Some New Results on Orthogonally Constrained Candecomp

Bennani Dosse, Mohammed; ten Berge, Jos M. F.; Tendeiro, Jorge N.

Published in:
Journal of Classification

DOI:
10.1007/s00357-011-9086-8

Journal of Classification 28:144-155 (2011)
DOI: 10.1007/s00357-011-9086-8

Some New Results on Orthogonally Constrained Candecomp

Mohammed Bennani Dosse

CNRS UMR 6625, University of Rennes 2, France

Jos M.F. ten Berge

University of Groningen, The Netherlands

Jorge N. Tendeiro

University of Groningen, The Netherlands

Abstract: The use of Candecomp to fit scalar products in the context of Indscal is based on the assumption that, due to the symmetry of the data matrices involved, two components matrices will become equal when Candecomp converges. Bennani Dosse and Ten Berge (2008) have shown that, in the single component case, the assumption can only be violated at saddle points in the case of Gramian matrices. This paper again considers Candecomp applied to symmetric matrices, but with an orthonormality constraint on the components. This constrained version of Candecomp, when applied to symmetric matrices, has long been known under the acronym Indort. When the data matrices are positive definite, or have become positive semidefinite due to double centering, and the saliences are nonnegative – by chance or by constraint –, the component matrices resulting from Indort are shown to be equal. Because Indort is also free from so-called degeneracy problems, it is a highly attractive alternative to Candecomp in the present context. We also consider a well-known successive approach to the orthogonally constrained Indscal problem and we compare, from simulated and real data sets, its results with those given by the simultaneous (Indort) approach.

Keywords: Simultaneous approach; Successive approach; Indort; Three-way Data; Candecomp; Indscal.

The authors are obliged to the editor and referees for helpful suggestions.

Authors' Addresses: M. Bennani Dosse, University of Rennes 2, Place du Recteur Henri Le Moal, CS 24307, 35043 Rennes Cedex, France, e-mail: mohammed.bennani@univ-rennes2.fr; J.Ten Berge and J. Tendeiro, Heijmans Institute of Psychological Research, University of Groningen, Grote Kruisstrat, 9712 TS Groningen, e-mails: j.m.f.ten.berge@rug.nl, j.n.tendeiro@rug.nl.

Published online 5 July 2011

1. Introduction

Carroll and Chang (1970) developed Candecomp as a method to minimize the residual sum of squares in a scalar products fitting problem derived from Indscal. Specifically, given m symmetric $p \times p$ matrices $\mathbf{S}_1, \ldots, \mathbf{S}_m$, they sought to minimize the function

$$g(\mathbf{X}, \mathbf{D}) = \sum_{i=1}^{m} \|\mathbf{S}_i - \mathbf{X}\mathbf{D}_i\mathbf{X}^t\|^2, \tag{1}$$

where $\mathbf{X}$ is a $p \times r$ matrix of components, $\mathbf{D}_i$ is a diagonal $r \times r$ matrix of saliences, with elements of row i of $\mathbf{D}$ in the diagonal, $i = 1, \ldots, m$, and $\|\cdot\|$ denotes the trace norm. Because direct minimization of g seems difficult, Carroll and Chang instead proposed Candecomp, to minimize

$$f(\mathbf{X}, \mathbf{Y}, \mathbf{D}) = \sum_{i=1}^{m} \|\mathbf{S}_i - \mathbf{X}\mathbf{D}_i\mathbf{Y}^t\|^2. \tag{2}$$

They assumed that, at the minimum, the symmetry of the slices $\mathbf{S}_i$ will cause $\mathbf{X}$ and $\mathbf{Y}$ to be equivalent (i.e. to be equal or to be column-wise proportional; in the latter case, columns of $\mathbf{Y}$ can be rescaled to become equal to those of $\mathbf{X}$, the inverse scaling being applied to the columns of $\mathbf{D}$).

The Candecomp algorithm minimizes f iteratively, by alternately optimizing $\mathbf{X}$ conditionally for fixed $\mathbf{Y}$ and $\mathbf{D}$, optimizing $\mathbf{Y}$ conditionally for fixed $\mathbf{X}$ and $\mathbf{D}$, and optimizing $\mathbf{D}$ conditionally for fixed $\mathbf{X}$ and $\mathbf{Y}$. In practice, the claim that symmetry of the slices will render $\mathbf{X}$ and $\mathbf{Y}$ equivalent seems warranted. However, for contrived data, counterexamples do exist, see Ten Berge and Kiers (1991).

Kroonenberg (1983, p. 118) discussed the possibility of minimizing the loss function (1) subject to $\mathbf{X}^t\mathbf{X} = \mathbf{I}_r$. At least two algorithms are available for this. First, it is straightforward to adapt Candecomp in such a way that the updates for $\mathbf{X}$ and $\mathbf{Y}$ are constrained to be columnwise orthonormal. That is, we can minimize $f(\mathbf{X}, \mathbf{Y}, \mathbf{D})$ subject to $\mathbf{X}^t\mathbf{X} = \mathbf{Y}^t\mathbf{Y} = \mathbf{I}_r$. The resulting algorithm, christened Indort by Kiers (1989), yields solutions for $\mathbf{X}$ and $\mathbf{Y}$ that may still be different.

