SORT 47 (2) July-December 2023, 1-22 DOI:

Fundamentals of convex optimization for compositional data

Jordi Saperas Riera[1], Josep Antoni Martín Fernández[3]
and Glòria Mateu Figueras[2]

Abstract

Many of the most popular statistical techniques incorporate optimisation problems in their inner workings. A convex optimisation problem is defined as the problem of minimising a convex function over a convex set. When traditional methods are applied to compositional data, misleading and incoherent results could be obtained. In this paper, we fill a gap in the specialised literature by introducing and rigorously defining novel concepts of convex optimisation for compositional data according to the Aitchison geometry. Convex sets and convex functions on the simplex are defined and illustrated.

MSC: *62F40, 62H15, 62H99, 62J10, 62J15, 62P25.*

Keywords: *Compositional data, logratio, simplex, proportion, function, convexity, optimisation.*

1. Introduction

Convex optimisation plays an important role in a wide range of scientific fields where statistical methods are applied. Thus, it has become particularly relevant in the design of experiments (Coetzer and Haines, 2017), variable selection (Susin et al., 2020), robust statistics (Boogaart et al., 2021), cluster analysis (Wang et al., 2020), and principal components analysis (Campbell and Wong, 2022).

In a broad sense, a convex optimisation problem in $\mathbb{R}^D$ is defined as (Boyd and Vandenberghe, 2004) the problem of minimising a convex function (objective function) over a convex set (feasible region). The feasible region is the set of all possible values of variables that verify the constraints of the problem. Commonly, it is assumed

[1,2,3] Department of Computer Science, Applied Mathematics and Statistics, University of Girona, Campus Montilivi, Edifici P-4, 17003 Girona, Spain.

[1] jordi.saperas@udg.edu

[2] gloria.mateu@udg.edu

[3] josepantoni.martin@udg.edu

Received: November 2022
Accepted: June 2023

that the convexity of the objective function and the feasible region corresponds to the ordinary convexity concept in real space, $\mathbb{R}^D$. However, in statistics, one has to take into account the particular geometry of the sample space of the variables. Compositional data (CoDa) (Aitchison, 1986) convey relative information because the variables describe relative contributions to a given total. These variables are called *parts* of a whole and are, usually, expressed in proportions, percentages or ppm. Historically (Aitchison, 1986), the sample space of CoDa is designed as the D-part unit simplex $\mathcal{S}^D = \{\mathbf{x} \in \mathbb{R}^D : x_j > 0; \sum x_j = 1; j = 1,\ldots,D\}$. The formal geometric framework for the analysis of CoDa was first introduced in Pawlowsky-Glahn and Egozcue (2001) and Billheimer, Guttorp and Fagan (2001). It was called the Aitchison geometry, and it was later formally established in Barceló-Vidal and Martín-Fernández (2016). Such geometry allows compositions to be expressed as coordinates on an orthonormal basis, formed by logratios and called olr-coordinates (Egozcue et al., 2003; Martín-Fernández, 2019). In short, the analysis of CoDa involves the use of standard techniques in olr-coordinates; what has been known as the principle of working on coordinates (Mateu-Figueras, Pawlowsky-Glahn and Egozcue, 2011).

Applications of the log-ratio methodology are increasingly found across a wide range of scientific fields such as geosciences (Martín-Fernández et al., 2018), chemistry and physics (Halim et al., 2021), and health (Bates and Tibshirani, 2019; Dumuid et al., 2020), among others. CoDa methods have become particularly relevant for the analysis of time-use data, especially in relation to public health studies (Chastin et al., 2015; Dumuid et al., 2021; Kitano et al., 2020; Fairclough et al., 2018; Gupta et al., 2020). In time-use data, the parts are the time spent on different activities such as sleep, work, and a range of physical activities. Some optimisation problems of practical relevance can be formulated in this context. For example, one can be interested in deciding what composition of daily activities agreeing with some medical recommendations (constraints) optimises a particular health biomarker (objective function). Alternatively, given a time-use composition of a patient not following some medical recommendation (feasible region), the question of interest is: what modification to the daily activity composition would enable a swift transition (objective function) into a healthy behaviour region?

The aim of this work is to adapt the concepts related to convex optimisation for problems involving CoDa and considering the Aitchison geometry. These definitions are essential to correctly identify and classify convex optimisation problems on the simplex. The paper is organised as follows. Section 2 summarises the basic concepts of CoDa. In Section 3, the concept of convex set on the simplex with the Aitchison geometry is defined. The basic sets on the simplex are analysed and their compositional convexity is studied. Section 4 develops the concept of compositional convex function and introduces the conditions to classify a problem as a compositional convex optimisation problem. These are illustrated in Section 5 through a case study based on time-use data. Finally, Section 6 concludes with some important remarks.

The analyses discussed in this article were carried out in R (R-Core-Team, 2022) and using the package *compositions* (van den Boogaart and Tolosana-Delgado, 2008).

2. Basic elements of Aitchison geometry

When analysing CoDa one assumes the property of *scale invariance*. That is, it is assumed that each D-part composition $\mathbf{w} \in \mathbb{R}_+^D$ is a member of an equivalence class (Barceló-Vidal and Martín-Fernández, 2016). In other words, the information contained in $\mathbf{w}$ is the same as in any other composition $k \cdot \mathcal{C}[\mathbf{w}]$ for any real scalar $k > 0$, where $\mathcal{C}[\mathbf{w}]$ is the closure operation defined by $\mathcal{C}[\mathbf{w}] = [w_1/\sum w_j, w_2/\sum w_j, \ldots, w_D/\sum w_j] = \mathbf{x} \in \mathcal{S}^D$.

The *perturbation operation*, $\mathbf{x} \oplus \mathbf{y} = \mathcal{C}[x_1y_1, x_2y_2, \ldots, x_Dy_D]$, defined on $\mathcal{S}^D \times \mathcal{S}^D$, and the *power transformation*, $\alpha \odot \mathbf{x} = \mathcal{C}[x_1^\alpha, x_2^\alpha, \ldots, x_D^\alpha]$, defined on $\mathbb{R} \times \mathcal{S}^D$, induce a vector space structure on the simplex $\mathcal{S}^D$ (Pawlowsky-Glahn and Egozcue, 2001). Another important element is the *logcontrast*, a log-linear combination

$$\sum_{i=1}^{D} \beta_i \ln x_i, \quad \text{with} \quad \sum_{i=1}^{D} \beta_i = 0, \quad \beta_i \in \mathbb{R} \tag{1}$$

which plays the typical role of the linear combination of variables (Aitchison, 1986).

Once we have a vector space structure, a metric structure is easily defined using the clr scores of a composition $\mathbf{x}$ (Aitchison, 1986):

$$\mathrm{clr}(\mathbf{x}) = \left(\ln \frac{x_1}{g(\mathbf{x})}, \ldots, \ln \frac{x_D}{g(\mathbf{x})} \right),$$

where $g(\cdot)$ means the geometric mean. Note that the $clr(\mathbf{x})_k$ score, being a logcontrast that involves all the parts of a composition, provides information about the relative importance of part x_k in the composition. The basic metric elements of the Aitchison geometry as inner product ($< \cdot, \cdot >_{\mathcal{A}}$), norm ($|| \cdot ||_{\mathcal{A}}$), and distance ($d_{\mathcal{A}}(\cdot, \cdot)$) can be defined as:

$$< \mathbf{x}, \mathbf{y} >_{\mathcal{A}} = < \mathrm{clr}(\mathbf{x}), \mathrm{clr}(\mathbf{y}) >_E \ , \quad ||\mathbf{x}||_{\mathcal{A}}^2 = < \mathbf{x}, \mathbf{x} >_{\mathcal{A}} \ , \quad d_{\mathcal{A}}(\mathbf{x}, \mathbf{y}) = ||\mathbf{x} \ominus \mathbf{y}||_{\mathcal{A}},$$

where "$\mathcal{A}$" means the Aitchison geometry, "E" means the typical Euclidean geometry, and "$\ominus$" is the perturbation difference $\mathbf{x} \ominus \mathbf{y} = \mathbf{x} \oplus ((-1) \odot \mathbf{y})$.

The metric elements are used to construct orthonormal basis and to calculate the corresponding log-ratio coordinates of a composition (olr($\mathbf{x}$)) (Egozcue et al., 2003; Martín-Fernández, 2019). The expression of these olr-coordinates depends on the selected basis. For example, following Egozcue and Pawlowsky-Glahn (2005) one can define particular olr-coordinates created through a sequential binary partition (SBP) of a complete composition $\mathbf{x} = (x_1, \ldots, x_D)$. In the first step of an SBP, when the first olr-coordinate is created, the complete composition $\mathbf{x} = (x_1, \ldots, x_D)$ is split into two groups of parts: one for the numerator and the other for the denominator. In the following steps, to create the following olr-coordinates, each group is in turn split into two groups. That is, in step k when the olr($\mathbf{x}$)$_k$-coordinate is created, the r_k parts $(x_{n1_k}, \ldots, x_{nr_k})$ in the first group are placed in the numerator; the s parts $(x_{d1_k}, \ldots, x_{ds_k})$ in the second group will appear in the denominator; and the rest of $D-(r_k+s_k)$ parts are not involved in the logratio. As a result, the $\text{olr}(\mathbf{x})_k$ is

$$\text{olr}(\mathbf{x})_k = \sqrt{\frac{r_k \cdot s_k}{r_k+s_k}} \ln \frac{(x_{n1_k} \cdots x_{nr_k})^{1/r_k}}{(x_{d1} \cdots x_{ds_k})^{1/s_k}},\ k=1,\dots,D-1. \tag{2}$$

where $\sqrt{\frac{r_k \cdot s_k}{r_k+s_k}}$ is the factor for normalising the coordinate. Note that $r_1+s_1=D$ for $k=1$. The $\text{olr}(\mathbf{x})_k$ coordinate, being a logcontrast that involves two groups of parts of a composition, informs, on average, about the relative importance of one group of parts with regard to the other.

