



On four-way CP model estimation efficiency

Violetta Simonacci¹ · Michele Gallo²

Received: 13 April 2022 / Accepted: 4 August 2022
© The Author(s) 2022

Abstract

The latent structure of four-dimensional tensors can be investigated by means of the four-way CANDECOMP/PARAFAC model. This technique is seldom used because its estimating design is challenging from an algorithmic and interpretational standpoint. Parameter estimation with a least-squares approach can be computationally costly, especially under difficult conditions such as factor collinearity and model over-specification. In this work, we implement a 4th-order extension of the efficient trilinear procedure INT-2 to tackle estimating setbacks and test it in a simulation study.

Keywords AQLD · CANDECOMP/PARAFAC · Computational efficiency · Multi-way data · QALS · 4th-order tensor

1 Introduction and background

The idea of simultaneously analyzing multiple two-way objects and studying the eigenstructure of higher-order tensors was formalized for the first time from a purely mathematical perspective in Hitchcock (1927, 1928) by showing how a tridimensional tensor can be represented in polyadic form.

Third-order extensions of bilinear component models for data analysis made their appearance in the sixties. The pioneering work of Tucker (1966), who created the Tucker3 model, was shortly followed by the introduction of a more restrictive technique proposed simultaneously by Carroll and Chang (1970) and Harshman (1970) with the name of CANDECOMP and PARAFAC, respectively. These latter

✉ Violetta Simonacci
violetta.simonacci@unina.it

Michele Gallo
mgallo@unior.it

¹ Department of Social Sciences, University of Naples “Federico II”, Vico Monte della Pietà, 1, 80134 Naples, Italy

² Department of Human and Social Sciences, University of Naples “L’Orientale”, Largo San Giovanni Maggiore, 30, 80134 Naples, Italy

techniques are often designated with the term of CANDECOMP/PARAFAC (CP) model as recommended by Kiers (2000) to standardize the technical vocabulary.

Many versions, extensions, and reinventions of the Tucker3 and CP models have been proposed, see Kroonenberg (2008) and Smilde et al. (2005) for a review. The natural extension to n th-order tensors with $n > 3$ of these three-way decompositions was envisaged from the beginning. Carroll and Chang (1970) already proposed a seven-way version of their algorithm.

The Tucker3 decomposition is generally perceived as the “true” higher-order extension of SVD because of its flexibility. This model is purely explorative, it can always be fitted to fully-crossed tridimensional data, and focuses only on maximizing explained variability. Tucker3 results are hard to interpret in terms of latent constructs because the model is characterized by sub-space uniqueness and rotational freedom. For this reason, it is the preferred method for dimensionality reduction and exploration of within-mode variability.

On the other hand, the CP model yields unique solutions under mild conditions. The model is component-unique because it imposes the restriction of the simultaneous simple structure assumption, namely, the idea that the underlying solution is unique for all samples (Cattell 1944). This characteristic makes the CP model appealing for exploring the latent structure of complex data but also harder to estimate. There still exists an open discussion on applicative and theoretical aspects of the CP model, rarely addressed in a higher-order setting, some of which represent the focus of this work.

Applications on third-order tensors have become more customary in recent years. After achieving quick acceptance in the field of psychometrics and chemometrics, they have seen growing applicability also in other disciplines (neuroscience, signal processing, text mining, etc.). A comprehensive overview is provided in Acar and Yener (2008) and Kolda and Bader (2009) for all multilinear tools. It is evident that, outside natural sciences, the diffusion of the CP model is still lacking (Kroonenberg et al. 2016).

For $n = 4$, applications are sporadic in all fields (Escandar et al. 2007) even though it has been demonstrated that using two- or three-dimensional tools on four-way tensors reduces the capability of extracting the intrinsic quadrilinear information from the data and badly affects estimation (Zhang et al. 2019). The reason for such scarce success is linked to interpretation issues, procedural complexity, and parameter estimation difficulties.

The benchmark procedure for fitting the CP model is PARAFAC-ALS (ALS), which, however, has the disadvantage of being slow at converging. Moreover, the ALS estimation process is adversely affected, both in terms of accuracy and efficiency, by specific data problems, namely factor collinearity, bad initialization, and wrong model specification. Possible solutions to these issues include model selection tools (Chen et al. 2001; Timmerman and Kiers 2000; Bro and Kiers 2003; Ceulemans and Kiers 2006; Xia et al. 2007b), and repeated random initialization. These fixes come at an additional computational cost, weighing down ALS convergence even more. For large datasets, this problem becomes even more relevant.

The reason ALS is still the procedure of choice lies in its stable convergence and well-defined properties, such as a monotonically decreasing fit function.

Estimation difficulties brought about the proliferation of alternative algorithms, some of which have been adapted to four-way CP with slight modifications. For details, see the APQLD, RSWAQLD, AWRCQLD, AQLD and SAQLD in Xia et al. (2007a), Fu et al. (2011), Kang et al. (2013), Qing et al. (2014), and Xie et al. (2017). These procedures are extensions of the best performing three-way alternatives to ALS, namely APTD (Xia et al. 2005), ATLD (Wu et al. 1998) and SWATLD (Chen et al. 2000). Comparative studies for third-order tensors confirm that ALS is the most reliable choice under general circumstances (Faber et al. 2003; Tomasi and Bro 2006; Yu et al. 2011; Zhang et al. 2015). Even in the four-way comparative study of Xie et al. (2017), quadrilinear ALS (QALS) appears more stable, especially under difficult data conditions, and proves to be superior in terms of model fitting. In brief, despite positive features such as speedy convergence and better performance under collinearity and over-specification, alternative algorithms struggle to match ALS precision.

A recent research thread demonstrated that a combinatory approach, integrating algorithms with complementary points of strength, could provide a suitable solution (Gallo et al. 2018; Simonacci and Gallo 2019, 2020). Two integrated algorithms INT and INT-2 were proposed which concatenate SWATLD and ATLD steps with ALS, respectively, to ensure faster convergence, stability, and insensitivity to wrong model specification. From this perspective, we will be implementing a quadrilinear integrated procedure (QINT-2), as a possible extension of this methodology, by also addressing the specificity of the four-way case. QINT-2 efficiency and stability performance is tested in a comparative Monte Carlo simulation study under varied conditions.

To conclude, we present an application in the social science field on Italian academics data. It will be illustrated how important issues such as gender, role, and regional differences can be easily studied by means of a four-way tool, which provides valuable and quickly interpretable insight for the implementation of educational policies for reducing gaps. Four-way results may be perceived as difficult to read for many reasons. Readability issues range from trivial aspects such as sign indeterminacy and oblique components to more substantial problems such as identifying the correct meaning and dimensionality of the inherent structure. Nonetheless, most of these problems are marginal when the research query is clear. To this end, the practicability of the four-way CP model is exemplified with help of visualization.

In Sect. 2 4th-order tensors are introduced, then the four-way CP model with QALS estimation and the QINT-2 procedure are laid out; in Sect. 3 the simulation study comparing QINT-2 with QALS is presented; in Sect. 4 the application on Italian academics is illustrated, and lastly, Sect. 5 includes the final discussion.

2 Methods

In this section, the four-way CP model is illustrated, after introducing 4th-order tensor notation. Afterward, an in-depth discussion on parameter estimation leads to the exposition of the proposed integrated methodology.

2.1 Notation

A 4th-order tensor $\mathcal{A}(I \times J \times K \times L)$ with generic element \mathcal{A}_{ijkl} is a data configuration where values are stored along four ordered dimensions, conventionally identified as first mode with index $[1, \dots, i, \dots, I]$, second-mode with index $[1, \dots, j, \dots, J]$, third-mode with index $[1, \dots, k, \dots, K]$ and fourth mode with index $[1, \dots, l, \dots, L]$.

