Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v6n2p202

A procedure for three analysis of compositions
By Gallo,Simonacci

October 14, 2013

Electronic Journal of Applied Statistical Analysis
Vol. 06, Issue 02, 2013, 202-210
DOI: 10.1285/i20705948v6n2p202

A procedure for three analysis of compositions

Michele Gallo *and Violetta Simonacci

University of Naples "L'Orientale" Department of Human and Social Sciences
P.zza S.Giovanni 30 Naples, It 80134

October 14, 2013

The Tucker3 model is one of the most widely used tools for factorial analysis of three-way data arrays. When orthogonal factors are extracted this model can be seen as a three-way PCA (principal component analysis). The Tucker3 model is characterized by extreme flexibility as it allows for the use of a different number of factors in each mode and it yields non-unique results. When this model is applied to vectors of non-negative values with a sum constraint all problems connected with the statistical analysis of compositions must be taken into consideration. Like other standard statistical techniques, this model cannot be directly applied. The aim of this paper is to present the theory behind the Tucker3 model on compositional data and to describe the TUCKALS3 algorithm.

keywords: Compositional data, simplex space, log-ratio transformation, Tucker models, TUCKALS3.

1 Introduction

Compositional data (CoDa) appear as proportions, percentages, concentrations, absolute and relative frequencies. They can often be found in many disciplines and in many scientific fields. A compositional vector is a vector made up of non negative values summing to a unit, or in general, to some fixed constant. The constant-sum constraint that characterizes compositions is, however frequently disregarded or improperly incorporated into statistical modeling and a misleading interpretation of the results is given. Due to these specifications, several difficulties arise when dealing with CoDa. The first word of warning came already in 1897 from Karl Pearson who showed the dangers of

*Michele Gallo: mgallo@unior.it

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

underestimating spurious correlations. Several attempts were done in the course of literature to identify the negative bias issue and to find a solution to this problem so that statistical modeling could reasonably be applied on compositional data.
Geometrically speaking, due to the constrained nature of compositional data, the sample space for compositional vectors is a subset of real space, the simplex. The simplex has been studied as an Euclidean linear vector space Pawlowsky-Glahn and Egozcue (2001) and Billheimer et al. (2001). In this paper, most of the elements that were introduced by Aitchison in the 1980s, such as perturbation and powering operations and orthogonal log-contrasts, have been organized into a systematic and coherent mathematical scheme. Recently, additional tools for the representation of compositions and their exploratory analysis have been developed Egozcue and Pawlowsky-Glahn (2005), Pawlowsky-Glahn and Egozcue (2011), Egozcue et al. (2003) and Egozcue et al. (2011). Aitchison showed how a statistical model applied on compositions will yield consistent results only if three conditions are fulfilled: scale invariance, permutation-invariance and sub-compositional coherence. Based on these assumptions he proposed to use log-ratio transformations to preprocess CoDa before conducting any statistical analysis. Through this transformation, a direct association between the simplex and the real space is found. This way it is possible to work in real space, where it is easier, and later, through the inverse function, the results can be taken back into simplex space for a correct interpretation.
Compositional data can be arranged into three-way matrices when, for example, I compositions of J parts have been collected in K occasions. When dealing with three-way data it is possible to apply two-way analysis techniques such as principal component analysis (PCA), but this generally resolves in loss of information due to the fact that two of the three modes are combined together. Therefore, PCA for CoDa, proposed by Aitchison (1982), is not adequate to analyze CoDa when arranged in three-way arrays. In recent statistical literature other papers have argued the use of three-mode analysis such as the PARAFAC/CANDECOMP, the Tucker3 and the weighted PCA models for studying three-way data, for more details see Gallo (2013a), Gallo (2013b) and Gallo and Buccianti (2013).
The purpose of this work is to provide a procedure to analyze compositions by Tucker3. Specifically, after presenting a set of convenient symbols, some basic concepts of compositional data analysis have been adapted for compositions arranged into a three-way array. Finally a procedure based on the least squares algorithm, known as TUCKALS3, for three-way compositional data analysis is given in point format.

2 Notations and elements of simplicial geometry

2.1 Three-way array definitions

The notation used in this paper is based on Gallo (2012) and Smilde et al. (2005). Boldface underlined letters designate three-way arrays; two-way arrays (matrices) are in boldface uppercase characters; vectors are in boldface lowercase characters (always a row vector) and scalars are in lowercase characters. Each three-way array can be seen as a collection of matrices, called slices. These slices, which are frontal, vertical

and horizontal, can be concatenated between them obtaining several kinds of matricized three-way arrays.

