JOURNAL OF CHEMOMETRICS
J. Chemometrics 2004; **18**: 2–11
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/cem.838

First-order perturbation analysis of the best rank-(R_1, R_2, R_3) approximation in multilinear algebra

Lieven De Lathauwer*
ETIS, CNRS UMR 8051, 6 Avenue du Ponceau, BP 44, F-95014 Cergy-Pontoise Cedex, France

Received 25 July 2003; Revised 21 December 2003; Accepted 26 December 2003

In this paper we perform a first-order perturbation analysis of the least squares approximation of a given higher-order tensor by a tensor having prespecified *n*-mode ranks. This work generalizes the classical first-order perturbation analysis of the matrix singular value decomposition. We will show that there are important differences between the matrix and the higher-order tensor case. We subsequently address (1) the best rank-1 approximation of supersymmetric tensors, (2) the best rank-(R_1, R_2, R_3) approximation of arbitrary tensors and (3) the best rank-(R_1, R_2, R_3) approximation of arbitrary tensors. Copyright © 2004 John Wiley & Sons, Ltd.

KEYWORDS: multilinear algebra; higher-order tensors; singular value decomposition; rank reduction; perturbation analysis

1. INTRODUCTION

1.1. Basic definitions and notation

Multilinear algebra is the algebra of higher-order tensors, which are the higher-order equivalents of vectors (first order) and matrices (second order), i.e. quantities of which the elements are addressed by more than two indices. Multilinear algebra is gaining more and more interest; it has important applications in chemometrics, psychometrics, spatial division multiple access in telecommunications, blind source separation and harmonic retrieval in signal processing, blind identification/deconvolution, higher-order statistics, etc.

Before defining the problem that will be treated in this paper, we introduce some basic definitions and notations.

For convenience our exposition is primarily limited to third-order real-valued tensors. The generalization to tensors of orders higher than three and/or complex-valued tensors is straightforward but more cumbersome from a notational point of view.

To facilitate the distinction between scalars, vectors, matrices and higher-order tensors, we will indicate the type of a given quantity by its representation: scalars are denoted by lower-case letters (a, b, …; α, β, …), vectors are written as italic capitals (A, B, …), matrices correspond to bold-face capitals ($\mathbf{A}$, $\mathbf{B}$, …) and tensors are written as script letters ($\mathcal{A}$, $\mathcal{B}$, …). This notation is consistently used for lower-order parts of a given structure. For instance, the entry with row index i and column index j in a matrix $\mathbf{A}$, i.e. $(\mathbf{A})_{ij}$, is symbolized by a_{ij} (also $(A)_i = a_i$ and $(\mathcal{A})_{i_1 i_2 i_3} = a_{i_1 i_2 i_3}$). To enhance the overall readability, we have made one exception to this rule: as we sometimes use the characters i, j, r and n in the meaning of indices (counters), I, J, R and N will be reserved to denote the respective index upper bounds, unless stated otherwise.

The 1-mode product [1] of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ by a matrix $\mathbf{U} \in \mathbb{R}^{J_1 \times I_1}$, denoted by $\mathcal{A} \times_1 \mathbf{U}$, is a $(J_1 \times I_2 \times I_3)$-tensor of which the entries are given by

$$(\mathcal{A} \times_1 \mathbf{U})_{j_1 i_2 i_3} = \sum_{i_1} a_{i_1 i_2 i_3} u_{j_1 i_1}$$

The 2-mode and the 3-mode product, represented by $\times_2$ and $\times_3$, are defined in a similar way, by a summation over the second and the third index of $\mathcal{A}$ respectively.

It is convenient to express certain results in matrix terms. To this end we must stack the elements of a higher-order tensor in a matrix. There are several ways to do so. One particular type of 'matricized version' will prove to be particularly useful, namely, the matrix representation of a given tensor in which all its column (row, …) vectors are simply stacked one after another. To avoid confusion, we will retain one particular ordering of the column (row, …) vectors; these matricization procedures are visualized in Figure 1.

Column and row vectors are specific examples of n-mode vectors. An n-mode vector of an $(I_1 \times I_2 \times I_3)$-tensor $\mathcal{A}$ is an I_n-dimensional vector obtained from $\mathcal{A}$ by varying the index i_n and keeping the other indices fixed. The n-rank of a higher-order tensor is the obvious generalization of the column (row) rank of matrices: it equals the dimension of the vector

*Correspondence to: L. De Lathauwer, ETIS, CNRS UMR 8051, 6 Avenue du Ponceau, BP 44, F-95014 Cergy-Pontoise Cedex, France.
E-mail: delathau@ensea.fr
Contract/grant sponsor: Research Council KU Leuven; Contract/grant number: GOA-MEFISTO-666.
Contract/grant sponsor: Flemish Government; Contract/grant number: G.0240.99.
Contract/grant sponsor: Belgian Federal Government; Contract/grant numbers: IUAP IV-02; IUAP V-22.

Copyright © 2004 John Wiley & Sons, Ltd.

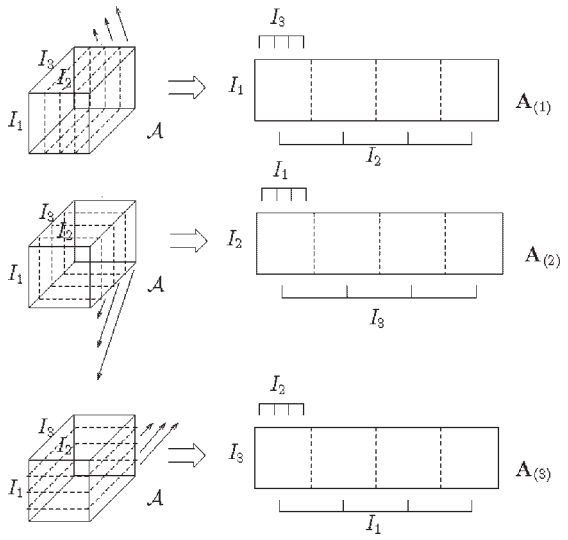


Figure 1. Mapping of the $(I_1 \times I_2 \times I_3)$-tensor $\mathcal{A}$ to the $(I_1 \times I_2 I_3)$-matrix $\mathbf{A}_{(1)}$, the $(I_2 \times I_3 I_1)$-matrix $\mathbf{A}_{(2)}$ and the $(I_3 \times I_1 I_2)$-matrix $\mathbf{A}_{(3)}$ $(I_1 = I_2 = I_3 = 4)$.

space spanned by the n-mode vectors. An important difference from the rank of matrices is that the different n-ranks of a higher-order tensor are not necessarily the same. A tensor of which the 1-rank, 2-rank and 3-rank are equal to R_1, R_2 and R_3 respectively is called a rank-(R_1, R_2, R_3) tensor. A rank-(1, 1, 1) tensor is briefly called a rank-1 tensor. A rank-1 tensor is formed by the outer product of vectors, i.e. $a_{ijk} = u_i v_j w_k$ for all values of the indices, which will be written as $\mathcal{A} = U \circ V \circ W$.

The scalar product $\langle \mathcal{A}, \mathcal{B} \rangle$ of two tensors $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ is defined in a straightforward way as $\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i_1} \sum_{i_2} \sum_{i_3} a_{i_1 i_2 i_3} b_{i_1 i_2 i_3}$. The Frobenius norm of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ is then defined as $\|\mathcal{A}\| = \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle}$.

A real-valued tensor is called supersymmetric when it is invariant under arbitrary permutations of its indices. The $(N \times N)$ identity matrix is denoted by $\mathbf{I}_{N \times N}$. Finally, $^{\mathrm{T}}$ and $^{\mathrm{H}}$ denote the transpose and the complex conjugated transpose of a matrix respectively.

1.2. Problem definition

The problem we start from in this paper has been proposed and studied in References [2–5] and has been further analyzed from a numerical algebraic viewpoint in Reference [6].

Given a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, we want to find a rank-(R_1, R_2, R_3) tensor $\hat{\mathcal{A}} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ such that the least squares cost function

$$f(\hat{\mathcal{A}}) = \|\mathcal{A} - \hat{\mathcal{A}}\|^2 \tag{1}$$

is minimized.

The n-rank conditions imply that $\hat{\mathcal{A}}$ can be decomposed as

$$\hat{\mathcal{A}} = \mathcal{T} \times_1 \mathbf{U} \times_2 \mathbf{V} \times_3 \mathbf{W} \tag{2}$$

in which $\mathbf{U} \in \mathbb{R}^{I_1 \times R_1}$, $\mathbf{V} \in \mathbb{R}^{I_2 \times R_2}$ and $\mathbf{W} \in \mathbb{R}^{I_3 \times R_3}$ each have orthonormal columns and in which $\mathcal{T} \in \mathbb{R}^{R_1 \times R_2 \times R_3}$ is a full rank-(R_1, R_2, R_3) tensor. Sometimes $\mathcal{T}$ is normalized to an all-orthogonal tensor, in which case Equation (2) corresponds to a Tucker-3 model of the best rank-(R_1, R_2, R_3) approximation [1,7]. However, this normalization is not necessary.

It can be proved that the minimization of cost function f is equivalent to the maximization of

$$g(\mathbf{U}, \mathbf{V}, \mathbf{W}) = \|\mathcal{A} \times_1 \mathbf{U}^{\mathrm{T}} \times_2 \mathbf{V}^{\mathrm{T}} \times_3 \mathbf{W}^{\mathrm{T}}\|^2 \tag{3}$$

After computation of the optimal $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$ the optimal $\mathcal{T}$ can be obtained by considering the minimization of (1) as a linear least squares problem in $\mathcal{T}$.

This problem is referred to as multimode principal component analysis (PCA) or as the best (least squares) rank-(R_1, R_2, R_3) approximation problem. It is actually a multilinear generalization of the best rank-R approximation problem in matrix algebra, which can be solved by a truncation of the singular value decomposition (SVD). In the multilinear case the computation is harder; techniques are studied in References [2–6].

Multimode PCA is a very important data processing tool. It allows for a least squares reduction of the given tensor such that the subsequent data manipulations can be realized in a low-dimensional space with minimum loss of accuracy [8–11].

In this paper we will carry out a first-order perturbation analysis of the best rank-(R_1, R_2, R_3) approximation problem. This means that we will examine the effect of small changes in $\mathcal{A}$ on the matrices $\mathbf{U}, \mathbf{V}$ and $\mathbf{W}$. Only the column spaces of $\mathbf{U}, \mathbf{V}$ and $\mathbf{W}$ are of importance, since the value of g in (3) does not change when $\mathbf{U}, \mathbf{V}$ or $\mathbf{W}$ is post-multiplied by an orthogonal matrix. Hence the problem is best studied on the (product of) Grasmann manifold(s). (The Stiefel manifold is the manifold of matrices of which the columns are orthonormal. The Grasmann manifold is the quotient of the Stiefel manifold over the equivalence relation $\mathbf{U}_1 \sim \mathbf{U}_2$ iff $\mathbf{U}_1 \sim \mathbf{U}_2 \cdot \mathbf{Q}$ with $\mathbf{Q}$ orthogonal, i.e. when the columns of $\mathbf{U}_1$ and $\mathbf{U}_2$ span the same subspace. Thus the Grasmann manifold can be imagined as the manifold of subspaces.) Because of the link with the best rank-R approximation of matrices, our results are a multilinear generalization of the well-known perturbation analysis for the SVD and the symmetric eigenvalue decomposition (EVD) [12–14].