Such tensor can be seen as a collection of first-order tensors or vectors called fibers by fixing all mode indices except one. By extending the concept of row and column vectors of a matrix, four types of fibers can be identified of I -, J -, K -, and L -dimensions. The total number of fibers of each type is obtained by the multiplication of the remaining indices. For example, there are IKL fibers $\mathcal{A}_{i:kl}$ with dimension J .

Similarly, by fixing two or one indices, the tensor can be expressed as a collection of matrices or 3rd-order tensors. For example, by fixing the third and the fourth mode, we obtain a collection of $K \times L$ matrices $\mathcal{A}_{::kl}(I \times J)$ and by fixing only the fourth mode, a collection of L third-order tensors $\mathcal{A}_{:::l}(I \times J \times K)$ is yielded. This latter operation is defined in Xie et al. (2017) as four-way slicing.

The information contained in a 4th-order tensor can also be rearranged in many ways. In detail, objects of smaller dimensions can be built by juxtaposing one or more modes of the analysis. This operation is defined as flattening or unfolding.

A pseudo-fully stretched (PFS) array is a third-order block obtained by flattening the original tensor along one dimension. The tensor \mathcal{A} can be rearranged in many PFS configurations by considering different mode combinations and ordering. Only four out of the possible PFS objects will be described, as conducive to the model illustrated in the next subsection. Let us define the following PFS arrays designated by juxtaposed modes: $\mathcal{A}^{JK}(I \times JK \times L)$, $\mathcal{A}^{KL}(J \times KL \times I)$, $\mathcal{A}^{LI}(K \times LI \times J)$ and $\mathcal{A}^{IJ}(L \times IJ \times K)$, see Kang et al. (2013) for details. The 2nd-order sections found by fixing the last index of each of these blocks can be referred to as PFS array frontal slices and denoted with $\mathcal{A}_{:::l}^{JK}(I \times JK)$, $\mathcal{A}_{:::i}^{KL}(J \times KL)$, $\mathcal{A}_{:::j}^{LI}(K \times LI)$ and $\mathcal{A}_{:::k}^{IJ}(L \times IJ)$.

2.2 Four-way CP mode and parameter estimation

A 4th-order tensor \mathcal{A} can be expressed in polyadic form by formulating its structural part $\hat{\mathcal{A}}$ as the sum of a finite number of 1st-order factors $\mathbf{a}_f \in \mathbb{R}^I$, $\mathbf{b}_f \in \mathbb{R}^J$, $\mathbf{c}_f \in \mathbb{R}^K$ and $\mathbf{d}_f \in \mathbb{R}^L$:

$$\mathcal{A} = \hat{\mathcal{A}} + \mathcal{E} = \sum_{f=1}^F \mathbf{a}_f \circ \mathbf{b}_f \circ \mathbf{c}_f \circ \mathbf{d}_f + \mathcal{E} \tag{1}$$

Each f set of factors is defined as a tetrad and $\mathcal{E}(I \times J \times K \times L)$ is the tensor of residuals. The minimal number of tetrads required to describe the tensor represents its rank, denoted with R .

Four factor matrices can be derived from the polyadic expression, each storing the F first-order objects of the same dimension. These matrices correspond to the first, second, third and fourth mode parameters and can be defined

as: $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_f, \dots, \mathbf{a}_F]$ with dimensions $(I \times F)$, $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_f, \dots, \mathbf{b}_F]$ with dimensions $(J \times F)$, $\mathbf{C} = [\mathbf{c}_1, \dots, \mathbf{c}_f, \dots, \mathbf{c}_F]$ with dimensions $(K \times F)$ and $\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_f, \dots, \mathbf{d}_F]$ with dimensions $(L \times F)$.

The polyadic decomposition is at the base of the CP model formulation. Given a noisy tensor, the CP model aims to find the F tetrads which ensure its best low-rank approximation. Ideally, the model is set by the user to extract $F = R$ tetrads, in this case, it is also called rank-decomposition. It is impossible to know the real rank of a tensor in advance. Often the model is over-specified with $F > R$, causing estimation issues.

The four-way CP model can be described using a PFS array notation in the following manner

$$\mathcal{F}_{::l}^{K} = \mathbf{A} \text{diag}(\mathbf{d}_l)(\mathbf{C} \odot \mathbf{B})^t + \mathcal{E}_{::l}^{K} \quad l = 1, \dots, L; \tag{2}$$

$$\mathcal{F}_{::i}^{L} = \mathbf{B} \text{diag}(\mathbf{a}_i)(\mathbf{D} \odot \mathbf{C})^t + \mathcal{E}_{::i}^{L} \quad i = 1, \dots, I; \tag{3}$$

$$\mathcal{F}_{::j}^{J} = \mathbf{C} \text{diag}(\mathbf{b}_j)(\mathbf{A} \odot \mathbf{D})^t + \mathcal{E}_{::j}^{J} \quad j = 1, \dots, J; \tag{4}$$

$$\mathcal{F}_{::k}^{I} = \mathbf{D} \text{diag}(\mathbf{c}_k)(\mathbf{B} \odot \mathbf{A})^t + \mathcal{E}_{::k}^{I} \quad k = 1, \dots, K. \tag{5}$$

In this formulation the symbol \odot identifies the Khatri–Rao product while the objects denoted as $\text{diag}(\mathbf{d}_l)$, $\text{diag}(\mathbf{a}_i)$, $\text{diag}(\mathbf{b}_j)$ and $\text{diag}(\mathbf{c}_k)$ are diagonal matrices extracting the l th, i th, j th and k th row of the corresponding factor matrices.

The CP model has the appealing property of being unique under mild conditions also in a four-way setting (Sidiropoulos and Bro 2000). However, the determinacy of the model comes at the price of a finicky estimation process.

The preferred estimating procedure for fitting the model is QALS. This method is based on a simple least-squares optimization criterion. Using the notation in Eq. 2, the QALS objective function can be formulated in terms of PFS arrays as follows, where the symbol $\| \cdot \|$ denotes the Frobenius norm:

$$\min_{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}} = \sum_{l=1}^L \|\mathcal{F}_{::l}^{K} - \mathcal{F}_{::l}^{K}\|^2 = \sum_{l=1}^L \|\mathcal{F}_{::l}^{K} - \mathbf{A} \text{diag}(\mathbf{d}_l)(\mathbf{C} \odot \mathbf{B})^t\|^2. \tag{6}$$

The QALS algorithm based on this function is an iterative procedure comprising four successive steps, each estimating one of the four sets of parameters. Conventionally, the algorithm converges when Loss of Fit relative changes ($rLoF$) become smaller than a user set threshold (e.g. $1e-06$).

As discussed in Sect. 1, the least-squares approach is designated as the procedure of choice in most comparison studies, both in a three-way (Faber et al. 2003; Tomasi and Bro 2006; Yu et al. 2011, 2012; Zhang et al. 2015) and four-way setting (Xie et al. 2017). Several benefits make QALS the reliable choice: (1) the algorithm is guaranteed to converge, (2) its convergence properties are clear; (3) it outperforms the competitors in terms of final fit and stability; (4) it is resistant to high noise

contamination. Nonetheless, it has been largely demonstrated that QALS records non-competitive convergence times. The algorithm has an inherent lack of efficiency connected to the usage of Khatri–Rao products on large matrices, which encumbers the convergence process, making it unsuitable for large tensors.

Moreover, specific conditions may cause the iterative process to slow down even more. Bad initialization values, collinearity and over-specification (Mitchell and Burdick 1993, 1994; Kiers 1998; Zhang et al. 2015) are likely to cause temporary degeneracies. In this scenario, the procedure progresses very slowly for many iterations. The process results inefficient but eventually finds a satisfactory solution. On occasion, permanent degenerate solutions may also occur when the procedure fails to emerge from the slow-down. A degeneracy is flagged when two factors present a high negative correlation.

A strategy to help reduce degeneracies is to repeat the procedure from different random starting points and select the best solution (random runs). In this manner, the problem of degenerate solutions is mostly solved, however, an additional strain is put on computational time. Similarly, procedures devised to select the correct rank of the model in advance can also be computationally expensive and do not ensure a correct outcome.