3. Convexity on the simplex

Following the basic definitions of convexity on $\mathbb{R}^D$ (Boyd and Vandenberghe, 2004), the counterpart definitions of convexity on the simplex $\mathcal{S}^D$ in a consistent manner with the Aitchison geometry (i.e. $\mathcal{A}$-convexity) are:

Definition 1. Let $\mathbf{x}_1,\ \mathbf{x}_2$ be two D-part compositions. The $\mathcal{A}$-segment $\overline{\mathbf{x}_1\mathbf{x}_2}$ is the set

$$\overline{\mathbf{x}_1\mathbf{x}_2} = \{\mathbf{y} \in \mathcal{S}^D | \mathbf{y} = \lambda \odot \mathbf{x}_1 \oplus (1-\lambda) \odot \mathbf{x}_2,\ \lambda \in [0,1]\}. \tag{3}$$

An $\mathcal{A}$-segment can be expressed in olr-coordinates as $\text{olr}(\overline{\mathbf{x}_1\mathbf{x}_2}) = \{\mathbf{z} \in \mathbb{R}^{(D-1)} | \mathbf{z} = \lambda \cdot \text{olr}(\mathbf{x}_1) + (1-\lambda) \cdot \text{olr}(\mathbf{x}_2),\ \lambda \in [0,1]\}$. That is the typical expression of a segment of a line on the real space. The definition of a compositional segment (Eq. (3)) can be used in the definition of a compositional convex set ($\mathcal{A}$-convex set).

Definition 2. A set $\mathcal{B} \subseteq \mathcal{S}^D$ is an $\mathcal{A}$-convex set if for all $\mathbf{x}_1,\ \mathbf{x}_2 \in \mathcal{B}$, the compositional segment $\overline{\mathbf{x}_1\mathbf{x}_2}$ is contained in $\mathcal{B}$. That is, for any $\mathbf{x}_1,\ \mathbf{x}_2 \in \mathcal{B}$ and any $\lambda \in [0,1]$, it holds that $\lambda \odot \mathbf{x}_1 \oplus (1-\lambda) \odot \mathbf{x}_2 \in \mathcal{B}$.

To illustrate the definition of an $\mathcal{A}$-convex set, 3-part compositions were selected in $\mathcal{S}^3$ for creating a compositional triangle using the Definition (1) of an $\mathcal{A}$-segment (Fig.1(a)). By construction, a compositional triangle is an $\mathcal{A}$-convex set. On the upper right, Fig.1(b) shows a typical strip (blue area). A compositional segment (red line) not entirely contained in the set shows the lack of the $\mathcal{A}$-convexity of the strip. Figure 1(b) shows how sets that look like convex sets in the simplex from a typical Euclidean point of view, are not compositional convex sets with the Aitchison geometry, and vice versa (Fig.1(a)). Figures 1(c) and (d), respectively, show the representation of Figures 1 (a) and (b) in olr-coordinates. In these figures, one recognises the typical form of a triangle and the shape of a non-convex set on the real space.

3.1. Convex hull on the simplex

The simplex endowed with the induced Euclidean geometry does not have the same structure and properties as the simplex with the Aitchison geometry. This difference

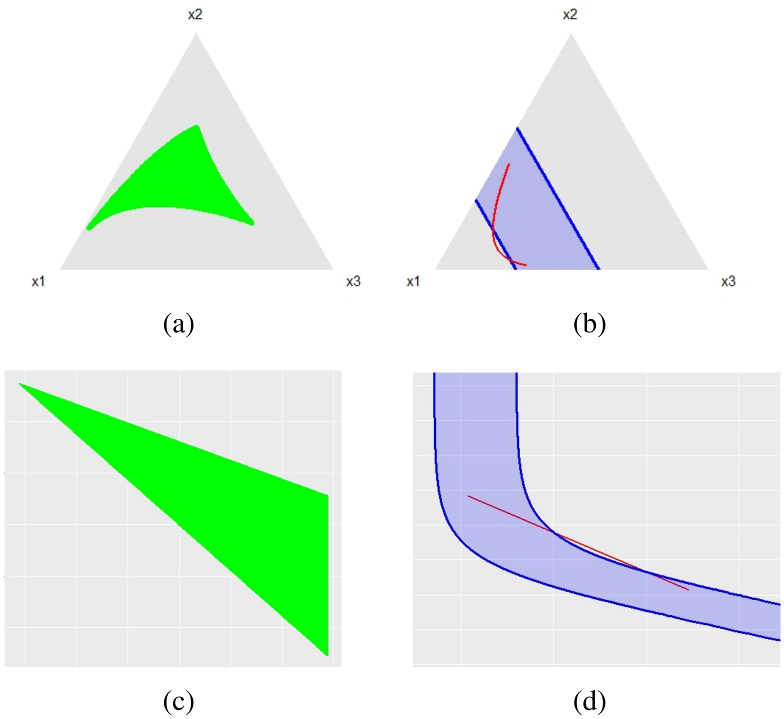


Figure 1. *Triangle and strip in $\mathbb{S}^3$: (a) $\mathcal{A}$-triangle (green area); (b) Strip (blue area) with an $\mathcal{A}$-segment (red line); (c) The $\mathcal{A}$-triangle (green area) in olr-coordinates ; and (d) The strip (blue area) with the $\mathcal{A}$-segment (red line) in olr-coordinates.*

between geometries has implications on the statistical techniques such as the peeling, which is a descriptive statistical technique based on the concept of convex hull (Caussinus, Ettinger and Tomassone, 2012; Small, 1990). Peeling can be described as an iterative algorithm that consists of removing layers of points. Each layer is formed by the points which form the border of the convex hull of the set of remaining points. The convex hull of a set of points $\{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$ in $\mathbb{R}^D$ is the set

$$conv_E(\mathbf{X}) = \{\mathbf{z}\in\mathbb{R}^D | \mathbf{z} = \lambda_1\mathbf{x}_1 + \ldots + \lambda_n\mathbf{x}_n, \sum_{i=1}^{n}\lambda_i = 1,\ \lambda_i \geq 0, i = 1,\ldots,n\},$$

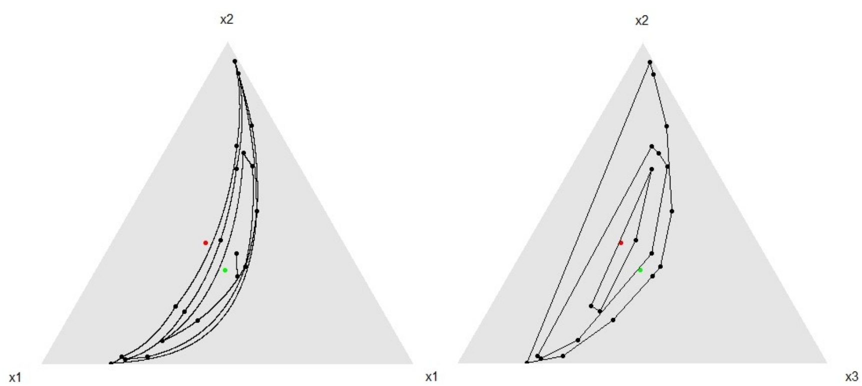
This definition is used by Tolosana-Delgado, von Eynatten and Karius (2011) for CoDa vectors. Among other applications, peeling allows us to graphically represent the centre of a set of points in the last *internal* layer and can be used for outlier detection (Harsh, Ball and Wei, 2016, chapter 4).