A matrix given by concatenating frontal slices is indicated in boldface uppercase characters with a subscript A. A matrix given by concatenating vertical slices is indicated in boldface uppercase characters with a subscript B. And finally, a matrix given by concatenating horizontal slices is indicated in boldface uppercase characters with a subscript C. For example, let $\underline{\mathbf{V}}$ (IxJxK) be a three-way array with I objects or compositions, J variables or parts of composition and K occasions. There are three types of slices, (IxJ) frontal slices $\mathbf{V}_k$ with k=1,...,K, (IxK) vertical slices $\mathbf{V}_j$ with j=1,...,J, and (KxJ) horizontal slices $\mathbf{V}_i$ with i=1,...,I. These slices can be concatenated between them obtaining the following matricization of the three-way array: $\mathbf{V}_A$ (IxJK) $\mathbf{V}_B$ (JxIK) $\mathbf{V}_C$ (KxJI), i.e. $\mathbf{V}_A$=[$\mathbf{V}_1$|...|$\mathbf{V}_k$|...|$\mathbf{V}_K$], $\mathbf{V}_B$=[$\mathbf{V}_1{}^t$|...|$\mathbf{V}_k{}^t$|...|$\mathbf{V}_K{}^t$] and $\mathbf{V}_C$=[$\mathbf{V}_1$|...|$\mathbf{V}_i$|...|$\mathbf{V}_I$]. In addition, the three-way array $\underline{\mathbf{V}}$ can be broken up into vectors, called fibers. The three different types of fibers are referred to as rows, columns and tubes. Thus, $\underline{\mathbf{V}}$ can be broken up into IK rows $\mathbf{v}_{ik}$, JK columns $\mathbf{v}_{jk}$, IJ tubes $\mathbf{v}_{ij}$, with dimension (1xJ), (1xI) and (1xK), respectively.

Traditionally, when a three-way array is unfolded into a matrix, all the scores of a subject are arranged into a sequence of records in order to create a wide combination-mode matrix, e.g., the ith row of $\mathbf{V}_A$ is $\underline{\mathbf{v}}_i$=[$\mathbf{v}_{i1}$|...|$\mathbf{v}_{ik}$|...|$\mathbf{v}_{iK}$]. Each slice of a three-way array can be converted into a column vector by vec-operator Kiers (2000) and Smilde et al. (2005). Thus, it is possible to arrange all the column vectors underneath each other. For example, let $\mathbf{V}_i$ be the ith horizontal slice, $Vec(\mathbf{V}_i{}^t)$=$\underline{\mathbf{v}}_i$=[$\mathbf{v}_{i1}$|...|$\mathbf{v}_{ik}$|...|$\mathbf{v}_{iK}$], which is the ith row of $\mathbf{V}_A$.

2.2 Basic concepts

Let $S_k{}^J$ be the simplex space with dimension J-1, defined as $S_k{}^J$=$\{\mathbf{v}_{\bullet k}$= ($v_{\bullet 1k}$,...,$v_{\bullet Jk}$): $v_{\bullet 1k}$> 0,...,$v_{\bullet Jk}$> 0;$\sum_j \mathbf{v}_{\bullet jk}$=$\kappa\}$, where κ is a given positive constant, which is usually 1 or 100, depending on whether the variables are measured in part per unit or as percentages, respectively. The typical simplex element, $\mathbf{v}_{\bullet k}$∈$S_k{}^J$, is called composition, and its components $v_{\bullet jk}$ (j=1,...,J) are called parts of $\mathbf{v}_{\bullet k}$.

Since absolute values of compositional data are not relevant, as they only carry relative information, we can scale any composition so that the fixed total of its components equals κ. This can indeed be achieved through an operator that projects a vector with all positive elements from real to a simplex space. In other words, let $\underline{\tilde{\mathbf{V}}}$ be a three-way array with all positive elements and $\tilde{\mathbf{v}}_{\bullet k}$ a row of the kth frontal slice of $\underline{\tilde{\mathbf{V}}}$, its closure is defined as $\mathbf{v}_{\bullet k}$=$\mathbb{C}(\tilde{\mathbf{v}}_{\bullet k})$= ($\kappa\tilde{\mathbf{v}}_{\bullet 1k}/\sum_j \tilde{\mathbf{v}}_{\bullet jk}$,...,$\kappa\tilde{\mathbf{v}}_{\bullet Jk}/\sum_j \tilde{\mathbf{v}}_{\bullet jk}$). The operator $\mathbb{C}$ is called the κ-closure operator.

