Psychometrika (2026), **91**, 789–813
doi:10.1017/psy.2025.10075

APPLICATION AND CASE STUDIES - ORIGINAL

Latent Functional PARAFAC for Modeling Multidimensional Longitudinal Data

Lucas Sort, Laurent Le Brusquet and Arthur Tenenhaus

Laboratoire des Signaux et Systèmes, Université Paris-Saclay CentraleSupélec, France

Corresponding author: Lucas Sort; Email: lucas.sort@riken.jp

(Received 16 December 2024; revised 20 November 2025; accepted 2 December 2025)

This manuscript is part of the special section on Integrating and analyzing complex high-dimensional data in social and behavioral sciences research. We thank Drs. Eric F. Lock and Katrijn Van Deun for serving as co-Guest Editors.

Abstract

In psychometric sciences, such as social or behavioral sciences, and, similarly, in medical sciences, it is increasingly common to deal with longitudinal data organized as high-dimensional multidimensional arrays, also known as tensors. Within this framework, the time-continuous property of longitudinal data often implies a smooth functional structure on one of the tensor modes. To help researchers investigate such data, we introduce a new tensor decomposition approach based on the PARAFAC decomposition. Our approach allows researchers to represent a high-dimensional functional tensor as a low-dimensional set of functions and feature matrices. Furthermore, to capture the underlying randomness of the statistical setting more efficiently, we introduce a probabilistic latent model in the decomposition. A covariance-based block-relaxation algorithm is derived to obtain estimates of model parameters. Thanks to the covariance formulation of the solving procedure and thanks to the probabilistic modeling, the method can be used in sparse and irregular sampling schemes, making it applicable in numerous settings. Our approach is applied in the psychometric setting to help characterize multiple neurocognitive scores observed over time in the Alzheimer's Disease Neuroimaging Initiative study. Finally, intensive simulations show a notable advantage of our method in reconstructing tensors.

Keywords: functional data; longitudinal data; tensor decomposition

1. Introduction

When handling high-dimensional tensor data, as often done in psychometrics sciences, it is common to seek a low-dimensional representation of the data to capture the latent information more efficiently and enhance characterization. Indeed, this low-dimensional representation often helps reduce the data's high-dimensionality-induced complexity. In this setting, it is increasingly common to manipulate tensors with some underlying smooth structure in one of the modes. Additionally, since data are often collected through a sampling process, tensors commonly feature a mode (the "sample mode") on which the induced sub-tensors can be seen as samples drawn from a random tensor distribution. In longitudinal studies involving tensor-valued data, these two properties are often encountered: the time-continuous property naturally implies that one of the modes has a smooth structure, while the sampling setting introduces randomness along one tensor mode, as described previously. Since integrating these smooth and probabilistic structures would lead to more interpretable results, extending the functional

© The Author(s), 2026. Published by Cambridge University Press on behalf of Psychometric Society.
This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/psy.2025.10075 Published online by Cambridge University Press

and multiway data analysis procedures appears essential to help psychometric researchers leverage the increasing complexity of their datasets.

For instance, we propose to consider in this article multiple neurocognitive markers observed over time to evaluate the cognitive abilities of Alzheimer's disease (AD) patients and cognitively normal (CN) patients observed in the Alzheimer's Disease Neuroimaging Initiative (ADNI) study. In this context, retrieving the most important modes of variation among patients can be particularly useful for medical researchers to investigate mechanisms behind cognitive decline among AD and CN patients. The data, which can be organized as a tensor with dimensions subject × time × marker, come with the properties described previously. Notably, the first mode can be seen as a "sample" mode, which comes with an underlying statistical structure, as its elements can be seen as samples drawn from some complex tensor distribution. On the other hand, the second mode, which represents time, comes with an inherent continuous property. Considering this, it appears essential to consider both properties to retrieve relevant results and provide researchers with more insightful information.

In the tensor literature, the PARAFAC decomposition Carroll & Chang, 1970; Harshman, 1970, also known as CANDECOMP/PARAFAC (CP) or the canonical polyadic decomposition (CPD), is a well-known dimension reduction approach used to represent a tensor as a sum of rank-1 tensors. The method effectively reduces the high-dimensionality of a tensor to a collection of matrices with a matching number of columns R, also called the rank. The Tucker decomposition Tucker, 1966 is a closely related technique reducing the dimension of a tensor to a collection of matrices, with a varying number of columns $r_1, \ldots, r_D$, and a (smaller) tensor, of dimensions $r_1 \times \cdots \times r_D$, called the tensor core. Both approaches have been studied intensively and applied to numerous statistical problems Sidiropoulos et al., 2017, including regression Lock, 2018; Zhou et al., 2013.

Over the past decades, numerous methods have been developed to analyze data exhibiting an underlying smooth structure. Such data are commonly referred to in the literature as functional data (see Ramsay & Silverman, 2006; Wang et al. 2016). In this context, functional principal component analysis (FPCA) has become a widely used technique for reducing the underlying infinite-dimensional space of the data to a low-dimensional subspace spanned by the most informative modes of variation, represented by the eigenfunctions of the covariance operator. To ensure smooth eigenfunctions, Rice and Silverman (1991) introduced smoothing penalties in the estimation of the mean and covariance functions. Building on this, Yao et al. (2005) proposed a local linear smoothing approach for estimating the covariance surface and extracting smooth eigenfunctions. FPCA has since been extended in various directions, including multilevel data Di et al., 2009 and multivariate data Happ & Greven, 2018, and has been applied to broader statistical problems such as regression Morris, 2015.

Despite these advances, there has been limited progress in bridging the gap between functional data analysis and tensor methods to handle tensors with inherent smooth structures. In this context, most approaches rely on smoothness penalties, particularly in the context of tensor decomposition Imaizumi & Hayashi, 2017; Reis & Ferreira, 2002; Timmerman & Kiers, 2002; Yokota et al., 2016. More recently, Guan (2023) proposed to extend the supervised PARAFAC framework of Lock and Li (2018) to decompose order-3 smooth tensors using a smoothness penalty. A key advantage of the method is its ability to handle missing data, considering shared missingness, and sparse settings, which are common in longitudinal studies. Alternatively, Han et al. (2023) extended functional singular value decomposition (SVD) to tensor-valued data using a reproducing kernel Hilbert space (RKHS) framework.

Nevertheless, a general and flexible framework that extends PARAFAC to functional tensors—capable of handling sparse and irregular sampling as well as the inherent randomness in observations—is still lacking. Notably, the method introduced in Han et al. (2023) does not accommodate irregular sampling and does not rely on a probabilistic model, while the approach introduced by Guan (2023) may lack of robustness for very sparse and irregular sampling schemes due to smoothing penalty considered. Finally, tensor methods grounded in time-series analysis have recently emerged Chang et al., 2023; Chen et al., 2022; Wang et al., 2024, offering potential to integrate smooth structures. However, these frameworks typically assume densely and regularly sampled data, which is not suited for the sparse and irregular settings considered here.

https://doi.org/10.1017/psy.2025.10075 Published online by Cambridge University Press

In this context, we introduce a new latent and functional PARAFAC model for functional tensors of any order with an underlying statistical structure. Our method uses a covariance-based algorithm that requires the estimation of (cross-)covariance surfaces of the functional entries of the tensor, allowing the decomposition of highly sparse and irregular data. Furthermore, we introduce a conditional approach, inspired by Yao et al. (2005), to predict individual-mode factors and fit the data more accurately under mild assumptions.

The article is organized as follows. In Section 2, we introduce notations and existing methods before presenting the proposed method. In Section 3, we provide the estimation setting and introduce the solving procedure for the method. An estimation method for sample-mode vectors is shown afterward. We present results obtained on simulated data in Section 4. As introduced above, we then provide an application study of our method in Section 5 to help characterize multiple neurocognitive scores observed over time in the ADNI study. Finally, we discuss our method's limitations and possible extensions in Section 6. All proofs are given in Appendix A. The code associated with this article can be found in Supplementary Materials.

2. Model

2.1. Notations

2.1.1. Tensor notations

In the following, we denote a vector by a *small* bold letter $\mathbf{a}$, a matrix by a *capital* bold letter $\mathbf{A}$, and a tensor by a *calligraphic* letter $\mathcal{A}$. Scalars are denoted so they do not fall in any of those categories. Additionally, we denote $[D] = \{1, \ldots, D\}$. Elements of a vector $\mathbf{a}$ are denoted by a_j, of a matrix $\mathbf{A}$ by $a_{jj'}$, and of a tensor $\mathcal{A}$ of order D by $a_{j_1 \ldots j_D}$, often abbreviated using the vector of indices $\mathbf{j} = (j_1, \ldots, j_D)$ as $a_{\mathbf{j}}$. The jth column of $\mathbf{A}$ is denoted by $\mathbf{a}_j$. The Frobenius norm of a matrix is defined as $\|\mathbf{A}\|_F^2 = \mathrm{tr}(\mathbf{A}^\top \mathbf{A}) = \sum_{j,j'} a_{jj'}^2$. We extend this definition to tensors as $\|\mathcal{A}\|_F^2 = \sum_{\mathbf{j}} a_{\mathbf{j}}^2$. We denote the mode-$d$ matricization of a tensor $\mathcal{X}$ as the matrix $\mathbf{X}_{(d)} \in \mathbb{R}^{p_d \times p_{(-d)}}$, where $p_{(-d)} = \prod_{d' \neq d} p_{d'}$, which columns correspond to mode-d fibers of the tensor $\mathcal{X}$. We denote the vectorization operator, stacking the columns of a matrix $\mathbf{A}$ into a column vector, as $\mathrm{vec}(\mathbf{A})$. In this context, we define the mode-d vectorization as $\mathbf{x}_{(d)} = \mathrm{vec}(\mathbf{X}_{(d)}) \in \mathbb{R}^{P \times 1}$, where $P = \prod_d p_d$. The mode-1 vectorization of $\mathcal{X}$ is sometimes simply denoted $\mathbf{x} = \mathbf{x}_{(1)}$. For two matrices $\mathbf{A} \in \mathbb{R}^{p \times d}$ and $\mathbf{B} \in \mathbb{R}^{p' \times d'}$, the Kronecker product between $\mathbf{A}$ and $\mathbf{B}$ is defined as

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \ldots & a_{1d}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{p1}\mathbf{B} & \ldots & a_{pd}\mathbf{B} \end{bmatrix} \in \mathbb{R}^{pp' \times dd'}.$$

The Khatri–Rao product between two matrices with a fixed number of columns $\mathbf{A} = [\mathbf{a}_1 \ldots \mathbf{a}_d] \in \mathbb{R}^{p \times d}$ and $\mathbf{B} = [\mathbf{b}_1 \ldots \mathbf{b}_d] \in \mathbb{R}^{p' \times d}$ is defined as a column-wise Kronecker product:

$$\mathbf{A} \odot \mathbf{B} = \begin{bmatrix} \mathbf{a}_1 \otimes \mathbf{b}_1 & \ldots & \mathbf{a}_d \otimes \mathbf{b}_d \end{bmatrix} \in \mathbb{R}^{pp' \times d}.$$

Finally, the Hadamard product between two matrices with similar dimensions $\mathbf{A} \in \mathbb{R}^{p \times d}$ and $\mathbf{B} \in \mathbb{R}^{p \times d}$ is defined as the element-wise product of the two matrices:

$$\mathbf{A} * \mathbf{B} = \begin{bmatrix} a_{11}b_{11} & \ldots & a_{1d}b_{1d} \\ \vdots & \ddots & \vdots \\ a_{p1}b_{p1} & \ldots & a_{pd}b_{pd} \end{bmatrix} \in \mathbb{R}^{p \times d}.$$

2.1.2. Functional notations

Considering an interval $\mathcal{I} \subset \mathbb{R}$, we consider now the space $L^2(\mathcal{I})$ of square integrable functions on $\mathcal{I}$ (i.e., $\int_{\mathcal{I}} f(s)^2 \mathrm{d}s < \infty$). The inner product in $L^2(\mathcal{I})$ is defined for any $f \in L^2(\mathcal{I})$ and $g \in L^2(\mathcal{I})$ by $\langle f, g \rangle_{L^2} = \int_{\mathcal{I}} f(s)g(s)\mathrm{d}s$. In the multivariate functional data setting (see Happ & Greven, 2018), we can

https://doi.org/10.1017/psy.2025.10075 Published online by Cambridge University Press

define the inner product between two multivariate functions $\mathbf{f} \in L^2(\mathcal{I})^p$ and $\mathbf{g} \in L^2(\mathcal{I})^p$ as

$$\langle \mathbf{f}, \mathbf{g} \rangle = \sum_{j=1}^{p} \langle f_j, g_j \rangle_{L^2} = \sum_{j=1}^{p} \int_{\mathcal{I}} f_j(s) g_j(s) \mathrm{d}s,$$

which defines a Hilbert Space. Considering now the space $L^2(\mathcal{I})^{p_1 \times \cdots \times p_D}$ of tensors of order D with entries in $L^2(\mathcal{I})$. We define the inner product between two functional tensors $\mathcal{F} \in L^2(\mathcal{I})^{p_1 \times \cdots \times p_D}$ and $\mathcal{G} \in L^2(\mathcal{I})^{p_1 \times \cdots \times p_D}$ as

$$\langle \mathcal{F}, \mathcal{G} \rangle = \sum_{\mathbf{j}} \langle f_{\mathbf{j}}, g_{\mathbf{j}} \rangle_{L^2} = \sum_{\mathbf{j}} \int_{\mathcal{I}} f_{\mathbf{j}}(s) g_{\mathbf{j}}(s) \mathrm{d}s, \tag{1}$$

where we denote $f_{\mathbf{j}}(s) = (f(s))_{\mathbf{j}}$.

2.2. Existing methods

The PARAFAC decomposition Carroll & Chang, 1970; Harshman, 1970 is a powerful and well-known method in the tensor literature. Its purpose is to represent any tensor as a sum of rank-1 tensors. In $\mathbb{R}^{p_1 \times \cdots \times p_D}$, the rank-$R$ PARAFAC decomposition is defined as

$$\sum_{r=1}^{R} \mathbf{a}_{1r} \circ \cdots \circ \mathbf{a}_{Dr} = [\![\mathbf{A}_1; \ldots; \mathbf{A}_D]\!], \tag{2}$$

where $\mathbf{a}_{dr} \in \mathbb{R}^{p_d}$ and $\mathbf{A}_d = [\mathbf{a}_{d1}, \ldots, \mathbf{a}_{dR}] \in \mathbb{R}^{p_d \times R}$. Here, each vector $\mathbf{a}_{dr}$ captures latent information along the mode d of the tensor. Consequently, the PARAFAC decomposition can be effectively employed as a dimension reduction technique, particularly compelling in the tensor setting, as the number of variables grows exponentially with the order of the tensor.

To approximate any tensor $\mathcal{X} \in \mathbb{R}^{p_1 \times \cdots \times p_D}$ using a PARAFAC decomposition, factor matrices $\mathbf{A}_d$ are estimated by minimizing the squared difference between the original tensor $\mathcal{X}$ and its decomposition approximation,

$$\min_{\mathbf{A}_1, \ldots, \mathbf{A}_D} \| \mathcal{X} - [\![\mathbf{A}_1; \ldots; \mathbf{A}_D]\!] \|_F^2. \tag{3}$$

The original optimization process used an alternating least squares (ALS) procedure Carroll & Chang, 1970; Harshman, 1970, where, iteratively, each factor matrix is updated while keeping the others fixed until convergence is reached. The procedure belongs to the more general class of block relaxation algorithms, known to be locally and globally convergent under mild assumptions Lange, 2013. Note that PARAFAC can be seen as a generalization of the factor analysis framework Bartholomew et al., 2011, which is commonly used in psychometrics. Indeed, the loading vector associated with the statistical/sample mode (which we will describe after) can be seen as factors, and all other parameters can be used to derive loading factors.

