JOURNAL OF CHEMOMETRICS
J. Chemometrics 2000; **14**: 261–274

On global, local and stationary solutions in three-way data analysis

René Henrion*

Weierstrass Institute Berlin, D-10117 Berlin, Germany

SUMMARY

The issue of global and local solutions to optimization problems is of much interest in the context of three-way analysis, in particular when dealing with the PARAFAC and Tucker3 models or core transformations within the latter. For clarity of statements, it is useful to consider the most simple yet reasonable situation, namely one-component PARAFAC decomposition or, closely related, maximization of the leading squared core entry in Tucker3. In the paper, necessary and *sufficient* conditions for global solutions are derived. Furthermore, it is shown that, in general, the usual cyclic co-ordinate optimization scheme of three-way methods does not converge towards a local minimum (or maximum) even if the iterates yield global solutions in each co-ordinate direction. Finally, an example for a proper local minimum in one-component PARAFAC is given. Copyright © 2000 John Wiley & Sons, Ltd.

1. INTRODUCTION

Most *N*-way methods are based on the solution of some optimization (minimization or maximization) problem. Typical examples are the PARAFAC and the Tucker3 decomposition of *N*-way arrays as well as the simple structure transformation of core arrays (within the Tucker3 model) which has attracted increasing attention recently. Since the underlying optimization problem usually lacks any convexity properties, one is faced with the occurrence of local solutions or even saddle points (stationary solutions not being a local minimizer or maximizer). The complexity increases with dimension. To illustrate this in the most simple case, consider the least squares problem of one-component approximation of some *N*-way array $\mathbf{X}$ of order $(n_1,\ldots, n_N)$:

$$\min\{\|\mathrm{vec}\,\mathbf{X} - \mathbf{a}_1 \otimes \ldots \otimes \mathbf{a}_N\|^2 \,|\, \mathbf{a}_i \in \mathbb{R}^{n_i}, i = 1,\ldots,N\} \tag{1}$$

It is frequently convenient to replace (1) by the structurally equivalent maximization problem

$$\max\left\{\left[(\mathrm{vec}\,\mathbf{X})^{\mathrm{T}}(\mathbf{a}_1 \otimes \ldots \otimes \mathbf{a}_N)\right]^2 \,|\, \|\mathbf{a}_1\| = \ldots = \|\mathbf{a}_N\| = 1\right\} \tag{2}$$

Structural equivalence means that, in particular, if $\mathbf{a} = (\mathbf{a}_1,\ldots,\mathbf{a}_N)$ is a stationary point, saddle point, local or global maximizer of (2), then

$$\mathbf{a}^* = (\beta_1\mathbf{a}_1, \beta_2\mathbf{a}_2, \ldots, \beta_N\mathbf{a}_N) \quad \text{with} \quad \beta_1\beta_2\ldots\beta_N = (\mathrm{vec}\,\mathbf{X})^{\mathrm{T}}(\mathbf{a}_1 \otimes \ldots \otimes \mathbf{a}_N)$$

* Correspondence to: R. Henrion, Weierstrass Institute Berlin, D-10117 Berlin, Germany.

 Accepted 2 November 1999

is a stationary point, saddle point, local or global minimizer respectively of (1). Conversely, if $\mathbf{a} = (\mathbf{a}_1,\ldots,\mathbf{a}_N)$ is a stationary point, saddle point, local or global minimizer of (1) with $\mathbf{a}_i \neq 0$ ($i = 1,\ldots,N$), then

$$\mathbf{a}' = (\mathbf{a}_1/\|\mathbf{a}_1\|, \ldots, \mathbf{a}_N/\|\mathbf{a}_N\|)$$

is a stationary point, saddle point, local or global maximizer respectively of (2). Passing from (1) to (2) has, among others, the advantage of working in a lower-dimensional space (by parametrization of the normalization constraints), which will be exploited later in this paper (see Section 4.4). Also note that two optimization problems may be equivalent from the viewpoint of global solutions but fail to be structurally equivalent in the sense mentioned above (see Section 4.4 and end of Section 3).

Now, considering (2) in the trivial one-dimensional case $N = 1$, one has to solve $\max\{(\mathbf{a}^T\mathbf{x})^2 | \ \|\mathbf{a}\| = 1\}$. Apart from the degenerate situation $\mathbf{x} = 0$, this problem has exactly two different global maximizers ($\mathbf{a} = \pm\mathbf{x}/\|x\|$), and there are neither proper local maximizers (local maximizers which are not global ones) nor saddle points in this situation. Next consider the case $N = 2$, where (2) reads as $\max\{(\mathbf{a}^T\mathbf{X}\mathbf{b})^2 | \ \|\mathbf{a}\| = \|\mathbf{b}\| = 1\}$. Restricting to the situation where rank $\mathbf{X} = n_1 \leq n_2$, it is well known that the stationary solutions for $\mathbf{a}$ and $\mathbf{b}$ are precisely given here by the normalized eigenvectors of $\mathbf{X}\mathbf{X}^T$ and $\mathbf{X}^T\mathbf{X}$ respectively, corresponding to a common positive eigenvalue λ, the root of which indicates the realized functional value in the maximization problem. As in the one-dimensional case, one has direct access to the global maximizers (eigenvectors corresponding to the largest eigenvalue). However, now one has to expect saddle points in the problem (eigenvectors corresponding to eigenvalues different from the largest and the smallest one). Of course, saddle points are no serious obstacle from the numerical point of view owing to the explicit solvability of the maximization problem via eigen-analysis. Furthermore, as for $N = 1$, there do not occur proper local maximizers for $N = 2$. Passing to $N = 3$, however, both saddle points and proper local maximizers become typical even in this simple constellation of one-component approximation. Corresponding examples will be given in this paper.

Being faced with the usual difficulties arising in non-convex optimization, one might be strongly interested in verifiable criteria for global solutions. While this is almost hopeless in general, there seems to be a chance in N-way analysis. In Section 4.3 we derive a criterion for global optimality in the most simple reasonable case of (2,2,2) arrays which applies both to core transformation and one-component PARAFAC. In principle, with increasing notational effort, this criterion may be extended to arrays of higher order. Furthermore, the derived criterion is far from being sharp and might be improved by more sophisticated arguments. It is our primary goal, however, to indicate the mere possibility of formulating such easily verifiable conditions and to stimulate more detailed investigations.

Another part of the paper is devoted to illustrating certain typical difficulties arising in the context of cycling co-ordinate optimization (alternating least squares), which is the standard approach in three-way algorithms. In Section 4.2 we give an example of core transformation (and one-component PARAFAC at the same time) where the solution scheme based on cycling co-ordinate optimization realizes a *global* optimization in each coordinate-related iteration but converges just towards a saddle point of the overall problem. In particular, the limiting point fails to be even a local maximizer (minimizer), much less a global one. The same example contradicts the statement that increasing and bounded-from-above functional values of iterates would guarantee convergence towards a local maximizer (or decreasing and bounded-from-below functional values would guarantee convergence towards a local minimizer). Finally, in Section 4.3 an example of proper local minima in one-component PARAFAC is given.

