Canonical Polyadic Decomposition of Third-Order Tensors of Which a Mode Rank is Equal to the Rank*

Lieven De Lathauwer

December 1, 2011

1 Introduction

We consider the Canonical Polyadic Decomposition (CPD) of a rank-R tensor $\mathcal{T} \in \mathbb{K}^{I_1 \times I_2 \times I_3}$, with $R = \text{rank}_3(\mathcal{T}) \leqslant I_3$:

$$\mathcal{T} = \sum_{r=1}^{R} \mathbf{a}_r \otimes \mathbf{b}_r \otimes \mathbf{c}_r. \tag{1}$$

Contributions:

- A new, relaxed, algebraic condition on $\mathbf{a}_r$, $\mathbf{b}_r$, $1 \leqslant r \leqslant R$ that is sufficient for essential uniqueness of the CPD.
- Under this condition the CP factors can be obtained by means of standard linear algebra (set of linear equations and matrix Eigenvalue Decomposition (EVD)).

*Available in SISTA publication engine as: L. De Lathauwer, *CPD Study, Part III*, Tech. Report 11-221, ESAT-SISTA, K.U.Leuven (Leuven, Belgium), 2011.

- Algorithm for the estimation of the CP factors in the case of an approximate CPD.
- Generic uniqueness for a given set (I_1, I_2, R) is established by checking the algebraic condition for one random sample.
- Verification for $(3 \times 7 \times 12)$ tensors of rank 12.
- Similar approach for $\mathbf{a}_r = \mathbf{b}_r$, $1 \leqslant r \leqslant R$. (Generic bound will be different.)

Preceding work. In matrix format we have:

$$\mathbf{T}_{[1,2;3]} = (\mathbf{A} \odot \mathbf{B}) \cdot \mathbf{C}^T. \tag{2}$$

A necessary condition for uniqueness of the CPD is that $\mathbf{A} \odot \mathbf{B}$ has full column rank. Our mode-3 rank assumption implies that $\text{rank}(\mathbf{C}) = R$. In [1] we showed that (1) can be obtained by means of standard linear algebra (set of linear equations and matrix Eigenvalue Decomposition (EVD)) if the tensors in the set $\mathsf{U} = \{\Phi(\mathbf{a}_{s_1}\mathbf{b}_{s_1}^T, \mathbf{a}_{s_2}\mathbf{b}_{s_2}^T) | 1 \leqslant s_1 < s_2 \leqslant R\}$ are linearly independent. In the case of an approximate decomposition, after solving a set of linear equations in least-squares sense, the computation of (1) can be reduced to the computation of a CPD of the form

$$\mathcal{M} = \sum_{r=1}^{R} \mathbf{w}_r \otimes \mathbf{w}_r \otimes \mathbf{l}_r, \tag{3}$$

in which $\mathcal{M} \in \mathbb{K}^{R \times R \times R}$ and in which $\mathbf{W} \in \mathbb{K}^{R \times R}$ and $\mathbf{L} \in \mathbb{K}^{R \times R}$ are nonsingular. Taking into account symmetries [2], the tensors in the set U can be represented by the columns of a matrix $\mathbf{U} \in \mathbb{K}^{C_{I_1}^2 C_{I_2}^2 \times C_R^2}$, which needs to have full rank. Consequently, in this approach R is bounded by $C_R^2 \leqslant C_{I_1}^2 C_{I_2}^2$. In the next section we propose a technique that is less restrictive.

2 Symmetrized mapping

Let $\mathbf{F} \in \mathbb{K}^{I_1 I_2 \times R}$ and $\mathbf{G} \in \mathbb{K}^{I_3 \times R}$ be such that

$$\mathbf{T}_{[1,2;3]} = \mathbf{F} \cdot \mathbf{G}^T \tag{4}$$

is a rank-revealing decomposition of $\mathbf{T}_{[1,2;3]}$. Then we have:

$$\begin{aligned} \mathbf{F} &= (\mathbf{A} \odot \mathbf{B}) \cdot \mathbf{W}^{-1}, & (5) \\ \mathbf{G}^T &= \mathbf{W} \cdot \mathbf{C}^T, & (6) \end{aligned}$$

in which $\mathbf{W} \in \mathbb{K}^{R \times R}$ is nonsingular.

