

Smooth and probabilistic PARAFAC model with auxiliary covariates

Leying Guan

To cite this article: Leying Guan (15 Sep 2023): Smooth and probabilistic PARAFAC model with auxiliary covariates, Journal of Computational and Graphical Statistics, DOI: [10.1080/10618600.2023.2257783](https://doi.org/10.1080/10618600.2023.2257783)

To link to this article: <https://doi.org/10.1080/10618600.2023.2257783>



© 2023 The Author(s). Published with license by Taylor and Francis Group, LLC



[View supplementary material](#)



Accepted author version posted online: 15 Sep 2023.



[Submit your article to this journal](#)



Article views: 133



[View related articles](#)



[View Crossmark data](#)

Smooth and probabilistic PARAFAC model with auxiliary covariates

Leying Guan*

Department of Biostatistics, Yale University

* leying.guan@yale.edu

Abstract

As immunological and clinical studies become more complex, there is an increasing need to analyze temporal immunophenotypes alongside demographic and clinical covariates, where each subject receives matrix-valued time series observations for potentially high-dimensional longitudinal features, as well as other static characterizations. Researchers aim to find the low-dimensional embedding of subjects using matrix-valued time series observations and investigate relationships between static clinical responses and the embedding. However, constructing these embeddings can be challenging due to high dimensionality, sparsity, and irregularity in sample collection over time. In addition, the incorporation of static auxiliary covariates is frequently desired during such a construction. To address these issues, we propose a smoothed probabilistic PARAFAC model with covariates (SPACO) that utilizes auxiliary covariates of interest. We provide extensive simulations to test different aspects of SPACO and demonstrate its application to an immunological dataset from patients with SARS-CoV-2 infection. Supplemental materials associated with this article are available online.

Keywords: Tensor decomposition; Time series; Missing data; Probabilistic model.

1 Introduction

Sparsely observed multivariate time series data are now common in immunological studies. For each subject or participant i ($i = 1, \dots, I$), we can collect multiple measurements on J features over time, but often at n_i different time stamps $\{t_{i,1}, \dots, t_{i,n_i}\}$. For example, for each subject, immune profiles are measured for hundreds of markers at irregular sampling times in Lucas et al. (2020) and Rendeiro et al. (2021). Let $X_i \in \mathbb{R}^{n_i \times J}$ be the longitudinal measurements for subject i , and $\mathcal{X} \in \mathbb{R}^{I \times T \times J}$ be the sparse three-way tensor collecting X_i for all I subjects, where T is the number of unique time points across all subjects, e.g., $T = |\cup_i \{t_{i,1}, \dots, t_{i,n_i}\}|$. A major computational task in practice is to construct low-dimensional embeddings that capture data variability in \mathcal{X} across subjects. Tensor decomposition is a widely used technique for achieving this goal. Figure 1 demonstrates the PARAFAC/CP tensor decomposition and utilization of the decomposition results (Harshman and Lundy, 1994), where we approximate a tensor by the sum of rank-one tensors. Apart from tensor reconstruction and denoising, the decomposition of the three-way tensor \mathcal{X} provides decomposed factors U , V , and Φ , which aim to capture major variations along the subject, feature, and time directions respectively. These factors can be utilized for further downstream analysis. For instance, factors U (also denoted as the subject scores) serve as the aforementioned subject embeddings, which can be used to investigate the relationship between immune profiles and various clinical/demographic covariates.

The immune profiles are usually highly sparse observed along the time dimensions, that is, the sets $\{t_{i,1}, \dots, t_{i,n_i}\}$ for different subjects i tend to be small in size and have low overlaps, resulting in a high missing rate along the time dimension of \mathcal{X} (Figure 1). In addition, researchers may have access to nontemporal covariates for each subject, such as medical conditions and demographic information, which may partially explain the variation in the temporal measurements of \mathcal{X} across subjects. As an example, the IMPACT study (Lucas et al., 2020) analyzed immune profiles of 98 COVID-19 infected individuals. Researchers measured immune profiles for more than 30 days after symptom onset, but only an average of 1.84 measured time points per subject. Auxiliary demographic information, including sex and age, and several

preexisting medical conditions, were also collected. To better estimate U in this example, a tensor decomposition tool that can properly model the longitudinal behavior of the factors and incorporate available auxiliary covariates is needed.

In this paper, we propose SPACO (smooth and probabilistic PARAFAC model with auxiliary covariates) to adapt to the sparsity long the time dimension in \mathbf{X} and utilize the auxiliary covariates \mathbf{Z} . SPACO assumes that \mathbf{X} is a noisy realization of some low-rank signal tensor whose time components are smooth and subject scores have a potential dependence on the auxiliary covariates \mathbf{Z} :

$$x_{ijt} = \sum_{k=1}^K u_{ik} \phi_{tk} v_{jk} + \epsilon_{ijt}, \quad \epsilon_{ijt} \sim \mathcal{N}(0, \sigma_j^2)$$

$$\mathbf{u}_i = (u_{ik})_{k=1}^K \sim \mathcal{N}(\boldsymbol{\eta}_i, \Lambda_f), \quad \boldsymbol{\eta}_i = \boldsymbol{\beta}^\top \mathbf{z}_i.$$

Here, (1) \mathbf{z}_i is the i th row of \mathbf{Z} and $\boldsymbol{\beta} \in \mathbb{R}^{q \times K}$ describes the dependence of the expected subject score for subject i on \mathbf{z}_i , (2) u_{ik} , ϕ_{tk} , v_{jk} are the subject score, trajectory value and feature loading for factor k in the PARAFAC model and the observation indexed by (i, t, j) where \mathbf{u}_i has a normal prior $\mathcal{N}(\boldsymbol{\eta}_i, \Lambda_f)$. We impose smoothness on time trajectories $(\phi_{tk})_{t=1}^\top$ and sparsity on $\boldsymbol{\beta}$ to deal with the irregular sampling along the time dimension and potentially high dimensionality in \mathbf{Z} .

Alongside the proposal of the model, we will delve into several issues pertinent to SPACO, including model initialization, auto-tuning of smoothness and sparsity in $\boldsymbol{\beta}$, and hypothesis testing on $\boldsymbol{\beta}$. Effectively addressing these concerns is crucial for applying SPACO in practice. In the remainder of the article, we summarize work closely related to our study in Section 1.1. We describe the SPACO model in Section 2 and discuss model parameter estimation with predetermined tuning parameters in Section 3. Section 4 is dedicated to the examination of the aforementioned issues that bear practical significance. A comparative analysis of SPACO with existing methods on synthetic data is presented in Section 5. Lastly, in Section 6, we employ SPACO on a highly sparse tensor containing immunological measurements from SARS-COV-2 infected patients. We also provide a Python package, *SPACO*, for researchers who are interested in utilizing the proposed method.

1.1 Related work

Matrix-valued observations have become increasingly important in modern applications, particularly in the context of dimension reduction or matrix-valued data clustering (Chang, 1983; Vichi et al., 2007; Viroli, 2011; Anderlucci and Viroli, 2015).

Tensor decomposition is a useful technique for modeling matrix-valued observations, and has been applied in various fields (Acar and Yener, 2008; Kolda and Bader, 2009; Sidiropoulos et al., 2017). In the study of multivariate longitudinal data, researchers in economics have combined tensor decomposition with auto-cross-covariance estimation and autoregressive models (Fan et al., 2008; Lam et al., 2011; Fan et al., 2011; Bai and Wang, 2015, 2016; Wang et al., 2019, 2021). However, these approaches are either not compatible with highly sparse data or do not scale well with the feature dimensions, both of which are important for medical applications.

Functional PCA (Besse and Ramsay, 1986; Yao et al., 2005) is often used for modeling sparse longitudinal data in the matrix form. It utilizes the smoothness of time trajectories to handle sparsity in longitudinal observations and estimates the eigenvectors and factor scores under this smoothness assumption. Yokota et al. (2016) and Imaizumi and Hayashi (2017) introduced smoothness into tensor decomposition and estimated the parameters by iteratively solving penalized regression problems. However, these methods do not consider auxiliary covariates.

It has been previously discussed that including auxiliary information could potentially improve our estimation. For example, Li et al. (2016) proposed the supervised sparse and functional principal component (SupSFPC) method and observed that auxiliary covariates improve signal estimation quality in the matrix setting for modeling multivariate longitudinal observations. In Chen et al. (2020), the authors observed that inclusion of auxiliary constraints improve the quality of tensor decomposition. Lock and Li (2018) proposed SupCP, a supervised multiway factorization model with complete observation that employs a probabilistic tensor model (Tipping and Bishop, 1999; Mnih and Salakhutdinov, 2007; Hinrich and Mørup, 2019). Although an extension to sparse tensor is straightforward, SupCP

does not model smoothness and can be highly affected by severe missingness along the time dimension.

To address these limitations, we propose SPACO, an extension of functional PCA and SupSFPC to the setting of three-way tensor decomposition using the parallel factor (PARAFAC) model (Carroll et al., 1980; Harshman and Lundy, 1994). SPACO jointly models smooth longitudinal data with potentially high-dimensional static covariates Z using a probabilistic model. When Z is unavailable, we refer to the SPACO model as SPACO-. SPACO- itself is also an attractive alternative to existing tensor decomposition implementations with probabilistic modeling, smoothness regularization, and automatic parameter tuning. In our numerical experiments, we demonstrate the advantages of SPACO and SPACO- over several state-of-the-art tensor decomposition methods and the improvement of SPACO over SPACO- by utilizing Z .