These shortcomings call upon the search for an alternative, efficient algorithm, less vulnerable to the degeneracy conditions detailed for QALS.

In a three-way setting, one of the procedures considered to be particularly strong with respect to ALS weaknesses is ATLD. ATLD was introduced with the declared goal to be strong with respect to ALS's major setbacks: sensitivity to over-factoring and slow convergence. The peculiar characteristic of this procedure is that it has a separate loss function for each set of parameters, focusing on the diagonal information in the data.

In a four-way setting, the loss functions for ATLD, referred to as AQLD, can be expressed in many notations. Here, coherently with previous formulations, a PFS array notation is used

$$\text{LF}(\mathbf{D}) = \sum_{l=1}^L \|\mathcal{J}_{::l}^{JK} - \mathbf{A} \text{diag}(\mathbf{d}_l)(\mathbf{C} \odot \mathbf{B})^t\|^2; \quad (7)$$

$$\text{LF}(\mathbf{A}) = \sum_{i=1}^I \|\mathcal{J}_{::i}^{KL} - \mathbf{B} \text{diag}(\mathbf{a}_i)(\mathbf{D} \odot \mathbf{C})^t\|^2; \quad (8)$$

$$\text{LF}(\mathbf{B}) = \sum_{j=1}^J \|\mathcal{J}_{::j}^{LI} - \mathbf{C} \text{diag}(\mathbf{b}_j)(\mathbf{A} \odot \mathbf{D})^t\|^2; \quad (9)$$

$$\text{LF}(\mathbf{C}) = \sum_{k=1}^K \|\mathcal{J}_{::k}^{IL} - \mathbf{D} \text{diag}(\mathbf{c}_k)(\mathbf{B} \odot \mathbf{A})^t\|^2. \quad (10)$$

Four distinct loss functions ensure different response surfaces, a faster exit from temporary degeneracies, and a steeper convergence curve. In addition, due to the differential properties of these objective functions, AQLD is insensitive to over-specification (Zhang et al. 2015).

Nevertheless, AQLD becomes unstable in presence of higher noise levels and greatly sacrifices precision. Both three-way procedures (SWATLD) and four-way extensions (RSWAQLD, AWRCQLD, and SAQLD) were implemented to compensate for this problem by adding additional weights and terms to the loss functions. These modifications were not sufficient to reach QALS stability (Xie et al. 2017).

A viable solution to this problem was presented in a three-way setting in Gallo et al. (2018) and Simonacci and Gallo (2019, 2020). This approach based on algorithm integration will be quickly recalled and then implemented in a novel four-way version in the following subsection.

2.3 Quadrilinear INtegrated algorithm

The integrated approach is based on the simple idea of combining the advantages of two fitting procedures and balancing out their specific performance issues. In detail, the main goal is to obtain an efficient estimation process that ensures the same stability of a least-squares method while dealing with collinearity and over-specification more suitably. A compromise between reliability and speed is reached by first optimizing parameters with an efficient procedure and then refining results with ALS steps to get optimal fit and further stability.

2.3.1 Integrated approach in a three-way setting

Two proposals were implemented in a three-way setting: the procedure INT, which concatenates SWATLD with ALS steps, and INT-2, which is more focused on boosting efficiency and uses the faster but less stable alternative ATLD. Both integrated algorithms consist of two optimization stages. For exemplifying, INT-2 can be described in this fashion:

- In Stage I, ATLD estimation is carried out, allowing quick jumps in the convergence process and helping retrieve the solution in case of difficult data conditions and over-specification. The procedure stops when the first stage *rLoF* criterion is met. The procedure allows the user to freely set the value for this interim convergence parameter as long as it is equal to or larger than the final convergence *rLoF* threshold. The authors' recommendation is to set interim convergence to $1e - 02$ under general conditions. Using a tighter parameter increases efficiency and over-specification tolerance, however, it may yield slightly noisier solutions.
- In Stage II, estimation is resumed with ALS steps to ensure desirable properties such as stability and least-squares results. It is important to note that this stage is mandatory. Even when the final *rLoF* is already reached at the end of Stage I, which can happen if the interim parameter is quite strict, the algorithm will still perform at least two ALS iterations.

INT and INT-2 behave quite similarly in simulation studies as, despite the volatility of stand-alone ATLD, INT-2 does not inherit this characteristic. Overall INT-2 appears faster than INT and just as reliable (Simonacci et al. 2019; Simonacci 2020).

2.3.2 QINT-2

In extending the integrated algorithm methodology to 4th-order tensors, the first issue is to decide which procedure to use in Stage I for the efficiency boost. As previously discussed, there are several ATLD/SWATLD extensions to a four-way setting, namely RSWAQLD, AWRCQLD, AQLD, and SAQLD. Each alternative was considered carefully. A comparative study was not carried out as already provided by Xie et al. (2017) in which the authors argue that SAQLD is the best option in terms of efficiency under general conditions, nonetheless it is unreliable for high noise and collinearity.

If all circumstances are considered, RSWAQLD and AQLD appear as the better compromise. The performance of these algorithms is similar but, after a quick comparison, it was found that RSWAQLD regularization parameters add improve stability but complicate the estimation process and may slightly decrease efficiency. This feature is desirable for a stand-alone procedure but not necessary for an integrated approach with a successive refinement stage. AQLD was thus selected for the four-way integrated alternative: QINT-2 was built with a starting AQLD stage followed by a QALS one, keeping the format of its three-way counterpart INT-2.

In writing the procedure, a second relevant issue arose concerning computations. As shown in Qing et al. (2014, pp. 9–10), there are different formulations of the four-way problem, which require alternative arrangements of the original tensor. AQLD is generally computed with a PFS notation, as described in this paper, while conventionally QALS is presented using fully stretched matrices (two-way unfolded PFS arrays).

The use of different arrangements affects estimation steps. Avoidance of flattening operations is generally conducive to better identification of parameters by preserving higher dimensionality while two-way unfolding can improve the speed of iterations but is more demanding in terms of memory due to larger Khatri–Rao products. Such aspects are rarely discussed in comparative studies, which limit themselves to the original formulations of the procedures.

In developing QINT-2, we decided to tackle this issue by presenting a consistent formulation. QALS was rewritten with a PFS notation, prioritizing memory usage. The revised QALS version is used for both QINT-2 second stage and the QALS algorithm in the comparative study. This choice seems like a sensible solution for a fair comparison, keeping in mind that the alternative notations are also feasible as long as consistently applied.

The full QINT-2 procedure is displayed in Algorithm 1. In the following section, the performance of QINT-2 is compared to QALS in a simulation study to assess its viability.

Algorithm 1 QINT-2

Require: \mathcal{F} with \mathcal{F}_{ijkl} ;
 initconv ▷ Interim convergence for AQLD step, default 1e-02
 conv ▷ Final convergence, default 1e-06
 maxit ▷ Maximum number of iterations, default 5000

Preparatory steps

- 1: **for** l in $1, \dots, L$ **do** $\text{list3}_l \leftarrow \mathcal{F}_{:::l}$
 $\mathcal{F}_{:::l}^{JK} \leftarrow \text{unfold}(\text{list3}_l)$ ▷ frontal slice of first mode PFS array \mathcal{F}^{JK}
 - 2: **end for**
 - 3: $\mathcal{F}^{IJ} \leftarrow \text{toArray}(\text{unfold}(\mathcal{F}^{JK}), L, I \times J, K)$ ▷ unfold with respect to third mode
 - 4: $\mathcal{F}^{LI} \leftarrow \text{toArray}(\text{unfold}(\mathcal{F}^{IJ}), K, L \times I, J)$
 - 5: $\mathcal{F}^{KL} \leftarrow \text{toArray}(\text{unfold}(\mathcal{F}^{LI}), J, K \times L, I)$
▷ toArray transforms a two-way object into a three-way one, with given dimensions
-

