PARALLEL FACTOR ANALYSIS BY MEANS OF SIMULTANEOUS MATRIX DECOMPOSITIONS

Lieven De Lathauwer

Lab. ETIS, UMR 8051
Cergy-Pontoise, France
delathau@ensea.fr

ABSTRACT

In this paper we consider simultaneous matrix decomposition approaches to Parallel Factor Analysis. We consider cases in which the rank is smaller than or equal to at least two, resp. at least one, tensor dimension(s).

1. INTRODUCTION

The Canonical Decomposition (CANDECOMP) or Parallel Factor model (PARAFAC) is a key concept in multilinear algebra and its applications. The PARAFAC of an $(I_1 \times I_2 \times I_3)$ tensor $\mathcal{A}$ is its decomposition in a linear combination of a minimal number of rank-1 terms:

$$\mathcal{A} = \sum_{r}^{R} \lambda_r \, U_r^{(1)} \circ U_r^{(2)} \circ U_r^{(3)}, \qquad (1)$$

in which R is called the rank of $\mathcal{A}$. Eq. (1) means that, entry-wise, we have

$$a_{i_1 i_2 i_3} = \sum_{r}^{R} \lambda_r \, u_{i_1 r}^{(1)} u_{i_2 r}^{(2)} u_{i_3 r}^{(3)}, \qquad \forall i_1, i_2, i_3.$$

The decomposition was independently introduced in Psychometrics [1] and Phonetics [2] around 1970. It is frequently used for the analysis of multi-way data in Chemometrics, food industry, etc. [3]. In wireless telecommunications, it provides powerful means for the exploitation of different types of diversity [4]. It also describes the basic structure of higher-order cumulants of multivariate data on which all algebraic methods for Independent Component Analysis (ICA) are based [5, 6].

We assume, without loss of generality, that all λ_r are equal to one. The standard way to compute the PARAFAC of a given tensor, is by direct minimization of the least-squares cost function associated with (1). It is possible to resort to an Alternating Least-Squares (ALS) algorithm, in which the vector estimates are updated mode per mode [1, 4, 3]. Define $\hat{\mathbf{U}}^{(n)} = [\hat{U}_1^{(n)} \ldots \hat{U}_R^{(n)}]$, $n = 1, 2, 3$. Now let us imagine that the matrices $\hat{\mathbf{U}}^{(m)}$, $m \neq n$, are fixed and that only $\hat{\mathbf{U}}^{(n)}$ is unknown. Because of the multilinearity of PARAFAC, the estimation of $\hat{\mathbf{U}}^{(n)}$ is a classical linear least squares problem. An ALS iteration consists of repeating this procedure for different mode numbers: in each step the estimate of one of the matrices $\mathbf{U}^{(1)}$, $\mathbf{U}^{(2)}$, $\mathbf{U}^{(3)}$ is optimized, while the other matrix estimates are kept constant. ALS iterations can be very slow. In addition, it is sometimes observed that the algorithm moves through a "swamp": the algorithm seems to converge, but then the convergence speed drastically decreases and remains small for several iteration steps, after which it may suddenly increase again. Recently, it has been understood that the multilinearity of PARAFAC allows for the determination of the optimal step size, which improves convergence [7]. Instead of the least-squares error, one can also minimize the least absolute error. To this end, an alternating linear programming algorithm as well as a weighted median filtering iteration are derived in [8].

In this paper we discuss new, simultaneous matrix decomposition based, computation schemes. We distinguish between two cases: (1) at least two dimensions are not smaller than the rank, and (2) at least one dimension is not smaller than the rank. These two cases from the subject of Sections 2 and 3, respectively. Section 4 shows the results of some numerical experiments. The results of Section 2 and 3 are described in more detail in [9] and [10], respectively.

