WORKING PAPERS

N° 1601

December 2024

"Log-Free Distance and Covariance Matrix for Compositional Data II: the Projective/Exterior Product Approach"

Olivier Faugeras

Toulouse School of Economics

Log-Free Distance and Covariance Matrix for Compositional Data II: the Projective/Exterior Product Approach

Olivier P. Faugeras
Toulouse School of Economics, Université Toulouse 1 Capitole

December 4, 2024

Abstract

Motivated by finding a way to deal with Compositional Data (CoDa) with or without zeroes in a unified way, we build upon the previous projective geometry viewpoint of Faugeras (2023) and use the tools provided by the exterior product and Grassmann's algebra. These allow to represent higher dimensional subspaces as linear objects, called multi-vectors, on which the usual Euclidean scalar product can be extended. Applied to CoDa seen as equivalence classes, this allows to define a pseudo-scalar product and pseudo-norm. Depending on the normalization chosen, it is remarkable that the pseudo-norm obtained is either the same barycentric divergence which was derived in Faugeras (2024a) from the affine geometry viewpoint, or becomes a new, orthogonally invariant, genuine distance on the full non-negative CoDa space.

These tools are then used to lay the foundations for further statistical analysis of CoDa: we show how the relative position of a pair of CoDa around their means can be decomposed along its components to form exterior covariance, variance and correlation matrices, along with their corresponding global scalar measure of (co)variation. Gaussian distributions, Mahalanobis distance, Fréchet means, etc.. can then be introduced and we sketch their potential statistical applications. Eventually, we establish some connections with various notions encountered in the literature, like divergences based on quantifying inequalities, or canonical angles between subspaces. The paper is preceded by a tutorial on the exterior product, based on intuitive geometric visualization and familiar linear algebra, in order to make the ideas of the paper accessible to non-specialists.

Contents

1 Introduction

1.1 Aims and scope

The present article is a continuation of the geometric approach to compositional data (CoDa), initiated in Faugeras (2023) and Faugeras (2024a). These papers were motivated by i) finding a unified way to reconcile several viewpoints on CoDa analysis and ii) proposing log-free divergence and covariance matrices allowing to effectively handle CoDa with zeroes.

More precisely, we proposed in Faugeras (2023) to view CoDa as projective points $[\mathbf{x}]_+$ in the space $\mathbb{P}^d_+$ obtained by projectivization of the non-negative orthant cone $\mathbb{R}^{d+1}_+ := \{\mathbf{x} \in \mathbb{R}^{d+1}, \boldsymbol{x} \geq \mathbf{0}\}$: CoDa elements $[\mathbf{x}]_+$ are equivalence classes of non-negative vectors $\boldsymbol{x} \geq \mathbf{0}$, $\mathbf{x} \in \mathbb{R}^{d+1}$, where $[.]_+$ denotes the equivalence class for the positive scaling relation, viz.

$$\mathbf{y} \in [\mathbf{x}]_+ \Longleftrightarrow \exists \lambda > 0, \quad y = \lambda x,$$

with homogeneous coordinates $[x_0 : x_1 : \ldots : x_d]$. Geometrically, a projective CoDa point $[\mathbf{x}]_+$ corresponds to a ray in the non-negative orthant. Faugeras (2024a) studied the simplex,

$$\Delta^d_+ := \{\mathbf{x} \in \mathbb{R}^{d+1} : \mathbf{x}^T \mathbf{1} = 1, \boldsymbol{x} \geq \mathbf{0}\} \tag{1}$$

as a particular affine model of $\mathbb{P}^d_+$, where simplex representatives $\mathbf{x} \in \Delta^d_+$ of $[\mathbf{x}]_+ \in \mathbb{P}^d_+$ are affine points expressed in (normalized) barycentric coordinates (and not Cartesian coordinates). This affine geometric perspective lead us to define, among others, i) a novel family of log-free barycentric divergences, to measure the proximity of pairs of CoDa points, and also ii) corresponding barycentric variance matrices, to measure the proportionality of CoDa components. These constructs were based on formulating the displacement vector between two points, in terms of barycentric coordinates.

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

Remarkably, both approaches, the affine one based on barycentric coordinates of Faugeras (2024a) and the projective one based on the exterior product of the present article, lead to closely related notions of distance/divergence and covariance matrices. Together, both approaches give a twin framework for the analysis of CoDa in a direct, log-free, unified way, effectively handling zeroes. Each setting capitalizes on an underlying linear structure, that of an affine space in Faugeras (2024a), and that of the exterior algebra of multi-vectors in the present paper.

In spite of its power, the exterior product and Grassmann's algebra appears to be relatively unknown to statisticians[1], partly due to its inherent abstract nature. Therefore, the main objective of this paper is to try to bridge the gap between the abstract theory and its practical application towards statistical analysis of CoDa. Hence, some parts of the paper are expository, intended to explain the basics of the exterior product and Grassmann's algebra in a geometrically intuitive way so that its statistical application to CoDa be made understandable to the widest possible audience.

1.2 Outline

More precisely, the paper is organized as follows: in Section 2, we provide a limited tutorial on the exterior product and bi-vectors. It is based on geometric intuition, envisioning simple bi-vectors as oriented parallelograms, and linear algebra, defining bi-vectors through their linear representations as matrices. We show how the usual Euclidean scalar product and norm extends to these bi-vectors. Alternatively to Section 2, the reader can find in Appendix A a short recap of the exterior product and Grassmann's algebra from the abstract algebraic viewpoint.

Section 3 applies the above constructs to a pair of equivalence classes of vectors, i.e. to a pair of directions $[\mathbf{x}], [\mathbf{y}]$, or a pair of CoDa $[\mathbf{x}]_+, [\mathbf{y}]_+$. They are represented algebraically by homogeneous bi-vectors, which can be reduced to a

[1] See e.g. Dieudonne (1979).

single bi-vector by choosing suitable normalized representatives. This allows to transfer the scalar product and norm of bi-vectors into a pseudo scalar product and pseudo-norm for pairs of equivalence classes. For CoDa, the pseudo-norm obtained interprets as a divergence or even a distance, depending on the norm chosen to standardize CoDa. For the ℓ_1 normalization, one obtains the same 2-barycentric divergence of Faugeras (2024a), which was obtained from the affine geometry viewpoint, with a reasoning based on barycentric coordinates. For the ℓ_2 normalization, one obtains a genuine, log-free, bounded distance (i.e. satisfying the triangle inequality) on the full CoDa space (hence, allowing for zeroes in the components), which has the additional property of being orthogonally invariant. A variant using a square root transform on the simplex is also suggested.

Section 4 introduces the main statistical objects of the paper for a CoDa analysis based on such projective viewpoint aided with the exterior product. By applying the pseudo-scalar product and norms of section 3 to homogeneous bi-vectors $[\mathbf{x}]_+ \wedge [\boldsymbol{\mu}^{\mathbf{x}}]_+$, $[\mathbf{y}]_+ \wedge [\boldsymbol{\mu}^{\mathbf{y}}]_+$, where $[\boldsymbol{\mu}^{\mathbf{x}}]_+$ and $[\boldsymbol{\mu}^{\mathbf{y}}]_+$ are deterministic means of random $[\mathbf{x}]_+$, $[\mathbf{y}]_+$, one defines exterior covariance, variance and correlation matrices, in a fashion similar to the barycentric covariance and variance matrices of Faugeras (2024a). These matrices give orthogonally invariant decomposition of the variation/covariation of CoDa along its pairs of components. We briefly show in Section 5 how these measures of statistical variation can be used for further CoDa analysis, e.g., for defining Fréchet means and their variants (useful for clustering and regression), Mahalanobis distances, Gaussian-type distributions, and Wasserstein type distance of CoDa distributions.

At last, Section 6 draws some connections between the norm and scalar product of multi-vectors and various notions encountered in the literature. In particular, we show how the norm of bi-vectors is related to divergences based on quantifications, either of the Cauchy-Schwarz inequality, or of the likelihood ratio order. Next is the relation between the bi-vector norm and the polar sine, a generalization of the sine function to ratio of volumes. At last, we explain how the scalar product and norm of multi-vectors are related to the relative position between subspaces, expressed either via their canonical angles, or via their projection matrices. This gives additional geometrical insight on the newly defined exterior covariance, exterior correlation and pseudo-norm between CoDa. We eventually conclude in Section 7.

We follow the notations of Faugeras (2023) and Faugeras (2024a), with the exception of simply[2] denoting by $||.||$ the usual Euclidean ℓ_2 norm of vectors. A list of the main notations and conventions is collected in Appendix B.

[2]instead of $||.||_2$.

2 A primer on the exterior product and Grassmann algebra

2.1 Introduction

The exterior (or wedge) product $\wedge$ is one of Grassmann's (Hermann Grassmann (1878), H. Grassmann (1995)) genial contribution to geometry. It is defined as a multilinear and antisymmetric product of any number of vectors, and thus can be thought of as a sort of a multivariate "rectangular determinant"[3]. The exterior product algebraicizes the notion of linear independence as vectors are linearly dependent if and only if their exterior product is zero. It provides a fundamental product operation for elements of a linear space V, of dimension n, which allows to represent algebraically vector subspaces as the "product" of their lower-dimensional parts (e.g. a vector plane is represented by the product of two vectors). It generates a series of new linear spaces, $\bigwedge^k(V)$, $0 \leq k \leq n$, whose elements are called k-vectors[4] and whose direct sum defines a (closed) algebra of multi-vectors, the exterior power $\bigwedge(V)$. If the linear space V possesses a metric, the measure or magnitude of such multi-vectors may be interpreted as a length, area, volume, or hyper-volume according to the grade of the product. The exterior product is also closely related to other algebraic products like the tensor and the Clifford (geometric) products, and to matrix notions like minors, Gram and compound matrices. It underpins geometric and conformal algebra and gives a coordinate free, unified treatment of vector calculus and differential forms (à la Cartan).

There are several ways to introduce the exterior product and Grassmann's algebra, based on algebra alone, (see e.g. Kung, Rota, and Yan (2009), Chapter 6.6, Spivak (1965) or Federer (1969)), geometry (Khosravi and Michael D. Taylor (2008), Mikusinski and Michael D Taylor (2012), Postnikov (1988) Chapter 7), a mixed of those (Browne (2012), Winitzki (2009), Rosén (2019) Chapter 2), or even from the geometric (Clifford) product (see e.g. Doran and Lasenby (2003), Hestenes and Sobczyk (1984), Dorst, Fontijne, and Mann (2009), Perwass (2009)). A difficulty of the topic, in addition to its variety of approaches and ramifications, is the inherent abstract nature, both in the definition of the exterior product as a formal combination of symbols, and in the exterior algebra as a formal sum of elements of disparate nature[5], scalars, vectors, bi-vectors, tri-vectors, etc. More fundamentally, as we will explain thereafter, not all $k-$vectors[6] can be given a geometric meaning.

Fortunately, we will only require basic facts about simple bi-vectors obtained from the exterior product of two vectors and the extension of the scalar product to those simple bi-vectors (which do have a geometric interpretation). Although

[3] i.e. determinant of fewer than n vectors in an $n-$dimensional space.

[4] 0-vectors are identified with scalars.

[5] i.e. elements of different grade.

[6] This is the important distinction between *simple* k-vectors $\mathbf{x}_1 \wedge \ldots \wedge \mathbf{x}_k \in \bigwedge^k(V)$, built from the exterior product of k (usual) vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k \in V$ and *indecomposable or compound* $k-$vectors made of irreducible sums of at least two simple k-vectors.

the whole construction of the general exterior product between elements of different grades and the full algebra of the exterior power space gives better insight and understanding, what is essential for our purposes of defining distance and covariance for CoDa is the geometric intuition of simple bi-vectors as oriented parallelograms and the scalar product between them. Our aim in this section is thus to make the exterior product geometrically intuitive so that its application to CoDa be made understandable to the widest possible audience.

2.2 An intuitive geometric approach to simple bi-vectors as oriented parallelograms

We thus proceed to give an inductive, concrete, geometric tutorial on the exterior product and algebra of bi-vectors, whose ideas are mainly inspired by Mikusinski and Michael D Taylor (2012) and Khosravi and Michael D. Taylor (2008). The general philosophy is similar to the idea that in linear algebra, the abstract definition of a vector space, based solely on algebraic properties, as a set V, with the two familiar operations $(+, \times)$, is better grasped after one is accustomed to manipulating the primitive high-school geometrical/intuitive notion of vector as an "arrow" in space.

Indeed, when one first encounters the concept of a vector $\mathbf{x}$, it is usually described in purely geometric/physical/sensory terms, as an "arrow", i.e. as (an equivalence class[7] of) a directed line segment, free "floating" in space: it has geometric predicates,

i) a direction (determined by the line $\text{span}(\mathbf{x})$ starting from the origin and parallel to $\mathbf{x}$),

ii) a sense or orientation (so that the line is divided into two rays, one in the sense of $\mathbf{x}$, and the opposite one in the sense of $-\mathbf{x}$),

iii) and a magnitude (the length $||\mathbf{x}||$ of the segment),

but is located nowhere in space (two directed line segments having same direction, orientation and magnitude, i.e. forming a parallelogram, are considered equal). An algebraic description is obtained by associating to the vector $\mathbf{x}$ a sequence of real numbers $x_i \in \mathbb{R}$, $i = 1, \ldots, n$, called the components, relative to an (orthogonal) coordinate system: each coordinate component x_i corresponds to the (signed) length of the projection of $\mathbf{x}$ on the coordinate axis $\text{span}(\mathbf{e}_i)$, where $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ is the canonical basis of $\mathbb{R}^n$. Pythagoras's theorem ensure that the squared length of the vector is the sum of the squared length of its projected components.

Such a geometric-to-analytic-to-abstract-algebraic approach to the linear algebra of vectors can similarly be employed for bi-vectors: we first apprehend a pair $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^n \times \mathbb{R}^n$ of vectors as a single geometric object, to which one can similarly associate notions of magnitude and coordinate components. In order to take into account that a pair of vectors is a bivariate object, we will have to

[7] i.e. equipollent.

replace the concepts of "arrow", (signed) length, and components w.r.t. the coordinate axis, with their suitable $2-$dimensional generalizations: parallelogram, (signed) area, coordinates w.r.t. to planes made of pairs of coordinate axis.

We thus conceives geometrically a pair of vectors $(\mathbf{x}, \mathbf{y})$ as the parallelogram $\{s\mathbf{x} + t\mathbf{y}, 0 \leq s, t \leq 1\}$ induced by them, with one vertex at the origin. We will denote by

$$\mathbf{x} \wedge \mathbf{y}$$

such a parallelogram, for reasons which will become clear below and will call it a simple bi-vector. Geometrically, $\mathbf{x} \wedge \mathbf{y}$ has

i) a direction, the plane $\text{span}(\mathbf{x}, \mathbf{y})$ corresponding to the subspace spanned by the vectors $(\mathbf{x}, \mathbf{y})$,

ii) an orientation, from $\mathbf{x}$ to $\mathbf{y}$ (which can be symbolically represented as a turning arrow from $\mathbf{x}$ to $\mathbf{y}$),

iii) a magnitude, $||\mathbf{x} \wedge \mathbf{y}||$, defined as the area of the parallelogram (it will be defined precisely below.)

Components to this parallelogram/ simple bi-vector are attached as follows: for each ordered pair $i < j$ of indices, the coordinate axis $\text{span}(\mathbf{e}_i, \mathbf{e}_j)$ determines an (i, j)-plane. For $\mathbf{x} = (x_1, \ldots, x_n)$, $\mathbf{y} = (y_1, \ldots, y_n)$, the projection of the parallelogram into this plane is the parallelogram

$$\begin{pmatrix} x_i \\ x_j \end{pmatrix} \wedge \begin{pmatrix} y_i \\ y_j \end{pmatrix},$$

generated by the two dimensional projected vectors (x_i, x_j) and (y_i, y_j) on this $(i, j)-$plane. The (i, j) component of the pair $\mathbf{x} \wedge \mathbf{y}$ will be defined as the oriented[8] area of this orthogonal projection. Since (x_i, x_j) and (y_i, y_j) are two vectors in the plane $\mathbb{R}^2$, the oriented area of the parallelogram $(x_i, x_j) \wedge (y_i, y_j)$ is given by the determinant

$$p_{ij} := p_{ij}(\mathbf{x} \wedge \mathbf{y}) := \det \begin{vmatrix} x_i & y_i \\ x_j & y_j \end{vmatrix} = x_i y_j - x_j y_i. \tag{2}$$

This yields $\binom{n}{2}$ components, which we call Plücker components.

2.3 Matrix descriptions of $\mathbf{x} \wedge \mathbf{y}$

2.3.1 Plücker matrix

These Plücker components can be arranged in an $n \times n$ anti-symmetric matrix, which we call the Plücker matrix:

[8] We consider oriented areas in order to allow for negative components.

Definition 2.1. *The Plücker matrix of the simple bi-vector* $\mathbf{x} \wedge \mathbf{y}$ *is the* $n \times n$ *anti-symmetric matrix*

$$P := P(\mathbf{x} \wedge \mathbf{y}) := (p_{ij})_{n \times n}, \tag{3}$$

with $p_{ij} = -p_{ji}$*, given by (2).*

Note that the Plücker matrix also writes as

$$P = \mathbf{x}\mathbf{y}^T - \mathbf{y}\mathbf{x}^T.$$

Remark 1. *The matrix* $\mathbf{x}\mathbf{y}^T$ *corresponds to the linear application* $\mathbf{x} \otimes \mathbf{y} : \mathbf{V} \to \mathbf{V}$*, given by* $\mathbf{u} \mapsto \mathbf{x}\langle \mathbf{y} | \mathbf{u} \rangle$*, which is the outer product of the two vectors* $\mathbf{x}, \mathbf{y}$*.* $\mathbf{x} \otimes \mathbf{y}$ *is bilinear w.r.t. to each components* $\mathbf{x}, \mathbf{y}$*. Its generalization to tensors, i.e. multidimensional arrays, gives the tensor product of two tensors and the general construction of tensor algebra. Hence,* $\mathbf{x} \wedge \mathbf{y}$*, identified with its Plücker matrix representation* $P(\mathbf{x} \wedge \mathbf{y})$*, corresponds to the anti-symmetrization of the outer product, formally*[9]*,* $\mathbf{x} \wedge \mathbf{y} = \mathbf{x} \otimes \mathbf{y} - \mathbf{y} \otimes \mathbf{x}$*. This is one possible abstract algebraic definition of the exterior product, as found e.g. in Spivak (1965). Closely related to the outer product is the Kronecker product, defined as the vectorization (i.e. stacking of the matrix components in a single long vector) of the outer product.*

2.3.2 Compound matrix

Another way to describe matricially this coordonatization of $\mathbf{x} \wedge \mathbf{y}$ is via what is known in the literature as compound matrices, see Aitken (1956) Chapter 5, Prells, Friswell, and Garvey (2003), Boutin, Gleeson, and Williams (1996). The general description of compound matrices is as follows: for a matrix $A = [a_{ij}] \in \mathbb{R}^{n \times m}$ and subsets $I \subset \{1, \ldots, n\}$, $J \subset \{1, \ldots, m\}$, denote by $A_{I,J}$ the submatrix of A with rows and columns taken from I, J, respectively. For I and J of same cardinality $p \leq \min(n, m)$, recall that the I, J-minor of order p of A is the number $\det A_{I,J}$. Then, the compound matrix $C_p(A)$ of order p of A is defined as the matrix of minors of order p of A, where the different subsets I, resp. J, are indexed in lexicographic order $\prec$, so that one can enumerate these subsets in increasing order, viz. $I_1 \prec I_2 \prec \ldots \prec I_{\binom{n}{p}}$, resp. $J_1 \prec J_2 \prec \ldots \prec J_{\binom{m}{p}}$:

$$\begin{aligned} C_p(A) := \big(\det A_{I,J}\big)\,, \quad & I \subset \{1, \ldots, n\}, J \subset \{1, \ldots, m\}, \\ & \text{card } I = \text{card } J = p. \end{aligned} \tag{4}$$

When $m = p$, the $p-$compound of the matrix $A \in \mathbb{R}^{n \times p}$ reduces to a vector of size $\binom{n}{p}$, since in this case the subsets J reduce to the single set $J = \{1, \ldots, p\}$.

In our case, the connection with Plücker matrices of simple bi-vectors is as follows: for a simple bi-vector $\mathbf{x} \wedge \mathbf{y}$, with $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, define its $A-$matrix as

$$A := A(\mathbf{x} \wedge \mathbf{y}) := (\mathbf{x} \quad \mathbf{y}), \tag{5}$$

[9] usually with a factor 1/2 added.

i.e. the $n \times 2$ matrix whose column vectors are $\mathbf{x}, \mathbf{y}$. Then, the $2,2-$minors of A are precisely the Plücker components (2): for $I = \{i,j\}$, $i < j$, in (4), $\det A_{\{i,j\},\{1,2\}} = x_i y_j - x_j y_i$. Hence, the 2-compound matrix/vector of A,

$$C_2(A) = \begin{pmatrix} \vdots \\ \det A_{\{i,j\},\{1,2\}} \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ x_i y_j - x_j y_i \\ \vdots \end{pmatrix}. \tag{6}$$

of size $\binom{n}{2} \times 1$ corresponds to the vectorization of the Plücker components (2), equivalently, the vectorization of the upper triangle of the Plücker matrix (3). Let us formalize this in a definition.