A good property of the PARAFAC decomposition is that aside from *scaling* and *order* indeterminacies, the decomposition is generally unique Sidiropoulos & Bro, 2000. The scaling indeterminacy stems from the fact that, for $r = 1, \ldots, R$, when considering any set of scalars $c_{1r}, \ldots, c_{Dr}$ such that $\prod_d c_{dr} = 1$, using $\tilde{\mathbf{a}}_{dr} = c_{dr} \mathbf{a}_{dr}$ leads to a similar tensor as when using $\mathbf{a}_{dr}$. This indeterminacy is usually fixed by scaling vectors $\mathbf{a}_{dr}$ for $d = 1, \ldots, D-1$ and $r = 1, \ldots, R$, such that they have unit-norm. On the other hand, the order indeterminacy is caused by the fact that permuting columns of matrices $\mathbf{A}_d$ jointly does not change the resulting tensor. This indeterminacy can be fixed by ordering matrix columns such that $\|\mathbf{a}_{D1}\| < \|\mathbf{a}_{D2}\| < \cdots < \|\mathbf{a}_{DR}\|$

Applying the PARAFAC model straightforwardly to decompose a tensor with smooth functional entries may be inefficient as the method does not integrate the smooth property. The PARAFAC decomposition also assumes a consistent structure across tensor modes. Therefore, it cannot handle schemes where functional entries are sampled irregularly, e.g., at different time points. To answer this

https://doi.org/10.1017/psy.2025.10075 Published online by Cambridge University Press

last problem, several alternatives to the PARAFAC model were introduced, especially the PARAFAC2 model Harshman, 1972; Kiers et al., 1999, which was widely employed but still does not integrate any smoothness structure into the reconstructed tensor.

Another limitation of the PARAFAC model in our setting is that it is not a statistical model. Consequently, it does not integrate the potential sampling structure of the data. In the tensor literature, several approaches are able to integrate this property (see Giordani et al. (2020) for references). The SupCP model of Lock and Li (2018) or the SupSVD model Li et al. (2016) recently went even further in this direction by allowing the introduction of covariates in the sampling modeling.

In the functional setting, the Karhunen–Loève theorem is often used to retrieve a low-dimensional representation of a high (or infinite) dimensional random process. The theorem guarantees the existence of an orthonormal basis Φ_k and a sequence ξ_k of independent random variables for any squared integrable random process X, such that

$$X(t) = \sum_{k}^{\infty} \xi_k \Phi_k(t). \tag{4}$$

By truncating the above sum to a limited number of functions, we can recover a good enough representation of X. More recently, this decomposition was extended to multivariate functional data, allowing the representation of any multivariate random process as a sum of multivariate orthonormal functions and scores. By arranging the sequence of independent random variables ξ_k in descending order of variance, this decomposition can be viewed as an extension of the well-known principal component analysis (PCA). Similarly to the PCA, estimating parameters in model (4) relies on the eigen decomposition of the covariance operator of the random process. As evoked previously, numerous approaches have been proposed to achieve this. The most notable of which, Yao et al. (2005), proposed a linear least squares smoothing procedure with asymptotic guarantees, allowing estimation of the covariance surface in sparse and irregular data sampling settings. Furthermore, using simple assumptions on the joint distribution of principal components and residual errors, they introduce an efficient approach for estimating principal components.

When dealing with multivariate or tensor-structured functional data, it appears natural to extend this decomposition. As evoked in Section 1, various approaches have been introduced to extend this decomposition to multivariate functional data. In the tensor-structured functional data setting, vectorizing the tensor appears as a straightforward way to proceed. However, similar to before, as the number of processes grows exponentially with the order of the tensor, this procedure would be computationally unfeasible for large tensors, as it would require estimating orthonormal functions for each functional feature. In addition, the vectorization of the tensor would lead to the loss of the tensor structure, and the resulting representation would still be high-dimensional, causing the approach to be inefficient.

2.3. Proposed model

We propose to extend the rank R PARAFAC decomposition introduced in (2) to functional tensors with elements in $L^2(\mathcal{I})$ by further multiplying each rank-1 sub-tensor $\mathbf{a}_{1r} \circ \cdots \circ \mathbf{a}_{Dr}$ by a functional term $\phi_r \in \mathscr{H}$, where $\mathscr{H}$ is a subset of $L^2(\mathcal{I})$,

$$\sum_{r=1}^{R} \phi_r(t)(\mathbf{a}_{1r} \circ \cdots \circ \mathbf{a}_{Dr}) = [\![\boldsymbol{\phi}(t); \mathbf{A}_1; \ldots; \mathbf{A}_D]\!], \qquad t \in \mathcal{I}, \tag{5}$$

denoting $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_r) \in \mathscr{H}^R$. We propose to name this decomposition the functional PARAFAC (F-PARAFAC) decomposition.

As discussed earlier, it is common for a functional tensor to have a sample mode, i.e., a mode on which sub-tensors can be seen as samples drawn from an underlying tensor distribution. Similar to the structural PARAFAC model Giordani et al., 2020, we propose to account for the inherent randomness

https://doi.org/10.1017/psy.2025.10075 Published online by Cambridge University Press

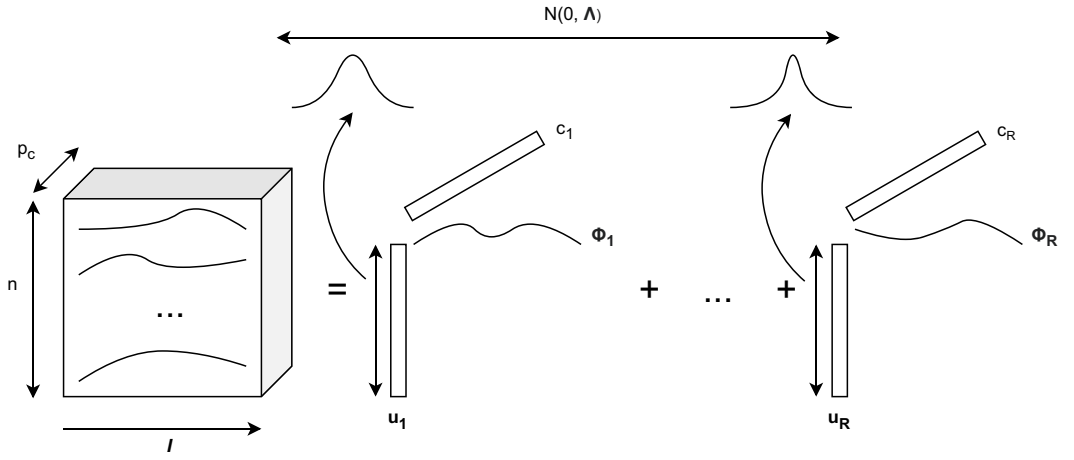


Figure 1. Latent functional PARAFAC decomposition for an order-3 functional tensor.

of random functional tensors with elements in $L^2(\mathcal{I})$ by multiplying each rank-1 functional tensor $\phi_r(t)(\mathbf{a}_{1r} \circ \cdots \circ \mathbf{a}_{Dr})$ by a random variable u_r,

$$\sum_{r=1}^{R} u_r \phi_r(t)(\mathbf{a}_{1r} \circ \cdots \circ \mathbf{a}_{Dr}) = [\![\mathbf{u}; \boldsymbol{\phi}(t); \mathbf{A}_1; \ldots; \mathbf{A}_D]\!], \qquad t \in \mathcal{I}, \tag{6}$$

denoting $\mathbf{u} = (u_1, \ldots, u_R) \in \mathbb{R}^R$. We propose naming this model the latent functional PARAFAC (LF-PARAFAC) decomposition. A visualization of the model is given in Figure 1. Assumptions on the distribution of $\mathbf{u}$ are further discussed in Section 3.

Given a random functional tensor $\mathcal{X}$, which we assume to be centered throughout this section, we are interested in finding the best LF-PARAFAC decomposition approximating $\mathcal{X}$:

$$(\boldsymbol{\psi}^*, \boldsymbol{\phi}^*, \mathbf{A}_1^*, \ldots, \mathbf{A}_D^*) = \underset{\boldsymbol{\psi}, \boldsymbol{\phi}, \mathbf{A}_1, \ldots, \mathbf{A}_D}{\operatorname{argmin}} \mathbb{E}[\|\mathcal{X} - [\![\boldsymbol{\psi}(\mathcal{X}); \boldsymbol{\phi}; \mathbf{A}_1; \ldots; \mathbf{A}_D]\!]\|^2]. \tag{7}$$

In this optimization problem, $\boldsymbol{\psi}^*$ is the operator $\mathcal{X} \mapsto \boldsymbol{\psi}^*(\mathcal{X}) = \mathbf{u}^*$ outputting the optimal (over the expectation) random vector $\mathbf{u}^* \in \mathbb{R}^{1 \times R}$ for the random functional tensor distribution $\mathscr{D}_{\mathcal{X}}$ followed by $\mathcal{X}$. It implies that for any sample $\mathcal{X}_n$ obtained from $\mathscr{D}_{\mathcal{X}}$, the criterion is optimal, in average, using the sample mode vector $\mathbf{u}_n = \boldsymbol{\psi}^*(\mathcal{X}_n)$ in its decomposition along with the other optimal parameters $\boldsymbol{\phi}^*, \mathbf{A}_1^*, \ldots, \mathbf{A}_D^*$. The asymmetry between $\mathbf{u}$, which is random, and $\boldsymbol{\phi}, \mathbf{A}_1, \ldots, \mathbf{A}_D$, which are fixed, allows us to rewrite the optimization problem as

$$\min_{\boldsymbol{\phi}, \mathbf{A}_1, \ldots, \mathbf{A}_D} \mathbb{E}[\min_{\mathbf{u}} \|\mathcal{X} - [\![\mathbf{u}; \boldsymbol{\phi}; \mathbf{A}_1; \ldots; \mathbf{A}_D]\!]\|^2]. \tag{8}$$

Proposition 1 gives the expression of Ψ^*, corresponding to the optimal matrix $\mathbf{u}$ of the optimization sub-problem on the right-hand side of (8).

Proposition 1. *Denoting $\boldsymbol{A}_{(D)} = (\boldsymbol{A}_D \odot \cdots \odot \boldsymbol{A}_1)$, and fixing $\boldsymbol{\phi}, \boldsymbol{A}_1, \ldots, \boldsymbol{A}_D$, for any $\mathcal{X}$, the optimal operator $\boldsymbol{\psi}^*$ is given by*

$$\boldsymbol{\psi}^*(\mathcal{X}) = \left(\int \boldsymbol{x}(t)^\top (\boldsymbol{A}_{(D)} \odot \boldsymbol{\phi}(t)) dt \right) \left(\int (\boldsymbol{A}_{(D)} \odot \boldsymbol{\phi}(t))^\top (\boldsymbol{A}_{(D)} \odot \boldsymbol{\phi}(t)) dt \right)^{-1}. \tag{9}$$

https://doi.org/10.1017/psy.2025.10075 Published online by Cambridge University Press

Plugging the optimal operator Ψ^* in model (6) eliminates the dependency on $\mathbf{u}$, and the problem can now be seen as depending only on $\boldsymbol{\phi}, \mathbf{A}_1, \ldots, \mathbf{A}_D$:

$$(\boldsymbol{\phi}^*, \mathbf{A}_1^*, \ldots, \mathbf{A}_D^*) = \underset{\boldsymbol{\phi}, \mathbf{A}_1, \ldots, \mathbf{A}_D}{\operatorname{argmin}} \mathbb{E}[\|\mathcal{X} - [\![\psi^*(\mathcal{X}); \boldsymbol{\phi}; \mathbf{A}_1; \ldots; \mathbf{A}_D]\!]\|^2]. \tag{10}$$

Denoting collapsed functional covariance matrices $\boldsymbol{\Sigma}_{[f]}(s,t) = \mathbb{E}[\mathbf{x}(s)(\mathbf{I}_P \otimes \mathbf{x}(t)^\top)]$ and $\boldsymbol{\Sigma}_{[d]}(s,t) = \mathbb{E}[\mathbf{X}_{(d)}(s)(\mathbf{I}_{p_{(-d)}} \otimes \mathbf{x}(t)^\top)]$, where $\mathbf{I}_P$ and $\mathbf{I}_{p_{(-d)}}$ denote the identity matrix of order P and $p_{(-d)}$ respectively, for which an interpretation of $\boldsymbol{\Sigma}_{[f]}(s,t)$ and $\boldsymbol{\Sigma}_{[d]}(s,t)$ is given in Section 3.2, we can derive Corollary 1 from Proposition 1.

Corollary 1. *Denoting* $\boldsymbol{\Lambda}^* = \mathbb{E}[\boldsymbol{u}^* \boldsymbol{u}^{*\top}]$ *the covariance matrix of* $\boldsymbol{u}^*$ *and* $\boldsymbol{K}(t) = (\boldsymbol{A}_{(D)} \odot \boldsymbol{\phi}(t)) \left(\int (\boldsymbol{A}_{(D)} \odot \boldsymbol{\phi}(t))^\top (\boldsymbol{A}_{(D)} \odot \boldsymbol{\phi}(t))\right)^{-1}$, *we have*

$$\boldsymbol{\Lambda}^* = \int \int \boldsymbol{K}(s)^\top \boldsymbol{\Sigma}_{[f]}(s,t) \boldsymbol{K}(t) ds dt. \tag{11}$$

Using Corollary 1, it is possible to show that the optimization problem only depends on the optimal covariance $\boldsymbol{\Lambda}^*$ and other operators derive via an operator C. Expressions of this operator C are derived in Lemma A1 in the Appendix.

For a given tensor, this operator C depends on the collapsed covariance $\boldsymbol{\Sigma}_{[f]}$. This dependency can be replaced by a dependency on $\boldsymbol{\Sigma}_{[d]}$ for any $d \in [D]$, which gives multiple expressions for C, introduced in Lemma A1. Lemma A2 in the Appendix allows us to derive simple stationary equations, which are introduced in Proposition 2.