2. CASE 1: $R \leqslant$ MIN(I_1, I_2)

In this section we assume that: (i) the set $\{U_r^{(1)}\}_{1 \leqslant r \leqslant R}$ is linearly independent, (ii) the set $\{U_r^{(2)}\}_{1 \leqslant r \leqslant R}$ is linearly independent, and (iii) that the set $\{U_r^{(3)}\}_{1 \leqslant r \leqslant R}$ does not contain collinear vectors. These conditions are sufficient to

The research reported in this paper is supported by several institutions: (1) the Flemish Government: (a) Research Council K.U.Leuven: GOA-Ambiorics, (b) F.W.O. project G.0240.99, (c) F.W.O. Research Communities ICCoS and ANMMM, (d) Tournesol 2005, (2) the Belgian Federal Science Policy Office: IUAP P5/22.

make PARAFAC essentially unique [11].

Let us denote the subsequent $(I_1 \times I_2)$ matrix slices of $\mathcal{A}$ by $\mathbf{A}_{i_3}$. Then Eq. (1) is equivalent to the following simultaneous matrix decomposition:

$$\begin{aligned} \mathbf{A}_1 &= \mathbf{U}^{(1)} \cdot \mathbf{D}_1 \cdot \mathbf{U}^{(2)^T} \\ &\vdots \\ \mathbf{A}_{I_3} &= \mathbf{U}^{(1)} \cdot \mathbf{D}_{I_3} \cdot \mathbf{U}^{(2)^T} \end{aligned} \tag{2}$$

in which $\mathbf{U}^{(1)} = [U_1^{(1)}, \ldots, U_R^{(1)}]$, $\mathbf{U}^{(2)} = [U_1^{(2)}, \ldots, U_R^{(2)}]$ and in which $\mathbf{D}_{i_3} = \mathrm{diag}(u_{i_3 1}^{(3)}, \ldots, u_{i_3 R}^{(3)})$. Hence, PARAFAC amounts to the computation of matrices $\mathbf{U}^{(1)}$ and $\mathbf{U}^{(2)}$ that simultaneously diagonalize the set $\mathbf{A}_1, \ldots, \mathbf{A}_{I_3}$ by equivalence. Along the lines of [12], one can substitute the QR-factorization $\mathbf{U}^{(1)} = \mathbf{Q}^T \mathbf{R}'$ and the RQ-decomposition $\mathbf{U}^{(2)^T} = \mathbf{R}'' \mathbf{Z}^T$ to obtain a Simultaneous Generalized Schur Decomposition (SGSD):

$$\begin{aligned} \mathbf{Q} \cdot \mathbf{A}_1 \cdot \mathbf{Z} &= \mathbf{R}_1 = \mathbf{R}' \cdot \mathbf{D}_1 \cdot \mathbf{R}'' \\ &\vdots \\ \mathbf{Q} \cdot \mathbf{A}_{I_3} \cdot \mathbf{Z} &= \mathbf{R}_K = \mathbf{R}' \cdot \mathbf{D}_{I_3} \cdot \mathbf{R}'' \end{aligned} \tag{3}$$

in which $\mathbf{Q}$ and $\mathbf{Z}$ are unitary and in which $\mathbf{R}'$,$\mathbf{R}''$,$\mathbf{R}_1$, $\mathbf{R}_2$, $\ldots$, $\mathbf{R}_{I_3}$ are upper triangular. The upper triangularization can be realized by means of an extended QZ-iteration [12] or, in the case of real data, by means of a Jacobi-type algorithm [9]. Note that, contrary to the ALS iteration, the unknowns are unitary, which may be numerically advantageous. The iteration can be initialized with the result obtained from two of the equations in (3), i.e., by means of a so-called "Generalized Schur Decomposition".

In a final step, the parallel factors are obtained from the simultaneous generalized Schur component matrices as follows. If $I_3 \geqslant R$, construct $\mathbf{B} = [(\mathbf{B})_{i_3,r}] = [(\mathbf{R}_{i_3})_{r,r}] \in \mathbb{C}^{I_3 \times R}$ and compute $\mathbf{C} = \mathbf{B}^\dagger$. Then $U_r^{(1)}$ and $U_r^{(2)}$ are respectively given by the dominant left and right singular vector of $\sum_{i_3=1}^{I_3} c_{r i_3} \mathbf{A}_{i_3}$ [12]. The vectors $U_r^{(3)}$ directly follow from (1), interpreted as a linear set of equations in the remaining unknowns. In [9] a procedure is described that also applies when $I_3 < R$. Contrary to what is stated in [9], error accumulation makes that the latter procedure is generally less reliable than the technique explained above when $I_3 \geqslant R$.

