

## Isometric Logratio Transformations for Compositional Data Analysis<sup>1</sup>

J. J. Egozcue,<sup>2</sup> V. Pawlowsky-Glahn,<sup>3</sup> G. Mateu-Figueras,<sup>3</sup>  
and C. Barceló-Vidal<sup>3</sup>

---

*Geometry in the simplex has been developed in the last 15 years mainly based on the contributions due to J. Aitchison. The main goal was to develop analytical tools for the statistical analysis of compositional data. Our present aim is to get a further insight into some aspects of this geometry in order to clarify the way for more complex statistical approaches. This is done by way of orthonormal bases, which allow for a straightforward handling of geometric elements in the simplex. The transformation into real coordinates preserves all metric properties and is thus called isometric logratio transformation (ilr). An important result is the decomposition of the simplex, as a vector space, into orthogonal subspaces associated with nonoverlapping subcompositions. This gives the key to join compositions with different parts into a single composition by using a balancing element. The relationship between ilr transformations and the centered-logratio (clr) and additive-logratio (alr) transformations is also studied. Exponential growth or decay of mass is used to illustrate compositional linear processes, parallelism and orthogonality in the simplex.*

---

**KEY WORDS:** Aitchison distance, Aitchison geometry, geodesic, orthogonal subcompositions, ternary diagram.

### INTRODUCTION

The statistical analysis of compositional data got an independent status when Aitchison (1986) introduced the additive-logratio (alr) and centered-logratio (clr) transformations, as well as an appropriate distance in the simplex, the sample space of compositional data without zero parts. He showed that these transformations from the simplex into multidimensional real space exhibit important properties. Among them, invariance under scale group of transformations, subcompositional consistency, and correspondence of the group of translations in multidimensional

---

<sup>1</sup>Received 5 April 2002; accepted 17 February 2002

<sup>2</sup>Dept. de Matemàtica Aplicada III, Universitat Politècnica de Catalunya, Barcelona, Spain; e-mail: juan.jose.egozcue@upc.es.

<sup>3</sup>Dept. d'Informàtica i Matemàtica Aplicada, Universitat de Girona, Girona, Spain; e-mail: vera.pawlowsky@udg.es; gloria.mateu@udg.es; carles.barcelo@udg.es.

real space with the group of perturbations in the simplex were enough to consider those transformations as firm candidates to define the standard way of analyzing compositional data. He also developed special strategies for dealing with data sets including zeros, hence all possible types of compositional data could be studied using his approach. As a consequence, in order to analyze a given data set, a widely extended approach now is to apply first an appropriate strategy when zeroes are present (see Martín-Fernández, Barceló-Vidal and Pawłowsky-Glahn, 2003, for details), and then either to alr- or clr-transform the data. Transformed data are then analyzed in the same way as standard data. Statistical models based on those transformations were developed successfully in most cases, as can be seen in the monograph by Aitchison (1986) and later developments (Aitchison, 1989, 1990, 1992, 1997, 2002; Aitchison et al., 2000; Aitchison and Thomas, 1998; Barceló-Vidal, Martín-Fernández, and Pawłowsky-Glahn, 2001; Billheimer, Guttorp, and Fagan, 2001; Pawłowsky-Glahn, 2002; Pawłowsky-Glahn and Egozcue, 2001, 2002; von Eynatten, Pawłowsky-Glahn, and Egozcue, 2002). However, there are a number of researchers who argue against Aitchison's methods, mainly due to difficulties in interpretation of the transformed data.

In fact, the alr and clr transformations have some shortcomings that make interpretation difficult. The alr transformation—defined as the logarithm of the ratio of parts over a given one—is asymmetric in the parts of the composition. By changing the part in the denominator, we obtain different alr transformations. But the main drawback of the alr transformation is not the asymmetry in its definition, but the fact that it is not an isometric transformation from the simplex, with the Aitchison metric, onto the real alr-space, with the ordinary Euclidean metric. Certainly, to overcome this problem one could use an appropriate metric with oblique coordinates in real alr-space, but that is not standard practice. As a consequence, when one does not take into account the absence of isometry, the interpretation of transformed data is not intuitive at all, as results occasionally do not match properties expected in the simplex (see Martín-Fernández, Olea-Meneses, and Pawłowsky-Glahn, 2001, for an example). Nevertheless, it is important to note that consistency of results is guaranteed, even in the presence of zeros, when inference is based on the likelihood of the additive logistic normal distribution (Aitchison, 1986).

The clr transformation is symmetric with respect to the compositional parts and keeps the same number of components as the number of parts in the composition. A composition with  $D$  parts is transformed into  $D$  real components adding up to 0. Thus, for example, for  $D = 3$ , clr-transformed data should be represented in a three-dimensional real space, although the linear restriction in the components would allow a two-dimensional representation. The advantage of the clr is that it is an isometric transformation of the simplex with the Aitchison metric, onto a subspace of real space with the ordinary Euclidean metric. Nevertheless, orthogonal references in that subspace are not obtained in a straightforward manner.

Our aim is to overcome these shortcomings giving new tools for interpreting compositional data by introducing a new family of logratio transformations of the simplex. This family is called *isometric* because it allows us to associate *angles* and *distances* in the simplex (following the Aitchison geometry) to *angles* and *distances* in real space, where we feel more comfortable from an intuitive point of view. This is of particular interest with respect to concepts of orthogonality.

In order to motivate the definition of the new family of isometric logratio (ilr) transformations and exploit the underlying intuitive ideas, we first introduce simple compositional models. These models help us understand the meaning of orthogonality in the simplex and, simultaneously, the role of perturbation.

There are also mathematical reasons for introducing this new family of transformations. In fact, all Hilbert spaces are isometric whenever they have the same dimension (Berberian, 1961). Thus, a natural undertaking is to find an isometry between the simplex, which is a Hilbert space—and hence an Euclidean space—and a real space of the same dimension (Billheimer, Guttorp, and Fagan, 2001; Pawlowsky-Glahn and Egozcue, 2001, 2002). Note that this task is neither solved by alr—it is not isometric—nor by clr—it is not expressed in terms of a basis neither of the simplex nor of the image subspace.

Second, we recall some important aspects of the Aitchison geometry in the simplex. Then introduce simple processes in the simplex, to help us understand orthogonality and perturbation. The section that follows builds up orthonormal bases that allow the explicit definition of the family of ilr transformations. Finally, we give the relationships between a given ilr, the clr and a given alr transformation.

## AITCHISON GEOMETRY IN THE SIMPLEX

### Vector Space Structure

The  $D$ -part simplex is defined as

$$\mathcal{S}^D = \left\{ \mathbf{x} = [x_1, x_2, \dots, x_D]; \quad x_i > 0, \quad i = 1, 2, \dots, D; \quad \sum_{i=1}^D x_i = \right\}, \quad (1)$$

where  $\quad$  is an arbitrary positive constant, e.g. 1 if measurements are performed in parts per 1 or 100 if they are in percentages. The elements of  $\mathcal{S}^D$ , which are considered to be row vectors, will be called  $D$ -part compositions. The  $(D - 1)$ -dimensional vector space structure of  $\mathcal{S}^D$  on  $\mathfrak{R}$  is based on the *perturbation* operation, defined for any two compositions  $\mathbf{x}, \mathbf{y} \in \mathcal{S}^D$ , as

$$\mathbf{x} \oplus \mathbf{y} = \mathcal{C}[x_1 y_1, x_2 y_2, \dots, x_D y_D], \quad (2)$$

and on an external operation called *power transformation*, given for  $\mathbf{x} \in \mathcal{S}^D$  and  $\alpha \in \Re$  by

$$\otimes \mathbf{x} = \mathcal{C}[x_1, x_2, \dots, x_D], \quad (3)$$

where  $\mathcal{C}(\cdot)$  denotes the *closure* of the vector in the argument (i.e. each component is divided by the sum of all the components and then multiplied by  $\alpha$ , so that the resulting row vector is in  $\mathcal{S}^D$ ) (Aitchison, 2002; Billheimer, Guttorp, and Fagan, 2001; Pawlowsky-Glahn and Egozcue, 2001, 2002). Note that perturbation can be defined between positive vectors that are not necessarily normalized to 1, because the closure operator is explicitly included in definition (2), and that the neutral element with respect to perturbation is  $\mathcal{C}[1, 1, \dots, 1]$ . In what follows  $\mathbf{x} \ominus \mathbf{y}$  will denote perturbation of  $\mathbf{x}$  with the opposite element of  $\mathbf{y}$ , which is  $(-1) \otimes \mathbf{y} = \mathcal{C}(\mathbf{y}^{-1})$ , i.e.

