to \mathbb{R}^d$ be obtained by dividing $\mathbf{x}$ by x_i and dropping the constant ith component, as

$$\pi_i([x_0 : \ldots : x_d]) = (\frac{x_0}{x_i}, \ldots, \frac{x_{i-1}}{x_i}, \frac{x_{i+1}}{x_i}, \ldots, \frac{x_d}{x_i}) \in \mathbb{R}^d.$$

Then, (i) the points of each $\mathbb{U}_i$ are in one-to-one correspondence with the points of $\mathbb{R}^d$, (ii) the complements $\mathbb{H}_i := \mathbb{P}^d \setminus \mathbb{U}_i$ may be identified with $\mathbb{P}^{d-1}$, (iii) $\mathbb{P}^d = \cup_{i=0}^d \mathbb{U}_i$. (See [20] Corollary 3 p. 369.) One has thus a covering of the projective space by an atlas of affine charts, and $\pi_i([\mathbf{x}])$ are the affine/inhomogeneous coordinates of $[\mathbf{x}]$.

A.6. Historical-Straight model . In this model, $\mathbb{P}^d$ is thought as the affine space[18] $\mathbb{R}^d$, completed with some "points at infinity". In the plane $\mathbb{R}^2$, these points at infinity corresponds to the direction of each set of parallel lines, so that in the completion, theses lines all pass through the point they define. This approach allows to overcome the deficiency of the incidence axiom of affine planes, in the sense that two lines in the projective plane will now always intersect at a unique point. In higher dimensions, the idea is similar,

$$\mathbb{P}(\mathbb{E}) = \mathbb{R}^d \cup \mathcal{H}^\infty$$

where $\mathcal{H}^\infty$ is a set of points at infinity. Historically, this approach was motivated by perspective drawings.

A.7. Grassmannian/ Matrix representations. More generally, the set of k dimensional subspaces of a $d+1$ dimensional vector space is called the Grassmannian $\mathcal{G}(k, d+1)$, Thus, for $k = 1$, $\mathcal{G}(1, d+1)$ corresponds to the projective space $\mathbb{P}^d$, and the CoDa space $\mathbb{P}^d_+$ corresponds to the non-negative part of such Grassmannian. Grassmannian admit a variety of representations by (classes) of matrices, see e.g. [19]. In particular, a projective, resp. CoDa, point, $[\mathbf{x}]$, resp. $[\mathbf{x}]_+$, can be represented by the orthogonal projection matrix $(\mathbf{x}\mathbf{x}^T)/(\mathbf{x}^T\mathbf{x})$, or dually, by the corresponding orthogonal projection matrix on the vector hyperplane orthogonal to $\mathbf{x}$.

APPENDIX B: PROOFS.

PROOF OF LEMMA 2.2. That $\mathbb{PGL}^d_+$ is a group is trivial. For $\mathbb{PGL}^d_{++}$, simply note that if $[A]_+, [B]_+ \in \mathbb{PGL}^d_{++}$, then, setting $b_k := \sum_{j=0}^d b_{kj} > 0$, one has that for all $0 \le j \le d$,

$$\sum_j \sum_k a_{ik} b_{kj} = \sum_k b_k a_{ik} \ge \min_k b_k \sum_k a_{ik} > 0.$$

Thus, $[AB]_+ \in \mathbb{PGL}^d_{++}$. □

PROOF OF LEMMA 3.1. i) Injectivity: let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{d+1}_{++}$ s.t. $[\mathbf{ln}\, \boldsymbol{x}]_\sim = [\mathbf{ln}\, \boldsymbol{y}]_\sim$. This means that there exists some $\mu \in \mathbb{R}$ s.t. $\mathbf{ln}\, \boldsymbol{x} = \mathbf{ln}\, \boldsymbol{y} + \mu\mathbf{1}$. Thus, $\ln x_i = \ln y_i + \mu = \ln(y_i e^\mu)$, $i = 0, \ldots, d$, which implies $x_i = \lambda y_i$, with $\lambda = e^\mu > 0$, that is to say $[\mathbf{x}]_+ = [\mathbf{y}]_+$.

ii) Surjectivity: let $[\mathbf{z}]_\sim \in \mathbb{R}^{d+1}/\sim$. Then, $[\mathbf{z}]_\sim$ is represented by vectors $\boldsymbol{\zeta} \in \mathbb{R}^{d+1}$ of the form $\boldsymbol{\zeta} = \mathbf{z} + \mu\mathbf{1}$, where $\mu \in \mathbb{R}$ can be chosen arbitrarily. Taking $\mu = -\ln(\sum_{i=0}^d e^{z_i})$ yields

$$\zeta_i = z_i - \ln\left(\sum_{i=0}^d e^{z_i}\right) = \ln\left(\frac{e^{z_i}}{\sum_{i=0}^d e^{z_i}}\right), \quad i = 0, \ldots, d.$$

[18] and not $\mathbb{R}^{d+1}$!