3. CASE 2: $R \leqslant I_3$

In this section we assume that $\mathbf{U}^{(3)}$ and the Khatri-Rao or column-wise Kronecker product $\mathbf{U}^{(1)} \odot \mathbf{U}^{(2)} \in \mathbb{C}^{I_1 I_2 \times R}$ are full column rank. The derivation is based on the reasoning in [13]. Consider a matrix $\mathbf{A} \in \mathbb{C}^{IJ \times K}$ in which the entries of $\mathcal{A}$ are stacked as follows:

$$(\mathbf{A})_{(i-1)J+j,k} = a_{ijk}.$$

We have

$$\mathbf{A} = (\mathbf{U}^{(1)} \odot \mathbf{U}^{(2)}) \cdot \mathbf{U}^{(3)^T}. \tag{4}$$

Note that the rank of the tensor $\mathcal{A}$ is equal to the rank of its matrix representation $\mathbf{A}$. Consider a factorization of $\mathbf{A}$ of the form

$$\mathbf{A} = \mathbf{E} \cdot \mathbf{F}^T, \tag{5}$$

with $\mathbf{E} \in \mathbb{C}^{IJ \times R}$ and $\mathbf{F} \in \mathbb{C}^{K \times R}$ full column rank. Because of (4) and (5), we have:

$$\mathbf{U}^{(1)} \odot \mathbf{U}^{(2)} = \mathbf{E} \cdot \mathbf{W} \tag{6}$$

for some nonsingular $\mathbf{W} \in \mathbb{C}^{R \times R}$. The task is now to find $\mathbf{W}$ such that the columns of $\mathbf{E} \cdot \mathbf{W}$ are Kronecker products. We will use the bilinear mapping introduced in the following theorem. For a proof we refer to [10].

Theorem 1 *Consider the mapping* $\Phi : (\mathbf{X}, \mathbf{Y}) \in \mathbb{C}^{I \times J} \times \mathbb{C}^{I \times J} \rightarrow \Phi(\mathbf{X}, \mathbf{Y}) \in \mathbb{C}^{I \times I \times J \times J}$ *defined by*

$$(\Phi(\mathbf{X}, \mathbf{Y}))_{ijkl} = x_{ik} y_{jl} + y_{ik} x_{jl} - x_{il} y_{jk} - y_{il} x_{jk}.$$

Then we have $\Phi(\mathbf{X}, \mathbf{X}) = 0$ *iff* $\mathbf{X}$ *is at most rank-1.*

Let $\mathcal{V}_{tu} = \Phi(U_t^{(1)} U_t^{(2)^T},\ U_u^{(1)} U_u^{(2)^T})$. Define matrices $\mathbf{E}_1, \ldots, \mathbf{E}_R \in \mathbb{R}^{I \times J}$ by

$$(\mathbf{E}_r)_{ij} = e_{(i-1)J+j,r} \qquad \forall i, j, r.$$

Let $\mathcal{P}_{rs} = \Phi(\mathbf{E}_r, \mathbf{E}_s)$. Since Φ is bilinear, we have from (6):

$$\mathcal{P}_{rs} = \sum_{t,u=1}^{R} (\mathbf{W}^{-1})_{tr} (\mathbf{W}^{-1})_{us} \mathcal{V}_{tu}. \tag{7}$$

Assume at this point that there exists a symmetric matrix $\mathbf{M}$ of which the entries satisfy the following set of homogeneous linear equations (we will justify this below):

$$\sum_{r,s=1}^{R} m_{rs} \mathcal{P}_{rs} = 0. \tag{8}$$

Substitution of (7) in (8) yields:

$$\sum_{r,s=1}^{R} \sum_{t,u=1}^{R} (\mathbf{W}^{-1})_{tr} (\mathbf{W}^{-1})_{us} m_{rs} \mathcal{V}_{tu} = 0.$$