The results of this paper may be used for a sensitivity analysis as follows. Consider a first-order perturbation $\tilde{\mathcal{A}} = \mathcal{A} + \epsilon \mathcal{B}$. Let the best rank-$(R_1, R_2, R_3)$ approximation of $\tilde{\mathcal{A}}$ be characterized by the perturbed matrices $\mathbf{U} + \epsilon \overline{\mathbf{U}} \cdot \mathbf{X}$, $\tilde{\mathbf{V}} = \mathbf{V} + \epsilon \overline{\mathbf{V}} \cdot \mathbf{Y}$ and $\tilde{\mathbf{W}} = \mathbf{W} + \epsilon \overline{\mathbf{W}} \cdot \mathbf{Z}$, in which the columns of $\overline{\mathbf{U}}$, $\overline{\mathbf{V}}$ and $\overline{\mathbf{W}}$ form orthonormal bases for the orthogonal complements of the column spaces of $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$ respectively. We will show that, except for some special cases, the entries of the matrices $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$, quantifying the effect of the perturbation, can be computed from a linear set of equations. More precisely, if we stack the entries of $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ in a parameter vector Ψ, then Ψ follows from a set of the form

$$\mathbf{M}(\mathcal{A}, \mathbf{U}, \mathbf{V}, \mathbf{W}) \cdot \Psi = N(\mathcal{A}, \mathcal{B}, \mathbf{U}, \mathbf{V}, \mathbf{W}) \tag{4}$$

in which $\mathbf{M}$ does not depend on $\mathcal{B}$. Thus, for a certain variation of the entries of $\mathcal{A}$, described by $\mathcal{B}$, the first-order effect on $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$ immediately follows from (4). Depending on the application, one may not know the precise form of the perturbing tensor $\mathcal{B}$. Instead, one may be interested in a general notion of the sensitivity of the best rank-(R_1, R_2, R_3) approximation. In such a context, inspection of the singular value spectrum of $\mathbf{M}$ may be useful. Indeed, if all the singular values of $\mathbf{M}$ are big, then we have $\|\Psi\| < \|N\|$, i.e. the

Copyright © 2004 John Wiley & Sons, Ltd.

effect of the perturbation is smaller than the perturbation itself, and the problem is well-conditioned. On the other hand, if all the singular values of $\mathbf{M}$ are small, then we have $\|\Psi\| > \|N\|$ and the problem is usually ill-conditioned. (In Section 4 it will be explained that there are some exceptions.) When there are big as well as small singular values, one may actually predict which perturbations can have a big effect on the best rank-(R_1, R_2, R_3) approximation by considering the whole SVD of $\mathbf{M}$. Such perturbations are characterized by a vector N that has significant components in the direction of the left singular vectors corresponding to small singular values; in other words, large entries of Ψ will be required to generate N as a linear combination of the columns of $\mathbf{M}$.

Instead of the general form of (2), in which the components are known up to an orthogonal transformation, one may prefer the normalized Tucker-3 representation. After carrying out the perturbation analysis on the Grasmann manifold, the perturbation of the individual columns of $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$ and of the tensor $\mathcal{T}$ then follows from the perturbation expressions proposed in Reference [15].

This paper is organized as follows. In the next section we first briefly recall some crucial facts related to the (perturbation of the) SVD of a matrix. In Section 3 we show that these basic facts cannot be readily taken over for higher-order tensors; the multilinear case is more complex. In Section 4 we derive and discuss perturbation expressions for the best supersymmetric rank-1 approximation of supersymmetric tensors. Sections 5 and 6 deal with the best rank-(R, R, R) and the best rank-(R_1, R_2, R_3) approximation of supersymmetric and arbitrary tensors respectively. Rewriting the core results in the form of (4) requires some technical formula manipulations; this derivation is given in the Appendix. Section 7 is the conclusion.

2. THE MATRIX CASE REVISITED

First we recall the well-known fact that the best rank-R approximation of a symmetric matrix is itself symmetric.

Secondly, it is also well known that, in the matrix case, the cost function f has no local optima. The subspaces corresponding to the dominant singular values yield the global minimum, the subspaces corresponding to the weakest singular values yield the global maximum and the subspaces corresponding to other subsets of singular values yield saddle points. This is visualized in the following example.

Example 1

Consider a (3×3) diagonal matrix $\mathbf{A}$ with diagonal elements 2, 1.5 and 1. Let $U = (\cos\alpha \;\; \sin\alpha\cos\beta \;\; \sin\alpha\sin\beta)^{\mathrm{T}}$ represent an arbitrary unit-norm vector in $\mathbb{R}^3$. In Figure 2 we plot $\lambda = U^{\mathrm{T}}\mathbf{A}U$ as a function of α and β. We have $g(\alpha, \beta) = \lambda^2$, with function g in the matrix case defined in analogy with (3). It can be seen that the global maximum of g (i.e. the global minimum of f) is reached for $U = (1\ 0\ 0)^{\mathrm{T}}$, that the global minimum of g is reached for $U = (0\ 0\ 1)^{\mathrm{T}}$ and that $U = (0\ 1\ 0)^{\mathrm{T}}$ corresponds to a saddle point. There are no local optima. □

Let now $\mathbf{A}(x)$ be a matrix, the elements of which are analytic functions of a real parameter x. Note that the first-order perturbation $\mathbf{A}(x) = \mathbf{A}(0) + x\,\mathbf{B}$ is a special case. It can be proven that the singular values and singular vectors of $\mathbf{A}(x)$ are also analytic functions of x, provided the terms are not ordered and the singular values are not constrained to be positive. Claiming that the singular values should be non-negative or putting them in order of decreasing absolute value may break the analyticity. In simple words, if a matrix $\mathbf{A}$ varies smoothly as a function of a parameter, its singular values and vectors also vary smoothly as a function of that parameter, provided the singular values are not constrained to be positive or put in a prespecified order. This is illustrated in the next example, taken from Reference [13].

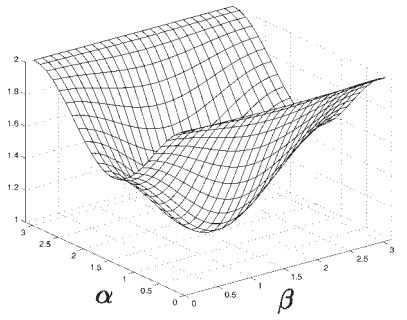


Figure 2. The matrix best rank-1 approximation problem has no local optima. There is one global maximum and one global minimum; the other critical points are saddle points.

Example 2

Consider a (3×3) diagonal matrix $\mathbf{A}(x)$ with diagonal elements

$$f_1(x) = \frac{x^2(3x^2+2)}{1+4x^2}$$
$$f_2(x) = \frac{3+2x^2}{4+x^4}$$
$$f_3(x) = \frac{1}{16}(x+1)^2 - \frac{1}{3}$$

which are analytic functions of a real parameter x. $f_1(x)$, $f_2(x)$ and $f_3(x)$ can be considered as unordered-signed singular value functions of $\mathbf{A}(x)$ (Figure 3). Imposing non-negativity and putting the singular values in order of decreasing magnitude leads to the functions in Figure 4 for the first and the third singular value. The latter functions are not analytic. □

When one singular value function becomes more dominant than another, i.e. at the intersection of (unsigned) singular value functions, the corresponding singular vectors can

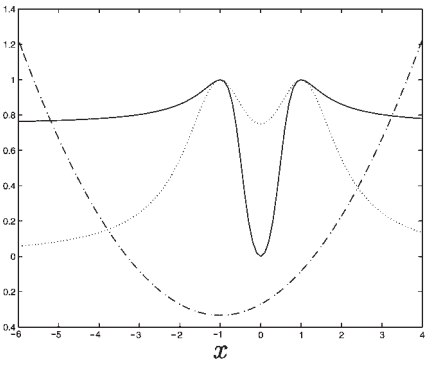


Figure 3. Unordered-signed singular value functions in Example 2.

Copyright © 2004 John Wiley & Sons, Ltd.

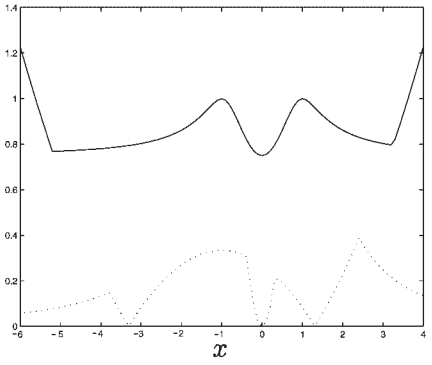


Figure 4. Non-negative ordered first and third singular value functions in Example 2.

be chosen as an arbitrary orthonormal basis of a *transition subspace* of which the dimension is equal to the multiplicity of the singular value. The vectors that form the analytic continuation of the singular vector functions are called preferred singular vectors. When singular values are close, the corresponding singular vectors are highly sensitive to perturbations, because the definition of the preferred singular vectors depends on the perturbation.

3. THE TENSOR CASE: SOME PRELIMINARY REMARKS

A first remarkable difference between matrices and tensors is that the best rank-(R, R, R) approximation of a supersymmetric tensor is itself not necessarily supersymmetric. We prove this by counterexample.

Example 3

Consider a tensor $\mathcal{A} = X \circ Y \circ Z + Z \circ X \circ Y + Y \circ Z \circ X$, in which X, Y and Z are orthonormal. By construction, $\mathcal{A}$ is supersymmetric. As explained in Section 1, the best supersymmetric rank-1 approximation $\hat{\mathcal{A}} = \lambda U \circ U \circ U$, in which λ is a scalar and U a unit-norm vector, can be found by maximization of $g(U) = |\mathcal{A} \times_1 U^{\mathrm{T}} \times_2 U^{\mathrm{T}} \times_3 U^{\mathrm{T}}|^2$. Let $U = \alpha_1 X + \alpha_2 Y + \alpha_3 Z$, with $|\alpha_1|^2 + |\alpha_2|^2 + |\alpha_3|^2 = 1$ (components that are not in the span of X, Y and Z are cancelled in the calculation of $g(U)$). We have

$$g(U) = |3\alpha_1\alpha_2\alpha_3|^2 = 9|\alpha_1|^2|\alpha_2|^2(1 - |\alpha_1|^2 - |\alpha_2|^2)$$

Straightforward optimization shows that the maximum is given by $|\alpha_1|^2 = |\alpha_2|^2 = |\alpha_3|^2 = \frac{1}{3}$. The corresponding value of $g(U)$ is $\frac{1}{3}$. However, this is not the best rank-1 approximation, because, for instance, $\hat{\mathcal{A}} = X \circ Y \circ Z$ is a better approximation ($g(X, Y, Z) = 1$).

However, in most applications involving supersymmetric tensors, one is only interested in approximations that can be decomposed as in (2) under the constraints that $\mathcal{T}$ is supersymmetric and $\mathbf{U} = \mathbf{V} = \mathbf{W}$. When we discuss the approximation of supersymmetric tensors, further in this section and in Sections 4 and 5, we implicitly assume that these constraints are satisfied. The general case of unsymmetric approximations is covered by the derivation in Section 6.

A second important difference between the tensor and the matrix case is that, for higher-order tensors, the cost function f can have local optima. This is proved by the next example.