The closure operation implies that two rows of the kth frontal slice of $\underline{\tilde{\mathbf{V}}}$, $\tilde{\mathbf{v}}_{\bullet k}$,$\tilde{\mathbf{v}}_{\circ k}$∈$\mathbb{R}_k{}^J$, are compositionally equivalent if $\mathbb{C}(\tilde{\mathbf{v}}_{\bullet k})$=$\mathbb{C}(\tilde{\mathbf{v}}_{\circ k})$ or, in other words, if there exists a positive scalar λ∈$\mathbb{R}_+{}^J$ so that $\tilde{\mathbf{v}}_{\bullet k}$=$\lambda\tilde{\mathbf{v}}_{\circ k}$. In this case we say that $\tilde{\mathbf{v}}_{\bullet k}$ and $\tilde{\mathbf{v}}_{\circ k}$ belong to the same compositional class. These equivalent vectors are connected to the origin of $\mathbb{R}_+{}^J$ by the same ray. The intersection point of this ray with the κ simplex is representative of that compositional class. Given all these specifications concerning the nature

of compositional data, a statistical method applied on compositions will yield consistent results only if three conditions are fulfilled: scale invariance, permutation-invariance and sub-compositional coherence. A function fulfills the scale invariance condition if, for any real positive value of λ and any composition $\mathbf{v}_{\bullet k}\in {S_k}^J$ we have $f(\mathbf{v}_{\bullet k})=f(\lambda\mathbf{v}_{\bullet k})$ while it is permutation-invariant if rearranging the parts in the composition does not modify the outcome. Finally the function must also be sub-compositionally coherent which, from a geometric stand point, implies that if we have a C part sub-composition ${\mathbf{v}_{\bullet k}}^s=(v_{\bullet sk},...,v_{\bullet sk})$ of a starting J part composition $\mathbf{v}_{\bullet k}$ this sub-composition ($S<J$) must behave as an orthogonal projection of the corresponding whole composition.

The Euclidean geometric structure has proven unfit to guarantee consistent results when applied to compositions as the unit-simplex is characterized by own structure known as Aitchison geometry. At the base of this geometry there are the two operations of perturbation and powering which give the simplex a vector space structure. The perturbation operation transforms compositions the same way the translation operation would in real space and it can be used to measure differences between compositions. The perturbation between two compositions $\mathbf{v}_{\bullet k},\mathbf{v}_{\circ k}\in {S_k}^J$ results in a new composition $\mathbf{v}_{*k}$ defined by: $\mathbf{v}_{*k}=\mathbf{v}_{\bullet k}\oplus\mathbf{v}_{\circ k}= \mathbb{C}(v_{\bullet 1k}v_{\circ 1k},v_{\bullet 2k}v_{\circ 2k},...,v_{\bullet Jk}v_{\circ Jk})$. The operation of powering is instead analogous to multiplication by a scalar and it can be defined for any $\alpha\in\mathbb{R}_+$ and any composition $\mathbf{v}_{\bullet k}\in {S_k}^J$ as it follows: $\alpha\odot\mathbf{v}_{\bullet k}\mathbb{C}({v_{\bullet 1k}}^\alpha,{v_{\bullet 2k}}^\alpha,...,{v_{\bullet Jk}}^\alpha)$. After defining the perturbation operation and the power transformation, the simplex ${S_k}^J$ can be considered a vector space with dimension *J-1* on $\mathbb{R}$. These operations are indeed equivalent to translation and scalar multiplication in real space as the following properties apply.

Property 1: Let $\mathbf{v}_{\bullet k},\mathbf{v}_{\circ k},\mathbf{v}_{+k}$ be compositions in ${S_k}^J$ and $\alpha\in\mathbb{R}_+$. Then

