Réponses au relecteur commun des papiers :

1. Uniqueness of Tensor Train Decomposition with Linear Dependencies
2. Formulation du modèle CP d'ordre élevé en un train de tenseurs (TT) d'ordre réduit

Par (ordre alphabétique) Rémy Boyer, David Brie, André L.F. De Almeida, Gérard Favier et Sebastian Miron

---

Papier n°1 : Uniqueness of Tensor Train Decomposition with Linear Dependencies

----
Commentaires généraux

L'article propose une étude du lien entre l'ensemble des décompositions canonique polyadiques d'un tenseur et l'ensemble des décompositions en Train de Tenseurs (TT) de ce même tenseur. En particulier, cet article se concentre sur le cas où un des facteurs de la décomposition CPD a un rang colonne déficient.

Réponse : L'article aborde un cas plus général que celui évoqué par le relecteur. En effet, la totalité des facteurs peuvent admettre des colinéarités.

D'une part, la question de l'unicité abordée par l'article est mal introduite. En effet les auteurs ne prouvent pas l'unicité de la solution d'un certain problème, mais plutôt une condition de reconstruction exacte de paramètres (les facteurs du modèle PARALIND) à partir d'une représentation compressée (le format TT). Les conditions de l'unicité de la factorisation PARALIND ne sont d'ailleurs pas abordée ; on comprend donc mal comment une reconstruction pourrait être exacte si les paramètres à estimer ne sont pas eux-mêmes uniques.

Réponse : contrairement à ce qu'annoncer par le relecteur, format TT n'est pas une version compressée de PARALIND mais une relation algébrique exacte. Pour un tenseur donné, sa représentation en format TT est UNIQUE (sous les hypothèses de rang plein données) car construite à partir de la SVD des dépliement séquentiels du tenseur. En conséquence, la remarque du relecteur nous est difficilement compréhensible. L'intérêt de ce travail est de montrer qu'on peut sous les conditions du th2 garantir l'identifiabilité partielle et sous celle du th3 l'identifiabilité des facteurs de la CPD à partir de la décomposition TT. D'ailleurs comme mentionné dans la conclusion, les conditions du th 3 sont plus restrictives que la condition de type Kruskal.

D'autre part, la rigueur mathématique de l'article ne me convainc pas :

Réponse : Merci de préciser cette remarque. Il s'agit de fournir une relecture étayée et précise.

- la définition du TT n'est pas standard (il n'est jamais question du rang des cœurs dans l'article~\cite{Oseledets2011Tensor} qui pose les bases du modèle TT).

Réponse : Cette remarque est vraie mais elle reflète à notre avis le manque d'ouverture d'esprit du relecteur. Le papier ~\cite{Oseledets2011Tensor} est en effet le papier fondateur des modèles TT dans la communauté math-appli. Cet article ne contraint pas les rangs multilinéaires des cœurs. A l'inverse, dans notre travail et notamment dans le papier EUSIPCO 2018, nous abordons les modèles

TT sous l'angle algorithmique. Notamment nous montrons en nous basant sur une preuve constructive qu'un CPD est équivalent à un TT faisant intervenir des matrices latentes de couplage et les facteurs. Du fait de la nature constructive de notre preuve basée sur un enchainement séquentiel de « reshaping » et de SVD tronquées, les cœurs sont par construction à rang multilinéaire minimal. Une autre différence notable avec l'article ~\cite{Oseledets2011Tensor} est que nous proposons un schéma d'optimisation exploitant le couplage entre les matrices latentes et que nous montrons que les indéterminés (ambiguïté d'échelle et de permutation des colonnes) sont identiques au modèle CPD. Ce résultat est fondamental dans le contexte du design d'un estimateur. L'ensemble de ces résultats formant le cœur de la contribution de l'article semblent avoir complètement échappés au relecteur.

- l'argument ``d'identification'' des termes d'un TT ne parait pas évident. Il dépend peut-être de l'hypothèse de rang plein des dépliements des cœurs, mais en tout cas ce résultat n'est à ma connaissance pas mentionné dans la littérature et mérite qu'on s'y attarde.

Réponse : Absolument, d'ailleurs c'est pour cela que nous nous intéressons à ce problème.

- L'argument `` One may note that ... which justifies that matrices P_1 and P_Q must be full columns rank'' est également délicat (preuve du théorème 1).

- De même, je ne comprends pas l'argument ``... implies that the square matrices U_q in theorem 1 are full column rank'' (preuve du théorème 2).

Tout cela est peut-être correct, mais dans ce cas cela nécessite un travail de rédaction un peu plus précis.

Réponse : Encore une fois, le relecteur fait preuve de légèreté et d'imprécision dans son analyse. Que signifie « délicat » ??? Que ne comprend-t-il dans la seconde phrase ? En l'absence d'une formulation précise, ces remarques sont complètement inexploitables. Une relecture d'article doit être précise et argumentée.

Enfin, des résultats récents de Phan et coauteurs~\cite{phan2018tensor} donnent un résultat similaire à ce qui est proposé ici, sans être référencé dans cet article. A l'inverse de TT-PARALIND, leur résultat ne se préoccupe pas de la structure de corrélation des facteurs mais repose sur l'unicité de la CPD. Comme discuté plus haut, une reconstruction exacte n'a pas vraiment de sens si les paramètres à reconstruire ne sont pas uniques, donc cette hypothèse semble très raisonnable. Ainsi le travail présenté ici n'a que d'intérêt que s'il arrive à se situer par rapport à~\cite{phan2018tensor} qui semble plus général ou bien s'il arrive à mettre en défaut leur résultat.

Réponse : L'article ~\cite{phan2018tensor} a été posté sur arXiv et publié à notre connaissance nul part. Celui-ci n'a pas suivi le parcourt d'évaluation standard d'une revue (ou d'une conférence…). Le relecteur ne peut pas à lui seul se substituer au travail d'analyse de plusieurs relecteurs et d'un AE et décider que cet article est digne d'intérêt et juste... même s'il estime les travaux des auteurs. Cela décrédibilise son argumentaire et le rend difficilement audible. A noter que notre papier publié à EUSIPCO de manière antérieur à ~\cite{phan2018tensor} n'est pas cité. De plus, l'article ~\cite{phan2018tensor} propose des algorithmes n'exploitant pas les propriétés de couplage entre cœurs TT. Dans le Lemme 2 de ~\cite{phan2018tensor}, la déficience de rang est envisagée au travers du modèle PARALIND mais aucune analyse d'identifiabilité n'est fournie. En conclusion, l'analyse du relecteur nous semble superficielle, approximative et souffre du défaut de baser son argumentaire sur un article non publié.