Proposition 2. *For the functional mode, for fixed* $\boldsymbol{\phi}$, *we have*

$$\boldsymbol{\phi}^*(s) = \left(\int \boldsymbol{\Sigma}_{[f]}(s,t)(\boldsymbol{A}_{(D)} \odot \boldsymbol{K}(t)) dt\right) \left((\boldsymbol{A}_{(D)}^\top \boldsymbol{A}_{(D)}) * \boldsymbol{\Lambda}^*\right)^{-1}. \tag{12}$$

Denoting $\boldsymbol{A}_{(-d)} = (\boldsymbol{A}_D \odot \cdots \odot \boldsymbol{A}_{d+1} \odot \boldsymbol{A}_{d-1} \odot \cdots \odot \boldsymbol{A}_1)$, *and* $\boldsymbol{A}_{(-d)}^{\boldsymbol{\phi}}(t) = \boldsymbol{A}_{(-d)} \odot \boldsymbol{\phi}(t)$, *we have for any* $d \in [D]$ *tabular mode, for fixed* $\boldsymbol{A}_d$,

$$\boldsymbol{A}_d^* = \left(\int \int \boldsymbol{\Sigma}_{[d]}(s,t)(\boldsymbol{A}_{(-d)}^{\boldsymbol{\phi}}(s) \odot \boldsymbol{K}(t)) ds dt\right) \left(\int \boldsymbol{A}_{(-d)}^{\boldsymbol{\phi}}(t)^\top \boldsymbol{A}_{(-d)}^{\boldsymbol{\phi}}(t) dt * \boldsymbol{\Lambda}^*\right)^{-1}. \tag{13}$$

2.4. Identifiability

As said previously, a well-known advantage of the PARAFAC decomposition is the identifiability of model parameters under mild assumptions. In the probabilistic setting, we define identifiability as the fact that if for two sets of model parameters $\boldsymbol{\Theta} = (\boldsymbol{\Lambda}, \boldsymbol{\phi}, \mathbf{A}_1, \ldots, \mathbf{A}_D)$ and $\tilde{\boldsymbol{\Theta}} = (\tilde{\boldsymbol{\Lambda}}, \tilde{\mathbf{A}}_f, \tilde{\mathbf{A}}_1, \ldots, \tilde{\mathbf{A}}_D)$ we have, $L(\boldsymbol{\Theta}|\mathcal{X}) = L(\tilde{\boldsymbol{\Theta}}|\mathcal{X})$, for any data $\mathcal{X}$, where $L(\boldsymbol{\Theta}|\mathcal{X})$ stands for the model likelihood, then $\boldsymbol{\Theta} = \tilde{\boldsymbol{\Theta}}$ Lock & Li, 2018. In this context, similar to the standard setting, two well-known indeterminacies preventing parameters from being identifiable can be noted: the *scaling* and *order* indeterminacies. To remove these, we propose to consider the following assumptions:

(C1) For any $d \in [D]$ and $r \in [R]$, $\|\mathbf{a}_{dr}\|_2 = 1$, and the first non-zero value of $\mathbf{a}_{dr}$ is positive. Similarly, $\|\phi_r\|_{\mathcal{H}} = 1$ and ϕ_r is positive at the lower bound of its support.

(C2) Function ϕ_r and vectors $\mathbf{a}_{dr}$ are ordered such that sample-mode random variables u_r have decreasing variance: $\boldsymbol{\Lambda}_{rr} < \boldsymbol{\Lambda}_{r'r'}$ for any $r < r'$ (we suppose no equal variance).

Considering the scaling and order indeterminacies fixed, a simple condition on the k-rank of the feature matrices and R was derived Sidiropoulos & Bro, 2000 for tensors of any order to guarantee identifiability. In the probabilistic setting, it can be shown that the any-order structural PARAFAC

https://doi.org/10.1017/psy.2025.10075 Published online by Cambridge University Press

decomposition model Giordani et al., 2020, with just the probabilistic property considered, is identifiable (see Lemma A3 in the Appendix). We propose to extend this identifiability result in the LF-PARAFAC setting. Proposition 3 gives conditions for which model (6) is identifiable. We recall that we denote $\mathrm{kr}(\mathbf{A}_d)$ the k-rank of the feature matrix $\mathbf{A}_d$ and define it as the maximum number k such that any k columns of $\mathbf{A}_d$ are linearly independent.

Proposition 3. *Denoting $\mathscr{H}$ the space to which belong functions ϕ_r, assuming assumptions (C1) and (C2) verified, model (6) is identifiable if $\mathscr{H}$ is infinite-dimensional and if the following inequality is verified:*

$$\sum_{d=1}^{D} kr(\boldsymbol{A}_d) \geq D-1. \tag{14}$$

Conversely, assuming $\mathscr{H}$ is finite-dimensional, and, considering $\{\psi_k\}_{1\leq k\leq p}$ any basis of $\mathscr{H}$, we can denote $\boldsymbol{C}$ the $R\times p$ basis decomposition matrix for functions $\{\phi_r\}_{1\leq r\leq R}$ (i.e., with entries corresponding to $\langle\psi_k,\phi_r\rangle_{\mathscr{H}}$). In this context, model (6) is identifiable if deleting any row of $\boldsymbol{C}\odot\boldsymbol{A}_1\odot\cdots\odot\boldsymbol{A}_D$ results in a matrix having two distinct full-rank submatrices and if we have:

$$kr(\boldsymbol{C})+\sum_{d=1}^{D} kr(\boldsymbol{A}_d) \geq R+(D-1). \tag{15}$$

Proposition 3 ensures that for a small enough R, the LF-PARAFAC decomposition will be identifiable. This result also implies that using high values of R often leads to unidentifiable and, therefore, poorly conditioned decomposition. Consequently, we recommend in practice considering not too large rank values for the decomposition to ensure identifiability and therefore stability of the interpretation. Additionally, note that since we enforce unit-norm constraints (C2), the number of free parameters in the decomposition is therefore reduced by a factor RD.

3. Estimation

3.1. Observation model

We now consider N samples of a random functional tensor $\mathcal{X}$ of order D, denoted $\{\mathcal{X}_n\}_{1\leq n\leq N}$. For a given $\mathcal{X}_n$, we assume that entries of $\mathcal{X}_n$ are observed at similar time points, e.g., all functional entries of $\mathcal{X}_n$ are observed at the same time points. Observation time points of $\mathcal{X}_n$ are denoted t_{nk} with $1\leq k\leq K_n$. The number of observations K_n is allowed to be different across samples. In general, observations are contaminated with noise. Denoting $\mathcal{E}$ the random functional tensor modeling this noise, and $\mathcal{E}_{nk}$ the measurement error associated with $\mathcal{X}_n(t_{nk})$, for any $t\in\mathcal{I}$, we consider that $\mathcal{E}(t)$ has i.i.d. entries drawn from a distribution $\mathcal{N}(0,\sigma^2)$. The kth observation of the sample i, denoted $\mathcal{Y}_{nk}$, is thus modeled as

$$\mathcal{Y}_{nk}=\mathcal{X}_n(t_{nk})+\mathcal{E}_{nk}. \tag{16}$$

As we aim to decompose $\mathcal{X}$, we model the random functional tensor using the mean functional tensor $\mathcal{M}=\mathbb{E}[\mathcal{X}]$ and the LF-PARAFAC decomposition model as introduced in (6),

$$\mathcal{X}(t)=\mathcal{M}(t)+[\![\mathbf{u};\boldsymbol{\phi}(t);\mathbf{A}_1;\ldots;\mathbf{A}_D]\!]. \tag{17}$$

In practice, as introduced in Section 2.3, we, therefore, seek parameters of the LF-PARAFAC decomposition model that best fit the data given by observations $\{\mathcal{Y}_{nk}\}_{i,k}$. The noise term $\mathcal{E}$ can also be seen as the unexplained part of the LF-PARAFAC decomposition. Furthermore, under the normality assumption on the error term $\mathcal{E}_{nk}$ entries, the identifiability result of Section 2.4, derived in the maximum likelihood setting, is similar for estimates obtained in the least squares setting (as seen in Section 2.3).

https://doi.org/10.1017/psy.2025.10075 Published online by Cambridge University Press

3.2. Covariance terms

In (12) and (13), we must estimate terms $\mathbf{\Sigma}_{[f]}(s,t) = \mathbb{E}[\mathbf{x}(s)(\mathbf{I}_P \otimes \mathbf{x}(t)^\top)]$ and $\mathbf{\Sigma}_{[d]}(s,t) = \mathbb{E}[\mathbf{X}_{(d)}(s)(\mathbf{I}_{p_{(-d)}} \otimes \mathbf{x}(t)^\top)]$ to retrieve optimal functions $\boldsymbol{\phi}^*$ and factor matrices $\mathbf{A}_d^*$. As evoke previously, for any $d \in [D]$, the term $\mathbf{\Sigma}_{[d]}(s,t)$ can be seen as a transformation of the mode-d functional covariance matrix $\mathbf{\Sigma}_{(d)}(s,t) = \mathbb{E}[\mathbf{x}_{(d)}(s)\mathbf{x}_{(d)}(t)^\top] \in \mathbb{R}^{P\times P}$. Recalling that $P = \prod_{d'} p_{d'}$ and $p_{(-d)} = \prod_{d'\neq d} p_{d'}$, the transformation consists of a row-space collapse of non-mode-d entries into the column space, resulting in a matrix of dimensions $p_d \times P \cdot p_{(-d)}$.

To better grasp the idea, let's consider an order-2 functional tensor of dimensions $p_1 \times p_2$. The mode-1 functional covariance matrix can be decomposed as

$$\mathbf{\Sigma}_{(1)}(s,t) = \begin{bmatrix} \mathbf{\Sigma}_{(1,1)}(s,t) & \dots & \mathbf{\Sigma}_{(1,p_2)}(s,t) \\ \vdots & \ddots & \vdots \\ \mathbf{\Sigma}_{(p_2,1)}(s,t) & \dots & \mathbf{\Sigma}_{(p_2,p_2)}(s,t) \end{bmatrix} \in \mathbb{R}^{p_1p_2\times p_1p_2},$$

where each block functional matrix $\mathbf{\Sigma}_{(j,j')}(s,t) = \mathbb{E}[\mathcal{X}_{\cdot j}(s)\mathcal{X}_{\cdot j'}(t)^\top] \in \mathbb{R}^{p_1\times p_1}$, where $\mathcal{X}_{\cdot j}$ is the jth column of $\mathcal{X}$, corresponds to functional covariance of mode-1 variables with fixed mode-2 variables $j,j' \in [p_2]$. The functional matrix $\mathbf{\Sigma}_{[1]}(s,t)$ can then be expressed as

$$\mathbf{\Sigma}_{[1]}(s,t) = \left[\left[\mathbf{\Sigma}_{(1,1)}(s,t) \ \dots \ \mathbf{\Sigma}_{(1,p_2)}(s,t)\right] \dots \left[\mathbf{\Sigma}_{(p_2,1)}(s,t) \ \dots \ \mathbf{\Sigma}_{(p_2,p_2)}(s,t)\right]\right] \in \mathbb{R}^{p_1\times p_1p_2p_2}.$$

For $\mathbf{\Sigma}_{[2]}(s,t) \in \mathbb{R}^{p_2\times p_1p_2p_1}$, a similar reasoning can help see the structure of the matrix with respect to the functional block entries of $\mathbf{\Sigma}_{(2)}$.

Since functional matrices $\mathbf{\Sigma}_{(d)}$ can be obtained by reordering $\mathbf{\Sigma}$, only $\mathbf{\Sigma}$ must be estimated to obtain an estimate of $\mathbf{\Sigma}_{[d]}$. For this purpose, we rely on a local linear smoothing procedure outlined in Fan and Gijbels (2018). First, assuming functional tensor entries are not centered ($\mathcal{M} \neq 0$), a centering pre-processing step is carried out. The mean function $m_{\mathbf{j}} = \mathbb{E}[y_{\mathbf{j}}]$, which is estimated by aggregating observations and using the smoothing procedure evoked earlier, is removed from each functional entry of $\mathcal{Y}$. Next, for each pair of *centered* functional entries $\mathcal{Y}_{\mathbf{j}}$ and $\mathcal{Y}_{\mathbf{j}'}$, we estimate the (cross-)covariance surface $\Sigma_{\mathbf{jj}'}(s,t)$ using a similar approach. The raw-(cross-)covariance is computed for each pair of observations within each sample. Raw (cross-)covariances are then aggregated, and the values are smoothed on a regular two-dimensional grid using the same smoothing procedure as before.

Additionally, we can estimate σ by averaging the residual variance of functional entry. This residual variance is obtained for each process by running a one-dimensional smoothing of the diagonal using the previously removed pairs of observations and then averaging the difference between this smoothed diagonal and the previously computed smoothed diagonal (without the noise). For more details on this procedure, we refer the reader to the seminal paper by Yao et al. (2005).

This traditional smoothing approach was chosen for its simplicity and theoretical guarantees Yao et al., 2005. However, several other approaches exist to estimate (cross-)covariance surfaces. Most notably, different smoothing procedures can be used Eilers & Marx, 2003; Wood, 2003. For a univariate process, the fast covariance estimation (FACE) proposed by Xiao et al. (2014) allows estimating the covariance surface with linear complexity (with respect to the number of observations). An alternative FACE method, robust to sparsely observed functional data, was introduced in Xiao et al. (2017). More recently, the method was adapted to multivariate functional data in Li et al. (2020).

3.3. Solving procedure

Solution equations introduced in Proposition 2 along the expression of sample mode optimal vector covariance matrix introduced in Corollary 1 emphasize the use of a block relaxation algorithm outlined in Algorithm 1 to estimate parameters in (6).

https://doi.org/10.1017/psy.2025.10075 Published online by Cambridge University Press

Algorithm 1: Block relaxation procedure for solving (6).

Result: $\hat{\boldsymbol{\Theta}} = \{\hat{\boldsymbol{\Lambda}}, \hat{\boldsymbol{\phi}}, \hat{\mathbf{A}_1}, \ldots, \hat{\mathbf{A}_D}\}$

Initialize $\hat{\boldsymbol{\Theta}}^{(0)} = \{\boldsymbol{\Lambda}^{(0)}, \boldsymbol{\phi}^{(0)}, \mathbf{A}_1^{(0)}, \ldots, \mathbf{A}_D^{(0)}\}$ and $i = 0$

repeat

$$\boldsymbol{\Lambda}^{(i+1)} \leftarrow \iint \mathbf{K}(s)^\top \hat{\boldsymbol{\Sigma}}(s,t)\mathbf{K}(t)\mathrm{d}s\mathrm{d}t$$

$$\boldsymbol{\phi}^{(i+1)} \leftarrow \underset{\boldsymbol{\phi}}{\operatorname{argmin}}\, C(\boldsymbol{\Lambda}^{(i+1)}, \boldsymbol{\phi}, \mathbf{A}_1^{(i)}, \ldots, \mathbf{A}_D^{(i)})$$

for $d = 1, \ldots, D$ **do**

$$\mathbf{A}_d^{(i+1)} \leftarrow \underset{\mathbf{A}_d}{\operatorname{argmin}}\, C(\boldsymbol{\Lambda}^{(i+1)}, \boldsymbol{\phi}^{(i+1)}, \mathbf{A}_1^{(i+1)}, \ldots, \mathbf{A}_{d-1}^{(i+1)}, \mathbf{A}_d, \mathbf{A}_{d+1}^{(i)}, \ldots, \mathbf{A}_D^{(i)})$$

end

$i \leftarrow i + 1$

until $C(\hat{\boldsymbol{\Theta}}^{(i+1)}) - C(\hat{\boldsymbol{\Theta}}^{(i)}) < \varepsilon$;

This algorithm uses all non-mode-d collapsed functional covariance matrices $\boldsymbol{\Sigma}_{[d]}$ as input. It produces estimates $\hat{\boldsymbol{\Lambda}}, \hat{\boldsymbol{\phi}}, \hat{\mathbf{A}_1}, \ldots, \hat{\mathbf{A}_D}$ of the parameters in model (6).

Initialization can significantly influence the algorithm's output, particularly in high-dimensional settings. In practice, random initialization is often chosen for simplicity, but it comes with the cost of running the algorithm multiple times. In our setting, we suggest initializing feature functions and matrices with the ones obtained from a standard PARAFAC decomposition. Experience shows that this simple approach is stable and provides good results. Note that alternative approaches were proposed in Guan (2023) and Han et al. (2023), but were not studied here.