That is to say $\boldsymbol{\zeta} = \mathbf{ln}\,([\boldsymbol{e}^{\mathbf{z}}]_{\mathbf{1}})$, where $[.]_1$ is the unit-sum rescaling/closure operation (7). Thus, every element $[\mathbf{z}]_\sim$ of $\mathbb{R}^{d+1}/\sim$ writes as the $\mathbf{ln}$ of a normalised CoDa element $[e^{\mathbf{z}}]_1$ in Δ^d_{++}, a fortiori as an element $[e^{\mathbf{z}}]_+$ of $\mathbb{P}^d_{++}$.

□

PROOF OF DEFINITION 4.3. Finiteness follows from Proposition 4.2 i) and independence of representatives from Proposition 4.2 iv). Completeness follows from e.g. [17] Theorem 4.1. □

PROOF OF PROPOSITION 5.1. i) By definition,

$$\begin{aligned} d_H([\mathbf{x}]_+ \oplus [\mathbf{p}]_+, [\mathbf{y}]_+ \oplus [\mathbf{p}]_+) &= d_H(\boldsymbol{x} \times \boldsymbol{p}, \boldsymbol{y} \times \boldsymbol{p}) \\ &= \ln \max_{i,j} \frac{x_i p_i y_j p_j}{x_j p_j y_i p_i} = \ln \max_{i,j} \frac{x_i y_j}{x_j y_i} \\ &= d_H([\mathbf{x}]_+, [\mathbf{y}]_+) \end{aligned}$$

and similarly for the other equality.

ii) Also clear from (23).

□

PROOF OF PROPOSITION 5.2. That $||[\mathbf{x}]_+||_H \geq 0$ and $||[\mathbf{x}]_+||_H = 0$ iff $[\mathbf{x}]_+ = [\mathbf{1}]_+$ are obvious from the definition of $||.||_H$. It is also clear that

$$||\lambda \odot [\mathbf{x}]_+||_H = ||[\mathbf{x}^\lambda]_+||_H = \max_{i,j}(\lambda \ln(x_i/x_j)) = |\lambda| \times ||[\mathbf{x}]_+||_H.$$

The triangle inequality $||[\mathbf{x}]_+ \oplus [\mathbf{y}]_+||_H \leq ||[\mathbf{x}]_+||_H + ||[\mathbf{y}]_+||_H$ follows from the vector isomorphism of the log map and the elementary inequality $\max_i(a_i + b_i) \leq \max_i a_i + \max_i b_i$. □

PROOF OF PROPOSITION 5.7. 1. i) The fact that m-segments, i.e. straight line segments in the simplex Δ^d_+, are geodesics, follows from [33] Theorem 2.6.3. We adapt their proof to our setting and notations, for the convenience of the reader. Let $[\mathbf{p}]_+, [\mathbf{q}]_+ \in \mathbb{P}^d_{++}$, and set $\delta = d_H([\mathbf{p}]_+, [\mathbf{q}]_+)$. For $0 \leq t \leq 1$, set $[\mathbf{p}(t)]_+$ defined by linear interpolation on the simplex as $[\mathbf{p}(t)]_1 = (1-t)[\mathbf{p}]_1 + t[\mathbf{q}]_1$.

The properties of the cross-ratio for aligned points entails that, for $0 \leq s < t \leq 1$,

$$d_H([\mathbf{p}]_+, [\mathbf{p}(s)]_+) + d_H([\mathbf{p}(s)]_+, [\mathbf{p}(t)]_+) = d_H([\mathbf{p}]_+, [\mathbf{p}(t)]_+). \tag{41}$$

Set $\alpha(t) := d_H([\mathbf{p}]_+, [\mathbf{p}(t)]_+)$, so that (41) writes

$$\alpha(s) - \alpha(t) = -d_H([\mathbf{p}(s)]_+, [\mathbf{p}(t)]_+) < 0$$

for $s < t$. Thus, $\alpha : [0,1] \to [0,\delta]$ is strictly increasing and continuous, hence a bijection with inverse mapping α^{-1}. Rescale the $m-$segment by setting, for $0 \leq u \leq 1$,

$$[\gamma^m(u)]_+ = [\mathbf{p}(\alpha^{-1}(\delta u)]_+.$$

Then, for $0 \leq u < v \leq 1$ and $s := \alpha^{-1}(\delta u)$, $t := \alpha^{-1}(\delta v)$, (41) yields