A noter que d'autres commentaires sont complémentaires avec ceux exposés ici dans la review d'une autre soumission au GRETSI par une partie des même auteurs, ``Formulation d'un modèle CP d'ordre élevé en train de tenseur''.

Réponse : Pour cette raison, nous ajoutons ci-dessous nos réponses concernant le papier en question.

Papier n°2 : Formulation du modèle CP d'ordre élevé en un train de tenseurs (TT) d'ordre réduit

----
Commentaires généraux

L'article propose une étude du lien entre l'ensemble des décompositions canonique polyadiques d'un tenseur et l'ensemble des décompositions en Train de Tenseurs (TT) de ce même tenseur. L'article propose deux contributions : 1/ Montrer une ``équivalence'' entre CPD et TT, 2/ Proposer un algorithme pour obtenir une CPD depuis un tenseur stocké sous format TT.

J'émets un avis réservé sur cet article, en particulier concernant la nouveauté des résultats proposés, voir plus bas. De plus, les preuves étant manquantes, on ne peut juger de la justesse des résultats avancés, mais ceux-ci diffèrent légèrement d'autres résultats de la littérature qui s'intéressent aux mêmes propriétés. Je suis en particulier assez surpris de la contrainte de rang sur les dépliements des cœurs dans la définition du modèle TT, qui ne figure nul part ailleurs dans la littérature. En se passant de ces contraintes de rang, Phan et coauteurs~\cite{phan2018tensor} ont montrés un lien entre CPD et TT qui semble plus général que celui proposé ici. Par ailleurs, le théorème 1 est déjà prouvé dans l'article de 2011 d'Oseledets (section 3.1).

Réponse : Ici encore, il n'est pas acceptable de contextualiser nos travaux par rapport à un papier déposé sur arXiv. L'absence de contrainte sur le rang multilinéaire des cœurs est interprétée comme un surcroit de généralité. C'est une opinion tout à fait critiquable du point de vue du problème qui nous intéresse, c'est-à-dire l'estimation des facteurs à partir des cœurs TT. La contribution principale de cet article est de proposer l'algorithme JIRAFE pour l'estimation des paramètres d'un modèle CPD. Cet algorithme est basé sur une première étape de construction des cœurs TT et ensuite d'une estimation couplée des facteurs et des matrices latentes. Notre preuve est constructive et basée sur la SVD. Il est bien connu que la troncature de la SVD fournit un rang dit minimal. Le relecteur semble avoir complètement ignoré cette partie du travail. Dans ce même esprit, le rappel que le th1 a déjà été énoncé dans ~\cite{Oseledets2011Tensor} est superflu compte tenu que nous donnons cette information dans le texte précèdent le th1. Le relecteur aurait plutôt du se pencher sur le th2, l'algorithme JIRAFE et ses performances.

Commentaires détaillés

Les commentaires ci-dessous ne sont pas ordonnés par ordre d'importance.

1/ Cet article est une traduction de l'article EUSIPCO de 2018 des mêmes auteurs. L'article de Phan reprenant les mêmes thématiques mais avec des résultats peut-être plus généraux datant également de 2018, il était certes normal de ne pas en avoir connaissance en 2018, mais il est dommage de ne pas le voir au moins mentionné ici.

Réponse : L'article de Pham et al. est postérieur à notre publication d'EUSIPCO. De plus, une fois encore, il n'est pas sérieux de prendre pour référence un article déposé sur un site de partage et n'ayant pas encore fait l'objet d'une publication après évaluation par des experts.

2/ Après avoir cherché sur plusieurs publications concernant les trains de tenseurs, je n'ai pas trouvé la définition exacte donnée ici par les auteurs, en particulier le fait de contraindre le rang des dépliements des cœurs. Dans la définition de~\cite{Oseledets2011Tensor}, les cœurs sont autorisés à être de rang multilinéaire déficient à priori.

Réponse : S'il fallait reprendre exactement l'existant alors il n'y aurait plus de place pour l'originalité et la nouveauté.

Si les auteurs souhaitent s'intéresser à la question des TT-rangs minimaux, alors il faudrait séparer la définition de TT de la définition du TT minimal qui est moins général. Une faiblesse de l'article est de ne pas s'intéresser proprement à cette contrainte supplémentaire sur les cœurs.

Réponse : Dans cet article, l'hypothèse de rangs TT minimaux est clairement énoncée. Il n'y a donc aucune ambiguïté.

Ainsi, on peut imaginer qu'avec la définition des auteurs, on puisse montrer que pour un tenseur décomposé en deux TT, les cœurs soient reliés par des produits sur les modes 1 et 3 par des matrices unitaires; ce n'est certainement pas le cas avec la définition générale de TT, et dans tous les cas ce n'est pas prouvé, à ma connaissance, dans la littérature. Au fond, c'est l'argument de "l'identification" des termes du modèle TT qui me dérange ici (argument utilisé dans la preuve du papier original pour le GRETSI 2018, ainsi que pour le papier sur TT-PARALIND) et que je ne trouve pas ailleurs dans la littérature.

Réponse : L'hypothèse envisagée par le relecteur d'unitarité des matrices latentes n'est que pure spéculation ne reposant sur aucun argument rigoureux. De plus, nous ne voyons pas ce que veut montrer le relecteur. Notre impression est encore renforcée par la dernière phrase qui exprime un sentiment - « qui me dérange » - et non une remarque construite et argumentée.

