DIFFERENTIAL CALCULUS ON THE SIMPLEX

C. Barceló-Vidal, J. A. Martín-Fernández
Depart. d'Informàtica i Matemàtica Aplicada
Universitat de Girona - *e-mail*: carles.barcelo@udg.es

Summary. The purpose of this paper is to introduce the concept of differentiability of a vector-valued function on the simplex. In particular, the concepts of compositional gradient and compositional directional derivative of a real-valued function on the simplex are exposed and discussed.

1. Introduction

In some cases a response variable y is assumed to depend only on the proportions $x_1, \ldots, x_D$ of D ingredients or parts present in a specific mixture and not on the amount of the mixture. These proportions are often expressed by volume, by weight, by mole fraction, etc. In mathematical terms the response variable y can be interpreted as a real or vector-valued function φ whose domain is a subset of the simplex space.

In many practical situations the expression $y = \varphi(x_1, \ldots, x_D)$ is unknown and the emphasis is on fitting the simplest model to sample or experimental data. However, in some cases, the function φ can be deduced from physical laws. One example is the rubidium-strontium method of dating *Rb*-minerals —based on the law of radioactivity— which uses the following function to determine the "age" t of a mineral:

$$t = \frac{1}{\lambda} \log \left[\frac{\frac{^{87}Sr}{^{86}Sr} - \xi_0}{\frac{^{87}Rb}{^{86}Sr}} + 1 \right], \tag{1}$$

where ^{86}Sr, ^{87}Sr and ^{87}Rb are, respectively, the total number of atoms of these isotopes in a unit weight of the mineral at the present time; ξ_0 is the $^{87}Sr/^{86}Sr$ ratio of strontium that was incorporated into the mineral at the time of its formation; λ is the decay constant of ^{87}Rb in units of reciprocal years ($\lambda = 1.42 \times 10^{-11}$ y^{-1}); and t is the time elapsed in years since the time of formation of mineral, that is, the "age" of the mineral. The number of ^{86}Sr atoms is constant because this isotope is stable and not produced by decay of a naturally occurring isotope of another element. The function (1) is the basis for age determination by the *Rb*-*Sr* method when the mineral has remained a "closed system" with respect to rubidium and strontium, and when the assumed value ξ_0 of the initial $^{87}Sr/^{86}Sr$ ratio is appropriate. From a mathematical point of view the expression (1) is a real-valued function defined on a subset in the simplex $\mathcal{S}^3$ of 3-parts ^{86}Sr-^{87}Sr-^{87}Rb.

In the next sections we introduce the main topics on differential calculus of the vector-valued functions on the compositional space. We suppose known the terminology associated to the metric vector space structure of the compositional space $\mathcal{C}^d$ introduced in Barceló-Vidal et al. (2001) and the notation introduced in Aitchison et al. (2002).

2. Vector-valued functions on $\mathcal{C}^d$

2.1 *Scale-invariant vector-valued functions on $\mathbb{R}^D_+$*

Let f be a vector-valued function whose domain is a subset A in $\mathbb{R}^D_+$ and with range contained in $\mathbb{R}^m$. Thus f assigns each $\mathbf{w}$ in A a value $f(\mathbf{w}) = (f_1(\mathbf{w}), \ldots, f_m(\mathbf{w}))'$, an m-tuple in $\mathbb{R}^m$.

We say f is *scale-invariant* —or *homogeneus of degree* 0— if $f(k\mathbf{w}) = f(\mathbf{w})$ for every positive real k, and for every $\mathbf{w}$ in A for which $k\mathbf{w}$ in A. Obviously, if a vector-valued function $f(\mathbf{w}) = (f_1(\mathbf{w}), \ldots, f_m(\mathbf{w}))'$, with domain $A \subset \mathbb{R}^D_+$, is scale-invariant, its components $f_1, \ldots, f_m$ are also real-valued scale-invariant functions on A. And inversely, if all the components $f_1, \ldots, f_m$ are real-valued scale-invariant functions, the vector-valued function $f(\mathbf{w}) = (f_1(\mathbf{w}), \ldots, f_m(\mathbf{w}))'$ is scale-invariant.