Example 4

Consider a $(2 \times 2 \times 2)$ supersymmetric tensor $\mathcal{A}$ of which the matricized version $\mathbf{A}_{(1)}$ is given by

$$\mathbf{A}_{(1)} = \left(\begin{array}{cc|cc} 0 & 3 & 3 & 1 \\ 3 & 1 & 1 & -1 \end{array}\right)$$

Let $U = (\cos\alpha \ \sin\alpha)^{\mathrm{T}}$ represent an arbitrary unit-norm vector in $\mathbb{R}^2$. In Figure 5 we plot $\lambda = \mathcal{A} \times_1 U^{\mathrm{T}} \times_2 U^{\mathrm{T}} \times_3 U^{\mathrm{T}}$ as a function of α ($g(\alpha) = \lambda^2$). The function clearly has local optima. □

Let now $\mathcal{A}(x)$ be a real-valued tensor, the elements of which are analytic functions of a real parameter x. The critical points (where the gradient is zero) of the cost function f in (1) can be found via the real solutions of a set of polynomial equations with real coefficients. (For instance, for the best real rank-1 approximation of real supersymmetric tensors in Section 4, this set of equations will be given by (9), together with the normalization constraint $U^{\mathrm{T}}U = 1$.) This even applies to possibly complex-valued critical points of f. Here the unknowns of the set of polynomial equations can be taken equal to the real and imaginary parts of the complex unknowns. Note that only the real-valued solutions of the set of polynomial equations have a meaning. In general, real and complex solutions of a set of polynomial equations of which the coefficients are analytic in a real parameter x are smooth functions of x, provided they are not ordered or subject to constraints such as positivity. However, there are big differences between matrices and tensors here. In the matrix case we know that, if $\mathbf{A}(x)$ is real-valued, the critical points are also real-valued. In the tensor case there may be complex-valued critical points. Moreover, in the matrix case the total number of critical points, taking multiplicities into account, is fixed by the size of the matrix. In the tensor case the number of critical points may depend on the nominal value of the parameter x. We may for instance have that, at a nominal value of x, a double real-valued solution of the set of polynomial equations smoothly passes to two complex conjugated solutions. Because the complex-valued solutions, having no meaning, can be omitted, the number of critical points of f may change at such a transition point. Hence the critical points, considered as a function of x, may only be defined

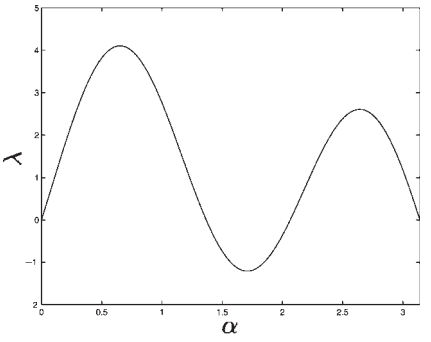


Figure 5. The tensor best rank-1 approximation problem can have local optima.

Copyright © 2004 John Wiley & Sons, Ltd.

over a certain interval of x, in contrast to the matrix case. This does not pose a problem for the further analysis. Transition points will be detected by the perturbation analysis in the following sections. For notational convenience we limit the exposition to real-valued critical points obtained for real-valued tensors. The discussion of complex-valued critical points for real-valued tensors and the analysis for complex-valued tensors are completely similar.

Example 5

Consider a supersymmetric tensor $\mathcal{A} \in \mathbb{R}^{2\times 2\times 2}$ and let a possibly complex-valued rank-1 approximation be represented by $\hat{\mathcal{A}}$. In this particular case the parametrization $\hat{\mathcal{A}} = \lambda\, U \circ U \circ U$, with scalar $\lambda \in \mathbb{C}$ and unit-norm vector $U = (\cos\alpha \;\; \sin\alpha\, \mathrm{e}^{\mathrm{i}\phi})^{\mathrm{T}}$, will prove to be useful. The critical points of f are the solutions of the equation

$$\mathcal{A} \times_2 U^{\mathrm{H}} \times_3 U^{\mathrm{H}} = \lambda U \tag{5}$$

which is the general form of (9) in Section 4 when also complex-valued solutions are allowed. After some goniometric manipulations we obtain the following solutions.

1. $U = (1 \;\; 0)^{\mathrm{T}}$ yields a solution if $a_{211} = 0$.
2. $U = (0 \;\; 1)^{\mathrm{T}}$ yields a solution if $a_{221} = 0$.
3. For $\cos\alpha \neq 0 \neq \sin\alpha$ there are two cases.

Case 1: $\phi = 0$ and α satisfies

$$\begin{aligned} &a_{211} + (2a_{221} - a_{111})\tan\alpha \\ &\quad + (a_{222} - 2a_{211})\tan^2\alpha - a_{221}\tan^3\alpha = 0 \end{aligned} \tag{6}$$

Case 2:

$$\cos\phi = \frac{a_{122}\sin^2\alpha - (a_{111} + 2a_{221})\cos^2\alpha}{2a_{222}\cos\alpha\sin\alpha} \tag{7}$$

and α satisfies

$$\begin{aligned} &(a_{222}^2 + 2a_{112}a_{222} + a_{111}a_{122})\tan^2\alpha \\ &= (a_{111}^2 + a_{211}a_{222} + 2a_{111}a_{221}) \end{aligned} \tag{8}$$

Equation (6) is just the condition for real-valued critical points, which has also been presented in Reference [6]. It is a polynomial of degree 3 in $\tan\alpha$ of which the roots have to be computed. Complex roots can be discarded, because they do not allow for an interpretation in terms of a critical point of the form $\lambda\, U \circ U \circ U$. If the entries of $\mathcal{A}$ depend on a parameter x, it is well possible that, for some values of x, Equation (6) admits three real roots, whereas, for other values of x, one real and two complex conjugated roots are obtained. Hence the number of real-valued critical points may depend on the value of x. On the other hand, Equation (8) may well have real-valued solutions in $\tan\alpha$. In combination with (7), these lead to complex-valued critical points that would not have been found from merely a real analysis. □

If the entries of a tensor are functions of a parameter, one local optimum may grow bigger than the original global optimum under variation of the parameter. When the two optima are exactly equal, there is not necessarily a transition subspace for the components of the approximation as in the matrix case. Instead, variation of the parameter may cause 'jumps' between distinct local optima. This is illustrated in the next example.

Example 6

Consider a $(2 \times 2 \times 2)$ supersymmetric tensor $\mathcal{A}(x)$ of which the matricized version $\mathbf{A}_{(1)}$ is given by

$$\mathbf{A}_{(1)} = \left(\begin{array}{cc|cc} 1+x & 0 & 0 & 0 \\ 0 & 0 & 0 & 1-x \end{array}\right)$$

For $x = 0$, $g(U)$ has two equivalent maxima, namely $U_1 = (1 \;\; 0)^{\mathrm{T}}$ and $U_2 = (0 \;\; 1)^{\mathrm{T}}$. For $x > 0$, U_1 yields the best rank-1 approximation and, for $x < 0$, U_2 yields the best rank-1 approximation. The transition between the two does not go via a subspace in which one can choose an arbitrary linear combination of U_1 and U_2. □

The way in which the best rank-(R_1, R_2, R_3) approximation can be underdetermined deserves further attention. In the matrix case, indeterminacies can only arise from coinciding singular values, and the number of degrees of freedom is linked with the multiplicity of the singular value. To show that the situation is much more complicated for higher-order tensors, we mention some results related to the best rank-1 approximation of $(2 \times 2 \times \cdots \times 2)$-tensors. First we have the following theorem.

Theorem 1

For supersymmetric $(2 \times 2 \times \cdots \times 2)$-tensors of odd order the best rank-1 approximation problem can only have a *discrete* number of equivalent solutions. For supersymmetric $(2 \times 2 \times \cdots \times 2)$-tensors of even order there are special cases in which all unit-norm vectors U are equivalent.

Proof

Consider a supersymmetric Nth-order $(2 \times 2 \times \cdots \times 2)$-tensor $\mathcal{A}$ and let a_n be a shorthand notation for the entry of which n indices are equal to 2 and $N - n$ indices are equal to 1. Let $U = (\cos\alpha \;\; \sin\alpha)^{\mathrm{T}}$ represent an arbitrary unit-norm vector in $\mathbb{R}^2$. Define $\lambda(\alpha) = \mathcal{A} \times_1 U^{\mathrm{T}} \times_2 U^{\mathrm{T}} \times \cdots \times_N U^{\mathrm{T}}$ $(g(\alpha) = \lambda^2(\alpha))$. We have

$$\lambda(\alpha) = \sum_{n=0}^{N} \frac{N!}{n!(N-n)!}\, a_n \cos^{N-n}\alpha \sin^n\alpha$$

In this expression the variational coefficient corresponds to the number of times the entry a_n appears in the tensor. Vectors $U(\alpha)$ are equivalent over a certain interval if $g(\alpha)$ is constant, hence if $\mathrm{d}\lambda(\alpha)/\mathrm{d}\alpha = \lambda'(\alpha) \equiv 0$. We have

$$\begin{aligned} \lambda'(\alpha) = {} & Na_1\cos^N\alpha + \sum_{n=1}^{N-1}\left(\frac{N!}{n!(N-n-1)!}\, a_{n+1}\right. \\ & \left. - \frac{N!}{(n-1)!(N-n)!}\, a_{n-1}\right)\cos^{N-n}\alpha\sin^n\alpha \\ & - Na_{N-1}\sin^N\alpha \end{aligned}$$

By writing out the coefficients of this expansion, it can be verified that, for odd tensor orders, $\lambda'(\alpha) \equiv 0$ implies that all the entries a_n should be equal to zero. On the other hand, if the tensor order is even, there are non-trivial values of a_n that make all the coefficients of $\lambda'(\alpha)$ vanish. ■

Example 7

In the fourth-order case the coefficients of $\lambda'(\alpha)$ are given by $4a_1$, $-4a_0 + 12a_2$, $-12a_1 + 12a_3$, $4a_4 - 12a_2$ and $-4a_3$. These coefficients can be set equal to zero by chosing $a_1 = a_3 = 0$

Copyright © 2004 John Wiley & Sons, Ltd.

and $a_0 = a_4 = 3a_2$. For instance, for the supersymmetric tensor $\mathcal{A}$ defined by

$$a_{1112} = a_{1222} = 0$$
$$a_{1122} = \frac{1}{3}$$
$$a_{1111} = a_{2222} = 1$$

we have

$$g(\alpha) = (\cos^4\alpha + 2\cos^2\alpha\sin^2\alpha + \sin^4\alpha)^2 \equiv 1$$

In the sense that the choice of U does not matter, $\mathcal{A}$ is the fourth-order generalization of the (2×2) identity matrix: just like we have $U^{\mathrm{T}} \cdot \mathbf{I}_{2\times 2} \cdot U = 1$ for arbitrary unit-norm vectors U, we have $\mathcal{A} \times_1 U^{\mathrm{T}} \times_2 U^{\mathrm{T}} \times_3 U^{\mathrm{T}} \times_4 U^{\mathrm{T}} = 1$ for arbitrary unit-norm U. □

For (2×2)-matrices, if the best rank-1 approximation is not unique, the solution manifold is at most of dimension 1. This is the case where the matrix has coinciding singular values. Let σ denote this singular value. The equivalent approximations take the form σUV^{T}, in which V is an arbitrary unit-norm vector and U is the corresponding left singular vector. Theorem 1 shows that for higher-order supersymmetric $(2 \times 2 \times \cdots \times 2)$-tensors the solution manifold is also at most of dimension 1; a one-dimensional manifold can only occur for tensors of even order. However, for higher-order $(2 \times 2 \times \cdots \times 2)$-tensors the manifold of equivalent solutions can be of dimension higher than 1 if one does not constrain the tensor to be supersymmetric. This is illustrated in the next example.