$$\begin{aligned} d_H([\gamma^m(u)]_+, [\gamma^m(v)]_+) &= d_H([\mathbf{p}(\alpha^{-1}(\delta u)]_+, [\mathbf{p}(\alpha^{-1}(\delta v)]_+) \\ &= d_H([\mathbf{p}(s]_+, [\mathbf{p}(t]_+) \\ &= \alpha(t) - \alpha(s) \\ &= \alpha(\alpha^{-1}(\delta v)) - \alpha(\alpha^{-1}(\delta u)) \\ &= \delta(v-u) \\ &= |u-v| d_H([\mathbf{p}]_+, [\mathbf{q}]_+) \end{aligned}$$

2. ii) e-segments (i.e. Aitchison's segments) are geodesics for Hilbert metric, as follows directly from the normed vector space structure of $(\mathbf{P}^d_{++}, \oplus, \odot, ||.||_H)$ of Proposition 5.2: for $0 \le s, t \le 1$,

$$\begin{aligned} d_H([\gamma^e(t)]_+, \gamma^e(s)]_+) &= d_H((1-t)\odot[\mathbf{p}]_+ \oplus t\odot[\mathbf{q}]_+, (1-s)\odot[\mathbf{p}]_+ \oplus s\odot[\mathbf{q}]_+) \\ &= ||[\mathbf{p}]_+ \oplus t\odot([\mathbf{q}]_+ \ominus [\mathbf{p}]_+) \ominus ([\mathbf{p}]_+ \oplus s\odot([\mathbf{q}]_+ \ominus [\mathbf{p}]_+))||_H \\ &= ||(t-s)\odot([\mathbf{q}]_+ \ominus [\mathbf{p}]_+)||_H \\ &= |t-s|\,||[\mathbf{q}]_+ \ominus [\mathbf{p}]_+||_H = |t-s| d_H([\mathbf{p}]_+, [\mathbf{q}]_+)) \end{aligned}$$

□

PROOF OF PROPOSITION 6.1. For some fixed $\mathbf{y} \in \mathbb{R}^{d+1}_{++}$, let $\phi_i : \mathbb{R}^{d+1}_{++} \to \mathbb{R}$ be defined by

$$\phi_i(\mathbf{x}) = \ln(x_i/y_i) = \ln x_i - \ln y_i$$

Formula (30) writes

$$f(\mathbf{x}) = \max_i \phi_i(\mathbf{x}) - \min_i \phi_i(\mathbf{x}) = \max_i \phi_i(\mathbf{x}) + \max_i(-\phi_i(\mathbf{x})).$$

ϕ_i is $\mathcal{C}_1-$differentiable on $\mathbb{R}^{d+1}_{++}$, hence directionally differentiable with directional derivative

$$\phi_i'(\mathbf{x}, \mathbf{v}) = \langle \nabla\phi_i(\mathbf{x}), \mathbf{v}\rangle = v_i/x_i.$$

By Theorem 2.4.1 in [30] p. 41 or Proposition 3.5 in [22], f is directionally differentiable, and formula (32) follows. □

PROOF OF PROPOSITION 6.2. The function $\phi : \mathbb{P}^d_{++} \to \mathbb{R}^{d+1}_+ +$ defined by $\phi([\mathbf{x}]_+) = \ln \boldsymbol{x} - \ln \boldsymbol{y}$ is affine w.r.t. to $\oplus, \odot$ (it expresses the translation from the line $[\ln \boldsymbol{y}]_\sim$ to the line $[\ln \boldsymbol{x}]_\sim$). Thus,

$$\frac{\phi([\mathbf{x}]_+ \oplus \lambda \odot [\mathbf{v}]_+) - \phi([\mathbf{x}]_+)}{\lambda} = \ln \boldsymbol{v} = \phi'^{\oplus}([\mathbf{x}]_+, [\mathbf{v}]_+).$$

Applying again Theorem 2.4.1 in [30] p. 41 or Proposition 3.5 in [22] yields the result. □