$$\mathbf{x} \ominus \mathbf{y} = \mathbf{x} \oplus ((-1) \otimes \mathbf{y}) = \mathbf{x} \oplus \mathbf{y}^{-1}.$$

### Hilbert Space Structure

It is well known that a finite dimensional vector space with an inner product is an Euclidean space and is complete. Now, for any two compositions  $\mathbf{x}, \mathbf{y} \in \mathcal{S}^D$ , we can define

$$\langle \mathbf{x}, \mathbf{y} \rangle_a = \frac{1}{D} \sum_{i < j} \ln \frac{x_i}{x_j} \ln \frac{y_i}{y_j} = \sum_{i=1}^D \ln \frac{x_i}{g(\mathbf{x})} \ln \frac{y_i}{g(\mathbf{y})}, \quad (4)$$

where  $g(\mathbf{x}) = [x_1 x_2 \cdots x_D]^{1/D}$  is the geometric mean of the parts of  $\mathbf{x}$ . It is straightforward to show that Equation (4) is an inner product, and thus  $\mathcal{S}^D$  is a  $(D - 1)$ -dimensional Hilbert space (Pawlowsky-Glahn and Egozcue, 2001, 2002). The expression (4) involving logratios of parts over geometric means of parts is related to the clr transformation defined by Aitchison (1986):

$$\text{clr}(\mathbf{x}) = \left[ \ln \frac{x_1}{g(\mathbf{x})}, \ln \frac{x_2}{g(\mathbf{x})}, \dots, \ln \frac{x_D}{g(\mathbf{x})} \right] = \ln \left( \frac{\mathbf{x}}{g(\mathbf{x})} \right), \quad (5)$$

whereas an expression using the alr transformation, i.e.

$$\text{alr}(\mathbf{x}) = \left[ \ln \frac{x_1}{x_D}, \ln \frac{x_2}{x_D}, \dots, \ln \frac{x_{D-1}}{x_D} \right], \quad (6)$$

was found by Billheimer, Guttorp, and Fagan (2001), who developed the algebra for compositions and the proof of the Hilbert space structure of the simplex in terms of this transformation.

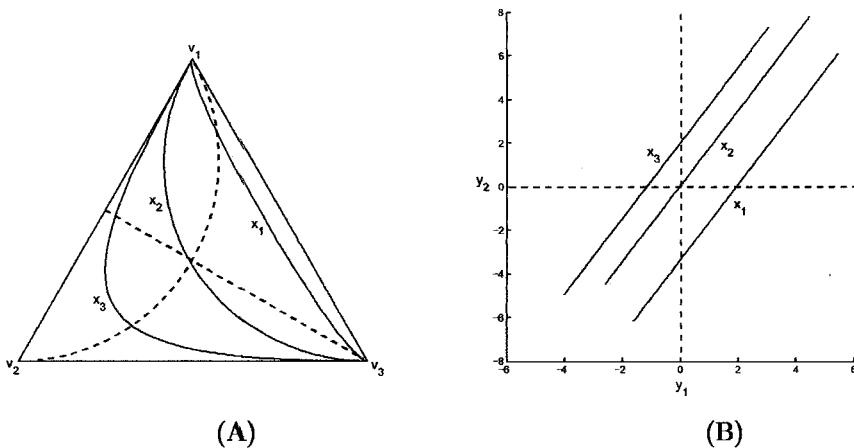
The inner product (4) induces a norm and a distance in  $S^D$  in the standard way,

$$\|\mathbf{x}\|_a^2 = \langle \mathbf{x}, \mathbf{x} \rangle_a, \quad d_a(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} \ominus \mathbf{y}\|_a. \tag{7}$$

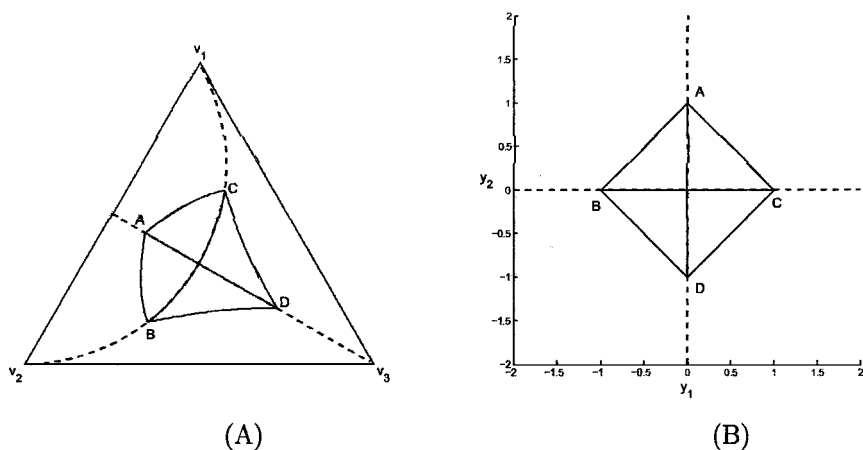
Note that the inner product (4), and thus the norm and the distance (7), are invariant under permutation of the parts of a composition. The extensive expression for  $d_a$  is

$$d_a(\mathbf{x}, \mathbf{y}) = \left[ \frac{1}{D} \sum_{i < j} \left( \ln \frac{x_i}{x_j} - \ln \frac{y_i}{y_j} \right)^2 \right]^{1/2} = \left[ \sum_{i=1}^D \left( \ln \frac{x_i}{g(\mathbf{x})} - \ln \frac{y_i}{g(\mathbf{y})} \right)^2 \right]^{1/2}. \tag{8}$$

The distance (8) is known as Aitchison distance in  $S^D$  and, therefore, the Hilbert geometry mentioned here will be called Aitchison geometry. The Aitchison geometry in  $S^D$  is an Euclidean geometry because the distance comes from an inner product. However, when representing *straight lines*, *rectangles*, and *circles* in this geometry (Fig. 1(A)–5(A)) one can observe features that differ from what commonly feels an Euclidean geometry (Fig. 1(B)–5(B)). To avoid confusion, from now on, we term Euclidean distance, geometry, etc., only those which are standard in real space.

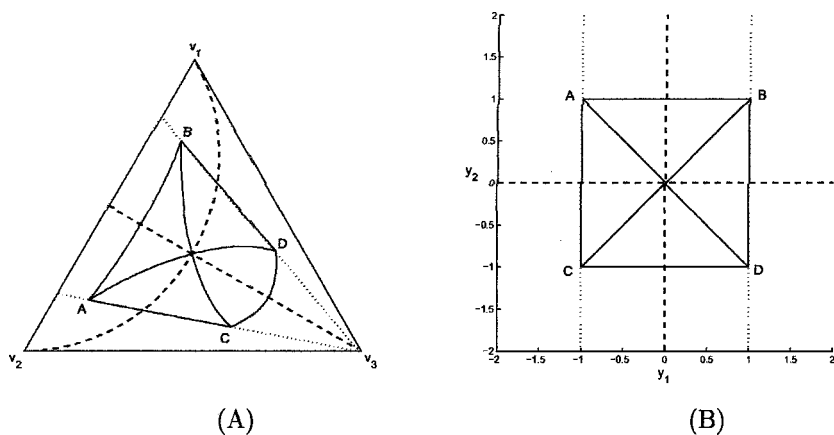


**Figure 1.** (A) Representation of compositional parallel processes in vector space  $S^3$ ; dashed lines correspond to orthogonal coordinate axes. (B) Same processes after ilr transformation;  $y_1 = 2^{-2} \ln(v_1/v_2)$ ,  $y_2 = 6^{-2} \ln((v_1 v_2)/(v_3 v_3))$ .