Let A be a subset of $\mathbb{R}^D_+$. We define the *scale closure* of A, denoted A^*, as the set

$$A^* = \{k\mathbf{w} : \ k > 0, \mathbf{w} \in A\}.$$

The set A is said to be *scale-closed*, if $A^* = A$.

It is clear that any scale-invariant vector-valued function f with domain $A \subset \mathbb{R}^D_+$ can be extended to the scale closure A^* of A. It suffices to define $f(k\mathbf{w}) = f(\mathbf{w}) \quad (k > 0)$ $(\mathbf{w} \in A)$. Obviously, this extended function with domain A^* is also scale-invariant. Therefore, throughout this paper, we shall implicitly suppose that the domain of any scale-invariant function is always an scale-closed subset.

Let $f : A \longrightarrow \mathbb{R}^m$ be an scale-invariant vector-valued function defined on a subset A in $\mathbb{R}^D_+$ with values in $\mathbb{R}^m$. It induces a vector-valued function $\underline{f} : \mathcal{A} \longrightarrow \mathbb{R}^m$ defined on $\mathcal{A} = \mathrm{ccl}A$ in $\mathcal{C}^d$. Certainly, if $\mathbf{c}$ is a composition which belongs to $\mathcal{A}$, there exists at least a vector $\mathbf{w}$ in A such that $\mathrm{ccl}\mathbf{w} = \mathbf{c}$. Then, we define $\underline{f}(\mathbf{c})$ as equal to $f(\mathbf{w})$. As f is scale-invariant, it is clear that $\underline{f}(\mathbf{c})$ is univocally defined. We will denote by $\underline{f}$ the vector-valued function defined on $\mathcal{A} = \mathrm{ccl}A$ in $\mathcal{C}^d$ associated to the scalar invariant vector-function f defined on the scalar-closed subset A in $\mathbb{R}^D_+$. Inversely, if $\varphi : \mathcal{A} \longrightarrow \mathbb{R}^m$ is a vector-valued function defined on a set $\mathcal{A}$ in $\mathcal{C}^d$ with values in $\mathbb{R}^m$, it can be proved that there exist an unique scale-invariant function f defined on an scale-closed set A in $\mathbb{R}^D_+$, such that $\underline{f} = \varphi$. Therefore we can suppose that any vector-valued function defined on a set in $\mathcal{C}^d$ is derived from an scale-invariant vector-valued function defined on an scale-closed set in $\mathbb{R}^D_+$.

2.2 *$\mathcal{C}$-linear functions on $\mathcal{C}^d$*

According to the vector space structure of $(\mathcal{C}^d, \oplus, \otimes)$, a vector-valued function $\varphi : \mathcal{C}^d \longrightarrow \mathbb{R}^m$ is *$\mathcal{C}$-linear* if

$$\varphi(\underline{\mathbf{w}} \oplus \underline{\mathbf{w}^*}) = \varphi(\underline{\mathbf{w}}) + \varphi(\underline{\mathbf{w}^*}) \quad (\underline{\mathbf{w}}, \underline{\mathbf{w}^*} \in \mathcal{C}^d),$$

and

$$\varphi(\lambda \otimes \underline{\mathbf{w}}) = \lambda\varphi(\underline{\mathbf{w}}) \quad (\underline{\mathbf{w}} \in \mathcal{C}^d) \ (\lambda \in \mathbb{R}).$$

It can be proved that, φ is *$\mathcal{C}$-linear* if and only if there exists a constant $m \times D$ matrix $\mathbf{A}$, constrained by the equality $\mathbf{A}\mathbf{1}_D = \mathbf{0}_m$, such that $\underline{f} = \varphi$, where $f : \mathbb{R}^D_+ \longrightarrow \mathbb{R}^m$ is the real-valued function defined by the expression

$$f(\mathbf{w}) = \mathbf{A} \log \mathbf{w} \quad (\mathbf{w} \in \mathbb{R}^D_+).$$