Definition 2.2. *The compound vector representation of* $\mathbf{x} \wedge \mathbf{y}$ *is the vector* $C_2(A(\mathbf{x} \wedge \mathbf{y})) \in \mathbb{R}^{\binom{n}{2}}$*, given by (6).*

2.4 The vector space $\bigwedge^2(V)$ of bi-vectors and the exterior product.

2.4.1 Vectorization of parallelograms from their linear representations and the exterior product

So far, we have just defined $\mathbf{x} \wedge \mathbf{y}$ as geometric objects, oriented parallelograms, with attributes like direction, orientation, magnitude and components. In particular, these objects do not possess yet a vector space structure (which would make possible to add simple bi-vectors together). However, since the components of simple bi-vectors are represented by linear objects, either antisymmetric (Plücker) matrices (3) or (compound) vectors (6), which are elements of vector spaces, it is natural to endow the set of simple bi-vectors with the corresponding addition and scalar multiplication of their components: we simply define the addition $+$ and scalar multiplication . on simple bi-vectors, as the addition and scalar multiplication of their Plücker matrices, equivalently their compound vectors. Formally, we set, for $i < j$, $\mathbf{x}, \mathbf{y}, \mathbf{r}, \mathbf{s} \in \mathbb{R}^n$, and $\lambda \in \mathbb{R}$,

$$\begin{aligned} p_{ij}(\mathbf{x} \wedge \mathbf{y} + \mathbf{r} \wedge \mathbf{s}) &:= p_{ij}(\mathbf{x} \wedge \mathbf{y}) + p_{ij}(\mathbf{r} \wedge \mathbf{s}) \\ p_{ij}(\lambda(\mathbf{x} \wedge \mathbf{y})) &:= \lambda p_{ij}(\mathbf{x} \wedge \mathbf{y}) \end{aligned}$$

The set of objects thus obtained by finite linear combination of simple bi-vectors by these operations are called bi-vectors and belong, by definition, to a vector space, the exterior power $\bigwedge^2(\mathbb{R}^n)$, of dimension $\binom{n}{2}$.

This extension of simple bi-vectors into elements of a linear space[10] allows to turn the formal symbol $\wedge$ into an operation between vectors: by bilinearity and antisymmetry of the 2×2 determinant defining p_{ij} (or by direct verification),

[10] The vector space structure of $\bigwedge^2(V)$ explains the terminology bi-*vectors* for $\mathbf{x} \wedge \mathbf{y}$ and linear combinations thereof.

one has that

$$
\begin{aligned}
p_{ij}((\mathbf{x}+\mathbf{y})\wedge\mathbf{z})) &= p_{ij}(\mathbf{x}\wedge\mathbf{z}) + p_{ij}(\mathbf{y}\wedge\mathbf{z})) \\
p_{ij}(\mathbf{x}\wedge(\mathbf{y}+\mathbf{z})) &= p_{ij}(\mathbf{x}\wedge\mathbf{y}) + p_{ij}(\mathbf{x}\wedge\mathbf{z}) \\
p_{ij}(\lambda(\mathbf{x}\wedge\mathbf{y})) &= p_{ij}((\lambda\mathbf{x})\wedge\mathbf{y}) = p_{ij}(\mathbf{x}\wedge(\lambda\mathbf{y})) \\
p_{ij}(\mathbf{y}\wedge\mathbf{x}) &= -p_{ij}(\mathbf{x}\wedge\mathbf{y})
\end{aligned}
$$

Thus, $\wedge$ amounts to a product on V,

$$\wedge : V \times V \to \bigwedge^2(V)$$

called the exterior product, which is bilinear and antisymmetric on V.

As explained in the introduction, an important distinction is between *simple* bi-vectors, i.e. bi-vectors which can be reduced to a single expression $\mathbf{x}_1 \wedge \mathbf{x}_2$, where each $\mathbf{x}_1$, $\mathbf{x}_2$ are two (regular) vectors in $\mathbb{R}^n$, and general *indecomposable* or *compound*[11] bi-vectors, made of sums (or linear combinations) of at least two simple bi-vectors: only simple bi-vectors have a geometric interpretation[12] as oriented parallelograms, with corresponding attributes. This is the unavoidable abstraction in the construction of the vector space $\bigwedge^2(V)$: in general, parallelograms can not be "added" together to give a parallelogram. They only give bi-vectors. Algebraically speaking, not every antisymmetric matrix, resp. vector of size $\binom{n}{2}$, can be obtained from a Plücker matrix $P(\mathbf{x}\wedge\mathbf{y})$ of two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, resp. as the compound vector of an $A-$matrix $A(\mathbf{x}\wedge\mathbf{y})$.

Remark 2. *i) In a three-dimensional space, i.e. for $n = 3$, $\binom{n}{2} = 3$. Hence, $C_2(A) \in \mathbb{R}^3$ identifies with a vector*[13] *of the original space V. This allows to view $\wedge$ as a binary operation $V \times V \to V$. Then, $\mathbf{x}\wedge\mathbf{y}$ corresponds, up to some sign change and permutation of the components*[14]*, to the well-known cross-product $\mathbf{x} \times \mathbf{y}$ of three-dimensional vectors.*

ii) In addition, for $n = 3$, every nonzero bi-vector of $\bigwedge^2(\mathbb{R}^3)$ reduces to a simple bi-vector: by bilinearity and antisymmetry of $\wedge$, one can write any bi-vector $\mathbf{x} \in \bigwedge^2(\mathbb{R}^3)$ on the basis $(\mathbf{e}_1\wedge\mathbf{e}_2, \mathbf{e}_1\wedge\mathbf{e}_3, \mathbf{e}_2\wedge\mathbf{e}_3)$ of $\bigwedge^2(\mathbb{R}^3)$ as

$$\mathbf{x} = x_1\mathbf{e}_1\wedge\mathbf{e}_2 + x_2\mathbf{e}_1\wedge\mathbf{e}_3 + x_3\mathbf{e}_2\wedge\mathbf{e}_3,$$

where $x_1, x_2, x_3 \in \mathbb{R}$. Assume w.l.o.g. that $x_1 \neq 0$ (otherwise pick another component which is nonzero). Then, again by bilinearity and antisymmetry, $\mathbf{x}$ writes as the simple bi-vector

$$\mathbf{x} = x_1\left(\mathbf{e}_1 - \frac{x_3}{x_1}\mathbf{e}_3\right)\wedge\left(\mathbf{e}_2 + \frac{x_2}{x_1}\mathbf{e}_3\right).$$

[11] One should not confuse the notion of compound bi-vectors with the compound vector (6) derived from the compound matrix (4). The terminology is somehow unfortunate but well-established in the literature.

[12] This is in contrast with classical linear algebra where all abstract vector elements can be given geometric meaning as an "arrow" in space.

[13] often called a pseudo-vector in physics.

[14] corresponding to Hodge dualization of the bi-vector, i.e. $\mathbf{x}\times\mathbf{y} := \star(\mathbf{x}\wedge\mathbf{y})$, see e.g. Kanatani (2015) Proposition 5.7 p. 71. for details.

2.4.2 Geometric implications for simple bi-vectors

For simple bi-vectors, the algebraic properties of $\wedge$ have geometric implications: let $\lambda \in \mathbb{R}$, then,

$$(\mathbf{x} + \lambda\mathbf{y}) \wedge \mathbf{y} = \mathbf{x} \wedge \mathbf{y} + \lambda\mathbf{y} \wedge \mathbf{y} = \mathbf{x} \wedge \mathbf{y},$$

and for $\lambda \neq 0$,

$$(\lambda\mathbf{x}) \wedge \left(\frac{1}{\lambda}\mathbf{y}\right) = \frac{\lambda}{\lambda}\mathbf{x} \wedge \mathbf{y} = \mathbf{x} \wedge \mathbf{y}.$$

The geometric meaning is that simple bi-vectors remain the same under shear mappings $(\mathbf{x}, \mathbf{x}) \mapsto (\mathbf{x} + \lambda\mathbf{y}, \mathbf{y})$, and squeeze mappings $(\mathbf{x}, \mathbf{x}) \mapsto (\lambda\mathbf{x}) \wedge \left(\frac{1}{\lambda}\mathbf{y}\right)$ (which both preserve the area of the parallelogram): this corresponds to defining simple bi-vectors as equivalence classes of parallelograms, where the equivalence relation is defined by a set of elementary transformations composed of shear and squeeze mappings, see Postnikov (1988) Chapter 7. One thus has an analogue for simple bi-vectors of the definition of vectors as equivalence class of directed line segments under translations. In other words, the shape of the parallelogram does not matter, only its direction, orientation and area-magnitude (yet to be precisely defined). In particular, one can reshape the parallelogram $\mathbf{x} \wedge \mathbf{y}$ in order to make it rectangular. Another way of saying the same thing is that only the part $\mathbf{y}^{\perp}$ of $\mathbf{y}$ perpendicular to $\mathbf{x}$ is involved in the exterior product, i.e.

$$\mathbf{x} \wedge \mathbf{y} = \mathbf{x} \wedge \mathbf{y}^{\perp}, \quad \text{with } \mathbf{y}^{\perp} := \mathbf{y} - \frac{\langle \mathbf{y}|\mathbf{x}\rangle}{||\mathbf{x}||^2}\mathbf{x}.$$

At last, changing the order of the terms in the product, i.e. the identity

$$\mathbf{y} \wedge \mathbf{x} = -\mathbf{x} \wedge \mathbf{y}$$

interprets as reversing the orientation of the parallelogram (from $\mathbf{y}$ to $\mathbf{x}$).

2.5 Norm and scalar product of bi-vectors

2.5.1 Geometric approach to the magnitude of simple bi-vectors via Gram matrices

Following our general philosophy, we first define an intuitive notion of magnitude for simple bi-vectors from geometric intuition, before proceeding to show it corresponds to a norm derived from a scalar product on bi-vectors.

Recall that for two rectangular matrices $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times m}$, with $m \leq n$, the Cauchy-Binet formula yields

$$\det(AB) = \sum_S \det(A_{\{1,\ldots,m\},S}) \det(B_{S,\{1,\ldots,m\}}), \tag{7}$$

where S is a subset of $\{1, \ldots, n\}$ with m-elements, $\det(A_{\{1,\ldots,m\},S})$, resp., $\det(B_{S,\{1,\ldots,m\}})$ are the $m-$minors of A, resp. B, and the sum is over all subsets S (see e.g. Gantmacher (1998)).

Now, let $A = A(\mathbf{x} \wedge \mathbf{y})$ be the $A-$matrix of (5). From A, one can construct

$$G := G(\mathbf{x} \wedge \mathbf{y}) := A^T A = \begin{pmatrix} \langle \mathbf{x}|\mathbf{x}\rangle & \langle \mathbf{x}|\mathbf{y}\rangle \\ \langle \mathbf{x}|\mathbf{y}\rangle & \langle \mathbf{y}|\mathbf{y}\rangle \end{pmatrix} \tag{8}$$

the symmetric, positive semi-definite, Gram matrix of size 2×2 associated with the pair of vectors $\mathbf{x} \wedge \mathbf{y}$. Then, as is well-known (see e.g. Mikusinski and Michael D Taylor (2012) p. 36), $\det(G)$ is the square of the area (the 2-dimensional volume or measure of magnitude) of the parallelogram $\mathbf{x} \wedge \mathbf{y}$.

Applied to the Gram matrix $G = A^T A$, the Cauchy-Binet formula (7), together with (2), yields

$$\begin{aligned} \det(G) &= \sum_{1 \le i < j \le n} \det \begin{pmatrix} x_i & y_i \\ x_j & y_j \end{pmatrix}^T \det \begin{pmatrix} x_i & y_i \\ x_j & y_j \end{pmatrix} \\ &= \sum_{1 \le i < j \le n} \det G \left(\begin{pmatrix} x_i \\ x_j \end{pmatrix} \wedge \begin{pmatrix} y_i \\ y_j \end{pmatrix} \right) \\ &= \sum_{1 \le i < j \le n} p_{ij}^2 = \sum_{1 \le i < j \le n} (x_i y_j - x_j y_i)^2 \end{aligned} \tag{9}$$

The geometric meaning of (9) is that the squared area of the parallelogram $\mathbf{x} \wedge \mathbf{y}$ is equal to the sum of the squares of the areas of its projections on all coordinate planes. One thus have a generalization of Pythagoras's theorem for length of vectors to areas of parallelogram, via Plücker components.

2.5.2 Scalar product and norms of parallelograms from Froebenius scalar product of Plücker matrices

Having linear representations of bi-vectors as Plücker matrices (3), or as compound vectors (6), it is natural to use the classical scalar product for matrices or vectors and corresponding norms to endow $\bigwedge^2(V)$ with a scalar product, effectively turning $\bigwedge^2(V)$ into a finite dimensional Hilbert space. In particular, the geometric interpretation of (9) suggest to use the Froebenius inner product between 2 matrices $A = (a_{ij}), B = (b_{ij})$ of the same size,

$$\langle A|B\rangle_F := \mathrm{Trace}(A^T B) = \sum_{i,j} a_{ij} b_{ij},$$

which is the usual generalization to matrices of the Euclidean scalar product of vectors. We thus define:

Definition 2.3. *The Plücker scalar product between two pairs of simple bi-vectors* $\mathbf{x} \wedge \mathbf{y}$*,* $\mathbf{r} \wedge \mathbf{s}$ *is defined as the Froebenius scalar product of their Plücker* *matrices, viz.*

$$
\begin{aligned}
\langle \mathbf{x}\wedge\mathbf{y}|\mathbf{r}\wedge\mathbf{s}\rangle &:= \frac{1}{2}\langle P(\mathbf{x}\wedge\mathbf{y})|P(\mathbf{r}\wedge\mathbf{s})\rangle_F && (10)\\
&= \sum_{i<j} p_{ij}(\mathbf{x},\mathbf{y})p_{ij}(\mathbf{r},\mathbf{s})\\
&= \det\left(A(\mathbf{x}\wedge\mathbf{y})^T A(\mathbf{r}\wedge\mathbf{s})\right) && (11)\\
&= \det\begin{vmatrix}\langle\mathbf{x}|\mathbf{r}\rangle & \langle\mathbf{x}|\mathbf{s}\rangle\\ \langle\mathbf{y}|\mathbf{r}\rangle & \langle\mathbf{y}|\mathbf{s}\rangle\end{vmatrix} = \langle\mathbf{x}|\mathbf{r}\rangle\langle\mathbf{y}|\mathbf{s}\rangle - \langle\mathbf{x}|\mathbf{s}\rangle\langle\mathbf{y}|\mathbf{r}\rangle, && (12)
\end{aligned}
$$

where we have normalized the Froebenius product by $1/2$*, so that (11) follows from the Cauchy-Binet-Formula (7). It is then extended by linearity to a genuine scalar product on (possibly indecomposable) bi-vectors of* $\bigwedge^2(V)$*.*

Obviously, this gives the same scalar product on $\bigwedge^2(V)$ which would have resulted from using the usual Euclidean scalar product $\langle .|.\rangle$ of $\mathbb{R}^{\binom{n}{2}}$ and the representation of bi-vectors via compound matrices, i.e.

$$
\langle \mathbf{x}\wedge\mathbf{y}|\mathbf{r}\wedge\mathbf{s}\rangle = \langle C_2(A(\mathbf{x}\wedge\mathbf{y}))|C_2(A(\mathbf{r}\wedge\mathbf{s}))\rangle,
$$

where $A(\mathbf{x}\wedge\mathbf{y})$, resp., $A(\mathbf{r}\wedge\mathbf{s})$ are the $A-$matrices (5) for $\mathbf{x}\wedge\mathbf{y}$, resp. $\mathbf{r}\wedge\mathbf{s}$.

Definition 2.3 yields a corresponding norm on $\bigwedge^2(V)$, which we call the Plücker norm:

$$
\begin{aligned}
||\mathbf{x}\wedge\mathbf{y}|| &:= \sqrt{\langle\mathbf{x}\wedge\mathbf{y}|\mathbf{x}\wedge\mathbf{y}\rangle} && (13)\\
&= \sqrt{\det\, G(\mathbf{x}\wedge\mathbf{y})} = \sqrt{||\mathbf{x}||^2||\mathbf{y}||^2 - \langle\mathbf{x}|\mathbf{y}\rangle^2} && (14)\\
&= \sqrt{\frac{1}{2}\mathrm{trace}(P\,(\mathbf{x}\wedge\mathbf{y})^T P(\mathbf{x}\wedge\mathbf{y}))} = \sqrt{\frac{1}{2}||P(\mathbf{x}\wedge\mathbf{y})||_F^2}\\
&= \sqrt{\sum_{1\le i<j\le n} p_{ij}^2} = \sqrt{\sum_{1\le i<j\le n}(x_i y_j - x_j y_i)^2} && (15)\\
&= ||C_2(A(\mathbf{x}\wedge\mathbf{y}))|| && (16)
\end{aligned}
$$

where the norm in (16) is the Euclidean norm on $\mathbb{R}^{\binom{n}{2}}$. Equation (14) shows that one obtains by linear algebra the same norm for simple bi-vectors as the one derived in (9) from the geometric standpoint. Hence, the Plücker norm of simple bi-vectors interprets geometrically as the area of the parallelogram, as explained before. Such geometric interpretation is also clear from (14): if one introduces the angle $\theta = \theta(\mathbf{x},\mathbf{y}) \in [0,\pi]$ between $\mathbf{x}$ and $\mathbf{y}$, defined via

$$
\langle\mathbf{x}|\mathbf{y}\rangle = ||\mathbf{x}||.||\mathbf{y}||.\cos(\theta),
$$

then, (14) gives

$$
||\mathbf{x}\wedge\mathbf{y}|| = ||\mathbf{x}||.||\mathbf{y}||.\sin(\theta) \qquad (17)
$$

and one recovers the familiar formula of the area of a parallelogram, base length times height.

Obviously, it is readily seen from the linear representation of bi-vectors that Cauchy-Schwarz's inequality for vectors extends to bi-vectors, viz.

$$|\langle \mathbf{x} \wedge \mathbf{y} | \mathbf{r} \wedge \mathbf{s} \rangle| \leq ||\mathbf{x} \wedge \mathbf{y}|| \, ||\mathbf{r} \wedge \mathbf{s}||.$$

In addition, the definition of the Plücker norm is consistent in the sense that it is invariant w.r.t. orthogonal transformation: if U is an orthogonal matrix, then

$$||U\mathbf{x} \wedge U\mathbf{y}|| = ||\mathbf{x} \wedge \mathbf{y}|| \tag{18}$$

This follows from the expression (14) and the orthogonal invariance of the Euclidean norm and scalar product. It can also be seen directly from the definition of the Plücker norm in terms of A-matrix (5): if $\mathbf{x}' := U\mathbf{x}$, $\mathbf{y}' := U\mathbf{y}$, then $A(\mathbf{x}' \wedge \mathbf{y}') = UA(\mathbf{x} \wedge \mathbf{y})$. Thus, by (14)

$$\begin{aligned} ||U\mathbf{x} \wedge U\mathbf{y}|| &= \sqrt{\det G((\mathbf{x}' \wedge \mathbf{y}')} = \sqrt{\det(A^T(\mathbf{x}' \wedge \mathbf{y}')A(\mathbf{x}' \wedge \mathbf{y}'))} \\ &= \sqrt{\det(A^T(\mathbf{x} \wedge \mathbf{y})U^T UA(\mathbf{x} \wedge \mathbf{y}))} = \sqrt{\det G(\mathbf{x} \wedge \mathbf{y})} \end{aligned}$$

Remark 3. *For a single vector* $\mathbf{x}$*, its Plücker norm, understood from its definition (14) in terms of* $A-$*matrix, yields the usual Euclidean* ℓ_2 *norm of* $\mathbf{x}$*, as*

$$||\mathbf{x}|| := \sqrt{\det(A(\mathbf{x})^T A(\mathbf{x}))} = \sqrt{\mathbf{x}^T\mathbf{x}} = \sqrt{\langle \mathbf{x}|\mathbf{x}\rangle} = ||\mathbf{x}||.$$

This explains why we use the same notation $||.||$ *for the Plücker norm of bi-vectors and for the Euclidean norm of vectors. Similar comments apply to the scalar product of bi-vectors.*

Remark 4. *These very basic parts of the theory are sufficient for our purposes of using the exterior product to CoDa analysis. Nevertheless, it is possible to generalize, first to the construction of simple* $k-$*vectors,* $\mathbf{x}_1 \wedge \ldots \wedge \mathbf{x}_k$*, geometrically interpreted as* $k-$*dimensional parallelotopes, then extended by linearity to the corresponding exterior powers vector space* $\wedge^k V$*, and eventually to the exterior algebra* $\bigwedge V$*, so that* $\wedge$ *becomes a genuine multi-linear and anti-symmetric operation between multi-vectors. The interested reader may consult Mikusinski and Michael D Taylor (2012) and/or Khosravi and Michael D. Taylor (2008) for details. Alternatively, we summarize, for the convenience and comparison purposes of the interested reader, the abstract algebraic approach to Grassmann's algebra in Appendix A.*

3 From pairs of vectors to pairs of directions and CoDa

3.1 Plücker coordinates for pairs of directions and CoDa

For the application of the exterior product to CoDa, we first need to homogenize vectors and bi-vectors. Indeed, in the projective viewpoint of Faugeras (2023), owing to the scale invariance, a CoDa element $[\mathbf{x}]_+$, with vector representative[15] $\mathbf{x} \in \mathbb{R}_+^{d+1}$, is an equivalence class for the positive scaling relation, as explained in the introduction. Hence, $[\mathbf{x}]_+$ represents a direction (in the non-negative orthant), stripped of its magnitude content. Similarly, for a vector $\mathbf{x} \in \mathbb{R}^{d+1}$, the equivalence class $[\mathbf{x}] = \{\lambda\mathbf{x}, \lambda \in \mathbb{R}^*\}$ for the collinearity equivalence relation $\sim$,

$$\mathbf{x} \sim \mathbf{y} \Leftrightarrow \exists \lambda \in \mathbb{R}^* \text{ s.t. } \mathbf{x} = \lambda\mathbf{y},$$

represents the directional part $\mathrm{span}(\mathbf{x})$ attached to the vector $\mathbf{x}$ and is pictured geometrically as the (two-sided) line through the origin parallel to $\mathbf{x}$.