3/ Le théorème 1 est une réécriture presque immédiate de la propriété prouvée dans~\cite{Oseledets2011Tensor} (paragraphe 3.1), sauf pour la discussion sur le rang de P_k. Encore une fois, cette discussion intervient car les auteurs ont supposés une définition un peu différente de TT par rapport à~\cite{Oseledets2011Tensor}. Le Théorème 1 est également reprouvé dans~\cite{phan2018tensor}, également sans discussion sur le rang des facteurs de la CPD. La littérature semble donc pointer vers la non nécessité de cette contrainte. Ainsi, étant donné un modèle TT exact pour un tenseur (pas forcement minimal), la CPD exact du tenseur s'obtient en estimant les deuxièmes facteurs des CPD d'ordre 3 des cœurs, peu importe le rang colonne des facteurs. Cela réduit également l'impact de l'autre proposition de papier des auteurs pour le GRETSI 2019.

Réponse : L'hypothèse de rang minimal est fondamentale pour le problème d'estimationauquel cet article est consacré. Dans ~\cite{Oseledets2011Tensor} le TT est vu sous l'angle modélisation et non de l'estimation. Le résultat donné dans~\cite{Oseledets2011Tensor} et rappelé dans notre th1 est inexploitable du point de vue de l'estimation. Il ne mentionne pas les quantités latentes de couplage des cœurs TT essentielles au recouvrement des facteurs à partir des cœurs. De plus, contrairement à notre travail, les indéterminations d'échelle et de permutation des colonnes ne sont pas non plus étudiées dans ~\cite{Oseledets2011Tensor}.

4/ Le théorème 2 est prouvé dans~\cite{phan2018tensor} dans une version similaire mais qui inclue des facteurs d'échelle. Si identification des termes TT et CPD il doit y avoir, ces facteurs d'échelle doivent probablement apparaître dans le théorème 2?

Réponse : Nous rappelons au relecteur que nos travaux sont antérieurs à ~\cite{phan2018tensor}. Les facteurs d'échelle et de permutation des colonnes peuvent être absorbés dans les matrices de changement de base.

5/ L'algorithme JIRAFE est également déjà étudié dans~\cite{phan2018tensor}, ou du moins un algorithme qui semble très proche. Si le résultat du théorème 2 énoncé ici est correct, JIRAFE a l'avantage de pouvoir se passer des facteurs d'échelle ce qui permet l'approche par bi-ALS successif.

Réponse : Nous rappelons au relecteur que nos travaux sont antérieurs à ~\cite{phan2018tensor}. Il est donc étrange voire malhonnête intellectuellement de dire que l'algorithme JIRAFE est déjà publié ailleurs. De plus, une comparaison sérieuse aurait mis à jour des différences importantes. Dans ~\cite{phan2018tensor}, le couplage des cœurs TT n'est pas exploité, ce qui nécessite d'estimer cœurs par cœurs les indéterminations d'échelle et de permutation. Ce qui n'est pas satisfaisant du point de vue de l'estimation car il faut des connaissances à priori sur chaque facteur et présente un coup algorithmique plus élevé. En conclusion, les algorithmes proposés dans ~\cite{phan2018tensor} ignorent une bonne partie de la richesse structurelle des cœurs TT.

6/ Dans les simulations, le temps rapporté prend-il en compte le temps de calcul du format TT?

Réponse : Bien-entendu ! Le temps mesuré par les fonctions tic-toc de Matlab prennent en compte la complexité en nombre d'opération des algorithmes mais aussi le nombre d'accès et d'écriture mémoire reflétant le cout de stockage.

Note: les commentaires faits ici sont également valable en grande partie pour la review du papier "Uniqueness of Tensor Train Decompositon with Linear Dependencies" également soumis au GRETSI 2019.

Refs:

[1] Ivan V Oseledets. Tensor-train decomposition. SIAM Journal on Scientific Computing, 33(5):2295–2317,2011.

[2] Anh-Huy Phan, Andrzej Cichocki, Ivan Oseledets, Salman Ahmadi Asl, Giuseppe Calvi, and Danilo Mandic. Tensor networks for latent variable analysis: Higher order canonical polyadic decomposition. arXiv preprint arXiv:1809.00535, 2018.

Résumé

- L'argumentation du relecteur est principalement basée sur un draft déposé sur ArXiv (non publié dans une revue ou une conférence, donc non expertisé, et dont nous n'avions pas connaissance pour cette raison).

- Ce draft est POSTERIEUR à nos propres résultats qui ne sont pas cités dans ce draft.

- Nous considérons que l'évaluation est partielle, focalisée essentiellement sur l'aspect modélisation, et plus spécialement sur les hypothèses de rang TT, faisant nullement référence aux contributions majeures qui portent d'une part sur la mise en évidence de couplages des cœurs TT, et sur l'algorithme JIRAFE et son exploitation pour l'estimation des paramètres d'un modèle CPD.

  En résumé, nous pensons qu'il s'agit d'une évaluation incomplète et non rigoureuse scientifiquement, témoignant d'une réelle subjectivité.

Uniqueness of Tensor Train Decomposition with Linear Dependencies

Yassine ZNIYED[1], Sebastian MIRON[2], Remy BOYER[3], David BRIE[2]

[1]Laboratoire des Signaux et Systèmes, CentraleSupelec, Gif-Sur-Yvette, France

[2]CRAN, Université de Lorraine, CNRS, Vandœuvre-lès-Nancy, France

[3]Laboratoire CRIStAL, Univeristé de Lille, Villeneuve d'Ascq, France

yassine.zniyed@l2s.centralesupelec.fr, sebastian.miron@univ-lorraine.fr
remy.boyer@univ-lille.fr, david.brie@univ-lorraine.fr

Résumé – Avec l'accroissement des moyens d'acquisition et de mesures, les données d'intérêt sont par essence de nature multidimensionnelle. Ceci peut s'interpréter comme un accroissement de la dimension/ordre des modèles tensoriels associés. Il y a donc là un besoin crucial d'avoir à notre disposition des représentations équivalentes d'un tenseur d'ordre élevé en un graphe de tenseurs d'ordre réduit. Dans ce travail nous considérons un graphe de type "train", c'est-à-dire, qu'un tenseur d'ordre Q sera représenté par un train de tenseur (TT) composé de $Q-2$ coeurs tensoriels d'ordre 3 et deux coeurs matriciels. Dans ce cadre, il a été démontré qu'un modèle CPD/PARAFAC de rang canonique R peut être toujours représenté de manière exacte par un modèle TT dont les coeurs sont eux-même CPD/PARAFAC de rang canonique R. Ce modèle est nommé TT-CPD. Nous généralisons cette équivalence au modèle PARALIND afin de prendre en compte des potentielles dépendances linéaires dans les facteurs. Nous donnons et discutons ici les conditions d'unicité dans le cas du modèle TT-PARALIND.