PROOF OF LEMMA 6.4. i) Follows from the inequality

$$\max e^{cx_i} = e^{c\max x_i} \le \sum_i e^{cx_i} \le m \max e^{cx_i} = me^{c\max x_i}.$$

ii) Monotonicity is obvious. Regarding convexity, let $S(\mathbf{x}) = (e^{cx_1}, \ldots, e^{cx_m})$, and $\mathbf{1} = (1, \ldots, 1)$. Then, the gradient writes $\nabla \text{lse}_c(\mathbf{x}) = S(\mathbf{x})/(\mathbf{1}^T S(\mathbf{x}))$, and the Hessian is

$$\nabla^2 \text{lse}_c(\mathbf{x}) = c\frac{\text{diag}(S(\mathbf{x}))\mathbf{1}^T S(\mathbf{x}) - S(\mathbf{x}))S(\mathbf{x})^T}{(\mathbf{1}^T S(\mathbf{x}))^2} =: \mu A,$$

where we have set the matrix $A = [a_{ij}] := \text{diag}(S(\mathbf{x}))\mathbf{1}^T S(\mathbf{x}) - S(\mathbf{x}))S(\mathbf{x})^T$ and $\mu := c/(\mathbf{1}^T S(\mathbf{x}))^2$ is a scalar. The diagonal terms of the Hessian are $a_i := \mu a_{ii} = \mu e^{cx_i}\sum_{k\neq i} e^{cx_k}$ and the sum of the absolute value of the non-diagonal terms for row i of the Hessian are $R_i := \mu \sum_{j\neq i}|a_{ij}| = a_i$. Therefore, it follows that the Hessian is diagonally dominant thus positive semi-definite by Gershgorin's circle Theorem [27].[19] Hence, lse_c is convex.

[19] Gershgorin's circle Theorem state that all (complex) eigenvalues of a square matrix lies within at least one of the closed discs of center a_i and radius R_i. Here, since $a_i = R_i$, this implies that all eigenvalues are ≥ 0.

iii) One has

$$\mathrm{lse}_c(\mathbf{x}+\mathbf{y}) = \frac{1}{c}\ln\left(\sum_i e^{cx_i}e^{cy_i}\right) \leq \frac{1}{c}\ln\left(\max_i e^{cy_i}\sum_i e^{cx_i}\right)$$
$$= \max(\mathbf{y}) + \mathrm{lse}_c(\mathbf{x}) \leq \mathrm{lse}_c(\mathbf{y}) + \mathrm{lse}_c(\mathbf{x}),$$

where the last step follows from i).

iv) Follows from calculation.

□

PROOF OF PROPOSITION 6.5. i) follows from the definition (34) and Lemma 6.4 i).

ii) Non-negativity follows from i) and non-negativity of Hilbert's distance. In fact, Remark 11 ii) yields $d_{H,c}([\mathbf{x}]_+,[\mathbf{y}]_+) \geq d_{H,c}([\mathbf{x}]_+,[\mathbf{x}]_+) = \frac{2\ln(d+1)}{c}$. Symmetry is obvious. Since

$$\ln\left(\frac{x_i z_j}{x_j z_i}\right) = \ln\left(\frac{x_i y_j}{x_j y_i}\right) + \ln\left(\frac{y_i z_j}{y_j z_i}\right),$$

the triangle inequality follows from the sub-additivity of lse_c, Lemma 6.4 iii).

iii) a) Obvious from the definition (34).

b) By definition of $\odot$,

$$d_{H,c}(\lambda\odot[\mathbf{x}]_+,\lambda\odot[\mathbf{y}]_+) = d_{H,c}([\mathbf{x}^\lambda]_+,[\mathbf{y}^\lambda]_+)$$
$$= \frac{1}{c}\ln\left(\sum_{i,j} e^{c\ln R(i,j)^\lambda}\right),$$

where $R(i,j)$ was defined in Remark 11 ii). If $\lambda > 0$,

$$d_{H,c}(\lambda\odot[\mathbf{x}]_+,\lambda\odot[\mathbf{y}]_+) = \frac{\lambda}{\lambda c}\ln\left(\sum_{i,j} e^{\lambda c\ln R(i,j)}\right)$$
$$= \lambda \times d_{H,\lambda c}([\mathbf{x}]_+,[\mathbf{y}]_+).$$

If $\lambda < 0$, write $\lambda = -|\lambda|$, so that

$$d_{H,c}(\lambda\odot[\mathbf{x}]_+,\lambda\odot[\mathbf{y}]_+) = \frac{|\lambda|}{|\lambda|c}\ln\left(\sum_{i,j} e^{-|\lambda|c\ln R(i,j)}\right)$$
$$= \frac{|\lambda|}{|\lambda|c}\ln\left(\sum_{i,j} e^{|\lambda|c\ln R(j,i)}\right)$$
$$= |\lambda| \times d_{H,|\lambda|c}([\mathbf{x}]_+,[\mathbf{y}]_+)$$

since $-\ln R(i,j) = \ln R(j,i)$, as shown in the aforementioned remark. Note that for $\lambda = 0$, one has only $d_{H,c}(\lambda\odot[\mathbf{x}]_+,\lambda\odot[\mathbf{y}]_+) = \frac{2\ln(d+1)}{c}$, which is the lower bound of $d_{H,c}$ by ii) (a).