We propose the corresponding definition of the convex hull on the simplex in terms of Aitchison geometry:

Definition 3. The $\mathcal{A}$-convex hull of the set of compositions $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \in \mathcal{S}^D$ is

$$conv_{\mathcal{A}}(\mathbf{X}) = \{\mathbf{y} \in \mathcal{S}^D | \mathbf{y} = \lambda_1 \odot \mathbf{x}_1 \oplus \ldots \oplus \lambda_n \odot \mathbf{x}_n, \sum_{i=1}^{n} \lambda_i = 1,\ \lambda_i \geq 0, i = 1, \ldots, n\}.$$

Note that the compositional triangle created in Fig. 1(a) is the most simple example of an $\mathcal{A}$-convex hull using the minimum number of compositions. The usefulness of this concept can be illustrated by means of a more complex example in $\mathcal{S}^3$. Let $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_{20}\} \in \mathcal{S}^3$ be a set of compositions randomly generated using a normal distribution on the simplex (Mateu-Figueras, Pawlowsky-Glahn and Egozcue, 2013). Figure 2 compares the $\mathcal{A}$-convex hull $conv_{\mathcal{A}}(\mathbf{X})$ (on the left) to the Euclidean convex hull $conv_E(\mathbf{X})$ (on the right). Figure 2(a) shows the successive layers created when peeling is applied to $\mathbf{X}$ using the $\mathcal{A}$-convex hull. The last internal layer (i.e., the smallest convex hull) is close to the compositional centre of $\mathbf{X}$, $g(\mathbf{X}) = \frac{1}{20} \odot (\oplus_{i=1}^{20} \mathbf{x}_i)$ (green dot), the column-wise geometric mean (e.g., Pawlowsky-Glahn, Egozcue and Tolosana-Delgado, 2015). On the other hand, the typical Euclidean centre of $\mathbf{X}$, the arithmetical mean $\overline{\mathbf{X}} = \frac{1}{20} \sum_{i=1}^{20} \mathbf{x}_i$ (red dot) is far from the last *external* layer of the peeling, suggesting a potential outlier. Fig. 2(b) shows the layers of the peeling using the E-convex hull. In this case, the Euclidean centre of $\mathbf{X}$ (red dot) is inside the last layer, whereas the compositional centre $g(\mathbf{X})$ (green dot) is far from it. In addition, when comparing the first external layers (i.e., the largest convex hull) in both geometries, one concludes that the $\mathcal{A}$-convex hull fits better with the common arch shape of a normally distributed CoDa set.



(a) (b)

Figure 2. *Peeling applied to a CoDa set $\mathbf{X} \in \mathcal{S}^3$: (a) with $\mathcal{A}$-convex hull; (b) with E-convex hull. Geometric centre $g(\mathbf{X})$ (green dot) and arithmetic centre $\overline{\mathbf{X}}$ (red dot) of CoDa set are plotted.*

3.2. Some basic compositional sets

The most common sets in convex optimisation in areas such as, among others, design of experiments (DOE), optimisation with mixtures, or time-use data are defined by constraining some parts of the composition $\mathbf{x}$: $\{0 < l_i \leq x_i \leq u_i < 1;\ i = 1,\ldots,D\}$ (Chen et al., 2010) or by constraining the ratio of two parts: $\{0 < l_{ij} \leq \frac{x_i}{x_j} \leq u_{ij};\ i \neq j = 1,\ldots,D\}$(Lo Huang and Huang, 2009).

3.2.1. Constraining the ratios between parts: a logcontrast

The equation $\{\frac{x_i}{x_j} = k\ ,1 \leq i \neq j \leq D;\ k > 0\}$ is equivalent to the logcontrast $\{\ln x_i - \ln x_j = \ln k;\ 1 \leq i \neq j \leq D\}$. This equation describes an affine subspace of dimension $D-2$ on the simplex with the Aitchison geometry (Egozcue, Pawlowsky-Glahn and Gloor, 2018). Consequently, $l \leq \frac{x_i}{x_j}$ or $\frac{x_i}{x_j} \leq u$ define closed half-spaces of the simplex, and both sets, $\Pi^+ = \{\mathbf{x} \in \mathcal{S}^D | l \leq \frac{x_i}{x_j}\}$ and $\Pi^- = \{\mathbf{x} \in \mathcal{S}^D | \frac{x_i}{x_j} \leq u\}$, verify the condition of being an $\mathcal{A}$-convex set (Definition 2).

In general, affine subspaces such as $\mathcal{A}$-lines, $\mathcal{A}$-planes or $\mathcal{A}$-hyperplanes are defined by logcontrasts (Eq. (1)) (Egozcue et al., 2018). A logcontrast splits the simplex into two closed half-spaces, $\Pi^+ = \{\mathbf{x} \in \mathcal{S}^D | \sum_{j=1}^{D} \beta_j \ln x_j \geq k\}$ and $\Pi^- = \{\mathbf{x} \in \mathcal{S}^D | \sum_{j=1}^{D} \beta_j \ln x_j \leq k\}$. By construction, both half-spaces are $\mathcal{A}$-convex sets (Definition 2). Figure 3 shows four different logcontrasts (blue lines) whose intersection determines a quadrilateral (green area). Because the intersection of convex sets is a convex set (Boyd and Vandenberghe, 2004) it follows that the quadrilateral is an $\mathcal{A}$-convex set.

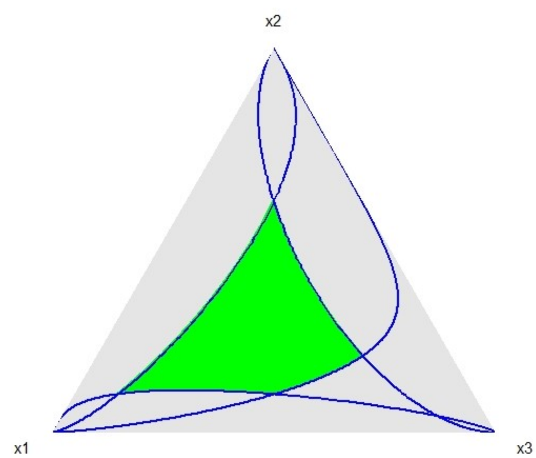


Figure 3. *An $\mathcal{A}$-convex quadrilateral in $\mathcal{S}^3$ (green area) determined by the intersection of four half-spaces defined by logcontrasts (blue lines).*

3.2.2. Constraining the parts of a composition

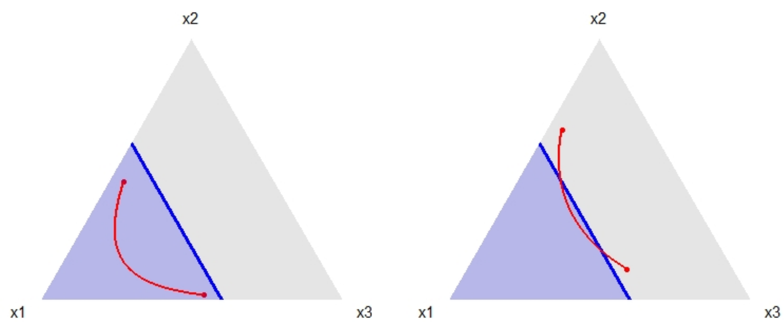
Despite the fact that an equation such as $\{x_i = k,\ k \in (0,1)\}$, for any $i = 1,\ldots,D$, cannot be expressed in terms of a logcontrast, this type of equation also splits the simplex into two sets: the upper set $\Sigma^+ = \{\mathbf{x} \in \mathcal{S}^D | x_i \geq k,\ k \in (0,1)\}$, and the lower set $\Sigma^- = \{\mathbf{x} \in$

$\mathcal{S}^D | x_i \leq k,\ k \in (0,1)\}$. However, in this case, the two sets are different with regard to their compositional convexity.

Proposition 1. *For any $i = 1,\ldots,D$, the set $\Sigma^+ = \{\mathbf{x} \in \mathcal{S}^D | x_i \geq k,\ k \in (0,1)\}$ is an $\mathcal{A}$-convex set.*

Proof. See Appendix. ■

In contrast, the lower set Σ^- is not an $\mathcal{A}$-convex set as is illustrated in Fig. 4(b). Figure 4 shows an example in S^3 of the type Σ^+ (blue area) and its complementary set, Σ^- (grey area), for part x_1 with $k = 0.4$. For each set, two compositions were selected and the corresponding $\mathcal{A}$-segment plotted (red line). Because the red $\mathcal{A}$-segment on Fig. 4(b) is not entirely contained in the set Σ^- (grey area), this set is not an $\mathcal{A}$-convex set.



(a) (b)

Figure 4. *A 3-part composition constrained for $i = 1$ and $k = 0.4$: (a) The set Σ^+ (blue area); (b) The set Σ^- (grey area). The red line represents the $\mathcal{A}$-segment for two compositions in each set.*

The sets Σ^+ and Σ^- can be generalised as follows:

$$\Sigma_i^+(\boldsymbol{\alpha}) = \{\mathbf{x} \in \mathcal{S}^D | \alpha_1 x_1 + \ldots + \alpha_D x_D \geq 0, \alpha_i > 0, \alpha_j \leq 0,\ 1 \leq j \neq i \leq D\}$$

$$\Sigma_i^-(\boldsymbol{\alpha}) = \{\mathbf{x} \in \mathcal{S}^D | \alpha_1 x_1 + \ldots + \alpha_D x_D \leq 0, \alpha_i > 0, \alpha_j \leq 0,\ 1 \leq j \neq i \leq D\}$$

Note that $\Sigma_i^+(\boldsymbol{\alpha})$ and $\Sigma_i^-(\boldsymbol{\alpha})$ for $\boldsymbol{\alpha} = (-k,\ldots,-k,\underbrace{(1-k)}_{i},-k,\ldots,-k),\ k \in (0,1)$, become the sets Σ^+ and Σ^-, respectively. Importantly, an analogous proof to Proposition 1 states that a set $\Sigma_i^+(\boldsymbol{\alpha})$ is $\mathcal{A}$-convex. On the other hand, $\Sigma_i^-(\boldsymbol{\alpha})$ is not an $\mathcal{A}$-convex set. To illustrate these properties, Figure 5 shows the set $\Sigma_1^+(\boldsymbol{\alpha}) = \{\mathbf{x} \in \mathcal{S}^4 | x_1 - \frac{2}{3}x_2 - \frac{2}{3}x_3 - \frac{1}{3}x_4 \geq 0\}$ (blue area) in $\mathcal{S}^4$, which generalises a set of type $\Sigma^+ = \{\mathbf{x} \in \mathcal{S}^D | x_1 \geq k,\ k \in (0,1)\}$.