From Theorem 1, we have $\mathcal{V}_{tt} = 0$, $1 \leqslant t \leqslant R$. Invoking the symmetry of Φ and $\mathbf{M}$ we obtain:

$$\sum_{r,s=1}^{R} \sum_{\substack{t,u=1 \\ t<u}}^{R} (\mathbf{W}^{-1})_{tr} (\mathbf{W}^{-1})_{us} m_{rs} \mathcal{V}_{tu} = 0. \tag{9}$$

We now make the crucial assumption that the tensors $\mathcal{V}_{tu}$, $1 \leqslant t < u \leqslant R$, are linearly independent. Then (9) implies:

$$\sum_{r,s=1}^{R} (\mathbf{W}^{-1})_{tr} (\mathbf{W}^{-1})_{us} m_{rs} = 0 \qquad \text{if} \quad t \neq u.$$

This can be written in a matrix format as

$$\mathbf{M} = \mathbf{W} \cdot \mathbf{\Lambda} \cdot \mathbf{W}^T,$$

in which $\mathbf{\Lambda}$ is diagonal. Actually, one can see that *any* diagonal matrix $\mathbf{\Lambda}$ generates a matrix $\mathbf{M}$ that satisfies Eq. (8). Hence, if the tensors $\{\mathcal{V}_{tu}\}_{t<u}$ are linearly independent, these matrices form an R-dimensional subspace of the symmetric $(R \times R)$ matrices. Let $\{\mathbf{M}_r\}$ represent a basis of this subspace. We have:

$$\begin{aligned} \mathbf{M}_1 &= \mathbf{W} \cdot \mathbf{\Lambda}_1 \cdot \mathbf{W}^T \\ &\vdots \\ \mathbf{M}_R &= \mathbf{W} \cdot \mathbf{\Lambda}_R \cdot \mathbf{W}^T \end{aligned} \tag{10}$$

in which $\mathbf{\Lambda}_1, \ldots, \mathbf{\Lambda}_R$ are diagonal. Eq. (10) is of the form (2), so we are back in Case 1. After computation of $\mathbf{W}$, $\mathbf{U}^{(1)}$ and $\mathbf{U}^{(2)}$ follow from (6). Finally, $\mathbf{U}^{(3)} = \mathbf{F} \cdot \mathbf{W}^{-T}$.

In [10] we prove that the set $\{\mathcal{V}_{tu}\}_{t<u}$ is linearly independent with probability one if the parallel factors can be considered drawn from continuous probability densities, provided $R(R-1) \leqslant I_1(I_1-1)I_2(I_2-1)/2$. This shows that linear independence of $\{\mathcal{V}_{tu}\}_{t<u}$ is a significantly more relaxed sufficient condition for PARAFAC uniqueness than Kruskal's condition [14] when applied to "Case 2" tensors. The condition was also obtained in [15], in a completely different manner.

4. NUMERICAL EXPERIMENTS

We generate tensors $\mathcal{A}$, of which the PARAFAC components will afterwards be estimated, in the following way:

$$\mathcal{A} = \tilde{\mathcal{A}}/\|\tilde{\mathcal{A}}\| + \sigma_N \tilde{\mathcal{N}}/\|\tilde{\mathcal{N}}\|, \tag{11}$$

in which $\tilde{\mathcal{A}}$ exactly satisfies the PARAFAC model.