- (associative) $(\mathbf{v}_{\bullet k}\oplus\mathbf{v}_{\circ k})\oplus\mathbf{v}_{+k} =\mathbf{v}_{\bullet k}\oplus(\mathbf{v}_{\circ k}\oplus\mathbf{v}_{+k})$;
- (commutative) $\mathbf{v}_{\bullet k}\oplus\mathbf{v}_{\circ k}=\mathbf{v}_{\circ k}\oplus\mathbf{v}_{\bullet k}$;
- (opposite element) $\mathbf{v}_{\bullet k}\oplus(\text{-}1\odot\mathbf{v}_{\circ k})=\eta$;
- (neutral element) $\eta=\mathbb{C}(1,1,...,1)=(1/J,1/J,...,1/J)$;
- (distributive) $(\alpha\odot\mathbf{v}_{\bullet k})\oplus(\alpha\odot\mathbf{v}_{\circ k})=\alpha\odot(\mathbf{v}_{\bullet k}\oplus\mathbf{v}_{\circ k})$;
- (unit) $(1\odot\mathbf{v}_{\bullet k})=\mathbf{v}_{\bullet k}$.

Note that formally we handle the operations $\oplus$ and $\odot$ in the simplex the same way we handle standard vector operations of addition, subtraction and multiplication in real space. All the fibers of a three-way array $\underline{\tilde{\mathbf{V}}}$ can be transformed into compositions by the closure operator $\mathbb{C}$, but it would really only make sense to do this for the objects. Thus, when it is applied to a row of $\tilde{\mathbf{V}}_{\mathrm{k}}$ ($k=1,...,K$), it then defines a transformation ${\mathbb{R}_k}^J\rightarrow {S_k}^J$ with ${S_k}^J$ as previously defined. Afterwards, we proceed to indicate the IK rows of three-way array as compositions. Thus, for each frontal slice $\mathbf{V}_{\mathrm{k}}$ ($k=1,...,K$) we have I compositions with the relative sample space ${S_k}^J$ ($k=1,...,K$). Therefore, in each simplex ${S_k}^J$ there are I points with coordinates $\mathbf{v}_{1k}...\mathbf{v}_{ik}...\mathbf{v}_{Ik}$. In three-way arrays, each

object is observed on several occasions, in other words the data of the ith object are arranged in the horizontal slice $\tilde{\mathbf{V}}_{\mathrm{i}}$ and often the values that the object assumes on different occasions are plotted in the same space as a trajectory. In this case, for each object, K points are plotted and linked between them in the same real space ${\mathbb{R}_+}^J$. In the same way, if the rows of a horizontal slice are compositions the sample space can be defined as S^J and for the representation of each composition K points are linked between them to obtain the trajectory for each composition through the K occasions. On the other hand, each object observed on several occasions is often summarized in real space with only one point. In this case, by vec-operator, $vec(\ .\ )$, horizontal slices can be vectored. This operator can be applied to the horizontal slices after the closure operator. Thus, it is possible to define the following vector: $\underline{\mathbf{v}}_i=\mathbb{C}({\tilde{\mathbf{V}}_i}^t)= (\kappa\tilde{v}_{i11}/\sum_j \tilde{v}_{ij1},...,\kappa\tilde{v}_{iJ1}/\sum_j \tilde{v}_{ij1}$ $|...|\kappa\tilde{v}_{i1K}/\sum_j \tilde{v}_{ijK},...,\kappa\tilde{v}_{iJK}/\sum_j \tilde{v}_{ijK})=[\mathbf{v}_{i1}|...|\mathbf{v}_{ik}|...|\mathbf{v}_{iK}]$. For each horizontal slice $\tilde{\mathbf{V}}_i$, $vec(\mathbb{C}(\ .\ ))$ defines the transformation onto simplex space ${S_k}^1$x...x ${S_k}^K=\prod_{k=1}^K {S_k}^J=S^{JK}$.

In accordance with perturbation and powering definitions, it is possible to verify the following properties for the ternary $(S^{JK}, \oplus,\odot)$.

Property 2: Let $\underline{\mathbf{v}}_i,\underline{\mathbf{v}}_{i'},\underline{\mathbf{v}}_{i''}$ be compositions in S^{JK} and $\alpha\in\mathbb{R}_+$. Then

- (associative) $(\underline{\mathbf{v}}_i\oplus\underline{\mathbf{v}}_{i'})\oplus\underline{\mathbf{v}}_{i''} =\underline{\mathbf{v}}_i\oplus(\underline{\mathbf{v}}_{i'}\oplus\underline{\mathbf{v}}_{i''})$;
- (commutative) $\underline{\mathbf{v}}_i\oplus\underline{\mathbf{v}}_{i'}=\underline{\mathbf{v}}_{i'}\oplus\underline{\mathbf{v}}_i$;
- (opposite element) $\underline{\mathbf{v}}_i\oplus(\text{-}1\odot\underline{\mathbf{v}}_{i'})=\eta^*$;
- (neutral element) $\eta^*=(1/J,...,1/J|...|1/J,...,1/J)$;
- (distributive) $(\alpha\odot\underline{\mathbf{v}}_i)\oplus(\alpha\odot\underline{\mathbf{v}}_{i'})=\alpha\odot(\underline{\mathbf{v}}_i\oplus\underline{\mathbf{v}}_{i'})$;
- (unit) $(1\odot\underline{\mathbf{v}}_i)=(\underline{\mathbf{v}}_i)$.