Example 8

Consider a $(2 \times 2 \times 2)$-tensor $\mathcal{A}$ of which the matricized version $\mathbf{A}_{(1)}$ is given by

$$\mathbf{A}_{(1)} = \left(\begin{array}{cc|cc} a & b & -b & a \\ -b & a & -a & -b \end{array}\right)$$

in which a and b are real scalars. Let $U = (\cos\alpha \;\; \sin\alpha)^{\mathrm{T}}$, $V = (\cos\beta \;\; \sin\beta)^{\mathrm{T}}$ and $W = (\cos\gamma \;\; \sin\gamma)^{\mathrm{T}}$ represent unit-norm vectors. For an arbitrary choice of W we have

$$\mathcal{A} \times_3 W = (a^2 + b^2)\,\mathbf{Q}$$

in which $\mathbf{Q}$ is an orthogonal matrix. Hence $g(U, V, W)$ can be maximized by picking an arbitrary value of β and γ (the solution manifold is two-dimensional) and taking $U = \mathbf{Q}V$. The maximal value is equal to $(a^2 + b^2)^2$. □

4. BEST RANK-1 APPROXIMATION OF SUPERSYMMETRIC TENSORS

4.1. Derivation

We first derive perturbation expressions for the best supersymmetric rank-1 approximation of supersymmetric tensors. A rank-1 tensor $\lambda\, U \circ U \circ U$ is a critical point of the cost function (1) for a given supersymmetric tensor $\mathcal{A} \in \mathbb{R}^{I\times I\times I}$ if it satisfies the Karush–Kuhn–Tucker (KKT) conditions [2–6]

$$\mathcal{A} \times_2 U^{\mathrm{T}} \times_3 U^{\mathrm{T}} = \lambda U \tag{9}$$

$$\|U\| = 1 \tag{10}$$

$$\lambda = \mathcal{A} \times_1 U^{\mathrm{T}} \times_2 U^{\mathrm{T}} \times_3 U^{\mathrm{T}} \tag{11}$$

Consider a first-order perturbation of $\mathcal{A}$, denoted by $\tilde{\mathcal{A}} = \mathcal{A} + \epsilon\mathcal{B}$, in which $\epsilon \in \mathbb{R}$ is small. Define $\tilde{\lambda} = \lambda + \epsilon\omega$ and $\tilde{U} = U + \epsilon\overline{\mathbf{U}}X$, in which the columns of $\overline{\mathbf{U}} \in \mathbb{R}^{I\times I-1}$ form an orthonormal basis for the orthogonal complement of U. We will examine under which conditions on X and ω the rank-1 tensor $\tilde{\lambda}\tilde{U} \circ \tilde{U} \circ \tilde{U}$ is a critical point of the cost function for $\tilde{\mathcal{A}}$. This allows us to characterize the effect of the perturbation of $\mathcal{A}$ on the critical point. The analysis applies to any critical point and hence also to the best rank-1 approximation.

The tensor $\tilde{\lambda}\tilde{U} \circ \tilde{U} \circ \tilde{U}$ is a critical point if it satisfies the KKT conditions for $\tilde{\mathcal{A}}$. Note that the specific form of $\tilde{U}$ ensures that this vector lies on the manifold of unit-norm vectors (up to first-order terms):

$$\tilde{U}^{\mathrm{T}}\tilde{U} = 1 + O(\epsilon^2)$$

Consider

$$\tilde{\mathcal{A}} \times_1 \tilde{U}^{\mathrm{T}} \times_2 \tilde{U}^{\mathrm{T}} = \tilde{\lambda}\tilde{U}$$

The zeroth-order terms vanish because of (9). The first-order terms vanish if

$$\begin{aligned}\tilde{\mathcal{A}} \times_2 (\overline{\mathbf{U}} \cdot X)^{\mathrm{T}} \times_3 U^{\mathrm{T}} + \tilde{\mathcal{A}} \times_2 U^{\mathrm{T}} \times_3 (\overline{\mathbf{U}} \cdot X)^{\mathrm{T}} \\ + \mathcal{B} \times_2 U^{\mathrm{T}} \times_3 U^{\mathrm{T}} = \omega U + \lambda\overline{\mathbf{U}} \cdot X\end{aligned} \tag{12}$$

Then 1-mode multiplication of this equation by $\overline{\mathbf{U}}^{\mathrm{T}}$ yields an expression that can be written as

$$\begin{aligned}[(\mathcal{A} \times_1 U^{\mathrm{T}} \times_2 U^{\mathrm{T}} \times_3 U^{\mathrm{T}})\,\mathbf{I}_{(I-1)\times(I-1)} \\ - 2\,\overline{\mathbf{U}}^{\mathrm{T}} \cdot \mathbf{A}' \cdot \overline{\mathbf{U}}] \cdot X = \overline{\mathbf{U}}^{\mathrm{T}} \cdot \mathbf{B}' \cdot U\end{aligned} \tag{13}$$

with $\mathbf{A}' = \mathcal{A} \times_3 U$ and $\mathbf{B}' = \mathcal{B} \times_3 U$. X can be computed from this linear set of equations, and ω can subsequently be determined by multiplying (12) by U^{T}. We obtain

$$\begin{aligned}\omega = 2\,\mathcal{A} \times_1 U^{\mathrm{T}} \times_2 (\overline{\mathbf{U}} \cdot X)^{\mathrm{T}} \times_3 U^{\mathrm{T}} \\ + \mathcal{B} \times_1 U^{\mathrm{T}} \times_2 U^{\mathrm{T}} \times_3 U^{\mathrm{T}}\end{aligned} \tag{14}$$

Note that (13) is indeed of the form (4) (in this particular case the right-hand side does not depend on $\mathcal{A}$).

4.2. Interpretation

Let us first illustrate the results obtained in the previous subsection by means of an example.

Example 9

Consider tensors $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{2\times 2\times 2}$, given in matricized form by

$$\mathbf{A}_{(1)} = \left(\begin{array}{cc|cc} 0.8 & 0.7 & 0.7 & 1.3 \\ 0.7 & 1.3 & 1.3 & 0.7 \end{array}\right)$$

$$\mathbf{B}_{(1)} = \left(\begin{array}{cc|cc} 1.2 & -1.2 & -1.2 & 0.3 \\ -1.2 & 0.3 & 0.3 & -0.2 \end{array}\right)$$

Let $U = (\cos\alpha \;\; \sin\alpha)^{\mathrm{T}}$ represent an arbitrary unit-norm vector in $\mathbb{R}^2$. In Figure 6 the full line shows the value α_{opt} that maximizes

$$g(U) = \|(\mathcal{A} + \epsilon\mathcal{B}) \times_1 U^{\mathrm{T}} \times_2 U^{\mathrm{T}} \times_3 U^{\mathrm{T}}\|^2$$

for varying values of ϵ. The broken line is the estimate of α_{opt} obtained by a first-order perturbation analysis around $\epsilon = 0$. The prediction of α_{opt} remains accurate over quite a big interval.

Copyright © 2004 John Wiley & Sons, Ltd.

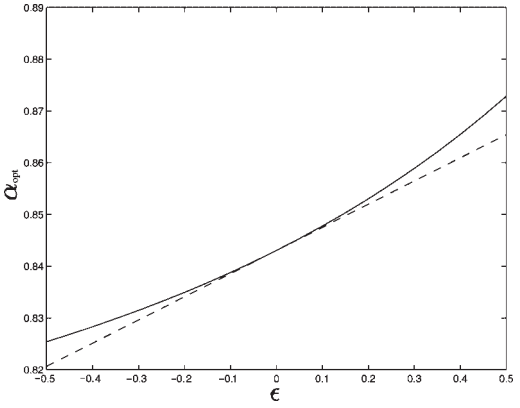


Figure 6. Best rank-1 approximation in Example 9: full line, optimal value; broken line, value predicted by first-order perturbation analysis.

The broken line was derived as follows. First, the best rank-1 approximation of $\mathcal{A}$ was computed. The optimal vector U was equal to $(0.67 \quad 0.75)^{\mathrm{T}}$. Its orthogonal complement $\overline{U}$ is given by $(-0.75 \quad 0.67)^{\mathrm{T}}$. In this $(2 \times 2 \times 2)$-dimensional example, Equation (13) reduces to a scalar equation: $3.12\,x = 0.14$, or $x = 0.045$. Hence a first-order perturbation analysis predicts that, if $\epsilon\,\mathcal{B}$ is added to $\mathcal{A}$, U will be replaced by $\tilde{U} = U + 0.045\,\epsilon\,\overline{U}$. Normalization to unit-length yields the angle displayed in Figure 6. □

The matrix counterpart of (13) is

$$[(U^{\mathrm{T}}\mathbf{A}U)\,\mathbf{I}_{(I-1)\times(I-1)} - \overline{\mathbf{U}}^{\mathrm{T}} \cdot \mathbf{A} \cdot \overline{\mathbf{U}}]X = \overline{\mathbf{U}}^{\mathrm{T}} \cdot \mathbf{B} \cdot U \qquad (15)$$

If we take U equal to the first and the columns of $\overline{\mathbf{U}}$ equal to the other eigenvectors of $\mathbf{A}$, then we obtain the well-known perturbation expression [14]

$$x_{i-1} = \frac{(\overline{\mathbf{U}}^{\mathrm{T}} \cdot \mathbf{B} \cdot U)_i}{\lambda_1 - \lambda_i} \qquad (16)$$

in which λ_i is the ith eigenvalue of $\mathbf{A}$. From this equation it is clear that eigenvectors are sensitive to perturbations when the corresponding eigenvalues are close: this leads to big entries of X and hence to a big perturbation term in $\tilde{U} = U + \epsilon\,\overline{\mathbf{U}}X$. In particular, when eigenvalues coincide, the coefficient matrix of X in (15) becomes singular and the system admits only infinite solutions in general. The reason is that the analytic continuation of the eigenvector functions, for varying ϵ, involves only the preferred eigenvectors for $\epsilon = 0$ (see Section 2). The other eigenvectors in the transition space are subject to a change of order $O(1)$ when the matrix is perturbed, and hence X in (15) and (16) takes infinite values. The preferred eigenvectors can be calculated as the eigenvectors for which the column vector on the right-hand side of (15) is in the span of the coefficient matrix on the left-hand side. Equivalently, the denominator in (16) should be equal to zero whenever the eigenvalues in the numerator are equal. One can verify that the solution can be obtained as follows. Let the columns of the matrix $\underline{\mathbf{U}}$ consist of an orthonormal basis of an eigenvalue with multiplicity $J > 1$. Define $\mathbf{C} = \underline{\mathbf{U}} \cdot \mathbf{B} \cdot \underline{\mathbf{U}}^{\mathrm{T}}$ and let the eigenvectors of $\mathbf{C}$ be represented by E_j $(1 \leq j \leq J)$. Then the preferred eigenvectors are given by $\underline{\mathbf{U}} \cdot E_j$ [13].

For higher-order tensors, however, a critical point of the cost function is not necessarily ill-conditioned if Equation (13) admits no solution. This is linked with the fact that in the higher-order case the cost function can have local optima and with the fact that indeterminacies can be discrete (not linked by a transition subspace). This is clarified in the next example.