Ten Berge, Knol and Kiers (1988), also see Bolla, Michaletzky, Tusnády and Ziermann (1998), used a method which updates just $\mathbf{X}$. Although the latter method cannot yield $\mathbf{X}$ and $\mathbf{Y}$ different, it seems to be more liable than Indort to arriving at local minima or saddle points of (1), see Tendeiro, Bennani Dosse & Ten Berge (2010), and hence requires a large number of random starts. In the present paper, this method will be ignored.

Ten Berge and Kiers (1991) have considered a third approach to minimizing f subject to $\mathbf{X}^t\mathbf{X} = \mathbf{Y}^t\mathbf{Y} = \mathbf{I}_r$. They noted that, for fixed $\mathbf{X}$ and $\mathbf{Y}$, both columnwise orthogonal, the optimal $\mathbf{D}$ must have rows satisfying

$$\mathbf{D}_i = \text{Diag}(\mathbf{X}^t\mathbf{S}_i\mathbf{Y}), \tag{3}$$

$i = 1, \ldots, m$ where Diag$(\cdot)$ denotes the diagonal matrix formed by the diagonal elements of the given matrix, whence minimizing f becomes equivalent to maximizing

$$h(\mathbf{X}, \mathbf{Y}) = \sum_{i=1}^{m} \text{tr}\left\{\left(\text{Diag}\mathbf{X}^t\mathbf{S}_i\mathbf{Y}\right)^2\right\}. \tag{4}$$

Although this function seems simpler than *f*, it is not clear how to update $\mathbf{X}$ and $\mathbf{Y}$, except when $r = 1$. In that case, $\mathbf{X}$ and $\mathbf{Y}$ become vectors, and hence *h* reduces to

$$h(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{m} \mathbf{x}^t\mathbf{S}_i\mathbf{y}\mathbf{y}^t\mathbf{S}_i\mathbf{x}. \tag{5}$$

Clearly, the optimal choice for $\mathbf{x}$ given $\mathbf{y}$ and of $\mathbf{y}$ given $\mathbf{x}$ are the principal eigenvectors of $\sum \mathbf{S}_i\mathbf{y}\mathbf{y}^t\mathbf{S}_i$ and $\sum \mathbf{S}_i\mathbf{x}\mathbf{x}^t\mathbf{S}_i$, respectively.

Ten Berge and Kiers (1991, Result 6) claimed that we would have $\mathbf{X} = \mathbf{Y}$ at every stationary point of $h(\mathbf{X}, \mathbf{Y})$ but this has been refuted by Bennani Dosse and Ten Berge (2008). They have shown for the $r = 1$ case that stationary points (saddle points, in fact) of *h* exist without conditional optimality for $\mathbf{X}$ and $\mathbf{Y}$, and with $\mathbf{X}$ and $\mathbf{Y}$ non-proportional. Nevertheless, it should be noted that any algorithm converging to a point of *h* where $\mathbf{X}$ is conditionally optimal given $\mathbf{Y}$, and vice versa, will return $\mathbf{X} = \mathbf{Y}$ if the matrices $\mathbf{S}_1, \ldots, \mathbf{S}_m$ are positive semidefinite (psd), and at least one of them is positive definite (pd). The proof of this can be found in Result 6 of Ten Berge and Kiers (1991 p. 325). It means that the algorithm described below (5) will return $\mathbf{x} = \mathbf{y}$. However, this result has no bearing on Indort, because Indort is not based on eliminating $\mathbf{D}$ from f by (3). Indeed, Indort may converge to points where $\mathbf{X}$ and $\mathbf{Y}$ differ, even when $\mathbf{S}_1, \ldots, \mathbf{S}_m$ are psd, and at least one of them is pd. Examples can be found in Bennani Dosse and Ten Berge (2008).

Throughout the results discussed so far, the signs of the elements in $\mathbf{D}$ have been ignored. In all reasonable applications of Indort, saliences (elements of $\mathbf{D}$) are tacitly assumed to be nonnegative, because negative saliences would be embarrassing. Ten Berge, Kiers and Krijnen (1993) have shown how negative saliences in Indscal can be suppressed by using nonnegative least squares for updating $\mathbf{D}$. In the present paper, we examine what happens when Indort is similarly adapted. Our main result is that, under weak nonsingularity assumptions, Indort will have $\mathbf{X} = \mathbf{Y}$ when $\mathbf{D}$ contains no negative elements.

The organization of this paper is as follows. First, we recall the Indort algorithm and repeat how to update all columns of $\mathbf{X}, \mathbf{Y}$ and $\mathbf{D}$ in a

simultaneous approach. Then in Section 3, our main result is presented. In Section 4, we consider a successive approach which finds one component at a time and compare its performance with the simultaneous approach. The paper ends with some conclusions.