For simple bi-vectors, the Plücker coordinates p_{ij} of $\mathbf{x} \wedge \mathbf{y}$ induces homogeneous coordinates for the homogenized versions of (ordered) pairs $([\mathbf{x}], [\mathbf{y}])$ of directions, resp. $([\mathbf{x}]_+, [\mathbf{y}]_+)$ of pairs of CoDa. Indeed, if $\mathbf{x}, \mathbf{y}$ are replaced by $\lambda\mathbf{x}, \mu\mathbf{y}$, $\lambda, \mu \neq 0$, then

$$p_{ij}(\lambda\mathbf{x}, \mu\mathbf{y}) = \lambda\mu . p_{ij}(\mathbf{x}, \mathbf{y}),$$

i.e. $p_{ij}(\lambda\mathbf{x}, \mu\mathbf{y})$ is a scalar multiple of $p_{ij}(\mathbf{x}, \mathbf{y})$. Considering the set of homogeneous coordinates

$$[P] := \{\lambda P, \lambda \in \mathbb{R}^*\},$$

i.e. scalar multiples of the Plücker matrix (3), gives a coordonatization of a pair of directions $[\mathbf{x}] \wedge [\mathbf{y}]$. Equivalently, one can take as coordonatization scalar multiples of the compound vector (6), i.e.

$$[C_2(A(\mathbf{x} \wedge \mathbf{y}))] := \{\lambda C_2(A(\mathbf{x} \wedge \mathbf{y})), \lambda \in \mathbb{R}^*\},$$

Here, $[\mathbf{x}] \wedge [\mathbf{y}]$ interprets as the set of parallelogram $(\lambda\mathbf{x}) \wedge (\mu\mathbf{y})$ obtained from $\mathbf{x}, \mathbf{y}$ by changing their amplitudes (and orientation for $\lambda\mu < 0$) and thus can be identified with the whole vector subspace $\mathrm{span}(\mathbf{x}, \mathbf{y})$. Taking (non-zero) scalar multiple of elements of a vector space corresponds to the operation of projectivization of a vector space in projective geometry. We have just described what is known in the literature as the Plücker embedding of the Grassmannian $G(2, d+1)$ into the projectivization of the exterior algebra, here $\mathbb{P}(\bigwedge^2 \mathbb{R}^{d+1})$.

For compositional data, taking the non-negativity constraint into account, one obtains

$$[P]_+ = \{\lambda P, \lambda > 0\},$$

with P the Plücker matrix (3), or, equivalently,

$$[C_2(A)]_+ = \{\lambda C_2(A), \lambda > 0\},$$

with $C_2(A)$ the compound vector (6), as homogeneous coordinates for the pair $[\mathbf{x}]_+ \wedge [\mathbf{y}]_+$. The latter pair identifies geometrically with the convex (polyhedral) cone $\mathrm{cone}(\mathbf{x}, \mathbf{y}) = \{\lambda\mathbf{x} + \mu\mathbf{y}, \lambda, \mu > 0\}$.

Summarizing, one has the formal statement on pairs of directions and CoDa:

$$[\mathbf{x}] \wedge [\mathbf{y}] = [\mathbf{x} \wedge \mathbf{y}], \quad \text{and } [\mathbf{x}]_+ \wedge [\mathbf{y}]_+ = [\mathbf{x} \wedge \mathbf{y}]_+.$$

[15] From now on, we set the dimension $n = d + 1$, in order to match the notations of Faugeras (2023) and Faugeras (2024a).

3.2 Non-homogeneous coordinates of pairs of directions and CoDa

Homogeneous systems of coordinates, being scalar multiples, do not associate to a direction $[\mathbf{x}]$, resp. a CoDa point $[\mathbf{x}]_+$, a unique set of numbers. One can de-homogenize, i.e. get a single set of coordinate numbers by taking a representative of the equivalence class $[\mathbf{x}]$, resp. $[\mathbf{x}]_+$, obtained by standardizing the vector $\mathbf{x}$ by a norm N. Indeed, by the property of a norm, for any scalar $\lambda \neq 0$, and $\mathbf{x} \neq \mathbf{0}$,

$$\frac{\lambda \mathbf{x}}{N(\lambda \mathbf{x})} = \frac{\lambda \mathbf{x}}{|\lambda| N(\mathbf{x})} = \operatorname{sign}(\lambda) \frac{\mathbf{x}}{N(\mathbf{x})} = \pm \frac{\mathbf{x}}{N(\mathbf{x})},$$

so that the vector $\mathbf{x}/N(\mathbf{x})$ is stripped of the magnitude content of $\mathbf{x}$, retaining only its direction (up to sign).

In turn, one obtains non-homogeneous coordinates of pairs of directions $[\mathbf{x}] \wedge [\mathbf{y}]$ by taking as representative the single $2-$vector obtained by taking the exterior product of the normalized representatives of $[\mathbf{x}], [\mathbf{y}]$, viz.

$$\frac{\mathbf{x}}{N(\mathbf{x})} \wedge \frac{\mathbf{y}}{N(\mathbf{y})}.$$

For directional data, it makes sense to take as norm the usual Euclidean ℓ_2 norm $N(.) = ||.||$, i.e. to consider the radial projection $\mathcal{S}$ on the unit sphere, viz.

$$\mathcal{S}(\mathbf{x}) := \frac{\mathbf{x}}{||\mathbf{x}||}, \quad \mathbf{x} \neq \mathbf{0}. \tag{19}$$

Thus, a direction $[\mathbf{x}]$ is represented by the opposite pair $\{\mathbf{x}/||\mathbf{x}||, -\mathbf{x}/||\mathbf{x}||\}$ of unit-norm vectors, i.e. a pair of antipodal points on the unit sphere. This yields as normalized representative of the pair of directions $[\mathbf{x}] \wedge [\mathbf{y}]$, the pair of simple bi-vectors with opposite orientation

$$\left\{ \frac{\mathbf{x}}{||\mathbf{x}||} \wedge \frac{\mathbf{y}}{||\mathbf{y}||}, - \frac{\mathbf{x}}{||\mathbf{x}||} \wedge \frac{\mathbf{y}}{||\mathbf{y}||} \right\} = \left\{ \frac{\mathbf{x}}{||\mathbf{x}||} \wedge \frac{\mathbf{y}}{||\mathbf{y}||}, \frac{\mathbf{y}}{||\mathbf{y}||} \wedge \frac{\mathbf{x}}{||\mathbf{x}||} \right\}. \tag{20}$$

For compositional data, the traditional view of CoDa as points on the probability simplex suggests to take as norm the ℓ_1 norm, $N(.) = ||.||_1$. The non-negativity constraint $\mathbf{x} \geq \mathbf{0}$ imposed on CoDa points entails that such ℓ_1 normalization corresponds to taking as representative of the (positively oriented) direction $[\mathbf{x}]_+$ the radial projection $\mathbf{x}/||\mathbf{x}||_1$ on the unit-sum affine hyperplane $\sum_i x_i = 1$. Such an operation is called "closure" in the CoDa literature, and is denoted by

$$\mathcal{C}(\mathbf{x}) := \frac{\mathbf{x}}{||\mathbf{x}||_1}.$$

This yields, as normalized representative of the pair of CoDa $[\mathbf{x}]_+ \wedge [\mathbf{y}]_+$, the sole bi-vector

$$\mathcal{C}(\mathbf{x}) \wedge \mathcal{C}(\mathbf{y}) = \frac{\mathbf{x} \wedge \mathbf{y}}{||\mathbf{x}||_1 ||\mathbf{y}||_1}, \tag{21}$$

However, the choice of the norm in the normalization is somehow conventional. Thus, an interesting alternative is to normalize by the usual ℓ_2 norm, i.e. to consider the bi-vector

$$\mathcal{S}(\mathbf{x}) \wedge \mathcal{S}(\mathbf{y}) = \frac{\mathbf{x} \wedge \mathbf{y}}{||\mathbf{x}|| ||\mathbf{y}||}, \tag{22}$$

where $\mathcal{S}$ corresponds to (19). This choice corresponds to a spherical representation of CoDA, see Faugeras (2023). Both choices will lead to interesting distance/divergences and measures of variation, as we will see.

3.3 Pseudo scalar product and norm for pairs of directions/CoDa

Having associated normalized representatives (20), resp., (21) or (22), to pairs of directions, resp., pairs of CoDa, one can then apply the Plücker scalar product (10) and norm (13) defined for bi-vectors to these normalized simple bi-vectors. This allows to partially transfer[16] the concepts of scalar product and norm to pairs of directions/CoDa. These will prove useful to define quantities of statistical interest, like divergences, means and variance/covariance measures for directional data/CoDa.

For directional data, in view of the sign indeterminacy of the normalized representatives of $[\mathbf{x}] \wedge [\mathbf{y}]$ in (20), only the absolute value of the Plücker scalar product (10) makes sense between them. For compositional data, there is no sign ambiguity, but choosing between normalizations (21) by the ℓ_1 norm, or (22) by the ℓ_2 norm, leads to two different pseudo-scalar products. We thus define:

Definition 3.1. *i) For directional data, the pseudo-scalar product* $\mathfrak{S}$ *between pairs of directions* $[\mathbf{x}] \wedge [\mathbf{y}]$ *and* $[\mathbf{r}] \wedge [\mathbf{s}]$*, i.e. between the planes* $span(\mathbf{x}, \mathbf{y})$ *and* $span(\mathbf{r}, \mathbf{s})$*, is defined as*

$$\mathfrak{S}(span(\mathbf{x}, \mathbf{y})), span(\mathbf{r}, \mathbf{s})) := \frac{|\langle \mathbf{x} \wedge \mathbf{y} | \mathbf{r} \wedge \mathbf{s} \rangle|}{||\mathbf{x}|| \, ||\mathbf{y}|| \, ||\mathbf{r}|| \, ||\mathbf{s}||} \tag{23}$$

ii) For CoDa, the pseudo scalar product $\langle .|. \rangle_1$ *between pairs* $[\mathbf{x}]_+ \wedge [\mathbf{y}]_+$ *and* $[\mathbf{r}]_+ \wedge [\mathbf{s}]_+$ *of CoDa, based on normalization (21) by the* ℓ_1 *norm, is defined as*

$$\begin{aligned} \langle [\mathbf{x}]_+ \wedge [\mathbf{y}]_+ | [\mathbf{r}]_+ \wedge [\mathbf{s}]_+ \rangle_1 &:= \langle \mathcal{C}(\mathbf{x}) \wedge \mathcal{C}(\mathbf{y}) | \mathcal{C}(\mathbf{r}) \wedge \mathcal{C}(\mathbf{s}) \rangle, \\ &= \frac{\langle \mathbf{x} \wedge \mathbf{y} | \mathbf{r} \wedge \mathbf{s} \rangle}{||\mathbf{x}||_1 \, ||\mathbf{y}||_1 \, ||\mathbf{r}||_1 \, ||\mathbf{s}||_1} \end{aligned} \tag{24}$$

and the pseudo scalar product $\langle .|. \rangle_2$*, based on normalization (22) by the* ℓ_2 *norm, is defined as*

$$\langle [\mathbf{x}]_+ \wedge [\mathbf{y}]_+ | [\mathbf{r}]_+ \wedge [\mathbf{s}]_+ \rangle_2 := \frac{\langle \mathbf{x} \wedge \mathbf{y} | \mathbf{r} \wedge \mathbf{s} \rangle}{||\mathbf{x}|| \, ||\mathbf{y}|| \, ||\mathbf{r}|| \, ||\mathbf{s}||} \tag{25}$$

[16] The spaces of pairs of CoDa or directions do not have a global vector space structure. (The Grassmannian $G(2, d+1)$ is a manifold).

Correspondingly, a pseudo-norm of a pair of equivalence classes can be defined classically as the pseudo- scalar product of such a pair of equivalence classes with itself. This gives, for directional data, in view of (23),

$$||[\mathbf{x}] \wedge [\mathbf{y}]|| := \sqrt{\mathfrak{S}(\mathrm{span}(\mathbf{x},\mathbf{y})), \mathrm{span}(\mathbf{x},\mathbf{y}))} = \frac{||\mathbf{x} \wedge \mathbf{y}||}{||\mathbf{x}||\,||\mathbf{y}||} \tag{26}$$

where $||.||$ in the numerator is the Plücker norm (13). For CoDa, applying (24), resp. (25), yields two pseudo-norms N_1, resp. N_2, defined as follows:

Definition 3.2. *For two CoDa elements* $[\mathbf{x}]_+, [\mathbf{y}]_+ \in \mathbb{P}^d_+$,

i) the pseudo-norm N_1 *of* $[\mathbf{x}]_+ \wedge [\mathbf{y}]_+$, *obtained from the* ℓ_1 *normalization of the pseudo-scalar product (24), is defined as*

$$\begin{aligned} N_1([\mathbf{x}]_+ \wedge [\mathbf{y}]_+) &:= \sqrt{\langle [\mathbf{x}]_+ \wedge [\mathbf{y}]_+ | [\mathbf{x}]_+ \wedge [\mathbf{y}]_+ \rangle_1} \\ &= ||\mathcal{C}(\mathbf{x}) \wedge \mathcal{C}(\mathbf{y})|| = \frac{||\mathbf{x} \wedge \mathbf{y}||}{||\mathbf{x}||_1\,||\mathbf{y}||_1}, \end{aligned} \tag{27}$$

ii) the pseudo-norm N_2 *of* $[\mathbf{x}]_+ \wedge [\mathbf{y}]_+$, *obtained from the* ℓ_2 *normalization of the pseudo-scalar product (25), is defined as*

$$N_2([\mathbf{x}]_+ \wedge [\mathbf{y}]_+) := \sqrt{\langle [\mathbf{x}]_+ \wedge [\mathbf{y}]_+ | [\mathbf{x}]_+ \wedge [\mathbf{y}]_+ \rangle_2} = \frac{||\mathbf{x} \wedge \mathbf{y}||}{||\mathbf{x}||\,||\mathbf{y}||} \tag{28}$$

Note that (28) for CoDa is the same as (26) for directional data[17].

3.4 Interpretation of the norm of pairs of equivalence classes as a divergence/distance between directions/CoDa

The pseudo norms $||.||$, resp. N_1, N_2 of the higher dimensional object $[\mathbf{x}] \wedge [\mathbf{y}]$, resp. $[\mathbf{x}]_+ \wedge [\mathbf{y}]_+$, is a scalar quantity which interprets geometrically as a proximity measure or divergence between the lower dimensional objects, i.e. directions $[\mathbf{x}], [\mathbf{y}]$, resp., CoDa $[\mathbf{x}]_+, [\mathbf{y}]_+$.

This is clear for directional data: in view of (17), Equation (26) simply writes

$$||[\mathbf{x}] \wedge [\mathbf{y}]|| = \sin\theta(\mathbf{x},\mathbf{y}) \in [0,1].$$

Thus, pseudo-norm depends explicitly on the angle between $\mathbf{x}$ and $\mathbf{y}$, via the sin function and thus quantifies the spread of the angular separation in the sine distance (see Theorem 3.4 below for detailed properties). Note that on can recover the (acute) angular/spherical distance between lines $[\mathbf{x}]$, $[\mathbf{y}]$,

$$d_{\mathrm{angular}}([\mathbf{x}],[\mathbf{y}]) := \theta([\mathbf{x}],[\mathbf{y}]) := \arcsin(||[\mathbf{x}] \wedge [\mathbf{y}]||) \in [0, \pi/2]$$

[17]This explains why we treat simultaneously directional and compositional data, although we are primarily interested in CoDa.

from such a exterior product approach, thus in a manner dual to the classical scalar product definition of the angle.

For CoDa, it is remarkable that the N_1 pseudo-norm (27) of simple bi-vectors gives the same $2-$barycentric divergence d_2 of Faugeras (2024a) (see Definition 4.1 and Lemma 4.3 in Faugeras (2024a)),

$$\begin{aligned} N_1([\mathbf{x}]_+ \wedge [\mathbf{y}]_+) = ||\mathcal{C}(\mathbf{x}) \wedge \mathcal{C}(\mathbf{y})|| &= \frac{\sqrt{||\mathbf{x}||^2||\mathbf{y}||^2 - \langle \mathbf{x}|\mathbf{y}\rangle^2}}{||\mathbf{x}||_1||\mathbf{y}||_1} \\ &= \frac{||\mathbf{x}||\,||\mathbf{y}||}{||\mathbf{x}||_1||\mathbf{y}||_1} \sin\theta(\mathbf{x},\mathbf{y}) = d_2([\mathbf{x}]_+, [\mathbf{y}]_+). \end{aligned} \tag{29}$$

The family of $\alpha-$barycentric divergence was derived from an affine geometry perspective, using completely different arguments. Thus, N_1 satisfy all properties of Theorem 4.2 in Faugeras (2024a): it is a well-defined, bounded, symmetric, permutation invariant, divergence on the full CoDa space $\mathbb{P}^d_+$, able to deal with zeroes. In addition, one has the following property on the attainment of the upper bound:

Lemma 3.3. *$N_1([\mathbf{x}]_+ \wedge [\mathbf{y}]_+) = 1$ iff $\mathcal{C}(\mathbf{x}) = \mathbf{e}_i$, $\mathcal{C}(\mathbf{y}) = \mathbf{e}_j$ for some $i \neq j$.*

Proof. From (29), $N_1([\mathbf{x}]_+ \wedge [\mathbf{y}]_+)$ writes as a product of three numbers $\frac{||\mathbf{x}||}{||\mathbf{x}||_1}$, $\frac{||\mathbf{y}||}{||\mathbf{y}||_1}$, $\sin\theta(\mathbf{x},\mathbf{y})$, bounded between zero and one. Thus, N_1 attains the upper bound 1 iff

$$\begin{cases} ||\mathbf{x}|| = ||\mathbf{x}||_1, \\ ||\mathbf{y}|| = ||\mathbf{y}||_1, \\ \theta(\mathbf{x},\mathbf{y}) = \pi/2. \end{cases} \tag{30}$$

Note that

$$\begin{aligned} ||\mathbf{x}|| = ||\mathbf{x}||_1 \Leftrightarrow \sum_i x_i^2 = \left(\sum_i x_i\right)^2 &= \sum_i x_i^2 + 2\sum_{i<j} x_i x_j \\ \Leftrightarrow \sum_{i<j} x_i x_j &= 0. \end{aligned}$$

Combined with $\mathbf{x} \geq 0$ and $\mathbf{x} \neq \mathbf{0}$, this is equivalent to $\mathbf{x} = \lambda \mathbf{e}_i$ for some $\lambda \neq 0$, and some $1 \leq i \leq n$, where $\mathbf{e}_i$ is the canonical basis vector $\mathbf{e}_i = (0,\dots,0,1,0\dots,0)$ of $\mathbb{R}^{d+1}$. As a result, (30) is equivalent to $\mathcal{C}(\mathbf{x}) = \mathbf{e}_i$, $\mathcal{C}(\mathbf{y}) = \mathbf{e}_j$ for some $i \neq j$. Geometrically, $\mathcal{C}(\mathbf{x})$ and $\mathcal{C}(\mathbf{y})$ are two different points on the intersection of the unit sphere and the unit-sum hyperplane. □

The interpretation of Lemma 3.3 is as follows: the N_1 norm, equivalently the 2-barycentric divergence, is maximal for CoDa which reduces to two different single-component CoDa, i.e. for CoDa points that are maximally spread apart and maximally sparse.

Remark 5. *The N_1 norm involves the ratio of norms ℓ_2/ℓ_1. Its inverse ℓ_1/ℓ_2 is widely used in machine learning as a nonconvex but scale-invariant surrogate of the ℓ_0 penalty for encouraging sparsity, see e.g. Yin, Esser, and Xin (2014) or Xu et al. (2021). Since $\max_{\mathbf{x}\neq\mathbf{0}} \frac{||\mathbf{x}||}{||\mathbf{x}||_1} = (\min_{\mathbf{x}\neq\mathbf{0}} \frac{||\mathbf{x}||_1}{||\mathbf{x}||})^{-1}$, this explains why maximizing N_1 in Lemma 3.3 makes appears the sparsity property induced by the minimization of the ratio ℓ_1/ℓ_2.*

This sparsity property of N_1 is in contrast with the N_2 pseudo-norm. For example, $N_1([1:0:0]_+\wedge[0:1:0]_+) = 1$, and $N_1([1:0:0]_+\wedge[0:1/2:1/2]_+) = 1/\sqrt{2}$, while $N_2([1:0:0]_+\wedge[0:1:0]_+) = N_2([1:0:0]_+\wedge[0:1/2:1/2]_+) = 1$: it suffices for two CoDa to be orthogonal to be maximally separated in the N_2 pseudo-norm.

Remark 3 in Faugeras (2024a) hinted at the possibility of removing the ratio of ℓ_2/ℓ_1 norms by normalizing CoDa with the ℓ_2 norm instead. This is precisely achieved by the N_2 pseudo-norm (28) of simple bi-vectors. By (17), the pseudo norm N_2 of formula (28) writes

$$N_2([\mathbf{x}]_+ \wedge [\mathbf{y}]_+) = \sin\theta(\mathbf{x}, \mathbf{y}), \tag{31}$$

where the non-negativity constraint of CoDa now entails that $\theta(\mathbf{x}, \mathbf{y}) \in [0, \pi/2]$: $\theta(\mathbf{x}, \mathbf{y})$ is equal to the acute angle $\theta([\mathbf{x}]_+, [\mathbf{y}]_+) \in [0, \pi/2]$ between the rays $[\mathbf{x}]_+, [\mathbf{y}]_+$ in the non-negative orthant. Compared to the $2-$barycentric divergence/ pseudo-norm N_1, N_2 enjoys improved properties: N_2 is a genuine distance, i.e. it satisfies the triangle inequality, is defined on the full CoDa space (thus allowing for zeroes), and is also orthogonally invariant, as shown in the next theorem.