2 SPACO Model

2.1 Notations

We use (1) bolded uppercase letters to represent three-way tensors, (2) regular uppercase letters to represent a matrix, and (3) bolded lowercase to represent vectors. Following this convention, let $\mathbf{X} \in \mathbb{R}^{I \times T \times J}$ be the tensor for some sparse multivariate longitudinal observations, where I is the number of subjects, J is the number of features, and T is the number of total unique time points. Let

$$\mathbf{X}_I := (\mathbf{X}_{:,1} \cdots \mathbf{X}_{:,J}) \in \mathbb{R}^{I \times (TJ)}, \mathbf{X}_T := (\mathbf{X}_{:,1}^\top \cdots \mathbf{X}_{:,J}^\top) \in \mathbb{R}^{T \times (IJ)}, \mathbf{X}_J := (\mathbf{X}_{:,1}^\top \cdots \mathbf{X}_{:,T}^\top) \in \mathbb{R}^{J \times (IT)}$$

be the matrix unfolding of \mathbf{X} in the subject/feature/time dimension respectively. For any matrix A , we let $A_{i\cdot} / A_{\cdot j}$ denote its i^{th} row/column, and often write $A_{i\cdot}$ as A_i for the i^{th} column for convenience. We also define:

Tensor product \odot : $\mathbf{a} \in \mathbb{R}^I, \mathbf{b} \in \mathbb{R}^T, \mathbf{c} \in \mathbb{R}^J$, then, $\mathbf{A} = \mathbf{a} \odot \mathbf{b} \odot \mathbf{c} \in \mathbb{R}^{I \times T \times J}$ with $A_{ijt} = a_i b_t c_j$.

Kronecker product \otimes : $A \in \mathbb{R}^{I_1 \times K_1}, B \in \mathbb{R}^{I_2 \times K_2}$, then $C = A \otimes B \in \mathbb{R}^{(I_1 I_2) \times (K_1 K_2)}$ is the

$$\text{Kronecker product of } A \text{ and } B: C = A \otimes B = \begin{pmatrix} A_{11}B & \cdots & A_{1K_1}B \\ \vdots & \ddots & \vdots \\ A_{I_1 1}B & \cdots & A_{I_1 K_1}B \end{pmatrix} \in \mathbb{R}^{(I_1 I_2) \times (K_1 K_2)}.$$

Column-wise Khatri-Rao product \odot : $A \in \mathbb{R}^{I_1 \times K}$, $B \in \mathbb{R}^{I_2 \times K}$, then $C = A \odot B \in \mathbb{R}^{(I_1 I_2) \times K}$ with $C_{:,k} = (A_{:,k} \otimes B_{:,k})$ for $k = 1, \dots, K$.

Element-wise multiplication \cdot : $A, B \in \mathbb{R}^{I \times K}$, then $C = A \cdot B \in \mathbb{R}^{I \times K}$ with $C_{ik} = (A_{ik} B_{ik})$; for $\mathbf{b} \in \mathbb{R}^K$, $C = A \cdot \mathbf{b} = A \text{diag}\{\mathbf{b}_1, \dots, \mathbf{b}_K\}$; for $\mathbf{b} \in \mathbb{R}^I$, $C = \mathbf{b} \cdot A = \text{diag}\{\mathbf{b}_1, \dots, \mathbf{b}_I\} A$.

2.2 Smooth and probabilistic PARAFAC model with covariates

We assume \mathbf{X} to be a noisy realization of an underlying signal array \mathbf{F} which is a rank K tensor with $U / \Phi / V$ be the decomposed factor matrices, e.g.,

$\mathbf{F} = \sum_{k=1}^K U_k \odot \Phi_k \odot V_k$, where $U_k / \Phi_k / V_k$ are k^{th} columns of $U / \Phi / V$. We denote the rows of $U / \Phi / V$ by $\mathbf{u}_i / \phi_i / \mathbf{v}_j$, and their entries by $u_{ik} / \phi_{ik} / v_{jk}$. We let x_{ijt} denote the (i, t, j) -entry of \mathbf{X} . Then,

$$x_{ijt} = \sum_{k=1}^K u_{ik} \phi_{ik} v_{jk} + \varepsilon_{ijt}, \quad \mathbf{u}_i \sim \mathcal{N}(\boldsymbol{\eta}_i, \Lambda_f), \quad \varepsilon_{ijt} \sim \mathcal{N}(0, \sigma_j^2), \quad (1)$$

where $\Lambda_f = \text{diag}\{s_1^2, \dots, s_K^2\}$ is a $K \times K$ diagonal covariance matrix and $\boldsymbol{\eta}_i$ is a mean vector of length K . In practice, although the observation tensor \mathbf{X} is almost always of high-rank, it is often assumed that \mathbf{F} is (or approximately) low-rank with K being small.

We set $\boldsymbol{\eta}_i = \mathbf{0}$ in the absence of auxiliary covariates $Z \in \mathbb{R}^{I \times q}$. In the setting where we are interested in explaining the heterogeneity $\boldsymbol{\eta}_i$ across subjects using Z , we may model $\boldsymbol{\eta}_i$ as a function of $\mathbf{z}_i := Z_{i,:}$. Here, we consider a linear model $\eta_{ik} = \mathbf{z}_i^\top \boldsymbol{\beta}_k$ for $k = 1, \dots, K$, where η_{ik} is the k^{th} entry of $\boldsymbol{\eta}_i$ and $\boldsymbol{\beta} \in \mathbb{R}^{q \times K}$ is the coefficients matrix describing the dependence of subject scores on Z . To avoid confusion, we will always call \mathbf{X} the “features”, and Z the “covariates” or “variables”.

Recalling that $X_I \in \mathbb{R}^{I \times (TJ)}$ is the unfolding of \mathbf{X} in the subject direction, we write \vec{i} for the indices of observed values in the i^{th} row of X_I , and $X_{I, \vec{i}}$ for the vector of these observed values. Each such observed value x_{ijt} has noise variance σ_j^2 , and we write $\Lambda_{\vec{i}}$ to represent the diagonal covariance matrix with diagonal values σ_j^2 being the corresponding variances for ε_{ijt} at indices in \vec{i} . Similarly, we define $\{\vec{i}, X_{I, \vec{i}}, \Lambda_{\vec{i}}\}$ for

the unfolding $X_T \in \mathbb{R}^{T \times (IJ)}$, and $\{\vec{j}, X_{J,\vec{j}}, \Lambda_{\vec{j}}\}$ for the observed indices, the associated observed vector and diagonal covariance matrix for the j^{th} row in $X_J \in \mathbb{R}^{J \times (IT)}$. We set $\Theta = \{V, \Phi, \beta, (\sigma_j^2, j=1, \dots, J), (s_k^2, k=1, \dots, K)\}$ to denote all model parameters. Set $H = V \odot \Phi \in \mathbb{R}^{(IJ) \times K}$ and $\mathbf{f}_i = X_{I,\vec{i}} - H_{\vec{i},\cdot} \boldsymbol{\eta}_i$ as the residual vector after removing mean signals from observations for subject i . If U is observed, the complete log-likelihood is

$$L(X, U | \Theta) = -\frac{1}{2} \sum_i \left(\mathbf{f}_i^\top \Lambda_{\vec{i}}^{-1} \mathbf{f}_i + (\mathbf{u}_i - \boldsymbol{\eta}_i)^\top \Lambda_f^{-1} (\mathbf{u}_i - \boldsymbol{\eta}_i) + \log |\Lambda_{\vec{i}}| + I \log |\Lambda_f| \right). \quad (2)$$

Set $\tilde{\Sigma}_i = \Lambda_{\vec{i}} + H_{\vec{i}} \Lambda_f H_{\vec{i}}^\top$. The marginalized log likelihood integrating out the randomness in U enjoys a closed form (Lock and Li, 2018):

$$L(X | \Theta) \propto -\frac{1}{2} \left(\sum_i \mathbf{f}_i^\top \tilde{\Sigma}_i^{-1} \mathbf{f}_i + \log |\tilde{\Sigma}_i| \right). \quad (3)$$

Model parameters in eq. (3) are not identifiable due to (1) parameters rescaling from $(\Phi_k, V_k, \beta_k, s_k^2)$ to $(c_1 \Phi_k, c_2 V_k, c_3 \beta_k, c_3^2 s_k^2)$ for any $c_1 c_2 c_3 = 1$, and (2) reordering of different component k for $k = 1, \dots, K$. More discussions of the model identifiability can be found in Lock and Li (2018). Hence, adopting similar rules from Lock and Li (2018), we require

$$(C.1) \quad \|V_k\|_2^2 = 1, \|\Phi_k\|_2^2 = T.$$

$$(C.2) \quad \text{The latent components are in decreasing order based on their overall variances } \Lambda_{f,kk} + \|Z\beta_k\|_2^2 / I, \text{ and the first non-zero entries in } V_k \text{ and } \Phi_k \text{ to be positive, e.g., } v_{k1} > 0 \text{ and } \phi_{k1} > 0 \text{ if they are non-zero.}$$

In the immunological application being considered, U represents subject scores, which are latent variables characterizing differences across subjects, V represents feature loadings, revealing the composition of factors using the original features and aiding downstream interpretation, and Φ represents time trajectories interpreted as function values sampled from some underlying smooth functions $\phi_k(t)$ at a set of discrete-time points where the sampling in the time direction is often highly sparse.

We address the challenge of sparse sampling utilizing the smoothness assumption in $\phi_k(t)$ and encourage it by directly penalizing the function values via a penalty term $\sum_k \lambda_{1k} \Phi_k^\top \Omega \Phi_k$. This paper considers a Laplacian smoothing (Sorkine et al., 2004) with a weighted adjacency matrix Γ . Let $\mathcal{T}[t]$ be the time associated with the t^{th} slice along the time dimension, e.g., time associated with $X_{:,t}$. We define Ω and Γ as

$$\Omega = \Gamma^\top \Gamma, \quad \Gamma = \begin{pmatrix} \frac{1}{\mathcal{T}[2]-\mathcal{T}[1]} & -\frac{1}{\mathcal{T}[2]-\mathcal{T}[1]} & \cdots & 0 & 0 \\ 0 & \frac{1}{\mathcal{T}[3]-\mathcal{T}[2]} & \ddots & 0 & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\mathcal{T}[T]-\mathcal{T}[T-1]} & -\frac{1}{\mathcal{T}[T]-\mathcal{T}[T-1]} \end{pmatrix} \in \mathbb{R}^{T \times (T-1)}$$

The Laplacian smoothing discourages abrupt trajectory changes in observations from close-by time points, which is usually a reasonable assumption biologically. Practitioners may choose other forms for Ω based on their applications. For example, if practitioners want $\phi_k(t)$ to have slowly varying derivatives, they can also use a penalty matrix that penalizes changes in gradients over time.

