



HAL
open science

Extension of Correspondence Analysis to multiway data-sets through High Order SVD: a geometric framework

Olivier Coulaud, Alain Franc, Martina Iannacito

► **To cite this version:**

Olivier Coulaud, Alain Franc, Martina Iannacito. Extension of Correspondence Analysis to multiway data-sets through High Order SVD: a geometric framework. [Research Report] RR-9429, Inria Bordeaux - Sud-Ouest; Inrae. 2021. hal-03418404

HAL Id: hal-03418404

<https://inria.hal.science/hal-03418404>

Submitted on 7 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Inria

Extension of Correspondence Analysis to multiway data-sets through High Order SVD: a geometric framework

Olivier Coulaud, Alain Franc, Martina Iannacito

**RESEARCH
REPORT**

N° 9429

November 8, 2021

Project-Teams HiePACS and
Pleiade

ISRN INRIA/RR--9429--FR+ENG

ISSN 0249-6399



Extension of Correspondence Analysis to multiway data-sets through High Order SVD: a geometric framework

Olivier Coulaud*, Alain Franc^{†‡}, Martina Iannacito*

Project-Teams HiePACS and Pleiade

Research Report n° 9429 — November 8, 2021 — 20 pages

Abstract: This paper presents an extension of Correspondence Analysis (CA) to tensors through High Order Singular Value Decomposition (HOSVD) from a geometric viewpoint. Correspondence analysis is a well-known tool, developed from principal component analysis, for studying contingency tables. Different algebraic extensions of CA to multi-way tables have been proposed over the years, nevertheless neglecting its geometric meaning. Relying on the Tucker model and the HOSVD, we propose a direct way to associate with each tensor mode a point cloud. We prove that the point clouds are related to each other. Specifically using the CA metrics we show that the barycentric relation is still true in the tensor framework. Finally two data sets are used to underline the advantages and the drawbacks of our strategy with respect to the classical matrix approaches.

Key-words: Correspondence analysis, Tucker model, HOSVD, tensors, barycentric relation, point clouds

* HiePACS team, Inria Bordeaux-Sud-Ouest, France

† BIOGECO, INRAE, Univ. Bordeaux, 33610 Cestas, France

‡ Pleiade team - Inria\INRAE Bordeaux-Sud-Ouest, France

**RESEARCH CENTRE
BORDEAUX – SUD-OUEST**

200 avenue de la Vieille Tour
33405 Talence Cedex

Extension de l'analyse des correspondances à des données multi-dimensionnelles : un point de vue géométrique

Résumé : Ce document présente une extension de l'analyse des correspondances aux tenseurs par la décomposition en valeurs singulières d'ordre élevé (HOSVD) d'un point de vue géométrique. L'analyse des correspondances est un outil bien connu, développé à partir de l'analyse en composantes principales, pour étudier les tables de contingence. Différentes extensions algébriques de l'analyse des correspondances aux tables à voies multiples ont été proposées au fil des ans. En nous appuyant sur le modèle de Tucker et la HOSVD, nous proposons d'associer à chaque mode d'un tenseur un nuage de points. Nous établissons un lien entre les coordonnées de ces différents nuages. Une telle relation est classique en Analyse Factorielle des Correspondances (AFC) pour justifier la projection simultanée des profils lignes et profils colonnes d'une table de contingence (d'où le nom de *correspondance*). Nous étendons une telle relation barycentrique aux liens entre les nuages de points associés aux différents modes de l'Analyse Factorielle des Correspondances Multiple d'un tenseur, construite via la HOSVD avec les métriques de l'AFC.

Mots-clés : Analyse des correspondances, modèle de Tucker, HOSVD, tenseurs, relation barycentrique, nuages de points

1 Introduction

Dimension reduction has been and still is one of the most widely used method to address modeling questions on data sets living in large dimensional spaces [1, 2, with further detailed references within]. One of the most popular and used method is Principal Component Analysis (PCA) [3, 4, 5, 6, 7, 8, 9, 10, 11]. It has three guises

1. an algebraic framework, working with vector spaces and matrix algebra, based on the Singular Value Decomposition (SVD) of the data matrix, see, e.g., [12]
2. a geometric framework, by associating a point cloud with the matrix representing the data set (one point per row, one dimension per feature), see, e.g., [13, 14]
3. a statistical modeling approach, see, e.g., [15, 16].

PCA has been extended to a diversity of situations, like Correspondence Analysis (CA) [17, 18, 19, 20, 21, 22, 23] for the analysis of contingency tables with metrics associated with χ^2 distances [23, 24]. From an algebraic perspective, CA is the PCA of a contingency table with the metric associated with the inverse of row and column marginals. CA has been explained in a geometric framework as well [25, 26]. Indeed, as for PCA, in the geometric view of CA each row of a contingency table corresponds to the coordinates of a point in a specific vector space. Since a similar argument holds for both rows and columns, a contingency table is naturally associated with a two point clouds. Thanks to the underlying dimension reduction techniques, CA makes possible the simultaneous visualization and interpretation of the two point clouds in a low dimension space where a specific barycentric relation links the two objects. The dictionary between the geometric and algebraic framework relies on the selection of weights and metrics in the spaces where the point clouds live, see [13, 14].

Multiway data have appeared in a wide range of domains, requesting a further development of these analysis techniques. PCA and associated methods have been extended algebraically to dimension reduction in multiway array, see, e.g., [27, 28, 29, 30, 31, 32, 33] and many references therein, with a link with tensor algebra [34]. These are algebraic extensions, relying on numerical computations based on elementary operations in tensor algebra in the same way that PCA and CA rely on numerical linear algebra. Starting from the generalization of PCA to multiway data through the Tucker model [35], with High Order Singular Value Decomposition (HOSVD) [30], an algebraic development of CA to MultiWay Correspondence Analysis (MWCA) follows naturally. Indeed MWCA is the HOSVD of a multiway contingency table associated with metric of the marginal inverse per each mode, see [29]. Our work aims to study the geometric perspective of MWCA, which is still poorly investigated. More precisely, we show that a point cloud is associated with each mode of a tensor in MWCA in the same way that a point cloud is associated with either rows or columns of a matrix. Finally, the correspondence between two point clouds based on scaling and computing barycenters holding in the classical CA is generalized to d point clouds in MWCA.

The remainder of this paper is organized as follows. Section 2 introduces all the preliminary tensor definitions, spaces and operations, clarifying the chosen notations. In Section 3 the algebraic link between the principal components is presented. In particular, Subsection 3.3 extends the results of the previous subsections in a space with generic metric. Section 4 presents the barycentric relation characterizing Correspondence Analysis in the tensor case. Finally we highlight the different outcome on two data sets using the previous theoretical results.

2 Preliminaries

2.1 Notations

Here real numbers and integers are denoted by small Latin letters, vectors by boldface Latin letters, matrices by capital Latin letters and tensors by boldface capital Latin letters. $\mathbb{O}(m \times n)$ denotes the set of all the real orthogonal matrices of m rows and n columns. Let us have d finite dimensional vector spaces E_1, \dots, E_d on a same field \mathbb{R} . Let \mathbf{A} be a tensor in $E_1 \otimes \dots \otimes E_d$ with $\dim E_\mu = n_\mu$ for $\mu \in \{1, \dots, d\}$. The integer d is called the order of the tensor \mathbf{A} , each vector space E_μ is called a mode and n_μ is the dimension of the tensor \mathbf{A} for mode μ . We denote as well $\mathbf{A} \in \mathbb{R}^{n_1 \times \dots \times n_d}$. The matricization or unfolding of \mathbf{A} with respect to the mode μ denoted by $A^{(\mu)}$ is a matrix of n_μ rows and $n_{\neq\mu}$ columns with $n_{\neq\mu} = \prod_{\alpha \neq \mu} n_\alpha$, see [30, Definition 1]. Notice that the matricization with respect to mode μ maps the (i_1, \dots, i_d) -th tensor element to the $(i_\mu, \overline{i_1, \dots, i_{\mu-1}, i_{\mu+1}, \dots, i_d})$ -th matrix element with

$$\overline{i_1, \dots, i_{\mu-1}, i_{\mu+1}, \dots, i_d} = 1 + \sum_{\substack{\alpha=1 \\ \alpha \neq \mu}}^d (i_\alpha - 1) m_\alpha \quad \text{with} \quad m_\alpha = \prod_{\substack{\beta=1 \\ \beta \neq \mu}}^{\alpha-1} n_\beta, \quad (1)$$

see [33]. Finding an approximation of Tucker model is facilitated by using an elementary operations on tensors: the Tensor-Times-Matrix product (TTM). Let $\mathbf{A} \in E_1 \otimes \dots \otimes E_d$, and $M_\mu \in \mathcal{L}(E_\mu, E_\mu) \simeq E_\mu \otimes E_\mu$ for any mode μ . Let \mathbf{A} be an elementary tensor $\mathbf{A} = \mathbf{a}_1 \otimes \dots \otimes \mathbf{a}_d$ with $\mathbf{a}_\mu \in E_\mu$. Then, TTM product of (M_1, \dots, M_d) with \mathbf{A} , denoted $(M_1, \dots, M_d)\mathbf{A}$, is the tensor $M_1\mathbf{a}_1 \otimes \dots \otimes M_d\mathbf{a}_d$. This is extended to any tensor by linearity, see, e.g., [30]. Let \mathbf{A} be a tensor of $E_1 \otimes \dots \otimes E_d$ and let (M_1, \dots, M_d) be a d -tuple of matrices of compatible dimension with \mathbf{A} . If $\mathbf{B} = (M_1, \dots, M_d)\mathbf{A}$, then the μ -matricization of \mathbf{B} is

$$B^{(\mu)} = M_\mu A^{(\mu)} (M_d \otimes_K \dots \otimes_K M_{\mu+1} \otimes_K M_{\mu-1} \otimes_K \dots \otimes_K M_1)^T \quad (2)$$

for every $\mu \in \{1, \dots, d\}$ and, \otimes_K the Kronecker product, we refer to [36, Proposition 3.7] for further details.