The paper is organized as follows. After a brief introduction to the Tucker3 model and to the issue

of core transformation in Section 2, some recently considered criteria of core simplicity are introduced along with the 'leading-entry' criterion, which will serve as a simple prototype example in our argumentation. Section 3 presents a solution algorithm for maximizing the 'leading-entry' criterion, which apart from its practical meaning mainly serves as a basis for deriving necessary and sufficient conditions for global maximizers in Section 4. All the derivations made in the framework of core transformations are reinterpreted as characterizations in the context of one-component PARAFAC.

2. THE TUCKER3 MODEL AND CORE TRANSFORMATIONS

The Tucker3 model of N-way principal components analysis, introduced in Reference [1], belongs to the most important approaches to multilinear decomposition of N-dimensional data arrays. Given an N-way array $\mathbf{X}$ of order $(n_1,\ldots, n_N)$, one looks for N component matrices $\mathbf{A}_i$ of orders (n_i, s_i) with $s_i \ll n_i$ (for $i = 1,\ldots, N$) such that the approximation

$$\mathrm{vec}\mathbf{X} \approx (\mathbf{A}_1 \otimes \ldots \otimes \mathbf{A}_N)\mathrm{vec}\,\mathbf{C} \tag{3}$$

is optimal in the least squares sense. Here 'vec' and '$\otimes$' denote the vectorization and Kronecker product respectively, while $\mathbf{C}$ refers to the so-called core array of order $(s_1,\ldots s_N)$. An algorithm for finding the $\mathbf{A}_i$ and $\mathbf{C}$, given $\mathbf{X}$, is contained in Reference [2]. In the case where $\mathbf{C}$ is diagonal (i.e. $C_{i_i},\ldots,i_N = 0$ unless $i_1 = \ldots = i_N$), the Tucker3 model reduces to the PARAFAC model [3]. In many applications the Tucker3 model is understood in a latent structure environment: restricting to three-way arrays, $\mathbf{X}$ typically contains data of real variables measured for real objects under varying real conditions, while the component matrices $\mathbf{A}_i$ contain the scores of the real entities with respect to a small number of latent ones and $\mathbf{C}$ is an array of virtual measurements for the different possible combinations of latent objects, latent variables and latent conditions. If the component matrices are required to be orthonormal (i.e. $\mathbf{A}_i^{\mathrm{T}}\mathbf{A}_i = \mathbf{I}_{s_i}$), then the squared entry c_{ijk}^2 of the core array indicates the data variation covered by a combination of latent object i with latent variable j and latent condition k. Therefore the core array can be understood as a generalization of the diagonal eigenvalue matrix in conventional principal component analysis (PCA) of data tables. However, in contrast to the latter, the core array is usually completely filled.

In the worst case, all contributions recorded in the core array are of comparable magnitude. Then, even in the simplest situation of two latent factors for each of the, say, three dimensions of the data array, one would arrive at eight different possible factor combinations to be interpreted. On the other hand, a lot of artefact contributions can be contained herein, whereas the probably small number of meaningful factor combinations remains hidden. This phenomenon results from the non-uniqueness of the decomposition (3). Indeed, applying the oblique transformation $\tilde{\mathbf{A}}_i = \mathbf{A}_i\mathbf{P}_i$ to the component matrices by means of non-singular transformation matrices $\mathbf{P}_i$, a different decomposition of the same reproducing data array can be found after counter-transforming the core array $\mathbf{C}$ to $\tilde{\mathbf{C}}$ via

$$\mathrm{vec}\ \tilde{\mathbf{C}} = (\mathbf{P}_1^{-1} \otimes \cdots \otimes \mathbf{P}_N^{-1})\ \mathrm{vec}\ \mathbf{C} \tag{4}$$

Now it frequently happens that a numerical solution of (3) provides a core array with many elements of comparable magnitudes. Exploiting the transformation degree of (4), there is hope, however, to separate artefacts from meaningful contributions. To give an extreme example, consider the core array

$$\mathbf{C} = \left(\begin{array}{cc|cc} c_{111} & c_{112} & c_{211} & c_{212} \\ c_{121} & c_{122} & c_{221} & c_{222} \end{array}\right) = \left(\begin{array}{cc|cc} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{array}\right)$$

of order (2,2,2). Choosing the orthogonal transformation matrices

$$\mathbf{P}_1 = \mathbf{P}_2 = \mathbf{P}_3 = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

one arrives via (4) at the transformed core array

$$\tilde{\mathbf{C}} = \left(\begin{array}{cc|cc} \sqrt{8} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Owing to the orthogonality of transformation matrices, the sum of squares remains invariant (no data variation can be gained or lost by transformation (4)). However, the structure of $\tilde{\mathbf{C}}$ suggests that there is only one single significant combination of latent factors left in contrast to eight equally weighted contributions in $\mathbf{C}$. Of course, the interpretation of the significant factor combinations indicated in $\tilde{\mathbf{C}}$ has to be carried out with respect to the transformed component matrices $\tilde{\mathbf{A}}_i$ rather than the original ones.

In the following we restrict our considerations to orthogonal transformation matrices and, without loss of generality, to three-dimensional data arrays. Accordingly, (4) takes the form $\text{vec}\,\tilde{\mathbf{C}} = (\mathbf{P}^{\mathrm{T}} \otimes \mathbf{Q}^{\mathrm{T}} \otimes \mathbf{R}^{\mathrm{T}})\,\text{vec}\,\mathbf{C}$, which in single entries of the respective matrices and arrays is written as

$$\tilde{c}_{ijk} = \sum_{\alpha=1}^{s_1} p_{\alpha i} \sum_{\beta=1}^{s_2} q_{\beta j} \sum_{\gamma=1}^{s_3} r_{\gamma k} c_{\alpha\beta\gamma} \quad (i = 1, \ldots, s_1; j = 1, \ldots, s_2; k = 1, \ldots, s_3) \tag{5}$$

Although $\sum_{ijk} \tilde{c}_{ijk}^2 = \sum_{ijk} c_{ijk}^2$ (i.e. the sum of squared core elements remains invariant under (4)), the converse does not hold true: core arrays having equal sum of squared elements need not be transformable into each other via (4). As an example, one may take

$$\mathbf{C}' = \left(\begin{array}{cc|cc} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \end{array}\right)$$

which by no orthogonal matrices $\mathbf{P}$, $\mathbf{Q}$ and $\mathbf{R}$ can be transformed into the core $\tilde{\mathbf{C}}$ specified above, even though the sum of squared elements equals eight in both cases. Hence one has to distinguish a variety of disjoint orbits of equivalent cores which may be transformed into each other via (4). In these terms, $\mathbf{C}$ and $\tilde{\mathbf{C}}$ share the same orbit, whereas $\mathbf{C}'$ is the representative of a different orbit. As a consequence, the number of two non-zero elements is minimal in the core $\mathbf{C}'$ and there is no chance to reduce it to one single non-zero element by placing the whole sum of squares onto just one position as in $\tilde{\mathbf{C}}$. In other words, $\mathbf{C}'$ is a representative of its orbit with simplest possible structure and the same holds true for $\tilde{\mathbf{C}}$. Now the obvious aim of core transformations will be to find the 'best' (in a sense to be made precise below) representative of an orbit which some given starting core belongs to.