Example 10

Consider a supersymmetric tensor $\mathcal{A} \in \mathbb{R}^{2\times2\times2}$, given in matricized form by

$$\mathbf{A}_{(1)} = \left(\begin{array}{cc|cc} 1 & 0 & 0 & 0.5 \\ 0 & 0.5 & 0.5 & 0 \end{array}\right)$$

Let $U = (\cos\alpha \;\; \sin\alpha)^{\mathrm{T}}$ represent an arbitrary unit-norm vector in $\mathbb{R}^2$. Define $\lambda = \mathcal{A} \times_1 U^{\mathrm{T}} \times_2 U^{\mathrm{T}} \times_3 U^{\mathrm{T}}$ $(g(\alpha) = \lambda^2)$. We have

$$\lambda(\alpha) = \cos^3\alpha + 1.5\cos\alpha\sin^2\alpha$$

and the derivative

$$\lambda'(\alpha) = -1.5\sin^3\alpha$$

Hence $\lambda(\alpha)$ has a single critical point (with multiplicity 3), namely $\alpha = 0$. The best rank-1 approximation of $\mathcal{A}$ is given by $U \circ U \circ U$, with $U = (1 \;\; 0)^{\mathrm{T}}$. Now consider a first-order perturbation $\mathcal{A} + \epsilon\,\mathcal{B}$. For $\alpha = 0$, Equation (13) takes the form

$$\left[1 - 2\,(0 \quad 1) \cdot \begin{pmatrix} 1 & 0 \\ 0 & 0.5 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right] x = (0 \quad 1) \cdot \mathbf{B}' \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} \qquad (17)$$

The coefficient of x is equal to zero, while the right-hand side of the equation is in general different from zero. The reason is that, by the perturbation, $\lambda'(\alpha)$ is replaced by a homogeneous polynomial $\tilde{\lambda}'(\alpha)$ of degree 3 in $\sin\alpha$ and $\cos\alpha$. The difference between the coefficients of $\lambda'(\alpha)$ and $\tilde{\lambda}'(\alpha)$ is of the order of ϵ. After perturbation the critical points can be obtained by dividing $\tilde{\lambda}'(\alpha)$ by $\cos^3\alpha$ and rooting a polynomial of degree 3 in $t = \tan\alpha$. As ϵ goes to zero, the roots converge to $\alpha = 0$, such that the best rank-1 approximation of $\mathcal{A}$ is well-conditioned. The difference between $\epsilon = 0$ and $\epsilon \neq 0$ is that in the first case there is a single critical point, with multiplicity 3, whereas in the latter there are (generally) three different critical points, with multiplicity 1. Hence the solution for $\epsilon = 0$ can be tracked in three different ways for varying ϵ. This ambiguity is reflected by the fact that Equation (17) does not have a solution for $\epsilon = 0$. □

Of course, it is still possible that Equation (13) not having a solution means that there is a transition subspace and that an analytic continuation is only possible for preferred basis vectors as in the matrix case. The computation of preferred basis vectors may be more difficult than in the matrix case. We give an example.

Copyright © 2004 John Wiley & Sons, Ltd.

Example 11

Reconsider the supersymmetric fourth-order tensor defined in Example 7. The fourth-order equivalent of (13) is given by

$$[(\mathcal{A} \times_1 U^{\mathrm{T}} \times_2 U^{\mathrm{T}} \times_3 U^{\mathrm{T}} \times_4 U^{\mathrm{T}})\, 1 - 3\,\overline{\mathbf{U}}^{\mathrm{T}} \cdot \mathbf{A}' \cdot \overline{\mathbf{U}}]X = \overline{\mathbf{U}}^{\mathrm{T}} \cdot \mathbf{B}' \cdot U \tag{18}$$

with $\mathbf{A}' = \mathcal{A} \times_3 U \times_4 U$ and $\mathbf{B}' = \mathcal{B} \times_3 U \times_4 U$. It can easily be verified that in our example the indeterminacy of the best rank-1 approximation is reflected by the fact that the coefficient on the left-hand side of (18) is equal to zero. The preferred basis vectors, for a given perturbation $\mathcal{B}$, can be computed by claiming that the right-hand side should also vanish. Using the parametrization $U = (\cos\alpha \;\; \sin\alpha)^{\mathrm{T}}$ and $\overline{U} = (-\sin\alpha \;\; \cos\alpha)^{\mathrm{T}}$, the right-hand side can be written as a homogeneous polynomial of degree 4 in $\sin\alpha$ and $\cos\alpha$. After division by $\cos^4\alpha$ the preferred basis vectors can be computed by rooting a polynomial of degree 4 in $t = \tan\alpha$. □

The analysis conducted in this section is a local analysis in the sense that we look at how a local critical point changes under perturbation. If we are interested in the global best rank-1 approximation, we should be aware that in the tensor case there can be jumps between different local optima, as explained in Section 3. For a perturbation analysis of the overall optimum we should repeat the calculations for the different candidate global optima and compare the results.

5. BEST RANK-(*R*, *R*, *R*) APPROXIMATION OF SUPERSYMMETRIC TENSORS

Consider the least squares approximation of a supersymmetric tensor $\mathcal{A} \in \mathbb{R}^{I\times I\times I}$ by a rank-(R, R, R) tensor $\mathcal{T} \times_1 \mathbf{U} \times_2 \mathbf{U} \times_3 \mathbf{U}$, in which $\mathcal{T} \in \mathbb{R}^{R\times R\times R}$ is supersymmetric and full rank-(R, R, R) and in which $\mathbf{U} \in \mathbb{R}^{I\times R}$ is column-wise orthonormal. The KKT condition (9) is now replaced by the optimality condition [2–6]

$$\mathbf{U} = \max_{\mathbf{U}^{\mathrm{T}}\mathbf{U}=\mathbf{I}} \|\mathbf{U}^{\mathrm{T}} \cdot \left[\mathbf{A}_{(1)} \cdot (\mathbf{U} \otimes \mathbf{U})\right]\|^2 \tag{19}$$

in which $\otimes$ denotes the Kronecker product. This makes the problem more complicated in comparison with the best rank-1 approximation of higher-order tensors and the best rank-R approximation of matrices. Equation (19) implies that critical points now correspond to invariant subspaces of the matrix

$$\mathbf{F}_{\mathbf{U}} = \mathbf{A}_{(1)} \cdot (\mathbf{U} \otimes \mathbf{U}) \cdot (\mathbf{U} \otimes \mathbf{U})^{\mathrm{T}} \cdot \mathbf{A}_{(1)}^{\mathrm{T}}$$

Suppose we have

$$\mathbf{F}_{\mathbf{U}} \cdot \mathbf{U} = \mathbf{U} \cdot \mathbf{\Omega} \tag{20}$$

with $\mathbf{\Omega} \in \mathbb{R}^{R\times R}$. Consider the perturbation $\mathcal{A} + \epsilon\,\mathcal{B}$, leading to a new critical point, characterized by $\tilde{\mathbf{U}} = \mathbf{U} + \epsilon\,\overline{\mathbf{U}} \cdot \mathbf{X}$ (the columns of $\overline{\mathbf{U}} \in \mathbb{R}^{I\times(I-R)}$ form an orthonormal basis for the orthogonal complement of the column space of $\mathbf{U}$) and $\tilde{\mathbf{\Omega}} = \mathbf{\Omega} + \epsilon\,\mathbf{\Theta}$, in which $\mathbf{X} \in \mathbb{R}^{(I-R)\times R}$ and $\mathbf{\Theta} \in \mathbb{R}^{R\times R}$ are unknown. Denoting

$$\tilde{\mathbf{F}}_{\tilde{\mathbf{U}}} = (\mathbf{A}_{(1)} + \epsilon\,\mathbf{B}_{(1)}) \cdot (\tilde{\mathbf{U}} \otimes \tilde{\mathbf{U}}) \cdot (\tilde{\mathbf{U}} \otimes \tilde{\mathbf{U}})^{\mathrm{T}} \cdot (\mathbf{A}_{(1)} + \epsilon\,\mathbf{B}_{(1)})^{\mathrm{T}}$$

we should have

$$\tilde{\mathbf{F}}_{\tilde{\mathbf{U}}} \cdot \tilde{\mathbf{U}} = \tilde{\mathbf{U}} \cdot \tilde{\mathbf{\Omega}}$$

The zeroth-order terms vanish because of (20). The first-order terms vanish if

$$\begin{aligned}
&\mathbf{B}_{(1)} \cdot (\mathbf{U} \otimes \mathbf{U}) \cdot (\mathbf{U} \otimes \mathbf{U})^{\mathrm{T}} \cdot \mathbf{A}_{(1)}^{\mathrm{T}} \cdot \mathbf{U} \\
&\quad + \mathbf{A}_{(1)} \cdot (\mathbf{U} \otimes \mathbf{U}) \cdot (\mathbf{U} \otimes \mathbf{U})^{\mathrm{T}} \cdot \mathbf{B}_{(1)}^{\mathrm{T}} \cdot \mathbf{U} \\
&\quad + \mathbf{A}_{(1)} \cdot \left[(\overline{\mathbf{U}} \cdot \mathbf{X}) \otimes \mathbf{U} + \mathbf{U} \otimes (\overline{\mathbf{U}} \cdot \mathbf{X})\right] \\
&\quad \cdot (\mathbf{U} \otimes \mathbf{U})^{\mathrm{T}} \cdot \mathbf{A}_{(1)}^{\mathrm{T}} \cdot \mathbf{U} \\
&\quad + \mathbf{A}_{(1)} \cdot (\mathbf{U} \otimes \mathbf{U}) \cdot [(\overline{\mathbf{U}} \cdot \mathbf{X}) \otimes \mathbf{U} \\
&\quad + \mathbf{U} \otimes (\overline{\mathbf{U}} \cdot \mathbf{X})] \cdot \mathbf{A}_{(1)}^{\mathrm{T}} \cdot \mathbf{U} \\
&\quad + \mathbf{A}_{(1)} \cdot (\mathbf{U} \otimes \mathbf{U}) \cdot (\mathbf{U} \otimes \mathbf{U})^{\mathrm{T}} \cdot \mathbf{A}_{(1)}^{\mathrm{T}} \cdot \overline{\mathbf{U}} \cdot \mathbf{X} \\
&= \overline{\mathbf{U}} \cdot \mathbf{X} \cdot \mathbf{\Omega} + \mathbf{U} \cdot \mathbf{\Theta}
\end{aligned} \tag{21}$$