2. The Indort Algorithm

The Indort algorithm, which has been around for several decades, minimizes $f(\mathbf{X}, \mathbf{Y}, \mathbf{D})$ subject to $\mathbf{X}^t\mathbf{X} = \mathbf{Y}^t\mathbf{Y} = \mathbf{I}_r$ by alternating least squares. After initializing $\mathbf{X}, \mathbf{Y}$ and $\mathbf{D}$, we cycle through the following steps:

1. For fixed $\mathbf{Y}$ and $\mathbf{D}_i$, the global maximum of $\sum_i \text{tr}(\mathbf{S}_i\mathbf{X}\mathbf{D}_i\mathbf{Y}^t)$ subject to $\mathbf{X}^t\mathbf{X} = \mathbf{I}_r$ is given by $\mathbf{X} = \mathbf{U}\mathbf{V}^t$ where $\mathbf{U}$ and $\mathbf{V}$ are derived from the singular value decomposition $\sum_i \mathbf{S}_i\mathbf{Y}\mathbf{D}_i = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^t$.
2. For fixed $\mathbf{X}$ and $\mathbf{D}_i$, the global maximum of $\sum_i \text{tr}(\mathbf{S}_i\mathbf{X}\mathbf{D}_i\mathbf{Y}^t)$ subject to $\mathbf{Y}^t\mathbf{Y} = \mathbf{I}_r$ is given by $\mathbf{Y} = \mathbf{U}\mathbf{V}^t$ where $\mathbf{U}$ and $\mathbf{V}$ are derived from the singular value decomposition $\sum_i \mathbf{S}_i\mathbf{X}\mathbf{D}_i = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^t$.
3. For fixed $\mathbf{X}$ and $\mathbf{Y}$, the global minimum of $\sum_i \|\mathbf{D}_i\|^2 - 2\sum_i \text{tr}(\mathbf{S}_i\mathbf{X}\mathbf{D}_i\mathbf{Y}^t)$ subject to the constraint that $\mathbf{D}_i$ is a diagonal matrix is given by $\mathbf{D}_i = \text{Diag}(\mathbf{Y}^t\mathbf{S}_i\mathbf{X})$.

Indort has become quite popular as a method to suppress so-called degenerate solutions, where two or more columns are proportional in the component matrices (Stegeman 2007).

3. Combining Indort with Non-Negative Least Squares to Update the Saliences

When Indort converges, we have at least conditional optimality for $\mathbf{X}$ and $\mathbf{Y}$, with $\mathbf{X}^t\mathbf{X} = \mathbf{Y}^t\mathbf{Y} = \mathbf{I}_r$. Consider the following nonsingularity assumptions:

a. A subset of n of the matrices $\mathbf{S}_1, \ldots, \mathbf{S}_m$, $1 \leqslant n \leqslant m$, is positive definite and the corresponding matrices $\mathbf{D}_i$ admit a nonsingular linear combination.

b. A subset of n of the matrices $\mathbf{D}_1, \ldots, \mathbf{D}_m$, $1 \leqslant n \leqslant m$, is positive definite and the corresponding matrices $\mathbf{S}_i$ admit a nonsingular linear combination.

When only one $\mathbf{S}_i$ or only one $\mathbf{D}_i$ is positive definite, the corresponding "linear combination" involves just the corresponding matrix $\mathbf{D}_i$ or $\mathbf{S}_i$, respectively.

Result 1. *When* $\mathbf{S}_1, \ldots, \mathbf{S}_m$ *and* $\mathbf{D}_1, \ldots, \mathbf{D}_m$ *are psd, and at least one of the nonsingularity assumptions is met, Indort will return* $\mathbf{X}$ *and* $\mathbf{Y}$ *equal.*

Proof. Suppose that Indort converges to $(\mathbf{X}, \mathbf{Y}, \mathbf{D})$. Since $\mathbf{X}$ is a global maximum of $\mathrm{h}_{\mathbf{Y}}(\mathbf{Z}) = \sum_i \mathrm{tr}(\mathbf{S}_i\mathbf{Y}\mathbf{D}_i\mathbf{Z}^t)$, with $\mathbf{Y}$ fixed, we have

$$\mathrm{tr} \sum_{i=1}^{m} \mathbf{S}_i\mathbf{Y}\mathbf{D}_i\mathbf{X}^t \geqslant \mathrm{tr} \sum_{i=1}^{m} \mathbf{S}_i\mathbf{Y}\mathbf{D}_i\mathbf{Y}^t,$$

and since $\mathbf{Y}$ is a global maximum of $\mathrm{h}_{\mathbf{X}}(\mathbf{Z}) = \sum_i \mathrm{tr}(\mathbf{S}_i\mathbf{X}\mathbf{D}_i\mathbf{Z}^t)$, with $\mathbf{X}$ fixed, we have

$$\mathrm{tr} \sum_{i=1}^{m} \mathbf{S}_i\mathbf{X}\mathbf{D}_i\mathbf{Y}^t \geqslant \mathrm{tr} \sum_{i=1}^{m} \mathbf{S}_i\mathbf{X}\mathbf{D}_i\mathbf{X}^t.$$