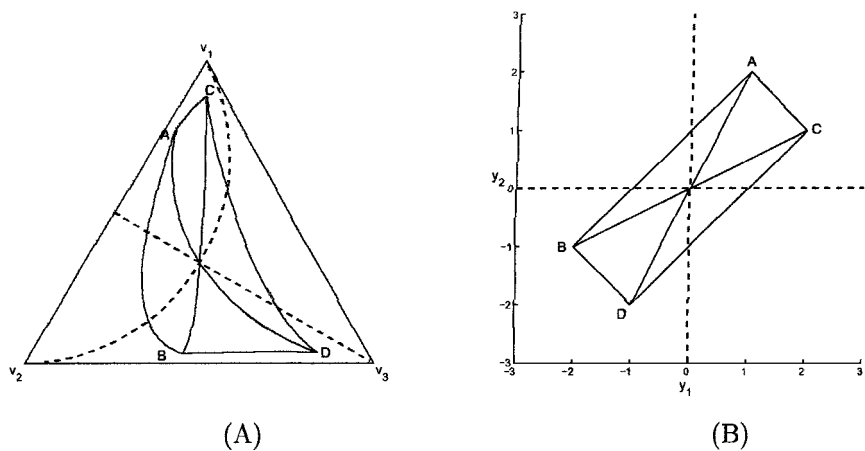


**Figure 2.** (A) Representation of centered compositional square in vector space  $\mathcal{S}^3$ , dashed lines correspond to orthogonal coordinate axes. Diagonals of square coincident with coordinate axes. (B) Same square after ilr transformation;  $y_1 = 2^{-2} \ln(v_1/v_2)$ ,  $y_2 = 6^{-2} \ln((v_1 v_2)/(v_3 v_3))$ .

The  $(D - 1)$ -Hilbert space structure of the simplex allows us to define compositional straight lines, orthogonal lines, angles between lines and all standard elements in an Euclidean geometry. Special attention is now paid to compositional lines and relations of orthogonality and parallelism between them.



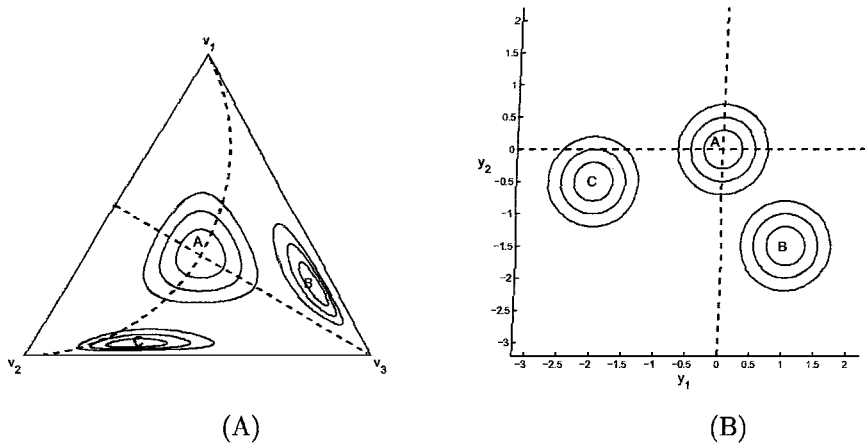
**Figure 3.** (A) Representation of centered compositional square in vector space  $\mathcal{S}^3$ , dashed lines correspond to orthogonal coordinate axes. Sides of square parallel to coordinate axes. (B) Same square after ilr transformation;  $y_1 = 2^{-2} \ln(v_1/v_2)$ ,  $y_2 = 6^{-2} \ln((v_1 v_2)/(v_3 v_3))$ .



**Figure 4.** (A) Representation of centered compositional rectangle in vector space  $\mathcal{S}^3$ , dashed lines correspond to orthogonal coordinate axes. Sides rotated with respect to coordinate axes. (B) Same rectangle after ilr transformation;  $y_1 = 2^{-2} \ln(v_1/v_2)$ ,  $y_2 = 6^{-2} \ln((v_1 v_2)/(v_3 v_3))$ .

**Straight, Parallel and Orthogonal Lines**

Recall that in any finite dimensional Hilbert space, a geodesic curve connecting two points is understood to be the only continuous curve whose length is minimum with respect to the distance of the space. Such a geodesic is a segment of



**Figure 5.** (A) Representation of compositional circles at different positions in vector space  $\mathcal{S}^3$ , dashed lines correspond to orthogonal coordinate axes. (B) Same circles after ilr transformation;  $y_1 = 2^{-2} \ln(v_1/v_2)$ ,  $y_2 = 6^{-2} \ln((v_1 v_2)/(v_3 v_3))$ .

a straight line within the geometry of the space considered. The parametric curve

$$\mathbf{x}(t) = \mathbf{x}_0 \oplus (t \otimes \mathbf{p}); \quad t \in \mathfrak{R}, \mathbf{x}_0, \mathbf{p} \in \mathcal{S}^D, \quad (9)$$

is a straight line containing the geodesic—with respect to the Aitchison distance (8)—going from  $\mathbf{x}_0$  to  $\mathbf{x}(t)$  in  $\mathcal{S}^D$ , for a given  $t$ . This will become self-evident after defining the family of isometric logratio transformations, *ilr*. (see under Isometric Logratio Transformations). We will say that the composition  $\mathbf{p}$  is the direction of such a geodesic or of the line containing it. All curves in  $\mathcal{S}^3$  shown in Figure 1(A) are straight lines with respect to the Aitchison geometry, which we call *compositional lines* to avoid confusion.

We can define orthogonal compositional lines at  $\mathbf{x}_0 \in \mathcal{S}^D$  as follows. Let  $\mathbf{p}_1, \mathbf{p}_2 \in \mathcal{S}^D$  be two orthogonal directions in  $\mathcal{S}^D$ , i.e.  $\langle \mathbf{p}_1, \mathbf{p}_2 \rangle_a = 0$ . Then,  $\mathbf{x}_1(t) = \mathbf{x}_0 \oplus (t \otimes \mathbf{p}_1)$  and  $\mathbf{x}_2(t) = \mathbf{x}_0 \oplus (t \otimes \mathbf{p}_2)$  are orthogonal compositional lines at  $\mathbf{x}_0$  because for any  $t_1, t_2 \in \mathfrak{R}$ ,

$$\langle \mathbf{x}_1(t_1) \ominus \mathbf{x}_0, \mathbf{x}_2(t_2) \ominus \mathbf{x}_0 \rangle_a = t_1 t_2 \langle \mathbf{p}_1, \mathbf{p}_2 \rangle_a = 0. \quad (10)$$

In our ordinary Euclidean space, such a statement means that segments along orthogonal straight lines (orthogonal directions) are perpendicular.

Parallelism between compositional lines is also easily defined. For any composition  $\mathbf{p} \in \mathcal{S}^D$  and two fixed points  $\mathbf{x}_0, \mathbf{x}_0^* \in \mathcal{S}^D$ , the compositional lines  $\mathbf{x}(t) = \mathbf{x}_0 \oplus (t \otimes \mathbf{p})$  and  $\mathbf{x}^*(t) = \mathbf{x}_0^* \oplus (t \otimes \mathbf{p})$ , are parallel, i.e. either they are coincident or they do not intersect in the simplex. Graphically, for  $D = 3$ , these parallel compositional lines converge either to the same vertices (Fig. 1(A)), or to one single vertex (Fig. 3(A)) (vertices and edges are placed at an infinite Aitchison distance from any point in the simplex).

After introducing these concepts about compositional lines in the simplex, we look for examples of natural processes whose compositional evolution is described by compositional lines and for the meaning of orthogonality and parallelism of such processes.

## EXPONENTIAL GROWTH AND DECAY OF MASS AS COMPOSITIONAL PROCESSES

### Linear Compositional Processes

Nature presents a large number of cases where some mass is lost or increases (approximately) exponentially in time. There are important examples in biology (population growth), in physics (radioactive decomposition of isotopes), or in geology (sedimentation or deposition). The mass deposition of different materials suspended in a steady water flow can be used as an example. If the conditions are



stationary and the different types of material do not interact, we can assume that the total mass of each material deposited instantaneously is proportional to the mass of this material suspended at the same instant. This leads to an exponential decrease of the total mass suspended in the water flow for each type of material. Denote by  $\mathbf{z}(t) = [z_1(t), \dots, z_D(t)]$  the mass suspended at time  $t$  for each material type. Then, the exponential loss of material can be modelled as

$$z_i(t) = z_i(0) \exp[a_i t], \quad i = 1, 2, \dots, D, \quad (11)$$

where  $[a_1, a_2, \dots, a_D] = \mathbf{a}$  is a vector of negative constants representing the mass loss rate. We can assume the components of  $\mathbf{a}$  to be all negative, i.e. all masses decrease. The model (11) for masses allows us to obtain the compositional evolution in time of the different types of material by simply applying the closure operator to Equation (11), i.e. by dividing each component by the total sum of the vector:

$$\mathbf{x}(t) = \mathcal{C}(\mathbf{z}(t)) = \mathbf{x}_0 \oplus \exp[\mathbf{a}t], \quad \mathbf{x}_0 = \mathcal{C}(\mathbf{z}(0)), \quad (12)$$

where the closure in the exponential vector has been removed because perturbation includes it. We immediately identify the compositional evolution (12) with the expression of a compositional line (9) in the simplex by defining  $\mathbf{p} = \mathcal{C}(\exp[\mathbf{a}])$ , and thus obtaining  $t \otimes \mathbf{p} = t \otimes \mathcal{C}(\exp[\mathbf{a}]) = \mathcal{C}(\exp[\mathbf{a}t])$ .