Abstract – With the increase in measurement/sensing technologies, the collected data are in intrinsically multidimensional in a large number of applications. This can be interpreted as a growth of the dimensionality/order of the associated tensor. There exists therefore a crucial need to derive equivalent and alternative models of a high-order tensor as a graph of low-order tensors. In this work we consider a " train " graph, *i.e.*, a Q-order tensor will be represented as a Tensor Train (TT) composed of $Q-2$ 3-order core tensors and two core matrices. In this context, it has been shown that a canonical rank-R CPD / PARAFAC model can always be represented exactly by a TT model whose cores are canonical rank-R CPD/PARAFAC. This model is called TT-CPD. We generalize this equivalence to the PARALIND model in order to take into account potential linear dependencies in factors. We derive and discuss here uniqueness conditions for the case of the TT-PARALIND model.

1 Introduction

Canonical Polyadic Decomposition (CPD) [6] is one of the most used tensor decompositions in signal processing. The CPD and its variants are attractive tools due to their ability to decompose tensors into physically interpretable quantities, called factors. Its uniqueness has been studied in several state-of-art articles such as [7, 10, 5]. Uniqueness and compactness are two of the advantages that make the CPD widespread. Indeed, the CPD is usually unique under mild conditions and its storage cost grows linearly with respect to the order. Recently, tensor networks (TNs) [4] have been subject of increasing interest, especially for high-order tensors, allowing more flexible tensor modelling. TNs split high-order ($Q > 3$) tensors into a set of lower-order tensors. Tensor train decomposition (TTD) [8] is one the most compact and simple TNs. Indeed, TTD breaks a high Q-order tensor into a set of Q lower-order tensors, called TT-cores. These TT-cores have orders at most equal to 3. In this sense, TNs are able to break the "curse of dimensionality".

In a recent work [12], an equivalence between the CPD and the TTD was proposed. In fact, it has been shown that a Q-order CPD of rank-R is equivalent to a train of 3-order CPD(s) of rank-R. This result makes it easier to jointly reduce the dimension and estimate the CPD factors using the TT-cores when the original tensor has a high order. Otherwise, when Q is high, the CPD factors estimation becomes a difficult task using ALS-based techniques [2]. At the same time, the existing results on the equivalence between CPD and TTD are based on assumption that the CPD factor matrices are all full column rank, in which case, estimating the factor matrices from the TT-cores is straightforward.

In this work, we focus on the case where linear dependencies are present between the columns on the factor matrices leading to high-order PARALIND (PARAllel profiles with LINear Dependences) model [3]. PARALIND is a variant of the CPD with constrained factor/loading matrices, that models a linearly dependent factor $\boldsymbol{P}$ as a product of a full column rank matrix $\tilde{\boldsymbol{P}}$ and an interaction matrix $\boldsymbol{\Phi}$. Matrix $\boldsymbol{\Phi}$ introduces the linear dependency and rank deficiency in $\boldsymbol{P}$. Linear dependencies in factor matrices are of great interest in real scenarios and can be encountered in chemometrics applications [3] or in array signal processing [11], to mention a few. In this work, some new equivalence results between the TTD and PARALIND are presented. The TT-cores structure is exposed when the Q-order PARALIND has only two full column rank factor matrices. Partial and full uniqueness conditions for the new TT-PARALIND model are also studied.

The notations used in this paper are as follows. Scalars, vectors, matrices and tensors are represented by x, $\boldsymbol{x}$, $\boldsymbol{X}$ and $\boldsymbol{\mathcal{X}}$, respectively. The symbols $(\cdot)^{\mathsf{T}}$ and $(\cdot)^{-1}$ denote, respectively, the transpose and the inverse. $\boldsymbol{\mathcal{I}}_{k,R}$ denotes the k-order identity tensor of size $R \times \cdots \times R$, and $\boldsymbol{\mathcal{I}}_{2,R} = \boldsymbol{I}_R$. The matrix $\text{unfold}_k \boldsymbol{\mathcal{X}}$ of size $N_k \times N_1 \cdots N_{k-1} N_{k+1} \cdots N_Q$ refers to the k-mode unfolding of $\boldsymbol{\mathcal{X}}$ of size $N_1 \times \cdots \times N_Q$. The n-mode product is denoted by $\underset{n}{\bullet}$. The contraction $\overset{p}{\underset{q}{\bullet}}$ between two tensors $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$ of size $N_1 \times \cdots \times N_Q$ and $M_1 \times \cdots \times M_P$, with $N_q = M_p$ is a tensor of order $(Q + P - 2)$ such that

$$[\boldsymbol{\mathcal{A}} \overset{p}{\underset{q}{\bullet}} \boldsymbol{\mathcal{B}}]_{n_1,\ldots,n_{q-1},n_{q+1},\ldots,n_Q,m_1,\ldots,m_{p-1},m_{p+1},\ldots,m_P} = \sum_{k=1}^{N_q} [\boldsymbol{\mathcal{A}}]_{n_1,\ldots,n_{q-1},k,n_{q+1},\ldots,n_Q} [\boldsymbol{\mathcal{B}}]_{m_1,\ldots,m_{p-1},k,m_{p+1},\ldots,m_P}.$$

2 Equivalence between the PARALIND and the TTD

2.1 Tensor-Train Decomposition (TTD)

Definition 1. *A Q-order tensor of size $N_1 \times \ldots \times N_Q$ that follows a Tensor Train decomposition (TTD) [8] of TT-ranks $\{R_1, \ldots, R_{Q-1}\}$ admits the following definition:*

$$\boldsymbol{\mathcal{X}} = \boldsymbol{G}_1 \overset{1}{\underset{2}{\bullet}} \boldsymbol{\mathcal{G}}_2 \overset{1}{\underset{3}{\bullet}} \boldsymbol{\mathcal{G}}_3 \overset{1}{\underset{4}{\bullet}} \cdots \overset{1}{\underset{Q-1}{\bullet}} \boldsymbol{\mathcal{G}}_{Q-1} \overset{1}{\underset{Q}{\bullet}} \boldsymbol{G}_Q, \tag{1}$$

where the TT-cores $\boldsymbol{G}_1, \boldsymbol{\mathcal{G}}_q$, and $\boldsymbol{G}_Q$ are, respectively, of dimensions $N_1 \times R_1, R_{q-1} \times N_q \times R_q$, and $R_{Q-1} \times N_Q$, for $2 \le q \le Q-1$, and we have $\text{rank}\{\boldsymbol{G}_1\} = R_1$, $\text{rank}\{\boldsymbol{G}_Q\} = R_{Q-1}$, $\text{rank}\{\text{unfold}_1 \boldsymbol{\mathcal{G}}_q\} = R_{q-1}$, *and* $\text{rank}\{\text{unfold}_3 \boldsymbol{\mathcal{G}}_q\} = R_q$.

