

THE ASSUMPTION OF PROPORTIONAL COMPONENTS WHEN CANDECOMP IS APPLIED TO SYMMETRIC MATRICES IN THE CONTEXT OF INDSCAL

MOHAMMED BENNANI DOSSE

UNIVERSITY OF RENNES 2

JOS M.F. TEN BERGE

UNIVERSITY OF GRONINGEN

The use of Candecomp to fit scalar products in the context of INDSCAL is based on the assumption that the symmetry of the data matrices involved causes the component matrices to be equal when Candecomp converges. Ten Berge and Kiers gave examples where this assumption is violated for Gramian data matrices. These examples are believed to be local minima. It is now shown that, in the single-component case, the assumption can only be violated at saddle points. Chances of Candecomp converging to a saddle point are small but still nonzero.

Key words: Candecomp, Parafac, Indscal

Introduction

Carroll and Chang (1970) developed Candecomp as a method of fitting scalar products derived from INDSCAL. Specifically, they sought to minimize the function

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

where \mathbf{S}_i is a symmetric $p \times p$ matrix of (true or pseudo) scalar products, \mathbf{X} is a $p \times r$ matrix of components, and \mathbf{D}_i is a diagonal $r \times r$ matrix of saliencies, with elements of row i of \mathbf{D} in the diagonal, $i = 1, \dots, m$. Because direct minimization of g seems difficult, Carroll and Chang instead proposed Candecomp, to minimize

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

They assumed that, at the minimum, the symmetry of the slices \mathbf{S}_i will cause \mathbf{X} and \mathbf{Y} to be equal or at least 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 columns of \mathbf{D} . The Candecomp algorithm minimizes f 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} column-wise proportional at the minimum seems warranted. However, for contrived data, counterexamples do exist. In particular, the single-component case ($r = 1$) has played a role in contriving counterexamples.

Requests for reprints should be sent to Mohammed Bennani Dosse, Statistics Research Team, University of Rennes 2, Rennes, France. E-mail: mohamed.bennani@uhb.fr

Ten Berge and Kiers (1991) examined the single-component case, where \mathbf{X} , \mathbf{Y} , and \mathbf{D} can be represented as vectors \mathbf{x} , \mathbf{y} , \mathbf{d} . They constructed Gramian matrices \mathbf{S}_1 and \mathbf{S}_2 , and a solution $\{\mathbf{x}, \mathbf{y}, \mathbf{d}\}$ for which the derivatives of f vanish, yet the claim that \mathbf{x} and \mathbf{y} will be proportional is unwarranted. They also showed that such cases could not possibly represent a global minimum for f , and conjectured that they could arise only at local minima. However, they seem to have ignored saddle points as another possibility. In the present note it is shown that, when $\mathbf{S}_1, \dots, \mathbf{S}_m$ are Gramian, nonproportional components in the single-component case can be obtained only for saddle points. This is more than merely a theoretical exercise: Because saddle points seem far more difficult to attain by Candecomp than local minima, the practical relevance of counterexamples where \mathbf{x} and \mathbf{y} are nonproportional is thus further reduced. We start by examining derivatives. Throughout, we use \sum to denote $\sum_{i=1}^m$.

First- and Second-Order Derivatives of f

The conditionally optimal solutions for \mathbf{X} , \mathbf{Y} , and \mathbf{D} are those that satisfy the equations

$$\frac{\partial f}{\partial \mathbf{X}} = 2\mathbf{X} \sum \mathbf{D}_i \mathbf{Y}' \mathbf{Y} \mathbf{D}_i - 2 \sum \mathbf{S}_i \mathbf{Y} \mathbf{D}_i = \mathbf{O}, \tag{3}$$

$$\frac{\partial f}{\partial \mathbf{Y}} = 2\mathbf{Y} \sum \mathbf{D}_i \mathbf{X}' \mathbf{X} \mathbf{D}_i - 2 \sum \mathbf{S}_i \mathbf{X} \mathbf{D}_i = \mathbf{O}, \tag{4}$$

and

$$\frac{\partial f}{\partial \mathbf{D}_i} = 2(\mathbf{X}' \mathbf{X} * \mathbf{Y}' \mathbf{Y}) \text{Vec}(\text{Diag}(\mathbf{D}_i)) - 2 \text{Vec}(\text{Diag}(\mathbf{X}' \mathbf{S}_i \mathbf{Y})) = \mathbf{O}, \tag{5}$$

where $*$ is the element-wise (Hadamard) product of matrices, and $\text{Diag}(\text{Vec}(\mathbf{D}_i))$ the vector holding the diagonal elements of \mathbf{D}_i . In the $r = 1$ case, these equations simplify to