Combining these yields

$$2\,\mathrm{tr} \sum_{i=1}^{m} \mathbf{S}_i\mathbf{Y}\mathbf{D}_i\mathbf{X}^t \geqslant \mathrm{tr} \sum_{i=1}^{m} \mathbf{S}_i\mathbf{X}\mathbf{D}_i\mathbf{X}^t + \mathrm{tr} \sum_{i=1}^{m} \mathbf{S}_i\mathbf{Y}\mathbf{D}_i\mathbf{Y}^t,$$

which implies that

$$\sum_{i=1}^{m} \|\mathbf{S}_i^{1/2}\mathbf{X}\mathbf{D}_i^{1/2} - \mathbf{S}_i^{1/2}\mathbf{Y}\mathbf{D}_i^{1/2}\|^2 \leqslant 0. \tag{6}$$

Because (6) is a sum of squares, it follows that $\mathbf{S}_i^{1/2}\mathbf{X}\mathbf{D}_i^{1/2} = \mathbf{S}_i^{1/2}\mathbf{Y}\mathbf{D}_i^{1/2}$, and hence $\mathbf{S}_i\mathbf{X}\mathbf{D}_i = \mathbf{S}_i\mathbf{Y}\mathbf{D}_i$, $i = 1, \ldots, m$, so

$$\mathbf{S}_i(\mathbf{X} - \mathbf{Y})\mathbf{D}_i = \mathbf{0}, \tag{7}$$

$i = 1, \ldots, m$. It is straightforward that each of the two nonsingularity assumptions are now sufficient to infer $\mathbf{X} = \mathbf{Y}$.
■

The assumptions of Result 1 need further qualification. First, when the slices $\mathbf{S}_1, \ldots, \mathbf{S}_m$ are scalar product or covariance matrices derived from data matrices $\mathbf{Z}_1, \ldots, \mathbf{Z}_m$ of rank p, that have been randomly sampled from a continuous distribution, the slices will be nonsingular almost surely. The matrices $\mathbf{D}_1, \ldots, \mathbf{D}_m$ will admit a nonsingular linear combination, when $\mathbf{D}$ has no zero columns (Ten Berge and Tendeiro 2009). Clearly, when $\mathbf{D}$ has a zero column, that would constitute a case of overfactoring because the same fit can then be obtained upon removing that component. Therefore, the first

nonsingularity assumption is typically met when the slices $\mathbf{S}_1, \ldots, \mathbf{S}_m$ are obtained from random sampling.

In some applications, however, the slices are double centered, and neither of the nonsingularity assumptions is met. Fortunately, when double centering is the *only* cause of singularity, Result 1 can still be preserved. Specifically, let $\mathbf{S}_i = \mathbf{J}\mathbf{S}_i\mathbf{J}$, where $\mathbf{J} = \mathbf{I}_p - \mathbf{1}\mathbf{1}^t/p$, the centering operator, and let $\mathbf{S}_i$ have rank $p-1$, $i = 1, \ldots, m$. Then $\sum_i \mathbf{S}_i\mathbf{Y}\mathbf{D}_i$ and $\sum_i \mathbf{S}_i\mathbf{X}\mathbf{D}_i$ are column centered, and the Indort algorithm, using singular vectors of those matrices, will return column-centered matrices $\mathbf{X}$ and $\mathbf{Y}$. Using $\mathbf{1}^t\mathbf{X} = \mathbf{1}^t\mathbf{Y} = \mathbf{0}^t$ in (7) and noting that $\mathbf{S}_i + \mathbf{1}\mathbf{1}^t$ has rank p yields $\mathbf{S}_i(\mathbf{X}-\mathbf{Y})\mathbf{D}_i = (\mathbf{S}_i+\mathbf{1}\mathbf{1}^t)(\mathbf{X}-\mathbf{Y})\mathbf{D}_i = \mathbf{0}, i = 1, \ldots, m$, so $\mathbf{X}\mathbf{D}_i = \mathbf{Y}\mathbf{D}_i$, $i = 1, \ldots, m$. It follows that $\mathbf{X} = \mathbf{Y}$ except in case of overfactoring. It can be concluded that the nonsingularity assumptions are very weak indeed.

The situation is different when it comes to the other assumptions of Result 1. Although the assumption that $\mathbf{S}_1, \ldots, \mathbf{S}_m$ are psd is guaranteed in the context of Indscal, when the slices are scalar product matrices, there is no guarantee that $\mathbf{D}_1, \ldots, \mathbf{D}_m$ will be psd at the global minimum of Candecomp/Parafac (CP) and Indort. Violations of the latter assumption occur quite frequently. For CP, Ten Berge, Kiers and Krijnen (1993) replaced the unconstrained update of $\mathbf{D}$ by a Non Negative Least Squares (NNLS) algorithm. Because this still did not guarantee $\mathbf{X} = \mathbf{Y}$ at the minimum, they proposed a symmetry preserving algorithm (Sympres) based on updating the columns of $\mathbf{X}$ one by one. In the context of Indscal, we can similarly impose nonnegativity on $\mathbf{D}$, by adopting NNLS to supplant step 3 of the Indort algorithm. However, when the slices are psd, there is no need for a symmetry-preserving alternative algorithm, because Result 1 already guarantees that Indort, adapted to include NNLS, yields $\mathbf{X} = \mathbf{Y}$.