Table I collects some criteria measuring simplicity of core arrays. Obviously, all of them favour the solution $\tilde{\mathbf{C}}$ among the three cores considered above. The first two criteria were the first historically. Both aim at maximizing some diagonal structure of the core. Their theoretical importance relies on the fact that they measure the closeness of the Tucker3 model to the simpler PARAFAC model. However, the diagonality criteria suffer from certain drawbacks: first, they assume a more or less

Table I. Summary of criteria for core simplicity. Arrows indicate whether the criterion has to be maximized or minimized. On the right-hand side, corresponding values for the three exemplary cores given in the text are indicated ($^{*}m = 6$; $^{**}w_1 = w_2 = w_3 = \frac{1}{3}$, $\gamma_1 = \gamma_2 = \gamma_3 = 0$)

Criterion	Definition	Reference	$\mathbf{C}$	$\tilde{\mathbf{C}}$	$\mathbf{C}'$
Slice diagonality ↑	$\sum_{i=1}^{s_1} \sum_{j=1}^{s_2} c_{ijj}^2$	[4]	4	8	4
Body diagonality ↑	$\sum_{i=1}^{s_1} c_{iii}^2$	[5]	2	8	4
Simplimax (m) ↓	$\min\{\sum_{(i,j,k)\in I} c_{ijk}^2 \mid \#I = m\}$	[6]	6^{*}	0	0
Orthomax (w,γ) ↑	$(w_1 + w_2 + w_3) \sum_{i=1}^{s_1} \sum_{j=1}^{s_2} \sum_{k=1}^{s_3} c_{ijk}^4$	[7]	8^{**}	64	32
(γ_i, w_i free parameters)	$-\frac{\gamma_1 w_1}{s_2 s_3} \sum_{i=1}^{s_1} \left(\sum_{j=1}^{s_2} \sum_{k=1}^{s_3} c_{ijk}^2\right)^2$				
	$-\frac{\gamma_2 w_2}{s_1 s_3} \sum_{j=1}^{s_2} \left(\sum_{i=1}^{s_1} \sum_{k=1}^{s_3} c_{ijk}^2\right)^2$				
	$-\frac{\gamma_3 w_3}{s_1 s_2} \sum_{k=1}^{s_3} \left(\sum_{j=1}^{s_2} \sum_{i=1}^{s_1} c_{ijk}^2\right)^2$				
Variance of squares ↑	$\sum_{i=1}^{s_1} \sum_{j=1}^{s_2} \sum_{k=1}^{s_3} \left(c_{ijk}^2 - \bar{c}\right)^2$	[8]	0	56	24
	$(\bar{c} = \frac{1}{s_1 s_2 s_3} \sum_{i=1}^{s_1} \sum_{j=1}^{s_2} \sum_{k=1}^{s_3} c_{ijk}^2)$				

restrictive shape of the core (cubic or slice-wise cubic); but second, and maybe even more important, they distribute a maximum of squared elements along prescribed positions in the core. Giving up this goal, one may hope for concentrating more 'mass' of the core on even fewer positions although possibly not on the diagonal(s) of the array. In this way the effort of interpreting Tucker3 solutions may be reduced considerably. The issue of arbitrary structure simplification of cores was investigated in numerous papers (e.g. References [6–10]), leading, among others, to the orthomax and variance-of-squares criteria listed in Table I.

The simplimax criterion gives a different view on the same problem. It minimizes the sum of the m smallest squared elements of the core and hence, since the total sum of squares is constant in our considerations here, maximizes the sum of the $s_1 s_2 s_3 - m$ largest squared elements in the core. A peculiarity of the simplimax criterion is its non-differentiability. In the extreme case of $m = s_1 s_2 s_3 - 1$, just the largest squared core element is maximized. In the sequel we will call this the 'leading-entry' criterion.

In general, none of the criteria may be expected to achieve its theoretical optimum on the orbit given by the starting core. The theoretical optimum for body diagonality is achieved if there are only zero elements off the diagonal left. Similarly, the theoretical optimum for the variance-of-squares and leading-entry criteria would imply all elements but one to be zero in both cases. Algorithms for optimizing the single criteria are found in the references cited in Table I. However, they all share the same abstract problem structure, namely the optimization of some differentiable (apart from simplimax) goal function depending on a set of orthogonal matrices. A general algorithm dealing with such problems was recently proposed in Reference [10]. It relies on the iterative improvement of the goal function by means of singular value decomposition of certain (criteria-dependent) derivative matrices. Then switching from one criterion to another is realized by simply changing one command line in the implementation of the algorithm.

3. MAXIMIZATION OF THE 'LEADING-ENTRY' CRITERION

In the following we want to focus on the leading-entry criterion, which as a function of the considered core is written as

$$\lambda(\mathbf{C}) = \max\{c_{ijk}^2 |\ i \in \{1,\ldots,s_1\}; j \in \{1,\ldots,s_2\}; k \in \{1,\ldots,s_3\}\} \tag{6}$$

Maximizing $\lambda(\mathbf{C})$ means to identify by core transformation a single factor combination with maximum variance. Our primary intention of doing so is not to introduce a new criterion which we expect certain advantages of, but rather to exploit its simplicity for highlighting aspects of optimality which are more difficult in the context of other criteria. To our feeling, it is the most simple yet meaningful (e.g. as a special case of simplimax) criterion of core simplicity. Furthermore, it offers the possibility of translating all results to the problem of one-component PARAFAC (see end of this section).