Let $N_\mu = M_\mu^2$ be a Symmetric Positive Definite (SPD) matrix defining an inner product on E_μ by $\langle \mathbf{a}, \mathbf{a}' \rangle_{M_\mu} = \langle M_\mu \mathbf{a}, M_\mu \mathbf{a}' \rangle = \langle N_\mu \mathbf{a}, \mathbf{a}' \rangle = \langle \mathbf{a}, N_\mu \mathbf{a}' \rangle$. It is extended to the whole spaces by linearity. This induces an inner product on $E_1 \otimes \dots \otimes E_d$ on elementary tensors by $\langle \mathbf{a}_1 \otimes \dots \otimes \mathbf{a}_d, \mathbf{a}'_1 \otimes \dots \otimes \mathbf{a}'_d \rangle_{M_1 \otimes \dots \otimes M_d} = \langle M_1 \mathbf{a}_1 \otimes \dots \otimes M_d \mathbf{a}_d, M_1 \mathbf{a}'_1 \otimes \dots \otimes M_d \mathbf{a}'_d \rangle$. Let us denote by $\mathcal{S} = (\otimes_\mu E_\mu, \otimes_\mu \mathbb{I}_\mu)$ the tensor space $E_1 \otimes \dots \otimes E_d$ endowed with standard inner product with \mathbb{I}_μ being the identity matrix on E_μ , and by $\mathcal{S}_M = (\otimes_\mu E_\mu, \otimes_\mu M_\mu)$ the same tensor space when endowed with inner product induced by the matrices M_μ . We will use the observation that the map

$$\nu : \mathcal{S}_M \rightarrow \mathcal{S} \quad \text{with} \quad \nu(\otimes_\mu \mathbf{a}_\mu) = \otimes_\mu M_\mu \mathbf{a}_\mu$$

is an isometry, because $\|\otimes_\mu M_\mu \mathbf{a}_\mu\| = \prod_\mu \|M_\mu \mathbf{a}_\mu\| = \prod_\mu \|\mathbf{a}_\mu\|_{M_\mu} = \|\otimes_\mu \mathbf{a}_\mu\|_{M_1 \otimes \dots \otimes M_d}$ where $\|\cdot\|$ denotes the Frobenius norm. This can be extended to the whole tensor space $E_1 \otimes \dots \otimes E_d$ by linearity (see [29] for details). The isometry can be written as $\nu(\mathbf{A}) = (M_1, \dots, M_d)\mathbf{A}$.

2.2 Tucker model

The Tucker decomposition [35, 27, 37, 32, 33] of tensor $\mathbf{A} \in E_1 \otimes \dots \otimes E_d$ is

$$\mathbf{A} = \sum_{i_1=1}^{r_1} \dots \sum_{i_d=1}^{r_d} C_{i_1 \dots i_d} \mathbf{u}_{i_1}^1 \otimes \dots \otimes \mathbf{u}_{i_d}^d \quad (3)$$

where the array of the $(C_{i_1 \dots i_d})_{i_1, \dots, i_d}$ is the *core tensor* of \mathbf{A} and $(\mathbf{u}_{i_\mu}^\mu)_{i_\mu}$ is an orthonormal basis of $U_\mu \subseteq E_\mu$ with r_μ minimal for $i_\mu \in \{1, \dots, r_\mu\}$ and $\mu \in \{1, \dots, d\}$. So if \mathbf{A} belongs to $U_1 \otimes \dots \otimes U_d$ with $U_\mu \subseteq E_\mu$ and $\dim U_\mu = r_\mu$, then $\mathbf{r} = (r_1, \dots, r_d)$ is the *multilinear rank* of \mathbf{A} [30]. In a synthetic way Tucker model is expressed with the TTM product as $\mathbf{T} = (U_1, \dots, U_d)\mathbf{C}$. We identify each Tucker subspace with its orthonormal basis, denoting $U_\mu \in \mathbb{O}(n_\mu \times r_\mu)$ for every mode $\mu \in \{1, \dots, d\}$.

Starting from this Tucker model, we formulate an approximation problem as follows. Given a tensor \mathbf{A} , its best Tucker approximation at multilinear rank \mathbf{r} is the tensor \mathbf{T} of multi-linear rank \mathbf{r} such that $\|\mathbf{A} - \mathbf{T}\|$ is minimal. This is a natural extension of PCA because the unknowns are the spaces $(U_\mu)_\mu$ under constraints of dimensions. Historically speaking, finding a solution to Tucker best approximation has a long history which can be found, e.g., in [35, 29, 33, 38]. There is no known algorithm yielding the best solution of the approximation problem, although several algorithms provide good quality results, known nowadays as High Order Singular Value Decomposition (HOSVD) [30], its Truncated version (T-HOSVD) [39] and High Order Orthogonal Iterations (HOOI) [31, 33]. The seminal paper for what is now called HOOI has been published as Tuckals3 for 3-order tensors is [40] (see as well [32]). The extension to four mode tensors has been made by Lastovicka in [41], and more generally to d -order multi-arrays by a group in Groningen in 1986 [37]. These works used Kronecker product as an algebraic framework, and those results have been put into a common framework of tensor algebra in [29]. Both approaches (decomposition and best approximation) have been popularized by two papers by de Lathauwer et al. in 2000, who derived HOSVD [30] and HOOI [31] relying on matricization and matrix algebra. Matricization, called unfolding as well, is building a matrix with one mode in row, and a combination of the remaining ones in columns.

2.3 Multiway correspondence analysis

PCA is solving dimension reduction problem for a matrix in $E \otimes F$, with $E = \mathbb{R}^m$ and $F = \mathbb{R}^n$ as natural choice in data science. In the geometric context, a cloud \mathcal{A} of m points in \mathbb{R}^n is associated with a matrix A , where point i is row i of A (points are in F). PCA of A is building a new orthonormal basis in F , called principal axis, and computing the coordinates of the points in this new basis, called principal components. Classically PCA is realized with a SVD of the given matrix, i.e., $A = U\Sigma V^T$. Then the principal axis are defined as the columns of V , and the array of coordinates is given by $Y = U\Sigma$. It can be developed *mutatis mutandis* by selecting some inner products associated with SPD matrices in E and/or F . Often in data analysis, those SPD matrices are diagonal, and the metrics are defined by weights. Indeed given M in $\mathbb{R}^{m \times m}$ and Q in $\mathbb{R}^{n \times n}$, we define the inner product $\langle A, B \rangle_{M \otimes Q} = \langle MAQ, MBQ \rangle$. Then, PCA of A at rank r with this inner product is finding a rank r matrix $A_r \in \mathbb{R}^{m \times n}$ such that $\|A - A_r\|_{M \otimes Q}$ is minimal. In other words, we compute the SVD at rank r of $X = \nu_M(A) = MAQ$. If (Y, Λ, V) is the set of principal components Y of X , eigenvalues $\Lambda = \Sigma^2$ and principal axis V , then, principal components and axis of PCA of A with metrics so defined are $(M^{-1}Y, \Lambda, Q^{-1}V)$. In particular, as Figure 1a shows, CA is a PCA on a contingency table with metrics associated with inverse of the marginals as weights [13].

This is extended naturally to a multiway contingency table \mathbf{T} [32], and formalized through HOSVD. In Figure 1b we sketch this approach, which we call MWCA as it extends with HOSVD the method of CA. As preprocessing step we compute the relative frequency tensor \mathbf{F} dividing each tensor \mathbf{T} entry by the sum of all its entries. First step is to compute all marginals of \mathbf{F} for all indices for all modes. This yields a vector of weights $\mathbf{w}^\mu = (w_{i_1}^\mu, \dots, w_{i_\mu}^\mu)$ for mode μ . In the second step the isometry ν_T is defined by the diagonal matrix M_μ whose diagonal elements are inverse of the weights. Third step is to perform HOSVD of $\mathbf{X} = \nu_T(\mathbf{F})$. Finally the HOSVD

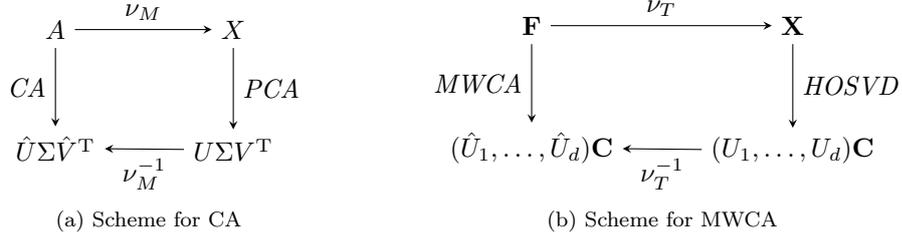


Figure 1: Correspondence analysis matrix and tensor approach

decomposition is transported back with the isometry inverse ν_T^{-1} .