Theorem 3.4. *N_2 is a bounded, orthogonally invariant distance on the full CoDa space $\mathbb{P}^d_+$: it satisfies the following properties:*

i) scale-invariance: $N_2([\mathbf{x}]_+ \wedge [\mathbf{y}]_+) = N_2([\lambda\mathbf{x}]_+ \wedge [\mu\mathbf{y}]_+)$, $\lambda, \mu > 0$.

ii) symmetry: $N_2([\mathbf{x}]_+ \wedge [\mathbf{y}]_+) = N_2([\mathbf{y}]_+ \wedge [\mathbf{x}]_+)$.

iii) boundedness: $0 \leq N_2([\mathbf{x}]_+ \wedge [\mathbf{y}]_+) \leq 1$.

iv) Positive-definiteness: $N_2([\mathbf{x}]_+ \wedge [\mathbf{y}]_+) = 0 \Leftrightarrow [\mathbf{x}]_+ = [\mathbf{y}]_+$.

v) Triangle inequality: for all $[\mathbf{x}]_+, [\mathbf{y}]_+[\mathbf{z}]_+ \in \mathbb{P}^d_+$,

$$N_2([\mathbf{x}]_+ \wedge [\mathbf{z}]_+) \leq N_2([\mathbf{x}]_+ \wedge [\mathbf{y}]_+) + N_2([\mathbf{y}]_+ \wedge [\mathbf{z}]_+).$$

vi) Orthogonal invariance. Let U be an orthogonal matrix. Then,

$$N_2([U\mathbf{x}]_+ \wedge [U\mathbf{y}]_+) = N_2([\mathbf{x}]_+ \wedge [\mathbf{y}]_+).$$

vii) In particular, vi) entails permutation invariance: let $\boldsymbol{\sigma}$ be a permutation of $\{0, 1, \ldots, d\}$ and $x_{\boldsymbol{\sigma}}$ be the vector obtained by permuting the coordinates of $\mathbf{x}$ by $\boldsymbol{\sigma}$. Then, $N_2([\mathbf{x}_{\boldsymbol{\sigma}}]_+ \wedge [\mathbf{y}_{\boldsymbol{\sigma}}]_+) = N_2([\mathbf{x}]_+ \wedge [\mathbf{y}]_+)$.

Proof. The proofs of i)-iv) are direct consequences of the definition (28) and formulas (14) and (17). See also Faugeras (2024a) Theorem 4.2. vi) follows from the orthogonal invariance of the Euclidean and Plucker norms, see (18). Only the triangle inequality v) deserves a more detailed proof. We provide two proofs.

1. first proof of v): w.l.o.g. consider normalized representatives $\mathbf{x}, \mathbf{y}, \mathbf{z}$ s.t. $||\mathbf{x}|| = ||\mathbf{y}|| = ||\mathbf{z}|| = 1$. Let $\text{proj}_{\mathbf{x}} := \mathbf{x}\mathbf{x}^T$ be the orthogonal projection matrix on $\mathbf{x}$. Then, in view of (14), one has, by expanding the square,

$$\frac{1}{2}||\text{proj}_{\mathbf{x}} - \text{proj}_{\mathbf{y}}||_F^2 = \frac{1}{2}\sum_{i,j}(x_i x_j - y_i y_j)^2 = \frac{1}{2}\sum_{i,j} x_i^2 x_j^2 + y_i^2 y_j^2 - 2x_i y_i x_j y_j$$
$$= 1 - \langle \mathbf{x}|\mathbf{y}\rangle^2 = N_2^2([\mathbf{x}]_+ \wedge [\mathbf{y}]_+)$$

Thus, $N_2([\mathbf{x}]_+ \wedge [\mathbf{y}]_+) = \frac{1}{\sqrt{2}}||\text{proj}_{\mathbf{x}} - \text{proj}_{\mathbf{y}}||_F$, and the triangle inequality for N_2 follows from the triangle inequality for the Froebenius distance, viz.

$$||\text{proj}_{\mathbf{x}} - \text{proj}_{\mathbf{z}}||_F \leq ||\text{proj}_{\mathbf{x}} - \text{proj}_{\mathbf{y}}||_F + ||\text{proj}_{\mathbf{y}} - \text{proj}_{\mathbf{z}}||_F.$$

2. second proof of v): By (31), the triangle inequality follows from the triangle inequality for the angular (i.e. spherical) distance. Indeed, one has:

   (a) Case 1: if $0 \leq \theta(\mathbf{x}, \mathbf{y}) + \theta(\mathbf{y}, \mathbf{z}) \leq \pi/2$, then, by the triangle inequality for the spherical/angular distance (e.g. corollary 18.6.10 in Berger (1987)) in the spherical triangle,

   $$\theta(\mathbf{x}, \mathbf{z}) \leq \theta(\mathbf{x}, \mathbf{y}) + \theta(\mathbf{y}, \mathbf{z}).$$

   Therefore, since sin is non-decreasing on $[0, \pi/2]$,

   $$\begin{aligned}\sin\theta(\mathbf{x}, \mathbf{z}) &\leq \sin(\theta(\mathbf{x}, \mathbf{y}) + \theta(\mathbf{y}, \mathbf{z})) \\ &= \sin\theta(\mathbf{x}, \mathbf{y})\cos\theta(\mathbf{y}, \mathbf{z}) + \cos\theta(\mathbf{x}, \mathbf{y})\sin\theta(\mathbf{y}, \mathbf{z}) \\ &\leq \sin\theta(\mathbf{x}, \mathbf{y}) + \sin\theta(\mathbf{y}, \mathbf{z}).\end{aligned}$$

   (b) Case 2: if $\pi \geq \theta(\mathbf{x}, \mathbf{y}) + \theta(\mathbf{y}, \mathbf{z}) > \pi/2$. It is easy to see (e.g. via KKT conditions) that the function

   $$\begin{aligned}&[0, \pi/2] \times [0, \pi/2] \to \mathbb{R} \\ &\qquad (x_1, x_2) \mapsto \sin x_1 + \sin x_2 - \frac{2}{\pi}(x_1 + x_2)\end{aligned}$$

   attains its minimum value 0 on $[0, \pi/2] \times [0, \pi/2]$ at the four corners $(0, 0)$, $(0, \pi/2)$, $(\pi/2, 0)$, $(\pi/2, \pi/2)$. Therefore, $\theta(\mathbf{x}, \mathbf{y}) + \theta(\mathbf{y}, \mathbf{z}) > \pi/2$ entails

   $$\sin\theta(\mathbf{x}, \mathbf{y}) + \sin\theta(\mathbf{y}, \mathbf{z}) \geq \frac{2}{\pi}(\theta(\mathbf{x}, \mathbf{y}) + \theta(\mathbf{y}, \mathbf{z})) > 1 \geq \sin\theta(\mathbf{x}, \mathbf{z}).$$

□

Remark 6 (On the tightness of the triangle inequality)**.** *Let* $\mathbf{z} = \lambda \frac{\mathbf{x}}{||\mathbf{x}||} + (1 - \lambda)\frac{\mathbf{y}}{||\mathbf{y}||}$*, with* $0 \leq \lambda \leq 1$*. Then, by anti-symmetry of* $\wedge$*,* $\mathbf{x} \wedge \mathbf{x} = \mathbf{y} \wedge \mathbf{y} = \mathbf{0}$*. Therefore,*

$$\begin{aligned}\Big\|\mathbf{z} \wedge \frac{\mathbf{y}}{||\mathbf{y}||}\Big\| &= \lambda \Big\|\frac{\mathbf{x}}{||\mathbf{x}||} \wedge \frac{\mathbf{y}}{||\mathbf{y}||}\Big\|,\\ \Big\|\frac{\mathbf{x}}{||\mathbf{x}||} \wedge \mathbf{z}\Big\| &= (1-\lambda)\Big\|\frac{\mathbf{x}}{||\mathbf{x}||} \wedge \frac{\mathbf{y}}{||\mathbf{y}||}\Big\|.\end{aligned}$$

Thus,

$$\begin{aligned}\Big\|\frac{\mathbf{x}}{||\mathbf{x}||} \wedge \frac{\mathbf{y}}{||\mathbf{y}||}\Big\| &= \Big\|\mathbf{z} \wedge \frac{\mathbf{y}}{||\mathbf{y}||}\Big\| + \Big\|\frac{\mathbf{x}}{||\mathbf{x}||} \wedge \mathbf{z}\Big\|\\ &= ||\mathbf{z}|| \left(\Big\|\frac{\mathbf{x}}{||\mathbf{x}||} \wedge \frac{\mathbf{z}}{||\mathbf{z}||}\Big\| + \Big\|\frac{\mathbf{z}}{||\mathbf{z}||} \wedge \frac{\mathbf{y}}{||\mathbf{y}||}\Big\|\right),\end{aligned}$$

which is

$$N_2([\mathbf{x}]_+ \wedge [\mathbf{y}]_+) = ||\mathbf{z}|| \left(N_2([\mathbf{x}]_+ \wedge [\mathbf{z}]_+) + N_2([\mathbf{z}]_+ \wedge [\mathbf{y}]_+)\right)$$

Since $\mathbf{z}$ *lies on the chord of the sphere,* $||\mathbf{z}|| < 1$*, unless* $[\mathbf{x}]_+ = [\mathbf{y}]_+$*, so that the triangle inequality does not reduce to an equality when* $\mathbf{x}, \mathbf{y}, \mathbf{z}$ *are on the same plane, i.e. when* $[\mathbf{x}]_+, [\mathbf{y}]_+, [\mathbf{z}]_+$ *are aligned on the same projective line. In other words,* N_2 *is not a projective metric, contrary to Hilbert's projective metric (see Faugeras (2023)) or the spherical distance.*

3.5 A variant via square root transform

3.5.1 Spherical representation of CoDa via square root transform

In Faugeras (2023) Appendix A, we explained that CoDa admits (at least) two normalized representatives on the (non-negative pat of the) Euclidean (ℓ_2) unit sphere $S^d_+ := S^d \cap \mathbb{R}^{d+1}_+$: given some vector $\mathbf{x} \in \mathbb{R}^{d+1}_+$ of raw amounts, its CoDa part $[\mathbf{x}]_+$ can be represented on S^d_+, either as

$$\mathcal{S}([\mathbf{x}]_+) = \mathcal{S}(\mathbf{x}) := \frac{\mathbf{x}}{||\mathbf{x}||}, \quad \text{or as} \quad \mathcal{R}([\mathbf{x}]_+) = \mathcal{R}(\mathbf{x}) := \sqrt{\mathcal{C}(\mathbf{x})} = \sqrt{\frac{\mathbf{x}}{||\mathbf{x}||_1}}, \quad (32)$$

see Figure 1. The commutative diagram expresses that $\mathcal{C} \circ \mathcal{S} = \mathcal{C}$ and $\mathcal{S} \circ \mathcal{C} = \mathcal{S}$.

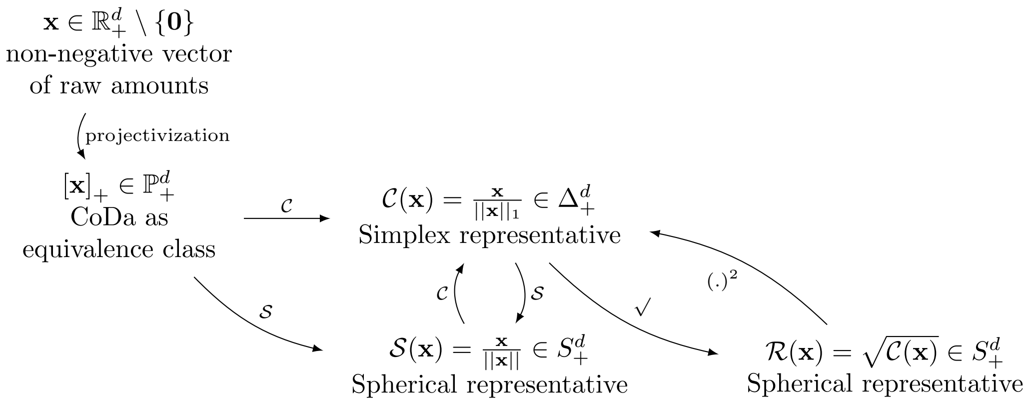

Figure 1: Normalized spherical representations $\mathcal{S}(\mathbf{x})$ or $\mathcal{R}(\mathbf{x})$ on the non-negative unit Euclidean sphere S^d_+ of a CoDa $[\mathbf{x}]_+$.

In the classical CoDa literature, where CoDa elements (equivalence classes $[\mathbf{x}]_+ \in \mathbb{P}^d_+$) are identified with their simplex representative $\mathbf{x} \in \Delta^d_+$, the transformation $\mathcal{R}$ simply corresponds to the square-root transform of e.g. Watson and Philip (1989). Generalizations of such transforms to a power (i.e. a Box-Cox) give the α-transform of Tsagris, Preston, and Wood (2011), and their variants.

The following (toy) example illustrates the differences between the $\mathcal{R}$ and $\mathcal{S}$ spherical representatives.

Example 1. *For $d = 1$, let $\mathbf{x} = (0.5, 1.5)$ be a bivariate vector of raw amounts. Its simplex representative is $\mathcal{C}(\mathbf{x}) = (1/4, 3/4)$, and the spherical ones $\mathcal{S}(\mathbf{x}) = \big(1/\sqrt{10}, 3/\sqrt{10}\big)$, and $\mathcal{R}(\mathbf{x}) = \big(1/2, \sqrt{3}/2\big)$, as illustrated in Figure 2. The picture clearly shows the effect of such power transformation: whereas the simplex representative $\mathcal{C}(\mathbf{x})$ and spherical one $\mathcal{S}(\mathbf{x})$ remain on the same ray $[\mathbf{x}]_+$, the spherical $\mathcal{R}(\mathbf{x})$ is moved toward the central direction $\pi/4$. $\mathcal{R}$ modifies the directional/compositional content of $\mathbf{x}$ as the square root largely increases the component values close to zero (here, for the first component). The angular distance from the x_0 axis thus decreases from $\arccos(1/\sqrt{10}) \approx 71.6$ degrees to $\arccos(1/2) = \pi/3 = 60$ degrees. To the contrary, for the vector $\mathbf{y} = (1, 1)$, corresponding to $\mathcal{C}(\mathbf{y}) = (.5, .5) \in \Delta^d_+$, both spherical representations coincide, i.e. $\mathcal{S}(y) = \mathcal{R}(y)$.*

Remark 7. *The fact that $\mathcal{R}$ modifies the compositional part of the data can be regarded either as an advantage, or as a drawback. On the positive side, by moving the CoDa components away from small values, it allows to reduce the influence of outliers (in particular for CoDa close to zeroes), or make the transformed data more Normally distributed. Gaussian models are then easier to fit or Gaussian assumptions are more likely to be met in statistical tests, in a way similar to the classical Box Cox transforms for Euclidean data. In addition, the $\mathcal{R}$ transform (the square root) acts component-wise from the simplex*

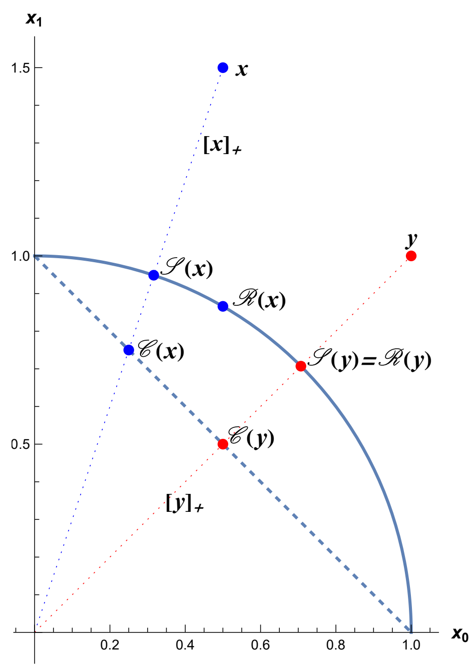

Figure 2: Comparison of the $\mathcal{R}$ and $\mathcal{S}$ spherical representations of CoDa, for $d = 1$. The $\mathcal{S}$ transform does not change the directional/CoDa part $[\mathbf{x}]_+$ (blue dotted ray) of $\mathbf{x}$, as it remains on the ray $[\mathbf{x}]_+$ where the simplex (blue dashed line) representative $\mathcal{C}(x)$ (blue point) also sits, whereas $\mathcal{R}(\mathbf{x})$ (blue point) is moved on the unit circle (blue solid line) towards the first diagonal. For $\mathbf{y}$ on the first diagonal, $\mathcal{S}(x)$ and $\mathcal{R}(x)$ (red point) coincide, and also match the directional/compositional content $[\mathbf{y}]_+$ (red dotted line).

representative and thus does not bring the influence of other components in the transformed data. On the negative side, by possibly changing the directions, its distort the geometrical configuration of points. Thus, measures of distances and variations in the transformed data do not reflect those of the original data.

3.5.2 Sine square root distance

As explained above, the $\mathcal{R}$ transform thus gives another way to obtain a normalized representative of a CoDa $[\mathbf{x}]_+$ on the non-negative Euclidean unit sphere. In turn, this $\mathcal{R}$ transform induces $\mathcal{R}(x) \wedge \mathcal{R}(\mathbf{y})$ as normalized simple bi-vector representative for the pair of equivalence classes $[\mathbf{x}]_+ \wedge [\mathbf{y}]_+$. One can then apply the bi-vector norm to such representative to obtain the following variant of the N_2 pseudo-norm:

Definition 3.5. *The sine square-root distance $N_{1/2}$ between two CoDa $[\mathbf{x}]_+$ and $[\mathbf{y}]_+$ is defined as the norm of the bi-vector $\mathcal{R}(\mathbf{x}) \wedge \mathcal{R}(\mathbf{y})$, i.e.*

$$N_{1/2}([\mathbf{x}]_+ \wedge [\mathbf{y}]_+) := ||\mathcal{R}(\mathbf{x}) \wedge \mathcal{R}(\mathbf{y})||$$

where $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^{d+1}$ are any vector representative of $[\mathbf{x}]_+$ and $[\mathbf{y}]_+$. In particular, for simplex representatives $\mathbf{x}, \mathbf{y} \in \Delta_+^d$, it writes, in coordinates, as

$$N_{1/2}([\mathbf{x}]_+ \wedge [\mathbf{y}]_+) = \sqrt{\sum_{0 \le i < j \le d} (\sqrt{x}_i \sqrt{y}_j - \sqrt{x}_j \sqrt{y}_i)^2}.$$

This approach differs from the ones obtained in Section 3.2, where we normalized $\mathbf{x}$ by its ℓ_1 or ℓ_2 norm. Properties of such a distance are similar to those of Theorem 3.4, and are thus omitted.

4 Covariance, variance and correlation for CoDa

4.1 Basic principle

Having now at our disposal measures of distance/divergence and pseudo-scalar product on the full CoDa space, we can now proceed to define measures of statistical dispersion and covariation, in a manner similar to Faugeras (2024a), but now from the projective viewpoint with the exterior product.

Indeed, let $([\mathbf{x}]_+, [\mathbf{y}]_+) \in \mathbb{P}_+^d \times \mathbb{P}_+^d$ be a pair of random CoDa (projective) points. Given some deterministic mean points $[\boldsymbol{\mu}^{\mathbf{x}}]_+$, resp. $[\boldsymbol{\mu}^{\mathbf{y}}]_+$, of $[\mathbf{x}]_+$, resp. $[\mathbf{y}]_+$, what is meaningful in the projective viewpoint is to consider the pair of homogeneous bi-vectors $[\mathbf{x}]_+ \wedge [\boldsymbol{\mu}^{\mathbf{x}}]_+$ and $[\mathbf{y}]_+ \wedge [\boldsymbol{\mu}^{\mathbf{y}}]_+$. Following Section 3, $[\mathbf{x}]_+ \wedge [\boldsymbol{\mu}^{\mathbf{x}}]_+$, resp. $[\mathbf{y}]_+ \wedge [\boldsymbol{\mu}^{\mathbf{y}}]_+$, represent a (convex) cone included in the vector plane $\mathrm{span}(\mathbf{x}, \boldsymbol{\mu}^{\mathbf{x}})$, resp. $\mathrm{span}(\mathbf{y}, \boldsymbol{\mu}^{\mathbf{y}})$, in the ambient space $\mathbb{R}^{d+1}$. The (average) relative orientation between these planes can serve as a basis upon which one can define a notion of covariance and correlation between random CoDa points. This is accomplished by taking the expectation of the pseudo-scalar products $\langle .|. \rangle_1$ and $\langle .|. \rangle_2$ of (24) and (25), that is to say by taking the expected scalar product (10) for ℓ_1 or ℓ_2-normalized bi-vectors (21) or (22). In view of the linear structure of bi-vectors and its corresponding scalar-product, such expected scalar product between pairs of normalized bi-vectors decomposes along its Plücker components p_{ij}, which allows to define a whole matrix of co-variations along the $d(d+1)/2$ pairs $i<j$ of Coda parts. Indeed, for the ℓ_2 normalization, one has

$$E\left(\langle [\mathbf{x}]_+ \wedge [\boldsymbol{\mu}^{\mathbf{x}}]_+ | [\mathbf{y}]_+ \wedge [\boldsymbol{\mu}^{\mathbf{y}}]_+ \rangle_2\right) = E\left(\frac{\langle \mathbf{x} \wedge \boldsymbol{\mu}^{\mathbf{x}} | \mathbf{y} \wedge \boldsymbol{\mu}^{\mathbf{y}} \rangle}{||\mathbf{x}||\,||\boldsymbol{\mu}^{\mathbf{x}}||\,||\mathbf{y}||\,||\boldsymbol{\mu}^{\mathbf{y}}||}\right)$$

$$= E\left(\sum_{i<j} \frac{p_{ij}(\mathbf{x} \wedge \boldsymbol{\mu}^{\mathbf{x}}) p_{ij}(\mathbf{y} \wedge \boldsymbol{\mu}^{\mathbf{y}})}{||\mathbf{x}||\,||\boldsymbol{\mu}^{\mathbf{x}}||\,||\mathbf{y}||\,||\boldsymbol{\mu}^{\mathbf{y}}||}\right) = \sum_{i<j} E\left(\frac{p_{ij}(\mathbf{x} \wedge \boldsymbol{\mu}^{\mathbf{x}}) p_{ij}(\mathbf{y} \wedge \boldsymbol{\mu}^{\mathbf{y}})}{||\mathbf{x}||\,||\boldsymbol{\mu}^{\mathbf{x}}||\,||\mathbf{y}||\,||\boldsymbol{\mu}^{\mathbf{y}}||}\right)$$

$$= \sum_{i<j} E\left(\frac{\det\begin{vmatrix} x_i & \mu_i^{\mathbf{x}} \\ x_j & \mu_j^{\mathbf{x}} \end{vmatrix} \det\begin{vmatrix} y_i & \mu_i^{\mathbf{y}} \\ y_j & \mu_j^{\mathbf{y}} \end{vmatrix}}{||\mathbf{x}||\,||\boldsymbol{\mu}^{\mathbf{x}}||\,||\mathbf{y}||\,||\boldsymbol{\mu}^{\mathbf{y}}||}\right).$$

We only treat the ℓ_2-normalized pseudo-scalar product $\langle .|.\rangle_2$ of (25), since, for the ℓ_1-normalized pseudo scalar product $\langle .|.\rangle_1$ of (24), the construction coincide with the $2-$barycentric divergence in Faugeras (2024a), which was obtained in that paper from the affine barycentric viewpoint. (We leave the case of the square root spherical representative $\mathcal{R}$ of Section 3.5 to the reader.)