c) Set $\widetilde{\Psi} : \mathbb{P}^d_{++} \to \mathbb{R}^{(d+1)\times(d+1)}$ the version of Ψ of (27) extended to all pairs of indices, i.e. defined for $0 \leq i,j \leq d$, by $\widetilde{\Psi}([\mathbf{x}]_+)_{ij} = \ln(x_i/x_j)$. Then, the smoothed Hilbert metric writes $d_{H,c}([\mathbf{x}]_+,[\mathbf{y}]_+) = \mathrm{lse}_c(\widetilde{\Psi}([\mathbf{x}]_+) - \widetilde{\Psi}([\mathbf{y}]_+))$. By the vector space isomorphism of the logarithm, $\widetilde{\Psi}([\mathbf{x}]_+\oplus[\mathbf{y}]_+) = \widetilde{\Psi}([\mathbf{x}]_+) + \widetilde{\Psi}([\mathbf{y}]_+)$ and $\widetilde{\Psi}(\lambda\odot[\mathbf{x}]_+) =$

$\lambda\widetilde{\Psi}([\mathbf{x}]_+)$. Therefore, $[\mathbf{x}]_+ \to d_{H,c}([\mathbf{x}]_+, [\mathbf{y}]_+)$ is the composition of the lse_c function, which is convex by Lemma 6.4 ii), with an affine function, hence is convex w.r.t. the vector space operations $\oplus, \odot$. Moreover, since $\widetilde{\Psi}([\mathbf{x}]_+)_{ii} = 0$ for all $0 \le i \le d$. Therefore, and as explained in Remark 11 ii), the lse function computes a smooth approximation of the maximum of $(d+1)^2$ terms with $d+1$ constant terms 1, which results in the mapping being strictly convex.

iv) Obvious.

□

PROOF OF LEMMA 7.2. i) By the reverse triangle inequality,

$$|\mathcal{L}_\mu([\mathbf{m}]_+) - \mathcal{L}_\mu([\mathbf{p}]_+)| \le E|d_H([\mathbf{X}]_+, [\mathbf{m}]_+) - d_H([\mathbf{X}]_+, [\mathbf{p}]_+)| \le d_H([\mathbf{m}]_+, [\mathbf{p}]_+).$$

Similarly for $\mathcal{F}_\mu$,

$$|\mathcal{F}_\mu([\mathbf{m}]_+) - \mathcal{F}_\mu([\mathbf{p}]_+)| \le d_H([\mathbf{m}]_+, [\mathbf{p}]_+)(\mathcal{L}_\mu([\mathbf{m}]_+) + \mathcal{L}_\mu([\mathbf{p}]_+)).$$

By Weierstrass theorem, $\mathcal{L}_\mu$ is bounded on each compact set of $\mathbb{P}^d_{++}$, thus $\mathcal{F}_\mu$ is Lipschitz on compact sets.

ii) By Cauchy-Schwarz, $\mathcal{F}_\mu([\mathbf{m}]_+) < \infty$ implies $\mathcal{L}_\mu([\mathbf{m}]_+) < \infty$. By the (local)-Lipschitz property of $\mathcal{L}_\mu$, resp. $\mathcal{F}_\mu$, $\mathcal{L}_\mu([\mathbf{m}]_+) < \infty$, resp. $\mathcal{F}_\mu([\mathbf{m}]_+) < \infty$ for some $[\mathbf{m}]_+ \in \mathbb{P}^d_{++}$ implies $\mathcal{L}_\mu([\mathbf{m}]_+) < \infty$, resp. $\mathcal{F}_\mu([\mathbf{m}]_+) < \infty$ for all $[\mathbf{m}]_+ \in \mathbb{P}^d_{++}$.

iii) Obvious from the Definition 4.3.

iv) Follows from the fact that $(\mathbb{P}^d_{++}, \oplus, \odot, ||.||_H)$ is a normed vector space, and that the norm is a convex function.