Further, when the number of covariates q in Z is moderately large, we may wish to impose sparsity in the β parameter. We encourage such sparsity by including a lasso penalty (Tibshirani, 2011) in the model. In summary, our goal is then to find parameters maximizing the expected penalized log-likelihood, or minimizing the penalized expected deviance loss $J(\Theta)$, under norm constraints:

$$\begin{aligned} \min J(\Theta) &= -\frac{1}{2} L(\mathbf{X} | \Theta) + \sum_{k=1}^K \lambda_{1k} \Phi_k^\top \Omega \Phi_k + \sum_k \lambda_{2k} |\beta_k| \\ \text{s.t. } & \|\mathbf{V}_k\|_2^2 = 1, \|\Phi_k\|_2^2 = T, \text{ for all } k=1, \dots, K. \end{aligned} \quad (4)$$

The objective only incorporates the identifiability constraint (C.1). By changing the signs in \mathbf{V} , Φ , β and reordering them, we can always ensure (C.2) without altering the attained objective value.

Equation (4) characterizes a non-convex problem. We will obtain local optimal solutions via an alternating update procedure: (1) we update β through lasso regressions while keeping other parameters fixed, (2) we update other model parameters using the EM algorithm while keeping β fixed. See Section 3 for more details.

3 Model parameter estimation

Given the model rank K and penalty terms $\lambda_{1k}, \lambda_{2k}$, we alternately update parameters β , V , U , \mathcal{S}^2 and σ^2 with a mixed EM procedure described in Algorithm 1. We briefly explain the updating steps here:

(1): Given other parameters, we find β to directly minimize the objective by solving a least-squares regression problem with lasso penalty.

(2): Fixing β , we update the other parameters using an EM procedure. Denote the current parameters as Θ_0 . At the M-step, we minimize the penalized expected negative log-likelihood

$$J(\Theta; \Theta_0) := \mathbb{E}_{U|\Theta_0} \left\{ -L(X, U | \Theta) + \sum_k \lambda_{1k} \Phi_k^\top \Omega \Phi_k + \sum_k \lambda_{2k} |\beta_k| \right\}, \quad (5)$$

under the current posterior distribution $U | \Theta_0, X$. We adopt a block-wise parameter updating scheme where we update V_k , Φ_k , Λ_f and σ_j^2 sequentially.

Algorithm 1: SPACO with fixed penalties

Data: $X, \Omega, \lambda_1, \lambda_2, K$

Result: Estimated $V, \Phi, \beta, \mathcal{S}^2, \sigma^2$ and posterior distribution of $U | \Theta, X$ and the marginalized density $\mathbf{P}(X | \Theta)$.

1 Initialization of $V, \Phi, \beta, \mathcal{S}^2, \sigma^2$ and the posterior distribution of U ,

2 **while** *Not converged* **do**

3 **for** $k = 1, \dots, K$ **do**

4 $\beta_k \leftarrow \arg \min_{\beta_k} \{-L(X | \Theta) + \lambda_{2k} |\beta_k| \}$

5 $V_k \leftarrow \arg \min_{V_k} \left[J(\Theta; \Theta_0) + \nu_V \|V_k\|_2^2 \right]$, ν_V is the largest value leading to the minimizer having $\|V_k\|_2^2 = 1$
 6 $\Phi_k \leftarrow \arg \min_{\Phi_k} \left[J(\Theta; \Theta_0) + \nu_\Phi \|\Phi_k\|_2^2 \right]$, ν_Φ is the largest value leading to the minimizer having $\|\Phi_k\|_2^2 = T$
 7 $s_k^2 \leftarrow \arg \max_{s_k^2} \mathbb{E}_{U|\Theta_0} L(\mathbf{X}, U | \Theta)$.
 8 end
 9 For $j = 1, \dots, J$: $\sigma_j^2 \leftarrow \arg \max_{\sigma_j^2} \mathbb{E}_{U|\Theta_0} L(\mathbf{X}, U | \Theta)$.
 10 Update the posterior distribution (posterior mean and covariance) of U .
 11 end

Algorithm 1 describes the high level ideas of our updating schemes. The posterior distribution of \mathbf{u}_i for each row in U is Gaussian, with posterior covariance Σ_i and posterior mean $\boldsymbol{\mu}_i$ given below,

$$\Sigma_i = (H_i^\top \Lambda_i^{-1} H_i + \Lambda_f^{-1})^{-1} \quad (6)$$

$$\boldsymbol{\mu}_i = \Sigma_i \left(\Lambda_f^{-1} \boldsymbol{\beta}^\top \mathbf{z}_i + H_i^\top \Lambda_i^{-1} X_{i,\bar{i}} \right). \quad (7)$$

In line 5 and 6, we guarantee the norm constraints on V_k and Φ_k by adding an additional quadratic term and set the coefficient ν to guarantee the norm requirements. Although the problem is not convex, our proposed approaches yield optimal solutions for updating different parameter blocks in the sub-routines, and the penalized (marginalized) deviance loss is non-increasing over the iterations.

Theorem 3.1. *In Algorithm 1, let Θ_ℓ and $\Theta_{\ell+1}$ are the estimated parameters at the beginning and end of the ℓ^{th} iteration of the outer **while** loop. We have $J(\Theta^\ell) \geq J(\Theta^{\ell+1})$.*

Proof of Theorem 3.1 is given in Appendix B.1. Derivations and explicit steps for carrying subroutines are deferred to Appendix A.1.

4 Initialization, tuning and testing

4.1 Initialization

One popular initialization approach for PARAFAC decomposition is through the Tucker decomposition $[U_{\perp}, \Phi_{\perp}, V_{\perp}; \mathbf{G}]$ of \mathbf{X} using HOSVD/MLSVD (De Lathauwer et al., 2000) where $\mathbf{G} \in \mathbb{R}^{K_1 \times K_2 \times K_3}$ is the core tensor and $U_{\perp} \in \mathbb{R}^{I \times K_1}$, $\Phi_{\perp} \in \mathbb{R}^{T \times K_2}$, $V_{\perp} \in \mathbb{R}^{J \times K_3}$ are unitary matrices multiplied with the core tensors along the subject, time and feature directions respectively ($K_1 / K_2 / K_3$ is the smallest between K and $I / T / J$), and then perform PARAFAC decomposition on the small core tensor \mathbf{G} (Bro and Andersson, 1998; Phan et al., 2013). Here, we propose to initialize SPACO with Algorithm 2, which combines the aforementioned approach with functional PCA (Yao et al., 2005) to handle sparse longitudinal data. Algorithm 2 proceeds as follows: (1) perform SVD on \mathbf{X}_J to get V_{\perp} ; (2) project \mathbf{X}_J onto each column of V_{\perp} and perform functional PCA to estimate Φ_{\perp} ; (3) a ridge-penalized regression regressing rows of \mathbf{X}_I on $V_{\perp} \otimes \Phi_{\perp}$, and estimate U_{\perp} and \mathbf{G} from the regression coefficients.

Algorithm 2: Initialization of SPACO

- 1 Let V_{\perp} be the top K_3 left singular vectors of \mathbf{X}_J using only the observed columns.
- 2 Set $Y(k) = (Y_1(k), \dots, Y_T(k)) \in \mathbb{R}^{I \times T}$, where $Y_t(k) = \mathbf{X}_{:,t,:}(V_{\perp})_k \in \mathbb{R}^I$.
- 3 Let Φ_{\perp} be the top K_2 eigenvectors from functional PCA on the row aggregation of matrices $Y(k)$ $k = 1, \dots, K_3$. (see Appendix A.2 for details on functional PCA.)
- 4 Let $\tilde{U} = \arg \min_U \{ \|\mathbf{X}_I - U(V_{\perp} \otimes \Phi_{\perp})^{\top}\|_F^2 + \delta \|U\|_F^2 \} \in \mathbb{R}^{I \times K^2}$, where δ is a small regularization parameter to avoid severe over-fitting to the noise.
- 5 Let U_{\perp} be the top K left singular eigenvectors of \tilde{U} , and $\tilde{G} = U_{\perp}^{\top} \tilde{U} \in \mathbb{R}^{K \times K^2}$. Let $\mathbf{G} \in \mathbb{R}^{K \times K \times K}$ be the estimated core array from rearranging \tilde{G} .

6 Let $\sum_{k=1}^K A_k \odot B_k \odot C_k$ be the rank-K CP approximation of \mathbf{G} . Stack these as the columns of $A, B, C \in \mathbb{R}^{K \times K}$, and set $[U, \Phi, V] = [U_{\perp} A, \Phi_{\perp} B, V_{\perp} C]$.

7 For each $k = 1, \dots, K$, rescale the initializers to satisfy constraints on V and Φ .

In a noiseless model with $\delta = 0$ and complete temporal observations, one may replace the functional PCA step of Algorithm 2 with standard PCA. Then $[U, \Phi, V]$ becomes a PARAFAC decomposition of $\frac{1}{1+\delta} X$.

Lemma 4.1. *Suppose $X = \sum_{k=1}^K U_k^* \odot \Phi_k^* \odot V_k^*$ and is completely observed. Replace Φ_{\perp} in Algorithm 2 by the top K eigenvectors of $W = \frac{1}{I} \sum_{k=1}^K Y(k)^{\top} Y(k)$. Then, the outputs U, Φ, V of Algorithm 2 satisfy that $X = (1+\delta) \sum_{k=1}^K U_k \odot \Phi_k \odot V_k$.*

By default, we set $\delta = \frac{1}{\sqrt{J \times T}}$. Proof of Lemma 4.1 is given in Appendix B.2.

4.2 Auto-selection of tuning parameters

Selection of regularizers λ_1 and λ_2 : One possible approach to select the tuning parameters λ_{1k} and λ_{2k} is through cross-validation. However, this can be computationally expensive even when tuning each parameter sequentially. Additionally, determining a suitable set of candidate values for the parameters before running the SPACO algorithm can be challenging. To address these issues, we adopt nested cross-validation, which has been empirically demonstrated to be useful (Huang et al., 2008; Li et al., 2016). Specifically, we tune the parameters within their corresponding subroutines as follows:

(1) In the update for Φ_k , we conduct column-wise leave-one-out cross-validation to select λ_{1k} , where all observations from a single time point are left out (See Appendix A.3.)

(2) In the update for β_k , we perform K-fold cross-validation to select λ_{2k} .