3 Multiway principal components analysis

Starting from the extension of principal component analysis to tensors with HOSVD, we associate a point cloud with each mode, and show the existence of an algebraic link between them. The aim of this section is interpreting from a geometrical point of view this relation. For the sake of simplicity in the result explicit verification, we first prove a link between the point clouds in the standard Euclidean 3-order tensor space, and then we generalize to d -order tensors. This structure choice is kept throughout the document. We will focus especially on the geometric interpretation of these results. We naturally attach a point cloud to each matricization of 3-order tensor. Each point cloud is the optimal projection of the mode matricization in low dimension space. Finally, we show how their coordinates are linked and we extend this result to general metric spaces of d -order tensors.

Let $\mathbf{X} \in E_1 \otimes \dots \otimes E_d$ be a tensor with $\dim(E_\mu) = n_\mu$ and let $(U_\mu)_\mu \in \mathbb{O}(n_\mu \times r_\mu)$ be the rank $\mathbf{r} = (r_1, \dots, r_d)$ Tucker decomposition basis obtained from the HOSVD algorithm. The tensor \mathbf{X} is expressed as

$$\mathbf{X} = \sum_{i_1, \dots, i_d=1}^{r_1, \dots, r_d} C_{i_1 \dots i_d} \mathbf{u}_{i_1}^1 \otimes \dots \otimes \mathbf{u}_{i_d}^d \quad (4)$$

with \mathbf{C} the HOSVD core tensor and $\mathbf{u}_{i_\mu}^\mu$ the i_μ -th column of U_μ . The condition for Equation (4) to be a decomposition is $r_\mu = \text{rank}(X^{(\mu)})$ for $\mu \in \{1, \dots, d\}$. Let Σ_μ be the diagonal singular value matrix of the matricization of \mathbf{X} with respect to mode μ . For simplicity the i_μ -th diagonal element of Σ_μ is denoted by $\sigma_{i_\mu}^{(\mu)}$ for every $i_\mu \in \{1, \dots, r_\mu\}$ and for every $\mu \in \{1, \dots, d\}$. The principal component of mode μ is $Y_\mu \in \mathbb{R}^{n_\mu \times r_\mu}$ defined as $Y_\mu = U_\mu \Sigma_\mu$ for each $\mu \in \{1, \dots, d\}$.

3.1 The 3-order tensor case in the Euclidean space

For the sake of simplicity and clarity, we assume d equal to 3. The following proposition states a relation linking the three sets of principal coordinates.

Proposition 1. *Let \mathbf{X} be a tensor of $E_1 \otimes E_2 \otimes E_3$ with $\dim(E_\mu) = n_\mu$ and let \mathbf{C} be its HOSVD core at multi-linear rank \mathbf{r} . Let Y_μ be the principal components of mode μ . If $r_\mu \leq \text{rank}(A^{(\mu)})$ for $\mu \in \{1, 2, 3\}$, then*

$$\begin{aligned} Y_1 &= X^{(1)}(Y_3 \otimes_K Y_2)(B^{(1)})^T \\ Y_2 &= X^{(2)}(Y_3 \otimes_K Y_1)(B^{(2)})^T \\ Y_3 &= X^{(3)}(Y_2 \otimes_K Y_1)(B^{(3)})^T \end{aligned}$$

with $\mathbf{B} = (\Sigma_1^{-1}, \Sigma_2^{-1}, \Sigma_3^{-1})\mathbf{C}$.

Proof: We start the proof for the first mode principal components. Let \mathbf{X} be expressed in the HOSVD basis as in Equation (4), i.e.,

$$\mathbf{X} = \sum_{i,j,k=1}^{r_1, r_2, r_3} C_{ijk} \mathbf{u}_i^1 \otimes \mathbf{u}_j^2 \otimes \mathbf{u}_k^3.$$

Then the matricization of \mathbf{X} with respect to mode 1 in the Tucker basis is expressed with the Kronecker product as

$$X^{(1)} = \sum_{i=1}^{r_1} \mathbf{u}_i^1 \otimes \left(\sum_{j,k=1}^{r_2, r_3} C_{ijk} \mathbf{u}_k^3 \otimes_K \mathbf{u}_j^2 \right). \quad (5)$$

The PCA of $X^{(1)}$ is

$$X^{(1)} = Y_1 V_1^T = \sum_{i=1}^{r_1} \sigma_i^{(1)} \mathbf{u}_i^1 \otimes \mathbf{v}_i^1 \quad (6)$$

with \mathbf{v}_i^1 the i -th column of $V_1 \in \mathbb{O}(n_2 n_3 \times r_1)$ and $\sigma_i^{(1)} \mathbf{u}_i^1$ i -th column of $Y_1 \in \mathbb{R}^{n_1 \times r_1}$. By comparing Equations (5) and (6) for a fixed index i , we get

$$\sigma_i^{(1)} \mathbf{u}_i^1 \otimes \mathbf{v}_i^1 = \mathbf{u}_i^1 \otimes \left(\sum_{j,k=1}^{r_2, r_3} C_{ijk} \mathbf{u}_k^3 \otimes_K \mathbf{u}_j^2 \right).$$

Remarking that Σ_1 is invertible, we identify \mathbf{v}_i^1 with a linear combination of the Kronecker product of \mathbf{u}_j^2 and \mathbf{u}_k^3 scaled by $\sigma_i^{(1)}$ as

$$\mathbf{v}_i^1 = \frac{1}{\sigma_i^{(1)}} \sum_{j,k=1}^{r_2, r_3} C_{ijk} \mathbf{u}_k^3 \otimes_K \mathbf{u}_j^2. \quad (7)$$

Notice that the j -th and k -th column of $Y_2 = U_2 \Sigma_2$ and $Y_3 = U_3 \Sigma_3$ are $\sigma_j^{(2)} \mathbf{u}_j^2$ and $\sigma_k^{(3)} \mathbf{u}_k^3$ respectively. So introducing in Equation (7) the singular values $\sigma_j^{(2)}$ and $\sigma_k^{(3)}$, we express the i -th column of V_1 as a linear combination of the Kronecker product of the j -th and k -th column of Y_2 and Y_3 , i.e.

$$\begin{aligned} \mathbf{v}_i^1 &= \frac{1}{\sigma_i^{(1)}} \sum_{j,k=1}^{r_2, r_3} C_{ijk} \frac{\sigma_j^{(2)} \sigma_k^{(3)}}{\sigma_j^{(2)} \sigma_k^{(3)}} \mathbf{u}_k^3 \otimes_K \mathbf{u}_j^2 \\ &= \sum_{j,k=1}^{r_2, r_3} \frac{C_{ijk}}{\sigma_i^{(1)} \sigma_j^{(2)} \sigma_k^{(3)}} \mathbf{y}_k^3 \otimes_K \mathbf{y}_j^2 \\ &= \sum_{j,k=1}^{r_2, r_3} B_{ijk} \mathbf{y}_k^3 \otimes_K \mathbf{y}_j^2 \end{aligned}$$

with $\mathbf{B} = (\Sigma_1^{-1}, \Sigma_2^{-1}, \Sigma_3^{-1})\mathbf{C}$, \mathbf{y}_j^2 and \mathbf{y}_k^3 the j -th and k -th column of Y_2 and Y_3 respectively. Remark that $Y_3 \otimes Y_2$ is a matrix of $n_2 n_3$ rows and $r_2 r_3$ columns whose ℓ -th column is $\mathbf{y}_k^3 \otimes_K \mathbf{y}_j^2$ with $\ell = \overline{jk}$ for every $j \in \{1, \dots, r_2\}$, $k \in \{1, \dots, r_3\}$ and $\ell \in \{1, \dots, r_2 r_3\}$, as defined in Equation (1). The tensor \mathbf{B} matricized with respect to mode 1 is a matrix of r_1 rows and $r_2 r_3$ columns, whose (i, \overline{jk}) -th element is b_{ijk} for all $j \in \{1, \dots, r_2\}$, $k \in \{1, \dots, r_3\}$. So the sum in the right-hand side of Equation (3.1) can be expressed as the matrix-product between $Y_3 \otimes_K Y_2$ and tensor \mathbf{B} matricized with respect to mode 1 as

$$V_1 = (Y_3 \otimes_K Y_2)(B^{(1)})^T. \quad (8)$$

Multiplying Equation (6) on the right by V_1 yields $Y_1 = X^{(1)}V_1$. Therefore multiplying Equation (8) by the matricization of \mathbf{X} with respect to mode 1, the principal component Y_1 is expressed as linear combination of the Kronecker product of principal components Y_2 and Y_3 , i.e.

$$Y_1 = X^{(1)}V_1 = X^{(1)}(Y_3 \otimes_K Y_2)(B^{(1)})^T.$$

The other relations follow straightforwardly from this one, permuting the indices coherently. \square

From an algebraic view this first proposition shows that the principal components of each mode can be expressed as a linear combination of the principal components of the two other modes. However as pointed out in the preliminary section, principal components can be seen from different viewpoints.

From a geometric viewpoint a point cloud \mathcal{X}_μ is attached to each mode μ matricization of tensor \mathbf{X} for $\mu \in \{1, 2, 3\}$. Indeed the i_μ -th row of $X^{(\mu)}$ represents the coordinates of the i_μ -th element of mode μ point cloud living in the space $\mathbb{R}^{n_{\neq\mu}}$ where $n_{\neq\mu} = \frac{n_1 n_2 n_3}{n_\mu}$. Given a multi-linear rank \mathbf{r} , we reformulate the problem of the Tucker approximation as a problem of dimensional reduction. Indeed we look for the subspace of $\mathbb{R}^{n_{\neq\mu}}$ of dimension r_μ which minimizes in norm the projection of point cloud \mathcal{X}_μ on it. This problem is solved with the HOSVD algorithm, which provides three orthogonal basis (U_1, U_2, U_3) of the corresponding subspaces. Therefore the i_μ -th row of $Y_\mu = U_\mu \Sigma_\mu$ represents the coordinates of the i_μ -th element of \mathcal{X}_μ projected into the subspace of $\mathbb{R}^{n_{\neq\mu}}$ of dimension r_{i_μ} for $\mu \in \{1, 2, 3\}$. The Proposition 1 result is interpreted geometrically as each point cloud living in the linear subspace built from the Kronecker product of the other two.