Therefore in the simplex space S^{JK} we use the same operations $\oplus$ and $\odot$ of ${S_K}^J$.

2.3 From simplex to real space

Studying compositional data within the framework of their sample space implies having to completely rethink any statistical tool one wishes to apply. Alternatively, the constant sum constraint can be removed with an appropriate transformation of compositional data so that it is possible to work in real space and standard unconstrained multivariate techniques can be applied.

As compositional data only carry relative information, this can be achieved considering logarithms of ratios, known as log-ratios, which allow for a bi-univocal correspondence between simplex and real space representation of data. We can briefly recall the formulas for the four most used log-ratio transformations proposed in literature: pairwise log-ratio (plr), additive log-ratio (alr), centered log-ratio (clr) and isometric log-ratio (ilr) where the plr, alr and clr were introduced by Aitchison (1982) and Aitchison (1986), while the ilr was introduced by Egozcue et al. (2003). Unfortunately, the alr transformation has the inconvenient of not being invariant under permutation of components therefore some

statistical procedure may fail. For this reason it will not be taken into consideration in this work (for more detail see Egozcue et al. (2011)).

A pairwise transformation applied to a vector $\mathbf{v}_{ik}$ gives a new vector with *(J-1)J/2* pairwise log-ratios, where the generic element is $\log(v_{ijk}/v_{ij'k})$ with $(j<j')$. The centered and isometric transformations applied to the same vector $\mathbf{v}_{ik}$ can be defined as:

- $clr(\mathbf{v}_{ik})=(\log(v_{i1k}/g(\mathbf{v}_{ik})),...,\log(v_{ijk}/g(\mathbf{v}_{ik})),...,\log(v_{iJk}/g(\mathbf{v}_{ik})))$ with dimension J and where $g(\mathbf{v}_{ik})=\prod_{j=1...J}^{J} v_{ijk}$;
- $ilr(\mathbf{v}_{ik})=(ilr(v_{i1k},...,ilr(v_{ijk},...,ilr(v_{i(J-1)k}))$, with dimension *(J-1)* and generic element $ilr(v_{ijk})=((J-j)/(J-j+1))^{1/2}\log(v_{ijk}/(\prod_{h=j+1}^{J} v_{ihk})^{1/(J-j)})$.

These transformations can be used to define a metric structure in the simplex. The inner product, norm and distance for the *clr* representation of compositions in ${S_K}^J$ are

- $<\mathbf{v}_{ik},\mathbf{v}_{i'k}>_a= \sum_{j=1}^{J}\log(v_{ijk}/g(\mathbf{v}_{ik}))*\log(v_{i'jk}/g(\mathbf{v}_{i'k}))$;
- $(||\mathbf{v}_{ik}||_a=(\sum_{j=1}^{J}\log(v_{ijk}/g(\mathbf{v}_{ik}))^2)^{1/2}$;
- $d(\mathbf{v}_{ik},\mathbf{v}_{i'k})_a= (\sum_{j=1}^{J}(\log(v_{ijk}/g(\mathbf{v}_{ik}))\text{-}\log(v_{i'jk}/g(\mathbf{v}_{i'k})))^2)^{1/2}$.

where $<.,.>_a$, $||.||_a$ and $d(.,.)_a$ denote the Aitchison inner product, norm and distance. With these properties, we have that the *clr* is an isometric transformation from simplex to real space. The same can be said for the *plr* and *ilr* transformations.