We start by deriving a solution procedure for maximizing the 'leading-entry' criterion on a given orbit defined by some starting core $\mathbf{C}$. Since being on the same orbit means the existence of orthogonal transformation matrices $\mathbf{P}$, $\mathbf{Q}$ and $\mathbf{R}$ satisfying (5), this amounts to maximizing $\lambda(\tilde{\mathbf{C}})$ as a function of these transformation matrices. In other words, defining $\lambda^*(\mathbf{P}, \mathbf{Q}, \mathbf{R}) = \lambda(\tilde{\mathbf{C}}(\mathbf{P}, \mathbf{Q}, \mathbf{R}))$, where the functional dependence $\tilde{\mathbf{C}}(\mathbf{P}, \mathbf{Q}, \mathbf{R})$ refers to (5), we wish to solve the optimization problem

$$\max\{\lambda^*(\mathbf{P},\mathbf{Q},\mathbf{R}) | \mathbf{P}^{\mathrm{T}}\mathbf{P} = \mathbf{I}_{s_1}, \mathbf{Q}^{\mathrm{T}}\mathbf{Q} = \mathbf{I}_{s_2}, \mathbf{R}^{\mathrm{T}}\mathbf{R} = \mathbf{I}_{s_3}\} \tag{7}$$

A common approach to solve problems like (7) is to apply the following cycling co-ordinate scheme of optimization.

1. Set $\mathbf{P}^0 := \mathbf{I}_{s_1}$, $\mathbf{Q}^0 := \mathbf{I}_{s_2}$, $\mathbf{R}^0 := \mathbf{I}_{s_3}$ and $k := 0$.
2. Choose an orthogonal $\mathbf{P}^{k+1}$ which maximizes $\lambda^*(\mathbf{P}, \mathbf{Q}^k, \mathbf{R}^k)$ as a function of $\mathbf{P}$.
3. Choose an orthogonal $\mathbf{Q}^{k+1}$ which maximizes $\lambda^*(\mathbf{P}^{k+1}, \mathbf{Q}, \mathbf{R}^k)$ as a function of $\mathbf{Q}$.
4. Choose an orthogonal $\mathbf{R}^{k+1}$ which maximizes $\lambda^*(\mathbf{P}^{k+1}, \mathbf{Q}^{k+1}, \mathbf{R})$ as a function of $\mathbf{R}$.
5. If $\lambda^*(\mathbf{P}^{k+1}, \mathbf{Q}^{k+1}, \mathbf{R}^{k+1})$ differs significantly from $\lambda^*(\mathbf{P}^k, \mathbf{Q}^k, \mathbf{R}^k)$, then set $k := k+1$ and go to step 2; else stop.

We want to check now how step 2, as the first meaningful step in this algorithm, works. This will be sufficient, since the following steps are completely analogous. Thus we seek an orthogonal matrix $\mathbf{P}^1$ that maximizes $\lambda^*(\mathbf{P}, \mathbf{I}_{s_2}, \mathbf{I}_{s_3})$. In view of (5), one has

$$\max_{\mathbf{P}^{\mathrm{T}}\mathbf{P}=\mathbf{I}_{s_1}} \lambda^*(\mathbf{P},\mathbf{I}_{s_2},\mathbf{I}_{s_3}) = \max_{\mathbf{P}^{\mathrm{T}}\mathbf{P}=\mathbf{I}_{s_1}} \max_{i,j,k} \tilde{c}_{ijk}^2(\mathbf{P},\mathbf{I}_{s_2},\mathbf{I}_{s_3}) = \max_{i,j,k} \max_{\mathbf{P}^{\mathrm{T}}\mathbf{P}=\mathbf{I}_{s_1}} \left(\sum_{\alpha=1}^{s_1} p_{\alpha i} c_{\alpha jk}\right)^2$$

$$= \max_{i,j,k} \max_{\hat{\mathbf{p}}_i^{\mathrm{T}}\hat{\mathbf{p}}_i=1} (\hat{\mathbf{p}}_i^{\mathrm{T}}\hat{\mathbf{c}}_{jk})^2$$

where we have used the abbreviations $\hat{\mathbf{p}}_i = (p_{1i}, \ldots, p_{s_1 i})^{\mathrm{T}}$ and $\hat{\mathbf{c}}_{jk} = (c_{1jk}, \ldots, c_{s_1 jk})^{\mathrm{T}}$. We recall that, given some $\mathbf{y} \neq 0$, the function $(\mathbf{x}^{\mathrm{T}}\mathbf{y})^2$ of $\mathbf{x}$ achieves its global maximum on the unit sphere at the point $\mathbf{x} = \mathbf{y}/\|\mathbf{y}\|$. Translating this to our problem, we can continue the last equation to

$$\max_{\mathbf{P}^{\mathrm{T}}\mathbf{P}=\mathbf{I}_{s_1}} \lambda^*(\mathbf{P},\mathbf{I}_{s_2},\mathbf{I}_{s_3}) = \max_{i,j,k} \|\hat{\mathbf{c}}_{jk}\|^2 = \max_{j,k} \sum_{\alpha=1}^{s_1} c_{\alpha jk}^2 \tag{8}$$

In this way we can read off the (global!) maximum value to be expected in this first step of iteration directly from the starting core. Furthermore, this maximum value is realized by any orthogonal matrix $\mathbf{P}$ having the vector $\hat{\mathbf{p}} := \hat{\mathbf{c}}_{j^*k^*} / \|\hat{\mathbf{c}}_{j^*k^*}\|$ as an arbitrary vector of its columns. Here j^* and k^* refer to those indices maximizing the expression on the right-hand side of (8). Formally, one would have to exclude the case $\hat{\mathbf{c}}_{j^*k^*} = 0$ in order to apply the above normalization, but in this case, by definition, the whole starting core has to be the trivial zero core, which is not altered by any transformation. This degenerate case is of no practical interest and can be ignored from the very beginning.

Next, according to the cycling co-ordinate scheme of optimization described above, we want to maximize the leading-entry criterion by variation of the second transformation matrix $\mathbf{Q}$ while keeping $\mathbf{P}$ and $\mathbf{R}$ fixed. However, rather than solving the problem $\max\{\lambda^*(\mathbf{P}, \mathbf{Q}, \mathbf{I}_{s_3}) | \mathbf{Q}^{\mathrm{T}}\ \mathbf{Q} = \mathbf{I}_{s_2}\}$ according to step 3 of the above scheme, which relates to the very starting core, we now update the 'starting core' as $\mathbf{C} := \tilde{\mathbf{C}}$ and solve the problem $\max\{\lambda^*(\mathbf{I}_{s_1}, \mathbf{Q}, \mathbf{I}_{s_3}) | \mathbf{Q}^{\mathrm{T}}\ \mathbf{Q} = \mathbf{I}_{s_2}\}$. Of course, the solution is the same, but updating the core rather than the transformation matrices has the advantage that two out of three transformation matrices are identities in each step, which facilitates the handling of the maximization and actually reduces it each time to the very first step we described above up to permutation of indices. In particular, we have to find in the second and third iterations (on $\mathbf{Q}$ and $\mathbf{R}$) the analogous quantities to (8):

$$\max_{i,k} \sum_{\alpha=1}^{s_2} c_{i\alpha k}^2 \quad \text{and} \quad \max_{i,j} \sum_{\alpha=1}^{s_3} c_{ij\alpha}^2$$

Each time the optimal $\mathbf{Q}$ and $\mathbf{R}$ are found exactly the same way as $\mathbf{P}$ before, and each time the core array has to be updated by applying the respective transformation found to be optimal. Note that the algorithm presented here provides a global maximum along each of the three co-ordinate directions just by one explicit step.