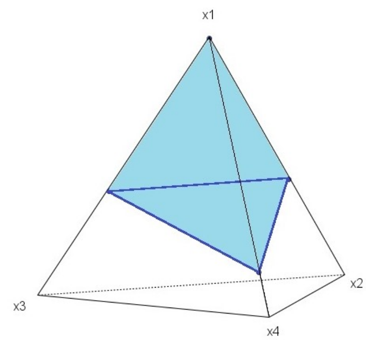


Figure 5. *The $\mathcal{A}$-convex set* $\Sigma_1^+(\alpha) = \{\mathbf{x} \in \mathcal{S}^4 | x_1 - \frac{2}{3}x_2 - \frac{2}{3}x_3 - \frac{1}{3}x_4 \geq 0\}$ *(blue area) in* $\mathcal{S}^4$.

Note that the intersection of sets of type Σ^+ or $\Sigma^+(\alpha)$ is an $\mathcal{A}$-convex set. However, when an optimisation problem includes a set of type $\Sigma^-(\alpha)$ the feasible region might be a non $\mathcal{A}$-convex set. In such a case, one should be cautious when applying convex optimisation techniques.

4. Convex functions on the simplex

The definition of a convex function in the Euclidean space can be adapted to the Aitchison geometry following the schema introduced in Luenberger and Ye (2008) and in Boyd and Vandenberghe (2004).

Definition 4. Let $W \subset \mathcal{S}^D$ be an $\mathcal{A}$-convex set. A function $f : W \to \mathbb{R}$ is an $\mathcal{A}$-convex function if for all $\mathbf{x}_1,\ \mathbf{x}_2 \in W$ and $\lambda \in [0,\ 1]$:

$$f((1-\lambda)\odot\mathbf{x}_1 \oplus \lambda\odot\mathbf{x}_2) \leq (1-\lambda)f(\mathbf{x}_1) + \lambda f(\mathbf{x}_2)$$

Definition 5. Let $W \subset \mathcal{S}^D$ be an $\mathcal{A}$-convex set. A function $g : W \to \mathbb{R}$ is an $\mathcal{A}$-concave function if $f = -g$ is an $\mathcal{A}$-convex function.

Note that, as expected, a constant function $f(\mathbf{x}) = k$, with k a real number, is simultaneously an $\mathcal{A}$-convex and $\mathcal{A}$-concave function.

Importantly, the classification of convex function through the gradient or the Hessian matrix also applies for CoDa. That is, the common rule *A twice differentiable function of several variables is convex on a convex set if and only if its Hessian matrix of second partial derivatives is positive semidefinite on the interior of the convex set* can be used to classify $\mathcal{A}$-convex functions by means of the basic concepts of compositional differential calculus (Barceló-Vidal, Martín-Fernández and Mateu-Figueras, 2011) working in olr-coordinates. Whereas, using the $\mathcal{A}$-gradient, one has to check the usual expression

$$f(\mathbf{x}) \geq f(\mathbf{y}) + \nabla_{\mathcal{A}}(f)(\mathbf{y})(\ln(\mathbf{x}) - \ln(\mathbf{y}))$$

for all $\mathbf{x},\ \mathbf{y} \in W$, where $\nabla_{\mathcal{A}}(f)$ is the $\mathcal{A}$-gradient vector in clr-coordinates Barceló-Vidal et al. (2011).

The following two propositions show that the sum of functions and product by a scalar are basic operations for creating more complex $\mathcal{A}$-convex functions.

Proposition 2. *Let f_1 and f_2 be two $\mathcal{A}$-convex functions on the $\mathcal{A}$-convex set $W \subset \mathcal{S}^D$. The function $f_1 + f_2$ is $\mathcal{A}$-convex on W.*

Proof. See Appendix. ■

Proposition 3. *Let f be an $\mathcal{A}$-convex function on the $\mathcal{A}$-convex set $W \subset \mathcal{S}^D$. For any $a \geq 0$, the function af is $\mathcal{A}$-convex on W*

Proof. This proof is immediate from the definition of convex function. ■

Using the two previous propositions, it follows that a positive linear combination of $\mathcal{A}$-convex functions is an $\mathcal{A}$-convex function. That is, given a set of $\mathcal{A}$-convex functions $f_1, ..., f_n$ on the $\mathcal{A}$-convex set $W \subset \mathcal{S}^D$, then the function $a_1 f_1 + \ldots + a_n f_n$ is an $\mathcal{A}$-convex function for any $a_j > 0,\ j = 1, \ldots, n$. This property is useful for creating more complex $\mathcal{A}$-convex functions using basic functions (see Section 4.1).

In addition, the following results state that the sublevel sets of an $\mathcal{A}$-convex function on the simplex verify the usual properties as regard to the convex sets.

Proposition 4. *Let f be an $\mathcal{A}$-convex function on the $\mathcal{A}$-convex set $W \subset \mathcal{S}^D$. The sublevel set $\mathcal{G}_\alpha^- = \{\mathbf{x} |\ \mathbf{x} \in W, f(\mathbf{x}) \leq \alpha\}$ is $\mathcal{A}$-convex for any real number α.*

Proof. See Appendix. ■

Even though for any $\mathcal{A}$-convex function, its sublevel sets $\mathcal{G}_\alpha^-$ are $\mathcal{A}$-convex sets, the converse is not true. This fact motivates the following definitions.

Definition 6. A function $f : \mathcal{S}^D \to \mathbb{R}$ is an $\mathcal{A}$-quasiconvex function if its domain and all its sublevel sets, $\mathcal{G}_\alpha^- = \{\mathbf{x} |\ \mathbf{x} \in dom(f), f(\mathbf{x}) \leq \alpha\}$ are $\mathcal{A}$-convex sets for any real number α.

Definition 7. A function $f : \mathcal{S}^D \to \mathbb{R}$ is an $\mathcal{A}$-quasiconcave function if its domain and all its superlevel sets, $\mathcal{G}_\alpha^+ = \{\mathbf{x} |\ \mathbf{x} \in dom(f), f(\mathbf{x}) \geq \alpha\}$ are $\mathcal{A}$-convex sets for any real number α.

As with convex optimisation problems in the Euclidean space when defining the feasible region (Boyd and Vandenberghe, 2004), it is recommended to represent the sublevels of an $\mathcal{A}$-quasiconvex function (or the superlevels of a $\mathcal{A}$-quasiconcave function) through inequalities of $\mathcal{A}$-convex functions. Therefore, an $\mathcal{A}$-quasiconvex function f, should be expressed by means of $\mathcal{A}$-convex functions, Φ_α such that,

$$f(\mathbf{x}) \leq \alpha \iff \Phi_\alpha(\mathbf{x}) \leq 0.$$

4.1. Some basic functions on the simplex

With the following examples, the $\mathcal{A}$-convexity of some popular functions on $\mathcal{S}^D$ is reviewed.

Example 1. The function $f(\mathbf{x}) = \frac{x_i}{x_j}$, $1 \leq i, j \leq D$ is an $\mathcal{A}$-convex function over its domain, $dom(f) = \mathcal{S}^D$.

Proof. See Appendix. ■

Example 2. For any $i = 1, \ldots, D$, the function $f(\mathbf{x}) = x_i$ is an $\mathcal{A}$-quasiconcave function over its domain, $dom(f) = \mathcal{S}^D$. Moreover, using the $\mathcal{A}$-convex function

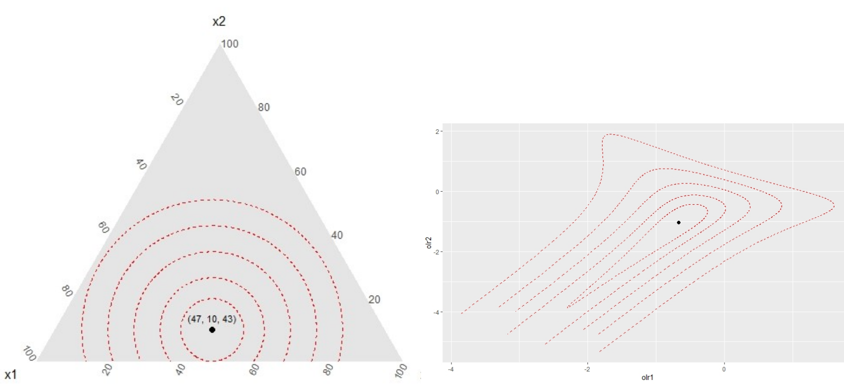
$$\Phi_\alpha(\mathbf{x}) = \alpha \sum_{j=1}^{D} \frac{x_j}{x_i} - 1 \,,$$

a superlevel $\mathcal{G}_\alpha^+ = \{\mathbf{x} \in \mathcal{S}^D |\, x_i \geq \alpha\}$ for any $\alpha \in (0,\ 1)$ can be represented by means of $\Phi_\alpha(\mathbf{x}) \leq 0$.