It is, now possible to define the pairwise, centered and isometric transformations for a vector $\underline{\mathbf{v}}_i$. These transformations can be defined easily by the following formulations:

- $\hat{plr}(\underline{\mathbf{v}}_i)=(\ plr(\mathbf{v}_{i1}),...,plr(\mathbf{v}_{ik}),...,plr(\mathbf{v}_{iK}))$;
- $\hat{clr}(\underline{\mathbf{v}}_i)=(\ clr(\mathbf{v}_{i1}),...,clr(\mathbf{v}_{ik}),...,clr(\mathbf{v}_{iK}))$;
- $\hat{ilr}(\underline{\mathbf{v}}_i)=(\ ilr(\mathbf{v}_{i1}),...,ilr(\mathbf{v}_{ik}),...,ilr(\mathbf{v}_{iK}))$.

Accordingly, the Aitchison geometry on the simplex ${S_K}^J$ can be used for S^{JK} as well.

The listed transformations are all viable tools for working with compositional data as long as we bear in mind the properties of that specific transformation when interpreting results. Choosing which to apply depends largely on the available data and on the kind of analysis one wishes to perform. Due to the asymmetric property of the *alr* it is hardly applied when conducting a multidimensional analysis. Therefore, it is not discussed in this work.

3 Three-way modeling by Tucker3 for CoDa

The Tucker3 model is one of the most basic multi-way models used in psychometrics and chemometrics Tucker (1966). The model is defined by the decomposition of a three-way array into a three-way core array and three two-way loadings matrices. In scalar notation, the Tucker3 model can be written as:

$$v_{ijk} = \sum_{p=1}^{P}\sum_{q=1}^{Q}\sum_{r=1}^{R} g_{pqr}(a_{ip}b_{jq}c_{kr}) + e_{ijk} \quad (1)$$

where e_{ijk} is an element of the residual array $\underline{\mathbf{E}}$ (IxJxK); a_{ip}, b_{jq}, c_{kr} are the typical elements of the loadings matrices $\mathbf{A}$ (IxP), $\mathbf{B}$ (JxQ) and $\mathbf{C}$ (KxR); and g_{pqr} is the typical element of the core-array $\underline{\mathbf{G}}$ (PxQxR), where the notation (P,Q,R) is used to indicate that the model has P, Q and R extracted factors for the first, the second and the third mode respectively, for full details see Kiers (2000) and Kroonenberg (2008).

According to the strategy proposed by Aitchison (1986), in order to analyze compositional data we have to move from simplex space to real space and then, after the multidimensional analysis has been carried out, move back to the simplex for interpretation of results. To achieve this purpose, the logarithmic transformations discussed in Section 2.3 can be applied to the three-way array $\underline{\mathbf{V}}$. The first step is to work out $\underline{\mathbf{L}}$ (IxJxK), an array with typical element $\log(v_{ijk})$, $\mathbf{L}_k$ indicates the kth frontal slice of this array. Thus, the kth frontal slice of the *clr* can be written as $\mathbf{L}_k\mathbf{P}_J{}^{\perp}$, where $\mathbf{P}_J{}^{\perp}$=($\mathbf{I}_J$-$\mathbf{1}_J\mathbf{1}_J{}^t/J$) is the symmetric and idempotent centering matrix, $\mathbf{I}_J$ is (JxJ) identity matrix and $\mathbf{1}$ is a I-dimensional vector of ones. The kth frontal slice of the *plr* can be written as $\mathbf{L}_k\Xi$, where Ξ is a (JxJ(J-1)/2) matrix with 0s in each column except for a 1 and -1 in two rows, since $\Xi\Xi^t$=$J\mathbf{P}_J{}^{\perp}$. Finally, the kth frontal slice of the *ilr* can be written as $\mathbf{L}_k\mathbf{P}_J{}^{\perp}\Psi$ where $\Psi^t\Psi$=$\mathbf{I}_{J\text{-}1}$ and $\Psi\Psi^t$=($\mathbf{I}_J$-$\mathbf{1}_k\mathbf{1}_k{}^t/J$). Moreover, to ensure that log-ratios are centered respect to column means each frontal slice is premultiplied by the symmetric and idempotent centering matrix $\mathbf{P}_I{}^{\perp}$=($\mathbf{I}_I$-$\mathbf{1}_I\mathbf{1}_I{}^t/I$), where $\mathbf{I}_I$ is the (IxI) identity matrix and $\mathbf{1}_I$ is a I-dimensional vector of ones. Thus, the kth frontal slice of the *clr* transformation is $\mathbf{Y}_k$=$\mathbf{P}_I{}^{\perp}\mathbf{L}_k\mathbf{P}_J{}^{\perp}$, while the columnwise-centered kth frontal slice for *plr* and *ilr* are $\hat{\mathbf{Y}}_k$=$\mathbf{P}_I{}^{\perp}\mathbf{L}_k\Xi$ and $\check{\mathbf{Y}}_k$= $\mathbf{P}_I{}^{\perp}\mathbf{L}_k\mathbf{P}_J{}^{\perp}\Psi$, respectively.