At the end of this section we want to point to an analogy between maximization of the core criterion λ and the one-component PARAFAC approximation of a three-way data array. To this aim, consider the core criterion $\mu(\mathbf{C}) = c_{111}^2$, which simply measures the contribution of the very first core element. The maximization of μ over all orthogonal transformation matrices is equivalent to that of λ, since the leading core entry can always be placed at the first position by permutation matrices which, being orthogonal, are admissible transformation matrices. Now, according to (5), the transformed first element is written as

$$\tilde{c}_{111}^2 = [(\mathrm{vec}\ \mathbf{C})^{\mathrm{T}} (\mathbf{p}_1 \otimes \mathbf{q}_1 \otimes \mathbf{r}_1)]^2$$

where $\mathbf{p}_1$, $\mathbf{q}_1$ and $\mathbf{r}_1$ are the first columns of orthogonal matrices $\mathbf{P}$, $\mathbf{Q}$ and $\mathbf{R}$ respectively. Comparing this with (2) and (1), we conclude that maximization of μ and hence of λ is formally equivalent to finding the best one-component PARAFAC approximation of C, when considered as a usual data array. However, this equivalence is lost on the level of local solutions and of cycling co-ordinate optimization: we shall see in Section 4·4 that a local minimizer in the one-component PARAFAC approximation of $\mathbf{C}$ may correspond to a non-stationary point in the context of maximizing the leading-entry criterion. In particular, the usual PARAFAC algorithm would stop at that local solution (and no small perturbation would help here) whereas the leading-entry criterion is improved in the next iteration. Furthermore, it turns out that the PARAFAC algorithm will not help us in finding the desired optimality conditions, but, once these are derived in the framework of core transformations and using the algorithm described before, analogous optimality conditions may be translated to one-component PARAFAC decompositions of arbitrary data arrays as well.

4. GLOBAL, LOCAL AND STATIONARY SOLUTIONS

In this section, necessary and sufficient conditions for global maximizers of the leading-entry criterion λ are formulated and translated to the one-component PARAFAC approximation. Furthermore, examples of saddle points as limits of three-way-methods and of proper local minimizers in PARAFAC are given.

4.1. Necessary conditions for global maxima

A necessary condition for a global maximum of the leading-entry criterion is readily derived from the algorithm itself. Suppose that **C** is a core realizing a maximum leading-entry criterion on its orbit. Then no transformation (5) with orthogonal **P**, **Q** and **R** can provide a higher value of the criterion. Taking **C** as the starting core, which the function λ^* defined in (7) refers to, this implies in particular with respect to the **P** direction of optimization that $\lambda^*(\mathbf{I}_{s_1}, \mathbf{I}_{s_2}, \mathbf{I}_{s_3}) = \max\{\lambda^*(\mathbf{P}, \mathbf{I}_{s_2}, \mathbf{I}_{s_3}) \mid \mathbf{P}^{\mathrm{T}}\mathbf{P} = \mathbf{I}_{s_1}\}$. Then (8) entails

$$\max_{j,k} \sum_{\alpha=1}^{s_1} c_{\alpha jk}^2 = \lambda^*(\mathbf{I}_{s_1}, \mathbf{I}_{s_2}, \mathbf{I}_{s_3}) = \lambda(\mathbf{C}) = \max_{i,j,k} c_{ijk}^2 = c_{i^*j^*k^*}^2$$

where (i^*,j^*,k^*) refers to the position of (one of) the leading squared element(s). It follows that

$$\sum_{\alpha=1}^{s_1} c_{\alpha j^* k^*}^2 \leq c_{i^* j^* k^*}^2, \quad \text{hence} \quad c_{\alpha j^* k^*} = 0 \quad \text{for all} \quad \alpha \neq i^*$$

On the other hand, because $i^* \in \{1, \ldots, s_1\}$, one derives that $c_{\alpha j^* k^*} = 0$ for all $\alpha \neq i^*$. Repeating the analogous argumentation for the **Q** and **R** steps of the co-ordinate-wise optimization, one arrives at the following necessary condition for a core **C** to realize a global maximum of the leading-entry criterion on its orbit:

$$c_{\alpha j^* k^*} = 0 \quad \text{for all} \quad \alpha \neq i^*, \qquad c_{i^* \alpha k^*} = 0 \quad \text{for all} \quad \alpha \neq j^*, \qquad c_{i^* j^* \alpha} = 0 \quad \text{for all} \quad \alpha \neq k^* \tag{9}$$

To give some examples, consider the cores

$$\mathbf{C}_1 = \left(\begin{array}{cc|cc} 1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 1 \end{array}\right), \quad \mathbf{C}_2 = \left(\begin{array}{cc|cc} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{array}\right), \quad \mathbf{C}_3 = \left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 \cdot 3 \\ 0 & -0 \cdot 3 & 0 \cdot 2 & -0 \cdot 1 \end{array}\right)$$

$\mathbf{C}_1$ is easily recognized not to realize a global optimum, since $(i^*,j^*,k^*) = (2,1,2)$, but $c_{222} \neq 0$, hence the second part of (9) is violated. Consequently, the algorithm proposed above would not stop at this iterate. In contrast, $\mathbf{C}_2$ and $\mathbf{C}_3$ satisfy the necessary conditions. This, however, does not mean that they really provide global maxima; all one can say at this point is that (9) does not contradict this property. Indeed, we shall see later that $\mathbf{C}_2$ is not a global maximum (not even a local one), whereas $\mathbf{C}_3$ satisfies the sufficient conditions to be derived below and hence is a global maximum. As a by-product of (9), we see that the algorithm proposed above will converge towards solutions with at least $s_1 + s_2 + s_3 - 3$ zeros in the core.

At the same time we can formulate the necessary optimality conditions for the one-component PARAFAC decomposition in (1) (restricted to three-way arrays, N = 3, for notational convenience). Given a data array X, assume that a certain triple $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ of non-trivial component vectors

approximating X is a global minimizer of the least squares problem (1). Then the structural equivalence of problems (1) and (2) implies

$$(\mathbf{p}, \mathbf{q}, \mathbf{r}) := (\mathbf{a}_1/\|\mathbf{a}_1\|, \mathbf{a}_2/\|\mathbf{a}_2\|, \mathbf{a}_3/\|\mathbf{a}_3\|)$$

to be a global maximizer of (2). Now we choose arbitrary orthogonal matrices (**P**, **Q**, **R**) having (**p**, **q**, **r**) as their first columns and calculate an array $\tilde{\mathbf{C}}$ from **C**:= **X** via (5). According to the remarks at the end of Section 3, (**P**, **Q**, **R**) must realize a global maximum of the leading-entry criterion, hence the transformed array $\tilde{\mathbf{C}}$ has to satisfy the relations (9).