3.2 Generalization to d -order tensors

Now, we generalize Proposition 1 to the d -case as follows.

Proposition 2. *Let \mathbf{X} be a tensor of $E_1 \otimes \dots \otimes E_d$ with $\dim(E_\mu) = n_\mu$ and let \mathbf{C} be its HOSVD core at multi-linear rank \mathbf{r} . Let Y_μ be the principal components of mode μ . If $r_\mu \leq \text{rank}(A^{(\mu)})$ for $\mu \in \{1, \dots, d\}$, then*

$$Y_\mu = X^{(\mu)}(Y_d \otimes_K \dots \otimes_K Y_{\mu+1} \otimes_K Y_{\mu-1} \otimes_K \dots \otimes_K Y_1)(B^{(\mu)})^T$$

with $\mathbf{B} = (\Sigma_1^{-1}, \dots, \Sigma_d^{-1})\mathbf{C}$ for every $\mu \in \{1, \dots, d\}$.

Proof: The proof is very similar to that of Proposition 1, so we give only the main steps. Let start by the first mode principal components. The 1-mode matricization of \mathbf{X} in the Tucker basis is expressed with the Kronecker product, getting

$$X^{(1)} = \sum_{i_1=1}^{r_1} \mathbf{u}_{i_1}^1 \otimes \left(\sum_{i_2, \dots, i_d=1}^{r_2, \dots, r_d} C_{i_1 \dots i_d} \mathbf{u}_{i_d}^d \otimes_K \dots \otimes_K \mathbf{u}_{i_2}^2 \right). \quad (9)$$

The PCA of $X^{(1)}$ leads to

$$X^{(1)} = Y_1 V_1^T = \sum_{i_1=1}^{r_1} \sigma_{i_1}^{(1)} \mathbf{u}_{i_1}^1 \otimes \mathbf{v}_{i_1}^1 \quad (10)$$

with $\mathbf{v}_{i_1}^1$ the i_1 -th unitary column of $V_1 \in \mathbb{O}(n_{\neq 1} \times r_1)$ where $n_{\neq 1} = \prod_{\mu=2}^d n_\mu$ and $\sigma_{i_1}^{(1)} \mathbf{u}_{i_1}^1$ the i_1 -th column of $Y_1 \in \mathbb{R}^{n_1 \times r_1}$. Comparing Equations (9) and (10) for a fixed index i_1 , we identify $\mathbf{v}_{i_1}^1$ with a linear combination of the Kronecker product of $\mathbf{u}_{i_\mu}^\mu$ for $\mu \in \{2, \dots, d\}$ scaled by $\sigma_{i_1}^{(1)}$ as

$$\mathbf{v}_{i_1}^1 = \frac{1}{\sigma_{i_1}^{(1)}} \sum_{i_2, \dots, i_d=1}^{r_2, \dots, r_d} C_{i_1 \dots i_d} \mathbf{u}_{i_d}^d \otimes_K \dots \otimes_K \mathbf{u}_{i_2}^2. \quad (11)$$

Introducing in Equation (11) the singular values $\sigma_{i_\mu}^{(\mu)}$, we express the i_1 -th column of V_1 as a linear combination of the Kronecker product of the i_μ -th column of Y_μ for every $i_\mu \in \{2, \dots, d\}$ as

$$\begin{aligned} \mathbf{v}_{i_1}^1 &= \sum_{i_2, \dots, i_d=1}^{r_2, \dots, r_d} \frac{C_{i_1 \dots i_d}}{\sigma_{i_1}^{(1)} \sigma_{i_2}^{(2)} \dots \sigma_{i_d}^{(d)}} \sigma_{i_2}^{(2)} \dots \sigma_{i_d}^{(d)} \mathbf{u}_{i_d}^d \otimes_K \dots \otimes_K \mathbf{u}_{i_2}^2 \\ &= \sum_{i_2, \dots, i_d=1}^{r_2, \dots, r_d} B_{i_1 \dots i_d} \mathbf{y}_{i_d}^d \otimes_K \dots \otimes_K \mathbf{y}_{i_2}^2 \end{aligned} \quad (12)$$

with $\mathbf{B} = (\Sigma_1^{-1}, \dots, \Sigma_d^{-1})\mathbf{C}$ and $\mathbf{y}_{i_\mu}^\mu$ the i_μ -th column of Y_μ for $\mu \in \{2, \dots, d\}$. Thanks to the correspondence between Kronecker product and matricization, the right-hand-side of Equation (12) is expressed as the matrix-product between $Y_d \otimes_K \dots \otimes_K Y_2$ and tensor \mathbf{B} matricized with respect to mode 1 and transposed as

$$V_1 = (Y_d \otimes_K \dots \otimes_K Y_2)(B^{(1)})^T. \quad (13)$$

Multiplying Equation (10) on the right by V_1 yields $Y_1 = X^{(1)}V_1$ and replacing V_1 by its expression of Equation (13), it finally follows

$$Y_1 = X^{(1)}V_1 = X^{(1)}(Y_d \otimes_K \dots \otimes_K Y_2)(B^{(1)})^T.$$

The other relations follow straightforwardly from this one, permuting the indices coherently. \square

3.3 Extension to generic metric space for d -order tensors

As discussed in Section 2, the minimization problem faced with the HOSVD algorithm is expressed by the Frobenious norm, induced by an inner product. The standard inner product is defined by the identity matrix. However, whatever SPD matrix induces an inner product and the associated metric norm on a vector space, which is therefore isomorphic to the standard Euclidean space. We emphasize in this section the role of the metric on the relationships between point clouds, using this isomorphic relationship between Euclidean spaces with different inner products.

Let $\mathbf{F} \in \mathcal{S}_M$ where \mathcal{S}_M is the tensor space $E_1 \otimes \dots \otimes E_d$ with $\dim(E_\mu) = n_\mu$ endowed with the inner product induced by SPD matrices M_μ of size n_μ . Let \mathcal{S} be the Euclidean tensor space $E_1 \otimes \dots \otimes E_d$ endowed with the standard inner product. As already mentioned, we can move back to \mathcal{S} , thanks to the isometry define as $\nu : \mathcal{S}_M \rightarrow \mathcal{S}$, such that $\mathbf{X} = \nu(\mathbf{F}) = (M_1, \dots, M_d)\mathbf{F}$. Let now $\mathbf{X} \in \mathcal{S}$ be the HOSVD approximation of $\nu(\mathbf{F})$ at multi-linear rank \mathbf{r} and let $(U_\mu)_\mu \in$

$\mathbb{O}(n_\mu \times r_\mu)$ be the associated basis. Σ_μ is the singular value matrix of $X^{(\mu)}$ and define $Y_\mu = U_\mu \Sigma_\mu$ the principal components of mode μ for tensor \mathbf{X} for every $\mu \in \{1, \dots, d\}$. Let $W_\mu = M_\mu^{-1} Y_\mu$ be the principal components of \mathbf{F} in the tensor space \mathcal{S}_M . In the following proposition, we link the sets of principal components in the metric space \mathcal{S}_M . As previously, for the sake of clarity the result is first proved for $d = 3$ and afterwards generalized to whatever order d .

Proposition 3. *Let \mathbf{F} be a tensor in \mathcal{S}_M and let \mathbf{X} be its image through the isometry ν in the standard tensor space \mathcal{S} . Let \mathbf{C} be the HOSVD core tensor of \mathbf{X} at multi-linear rank \mathbf{r} such that $r_\mu \leq \text{rank}(X^{(\mu)})$ for $\mu \in \{1, 2, 3\}$. Then for W_μ the principal component of mode μ of \mathbf{F} in the metric tensor space \mathcal{S}_M it holds*

$$\begin{aligned} W_1 &= F^{(1)}(M_3^2 W_3 \otimes_K M_2^2 W_2)(B^{(1)})^T \\ W_2 &= F^{(2)}(M_3^2 W_3 \otimes_K M_1^2 W_1)(B^{(2)})^T \\ W_3 &= F^{(3)}(M_2^2 W_2 \otimes_K M_1^2 W_1)(B^{(3)})^T \end{aligned}$$

with $\mathbf{B} = (\Sigma_1^{-1}, \Sigma_2^{-1}, \Sigma_3^{-1})\mathbf{C}$.