3. Derivatives on $\mathcal{C}^d$

3.1 Derivative of scale-invariant vector-valued functions on $\mathbb{R}^D_+$

Let $f = (f_1, \ldots, f_m) : A \longrightarrow \mathbb{R}^m$ be a vector-valued function whose domain is an open subset A in $\mathbb{R}^D_+$. If f is differentiable at $\mathbf{w}^* \in A$, we will denote by

$$\mathbf{D}f(\mathbf{w}^*) = \begin{bmatrix} \frac{\partial f_1}{\partial w_1}(\mathbf{w}^*) & \ldots & \frac{\partial f_1}{\partial w_D}(\mathbf{w}^*) \\ \vdots & \ldots & \vdots \\ \frac{\partial f_m}{\partial w_1}(\mathbf{w}^*) & \ldots & \frac{\partial f_m}{\partial w_D}(\mathbf{w}^*) \end{bmatrix}$$

the *derivative* of f at $\mathbf{w}^*$. Therefore, this derivative is a $m \times D$ matrix referred to as the *matrix of partial derivatives of f at $\mathbf{w}^*$*. In particular, if f is a real-valued function, its derivative at $\mathbf{w}^*$

$$\mathbf{D}f(\mathbf{w}^*) = \begin{bmatrix} \frac{\partial f}{\partial w_1}(\mathbf{w}^*) & \ldots & \frac{\partial f}{\partial w_D}(\mathbf{w}^*) \end{bmatrix}$$

is called the *gradient* of f at $\mathbf{w}^*$ and denoted by $\nabla f(\mathbf{w}^*)$.

According to the well-known *Euler's theorem* for homogeneus functions, if f is scale-invariant and differentiable at $\mathbf{w}^* \in A$, then

$$\sum_{j=1}^{D} w_j^* \frac{\partial f_i}{\partial w_j}(\mathbf{w}^*) = 0$$

for each $i = 1, \ldots, m$. Furthermore, f is also differentiable at $k\mathbf{w}^*$, for each $k > 0$, and

$$\frac{\partial f_i}{\partial w_j}(k\mathbf{w}^*) = \frac{1}{k} \frac{\partial f_i}{\partial w_j}(\mathbf{w}^*)$$

for each $i = 1, \ldots, m$, and $j = 1, \ldots, D$.

3.2 Derivative of vector-valued functions on $\mathcal{C}^d$

Let $f = (f_1, \ldots, f_m) : A \longrightarrow \mathbb{R}^m$ be a scale-invariant vector-valued function whose domain is an open and scale-closed subset A in $\mathbb{R}^D_+$. Let $\underline{f} = (\underline{f_1}, \ldots, \underline{f_m})$ be the vector-valued function defined on $\mathcal{A} = \text{ccl}A$ in $\mathcal{C}^d$ associated to function f.

We say that $\underline{f}$ is *$\mathcal{C}$-differentiable* at $\underline{\mathbf{w}^*}$ in $\mathcal{A}$ if it exists a $m \times D$ matrix $\mathbf{A}$, constrained by the equality $\mathbf{A}\mathbf{1}_D = \mathbf{0}_m$, such that

$$\lim_{\underline{\mathbf{u}} \xrightarrow{\mathcal{C}} \underline{\mathbf{1}_D}} \frac{\| \underline{f}(\underline{\mathbf{w}^*} \oplus \underline{\mathbf{u}}) - \underline{f}(\underline{\mathbf{w}^*}) - \mathbf{A} \log \mathbf{u} \|}{\| \underline{\mathbf{u}} \|_{\mathcal{C}}} = 0,$$

where $\underline{\mathbf{1}_D} = \text{ccl}(1, \ldots, 1)'$ is the neutral element of $(\mathcal{C}^d, \oplus)$.

The element (i,j)-th of matrix $\mathbf{A}$ will referred to as the *$\mathcal{C}$-partial derivative* of $\underline{f}_i$ with respect to the j-th coordinate at $\underline{\mathbf{w}}^*$ and is denoted by $\frac{\partial_{\mathcal{C}} \underline{f}_i}{\partial w_j}(\underline{\mathbf{w}}^*)$. It is easy to prove that

$$\frac{\partial_{\mathcal{C}} \underline{f}_i}{\partial w_j}(\underline{\mathbf{w}}^*) = \lim_{t\to 0} \frac{f_i(w_1^*, \ldots, w_{j-1}^*, w_j^* \exp t, w_{j+1}^*, \ldots, w_D^*) - f_i(\mathbf{w}^*)}{t}.$$