It is straightforward to see that the TTD of $\boldsymbol{\mathcal{X}}$ in eq. (1) is not unique since

$$\boldsymbol{\mathcal{X}} = \boldsymbol{A}_1 \overset{1}{\underset{2}{\bullet}} \boldsymbol{\mathcal{A}}_2 \overset{1}{\underset{3}{\bullet}} \boldsymbol{\mathcal{A}}_3 \overset{1}{\underset{4}{\bullet}} \cdots \overset{1}{\underset{Q-1}{\bullet}} \boldsymbol{\mathcal{A}}_{Q-1} \overset{1}{\underset{Q}{\bullet}} \boldsymbol{A}_Q,$$

where

$$\begin{aligned}\boldsymbol{A}_1 &= \boldsymbol{G}_1 \boldsymbol{U}_1^{-1},\\ \boldsymbol{A}_Q &= \boldsymbol{U}_{Q-1} \boldsymbol{G}_Q,\\ \boldsymbol{\mathcal{A}}_q &= \boldsymbol{U}_{q-1} \overset{1}{\underset{2}{\bullet}} \boldsymbol{\mathcal{G}}_q \overset{1}{\underset{3}{\bullet}} \boldsymbol{U}_q^{-1}.\end{aligned}$$

For $1 \le q \le Q-1$, $\boldsymbol{U}_q$ are square nonsingular matrices of dimension $R_q \times R_q$. In practice, the TTD is performed thanks to the state-of-art TT-SVD algorithm [8]. It is a sequential algorithm that recovers the TT-cores $\boldsymbol{\mathcal{G}}_q$ based on $(Q-1)$ SVDs applied to several matrix-based reshapings using the original tensor $\boldsymbol{\mathcal{X}}$. This algorithm allows to recover the true TT-cores up to a post and pre-multiplication by transformation (*change-of-basis*) matrices due to the extraction of dominant subspaces when using the SVD. In the next section, we will derive the structure of the estimated TT-cores when the original tensor $\boldsymbol{\mathcal{X}}$ follows a CPD with linear dependencies between the columns of the loading matrices.

2.2 PARALIND-TTD equivalence

Consider Q-order tensor $\boldsymbol{\mathcal{X}}$ of size $N_1 \times \cdots \times N_Q$ that follows a rank-R CPD:

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{I}}_{Q,R} \underset{1}{\bullet} \boldsymbol{P}_1 \underset{2}{\bullet} \boldsymbol{P}_2 \cdots \underset{Q}{\bullet} \boldsymbol{P}_Q, \tag{2}$$

where the loading matrices $\boldsymbol{P}_q$ are of size $N_q \times R$. It was shown in [12] that if the loading matrices $\boldsymbol{P}_q$ are full-column rank for $1 \le q \le Q$, then they can be recovered from the TT-cores by order-3 CPD decompositions.

In this section we study the case where linear dependencies are present between the columns of the loading matrices of (2). Thus, a loading matrix $\boldsymbol{P}_q$ can be expressed as:

$$\boldsymbol{P}_q = \tilde{\boldsymbol{P}}_q \boldsymbol{\Phi}_q, \tag{3}$$

where $\tilde{\boldsymbol{P}}_q$ is full column rank of size $N_q \times R_q$ ($R_q \le R$) and $\boldsymbol{\Phi}_q$ is a rank deficient matrix of size $R_q \times R$ containing the dependency pattern between the columns of $\tilde{\boldsymbol{P}}_q$. This CPD model with linear dependencies is also known as *PARALIND (PARAllel profiles with LINear Dependences)* [3].

Theorem 1 (PARALIND - TTD equivalence)**.** *Decomposing tensor $\boldsymbol{\mathcal{X}}$ in (2) into a TT format, where $\boldsymbol{P}_1$ and $\boldsymbol{P}_Q$ are full column rank matrices, and $\boldsymbol{P}_q$ ($2 \le q \le Q-1$) follow (3), recovers the estimated TT-cores such that*

$$\begin{aligned}\boldsymbol{G}_1 &= \boldsymbol{P}_1 \boldsymbol{U}_1^{-1},\\ \boldsymbol{\mathcal{G}}_q &= \boldsymbol{\mathcal{I}}_{3,R} \underset{1}{\bullet} \boldsymbol{U}_{q-1} \underset{2}{\bullet} (\tilde{\boldsymbol{P}}_q \boldsymbol{\Phi}_q) \underset{3}{\bullet} \boldsymbol{U}_q^{-\mathsf{T}},\ 2 \le q \le Q-1\\ \boldsymbol{G}_Q &= \boldsymbol{U}_{Q-1} \boldsymbol{P}_Q^{\mathsf{T}},\end{aligned}$$

where, for $1 \le q \le Q-1$, $\boldsymbol{U}_q$ is a square $R \times R$ nonsingular matrix. The TT-cores $\boldsymbol{G}_1, \boldsymbol{\mathcal{G}}_q$, and $\boldsymbol{G}_Q$ are, respectively, of dimensions $N_1 \times R, R \times N_q \times R$, and $R \times N_Q$, given TT-ranks all equal to R.