A second example would be the growth of noninteracting species of bacteria in a rich medium, which can be assumed proportional to the existing population of each species, thus producing an exponential growth in each component. The mass model (11) is still valid simply taking positive values in the rate vector  $\mathbf{a}$ . Obviously, the compositional evolution is again that represented by Equation (12) and, therefore, this kind of process also follows a compositional line in the simplex. Naturally, we can imagine physical processes in which some components increase exponentially their mass, while others decrease exponentially and some of them remain constant. All these processes, when viewed as compositions, evolve following compositional lines in the simplex whatever the rate vector  $\mathbf{a}$ .

### Parallel Compositional Processes

An important feature of linear compositional processes is that the direction  $\mathbf{p} = \mathcal{C}(\exp[\mathbf{a}])$  does not depend on additive terms in the rate vector. This is due to the fact that

$$\mathbf{p} = \mathcal{C}(\exp[\mathbf{a}]) = \mathcal{C}(\exp[\mathbf{a} + \mathbf{1}]), \quad (13)$$

for any real  $\alpha$  and for  $\mathbf{1} = [1, 1, \dots, 1]$ . Thus, different exponential evolutions of mass represent parallel compositional processes whenever their rate vectors differ in the same constant in all the components. Moreover, they follow the same compositional line whenever the initial condition is equal for the two processes or whenever they coincide in some evolutionary state. For instance, the following four rate vectors correspond to the same direction in  $\mathcal{S}^3$ :  $\mathbf{a}_1 = [1, 2, 3]$ ,  $\mathbf{a}_2 = [0, 1, 2]$ ,  $\mathbf{a}_3 = [-1, 0, 1]$ ,  $\mathbf{a}_4 = [-2, -1, 0]$ . Compositional lines following directions  $\exp[\mathbf{a}_i]$ ,  $i = 1, 2, 3$ , are shown in Figure 1(A) and they are parallel.

### Constant Compositional Processes

Constant processes have also different meanings for mass processes and their compositional counterpart. For instance, the total mass in process (11) cannot be constant except for  $\mathbf{a}$  being null. If  $\mathbf{a}$  is null, both the mass process (11) and the compositional process (12) are constant. But, if  $\mathbf{a}$  has the same value in each component, the mass process changes in time, but the composition remains constant.

### Orthogonal Compositional Processes

Different interpretations of orthogonality are required when observing processes in real space, with the standard Euclidean geometry, and the corresponding compositional processes. For instance, again in  $\mathcal{S}^3$ , consider two rate vectors  $\mathbf{a}_1 = [1, 0, 0]$  and  $\mathbf{a}_2 = [0, -1, -1]$  and the corresponding exponential mass processes  $\mathbf{z}_1(t)$ ,  $\mathbf{z}_2(t)$  as defined in Equation (11). One can observe that for any  $t_1$  and  $t_2$ ,  $\mathbf{z}_1(t_1) - \mathbf{z}_1(0)$  is an orthogonal vector (ordinary Euclidean sense) to  $\mathbf{z}_2(t_2) - \mathbf{z}_2(0)$ . In fact,

$$\begin{aligned}\mathbf{z}_1(t_1) - \mathbf{z}_1(0) &= (e^{t_1} - 1) [z_{11}(0), 0, 0], \\ \mathbf{z}_2(t_2) - \mathbf{z}_2(0) &= (e^{-t_2} - 1) [0, z_{22}(0), z_{23}(0)],\end{aligned}$$

and they are orthogonal (ordinary Euclidean sense) for any  $t_1$  and  $t_2$ . Therefore, intuitively speaking, the mass change in the first component would be considered not to be related (linearly) to the variation of mass in the second and third components. But this orthogonality between the two mass processes is completely lost when the two processes are viewed as compositional: since the rate vectors  $\mathbf{a}_1$  and  $\mathbf{a}_2$  differ in an additive constant, they follow parallel compositional lines and, if initial conditions are equal or they coincide in some evolutionary state, they follow the same compositional line. The conclusion is clear: ordinary Euclidean orthogonality of mass processes does not imply orthogonality of the corresponding

compositional processes. Furthermore, ordinary Euclidean geometry of mass processes is not adapted to previously stated requirements on the geometry of compositional processes. We remark that mass processes are real positive valued and, following ideas in Pawlowsky-Glahn and Egozcue (2001), it is possible to define an alternative geometry for mass processes compatible with the compositional geometry.

## ORTHOGONALITY AND SUBCOMPOSITIONS

After discussing what *should not be* orthogonality in  $\mathcal{S}^D$ , a natural question is what should orthogonality be? Some intuitive ideas about subcompositions may point out desirable properties of orthogonality in the simplex. To this end, recall the example of deposition of masses. The statement *different materials do not interact* implies that separating  $\mathbf{x}(t)$  into two subvectors, each one containing at least two components,  $\mathbf{x}(t) = (\mathbf{r}(t), \mathbf{s}(t))$ , should give two subcompositions,  $\mathcal{C}(\mathbf{r}(t)) \in \mathcal{S}^r$ ,  $\mathcal{C}(\mathbf{s}(t)) \in \mathcal{S}^s$ ,  $D = r + s$ , that intuitively are not related. In our approach to the Aitchison geometry, this fact is equivalent to the statement  $\mathcal{C}(\mathbf{r}(t))$  and  $\mathcal{C}(\mathbf{s}(t))$  are orthogonal.

In order to establish an orthogonality relationship between nonoverlapping compositions in  $\mathcal{S}^r$  and  $\mathcal{S}^s$ , it is necessary to imbed both spaces in a larger one, which we identify as  $\mathcal{S}^D$ , with  $D = r + s$ . The way of doing this is to define compositions in  $\mathcal{S}^D$  associated with  $\mathcal{C}(\mathbf{r}(t))$  and  $\mathcal{C}(\mathbf{s}(t))$ , respectively,

$$\mathbf{R}(t) = \mathcal{C}(\mathbf{r}(t), \mathbf{c}_s), \quad \mathbf{S}(t) = \mathcal{C}(\mathbf{c}_r^*, \mathbf{s}(t)), \quad (14)$$

where  $\mathbf{c}_s$ , denotes a vector whose  $s$  components are equal to an arbitrary constant  $c > 0$ , and analogously for  $\mathbf{c}_r^*$ , made up of  $r$  equal constants  $c^* > 0$ . This construction of  $\mathbf{R}(t)$  and  $\mathbf{S}(t)$  corresponds to compositional linear processes in  $\mathcal{S}^D$  in which some components reproduce the evolution of the selected subcomposition but the remaining components are constant in time (i.e. they do not play any role both in the deposition of masses and in the corresponding compositional process). A requirement on  $\mathbf{R}(t)$  and  $\mathbf{S}(t)$  would be that the compositions

$$\mathbf{R}_0(t) = \mathbf{R}(t) \ominus \mathbf{R}(0), \quad \mathbf{S}_0(t) = \mathbf{S}(t) \ominus \mathbf{S}(0), \quad (15)$$

should be orthogonal whenever  $\mathbf{R}_0(t)$  and  $\mathbf{S}_0(t)$  are not constant processes. A valid question, which arises at this point, is the possible connection between this subcompositional orthogonality and partition independence (Aitchison, 1986), at least under certain model assumptions, particularly of additive logistic normality. Although we have preliminary results in this direction, they are not mature enough to be included in the present paper.

## ORTHONORMAL BASES IN THE SIMPLEX

The aim of this section is to build up orthonormal bases—both general and associated to a partition of the full composition into subcompositions—of  $\mathcal{S}^D$  in a practical way; these orthonormal bases will be the starting point to define isometric logratio transformations mapping  $\mathcal{S}^D$  onto  $\mathfrak{N}^{D-1}$ .

### General Expression

Orthonormal bases can be readily obtained by the standard Gram–Schmidt procedure once an independent set of  $D - 1$  compositions in  $\mathcal{S}^D$  is given. This is known to be so for any vector space with an inner product. But we are interested in carrying out this process explicitly so that the orthonormal basis can be written immediately for any dimension. Once an orthonormal basis is obtained, we will study a special kind of orthonormal basis associated with a subcomposition.