Proof. See Appendix. ■

Example 3. For any $\mathbf{x}_0 \in \mathcal{S}^D$, the function squared Euclidean distance $f(\mathbf{x}) = d_E^2(\mathbf{x}, \mathbf{x}_0) = \sum_{j=1}^{D} (x_j - x_{0j})^2$ is not $\mathcal{A}$-convex.

Figure 6(a) shows the contour lines of the function $f(\mathbf{x}) = d_E^2(\mathbf{x}, \mathbf{x}_0)$ on $\mathcal{S}^3$ for $\mathbf{x}_0 = (47, 10, 43)$. The sublevel sets are not $\mathcal{A}$-convex sets, and therefore, the function $f(\mathbf{x})$ is not an $\mathcal{A}$-convex function (Proposition 4). Because it may be difficult to see the lack of convexity of a set on the ternary diagram, Figure 6(b) shows the sublevel sets in olr-coordinates.



(a) (b)

Figure 6. *Contour lines of the function* $f(\mathbf{x}) = d_E^2(\mathbf{x}, \mathbf{x}_0)$ *on* $\mathcal{S}^3$ *for* $\mathbf{x}_0 = (47, 10, 43)$*: (a) Ternary diagram; (b) olr-coordinates.*

Note that the function Euclidean distance and the function (squared) $d_E^2(\mathbf{x},\mathbf{x}_0)$ share the same contour lines. Consequently, the function (non-squared) $d_E(\mathbf{x},\mathbf{x}_0)$ is not $\mathcal{A}$-convex. The lack of convexity means that the Euclidean distance does not satisfy the triangular inequality, that is, it is not a distance function on the simplex endowed with the Aitchison geometry.

Example 4. For any $\mathbf{x}_0 \in \mathcal{S}^D$, the function squared Aitchison distance $f(\mathbf{x}) = d_{\mathcal{A}}^2(\mathbf{x},\mathbf{x}_0) = d_E^2(\text{clr}(\mathbf{x}),\text{clr}(\mathbf{x}_0))$ is $\mathcal{A}$-convex.

Proof. The proof is immediate because the function squared Euclidean distance is convex on the Real space. ■

4.2. Convex optimisation on the simplex

In a broad sense, a convex optimisation problem in $\mathbb{R}^D$ is defined as (Boyd and Vandenberghe, 2004):

$$\begin{aligned} &\textit{minimise} \quad f_0(\mathbf{x}) \\ &\textit{subject to} \;\; f_j(\mathbf{x}) \le 0 \;\; j = 1,\dots,m \\ &\qquad\qquad\quad g_k(\mathbf{x}) = b_k \; k = 1,\dots,n \end{aligned} \tag{4}$$

where $\mathbf{x} \in \mathbb{R}^D$, $f_0,\dots,f_m : \mathbb{R}^D \to \mathbb{R}$ are convex functions and $g_1,\dots,g_n : \mathbb{R}^D \to \mathbb{R}$ are linear functions.

For CoDa, the above definition is adapted as:

Definition 8. An $\mathcal{A}$-convex optimisation problem in standard form is defined as

$$\begin{array}{lll} \text{minimise} & f_0(\mathbf{x}) & \\ \text{subject to} & f_j(\mathbf{x}) \le 0 & j = 1,\dots,m \\ & \boldsymbol{\beta}_k^\intercal \ln \mathbf{x} = b_k & k = 1,\dots,n \end{array}$$

where $\mathbf{x} \in \mathcal{S}^D$, $f_0,\ f_1,\dots,\ f_m$ are $\mathcal{A}$-convex functions and $\boldsymbol{\beta}_k^\intercal \ln \mathbf{x}$ are logcontrasts, that is, $\sum_{j=1}^D \beta_{k,j} = 0,\ k = 1,\dots,n$.

An important case of convex optimisation problems is that of *linear programming*, that is, when the objective and all constraining functions are linear. For CoDa, an $\mathcal{A}$-linear programming problem is defined in terms of logcontrasts as follows:

$$\begin{array}{lll} \text{minimise} & \boldsymbol{\beta}_0^\intercal \ln \mathbf{x} & \\ \text{subject to} & \boldsymbol{\beta}_k^\intercal \ln \mathbf{x} \le b_k & k = 1,\dots,m \\ & \boldsymbol{\beta}_k^\intercal \ln \mathbf{x} = b_k & k = m+1,\dots,n \end{array}$$

where $\sum_{j=1}^D \beta_{k,j} = 0,\ k = 0,\dots,n$.

5. Case Study

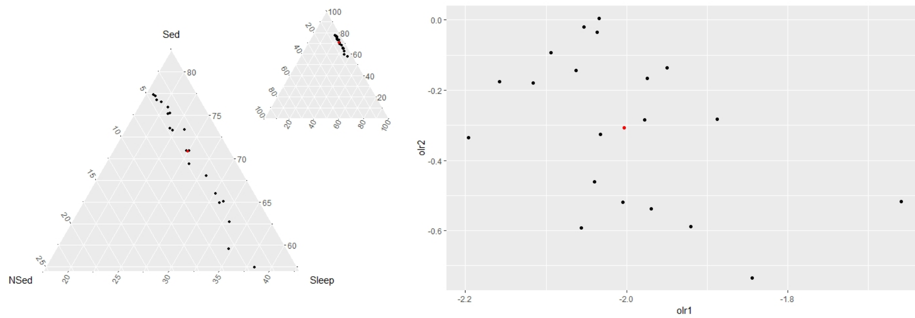
The example we present here is based on data from Aitchison (1986) and only serves for illustrative purposes. We consider a 3-part time-use composition of mutually exclusive and exhaustive parts: non-sedentary time (NSed), sedentary time (Sed), and sleeping time (Sleep). Table 1 shows the 3-part time-use compositions of a university associate professor with unhealthy physical activity habits. Figure 7 shows the data set in the ternary diagram (Fig. 7a) and in the olr-space (Fig. 7b), where the olr-basis used is $\mathrm{olr}_1(\mathbf{x}) = \frac{\sqrt{2}}{2}\ln\frac{NSed}{Sed}$ and $\mathrm{olr}_2(\mathbf{x}) = \sqrt{\frac{2}{3}}\ln\frac{\sqrt{NSed \cdot Sed}}{Sleep}$.

Table 1. *3-part time-use composition over 20 days.*

	Non-sedentary	Sedentary	Sleep
D1	0.04234	0.77218	0.18547
D2	0.03772	0.75235	0.20993
D3	0.04807	0.69388	0.25805
D4	0.05705	0.59596	0.34699
D5	0.04306	0.76733	0.18961
D6	0.03592	0.75916	0.20493
D7	0.03797	0.67973	0.28231
D8	0.03959	0.76519	0.19522
D9	0.04321	0.70868	0.24811
D10	0.04000	0.70886	0.25114
D11	0.04060	0.75101	0.20838
D12	0.04148	0.62683	0.33169
D13	0.04003	0.64864	0.31133
D14	0.04357	0.77365	0.18277
D15	0.04488	0.73273	0.22239
D16	0.04665	0.73483	0.21853
D17	0.03873	0.65937	0.30190
D18	0.03282	0.73313	0.23405
D19	0.03552	0.65058	0.31390
D20	0.04231	0.57445	0.38324

The centre or mean ($\mathbf{x}_0$) of a CoDa set is the vector of geometric means of its parts, scaled to sum 1 in order to obtain its representative on the unit simplex. Therefore, on dayly average, the associate professor engages in physical activity for one hour, exhibits sedentary behaviour for seventeen hours, and sleeps for six hours, $\mathbf{x}_0 = (1/24, 17/24, 6/24)$ (see the red point in Figure 7).