If we denote

$$\begin{aligned}
\mathbf{C}_1 &= \overline{\mathbf{U}}^{\mathrm{T}} \cdot \mathbf{A}_{(1)} \cdot (\mathbf{U} \otimes \mathbf{U}) \\
\mathbf{C}_2 &= (\mathbf{U} \otimes \mathbf{U})^{\mathrm{T}} \cdot \mathbf{A}_{(1)}^{\mathrm{T}} \cdot \mathbf{U}
\end{aligned}$$

then left multiplication of (21) by $\overline{\mathbf{U}}^{\mathrm{T}}$ leads to

$$\begin{aligned}
&\Big\{-\mathbf{X} \cdot \mathbf{U}^{\mathrm{T}} \cdot \mathbf{A}_{(1)} \cdot (\mathbf{U} \otimes \mathbf{U}) + \overline{\mathbf{U}}^{\mathrm{T}} \cdot \mathbf{B}_{(1)} \cdot (\mathbf{U} \otimes \mathbf{U}) \\
&\quad + 2\,\overline{\mathbf{U}}^{\mathrm{T}} \cdot \mathbf{A}_{(1)} \cdot \left[(\overline{\mathbf{U}} \cdot \mathbf{X}) \otimes \mathbf{U}\right]\Big\} \cdot \mathbf{C}_2 \\
&\quad + \mathbf{C}_1 \cdot \Big\{2\left[(\overline{\mathbf{U}} \cdot \mathbf{X}) \otimes \mathbf{U}\right]^{\mathrm{T}} \cdot \mathbf{A}_{(1)}^{\mathrm{T}} \cdot \mathbf{U} \\
&\quad + (\mathbf{U} \otimes \mathbf{U})^{\mathrm{T}} \cdot \mathbf{A}_{(1)}^{\mathrm{T}} \cdot \overline{\mathbf{U}} \cdot \mathbf{X} + (\mathbf{U} \otimes \mathbf{U})^{\mathrm{T}} \cdot \mathbf{B}_{(1)}^{\mathrm{T}} \cdot \mathbf{U}\Big\} = \mathbf{0}
\end{aligned} \tag{22}$$

Although it looks complicated, this expression is just a linear set of equations in the unknown $\mathbf{X}$, from which $\mathbf{X}$ may be computed. It is explicitly rewritten in the form of (4) in the Appendix. If, after perturbation, the new critical point is given by $\tilde{\mathcal{T}} \times_1 \tilde{\mathbf{U}} \times_2 \tilde{\mathbf{U}} \times_3 \tilde{\mathbf{U}}$, then $\tilde{\mathcal{T}}$ may be computed by solving

$$\mathcal{A} = \tilde{\mathcal{T}} \times_1 \tilde{\mathbf{U}} \times_2 \tilde{\mathbf{U}} \times_3 \tilde{\mathbf{U}}$$

in the least squares sense, as we have already indicated in Section 1.2. Because of the orthonormality of the columns of $\tilde{\mathbf{U}}$, the solution is

$$\tilde{\mathcal{T}} = \mathcal{A} \times_1 \tilde{\mathbf{U}}^{\mathrm{T}} \times_2 \tilde{\mathbf{U}}^{\mathrm{T}} \times_3 \tilde{\mathbf{U}}^{\mathrm{T}}$$

6. BEST RANK-(R_1, R_2, R_3) APPROXIMATION OF ARBITRARY TENSORS

Finally we treat the general problem in which the original tensor can be unsymmetric and the approximation can be subject to arbitrary n-mode rank constraints.

Let the tensor to be approximated be represented by $\mathcal{A} \in \mathbb{R}^{I_1\times I_2\times I_3}$ and let a critical point of cost function f in (1) be parametrized as in (2). Instead of (19) we have now three conditions:

$$\mathbf{U} = \max_{\mathbf{U}^{\mathrm{T}}\mathbf{U}=\mathbf{I}} \|\mathbf{U}^{\mathrm{T}} \cdot \left[\mathbf{A}_{(1)} \cdot (\mathbf{V} \otimes \mathbf{W})\right]\|^2 \tag{23}$$

$$\mathbf{V} = \max_{\mathbf{V}^{\mathrm{T}}\mathbf{V}=\mathbf{I}} \|\mathbf{V}^{\mathrm{T}} \cdot \left[\mathbf{A}_{(2)} \cdot (\mathbf{W} \otimes \mathbf{U})\right]\|^2 \tag{24}$$

$$\mathbf{W} = \max_{\mathbf{W}^{\mathrm{T}}\mathbf{W}=\mathbf{I}} \|\mathbf{W}^{\mathrm{T}} \cdot \left[\mathbf{A}_{(3)} \cdot (\mathbf{U} \otimes \mathbf{V})\right]\|^2 \tag{25}$$

Define

$$\mathbf{F}_{\mathbf{V},\mathbf{W}} = \mathbf{A}_{(1)} \cdot (\mathbf{V} \otimes \mathbf{W}) \cdot (\mathbf{V} \otimes \mathbf{W})^{\mathrm{T}} \cdot \mathbf{A}_{(1)}^{\mathrm{T}} \tag{26}$$

Copyright © 2004 John Wiley & Sons, Ltd.

$$\mathbf{F}_{\mathbf{W},\mathbf{U}} = \mathbf{A}_{(2)} \cdot (\mathbf{W} \otimes \mathbf{U}) \cdot (\mathbf{W} \otimes \mathbf{U})^{\mathrm{T}} \cdot \mathbf{A}_{(2)}^{\mathrm{T}} \tag{27}$$

$$\mathbf{F}_{\mathbf{U},\mathbf{V}} = \mathbf{A}_{(3)} \cdot (\mathbf{U} \otimes \mathbf{V}) \cdot (\mathbf{U} \otimes \mathbf{V})^{\mathrm{T}} \cdot \mathbf{A}_{(3)}^{\mathrm{T}} \tag{28}$$

Conditions for a critical point are that $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$ represent invariant subspaces of $\mathbf{F}_{\mathbf{V},\mathbf{W}}$, $\mathbf{F}_{\mathbf{W},\mathbf{U}}$ and $\mathbf{F}_{\mathbf{U},\mathbf{V}}$ respectively. Let the perturbed matrices be represented by $\tilde{\mathbf{U}} = \mathbf{U} + \epsilon\,\overline{\mathbf{U}} \cdot \mathbf{X}$, $\tilde{\mathbf{V}} = \mathbf{V} + \epsilon\,\overline{\mathbf{V}} \cdot \mathbf{Y}$ and $\tilde{\mathbf{W}} = \mathbf{W} + \epsilon\,\overline{\mathbf{W}} \cdot \mathbf{Z}$, in which the columns of $\overline{\mathbf{U}} \in \mathbb{R}^{I_1 \times (I_1 - R_1)}$, $\overline{\mathbf{V}} \in \mathbb{R}^{I_2 \times (I_2 - R_2)}$ and $\overline{\mathbf{W}} \in \mathbb{R}^{I_3 \times (I_3 - R_3)}$ form orthonormal bases for the orthogonal complements of the column spaces of $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$ respectively. By working in a similar way as in the previous section, we obtain the following conditions on the unknowns $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$:

$$\begin{aligned}
\Big\{ & -\mathbf{X} \cdot \mathbf{U}^{\mathrm{T}} \cdot \mathbf{A}_{(1)} \cdot (\mathbf{V} \otimes \mathbf{W}) + \overline{\mathbf{U}}^{\mathrm{T}} \cdot \mathbf{B}_{(1)} \cdot (\mathbf{V} \otimes \mathbf{W}) \\
& + \overline{\mathbf{U}}^{\mathrm{T}} \cdot \mathbf{A}_{(1)} \cdot [(\overline{\mathbf{V}} \cdot \mathbf{Y}) \otimes \mathbf{W}] \\
& + \overline{\mathbf{U}}^{\mathrm{T}} \cdot \mathbf{A}_{(1)} \cdot [\mathbf{V} \otimes (\overline{\mathbf{W}} \cdot \mathbf{Z})] \} \cdot \mathbf{C}_{1,2} \\
& + \mathbf{C}_{1,1} \cdot \Big\{ \big[(\overline{\mathbf{V}} \cdot \mathbf{Y}) \otimes \mathbf{W}\big]^{\mathrm{T}} \cdot \mathbf{A}_{(1)}^{\mathrm{T}} \cdot \mathbf{U} \\
& + \big[\mathbf{V} \otimes (\overline{\mathbf{W}} \cdot \mathbf{Z})\big]^{\mathrm{T}} \cdot \mathbf{A}_{(1)}^{\mathrm{T}} \cdot \mathbf{U} \\
& + (\mathbf{V} \otimes \mathbf{W})^{\mathrm{T}} \cdot \mathbf{A}_{(1)}^{\mathrm{T}} \cdot \overline{\mathbf{U}} \cdot \mathbf{X} + (\mathbf{V} \otimes \mathbf{W})^{\mathrm{T}} \cdot \mathbf{B}_{(1)}^{\mathrm{T}} \cdot \mathbf{U} \Big\} = \mathbf{0}
\end{aligned} \tag{29}$$

$$\begin{aligned}
\Big\{ & -\mathbf{Y} \cdot \mathbf{V}^{\mathrm{T}} \cdot \mathbf{A}_{(2)} \cdot (\mathbf{W} \otimes \mathbf{U}) + \overline{\mathbf{V}}^{\mathrm{T}} \cdot \mathbf{B}_{(2)} \cdot (\mathbf{W} \otimes \mathbf{U}) \\
& + \overline{\mathbf{V}}^{\mathrm{T}} \cdot \mathbf{A}_{(2)} \cdot \big[(\overline{\mathbf{W}} \cdot \mathbf{Z}) \otimes \mathbf{U}\big] \\
& + \overline{\mathbf{V}}^{\mathrm{T}} \cdot \mathbf{A}_{(2)} \cdot \big[\mathbf{W} \otimes (\overline{\mathbf{U}} \cdot \mathbf{X})\big] \Big\} \cdot \mathbf{C}_{2,2} \\
& + \mathbf{C}_{2,1} \cdot \Big\{ \big[(\overline{\mathbf{W}} \cdot \mathbf{Z}) \otimes \mathbf{U}\big)^{\mathrm{T}} \cdot \mathbf{A}_{(2)}^{\mathrm{T}} \cdot \mathbf{V} \\
& + \big[\mathbf{W} \otimes (\overline{\mathbf{U}} \cdot \mathbf{X})\big]^{\mathrm{T}} \cdot \mathbf{A}_{(2)}^{\mathrm{T}} \cdot \mathbf{V} \\
& + (\mathbf{W} \otimes \mathbf{U})^{\mathrm{T}} \cdot \mathbf{A}_{(2)}^{\mathrm{T}} \cdot \overline{\mathbf{V}} \cdot \mathbf{Y} + (\mathbf{W} \otimes \mathbf{U})^{\mathrm{T}} \cdot \mathbf{B}_{(2)}^{\mathrm{T}} \cdot \mathbf{V} \Big\} = \mathbf{0}
\end{aligned} \tag{30}$$

$$\begin{aligned}
\Big\{ & -\mathbf{Z} \cdot \mathbf{W}^{\mathrm{T}} \cdot \mathbf{A}_{(3)} \cdot (\mathbf{U} \otimes \mathbf{V}) + \overline{\mathbf{W}}^{\mathrm{T}} \cdot \mathbf{B}_{(3)} \cdot (\mathbf{U} \otimes \mathbf{V}) \\
& + \overline{\mathbf{W}}^{\mathrm{T}} \cdot \mathbf{A}_{(3)} \cdot \big[(\overline{\mathbf{U}} \cdot \mathbf{X}) \otimes \mathbf{V}\big] \\
& + \overline{\mathbf{W}}^{\mathrm{T}} \cdot \mathbf{A}_{(3)} \cdot \big[\mathbf{U} \otimes (\overline{\mathbf{V}} \cdot \mathbf{Y})\big] \Big\} \cdot \mathbf{C}_{3,2} \\
& + \mathbf{C}_{3,1} \cdot \Big\{ \big[(\overline{\mathbf{U}} \cdot \mathbf{X}) \otimes \mathbf{V}\big]^{\mathrm{T}} \cdot \mathbf{A}_{(3)}^{\mathrm{T}} \cdot \mathbf{W} \\
& + \big[\mathbf{U} \otimes (\overline{\mathbf{V}} \cdot \mathbf{Y})\big]^{\mathrm{T}} \cdot \mathbf{A}_{(3)}^{\mathrm{T}} \cdot \mathbf{W} \\
& + (\mathbf{U} \otimes \mathbf{V})^{\mathrm{T}} \cdot \mathbf{A}_{(3)}^{\mathrm{T}} \cdot \overline{\mathbf{W}} \cdot \mathbf{Z} + (\mathbf{U} \otimes \mathbf{V})^{\mathrm{T}} \cdot \mathbf{B}_{(3)}^{\mathrm{T}} \cdot \mathbf{W} \Big\} = \mathbf{0}
\end{aligned} \tag{31}$$