From this interpretation of the $\mathcal{C}$-partial derivatives, it is not difficult to prove that $\underline{f}$ is $\mathcal{C}$-differentiable at $\underline{\mathbf{w}}^*$ in $\mathcal{A}$ if and only if f is differentiable at $\mathbf{w}^*$ in A. Furthermore, the $\mathcal{C}$-partial derivatives of $\underline{f}$ at $\underline{\mathbf{w}}^*$ and the partial derivatives of f at $\mathbf{w}^*$ are related by the equality

$$\frac{\partial_{\mathcal{C}} \underline{f}_i}{\partial w_j}(\underline{\mathbf{w}}^*) = w_j^* \frac{\partial f_i}{\partial w_j}(\mathbf{w}^*),$$

for each $i = 1, \ldots, m$, and $j = 1, \ldots, D$.

The matrix

$$\mathbf{A} = \begin{bmatrix} \frac{\partial_{\mathcal{C}} \underline{f}_1}{\partial w_1}(\underline{\mathbf{w}}^*) & \cdots & \frac{\partial_{\mathcal{C}} \underline{f}_1}{\partial w_D}(\underline{\mathbf{w}}^*) \\ \vdots & \cdots & \vdots \\ \frac{\partial_{\mathcal{C}} \underline{f}_m}{\partial w_1}(\underline{\mathbf{w}}^*) & \cdots & \frac{\partial_{\mathcal{C}} \underline{f}_m}{\partial w_D}(\underline{\mathbf{w}}^*) \end{bmatrix}$$

will be referred to as the *matrix of $\mathcal{C}$-partial derivatives of $\underline{f}$ at $\underline{\mathbf{w}}^*$* and will be denoted by $\mathbf{D}_{\mathcal{C}} \underline{f}(\underline{\mathbf{w}}^*)$.

In particular, if $\underline{f}$ is a real-valued function, the matrix of $\mathcal{C} - partial$ derivatives of $\underline{f}$ at $\underline{\mathbf{w}}^*$

$$\mathbf{D}_{\mathcal{C}} \underline{f}(\underline{\mathbf{w}}^*) = \left[\frac{\partial_{\mathcal{C}} \underline{f}}{\partial w_1}(\underline{\mathbf{w}}^*), \ldots, \frac{\partial_{\mathcal{C}} \underline{f}}{\partial w_D}(\underline{\mathbf{w}}^*)\right]$$

is called the *$\mathcal{C}$-gradient* of $\underline{f}$ at $\underline{\mathbf{w}}^*$ and denoted by $\nabla_{\mathcal{C}} \underline{f}(\underline{\mathbf{w}}^*)$. The *first-order Taylor approximation* of $\underline{f}$ in a neighbourhood of $\underline{\mathbf{w}}^*$ is given by

$$\underline{f}(\underline{\mathbf{w}}^*; \underline{\mathbf{u}}) = \underline{f}(\underline{\mathbf{w}}^*) + \sum_{j=1}^{D} \left(\frac{\partial_{\mathcal{C}} \underline{f}}{\partial w_j}(\underline{\mathbf{w}}^*)\right) \log u_j + R_1(\underline{\mathbf{w}}^*; \underline{\mathbf{u}}),$$

where $\underline{\mathbf{u}} = \mathrm{ccl}(u_1, \ldots, u_D)'$, and $R_1(\underline{\mathbf{w}}^*; \underline{\mathbf{u}}) / \parallel \underline{\mathbf{u}} \parallel_{\mathcal{C}} \to 0$ in $\mathbb{R}$ as $\underline{\mathbf{u}} \overset{\mathcal{C}}{\to} \underline{\mathbf{1}_D}$ in $\mathcal{C}^d$.

4. Directional derivatives on $\mathcal{C}^d$

Let $f : \mathbb{R}_+^D \to \mathbb{R}$ be a scale-invariant real-valued function defined on $\mathbb{R}_+^D$ and $\underline{f}$ be the real-valued function defined on $\mathcal{C}^d$ associated to function f. Let $\underline{\mathbf{w}}^*, \underline{\mathbf{u}} \in \mathcal{C}^d$.