The first step is to get a simplified expression for the orthogonality of two compositions. The clr transformation (5) gives an adequate way. The clr transformation assigns each composition in  $\mathcal{S}^D$  to a row vector  $\mathbf{a} = \text{clr}(\mathbf{x})$  in  $\mathfrak{N}^D$ , satisfying  $\sum_{i=1}^D a_i = 0$ . Vectors satisfying such a condition constitute a  $(D - 1)$ -dimensional subspace of  $\mathfrak{N}^D$ , denoted here by  $V_{\mathcal{S}}$ . We remark that the ratio of the compositional parts to their geometric mean in the clr transformation is convenient because the clr image of the neutral composition  $\mathbf{e} = \mathcal{C}[1, 1, \dots, 1]$  is  $\text{clr}(\mathbf{e}) = [0, 0, \dots, 0]$ , the neutral element in  $\mathfrak{N}^D$ ; this is necessary to define an isomorphism of linear spaces between  $\mathcal{S}^D$  and  $V_{\mathcal{S}}$ . The inner product (4) can be expressed in the clr-transformed space as follows: for  $\mathbf{a} = \text{clr}(\mathbf{x})$  and  $\mathbf{b} = \text{clr}(\mathbf{y})$

$$\langle \mathbf{x}, \mathbf{y} \rangle_a = \frac{1}{D} \sum_{i < j} (a_i - a_j)(b_i - b_j) = \frac{1}{D} \mathbf{a} \mathbf{M} \mathbf{b}', \quad (16)$$

where the  $D \times D$  matrix  $\mathbf{M}$  is

$$\mathbf{M} = \begin{pmatrix} D-1 & -1 & -1 & \cdots & -1 \\ -1 & D-1 & -1 & \cdots & -1 \\ -1 & -1 & D-1 & \cdots & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & -1 & \cdots & D-1 \end{pmatrix}.$$

The right hand side of Equation (16) defines a degenerate inner product in  $\mathfrak{N}^D$ , whose metric matrix is  $\mathbf{M}$ . Degeneracy of  $\mathbf{M}$  is easily checked. It has two eigenvalues: 0 with multiplicity 1, and  $D$  with multiplicity  $D - 1$  and, consequently,  $\mathbf{M}$  is positive semidefinite. The eigenspace associated with the eigenvalue  $D$  is

precisely the linear subspace  $V_S$ , defined by the condition  $\sum_{i=1}^D a_i = 0$  and, consequently,  $\text{clr}(\mathbf{y})\mathbf{M} = D \text{clr}(\mathbf{y})$ . This means that, when carrying out inner products by means of Equation (16), the result does not depend on the scaling value  $g(\mathbf{x})$  used in the clr transformation. Consequently, for computing Aitchison inner products as indicated in Equation (16), we can change  $\mathbf{a} = \text{clr}(\mathbf{x})$  simply to  $\mathbf{a} = \ln(\mathbf{x}) = [\ln x_1, \ln x_2, \dots, \ln x_D]$  and  $\langle \mathbf{x}, \mathbf{y} \rangle_a$  remains unaltered. This result is consistent with the fact that a composition  $\mathbf{x}$  is just a representant of an equivalence class, as stated by Barceló-Vidal, Martín-Fernández, and Pawłowsky-Glahn (2001).

In order to obtain an orthonormal basis of the linear subspace associated with the eigenvalue  $D$  (i.e.  $V_S$ ), we select a set of  $D - 1$  linearly independent vectors in that subspace. Let them be  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{D-1}$  defined as  $\mathbf{v}_i = [0, \dots, 0, 1, -1, 0, \dots, 0]$ , the first non-null element being placed in the  $i$ -th column. For any two vectors,  $\mathbf{a}$  and  $\mathbf{b}$  in  $V_S$ , the ordinary Euclidean inner product is  $\langle \mathbf{a}, \mathbf{b} \rangle = D^{-1} \mathbf{a} \mathbf{M} \mathbf{b}'$  as pointed out before. Therefore, the Gram–Schmidt procedure, with respect to the ordinary Euclidean inner product, can be applied to  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{D-1}$  and we obtain the following result.

*Proposition 1.* Let  $\mathbf{u}_i \in \Re^D, i = 1, 2, \dots, D - 1$ , be the vectors

$$\mathbf{u}_i = \sqrt{\frac{i}{i+1}} \left[ \underbrace{\frac{1}{i}, \dots, \frac{1}{i}}_{i \text{ elements}}, -1, 0, \dots, 0 \right]. \quad (17)$$

The vectors  $\mathbf{u}_i$  are orthonormal with respect to the ordinary Euclidean inner product in  $\Re^D$  and constitute a basis of  $(D - 1)$ -dimensional linear subspace  $V_S$ .

*Proof:* The vector  $\mathbf{u}_1 = \mathbf{v}_1 / \|\mathbf{v}_1\|$  is in  $V_S$  because its components add to zero. Assume now that the vectors  $\mathbf{u}_i, i = 1, 2, \dots, k$ , are in  $V_S$  and that they are orthonormal. Applying the Gram–Schmidt procedure we obtain the next orthonormal vector

$$\mathbf{w} = \mathbf{v}_{k+1} - \sum_{i=1}^{k-1} \langle \mathbf{v}_{k+1}, \mathbf{u}_i \rangle \mathbf{u}_i - \langle \mathbf{v}_{k+1}, \mathbf{u}_k \rangle \mathbf{u}_k, \quad \mathbf{u}_{k+1} = \frac{\mathbf{w}}{\|\mathbf{w}\|},$$

where  $\langle \cdot, \cdot \rangle$  denotes the ordinary Euclidean inner product in  $\Re^D$ . The inner products in the summation vanish and the resulting expression defines  $\mathbf{u}_{k+r} \in V_S$  as stated in Equation (17). This completes the proof by induction.  $\square$

As a consequence of the previous proposition, we can obtain an Aitchison-orthonormal basis in the simplex by applying  $\text{clr}^{-1}$  to the orthonormal vectors in  $V_S$ .

*Proposition 2.* Let  $\mathbf{e}_i, i = 1, 2, \dots, D - 1$ , be the following compositions in  $\mathcal{S}^D$ :

$$\mathbf{e}_i = \mathcal{C}(\exp \mathbf{u}_i) = \mathcal{C} \left[ \exp \left( \underbrace{\left( \sqrt{\frac{1}{i(i+1)}}, \dots, \sqrt{\frac{1}{i(i+1)}} \right)}_{i \text{ elements}}, -\sqrt{\frac{i}{i+1}}, 0, \dots, 0 \right) \right]. \quad (18)$$

The compositions  $\mathbf{e}_i, i = 1, \dots, D - 1$ , are orthonormal with respect to the Aitchison inner product (4) and they are a basis of  $\mathcal{S}^D$ .

*Proof:* The Aitchison inner product between  $\mathbf{e}_i, \mathbf{e}_j$  can be computed using Equation (16):

$$\langle \mathbf{e}_i, \mathbf{e}_j \rangle_a = \frac{1}{D} \mathbf{u}_i \mathbf{M} \mathbf{u}'_j = \mathbf{u}_i \mathbf{u}'_j = 0, \quad (19)$$

for  $i \neq j$  due to the orthogonality of  $\mathbf{u}_i, \mathbf{u}_j$  in  $\mathfrak{N}^D$  with respect to both the standard metric in  $\mathfrak{N}^D$  and the metric  $D^{-1} \mathbf{M}$  as stated in Proposition 1. Normalization to unity of  $\mathbf{e}_i$  follows from the same expression by taking  $i = j$ .  $\square$

Representations in  $\mathcal{S}^3$  of the axes associated with the orthonormal basis obtained are plotted as dashed lines in Figures 1(A)–5(A), with several geometric figures (squares, rectangle, circles) for illustration.

Alternative choices of  $\mathbf{v}_i$  can produce different orthonormal bases in the simplex. For instance, permutation of parts in the simplex permute the components of the basis, but we remark that different choices of  $\mathbf{v}_i$  may lead to the same orthogonal basis. See e.g. Barceló-Vidal, Martín-Fernández, and Pawlowsky-Glahn (2001) and Pawlowsky-Glahn (2002).