Rank selection: Rank selection can be performed through cross-validation, as suggested in SupCP (Lock and Li, 2018). To reduce computational costs, we adopt two strategies (see Appendix A.4 for more details):

(1) We initialize the cross-validation parameters with estimates obtained from running a full SPACO/SPACO- analysis and only carry out a limited number of iterations to update the parameters. We find that using 5 or 10 iterations is sufficient in our experiments (the default maximum number of iterations is 10).

(2) We start the analysis with the smallest possible rank and gradually increase it. We terminate the analysis when either the cross-validated marginal log-likelihood stops improving or when we reach the maximum rank that we are willing to consider.

4.3 Covariate importance

In Section 5 of our synthetic experiments, we found that incorporating Z could improve the estimation of subject scores when Z strongly influenced the subject scores. This leads us to question how to determine the significance of such covariates when they are present. Here, we consider the construction of approximated p-values from partial independence/marginal independence tests between Z_j and η_k :

$$H_{0k}^{partial} : Z_j \perp\!\!\!\perp \eta_k \mid Z_{j^c}, \quad H_{0k}^{margin} : Z_j \perp\!\!\!\perp \eta_k,$$

both of which are practical interest.

Recap on randomization-based hypothesis testing: Before introducing our proposal, we will revisit concepts of hypothesis testing via randomization in the traditional setting where we investigate the relationship between a response variable y and a covariate variable z .

First, we investigate the marginal independence between y and z , with the null hypothesis $H_0^{margin} : z \perp\!\!\!\perp y$. The randomization test, widely employed for independence testing, offers a valid p-value without assuming a specific dependence

structure between y and z (Fisher, 1936; Pitman, 1937; Efron et al., 2001; Ernst, 2004). Here, we outline the randomization test for marginal independence. Let $t(\mathbf{z}, \mathbf{y})$ be a test statistic where \mathbf{z} and \mathbf{y} are the observation vectors for z and response y . For instance, one may set $t(\mathbf{z}, \mathbf{y}) = |\text{cor}(\mathbf{z}, \mathbf{y})|$. Let $T := t(\mathbf{z}, \mathbf{y})$ and $\underline{T}_b := t(\underline{\mathbf{z}}^b, \mathbf{y})$, for $b = 1, \dots, B$, where $\underline{\mathbf{z}}^b$ is either a permutation of \mathbf{z} or a vector of independent realizations from the marginal distribution of z . There are subtle differences between using $\underline{\mathbf{z}}^b$ generated from a permutation and using one generated from the marginal distribution (Onghena, 2017); however, we will not distinguish them in this context. Under the null hypothesis H_0^{margin} , $T, \underline{T}_1, \dots, \underline{T}_B$ are exchangeable. Thus, for any $\alpha \in (0, 1)$, we have $P(T > t_{1-\alpha}^* | \mathbf{y}) \leq \alpha$, where $t_{1-\alpha}^*$ is the $(1 - \alpha)$ -percentile of the empirical distribution constituted by $(T, \underline{T}_1, \dots, \underline{T}_B)$.

In addition to marginal independence, many statistical inquiries focus on the conditional or partial independence between y and z given additional variable(s) z_o , with the null hypothesis $H_0^{\text{partial}} : z \perp\!\!\!\perp y | z_o$. For illustration, one may consider a linear regression model expressed by the following equation:

$$y = \mathbf{z}\beta + \mathbf{z}_o^\top \gamma + \zeta,$$

where ζ is a mean-zero noise term independent of z and z_o . In recent years, the conditional randomization test has emerged as a popular method for testing the partial independence between variables z and y , given z_o . If the conditional distribution of $z | z_o$ is known, the conditional randomization test generates new copies of z from the distribution $z | z_o$, allowing the construction of randomized test statistics that are exchangeable with the original one under H_0^{partial} . This approach leverages the known conditional distribution to construct valid testing procedures without imposing additional constraints on the relationship between y and (z, z_o) (Candes et al., 2018; Berrett et al., 2020). To be specific, we define the test statistic $t(\mathbf{z}, \mathbf{z}_o, \mathbf{y})$ where \mathbf{z}_o represents the observations for variables z_o . The test statistics $t(\mathbf{z}, \mathbf{z}_o, \mathbf{y})$ can take any function form, e.g., we can set $t(\mathbf{z}, \mathbf{z}_o, \mathbf{y}) = |\text{cor}(\mathbf{z}, \mathbf{y})|$. Let $T := t(\mathbf{z}, \mathbf{z}_o, \mathbf{y})$ and $\underline{T}_b := t(\underline{\mathbf{z}}^b, \mathbf{z}_o, \mathbf{y})$, where $\underline{\mathbf{z}}^b$ is an independent copy of the observation vector generated from the distribution of $z | z_o$, for $b = 1, \dots, B$. Under the null hypothesis H_0^{partial} , the test statistics T and \underline{T}^b have the same conditional

distribution given \mathbf{z}_o and \mathbf{y} . Hence, $P(T > t_{1-\alpha}^* | \mathbf{z}_o, \mathbf{y}) \leq \alpha$, where $t_{1-\alpha}^*$ is the $(1-\alpha)$ -percentile of the conditional distribution of $(T, \underline{T}_1, \dots, \underline{T}_B)$ (Candes et al., 2018).

In this paper, we propose to adapt the idea of randomization tests to construct robust p-values for testing independence/conditional independence between factors and auxiliary covariates in SPACO, where the responses are not given beforehand.

Oracle randomization test in SPACO: Returning to SPACO, we first consider the ideal scenario where V , Φ , s^2 and σ^2 are given. Lemma 4.2 forms the basis of our proposal.

Lemma 4.2. *Given V, Φ, s^2, σ^2 , and let β_ℓ^* be the true regression coefficients on the covariates Z for $\mu_\ell, \ell = 1, \dots, K$. For any $k = 1, \dots, K$, set*

$$\begin{aligned}\underline{\Sigma}_i &= \left((V \odot \Phi)_{\bar{i}}^\top \Lambda_{\bar{i}} (V \odot \Phi)_{\bar{i}} \right)^{-1}, \\ w_i &= s_k^2 + (\underline{\Sigma}_i)_{kk}, \\ y_i &= (\underline{\Sigma}_i)_{k,:} H_{\bar{i},:}^\top \Lambda_{\bar{i}} X_{I,\bar{i}}.\end{aligned}$$

Then, we have $y_i = \mathbf{z}_i^\top \beta_k^ + \xi_i$, where $\xi_i \sim \mathcal{N}(0, w_i)$ is independent of auxiliary covariates.*

Lemma 4.2 is proved via direct calculation in Appendix B.3. Readers may recognize that y and Z match the regression model, potentially allowing the application of the standard randomization technique. Proposition 4.3 details our construction of both observed and randomized test statistics, the validity of which is a result of Lemma 4.2 and Lemma F.1 in Candes et al. (2018). A proof is provided in Appendix B.4 for completeness.

Proposition 4.3. *Set $T_{\text{partial}} = \frac{\sum_i \frac{1}{w_i} (y_i - Z_{i,j^c}^\top \beta_{j^c,k}^*) Z_{ij}}{\sum_i \frac{1}{w_i} Z_{ij}^2}$ for $\beta_{j^c,k}^*$ estimated without the j^{th} covariate Z_j , $T_{\text{margin}} = \frac{\sum_i \frac{1}{w_i} y_i Z_{ij}}{\sum_i \frac{1}{w_i} Z_{ij}^2}$. Replacing Z_j with the properly randomized \underline{Z}_j to*

create \underline{T}_{margin} and $\underline{T}_{partial}$, where \underline{Z}_j is a permutation of Z_j for \underline{T}_{margin} and \underline{Z}_j is independently generated from $Z_j | Z_{j^c}$ for $\underline{T}_{partial}$, we have

$$T_{margin} | \mathbf{y} \stackrel{d}{=} \underline{T}_{margin} | \mathbf{y}, \quad \text{under } H_{0k}^{margin}, \quad (8)$$

$$T_{partial} | (\mathbf{y}, Z_{j^c}) \stackrel{d}{=} \underline{T}_{partial} | (\mathbf{y}, Z_{j^c}), \quad \text{under } H_{0k}^{partial}. \quad (9)$$

Algorithm 3 outlines our proposal. The conditional randomization test involves generating \underline{Z}_j from the conditional distribution $Z_j | Z_{j^c}$, which is estimated using an appropriate exponential family distribution. Further details on the generations of \underline{Z}_j can be found in Appendix A.5.

Algorithm 3: Randomization test for Z_j

1 for $k = 1, \dots, K$ do

3 Construct responses and features as described in Lemma 4.2.

5 Define $\hat{\beta}_k$ by

$$\hat{\beta}_k = \arg \min_{\beta_k, \beta_{k,j}=0} \left\{ \sum_{i=1}^I \frac{1}{w_i} (\mathbf{z}_i^\top \beta_k - \tilde{y}_i)^2 + \lambda_{2k} \|\beta_k\|_1 \right\}.$$

7 Compute the designed test statistic T using $(Z_j, Z_{j^c}, \mathbf{y}, \hat{\beta}_{j^c,k})$.

9 Compute randomized statistics \underline{T}^b using $(\underline{Z}_j^b, Z_{j^c}, \mathbf{y}, \hat{\beta}_{j^c,k})$, where \underline{Z}_j^b for $b = 1, \dots, B$ are (conditionally or marginally) independent copies of Z_j .

11 Let $\hat{G}(\cdot)$ be the empirical estimate of the CDF of T under H_0 using $\{\underline{T}_1, \dots, \underline{T}^B, T\}$, and return the two-sided p-value $p = [1 - \hat{G}(|T|)] + \hat{G}(-|T|)$.

12 end

Approximated p-value construction with estimated model parameters: The true model parameters V , Φ , \mathcal{S}^2 , and σ^2 are unknown, so we will substitute the true parameters with their empirical estimates. However, the empirical estimates from a full SPACO fit tend to suffer from fitting towards the noise. To mitigate such influence, we utilize V, Φ from cross-validation described in Section 4.2. Specifically, for $i \in \mathcal{V}_m$, the index set for fold m , we construct \tilde{y}_i using estimates V^{-m}, Φ^{-m} and the prior covariance estimate Λ_f^{-m} from other folds, by setting $\tilde{y}_i = (\underline{\Sigma}_i)_{k,:} (V^{-m} \odot \Phi^{-m})^\top \Lambda_f^{-m} X_{I,\bar{i}}$ with $\underline{\Sigma}_i$ is also estimated using V^{-m}, Φ^{-m} , and Λ_f^{-m} . This procedure, referred to as cross-fit, employs data from other folds to estimate certain parameters, and forms the final quantity of interest with them and data from current fold. Additionally, initializing each fold at the global solution enhances the comparability of the constructed \tilde{y} and \tilde{z} across different folds. This cross-fit approach yields improved Type I error control compared to the naive incorporation of full estimates in our empirical experiments.