Initialize

Randomly generate $\mathbf{A}^{(0)}$; $\mathbf{B}^{(0)}$; $\mathbf{C}^{(0)}$; $\mathbf{D}^{(0)}$ from uniform distribution
 $it \leftarrow 0$; $fit \leftarrow 100$; $\text{oldfit} \leftarrow fit + 2\text{conv} * fit$ ▷ Starting parameters for loop
▷ it = iteration, fit =current residual sum of squares, oldfit =previous residual sum of squares

STAGE I: AQLD

- 6: **while** $\text{abs}((fit - \text{oldfit})/fit) > \text{initconv}$ & $it < \text{maxit}$ **do**
 $it \leftarrow it + 1$; $\text{oldfit} \leftarrow fit$
 - 7: **for** l in $1, \dots, L$ **do** $\mathbf{D}_{l:(it)} = \text{diag}[(\mathbf{A}_{(it-1)}^+ \mathcal{F}_{:::l}^{JK} ((\mathbf{C}_{(it-1)} \odot \mathbf{B}_{(it-1)})^+)^l)^l]$
 - 8: **end for**
 - 9: **for** k in $1, \dots, K$ **do** $\mathbf{C}_{k:(it)} = \text{diag}[(\mathbf{D}_{(it)}^+ \mathcal{F}_{:::k}^{IJ} ((\mathbf{B}_{(it-1)} \odot \mathbf{A}_{(it-1)})^+)^l)^l]$
 - 10: **end for**
 - 11: **for** j in $1, \dots, J$ **do** $\mathbf{B}_{j:(it)} = \text{diag}[(\mathbf{C}_{(it)}^+ \mathcal{F}_{:::j}^{LI} ((\mathbf{A}_{(it-1)} \odot \mathbf{D}_{(it)})^+)^l)^l]$
 - 12: **end for**
 - 13: **for** i in $1, \dots, I$ **do** $\mathbf{A}_{i:(it)} = \text{diag}[(\mathbf{B}_{(it)}^+ \mathcal{F}_{:::i}^{KL} ((\mathbf{D}_{(it)} \odot \mathbf{C}_{(it)})^+)^l)^l]$
 - 14: **end for**
▷ here + denotes the Moore–Penrose pseudo inverse
▷ At the end of each estimation step the computed matrix is normalized
 - 15: $\hat{\mathcal{F}}_{:::l}^{JK} \leftarrow \mathbf{A}_{(it)} \text{diag}(\mathbf{d}_{l(it)}) (\mathbf{C}_{(it)} \odot \mathbf{B}_{(it)})^l$
 - 16: $fit \leftarrow \text{sum}[(\mathcal{F}_{:::l}^{JK} - \hat{\mathcal{F}}_{:::l}^{JK})^2]$
 - 17: **end while**
 - 18: $\text{iteropt} \leftarrow it$ ▷ stores the number of first stage iterations
-

STAGE II: QALS

- 19: **while** $\text{abs}((fit - \text{oldfit})/fit) > \text{conv}$ OR $it \leq \text{iteropt} + 1$ & $(it < \text{maxit})$ & $(fit > \text{conv}^2)$ **do**
 $it \leftarrow it + 1$; $\text{oldfit} \leftarrow fit$
 - 20: $\mathbf{A}_{(it)} = \sum_{l=1}^L (\mathcal{F}_{:::l}^{JK} (\mathbf{C}_{(it-1)} \odot \mathbf{B}_{(it-1)})) \text{diag}(\mathbf{d}_{l(it)}) [\mathbf{D}_{(it-1)}' \mathbf{D}_{(it-1)} (\mathbf{C}_{(it-1)} \odot \mathbf{B}_{(it-1)})^l (\mathbf{C}_{(it-1)} \odot \mathbf{B}_{(it-1)})]^{-1}$
 - 21: $\mathbf{B}_{(it)} = \sum_{l=1}^L (\mathcal{F}_{:::l}^{KL} (\mathbf{D}_{(it-1)} \odot \mathbf{C}_{(it-1)})) \text{diag}(\mathbf{b}_{l(it)}) [\mathbf{A}_{(it)}' \mathbf{A}_{(it)} (\mathbf{D}_{(it-1)} \odot \mathbf{C}_{(it-1)})^l (\mathbf{D}_{(it-1)} \odot \mathbf{C}_{(it-1)})]^{-1}$
 - 22: $\mathbf{C}_{(it)} = \sum_{j=1}^J (\mathcal{F}_{:::j}^{LI} (\mathbf{A}_{(it)} \odot \mathbf{D}_{(it-1)})) \text{diag}(\mathbf{b}_{j(it)}) [\mathbf{B}_{(it)}' \mathbf{B}_{(it)} (\mathbf{A}_{(it)} \odot \mathbf{D}_{(it-1)})^l (\mathbf{A}_{(it)} \odot \mathbf{D}_{(it-1)})]^{-1}$
 - 23: $\mathbf{D}_{(it)} = \sum_{k=1}^K (\mathcal{F}_{:::k}^{IJ} (\mathbf{B}_{(it)} \odot \mathbf{A}_{(it)})) \text{diag}(\mathbf{c}_{k(it)}) [\mathbf{C}_{(it)}' \mathbf{C}_{(it)} (\mathbf{B}_{(it)} \odot \mathbf{A}_{(it)})^l (\mathbf{B}_{(it)} \odot \mathbf{A}_{(it)})]^{-1}$
 - 24: Compute fit as in step 15
 - 25: **end while**
-

Final output \mathbf{A} ; \mathbf{B} ; \mathbf{C} ; \mathbf{D} ; $\hat{\mathcal{F}}$; percentage of model fit

3 Comparing QALS and QINT-2

3.1 Simulation design

A Monte Carlo simulation study has been set up to appraise the efficiency gain ensured by QINT-2 versus QALS while monitoring stability. A comprehensive set of data conditions is considered to check performance in general and with respect to the specific problematic aspects of noise contamination, factor collinearity, and over-specification.

The following steps are implemented to generate data for each simulated 4th-order tensor. The real solution factor matrices $\mathbf{A}(I \times R)$, $\mathbf{B}(J \times R)$, $\mathbf{C}(K \times R)$ and $\mathbf{D}(L \times R)$ are generated randomly from a uniform distribution. A predetermined level of factor collinearity (CONG) is then forced on them using the QR decomposition to impose a given upper triangular matrix.

At this point, a pure 4th-order tensor is computed and then contaminated with set percentages of homoscedastic noise HO and heteroscedastic noise HE. Error tensors are created as normally distributed values. The heteroscedastic noise tensor is then multiplied by the pure tensor to provide distinct weights. Noise percentages (NOISE) are expressed in terms of total tensor inertia. For a more detailed explanation of data generation please refer to the appendix in Simonacci and Gallo (2020) where similar parameters are described in a three-way setting.

This flexible design allows us to consider different combinations of values for the described parameters so that the 4th-order artificial tensors can replicate a variety of realistic conditions. The parameter values selected for this study are reported in Table 1.

All the possible combinations between three levels of CONG, three percentages of HE, and three percentages of HO were considered for a total of 27 experimental conditions. For each condition, 50 datasets were generated to stabilize estimates. A total of 1350 datasets were artificially created.

QINT-2 and QALS were carried out on all simulated datasets by imposing both $F = R$ and $F = R + 1$ in order to assess their performance in the case of rank-decomposition and over-specification. A final $rLoF$ convergence of $1e - 06$ was set in all cases. For QINT-2 the interim convergence was set to $1e - 02$ as recommended.

Both procedures are computed using a 10 random runs initialization strategy. Without this approach, ALS struggles to converge when over-specified and, at times, encounters permanent degenerate solutions. It is demonstrated in a three-way setting

Table 1 Parameters for comparative simulations

R	3
I, J, K, L	50, 50, 50, 50
HO	15% 20% 25%
HE	10% 15% 20%
CONG	0.2 0.5 0.9
Final $rLoF$	10^{-6}
Repetitions	50

that INT and INT-2 are more stable in this respect but occasionally also degenerate (Simonacci and Gallo 2019, 2020). Random runs nearly eliminate the permanent degeneracy problem; thus, in this work, 10 random runs are performed to ensure a fair comparison in terms of speed.