### Orthonormal Bases Associated to a Partition

Although Equation (18) gives an explicit Aitchison-orthonormal basis in  $\mathcal{S}^D$ , its meaning is hardly intuitive. An alternative and interesting example is obtained whenever one tries to associate an orthonormal basis with a partition of the compositional vector  $\mathbf{x}$  into two subcompositions,  $\mathbf{x} = (\mathbf{r}, \mathbf{s})$ , where  $\mathcal{C}(\mathbf{r}) \in \mathcal{S}^r, \mathcal{C}(\mathbf{s}) \in \mathcal{S}^s, D = r + s, r \geq 2, s \geq 2$ . Note that for  $r = D - 1$  we get  $s = 1$ , which corresponds to a degenerate case in the sense that we have a subcomposition with only one part. In the nondegenerate case, we look for an orthonormal basis such that the compositions of the form

$$\mathcal{C}(\mathbf{r}, \mathbf{c}_s) = \left[ x_1, x_2, \dots, x_r, \underbrace{c, c, \dots, c}_{s \text{ components}} \right],$$

$$\mathcal{C}(\mathbf{c}_r^*, \mathbf{s}) = \left[ \underbrace{c^*, c^*, \dots, c^*}_{r \text{ components}}, x_{r+1}, x_{r+2}, \dots, x_D \right],$$

can be expressed by using  $r - 1$  elements of the orthonormal basis, denoted by  $\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_{r-1}$  for the first one, and  $s - 1$  elements, denoted by  $\mathbf{h}_{r+1}, \mathbf{h}_{r+2}, \dots, \mathbf{h}_{D-1}$ , for the second one. In this way, we associate the first  $r - 1$  elements of the basis with the subcomposition  $\mathcal{C}(\mathbf{r})$  and the last  $s - 1$  with the complementary subcomposition  $\mathcal{C}(\mathbf{s})$ . To complete the orthonormal basis, we need an additional orthonormal vector, which we call *balancing element* and denote by  $\mathbf{h}_r$ ; the orthogonal projection—within the Aitchison geometry—of  $\mathbf{x}$  on this element defines the mass ratio between the subcompositions  $\mathbf{r}$  and  $\mathbf{s}$  to reconstruct  $\mathbf{x}$ . We claim that such an Aitchison-orthonormal basis can be defined as follows.

*Proposition 3.* Let  $r \geq 2, s \geq 2$ , be such that  $D = r + s$ . The compositions in  $\mathcal{S}^D$  defined as

$$\mathbf{h}_i = \mathcal{C} \left[ \exp \left( \underbrace{\left( \sqrt{\frac{1}{i(i+1)}}, \dots, \sqrt{\frac{1}{i(i+1)}} \right)}_{i \text{ elements}}, -\sqrt{\frac{i}{i+1}}, 0, \dots, 0 \right) \right], \quad (20)$$

for  $i = 1, 2, \dots, r - 1$ ,

$$\mathbf{h}_r = \mathcal{C} \left[ \exp \left( \underbrace{\left( \sqrt{\frac{s}{Dr}}, \dots, \sqrt{\frac{s}{Dr}} \right)}_{r \text{ elements}}, \underbrace{\left( -\sqrt{\frac{r}{Ds}}, \dots, -\sqrt{\frac{r}{Ds}} \right)}_{s \text{ elements}} \right) \right]; \quad (21)$$

and, for  $j = 1, 2, \dots, s - 1$ ,

$$\mathbf{h}_{D-j} = \mathcal{C} \left[ \exp \left( 0, \dots, 0, -\sqrt{\frac{j}{j+1}}, \underbrace{\left( \sqrt{\frac{1}{j(j+1)}}, \dots, \sqrt{\frac{1}{j(j+1)}} \right)}_{j \text{ elements}} \right) \right], \quad (22)$$

are Aitchison-orthonormal and constitute a basis of  $\mathcal{S}^D$  associated with a partition into two orthogonal subcompositions with, respectively,  $r$  and  $s$  components.

*Proof:* Elements  $\mathbf{h}_i$  in Equation (20) are orthonormal due to Proposition 2. The same argument applies to  $\mathbf{h}_{D-j}$  in Equation (22) after a reversion of components. We readily check orthogonality of the two sets of compositions by carrying out the inner product defined in Equation (16). Also the balancing  $\mathbf{h}_r$ , is shown to have Aitchison-norm equal to 1 and to be orthogonal to the other elements by using Equation (16).  $\square$

Proposition 3 is still valid in the degenerate case  $r = D - 1, s = 1$ . The orthonormal basis consists then of those vectors given in Equations (20) and (21), whereas the vectors given by Equation (22) do not appear at all. Note that the orthogonal bases represented in Figures 1(A)–5(A) correspond to the degenerate case  $r = 2, s = 1$ , the straight dashed axis in an Euclidean sense corresponding to the balancing element and the other axis associated to the subcomposition  $(x_1, x_2)$ .

The result can be easily generalized to more than two subcompositions. Thus, the main consequence of Proposition 3 is that any  $\mathbf{x} \in \mathcal{S}^D$ ,  $D > 3$ , can always be decomposed into three orthogonal parts: two of them in orthogonal subspaces associated with two subcompositions and a balancing one. This was stated as an intuitive requirement of orthogonality in the simplex in the preceding section. This decomposition is

$$\mathbf{x} = \mathcal{C}(\mathbf{r}, \mathbf{c}_s) \oplus \mathcal{C}(\mathbf{c}_r^*, \mathbf{s}) \oplus (\langle \mathbf{x}, \mathbf{h}_r \rangle_a \otimes \mathbf{h}_r),$$

$$\mathcal{C}(\mathbf{r}, \mathbf{c}_s) = \bigoplus_{i=1}^{r-1} (\langle \mathbf{x}, \mathbf{h}_i \rangle_a \otimes \mathbf{h}_i), \quad \mathcal{C}(\mathbf{c}_r^*, \mathbf{s}) = \bigoplus_{j=1}^{s-1} (\langle \mathbf{x}, \mathbf{h}_{D-j} \rangle_a \otimes \mathbf{h}_{D-j}),$$

where the symbol  $\oplus$  stands for repeated perturbation.

## ISOMETRIC LOGRATIO TRANSFORMATION

An isometry between  $\mathcal{S}^D$  and  $\mathfrak{N}^{D-1}$  can be obtained in a standard way if an Aitchison-orthonormal basis,  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{D-1}$ , is given. In fact, we look for a transformation, that we call isometric logratio transformation (ilr), such that  $\text{ilr}(\mathbf{e}_i) = \vec{e}_i$  for  $i = 1, 2, \dots, D - 1$ ;  $\vec{e}_i$  being the  $i$ th vector of the canonical basis in  $\mathfrak{N}^{D-1}$ . Therefore, the desired transformation is defined as follows:

*Definition 1.* For any composition  $\mathbf{x} \in \mathcal{S}^D$ , the isometric logratio (ilr) transformation associated to an Aitchison-orthonormal basis in  $\mathcal{S}^D$ ,  $\mathbf{e}_i, i = 1, 2, \dots, D - 1$ , is the transformation from  $\mathcal{S}^D$  to  $\mathfrak{N}^{D-1}$  given by

$$\mathbf{y} = \text{ilr}(\mathbf{x}) = [\langle \mathbf{x}, \mathbf{e}_1 \rangle_a, \langle \mathbf{x}, \mathbf{e}_2 \rangle_a, \dots, \langle \mathbf{x}, \mathbf{e}_{D-1} \rangle_a]. \quad (23)$$



The real vector  $\text{ilr}(\mathbf{x}) \in \Re^{D-1}$  can be identified with the coordinates of  $\mathbf{x} \in \mathcal{S}^D$  with respect to the basis  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{D-1}$ . It can also be identified with coordinates with respect to the canonical basis in  $\Re^{D-1}$ , leading to the equivalent expression:  $\text{ilr}(\mathbf{x}) = \sum \langle \mathbf{x}, \mathbf{e}_i \rangle_a \vec{e}_i$ . Furthermore, from Equation (23) and orthonormality of the basis in  $\mathcal{S}^D$ , we obtain the desired property  $\text{ilr}(\mathbf{e}_i) = \vec{e}_i$  for all  $i$ . Then, the inverse  $\text{ilr}$  transformation corresponds to the expression of  $\mathbf{x}$  in the reference basis of  $\mathcal{S}^D$

$$\mathbf{x} = \text{ilr}^{-1}(\mathbf{y}) = \bigoplus_{i=1}^{D-1} (\langle \mathbf{y}, \vec{e}_i \rangle \otimes \mathbf{e}_i), \quad (24)$$

where  $\langle \mathbf{y}, \vec{e}_i \rangle = \langle \mathbf{x}, \mathbf{e}_i \rangle_a = y_i$ . From definition (23) isomorphism holds, i.e. if  $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{S}^D$ ; and  $\mathbf{y}_1 = \text{ilr}(\mathbf{x}_1)$ ,  $\mathbf{y}_2 = \text{ilr}(\mathbf{x}_2)$ , then

$$\text{ilr}(\mathbf{x}_1 \oplus \mathbf{x}_2) = \mathbf{y}_1 + \mathbf{y}_2, \quad \text{ilr}(\mathbf{x}_1 \otimes \mathbf{x}_2) = \mathbf{y}_1 \otimes \mathbf{y}_2.$$