The process of double centering may turn slices that are psd into indefinite slices. When this happens, Result 1 does not apply unless a transformation is applied that restores the property of slices being psd. For instance see Bénasséni, Bennani Dosse and Joly (2007).

An anonymous referee pointed out that the first nonsingularity assumption discussed above cannot be met in special cases, such as when the slices fit the Indort model perfectly. This still does not imply that Indort will return $\mathbf{X}$ and $\mathbf{Y}$ different. Firstly, in that case some of the diagonal matrices $\mathbf{D}_1, \ldots, \mathbf{D}_m$ are likely to be positive definite and the second condition may still apply. But even when neither the first nor the second condition applies, we will have $\mathbf{X}$ and $\mathbf{Y}$ equal in the perfect Indort case. Specifically, suppose the slices satisfy the Indort model and are psd. If we find $\mathbf{S}_i = \mathbf{X}\mathbf{D}_i\mathbf{Y}^t$, with $\mathbf{X}$ and $\mathbf{Y}$ potentially different, then $\mathbf{S}_i^2 = \mathbf{X}\mathbf{D}_i\mathbf{Y}^t\mathbf{Y}\mathbf{D}_i\mathbf{X}^t = \mathbf{X}\mathbf{D}_i^2\mathbf{X}^t$ so $\mathbf{S}_i = \mathbf{X}\mathbf{D}_i\mathbf{X}^t$ because negative roots of elements of $\mathbf{D}_i^2$ would contradict that the $\mathbf{S}_i$ are psd. So we have a solution with $\mathbf{X} = \mathbf{Y}$.

4. Successive Approach

The direct minimization of (1) subject to the constraint $\mathbf{X}^t\mathbf{X} = \mathbf{I}_r$ is sometimes called a simultaneous approach since all the components of $\mathbf{X}$ are computed simultaneously. The idea of a successive approach is to calculate the first component, then remove the influence of this component from all matrices $\mathbf{S}_i$, $i = 1, \ldots, m$ and subsequently calculate the next component from the corrected matrices. These calculations are repeated until some specified criterion is reached. The successive approach is much used in chemometrics (Hanafi, Mazerolles, Dufour, and Qannari 2006; Mazerolles, Hanafi, Dufour, Qannari and Bertrand 2006) and sensory analysis (Qannari, Wakeling, Courcoux, and MacFie 2000). It is often believed that the successive approach is just a computational alternative to optimizing $\mathbf{X}$ simultaneously. This belief, however, is unwarranted. To clarify this, it will be shown in Section 4.1 and 4.2 that solutions given by the successive approach are not optimal for the function *g* in (1).

The successive approach to minimizing (1) works as follows. Minimizing (1) can be written as minimizing $\sum_i \|\mathbf{E}_i\|^2$, when

$$\mathbf{S}_i = \sum_{\ell=1}^{r} d_{i\ell}\mathbf{x}_\ell\mathbf{x}_\ell^t + \mathbf{E}_i, \qquad i = 1, \ldots, m.$$

The first step consists of seeking a unit length vector $\mathbf{x}_1$ and its associated saliences d_{i1}, $i = 1, \ldots, m$ that minimize the function

$$\sum_i \|\mathbf{S}_i - d_{i1}\mathbf{x}_1\mathbf{x}_1^t\|^2. \tag{8}$$

At step ν, $\nu = 2, \ldots, r$, a vector $\mathbf{x}_\nu$ and its associated saliences $d_{i\nu}$, $i = 1, \ldots, m$ are sought such that the following loss function is minimized :

$$\sum_i \|\widetilde{\mathbf{S}}_i^{(\nu)} - d_{i\nu}\mathbf{x}_\nu\mathbf{x}_\nu^t\|^2, \tag{9}$$

subject to $\mathbf{x}_\nu^t\mathbf{x}_\nu = 1$ and $\mathbf{x}_\nu^t\mathbf{x}_\ell = 0$, $\ell = 1, \ldots, \nu - 1$, where

$$\widetilde{\mathbf{S}}_i^{(\nu)} = \mathbf{S}_i - \sum_{\ell=1}^{\nu-1} d_{i\ell}\mathbf{x}_\ell\mathbf{x}_\ell^t. \tag{10}$$

The successive Indort algorithm differs from Sympres (Ten Berge, Kiers, and Krijnen 1993) in two respects. Firstly, Sympres has no constraint of orthogonality on $\mathbf{X}$. Secondly, Sympres starts with a random initial choice of $\mathbf{X}$, and then updates the components (columns of $\mathbf{X}$) one by one

in iterative cycles, till each component and its associated saliences are conditionally optimal given all other components and saliences. The successive Indort method, on the other hand, does impose the constraint $\mathbf{X}^t\mathbf{X} = \mathbf{I}_r$, and it determines the components one by one in a single sweep of r computations. Indeed, if the sweeps were to be repeated like in Sympres, the optimality properties to be derived in 4.1 would not be different from those of the simultaneous approach to Indort.