5 Numerical studies with synthetic data

This section presents an evaluation of SPACO using synthetic Gaussian data. The noise variance is fixed at 1, and the true rank is set to $K = 3$. The simulated data consists of (X, Z) with dimensions $(I, T, J, q) \in \{(100, 30, 10, 100), (100, 30, 500, 100)\}$. The observed rate (i.e., $1 -$ the missing rate) along the time dimension is set to $r \in \{100\%, 50\%, 10\%\}$, with observed time stamps chosen randomly for each subject.

To generate the data, we first set $v_{jk} \stackrel{i.i.d}{\sim} \mathcal{N}(0, \frac{1}{J})$ and $z_{i\ell} \stackrel{i.i.d}{\sim} \mathcal{N}(0, 1)$ for

$i = 1, \dots, I, j = 1, \dots, J$ and $\ell = 1, \dots, q$. Then, we set $\phi_1(t) = \theta_1, \phi_2(t) = \theta_2 \sqrt{1 - \left(\frac{t}{T}\right)^2}$, and

$\phi_3(t) = \theta_3 \cos(4\pi \frac{t}{T})$ with random parameters $\theta_1, \theta_2, \theta_3 \sim c_1 \cdot \mathcal{N}(0, \frac{\log J + \log T}{rIT})$ for

$c_1 \in \{1, 3, 5\}$. We also set $\beta_{\ell,k} \sim c_2 \cdot \mathcal{N}(0, \frac{\log q}{I})$ for $c_2 \in \{0, 3, 10\}$ for the first $\ell = 1, 2, 3$, and

set $\beta_{\ell k} = 0$ otherwise. Here, c_1 controls the signal-to-noise ratio (SNR) in the observed tensor X and is referred to as SNR1; c_2 captures the signal-to-noise ratio in the auxiliary covariates Z and is also referred to as SNR2. Each U_k is standardized to have mean 0 and variance 1 after generation. This generates 54 different simulation setups in total.

5.1 Reconstruction quality evaluation

We compare SPACO, SPACO-, plain CP from python package *tensorly*, and a proxy to SupCP by setting $\lambda_{1k} = \lambda_{2k} = 10^{-2}$ in SPACO as small fixed values (additional small penalties enhance numerical stability when dealing with large q and a high missing rate). Although this is not exactly SupCP, it bears high resemblance. This special case of SPACO is referred to as SupCP, which is included to assess gains from smoothness regularization on the time trajectory. Note that SPACO, SPACO-, and SupCP all employ our proposed initialization, improving stability and performance at high missing rates. An analysis contrasting SupCP with random and proposed initialization is presented in Appendix C.1. In our experiments, we used the true rank for all methods. The accuracy the rank estimation procedure proposed in section 4.2 is assessed in Appendix C.3. In our experiments, the rank estimation procedure closely approximates the true model rank when the signal is strong, with a tendency to underestimate in cases of weaker signal. This underestimation does not, however, lead to significant increases in reconstruction loss.

Figure 2 shows the achieved correlation between the reconstructed tensors and the underlying signal tensors across various setups and 20 random repetitions. The correlation between two tensors F and \hat{F} is defined as $\langle F_{demean}, \hat{F}_{demean} \rangle / (\|F_{demean}\|_2 \|\hat{F}_{demean}\|_2)$, where F_{demean} and \hat{F}_{demean} are the demeaned version the original and estimated tensors, respectively, and $\langle F_{demean}, \hat{F}_{demean} \rangle$ is their inner product. Each subplot corresponds to different signal-to-noise ratio combinations (SNR1, SNR2), as indicated by its row and column labels. The y-axis represents the achieved correlation, while the x-axis shows different combinations of J and observation rate. For example, x-axis label $J10_r0.1$ means the feature dimension is 10, and 10% of entries are observed. The “Raw” method, which correlates the true signal with the empirical observation on non-missing entries, is included to demonstrate signal level in different simulation setups, even though it is not directly comparable to other others in the presence of missing data. A comparison of reconstruction quality on missing entries is provided in Appendix C.2.

SPACO achieves better reconstruction than SPACO- when subject scores U depend strongly on Z . This is likely due to improved quality in estimating U . To confirm this, we evaluate the estimation quality of U at $J = 10$ and $SNR2 = 10$ and measure the

estimation quality by R^2 (regressing the true subject scores on the estimated ones). In Figure 3, we show the achieved $(1 - R^2)$ for SPACO and SPACO- (smaller is better), where the x-axis label represents the observing rate and column names represent the component, e.g., $\text{Comp1} \rightarrow U_1$.

SPACO /SPACO- are both top performers for our smooth tensor decomposition problem and achieve significantly better performance than CP and SupCP when the signal is weak and when the missing rate is high by utilizing the smoothness of the time trajectory. To see this, we compared the estimation quality of the time trajectories using SPACO and SupCP. In Figure 4, we show the achieved $(1 - R^2)$ for SPACO and SupCP at $J = 10$. The x-axis label represents different trajectory and observing rate, e.g., $C1_r1.0$ represents estimation of $\phi_1(t)$ at observing rate $r = 1.0$. When the signal is weak ($\text{SNR} = 1$), SPACO can approximate the constant trajectory component ($C1$) and begin to estimate other trajectories successfully as the signal increases. SPACO achieves significantly better estimation of the true underlying trajectories than SupCP for various signal-to-noise ratios.

5.2 Variable importance and hypothesis testing

We investigate the approximated p-values based on cross-fit for testing the partial and marginal associations of Z with U under the same simulation set-ups. Since our variables in Z are generated independently, the two null hypotheses coincide. Nevertheless, the two tests have different powers given the same p-value cut-off, as a result of different test statistics adopted. The proposed randomization tests for SPACO achieve reasonable Type I error controls. Utilizing Cross-fit is important for maintaining good Type I error control. In Appendix C.5, we present qq-plots comparing p-values using cross-fitted V and Φ to the naive construction. The latter exhibits noticeable deviations from the uniform distribution for the later construction when the signal-to-noise ratio is low. Figure 5 and Figure 6 show the achieved Type I error and power with p-value cut-offs at $\alpha = 0.01, 0.05$ and with an observing rate $r = 0.5$. The type I errors are also well controlled for $r \in \{1.0, 0.1\}$ (Appendix C.4).

6 Case study

SPACO was employed to analyze a longitudinal immunological dataset on COVID-19 obtained from the IMPACT study (Lucas et al., 2020). The original dataset comprised 180 samples collected from 98 COVID-19 infected participants, measuring 135 immunological features. Features with more than 20% missing observations were excluded, and the remaining missing values were imputed using MOFA (Argelaguet et al., 2018) (see Appendix D for more details), resulting a complete matrix of 111 features and 180 samples. This matrix was organized into a sparsely observed tensor of size $(I, T, J) = (98, 35, 111)$, where T is the number of unique days from symptom onset (DFSO). The average observing rate along the time dimension was 1.84. Additionally, we had access to auxiliary covariates Z containing eight risk factors (COVID_risk1 - COVID_risk5, age, sex, BMI), as well as four symptom measures (ICU, Clinical score, Length of Stay, Coagulopathy). We ran SPACO with X and Z and SPACO- with X only, selecting a model rank of $K = 3$ based on five-fold cross-validation. The three estimated components are denoted as C1/C2/C3. Integrating static covariates Z with the longitudinal measurements can sometimes improve the estimation quality of subject scores compared to SPACO-, as demonstrated in our synthetic data examples. While true subject scores are unobtainable in the real dataset, the clinical relevance of the estimated subject scores can still be assessed by comparing them with clinical responses.

SPACO and SPACO- yielded overall highly similar estimations and subject scores, as depicted in Figure 7A, with the second component exhibiting the largest discrepancy, yet maintaining a correlation above 0.98. Despite this similarity, we observed a noticeable increase in correlations between C2 from SPACO and ICU/Clinical scores/Length of Stay. The permutation p-values from tests on whether the association between these clinical responses are higher with C2 from SPACO than with C2 from SPACO- are 0.006, 0.008, and 0.002 for ICU, Clinical scores and Length of Stay, respectively, as shown in Figure 7B.

Through the use of the randomization test, we can assess the contribution of each Z_j to C2. Table 1 provides the p-value and adjusted p-value from the conditional independence test, with the number of randomization $B = 2000$. The top associated risk factor is COVIDRISK_3 (hypertension) with a p-value of 0.001 (adjusted p-value around 0.01). BMI is also weakly associated with C2, with a p-value of 0.07. Both

these risk factors displayed much weaker associations with symptom measures when analyzed separately. By including these risk factors in SPACO, the method not only outperforms SPACO-, but also achieves a stronger association with the clinical responses compared to the risk factors themselves (Figure 7B-C).

7 Discussion

We propose SPACO to model sparse multivariate longitudinal data and auxiliary covariates jointly. The smoothness regularization used in SPACO may lead to a significant improvement in estimation quality, particularly when the missing rate is high. Inclusion of informative auxiliary covariates can also enhance the estimation of subject scores. We applied our proposed pipeline to COVID-19 datasets and demonstrated its effectiveness in identifying components with subject scores that are closely associated with clinical outcomes of interest. Moreover, SPACO can identify static covariates that may contribute to severe symptoms. In the future, we plan to extend SPACO to model multi-omics data, characterized by differing data types, scales, and potentially measurement times. Such an extension will require a tailored model design that carefully integrates the different omics data, rather than a naive approach of blindly pooling them together.

8 Acknowledgments

We express our gratitude to the editor, associate editor, and two reviewers for their insightful feedback, which has substantially enhanced the article. L.G. was supported in part by the NSF award DMS-2310836.

9 Disclosure Statement

The author reports there are no competing interests to declare.

SUPPLEMENTARY MATERIALS

Online Appendices:

(Online Appendices.pdf, pdf file) provides additional details of the algorithms, more results from numerical experiments and technical proofs.