Also isometry properties of the  $\text{ilr}$  transformation are derived from definition (23). The Aitchison inner product is transformed into the ordinary Euclidean one

$$\langle \mathbf{x}_1, \mathbf{x}_2 \rangle_a = \langle \mathbf{y}_1, \mathbf{y}_2 \rangle = \sum_{i=1}^{D-1} y_{1i} y_{2i},$$

where  $y_{ji}$  is the  $i$ th component of row vector  $\mathbf{y}_j$  and  $\langle \cdot, \cdot \rangle$  stands for the usual Euclidean inner product in  $\Re^{D-1}$ . Aitchison norms and distances are translated into ordinary ones in  $\text{ilr}$ -space,

$$\|\mathbf{x}_j\|_a = \|\mathbf{y}_j\|, \quad d_a(\mathbf{x}_1, \mathbf{x}_2) = \|\mathbf{y}_1 - \mathbf{y}_2\|,$$

where  $\|\mathbf{y}_j\|^2 = \sum_i y_{ji}^2$  is the ordinary Euclidean norm of a vector in  $\Re^{D-1}$ .

The explicit expression of the  $\text{ilr}$  transformation is an important interpretative and computational tool. In order to obtain it, let us define

$$\mathbf{h} = \mathcal{C} \left[ \exp \left( \underbrace{\frac{1}{r}, \dots, \frac{1}{r}}_{r \text{ elements}}, \underbrace{-\frac{1}{s}, \dots, -\frac{1}{s}}_{s \text{ elements}}, \underbrace{0, \dots, 0}_{t \text{ elements}} \right) \right], \quad r + s + t = D.$$

The square norm of such a composition is  $\|\mathbf{h}\|_a^2 = (r + s)/rs$ , as obtained using

Equation (16). From Equation (16), and after some simple algebra we obtain

$$\frac{\langle \mathbf{x}, \mathbf{h} \rangle_a}{\|\mathbf{h}\|_a} = \sqrt{\frac{rs}{r+s}} \ln \left[ \frac{g(x_1, \dots, x_r)}{g(x_{r+1}, \dots, x_{r+s})} \right] \quad (25)$$

where  $g(\cdot)$  denotes geometric mean. The elements of the bases (18) and (20)–(22) have the form  $\|\mathbf{h}\|_a^{-1} \otimes \mathbf{h}$ ; then, ilr transformations with respect to these bases can be expressed using Equation (25). For instance, the components of  $\mathbf{y} = \text{ilr}(\mathbf{x})$  with respect to the basis (18) are

$$y_i = \sqrt{\frac{i}{i+1}} \ln \left[ \frac{g(x_1, \dots, x_i)}{x_{i+1}} \right], \quad i = 1, 2, \dots, D-1,$$

i.e. the ilr coordinates are logratios.

The definition of the ilr transformation applies to any Aitchison-orthonormal basis in  $\mathcal{S}^D$ . Particularly, this basis can be the one defined in Equation (18) in the previous section. Other bases can be used and they can be obtained from Equation (18) by using unitary transformations in  $\mathcal{S}^D$ . But unitary transformations in  $\mathcal{S}^D$  are easily obtained taking into account that ilr is actually an isometry. To see that this is so, let  $\mathbf{P}$  be the matrix associated with an unitary transformation in  $\mathfrak{R}^{D-1}$  (i.e.  $\mathbf{P}\mathbf{P}' = \mathbf{P}'\mathbf{P} = \mathbf{I}$  where  $\mathbf{I}$  stands for the identity matrix in  $\mathfrak{R}^{D-1}$ ). It transforms the canonical basis into another orthonormal basis  $\vec{h}_i = \vec{e}_i\mathbf{P}$ ,  $i = 1, 2, \dots, D-1$ . The inverse ilr transformation gives the transformed orthonormal basis in  $\mathcal{S}^D$ .

## RELATIONSHIP BETWEEN TRANSFORMATIONS

Since alr and clr transformation have been extensively used in compositional data analysis, the relationship between them and the ilr family of transformations may be useful. Given that Aitchison-orthonormal bases are related through unitary transformations, it suffices to work with one of them. Therefore, we will work in what follows with a generic orthonormal basis  $\mathbf{e}_i = \mathcal{C}(\exp \mathbf{u}_i)$  in  $\mathcal{S}^D$ . Particularly,  $\mathbf{u}_i$  can be taken as the basis (17) in  $\mathfrak{R}^{D-1}$  or  $\mathbf{e}_i$  as (18) in  $\mathcal{S}^D$ . A first important aspect is the fact that  $\text{clr}(\mathbf{e}_i) = \mathbf{u}_i$ . This can be easily checked by using the definition of the clr transformation (5). A second aspect, previously stated by Aitchison (1986), is that, for any  $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{S}^D$  and  $1, 2 \in \mathfrak{R}$ ,

$$\text{clr}((1 \otimes \mathbf{x}_1) \oplus (2 \otimes \mathbf{x}_2)) = 1 \text{clr}(\mathbf{x}_1) + 2 \text{clr}(\mathbf{x}_2). \quad (26)$$

Now, let  $\mathbf{x} \in \mathcal{S}^D$  and consider  $\text{ilr}(\mathbf{x}) = [y_1, y_2, \dots, y_{D-1}]$ . Then

$$\mathbf{x} = \bigoplus_{k=1}^{D-1} (y_k \otimes \mathbf{e}_k),$$

and its clr transformation is

$$\text{clr}(\mathbf{x}) = \sum_{k=1}^{D-1} y_k \text{clr}(\mathbf{e}_k) = \sum_{k=1}^{D-1} y_k \mathbf{u}_k = \text{ilr}(\mathbf{x})\mathbf{U}, \quad (27)$$

where the  $(D-1) \times D$  matrix  $\mathbf{U}$  has the vectors  $\text{clr}(\mathbf{e}_i) = \mathbf{u}_i$  as row vectors. Equation (27) allows us to go from the ilr representation of  $\mathbf{x}$  to its representation by means of clr. The inverse relationship is not easily obtained from Equation (27) because it requires a generalized inversion of the rectangular matrix  $\mathbf{U}$ . However, it has been implicitly found in Equation (16). In fact,

$$\text{ilr}(\mathbf{x}) = [\langle \mathbf{x}, \mathbf{e}_1 \rangle_a, \langle \mathbf{x}, \mathbf{e}_2 \rangle_a, \dots, \langle \mathbf{x}, \mathbf{e}_{D-1} \rangle_a] = \frac{1}{D} \text{clr}(\mathbf{x})\mathbf{M}\mathbf{U}' = \text{clr}(\mathbf{x})\mathbf{U}'. \quad (28)$$

In order to determine the relationship between the ilr transformation and the alr transformation—with respect to the last component—let us recall its definition (Eq. (6)).

The relationship between alr and clr is well known (Aitchison, 1986) and it is given by

$$\text{alr}(\mathbf{x}) = \text{clr}(\mathbf{x})\mathbf{F}, \quad \mathbf{F}' = [\mathbf{I}_{D-1} : -\mathbf{1}'_{D-1}], \quad (29)$$

where  $\mathbf{I}_{D-1}$  is the identity matrix of dimension  $(D-1)$  and  $\mathbf{1}_{D-1}$  is a  $(D-1)$  row vector of units. The inverse relationship between clr and alr can be expressed as

$$\text{clr}(\mathbf{x}) = \text{alr}(\mathbf{x})\mathbf{A}, \quad (30)$$

where the  $(D-1) \times D$  matrix  $\mathbf{A}$  is the Moore–Penrose generalized inverse of matrix  $\mathbf{F}$ ; that is

$$\mathbf{A} = \frac{1}{D} \begin{pmatrix} D-1 & -1 & -1 & \dots & -1 & -1 \\ -1 & D-1 & -1 & \dots & -1 & -1 \\ -1 & -1 & D-1 & \dots & -1 & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -1 & -1 & -1 & \dots & D-1 & -1 \end{pmatrix}.$$

Therefore, using Equations (27), (28), (29), and (30), we have

$$\text{alr}(\mathbf{x}) = \text{ilr}(\mathbf{x})\mathbf{UF}, \quad \text{ilr}(\mathbf{x}) = \text{alr}(\mathbf{x})\mathbf{AU}'.$$

Above results can be summarized as follows.