Proof. Note that tensor $\boldsymbol{\mathcal{I}}_{Q,R}$ in eq. (2) can be expressed as

$$\boldsymbol{\mathcal{I}}_{Q,R} = \mathbf{I}_R \overset{1}{\underset{2}{\bullet}} \boldsymbol{\mathcal{I}}_{3,R} \overset{1}{\underset{3}{\bullet}} \cdots \overset{1}{\underset{Q-1}{\bullet}} \boldsymbol{\mathcal{I}}_{3,R} \overset{1}{\underset{Q}{\bullet}} \mathbf{I}_R, \tag{4}$$

replacing eq. (4) into eq. (2), we get

$$\begin{aligned}\boldsymbol{\mathcal{X}} &= (\mathbf{I}_R \overset{1}{\underset{2}{\bullet}} \boldsymbol{\mathcal{I}}_{3,R} \overset{1}{\underset{3}{\bullet}} \cdots \overset{1}{\underset{Q}{\bullet}} \mathbf{I}_R) \underset{1}{\bullet} \boldsymbol{P}_1 \underset{2}{\bullet} \boldsymbol{P}_2 \underset{3}{\bullet} \cdots \underset{Q}{\bullet} \boldsymbol{P}_Q\\ &= (\mathbf{I}_R \overset{1}{\underset{2}{\bullet}} \boldsymbol{\mathcal{I}}_{3,R} \overset{1}{\underset{3}{\bullet}} \cdots \overset{1}{\underset{Q}{\bullet}} \mathbf{I}_R) \underset{1}{\bullet} \boldsymbol{P}_1 \underset{2}{\bullet} \tilde{\boldsymbol{P}}_2 \boldsymbol{\Phi}_2 \underset{3}{\bullet} \cdots \underset{Q}{\bullet} \boldsymbol{P}_Q\end{aligned}$$

Before introducing the ambiguity matrices $\boldsymbol{U}_q$, tensor $\boldsymbol{\mathcal{X}}$ can then be expressed into a TT format as

$$\begin{aligned}\boldsymbol{\mathcal{X}} = &\underbrace{\boldsymbol{P}_1}_{\boldsymbol{A}_1} \overset{1}{\underset{2}{\bullet}} \underbrace{(\boldsymbol{\mathcal{I}}_{3,R} \underset{2}{\bullet} \tilde{\boldsymbol{P}}_2 \boldsymbol{\Phi}_2)}_{\boldsymbol{\mathcal{A}}_2} \overset{1}{\underset{3}{\bullet}} \cdots \overset{1}{\underset{Q-2}{\bullet}} \underbrace{(\boldsymbol{\mathcal{I}}_{3,R} \underset{2}{\bullet} \tilde{\boldsymbol{P}}_{Q-2} \boldsymbol{\Phi}_{Q-2})}_{\boldsymbol{\mathcal{A}}_{Q-2}}\\ &\overset{1}{\underset{Q-1}{\bullet}} \underbrace{(\boldsymbol{\mathcal{I}}_{3,R} \underset{2}{\bullet} \tilde{\boldsymbol{P}}_{Q-1} \boldsymbol{\Phi}_{Q-1})}_{\boldsymbol{\mathcal{A}}_{Q-1}} \overset{1}{\underset{Q}{\bullet}} \underbrace{\boldsymbol{P}_Q^T}_{\boldsymbol{A}_Q}.\end{aligned} \tag{5}$$

One may note that for $2 \leq q \leq Q-1$, the considered TT-cores $\boldsymbol{A}_1$, $\boldsymbol{\mathcal{A}}_q$ and $\boldsymbol{A}_Q$ verify the definition of the TTD given in Definition 1, *i.e.,* $\text{rank}\{\boldsymbol{A}_1\} = \text{rank}\{\boldsymbol{A}_Q\} = \text{rank}\{\text{unfold}_1\boldsymbol{\mathcal{A}}_q\} = \text{rank}\{\text{unfold}_3\boldsymbol{\mathcal{A}}_q\} = R$, which justify that matrices $\boldsymbol{P}_1$ and $\boldsymbol{P}_Q$ must be of full column rank. By identifying the TT-cores $\boldsymbol{\mathcal{A}}_q$ in eq. (5), introducing the pre- and post-multiplication ambiguity matrices $\boldsymbol{U}_q$ presented in 2.1, and using the following equivalence

$$\boldsymbol{\mathcal{G}}_q = \boldsymbol{U}_{q-1} \overset{1}{\underset{2}{\bullet}} \boldsymbol{\mathcal{A}}_q \overset{1}{\underset{3}{\bullet}} \boldsymbol{U}_q^{-1} = \boldsymbol{\mathcal{A}}_q \underset{1}{\bullet} \boldsymbol{U}_{q-1} \underset{3}{\bullet} \boldsymbol{U}_q^{-T},$$

theorem 1 is proven. □

3 Uniqueness of the PARALIND-TTD

One of the most popular condition for the uniqueness of the CPD decomposition is the Kruskal's condition [7] relying on the concept of "Kruskal-rank", or simply krank. The krank of an $N \times R$ matrix $\boldsymbol{P}$, denoted by $\text{krank}\{\boldsymbol{P}\}$, is the maximum value of $\ell \in \mathbb{N}$ such that every ℓ columns of $\boldsymbol{P}$ are linearly independent. By definition, the krank of a matrix is less than or equal to its rank. Kruskal proved [7] that the condition

$$\text{krank}\{\boldsymbol{P}_1\} + \text{krank}\{\boldsymbol{P}_2\} + \text{krank}\{\boldsymbol{P}_3\} \geq 2R + 2 \quad (6)$$

is sufficient for uniqueness of the CPD decomposition in (2), with $Q = 3$. Furthermore, it becomes a necessary and sufficient condition in the cases $R = 2$ or 3 (see [10]). Herein, by uniqueness, we understand "essential uniqueness", meaning that if another set of matrices $\bar{\boldsymbol{P}}_1$, $\bar{\boldsymbol{P}}_2$ and $\bar{\boldsymbol{P}}_3$ verify (6), then there exists a permutation matrix $\boldsymbol{\Pi}$ and three invertible diagonal scaling matrices ($\boldsymbol{\Delta}_1$, $\boldsymbol{\Delta}_2$, $\boldsymbol{\Delta}_3$) satisfying $\boldsymbol{\Delta}_1\boldsymbol{\Delta}_2\boldsymbol{\Delta}_3 = \boldsymbol{I}_R$, where $\boldsymbol{I}_R$ is the R-th-order identity matrix, such that

$$\bar{\boldsymbol{P}}_1 = \boldsymbol{P}_1\boldsymbol{\Pi}\boldsymbol{\Delta}_1, \quad \bar{\boldsymbol{P}}_2 = \boldsymbol{P}_2\boldsymbol{\Pi}\boldsymbol{\Delta}_2, \quad \bar{\boldsymbol{P}}_3 = \boldsymbol{P}_3\boldsymbol{\Pi}\boldsymbol{\Delta}_3.$$