$$2 \sum d_i^2 (\mathbf{y}' \mathbf{y}) \mathbf{x} - 2 \sum d_i \mathbf{S}_i \mathbf{y} = \mathbf{0}, \tag{6}$$

$$2 \sum d_i^2 (\mathbf{x}' \mathbf{x}) \mathbf{y} - 2 \sum d_i \mathbf{S}_i \mathbf{x} = \mathbf{0}, \tag{7}$$

and

$$2d_i \mathbf{x}' \mathbf{x} \mathbf{y}' \mathbf{y} - 2\mathbf{x}' \mathbf{S}_i \mathbf{y} = 0, \tag{8}$$

where d_i is the only element of \mathbf{D}_i . The second-order derivative matrix for the $r = 1$ case is

$$\begin{bmatrix} 2\mathbf{y}' \mathbf{y} (\mathbf{d}' \mathbf{d}) \mathbf{I}_p & 4(\mathbf{d}' \mathbf{d}) \mathbf{x} \mathbf{y}' - 2 \sum d_i \mathbf{S}_i & 4\mathbf{y}' \mathbf{y} (\mathbf{x} \mathbf{d}') - 2[\mathbf{S}_1 \mathbf{y} \dots \mathbf{S}_m \mathbf{y}] \\ 4(\mathbf{d}' \mathbf{d}) \mathbf{y} \mathbf{x}' - 2 \sum d_i \mathbf{S}_i & 2\mathbf{x}' \mathbf{x} (\mathbf{d}' \mathbf{d}) \mathbf{I}_p & 4(\mathbf{x}' \mathbf{x}) \mathbf{y} \mathbf{d}' - 2[\mathbf{S}_1 \mathbf{x} \dots \mathbf{S}_m \mathbf{x}] \\ 4(\mathbf{y}' \mathbf{y}) \mathbf{d} \mathbf{x}' - 2[\mathbf{S}_1 \mathbf{y} \dots \mathbf{S}_m \mathbf{y}]' & 4(\mathbf{x}' \mathbf{x}) \mathbf{d} \mathbf{y}' - 2[\mathbf{S}_1 \mathbf{x} \dots \mathbf{S}_m \mathbf{x}]' & 2\mathbf{x}' \mathbf{x} \mathbf{y}' \mathbf{y} \mathbf{I}_m \end{bmatrix}, \tag{9}$$

where \mathbf{d} is the vector holding d_1, \dots, d_m . When this matrix is positive definite (i.e., having all eigenvalues positive), we have a local or global minimum. When it is indefinite (i.e., having positive and negative eigenvalues), we have a saddle point. Consider the following contrived data and solution of Ten Berge and Kiers (1991), with \mathbf{x} and \mathbf{y} nonproportional. Let

$$\mathbf{S}_1 = \begin{bmatrix} 3 & 1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{S}_2 = \begin{bmatrix} 3 & -1 & 0 \\ -1 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{10}$$

$$\mathbf{x} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \text{and} \quad \mathbf{d} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

It can be verified that the first-order derivatives vanish. Ten Berge and Kiers (1991, p. 324) conjectured from their Result 6 that such cases with \mathbf{x} and \mathbf{y} nonproportional occur only at local minima of f . However, the matrix of second-order derivatives (9) is

$$\left[\begin{array}{ccc|ccc|cc} 4 & 0 & 0 & 0 & 4 & 0 & 2 & -2 \\ 0 & 4 & 0 & -4 & 0 & 0 & -6 & -6 \\ \hline 0 & 0 & 4 & 0 & 0 & 2 & 0 & 0 \\ 0 & -4 & 0 & 4 & 0 & 0 & -6 & -6 \\ 4 & 0 & 0 & 0 & 4 & 0 & 2 & -2 \\ \hline 0 & 0 & 2 & 0 & 0 & 4 & 0 & 0 \\ 2 & -6 & 0 & -6 & 2 & 0 & 2 & 0 \\ \hline -2 & -6 & 0 & -6 & -2 & 0 & 0 & 2 \end{array} \right]$$

with eigenvalues 13.04, 10, 8, 6, 2, 0, 0, and -11.04 , so it is indefinite. Therefore, we are looking at a saddle point. This is not just a property of this particular example. All stationary points of f with \mathbf{x} and \mathbf{y} nonproportional are saddle points. To prove this, we might try to show that (9) is indefinite whenever (6), (7), and (8) are satisfied for a nonproportional pair \mathbf{x} and \mathbf{y} . It is, however, much easier to use an alternative approach to minimizing f .