in which

$$\begin{aligned}
\mathbf{C}_{1,1} &= \overline{\mathbf{U}}^{\mathrm{T}} \cdot \mathbf{A}_{(1)} \cdot (\mathbf{V} \otimes \mathbf{W}) \\
\mathbf{C}_{1,2} &= (\mathbf{V} \otimes \mathbf{W})^{\mathrm{T}} \cdot \mathbf{A}_{(1)}^{\mathrm{T}} \cdot \mathbf{U} \\
\mathbf{C}_{2,1} &= \overline{\mathbf{V}}^{\mathrm{T}} \cdot \mathbf{A}_{(2)} \cdot (\mathbf{W} \otimes \mathbf{U}) \\
\mathbf{C}_{2,2} &= (\mathbf{W} \otimes \mathbf{U})^{\mathrm{T}} \cdot \mathbf{A}_{(2)}^{\mathrm{T}} \cdot \mathbf{V} \\
\mathbf{C}_{3,1} &= \overline{\mathbf{W}}^{\mathrm{T}} \cdot \mathbf{A}_{(3)} \cdot (\mathbf{U} \otimes \mathbf{V}) \\
\mathbf{C}_{3,2} &= (\mathbf{U} \otimes \mathbf{V})^{\mathrm{T}} \cdot \mathbf{A}_{(3)}^{\mathrm{T}} \cdot \mathbf{W}
\end{aligned}$$

Equations (29)–(31) form a set of linear equations from which $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ may be calculated. They are explicitly rewritten in the form of (4) in the Appendix. The tensor $\tilde{\mathcal{T}}$ is subsequently given by

$$\tilde{\mathcal{T}} = \mathcal{A} \times_1 \tilde{\mathbf{U}}^{\mathrm{T}} \times_2 \tilde{\mathbf{V}}^{\mathrm{T}} \times_3 \tilde{\mathbf{W}}^{\mathrm{T}}$$

7. CONCLUSION

We have performed a first-order perturbation analysis of the best rank-(R_1, R_2, R_3) approximation of third-order tensors. The generalization to orders higher than three is straightforward. Our results generalize the classical first-order perturbation analysis of the matrix SVD in a non-trivial way. Besides the fact that the perturbation expressions are more complicated, the tensor problem involves a number of phenomena that do not appear at the matrix level. For instance, the least squares cost function can have local optima, and indeterminacies can have a completely different structure than in the matrix case.

Acknowledgements

This work is supported by several institutions: (1) Research Council KU Leuven: Concerted Research Action GOA-MEFISTO-666; (2) Flemish Government: FWO project G.0240.99 and FWO Research Communities ICCoS and ANMMM; (3) Belgian Federal Government: Interuniversity Poles of Attraction Programmes IUAP IV-02 and IUAP V-22. The scientific responsibility is assumed by the author.

APPENDIX

In this appendix we will write Equations (29)–(31) and (22) in the standard form (4).

First we introduce some definitions and notation. With a matrix $\mathbf{A} \in \mathbb{R}^{I \times J}$ we associate the vector $\mathrm{vec}(\mathbf{A}) \in \mathbb{R}^{IJ}$ defined by

$$\mathrm{vec}(\mathbf{A}) = [a_{11}, \ldots, a_{I1}, a_{12}, \ldots, a_{I2}, \ldots, a_{1J}, \ldots, a_{IJ}]^{\mathrm{T}}$$

$\mathbf{P}_{I,J} \in \mathbb{R}^{IJ \times IJ}$ is a permutation matrix of which the non-zero entries are given by

$$\mathbf{P}_{I,J}((i-1)J + j, (j-1)I + i) = 1 \qquad 1 \le i \le I, \quad 1 \le j \le J$$

$\mathbf{K}_{n,I,N} \in \mathbb{R}^{NI \times I}$ represents the nth block column of $\mathbf{I}_{NI \times NI}$. We also define

$$\mathbf{L}_{I,J,N} = [(\mathbf{I}_{J \times J} \otimes \mathbf{K}_{1,I,N})^{\mathrm{T}}, \ldots, (\mathbf{I}_{J \times J} \otimes \mathbf{K}_{N,I,N})^{\mathrm{T}}]^{\mathrm{T}}$$

$$\mathbf{S}_{I,N} = [\mathrm{vec}(\mathbf{K}_{1,N,I}), \ldots, \mathrm{vec}(\mathbf{K}_{1,N,I})]$$

The derivation is based on the repeated application of the following computation rules.

- For matrices of compatible dimensions we have
$$(\mathbf{A} \cdot \mathbf{B}) \otimes (\mathbf{C} \cdot \mathbf{D}) = (\mathbf{A} \otimes \mathbf{C}) \cdot (\mathbf{B} \otimes \mathbf{D})$$
- We have
$$(\mathbf{A} \otimes \mathbf{B})^{\mathrm{T}} = \mathbf{A}^{\mathrm{T}} \otimes \mathbf{B}^{\mathrm{T}}$$
- For matrices of compatible dimensions we have
$$\mathrm{vec}(\mathbf{A} \cdot \mathbf{B} \cdot \mathbf{C}^{\mathrm{T}}) = (\mathbf{C} \otimes \mathbf{A}) \cdot \mathrm{vec}(\mathbf{B})$$
- For $\mathbf{A} \in \mathbb{R}^{I \times J}$ we have
$$\mathrm{vec}(\mathbf{I}_{N \times N} \otimes \mathbf{A}) = \mathbf{L}_{I,J,N} \cdot \mathrm{vec}(\mathbf{A})$$
- For $\mathbf{A} \in \mathbb{R}^{I \times J}$ we have
$$\mathrm{vec}(\mathbf{A} \otimes \mathbf{I}_{N \times N}) = \big(\mathbf{I}_{J \times J} \otimes \mathbf{S}_{I,N}\big) \cdot \mathrm{vec}(\mathbf{A})$$
- For $\mathbf{A} \in \mathbb{R}^{I \times J}$ we have
$$\mathrm{vec}(\mathbf{A}^{\mathrm{T}}) = \mathbf{P}_{I,J} \cdot \mathrm{vec}(\mathbf{A})$$

Copyright © 2004 John Wiley & Sons, Ltd.

We now vectorize the subsequent terms of (29):

$$\begin{aligned}&\operatorname{vec}\left(-\mathbf{X}\cdot\mathbf{U}^{\mathrm{T}}\cdot\mathbf{A}_{(1)}\cdot(\mathbf{V}\otimes\mathbf{W})\cdot\mathbf{C}_{1,2}\right)\\&\quad=-\left\{\left[\mathbf{C}_{1,2}^{\mathrm{T}}\cdot(\mathbf{V}\otimes\mathbf{W})^{\mathrm{T}}\cdot\mathbf{A}_{(1)}^{\mathrm{T}}\cdot\mathbf{U}\right]\right.\\&\qquad\left.\otimes\,\mathbf{I}_{(I_1-R_1)\times(I_1-R_1)}\right\}\cdot\operatorname{vec}(\mathbf{X})\\&\quad=\mathbf{M}_{1,1a}\cdot\operatorname{vec}(\mathbf{X})\end{aligned}\tag{32}$$

$$\operatorname{vec}\left(\overline{\mathbf{U}}^{\mathrm{T}}\cdot\mathbf{B}_{(1)}\cdot(\mathbf{V}\otimes\mathbf{W})\cdot\mathbf{C}_{1,2}\right)=-N_{1a}\tag{33}$$

$$\begin{aligned}&\operatorname{vec}\left(\overline{\mathbf{U}}^{\mathrm{T}}\cdot\mathbf{A}_{(1)}\cdot\left[(\overline{\mathbf{V}}\cdot\mathbf{Y})\otimes\mathbf{W}\right]\cdot\mathbf{C}_{1,2}\right)\\&\quad=\operatorname{vec}\left(\overline{\mathbf{U}}^{\mathrm{T}}\cdot\mathbf{A}_{(1)}\cdot(\overline{\mathbf{V}}\cdot\mathbf{W})\cdot(\mathbf{Y}\otimes\mathbf{I}_{R_3\times R_3})\cdot\mathbf{C}_{1,2}\right)\\&\quad=\left\{\mathbf{C}_{1,2}^{\mathrm{T}}\otimes\left[\overline{\mathbf{U}}^{\mathrm{T}}\cdot\mathbf{A}_{(1)}\cdot(\overline{\mathbf{V}}\otimes\mathbf{W})\right]\right\}\cdot\operatorname{vec}(\mathbf{Y}\otimes\mathbf{I}_{R_3\times R_3})\\&\quad=\left\{\mathbf{C}_{1,2}^{\mathrm{T}}\otimes\left[\overline{\mathbf{U}}^{\mathrm{T}}\cdot\mathbf{A}_{(1)}\cdot(\overline{\mathbf{V}}\otimes\mathbf{W})\right]\right\}\\&\qquad\cdot\left(\mathbf{I}_{R_2\times R_2}\otimes\mathbf{S}_{I_2-R_2,R_3}\right)\cdot\operatorname{vec}(\mathbf{Y})\\&\quad=\mathbf{M}_{1,2a}\cdot\operatorname{vec}(\mathbf{Y})\end{aligned}\tag{34}$$

$$\begin{aligned}&\operatorname{vec}\left(\overline{\mathbf{U}}^{\mathrm{T}}\cdot\mathbf{A}_{(1)}\cdot\left[\mathbf{V}\otimes(\overline{\mathbf{W}}\cdot\mathbf{Z})\right]\cdot\mathbf{C}_{1,2}\right)\\&\quad=\operatorname{vec}\left(\overline{\mathbf{U}}^{\mathrm{T}}\cdot\mathbf{A}_{(1)}\cdot(\mathbf{V}\otimes\overline{\mathbf{W}})\cdot(\mathbf{I}_{R_2\times R_2}\otimes\mathbf{Z})\cdot\mathbf{C}_{1,2}\right)\\&\quad=\left\{\mathbf{C}_{1,2}^{\mathrm{T}}\otimes\left[\overline{\mathbf{U}}^{\mathrm{T}}\cdot\mathbf{A}_{(1)}\cdot(\mathbf{V}\otimes\overline{\mathbf{W}})\right]\right\}\cdot\operatorname{vec}(\mathbf{I}_{R_2\times R_2}\otimes\mathbf{Z})\\&\quad=\left\{\mathbf{C}_{1,2}^{\mathrm{T}}\otimes\left[\overline{\mathbf{U}}^{\mathrm{T}}\cdot\mathbf{A}_{(1)}\cdot(\mathbf{V}\otimes\overline{\mathbf{W}})\right]\right\}\cdot\mathbf{L}_{I_3-R_3,R_3,R_2}\cdot\operatorname{vec}(\mathbf{Z})\\&\quad=\mathbf{M}_{1,3a}\cdot\operatorname{vec}(\mathbf{Z})\end{aligned}\tag{35}$$