Theorem 1 *Consider the mapping* $\Phi^{(3)} : (\mathbf{X}^{(r_1)}, \mathbf{X}^{(r_2)}, \mathbf{X}^{(r_3)}) \in \mathbb{K}^{I \times J} \times \mathbb{K}^{I \times J} \times \mathbb{K}^{I \times J} \rightarrow \Phi^{(3)}(\mathbf{X}^{(r_1)}, \mathbf{X}^{(r_2)}, \mathbf{X}^{(r_3)}) \in \mathbb{K}^{I \times I \times I \times J \times J \times J}$ *defined by*

$$\begin{gathered} (\Phi^{(3)}(\mathbf{X}^{(r_1)}, \mathbf{X}^{(r_2)}, \mathbf{X}^{(r_3)}))_{i_1 i_2 i_3 j_1 j_2 j_3} = \\ \textstyle\sum_{\sigma \in \mathsf{P}(r_1, r_2, r_3)} (x_{i_1 j_1}^{(\sigma(r_1))} x_{i_2 j_2}^{(\sigma(r_2))} - x_{i_1 j_2}^{(\sigma(r_1))} x_{i_2 j_1}^{(\sigma(r_2))}) x_{i_3 j_3}^{(\sigma(r_3))}, \end{gathered}$$

in which $\mathsf{P}(r_1, r_2, r_3)$ *is the set of permutations of* (r_1, r_2, r_3)*. Then* $\Phi^{(3)}(\mathbf{X}, \mathbf{X}, \mathbf{X}) = \mathcal{O}$ *if and only if* $\mathbf{X}$ *has at most rank 1.*

Note that $\Phi^{(3)}(\cdot, \cdot, \cdot)$ is symmetric in its arguments. Define

$$\mathcal{P}_{r_1 r_2 r_3} = \Phi^{(3)}((\mathbf{f}_{r_1})_{[1;2]}, (\mathbf{f}_{r_2})_{[1;2]}, (\mathbf{f}_{r_3})_{[1;2]}), \quad 1 \leqslant r_1 \leqslant r_2 \leqslant r_3 \leqslant R. \tag{7}$$

We have that

$$\mathcal{P}_{r_1 r_2 r_3} = \sum_{s_1, s_2, s_3 = 1}^{R} (\mathbf{W}^{-1})_{s_1 r_1} (\mathbf{W}^{-1})_{s_2 r_2} (\mathbf{W}^{-1})_{s_3 r_3} \Phi^{(3)}(\mathbf{a}_{s_1} \mathbf{b}_{s_1}^T, \mathbf{a}_{s_2} \mathbf{b}_{s_2}^T, \mathbf{a}_{s_3} \mathbf{b}_{s_3}^T). \tag{8}$$

We look for symmetric tensors $\mathcal{X} \in \mathbb{K}^{R \times R \times R}$ that satisfy

$$\sum_{r_1, r_2, r_3 = 1}^{R} x_{r_1 r_2 r_3} \mathcal{P}_{r_1 r_2 r_3} = \mathcal{O}. \tag{9}$$

We stack $\mathcal{P}_{r_1 r_2 r_3}$, $1 \leqslant r_1 \leqslant r_2 \leqslant r_3 \leqslant R$, in a matrix $\bar{\mathbf{P}} \in \mathbb{K}^{I_1^3 I_2^3 \times C_{R+2}^3}$. We determine the vectors $\mathbf{x}$ in the null space of $\bar{\mathbf{P}}$, which can then be mapped to tensors $\mathcal{X}$. We make the crucial assumption that the set of tensors $\{\Phi^{(3)}(\mathbf{a}_{s_1} \mathbf{b}_{s_1}^T, \mathbf{a}_{s_2} \mathbf{b}_{s_2}^T, \mathbf{a}_{s_3} \mathbf{b}_{s_3}^T) | 1 \leqslant s_1 < s_2 \leqslant s_3 \leqslant R\}$ is linearly independent. Combination of (7)–(9) shows that, under this assumption, the null space of $\bar{\mathbf{P}}$ has dimension R. Moreover, the vectors in the null space represent

tensors that are linear combinations of $\mathbf{w}_r \otimes \mathbf{w}_r \otimes \mathbf{w}_r$, $1 \leqslant r \leqslant R$. If we stack the tensors obtained from a basis of the null space of $\bar{\mathbf{P}}$ in a super-tensor $\mathcal{Y} \in \mathbb{K}^{R \times R \times R \times R}$, then this super-tensor admits the CPD

$$\mathcal{Y} = \sum_{r=1}^{R} \mathbf{w}_r \otimes \mathbf{w}_r \otimes \mathbf{w}_r \otimes \mathbf{v}_r. \tag{10}$$

The vectors $\mathbf{v}_r$, $1 \leqslant r \leqslant R$, are linearly independent since they generate a basis of the null space of $\bar{\mathbf{P}}$. Since additionally $\mathbf{W}$ is nonsingular, CPD (10) is unique. In the case of an exact decomposition, (10) follows from a matrix EVD [3, 4]. In the case of an approximate decomposition, the EVD may be used to initialize an algorithm for fitting (10).