3.4. Sample-mode inference

In this section, we consider that $\mathbf{u}_n = (u_{n1}, \ldots, u_{nR})^\top \in \mathbb{R}^R$. For a given sample $\boldsymbol{\mathcal{Y}}_n$, when dealing with sparse and irregular observations, the estimation of $\mathbf{u}_n$ given by (9) can be unstable (or untracktable), since it requires integrating $\mathbf{x}_n$ on possibly very few observation time points. Furthermore, this estimator can give inconsistent estimates as it does not incorporate the error term $\mathcal{E}$. Therefore, similar to the FPCA PACE-based procedure of Yao et al. (2005), we propose using a conditional approach to predict the factors $\mathbf{u}_n$. In this context, we assume a joint Gaussian distribution on vectors $\mathbf{u}_i$ and residual errors $\mathbf{e}_i$. Then, considering $\mathbf{u}_n \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda})$, we propose using the conditional expectation as our predictor of factors $\mathbf{u}_n$. Denoting $\mathbf{y}_i$ the vector of observations of sample i, and $\mathbf{m}_i$ the vectorized mean tensor of $\mathcal{Y}$, at time points $\mathbf{t}_i$, Proposition 4 gives an expression of the conditional expectation of $\mathbf{u}_n$.

Proposition 4. *Denoting* $\boldsymbol{u}_n = (u_{n1}, \ldots, u_{nR}) \in \mathbb{R}^{1\times R}$ $\boldsymbol{A}_{f,n} = [\boldsymbol{\phi}(t_{n1})^\top, \ldots, \boldsymbol{\phi}(t_{nK_n})^\top]^\top \in \mathbb{R}^{K_n \times R}$, $\boldsymbol{F}_n = \boldsymbol{A}_{(D)} \odot \boldsymbol{A}_{f,n} \in \mathbb{R}^{K_n P \times R}$, *assuming joint Gaussian distribution on* $\boldsymbol{u}$ *and residual errors, a conditional estimator of* $\boldsymbol{u}_n$ *is*

$$\tilde{\boldsymbol{u}}_n = \mathbb{E}[\boldsymbol{u}_n | \boldsymbol{\mathcal{Y}}_n] = \boldsymbol{\Lambda}\boldsymbol{F}_n^\top (\boldsymbol{F}_n \boldsymbol{\Lambda} \boldsymbol{F}_n^\top + \sigma^2 \boldsymbol{I})^{-1} (\boldsymbol{y}_n - \boldsymbol{m}_n). \tag{18}$$

Coupled with estimations $\hat{\boldsymbol{\Lambda}}, \hat{\boldsymbol{\phi}}, \hat{\mathbf{A}_1}, \ldots, \hat{\mathbf{A}_D}$ obtained from Algorithm 1, and estimations $\hat{\sigma}$ and $\hat{\mathbf{m}}_n$ of σ and $\mathbf{m}_n$, obtained as described in Section 3.2, Proposition 4 allows to derive the following empirical Bayes estimator for $\mathbf{u}_n$:

$$\hat{\mathbf{u}}_n = \hat{\boldsymbol{\Lambda}}\hat{\mathbf{F}}_n^\top (\hat{\mathbf{F}}_n \hat{\boldsymbol{\Lambda}} \hat{\mathbf{F}}_n^\top + \hat{\sigma}^2 \mathbf{I})^{-1} (\boldsymbol{y}_n - \hat{\mathbf{m}}_n). \tag{19}$$

https://doi.org/10.1017/psy.2025.10075 Published online by Cambridge University Press

4. Simulations study

Using model (16), we generate $N = 100$ functional tensors of rank $R \in \{3,5,10\}$ on the interval $\mathcal{I} = [0,1]$. Functions ϕ_r are randomly generated using a Fourier basis, with $M = 5$ basis functions, and elements of factor matrices $\mathbf{A}_d$ are sampled from a uniform distribution. As assumed previously, sample mode vectors $\mathbf{u}_n$ are drawn from a distribution $\mathcal{N}(\mathbf{0}_R, \mathbf{\Lambda})$, where $\mathbf{\Lambda}$ is the diagonal matrix $\text{diag}(\{r^2\})_{R \geq r \geq 1}$. The functional tensor is sampled on a regular grid of $\mathcal{I}$ with size $K = 30$. The resulting array is then sparsified by removing a proportion $s \in \{0.0, 0.2, 0.5, 0.8\}$ of observations. We consider two settings for tensor dimensions: one of order $D = 2$, with $p_1 = 10$, and one of order $D = 3$ with $p_1 = p_2 = 5$ (in the Appendix). Observation tensors are contaminated with i.i.d. measurement errors of variance $\sigma^2 = 1$, and we adjust the signal-to-noise ratio (SNR) by multiplying the original tensors by a constant c_{SNR}. Three settings are considered: $\text{SNR} \in \{0.5, 1, 2\}$.

We compare our approach to the native PARAFAC decomposition (without smoothing and probabilistic modeling), the functional tensor SVD (FTSVD) Han et al., 2023, and the multivariate FPCA (MFPCA) Happ & Greven, 2018. Due to software and package issues, we were unable to include other functional approaches in our comparison, particularly that of Guan (2023). Two comparison tasks are considered: a reconstruction task and a parameter recovery task. Since the FTSVD does not handle missing values, it is not considered when $s \neq 0$. MFPCA is only considered for the reconstruction task as it is not based on the F-PARAFAC decomposition model. Furthermore, since the method only deals with multivariate functional data, we apply the mode on the vectorized tensor. To obtain an approximation of the original tensor, we then tensorize back the approximated functional vector. For the reconstruction task, the metric used to compare methods is the root mean squared error (RMSE) between original tensors $\mathcal{X}_n$ and approximations $\hat{\mathcal{X}}_i$. For the parameter recovery task, we compare the true functions ϕ_r and estimated functions $\hat{\phi}_r$ using the maximum principal angle Bjorck & Golub, 1973. MFPCA is carried out using the `R` package `MFPCA`, the PARAFAC decomposition is obtained using the `multiway` package, and the FTSVD is performed using the code from the associated paper. Simulations were run on a standard laptop.

As observed in Figure 2, we note that our approach outperforms all methods for $\text{SNR} \in \{0.5, 1\}$ for the reconstruction task. For SNR = 2, the naive PARAFAC decomposition seems to be the best approach. However, as seen in the parameter estimation task (Figure 3), smoothing approaches are more accurate for retrieving functions ϕ_r in all cases. Thanks to the smoothing procedure, our approach is particularly efficient when the SNR is low and the sparsity is high. The FTSVD performs well in retrieving the true feature functions but seems to be slightly worse than the PARAFAC decomposition at reconstructing the original tensor. The unfavorable setting may explain this observation, as we do not assume any additional regularity conditions on the space on which lie functions ϕ_r. Nevertheless, the good performances, and, especially, the ability of our approach to handle settings with 80% of missing values, as observed in the figures, come with an overall increased computational cost, as other approaches ran much faster most of the time.

In the higher-order setting, described in Appendix B.1, we observe overall the same results. One important difference is that all tensor methods outperform the MFPCA in most cases, as the vectorization leads to the loss of the tensor structure. In the misspecified model setting, described in Appendix B.2, we observe that the LF-PARAFAC model still performs well compared to the PARAFAC and FTSVD approaches in all settings and to MFPCA in the high-sparsity setting only.

5. Application study

To help characterize cognitive decline among AD and CN patients and illustrate our method in the psychometrics setting, we propose to apply the introduced method to various neurocognitive markers measured over several years in the ADNI study. In this context, we consider six different cognitive markers observed on 518 healthy controls with normal cognition and 370 patients diagnosed with AD, resulting in $N = 888$ total patients. Although some subjects were observed for more than 10 years, we propose first to restrict ourselves to the first two years of the study to ensure a uniform distribution

https://doi.org/10.1017/psy.2025.10075 Published online by Cambridge University Press

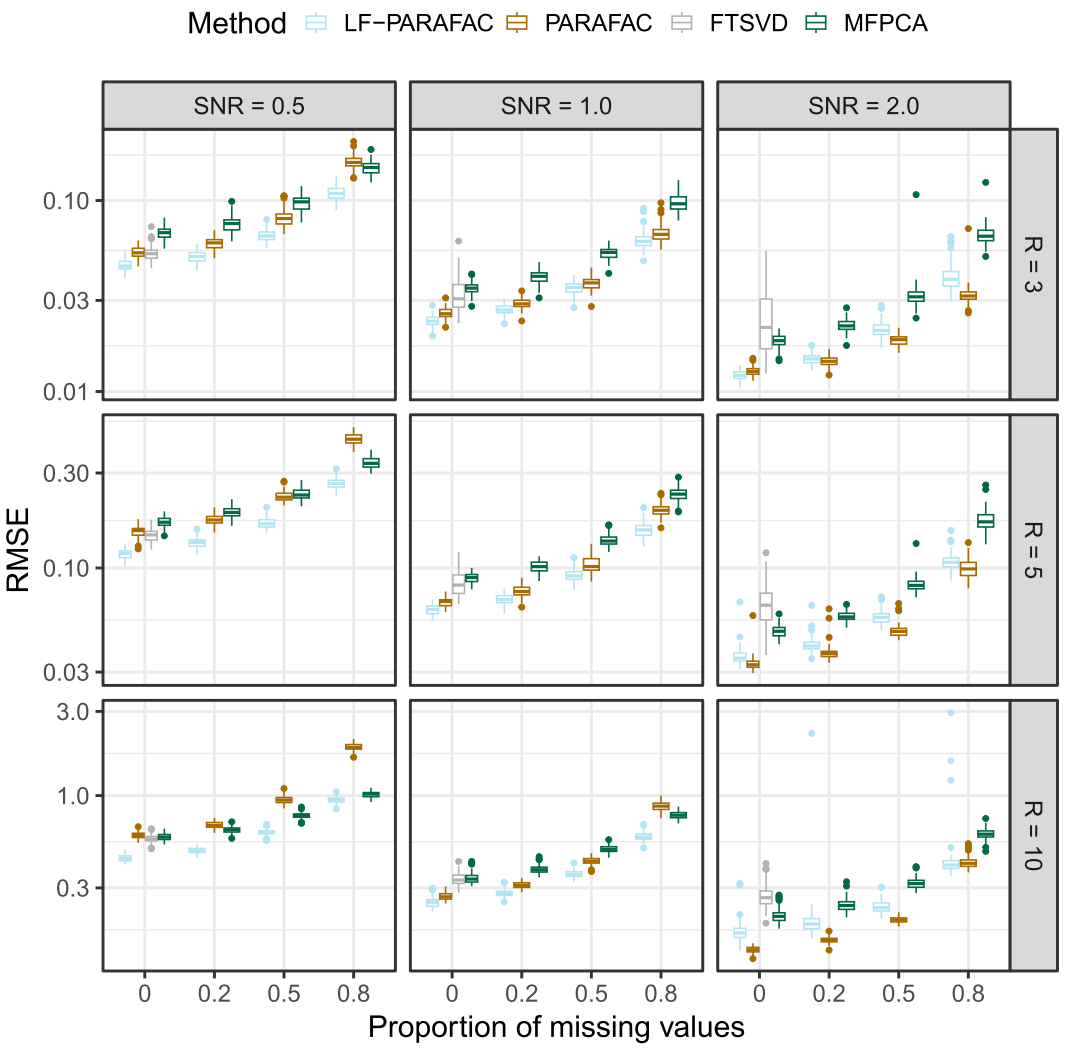


Figure 2. Root mean squared error (RMSE) of order-2 reconstructed functional tensors. Comparing latent functional PARAFAC (LF-PARAFAC), standard PARAFAC (PARAFAC), functional tensor singular value decomposition (FTSVD), and multivariate functional principal component analysis (MFPCA).

of time points. Neurocognitive tests are often used by physicians to help assess the cognitive abilities of patients. In the context of AD and, more generally, dementia, they play an important role in the diagnosis and follow-up. Most neurocognitive tests consist of more or less simple questions asked to patients regarding their everyday lives or to complete a task involving specific cognitive abilities. The six cognitive markers considered here are the Alzheimer's Disease Assessment Scale (ADAS), the Mini-Mental State Examination (MMSE), the Rey Auditory Verbal Learning Test (RAVLT), the Functional Activities Questionnaire (FAQ), the Montreal Cognitive Assessment (MOCA), and the logical memory delayed recall score (LDEL). Scores were normalized for improved readability. Markers were observed at different time points across individuals, resulting in a heavily irregular sampling scheme with 262 unique time points overall. Only subjects with at least two observations were considered, resulting in an average number of observations per subject of 3. Trajectories are represented in Figure 4.

The LF-PARAFAC decomposition was applied to the resulting random functional "tensor", which lies in $L^2(\mathcal{I})^6$, with $\mathcal{I} = [0,2]$ and can be seen as a tensor with dimensions subject × time × marker. As seen

https://doi.org/10.1017/psy.2025.10075 Published online by Cambridge University Press

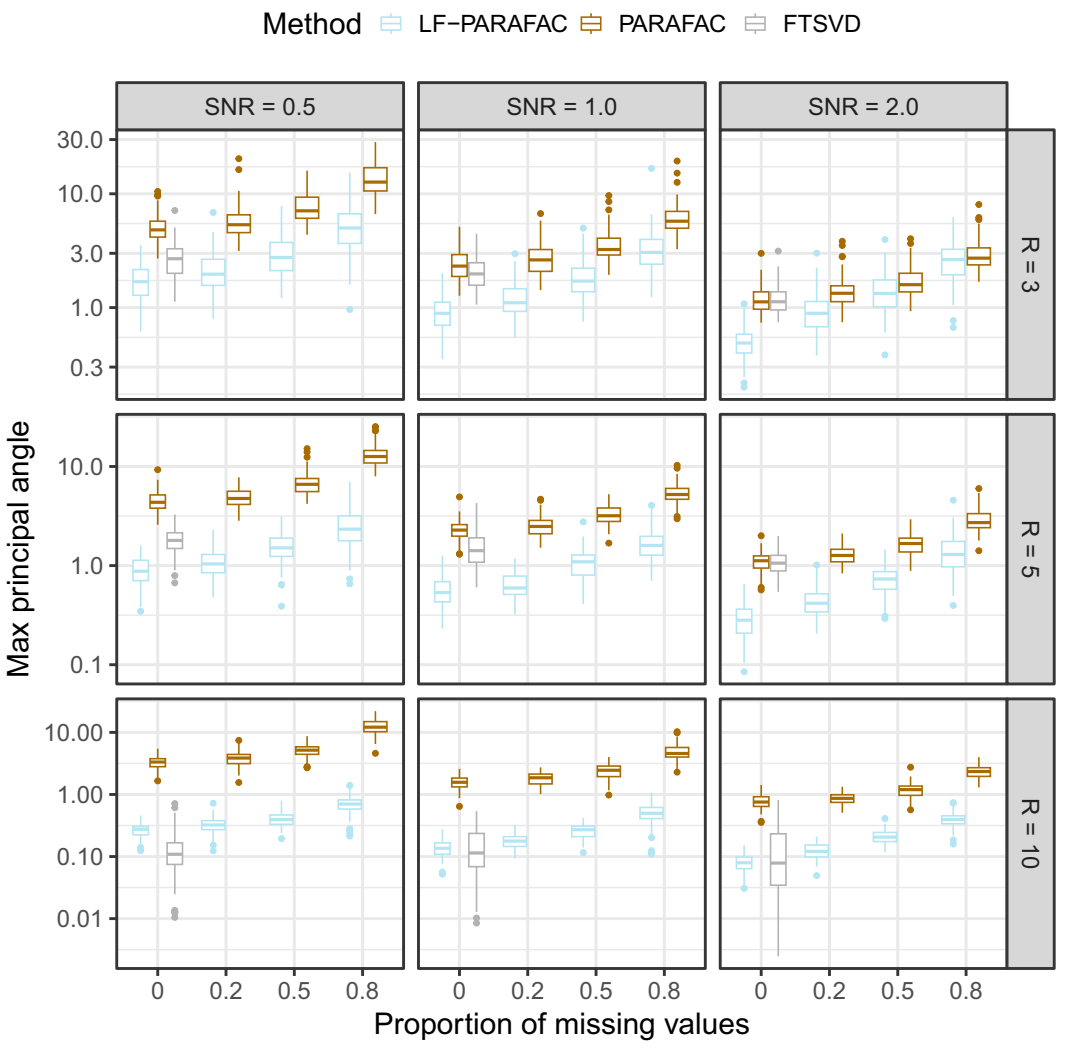


Figure 3. Max principal angle between Φ and $\hat{\Phi}$. Comparing latent functional PARAFAC (LF-PARAFAC), standard PARAFAC (PARAFAC), and functional tensor singular value decomposition (FTSVD).

in Figure 5, the rank $R = 4$ of the decomposition was chosen since the mean squared error between the predicted tensor and the observed tensor did not appear to provide a significantly better fit compared to $R > 4$. The smoothing procedure's bandwidths were manually set to 1.