The location and spread of a compositional data set are summarised in the variation array (Table 2) through pairwise logratios of parts. The elements above the first diagonal are the pairwise log-ratio variances, whereas the elements below it are the arithmetic



(a) (b)

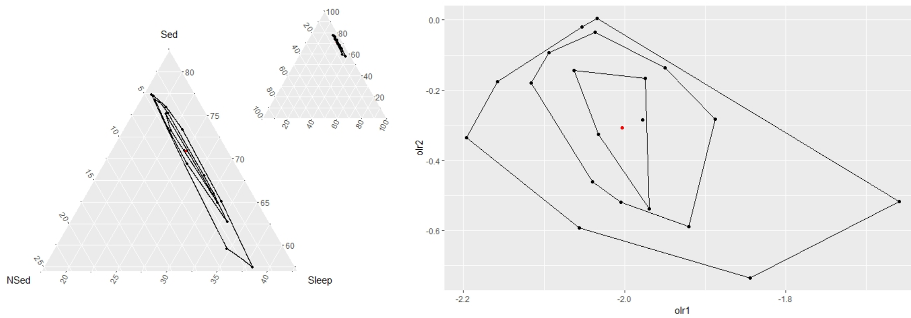
Figure 7. *(a) 3-part time-use compositions on the ternary diagram (small triangle). The large triangle zooms in on the data set region. In red, the centre of the data set* $\mathbf{x}_0 = (\frac{1}{24}, \frac{17}{24}, \frac{6}{24})$*; (b) the data set and its centre in the olr-space.*

means. As suggested by the ternary diagram (Fig. 7a), the largest log-ratio variance corresponds to $\{Sed, Sleep\}$, whereas the smallest value is $Var(\ln \frac{NSed}{Sed}) = 0.0265$. In this case, because the estimate of the log-ratio expectation is $E(\ln \frac{Sed}{NSed}) = 2.8332$, on average, sedentary time is approximately 17 times ($\approx \exp\{2.8332\}$) the non-sedentary time, as the centre $\mathbf{x}_0$ of the data set indicates.

Table 2. *Variation array of the 3-part time-use compositional data.*

	Pairwise log-ratio variance		
	NSed	Sed	Sleep
NSed		0.0265	0.0561
Sed	2.8332		0.0949
Sleep	1.7918	-1.0415	
	Pairwise log-ratio arithmetical mean		

The data set shown in the ternary diagram (Fig. 7a) suggests that the farthest point from the centre is the point located further down on the simplex: $D20 = (0.04231, 0.57445, 0.38324)$, with the smallest value in the part $NSed$ (Table 1). At first glance, this point may be considered a potential outlier. Moreover, when representing the data set in olr-coordinates (Fig. 7b), the outermost point is the point located further to the right: $\text{olr}(D4) = (-1.659, -0.516)$. This fact is corroborated by the peeling using the $\mathcal{A}$-convex hull (3). Figure 8b suggests that each layer has an elliptical shape, the typical shape of a normal probability distribution. A multivariate Anderson-Darling test confirms that one can fail to reject normality ($A^2 = 0.7760$; $p - value = 0.2199$). Indeed, under this assumption, the χ^2 atipicality index indicates that the sample $D4$ is a potential outlier ($\chi^2 percentile = 98.75\%$), whereas the sample $D20$ is not ($\chi^2 percentile = 86.87\%$).



(a) (b)

Figure 8. *Peeling with the Aitchison geometry of the 3-part time-use compositions: (a) on the ternary diagram (small triangle). The large triangle zooms into the data set region. The centre of the data set is represented in red colour; (b) in the olr-space.*

The professor was recommended to increase non-sedentary time up to, on average, at least 13 hours per day. The question is how to distribute the time for the rest of the activities (i.e., sedentary and sleeping). One criterion may be to move from the centre of the data set ($\mathbf{x}_0$) to the closest point in the $\mathcal{A}$-convex set $\Sigma^+ = \{\mathbf{x} \in \mathcal{S}^D | x_1 \geq \frac{13}{24}\}$, whereas the spread is preserved. This is one of the simplest examples of convex optimisation problem: to find the minimum distance from a given point $\mathbf{x}_0 \in \mathcal{S}^D$ to a convex set Σ^+.

From a Euclidean approach, the E-convex optimisation problem is:

$$\begin{array}{ll} \text{minimise} & d_E^2(\mathbf{x}_0, \mathbf{x}) = \sum_{j=1}^{3} (x_j - x_{0j})^2 \\ \text{subject to} & x_1 \geq 13/24 \\ & x_1 + x_2 + x_3 = 1 \\ & x_j \geq 0, \ j = 1, \ldots, 3 \end{array} \tag{5}$$

Figure 9(a) shows that the solution of the E-convex optimisation problem (Eq. 5) is $\mathbf{x} = (\frac{13}{24}, \frac{11}{24}, 0)$, where the proposed movement is $\mathbf{x} - \mathbf{x}_0 = (12/24, -6/24, -6/24)$. That is, in order to increase the fraction of non-sedentary time in $\frac{12}{24}$, the Euclidean approach subtracts from the rest of the parts (non-sedentary and sleeping time) the same amount of time, $\frac{6}{24}$ hours. Note that the sleeping time has to be zero, a solution that is not realistic in practice.

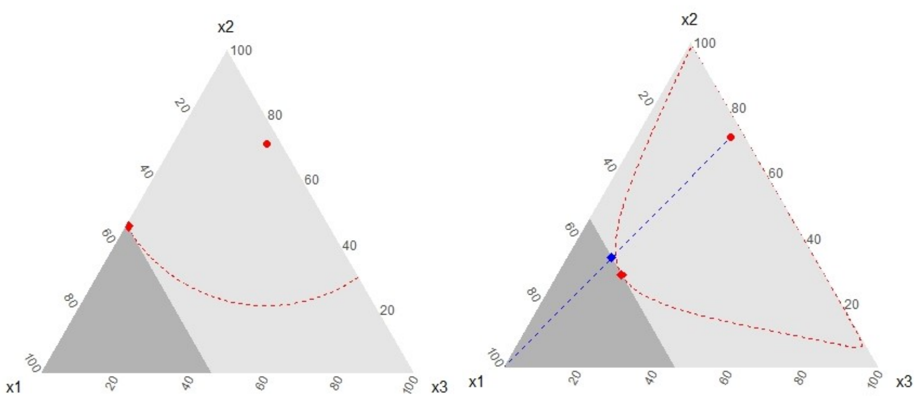
In this situation, one could consider applying an analogous procedure using the Aitchison geometry. That is, perturb by $\frac{13}{1}$ the non-sedentary time $\frac{1}{24}$ to verify that $x_1 = 13/24$, while perturbing the other two parts ($\{Sed, Sleep\}$) by the same factor to preserve its relative information. Following this idea, when the centre $\mathbf{x}_0 = (1/24, 17/24, 6/24)$ is perturbed by the vector $\mathbf{p} = (13/1, 11/23, 11/23)$ the composition obtained is $\mathbf{x} = (13/24, 8.13/24, 2.87/24)$, which verifies the constraint $x_1 \geq 13/24$ (see figure 9 (b)). When one calculates the (squared) Aitchison distance from $\mathbf{x}_0$ to the new centre $\mathbf{x}$ the result is 7.271. To confirm whether this distance is the minimum value one must formulate the convex optimisation problem using the Aitchison geometry:

$$\begin{array}{ll} \text{minimise} & d^2_{\mathcal{A}}(\mathbf{x}_0,\mathbf{x}) = \sum_{j=1}^{3}\left(\ln\frac{x_j}{g(\mathbf{x})} - \ln\frac{x_{0j}}{g(\mathbf{x_0})}\right)^2 \\ \text{subject to} & x_1 \geq \frac{13}{24} \end{array} \tag{6}$$

whose standard form is (Example 2)

$$\begin{array}{ll} \text{minimise} & d^2_{\mathcal{A}}(\mathbf{x}_0,\mathbf{x}) = \sum_{j=1}^{3}\left(\ln\frac{x_j}{g(\mathbf{x})} - \ln\frac{x_{0j}}{g(\mathbf{x_0})}\right)^2 \\ \text{subject to} & \frac{13}{24}\sum_{j=1}^{3}\dfrac{x_j}{x_1} - 1 \leq 0 \end{array}$$

Figure 9(b) shows that the solution to the $\mathcal{A}$-convex optimisation problem (Eq. 6) is $\mathbf{x} = \left(\frac{13}{24}, \frac{6.87}{24}, \frac{4.13}{24}\right)$. Note that, the Aitchison approach has a reasonable behaviour because the largest part of the initial composition (i.e., sedentary time) contributes more to increase the non-sedentary time. The way that the other parts contribute to increase the non-sedentary time is not proportional in any sense. In this case, the (squared) Aitchison distance from $\mathbf{x}_0$ to the new centre $\mathbf{x}$ is 6.989, smaller than the distance obtained when the new centre is created by perturbation. Importantly, the ratio $\frac{Sed}{Sleep}$ is not preserved when moving from $\mathbf{x}_0$ to the solution of the optimisation problem ($\mathbf{x}$). That is, using this approach, the parts $\{Sed,\ Sleep\}$ are not perturbed by the same factor.