In case of centered log-ratio transformation, Equation 1, in matrix notation, can be written:

$$\mathbf{Y}_A = \mathbf{A}\mathbf{G}_A(\mathbf{C}\otimes\mathbf{B})^t + \mathbf{E}_A \quad (2)$$

where $\otimes$ is the Kronecher product and $\mathbf{Y}_A$=[$\mathbf{Y}_1$|...|$\mathbf{Y}_k$|...|$\mathbf{Y}_K$]. In case of CoDa, the loadings matrices for the *first* and *second* mode should have a column sum equal to zero or in other words they must be centered, in mathematical term $\tilde{\mathbf{1}}_I\mathbf{A}$=$\tilde{\mathbf{0}}_P$ and $\tilde{\mathbf{1}}_J\mathbf{B}$=$\tilde{\mathbf{0}}_Q$ (where $\tilde{\mathbf{0}}_P$ and $\tilde{\mathbf{0}}_Q$ are vectors of zero with P- and Q- dimensions, respectively). Therefore, the Tucker3 model for CoDa has the following loss function:

$$\min_{\tilde{\mathbf{1}}_I\mathbf{A}=\tilde{\mathbf{0}}_P,\tilde{\mathbf{1}}_J\mathbf{B}=\tilde{\mathbf{0}}_Q,\mathbf{C},\mathbf{G}} \|\mathbf{Y}_A - \mathbf{A}\mathbf{G}_A(\mathbf{C}\otimes\mathbf{B})^t\|^2 \quad (3)$$

In Equation 3 the constrained $\tilde{\mathbf{1}}_I\mathbf{A}$=$\tilde{\mathbf{0}}_P$ and $\tilde{\mathbf{1}}_J\mathbf{B}$=$\tilde{\mathbf{0}}_Q$ will automatically be respected. In other words, the double-centering of each $\mathbf{Y}_k$ (k=1,...,K) assures that the loadings matrices $\mathbf{A}$ and $\mathbf{B}$ will always be centered. Therefore, the additional constraints do not cause any problem because traditional algorithms, as such TUCKALS3 can be applied for centered log-ratio data. This algorithm fits the model in a least squares sense with orthonormal loadings vectors.

Gallo (2013b) has shown that the loadings matrices of *clr* preprocessed data are strongly linked with the correspondent *plr* and *ilr* loadings matrices. In detail, the loadings matrix of *clr* data for the first mode is equivalent to the loadings matrix for the first mode obtained from *plr* and *ilr* transformed data. The same relationship exists among the loadings matrices of the third mode. On the other hand, let $\hat{\mathbf{B}}$ and $\breve{\mathbf{B}}$ be the loadings matrices for the second mode for *plr* and *ilr* preprocessed data, respectively, it can be shown that $\hat{\mathbf{B}}=\Xi^t\mathbf{B}$ and $\breve{\mathbf{B}}=\Psi^t\mathbf{B}$, where the matrices Ξ and Ψ have been previously defined. Thus, Tucker3 results on pairwise log-ratio data can be obtained by the analysis of the smaller three-way arrays of centered log-ratio data. In the same way, it is possible to obtain Tucker3 results on isometric log-ratio data by the loadings matrices of the *clr* preprocessed data.
To summarize, in order to find the final parameter estimated with orthogonal loadings, the more efficient TUCKALS3 can be used for CoDa too Smilde et al. (2005). A schematic overview of the proposed method can be found in Table 1, where (*P, Q, R*) are the sought dimensions for the three mode of a three-way compositional data set $\underline{\mathbf{V}}$ (IxJxK).