4.2. A sufficient condition for global maxima

We start by introducing the 'leading-pair' ratio π (**C**) for core arrays **C** as the ratio between the leading and the second-leading squared values in the core. In the case where the latter equals zero, we formally put π (**C**) = ∞. Now the following result holds true.

Theorem 1

If **C** is a core array of order (2,2,2) satisfying the conditions (9) as well as π (**C**) $\geq$ 9, then **C** realizes a global maximum of the leading-entry criterion on its orbit.

Proof

First we exclude the case π (**C**) = ∞, since this directly implies that at most one core element is different from zero, hence the assertion of the theorem follows trivially. Without loss of generality, after applying appropriate permutation matrices **P**, **Q** and **R** in (5), which do not alter the value of the leading-entry criterion, we may assume that c_{111}^2 is the leading squared element of **C**. We have to show that the leading squared element of any core $\tilde{\mathbf{C}}$ obtained from **C** via (5) is not larger than c_{111}^2. This amounts to checking that $\tilde{c}_{ijk}^2 \leq c_{111}^2$ holds true for any index triple (*i*,*j*,*k*). Now we fix an arbitrary such triple and abbreviate the coefficients in (5) by $p_\alpha := p_{\alpha i}$, $q_\beta := q_{\beta j}$ and $r_\gamma := r_{\gamma k}$. According to the assumed conditions (9), one has $c_{211} = c_{121} = c_{112} = 0$, such that with

$$I := \{(1,1,1), (1,2,2), (2,1,2), (2,2,1), (2,2,2)\}$$

(5) provides

$$\tilde{c}_{ijk}^2 = \left(\sum_{(\alpha,\beta,\gamma)\in I} p_\alpha q_\beta r_\gamma c_{\alpha\beta\gamma} \right)^2 = S_1 + S_2$$

where

$$S_1 = \sum_{(\alpha,\beta,\gamma)\in I} p_\alpha^2 q_\beta^2 r_\gamma^2 c_{\alpha\beta\gamma}^2 \qquad S_2 = 2 \sum p_{\alpha_1} p_{\alpha_2} q_{\beta_1} q_{\beta_2} r_{\gamma_1} r_{\gamma_2} c_{\alpha_1\beta_1\gamma_1} c_{\alpha_2\beta_2\gamma_2}$$

and the summation in S_2 runs over all $(\alpha_1,\beta_1,\gamma_1) \in I$ and $(\alpha_2,\beta_2,\gamma_2) \in I \setminus \{(\alpha_1,\beta_1,\gamma_1)\}$. Using the assumption π (**C**) $\geq$ 9, we get the estimate

$$S_1 \leq c_{111}^2 \left(p_1^2 q_1^2 r_1^2 + \tfrac{1}{9} p_1^2 q_2^2 r_2^2 + \tfrac{1}{9} p_2^2 q_1^2 r_2^2 + \tfrac{1}{9} p_2^2 q_2^2 r_1^2 + \tfrac{1}{9} p_2^2 q_2^2 r_2^2 \right)$$

Table II. Estimation from above of terms in sum S_2

$(\alpha_1, \beta_1, \gamma_1)$	$(\alpha_2, \beta_2, \gamma_2)$	Estimation from above
(1,1,1)	(1,2,2)	$\frac{1}{3}c_{111}^2\ (p_1^2q_2^2r_1^2 + p_1^2q_1^2r_2^2)$
(1,1,1)	(2,1,2)	$\frac{1}{3}c_{111}^2\ (p_2^2q_1^2r_1^2 + p_1^2q_1^2r_2^2)$
(1,1,1)	(2,2,1)	$\frac{1}{3}c_{111}^2\ (p_1^2q_2^2r_1^2 + p_2^2q_1^2r_1^2)$
(1,1,1)	(2,2,2)	$\frac{1}{3}c_{111}^2\ (p_1^2q_1^2r_2^2 + p_2^2q_2^2r_1^2)$
(1,2,2)	(2,1,2)	$\frac{1}{9}c_{111}^2\ (p_1^2q_2^2r_2^2 + p_2^2q_1^2r_2^2)$
(1,2,2)	(2,2,1)	$\frac{1}{9}c_{111}^2\ (p_1^2q_2^2r_2^2 + p_2^2q_2^2r_1^2)$
(1,2,2)	(2,2,2)	$\frac{1}{9}c_{111}^2\ (p_1^2q_2^2r_2^2 + p_2^2q_2^2r_2^2)$
(2,1,2)	(2,2,1)	$\frac{1}{9}c_{111}^2\ (p_2^2q_1^2r_2^2 + p_2^2q_2^2r_1^2)$
(2,1,2)	(2,2,2)	$\frac{1}{9}c_{111}^2\ (p_2^2q_1^2r_2^2 + p_2^2q_2^2r_2^2)$
(2,2,1)	(2,2,2)	$\frac{1}{9}c_{111}^2\ (p_2^2q_2^2r_1^2 + p_2^2q_2^2r_2^2)$

Setting $a := p_1q_2r_1\ \sqrt{|c_{111}|\,|c_{122}|}$ and $b := p_1q_1r_2\ \sqrt{|c_{111}|\,|c_{122}|}$ and exploiting the general relation $2ab \le a^2 + b^2$, one may estimate the first term (relating to the index triples (1,1,1) and (1,2,2)) in the sum S_2 by

$$2p_1^2q_1q_2r_1r_2c_{111}c_{122} \le 2p_1^2q_1q_2r_1r_2|c_{111}||c_{122}| \le |c_{111}||c_{122}|(p_1^2q_2^2r_1^2 + p_1^2q_1^2r_2^2)$$
$$\le \tfrac{1}{3}c_{111}^2(p_1^2q_2^2r_1^2 + p_1^2q_1^2r_2^2)$$

where again the assumption $\pi(\mathbf{C}) \ge 9$ (implying that $|c_{122}| \le \frac{1}{3}|c_{111}|$) was used. Proceeding like this for all 10 terms in S_2, one obtains the estimations from above collected in Table II. Summation of all terms then gives

$$S_2 \le c_{111}^2\left(p_1^2q_1^2r_2^2 + \tfrac{2}{3}p_1^2q_2^2r_1^2 + \tfrac{2}{3}p_2^2q_1^2r_1^2 + \tfrac{1}{3}p_1^2q_2^2r_2^2 + \tfrac{1}{3}p_2^2q_1^2r_2^2 + \tfrac{2}{3}p_2^2q_2^2r_1^2 + \tfrac{1}{3}p_2^2q_2^2r_2^2\right)$$