Experiment Code:

(SPACOexperiments.zip, zip file) It contains code for reproducing results in synthetic and real data experiments, and organized read data set. A README.md document has been included to for detailed instructions on how to reproduce results represented in this article. Contents in this zip file can also be found at <https://github.com/LeyingGuan/SPACOexperiments>.

Python package for SPACO:

(SPACO.zip, zip file) It contains the Python implementation of the SPACO package. Readers can also find and install SPACO from GitHub at <https://github.com/LeyingGuan/SPACO>.

References

- Acar, E. and B. Yener (2008). Unsupervised multiway data analysis: A literature survey. *IEEE transactions on knowledge and data engineering* 21 (1), 6–20.
- Anderlucci, L. and C. Viroli (2015). Covariance pattern mixture models for the analysis of multivariate heterogeneous longitudinal data. *The Annals of Applied Statistics* 9 (2), 777–800.
- Argelaguet, R., B. Velten, D. Arnol, S. Dietrich, T. Zenz, J. C. Marioni, F. Buettner, W. Huber, and O. Stegle (2018). Multi-omics factor analysis—a framework for unsupervised integration of multi-omics data sets. *Molecular systems biology* 14 (6), e8124.
- Bai, J. and P. Wang (2015). Identification and bayesian estimation of dynamic factor models. *Journal of Business & Economic Statistics* 33 (2), 221–240.
- Bai, J. and P. Wang (2016). Econometric analysis of large factor models. *Annual Review of Economics* 8, 53–80.
- Berrett, T. B., Y. Wang, R. F. Barber, and R. J. Samworth (2020). The conditional permutation test for independence while controlling for confounders. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 82 (1), 175–197.
- Besse, P. and J. O. Ramsay (1986). Principal components analysis of sampled functions. *Psychometrika* 51 (2), 285–311.
- Bro, R. and C. A. Andersson (1998). Improving the speed of multiway algorithms: Part ii: Compression. *Chemometrics and intelligent laboratory systems* 42 (1-2), 105–113.
- Candes, E., Y. Fan, L. Janson, and J. Lv (2018). Panning for gold: ‘model-x’ knockoffs for high dimensional controlled variable selection. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 80 (3), 551–577.

Carroll, J. D., S. Pruzansky, and J. B. Kruskal (1980). Candelinc: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters. *Psychometrika* 45 (1), 3–24.

Chang, W.-C. (1983). On using principal components before separating a mixture of two multivariate normal distributions. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 32 (3), 267–275.

Chen, E. Y., R. S. Tsay, and R. Chen (2020). Constrained factor models for high-dimensional matrix-variate time series. *Journal of the American Statistical Association* 115 (530), 775–793.

De Lathauwer, L., B. De Moor, and J. Vandewalle (2000). A multilinear singular value decomposition. *SIAM journal on Matrix Analysis and Applications* 21 (4), 1253–1278.

Efron, B., R. Tibshirani, J. D. Storey, and V. Tusher (2001). Empirical bayes analysis of a microarray experiment. *Journal of the American statistical association* 96 (456), 1151–1160.

Ernst, M. D. (2004). Permutation methods: a basis for exact inference. *Statistical Science* , 676–685.

Fan, J., Y. Fan, and J. Lv (2008). High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics* 147 (1), 186–197.

Fan, J., Y. Liao, and M. Mincheva (2011). High dimensional covariance matrix estimation in approximate factor models. *Annals of statistics* 39 (6), 3320.

Fisher, R. A. (1936). Design of experiments. *British Medical Journal* 1 (3923), 554.

Harshman, R. A. and M. E. Lundy (1994). Parafac: Parallel factor analysis. *Computational Statistics & Data Analysis* 18 (1), 39–72.

Hinrich, J. L. and M. Mørup (2019). Probabilistic tensor train decomposition. In *2019 27th European Signal Processing Conference (EUSIPCO)*, pp. 1–5. IEEE.

Huang, J. Z., H. Shen, A. Buja, et al. (2008). Functional principal components analysis via penalized rank one approximation. *Electronic Journal of Statistics* 2, 678–695.

Imaizumi, M. and K. Hayashi (2017). Tensor decomposition with smoothness. In *International Conference on Machine Learning*, pp. 1597–1606. PMLR.

Kolda, T. G. and B. W. Bader (2009). Tensor decompositions and applications. *SIAM review* 51 (3), 455–500.

Lam, C., Q. Yao, and N. Bathia (2011). Estimation of latent factors for high-dimensional time series. *Biometrika* 98 (4), 901–918.

Li, G., H. Shen, and J. Z. Huang (2016). Supervised sparse and functional principal component analysis. *Journal of Computational and Graphical Statistics* 25 (3), 859–878.

Lock, E. F. and G. Li (2018). Supervised multiway factorization. *Electronic journal of statistics* 12 (1), 1150.

Lucas, C., P. Wong, J. Klein, T. B. Castro, J. Silva, M. Sundaram, M. K. Ellingson, T. Mao, J. E. Oh, B. Israelow, et al. (2020). Longitudinal analyses reveal immunological misfiring in severe covid-19. *Nature* 584 (7821), 463–469.

Mnih, A. and R. R. Salakhutdinov (2007). Probabilistic matrix factorization. *Advances in neural information processing systems* 20, 1257–1264.

Onghena, P. (2017). Randomization tests or permutation tests? A historical and terminological clarification. In *Randomization, masking, and allocation concealment*, pp. 209–228. Chapman and Hall/CRC.

Phan, A.-H., P. Tichavský, and A. Cichocki (2013). Candecomp/parafac decomposition of high-order tensors through tensor reshaping. *IEEE transactions on signal processing* 61 (19), 4847–4860.

Pitman, E. J. G. (1937). Significance tests which may be applied to samples from any populations. II. The correlation coefficient test. *Supplement to the Journal of the Royal Statistical Society* 4 (2), 225–232.

Rendeiro, A. F., J. Casano, C. K. Vorkas, H. Singh, A. Morales, R. A. DeSimone, G. B. Ellsworth, R. Soave, S. N. Kapadia, K. Saito, et al. (2021). Profiling of immune dysfunction in covid-19 patients allows early prediction of disease progression. *Life Science Alliance* 4 (2).

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

Sorkine, O., D. Cohen-Or, Y. Lipman, M. Alexa, C. Rössl, and H.-P. Seidel (2004). Laplacian surface editing. In *Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing*, pp. 175–184.

Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73 (3), 273–282.

Tipping, M. E. and C. M. Bishop (1999). Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61 (3), 611–622.

Vichi, M., R. Rocci, and H. A. Kiers (2007). Simultaneous component and clustering models for three-way data: within and between approaches. *Journal of Classification* 24 (1), 71–98.

Viroli, C. (2011). Finite mixtures of matrix normal distributions for classifying three-way data. *Statistics and Computing* 21, 511–522.

Wang, D., X. Liu, and R. Chen (2019). Factor models for matrix-valued high-dimensional time series. *Journal of econometrics* 208 (1), 231–248.

Wang, D., Y. Zheng, H. Lian, and G. Li (2021). High-dimensional vector autoregressive time series modeling via tensor decomposition. *Journal of the American Statistical Association* , 1–19.

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

Yokota, T., Q. Zhao, and A. Cichocki (2016). Smooth parafac decomposition for tensor completion. *IEEE Transactions on Signal Processing* 64 (20), 5423–5436.

Accepted Manuscript

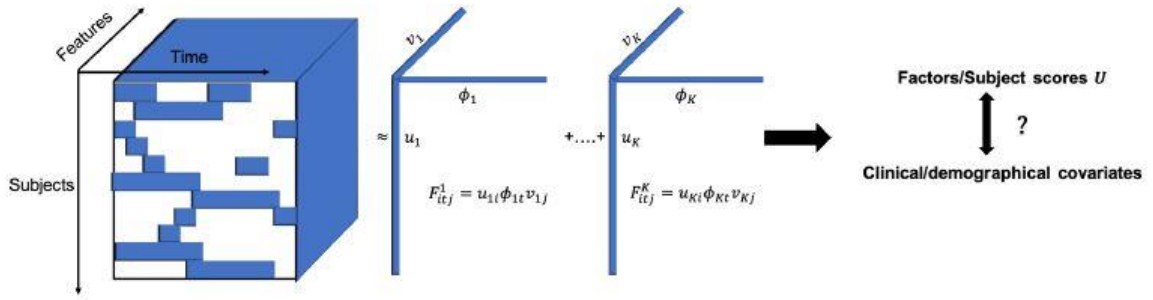


Fig. 1 Illustration of the PARAFAC/CP decomposition method for immunoprofiles measured longitudinally. The white color represents missing data in the time direction. The three-way tensor is approximated as the sum of K rank-one tensors F^1, \dots, F^K . Each rank-one tensor F^k ($k = 1, \dots, K$) can be expressed using a factor along the subject direction (U_k), a factor along the feature direction (Φ_k), and a factor along the time direction (Ψ_k). The factors associated with the tensor decomposition can be used in downstream analysis.