□

PROOF OF THEOREM 7.3. By assumption, there exists $[\mathbf{m}_0]_+$ s.t. $\mathcal{F}_\mu([\mathbf{m}_0]_+) < \infty$. Since F_μ is continuous, $\mathcal{X}_0 := \{[\mathbf{m}]_+ \in \mathbb{P}^d_{++} : \mathcal{F}_\mu([\mathbf{m}]_+) \le \mathcal{F}_\mu([\mathbf{m}_0]_+)\}$ is closed. Hence, $\mathcal{X}_0$ is non-empty, bounded by coercivity of $\mathcal{F}_\mu$, and closed, therefore compact since $\mathbb{P}^d_{++}$ is a finite dimensional vector space. Thus $\mathcal{F}_\mu$ admits a minimum $[\mathbf{m}^H]_+$ on $\mathcal{X}_0$ by Weierstrass theorem. For all $[\mathbf{m}]_+ \notin \mathcal{X}_0$, $\mathcal{F}_\mu([\mathbf{m}]_+) > \mathcal{F}_\mu([\mathbf{m}_0]_+) \ge \mathcal{F}_\mu([\mathbf{m}^H]_+)$. Therefore, $[\mathbf{m}^H]_+$ is a global minimum on $\mathbb{P}^d_{++}$.

The proof for the median is similar. □

Acknowledgments. The author thanks Christine Thomas-Agnan and her CoDa group at the Toulouse School of Economics for discussions and feedback on the ideas proposed in this manuscript.

Funding. The author 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

[1] AFSARI, B. (2011). Riemannian L^p center of mass: existence, uniqueness, and convexity. *Proc. Amer. Math. Soc.* **139** 655–673. https://doi.org/10.1090/S0002-9939-2010-10541-5 MR2736346

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

[3] AITCHISON, J. (1986). *The statistical analysis of compositional data*. *Monographs on Statistics and Applied Probability*. Chapman & Hall, London. https://doi.org/10.1007/978-94-009-4109-0 MR865647

[4] AITCHISON, J. and PAWLOWSKY-GLAHN, V. (1997). The one-hour course in compositional data analysis or compositional data analysis is simple. In *Proceedings of IAMG* **97** 3–35. Barcelona.

[5] AMARI, S.-I. (2016). *Information geometry and its applications*. *Applied Mathematical Sciences* **194**. Springer, [Tokyo]. https://doi.org/10.1007/978-4-431-55978-8 MR3495836

[6] AMARI, S.-I. and NAGAOKA, H. (2000). *Methods of information geometry*. *Translations of Mathematical Monographs* **191**. American Mathematical Society, Providence, RI; Oxford University Press, Oxford Translated from the 1993 Japanese original by Daishi Harada. MR1800071

[7] BAXTER, M., BEARDAH, C., COOL, H. and JACKSON, C. (2005). Compositional data analysis of some alkaline glasses. *Mathematical geology* **37** 183–196.

[8] BAXTER, M. J. (1995). Standardization and transformation in principal component analysis, with applications to archaeometry. *Journal of the Royal Statistical Society Series C: Applied Statistics* **44** 513–527.

[9] BAXTER, M. J. (2001). Statistical modelling of artefact compositional data. *Archaeometry* **43** 131–147.

[10] BAXTER, M. J. and FREESTONE, I. C. (2006). Log-ratio compositional data analysis in archaeometry. *Archaeometry* **48** 511–531.

[11] BERGER, M. (2009). *Geometry I*. *Universitext*. Springer-Verlag, Berlin Translated from the 1977 French original by M. Cole and S. Levy, Fourth printing of the 1987 English translation [MR0882541]. MR2724360

[12] BIRKHOFF, G. (1957). Extensions of Jentzsch's theorem. *Trans. Amer. Math. Soc.* **85** 219–227. https://doi.org/10.2307/1992971 MR87058

[13] BOYD, S. and VANDENBERGHE, L. (2004). *Convex optimization*. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511804441 MR2061575

[14] BULLEN, P. S. (2003). *Handbook of means and their inequalities*. *Mathematics and its Applications* **560**. Kluwer Academic Publishers Group, Dordrecht Revised from the 1988 original [P. S. Bullen, D. S. Mitrinović and P. M. Vasić, ıt Means and their inequalities, Reidel, Dordrecht; MR0947142]. https://doi.org/10.1007/978-94-017-0399-4 MR2024343

[15] BURAGO, D., BURAGO, Y. and IVANOV, S. (2001). *A course in metric geometry*. *Graduate Studies in Mathematics* **33**. American Mathematical Society, Providence, RI. https://doi.org/10.1090/gsm/033 MR1835418

[16] BUSEMANN, H. and KELLY, P. J. (1953). *Projective geometry and projective metrics*. Academic Press, Inc., New York, N.Y. MR0054980

[17] BUSHELL, P. J. (1973). Hilbert's metric and positive contraction mappings in a Banach space. *Arch. Rational Mech. Anal.* **52** 330–338. https://doi.org/10.1007/BF00247467 MR336473

[18] CHARLIER, B. (2013). Necessary and sufficient condition for the existence of a Fréchet mean on the circle. *ESAIM Probab. Stat.* **17** 635–649. https://doi.org/10.1051/ps/2012015 MR3126155