Remark 1. Instead of working with $\bar{\mathbf{P}} \in \mathbb{K}^{I_1^3 I_2^3 \times C_{R+2}^3}$, one may also estimate the null space of $\bar{\mathbf{P}}^H \cdot \bar{\mathbf{P}} \in \mathbb{K}^{C_{R+2}^3 \times C_{R+2}^3}$. The latter matrix can be computed efficiently by computing $(\mathbf{f}_{r_1} \otimes \mathbf{f}_{r_2} \otimes \mathbf{f}_{r_3})^H (\mathbf{f}_{s_1} \otimes \mathbf{f}_{s_2} \otimes \mathbf{f}_{s_3})$ as $(\mathbf{f}_{r_1}^H \mathbf{f}_{s_1})(\mathbf{f}_{r_2}^H \mathbf{f}_{s_2})(\mathbf{f}_{r_3}^H \mathbf{f}_{s_3})$. The disavantage is that the condition number is squared.

Example 1. Consider tensors of dimension $(3 \times 7 \times R)$ and rank R. The number of free parameters in the CPD equals $(3{+}7{+}R{-}2)R = (8{+}R)R$. The number of tensor entries is $21R$. Consequently, for $R = 13$ there generically exist a finite number of decompositions. Moreover, we expect that for $R \leqslant 12$ the CPD is generically unique. For $R = 12$ uniqueness is not guaranteed by the result in [1]. However, it can be verified that the 352 tensors in the set $\{\Phi^{(3)}(\mathbf{a}_{s_1}\mathbf{b}_{s_1}^T, \mathbf{a}_{s_2}\mathbf{b}_{s_2}^T, \mathbf{a}_{s_3}\mathbf{b}_{s_3}^T) | 1 \leqslant s_1 < s_2 \leqslant s_3 \leqslant 12\}$ are linearly independent for a random example. Hence, for $R = 12$ the decomposition is generically unique and it can be computed as explained above.

Remark 2. It is possible to generalize the trilinear mapping $\Phi^{(3)}$ to a K-linear mapping $\Phi^{(K)}$ that is supposed to yield a matrix $\bar{\mathbf{P}} \in \mathbb{K}^{I_1^K I_2^K \times C_{R+K-1}^K}$ that has an R-dimensional null space.

(The trick here is that by increasing K we work with a matrix of which the number of rows increases faster than the number of columns. Since the columns have a multiplicative rather than a linear structure, this will hopefully lead to more relaxed conditions and algorithms.)

3 Alternative mapping

Theorem 2 *Consider the mapping* $\Psi^{(3)} : (\mathbf{X}, \mathbf{Y}, \mathbf{Z}) \in \mathbb{K}^{I\times J} \times \mathbb{K}^{I\times J} \times \mathbb{K}^{I\times J} \rightarrow \Psi^{(3)}(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) \in \mathbb{K}^{I\times I\times J\times J\times 6}$ *defined by*

$$
\begin{aligned}
(\Psi^{(3)}(\mathbf{X}, \mathbf{Y}, \mathbf{Z}))_{i_1 i_2 j_1 j_2 1} &= x_{i_1 j_1} y_{i_2 j_2} - x_{i_1 j_2} y_{i_2 j_1}, \\
(\Psi^{(3)}(\mathbf{X}, \mathbf{Y}, \mathbf{Z}))_{i_1 i_2 j_1 j_2 2} &= y_{i_1 j_1} x_{i_2 j_2} - y_{i_1 j_2} x_{i_2 j_1}, \\
(\Psi^{(3)}(\mathbf{X}, \mathbf{Y}, \mathbf{Z}))_{i_1 i_2 j_1 j_2 3} &= x_{i_1 j_1} z_{i_2 j_2} - x_{i_1 j_2} z_{i_2 j_1}, \\
(\Psi^{(3)}(\mathbf{X}, \mathbf{Y}, \mathbf{Z}))_{i_1 i_2 j_1 j_2 4} &= z_{i_1 j_1} x_{i_2 j_2} - z_{i_1 j_2} x_{i_2 j_1}, \\
(\Psi^{(3)}(\mathbf{X}, \mathbf{Y}, \mathbf{Z}))_{i_1 i_2 j_1 j_2 5} &= y_{i_1 j_1} z_{i_2 j_2} - y_{i_1 j_2} z_{i_2 j_1}, \\
(\Psi^{(3)}(\mathbf{X}, \mathbf{Y}, \mathbf{Z}))_{i_1 i_2 j_1 j_2 6} &= z_{i_1 j_1} y_{i_2 j_2} - z_{i_1 j_2} y_{i_2 j_1}.
\end{aligned}
$$

