PARTENSOR project

Deliverable Π2.1: Tensor based models for fMRI data analysis

Authored by

Paris Karakasis (TSI), Athanasios P. Liavas (TSI), Panagiotis Simos (UoC), Efrosini Papadaki (UoC), Thomas Maris (UoC), Dionisios Christopoulos (TSI)

with contributions by

Nicholas Sidiropoulos (UVirginia), Eleftherios Kavroulakis (UoC)

June 2019

Contents

Chapter 1

Introduction

Functional magnetic resonance imaging or functional MRI (fMRI) is a non-invasive functional neuroimaging procedure that measures brain activity by detecting changes associated with blood flow, over time. It relies on the fact that cerebral blood flow and neuronal activation are coupled since, when neural activity increases in an brain area, the metabolic demands in this area rise. Thus, the vascular system concentrates oxygen (oxygenated hemoglobin) into the area.

In fMRI, a brain is represented by a finite set of volume elements (voxels). For each voxel, we have a time series that indicates the concentration of oxygen in this area over time. These time series are known as blood oxygenation level dependent (BOLD) signals. The purpose of task-based fMRI data analysis is to determine which brain areas are activated when a specific task is performed, based on the BOLD signals analysis. Hence, brain activation maps related to specific tasks can be obtained. This procedure is very useful for understanding how the human brain works. Also, the study of how brain activations maps, as well as how activation time patterns change over different trials, can be used for diagnostic purposes. For example, fMRI could provide an in vivo means to investigate alterations in brain function related to the earliest symptoms of Alzheimer's disease, possibly before development of significant irreversible structural damage.

The BOLD signal of a voxel corresponds to oxygenation changes that are not only related to the specific task that we study but also to irrelevant factors. Thus, in order to isolate the signals that we are interested in, we can consider the problem of fMRI data analysis as a blind source separation problem (BSS). BSS refers to the problem of extracting a set of source signals from a set of mixed signals, without using prior information (or with

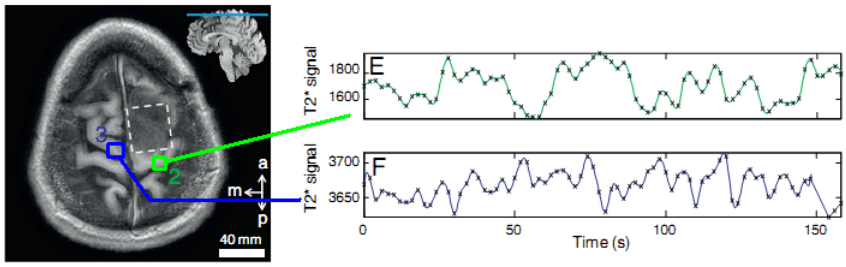


Figure 1.1: An example of fMRI data [1]

.

very little prior information) about the source signals or the mixing process. The classical example of a source separation problem is the cocktail party problem, where a number of people are talking simultaneously in a room (for example, at a cocktail party) and a listener is trying to follow one of the discussions. The human brain can handle this sort of auditory source separation problem, but this is a difficult problem in digital signal processing. Hence, BSS aims in enhancing noisy speech in real world environments and the applications are not just limited to speech/audio processing but also include topics in astronomical, satellite, econometric, and biomedical signal processing.

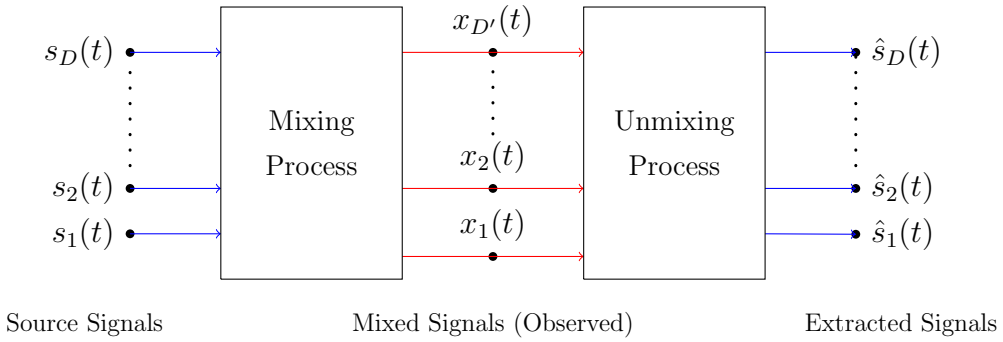


Figure 1.2: Abstract view of the BSS problem.