Proof: This result comes straightforwardly from Proposition 1 proof by introducing the metrics matrices. For completeness, we illustrate the proof focusing on the link for the principal components of the first mode. Let $(U_\mu)_{\mu=1,2,3}$ be the HOSVD basis of \mathbf{X} at multi-linear rank \mathbf{r} with $U_\mu \in \mathbb{O}(n_\mu \times r_\mu)$. Let $Y_\mu = U_\mu \Sigma_\mu$ be the principal components of mode μ for tensor \mathbf{X} , where Σ_μ is the singular values matrix of $X^{(\mu)}$. Proposition 1 yields

$$Y_1 = X^{(1)}(Y_3 \otimes_K Y_2)(B^{(1)})^T \quad (14)$$

with $\mathbf{B} = (\Sigma_1^{-1}, \Sigma_2^{-1}, \Sigma_3^{-1})\mathbf{C}$. Notice that $\mathbf{X} = \nu(\mathbf{F}) = (M_1, M_2, M_3)\mathbf{F}$. So thanks to Equation (2), we express $X^{(1)}$ in function of $F^{(1)}$ as

$$X^{(1)} = M_1 F^{(1)}(M_3^T \otimes_K M_2^T)$$

and replacing it into (14), it gets

$$Y_1 = M_1 F^{(1)}(M_3 \otimes_K M_2)(Y_3 \otimes_K Y_2)(B^{(1)})^T \quad (15)$$

since M_μ are SPD matrices. Remarking that $Y_\mu = M_\mu W_\mu$ from the W_μ definition, substituting it in the Equation (15) we obtain

$$M_1 W_1 = M_1 F^{(1)}(M_3 \otimes_K M_2)(M_3 W_3 \otimes_K M_2 W_2)(B^{(1)})^T. \quad (16)$$

Since M_1 is SPD and consequently invertible, from the previous equation it follows the thesis.

The other relations follow straightforwardly from this proof, permuting the indices coherently. \square

This result is easily generalized to d -order tensors as follows.

Proposition 4. *Let \mathbf{F} be a tensor in \mathcal{S}_M and let \mathbf{X} be its image through the isometry ν in the standard tensor space \mathcal{S} . Let \mathbf{C} be the HOSVD core tensor of \mathbf{X} at multi-linear rank \mathbf{r} such that $r_\mu \leq \text{rank}(X^{(\mu)})$ for $\mu \in \{1, \dots, d\}$. Then for W_μ the principal component of mode μ of \mathbf{F} in the metric tensor space \mathcal{S}_M it holds*

$$W_\mu = F^{(1)}(M_d^2 W_d \otimes_K \dots \otimes_K M_{\mu+1}^2 W_{\mu+1} \otimes_K M_{\mu-1}^2 W_{\mu-1} \otimes_K \dots \otimes_K M_1^2 W_1)(B^{(\mu)})^T$$

with $\mathbf{B} = (\Sigma_1^{-1}, \dots, \Sigma_d^{-1})\mathbf{C}$ for every $\mu \in \{1, \dots, d\}$.

Proof: The proof is a direct consequence of Proposition 2 and Proposition 3, so the details are omitted here. \square

4 Geometric view for multiway correspondence analysis

In this last section, we transport the previous results in the correspondence analysis framework. We firstly clarify the Euclidean space and its metric where we set our problem. Then we make explicit the point cloud relation in this particular context. As final outcome we are able to prove the correspondence between the point clouds attached to each mode.

In accordance with the correspondence analysis framework, we consider a d -way contingency table $\mathbf{T} \in \mathbb{N}^{n_1 \times \dots \times n_d}$. The first step for performing CA is scaling \mathbf{T} by the sum of all its components setting a new frequency tensor

$$\mathbf{F} = \frac{1}{\sum_{i_1, \dots, i_d=1}^{n_1, \dots, n_d} T_{i_1 \dots i_d}} \mathbf{T}.$$

We first clarify the tensor space we will work with. Let \mathbf{f}^μ be the marginal of mode μ , i.e., the vector whose components are the sums of the slices in mode μ for all $\mu \in \{1, \dots, d\}$. For example the i -th element of \mathbf{f}^1 is

$$f_{i_1}^1 = \sum_{i_2, \dots, i_d=1}^{n_2, \dots, n_d} F_{i_1 \dots i_d} \quad \text{for all } i_1 \in \{1, \dots, n_1\}.$$

We assume that \mathbf{f}^μ has no zero component and by construction $f_{i_\mu}^\mu > 0$ for every $\mu \in \{1, \dots, d\}$. Let define $D_\mu = \text{diag}(\sqrt{\mathbf{f}^\mu}) \in \mathbb{R}^{n_\mu \times n_\mu}$ for each $\mu \in \{1, \dots, d\}$ and let assume that \mathbf{F} belongs to $\mathbb{R}^{n_1 \times \dots \times n_d}$ endowed by the metric induced by the matrices $(D_1^{-1}, \dots, D_d^{-1})$, since D_μ^{-1} is SPD for every $\mu \in \{1, \dots, d\}$. We denote by \mathcal{S}_M this metric space and by \mathcal{S} the tensor space $\mathbb{R}^{n_1 \times \dots \times n_d}$ endowed with the standard inner product. Under this assumption, let ν the isometry between the spaces \mathcal{S}_M and \mathcal{S} and let $\mathbf{X} = \nu(\mathbf{F}) = (D_1^{-1}, \dots, D_d^{-1})\mathbf{F}$. The general element of tensor \mathbf{X} is written

$$X_{i_1 \dots i_d} = \frac{F_{i_1 \dots i_d}}{\sqrt{f_{i_1}^1 \dots f_{i_d}^d}}.$$

Performing the HOSVD over tensor \mathbf{X} at multi-linear rank \mathbf{r} leads to a new orthogonal basis $(U_\mu)_{\mu=1, \dots, d}$, to a core tensor \mathbf{C} and to the principal components $Y_\mu = U_\mu \Sigma_\mu$ for every $\mu \in \{1, \dots, d\}$ in the standard tensor space. Focusing on the principal components $W_\mu = D_\mu Y_\mu$ of tensor \mathbf{F} in \mathcal{S}_M , Proposition 4 entails

$$W_\mu = F^{(1)}(M_d^2 W_d \otimes_K \dots \otimes_K M_{\mu+1}^2 W_{\mu+1} \otimes_K M_{\mu-1}^2 W_{\mu-1} \otimes_K \dots \otimes_K M_1^2 W_1)(B^{(\mu)})^T$$

where $B^{(\mu)}$ is the matricization of $\mathbf{B} = (\Sigma_1^{-1}, \dots, \Sigma_d^{-1})\mathbf{C}$ for $\mu \in \{1, \dots, d\}$. Let $Z_\mu = D_\mu^{-2} W_\mu$ be the *principal components scaled* by the singular value inverse for $\mu \in \{1, \dots, d\}$ in tensor space \mathcal{S}_M . Henceforth, we denote by $\mathbf{z}_{i_\mu}^\mu$ the i_μ -th row of Z_μ . Now we prove that each component of vector $\mathbf{z}_{i_\mu}^\mu$ can be expressed as a scaling factor times the barycenter of the linear combinations of the other two scaled principal component rows. We assume d equal to 3 to facilitate the comprehension of the following proof.

Proposition 5. *Let \mathbf{F} be a tensor in the tensor space \mathcal{S}_M endowed with the norm induced by the inner product matrices $D_\mu = \text{diag}(\sqrt{\mathbf{f}^\mu})$ with \mathbf{f}^μ the μ mode marginal of \mathbf{F} for $\mu \in \{1, 2, 3\}$. Let $Z_\mu \in \mathbb{R}^{n_\mu \times r_\mu}$ be the scaled principal components for tensor \mathbf{F} of mode μ in \mathcal{S}_M . If $r_\mu = \text{rank}(F^{(\mu)})$*

for every $\mu \in \{1, 2, 3\}$, then

$$\begin{aligned} (z_i^1)_\ell &= \frac{1}{\sigma_\ell^1} \sum_{j,k=1}^{n_2, n_3} \sum_{m,p=1}^{r_2, r_3} \frac{F_{ijk}}{f_i^1} (z_k^3)_p (z_j^2)_m (B_1)_{\ell mp} \\ (z_j^2)_m &= \frac{1}{\sigma_m^2} \sum_{i,k=1}^{n_1, n_3} \sum_{\ell,p=1}^{r_1, r_3} \frac{F_{ijk}}{f_j^2} (z_k^3)_p (z_i^1)_\ell (B_2)_{\ell mp} \\ (z_k^3)_p &= \frac{1}{\sigma_p^3} \sum_{i,j=1}^{n_1, n_2} \sum_{\ell,m=1}^{r_1, r_2} \frac{F_{ijk}}{f_k^3} (z_i^1)_\ell (z_j^2)_m (B_3)_{\ell mp} \end{aligned}$$

where $B_\mu = \Sigma_\mu B^{(\mu)}$ from Proposition 3.

Proof: We describe the proof for the i -th row of Z_1 with $i \in \{1, \dots, n_1\}$. From Proposition 3, under the CA metric choice, it follows that the principal components of \mathbf{F} in the tensor space \mathcal{S}_M satisfy the relation

$$W_1 = F^{(1)}(D_3^{-2}W_3 \otimes_K D_2^{-2}W_2)(B^{(1)})^T = F^{(1)}(Z_3 \otimes_K Z_2)(B^{(1)})^T.$$

Multiplying on the left this last equation by D_1^{-2} , we obtain

$$Z_1 = D_1^{-2}W_1 = D_1^{-2}F^{(1)}(Z_3 \otimes_K Z_2)(B^{(1)})^T$$

and since $B^{(1)} = \Sigma_1^{-1}B_1^{(1)}$, it gets

$$Z_1 = D_1^{-2}F^{(1)}(Z_3 \otimes_K Z_2)(\Sigma_1^{-1}B_1^{(1)})^T. \quad (17)$$

Making explicit the ℓ -th component of \mathbf{z}_i^1 , the i -th row of Z_1 , from Equation (17), we have

$$\begin{aligned} (z_i^1)_\ell &= (Z_1)_{i\ell} = \sum_{j,k=1}^{n_2, n_3} \sum_{m,p=1}^{r_2, r_3} (D_1^{-2}F^{(1)})_{ijk} (Z_3 \otimes_K Z_2)_{\overline{jk} \overline{mp}} (\Sigma_1^{-1}B_1^{(1)})_{\ell \overline{mp}} \\ &= \sum_{j,k=1}^{n_2, n_3} \sum_{m,p=1}^{r_2, r_3} \frac{F_{ijk}}{f_i^1} (Z_3)_{kp} (Z_2)_{jm} \frac{(B_1)_{\ell mp}}{\sigma_\ell^1} \\ &= \frac{1}{\sigma_\ell^1} \sum_{j,k=1}^{n_2, n_3} \sum_{m,p=1}^{r_2, r_3} \frac{F_{ijk}}{f_i^1} (z_k^3)_p (z_j^2)_m (B_1)_{\ell mp} \end{aligned}$$

by the definition of \mathbf{z}_j^2 and \mathbf{z}_k^3 for every $i \in \{1, \dots, n_1\}$, $j \in \{1, \dots, n_2\}$, $k \in \{1, \dots, n_3\}$ and $\ell \in \{1, \dots, r_1\}$. This final equation can be read as a mutual barycenter relation scaled by the inverse of the corresponding singular value. Indeed there is a list of weights terms which sum to 1, i.e., $\sum_{j,k=1}^{n_2, n_3} \frac{F_{ijk}}{f_i^1} = \frac{f_i^1}{f_i^1} = 1$ times a linear combination expressed through \mathbf{B}_1 of \mathbf{z}_j^2 and \mathbf{z}_k^3 .