Then $\Psi^{(3)}(\mathbf{X}, \mathbf{X}, \mathbf{X}) = \mathcal{O}$ *if and only if* $\mathbf{X}$ *has at most rank 1.*

Define

$$
\mathcal{P}_{r_1 r_2 r_3} = \Psi^{(3)}((\mathbf{f}_{r_1})_{[1;2]}, (\mathbf{f}_{r_2})_{[1;2]}, (\mathbf{f}_{r_3})_{[1;2]}), \quad 1 \leqslant r_1 \leqslant r_2 \leqslant r_3 \leqslant R. \tag{11}
$$

We have that

$$
\mathcal{P}_{r_1 r_2 r_3} = \sum_{s_1, s_2, s_3 = 1}^{R} (\mathbf{W}^{-1})_{s_1 r_1} (\mathbf{W}^{-1})_{s_2 r_2} (\mathbf{W}^{-1})_{s_3 r_3} \Psi^{(3)}(\mathbf{a}_{s_1} \mathbf{b}_{s_1}^T, \mathbf{a}_{s_2} \mathbf{b}_{s_2}^T, \mathbf{a}_{s_3} \mathbf{b}_{s_3}^T). \tag{12}
$$

We look for tensors $\mathcal{X} \in \mathbb{K}^{R\times R\times R}$ that satisfy

$$
\sum_{r_1, r_2, r_3 = 1}^{R} x_{r_1 r_2 r_3} \mathcal{P}_{r_1 r_2 r_3} = \mathcal{O}. \tag{13}
$$

We stack $\mathcal{P}_{r_1 r_2 r_3}$, $1 \leqslant r_1, r_2, r_3 \leqslant R$, in a matrix $\bar{\mathbf{P}} \in \mathbb{K}^{6 I_1^2 I_2^2 \times R^3}$. We determine the vectors $\mathbf{x}$ in the null space of $\bar{\mathbf{P}}$, which can then be mapped to tensors $\mathcal{X}$. We make the crucial assumption that the set of tensors $\{\Psi^{(3)}(\mathbf{a}_{s_1} \mathbf{b}_{s_1}^T, \mathbf{a}_{s_2} \mathbf{b}_{s_2}^T, \mathbf{a}_{s_3} \mathbf{b}_{s_3}^T) | 1 \leqslant s_1, s_2, s_3 \leqslant R, \neg(s_1 = s_2 = s_3)\}$ is linearly independent. Combination of (11)–(13) shows that, under this assumption, the null space of $\bar{\mathbf{P}}$ has dimension R. Moreover, the vectors in the null space represent tensors that are linear combinations of $\mathbf{w}_r \otimes \mathbf{w}_r \otimes \mathbf{w}_r$, $1 \leqslant r \leqslant R$.

If we stack the tensors obtained from a basis of the null space of $\bar{\mathbf{P}}$ in a super-tensor $\mathcal{Y} \in \mathbb{K}^{R \times R \times R \times R}$, then this super-tensor admits the CPD

$$\mathcal{Y} = \sum_{r=1}^{R} \mathbf{w}_r \otimes \mathbf{w}_r \otimes \mathbf{w}_r \otimes \mathbf{v}_r. \tag{14}$$

The vectors $\mathbf{v}_r$, $1 \leqslant r \leqslant R$, are linearly independent since they generate a basis of the null space of $\bar{\mathbf{P}}$. Since additionally $\mathbf{W}$ is nonsingular, CPD (14) is unique. In the case of an exact decomposition, (14) follows from a matrix EVD [3, 4]. In the case of an approximate decomposition, the EVD may be used to initialize an algorithm for fitting (14).

Remark 3. Since all tensors $\mathcal{X}$ are symmetric, we might as well have worked with a matrix $\bar{\bar{\mathbf{P}}} \in \mathbb{K}^{6I_1^2 I_2^2 \times C_{R+2}^3}$ in which the columns corresponding to the same entry are added together. In other words, we might have symmetrized $\Psi^{(3)}$.

Remark 4. It seems that the mapping $\Psi^{(3)}$ does not yield a uniqueness condition that is more relaxed than the one in [1]. However, $\Psi^{(3)}$ might perhaps be used as a starting point for further research.

References

[1] L. De Lathauwer, "A Link between the Canonical Decomposition in Multilinear Algebra and Simultaneous Matrix Diagonalization", *SIAM J. Matrix Anal. Appl.*, Vol. 28, No. 3, 2006, pp. 642–666.

[2] A link between the decomposition of a third-order tensor in rank-(L,L,1) terms and simultaneous matrix diagonalization, Tech. report.

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

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