A priori, one could use any kind of mean points $[\boldsymbol{\mu}^{\mathbf{x}}]_+$, $[\boldsymbol{\mu}^{\mathbf{y}}]_+$, like Aitchison's geometric mean, the arithmetic mean, or Fréchet means based on various distances/divergences (see Section 5 of Faugeras (2024a)). However, as in Faugeras (2024a), the most interesting properties occur with the arithmetic mean, which corresponds to the barycenter (centroid) obtained by affine addition of points, see Faugeras (2024a). Such centroid also has a natural interpretation as mixing CoDa, see Scealy and Welsh (2014), and corresponds to the $m-$ affine connection of information geometry (see Faugeras (2023)). Hence, we set thereafter

$$[\boldsymbol{\mu}^{\mathbf{x}}]_+ := [E\mathbf{x}]_+, \quad [\boldsymbol{\mu}^{\mathbf{y}}]_+ := [E\mathbf{y}]_+. \tag{33}$$

Note that for CoDa with some zeroes components, Aitchison's component-wise geometric mean is ill-suited as it suffices that a CoDa has a zero in one of its components to make the corresponding component of the geometric mean also zero.

4.2 Exterior covariance and variance matrices for CoDa

For a pair of random CoDa, we thus define the following quantities of statistical interest:

Definition 4.1 (Exterior Covariance). *i) Exterior covariance matrix for a pair of CoDa:*

Let $([\mathbf{x}]_+, [\mathbf{y}]_+) \in \mathbb{P}_+^d \times \mathbb{P}_+^d$ be a pair of random CoDa, with corresponding deterministic barycenter mean points $[\boldsymbol{\mu}^{\mathbf{x}}]_+$, $[\boldsymbol{\mu}^{\mathbf{y}}]_+ \in \mathbb{P}_+^d$, defined by (33).

The exterior covariance matrix of $([\mathbf{x}]_+, [\mathbf{y}]_+)$*, based on the* ℓ_2 *normalized pseudo-scalar product* $\langle .|. \rangle_2$ *of (25), is defined as the following symmetric matrix (with null diagonal) of size* $d+1$*,*

$$Cov_2([\mathbf{x}]_+, [\mathbf{y}]_+) = \big(Cov_2([\mathbf{x}]_+, [\mathbf{y}]_+)_{i,j}\big) \in \mathbb{R}^{(d+1)\times(d+1)},$$

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

$$\begin{aligned} Cov_2([\mathbf{x}]_+, [\mathbf{y}]_+)_{i,j} &:= E\left(\frac{p_{ij}(\mathbf{x}\wedge\boldsymbol{\mu}^{\mathbf{x}})p_{ij}(\mathbf{y}\wedge\boldsymbol{\mu}^{\mathbf{y}})}{||\mathbf{x}||\,||\boldsymbol{\mu}^{\mathbf{x}}||\,||\mathbf{y}||\,||\boldsymbol{\mu}^{\mathbf{y}}||}\right) \\ &= E\left(\frac{\det\begin{vmatrix} x_i & \mu_i^{\mathbf{x}} \\ x_j & \mu_j^{\mathbf{x}} \end{vmatrix} \det\begin{vmatrix} y_i & \mu_i^{\mathbf{y}} \\ y_j & \mu_j^{\mathbf{y}} \end{vmatrix}}{||\mathbf{x}||\,||\boldsymbol{\mu}^{\mathbf{x}}||\,||\mathbf{y}||\,||\boldsymbol{\mu}^{\mathbf{y}}||}\right). \end{aligned} \tag{34}$$

ii) *Total exterior covariance: The total exterior covariance of* $([\mathbf{x}]_+, [\mathbf{y}]_+)$ *is the scalar*

$$TCov_2([\mathbf{x}]_+, [\mathbf{y}]_+) = \sum_{i<j} Cov_2([\mathbf{x}]_+, [\mathbf{y}]_+)_{i,j} \tag{35}$$

$$= E\left(\frac{\langle \mathbf{x}\wedge\boldsymbol{\mu}^{\mathbf{x}}|\mathbf{y}\wedge\boldsymbol{\mu}^{\mathbf{y}}\rangle}{||\mathbf{x}||\,||\boldsymbol{\mu}^{\mathbf{x}}||\,||\mathbf{y}||\,||\boldsymbol{\mu}^{\mathbf{y}}||}\right). \tag{36}$$

TCov_2 gives a global (scalar) measure of the covariation of two random CoDa around their respective means, while Cov_2 decomposes the latter along its (i,j) components and organizes it into a matrix. It is worth remarking that TCov_2 can be computed at once, via (36), without the need to summing all its $d(d+1)/2$ components.

Definition 4.2 (Exterior Variance). i) *Log-free exterior variance matrix of a CoDa:*

The exterior variance matrix of $[\mathbf{x}]_+$ *is defined as*

$$Var_2([\mathbf{x}]_+) := Cov_2([\mathbf{x}]_+, [\mathbf{x}]_+) = \big(Var_2([\mathbf{x}]_+)_{ij}\big) \in \mathbb{R}^{(d+1)\times(d+1)}, \tag{37}$$

whose (i,j) *component is*

$$Var_2([\mathbf{x}]_+)_{ij} := E\left(\frac{p_{ij}^2(\mathbf{x}\wedge\boldsymbol{\mu}^{\mathbf{x}})}{||\mathbf{x}||^2\,||\boldsymbol{\mu}^{\mathbf{x}}||^2}\right) = E\left(\frac{\det^2\begin{vmatrix} x_i & \mu_i^{\mathbf{x}} \\ x_j & \mu_j^{\mathbf{x}} \end{vmatrix}}{||\mathbf{x}||^2\,||\boldsymbol{\mu}^{\mathbf{x}}||^2}\right).$$

ii) *Total exterior variance:*

The total exterior variance of $([\mathbf{x}]_+$ *is the scalar*

$$TVar_2([\mathbf{x}]_+) := \sum_{i<j} Var_2([\mathbf{x}]_+)_{i,j} \tag{38}$$

$$= EN_2^2([\mathbf{x}]_+ \wedge [\boldsymbol{\mu}^{\mathbf{x}}]_+) = E\left(\frac{||\mathbf{x}\wedge\boldsymbol{\mu}^{\mathbf{x}}||^2}{||\mathbf{x}||^2\,||\boldsymbol{\mu}^{\mathbf{x}}||^2}\right). \tag{39}$$

Compared to the barycentric covariance and variance matrices (Definitions 7.1, 7.2, 7.3 of Faugeras (2024a)), it is striking that the exterior covariance (34) and variance matrices (37) only differ by the normalization, i.e. the ℓ_2 norm instead of the ℓ_1 norm. This change allows to remove the ratio of ℓ_2/ℓ_1 norm mentioned in Remark 5 and Remark 3 of Faugeras (2024a), and yields the improved properties of N_2 as an orthogonally invariant, genuine distance between CoDa (see Theorem 3.4). Indeed, one has the analogue of Proposition 7.6 and Theorem 7.7 in Faugeras (2024a):

Proposition 4.3 (Properties of the exterior variance)**.** *i) Measure of proportionality:*

At the level of components, $Var_2([\mathbf{x}]_+)_{i,j} = 0$, for some $i < j$, iff x_i and x_j are proportional or one of them is zero a.s.

ii) Global orthogonal invariance:

$TVar_2([\mathbf{x}]_+) = TVar_2([U^T\mathbf{x}]_+)$, for any orthogonal matrix U.

At the level of components, the components of the transformed exterior variance matrix writes

$$Var_2([U^T\mathbf{x}]_+)_{i,j} = E\left(\frac{\langle \mathbf{x} \wedge \boldsymbol{\mu}^{\mathbf{x}} | \mathbf{u}_i \wedge \mathbf{u}_j\rangle^2}{||\mathbf{x}||^2||\boldsymbol{\mu}^{\mathbf{x}}||^2}\right).$$

Proof. i) As in Proposition 7.6 of Faugeras (2024a).

ii) Orthogonal invariance of TVar_2 follows from orthogonal invariance of N_2, Theorem 3.4 vi).

Let U be an orthogonal matrix with column vectors $\mathbf{u}_i$, $i = 0, \ldots, d$, and $\mathbf{x}' = U^T\mathbf{x}$ be the coordinates of $\mathbf{x}$ in the new basis $(\mathbf{u}_i)$. Then, by linearity, $E\mathbf{x}' = U^T\boldsymbol{\mu}^{\mathbf{x}}$. Hence,

$$\det\begin{vmatrix} x_i' & Ex_i' \\ x_j' & Ex_j' \end{vmatrix} = \det\begin{vmatrix} \langle \mathbf{u}_i|\mathbf{x}\rangle & \langle \mathbf{u}_i|\boldsymbol{\mu}^{\mathbf{x}}\rangle \\ \langle \mathbf{u}_j|\mathbf{x}\rangle & \langle \mathbf{u}_j|\boldsymbol{\mu}^{\mathbf{x}}\rangle \end{vmatrix} = \langle \mathbf{x} \wedge \boldsymbol{\mu}^{\mathbf{x}} | \mathbf{u}_i \wedge \mathbf{u}_j\rangle$$

by the Binet-Cauchy identity (12) (or the definition of the scalar product of bi-vectors). Thus, by invariance of the Euclidean norm by an orthogonal transformation,

$$\mathrm{Var}_2([U^T\mathbf{x}]_+)_{i,j} = E\left(\frac{\langle \mathbf{x} \wedge \boldsymbol{\mu}^{\mathbf{x}} | \mathbf{u}_i \wedge \mathbf{u}_j\rangle^2}{||\mathbf{x}||^2||\boldsymbol{\mu}^{\mathbf{x}}||^2}\right)$$

□

Proposition 4.4 (Properties of the exterior covariance)**.** *i) Boundedness:*

One has

$$Cov_2^2([\mathbf{x}]_+, [\mathbf{y}]_+)_{i,j} \leq Var_2([\mathbf{x}]_+)_{i,j}\, Var_2([\mathbf{y}]_+)_{i,j}$$

ii) Assume $\mathbf{x}, \mathbf{y}$ are spherical representatives of $[\mathbf{x}]_+, [\mathbf{y}]_+$, i.e. $||\mathbf{x}|| = ||\mathbf{y}|| = 1$. If (x_i, x_j), $i \neq j$, is independent of (y_i, y_j), then

$$Cov_2([\mathbf{x}]_+, [\mathbf{y}]_+)_{i,j} = 0.$$

Proof. As Theorem 7.7 in Faugeras (2024a). □

4.3 Exterior correlation matrix

Eventually, as in Faugeras (2024a), a part-by-part measure of correlation is obtained by combining Definitions 4.1 and 4.2.

Definition 4.5 (Exterior correlation matrix). *The exterior correlation matrix of* $([\mathbf{x}]_+, [\mathbf{y}]_+)$ *is the matrix defined as*

$$Cor_2([\mathbf{x}]_+, [\mathbf{y}]_+) := \left(Cor_2([\mathbf{x}]_+, [\mathbf{y}]_+)_{ij}\right) \in \mathbb{R}^{(d+1)\times(d+1)}$$

with (i,j) *component set, for* $i \neq j$, *as*

$$Cor_2([\mathbf{x}]_+, [\mathbf{y}]_+)_{ij} := \frac{Cov_2([\mathbf{x}]_+, [\mathbf{y}]_+)_{i,j}}{\sqrt{Var_2([\mathbf{x}]_+)_{i,j}\, Var_2([\mathbf{y}]_+)_{i,j}}} \tag{40}$$

Definition 4.5 corresponds to the barycentric correlation matrix, Definition 7.4 in Faugeras (2024a), but with the ℓ_2 normalization instead of the ℓ_1 one. In (40), both the numerator and denominator are normalized inside the expectations. As in Faugeras (2024a), one can dispense with this simultaneous normalization, by defining a modified correlation matrix $r = \left(r_{ij}\right) \in \mathbb{R}^{(d+1)^2}$ as

$$\begin{aligned} r([\mathbf{x}]_+, [\mathbf{y}]_+)_{ij} &:= \frac{E\left(p_{ij}(\mathbf{x}\wedge\boldsymbol{\mu}^{\mathbf{x}})p_{ij}(\mathbf{y}\wedge\boldsymbol{\mu}^{\mathbf{y}})\right)}{\sqrt{Ep_{ij}^2(\mathbf{x}\wedge\boldsymbol{\mu}^{\mathbf{x}})}\sqrt{Ep_{ij}^2(\mathbf{y}\wedge\boldsymbol{\mu}^{\mathbf{y}})}} \qquad (41)\\ &= \frac{E\left\langle \begin{pmatrix} x_i \\ x_j \end{pmatrix} \wedge \begin{pmatrix} \mu_i^{\mathbf{x}} \\ \mu_j^{\mathbf{x}} \end{pmatrix} \middle| \begin{pmatrix} y_i \\ y_j \end{pmatrix} \wedge \begin{pmatrix} \mu_i^{\mathbf{y}} \\ \mu_j^{\mathbf{y}} \end{pmatrix} \right\rangle}{\sqrt{E\left\| \begin{pmatrix} x_i \\ x_j \end{pmatrix} \wedge \begin{pmatrix} \mu_i^{\mathbf{x}} \\ \mu_j^{\mathbf{x}} \end{pmatrix} \right\|^2}\sqrt{E\left\| \begin{pmatrix} y_i \\ y_j \end{pmatrix} \wedge \begin{pmatrix} \mu_i^{\mathbf{y}} \\ \mu_j^{\mathbf{y}} \end{pmatrix} \right\|^2}}\\ &= \frac{E\left(\det\begin{vmatrix} x_i & \mu_i^{\mathbf{x}} \\ x_j & \mu_j^{\mathbf{x}} \end{vmatrix} \det\begin{vmatrix} y_i & \mu_i^{\mathbf{y}} \\ y_j & \mu_j^{\mathbf{y}} \end{vmatrix}\right)}{\sqrt{E\left(\det^2\begin{vmatrix} x_i & \mu_i^{\mathbf{x}} \\ x_j & \mu_j^{\mathbf{x}} \end{vmatrix}\right)}\sqrt{E\left(\det^2\begin{vmatrix} y_i & \mu_i^{\mathbf{y}} \\ y_j & \mu_j^{\mathbf{y}} \end{vmatrix}\right)}} \end{aligned}$$

which corresponds exactly to the modified barycentric correlation matrix r, Definition 7.5 of Faugeras (2024a). The exterior correlation matrices (40) and (41) have properties similar to their barycentric analogue and their statements are thus omitted, see Theorem 7.7 in Faugeras (2024a). In particular, these exterior correlation matrices are conceived to be a standardization between -1 and 1 of the exterior covariance matrix (34), in the same way that Pearson's correlation coefficient standardizes the usual covariance in the Euclidean setting.

A total (scalar) measure of correlation between two CoDa $[\mathbf{x}]_+$ and $[\mathbf{y}]_+$ can be obtained by summing the components of (40) or (41). Alternatively, one can define a total scalar measure of correlation Tcor, directly from the exterior scalar product and norm as follows:

Definition 4.6. *The total exterior correlation between* $[\mathbf{x}]_+$ *and* $[\mathbf{y}]_+$ *in* $\mathbb{P}^d_+$ *is defined as*

$$Tcor([\mathbf{x}]_+, [\mathbf{y}]_+) := \frac{E\langle \mathbf{x} \wedge \boldsymbol{\mu}^{\mathbf{x}} | \mathbf{y} \wedge \boldsymbol{\mu}^{\mathbf{y}} \rangle}{\sqrt{E\left\|\mathbf{x} \wedge \boldsymbol{\mu}^{\mathbf{x}}\right\|^2}\sqrt{E\left\|\mathbf{y} \wedge \boldsymbol{\mu}^{\mathbf{y}}\right\|^2}} \tag{42}$$

Further discussion, in particular, the relation with canonical angles, is to be found in Section 6.4.

5 Sketch of other statistical applications

Summarizing the results so far, we have introduced a distance and a pseudo-scalar product between pairs of CoDa, along with covariance and variance matrices. One can further define statistical notions like Fréchet means, Gaussian distributions, etc. together with corresponding statistical applications like clustering or regression for random CoDa. We can mimic the constructs of Faugeras (2024a) obtained with an affine barycentric approach. In order not to repeat ourselves, we refer the reader to Faugeras (2024a) for more details and briefly sketch the difference and some possible new statistical applications derived from the exterior product approach.

5.1 Fréchet means and regression based on the exterior pseudo-norm N_2

As in Faugeras (2024a) Section 5, one can define Fréchet mean/median and variance based on the minimization of the expected (squared or elevated to some power) pseudonorm N_2, i.e. metric notions of center and dispersion. The Fréchet mean corresponds to minimizing the total exterior variance TVar_2 of (39). Several useful variants (medoid, clustering,...) are discussed in Remark 4 of Faugeras (2024a). In particular, weighted versions yield nonparametric estimators of the regression function of a CoDa given covariates: given a sample $([\mathbf{x}^1]_+, \mathbf{z}^1), \ldots, ([\mathbf{x}^n]_+, \mathbf{z}^n) \in \mathbb{P}^d_+ \times \mathbb{R}^k$, where $\mathbf{z}^1, \ldots, \mathbf{z}^n$ are the observed values of some covariate $\boldsymbol{\zeta} \in \mathbb{R}^k$, a Nadaraya-Watson nonparametric estimator $[\mathbf{m}]_+$ of the regression function $E([\mathbf{x}]_+ | \boldsymbol{\zeta} = \mathbf{z})$ is obtained by minimizing the the empirical weighted version of the Tvar,

$$\min_{[\mathbf{m}]_+ \in \mathbb{P}^d_+} \sum_k w_k N_2^2([\mathbf{x}^k]_+, [\mathbf{m}]_+),$$

where (w_k) are kernel weights measuring the proximity of the covariates $\boldsymbol{\zeta}$ to a fixed covariate value $\mathbf{z}$, see e.g. Faugeras (2023) Section 7.3 for an example using Hilbert's projective metric. Assuming the data is normalized on the unit sphere, this amounts to minimizing a quadratic function with the unit sphere constraint, i.e. with a quadratic constraint, and thus is computationally relatively straightforward. As in Faugeras (2024a), we stress the interest of the proposed approach in the ability of the distance N_2 to handle CoDa with zeroes, which is useful in some applications.

5.2 Weighted distance, Mahalanobis distance and classification

As a further generalization, one can define a weighted version of the N_2 pseudo-norm (28) between CoDa: given $W = \left(w_{ij}\right) \in \mathbb{R}^{(d+1)^2}$ some symmetric matrix with positive components, the W-weighted pseudo-norm $N_{2,W}$ between $[\mathbf{r}]_+$ and $[\mathbf{s}]_+$ is defined as

$$N_{2,W}^2([\mathbf{r}]_+ \wedge [\mathbf{s}]_+) := \frac{\sum_{i<j} w_{ij}^{-1} p_{ij}^2([\mathbf{r}]_+ \wedge [\mathbf{s}]_+])}{||\mathbf{r}||^2 \, ||\mathbf{s}||^2} = \frac{\sum_{i<j} w_{ij}^{-1} \det^2 \begin{vmatrix} r_i & s_i \\ r_j & s_j \end{vmatrix}}{||\mathbf{r}||^2 \, ||\mathbf{s}||^2}, \tag{43}$$

which is the analogue of the W-weighted 2-barycentric divergence, Definition 6.2, Equation (22), of Faugeras (2024a), but with the ℓ_2 normalization. This allows to define anisotropic Gaussian distributions, similar to Definition 6.3 in Faugeras (2024a).