Then, the *$\mathcal{C}$-directional derivative* of $\underline{f}$ at $\underline{\mathbf{w}}^*$ along the direction $\underline{\mathbf{u}}$ is given by

$$\left(\frac{d\underline{f}(\underline{\mathbf{w}}^* \oplus (t \otimes \underline{\mathbf{u}}))}{dt}\right)_{t=0}$$

if this exists. From this definition, it is no difficult to prove that the $\mathcal{C}$-directional derivative can also be defined by the formula

$$\lim_{t\to 0} \frac{f(w_1^* u_1^t, \ldots, w_D^* u_D^t) - f(w_1^*, \ldots, w_D^*)}{t}.$$

If f is differentiable, then all $\mathcal{C}$-directional derivatives of $\underline{f}$ exist. Furthermore, the $\mathcal{C}$-directional derivative at $\underline{\mathbf{w}^*} \in \mathcal{C}^d$ along the direction $\underline{\mathbf{u}} \in \mathcal{C}^d$ is given by

$$\sum_{j=1}^{D} w_j^* \frac{\partial f}{\partial w_j}(\mathbf{w}^*) \log u_j = \nabla_{\mathcal{C}} \underline{f}(\underline{\mathbf{w}^*}) \ \log \mathbf{u}. \tag{2}$$

Therefore, if $\nabla_{\mathcal{C}} \underline{f}(\underline{\mathbf{w}^*}) \neq \mathbf{0}_D^{'}$, the composition $\mathrm{ccl}\left\{\exp\left(\nabla_{\mathcal{C}} \underline{f}(\underline{\mathbf{w}^*})^{'}\right)\right\}$ points in the direction along which $\underline{f}$ is increasing the fastest in a $\mathcal{C}$-neighbourhood of $\underline{\mathbf{w}^*}$.

We can deduce from (2) that the $\mathcal{C}$-directional derivative of $\underline{f}$ at $\underline{\mathbf{w}^*} \in \mathcal{C}^d$ along the direction

$$\underline{\mathbf{u}_j} = \mathrm{ccl}(\underbrace{1,\ldots,1}_{j-1}, \mathrm{e}, \underbrace{1,\ldots,1}_{D-j})' \quad (j = 1,\ldots,D)$$

is equal to

$$\frac{\partial_{\mathcal{C}} \underline{f}}{\partial w_j}(\underline{\mathbf{w}^*}) = w_j^* \frac{\partial f}{\partial w_j}(\mathbf{w}^*). \tag{3}$$

If the $\mathcal{C}$-directional derivative of $\underline{f}$ along the direction $\underline{\mathbf{u}_j}$ is equal to 0 on $\mathcal{C}^d$, we say that $\underline{f}$ is $\mathcal{C}$-*independent* of the j-th part. It is clear from (3) that $\underline{f}$ is $\mathcal{C}$-independent of the j-th part if and only if f does not depend of the j-th coordinate w_j on $\mathbb{R}_+^D$.

5. Example

The function (1) giving the "age" t of a mineral from the $^{87}Sr/^{86}Sr$ and $^{87}Rb/^{86}Sr$ ratios can be interpreted as a scale-invariant real-valued function f

$$\left(^{86}Sr, ^{87}Sr, ^{87}Rb\right) \xrightarrow{f} \frac{1}{\lambda} \log \left[\frac{\frac{^{87}Sr}{^{86}Sr} - \xi_0}{\frac{^{87}Rb}{^{86}Sr}} + 1 \right]$$

defined on

$$A = \left\{ \left(^{86}Sr, ^{87}Sr, ^{87}Rb\right)' \in \mathbb{R}_+^3 : \frac{^{87}Sr}{^{86}Sr} - \xi_0 \geq 0 \right\}.$$

Therefore, the function f induces a real-valued function $\underline{f}$ defined on $\mathcal{A} = \mathrm{ccl}A$ in $\mathcal{C}^2$. This function $\underline{f}$ can also be interpreted as a function on the simplex $\mathcal{S}^3$. Figure 1 shows some level contours of this function for $\lambda = 1.42 \times 10^{-11}$ and $\xi_0 = 0.7071$. The projections of these level contours on the simplex are straight segments from a common point on the $^{86}Sr - ^{87}Sr$ edge.