$$\begin{aligned}&\operatorname{vec}\left(\mathbf{C}_{1,1}\cdot\left[(\overline{\mathbf{V}}\cdot\mathbf{Y})\otimes\mathbf{W}\right]^{\mathrm{T}}\cdot\mathbf{A}_{(1)}\cdot\mathbf{U}\right)\\&\quad=\operatorname{vec}\left(\mathbf{C}_{1,1}\cdot(\overline{\mathbf{V}}\otimes\mathbf{W})^{\mathrm{T}}\cdot(\mathbf{Y}^{\mathrm{T}}\otimes\mathbf{I}_{R_3\times R_3})\cdot\mathbf{A}_{(1)}\cdot\mathbf{U}\right)\\&\quad=\left\{\left(\mathbf{U}^{\mathrm{T}}\cdot\mathbf{A}_{(1)}^{\mathrm{T}}\right)\otimes\left[\mathbf{C}_{1,1}\cdot(\overline{\mathbf{V}}\otimes\mathbf{W})^{\mathrm{T}}\right]\right\}\cdot\operatorname{vec}(\mathbf{Y}^{\mathrm{T}}\otimes\mathbf{I}_{R_3\times R_3})\\&\quad=\left\{\left(\mathbf{U}^{\mathrm{T}}\cdot\mathbf{A}_{(1)}^{\mathrm{T}}\right)\otimes\left[\mathbf{C}_{1,1}\cdot(\overline{\mathbf{V}}\otimes\mathbf{W})^{\mathrm{T}}\right]\right\}\\&\qquad\cdot\left(\mathbf{I}_{(I_2-R_2)\times(I_2-R_2)}\otimes\mathbf{S}_{R_2,R_3}\right)\cdot\mathbf{P}_{I_2-R_2,R_2}\cdot\operatorname{vec}(\mathbf{Y})\\&\quad=\mathbf{M}_{1,2b}\cdot\operatorname{vec}(\mathbf{Y})\end{aligned}\tag{36}$$

$$\begin{aligned}&\operatorname{vec}\left(\mathbf{C}_{1,1}\cdot\left[\mathbf{V}\otimes(\overline{\mathbf{W}}\cdot\mathbf{Z})\right]^{\mathrm{T}}\cdot\mathbf{A}_{(1)}^{\mathrm{T}}\cdot\mathbf{U}\right)\\&\quad=\operatorname{vec}\left(\mathbf{C}_{1,1}\cdot(\mathbf{V}\otimes\overline{\mathbf{W}})^{\mathrm{T}}\cdot(\mathbf{I}_{R_2\times R_2}\otimes\mathbf{Z}^{\mathrm{T}})\cdot\mathbf{A}_{(1)}^{\mathrm{T}}\cdot\mathbf{U}\right)\\&\quad=\left\{(\mathbf{U}^{\mathrm{T}}\cdot\mathbf{A}_{(1)})\otimes\left[\mathbf{C}_{1,1}\cdot(\mathbf{V}\otimes\overline{\mathbf{W}})^{\mathrm{T}}\right]\right\}\cdot\operatorname{vec}(\mathbf{I}_{R_2\times R_2}\otimes\mathbf{Z}^{\mathrm{T}})\\&\quad=\left\{(\mathbf{U}^{\mathrm{T}}\cdot\mathbf{A}_{(1)})\otimes\left[\mathbf{C}_{1,1}\cdot(\mathbf{V}\otimes\overline{\mathbf{W}})^{\mathrm{T}}\right]\right\}\\&\qquad\cdot\mathbf{L}_{R_3,I_3-R_3,R_2}\cdot\mathbf{P}_{I_3-R_3,R_3}\cdot\operatorname{vec}(\mathbf{Z})\\&\quad=\mathbf{M}_{1,3b}\cdot\operatorname{vec}(\mathbf{Z})\end{aligned}\tag{37}$$

$$\begin{aligned}&\operatorname{vec}\left(\mathbf{C}_{1,1}\cdot(\mathbf{V}\otimes\mathbf{W})^{\mathrm{T}}\cdot\mathbf{A}_{(1)}^{\mathrm{T}}\cdot\overline{\mathbf{U}}\cdot\mathbf{X}\right)\\&\quad=\left\{\mathbf{I}_{R_1\times R_1}\otimes\left[\mathbf{C}_{1,1}\cdot(\mathbf{V}\otimes\mathbf{W})^{\mathrm{T}}\cdot\mathbf{A}_{(1)}^{\mathrm{T}}\cdot\overline{\mathbf{U}}\right]\right\}\cdot\operatorname{vec}(\mathbf{X})\\&\quad=\mathbf{M}_{1,1b}\cdot\operatorname{vec}(\mathbf{X})\end{aligned}\tag{38}$$

$$\operatorname{vec}\left(\mathbf{C}_{1,1}\cdot(\mathbf{V}\otimes\mathbf{W})^{\mathrm{T}}\cdot\mathbf{B}_{(1)}^{\mathrm{T}}\cdot\mathbf{U}\right)=-N_{1b}\tag{39}$$

By defining $\mathbf{M}_{1,1}=\mathbf{M}_{1,1a}+\mathbf{M}_{1,1b}$, $\mathbf{M}_{1,2}=\mathbf{M}_{1,2a}+\mathbf{M}_{1,2b}$, $\mathbf{M}_{1,3}=\mathbf{M}_{1,3a}+\mathbf{M}_{1,3b}$ and $N_1=N_{1a}+N_{1b}$, we obtain

$$\mathbf{M}_{1,1}\cdot\operatorname{vec}(\mathbf{X})+\mathbf{M}_{1,2}\cdot\operatorname{vec}(\mathbf{Y})+\mathbf{M}_{1,3}\cdot\operatorname{vec}(\mathbf{Z})=N_1$$

in which $\mathbf{M}_{1,1}=f_{1,1}(\mathbf{A}_{(1)},\mathbf{U},\mathbf{V},\mathbf{W})$, $\mathbf{M}_{1,2}=f_{1,2}(\mathbf{A}_{(1)},\mathbf{U},\mathbf{V},\mathbf{W})$, $\mathbf{M}_{1,3}=f_{1,3}(\mathbf{A}_{(1)},\mathbf{U},\mathbf{V},\mathbf{W})$ and $N_1=g(\mathbf{A}_{(1)},\mathbf{B}_{(1)},\mathbf{U},\mathbf{V},\mathbf{W})$.

By exploiting the symmetry in (29)–(31), we can easily write Equations (30) and (31) in a similar form. We define

$$\begin{aligned}\mathbf{M}_{2,1}&=f_{1,3}(\mathbf{A}_{(2)},\mathbf{V},\mathbf{W},\mathbf{U}) & \mathbf{M}_{3,1}&=f_{1,2}(\mathbf{A}_{(3)},\mathbf{W},\mathbf{U},\mathbf{V})\\\mathbf{M}_{2,2}&=f_{1,1}(\mathbf{A}_{(2)},\mathbf{V},\mathbf{W},\mathbf{U}) & \mathbf{M}_{3,2}&=f_{1,3}(\mathbf{A}_{(3)},\mathbf{W},\mathbf{U},\mathbf{V})\\\mathbf{M}_{2,3}&=f_{1,2}(\mathbf{A}_{(2)},\mathbf{V},\mathbf{W},\mathbf{U}) & \mathbf{M}_{3,3}&=f_{1,1}(\mathbf{A}_{(3)},\mathbf{W},\mathbf{U},\mathbf{V})\\N_2&=g(\mathbf{A}_{(2)},\mathbf{B}_{(2)},\mathbf{V},\mathbf{W},\mathbf{U}) & N_3&=g(\mathbf{A}_{(3)},\mathbf{B}_{(3)},\mathbf{W},\mathbf{U},\mathbf{V})\end{aligned}$$

Then we finally obtain

$$\begin{pmatrix}\mathbf{M}_{1,1}&\mathbf{M}_{1,2}&\mathbf{M}_{1,3}\\\mathbf{M}_{2,1}&\mathbf{M}_{2,2}&\mathbf{M}_{2,3}\\\mathbf{M}_{3,1}&\mathbf{M}_{3,2}&\mathbf{M}_{3,3}\end{pmatrix}\cdot\begin{pmatrix}\operatorname{vec}(\mathbf{X})\\\operatorname{vec}(\mathbf{Y})\\\operatorname{vec}(\mathbf{Z})\end{pmatrix}=\begin{pmatrix}N_1\\N_2\\N_3\end{pmatrix}\tag{40}$$

which is of the form of (4)

Equation (22) can be rewritten by just substituting $\mathbf{U}=\mathbf{V}=\mathbf{W}$ and $\mathbf{X}=\mathbf{Y}=\mathbf{Z}$ in (40). We obtain

$$(\mathbf{M}_{1,1}+\mathbf{M}_{1,2}+\mathbf{M}_{1,3})\cdot\operatorname{vec}(\mathbf{X})=N_1\tag{41}$$

REFERENCES

1. Tucker LR. Some mathematical notes on three-mode factor analysis. *Psychometrika* 1966; **31**: 279–311.
2. Kroonenberg PM, de Leeuw J. Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika* 1980; **45**: 69–97.
3. Kroonenberg PM. Three-mode principal component analysis: illustrated with an example from attachment theory. In *Research Methods for Multimode Data Analysis*, Law HG, Snyder CW, Hattie JA, McDonald RP (eds). Praeger: New York, 1984, 64–103.
4. Kroonenberg PM. *Three-mode Principal Component Analysis*. DSWO Press: Leiden, 1983.
5. Ten Berghe J, de Leeuw J, Kroonenberg PM. Some additional results on principal components analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika* 1987; **52**: 183–191.
6. De Lathauwer L, De Moor B, Vandewalle J. On the best rank-1 and rank-$(R_1, R_2, \ldots, R_N)$ approximation of higher-order tensors. *SIAM J. Matrix Anal. Appl.* 2000; **21**: 1324–1342.
7. Tucker LR. The extension of factor analysis to three-dimensional matrices. In *Contributions to Mathematical Psychology*, Gulliksen H, Frederiksen N (eds). Holt, Rinehart and Winston: New York, 1964; 109–127.
8. Bro R. PARAFAC. Tutorial and applications. *Chemometrics Intell. Lab. Syst.* 1997; **38**: 149–171.
9. De Lathauwer L. Signal processing based on multilinear algebra. *PhD Thesis*, KU Leuven, 1997.
10. De Lathauwer L, Vandewalle J. Dimensionality reduction in higher-order signal processing and rank-$(R_1, R_2, \ldots, R_N)$ reduction in multilinear algebra. *Linear Algebra Appl.* accepted.
11. Sidiropoulos N, Giannakis G, Bro R. Blind PARAFAC receivers for DS-CDMA systems. *IEEE Trans. Signal Process.* 2000; **48**: 810–823.
12. Dehaene J. Continuous-time matrix algorithms, systolic algorithms and adaptive neural networks. *PhD Thesis*, KU Leuven, 1995.
13. De Moor B, Boyd S. Analytic properties of singular values and vectors. *Tech. Rep. 89-28*, ESAT/SCD, KU Leuven, 1989.
14. Kato T. *A Short Introduction to Perturbation Theory for Linear Operators*. Springer: New York, 1982.
15. De Lathauwer L, De Moor B, Vandewalle J. A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl.* 2000; **21**: 1253–1278.

Copyright © 2004 John Wiley & Sons, Ltd.