In particular, it is interesting to use, as weight matrix in (43), the exterior variance matrix (37), i.e. $W = \mathrm{Var}_2([\mathbf{x}]_+)$. This allows to define an analogue of the Mahalanobis distance for CoDa as

$$N_{2,\mathrm{Var}_2([\mathbf{x}]_+)}([\mathbf{r}]_+ \wedge [\mathbf{s}]_+).$$

It measures the distance between pairs of component of $[\mathbf{r}]_+$ and $[\mathbf{s}]_+$ relatively to the scale of variation of the corresponding pairs of components of $[\mathbf{x}]_+$. Such Mahalanobis distance should be useful for measuring the outlyingness of a CoDa point and identifying outliers in a sample. In addition, it could also be applied for clustering and classification of CoDa: for classification into k classes, one estimates the variance matrix W_k and mean $[\boldsymbol{\mu}_k]_+$ of each class, based on samples known to belong to each class, and then one classify a test point $[\mathbf{y}]_+$ as belonging to the class for which the CoDa Mahalanobis distance $N_{2,W_k}([\mathbf{y}]_+, [\boldsymbol{\mu}_k]_+)$ is minimal.

5.3 Gaussian distributions

As in Faugeras (2023) and Faugeras (2024a), one can define families of Gaussian-type distributions based on the N_2 distance (28). This amounts to defining the density $f_{[\mathbf{X}]_+}$ of the exterior (isotropic) Gaussian distribution with parameters $([\mathbf{m}]_+, \sigma)$ as

$$f_{[\mathbf{X}]_+}([\mathbf{x}]_+; [\mathbf{m}]_+, \sigma) := Z^{-1}([\mathbf{m}]_+, \sigma) \exp\left(-N_2^2([\mathbf{x}]_+ \wedge [\mathbf{m}]_+)/\sigma^2\right)$$

where Z is a normalizing constant. Variants including anisotropic versions can be defined similarly to Faugeras (2024a), by using the W-weighted version (43) of the N_2 pseudo-norm. Details are omitted. We simply content ourselves with giving an illustration with $[\mathbf{m}]_+ = [1 : 0.5 : 0]_+$, $\sigma = 1$ in Figure 3: one gets similar level sets as the isotropic 2-Barycentric Gaussian distribution (Definition 6.1 in Faugeras (2024a)), but with a different normalizing constant, owing to the ℓ_2 rescaling instead of the ℓ_1 one. Note that in Figure 3, we chose a mean parameter on the boundary of the simplex, illustrating the ability of these distributions based on log-free divergences/distance to represent CoDa with zeroes.

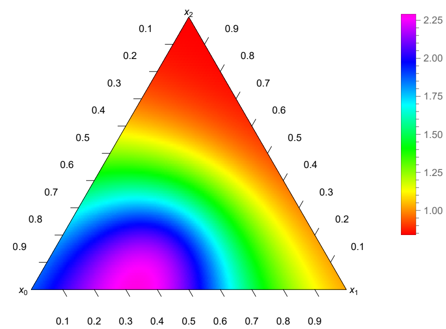

Figure 3: Isotropic Gaussian distribution based on the N_2 norm. $[\mathbf{m}]_+ = [1 : 0.5 : 0]_+$, $\sigma = 1$.

5.4 Wasserstein exterior distance between CoDa distributions via optimal transportation

The N_2 norm defines a bounded distance on the space of random CoDa variables,

$$EN_2([\mathbf{x}]_+ \wedge [\mathbf{y}]_+) = E\left(\frac{||\mathbf{x} \wedge \mathbf{y}||}{||\mathbf{x}|| \, ||\mathbf{y}||}\right),$$

which depends on the joint distribution $P^{[\mathbf{x}]_+,[\mathbf{y}]_+}$ of the pair $([\mathbf{x}]_+, [\mathbf{y}]_+)$ of CoDa. It can be turned into a Wasserstein-Kantotovich type probability metric (see e.g. Rachev (1991)), i.e. a distance $\mathcal{N}_2$ on the space of CoDa probability measures between the (marginal) distributions $P^{[\mathbf{x}]_+}$, $P^{[\mathbf{y}]_+}$, via optimal transportation,

$$\mathcal{N}_2(P^{[\mathbf{x}]_+}, P^{[\mathbf{y}]_+}) := \inf EN_2([\mathbf{x}]_+ \wedge [\mathbf{y}]_+),$$

where the infimum is over all joint distribution $P^{[\mathbf{x}]_+,[\mathbf{y}]_+}$ with given marginals $P^{[\mathbf{x}]_+}$, $P^{[\mathbf{y}]_+}$. The latter metricizes weak convergence[18], see e.g. Rachev and Rüschendorf (1998) or Villani (2009). Interpreting CoDa as a (discrete) probability distribution (see Faugeras (2024b)), this yields a genuine distance between distributions of distributions, with possibly unequal supports.

[18]Moment convergence is automatically implied since the distance is bounded.

6 Related notions and approaches

6.1 The norm of a simple bi-vector as a quantification of Cauchy-Schwarz inequality and relation with inequality divergences

Nielsen, K. Sun, and Marchand-Maillet (2017) show how one can build divergences and proximity measures by quantifying an inequality, in particular the Cauchy-Schwarz inequality (see also Budka, Gabrys, and Musial (2011)). The general principle is as follows: given an inequality

$$\mathrm{lhs}(\mathbf{x},\mathbf{y}) \leq \mathrm{rhs}(\mathbf{x},\mathbf{y}),$$

where lhs, resp. rhs, denote the left-hand side, resp. right-hand side, of the inequality, a divergence can be built by measuring the tightness in the inequality. Such measure of tightness can be performed either on an interval scale, through the difference gap

$$\mathrm{diff}(\mathbf{x},\mathbf{y}) := \mathrm{rhs}(\mathbf{x},\mathbf{y}) - \mathrm{lhs}(\mathbf{x},\mathbf{y}) \geq 0,$$

(or any monotone strictly increasing function thereof), or, when lhs > 0, on the ratio scale, via the log-ratio gap

$$\mathrm{lr}(\mathbf{x},\mathbf{y}) := -\log\left(\frac{\mathrm{lhs}(\mathbf{x},\mathbf{y})}{\mathrm{rhs}(\mathbf{x},\mathbf{y})}\right). \tag{44}$$

If $\mathrm{lhs}(\mathbf{x},\mathbf{y}) < \mathrm{rhs}(\mathbf{x},\mathbf{y})$ for $\mathbf{x} \neq \mathbf{y}$, and $\mathrm{lhs}(\mathbf{x},\mathbf{x}) = \mathrm{rhs}(\mathbf{x},\mathbf{x})$, then either diff or lr gives a divergence, i.e. a measure of proximity, between $\mathbf{x}$ and $\mathbf{y}$. In addition, if the inequality is homogeneous, i.e. invariant through rescalings $\mathbf{x} \leftarrow \lambda\mathbf{x}$, $\mathbf{y} \leftarrow \mu\mathbf{y}$, with $\lambda, \mu \neq 0$, then lhs > 0 entails, as sister inequality,

$$0 < \frac{\mathrm{lhs}(\mathbf{x},\mathbf{y})}{\mathrm{rhs}(\mathbf{x},\mathbf{y})} \leq 1,$$

which is dimensionless and scale-invariant. In turn, the latter can be quantified, either via the difference gap, which gives as variant of diff its rational version,

$$\mathrm{diff}'(\mathbf{x},\mathbf{y}) := 1 - \frac{\mathrm{lhs}(\mathbf{x},\mathbf{y})}{\mathrm{rhs}(\mathbf{x},\mathbf{y})}$$

or via the log-ratio gap, the latter giving the same divergence lr as (44). Both the rational difference gap, $\mathrm{diff}'(\mathbf{x},\mathbf{y})$, and the log-ratio gap, $\mathrm{lr}(\mathbf{x},\mathbf{y})$, give a projective divergence, i.e. a divergence which is invariant w.r.t. to rescaling of $\mathbf{x}, \mathbf{y}$. The log transformation is simply for stretching the range of value to $[0, \infty]$ (thus with a possible infinite value if lhs $= 0$.)

Applied to the the Cauchy-Schwarz inequality,

$$|\langle\mathbf{x}|\mathbf{y}\rangle| \leq ||\mathbf{x}||.||\mathbf{y}|| \Leftrightarrow ||\mathbf{x}||^2.||\mathbf{y}||^2 - \langle\mathbf{x}|\mathbf{y}\rangle^2 \geq 0,$$

this gives

$$
\begin{aligned}
\operatorname{diff}(\mathbf{x},\mathbf{y}) &= ||\mathbf{x}||^2.||\mathbf{y}||^2 - \langle\mathbf{x}|\mathbf{y}\rangle^2 \\
\operatorname{diff}'(\mathbf{x},\mathbf{y}) &= 1 - \frac{\langle\mathbf{x}|\mathbf{y}\rangle^2}{||\mathbf{x}||^2.||\mathbf{y}||^2} \\
D(\mathbf{x},\mathbf{y}) &= -\ln\left(\frac{\langle\mathbf{x}|\mathbf{y}\rangle^2}{||\mathbf{x}||^2.||\mathbf{y}||^2}\right)
\end{aligned}
$$

In view of (14), the (squared) norm of simple bi-vectors, resp. pseudo-norm N_2 of equivalence classes, writes,

$$
\begin{aligned}
||\mathbf{x}\wedge\mathbf{y}||^2 &= \operatorname{diff}(\mathbf{x},\mathbf{y}), \\
N_2^2([\mathbf{x}]_+\wedge[\mathbf{y}]+) &= \operatorname{diff}'(\mathbf{x},\mathbf{y}).
\end{aligned}
$$

Thus, one gets an algebraic interpretation of the Plücker norm $||\mathbf{x}\wedge\mathbf{y}||$ of the parallelogram $\mathbf{x}\wedge\mathbf{y}$ and of the N_2 pseudo-norm of equivalence classes $[\mathbf{x}]\wedge[\mathbf{y}]$ of planes, resp. $[\mathbf{x}]_+\wedge[\mathbf{y}]_+$ of CoDa pairs: they both represent a quantification of the Cauchy-Schwarz inequality, either on the interval scale for the square of the Plücker Norm $||\mathbf{x}\wedge\mathbf{y}||^2$, or on the ratio scale for $N_2^2([\mathbf{x}]_+\wedge[\mathbf{y}]+)$. Conversely, in view of their expression in terms of the sine distance (17), one gets geometric insight and interpretation of the divergences based on the Cauchy-Schwarz inequality of Nielsen, K. Sun, and Marchand-Maillet (2017) and Budka, Gabrys, and Musial (2011).

Remark 8. *i) The quantification of the Cauchy-Schwarz inequality in (14) is also known in the matrix algebra literature as Lagrange's identity,*

$$
||\mathbf{x}||^2||\mathbf{y}||^2 - \langle\mathbf{x}|\mathbf{y}\rangle^2 = \sum_{i<j}(x_iy_j - x_jy_i)^2.
$$

It is a special case, for $\mathbf{x}=\mathbf{r}$ and $\mathbf{y}=\mathbf{s}$, of the so-called Binet-Cauchy identity,

$$
\left(\sum_{i=1}^n x_ir_i\right)\left(\sum_{j=1}^n y_js_j\right) - \left(\sum_{i=1}^n x_is_i\right)\left(\sum_{j=1}^n y_jr_j\right) = \sum_{1\le i<j\le n}(x_iy_j - x_jy_i)(r_is_j - r_js_i),
$$

which itself is a special case of the Cauchy-Binet formula (7). The exterior algebra approach thus gives geometric insight on these algebraic identities, as the latter equation simply corresponds to the definition (12) of the scalar product of bi-vectors.

ii) Note that an elementary proof of Cauchy-Schwarz's inequality follows from Lagrange identity, or equivalently from the semi-positive-definiteness of the Gram Matrix of two vectors.

6.2 Barycentric divergence as a two-sided quantification of the likelihood ratio order

Let X, Y be continuous univariate real-valued random variables with densities f_X, f_Y. The likelihood ratio order is defined on the set of univariate absolutely continuous distributions as

$$X \leq_{\text{lr}} Y \Longleftrightarrow \frac{f_Y}{f_X} \text{ is non-decreasing.} \tag{45}$$

Such a concept is useful for constructing Uniformly Most Powerful tests or Median-Unbiased estimates on families of distributions having a monotone likelihood ratio, see Pfanzagl (1979), Brown, Cohen, and Strawderman (1976). To dispense with the cases where the ratios of densities are undefined, condition (45) can be rewritten more generally as: for $s, t \in \mathbb{R}$, $s \leq t$ implies

$$f_X(t)f_Y(s) \leq f_X(s)f_Y(t) \Longleftrightarrow g(s,t) := \det \begin{vmatrix} f_X(s) & f_Y(s) \\ f_X(t) & f_Y(t) \end{vmatrix} \geq 0,$$

The implication $s \leq t \Rightarrow g(s,t) \geq 0$ can further be written as a single inequality

$$\mathbb{1}_{s\leq t} g(s,t) \geq 0, \quad \forall s, t \in \mathbb{R}, \tag{46}$$

Following Faugeras and Rüschendorf (2018), one can build a one-sided risk excess measure, i.e. a quantitative measure encoding the likelihood ratio order, by setting

$$\mathcal{D}_+^{\leq_{\text{lr}}}(X,Y) := \iint \mathbb{1}_{s\leq t}(g(s,t))^+ dsdt. \tag{47}$$

where $x^+ = \max(x, 0)$ is the positive part of x. ($\mathcal{D}_+^{\leq_{\text{lr}}}(X,Y)$ is finite since, by the triangle inequality, $0 \leq \mathcal{D}_+^{\leq_{\text{lr}}}(X,Y) \leq 2$.) Indeed, taking the integral of the positive part of the l.h.s. of the inequality (46) gives a cumulative quantification of the strength of the (qualitative) likelihood ratio order relation: if $X \leq_{\text{lr}} Y$, the higher $\mathcal{D}_+^{\leq_{\text{lr}}}(X,Y)$, the more Y dominates X in the likelihood ratio order. On the other hand, if $Y \leq_{\text{lr}} X$, then $\mathcal{D}_+^{\leq_{\text{lr}}}(X,Y) = 0$, so that $\mathcal{D}^{\leq_{\text{lr}}}$ is one-sided, in the terminology introduced by Faugeras and Rüschendorf (2018): the zero value of $\mathcal{D}_+^{\leq_{\text{lr}}}(X,Y)$ encodes when the order relation $Y \leq_{\text{lr}} X$ occurs.

Similarly, one can quantify the extent to which the reverse relation $Y \leq_{\text{lr}} X$ occurs. Indeed, since $g(t,s) = -g(s,t)$,

$$\begin{aligned} Y \leq_{\text{lr}} X &\Longleftrightarrow \{s \leq t \Rightarrow f_Y(s)f_X(t) - f_Y(t)f_X(s) \geq 0\} \\ &\Longleftrightarrow \mathbb{1}_{s\leq t} g(t,s) \geq 0 \Longleftrightarrow \mathbb{1}_{s\leq t} g(s,t) \leq 0. \end{aligned}$$

By setting

$$\mathcal{D}_-^{\leq_{\text{lr}}}(X,Y) := \iint \mathbb{1}_{s\leq t}(g(s,t))^- dsdt \geq 0, \tag{48}$$

where $x^- := \max(-x, 0) \geq 0$ is the negative part of x, one obtains a one-sided risk excess measure of $X \geq_{\text{lr}} Y$. Since $|x| = x^+ + x^-$, one can combine (47) and

(48) into a single measure by setting

$$\mathcal{D}^{\leq_{\rm lr}}(X,Y) := \mathcal{D}_{+}^{\leq_{\rm lr}}(X,Y) + \mathcal{D}_{-}^{\leq_{\rm lr}}(X,Y) = \iint \mathbb{1}_{s\leq t}|g(s,t)|dsdt \geq 0.$$

This yields a quantification of the likelihood ratio order "in both directions".

In view of the expression of the N_1 divergence (27) via the determinant (15), resp. the determinantal expression of the α-barycentric divergence of Faugeras (2024a) Definition 4.1 and 4.4, it appears that the N_1 divergence, resp. the $\alpha-$barycentric divergence, is such a two-sided quantification in the L_2 norm, resp. the L_α norm, of how much two probability measures P^X, P^Y differ, where the measure of difference is based on the twin quantification of the likelihood ratio order: N_1 tells how much P^X and P^Y differ, either from quantifying one side $X \leq_{\rm lr} Y$, or from quantifying the other side $Y \leq_{\rm lr} X$. In other words, as in Definition 4.4 in Faugeras (2024a), one can extend the notion of N_1 divergence from CoDa, i.e. probability measures on a finite set of atoms, to probability measures on $\mathbb{R}$, as

$$N_1(P_X \wedge P_Y) := \sqrt{\iint \mathbb{1}_{s\leq t}|g(s,t)|^2 dsdt},$$

where $P_X \wedge P_Y$ is here understood as a formal expression[19].

Remark 9. *By integrating (46), $X \leq_{lr} Y$ implies $X \leq_{hr} Y$, i.e. comparison in the hazard rate order, and the latter is equivalent to $X|X > t \geq_{st} Y|Y > t$ for all t, i.e. stochastic dominance of the conditional distributions over a threshold, see e.g. Müller and Stoyan (2002). The latter property is key in defining the concept of neutrality of Connor and Mosimann (1969), which is an analogue of intra-independence of components for CoDa (see also Faugeras (2024b)). Thus, N_1 also appears as a quantification of neutrality.*

6.3 Relation to generalized sine functions

Related to the exterior product approach are several generalizations of the sine function of a planar angle to the solid angle of multi-vectors and corresponding subspaces. Since the planar sine is a ratio of length, the natural extension of the sine to solid angle is to consider the generalization of the sine as a ratio of volumes. Hence, Euler (1781), Eriksson (1978), Lerman and Whitehouse (2009), were lead to define the polar sine (originally defined for $d+1$ vectors $\mathbf{x}_0, \ldots, \mathbf{x}_d$ in $\mathbb{R}^{d+1}$), as the ratio of the signed volume of the parallellotope represented by $\mathbf{x}_0 \wedge \ldots \wedge \mathbf{x}_d$ w.r.t the volume of the rectangular parallellotope with edges equal to the magnitudes of the vectors $\mathbf{x}_0, \ldots, \mathbf{x}_d$,

$$\text{psin}(\mathbf{x}_0, \ldots, \mathbf{x}_d) := \frac{\det A(\mathbf{x}_0 \wedge \ldots \wedge \mathbf{x}_d)}{||\mathbf{x}_0|| \ldots ||\mathbf{x}_d||},$$

[19] It could be defined mathematically as the product signed measure $P^X \wedge P^Y := P^X \otimes P^Y - P^Y \otimes P^X$, with zero total mass.

where $A(\mathbf{x}_0 \wedge \ldots \wedge \mathbf{x}_d)$ is the $A-$matrix (5) with columns the vectors $\mathbf{x}_0, \ldots, \mathbf{x}_d$, see[20], e.g., Section 2.4 in Lerman and Whitehouse (2009). A non-negative version of the polar sine, which works also for k vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$ in $\mathbb{R}^{d+1}$, is defined as

$$\mathrm{psin}_+(\mathbf{x}_1, \ldots, \mathbf{x}_k) = \frac{\sqrt{\det G}}{||\mathbf{x}_1|| \ldots ||\mathbf{x}_k||} = \frac{||\mathbf{x}_1 \wedge \ldots \wedge \mathbf{x}_k||}{||\mathbf{x}_1|| \ldots ||\mathbf{x}_k||},$$

where $G = G(\mathbf{x}_1, \ldots, \mathbf{x}_k)$ is the Gram matrix (8) associated with $\mathbf{x}_1, \ldots, \mathbf{x}_k$. Hadamard's inequality yields $0 \leq \mathrm{psin}_+(\mathbf{x}_1, \ldots, \mathbf{x}_k) \leq 1$ and in the case $k = d+1$, $\mathrm{psin}_+(\mathbf{x}_0, \ldots, \mathbf{x}_d) = |\mathrm{psin}(\mathbf{x}_0, \ldots, \mathbf{x}_d)|$. The polar sine can also be understood as a quantification of Hadamard's inequality (in the spirit of Section 6.1) to obtain a standardized/dimensionless measure of linear dependence, being zero if the vectors are dependent and one if they are orthogonal ("maximally independent").

As a consequence, it appears that the spherical/ℓ_2 normalized representation (22) of homogeneous simple bi-vectors $[\mathbf{x}] \wedge [\mathbf{y}]$ and pair of CoDa $[\mathbf{x}]_+ \wedge [\mathbf{y}]_+$ of Section 3 is the particular case, for $k = 2$, of a more general possible "spherical" representation of homogeneous k-vectors $[\mathbf{x}_1] \wedge \ldots \wedge [\mathbf{x}_k] = [\mathbf{x}_1 \wedge \ldots \wedge \mathbf{x}_k]$, resp. of k CoDa $[\mathbf{x}_1]_+ \wedge \ldots \wedge [\mathbf{x}_k]_+ = [\mathbf{x}_1 \wedge \ldots \wedge \mathbf{x}_k]_+$, as the ratio of a simple $k-$vector to the product of the norms of its $1-$vectors components

$$\frac{\mathbf{x}_1 \wedge \ldots \wedge \mathbf{x}_k}{||\mathbf{x}_1|| \ldots ||\mathbf{x}_k||}.$$

The norm of this representative of a homogeneous simple k CoDa thus corresponds to the polar sine, viz.

$$\mathrm{psin}_+(\mathbf{x}_1, \ldots, \mathbf{x}_k) = ||[\mathbf{x}_1 \wedge \ldots \wedge \mathbf{x}_k]_+||.$$

6.4 Relation to canonical angles of subspaces and projection matrices

Being an algebraization of geometric notions, the exterior product naturally expresses in geometric language. In particular, the scalar product between multivectors is related to Jordan (1875)'s canonical/principal angles between subspaces and Hotelling (1936)'s canonical correlations. Since subspaces are in one to one correspondence with orthogonal projection matrices, relations between subspaces can also be studied via their analogue between projection matrices, see e.g. Afriat (1957), A. Galántai (2008).