Table 1: A schematic overview of Tucker3 analysis for compositions.

a) Preprocessing

For each frontal slice *k=1,...,K*

- Logarithmic transformation is applied on each element of $\mathbf{V}_k$, the results are in $\mathbf{L}_k$
- Create $\mathbf{P}_I{}^{\perp}$ and $\mathbf{P}_J{}^{\perp}$
- $\mathbf{Y}_k$ is given by $\mathbf{P}_I{}^{\perp}\mathbf{Y}_k\mathbf{P}_J{}^{\perp}$

b) Do TUCKALS3 algorithm

1. Initialize $\mathbf{B}$ and $\mathbf{C}$ (with the first Q and R left singular vector of $\mathbf{Y}_B$ and $\mathbf{Y}_C$)
2. $\mathbf{A}$ equal first P left singular vectors of $\mathbf{Y}_A(\mathbf{C}\otimes\mathbf{B})$
3. $\mathbf{B}$ equal first Q left singular vectors of $\mathbf{Y}_B(\mathbf{C}\otimes\mathbf{A})$
4. $\mathbf{C}$ equal first R left singular vectors of $\mathbf{Y}_C(\mathbf{B}\otimes\mathbf{A})$
5. Repeat steps 2-4 until relative changes are small
6. $\mathbf{G}_A=\mathbf{A}^t\mathbf{Y}_A(\mathbf{C}\otimes\mathbf{B})$

c) Get results for *plr* and *ilr*

- Create Ξ and Ψ
- $\hat{\mathbf{B}}$ and $\breve{\mathbf{B}}$ are given by $\Xi\mathbf{B}$ and $\Psi^t\mathbf{B}$
- $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{G}_A$ are the same for the three log-ratio transformations

4 Conclusions

Standard statistical methodology was traditionally developed on real space. In case of compositional data the sample space has a different algebraic and geometric structure know as simplex. Following Aitchisons approach it is however possible to move compositions from simplex space to real space by using log-ratio transformations. After one of these transformations has been used, the application of standard statistical techniques is indeed possible and algorithms proposed in statistical literature for fitting the parameters can be used for compositional data too. Finally, it is possible to return to the simplex using the inverse log-ratio transformation.

Following this approach, a schematic overview of the TUCKALS3 algorithm for study CoDa was given but not much has been said about interpretation of results nor about any of the plotting procedures used for visual representations of Tucker3 results, see for more details on these topics Gallo (2013a), Gallo (2013b) and Gallo and Buccianti (2013).

An alternative is working on orthonormal coordinates Barcelo-Vidal et al. (2011). The principle of working on coordinates described above implies that all standard methods can be applied to coordinates of any composition with respect to an orthonormal basis. These two alternative approaches are essentially the same. In this work the generalization for three-way arrays of the theoretical background can be used for working on orthonormal coordinates too.

Aknowledgment

Funding for this project was provided by the University of Naples "LOrientale".

References

Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 139–177.

Aitchison, J. (1986). *The statistical analysis of compositional data.* Monographs on Statistics and Applied Probability.

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

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

Egozcue, J. and Pawlowsky-Glahn, V. (2005). Groups of parts and their balances in compositional data analysis. *Mathematical Geology*, 37(7):795–828.

Egozcue, J. J., Barcelo-Vidal, C., Martín-Fernández, J. A., Jarauta-Bragulat, E., Díaz-

Barrero, J. L., and Mateu-Figueras, G. (2011). Elements of simplicial linear algebra and geometry. *Compositional Data Analysis: Theory and Applications*, pages 141–157.

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

Gallo, M. (2012). Coda in three-way arrays and relative sample spaces. *Electronic Journal of Applied Statistical Analysis*, 5(3):400–405.

Gallo, M. (2013a). Log-ratio and parallel factor analysis: an approach to analyze three-way compositional data. In *Advanced Dynamic Modeling of Economic and Social Systems*, pages 209–221. Springer.

Gallo, M. (2013b). Tucker3 model for compositional data. *Commun. Statist. Theor. Meth.*, in press.

Gallo, M. and Buccianti, A. (2013). Weighted principal component analysis for compositional data: application example for the water chemistry of the arno river (tuscany, central italy). *Environmetrics*.

Kiers, H. A. (2000). Towards a standardized notation and terminology in multiway analysis. *Journal of chemometrics*, 14(3):105–122.

Kroonenberg, P. M. (2008). *Applied multiway data analysis*, volume 702. John Wiley & Sons.

Pawlowsky-Glahn, V. and Egozcue, J. J. (2001). Geometric approach to statistical analysis on the simplex. *Stochastic Environmental Research and Risk Assessment*, 15(5):384–398.

Pawlowsky-Glahn, V. and Egozcue, J. J. (2011). Exploring compositional data with the coda-dendrogram. *Austrian Journal of Statistics*, 40(1-2):103–113.

Smilde, A., Bro, R., and Geladi, P. (2005). *Multi-way analysis: applications in the chemical sciences*. Wiley. com.

Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3):279–311.