Now, combining the results obtained so far, one arrives at

$$\tilde{c}_{ijk}^2 = S_1 + S_2 \le c_{111}^2 \sum_{\alpha=1}^{2}\sum_{\beta=1}^{2}\sum_{\gamma=1}^{2} p_\alpha^2q_\beta^2r_\gamma^2 = c_{111}^2\left(\sum_{\alpha=1}^{2}p_\alpha^2\right)\left(\sum_{\beta=1}^{2}q_\beta^2\right)\left(\sum_{\gamma=1}^{2}r_\gamma^2\right) = c_{111}^2$$

where the last equality relies on the orthogonality of the transformation matrices **P**, **Q** and **R**, which in particular implies that $\sum p_\alpha^2 = \sum q_\beta^2 = \sum r_\gamma^2 = 1$. This finishes the proof. ■

Applying Theorem 1 to the core $\mathbf{C}_3$ considered before in the context of necessary conditions, one verifies that $\mathbf{C}_3$ realizes a global maximum of the leading-entry criterion on its orbit (owing to $\pi(\mathbf{C}_3) = 1/(0{\cdot}3)^2 \ge 9$). In principle, one may extend the approach in Theorem 1 to core arrays of larger size than (2,2,2). However, apart from the increased notational effort, the required condition on $\pi(\mathbf{C})$ becomes more and more restrictive. On the other hand, the theoretical conditions which are necessary for an exact proof are far from being optimal. For instance, in the constellation of Theorem 1 the numerical solution of an appropriate optimization problem confirms that the theoretical condition $\pi(\mathbf{C}) \ge 9$ may be replaced by the much weaker condition $\pi(\mathbf{C}) \ge 5$. Similar improvements are

possible for larger-sized core arrays as well (e.g. $\pi(\mathbf{C}) \geq 25$ is sufficient in the (3,3,3) case).

Finally we note that the sufficient condition found here in the context of core transformation applies equally well to the one-component PARAFAC decomposition of a (2,2,2) data array along the same lines as discussed at the end of Section 4·1. In order to illustrate the procedure, assume that we want to check if the component vectors $\mathbf{a}_1 = (1,-1)^{\mathrm{T}}$, $\mathbf{a}_2 = (1,-1)^{\mathrm{T}}$ and $\mathbf{a}_3 = (3,0)^{\mathrm{T}}$ realize a global minimum in the PARAFAC decomposition of the data array

$$\mathbf{X} = \left(\begin{array}{cc|cc} 4 & 3 & -2 & 1 \\ -2 & 1 & 4 & -1 \end{array}\right)$$

First we choose orthogonal matrices having the normalized component vectors in their first columns:

$$\mathbf{P} = \mathbf{Q} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}, \qquad \mathbf{R} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Next, with $\mathbf{C} := \mathbf{X}$, the transformed array is calculated via (5) as

$$\widetilde{\mathrm{C}} = \left(\begin{array}{cc|cc} 6 & 0 & 0 & 2 \\ 0 & 2 & 2 & 2 \end{array}\right)$$

and Theorem 1 shows that $\tilde{\mathbf{C}}$ realizes a global maximum of the leading-entry criterion. Equivalently (see remarks at the end of Section 3), the first columns $\mathbf{p} = \mathbf{q} = (1/\sqrt{2}, -1/\sqrt{2})^{\mathrm{T}}$ and $\mathbf{r} = (1,0)^{\mathrm{T}}$ of **P, Q** and **R** realize a global maximum of (2). Finally, they differ from the original component vectors by the coefficients $\beta_1 = \beta_2 = \sqrt{2}$ and $\beta_3 = 3$, the product of which equals $6 = (\mathrm{vec}\ \mathbf{X})^{\mathrm{T}}\ (\mathbf{p} \otimes \mathbf{q} \otimes \mathbf{r})$. According to the remarks following (2),$(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ is a global minimizer of (1), as was to be checked.

4.3. Saddle points as limits of cycling co-ordinate optimization

As already pointed out in Section 3, the typical solution pattern of three-way algorithms is a cyclic co-ordinate optimization scheme where all but one transformation matrices are successively kept fixed and optimization is carried out for one single transformation matrix. Concerning the possibility in this way to reach local or global maxima of the criterion at hand, one might believe that finding global maxima in each co-ordinate has some impact on finding global maxima of the overall problem. This is by no means true and, much worse, finding co-ordinate-wise global maxima does not even guarantee the iterates of the algorithm to converge towards a local maximum. Actually, all one can hope for is to arrive at stationary points in general, i.e. at points at which the first-order necessary optimality conditions are satisfied for the given criterion. A simple example will illustrate these facts. Consider, for instance, the starting core

$$\mathrm{C}_0 = \left(\begin{array}{cc|cc} 1/\sqrt{2} & -1/\sqrt{2} & 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} & 1/\sqrt{2} & 1/\sqrt{2} \end{array}\right)$$

After performing the first step (optimization of the **P** matrix) of the algorithm for maximizing the leading-entry criterion as described in Section 3, one obtains the new core array

$$\mathbf{C}_1 = \left(\begin{array}{cc|cc} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{array}\right)$$

which now realizes a global maximum in all three co-ordinate directions according to (9); hence the algorithm stops at this solution after having increased the leading-entry criterion from 0·5 to 1. Now we define the following one-parameter family of orthogonal transformation matrices:

$$\mathbf{P}(t) = \mathbf{Q}(t) = \mathbf{R}(t) = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix} \quad (t \in \mathbb{R})$$

Corresponding to these, $\mathbf{C}(t)$ will denote the core array transformed from $\mathbf{C}_1$ by means of $\mathbf{P}(t)$, $\mathbf{Q}(t)$ and $\mathbf{R}(t)$ via (5). Evidently, $\mathbf{C}(0) = \mathbf{C}_1$. Furthermore, the squared first entry $c^2_{111}(t)$ of $\mathbf{C}(t)$ is calculated from $\mathbf{C}_1$, according to (5), as

$$c^2_{111}(t) = (\cos^3 t + 3 \cos t \sin^2 t)^2$$

The first and second derivatives of this function evaluated at $t = 0$ are 0 and 6 respectively. Hence this function realizes a strict local minimum at $t = 0$, i.e. $c^2_{111}(t) > c^2_{111}(0) = 1$ for t close to but different from 0. Now the leading squared entry of $\mathbf{C}(t)$ cannot be less than the specific value $c^2_{111}(t)$, which implies that for t arbitrarily close to but different from 0 the leading squared entry of $\mathbf{C}(t)$ is strictly larger than 1, which in turn is the value of the leading squared entry of $\mathbf{C}_1$. In other words, arbitrarily close to $\mathbf{C}_1$, there exists a better core $\mathbf{C}(t)$ on the same orbit; hence $\mathbf{C}_1$ does not realize a local maximum (much less a global maximum) of the leading-entry criterion on its orbit, although it realizes global maxima along all co-ordinate directions. Actually, $\mathbf{C}_1$ is a saddle point of the leading-entry criterion.