6.4.1 Canonical angles from Grassmann's viewpoint

Let us recall the definition of principal angles: let $\mathcal{V}, \mathcal{W}$ be nonzero subspaces, $p = \dim\mathcal{V}$, $q = \dim\mathcal{W}$, with respective basis $(\mathbf{v}_1, \ldots, \mathbf{v}_p)$ and $(\mathbf{w}_1, \ldots, \mathbf{w}_q)$. These subspaces are represented by matrices $V \in \mathbb{R}^{(d+1)\times p}$, $W \in \mathbb{R}^{(d+1)\times q}$ of

[20] psin is denoted $p_d \sin_0$ in Lerman and Whitehouse (2009) and polsin in Eriksson (1978)

full rank formed by the column vectors of the corresponding basis. Then, the relative position of $\mathcal{V}, \mathcal{W}$ is given by $m = \min(p, q)$ principal angles $0 \leq \theta_1 \leq \ldots \leq \theta_m \leq \pi/2$ and associated principal orthonormal basis $(\tilde{\mathbf{v}}_1, \ldots, \tilde{\mathbf{v}}_p)$ and $(\tilde{\mathbf{w}}_1, \ldots, \tilde{\mathbf{w}}_q)$ of V and W resp. s.t. for $1 \leq i, j \leq m$, $\langle \tilde{\mathbf{v}}_i | \tilde{\mathbf{w}}_j \rangle = \delta_{ij} \cos\theta_i$, where δ_{ij} stands for the Kronecker symbol, and $\langle \tilde{\mathbf{v}}_i | \tilde{\mathbf{w}}_j \rangle = 0$ for $i > m$ or $j > m$. The principal angles and basis can be obtained iteratively, by solving an eigensystem (Zassenhaus (1964)), or by the singular value decomposition of an oblique projection matrix or of a difference of projection (Björck and Golub (1973), Stewart and J. G. Sun (1990)).

Gluck (1967), Jiang (1996), Mandolesi (2021b), Mandolesi (2021a), Mandolesi (2020) have studied the relation between canonical angles and the scalar product and norm of multi-vectors (see also Miao and Ben-Israel (1992), Gunawan, Neswan, and Setya-Budhi (2005)). In particular, given a simple p-vector $\mathbf{v}_1 \wedge \ldots \wedge \mathbf{v}_p$ representing $\mathcal{V}$ and $P_{\mathcal{W}}$ the orthogonal projection matrix on $\mathcal{W}$, Mandolesi (2021a) defines the Grassmann (asymmetric) angle $\theta_{\mathcal{V},\mathcal{W}} \in [0, \pi/2]$ between $\mathcal{V}$ and $\mathcal{W}$ via its cosine as

$$\cos\theta_{\mathcal{V},\mathcal{W}} := \frac{||P_{\mathcal{W}}\mathbf{v}_1 \wedge \ldots \wedge P_{\mathcal{W}}\mathbf{v}_p||}{||\mathbf{v}_1 \wedge \ldots \wedge \mathbf{v}_p||}, \tag{49}$$

and the complementary Grassmann angle as $\theta^{\perp}_{\mathcal{V},\mathcal{W}} := \theta_{\mathcal{V},\mathcal{W}^{\perp}}$. Since the Plücker embedding represents multi-dimensional subspaces of $\mathbb{R}^{d+1}$ as one dimensional subspaces of $\bigwedge(\mathbb{R}^{d+1})$, $\theta_{\mathcal{V},\mathcal{W}}$ corresponds to a genuine angle between lines of $\bigwedge(\mathbb{R}^{d+1})$. $\theta_{\mathcal{V},\mathcal{W}}$ is an asymmetric measure of partial orthogonality: $\theta_{\mathcal{V},\mathcal{W}} = \pi/2$ if $p > q$ or if $\mathcal{V}$ is partially orthogonal to $\mathcal{W}$, i.e. if a nonzero vector of $\mathcal{V}$ is orthogonal to all vectors of $\mathcal{W}$, see e.g. Mandolesi (2020).

The relation with the canonical angles is (Proposition 2.6 in Mandolesi (2020)) that, if $m = p \leq q$, then

$$\cos\theta_{\mathcal{V},\mathcal{W}} = \prod_{i=1}^{m} \cos\theta_i, \quad \cos\theta^{\perp}_{\mathcal{V},\mathcal{W}} = \prod_{i=1}^{m} \sin\theta_i,$$

and the connection with the scalar product and norm of multi-vectors is that, if $m = p \leq q$, then

$$\begin{aligned} |\langle \mathbf{v}_1 \wedge \ldots \wedge \mathbf{v}_p | \mathbf{w}_1 \wedge \ldots \wedge \mathbf{w}_q \rangle| &= ||\mathbf{v}_1 \wedge \ldots \wedge \mathbf{v}_p||\, ||\mathbf{w}_1 \wedge \ldots \wedge \mathbf{w}_q|| \cos\theta_{\mathcal{V},\mathcal{W}} \\ ||\mathbf{v}_1 \wedge \ldots \wedge \mathbf{v}_p \wedge \mathbf{w}_1 \wedge \ldots \wedge \mathbf{w}_q|| &= ||\mathbf{v}_1 \wedge \ldots \wedge \mathbf{v}_p||\, ||\mathbf{w}_1 \wedge \ldots \wedge \mathbf{w}_q|| \cos\theta^{\perp}_{\mathcal{V},\mathcal{W}} \end{aligned}$$

These relations are multi-dimensional generalizations of the one-dimensional (symmetric) acute angle $\theta([\mathbf{x}], [\mathbf{y}]) \in [0, \pi/2]$ between lines $[\mathbf{x}] = \mathrm{span}(\mathbf{x})$ and $[\mathbf{y}] = \mathrm{span}(\mathbf{y})$:

$$\begin{aligned} |\langle \mathbf{x} | \mathbf{y} \rangle| &= ||\mathbf{x}||\, ||\mathbf{y}|| \cos\theta([\mathbf{x}], [\mathbf{y}]) \\ ||\mathbf{x} \wedge \mathbf{y}|| &= ||\mathbf{x}||\, ||\mathbf{y}|| \sin\theta([\mathbf{x}], [\mathbf{y}]) \end{aligned} \tag{50}$$

Note that an oriented version $\tilde{\theta}_{\mathcal{V},\mathcal{W}} \in [0, \pi]$ of the Grassmann angle $\theta_{\mathcal{V},\mathcal{W}} \in [0, \pi/2]$ can be defined by (see equation (8) of Mandolesi (2020))

$$\cos\tilde{\theta}_{\mathcal{V},\mathcal{W}} := \mathrm{sign}(\langle \mathbf{v}_1 \wedge \ldots \wedge \mathbf{v}_p | \mathbf{w}_1 \wedge \ldots \wedge \mathbf{w}_q \rangle) \cos\theta_{\mathcal{V},\mathcal{W}}, \tag{51}$$

so that the scalar product between multi-vectors writes

$$\langle \mathbf{v}_1 \wedge \ldots \wedge \mathbf{v}_p | \mathbf{w}_1 \wedge \ldots \wedge \mathbf{w}_q \rangle = ||\mathbf{v}_1 \wedge \ldots \wedge \mathbf{v}_p|| \, ||\mathbf{w}_1 \wedge \ldots \wedge \mathbf{w}_q|| \cos \tilde{\theta}_{\mathcal{V},\mathcal{W}} \quad (52)$$

6.4.2 Relation to the exterior covariance of CoDa

Applied to the bi-vectors $\mathbf{x} \wedge \boldsymbol{\mu}^{\mathbf{x}}$ and $\mathbf{y} \wedge \boldsymbol{\mu}^{\mathbf{y}}$ of Section 4, these formulas give the following geometrical interpretation of TCov_2 of Definition 4.1: by (52) and (50), (36) writes

$$\begin{aligned} \mathrm{TCov}_2([\mathbf{x}]_+, [\mathbf{y}]_+) &:= E\left(\frac{\langle \mathbf{x} \wedge \boldsymbol{\mu}^{\mathbf{x}} | \mathbf{y} \wedge \boldsymbol{\mu}^{\mathbf{y}} \rangle}{||\mathbf{x}|| \, ||\boldsymbol{\mu}^{\mathbf{x}}|| \, ||\mathbf{y}|| \, ||\boldsymbol{\mu}^{\mathbf{y}}||} \right) \\ &= E\left(\sin \theta(\mathbf{x}, \boldsymbol{\mu}^{\mathbf{x}}) \sin \theta(\mathbf{y}, \boldsymbol{\mu}^{\mathbf{y}}) \cos \tilde{\theta}_{\mathrm{span}(\mathbf{x},\boldsymbol{\mu}^{\mathbf{x}}),\mathrm{span}(\mathbf{y},\boldsymbol{\mu}^{\mathbf{y}})} \right) \end{aligned} \quad (53)$$

Such formula is the analogue for CoDa of the total covariance for Euclidean vectors. Indeed, for Euclidean vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{d+1}$, with covariance matrix $\Sigma_{\mathbf{xy}} = E((\mathbf{x} - E\mathbf{x})(\mathbf{y} - E\mathbf{y})^T)$, let us define the total covariance of $\mathbf{x}, \mathbf{y}$ as

$$\begin{aligned} \mathrm{TCov}(\mathbf{x}, \mathbf{y}) := \mathrm{trace}(\Sigma_{\mathbf{xy}}) &= E\left(\sum_i (x_i - Ex_i)(y_i - Ey_i) \right) \\ &= E(\langle \mathbf{x} - E\mathbf{x} | \mathbf{y} - E\mathbf{y} \rangle) \\ &= E(||\mathbf{x} - E\mathbf{x}|| \, ||\mathbf{y} - E\mathbf{y}|| \cos \theta(\mathbf{x} - E\mathbf{x}, \mathbf{y} - E\mathbf{y})). \end{aligned} \quad (54)$$

TCov writes as the expected product of the distances between $\mathbf{x}, \mathbf{y}$ and their respective means $E\mathbf{x}, E\mathbf{y}$, times the angle between these displacement vectors. Thus, it gives a valid global scalar measure, summarizing the covariation between two random vectors, while the full picture, i.e. the covariation components by components, is given by the covariance matrix $\Sigma_{\mathbf{xy}}$.

Formula (53) is the homogenized version of (54): $\sin \theta(\mathbf{x}, \boldsymbol{\mu}^{\mathbf{x}})$, $\sin \theta(\mathbf{y}, \boldsymbol{\mu}^{\mathbf{y}})$ now measures the scale invariant sine distance between CoDa $[\mathbf{x}]_+$, $[\mathbf{y}]_+$ and their respective means $[\boldsymbol{\mu}^{\mathbf{x}}]_+$ and $[\boldsymbol{\mu}^{\mathbf{y}}]_+$, and $\tilde{\theta}_{\mathrm{span}(\mathbf{x},\boldsymbol{\mu}^{\mathbf{x}}),\mathrm{span}(\mathbf{y},\boldsymbol{\mu}^{\mathbf{y}})}$ is the cosine of the oriented Grassmann angle (49), i.e. the product of the cosines (up to sign) of the two canonical angles between the projective lines corresponding to the planes $\mathrm{span}(\mathbf{x}, \boldsymbol{\mu}^{\mathbf{x}})$, and $\mathrm{span}(\mathbf{y}, \boldsymbol{\mu}^{\mathbf{y}})$.

For the components $\mathrm{Cov}_2([\mathbf{x}]_+, [\mathbf{y}]_+)_{i,j}$ of the exterior covariance matrix of Definition 4.1 i), such geometrical interpretation breaks down, as the ratio in (34) involves in the numerator the scalar product of the *projection* on the (i, j) plane of the two bi-vectors, whereas the denominator is the norm of the (non-projected) vectors. Still, the concept of the exterior covariance matrix remains meaningful: first, its construction is analogous to the Euclidean vector case. Second, it allows to decompose the total exterior covariance along its components, as the sum of the components of the exterior covariance matrix is the total exterior covariance.

6.4.3 Relation to exterior correlation of CoDa

Similar discussion can be carried on for the exterior correlation matrices, Definition 4.5. Regarding the exterior total correlation coefficient, Definition 4.6 mimics in equation (42) the definition of the correlation coefficient of Euclidean vectors, i.e. as a ratio of the total exterior covariance by the square root of the total exterior variances: one takes the expectation of the expected scalar product of bi-vectors, divided by the square root of the product of the expected norm of the bi-vectors. In view of (52), it also makes sense to consider directly the average of the Grassmann angle as a measure of correlation for CoDa. We are thus lead to the following definition.

Definition 6.1. *The Grassmann total correlation coefficient between* $[\mathbf{x}]_+$ *and* $[\mathbf{y}]_+$ *is defined as*

$$TCor_G([\mathbf{x}]_+, [\mathbf{y}]_+) := E\left(\frac{\langle \mathbf{x}\wedge\boldsymbol{\mu}^{\mathbf{x}} | \mathbf{y}\wedge\boldsymbol{\mu}^{\mathbf{y}}\rangle}{\|\mathbf{x}\wedge\boldsymbol{\mu}^{\mathbf{x}}\|\,\|\mathbf{y}\wedge\boldsymbol{\mu}^{\mathbf{y}}\|}\right) = E\cos\tilde{\theta}_{span(\mathbf{x},\boldsymbol{\mu}^{\mathbf{x}}),span(\mathbf{y},\boldsymbol{\mu}^{\mathbf{y}})},$$

where $\tilde{\theta}_{span(\mathbf{x},\boldsymbol{\mu}^{\mathbf{x}}),span(\mathbf{y},\boldsymbol{\mu}^{\mathbf{y}})}$ *is the oriented Grassmann angle (51) between the projective lines corresponding to the planes* $span(\mathbf{x},\boldsymbol{\mu}^{\mathbf{x}})$*, and* $span(\mathbf{y},\boldsymbol{\mu}^{\mathbf{y}})$*.*

6.4.4 Relation to the gap distance of projection matrices

Eventually, we hinted in the Appendix of Faugeras (2023), at the possibility of representing a CoDa point $[\mathbf{x}]_+$ as a projection matrix $\text{proj}_{\mathbf{x}} := \mathbf{x}\mathbf{x}^T/(\mathbf{x}^T\mathbf{x})$. Such representation was used in the first proof of the triangle inequality in Theorem 3.4 v), where we expressed the N_2 norm in term of the Froebenius distance between projection matrices

$$N_2([\mathbf{x}]_+ \wedge [\mathbf{y}]_+) = \frac{1}{\sqrt{2}}||\text{proj}_{\mathbf{x}} - \text{proj}_{\mathbf{y}}||_F.$$

For a unit norm $\mathbf{x}$, $||\text{proj}_{\mathbf{x}} - \text{proj}_{\mathbf{y}}||_F$ corresponds to $d(\mathbf{x}, \text{span}(\mathbf{y})) = \inf_{\lambda\in\mathbb{R}} ||\mathbf{x} - \lambda\mathbf{y}||$, see e.g. Qiu, Zhang, and Li (2005), Aurél Galántai (2004).

Let us mention that, more generally, the gap distance (used, notably in functional analysis, see e.g. Kato (1995), Aurél Galántai (2004)) between equidimensional subspaces $\mathcal{U}, \mathcal{V}$ is defined as

$$\text{gap}(\mathcal{U}, \mathcal{V}) := ||\text{proj}_{\mathcal{U}} - \text{proj}_{\mathcal{V}}||_F = \sin\theta_m$$

where $\text{proj}_{\mathcal{U}}$, resp. $\text{proj}_{\mathcal{V}}$, is the orthogonal projection matrix on $\mathcal{U}$, resp. $\mathcal{V}$ and is equal to the sine of their largest principal angle.

7 Conclusion

We have thus tried to give a guided tour on the projective geometry viewpoint to CoDa analysis based on the exterior product and Grassmann's algebra, so that these powerful tools can be made available to the CoDa community. We supplied a gentle, intuitive approach to these abstract objects, interpreting geometrically bi-vectors as oriented parallelograms, which admit linear representations as Plücker or Compound matrices. A scalar product and norm on bi-vectors, extending the usual ones, can be defined and transferred to equivalence classes. For CoDa, this results in a pseudo-scalar product and pseudo-norm. It is remarkable that the pseudo-norm obtained with the ℓ_1 normalization yields the 2-barycentric divergence of Faugeras (2024a), which was obtained in that paper from the affine geometry viewpoint, with a somehow heuristic reasoning based on barycentric coordinates. The present paper thus gives a theoretical justification to the constructs of Faugeras (2024a).

The ℓ_2 normalization of the pseudo-norm yields added benefits for measuring the closeness of CoDa: one gets a log-free, orthogonally invariant, bounded distance satisfying the triangle inequality on the full CoDa space (hence, allowing for zeroes). Elaborating further, we introduced key statistical constructs for measuring the dispersion and covariation of Coda, both at the level of pairs of components, and globally. We described the connections with related notions of the literature, like divergences based on quantifying the Cauchy-Schwarz inequality or the likelihood ratio order, the polar sine, canonical angles between subspaces and projections matrices.

In turn, the exterior distance and variance/covariance/correlation matrices introduced in the paper can serve as the backbone for further statistical analysis. For length reasons, we have barely scratched the surface of potential statistical applications and only sketched possible statistical analysis based on Fréchet means, weighted and Mahalanobis-type distances, Gaussian distributions, Optimal Transportation distances, etc. Our objective was to lay the theoretical foundations for a unified statistical analysis from the projective viewpoint aided with the exterior product. The main message is that the exterior product yields an underlying linear structure and a powerful algebraic tool for the analysis of CoDa in a geometric way. More detailed and applied statistical applications will be pursued elsewhere.

Appendix A: The exterior product and Grassmann's algebra in a nutshell by the abstract algebraic approach

It is instructive to look how the general, abstract algebraic, top-bottom, construction of the exterior product and Grassmann's algebra traditionally found in the literature is in agreement with the bottom-up, based on geometric intuition, approach of Section 2. We thus give a very quick introduction to the general theory, following Kung, Rota, and Yan (2009) Chapter 6.6. Let V be a n-dimensional real vector space. For $\mathbf{u}_1, \ldots, \mathbf{u}_k \in V$, their exterior product is the formal expression $\mathbf{u}_1 \wedge \ldots \wedge \mathbf{u}_k$, and the k-th exterior power $\Lambda^k(V)$ of V

is the vector space of all linear combination of such expressions, modulo[21] the following relations

$$\mathbf{u}_1 \wedge \ldots \wedge a\mathbf{u}_i + b\mathbf{v}_i \wedge \ldots \wedge \mathbf{u}_k = a\mathbf{u}_1 \wedge \ldots \wedge \mathbf{u}_i \wedge \ldots \wedge \mathbf{u}_k + b\,\mathbf{u}_1 \wedge \ldots \wedge \mathbf{v}_i \wedge \ldots \wedge \mathbf{u}_k$$
$$\mathbf{u}_1 \wedge \ldots \wedge \mathbf{u}_i \wedge \ldots \wedge \mathbf{u}_j \wedge \ldots \wedge \mathbf{u}_k = -\mathbf{u}_1 \wedge \ldots \wedge \mathbf{u}_j \wedge \ldots \wedge \mathbf{u}_i \wedge \ldots \wedge \mathbf{u}_k$$

expressing the multi-linearity of the product and its anti-symmetry. If $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is a basis of V, then the set of multi-vectors $\mathbf{e}_{i_1} \wedge \mathbf{e}_{i_2} \wedge \ldots \wedge \mathbf{e}_{i_k}$ with $1 \le i_1 < i_2 < \ldots i_k \le n$ form a basis of $\Lambda^k(V)$. The exterior algebra $\Lambda(V)$ is the algebra on the direct sum $\Lambda(V) = \bigoplus_{k=1}^n \Lambda^k(V)$ with (associative) multiplication $\wedge$ defined formally on exterior products by

$$(\mathbf{u}_1 \wedge \ldots \wedge \mathbf{u}_k) \wedge (\mathbf{v}_1 \wedge \ldots \wedge \mathbf{v}_j) := \mathbf{u}_1 \wedge \ldots \wedge \mathbf{u}_k \wedge \mathbf{v}_1 \wedge \ldots \wedge \mathbf{v}_j$$

and extended by linearity.

For two vectors $\mathbf{x} = \sum_{i=1}^n x_i \mathbf{e}_i$, $\mathbf{y} = \sum_{i=1}^n y_i \mathbf{e}_i$, their exterior product gives, by bilinearity and antisymmetry,

$$\begin{aligned}
\mathbf{x} \wedge \mathbf{y} &= \left(\sum_{i=1}^n x_i \mathbf{e}_i\right) \wedge \left(\sum_{j=1}^n y_j \mathbf{e}_j\right) \\
&= \sum_{i=1}^n \sum_{j=1}^n x_i y_j \mathbf{e}_i \wedge \mathbf{e}_j \\
&= \sum_{1 \le i < j \le n} x_i y_j \mathbf{e}_i \wedge \mathbf{e}_j + \sum_{1 \le j < i \le n} x_i y_j \mathbf{e}_i \wedge \mathbf{e}_j \\
&= \sum_{1 \le i < j \le n} x_i y_j \mathbf{e}_i \wedge \mathbf{e}_j + \sum_{1 \le i < j \le n} x_j y_i \mathbf{e}_j \wedge \mathbf{e}_i \\
&= \sum_{1 \le i < j \le n} (x_i y_j - y_j x_i) \mathbf{e}_i \wedge \mathbf{e}_j,
\end{aligned}$$

where we used the fact that $\mathbf{e}_i \wedge \mathbf{e}_i = 0$ and $\mathbf{e}_j \wedge \mathbf{e}_i = -\mathbf{e}_i \wedge \mathbf{e}_j$. Thus, the simple bi-vector $\mathbf{x} \wedge \mathbf{y}$ decomposes onto the bi-vector basis $(\mathbf{e}_i \wedge \mathbf{e}_j)_{i<j}$, into the $n(n-1)/2$ components

$$\det \begin{pmatrix} x_i & y_i \\ x_j & y_j \end{pmatrix} = p_{ij}(\mathbf{x}, \mathbf{y}),$$

in agreement with our definition of Plücker components (2).