Moving back to the geometric perspective, the Proposition 5 states that the scaled coordinates of a point cloud correspond to the barycenter of the other two point cloud scaled coordinates.

The two remaining barycentric relations follow from this proof, permuting coherently the indices. \square

Correspondence in CA refers to the correspondence of point cloud coordinates through the scaled barycentric relation. We proved that this well known relation in matrix framework is holding also in the tensor one, through HOSVD. Therefore we propose to refer to it as *correspondence analysis from HOSVD*.

This final proposition is extended and verified straightforwardly for d -order tensors.

Proposition 6. Let \mathbf{F} be a tensor in the tensor space \mathcal{S}_M endowed with the norm induced by the inner product matrices $D_\mu = \text{diag}(\sqrt{\mathbf{f}^\mu})$ with \mathbf{f}^μ the μ mode marginal of \mathbf{F} for every $\mu \in \{1, \dots, d\}$. Let $Z_\mu \in \mathbb{R}^{n_\mu \times r_\mu}$ be the scaled principal components for tensor \mathbf{F} of mode μ in \mathcal{S}_M . If $r_\mu = \text{rank}(F^{(\mu)})$ for $\mu \in \{1, \dots, d\}$, then

$$(z_{i_\mu}^\mu)_{\ell_\mu} = \frac{1}{\sigma_{\ell_\mu}^\mu} \sum_{\substack{\eta=1 \\ \eta \neq \mu}}^d \sum_{i_\eta=1}^{n_\eta} \sum_{\ell_\eta=1}^{r_\eta} \frac{F_{i_1 \dots i_d}}{f_{i_\mu}^\mu} (z_{i_d}^d)_{\ell_d} \cdots (z_{i_{\mu+1}}^{\mu+1})_{\ell_{\mu+1}} (z_{i_{\mu-1}}^{\mu-1})_{\ell_{\mu-1}} \cdots (z_{i_2}^2)_{\ell_2} (B_\mu)_{\ell_1 \dots \ell_d}$$

where $B_\mu = \Sigma_\mu B^{(\mu)}$ from Proposition 4.

Proof: From Proposition 4, under the CA metric choice, the 1st mode principal components of \mathbf{F} in \mathcal{S}_M are expressed as

$$W_1 = F^{(1)}(Z_d \otimes_K \cdots \otimes_K Z_2)(B^{(1)})^T.$$

If the previous equation is multiplied by equation by D_1^{-2} and it has $B^{(1)}$ replaced by the equivalent $\Sigma_1^{-1} B_1^{(1)}$, it gets

$$Z_1 = D_1^{-2} F^{(1)}(Z_d \otimes_K \cdots \otimes_K Z_2)(\Sigma_1^{-1} B_1^{(1)})^T. \quad (18)$$

Making explicit the ℓ_1 -th component of $\mathbf{z}_{i_1}^1$, the i_1 -th row of Z_1 , from Equation (18), it follows the thesis, i.e.

$$(z_{i_1}^1)_{\ell_1} = \frac{1}{\sigma_{\ell_1}^1} \sum_{i_2, \dots, i_d=1}^{n_2, \dots, n_d} \sum_{\ell_2, \dots, \ell_d=1}^{r_2, \dots, r_d} \frac{F_{i_1 \dots i_d}}{f_{i_1}^1} (z_{i_d}^d)_{\ell_d} \cdots (z_{i_2}^2)_{\ell_2} (B_1)_{\ell_1 \dots \ell_d}.$$

The proof for the other modes follows directly from this one permuting coherently the indices. \square

5 Applications

In this section we compare the MWCA based on the point cloud relation proved in Section 4 with the classical CA performed on the same data reorganized as a matrix. Indeed, if we assume $\mathbf{F} \in \mathbb{R}^{n_1 \times \dots \times n_d}$ to store the data relative frequencies as a tensor, then their matrix representation is given by A_k , the mode k -matricization of \mathbf{F} , once mode k has been selected. The MWCA is performed on \mathbf{F} , while CA is performed on A_k . Let denote by $\tilde{F}^{(k)}$ the k -matricization of \mathbf{F} after applying the isometry, similarly \tilde{A}_k is the outcome of isometry in the CA case. Since the two approaches are different, the isometries transporting \mathbf{F} and A_k differ and so do \tilde{A}_k and $\tilde{F}^{(k)}$. To estimate this discrepancy between the two objects we define the relative error $e(\tilde{F}^{(k)}, \tilde{A}_k)$

$$e(\tilde{F}^{(k)}, \tilde{A}_k) = \frac{\|\tilde{F}^{(k)} - \tilde{A}_k\|}{\|\tilde{F}^{(k)}\|}. \quad (19)$$

The SVD and HOSVD, over which CA and MWCA rely respectively, provide an orthogonal basis for the matrix or the decomposed tensor, which is unique up to an orthogonal rotation. As proposed in [42] and [43], we orient these new basis selecting as leading direction the one where the majority of the data points out. To perform CA and MWCA we used `python 3.6.9` and the library `TensorLy 0.6.0`, see [44].

We first analyse with these two techniques the data reported in [45]. This example is important since it shows that the multiway method results are coherent with those stated with correspondence analysis. Then the multiway barycentric relation is used to interpret an original data-set from the ecological domain [46].

Age group	Males					Females				
	Very good	Good	Regular	Bad	Very bad	Very good	Good	Regular	Bad	Very bad
16-24	145	402	84	5	3	98	387	83	13	3
25-34	112	414	74	13	2	108	395	90	22	4
35-44	80	331	82	24	4	67	327	99	17	4
45-54	54	231	102	22	6	36	238	134	28	10
55-64	30	219	119	53	12	23	195	187	53	18
65-74	18	125	110	35	4	26	142	174	63	16
+75	9	67	65	25	8	11	69	92	41	9

Table 1: Data from the Spanish National Health Survey of 1997, see [45]

Example 1. In [47] data are reported from a survey over 6731 people of both genders, from 16 to over 75 years old, who were asked to evaluate their health status. Then the answers were organized in a 3-way table with dimensions $n_1 = 2$, $n_2 = 7$ and $n_3 = 5$, as in Table 1. On mode 1 we have the two genders: male, ‘M’, and female, ‘F’ and on the second mode there are 7 age groups. On the last mode we set 5 health grades, from ‘Very Good’ to ‘Very bad’. The $(1, 1, 1)$ entry of the multiway table is the number of men between 16 and 24 years old who judge ‘Very good’ their health status. Let $\mathbf{F} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ represent the data relative frequencies in tensor format for MWCA. CA is realized over A_3 the matricization of \mathbf{F} with respect to mode 3, i.e. ‘health grade’, where on the row we set the health categories and on the columns all the possible combinations of ages and gender. In Figure 2a we recover the correspondence analysis of [47], while in 2b we display multiway correspondence analysis with the first two columns of $Y_\mu = U_\mu \Sigma_\mu$, defined in Section 4 for $\mu \in \{1, 2, 3\}$. As in [47], both plots in Figure 2 display a

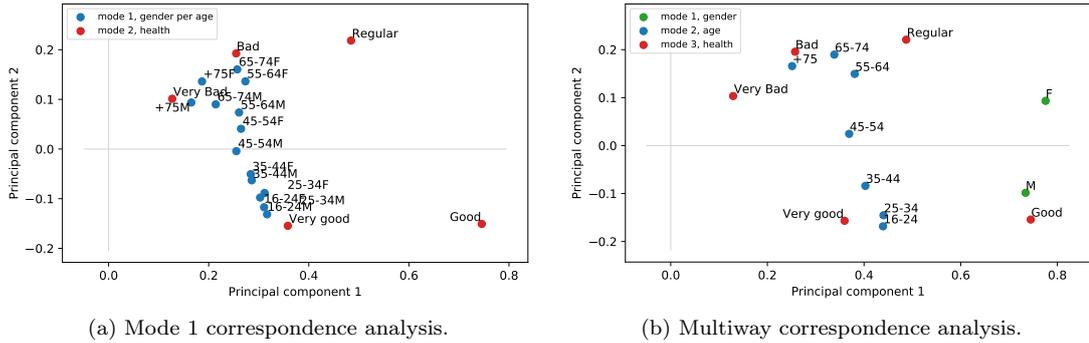


Figure 2: CA versus MWCA for data of [45].