4.1. Optimality Properties

In this section, it is shown that the solution given by the successive approach is not a local minimum of *g* subject to the orthogonality constraint. As a first step, we express *g* in a simpler form by eliminating $\mathbf{D}_1, \ldots, \mathbf{D}_m$. Specifically, when $\mathbf{X}^t\mathbf{X} = \mathbf{I}_r$, then

$$g = \sum_i \|\mathbf{S}_i\|^2 + \|\mathbf{D}_i\|^2 - 2\,\mathrm{tr}(\mathbf{S}_i\mathbf{X}\mathbf{D}_i\mathbf{X}^t),$$

but

$$\mathbf{X}\mathbf{D}_i\mathbf{X}^t = \sum_{\ell=1}^{r} d_{i\ell}\mathbf{x}_\ell\mathbf{x}_\ell^t,$$

so *g* can be written as

$$g = \sum_i \|\mathbf{S}_i\|^2 + \sum_i\sum_\ell d_{i\ell}^2 - 2\sum_i\sum_\ell d_{i\ell}\mathbf{x}_\ell^t\mathbf{S}_i\mathbf{x}_\ell.$$

For arbitrary but fixed $\mathbf{X}$, the conditionally optimal choice of $d_{i\ell}$ is

$$d_{i\ell} = \mathbf{x}_\ell^t\mathbf{S}_i\mathbf{x}_\ell,$$

so minimizing *g* is equivalent to maximizing

$$G(\mathbf{X}) = \sum_i\sum_\ell (\mathbf{x}_\ell^t\mathbf{S}_i\mathbf{x}_\ell)^2, \tag{11}$$

subject to $\mathbf{X}^t\,\mathbf{X} = \mathbf{I}_r$.

Result 2. *The maximization problem* (11) *admits a global maximum.*

Proof. Since the function *G* is continuous and the set of orthonormal matrices is compact we have the result.
■

Thus the minimization of (1) subject to orthogonality constraint $\mathbf{X}^t\mathbf{X} = \mathbf{I}_r$ has always at least one solution. This is an important difference with the general case (i.e nonorthogonal case.), see Ten Berge, Kiers, and De Leeuw (1988).

As a second step, we consider the first order necessary conditions for a maximum of Indort. Taking partial derivatives of (11) shows that it is necessary for a maximum of Indort that

$$\mathrm{A}(\mathbf{X}) = \mathbf{X}\mathbf{C}, \tag{12}$$

where $\mathbf{C}$ is a symmetric matrix and $\mathrm{A}(\mathbf{X})$ is a matrix whose columns $\mathbf{u}_\ell$ ($\ell = 1, \ldots, p$) are given by

$$\mathbf{u}_\ell = \sum_i (\mathbf{x}_\ell^t \mathbf{S}_i \mathbf{x}_\ell)\mathbf{S}_i \mathbf{x}_\ell. \tag{13}$$

So Indort has $\mathbf{X}^t\mathrm{A}(\mathbf{X})$ symmetric at every maximum of (11). Finally, the next result shows that a solution to the successive approach is usually neither a local nor a global maximum of (11).

Result 3. *If* $\mathbf{X}$ *is a solution to the successive approach then* $\mathbf{X}^t A(\mathbf{X})$ *is an upper-triangular matrix.*

Proof. First, we show that $\mathbf{x}_j^t \mathbf{u}_1 = 0$ for $j = 2, \ldots, p$. It is obvious that if $\mathbf{x}_1$ is the maximum of $\sum_i (\mathbf{x}^t \mathbf{S}_i \mathbf{x})^2$ subject to $\mathbf{x}^t\mathbf{x} = 1$ then there exists a scalar λ_1 such that

$$\mathbf{u}_1 = \sum_i \mathbf{S}_i \mathbf{x}_1 \mathbf{x}_1^t \mathbf{S}_i \, \mathbf{x}_1 = \lambda_1 \, \mathbf{x}_1. \tag{14}$$

So $\mathbf{x}_j^t \mathbf{u}_1 = 0$ for $j > 1$. At step ν, $\mathbf{x}_\nu$ is the maximum of

$$\mathbf{x}^t \left(\sum_i \widetilde{\mathbf{S}}_i^{(\nu)} \mathbf{x}\mathbf{x}^t \widetilde{\mathbf{S}}_i^{(\nu)} \right) \mathbf{x}, \tag{15}$$

subject to the constraints $\mathbf{x}^t\mathbf{x} = 1$ and $\mathbf{x}^t\mathbf{x}_\ell = 0$, $\ell = 1, \ldots, \nu - 1$. The first necessary optimality condition states that there exist a λ_2 and $\mu_1, \ldots, \mu_{\nu-1}$ such that

$$\sum_i \widetilde{\mathbf{S}}_i^{(\nu)} \mathbf{x}_\nu \mathbf{x}_\nu^t \widetilde{\mathbf{S}}_i^{(\nu)} \, \mathbf{x}_\nu = \lambda \, \mathbf{x}_\nu + \sum_{\ell=1}^{\nu-1} \mu_\ell \mathbf{x}_\ell. \tag{16}$$