In the simulations, efficiency is assessed by considering CPU time to convergence. CPU reports will refer to the performance of the algorithms on all random runs.

It is also critical to preventively make sure that the added efficiency of QINT-2 does not somehow affect the stability and goodness of the solutions. To this end, two reliability diagnostics are considered. Monitoring the minimum value reached by the loss function (FIT) is fundamental. This diagnostic always favors QALS due to the inherent structure of its least-squares loss function. Other algorithms generally struggle to compete. In this perspective, it is essential to check if QINT-2 manages to yield a least-squares solution like QALS.

Similarly, the MSE measure is calculated to assess the amount of excess modeled noise. The four loading matrices were scaled to have factors with the same norm then their average MSE is computed. For details on specific formulas and other computational aspects refer to Simonacci and Gallo (2019).

The occurrence of degeneracies is not discussed here as the random runs ensure that no failed recoveries are flagged throughout simulations.

Both procedures were written in-house with R 4.1.1 (R Core Team 2020) and Rstudio IDE v.1.4.1106 (RStudio Team 2019) using PFS arrays, as specified in Algorithm I, with the support of the package `rrcov3way` (Todorov et al. 2020). Simulations were carried out on a processing device with the following specifications: Intel(R) Xeon(R) Gold 6238 #8 CPU @ 2.10GHz 128GB RAM.

3.2 Comparative results

Starting from theoretical knowledge, QINT-2 is expected to yield least-squares results more efficiently than QALS, especially for problematic data features.

The simulation scheme was developed by creating wide-ranging data conditions to test this hypothesis and respond to three research queries. The following questions will be addressed: (1) is the capability of retrieving a least-squares solution of QINT-2 the same as QALS? (2) Is QINT-2 more efficient and stable than QALS in general terms? (3) How do different conditions such as noise, collinearity, and over-specification affect the performance gap?

The FIT and MSE diagnostics can be of help to check whether QINT-2 converges to a least-squares solution without modeling excessive noise. For all simulations, the difference in model fit is computed as $DIF_{FIT} = FIT_{QALS} - FIT_{QINT-2}$. If $abs(DIF_{FIT}) \leq 1e-04$, then the solutions are considered to be the same (Tomasi and Bro 2006).

The procedures differ more than $1e-04$ (but less than $1e-03$) only in a negligible percentage of simulations (4 instances out of the total) and, in this handful of cases, QINT-2 is slightly superior.

The MSE diagnostic gives similar information as in none of the simulations significant differences are detected. In response to (1), we conclude that QINT-2 is just as capable as QALS of retrieving a least-square solution and does not model excessive noise like the stand-alone AQLD does.

The first step in assessing efficiency was to test throughout simulations if the mean CPU time employed by QALS and QINT-2 is significantly different. Significance is confirmed by t-test results which yield a p value of ~ 0 . In the case of rank-decomposition, the efficiency gain is estimated in the interval [17%;32%], while for over-specification the range is [19%;29%]. In response to question 2, we can state that QINT-2 proved to be more efficient than QALS under general circumstances.

To better grasp the effect of data conditions, the CPU TIME distributions by NOISE and CONG are displayed in Fig. 1 with respect to rank decomposition results. The NOISE parameter shows the combinations of HO and HE.

In general, QINT-2 is far more efficient: there is no scenario where QALS is not surpassed by QINT-2. Focusing on distributions' shifts connected to NOISE levels, we find that NOISE does not appear to be detrimental. No specific effect of either HO or HE is detected, except for $CONG = 0.9$. Here we can see that for both algorithms, as the level of noise increases, estimation becomes speedier, possibly because the noise helps mitigate high collinearity.

Robustness to NOISE is a notable result for QINT-2. QALS is known to be insensitive to noise, whereas AQLD is badly affected. It is encouraging to see that QINT-2 inherits QALS stability rather than AQLD's problems in this matter.

The congruence level affects both procedures. Looking at plot scales, we notice that a slight loss in efficiency is detected between $CONG = 0.2$ and $CONG = 0.5$ while a big jump is recorded between $CONG = 0.5$ and $CONG = 0.9$. At low and high collinearity, distributions show that estimation is more unstable than for the $CONG = 0.5$ case, where both procedures yield well-separated and relatively small box plots. In the $CONG = 0.9$ case, in particular, it is possible to see how QALS

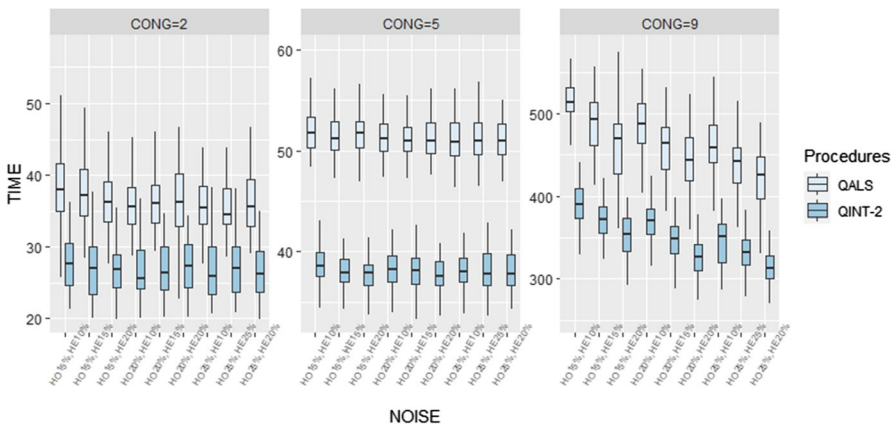


Fig. 1 Testing efficiency: CPU time by NOISE and CONG for rank-decompositions

becomes more and more unpredictable (wide-ranging distribution) if compared to QINT-2.

Let us now focus on the over-specification case displayed in Fig. 2. The first thing we notice is that the scale of all plots is increased by much. NOISE and CONG have a similar effect as described in Fig. 1. The main difference, which demonstrates the stability problems encountered by QALS due to over-specification, is given by looking at the distributions. QINT-2 appears to be more stable in computational performance throughout simulations as the range of the plots is, in general, smaller than the QALS ones, which display much longer upper whiskers and boxes. To further demonstrate this, an F test to compare variances was performed. In detail, we checked if QALS variance is significantly greater. In all the over-specification CONG/NOISE scenarios, the test yields a p-value of ~ 0 .

To conclude, we can thus answer the last query (3). The procedures are affected by NOISE in a similar way. Higher congruence appears to increase the efficiency gap, not so much in terms of median values but in terms of stability. Likewise, over-factoring increases QALS variability in convergence performance. This instability is due to QALS's propensity to degenerate with excess factors.

4 Four-way Italian academics application

In this section, we provide a demonstration of the four-way CP model's usefulness and applicability. A case study on the variability structure of Italian academics differentiated by gender-role and scientific areas throughout years and macro-regions is presented.