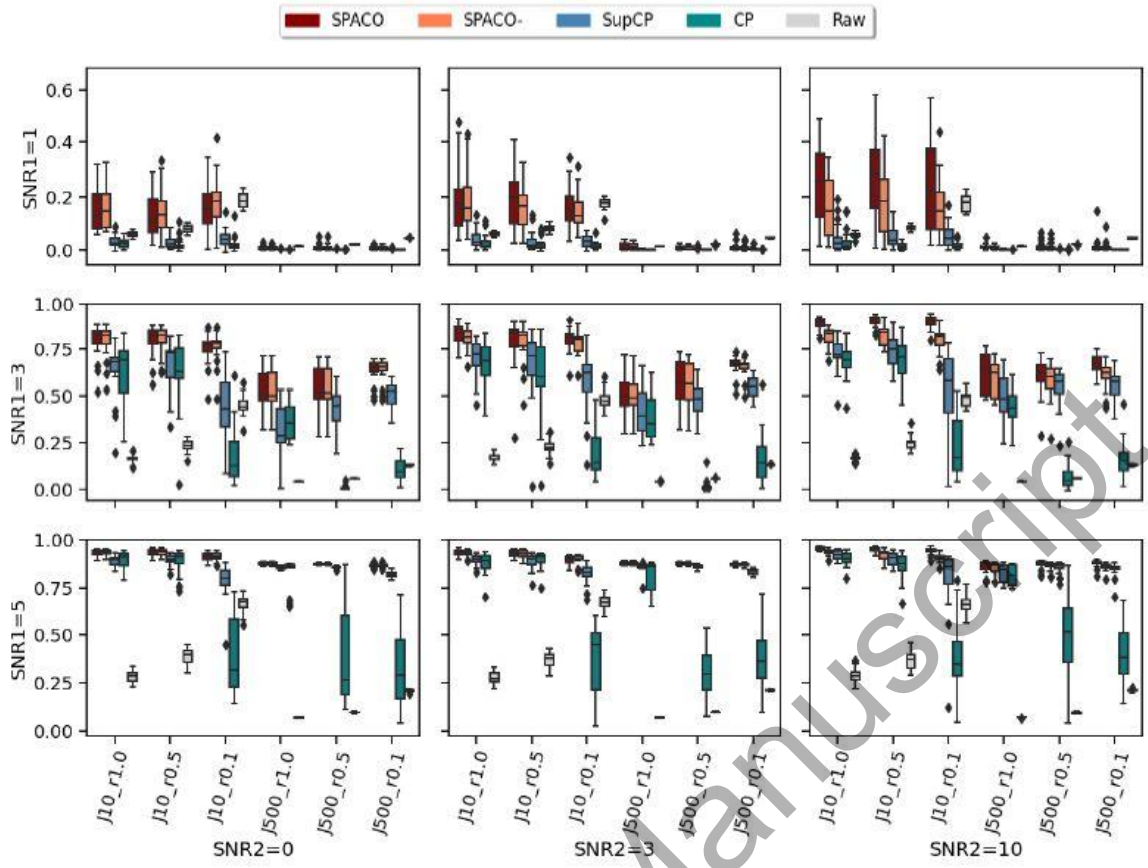


Fig. 2 Reconstruction evaluation by the correlation between the estimates and the true signal tensor. In each subplot, the x-axis label indicates different J and observing rate, the y-axis is the achieved correlation, and the box colors represent different methods. The corresponding subplot column/row name represents the signal-to-noise ratio SNR1/SNR2.

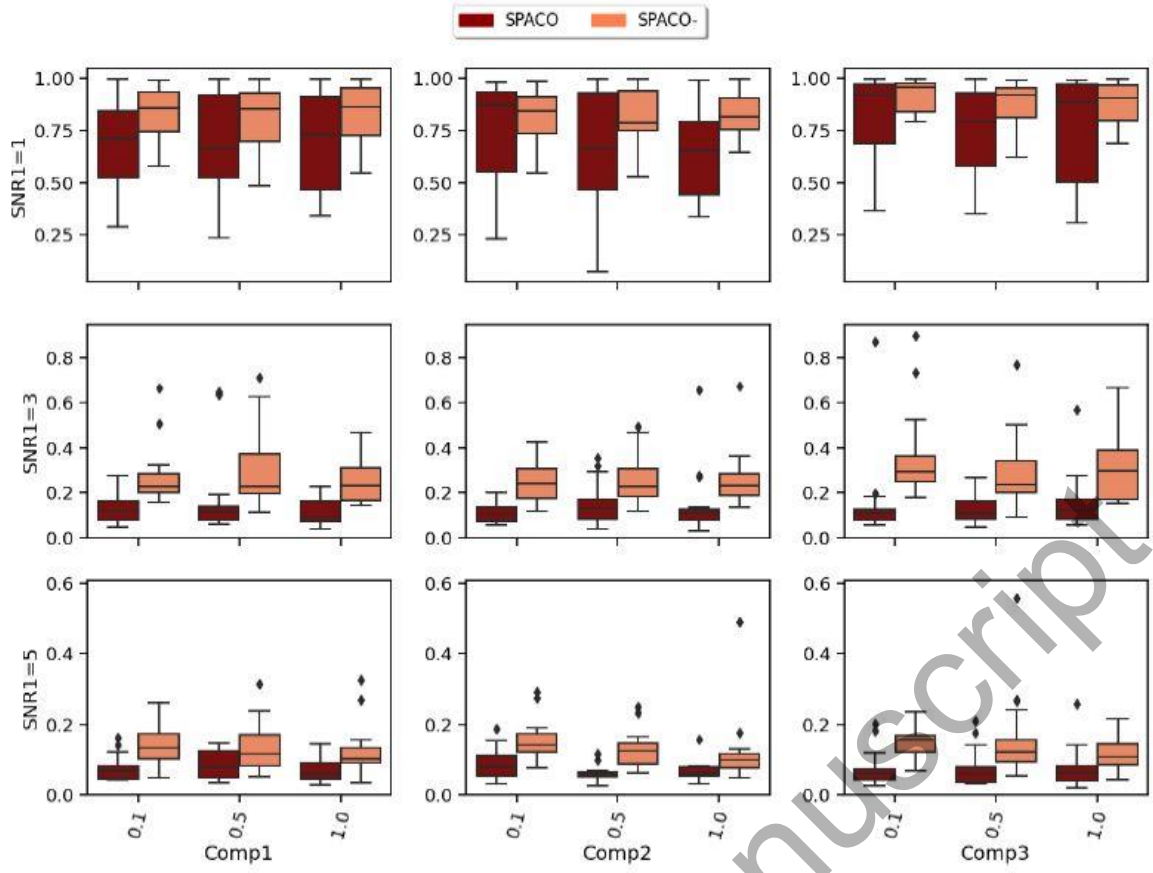


Fig. 3 Comparison of SPACO and SPACO- for reconstructing U at $J=10$, $\text{SNR}_2=10$. In each subplot, the x-axis label indicates different component and observing rate, the y-axis is the achieved $(1 - R^2)$, and the box colors represent different methods. The corresponding subplot column/row name represents the signal-to-noise ratio $\text{SNR}_1/\text{component}$.

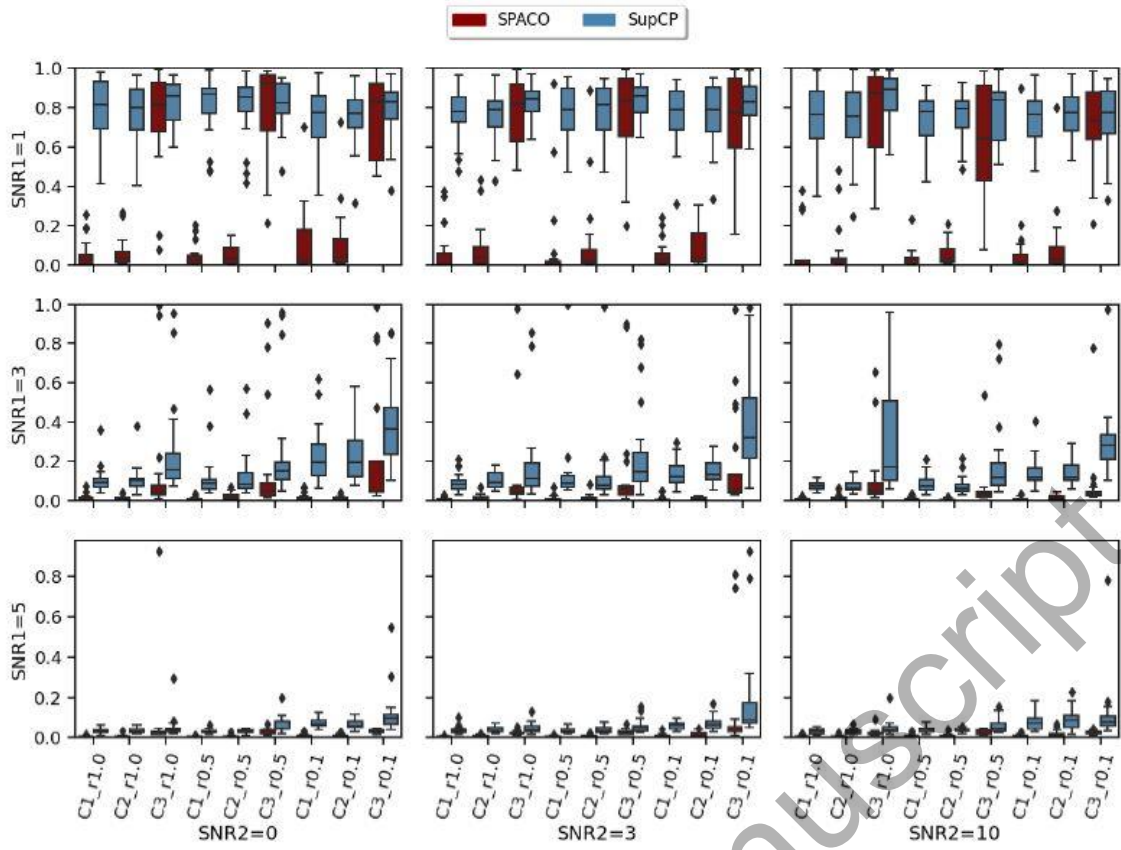


Fig. 4 Comparison of SPACO and SupCP for reconstructing Φ at $J=10$. In each subplot, the x-axis label indicates different component and observing rate, the y-axis is the achieved $(1 - R^2)$, and the box colors represent different methods. The corresponding subplot column/row name represents the signal-to-noise ratio SNR1/SNR2.