(a) (b)

Figure 9. *Time-use optimisation for the composition* $\mathbf{x}_0 = (\frac{1}{24}, \frac{17}{24}, \frac{6}{24})$ *(red dot). The dark grey area is the set* $\Sigma^+ = \{\mathbf{x} \in \mathbb{S}^3 | x_1 \geq \frac{13}{24}\}$*. The red diamond* $\mathbf{x} \in \Sigma^+$ *is the closest point to* $\mathbf{x}_0$*. The dashed red line is the contour line for the corresponding distance. (a) Euclidean geometry:* $\mathbf{x} = (\frac{13}{24}, \frac{11}{24}, \frac{0}{24})$*; (b) Aitchison geometry:* $\mathbf{x} = \left(\frac{13}{24}, \frac{6.87}{24}, \frac{4.13}{24}\right)$*. Dashed blue line is the perturbation direction from* $\mathbf{x}_0$ *to* $\mathbf{x} = (\frac{13}{24}, \frac{8.13}{24}, \frac{2.87}{24}) \in \Sigma^+$ *(blue diamond)*

In the optimisation problem, the movement from $\mathbf{x}_0$ to $\mathbf{x}$ is explained by the perturbation difference $\mathbf{x} \ominus \mathbf{x}_0 = (13/1, 6.87/17, 4.13/6)$. Figure 10 shows how the original data set (grey) is moved to the data set perturbed (black) by $\mathbf{x} \ominus \mathbf{x}_0 = (13/1, 6.87/17, 4.13/6)$, preserving the data spread. The centre of the perturbed data set fulfils the condition $x_1 \geq \frac{13}{24}$. With the Aitchison geometry approach, the solution is more realistic.

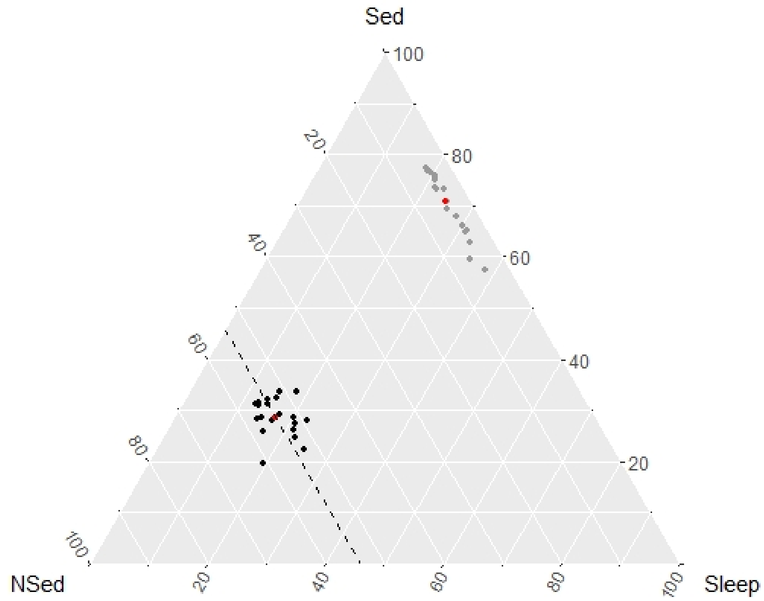


Figure 10. *Resolving the optimisation problem on the ternary diagram: the original data set (grey) is perturbed by* $\mathbf{x} \ominus \mathbf{x}_0 = (13/1, 6.87/17, 4.13/6)$ *for becoming the new data set (black). The dashed segment is the border of the set* $x_1 \geq \frac{13}{24}$. *The centre of each data set is shown in red.*

6. Conclusions and final remarks

The adequate definitions of convex set, convex hull and convex function in a compatible manner with the Aitchison geometry play an important role in the analysis of CoDa. The most popular statistical techniques include optimisation problems that are sensitive to the geometry of the sample space. We have compared the Euclidean geometry to the Aitchison geometry when solving a convex optimisation problem for CoDa. While the Euclidean approach found inconsistent solutions, the Aitchison geometry provided more satisfactory results. This concludes that, despite the fact that any method may give a reasonable solution in some scenarios, it is advisable to use methods that are consistent with the geometry of the sample space.

With basic concepts of constrained convex optimisation for CoDa on hand, a revision of some statistical techniques should be carried out. The pending challenge is to investigate the implications in popular techniques such as, among others, outlier detection, experimental design of mixtures, or Lasso regression.

Appendix: Proofs

Proposition 1 *For any $i = 1,\dots,D$, the set $\Sigma^+ = \{\mathbf{x} \in \mathcal{S}^D | x_i \geq k,\ k \in (0,1)\}$ is an $\mathcal{A}$-convex set.*

Proof. Let $\mathbf{x}_1 = (x_{11},\dots,x_{1D}),\ \mathbf{x}_2 = (x_{21},\dots,x_{2D})$ be two D-part compositions in Σ^+. We have to check that for any $\lambda \in [0,\ 1]$ it holds that

$$(1-\lambda)\odot\mathbf{x}_1 \oplus \lambda\odot\mathbf{x}_2 \in \Sigma^+.$$

Due to $\sum_{j=1}^{D} x_{1j} = \sum_{j=1}^{D} x_{2j} = 1$ then the equations $x_{1i} \geq k$ and $x_{2i} \geq k$ are respectively equivalent to

$$x_{1i} \geq k(x_{11}+\ldots+x_{1D}) \quad \text{and} \quad x_{2i} \geq k(x_{21}+\ldots+x_{2D}).$$

Using this equivalence, $\forall\lambda \in [0,\ 1]$ it holds

$$x_{1i}^{1-\lambda}x_{2i}^{\lambda} \geq k(x_{11}+\ldots+x_{1D})^{1-\lambda}(x_{21}+\ldots+x_{2D})^{\lambda} = k(\sum_{j=1}^{D} x_{1j})^{1-\lambda}(\sum_{j=1}^{D} x_{2j})^{\lambda},$$

where, with Hölder's inequality, the last term of the expression holds

$$k(\sum_{j=1}^{D} x_{1j})^{1-\lambda}(\sum_{j=1}^{D} x_{2j})^{\lambda} \geq k\sum_{j=1}^{D} x_{1j}^{1-\lambda}x_{2j}^{\lambda}.$$

That is, it holds that

$$x_{1i}^{1-\lambda}x_{2i}^{\lambda} \geq k\sum_{j=1}^{D} x_{1j}^{1-\lambda}x_{2j}^{\lambda}. \tag{7}$$

Finally, writing Eq. (7) in terms of the i^{th} part of the composition $(1-\lambda)\odot\mathbf{x}_1 \oplus \lambda\odot\mathbf{x}_2$:

$$((1-\lambda)\odot\mathbf{x}_1 \oplus \lambda\odot\mathbf{x}_2)_i = \frac{x_{1i}^{1-\lambda}x_{2i}^{\lambda}}{\sum_{j=1}^{D} x_{1j}^{1-\lambda}x_{2j}^{\lambda}} \geq k,$$

that is, $(1-\lambda)\odot\mathbf{x}_1 \oplus \lambda\odot\mathbf{x}_2 \in \Sigma^+$. ∎

Proposition 2 *Let f_1 and f_2 be two $\mathcal{A}$-convex functions on the $\mathcal{A}$-convex set $W \subset \mathcal{S}^D$. The function $f_1 + f_2$ is $\mathcal{A}$-convex on W.*

Proof. Let $\mathbf{x}_1,\ \mathbf{x}_2$ be two D-part compositions in W. For any $\lambda \in [0,1]$ it holds that

$$f_1((1-\lambda)\odot\mathbf{x}_1 \oplus \lambda\odot\mathbf{x}_2) + f_2((1-\lambda)\odot\mathbf{x}_1 \oplus \lambda\odot\mathbf{x}_2) \leq \\ (1-\lambda)[f_1(\mathbf{x}_1)+f_2(\mathbf{x}_1)] + \lambda[f_1(\mathbf{x}_2)+f_2(\mathbf{x}_2)]. \tag{8}$$

∎

Proposition 4 *Let f be an $\mathcal{A}$-convex function on the $\mathcal{A}$-convex set $W \subset \mathcal{S}^D$. The sublevel set $\mathcal{G}^-_\alpha = \{\mathbf{x}|\ \mathbf{x} \in W, f(\mathbf{x}) \leq \alpha\}$ is an $\mathcal{A}$-convex set for any real number α.*

Proof. Let $\mathbf{x}_1,\ \mathbf{x}_2$ be in $\mathcal{G}^-_\alpha$ and $\lambda \in [0,\ 1]$,

$$f((1-\lambda)\odot\mathbf{x}_1 \oplus \lambda\odot\mathbf{x}_2) \leq (1-\lambda)f(\mathbf{x}_1) + \lambda f(\mathbf{x}_2) \leq \alpha.$$

Consequently, $(1-\lambda)\odot\mathbf{x}_1 \oplus \lambda\odot\mathbf{x}_2 \in \mathcal{G}^-_\alpha$. ■

Example 1 The function $f(\mathbf{x}) = \frac{x_i}{x_j},\ 1 \leq i, j \leq D$ is an $\mathcal{A}$-convex over its domain, $dom(f) = \mathcal{S}^D$.