But it is clear (from (10)) that $\widetilde{\mathbf{S}}_i^{(\nu)} \mathbf{x}_\nu = \mathbf{S}_i \mathbf{x}_\nu$ for $i = 1, \ldots, m$ so we have

$$\mathbf{u}_\nu = \sum_i \mathbf{S}_i \mathbf{x}_\nu \mathbf{x}_\nu^t \mathbf{S}_i \, \mathbf{x}_\nu = \lambda \, \mathbf{x}_\nu + \sum_{\ell=1}^{\nu-1} \mu_\ell \mathbf{x}_\ell, \tag{17}$$

and consequently $\mathbf{x}_j^t \mathbf{u}_\nu = 0$ for $j > \nu$. It follows that $\mathbf{X}^t\mathrm{A}(\mathbf{X})$ is upper-triangular.

■

It has thus been shown that the successive approach cannot yield, in general, the optimal solution of the Indort function. The explanation for this is as follows. In both approaches, the columns of $\mathbf{X}$ are constrained to be of unit length and orthogonal. In the successive approach, however, there are additional constraints: The first column of $\mathbf{X}$ must be optimal for the $r = 1$ version of the problem; the second column of $\mathbf{X}$ must be optimal for the $r = 1$ version of the problem in the orthogonal complement space of the first column, and so on. These additional constraints explain why, in general, the successive approach cannot attain the same level of fit as Indort.

To conclude this subsection, it should be noted that there are situations in which the simultaneous and the successive approaches give identical result (up to permutation of the columns of $\mathbf{X}$). This is the case when $m = 1$ (i.e. one matrix $\mathbf{S}$) or when $\mathbf{S}_1, \ldots, \mathbf{S}_m$ can be simultaneously diagonalized. In those cases, it can be easily seen that the matrix $\mathbf{X}^t \mathrm{A}(\mathbf{X})$ involved in Result 3 will be diagonal (and therefore will be both symmetric and upper-triangular).

4.2. Empirical Comparison of the Two Approaches

To compare the results obtained from applying simultaneous and successive approaches, three Monte Carlo and three real data sets were used. The Monte Carlo data are $m = 6$ matrices of order 8×4 generated randomly from a uniform, normal and mixed (uniform and normal) distributions respectively and $\mathbf{S}_1, \ldots, \mathbf{S}_6$ were obtained as scalar product matrices, scaled to unit sums of squares. Both methods were initialized with $\mathbf{X}$ as an orthonormalized version of a matrix of random numbers from the uniform $[-1, +1]$ distribution and were run to convergence of (1) in nine decimal places. Convergence was extremely fast with both methods, never taking more than 57 iterations.

In Table 1 we report the residual sums of squares (1) for the simultaneous and the successive approaches and their relative differences, for 2, 3, and 4 components, respectively. It is clear that the simultaneous approach outperforms the successive approach throughout.

Next, three real data sets from the field of food and sensory analysis were used. In the first data set, provided by Williams and Langron (1984), 8 wines are evaluated by 6 judges. The second data set, discussed by Dijksterhuis and Punter (1990), deals with a study in which 7 assessors rated eight yogurts on different variables. The last data set, which can be found in the database of Chessel and Doledec (1995), consists of various measurements made in a number of sampling points in the French river "The Doubs". In Table 2 we report the results. Clearly, the simultaneous approach yields better fit again, albeit that the differences are puny. Moreover, the difference

Table 1. Residual sums of squares for Monte Carlo data.

Simulation	**1**			**2**			**3**		
r	2	3	4	2	3	4	2	3	4
Simultaneous	3.313	2.374	1.741	3.100	2.388	1.983	3.617	2.576	2.007
Successive	3.321	2.602	2.118	3.110	2.422	2.070	3.670	2.697	2.358
Rel. diff.	0.002	0.096	0.217	0.003	0.014	0.044	0.015	0.047	0.175

Table 2. Residual sums of squares for real life data.

Data set	**wines**			**yoghurts**			**river**		
r	2	3	4	2	3	4	2	3	4
Simultaneous	0.918	0.550	0.534	2.362	1.892	1.755	5.629	5.141	4.539
Successive	0.931	0.580	0.564	2.364	1.932	1.795	5.727	5.256	4.680
Rel. diff.	0.014	0.056	0.057	0.001	0.021	0.023	0.017	0.022	0.031

seems to grow with the number of components retained for the three data sets at hand.