The $\mathcal{C}$-gradient of $\underline{f}$ at $\underline{\mathbf{w}} = \mathrm{ccl}\,(^{86}Sr, ^{87}Sr, ^{87}Rb)'$ is equal to

$$\nabla_{\mathcal{C}} \underline{f}(\underline{\mathbf{w}}) = \frac{1/\lambda}{\left(\frac{^{87}Sr}{^{86}Sr} - \xi_0\right) + \frac{^{87}Rb}{^{86}Sr}} \left[-\xi_0, \frac{^{87}Sr}{^{86}Sr}, -\left(\frac{^{87}Sr}{^{86}Sr} - \xi_0\right) \right] \tag{4}$$

From (4) we deduce that $\frac{\partial_{\mathcal{C}} \underline{f}}{\partial\, ^{87}Sr}(\underline{\mathbf{w}})$ is always positive in each $\underline{\mathbf{w}} \in \mathcal{A}$. It means that the "age" of a mineral increases when ^{87}Sr increases and the $^{87}Rb/^{86}Sr$ ratio remains constant. Similarly, the $\mathcal{C}$-partial derivative $\frac{\partial_{\mathcal{C}} \underline{f}}{\partial\, ^{87}Rb}(\underline{\mathbf{w}})$ is negative in each $\underline{\mathbf{w}}$ in the $\mathcal{C}$-interior of $\mathcal{A}$. Therefore, the "age" of a mineral decreases when ^{87}Rb increases and the $^{87}Sr/^{86}Sr$ ratio remains constant.

Any of the three dotted paths in Figure (1) represents in each point the direction in which the values of function $\underline{f}$ change most rapidly, showing the evolution of the (^{86}Sr-^{87}Sr-^{87}Rb)-composition of a mineral from the time of its formation.

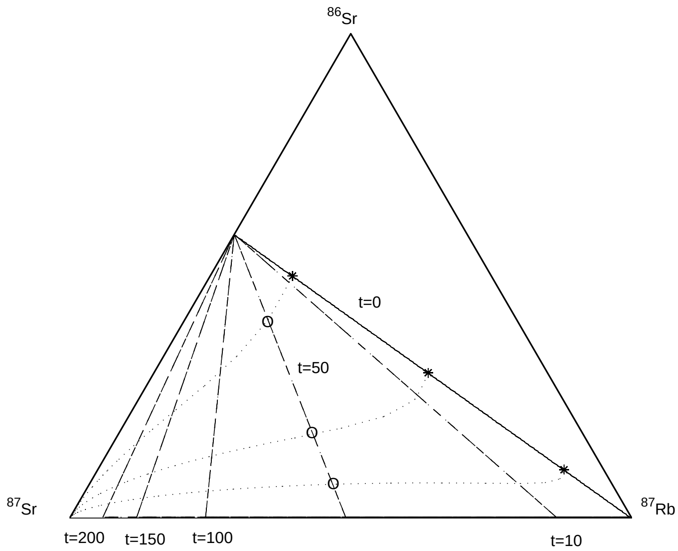

Figure 1: Level contours ($t = 0;\ 10 \times 10^9;\ 50 \times 10^9;\ 100 \times 10^9;\ 150 \times 10^9;\ 200 \times 10^9$ years) of the function $t = \frac{1}{1.42 \times 10^{-11}} \log \left[\frac{\frac{^{87}Sr}{^{86}Sr} - 0.7071}{\frac{^{87}Rb}{^{86}Sr}} + 1 \right]$ on the simplex of 3-parts ^{86}Sr-^{87}Sr-^{87}Rb. The dotted lines show the evolution of the (^{86}Sr-^{87}Sr-^{87}Rb)-composition of three minerals from the time of their formation

References

Aitchison, J., C. Barceló-Vidal, J. J. Egozcue, and V. Pawlowsky-Glahn (2002). A concise guide to the algebraic-geometric structure of the simplex, the sample space for compositional data analysis. In *Proceedings of IAMG'02*, Berlin.

Barceló-Vidal, C., J. A. Martín-Fernández, and V. Pawlowsky-Glahn (2001). Mathematical foundations of compositional data analysis. In G. Ross (Ed.), *Proceedings of IAMG'01 — The sixth annual conference of the International Association for Mathematical Geology*, Volume CD, pp. 20 p. electronic publication.