This data provides information on academic investment and University system diversification. By separately studying regional, time, and role variability, a four-way CP model allows the measurement of mode entities deviations with respect to

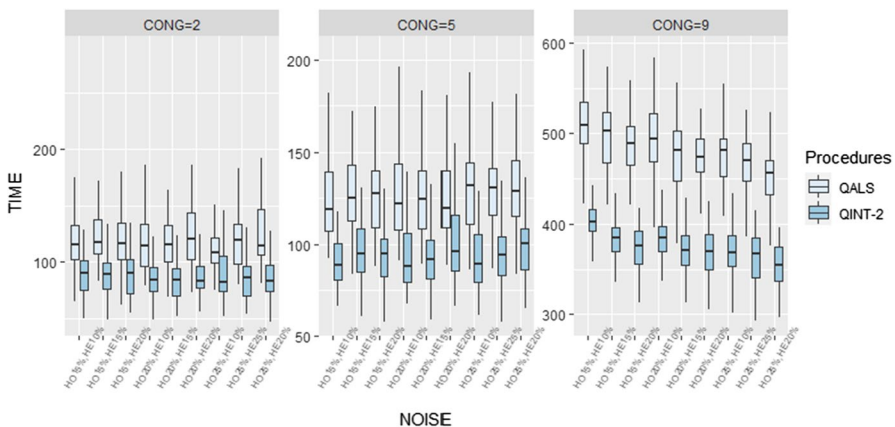


Fig. 2 Testing efficiency: CPU time by NOISE and CONG for over-specification

a common structure in terms of scientific areas. The goal of the application is to unveil relevant differences by considering all modes separately and together.

A dataset of 283,437 observations concerning Italian academics information officially recorded by the Ministry of Education from 2005 to 2020 has been arranged along four modes, creating the 4th-order tensor $\mathcal{A}(5 \times 14 \times 6 \times 5)$. In detail, the first mode entities correspond to the $I = 5$ macro-regions: North-West (abbrev. *NW*), North-East (*NE*), Central regions (*Central*), South (*South*), and Islands (*Islands*); the second mode includes the $J = 14$ scientific areas described in Table 2; the third mode considers the $K = 6$ gender-role combinations: Female Researcher (*Researcher_F*), Female Associate Professor (*Associate_F*), Female Full Professor (*Full_F*), Male Researcher (*Researcher_M*), Male Associate Professor (*Associate_M*), Male Full Professor (*Full_M*); and lastly the fourth mode selects the $L = 5$ years 2005, 2010, 2013, 2015 and 2020.

No additional pre-processing, such as column centering or normalization, was performed. This strategy was decided following Kroonenberg (2008), where it is recommended to pre-treat data with care in a multiway setting.

An $F = 1$ model was selected because it explains more than 90% of the total variability. Computations were carried out using QINT-2, however, the factor extracted by QALS was exactly the same. Given the small dimensions of the tensor, no algorithm would particularly struggle in this instance.

We display results using a powerful one-dimensional visualization tool, the per-component plot. This graphic allows plotting the loadings of all four modes together along the same direction, representing one of the F extracted factors. The main goal of this tool is to allow inter-modal comparisons and within-mode interpretation with respect to the latent measure. The per-component model for the $F = 1$ direction of the 4-way CP is displayed in Fig. 3.

Table 2 Scientific areas

Abbrev.	Scientific areas
01	Mathematical and computer sciences
02	Physical sciences
03	Chemical sciences
04	Geological sciences
05	Biological sciences
06	Medical sciences
07	Agricultural and veterinary sciences
08	Civil engineering and architecture
09	Industrial and information engineering
10	Antiquity, philological-literary and historical-artistic sciences
11	Historical, philosophical, pedagogical and psychological sciences
12	Legal sciences
13	Economics and statistics
14	Political and social sciences

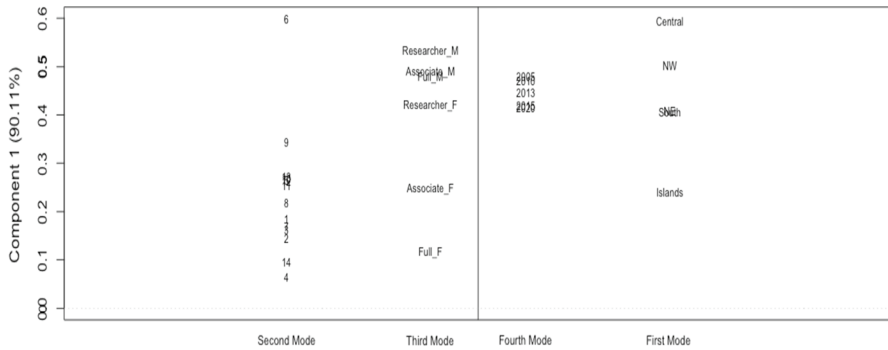


Fig. 3 $F = 1$ Per-component plot

The first step in CP interpretation is to give meaning to the latent construct by referring to the variable mode, here given by the scientific areas. After a quick assessment, it is easy to interpret the factor as a measure of academics investment scale. In other words, the ranking of the areas on the construct shows at a glance how the educational areas are prioritized in Italy. In detail, we can observe how the area with more academics is 6 followed by 9 while the areas with a smaller number of academics are 14 and 4. This can be interpreted as the typical distribution of educational and research investment in Italy.

Each of the remaining modes can then be assessed separately by referring to the common construct. For the first mode, academics are distributed in larger numbers in the macro-region *Central*, followed by *NW*. *NE* and *South* record similar values while *Islands* is quite distanced. The third mode coefficients show that the most numerous type of academics is *Researcher_M* followed by *Associate_M*, *Full_M*, and *Researcher_F*. The *Associate_F* and, even more, the *Full_F* categories are very detached. Lastly, the fourth mode gives us information on the overall number of academics present in the university system. The year 2005 recorded the highest number of academics employed. Over the years, a decreasing trend is documented with a stabilization between 2015 and 2020.

On the per-component plot, across-mode relationships can also be ascertained. By reading second and third mode coefficients together, for instance, it is possible to see that *Researcher_M* has the highest value for area 6 and the lowest for area 4; the same can be observed for *Associate_M* and *Researcher_F*.

Similarly, it is also possible to consider the loadings of all modes simultaneously. For example, we can observe that *Researcher_M* in the scientific area 6 for the macro-region *Central* in 2005 recorded the highest value ever, which progressively decreases over the period considered. Analogous readings can be carried out for any combination of different modes.

This presentation of four-way CP output yields a condensed and quick snapshot of investment differences in order to evaluate gender/role and regional disparities across time. This is done in the perspective of possibly implementing policies that may help reduce the gap in terms of geographic location, role, and gender differences.

A four-way model provides a more accurate and simple method of detection of such differences than standard bilinear tools because: (i) it allows the assessment of all modes together, (ii) it keeps variability separate for each mode, (iii) it allows to focus on one mode at the time as well as to combine information (Kroonenberg 2008).

The case study also demonstrates the ease of model interpretability. The per-component plot is an intelligible tool in which mode relations are easily detected. The number of modes does not complicate interpretation for the CP model, as it might for a four-way Tucker model, because the latent measure is the same throughout dimensions. The only interpretational challenge, no matter the order of the tensor, is to understand the phenomenon behind the underlying construct.

5 Discussion

This contribution aims at addressing the efficiency issues connected with the four-way CP model parameter estimation process. To this end, an alternative integrated estimation strategy is proposed and tested in simulations.

The most broadly used estimating algorithm QALS, albeit stable and well-defined, is not competitive in terms of computational time as its convergence slope quickly flattens. QALS efficiency is further hampered by other issues such as over-specification and collinearity. To address this difficulty, we propose to fit the quadri-linear decomposition through an integrated optimizing design called QINT-2. This procedure is a two-stage scheme extending the INT-2 algorithm (Simonacci and Gallo 2020) to a four-way setting. By estimating parameters in two steps, first with AQLD and then with QALS in a PFS array formulation, QINT-2 derives desirable properties from both algorithms.

The implemented simulation study allows verifying performance assumptions by testing QINT-2 against the baseline algorithm QALS. Realistic data conditions are ensured by considering different combinations of noise and factor congruence. In brief, the following considerations emerged from the tests on artificial data.

1. QINT-2 is more efficient than QALS in general and under all data conditions.
2. QINT-2 is more resistant to excess factors usage and collinearity than QALS, which records a less stable convergence behavior due to an increase in degenerate solutions during the performed random runs.
3. The boost in efficiency does not prevent QINT-2 from reaching a least-squares solution. This ability is demonstrated by an essentially identical performance of QINT-2 and QALS in terms of FIT and MSE. This is fundamental because it proves that the integrated approach does not inherit AQLD volatility and does not model excess noise.
4. High noise contamination does not badly affect QINT-2 as it does for AQLD.