5. Concluding Remarks

In this paper we have shown that Indort, a version of CP that imposes orthogonality on the columns of $\mathbf{X}$ and $\mathbf{Y}$, when applied to psd matrices, will return $\mathbf{X}$ and $\mathbf{Y}$ equal if the saliences are constrained to be non-negative. The resulting method is thus quite attractive for Indscal problems, because the problems of negative saliences, degenerate solutions, and different $\mathbf{X}$ and $\mathbf{Y}$ at convergence are all suppressed. Also, we have discussed a successive version of Indort, and shown that it has different minima and yields inferior fit to its simultaneous counterpart. This should remove a long standing source of confusion.

References

BENASSENI, J., BENNANI DOSSE, M., and JOLY, S. (2007), "On a General Transformation Making a Dissimilarity Matrix Euclidean", *Journal of Classification, 24(1)*, 33–51.

BENNANI DOSSE, M., and TEN BERGE, J.M.F. (2008), "The Assumption of Proportional Components When Candecomp Is Applied to Symmetric Matrices in the Context of Indscal", *Psychometrika, 73*, 303–308.

BOLLA, M., MICHALETZKY, G., TUSNÁDY, G., and ZIERMANN, M. (1998), "Extrema of Sums of Heterogeneous Quadratic Forms", *Linear Algebra And Its Applications, 269*, 331–365.

CARROLL, J.D., and CHANG, J.J. (1970), "Analysis of Individual Differences in Multidimensional Scaling via an n-way Generalization of Eckart-Young Decomposition", *Psychometrika, 35*, 283–319.

CHESSEL, D., and DOLEDEC, S. (1995), *ADE Software Hypercard Stacks and Quick-basic Microsoft Program Library for the Analysis of Environnemental Data* (Version 4 Documentation ESA), Lyon University: CNRS 5023.

DIJKSTERHUIS, G.B., and PUNTER, P.H. (1990), "Interpreting Generalized Procrustes Analysis 'Analysis of variance' Tables", *Food Quality and Preference, 2*, 255–265.

HANAFI, M., MAZEROLLES, G., DUFOUR, E., and QANNARI, E.M. (2006), "Common Components and Specific Weights Analysis and Multiple Co-inertia Analysis Applied to Coupling of Several Measurement Techniques", *Journal of Chemometrics, 20(5)*, 172–183.

KIERS, H.A.L. (1989), *Three-way Methods for the Analysis of Qualitative and Quantitative Two-way Data*, Leiden: DSWO Press.

KROONENBERG, P.M. (1983), *Three Mode Principal Component Analysis : Theory and Applications*, Leiden: DSWO Press.

MAZEROLLES, G., HANAFI, M., DUFOUR, E., QANNARI, E.M., and BERTRAND, D. (2006), "Common Components and Specific Weights Analysis : A Chemometric Method for Dealing with Complexity of Food Products", *Chemometrics and Intelligent Laboratory Systems, 81*, 41–49.

QANNARI, E.M., WAKELING, I., COURCOUX, P., and MACFIE, H.J.H. (2000), "Defining the Underlying Sensory Dimensions, *Food Quality and Preference, 11*, 151–154.

STEGEMAN, A. (2007), "Degeneracy in Candecomp/Parafac and Indscal Explained for Several Three-sliced Arrays with a Two-valued Typical Rank", *Psychometrika, 72*, 601–619.

TEN BERGE, J.M.F., and KIERS, H.A.L. (1991), "Some Clarifications of the Candecomp Algorithm Applied to Indscal", *Psychometrika, 56*, 317–326.

TEN BERGE, J.M.F., KIERS, H.A.L., and DE LEEUW, J. (1988), "Explicit Candecomp/Parafac Solutions for a Contrived $2 \times 2 \times 2$ Array of Rank Three", *Psychometrika, 53*, 579–584.

TEN BERGE, J.M.F., KIERS, H.A.L., and KRIJNEN, W.P. (1993), "Computational Solutions for the Problem of Negative Saliences and Nonsymmetry in Indscal", *Journal of Classification, 10*, 115–124.

TEN BERGE, J.M.F., KNOL, D. L., and KIERS, H.A.L. (1988), "A Treatment of the Orthomax Rotation Family in Terms of Diagonalization, and a Re-examination of a Singular Value Approach to Varimax Rotation", *Computational Statistics Quarterly, 3*, 207–217.

TEN BERGE, J.M.F., and TENDEIRO, J.N. (2009), "The Link Between Sufficient Conditions by Harshman and by Kruskal for Uniqueness in Candecomp/Parafac", *Journal of Chemometrics, 23*, 321–323.

TENDEIRO, J., BENNANI DOSSE, M., and TEN BERGE, J.M.F. (2010), "First and Second-order Derivatives for CP and INDSCAL", *Chemometrics and Intelligent Laboratory Systems*, doi:10.1016/j.chemolab.2010.05.013.

WILLIAMS, A., and LANGRON, S.P. (1984), "The Use of Free-choice Profiling for the Evaluation of Commercial Ports", *Journal of the Science of Food and Agriculture, 35*, 558–568.