An Alternative Approach for the $r = 1$ Case

Instead of using Candecomp to minimize f , the $r = 1$ case can be solved by constraining \mathbf{x} and \mathbf{y} to be of unit length, and expressing d_i as $\mathbf{x}'\mathbf{S}_i\mathbf{y}$, $i = 1, \dots, m$, see (8). Then minimizing f is reduced to maximizing

$$h(\mathbf{x}, \mathbf{y}) = \sum (\mathbf{x}'\mathbf{S}_i\mathbf{y})^2 \tag{11}$$

subject to the constraint $\mathbf{x}'\mathbf{x} = \mathbf{y}'\mathbf{y} = 1$. Clearly, for \mathbf{y} fixed, \mathbf{x} can be optimized conditionally as the eigenvector associated with the largest eigenvalue of $\sum \mathbf{S}_i\mathbf{y}\mathbf{y}'\mathbf{S}_i$ and, for \mathbf{x} fixed, the conditionally optimal \mathbf{y} is the eigenvector associated with the largest eigenvalue of $\sum \mathbf{S}_i\mathbf{x}\mathbf{x}'\mathbf{S}_i$. Because \mathbf{d} is updated implicitly, this algorithm optimizes $\{\mathbf{x}, \mathbf{d}\}$ and $\{\mathbf{y}, \mathbf{d}\}$, iteratively. The first-order derivatives of the Lagrangian $L(\mathbf{x}, \mathbf{y}) = \sum (\mathbf{x}'\mathbf{S}_i\mathbf{y})^2 - \lambda_1(\mathbf{x}'\mathbf{x} - 1) - \lambda_2(\mathbf{y}'\mathbf{y} - 1)$ are

$$\frac{\partial L}{\partial \mathbf{x}} = 2 \sum \mathbf{S}_i\mathbf{y}\mathbf{y}'\mathbf{S}_i\mathbf{x} - 2\lambda_1\mathbf{x}, \tag{12}$$

and

$$\frac{\partial L}{\partial \mathbf{y}} = 2 \sum \mathbf{S}_i\mathbf{x}\mathbf{x}'\mathbf{S}_i\mathbf{y} - 2\lambda_2\mathbf{y}, \tag{13}$$

and the matrix of second-order derivatives is

$$\frac{\partial^2 L}{\partial \mathbf{x}\partial \mathbf{y}} = 2 \left[\begin{array}{cc} \sum \mathbf{S}_i\mathbf{y}\mathbf{y}'\mathbf{S}_i - \lambda_1\mathbf{I}_3 & \sum (\mathbf{x}'\mathbf{S}_i\mathbf{y})\mathbf{S}_i + \sum \mathbf{S}_i\mathbf{y}\mathbf{x}'\mathbf{S}_i \\ \sum (\mathbf{x}'\mathbf{S}_i\mathbf{y})\mathbf{S}_i + \sum \mathbf{S}_i\mathbf{y}\mathbf{x}'\mathbf{S}_i & \sum \mathbf{S}_i\mathbf{x}\mathbf{x}'\mathbf{S}_i - \lambda_2\mathbf{I}_3 \end{array} \right]. \tag{14}$$

It is immediate from (12) and (13) that $\lambda_1 = \lambda_2 = \sum (\mathbf{x}'\mathbf{S}_i\mathbf{y})^2$ when first-order derivatives vanish. Let the columns of a $2p \times 2$ matrix \mathbf{W} span the subspace orthogonal to the columns of $\begin{bmatrix} 2\mathbf{x} & \mathbf{0} \\ \mathbf{0} & 2\mathbf{y} \end{bmatrix}$.

When the first-order derivatives of L vanish, we have a maximum if $\mathbf{W}'\frac{\partial^2 L}{\partial \mathbf{x}\partial \mathbf{y}}\mathbf{W}$ is negative definite, and a saddle point if it is indefinite. For the data of (10), that matrix is

$$\begin{bmatrix} 32 & 0 & 40 & 0 \\ 0 & -4 & 0 & -2 \\ 40 & 0 & 32 & 0 \\ 0 & -2 & 0 & -4 \end{bmatrix}, \tag{15}$$

with eigenvalues 72, -2, -6, and -8. It is indefinite, implying a saddle point like before. It will now be shown that *all* cases where \mathbf{x} and \mathbf{y} are nonproportional are saddle points of h .