Appendix B: Notations

- (Column) vectors $\mathbf{x} = (x_0, x_1, \ldots, x_d) \in \mathbb{R}^{d+1}$ are written in bold letters, and operations on vector are interpreted component wise. $\mathrm{span}(\mathbf{x}, \mathbf{y})$ vector subspace spanned by $\mathbf{x}, \mathbf{y}$.

[21] This corresponds to defining the exterior power as a quotient space of the tensor product space by these relations, see e.g. Bhatia (1997) and Federer (1969).

- $||.|| = ||.||_2$ denotes the usual norm (Euclidean/ℓ^2), $||\mathbf{x}||_1 = \sum_{i=0}^{d} |x_i|$ the ℓ^1 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, $S^d = \{\mathbf{x} \in \mathbb{R}^{d+1} : ||\mathbf{x}|| = 1\}$ the unit sphere.
- $\mathbb{P}^d$ real projective space of dimension d induced by $\mathbb{R}^{d+1}$, $\mathbb{P}_+^d$ the space of (non-negative) CoDa vectors as equivalence classes $[.]_+$, $\mathbb{P}_{++}^d$ the space of positive CoDa vectors.
- $[\mathbf{x}] \in \mathbb{P}^d$, $[\mathbf{x}]_+ \in \mathbb{P}_+^d$ are projective, resp. CoDa, equivalence classes of $\mathbf{x}$.
- $\mathcal{C}(\mathbf{x}) = \mathbf{x}/||\mathbf{x}||_1$ closure operation/normalization by the ℓ_1 norm. $\mathcal{S}(\mathbf{x}) = \mathbf{x}/||\mathbf{x}||$ spherical projection/normalization by the ℓ_2 norm. $\mathcal{R}(\mathbf{x})$ square root transformation.
- $\mathbf{x} \wedge \mathbf{y}$ parallelogram/exterior product of two vectors/simple bi-vector. $\bigwedge^k(V)$ kth exterior power of V, $\bigwedge(V)$ exterior algebra of V.
- $A(\mathbf{x} \wedge \mathbf{y})$ A-matrix of $\mathbf{x} \wedge \mathbf{y}$, $P(\mathbf{x} \wedge \mathbf{y})$ Plücker P-matrix of $\mathbf{x} \wedge \mathbf{y}$, $C_p(A)$ p-th Compound matrix of A. $G(\mathbf{x} \wedge \mathbf{y})$ Gram matrix of the $A(\mathbf{x} \wedge \mathbf{y})$ matrix.
- $\langle \mathbf{x} \wedge \mathbf{y} | \mathbf{r} \wedge \mathbf{s} \rangle$ scalar product of two simple bi-vectors, $||\mathbf{x} \wedge \mathbf{y}||$ norm of a simple bi-vector.
- $\langle [\mathbf{x}]_+ \wedge [\mathbf{y}]_+ | [\mathbf{r}]_+ \wedge [\mathbf{s}]_+ \rangle_1$, resp. $\langle [\mathbf{x}]_+ \wedge [\mathbf{y}]_+ | [\mathbf{r}]_+ \wedge [\mathbf{s}]_+ \rangle_2$ pseudo-scalar product between pairs of CoDa based on the ℓ_1, resp. ℓ_2 normalization. $N_1([\mathbf{x}]_+ \wedge [\mathbf{y}]_+)$, resp. $N_2([\mathbf{x}]_+ \wedge [\mathbf{y}]_+)$ corresponding pseudo-norm of a pair of CoDa. $N_{1/2}([\mathbf{x}]_+ \wedge [\mathbf{y}]_+)$ pseudo-norm based on the square root transform $\mathcal{R}$. $N_{2,W}([\mathbf{x}]_+ \wedge [\mathbf{y}]_+)$ weighted N_2 pseudo-norm with weight matrix W.
- $\mathrm{Cov}_2([\mathbf{x}]_+, [\mathbf{y}]_+)$, $\mathrm{Var}_2([\mathbf{x}]_+)$, $\mathrm{Cor}_2([\mathbf{x}]_+, [\mathbf{y}]_+)$ exterior covariance, variance, correlation matrices based on the ℓ_2 normalization. $\mathrm{TCov}_2([\mathbf{x}]_+, [\mathbf{y}]_+)$, $\mathrm{TVar}_2([\mathbf{x}]_+)$, $\mathrm{TCor}_2([\mathbf{x}]_+, [\mathbf{y}]_+)$ corresponding total (scalar) measure of exterior covariance, variance and correlation. $\mathrm{TCor}_G([\mathbf{x}]_+, [\mathbf{y}]_+)$ Grassmann total correlation coefficient.
- $\theta(\mathbf{x}, \mathbf{y}) \in [0, \pi]$ (symmetric) angle between vectors, $\theta([\mathbf{x}], [\mathbf{y}]) \in [0, \pi/2]$, $\theta([\mathbf{x}]_+, [\mathbf{y}]_+) \in [0, \pi/2]$ acute angle between lines/rays, $\theta_{\mathcal{V},\mathcal{W}} \in [0, \pi/2]$ Grassmann angle between subspaces, $\tilde{\theta}_{\mathcal{V},\mathcal{W}} \in [0, \pi]$ oriented Grassmann angle between subspaces, $\theta_{\mathcal{V},\mathcal{W}}^{\perp} \in [0, \pi/2]$ complementary Grassmann angle between subspaces.

Acknowledgements

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

References

Afriat, S. N. (1957). "Orthogonal and oblique projectors and the characteristics of pairs of vector spaces". In: *Proc. Cambridge Philos. Soc.* 53, pp. 800–816. ISSN: 0008-1981. DOI: 10.1017/s0305004100032916. URL: https://doi.org/10.1017/s0305004100032916.

Aitken, AC (1956). *Determinants and matrices.*

Berger, Marcel (1987). *Geometry. II.* Universitext. Translated from the French by M. Cole and S. Levy. Springer-Verlag, Berlin, pp. x+406. ISBN: 3-540-17015-4. DOI: 10.1007/978-3-540-93816-3. URL: https://doi.org/10.1007/978-3-540-93816-3.

Bhatia, Rajendra (1997). *Matrix analysis.* Vol. 169. Graduate Texts in Mathematics. Springer-Verlag, New York, pp. xii+347. ISBN: 0-387-94846-5. DOI: 10.1007/978-1-4612-0653-8. URL: https://doi.org/10.1007/978-1-4612-0653-8.

Björck, Åke and Gene H. Golub (1973). "Numerical methods for computing angles between linear subspaces". In: *Math. Comp.* 27, pp. 579–594. ISSN: 0025-5718. DOI: 10.2307/2005662. URL: https://doi.org/10.2307/2005662.

Boutin, Debra L, Ronald F Gleeson, and Robert M Williams (1996). *Wedge Theory, Compund Matrices: Properties and Applications.* Naval Air Warfare Center, Aircraft Division.

Brown, L. D., Arthur Cohen, and W. E. Strawderman (1976). "A complete class theorem for strict monotone likelihood ratio with applications". In: *Ann. Statist.* 4(4), pp. 712–722. ISSN: 0090-5364. URL: http://links.jstor.org/sici?sici=0090-5364(197607)4:4%3C712:ACCTFS%3E2.0.CO;2-W&origin=MSN.

Browne, John (2012). *Grassmann Algebra Volume 1: Foundations: Exploring Extended Vector Algebra with Mathematica.* Vol. 1. John M Browne.

Budka, Marcin, Bogdan Gabrys, and Katarzyna Musial (2011). "On accuracy of PDF divergence estimators and their applicability to representative data sampling". In: *Entropy* 13(7), pp. 1229–1266.

Connor, Robert J. and James E. Mosimann (1969). "Concepts of Independence for Proportions with a Generalization of the Dirichlet Distribution". In: *Journal of the American Statistical Association* 64(325), pp. 194–206. DOI: 10.1080/01621459.1969.10500963. eprint: https://www.tandfonline.com/doi/pdf/10.1080/01621459.1969.10500963. URL: https://www.tandfonline.com/doi/abs/10.1080/01621459.1969.10500963.

Dieudonne, J. (1979). "The tragedy of grassmann". In: *Linear and Multilinear Algebra* 8(1), pp. 1–14. DOI: 10.1080/03081087908817294. eprint: https://doi.org/10.1080/03081087908817294. URL: https://doi.org/10.1080/03081087908817294.

Doran, Chris and Anthony Lasenby (2003). *Geometric algebra for physicists.* Cambridge University Press, Cambridge, pp. xiv+578. ISBN: 0-521-48022-1; 978-0-521-71595-9. DOI: 10.1017/CBO9780511807497. URL: https://doi.org/10.1017/CBO9780511807497.

Dorst, Leo, Daniel Fontijne, and Stephen Mann (2009). *Geometric algebra for computer science (revised edition): An object-oriented approach to geometry.* Morgan Kaufmann.

Eriksson, Folke (1978). "The law of sines for tetrahedra and n-simplices". In: *Geometriae Dedicata* 7(1), pp. 71–80. DOI: 10.1007/BF00181352. URL: https://doi.org/10.1007/BF00181352.

Euler, Leonhard (1781). "De mensura angulorum solidorum". In: *Acta Academiae Scientiarum Imperialis Petropolitanae*, pp. 31–54.

Faugeras, Olivier P. (Dec. 2023). "An invitation to intrinsic compositional data analysis using projective geometry and Hilbert's metric". TSE Working Paper, no. 23-1496. URL: https://www.tse-fr.eu/fr/publications/invitation-intrinsic-compositional-data-analysis-using-projective-geometry-and-hilberts-metric.

Faugeras, Olivier P. (July 2024a). "Log-Free Divergence and Covariance matrix for Compositional Data I: The Affine/Barycentric approach". Submitted.

Faugeras, Olivier P. (Jan. 2024b). "The Stick-Breaking and Ordering Representation of Compositional Data: Copulas and Regression models". TSE Working Paper, n° 24-1500. URL: https://hal.science/hal-04409705.

Faugeras, Olivier P. and Ludger Rüschendorf (2018). "Risk excess measures induced by hemi-metrics". In: *Probability, Uncertainty and Quantitative Risk* 3(0), p. 6. ISSN: 2095-9672. DOI: 10.1186/s41546-018-0032-0. URL: https://www.aimsciences.org/article/id/b11caab2-557c-4d92-8e94-b4f3f2c6f662.

Federer, Herbert (1969). *Geometric measure theory.* Die Grundlehren der mathematischen Wissenschaften, Band 153. Springer-Verlag New York, Inc., New York, pp. xiv+676.

Galántai, A. (2008). "Subspaces, angles and pairs of orthogonal projections". In: *Linear Multilinear Algebra* 56(3), pp. 227–260. ISSN: 0308-1087. DOI: 10.1080/03081080600743338. URL: https://doi.org/10.1080/03081080600743338.

Galántai, Aurél (2004). *Projectors and projection methods.* Vol. 6. Advances in Mathematics (Dordrecht). Kluwer Academic Publishers, Boston, MA, pp. x+287. ISBN: 1-4020-7572-3. DOI: 10.1007/978-1-4419-9180-5. URL: https://doi.org/10.1007/978-1-4419-9180-5.

Gantmacher, F. R. (1998). *The theory of matrices. Vol. 1.* Translated from the Russian by K. A. Hirsch, Reprint of the 1959 translation. AMS Chelsea Publishing, Providence, RI, pp. x+374. ISBN: 0-8218-1376-5.

Gluck, Herman (1967). "Higher curvatures of curves in euclidean space. II". In: *Amer. Math. Monthly* 74, pp. 1049–1056. ISSN: 0002-9890. DOI: 10.2307/2313601. URL: https://doi.org/10.2307/2313601.

Grassmann, H. (1995). *A New Branch of Mathematics: The "Ausdehnungslehre" of 1844 and Other Works.* Open Court. ISBN: 9780812692754. URL: https://books.google.fr/books?id=978fAQAAIAAJ.

Grassmann, Hermann (1878). *Die Ausdehnungslehre von 1844 oder die lineale Ausdehnungslehre.* Wigand: Leipzig.

Gunawan, Hendra, Oki Neswan, and Wono Setya-Budhi (2005). "A formula for angles between subspaces of inner product spaces". In: *Beiträge Algebra Geom.* 46(2), pp. 311–320. ISSN: 0138-4821.

Hestenes, David and Garret Sobczyk (1984). *Clifford algebra to geometric calculus*. Fundamental Theories of Physics. A unified language for mathematics and physics. D. Reidel Publishing Co., Dordrecht, pp. xviii+314. ISBN: 90-277-1673-0. DOI: 10.1007/978-94-009-6292-7. URL: https://doi.org/10.1007/978-94-009-6292-7.

Hotelling, Harold (Dec. 1936). "Relations Between Two Sets of Variates". In: *Biometrika* 28(3-4), pp. 321–377. ISSN: 0006-3444. DOI: 10.1093/biomet/28.3-4.321. eprint: https://academic.oup.com/biomet/article-pdf/28/3-4/321/586830/28-3-4-321.pdf. URL: https://doi.org/10.1093/biomet/28.3-4.321.

Jiang, Sheng (1996). "Angles between Euclidean subspaces". In: *Geom. Dedicata* 63(2), pp. 113–121. ISSN: 0046-5755. DOI: 10.1007/BF00148212. URL: https://doi.org/10.1007/BF00148212.

Jordan, Camille (1875). "Essai sur la géométrie à n dimensions". In: *Bull. Soc. Math. France* 3, pp. 103–174. ISSN: 0037-9484. URL: http://www.numdam.org/item?id=BSMF_1875__3__103_2.

Kanatani, Kenichi (2015). *Understanding geometric algebra*. Hamilton, Grassmann, and Clifford for computer vision and graphics. CRC Press, Boca Raton, FL, pp. xv+192. ISBN: 978-1-4822-5950-6.

Kato, Tosio (1995). *Perturbation theory for linear operators*. Classics in Mathematics. Reprint of the 1980 edition. Springer-Verlag, Berlin, pp. xxii+619. ISBN: 3-540-58661-X.

Khosravi, Mehrdad and Michael D. Taylor (2008). "The wedge product and analytic geometry". In: *Amer. Math. Monthly* 115(7), pp. 623–644. ISSN: 0002-9890. DOI: 10.1080/00029890.2008.11920573. URL: https://doi.org/10.1080/00029890.2008.11920573.

Kung, Joseph P. S., Gian-Carlo Rota, and Catherine H. Yan (2009). *Combinatorics: the Rota way*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, pp. xii+396. ISBN: 978-0-521-73794-4. DOI: 10.1017/CBO9780511803895. URL: https://doi.org/10.1017/CBO9780511803895.

Lerman, Gilad and J. Tyler Whitehouse (2009). "On d-dimensional d-semimetrics and simplex-type inequalities for high-dimensional sine functions". In: *J. Approx. Theory* 156(1), pp. 52–81. ISSN: 0021-9045. DOI: 10.1016/j.jat.2008.03.005. URL: https://doi.org/10.1016/j.jat.2008.03.005.

Mandolesi, André L. G. (2020). *Grassmann angle formulas and identities*. arXiv: 2005.12700 [math.GM]. URL: https://arxiv.org/abs/2005.12700.

Mandolesi, André L. G. (2021a). "Blade products and angles between subspaces". In: *Adv. Appl. Clifford Algebr.* 31(5), Paper No. 69, 33. ISSN: 0188-7009. DOI: 10.1007/s00006-021-01169-w. URL: https://doi.org/10.1007/s00006-021-01169-w.

Mandolesi, André L. G. (2021b). *Grassmann angles between real or complex subspaces*. arXiv: 1910.00147 [math.MG]. URL: https://arxiv.org/abs/1910.00147.

Miao, Jian Ming and Adi Ben-Israel (1992). "On principal angles between subspaces in $\mathbf{R}^n$". In: *Linear Algebra Appl.* 171, pp. 81–98. ISSN: 0024-3795. DOI: 10.1016/0024-3795(92)90251-5. URL: https://doi.org/10.1016/0024-3795(92)90251-5.

Mikusinski, Piotr and Michael D Taylor (2012). *An introduction to multivariable analysis from vector to manifold.* Springer Science & Business Media.

Müller, Alfred and Dietrich Stoyan (2002). *Comparison methods for stochastic models and risks.* Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, pp. xii+330. ISBN: 0-471-49446-1.

Nielsen, Frank, Ke Sun, and Stéphane Marchand-Maillet (2017). "On Hölder projective divergences". In: *Entropy* 19(3), p. 122.

Perwass, Christian (2009). *Geometric algebra with applications in engineering.* Vol. 4. Geometry and Computing. Springer-Verlag, Berlin, pp. xiv+385. ISBN: 978-3-540-89067-6.

Pfanzagl, J. (1979). "On optimal median unbiased estimators in the presence of nuisance parameters". In: *Ann. Statist.* 7(1), pp. 187–193. ISSN: 0090-5364. URL: http://links.jstor.org/sici?sici=0090-5364(197901)7:1%3C187:OOMUEI%3E2.0.CO;2-K&origin=MSN.

Postnikov, M. (1988). *Geometria analítica.* Lições de geometria. Primeiro semestre. [Lectures in geometry. Semester I], Translated from the Russian by M. Dombrovskii'. "Mir", Moscow, p. 341. ISBN: 5-03-000510-2.

Prells, Uwe, Michael I. Friswell, and Seamus D. Garvey (2003). "Use of geometric algebra: compound matrices and the determinant of the sum of two matrices". In: *R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci.* 459(2030), pp. 273–285. ISSN: 1364-5021. DOI: 10.1098/rspa.2002.1040. URL: https://doi.org/10.1098/rspa.2002.1040.

Qiu, Li, Yanxia Zhang, and Chi-Kwong Li (2005). "Unitarily invariant metrics on the Grassmann space". In: *SIAM J. Matrix Anal. Appl.* 27(2), pp. 507–531. ISSN: 0895-4798. DOI: 10.1137/040607605. URL: https://doi.org/10.1137/040607605.

Rachev, Svetlozar T. (1991). *Probability metrics and the stability of stochastic models.* Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons, Ltd., Chichester, pp. xiv+494. ISBN: 0-471-92877-1.

Rachev, Svetlozar T. and Ludger Rüschendorf (1998). *Mass transportation problems. Vol. I.* Probability and its Applications (New York). Theory. Springer-Verlag, New York, pp. xxvi+508. ISBN: 0-387-98350-3.

Rosén, Andreas (2019). *Geometric multivector analysis.* Birkhäuser Advanced Texts: Basler Lehrbücher. [Birkhäuser Advanced Texts: Basel Textbooks]. From Grassmann to Dirac. Birkhäuser/Springer, Cham, pp. xii+465. ISBN: 978-3-030-31410-1; 978-3-030-31411-8. DOI: 10.1007/978-3-030-31411-8. URL: https://doi.org/10.1007/978-3-030-31411-8.

Scealy, J. L. and A. H. Welsh (2014). "Colours and cocktails: compositional data analysis 2013 Lancaster lecture". In: *Aust. N. Z. J. Stat.* 56(2), pp. 145–169. ISSN: 1369-1473. DOI: 10.1111/anzs.12073. URL: https://doi.org/10.1111/anzs.12073.

Spivak, Michael (1965). *Calculus on manifolds. A modern approach to classical theorems of advanced calculus.* W. A. Benjamin, Inc., New York-Amsterdam, pp. xii+144.

Stewart, G. W. and Ji Guang Sun (1990). *Matrix perturbation theory.* Computer Science and Scientific Computing. Academic Press, Inc., Boston, MA, pp. xvi+365. ISBN: 0-12-670230-6.

Tsagris, Michail T, Simon Preston, and Andrew TA Wood (2011). "A data-based power transformation for compositional data". In: *arXiv preprint arXiv:1106.1451*.

Villani, Cédric (2009). *Optimal transport.* Vol. 338. Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Old and new. Springer-Verlag, Berlin, pp. xxii+973. ISBN: 978-3-540-71049-3. DOI: 10.1007/978-3-540-71050-9. URL: https://doi.org/10.1007/978-3-540-71050-9.

Watson, D. F. and G. M. Philip (Feb. 1989). "Measures of variability for geological data". In: *Mathematical Geology* 21(2), pp. 233–254. ISSN: 1573-8868. DOI: 10.1007/BF00893217. URL: https://doi.org/10.1007/BF00893217.

Winitzki, Sergei (2009). *Linear algebra via exterior products.* lulu.com.

Xu, Yiming et al. (2021). "Analysis of the ratio of ℓ_1 and ℓ_2 norms in compressed sensing". In: *Appl. Comput. Harmon. Anal.* 55, pp. 486–511. ISSN: 1063-5203. DOI: 10.1016/j.acha.2021.06.006. URL: https://doi.org/10.1016/j.acha.2021.06.006.

Yin, Penghang, Ernie Esser, and Jack Xin (2014). "Ratio and difference of l_1 and l_2 norms and sparse representation with coherent dictionaries". In: *Commun. Inf. Syst.* 14(2), pp. 87–109. ISSN: 1526-7555. DOI: 10.4310/CIS.2014.v14.n2.a2. URL: https://doi.org/10.4310/CIS.2014.v14.n2.a2.

Zassenhaus, Hans (1964). "Advanced Problems and Solutions: Solutions: 5076". In: *Amer. Math. Monthly* 71(2), pp. 218–219. ISSN: 0002-9890. DOI: 10.2307/2311774. URL: https://doi.org/10.2307/2311774.