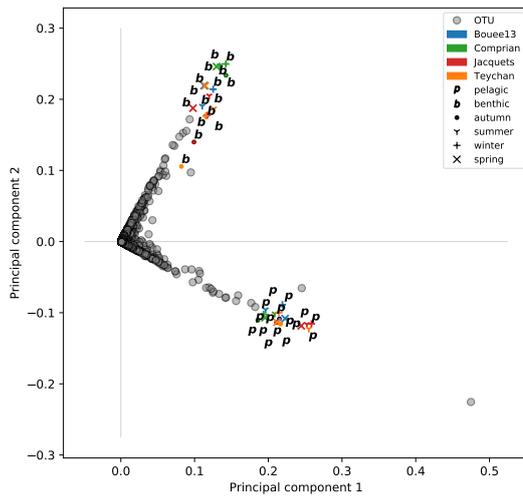
gradient for the age categories: they distribute from the youngest in the bottom to the oldest near the top. The 5 health categories are at the extremes of this gradient, the better health levels at the bottom, worst ones at the top. From this we infer that the health evaluation decreases for increasing ages. This phenomenon is highlighted by both the techniques, even if it is clearer in Figure 2b, thanks to the possibility of analyzing the variables separately. Taking into account

the gender in Figure 2b, let us remark that the male dot is close to the ‘Good’ one. Thanks to the barycentric relation, we know that the closer the points the more correlated they are. So men appear to provide optimistic evaluations of their health status. Women dot stands aside from the other, slightly closer to worst health status, suggesting that female health judging is balanced between the age classes, with a minor inclination toward pessimistic evaluation. In Figure 2a for increasing ages, male and female dots tend to increase their distances, suggesting that when men and women age, they tend respectively to be more optimistic and pessimistic in evaluating their health status. MWCA plot seems to point out that men are more optimistic than how much pessimistic women are. In contrast from Figure 2b we cannot conclude easily which age category presents the biggest difference in health evaluation for the two genders, information that can be inferred from 2a. Lastly comparing the health point cloud in Figures 2a and 2b, we observe that their principal coordinates are almost the same and we explain it in terms of relative error from Equation (19). Indeed for this data-set the relative error $e(\bar{T}^{(3)}, \bar{A}_3) \approx 0.035$ is quite small.

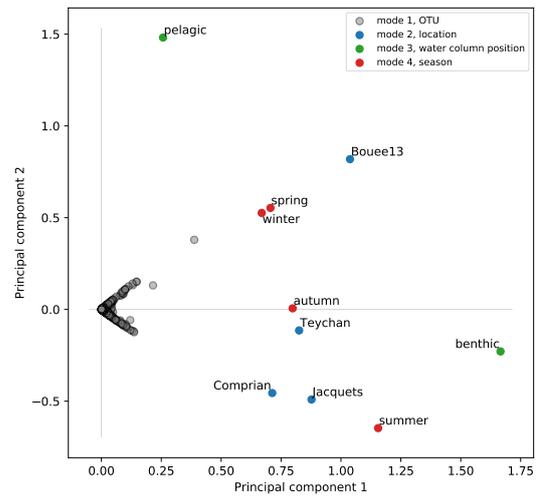
Example 2. The data-set provided by Malabar project (IFREMER, CNRS, INRAE, Labex COTE) [46] is a multiway contingency table of 4-order. In this metabarcoding project, 32 water samples have been collected in Arcachon Bay at four locations (Bouee13, Comprian, Jacquets, Teychan), during four seasons, and at two positions in water column (pelagic and benthic). DNA has been extracted, and Operational Taxonomic Units (OTUs) have been built. OTU are expected to correspond to species and, without entering too much into details here, the question which motivates Malabar project is to understand or quantify the role of location, season and water column in the diversity of protists. To address this question, a four way contingency table \mathbf{T} has been built, where $T_{i,j,k,\ell}$ is the number of sequences collected at location j , in season ℓ and position k which belong to OTU i . The size of the contingency tensor is $3539 \times 4 \times 2 \times 4$. Our contribution here is to make a first analysis of this data-set, comparing results provided by classical CA and MWCA in Figure 3.

In Figure 3a the result of CA performed over A_1 , the matricization of frequency tensor data \mathbf{F} with respect to mode 1, where on the rows there are the different OTUs and on the columns the locations, water column positions and seasons combined. Next to it, Figure 3b displays the result of MWCA over the tensor data. In the CA case OTU point cloud spreads over two orthogonal direction, led by the two positions in the water column. Similarly in MWCA plot the water column point cloud seem to organize and to affect the position and clustering of the others. Because of mode combination in Figure 3a, it is not evident which position in the water column affects which location or season. However thanks to the barycentric relation and the multiway approach, we can infer some relations among season, location and water column from Figure 3b, where they are independently displayed. Indeed pelagic point appears to affect the OTU present during spring and winter time. Similarly benthic position probably drives the distribution of OTU in autumn and summer. A similar argument can be repeated for location point cloud. Indeed the barycentric relation suggests that the benthic position leads the distribution of OTU in Comprian, Jacquets and Teychan, while the pelagic dot influences Bouee13 dot. The barycentric relation enables us to interpret the interaction among season and location point clouds. In Figure 3b winter and spring seem to be more correlated with Bouee13 and similarly summer and autumn are probably more correlated with the remaining three locations. Analysing the OTU and the season point clouds, we observe that winter and spring pull out one orthogonal direction over which part of the OTU points spread. Similarly summer pulls out the other OTU orthogonal direction. The autumn point appears to be neutral with respect to the OTU point cloud distribution. Similarly if we study the OTU and location point clouds interaction, one orthogonal direction is driven by Bouee13, one by Comprian and Jacquets, while Teychan seems to have a small effect on the OTU point distribution.

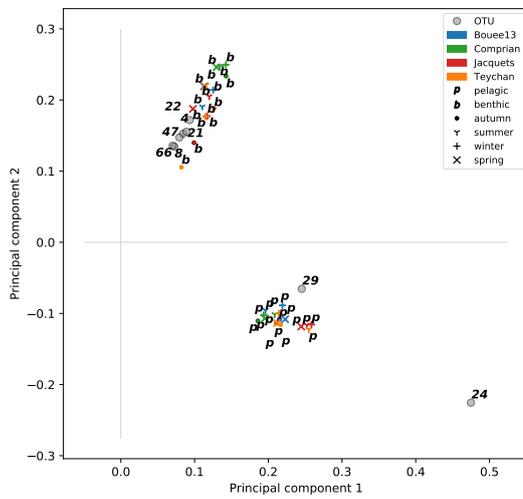
In Figures 3c and 3d, we filter the OTU by size, keeping just those with first principal



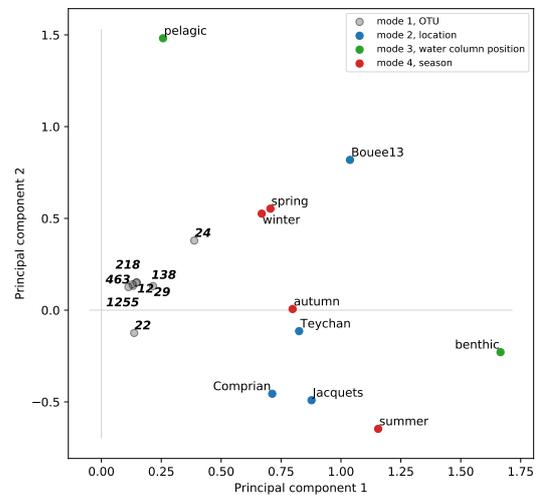
(a) Correspondence Analysis of mode 1.



(b) Multiway Correspondence Analysis.



(c) CA with filtered OTU.



(d) MWCA with filtered OTU.

Figure 3: CA vs MWCA biplot of Arcachon Bay data-set.

components greater than 0.2 and the second greater than 0.12. Figure 3c confirms the idea of water-column position driving the OTU point cloud distribution. However in Figure 3b most of the OTU points are close to the origin. Indeed in this left plot we filtered 8 OTU which are probably more correlated with benthic or pelagic condition. Comparing the two filtered figures, we see that only 3 of the 8 selected OTU are common between them, i.e., OTU ‘22’, ‘24’ and ‘29’. These common OTU share the same correlations between the two methods. Indeed in both case, ‘22’ is more related to benthic position, while ‘24’ and ‘29’ with pelagic one. Moreover most of the OTU in Figure 3d are concentrated around the origin. We may infer that the action of the water column point cloud is softened by the other modes, here independent, leading to a more balanced distribution of OTU. In conclusion notice that in MWCA the different point clouds are more spread than in the CA case, enabling a deeper analysis of the variables interactions. It could be linked to the significant relative error $e(\hat{F}^{(1)}, \hat{A}_1) \approx 0.14$, which may also explain why CA and MWCA figures present a flipped y -axis. Indeed the mode combinations including benthic have positive y -components in Figure 3a, while benthic point has negative y -component in Figure 3b. However since it affects in the same way all the point clouds, the two methods lead to a coherent interpretation.