The four functions and vectors retrieved from the decomposition are represented in Figure 6 along with population mode scores. Our approach took 2.4 seconds on average to provide estimates using an Apple M4 Pro chip. Additionally, note that with $R = 4$, we obtained approximately similar results across different initializations. Since it was not the case for larger values of R, it suggests that our selected rank was probably close to the maximum rank under which the identifiability criterion of Proposition 3 is met for the true underlying tensor.

Figures show that the LF-PARAFAC decomposition successfully captured the inherent information of the random functional tensor. Notably, as expected, the method has retrieved smooth continuous vectors for the functional mode, describing predominant tendencies among cognitive scores.When observed with feature mode loadings, it appears that the first rank (first row), which has the highest

https://doi.org/10.1017/psy.2025.10075 Published online by Cambridge University Press

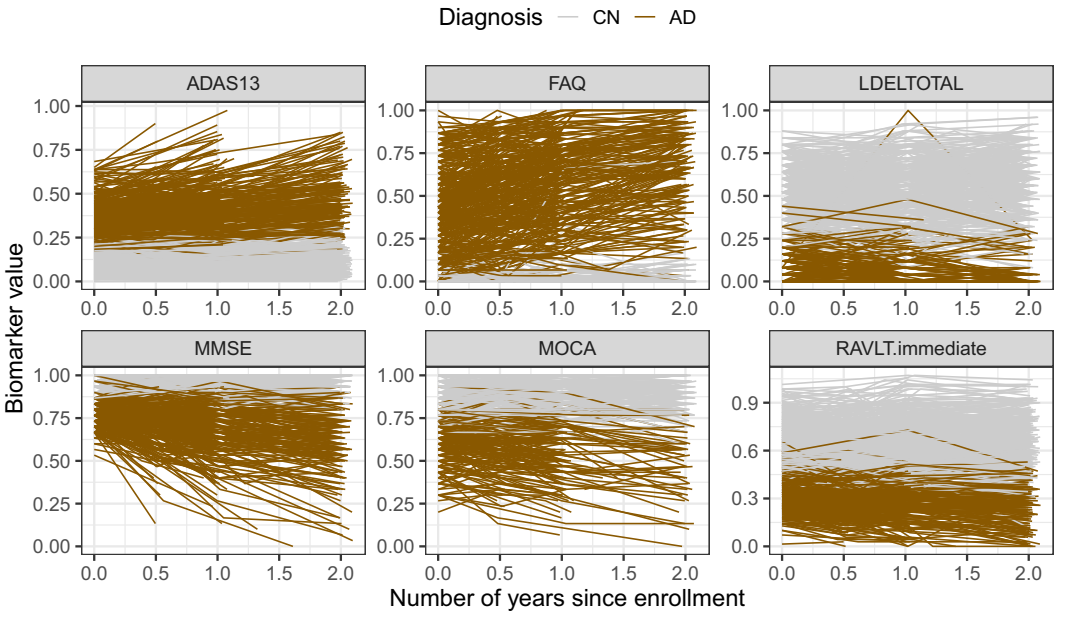


Figure 4. Trajectories of six cognitive scores measured over 10 years. Colors correspond to baseline diagnosis: cognitively normal (CN) and Alzheimer's disease (AD).

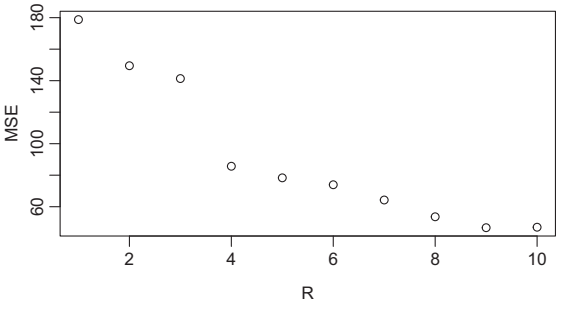


Figure 5. Mean squared error between predicted tensor and observed tensor in the ADNI application using various values for *R*.

associated sample-mode score variance, captures the average level of neurocognitive abilities: the higher the associated sample-mode score is, the better the cognitive abilities are, implying, as captured by the feature-mode loadings, high LDEL, MMSE, MOCA, and RAVLT, along with low ADAS and FAQ. This trend is also asserted by the distribution of the associated sample-mode scores: AD patients, who typically show poor cognitive abilities overall, have score values much lower than CN patients' scores. The next rank (second row) seems to describe an increase of cognitive measures over the first year followed by a decrease during the second year. The associated marker-mode vector tells us that LDEL particularly follows this trend. ADAS and FAQ scores appear not to be involved for this rank. The third rank (third row), which is still associated with a high variance sample-mode score, describes an increasing trend, starting at the end of the first year. This trend seems to describe a decline (or improvement, depending on the sign) in cognitive abilities over the second year. Finally, the fourth rank appears to capture a similar trend as the third rank, although the increase is softer. The feature-mode vector suggests that it describes a trend shared by both FAQ and LDEL scores, and in the opposite way for MMSE.

https://doi.org/10.1017/psy.2025.10075 Published online by Cambridge University Press

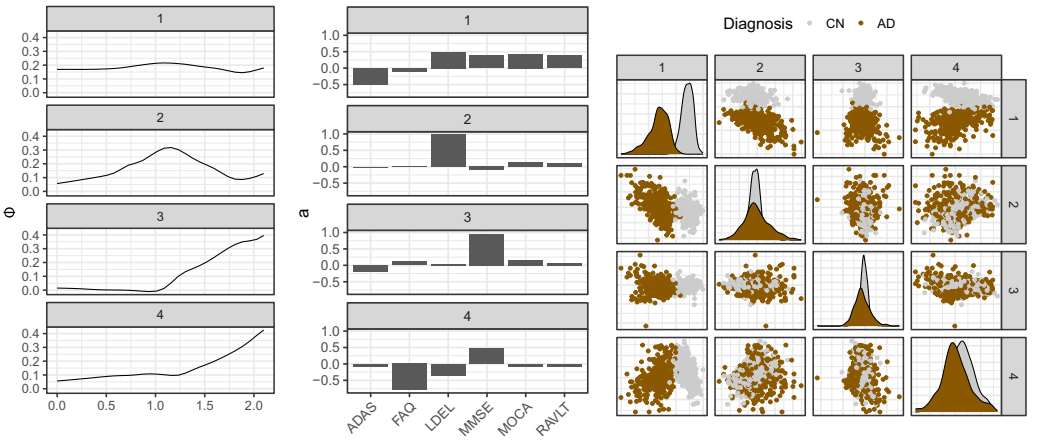


Figure 6. (left) Functions and vectors retrieved by a rank $R = 4$ LF-PARAFAC decomposition. (right) Population mode scores colored by diagnosis at baseline: cognitively normal (CN) and Alzheimer's disease (AD).

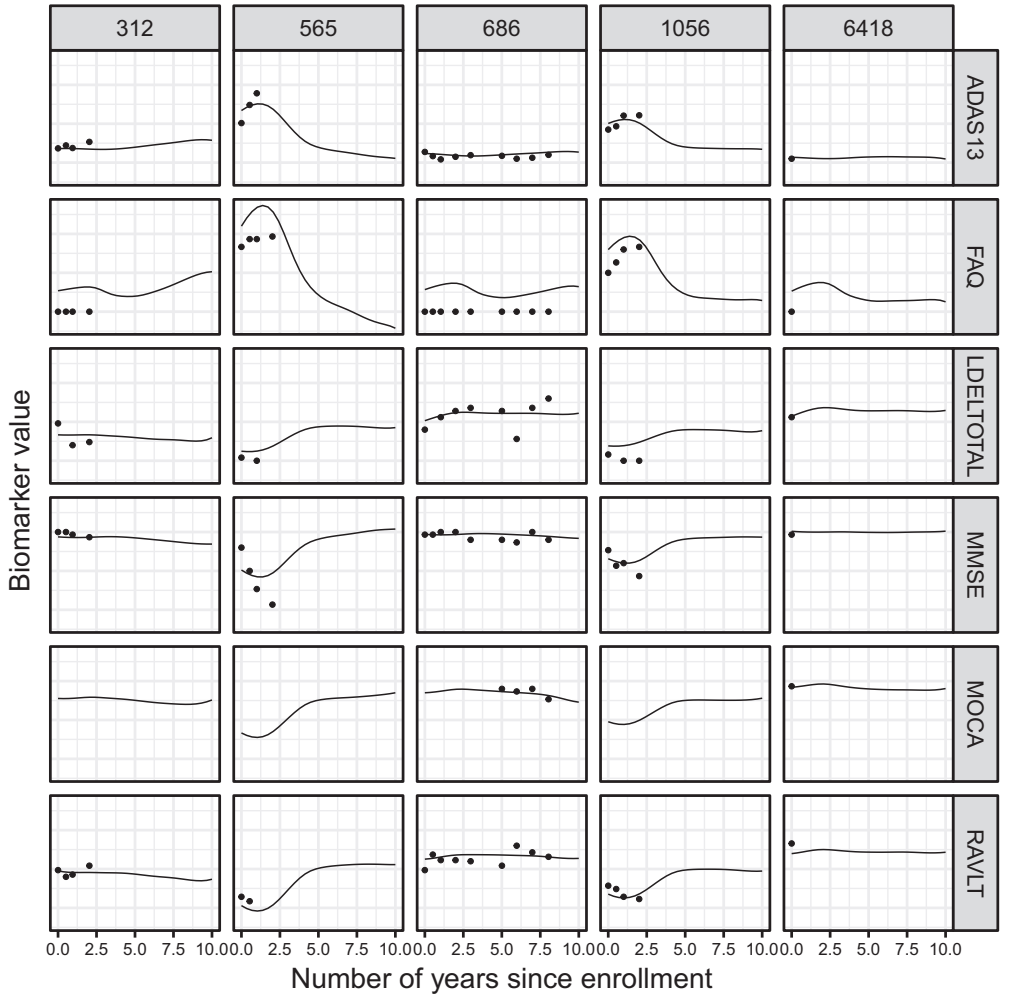


Figure 7. Observations as points and reconstructed trajectories as solid lines estimated using the latent functional PARAFAC decomposition.

https://doi.org/10.1017/psy.2025.10075 Published online by Cambridge University Press

Further investigation (Appendix B.3) shows that the first component is indeed associated with a bad prognosis: subjects in the first decile of the first sample-mode scores overall have low cognitive abilities. For the second component, we note that subjects with low sample-mode scores have worse cognitive abilities, particularly for LDEL (which explains the high value of the associated loading). As mentioned earlier, the third component appears to be associated with a slow deterioration of cognitive abilities.

Finally, we propose also to visualize the reconstruction capacity of the LF-PARAFAC decomposition using the parameters retrieved previously on a subset of 5 random subjects. For this analysis, we used the data retrieved over 10 years. Results are displayed in Figure 7. Overall, we observe that predicted curves are close to observations, except for FAQ, for which predicted trajectories seem to be inaccurate. An interesting property of using the LF-PARAFAC decomposition to reconstruct such tensor-structured data is that it can allow the prediction of trajectories of features for which no observations are made, as observed for trajectories of subjects 312, 565, and 1,056 for the MOCA score.

6. Discussion

We introduced a new approach to assist psychometric researchers by efficiently decomposing a functional tensor with an underlying random structure. Thanks to the (cross-)covariance formulation of the solving procedure, the method can be applied to a broad range of longitudinal settings, notably to sparse and irregular sampling schemes. Simulation results show that our approach retrieves feature functions closer to the true smooth functions, and the approximation made from those functions is indeed better.

The developments introduced in this article pave the way for a broad range of extensions, the more prominent of which is tensor regression, i.e., regression using tensor-structured covariates. A similar methodology to Zhou et al. (2013) is currently being investigated to perform linear functional tensor regression. Another interesting extension of our work would be to consider that not one but multiple modes have a smooth structure. This situation is often encountered in spatial statistics. As evoked earlier, various tensor decomposition frameworks have been introduced in such settings. However, those methods were developed from a signal-processing setting rather than a statistical setting. Therefore, no probabilistic modeling is made in the decomposition, and the random structure is not considered.

The most notable extension of our approach would be allowing to control for external factors, especially, confounders. This problem is particularly relevant in a longitudinal framework, as the time dimension is rarely easily comparable between subjects. In our setting for instance, subjects may have entered the study at different stages of the disease and, more importantly, at different ages. This information, which is strongly associated with the disease (aged patients are more likely to have AD), could act as a major confounder, explaining most of the variation observed in our sample mode vectors, and therefore significantly changing the interpretation. In this context, integrating this external information, similar to Guan (2023) and Lock and Li (2018), appears as a straightforward extension of our approach that would definitely improve our model's capacity to represent the data and avoid misinterpretation.

Additionally, our work emphasizes the need for multivariate high-dimensional FACE methods. Indeed, one major drawback of our approach is its computational complexity. Currently, the method requires estimating the (cross-)covariance surface between each pair of functional entries of the tensor, a number that grows exponentially with the order of the tensor. Assuming local linear smoothing is used to estimate (cross-)covariance surfaces, which has a complexity of $O(IJ^2)$ denoting I the number of samples and J the number of observations for each sample (to simplify, in a sparse and regular sampling setting), the overall complexity of multidimensional tensor covariance estimation is $O(IJ^2P^2)$, where we recall that $P = \prod_d p_d$. As already mentioned, other covariance estimation strategies such as FACE Xiao et al., 2014, which has a complexity $O(I^2J + I^3)$ (assuming the number of knots c in the spline modeling is $O(I)$), could be considered to speed up the process. Notably, this would allow us to more closely match the complexity of the method introduced in Guan (2023), which appears to be linear in J. Nevertheless,

https://doi.org/10.1017/psy.2025.10075 Published online by Cambridge University Press

the multidimensional structure of the data still prevents considering highly multidimensional data in practice due to the P^2 term in the complexity, and, similar to before, in some settings, the approach of Guan (2023) could be a better alternative due to its linear complexity in P. To solve this issue, we currently investigate a new estimation approach which considers a separable Kronecker covariance structure Hoff, 2011 on the covariance function $\mathbf{\Sigma}$. Under simple assumptions, assuming such a structure can significantly reduce the number of computations.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/psy.2025.10075.