The uniqueness condition (6) has been generalised to Q-order CPDs in [9]. It states that the loading matrices $\boldsymbol{P}_q$ ($q = 1, \ldots, Q$) in (2) can be uniquely estimated from $\boldsymbol{\mathcal{X}}$ if

$$\sum_{q=1}^{Q} \text{krank}\{\boldsymbol{P}_q\} \geq 2R + (Q-1). \quad (7)$$

This condition is sufficient but not necessary for the uniqueness of the CPD decomposition.

Based on Kruskal's uniqueness condition as well as on the results derived in [5], we formulate in the following a partial and a full uniqueness condition for the PARALIND-TTD of a Q-order tensor.

Theorem 2 (Partial uniqueness of TT-PARALIND)**.** *The loading matrix $\boldsymbol{P}_q$ can be uniquely recovered from the estimated TT decomposition of $\boldsymbol{\mathcal{X}}$ if there exist q_1 and q_2 ($q_1 \neq q_2 \neq q$), such that:*

$$\begin{cases} \text{rank}\{\boldsymbol{P}_{q_1}\} = \text{rank}\{\boldsymbol{P}_{q_2}\} = R, \\ \text{rank}\{\boldsymbol{P}_q\} \geq 2. \end{cases}$$

Proof. In the CPD (2) the order of the factor matrices is arbitrary and can be changed by a simple index permutation. Thus, in the following we will suppose, without loss of generality, that $q_1 = 1$ and $q_2 = Q$. The fact that $\text{rank}\{\boldsymbol{P}_1\} = \text{rank}\{\boldsymbol{P}_Q\} = R$ implies that the square matrices $\boldsymbol{U}_q$ in theorem 1 are all full rank R. Therefore, the $\boldsymbol{\mathcal{G}}_q$ tensor can be uniquely recovered from $\boldsymbol{\mathcal{X}}$ by the TT-SVD algorithm.

According to theorem 1, the tensor $\boldsymbol{\mathcal{G}}_q$ can be expressed as:

$$\boldsymbol{\mathcal{G}}_q = \boldsymbol{\mathcal{I}}_{3,R} \underset{1}{\bullet} \boldsymbol{U}_{q-1} \underset{2}{\bullet} \boldsymbol{P}_q \underset{3}{\bullet} \boldsymbol{U}_q^{-\mathsf{T}}. \quad (8)$$

Following Kruskal's uniqueness condition (6), the factor matrices in (8) can be recovered from $\boldsymbol{\mathcal{G}}_q$ if

$$\text{krank}\{\boldsymbol{U}_{q-1}\} + \text{krank}\{\boldsymbol{P}_q\} + \text{krank}\{\boldsymbol{U}_q^{-\mathsf{T}}\} \geq 2R+2. \quad (9)$$

However, in our case we are only interested in recovering $\boldsymbol{P}_q$, which allows to relax Kruskal's condition. It was proven in [5] that the matrix $\boldsymbol{P}_q$ can be be uniquely estimated from $\boldsymbol{\mathcal{G}}_q$ if

$$\text{krank}\{\boldsymbol{U}_{q-1}\} + \text{rank}\{\boldsymbol{P}_q\} + \text{krank}\{\boldsymbol{U}_q^{-\mathsf{T}}\} \geq 2R+2. \quad (10)$$

As $\boldsymbol{U}_{q-1}$ and $\boldsymbol{U}_q$ are full rank square matrices, and $\text{rank}\{\boldsymbol{P}_q\} \geq 2$, (9) is verified, which completes the proof. □

Theorem 3 (Full TT-PARALIND uniqueness)**.** *The loading matrices $\boldsymbol{P}_1, \ldots, \boldsymbol{P}_Q$ can be uniquely recovered from the estimated TT-cores $\boldsymbol{G}_1, \boldsymbol{\mathcal{G}}_2, \ldots, \boldsymbol{\mathcal{G}}_{Q-1}, \boldsymbol{G}_Q$ if:*

$$\begin{cases} \text{rank}\{\boldsymbol{P}_1\} = \text{rank}\{\boldsymbol{P}_Q\} = R \\ \text{rank}\{\boldsymbol{P}_q\} \geq 2, \quad 2 < q < Q-1 \\ \text{krank}\{\boldsymbol{P}_2\}, \text{krank}\{\boldsymbol{P}_{Q-1}\} \geq 2. \end{cases}$$

Proof. This result is a consequence of theorem 2. The uniqueness of factor matrices $\boldsymbol{P}_2, \ldots, \boldsymbol{P}_{Q-1}$ can be proven by repeatedly applying theorem 2 to the different TT-cores $\boldsymbol{\mathcal{G}}_q$, $2 \leq q \leq Q-1$. Meanwhile, condition (10) does not guarantee uniqueness of the change-of-basis matrices $\boldsymbol{U}_{q-1}$ and $\boldsymbol{U}_q$. In order to guarantee this, Kruskal's condition (9) must be verified.