With respect to Case 1, extensive simulations confirmed that ALS is reliable if the problem is well-conditioned. For some ill-conditioned problems however, the results are less satisfactory. We give an example. In each of 50 Monte Carlo runs, a tensor $\mathcal{A} \in \mathbb{R}^{2\times2\times10}$ of the form (11) is generated. The rank $R = 2$. $\mathbf{U}^{(2)}$ is a random matrix of which the singular values are replaced by $2, 1$. $\mathbf{U}^{(1)}$ is a random matrix of which the singular values are replaced by (i) 2,1, (ii) 10,1, (iii) 100,1. The entries of $\mathbf{U}^{(3)}$ are generated as $u_{ij}^{(3)} = 1 + g_{ij}/50$, in which g_{ij} is drawn from a Gaussian distribution with unit variance. In Fig. 1 we compare the mean value of $\|\hat{\mathbf{U}}^{(1)} - \mathbf{U}^{(1)}\|$ obtained by ALS initialized by means of (i) an EVD, as in [11], (ii) the SGSD. In Fig. 2 we compare the ALS-enhanced SGSD result to the best result obtained by ALS iteration, starting from 10 random initializations. An SGSD-initialized ALS iteration took about 50 s of CPU time, a randomly initialized ALS iteration about 100 s (several initializations were needed to find the optimum), and the SGSD itself about 1 s. Only when $\mathbf{U}^{(1)}$ was ill-conditioned, ALS was able to further improve the result obtained by means of the SGSD. We also mention that the extended QZ-iteration almost never suffers from swamp-type convergence.

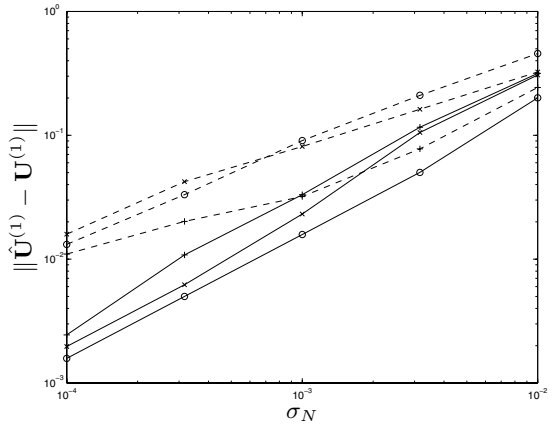


Fig. 1. Absolute error in the first experiment. SGSD followed by ALS iteration (solid) and EVD followed by ALS iteration (dashed). The condition number κ of $\mathbf{U}^{(1)}$ is equal to 2 ($\circ$), 10 ($\times$) or 100 ($+$).

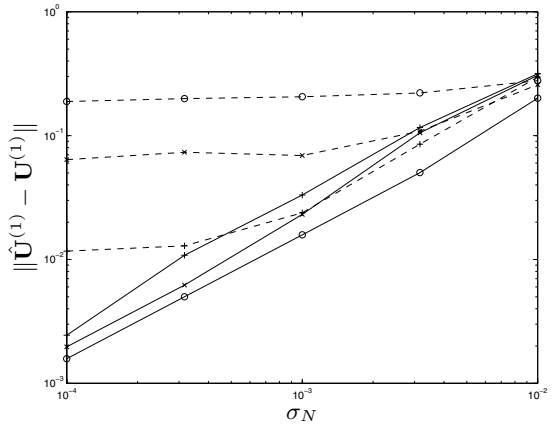


Fig. 2. Absolute error in the first experiment. ALS-enhanced SGSD (solid) and direct ALS started from 10 random initializations (dashed). The condition number κ of $\mathbf{U}^{(1)}$ is equal to 2 ($\circ$), 10 ($\times$) or 100 ($+$).

In a second experiment, we consider $\tilde{\mathcal{A}} \in \mathbb{R}^{3\times4\times12}$. The entries of $\mathbf{U}^{(1)}$, $\mathbf{U}^{(2)}$ and $\mathbf{U}^{(3)}$ are drawn from a zero-mean unit-variance Gaussian distribution. The rank $R = 6$. In this case, Kruskal's condition does not guarantee uniqueness. The results are plotted in Fig. 3. Clearly, ALS did not find the solution. Moreover, the computational cost of the best ALS iteration (out of 10), was typically three orders of magnitude higher than that of the simultaneous matrix diag-

onalization. The latter was realized by means of an extended QZ-iteration.