Funding statement. Data collection and sharing for the Alzheimer's Disease Neuroimaging Initiative (ADNI) is funded by the National Institute on Aging (National Institutes of Health Grant U19 AG024904). The grantee organization is the Northern California Institute for Research and Education. In the past, ADNI has also received funding from the National Institute of Biomedical Imaging and Bioengineering, the Canadian Institutes of Health Research, and private sector contributions through the Foundation for the National Institutes of Health (FNIH), including generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd. and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics.

Competing interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

Anderson, T. W., & Rubin, H. (1956). Statistical inference in factor analysis. *Proceedings of the third Berkeley symposium on mathematical statistics and probability*, *5*, 111–150.

Bartholomew, D. J., Knott, M., & Moustaki, I. (2011). *Latent variable models and factor analysis: A unified approach*. John Wiley & Sons.

Bjorck, A., & Golub, G. H. (1973). Numerical methods for computing angles between linear subspaces. *Mathematics of Computation*, *27*(123), 579. https://doi.org/10.2307/2005662

Carroll, J. D., & Chang, J.-J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition. *Psychometrika*, *35*(3), 283–319.

Chang, J., He, J., Yang, L., & Yao, Q. (2023). Modelling matrix time series via a tensor CP-decomposition. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *85*(1), 127–148. https://doi.org/10.1093/jrsssb/qkac011

Chen, R., Han, Y., Li, Z., Xiao, H., Yang, D., & Yu, R. (2022). Analysis of tensor time series: Tensorts. *Journal of Statistical Software*, *323*, 1–41.

Di, C.-Z., Crainiceanu, C. M., Caffo, B. S., & Punjabi, N. M. (2009). Multilevel functional principal component analysis. *The Annals of Applied Statistics*, *3*(1). https://doi.org/10.1214/08-aoas206

Eilers, P. H., & Marx, B. D. (2003). Multivariate calibration with temperature interaction using two-dimensional penalized signal regression. *Chemometrics and Intelligent Laboratory Systems*, *66*(2), 159–174.

Fan, J., & Gijbels, I. (2018). *Local polynomial modelling and its applications*. Routledge.

Giordani, P., Rocci, R., & Bove, G. (2020). Factor uniqueness of the structural parafac model. *Psychometrika*, *85*(3), 555–574. https://doi.org/10.1007/s11336-020-09715-4

Guan, L. (2023). Smooth and probabilistic PARAFAC model with auxiliary covariates. *Journal of Computational and Graphical Statistics*, 1–13. https://doi.org/10.1080/10618600.2023.2257783

Han, R., Shi, P., & Zhang, A. R. (2023). Guaranteed functional tensor singular value decomposition. *Journal of the American Statistical Association*, 1–13. https://doi.org/10.1080/01621459.2022.2153689

Happ, C., & Greven, S. (2018). Multivariate functional principal component analysis for data observed on different (dimensional) domains. *Journal of the American Statistical Association*, *113*(522), 649–659.

Harshman, R. A. (1972). PARAFAC2: Mathematical and technical notes. *UCLA Working Papers in Phonetics*, *22*, 30–44.

Harshman, R. A. (1970). Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis. *UCLA working papers in phonetics*, *16*(1), 84.

Hoff, P. D. (2011). Separable covariance arrays via the tucker product, with applications to multivariate relational data. *Bayesian Analysis*, *6*(2). https://doi.org/10.1214/11-ba606

https://doi.org/10.1017/psy.2025.10075 Published online by Cambridge University Press

Imaizumi, M., & Hayashi, K. (2017). Tensor decomposition with smoothness, in International Conference on Machine Learning, *70*, 1597–1606.

Kiers, H. A. L., ten Berge, J. M. F., & Bro, R. (1999). PARAFAC2—Part I. A direct fitting algorithm for the PARAFAC2 model. *Journal of Chemometrics*, *13*(3–4), 275–294. https://doi.org/10.1002/(SICI)1099-128X(199905/08)13:3/4<275::AID-CEM543>3.0.CO;2-B

Lange, K. (2013). *Optimization*. Springer. https://doi.org/10.1007/978-1-4614-5838-8

Li, C., Xiao, L., & Luo, S. (2020). Fast covariance estimation for multivariate sparse functional data. *Stat*, *9*(1). https://doi.org/10.1002/sta4.245

Li, G., Yang, D., Nobel, A. B., & Shen, H. (2016). Supervised singular value decomposition and its asymptotic properties. *Journal of Multivariate Analysis*, *146*, 7–17. https://doi.org/10.1016/j.jmva.2015.02.016

Lock, E. F. (2018). Tensor-on-tensor regression. *Journal of Computational and Graphical Statistics*, *27*(3), 638–647. https://doi.org/10.1080/10618600.2017.1401544

Lock, E. F., & Li, G. (2018). Supervised multiway factorization. *Electronic Journal of Statistics*, *12*(1). https://doi.org/10.1214/18-ejs1421

Morris, J. S. (2015). Functional regression. *Annual Review of Statistics and its Application*, *2*(1), 321–359.

Ramsay, J. O., & Silverman, B. W. (2006). *Functional data analysis*. Springer.

Reis, M. M., & Ferreira, M. M. C. (2002). PARAFAC with splines: A case study. *Journal of Chemometrics*, *16*(8–10), 444–450. https://doi.org/10.1002/cem.749

Rice, J. A., & Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society: Series B (Methodological)*, *53*(1), 233–243. https://doi.org/10.1111/j.2517-6161.1991.tb01821.x.

Sidiropoulos, N. D., & Bro, R. (2000). On the uniqueness of multilinear decomposition of N-way arrays. *Journal of Chemometrics*, *14*(3), 229–239. https://doi.org/10.1002/1099-128X(200005/06)14:3<229::AID-CEM587>3.0.CO;2-N

Sidiropoulos, N. D., De Lathauwer, L., Fu, X., Huang, K., Papalexakis, E. E., & Faloutsos, C. (2017). Tensor decomposition for signal processing and machine learning. *IEEE Transactions on Signal Processing*, *65*(13), 3551–3582. https://doi.org/10.1109/tsp.2017.2690524

Timmerman, M. E., & Kiers, H. A. (2002). Three-way component analysis with smoothness constraints. *Computational Statistics & Data Analysis*, *40*(3), 447–470. https://doi.org/10.1016/s0167-9473(02)00059-2

Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, *31*(3), 279–311.

Wang, D., Zheng, Y., & Li, G. (2024). High-dimensional low-rank tensor autoregressive time series modeling. *Journal of Econometrics*, *238*(1), 105544. https://doi.org/10.1016/j.jeconom.2023.105544

Wang, J.-L., Chiou, J.-M., & Müller, H.-G. (2016). Functional data analysis. *Annual Review of Statistics and Its Application*, *3*(1), 257–295.

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *65*(1), 95–114.

Xiao, L., Zipunnikov, V., Ruppert, D., & Crainiceanu, C. (2014). Fast covariance estimation for high-dimensional functional data. *Statistics and Computing*, *26*(1–2), 409–421. https://doi.org/10.1007/s11222-014-9485-x

Xiao, L., Li, C., Checkley, W., & Crainiceanu, C. (2017). Fast covariance estimation for sparse functional data. *Statistics and Computing*, *28*(3), 511–522. https://doi.org/10.1007/s11222-017-9744-8

Yao, F., Müller, H. G., & Wang, J. L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, *100*(470), 577–590.

Yokota, T., Zhao, Q., & Cichocki, A. (2016). Smooth parafac decomposition for tensor completion. *IEEE Transactions on Signal Processing*, *64*(20), 5423–5436. https://doi.org/10.1109/tsp.2016.2586759

Zhou, H., Li, L., & Zhu, H. (2013). Tensor regression with applications in neuroimaging data analysis. *Journal of the American Statistical Association*, *108*(502), 540–552. https://doi.org/10.1080/01621459.2013.776499

https://doi.org/10.1017/psy.2025.10075 Published online by Cambridge University Press

A. Proofs

Proposition 1. To obtain Ψ^*, we look for solutions of

$$\min_{\mathbf{u}} \|\mathcal{X} - [\![\mathbf{u};\boldsymbol{\phi};\mathbf{A}_1;\ldots;\mathbf{A}_D]\!]\|^2$$

for fixed $\boldsymbol{\phi},\mathbf{A}_1,\ldots,\mathbf{A}_D$ and a given $\mathcal{X}$. Developing the criterion, we get

$$\begin{aligned}\|\mathcal{X} - [\![\mathbf{u};\boldsymbol{\phi};\mathbf{A}_1;\ldots;\mathbf{A}_D]\!]\|^2 &= \int \|\mathbf{x}(t) - (\mathbf{A}_D \odot \cdots \odot \mathbf{A}_1 \odot \boldsymbol{\phi}(t))\mathbf{u}^\top\|^2 \\ &= \int \|\mathbf{x}(t) - (\mathbf{A}_{(D)} \odot \boldsymbol{\phi}(t))\mathbf{u}^\top\|^2.\end{aligned}$$

Deriving this expression with respect to U and setting the derivative to 0 leads to

$$-\int \mathbf{x}(t)^\top(\mathbf{A}_{(D)} \odot \boldsymbol{\phi}(t)) + \mathbf{u}^*\left(\int (\mathbf{A}_{(D)} \odot \boldsymbol{\phi}(t))^\top(\mathbf{A}_{(D)} \odot \boldsymbol{\phi}(t))\right) = 0.$$

Which, under mild assumptions, leads to

$$\mathbf{u}^* = \left(\int \mathbf{x}(t)^\top(\mathbf{A}_{(D)} \odot \boldsymbol{\phi}(t))\right)\left(\int (\mathbf{A}_{(D)} \odot \boldsymbol{\phi}(t))^\top(\mathbf{A}_{(D)} \odot \boldsymbol{\phi}(t))\right)^{-1}.$$

Corollary 1. From (9), we have

$$\begin{aligned}\mathbb{E}[\mathbf{u}^{*\top}\mathbf{u}^*] &= \mathbb{E}[\left(\int \mathbf{x}(s)^\top\mathbf{K}(s)\right)^\top\left(\int \mathbf{x}(t)^\top\mathbf{K}(t)\right)] \\ &= \int\int \mathbf{K}(s)^\top\mathbb{E}[\mathbf{x}(s)\mathbf{x}(t)^\top]\mathbf{K}(t) \\ &= \int\int \mathbf{K}(s)^\top\boldsymbol{\Sigma}_{[f]}(s,t)\mathbf{K}(t).\end{aligned}$$

Lemma A1. *The objective function of optimization problem (10) only depends on covariance $\boldsymbol{\Lambda}^*$ and other parameters via an operator C:*

$$(\boldsymbol{\phi}^*,\boldsymbol{A}_1^*,\ldots,\boldsymbol{A}_D^*) = \underset{\boldsymbol{\phi},\boldsymbol{A}_1,\ldots,\boldsymbol{A}_D}{argmin}\ C(\boldsymbol{\Lambda}^*,\boldsymbol{\phi},\boldsymbol{A}_1,\ldots,\boldsymbol{A}_D).$$

Proof. We have

$$\begin{aligned}\mathbb{E}[\|\mathcal{X} - [\![\mathbf{u};\boldsymbol{\phi};\mathbf{A}_1;\ldots;\mathbf{A}_D]\!]\|^2] &= \int \mathbb{E}[\|\mathbf{x}(t) - (\mathbf{A}_D \odot \cdots \odot \mathbf{A}_1 \odot \mathbf{u}^*)\boldsymbol{\phi}(t)^\top\|^2] \\ &= \int \mathbb{E}[\|\mathbf{x}(t) - (\mathbf{A}_{(D)} \odot \mathbf{u}^*)\boldsymbol{\phi}(t)^\top\|^2] \\ &= \int \mathbb{E}[\|\mathbf{x}(t)\|^2] - 2\int \mathbb{E}[\mathbf{x}(t)^\top(\mathbf{A}_{(D)} \odot \mathbf{u}^*)\boldsymbol{\phi}(t)^\top] \\ &\quad + \int \mathbb{E}[\|(\mathbf{A}_{(D)} \odot \mathbf{u}^*)\boldsymbol{\phi}(t)^\top\|^2].\end{aligned}$$

In this expression, the first term does not depend on the parameters and can thus be removed for optimization. Next, using the property $(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\odot\mathbf{D}) = (\mathbf{AC})\odot(\mathbf{BD})$, we can write

$$\begin{aligned}\int \mathbb{E}[\mathbf{x}(t)^\top(\mathbf{A}_{(D)} \odot \mathbf{u}^*)\boldsymbol{\phi}(t)^\top] &= \int \mathbb{E}[\mathbf{x}(t)^\top(\mathbf{A}_{(D)} \odot \int \mathbf{x}(t)^\top\mathbf{K}(s))\boldsymbol{\phi}(t)^\top] \\ &= \int\int \mathbb{E}[\mathbf{x}(t)^\top(\mathbf{A}_{(D)} \odot \mathbf{x}(t)^\top\mathbf{K}(s))\boldsymbol{\phi}(t)^\top] \\ &= \int\int \mathbb{E}[\mathbf{x}(t)^\top(\mathbf{I}_{p_{(D)}} \otimes \mathbf{x}(t)^\top)](\mathbf{A}_{(D)} \odot \mathbf{K}(s))\boldsymbol{\phi}(t)^\top \\ &= \int\int \boldsymbol{\Sigma}_{[1]}(s,t)(\mathbf{A}_{(D)} \odot \mathbf{K}(s))\boldsymbol{\phi}(t)^\top\end{aligned}$$

and

$$\begin{aligned}\int \mathbb{E}[\|(\mathbf{A}_{(D)} \odot \mathbf{u}^*)\boldsymbol{\phi}(t)^\top\|^2] &= \int \mathbb{E}[\boldsymbol{\phi}(t)(\mathbf{A}_{(D)} \odot \mathbf{u}^*)^\top(\mathbf{A}_{(D)} \odot \mathbf{u}^*)\boldsymbol{\phi}(t)^\top] \\ &= \int \mathbb{E}[\boldsymbol{\phi}(t)(\mathbf{A}_{(D)}^\top\mathbf{A}_{(D)} * \mathbf{u}^{*\top}\mathbf{u}^*)\boldsymbol{\phi}(t)^\top] \\ &= \int \boldsymbol{\phi}(t)(\mathbf{A}_{(D)}^\top\mathbf{A}_{(D)} * \mathbb{E}[\mathbf{u}^{*\top}\mathbf{u}^*])\boldsymbol{\phi}(t)^\top \\ &= \int \boldsymbol{\phi}(t)(\mathbf{A}_{(D)}^\top\mathbf{A}_{(D)} * \boldsymbol{\Lambda}^*)\boldsymbol{\phi}(t)^\top,\end{aligned}$$

https://doi.org/10.1017/psy.2025.10075 Published online by Cambridge University Press

which leads to

$$C(\boldsymbol{\Lambda}^*,\boldsymbol{\phi},\mathbf{A}_1,\ldots,\mathbf{A}_D) = \int \boldsymbol{\phi}(t)(\mathbf{A}_{(D)}^\top \mathbf{A}_{(D)} * \boldsymbol{\Lambda}^*)\boldsymbol{\phi}(t)^\top - 2\int\int \boldsymbol{\Sigma}_{[f]}(s,t)(\mathbf{A}_{(D)} \odot \mathbf{K}(s))\boldsymbol{\phi}(t)^\top. \tag{A1}$$

Alternatively, for any $d \in [D]$, we can rewrite (A1) as

$$C(\boldsymbol{\Lambda}^*,\boldsymbol{\phi},\mathbf{A}_1,\ldots,\mathbf{A}_D) = \int \mathbf{A}_d(\mathbf{A}_{(-d),f}(t)^\top \mathbf{A}_{(-d),f}(t) * \boldsymbol{\Lambda}^*)\mathbf{A}_d^\top - 2\int\int \boldsymbol{\Sigma}_{[d]}(s,t)(\mathbf{A}_{(-d),f}(t) \odot \mathbf{K}(s))\mathbf{A}_d^\top.$$

□

In the following, we denote $C_f : \boldsymbol{\phi} \mapsto C(\boldsymbol{\Lambda}^*,\boldsymbol{\phi},\mathbf{A}_1,\ldots,\mathbf{A}_D)$ for any fixed $\boldsymbol{\Lambda}^*,\mathbf{A}_1,\ldots,\mathbf{A}_D$, and similarly $C_d : \mathbf{A}_d \mapsto C(\boldsymbol{\Lambda}^*,\boldsymbol{\phi},\mathbf{A}_1,\ldots,\mathbf{A}_D)$ for fixed $\boldsymbol{\Lambda}^*,\boldsymbol{\phi},\mathbf{A}_1,\ldots,\mathbf{A}_{d-1},\mathbf{A}_{d+1},\ldots,\mathbf{A}_D$.