[19] CHIKUSE, Y. (2003). *Statistics on special manifolds*. *Lecture Notes in Statistics* **174**. Springer-Verlag, New York. https://doi.org/10.1007/978-0-387-21540-2 MR1960435

[20] COX, D., LITTLE, J. and O'SHEA, D. (2007). *Ideals, varieties, and algorithms*, third ed. *Undergraduate Texts in Mathematics*. Springer, New York An introduction to computational algebraic geometry and commutative algebra. https://doi.org/10.1007/978-0-387-35651-8 MR2290010

[21] DEB, N., GHOSAL, P. and SEN, B. (2020). Measuring Association on Topological Spaces Using Kernels and Geometric Graphs.

[22] DEM'YANOV, V. F. and RUBINOV, A. M. (1986). *Quasidifferential calculus*. *Translations Series in Mathematics and Engineering*. Optimization Software.

[23] EICHFELDER, G., HOTZ, T. and WIEDITZ, J. (2019). An algorithm for computing Fréchet means on the sphere. *Optim. Lett.* **13** 1523–1533. https://doi.org/10.1007/s11590-019-01415-y MR4002314

[24] ERB, I. and AY, N. (2021). The information-geometric perspective of Compositional Data Analysis. In *Advances in Compositional Data Analysis: Festschrift in Honour of Vera Pawlowsky-Glahn* 21–43. Springer.

[25] FILZMOSER, P., HRON, K. and TEMPL, M. (2018). *Applied compositional data analysis*. *Springer Series in Statistics*. Springer, Cham With worked examples in R. https://doi.org/10.1007/978-3-319-96422-5 MR3839314

[26] FRÉCHET, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. In *Annales de l'institut Henri Poincaré* **10** 215–310.

[27] GERSHGORIN, S. (1931). Über die Abgrenzung der Eigenwerte einer Matrix. *Bull. Acad. Sci. URSS* **1931** 749–754.

[28] GLOCK, M. and HOTZ, T. (2017). Constructing universal, non-asymptotic confidence sets for intrinsic means on the circle. In *Geometric science of information*. *Lecture Notes in Comput. Sci.* **10589** 477–485. Springer, Cham. MR3737761

[29] HILBERT, D. (1895). Über die gerade Linie als kürzeste Verbindung zweier Punkte. *Mathematische Annalen* **46** 91–96.

[30] HIRIART-URRUTY, J.-B. (1991). *L'optimisation*. *Que sais-je?* Presses Universitaires de France.

[31] HOTZ, T., KELMA, F. and WIEDITZ, J. (2015). Universal, non-asymptotic confidence sets for circular means. In *Geometric science of information*. *Lecture Notes in Comput. Sci.* **9389** 635–642. Springer, Cham. MR3442246

[32] KOHLBERG, E. and PRATT, J. W. (1982). The contraction mapping approach to the Perron-Frobenius theory: why Hilbert's metric? *Math. Oper. Res.* **7** 198–210. https://doi.org/10.1287/moor.7.2.198 MR665557

[33] LEMMENS, B. and NUSSBAUM, R. (2012). *Nonlinear Perron-Frobenius theory*. *Cambridge Tracts in Mathematics* **189**. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9781139026079 MR2953648

[34] LEMMENS, B. and NUSSBAUM, R. (2013). Birkhoff's version of Hilbert's metric and its applications in analysis. *arXiv preprint arXiv:1304.7921*.

[35] LYONS, R. (2013). Distance covariance in metric spaces. *Ann. Probab.* **41** 3284–3305. https://doi.org/10.1214/12-AOP803 MR3127883

[36] MARZIO, M. D., PANZERA, A. and VENIERI, C. (2015). Non-parametric regression for compositional data. *Statistical Modelling* **15** 113-133. https://doi.org/10.1177/1471082X14535522

[37] MODENOV, P. S. and PARKHOMENKO, A. S. (1965). *Geometric transformations. Vol. 2: Projective transformations*. Academic Press, New York-London Translated and adapted from the first Russian edition by Michael B. P. Slater, Published in cooperation with the Survey of Recent East European Mathematical Literature. MR0193558

[38] NIELSEN, F. (2019). *Geometric Structures of Information*. Springer.

[39] NIELSEN, F. and HADJERES, G. (2018). Monte Carlo Information Geometry: The dually flat case.

[40] NIELSEN, F. and SUN, K. (2018). *Clustering in Hilbert's Projective Geometry: The Case Studies of the Probability Simplex and the Elliptope of Correlation Matrices* In *Signals and Communication Technology* 297–331. Springer International Publishing. https://doi.org/10.1007/978-3-030-02520-5_11

[41] NIELSEN, F. and SUN, K. (2023). Non-linear Embeddings in Hilbert Simplex Geometry.