To summarize, simulations indicate that QINT-2 is highly desirable because it is just as stable as QALS but more efficient.

Compression tools can be used to boost efficiency (Kiers and Harshman 1997; Bro and Andersson 1998; Kiers 1998). They can be combined with both QALS and QINT-2 because they act on the original tensor rather than on the estimation process. For this reason, compression would not affect the recorded performance differences in the simulation study.

It is also important to note that we preferred a conservative approach for the interim convergence parameter of QINT-2 Stage I, setting it to $1e - 02$. Stricter values may, however, strengthen QINT-2 computational efficiency. For larger datasets and cases at high risk of over-specification, stricter thresholds represent the best choice.

A short discussion on the formulation of the four-way model was also presented in the methodological section. We found in preliminary simulations on this matter that the algebraic steps to the solution are affected both in terms of efficiency and accuracy by the type of flattening and data arrangement selected. Here we simply decided to use a formulation consistent with AQLD's original format to ensure a reliable comparison. Nonetheless, this non-trivial issue can be investigated even further with an in-depth study on the computational consequences of these choices.

The four-way CP approach is exemplified in the application section. The principal merit of the case study is to show how the four-way CP model can be a useful tool in social sciences, especially with respect to the evaluation of individual differences. Thanks to the visual support provided by the per-component plot, the model yields a clear and powerful representation of the phenomenon by identifying a common latent direction which grants a quick assessment of the academic employment system in Italy with its disparities. In detail, the model aims to evaluate the gender/role and location bias in academic employment across time. Many similar applications in the service evaluation area can be envisaged. To provide one relatable example, let us consider an educational quality study in which the multiple aspects of educational quality are differentiated by the type of course, location, and type of University.

From an algorithmic standpoint, it is clear that QINT-2 efficiency and insensitivity to over-specification make it particularly effective for parameter estimation in social sciences problems. In this context, it is more difficult to assess the rank of the true quadrilinear solution. Additionally, complex data applications could present conditions that further inhibit estimation, making QINT-2 an even stronger option. This is the case of Compositional Data, vectors of relative information with a biased covariance structure. In Gallo et al. (2018) is shown that an integrated approach works well for the specific challenges of Compositional Data. This aspect is easily extendable to a four-way setting.

The granted computational advancement can also be particularly beneficial for larger data. Nonetheless, the machine requirement may still become prohibitive, especially in terms of memory usage, as the Khatri–Rao products became excessively large and hard to store. In this instance, more complex solutions may be required, such as sub-partitioning of the data (Phan and Cichocki 2011). It is also important to remember that the CP model aims at discovering the “real” multilinear solution in the data rather than simply finding a subspace that maximizes variability. This can hardly be presumed for extremely large datasets and a size reduction through a Tucker decomposition is generally a better option. Similarly, even if from

an algebraic standpoint higher-order versions can be considered, computational requirements increase in terms of the number of operations and/or size of objects to store, and care should be used in assuming the existence of a real multicollinear structure.

The efficiency gain of QINT-2 is clear in this paper, however, a confrontation with the results obtained by INT-2 in a three-way setting, suggests that this improvement is less marked for four-way data. The simulation study conditions are however not the same, so we cannot consider the results directly comparable and should limit ourselves to consider all output only with respect to the given data contingencies. From a broader perspective, it emerged that there is a need for a comprehensive simulation study of all four-way algorithms to better specify the points of strength of all procedures proposed so far, as it is still lacking for $n > 3$ tensors.

Funding Open access funding provided by Università degli Studi di Napoli Federico II within the CRUI-CARE Agreement.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Acar E, Yener B (2008) Unsupervised multiway data analysis: a literature survey. *IEEE Trans Knowl Data Eng* 21(1):6–20. <https://doi.org/10.1109/TKDE.2008.112>
- Bro R, Andersson CA (1998) Improving the speed of multiway algorithms: part II: compression. *Chemom Intell Lab Syst* 42(1–2):105–113. [https://doi.org/10.1016/S0169-7439\(98\)00011-2](https://doi.org/10.1016/S0169-7439(98)00011-2)
- Bro R, Kiers HA (2003) A new efficient method for determining the number of components in PARAFAC models. *J Chemom* 17(5):274–286. <https://doi.org/10.1002/cem.801>
- Carroll JD, Chang JJ (1970) Analysis of individual differences in multidimensional scaling via an n-way generalization of Eckart–Young decomposition. *Psychometrika* 35(3):283–319. <https://doi.org/10.1007/BF02310791>
- Cattell RB (1944) parallel proportional profiles and other principles for determining the choice of factors by rotation. *Psychometrika* 9(4):267–283. <https://doi.org/10.1007/BF02288739>
- Ceulemans E, Kiers HA (2006) Selecting among three-mode principal component models of different types and complexities: a numerical convex hull based method. *Br J Math Stat Psychol* 59(1):133–150. <https://doi.org/10.1348/000711005X64817>
- Chen ZP, Wu HL, Jiang JH, Li Y, Yu RQ (2000) A novel trilinear decomposition algorithm for second-order linear calibration. *Chemom Intell Lab Syst* 52(1):75–86. [https://doi.org/10.1016/S0169-7439\(00\)00081-2](https://doi.org/10.1016/S0169-7439(00)00081-2)
- Chen ZP, Liu Z, Cao YZ, Yu RQ (2001) Efficient way to estimate the optimum number of factors for trilinear decomposition. *Anal Chim Acta* 444(2):295–307. [https://doi.org/10.1016/S0003-2670\(01\)01179-5](https://doi.org/10.1016/S0003-2670(01)01179-5)
- Escandar GM, Olivieri AC, Faber NKM, Goicoechea HC, de la Peña AM, Poppi RJ (2007) Second- and third-order multivariate calibration: data, algorithms and applications. *TrAC Trends Anal Chem* 26(7):752–765. <https://doi.org/10.1016/j.trac.2007.04.006>