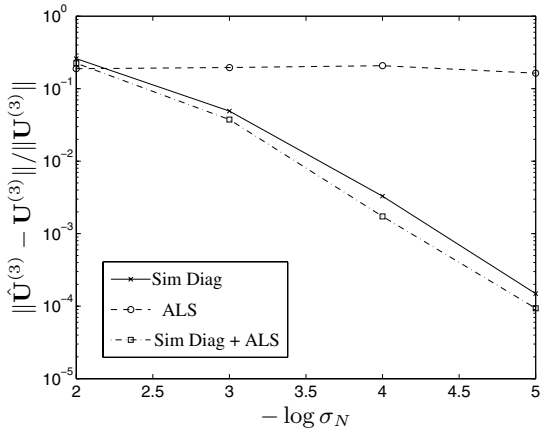


Fig. 3. Relative error in the second experiment.

5. CONCLUSION

In Case 1, ALS may not be reliable when the problem is ill-conditioned. PARAFAC can be reformulated as a simultaneous matrix decomposition. This may lead to new numerical linear algebra based algorithms. In particular, we discussed a link with the SGSD. We observed that the extended QZ-iteration for the computation of the SGSD almost never suffers from swamps. Case 2 may be reformulated in terms of Case 1. The noise-free solution can thus be written in terms of a standard matrix decomposition. We presented a new bound on the number of rank-1 terms that can be admitted while preserving uniqueness. For the tensors under consideration, this bound is more relaxed than Kruskal's.

6. REFERENCES

[1] J. Carroll and J. Chang, "Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition," *Psychometrika*, vol. 9, pp. 267–283, 1970.

[2] R. Harshman, "Foundations of the PARAFAC procedure: Model and conditions for an "explanatory" multi-mode factor analysis," *UCLA Working Papers in Phonetics*, vol. 16, pp. 1–84, 1970.

[3] A. Smilde, R. Bro, and P. Geladi, *Multi-way Analysis. Applications in the Chemical Sciences*, John Wiley and Sons, Chichester, U.K., 2004.

[4] N.D. Sidiropoulos, G.B. Giannakis, and R. Bro, "Blind PARAFAC receivers for DS-CDMA systems," *IEEE Trans. on Signal Processing*, vol. 48, pp. 810–823, March 2000.

[5] P. Comon, "Independent component analysis, a new concept?," *Signal Process.*, vol. 36, pp. 287–314, 1994.

[6] L. De Lathauwer, B. De Moor, and J. Vandewalle, "An introduction to independent component analysis," *J. Chemometrics*, vol. 14, pp. 123–149, 2000.

[7] M. Rajih and P. Comon, "Enhanced line search: A novel method to accelerate parafac," in *Proc. Eusipco'05*, Antalya, Turkey, Sept. 2005.

[8] S.A. Vorobyov, Y. Rong, N.D. Sidiropoulos, and A.B. Gershman, "Robust iterative fitting of multilinear models," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 2678–2689, August 2005.

[9] L. De Lathauwer, B. De Moor, and J. Vandewalle, "Computation of the canonical decomposition by means of a simultaneous generalized schur decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 26, no. 2, pp. 295–327, 2004.

[10] L. De Lathauwer, "A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization," Tech. Rep., ETIS, CNRS UMR 8051, Cergy-Pontoise, France, 2005, submitted.

[11] S.E. Leurgans, R.T. Ross, and R.B. Abel, "A decomposition for three-way arrays," *SIAM J. Matrix Anal. Appl.*, vol. 14, pp. 1064–1083, 1993.

[12] A.-J. van der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Trans. Signal Processing*, vol. 44, no. 5, pp. 1136–1155, May 1996.

[13] J.-F. Cardoso, "Super-symmetric decomposition of the fourth-order cumulant tensor. Blind identification of more sources than sensors," in *Proc. ICASSP '91*, Toronto, Canada, 1991, pp. 3109–3112.

[14] J.B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Lin. Alg. Appl.*, vol. 18, pp. 95–138, 1977.

[15] T. Jiang and N.D. Sidiropoulos, "Kruskal's permutation lemma and the identification of CANDECOMP/PARAFAC and bilinear models with constant modulus constraints," *IEEE Trans. on Signal Processing*, vol. 52, pp. 2625–2636, Sept. 2004.