Lemma A2. *The operator C_f is Gâteau-differentiable and at any $\boldsymbol{\phi} \in \mathcal{H}^{1\times R}$ and $s \in \mathcal{I}$, the derivative is given by*

$$C_f'(\boldsymbol{\phi})(s) = 2\boldsymbol{\phi}(s)(\boldsymbol{A}_{(D)}^\top \boldsymbol{A}_{(D)} * \boldsymbol{\Lambda}^*) - 2\int \boldsymbol{\Sigma}_{[f]}(s,t)(\boldsymbol{A}_{(D)} \odot \boldsymbol{K}(t)).$$

Proof. Let $\boldsymbol{\phi} \in \mathcal{H}^{1\times R}$. Let's show that there is a continuous (bounded) linear operator $C_f'(\boldsymbol{\phi})$ such that

$$\forall \mathbf{H} \in \mathcal{H}^{1\times R}, \langle C_f'(\boldsymbol{\phi}),\mathbf{H}\rangle = \lim_{\substack{\alpha\to 0\\ \alpha>0}} \frac{C_f(\boldsymbol{\phi}+\alpha\mathbf{H}) - C_f(\boldsymbol{\phi})}{\alpha}. \tag{A2}$$

We have

$$C_f(\boldsymbol{\phi}+\alpha\mathbf{H}) = C_f(\boldsymbol{\phi}) + 2\alpha\int \boldsymbol{\phi}(t)(\mathbf{A}_{(D)}^\top \mathbf{A}_{(D)} * \boldsymbol{\Lambda}^*)\mathbf{H}(t)^\top - 2\alpha\int\int \boldsymbol{\Sigma}_{[f]}(s,t)(\mathbf{A}_{(D)} \odot \mathbf{K}(s))\mathbf{H}(t)^\top + \alpha^2\int \mathbf{H}(t)(\mathbf{A}_{(D)}^\top \mathbf{A}_{(D)} * \boldsymbol{\Lambda}^*)\mathbf{H}(t)^\top.$$

Therefore,

$$\lim_{\substack{\alpha\to 0\\ \alpha>0}} \frac{C_f(\boldsymbol{\phi}+\alpha\mathbf{H}) - C_f(\boldsymbol{\phi})}{\alpha} = 2\int \boldsymbol{\phi}(t)(\mathbf{A}_{(D)}^\top \mathbf{A}_{(D)} * \boldsymbol{\Lambda}^*)\mathbf{H}(t)^\top - 2\int\int \boldsymbol{\Sigma}_{[f]}(s,t)(\mathbf{A}_{(D)} \odot \mathbf{K}(s))\mathbf{H}(t)^\top.$$

The term on the right-hand side can be rewritten $\langle C_f'(\boldsymbol{\phi}),\mathbf{H}\rangle$, where $C_f'(\boldsymbol{\phi})$ is defined for any $s \in \mathcal{I}$ as

$$C_f'(\boldsymbol{\phi})(s) = 2\boldsymbol{\phi}(s)(\mathbf{A}_{(D)}^\top \mathbf{A}_{(D)} * \boldsymbol{\Lambda}^*) - 2\int \boldsymbol{\Sigma}_{[f]}(s,t)(\mathbf{A}_{(D)} \odot \mathbf{K}(t)),$$

which is bounded under mild conditions. □

Proposition 2. We have

- Using Lemma A2, we have for any $\boldsymbol{\phi}\mathcal{H}^{1\times R}$ and $s \in \mathcal{I}$,

$$C_f'(\boldsymbol{\phi})(s) = 2\boldsymbol{\phi}(s)(\mathbf{A}_{(D)}^\top \mathbf{A}_{(D)} * \boldsymbol{\Lambda}^*) - 2\int \boldsymbol{\Sigma}_{[f]}(s,t)(\mathbf{A}_{(D)} \odot \mathbf{K}(t))$$

  Setting the derivative to zero leads to

$$\boldsymbol{\phi}(s)(\mathbf{A}_{(D)}^\top \mathbf{A}_{(D)} * \boldsymbol{\Lambda}^*) - \int \boldsymbol{\Sigma}_{[f]}(s,t)(\mathbf{A}_{(D)} \odot \mathbf{K}(t)) = 0,$$

  which implies under nonsingular assumption,

$$\boldsymbol{\phi}(s) = \int \boldsymbol{\Sigma}_{[f]}(s,t)(\mathbf{A}_{(D)} \odot \mathbf{K}(t))(\mathbf{A}_{(D)}^\top \mathbf{A}_{(D)} * \boldsymbol{\Lambda}^*)^{-1}.$$

https://doi.org/10.1017/psy.2025.10075 Published online by Cambridge University Press

- Similarly, we have

$$C'_d(\mathbf{A}_d) = 2\int \mathbf{A}_d(\mathbf{A}_{(-d),f}(t)^\top \mathbf{A}_{(-d),f}(t) * \mathbf{\Lambda}^*) - 2\int\int \mathbf{\Sigma}_{[d]}(s,t)(\mathbf{A}_{(-d),f}(t) \odot \mathbf{K}(s)).$$

  Setting the derivative to zero leads to

$$\mathbf{A}_d \int \mathbf{A}_{(-d),f}(t)^\top \mathbf{A}_{(-d),f}(t) * \mathbf{\Lambda}^* - \int\int \mathbf{\Sigma}_{[d]}(s,t)(\mathbf{A}_{(-d),f}(t) \odot \mathbf{K}(s)) = 0,$$

  which implies under nonsingular assumption,

$$\mathbf{A}_d = \int\int \mathbf{\Sigma}_{[d]}(s,t)(\mathbf{A}_{(-d),f}(t) \odot \mathbf{K}(s)) \left(\int \mathbf{A}_{(-d),f}(t)^\top \mathbf{A}_{(-d),f}(t) * \mathbf{\Lambda}^*\right)^{-1}.$$

Lemma A3. *Consider the following rank-R latent PARAFAC decomposition model:*

$$\mathcal{X} = [\![\boldsymbol{u}; \boldsymbol{V}_1; \ldots; \boldsymbol{V}_D]\!] + \mathcal{E}, \tag{A3}$$

$$\boldsymbol{u} \sim \mathcal{N}(0, \mathbf{\Sigma}), \tag{A4}$$

with $\boldsymbol{V}_d \in \mathbb{R}^{p_d \times R}$ *for any* $d = 1, \ldots, D$, $\mathbf{\Sigma} \in \mathbb{R}^{R \times R}$ *a positive-definite matrix, and elements of* $\mathcal{E}$ *i.i.d. normally distributed with variance* σ^2. *The model is identifiable, up to scaling and order, under the following constraints:*

(A1) *When deleting any row of* $\boldsymbol{V}_M = \boldsymbol{V}_1 \odot \cdots \odot \boldsymbol{V}_D$, *the resulting matrix always has two distinct submatrices of rank R.*
(A2) *The following inequality is verified:*

$$\sum_{d=1}^{D} kr(\boldsymbol{V}_d) \geq R + (D-1). \tag{A5}$$

Proof. The proof is largely based on Giordani et al. (2020, Section 4.1). Identifiability is defined as the identifiability of the covariance matrix $\mathbf{\Sigma}_\mathbf{x}$ of $\mathbf{x}$. The residual variance σ is identifiable since, using the fact that $\text{rank}(\mathbf{\Sigma}) = R$, it is defined by the rank R of the decomposition, such as

$$\text{rank}(\mathbf{\Sigma}_\mathbf{x} - \sigma^2 \mathbf{I}) = R. \tag{A6}$$

Now, since $\mathbf{\Sigma}_\mathbf{x} - \sigma^2 \mathbf{I} = \mathbf{V}_M \mathbf{\Sigma} \mathbf{V}_M^\top$ assume we have $\mathbf{V}'_1, \ldots \mathbf{V}'_D$ and $\mathbf{\Sigma}'$ such that

$$\mathbf{V}_M \mathbf{\Sigma} \mathbf{V}_M^\top = \mathbf{V}'_M \mathbf{\Sigma}' \mathbf{V}'^\top_M, \tag{A7}$$

which is equivalent to

$$\mathbf{V}_M \mathbf{\Sigma}^{1/2} \mathbf{\Sigma}^{1/2} \mathbf{V}_M^\top = \mathbf{V}'_M \mathbf{\Sigma}'^{1/2} \mathbf{\Sigma}'^{1/2} \mathbf{V}'^\top_M. \tag{A8}$$

Under condition (A2), since $\text{rank}(\mathbf{\Sigma}) = R$, if any row of $\mathbf{V}_M \mathbf{\Sigma}^{1/2}$ is deleted, there remain two distinct submatrices of rank R. By considering Theorem 5.1 in Anderson and Rubin (1956), we known there exists an orthogonal matrix $\mathbf{M}$ such that $\mathbf{V}_M \mathbf{\Sigma}^{1/2} = \mathbf{V}'_M \mathbf{\Sigma}'^{1/2} \mathbf{M}$. Finally, using condition (A2) and the fact that $\text{rank}(\mathbf{\Sigma}) = R$, we have

$$\text{kr}(\mathbf{\Sigma}^{1/2}) + \sum_{d=1}^{D} \text{kr}(\mathbf{V}_d) \geq 2R + (D-1). \tag{A9}$$

Using the PARAFAC identifiability result provided in Sidiropoulos and Bro (2000), up to permutation and scaling defined by a matrix $\mathbf{P}$, we can assert that $\mathbf{V}_d = \mathbf{V}'_M \mathbf{P}$ for any $d = 1, \ldots, D$ and $\mathbf{\Sigma}^{1/2} = \mathbf{M} \mathbf{\Sigma}'^{1/2} \mathbf{P}$, which implies $\mathbf{\Sigma} = \mathbf{P}^\top \mathbf{\Sigma}' \mathbf{P}$. Identifiability is therefore verified. □

Proposition 3. We follow the same steps as in Han et al. (2023):

- $\mathscr{H}$ is finite dimensional. Denoting p_ϕ the dimension of $\mathscr{H}$. We can decompose ϕ_r and $\tilde{\phi}_r$ using an orthonormal basis $\{\psi_k\}_{1 \leq k \leq p_\phi}$ of $\mathscr{H}$,

$$\phi_r = \sum_{k=1}^{p_\phi} c_{kr} \psi_k \quad \tilde{\phi}_r = \sum_{k=1}^{p_\phi} \tilde{c}_{kr} \psi_k.$$

  Considering the following latent factor PARAFAC decomposition:

$$\mathcal{B} = \sum_{r=1}^{R} u_r (\mathbf{c}_r \circ \mathbf{a}_{1r} \circ \cdots \circ \mathbf{a}_{Dr}) \quad \tilde{\mathcal{B}} = \sum_{r=1}^{R} \tilde{u}_r (\tilde{\mathbf{c}}_r \circ \tilde{\mathbf{a}}_{1r} \circ \cdots \circ \tilde{\mathbf{a}}_{Dr}).$$

https://doi.org/10.1017/psy.2025.10075 Published online by Cambridge University Press

We can rewrite (6)

$$\begin{aligned}\sum_{r=1}^{R} u_r \phi_r(t)(\mathbf{a}_{1r} \circ \cdots \circ \mathbf{a}_{Dr}) &= \sum_{r=1}^{R} u_r \sum_{k=1}^{p_\phi} c_{kr}\psi_k(t)(\mathbf{a}_{1r} \circ \cdots \circ \mathbf{a}_{Dr}) \\ &= \sum_{k=1}^{p_\phi}\sum_{r=1}^{R} u_r (c_{kr}(\mathbf{a}_{1r} \circ \cdots \circ \mathbf{a}_{Dr}))\psi_k(t) \\ &= \sum_{k}^{p_\phi} \mathcal{B}_k \psi_k(t),\end{aligned}$$

where $\mathcal{B}_k$ denotes the sub-tensor of $\mathcal{B}$ defined by fixing the first mode (associated with basis coefficients $\mathbf{C}$) to the kth dimension. Similarly, we have

$$\sum_{r=1}^{R} \tilde{u}_r \tilde{\phi}_r(t)(\tilde{\mathbf{a}}_{1r} \circ \cdots \circ \tilde{\mathbf{a}}_{Dr}) = \sum_{k=1}^{p_\phi} \tilde{\mathcal{B}}_k \psi_k(t).$$

From the equality,

$$\sum_{k=1}^{p_\phi} \mathcal{B}_k \psi_k(t) = \sum_{k=1}^{p_\phi} \tilde{\mathcal{B}}_k \psi_k(t)$$

and the orthonormality of $\{\psi_k\}_{1\le k\le p_\phi}$, we have that $\mathcal{B} = \tilde{\mathcal{B}}$. From the identifiability result given in Lemma A3, assuming scaling and order indeterminacies fixed, we have $\boldsymbol{\Theta} = \tilde{\boldsymbol{\Theta}}$ under the assumption that deleting any row of $\mathbf{C} \odot \mathbf{A}_1 \odot \cdots \odot \mathbf{A}_D$ results in a matrix having two distinct full-rank submatrices and considering:

$$\mathrm{kr}(\mathbf{C}) + \sum_{d=1}^{D} \mathrm{kr}(\mathbf{A}_d) \ge R + (D-1).$$

- $\mathscr{H}$ is infinite-dimensional. Considering a regular grid of $\mathcal{I}$ with size K, we denote $\mathcal{X}^{(K)}$ and $\tilde{\mathcal{X}}^{(K)}$, the values of $\mathcal{X}$ and $\tilde{\mathcal{X}}$, respectively, at the time points of the grid. Since $\mathcal{X} = \tilde{\mathcal{X}}$, we have $\mathcal{X}^{(K)} = \tilde{\mathcal{X}}^{(K)}$. Considering $\boldsymbol{\phi}_r^{(K)} \in \mathbb{R}^K$ the vector of values of ϕ_r at the time points of the grid, we have from the previous equality,

$$\sum_{r=1}^{R} u_r(\boldsymbol{\phi}_r^{(K)} \circ \mathbf{a}_{1r} \circ \cdots \circ \mathbf{a}_{Dr}) = \sum_{r=1}^{R} \tilde{u}_r(\tilde{\boldsymbol{\phi}}_r^{(K)} \circ \tilde{\mathbf{a}}_{1r} \circ \cdots \circ \tilde{\mathbf{a}}_{Dr}),$$

which also corresponds to a SupCP model. In this context, for K large enough, we are guaranteed that $\mathrm{kr}(\boldsymbol{\phi}^{(K)}) = \mathrm{kr}(\boldsymbol{\phi})$. Using identifiability result from Lemma A3, we can assert that $\boldsymbol{\phi}^{(K)} = \tilde{\boldsymbol{\phi}}^{(K)}$ as long as deleting any row of $(\boldsymbol{\phi}^{(K)} \odot \mathbf{A}_1 \odot \cdots \odot \mathbf{A}_D)$ results in a matrix which has two full rank distinct submatrices and

$$\mathrm{kr}(\boldsymbol{\phi}^{(K)}) + \sum_{d=1}^{D} \mathrm{kr}(\mathbf{A}_d) \ge R + (D-1).$$

Finally, we can assert that there exists K_T such that for all $K \ge K_T$, the first condition is necessarily verified and $\mathrm{kr}(\boldsymbol{\phi}^{(K)}) = R = \mathrm{kr}(\boldsymbol{\phi})$.