Thus, the condition $\text{krank}\{\boldsymbol{P}_2\}, \text{krank}\{\boldsymbol{P}_{Q-1}\} \geq 2$ implies uniqueness of the CPD decomposition of TT-cores $\boldsymbol{\mathcal{G}}_2$ and $\boldsymbol{\mathcal{G}}_{Q-1}$ and consequently, the uniqueness of the $R \times R$ nonsingular matrices $\boldsymbol{U}_1$ and $\boldsymbol{U}_{Q-1}$. From theorem 1 we get:

$$\boldsymbol{P}_1 = \boldsymbol{G}_1\boldsymbol{U}_1 \text{ and } \boldsymbol{P}_Q = \boldsymbol{G}_Q^{\mathsf{T}}\boldsymbol{U}_{Q-1}^{-\mathsf{T}}.$$

Thus, the unique recovery of $\boldsymbol{G}_1$ and $\boldsymbol{G}_Q$ from $\boldsymbol{\mathcal{X}}$ along with uniqueness of $\boldsymbol{U}_1$ and $\boldsymbol{U}_{Q-1}$ implies uniqueness of factor matrices $\boldsymbol{P}_1$ and $\boldsymbol{P}_Q$, which completes the proof. □

4 Discussion and conclusions

4.1 More restrictive conditions

Compared to Kruskal's condition (7) for order-Q CPD, the uniqueness condition of theorem 3 is more restrictive. For example, in the case of a fourth-order tensor ($Q = 4$), the condition of theorem 3 implies $\sum_{q=1}^{4} \text{krank}\{\boldsymbol{P}_q\} \geq 2R + 4$, while Kruskal's condition requires $\sum_{q=1}^{4} \operatorname{krank}\{\boldsymbol{P}_q\} \geq 2R+3$. This is a direct consequence of imposing simultaneous (partial) uniqueness on all the order-3 TT-cores. More restrictive uniqueness conditions is the price to pay for having a numerically efficient algorithm, that guarantees recovery of the loading matrices for a wide variety of scenarios.

4.2 Estimation scheme architecture

It is worth noting that, from an algorithmic point of view, the estimation of the loading matrices $\boldsymbol{P}_q$ can be done either in parallel or sequentially. For a parallel estimation scheme, the conditions of theorem 3 are sufficient. In [12], a sequential scheme was proposed, based on a sequential retrieval of both matrices $\boldsymbol{P}_q$ and $\boldsymbol{U}_q$. It requires at each step the knowledge of $\boldsymbol{U}_{q-1}$ for decomposing $\boldsymbol{\mathcal{G}}_q$. To use a similar sequential scheme for the TT-PARALIND model, it is necessary to also ensure the uniqueness of matrices $\boldsymbol{U}_q$. This can be done by replacing condition $\operatorname{rank}\{\boldsymbol{P}_q\} \geq 2$ $(2 < q < Q-1)$ in theorem 3 by a stronger one, $\operatorname{krank}\{\boldsymbol{P}_q\} \geq 2$ $(2 < q < Q-1)$.

4.3 Perspectives

1. The condition $\operatorname{rank}\{\boldsymbol{P}_1\} = \operatorname{rank}\{\boldsymbol{P}_Q\}$ in theorem 1 requires the knowledge of the indices of full-rank modes of tensor $\boldsymbol{\mathcal{X}}$, which are then arbitrarily fixed to 1 and Q; once these two modes are fixed, the order in which the remaining modes are processed is arbitrary. It is certainly possible to obtain a condition involving only one full rank matrix, but in this case the order in which the other modes are processed must be carefully chosen to guarantee the required rank conditions for the TT-SVD algorithm. This aspect is currently under investigation.

2. A very promising application domain of these results is the low-rank approximation of high-dimensional probability mass functions. In this case, these uniqueness results are of upmost importance as the linear dependencies in the model could account for the random variables correlations. A potential application is represented by the flow cytometry data analysis, as shown in [1].

5 Conclusion

The factorisation of a high-order tensor into a collection of low-order tensors, called cores, is an important research topic. Indeed, this family of methods called tensor Networks is an efficient way to mitigate the well-known “curse of dimentionality” problem. In this work, we prove that a Q-order PARALIND of rank R can be reformulated as a $Q-2$ train of tensors possibly column-deficiency and two full column rank matrices. The condition of partial and full uniqueness are exposed and discussed.

References

[1] David Brie, Rémi Klotz, Sebastian Miron, Saïd Moussaoui, Christian Mustin, Philippe Bécuwe, and Stéphanie Grandemange. Joint analysis of flow cytometry data and fluorescence spectra as a non-negative array factorization problem. *Chemometrics and Intelligent Laboratory Systems*, 137:21–32, 2014.

[2] R. Bro, N. D. Sidiropoulos, and G. B. Giannakis. A fast least squares algorithm for separating trilinear mixtures. In *ICA99 - Int. Workshop on Independent Component Analysis and Blind Separation.*, 1999.

[3] Rasmus Bro, Richard A. Harshman, Nicholas D. Sidiropoulos, and Margaret E. Lundy. Modeling multi-way data with linearly dependent loadings. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 23(7-8):324–340, 2009.

[4] A. Cichocki. Era of big data processing: A new approach via tensor networks and tensor decompositions. *CoRR*, 2014.

[5] Xijing Guo, Sebastian Miron, David Brie, and Alwin Stegeman. Uni-mode and partial uniqueness conditions for CANDECOMP/PARAFAC of three-way arrays with linearly dependent loadings. *SIAM Journal on Matrix Analysis and Applications*, 33(1):111–129, 2012.

[6] R. A. Harshman. Foundations of the PARAFAC procedure: Models and conditions for an explanatory multimodal factor analysis. *UCLA Working Papers in Phonetics*, 16:1–84, 1970.

[7] J. B. Kruskal. Three-way arrays: Rank and uniqueness of trilinear decompositions with application to arithmetic complexity and statistics. *Linear Algebra Appl.*, 18:95–138, 1977.

[8] I. V. Oseledets. Tensor-Train decomposition. *SIAM J. Scientific Computing*, 33(5):2295–2317, 2011.

[9] Nicholas D. Sidiropoulos and Rasmus Bro. On the uniqueness of multilinear decomposition of N-way arrays. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 14(3):229–239, 2000.

[10] J. M. F. ten Berge and N. D. Sidiropoulos. On uniqueness in CANDECOMP/PARAFAC. *Psychometrika*, 67(3):399–409, September 2002.

[11] L. Xu, T. Jing, Y. Longxiang, and Z. Hongbo. PARALIND-based identifiability results for parameter estimation via uniform linear array. *EURASIP Journal on Advances in Signal Processing*, 2012.

[12] Y. Zniyed, R. Boyer, A. L. F. de Almeida, and G. Favier. Formulation du modèle CP d'ordre élevé en un train de tenseurs (TT) d'ordre réduit. *Article soumis au GRETSI'19*.