Proof. Let $\mathbf{x}_1,\ \mathbf{x}_2$ be two D-part compositions and for any $\lambda \in [0,\ 1]$,

$$f((1-\lambda)\odot\mathbf{x}_1 \oplus \lambda\odot\mathbf{x}_2) = \left(\frac{x_{1i}}{x_{1j}}\right)^{1-\lambda}\left(\frac{x_{2i}}{x_{2j}}\right)^{\lambda}.$$

We apply Young's inequality, $a^{1-\lambda}b^\lambda \leq (1-\lambda)a + \lambda b$ for any $a, b \geq 0$ and $\lambda \in [0,\ 1]$, to the values $a = \dfrac{x_{1i}}{x_{1j}}$ and $b = \dfrac{x_{2i}}{x_{2j}}$,

$$\begin{aligned} f((1-\lambda)\odot\mathbf{x}_1 \oplus \lambda\odot\mathbf{x}_2) = \left(\frac{x_{1i}}{x_{1j}}\right)^{1-\lambda}\left(\frac{x_{2i}}{x_{2j}}\right)^{\lambda} \leq \\ (1-\lambda)\frac{x_{1i}}{x_{1j}} + \lambda\frac{x_{2i}}{x_{2j}} = (1-\lambda)f(\mathbf{x}_1) + \lambda f(\mathbf{x}_2). \quad (9) \end{aligned}$$

■

Example 2 For any $i = 1, \ldots, D$, the function $f(\mathbf{x}) = x_i$ is an $\mathcal{A}$-quasiconcave function over all its domain, $dom(f) = \mathcal{S}^D$. Moreover, using the $\mathcal{A}$-convex function

$$\Phi_\alpha(\mathbf{x}) = \alpha\sum_{j=1}^{D}\frac{x_j}{x_i} - 1,$$

a superlevel $\mathcal{G}^+_\alpha = \{\mathbf{x} \in \mathcal{S}^D|\ x_i \geq \alpha\}$ for any $\alpha \in (0,\ 1)$ can be represented by means of $\Phi_\alpha(\mathbf{x}) \leq 0$.

Proof. Through proposition 1, the set $\mathcal{G}^+_\alpha = \{\mathbf{x} \in \mathcal{S}^D|\ x_i \geq \alpha\}$ is $\mathcal{A}$-convex for any real number α.

Because $\mathbf{x} \in \mathcal{S}^D$, the equality $\sum_{j=1}^D x_j = 1$ holds. So, $x_i \geq \alpha \iff \alpha\sum_{j=1}^D \frac{x_j}{x_i} \leq 1$. And the function $\Phi_\alpha(\mathbf{x}) = \alpha\sum_{j=1}^D \frac{x_j}{x_i} - 1$ is $\mathcal{A}$-convex because it is a positive linear combination of $\mathcal{A}$-convex functions $\frac{x_j}{x_i}$. ■

References

Aitchison, J. (1986). *The statistical analysis of compositional data*. Chapman & Hall, London. Reprinted 2003 with additional material by The Blackburn Press, London, UK.

Barceló-Vidal, C. and Martín-Fernández, J. A. (2016). The mathematics of compositional analysis. *Austrian Journal of Statistics*, 45(4), 57-71.

Barceló-Vidal, C., Martín-Fernández, J. A. and Mateu-Figueras, G. (2011). Compositional differential calculus on the simplex. In: Pawlowsky, V. and Buccianti, A. (eds) *Compositional Data Analysis: Theory and Applications*. Chichester (UK), John Wiley & Sons, Chapter 13, 176-190.

Bates, S. and Tibshirani, R. (2019). Log-ratio lasso: Scalable, sparse estimation for log-ratio models. *Biometrics*, 75(2), 613-624.

Billheimer, D., Guttorp, P. and Fagan, W. F. (2001). Statistical interpretation of species composition. *Journal of the American Statistical Association*, 96(456), 1205-1214.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge university press.

Campbell, S. and Wong, T. (2022). Eficient convex pca with applications to wasserstein geodesic pca and ranked data. (in press).

Caussinus, H., Ettinger, P. and Tomassone, R. (2012). *COMPSTAT 1982 5th Symposium Held at Toulouse 1982: Part I: Proceedings in Computational Statistics*. Springer Science & Business Media.

Chastin, S.F.M., Palarea-Albaladejo, J., Dontje, M.L. and Skelton, D.A. (2015). Combined effects of time spent in physical activity, sedentary behaviors and sleep on obesity and cardio-metabolic markers: A novel compositional data analysis approach. *PLoS ONE*, 10.

Chen, R., Zhang, Z., Feng, C., Hu, K., Li, M., Li, Y., Shimizu, K., Chen, N., and Sugiura, N. (2010). Application of simplex-centroid mixture design in developing and optimizing ceramic adsorbent for as(v) removal from water solution microporous and mesoporous materials. *Microporous and Mesoporous Materials*, 131, 115-121.

Coetzer, R. and Haines, L. (2017). The construction of d- and i-optimal designs for mixture experiments with linear constraints on the components. *Chemometrics and Intelligent Laboratory Systems*, 171, 112-124.

Dumuid, D., Pedi**v**sić, Z., Palarea-Albaladejo, J., Martín-Fernández, J. A., Hron, K., and Olds, T. (2020). Compositional data analysis in time-use epidemiology: what, why, how. *International Journal of Environmental Research and Public Health*, 17(2220).

Dumuid, D., Wake, M., Burgner, D., Tremblay, M., Okely, A., Edwards, B., Dwyer, T., and Olds, T. (2021). Balancing time use for children's fitness and adiposity: Evidence to inform 24-hour guidelines for sleep, sedentary time and physical activity. *PLoS ONE*, 16(1).

Egozcue, J. J. and Pawlowsky-Glahn, V. (2005). Groups of parts and their balances in compositional data analysis. *Mathematical Geology*, 37, 795-828.

Egozcue, J. J., Pawlowsky-Glahn, V. and Gloor, G. (2018). Linear association in compositional data analysis. *Austrian Journal of Statistics*, 47, 3.

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

Fairclough, S., Dumuid, D., Mackintosh, K., Stone, G., Dagger, R., Stratton, G., Davies, I. and Boddy, L. (2018). Adiposity, fitness, health-related quality of life and the reallocation of time between children's school day activity behaviours: A compositional data analysis. *Preventive medicine reports*, 11, 254-261.

Gupta, N., Rasmussen, C., Holtermann, A., and Mathiassen, S. (2020). Time-based data in occupational studies: The whys, the hows, and some remaining challenges in compositional data analysis (coda). *Annals of Work Exposures and Health*, 64(8):778–785.

Halim, N., Abidin, Z., Siajam, S., Hean, C. and Harun, M. (2021). Optimization studies and compositional analysis of subcritical water extraction of essential oil from citrus hystrix dc. leaves. *The Journal of Supercritical Fluids*, 178.

Harsh, A., Ball, J. and Wei, P. (2016). Onion-peeling outlier detection in 2-d data sets. *International Journal of Computer Applications*, 139(3), 26-31.

Kitano, N., Kay, Y., Jindo, T., Tsunoda, K. and Arao, T. (2020). Compositional data analysis of 24-hour movement behaviors and mental health in workers. *Preventive medicine reports*, 20.

Lo Huang, M.-N. and Huang, M.-K. (2009). ϕ_p-optimal designs for a linear log contrast model for experiments with mixtures. *Metrika*, 70, 239-256.

Luenberger, D. and Ye, Y. (2008). *Linear and Nonlinear Programming*. International Series in Operations Research & Management Science. Springer US.

Martín-Fernández, J. A. (2019). Comments on: Compositional data: the sample space and its structure. *TEST*, 28(3), 653-657.

Martín-Fernández, J. A., Pawlowsky-Glahn, V., Egozcue, J. J. and Tolosona-Delgado, R. (2018). Advances in principal balances for compositional data. *Mathematical Geosciences*, 50(3), 273-298.

Mateu-Figueras, G., Pawlowsky-Glahn, V. and Egozcue, J. J. (2011). The principle of working on coordinates. In *Compositional Data Analysis: Theory and Applications.*

Mateu-Figueras, G., Pawlowsky-Glahn, V. and Egozcue, J. J. (2013). The normal distribution in some constrained sample spaces. *SORT (Statistics and Operations Research Transactions)*, 37, 29-56.

Pawlowsky-Glahn, V. and Egozcue, J. J. (2001). Geometric approach to statistical analysis on the simplex. *Stochastic Environmental Research and Risk Assessment*, 15, 384-398.

Pawlowsky-Glahn, V., Egozcue, J. J. and Tolosana-Delgado, R. (2015). *Modeling and analysis of compositional data*. John Wiley & Sons, Chichester.

R-Core-Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Small, C. (1990). A survey of multidimensional medians. *International Statistical Review*, 58, 263-277.

Susin, A., Wang, Y., Lê Cao, K.-A. and Calle, M. L. (2020). Variable selection in microbiome compositional data analysis. *NAR Genomics and Bioinformatics*, 2(2), lqaa029.

Tolosana-Delgado, R., von Eynatten, H. and Karius, V. (2011). Constructing modal mineralogy from geochemical composition: A geometric-bayesian approach. *Computers & Geosciences*, 37(5), 677-691.

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

van den Boogaart, K., Filzmoser, P., Hron, K., Templ, M. and Tolosana-Delgado, R. (2021). Classical and robust regression analysis with compositional data. *Mathematical Geosciences*, 53, 823-858.

Wang, X., Wang, H., Wang, Z. and Yuan, J. (2020). Convex clustering method for compositional data modeling. *Soft Computing*, 25, 2965-2980.