- Faber NKM, Bro R, Hopke PK (2003) Recent developments in CANDECOMP/PARAFAC algorithms: a critical review. *Chemom Intell Lab Syst* 65(1):119–137. [https://doi.org/10.1016/S0169-7439\(02\)00089-8](https://doi.org/10.1016/S0169-7439(02)00089-8)
- Fu HY, Wu HL, Yu YJ, Yu LL, Zhang SR, Nie JF, Li SF, Yu RQ (2011) A new third-order calibration method with application for analysis of four-way data arrays. *J Chemom* 25(8):408–429. <https://doi.org/10.1002/cem.1386>
- Gallo M, Simonacci V, Di Palma MA (2018) An integrated algorithm for three-way compositional data. *Quality Quantity* 10:2353–2370. <https://doi.org/10.1007/s11135-018-0745-2>
- Harshman RA (1970) Foundations of the PARAFAC procedure: models and conditions for an explanatory multimodal factor analysis. *UCLA Work Pap Phon* 16:1–84
- Hitchcock FL (1927) The expression of a tensor or a polyadic as a sum of products. *J Math Phys* 6(1–4):164–189. <https://doi.org/10.1002/sapm192761164>
- Hitchcock FL (1928) Multiple invariants and generalized rank of a p-way matrix or tensor. *J Math Phys* 7(1–4):39–79. <https://doi.org/10.1002/sapm19287139>
- Kang C, Wu HL, Yu YJ, Liu YJ, Zhang SR, Zhang XH, Yu RQ (2013) An alternative quadrilinear decomposition algorithm for four-way calibration with application to analysis of four-way fluorescence excitation-emission-ph data array. *Anal Chim Acta* 758:45–57. <https://doi.org/10.1016/j.aca.2012.10.056>
- Kiers HA (1998) A three-step algorithm for CANDECOMP/PARAFAC analysis of large data sets with multicollinearity. *J Chemom* 12(3):155–171. [https://doi.org/10.1002/\(SICI\)1099-128X\(199805/06\)12:3<155::AID-CEM502>3.0.CO;2-5](https://doi.org/10.1002/(SICI)1099-128X(199805/06)12:3<155::AID-CEM502>3.0.CO;2-5)
- Kiers HA (2000) Towards a standardized notation and terminology in multiway analysis. *J Chemom* 14(3):105–122. [https://doi.org/10.1002/1099-128X\(200005/06\)14:3<105::AID-CEM582>3.0.CO;2-I](https://doi.org/10.1002/1099-128X(200005/06)14:3<105::AID-CEM582>3.0.CO;2-I)
- Kiers HA, Harshman RA (1997) Relating two proposed methods for speedup of algorithms for fitting two-and three-way principal component and related multilinear models. *Chemom Intell Lab Syst* 36(1):31–40. [https://doi.org/10.1016/S0169-7439\(96\)00074-3](https://doi.org/10.1016/S0169-7439(96)00074-3)
- Kolda TG, Bader BW (2009) Tensor decompositions and applications. *SIAM Rev* 51(3):455–500. <https://doi.org/10.1137/07070111X>
- Kroonenberg PM (2008) Applied multiway data analysis, vol 702. Wiley, New York. ISBN 978-0-470-23799-1
- Kroonenberg PM et al (2016) My multiway analysis: from Jan de Leeuw to TWPack and back. *J Stat Softw* 73:22. <https://doi.org/10.18637/jss.v073.i03>
- Mitchell BC, Burdick DS (1993) An empirical comparison of resolution methods for three-way arrays. *Chemom Intell Lab Syst* 20(2):149–161. [https://doi.org/10.1016/0169-7439\(93\)80011-6](https://doi.org/10.1016/0169-7439(93)80011-6)
- Mitchell BC, Burdick DS (1994) Slowly converging parafac sequences: swamps and two-factor degeneracies. *J Chemom* 8(2):155–168. <https://doi.org/10.1002/cem.1180080207>
- Phan AH, Cichocki A (2011) PARAFAC algorithms for large-scale problems. *Neurocomputing* 74(11):1970–1984. <https://doi.org/10.1016/j.neucom.2010.06.030>
- Qing XD, Wu HL, Yan XF, Li Y, Ouyang LQ, Nie CC, Yu RQ (2014) Development of a novel alternating quadrilinear decomposition algorithm for the kinetic analysis of four-way room-temperature phosphorescence data. *Chemom Intell Lab Syst* 132:8–17. <https://doi.org/10.1016/j.chemolab.2013.12.011>
- R Core Team (2020) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- RStudio Team (2019) RStudio: integrated development environment for R. RStudio, Inc., Boston. <http://www.rstudio.com/>
- Sidiropoulos ND, Bro R (2000) On the uniqueness of multilinear decomposition of N-way arrays. *J Chemom* 14(3):229–239. [https://doi.org/10.1002/1099-128X\(200005/06\)14:3<229::AID-CEM587>3.0.CO;2-N](https://doi.org/10.1002/1099-128X(200005/06)14:3<229::AID-CEM587>3.0.CO;2-N)
- Simonacci V (2020) Algorithms for compositional tensors of third-order. *Book of short papers SIS2020*
- Simonacci V, Gallo M (2019) Improving PARAFAC-ALS estimates with a double optimization procedure. *Chemom Intell Lab Syst* 192:103822. <https://doi.org/10.1016/j.chemolab.2019.103822>
- Simonacci V, Gallo M (2020) An ATLD-ALS method for the trilinear decomposition of large third-order tensors. *Soft Comput* 24(18):13535–13546. <https://doi.org/10.1007/s00500-019-04320-9>
- Simonacci V, Gallo M, Guarino M (2019) A PARAFAC-ALS variant for fitting large data sets. In: *Proceedings of the scientific meeting of the Italian statistical society—smart statistics for smart applications*

- Smilde A, Bro R, Geladi P (2005) Multi-way analysis: applications in the chemical sciences. Wiley, New York. ISBN 978-0-471-98691-1
- Timmerman ME, Kiers HA (2000) Three-mode principal components analysis: choosing the numbers of components and sensitivity to local optima. *Br J Math Stat Psychol* 53(1):1–16. <https://doi.org/10.1348/000711000159132>
- Todorov V, Palma MAD, Gallo M (2020) rrcov3way: robust methods for multiway data analysis, applicable also for compositional data. <https://CRAN.R-project.org/package=rrcov3way>. R package version 0.1-18
- Tomasi G, Bro R (2006) A comparison of algorithms for fitting the PARAFAC model. *Comput Stat Data Anal* 50(7):1700–1734. <https://doi.org/10.1016/j.csda.2004.11.013>
- Tucker LR (1966) Some mathematical notes on three-mode factor analysis. *Psychometrika* 31(3):279–311. <https://doi.org/10.1007/BF02289464>
- Wu HL, Shibukawa M, Oguma K (1998) An alternating trilinear decomposition algorithm with application to calibration of HPLC-DAD for simultaneous determination of overlapped chlorinated aromatic hydrocarbons. *J Chemom* 12(1):1–26. [https://doi.org/10.1002/\(SICI\)1099-128X\(199801/02\)12:1<1::AID-CEM492>3.0.CO;2-4](https://doi.org/10.1002/(SICI)1099-128X(199801/02)12:1<1::AID-CEM492>3.0.CO;2-4)
- Xia AL, Wu HL, Fang DM, Ding YJ, Hu LQ, Yu RQ (2005) Alternating penalty trilinear decomposition algorithm for second-order calibration with application to interference-free analysis of excitation-emission matrix fluorescence data. *J Chemom* 19(2):65–76. <https://doi.org/10.1002/cem.911>
- Xia AL, Wu HL, Li SF, Zhu SH, Hu LQ, Yu RQ (2007) Alternating penalty quadrilinear decomposition algorithm for an analysis of four-way data arrays. *J Chemom* 21(3–4):133–144. <https://doi.org/10.1002/cem.1051>
- Xia AL, Wu HL, Zhang Y, Zhu SH, Han QJ, Yu RQ (2007) A novel efficient way to estimate the chemical rank of high-way data arrays. *Anal Chim Acta* 598(1):1–11. <https://doi.org/10.1016/j.aca.2007.07.015>
- Xie LX, Wu HL, Zhang XH, Wang T, Zhu L, Xiang SX, Liu Z, Yu RQ (2017) slicing data array in quadrilinear component model: an alternative quadrilinear decomposition algorithm for third-order calibration method. *Chemom Intell Lab Syst* 167:12–22. <https://doi.org/10.1016/j.chemolab.2017.05.017>
- Yu YJ, Wu HL, Nie JF, Zhang SR, Li SF, Li YN, Zhu SH, Yu RQ (2011) A comparison of several trilinear second-order calibration algorithms. *Chemom Intell Lab Syst* 106(1):93–107. <https://doi.org/10.1016/j.chemolab.2010.03.006>
- Yu YJ, Wu HL, Kang C, Wang Y, Zhao J, Li YN, Liu YJ, Yu RQ (2012) Algorithm combination strategy to obtain the second-order advantage: simultaneous determination of target analytes in plasma using three-dimensional fluorescence spectroscopy. *J Chemom* 26(5):197–208. <https://doi.org/10.1002/cem.2442>
- Zhang SR, Wu HL, Yu RQ (2015) A study on the differential strategy of some iterative trilinear decomposition algorithms: PARAFAC-ALS, ATLD, SWATLD, and APTLD. *J Chemom* 29(3):179–192. <https://doi.org/10.1002/cem.2690>
- Zhang XH, Qing XD, Wu HL (2019) Discussion on the superiority of third-order advantage: analytical application for four-way data in complex system. *Microchem J* 145:1078–1085. <https://doi.org/10.1016/j.microc.2018.12.037>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.