6 Conclusion

The correspondence word in correspondence analysis comes from the relationship between the row and column point cloud coordinates, proven in the matrix case. We have proposed an extension of this correspondence to tensors by the High Order Singular Value Decomposition (HOSVD). Indeed, for each matricization of a given tensor, we computed the associated principal components. We first proved that each principal component is a linear combination of the Kronecker product of the others. Noting that, from the geometric viewpoint, principal components are the coordinates of point clouds, we provided a geometric interpretation of the algebraic link. The next step of our argument was to highlight the role of the metric in the principal component links. By finally choosing the CA metric, we recovered the barycentric link between the coordinates of the point clouds. The persistence of this relationship allows us to refer it as the extension of correspondence analysis to tensors through HOSVD. In Subsection 2.2 we observed that several algorithms were proposed over the years to compute the Tucker basis. We focused on the simplest one, e.g., HOSVD. However the statements of all propositions can be proved using Tucker basis obtained from other algorithms, as, for example, HOOI, [31]. In this case a further algorithm to compute the singular values will be necessary to get the point clouds. Numerical experiments have shown two comparisons between the classical Correspondence Analysis (CA) performed on matricized tensor data and the MultiWay Correspondence Analysis (MWCA) implemented with HOSVD working with data structured as tensor. Both techniques guide to the same interpretation, although with different nuances. The main advantage MWCA is to highlight the global properties of a variable category. Indeed, the coordinates of each point cloud are computed independently from the others. Therefore MWCA catches at the same time the global pattern of point clouds and the minor ones. On the contrary CA is able to state clearly only the most important relation among two points clouds. It is worthwhile underlining that in CA combining variables helps in understanding their peculiarities. Indeed in CA two or more variables are coupled, because of the tensor matricization, and the analysis may highlight relations specific for a certain combination of modes.

In conclusion, the choice of MWCA or CA for interpreting multiway tables depends strongly on the objective of the data analysis one wishes to perform. We recommend MWCA when the objective is to discover an overall reciprocal correlation of different variables. While we suggest

the CA when the objective is to identify specific combination patterns.

7 Acknowledgments.

The authors are very grateful to Luc Giraud for many helpful discussions and suggestions, which helped us improving this work. They thank also the teams of the Malabar project for sharing their data before their publication and letting them test their method.

References

- [1] A. J. Izenman, *Modern Multivariate Statistical Techniques*. NY: Springer, 2008.
- [2] J. Wang, *Geometric structure of high-dimensional data and dimensionality reduction*. Springer, 2012.
- [3] K. Pearson, “LIII. On lines and planes of closest fit to systems of points in space,” *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, no. 11, pp. 559–572, 1901.
- [4] H. Hotelling, “Analysis of a complex of statistical variables into principal components.,” *Journal of Educational Psychology*, vol. 24, no. 6, pp. 417–441, 1933.
- [5] K. R. Gabriel, “The biplot graphic display of matrices with application to principal component analysis,” *Biometrika*, vol. 58, pp. 453–467, 12 1971.
- [6] W. Atchley and E. Bryant, *Multivariate Statistical Methods: Among-Groups Covariation*. Benchmark Papers in Systematic and Evolutionary Biology, Elsevier Science & Technology Books, 1975.
- [7] M. G. Kendall, *Multivariate analysis*. Griffin London, 1975.
- [8] D. Morrison, D. Morrison, L. Marshall, M.-H. Incorporated, F. Morrison, H. Sahlin, N. Y. A. of Sciences, L. da Vinci Academy of Arts, and Sciences, *Multivariate Statistical Methods*. Annals of the New York Academy of Sciences, McGraw-Hill, 1976.
- [9] J. Chambers, *Computational Methods for Data Analysis*. Wiley Series in Probability and Mathematical Statistics, Wiley, 1977.
- [10] I. T. Jolliffe, *Principal Component Analysis*. Springer Series in Statistics, Springer-Verlag New York, second ed., 2002.
- [11] I. T. Jolliffe and J. Cadima, “Principal component analysis: a review and recent developments,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, p. 20150202, 2016.
- [12] K. V. Mardia, J. Kent, and J. M. Bibby, *Multivariate Analysis*. Probability and Mathematical Statistics, Academic Press, 1979.
- [13] L. Lebart, A. Morineau, and J.-P. Fénelon, *Traitement des données statistiques*. Dunod, Paris, 1982.
- [14] G. Saporta, *Probabilités, Analyse de Données et Statistique*. Editions Technip, 1990.

-
- [15] T. Anderson, *An Introduction to Multivariate Statistical Analysis*. Wiley Series in Probability and Statistics, Wiley, 2003.
- [16] C. R. Rao, *Linear statistical Inference and its Applications*. Wiley Series in Probability and Mathematical Statistics, Wiley, second ed., 1973.
- [17] J. Benzécri and L. Bellier, *L'analyse des données: Benzécri, J.-P. et al. L'analyse des correspondances*. L'analyse des données: leçons sur l'analyse factorielle et la reconnaissance des formes, et travaux du Laboratoire de statistique de l'Université de Paris VI, Dunod, 1973.
- [18] M. O. Hill, "Correspondence analysis: A neglected multivariate method," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 23, no. 3, pp. 340–354, 1974.
- [19] M. Greenacre and L. Degos, "Correspondence analysis of HLA gene frequency data from 124 population samples," *American journal of human genetics*, vol. 29, pp. 60–75, 02 1977.
- [20] M. J. Greenacre, "Some objective methods of graphical display of a data matrix," *Special Report*. UNISA, Pretoria, 1978.
- [21] H. G. Gauch, *Multivariate Analysis in Community Ecology*. Cambridge Studies in Ecology, Cambridge University Press, 1982.
- [22] A. Gifi, *Nonlinear Multivariate Analysis*. Wiley Series in Probability and Statistics, Wiley, 1990.
- [23] M. J. Greenacre, *Theory and Applications of Correspondence Analysis*. Academic Press, 1984.
- [24] M. J. Greenacre, *Correspondence Analysis in practice*. Chapman & Hall/CRC, second ed., 2007.
- [25] B. Escofier-Cordier, *L'analyse factorielle des correspondances*. PhD thesis, University of Rennes, 1965.
- [26] J. Benzécri, "Statistical analysis as a tool to make patterns emerge from data," in *Methodologies of Pattern Recognition* (S. Watanabe, ed.), pp. 35–74, Academic Press, 1969.
- [27] P. M. Kroonenberg, *Three-mode principal component analysis: Theory and applications*. DSWO Press, Leiden, 1983.
- [28] R. Coppi and S. Bolasco, eds., *Multiway Data Analysis*. North-Holland Publishing Co., 1989.
- [29] A. Franc, *Etude Algébrique des multitableaux: apports de l'algèbre tensorielle*. PhD thesis, University of Montpellier, 1992.
- [30] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1253–1278, 2000.
- [31] L. De Lathauwer, B. De Moor, and J. Vandewalle, "On the best rank-1 and rank- (r_1, r_2, \dots, r_n) approximation of high order tensors," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1324–1342, 2000.

- [32] P. M. Kroonenberg, *Applied Multiway Data Analysis*. Wiley, 2008.
- [33] T. G. Kolda and B. W. Bader, “Tensor decompositions and applications,” *SIAM Review*, vol. 51, pp. 455–500, September 2009.
- [34] A. Franc, “Multiway arrays: some algebraic remarks,” in *Multiway Data Analysis* (R. Coppi and S. Bolasco, eds.), ch. 2, pp. 19–29, NLD: North-Holland Publishing Co., 1989.
- [35] L. R. Tucker, “Some mathematical notes on three-modes factor analysis,” *Psychometrika*, vol. 31, no. 3, pp. 279–311, 1966.
- [36] T. G. Kolda, “Multilinear operators for higher-order decompositions,” Tech. Rep. SAND2006-2081, Sandia National Laboratories, April 2006.
- [37] K. Arie, H. Neudecker, and T. Wansbeek, “An approach to n -mode components analysis,” *Psychometrika*, vol. 51, no. 2, pp. 269–275, 1986.
- [38] L. Grasedyck, “Hierarchical singular value decomposition of tensors,” *SIAM Journal on Matrix Analysis and Applications*, vol. 31, no. 4, pp. 2029–2054, 2010.
- [39] N. Vannieuwenhoven, R. Vandebril, and K. Meerbergen, “A new truncation strategy for the higher-order singular value decomposition,” *SIAM Journal on Scientific Computing*, vol. 34, no. 2, pp. A1027–A1052, 2012.
- [40] P. Kroonenberg and J. de Leeuw, “Principal component analysis of three modes data by means of alternating least square algorithms,” *Psychometrika*, vol. 45, no. 1, pp. 69–97, 1980.
- [41] J. L. Lastovicka, “The extension of component analysis to four-mode matrices,” *Psychometrika*, vol. 46, no. 1, pp. 47–57, 1981.
- [42] R. Bro, E. Acar, and T. G. Kolda, “Resolving the sign ambiguity in the singular value decomposition,” *Journal of Chemometrics*, vol. 22, no. 2, pp. 135–140, 2008.
- [43] R. Bro, R. Leardi, and L. G. Johnsen, “Solving the sign indeterminacy for multiway models,” *Journal of Chemometrics*, vol. 27, no. 3-4, pp. 70–75, 2013.
- [44] J. Kossaifi, Y. Panagakis, A. Anandkumar, and M. Pantic, “Tensorly: Tensor learning in python,” *Journal of Machine Learning Research*, vol. 20, no. 26, pp. 1–6, 2019.
- [45] CARME-N, “Spanish national health survey.” <http://www.carme-n.org/?sec=data2>, 2007. [Data-set 3].
- [46] “Malabar project data paper (IFREMER, CNRS, INRAE, Labex COTE).” in prep., 202X.
- [47] M. Greenacre, *Correspondence Analysis in Practice*. Academic Press, 1993. Data-sets are available at.

Inria

**RESEARCH CENTRE
BORDEAUX – SUD-OUEST**

200 avenue de la Vieille Tour
33405 Talence Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399