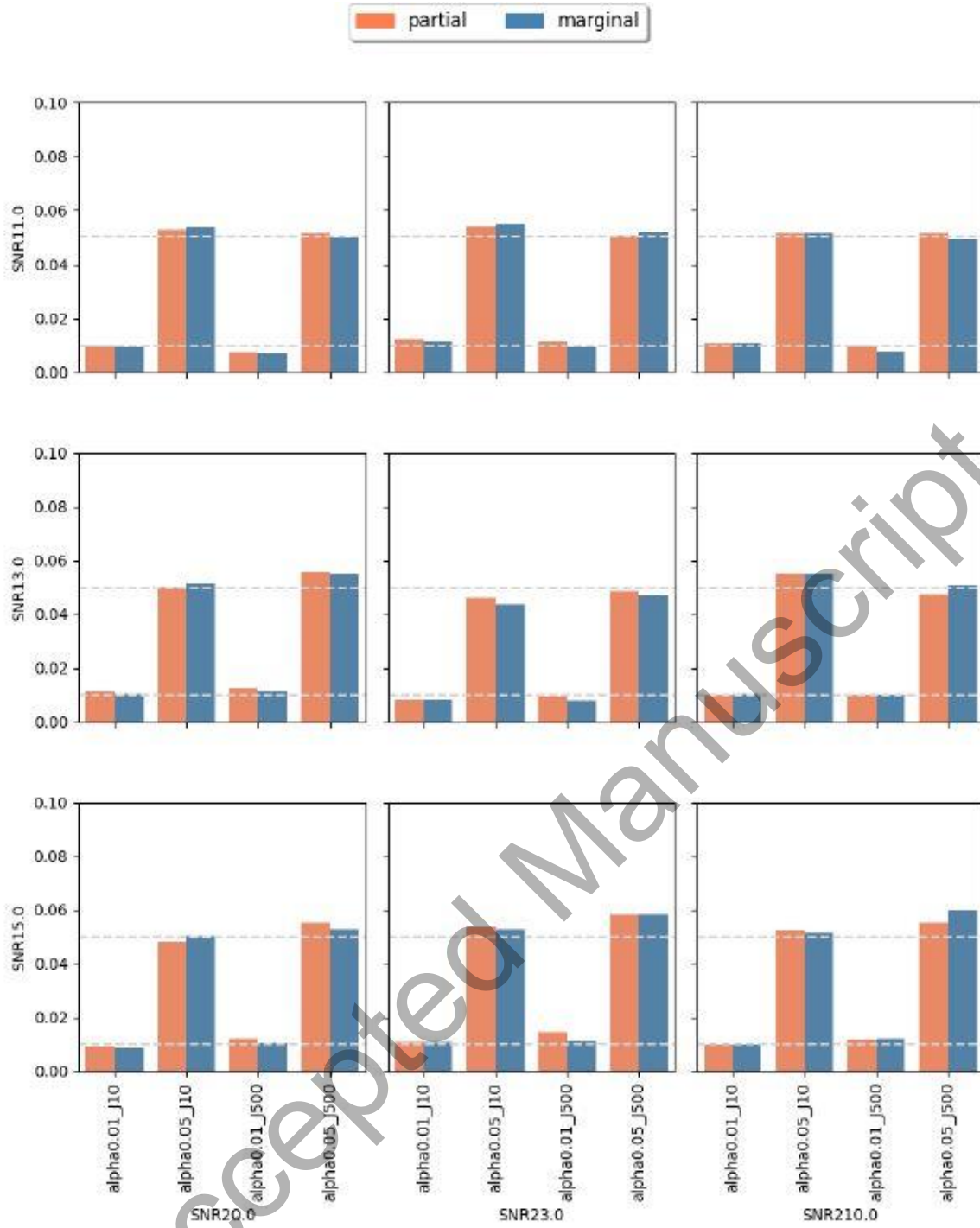


Fig. 5 Achieved type I errors at observing rate $r = 0.5$. In each subplot, x-axis label indicates different combination of feature dimension J and targeted level $\alpha \in \{0.01, 0.05\}$, while the y-axis represents the achieved type I errors. Different bar colors represent different tests (partial or marginal). The two dashed horizontal lines correspond to levels 0.01 and 0.05.

Accepted Manuscript

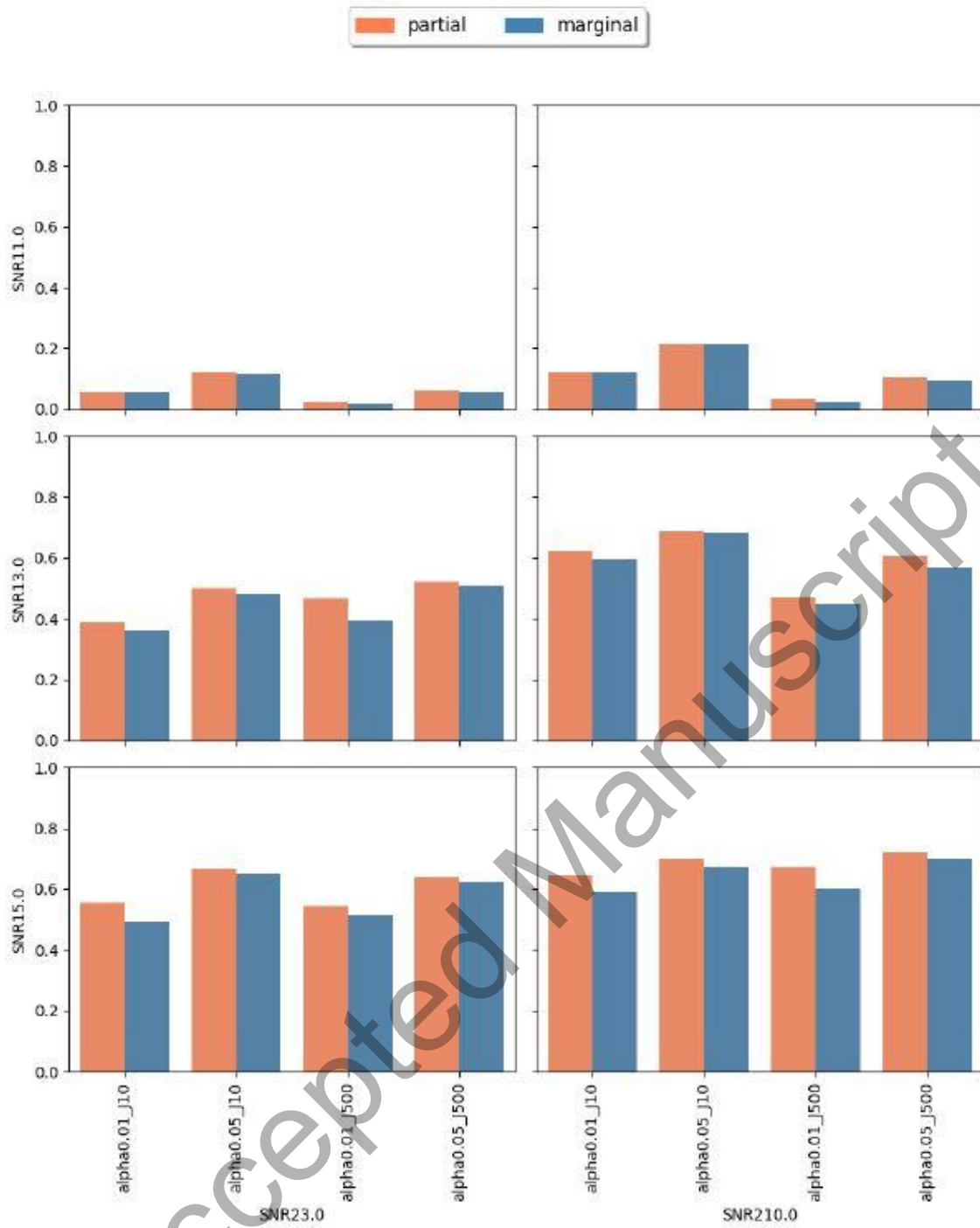


Fig. 6 Achieved power at observing rate $r = 0.5$. In each subplot, the x-axis label indicates different combinations of feature dimension J and targeted level $\alpha \in \{0.01, 0.05\}$, the y-axis indicates the achieved power. Different bar colors represent different tests (partial or marginal).

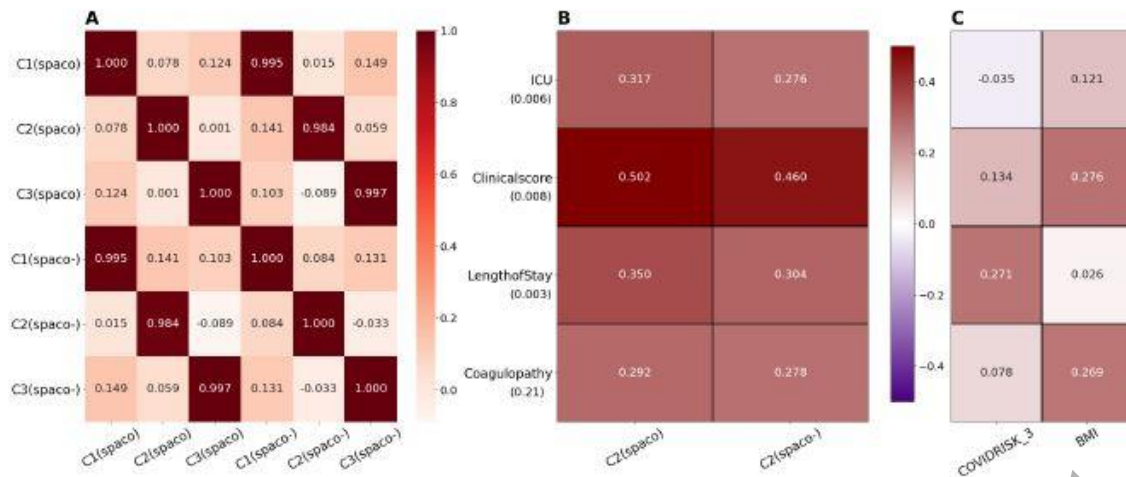


Fig. 7 Comparisons between subject scores estimated from SPACO and SPACO- as well as the static risk factors. Panel A displays the high concordance in the correlations between estimated subject scores from SPACO and SPACO-. Panel B shows the correlations with clinical responses (row label) using subject scores from the most distinct component, C2, estimated with SPACO and SPACO-. The associated permutation p-values, which assess the improvement in correlations with the four clinical responses using SPACO, are shown beneath the row label. Panel C shows the correlation between clinical responses and the two most significant risk factors identified through conditional independence testing, COVIDRISK_3 and BMI.

Table 1 Results from randomization test for the second component (C2). The column nonzero counts the number of 1 for binary covariate COVIDRISK_1 - COVIDRISK_5 and sex (1 = Male, 0 = Female). The adjusted p-values (padj) are based on BH correction.

	COVIDRISK_1	COVIDRISK_2	COVIDRISK_3	COVIDRISK_4	COVIDRISK_5	Age	sex	BMI
pval	0.958	0.809	0.001	0.268	0.73	0.518	0.762	0.07
padj	0.958	0.924	0.009	0.714	1.00	1.000	1.000	0.28
nz_count	7	24	50	23	5	NaN	47	NaN