Proposition 4. Denoting $\boldsymbol{e}_n$ the vector of measurement errors $\{\mathcal{E}_{nk}\}_{1\le k\le K_n}$ of sample n. We have

$$\mathbf{y}_n = \mathbf{F}_n \mathbf{u}_n + \boldsymbol{e}_n.$$

From this, we propose to rewrite jointly the observations and the scores as:

$$\begin{bmatrix}\mathbf{y}_n \\ \mathbf{u}_n\end{bmatrix} = \begin{bmatrix}0 \\ 0\end{bmatrix} + \begin{bmatrix}\mathbf{F}_n & \mathbf{I} \\ \mathbf{I} & 0\end{bmatrix}\begin{bmatrix}\mathbf{u}_n \\ \boldsymbol{e}_n\end{bmatrix}.$$

Using this expression, and the jointly Gaussian assumptions on $\mathbf{u}_n$ and $\boldsymbol{e}_n$ we clearly see that $\mathbf{u}_n$ and $\mathbf{y}_n$ are also jointly Gaussian with joint law:

$$\begin{aligned}\begin{bmatrix}\mathbf{y}_n \\ \mathbf{u}_n\end{bmatrix} &\sim N\left(\begin{bmatrix}\mathbf{F}_n & \mathbf{I} \\ \mathbf{I} & 0\end{bmatrix}\begin{bmatrix}\boldsymbol{\mu} \\ 0\end{bmatrix}, \begin{bmatrix}\mathbf{F}_n & \mathbf{I} \\ \mathbf{I} & 0\end{bmatrix}\begin{bmatrix}\boldsymbol{\Lambda} & 0 \\ 0 & \sigma^2\mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{F}_n^\top & \mathbf{I} \\ \mathbf{I} & 0\end{bmatrix}\right) \\ &\sim N\left(\begin{bmatrix}\mathbf{F}_n\boldsymbol{\mu} \\ \boldsymbol{\mu}\end{bmatrix}, \begin{bmatrix}\mathbf{F}_n\boldsymbol{\Lambda}\mathbf{F}_n^\top + \sigma^2\mathbf{I} & \mathbf{F}_n\boldsymbol{\Lambda} \\ \boldsymbol{\Sigma}\mathbf{F}_n^\top & \boldsymbol{\Lambda}\end{bmatrix}\right).\end{aligned}$$

https://doi.org/10.1017/psy.2025.10075 Published online by Cambridge University Press

To obtain the best predictions for the scores, we now consider the conditional distribution of $\mathbf{u}_n$ and suggest using the conditional expectation as the best predictor. Using the standard formulation of the Gaussian conditional distribution, we obtain

$$\mathbb{E}[\mathbf{u}_n|\boldsymbol{\mathcal{Y}}_n] = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{F}_n^\top(\mathbf{F}_n\boldsymbol{\Lambda}\mathbf{F}_n^\top + \sigma^2\mathbf{I})^{-1}\boldsymbol{y}_n$$
$$\boldsymbol{\Sigma}_{\mathbf{u}_n|\boldsymbol{\mathcal{Y}}_n} = \boldsymbol{\Lambda} - \boldsymbol{\Lambda}\mathbf{F}_n^\top(\mathbf{F}_n\boldsymbol{\Lambda}\mathbf{F}_n^\top + \sigma^2\mathbf{I})^{-1}\mathbf{F}_n\boldsymbol{\Lambda}.$$

B. Additional experiments

B.1. Simulations: Higher order

We propose to compare decomposition methods on order $D = 3$ tensors using the simulation setting of Section 4. Dimensions of tensors considered are $(30,5,5)$. Results are displayed in Figure B1. We still observe a notable advantage of our approach in high and medium noise settings. On the other hand, MFPCA seems to perform worse than before. We suppose that this is caused by the lack of assumptions on the underlying structure of tensors.

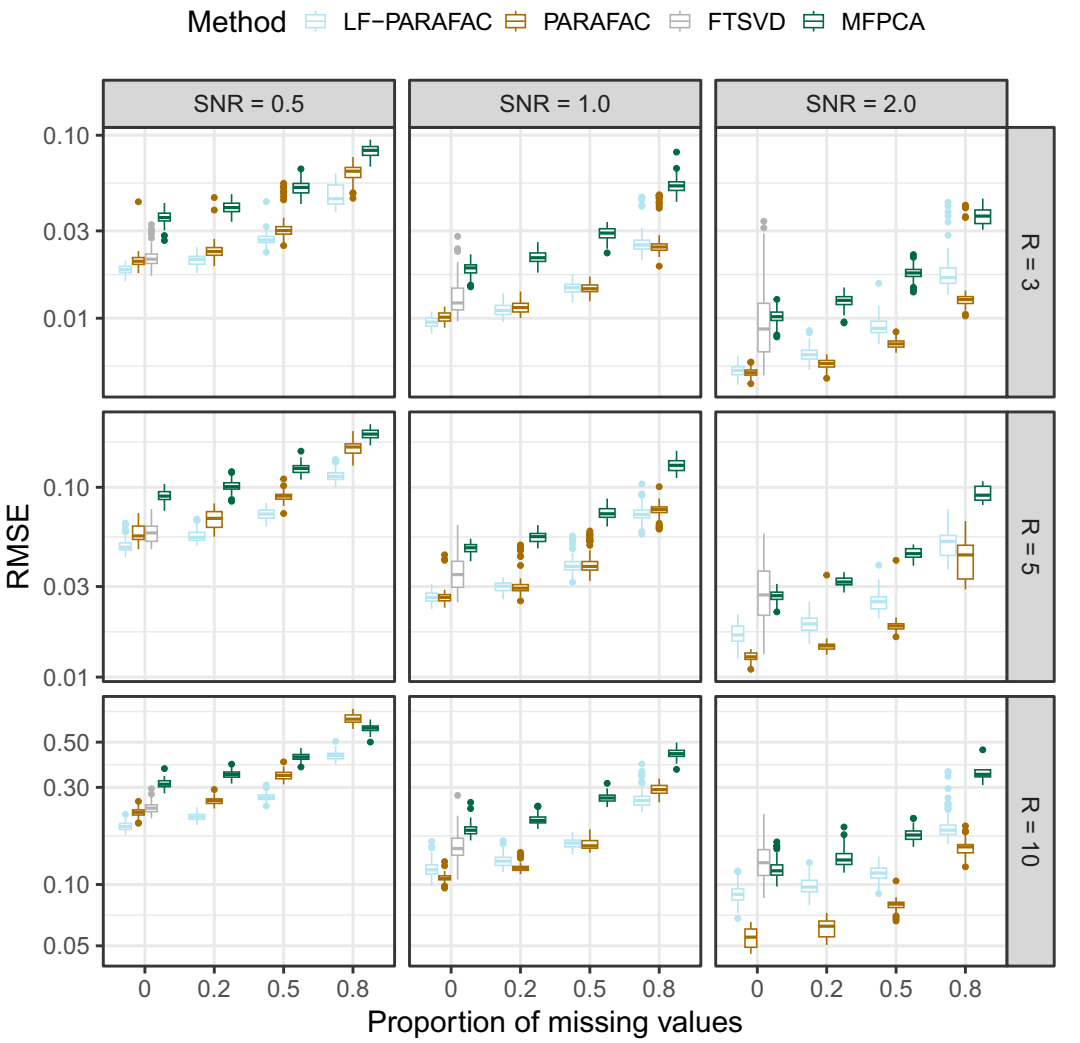


Figure B1. Root mean squared error (RMSE) of order-3 reconstructed functional tensors. Comparing latent functional PARAFAC (LF-PARAFAC), standard PARAFAC (PARAFAC), functional tensor singular value decomposition (FTSVD), and multivariate functional principal component analysis (MFPCA).

https://doi.org/10.1017/psy.2025.10075 Published online by Cambridge University Press

B.2. Simulations: Misspecified model

We propose here to investigate how decomposition methods perform on order $D = 3$ tensors assuming no underlying low-rank structure. In this context, data is generated using a Fourier basis of order 3 by sampling Fourier coefficients from a multivariate normal distribution for each tensor entry. The dimensions of tensors considered are $(30, 8, 8)$. For tensor decomposition approaches, a rank of $R = 3$ decomposition is considered. Experiments are repeated 10 times. Results are displayed in Figure B2. As expected, we observe that MFPCA performs better than tensor decomposition approaches in dense and low-sparsity settings since the method does not assume any underlying low-rank structure. Interestingly, our method still performs better than MFPCA for medium-sparsity and high-sparsity settings (Figure B3).

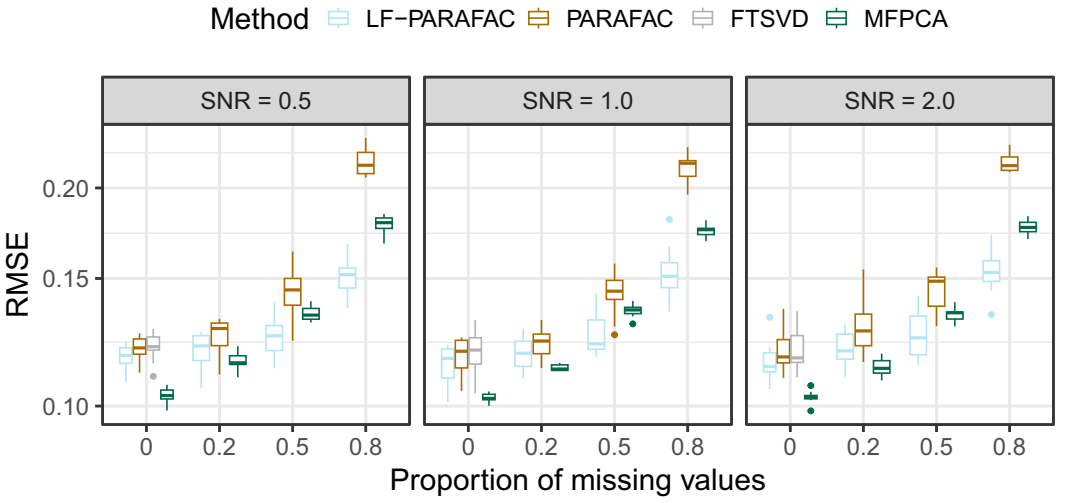


Figure B2. Reconstruction error on order $D = 3$ tensors using $p_1 = p_2 = 8$, with $N = 100$ samples. Comparing LF-PARAFAC, PARAFAC (standard), FTSVD, and MFPCA for different proportions of missing values (x-axis) and signal-to-noise ratios (SNR) (column-wise facets). Plot obtained from 10 simulation runs.

https://doi.org/10.1017/psy.2025.10075 Published online by Cambridge University Press

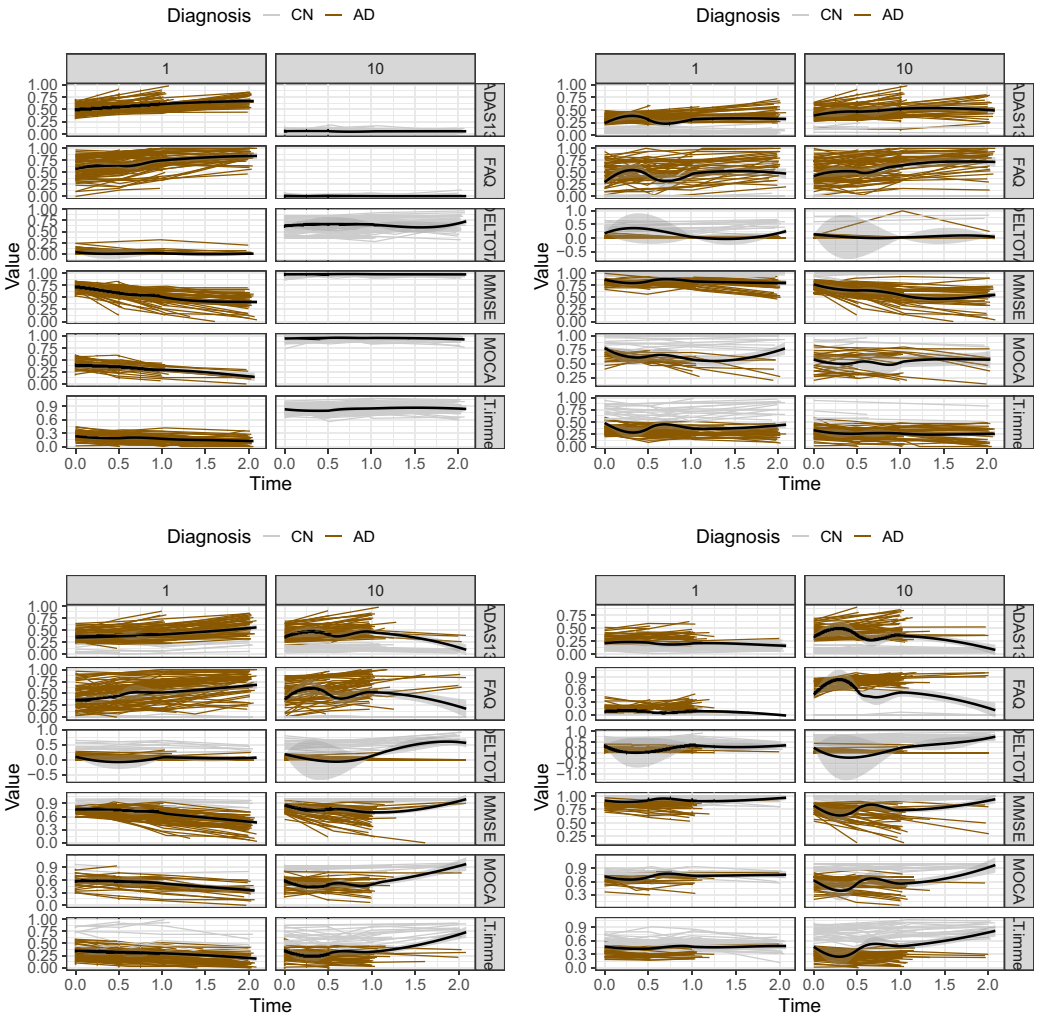


Figure B3. Cognitive score trajectories in first (left column) and last deciles (right column) of rank-1 sample-mode scores (top left), of rank-2 sample-mode scores (top right), of rank-3 sample-mode scores (bottom left), and of rank-4 sample-mode scores (bottom right).

Cite this article: Sort, L., Le Brusquet, L. and Tenenhaus, A. (2026). Latent Functional PARAFAC for Modeling Multidimensional Longitudinal Data. *Psychometrika*, 789–813. https://doi.org/10.1017/psy.2025.10075

https://doi.org/10.1017/psy.2025.10075 Published online by Cambridge University Press