[42] PANTAZIS, Y., TSAGRIS, M. and WOOD, A. T. A. (2019). Gaussian asymptotic limits for the α-transformation in the analysis of compositional data. *Sankhya A* **81** 63–82. https://doi.org/10.1007/s13171-018-00160-1 MR3982191

[43] PAPADOPOULOS, A. and TROYANOV, M., eds. (2014). *Handbook of Hilbert geometry*. *IRMA Lectures in Mathematics and Theoretical Physics* **22**. European Mathematical Society (EMS), Zürich. MR3309067

[44] PAWLOWSKY-GLAHN, V. and BUCCIANTI, A. (2011). *Compositional data analysis: Theory and applications*. John Wiley & Sons.

[45] PAWLOWSKY-GLAHN, V., EGOZCUE, J. J. and TOLOSANA DELGADO, R. (2007). Lecture notes on compositional data analysis.

[46] PAWLOWSKY-GLAHN, V., EGOZCUE, J. J. and TOLOSANA-DELGADO, R. (2015). *Modeling and analysis of compositional data*. *Statistics in Practice*. John Wiley & Sons, Ltd., Chichester. MR3328965

[47] PENNEC, X. (2006). Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision* **25** 127–154.

[48] RICHTER-GEBERT, J. (2011). *Perspectives on projective geometry*. Springer, Heidelberg A guided tour through real and complex geometry. https://doi.org/10.1007/978-3-642-17286-1 MR2791970

[49] ROBINSON, S. M. (1987). *Local structure of feasible sets in nonlinear programming, Part III: Stability and sensitivity*. Springer.

[50] SCEALY, J. L. and WELSH, A. H. (2011). Regression for compositional data by using distributions defined on the hypersphere. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **73** 351–375. https://doi.org/10.1111/j.1467-9868.2010.00766.x MR2815780

[51] SCEALY, J. L. and WELSH, A. H. (2014). Colours and cocktails: compositional data analysis 2013 Lancaster lecture. *Aust. N. Z. J. Stat.* **56** 145–169. https://doi.org/10.1111/anzs.12073 MR3226434

[52] SCHOLTES, S. (2012). *Introduction to piecewise differentiable equations*. Springer Science & Business Media.

[53] SHAFAREVICH, I. R. and REMIZOV, A. O. (2013). *Linear algebra and geometry*. Springer, Heidelberg Translated from the 2009 Russian original by David Kramer and Lena Nekludova. https://doi.org/10.1007/978-3-642-30994-6 MR2963561

[54] STOLFI, J. (1991). *Oriented projective geometry*. Academic Press, Inc., Boston, MA A framework for geometric computations. MR1121502

[55] SZÉKELY, G. J. and RIZZO, M. L. (2023). *The energy of data and distance correlation*. CRC Press.

[56] SZÉKELY, G. J., RIZZO, M. L. and BAKIROV, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35** 2769–2794. https://doi.org/10.1214/009053607000000505 MR2382665

[57] TSAGRIS, M. (2015). Regression analysis with compositional data containing zero values. *Chil. J. Stat.* **6** 47–57. MR3407274

[58] TSAGRIS, M., ALENAZI, A. and STEWART, C. (2023). Flexible non-parametric regression models for compositional response data with zeros. *Stat. Comput.* **33** Paper No. 106, 17. https://doi.org/10.1007/s11222-023-10277-5 MR4621141

[59] TSAGRIS, M. T., PRESTON, S. and WOOD, A. T. (2011). A data-based power transformation for compositional data. *arXiv preprint arXiv:1106.1451*.

[60] VAN DEN BOOGAART, K. G. and TOLOSANA-DELGADO, R. (2008). "Compositions": a unified R package to analyze compositional data. *Computers & Geosciences* **34** 320–338.

[61] VAN DEN BOOGAART, K. G. and TOLOSANA-DELGADO, R. (2013). *Analyzing compositional data with R*. *Use R!* Springer, Heidelberg. https://doi.org/10.1007/978-3-642-36809-7 MR3099409

[62] ČENCOV, N. N. (1982). *Statistical decision rules and optimal inference*. *Translations of Mathematical Monographs* **53**. American Mathematical Society, Providence, RI Translation from the Russian edited by Lev J. Leifman. MR645898

[63] WATSON, D. F. and PHILIP, G. M. (1989). Measures of variability for geological data. *Mathematical Geology* **21** 233–254. https://doi.org/10.1007/BF00893217