*Proposition 4.* Let  $\mathbf{x}$  be any composition in  $\mathcal{S}^D$  and consider an orthonormal basis  $\mathbf{e}_i = \mathcal{C}(\exp \mathbf{u}_i)$ ,  $i = 1, \dots, D - 1$ ,  $\mathbf{e}_i \in \mathcal{S}^D$ ,  $\mathbf{u}_i \in \mathfrak{R}^D$ . The logratio transformations,  $\text{ilr}(\mathbf{x})$ —with respect to the basis (Eq. 18)— $\text{clr}(\mathbf{x})$  and  $\text{alr}(\mathbf{x})$ —with respect to the last component—are related by the following formulae

$$\begin{aligned} \text{clr}(\mathbf{x}) &= \text{ilr}(\mathbf{x})\mathbf{U}, & \text{alr}(\mathbf{x}) &= \text{ilr}(\mathbf{x})\mathbf{UF}, \\ \text{ilr}(\mathbf{x}) &= \text{clr}(\mathbf{x})\mathbf{U}' = \text{alr}(\mathbf{x})\mathbf{AU}'. \end{aligned}$$

## DISCUSSION

Geometry in the simplex—the sample space of compositional data—has been developed in the last 15 years mainly based on the contributions due to J. Aitchison. The main goal was to develop analytical tools for the statistical analysis of this type of data. More recent contributions have identified the simplex as an Euclidean linear space, being the perturbation and the power transformation, the internal operation and the product by scalars of the linear space. Particularly, the association of the Aitchison distance with a norm and an inner product has been made evident. Once these geometrical elements for the simplex are available, further efforts in interpretation and practical application of Aitchison geometry in the simplex are required, just clarifying the way for more complex statistical models within the logratio approach.

Parallelism and orthogonality are properties of primary interest in Euclidean spaces. This motivated the interpretative results in the section on Exponential Growth and Decay of Mass as Compositional Processes, where the processes of exponential growth or decay of mass have been interpreted as linear in the simplex, i.e. as straight lines in the sense of Aitchison geometry. Parallelism and orthogonality are also briefly studied.

The main theoretical result is the presentation of an explicit orthonormal basis in the simplex. An orthonormal basis allows us to represent any element of the simplex, a composition, by its coordinates in orthogonal axes as we normally do in real spaces. We only need to project the composition on each element of the basis by using the inner product. This representation, associated to the particular orthonormal basis, may be viewed as a transformation that assigns to each element in the simplex its coordinates; this transformation is a bijection and is isometric, i.e. Aitchison distances and angles in the simplex are transformed into ordinary

Euclidean distances and angles in the real space of the coordinates. Moreover, the expression of the coordinates are log-contrasts, which motivates the name of *isometric logratio transformation*, *ilr*, for this representation.

The *ilr* coordinate expression of the elements of the simplex allows an effective and practical handling of compositional data. Once an orthonormal basis is selected, compositional data can be *ilr*-transformed, i.e. represented by their coordinates, and then analyzed as a standard multivariate data set in real space.

The classical transformations *clr* (centered logratio) and *alr* (additive logratio) have been used in a similar way to analyze compositional data without zero parts: they are bijections, thus allowing the representation of compositions in real space. However, the image of  $S^D$  using the *clr* transformation, which is isometric, remains constrained to a subspace; and the *alr* transformation, leading to a representation free of constraints, is not isometric. The only difficulty in using *ilr* representation is the selection of the orthonormal basis of reference.

However, the selection of a particular orthonormal basis may help interpretation of the *ilr* coordinates. The decomposition of the vector space  $S^D$  into orthogonal subspaces associated with nonoverlapping subcompositions is a particular but important result related to this issue. In fact, the presented result identifies a given subcomposition with an orthogonal projection in the simplex. It also gives the key to join compositions with different parts into a single composition by using a *balancing element*. This is an important point to define independence of random compositions.

Principal component analysis and singular value decomposition on the *ilr* coordinates (also on *clr*-transformed data) of compositional data without zero parts can also be interpreted as a particular selection of an orthonormal basis, which allows us dimension reduction and straightforward interpretations in data analysis.

## ACKNOWLEDGMENTS

This research has been supported by the Dirección General de Enseñanza Superior (DGES) of the Spanish Ministry for Education and Culture through the project BFM2000-0540.

The authors thank J. Aitchison for his encouragement and challenging discussion. The authors also thank the reviewers D. Billheimer and R. M. Renner for their constructive and valuable advising.

## REFERENCES

- Aitchison, J., 1986, The statistical analysis of compositional data: Monographs on statistics and applied probability: Chapman & Hall Ltd., London, 416 p.
- Aitchison, J., 1989, Measures of location of compositional data sets: Math. Geol., v. 21, no. 7, p. 787–790.

- Aitchison, J., 1990, Relative variation diagrams for describing patterns of compositional variability: *Math. Geol.*, v. 22, no. 4, p. 487–511.
- Aitchison, J., 1992, On criteria for measures of compositional difference: *Math. Geol.*, v. 24, no. 4, p. 365–379.
- Aitchison, J., 1997, The one-hour course in compositional data analysis or compositional data analysis is simple, *in* V. Pawlowsky-Glahn, ed., Proceedings of the third annual conference of the International Association for Mathematical Geology, Vol. I, II and addendum: International Center for Numerical Methods in Engineering (CIMNE), Barcelona, p. 3–35.
- Aitchison, J., 2002, Simplicial inference, *in* M. A. G. Viana and D. S. P. Richards, eds., Contemporary Mathematics Series, Vol. 287: Algebraic Methods in Statistics and Probability: American Mathematical Society, Providence, Rhode Island, p. 1–22.
- Aitchison, J., Barceló-Vidal, C., Martín-Fernández, J. A., and Pawlowsky-Glahn, V., 2000, Logratio analysis and compositional distance: *Math. Geol.* v. 32, no. 3, p. 271–275.
- Aitchison, J., and Thomas, C. W., 1998, Differential perturbation processes: A tool for the study of compositional processes, *in* A. Buccianti, G. Nardi, and R. Potenza, eds., Proceedings of the fourth annual conference of the International Association for Mathematical Geology, Vol. I and II: De Frede Editore, Napoli, p. 499–504.
- Barceló-Vidal, C., Martín-Fernández, J. A., and Pawlowsky-Glahn, V., 2001, Mathematical foundations of compositional data analysis: in G. Ross, ed., Proceedings of the sixth annual conference of the International Association for Mathematical Geology Cancun (Mexico) CD-Rom.
- Berberian, S. K., 1961, Introduction to Hilbert space: Oxford University Press, New York, NY, 206 p.
- Billheimer, D., Guttorp, P., and Fagan, W. F., 2001, Statistical interpretation of species composition: *J. Am. Stat. Assoc.*, v. 96, no. 456, p. 1205–1214.
- Martín-Fernández, J. A., Olea-Meneses, R. A., and Pawlowsky-Glahn, V., 2001, Criteria to compare estimation methods of regionalized compositions: *Math. Geol.*, v. 33, no. 8, p. 889–909.
- Martín-Fernández, J. A., Barceló-Vidal, C., and Pawlowsky-Glahn, V., 2003, Dealing with zeros and missing values in compositional data sets using nonparametric imputation: *Math. Geol.*, v. 35, no. 3, p. 253–278.
- Pawlowsky-Glahn, V., 2002, Lecture notes on statistical analysis of compositional data: Freie Universität Berlin.
- Pawlowsky-Glahn, V., and Egozcue, J. J., 2001, Geometric approach to statistical analysis on the simplex: *Stoch. Environ. Res. Risk Assess.*, v. 15, no. 5, p. 384–398.
- Pawlowsky-Glahn, V., and Egozcue, J. J., 2002, BLU estimators and compositional data: *Math. Geol.*, v. 34, no. 3, p. 259–274.
- von Eynatten, H., Pawlowsky-Glahn, V., and Egozcue, J. J., 2002, Understanding perturbation on the simplex: A simple method to better visualise and interpret compositional data in ternary diagrams: *Math. Geol.*, v. 34, no. 3, p. 249–275.