Result. *When a solution for $h(\mathbf{x}, \mathbf{y})$ is a local maximum, it has conditional optimality. Such a solution has $\mathbf{x} = \pm\mathbf{y}$ when any subset of $\mathbf{S}_1, \dots, \mathbf{S}_m$ are Gramian and admit a nonsingular linear combination.*

Proof: Suppose the solution (\mathbf{x}, \mathbf{y}) is a local maximum of h . Then \mathbf{x} is a local maximum of $h_{\mathbf{y}}(\mathbf{z}) = \mathbf{z}' \sum \mathbf{S}_i \mathbf{y} \mathbf{y}' \mathbf{S}_i \mathbf{z}$, with \mathbf{y} fixed. Define $\mathbf{A}_{\mathbf{y}} = \sum \mathbf{S}_i \mathbf{y} \mathbf{y}' \mathbf{S}_i$, with eigendecomposition $\mathbf{A}_{\mathbf{y}} = \mathbf{K} \mathbf{\Lambda} \mathbf{K}'$, where $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$. Define $\mathbf{u} = \mathbf{K}' \mathbf{z}$. Then \mathbf{u} is a local maximum of $\mathbf{u}' \mathbf{\Lambda} \mathbf{u}$ subject to $\mathbf{u}' \mathbf{u} = 1$. Because the first-order derivatives vanish we have

$$\mathbf{\Lambda} \mathbf{u} = \lambda \mathbf{u} \tag{16}$$

for some eigenvalue λ and eigenvector \mathbf{u} of $\mathbf{\Lambda}$. The second derivatives matrix is $2(\mathbf{\Lambda} - \lambda \mathbf{I})$. At a local maximum, it is at least negative semidefinite on the subspace orthogonal to \mathbf{u} . That is, for all \mathbf{h} orthogonal to \mathbf{u} we have $\mathbf{h}'(\mathbf{\Lambda} - \lambda \mathbf{I})\mathbf{h} \leq 0$. When we pick the largest eigenvalue and associated eigenvector \mathbf{e}_1 , the first column of the identity matrix, this inequality is satisfied. However, when we pick any other eigenvalue and associated eigenvector, $\mathbf{e}'_1(\mathbf{\Lambda} - \lambda \mathbf{I})\mathbf{e}_1$ will be strictly positive. It follows that a local maximum has $\mathbf{u} = \mathbf{e}_1$ hence \mathbf{x} equal to the principal eigenvector of $\mathbf{A}_{\mathbf{y}}$. That is, \mathbf{x} is conditionally optimal for h given \mathbf{y} . The entire derivation can be repeated with roles of \mathbf{x} and \mathbf{y} reversed, showing that we have conditional optimality also for \mathbf{y} at local maxima of h . Finally, Result 6 of Ten Berge and Kiers (1991) implies that every solution with conditional optimality for \mathbf{x} and \mathbf{y} has $(\mathbf{x}' \mathbf{S}_i \mathbf{x})^2 = (\mathbf{y}' \mathbf{S}_i \mathbf{y})^2$. When any subset of $\mathbf{S}_1, \dots, \mathbf{S}_m$ are Gramian, and admit a nonsingular linear combination, it follows that $\mathbf{x} = \pm\mathbf{y}$. \square

The meaning of this result is that at stationary points of h in the case of Gramian matrices, \mathbf{x} and \mathbf{y} can be nonproportional only at saddle points. This also extends to the more general function f . When that function has a local minimum, h has a local maximum, which means that $\mathbf{x} = \pm\mathbf{y}$ is guaranteed. As a result, a comment in Ten Berge and Kiers (1991, p. 324) can now be sharpened: Where it was stated that asymmetry in the present examples can occur only *at local maxima*, it is now clear that asymmetry can occur only *at saddle points*. This fact reduces the probability of ever finding asymmetry in practical applications. Still, it does happen now and then that Candecom converges to saddle points. For instance, when the data are \mathbf{S}_1 and \mathbf{S}_2 of (13), we have a saddle point at

$$\mathbf{x}_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{y}_0 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \text{and} \quad \mathbf{d}_0 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}. \tag{17}$$

When Candecom is started at points where \mathbf{x} is replaced by $\mathbf{x}_0 + t \times \text{rand}(-0.5, +0.5)$, updating \mathbf{y} and \mathbf{d} first, and t is picked as small as 0.0001, it converges to the saddle point in 7% of the cases.

Discussion

The assumption that Candecom, when applied to symmetric slices, will converge to solutions with two component matrices equal up to column scaling remains unsettled for $r > 1$. However, the counterexamples presented by Ten Berge and Kiers are saddle points rather than

local minima. This is reassuring, because Candecomp, for $r = 1$, tends to get stuck at local minima much easier than at saddle points. Still, the probability of convergence to a saddle point, when Candecomp starts very close to one, is strictly positive for $r = 1$.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Carroll, J.D., & Chang, J.J. (1970). Analysis of individual differences in multidimensional scaling via an n -way generalization of Eckart–Young decomposition. *Psychometrika*, *35*, 283–319.
- Ten Berge, J.M.F., & Kiers, H.A.L. (1991). Some clarifications of the Candecomp algorithm applied to Indscal. *Psychometrika*, *56*, 317–326.

Manuscript received 28 FEB 2007

Final version received 21 SEP 2007

Published Online Date: 25 JAN 2008