Recalling that the leading-entry criterion has a comparatively simple structure, the phenomenon of ending up at a saddle point is to be expected *a fortiori* for the more complicated criteria listed in Table I and even for all the other usual three-way algorithms. Indeed, the given example may be translated again to show that in the one-component PARAFAC decomposition of $\mathbf{C}_1$ (now considered as a data array) the component vectors $(1,0)^{\mathrm{T}}$, $(1,0)^{\mathrm{T}}$ and $(1,0)^{\mathrm{T}}$ represent a saddle point as well while realizing co-ordinate-wise global optima at the same time.

4.4. Local minima of one-component PARAFAC that are not stationary points of the leading-entry criterion

It is not difficult to find proper local minima in (1). To see this, let us check for which data arrays of order (2,2,2) the component vectors $(1,0)^{\mathrm{T}}$, $(1,0)^{\mathrm{T}}$ and $(1,0)^{\mathrm{T}}$ represent a local minimizer. One possibility is given by arrays of the form

$$\mathbf{X} = \left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & b \end{array}\right)$$

Indeed, parametrizing the normalized component vectors $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$ in (2) by

$$(\cos\alpha, \sin\alpha)^{\mathrm{T}}, \qquad (\cos\beta, \sin\beta)^{\mathrm{T}}, \qquad (\cos\gamma, \sin\gamma)^{\mathrm{T}}$$

problem (2) with the given $\mathbf{X}$ amounts to finding the maximum of the function

$$f(\alpha, \beta, \gamma) = (\cos\alpha \cos\beta \cos\gamma + b \sin\alpha \sin\beta \sin\gamma)^2$$

The concrete component vectors above relate to the case $\alpha = \beta = \gamma = 0$. Elementary analysis shows

that at this triple the derivative of f vanishes while the Hessian equals -2. (identity matrix). Hence f realizes a local maximum at the origin. By the structural equivalence of problems (1) and (2), the component vectors indicated above represent a local minimizer of PARAFAC in (1). Furthermore, these component vectors attain a proper local minimum (one that is not a global minimum) whenever $b^2 > 1$. This follows from the fact that the complementary component vectors $(0,1)^T$, $(0,1)^T$ and $(0,1)^T$ have a functional value of b^2 in (2); hence 1 cannot be the global maximum in (2) and thus the original component vectors cannot realize a global minimum in (1).

Now it becomes evident that the one-component PARAFAC decomposition is not structurally equivalent with the maximization of the leading entry in the context of core transformations: choosing $b = 3$ in $\mathbf{X}$ defined above, we have a proper local minimizer of PARAFAC at the unit component vectors, but at the same time, when $\mathbf{X}$ is considered to be a core array, Theorem 1 indicates a *global* maximum of the leading-entry criterion. Actually, the situation can be much worse: consider the data array

$$\mathrm{X} = \left(\begin{array}{cc|cc} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 \end{array}\right)$$

Although this shape of data array is different from the one considered above, elementary analysis shows again that the component vectors $(2,0)^T$, $(1,0)^T$ and $(1,0)^T$ realize a local minimum of the PARAFAC decomposition. Indeed, considering the corresponding normalized components in the context of problem (2), the function to be analysed now is

$$f(\alpha,\beta,\gamma) = (2\cos\alpha\cos\beta\cos\gamma + \cos\alpha\sin\beta\sin\gamma + 2\sin\alpha\sin\beta\sin\gamma)^2$$

Its derivative vanishes at the origin while the Hessian has eigenvalues -12, -8 and -4. Consequently, the normalized components constitute a local maximizer of (2) and the original components, which have the right scaling as compared to the normalized ones, represent a local minimizer of (1). Nevertheless, when $\mathbf{X}$ is considered as a core array by putting $\mathbf{C} := \mathbf{X}$, the necessary conditions (9) are violated. It follows that stationarity conditions of problem (1) do not carry all information about stationarity conditions for maximizing the leading-entry criterion. As a numerical consequence, the usual PARAFAC algorithm (based on alternating least squares only) would stop at the mentioned minimizer, since in each component direction a global minimum has been reached, whereas application of the algorithm in Section 3 immediately gives a next iterate with higher value of the leading-entry criterion. The shortcoming of PARAFAC cannot be compensated for in this situation by small perturbations of the iterate, since the local minimum is stable. In other words, starting the PARAFAC algorithm at any point close to the minimizer will lead to convergence towards the same proper local minimizer again. On the other hand, the algorithm for maximizing the leading-entry criterion ends up at a core with a leading squared entry of 5 (as compared to the original value of 4) under the optimal transformation matrices

$$\mathbf{P} = \begin{pmatrix} 1/\sqrt{5} & -2/\sqrt{5} \\ 2/\sqrt{5} & 1/\sqrt{5} \end{pmatrix}, \qquad \mathbf{Q} = \mathbf{R} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

We also know that the improved solution of the leading-entry criterion must translate to an improved solution in the PARAFAC context: taking the first columns of the above transformation matrices and multiplying the first column of $\mathbf{P}$ by the optimal scaling factor $\sqrt{5}$, we arrive at the component vectors $(1,2)^T$, $(0,1)^T$ and $(0,1)^T$ for which the least squares function in (1) reduces from 5 (corresponding to the proper local minimizer) to 4.

REFERENCES

1. Tucker LR. Some mathematical notes on three-mode factor analysis. *Psychometrika* 1966; **31**: 279.
2. Kroonenberg PM, de Leeuw J. Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika* 1980; **45**: 69.
3. Harshman RA. Foundations of the PARAFAC procedure: models and conditions for an exploratory multimode factor analysis. *UCLA Working Papers Phonet*. 1970; **16**: 1.
4. Kroonenberg PM. *Three-mode Principal Component Analysis: Theory and Applications*. DSWO Press: Leiden, 1983.
5. Kiers HAL. TUCKALS core rotations and constrained TUCKALS modelling. *Statist. Appl.* 1992; **4**: 659.
6. Kiers HAL. Three-way SIMPLIMAX for oblique rotation of the three-mode factor analysis core to simple structure. *Comput. Statist. Data Anal.* 1998; **28**: 307.
7. Kiers HAL. Three-mode orthomax rotation. *Psychometrika* 1997; **62**: 579.
8. Henrion R, Andersson CA. A new criterion for simple-structure transformations of core arrays in *N*-way PCA. *Chemometrics Intell. Lab. Syst.* 1999; **47**: 189.
9. Murakami T. Quasi three-mode principal component analysis— a method for assessing factor change. *Behaviormetrika* 1983; **14**: 27.
10. Andersson CA, Henrion R. A new general algorithmic approach for obtaining simple-structure *N*-way core arrays with an application to a 3-way data array from fluorometry. *Comput. Statist. Data Anal.* 1999; **31**: 255.