BSS problems are, in general, highly under-determined, since lack of prior knowledge (number of sources, characteristics of the source signals and the mixing procedure) may lead to a set of multiple solutions for the same problem. A variety of methods in addressing the BSS problem have been proposed in the literature. The most popular of them, among others, are principal component analysis (PCA), independent compontent analysis (ICA), nonnegative matrix factorization (NMF), as well as tensor factorization methods.

Tensors are mathematical objects that have recently gained great popularity due to their ability to model multi-*way* (multidimensional) data dependencies [2], [3], [4]. Tensors often offer more natural representations of data. For example, consider a video recording, which consists of frequently correlated images over time. Tensor factorization (or decomposition) into latent factors is very important for numerous tasks, such as BSS, feature selection, dimensionality reduction, multiway clustering, data visualization and interpretation, and others. The Canonical Decomposition or Canonical Polyadic Decomposition (CANDECOMP or CPD), also known as Parallel Factor Analysis (PARAFAC), and the Tucker decomposition are the two most widely used tensor factorization models.

1.1 Purpose

In this report, we focus on how tensor factorization models can be used in BSS problems and particularly for task-based multi-subject fMRI data analysis. We begin by presenting the matrix factorization problem in the framework of BSS problems. Then, we introduce the problem of tensor factorization under the PARAFAC and PARAFAC2 models and we examine their adequacy for fMRI data analysis, by applying them to real-world fMRI data collected at the University of Crete. Unfortunately, after extensive efforts, we concluded that classical tensor factorization methods do not seem very suitable for processing real-world data fMRI. The main reason for this seems to be the high sensitivity of these methods to the *unknown* tensor rank.

Thus, we tried another approach based on canonical correlation analysis (CCA). The first results are very encouraging. The derived method is very stable and offers very informative fMRI maps.

1.2 Notation

Scalars are denoted by small letters, vectors, matrices, and tensors are denoted by small, capital, and calligraphic capital bold letters, respectively; for example, x, $\mathbf{x}$, $\mathbf{X}$, and $\boldsymbol{\mathcal{X}}$. Sets are denoted by blackboard bold capital letter; for example, $\mathbb{U}$. Specifically, $\mathbb{Z}$, $\mathbb{R}$, and $\mathbb{C}$ denote the sets of integer, real, and complex numbers, respectively. $\mathbb{R}_+$ denotes the set of real nonnegative numbers, while $\mathbb{R}^*$ denotes the set of nonzero real numbers. $\mathbb{R}^{I\times J}$ denotes the set of $(I \times J)$ real matrices. $\mathbb{R}_+^{I\times J}$ denotes the set of $(I \times J)$ real nonnegative

matrices. $\mathbb{C}^{I\times J}$ denotes the set of $(I \times J)$ complex matrices. $\mathbb{R}^{I\times J\times K}$ denotes the set of $(I \times J \times K)$ real tensors. $\mathbb{R}_{+}^{I\times J\times K}$ denotes the set of $(I \times J \times K)$ real tensors. $\|\cdot\|_F$ denotes the Frobenius norm of the tensor or matrix argument. Inequality $\mathbf{A} \geq \mathbf{0}$ means that matrix $\mathbf{A}$ has nonnegative elements. The transpose, conjugate, and hermitian matrices of a matrix $\mathbf{A}$ are denoted by $\mathbf{A}^T$, $\mathbf{A}^*$, and $\mathbf{A}^H$, respectively. The outer product of two vectors $\mathbf{a}$ and $\mathbf{b}$ is denoted as $\mathbf{a} \circ \mathbf{b}$ (see Definition 3.2.1), the Kronecker product of two matrices $\mathbf{A}$ and $\mathbf{B}$ is denoted as $\mathbf{A} \otimes \mathbf{B}$ (see Definition 3.2.2), and the Khatri-Rao product of two matrices $\mathbf{A}$ and $\mathbf{B}$ is denoted as $\mathbf{A} \circledast \mathbf{B}$ (see Definition 3.2.3). Finally, we introduce some Matlab style notations. $\mathbf{A}_{:,l}$ and $\mathbf{A}_{k,:}$ denote the l^{th} column and and the k^{th} row of a matrix $\mathbf{A}$, respectively.

1.3 Report Outline

The report is organized as follows:

- In Chapter 2, we present the matrix factorization model in the context of BSS problems and fMRI data analysis.
- In Chapter 3, we present the PARAFAC and PARAFAC2 models in the framework of multi-subject fMRI data analysis.
- In Chapter 4, we consider the problem of common component extraction via means of canonical correlation analysis. As we shall explain in detail, this approach leads to a very effective data processing and is very promising.
- Finally, in Chapter 5, we conclude the report and describe future research directions.

Chapter 2

Matrix Factorization Models

2.1 Introduction

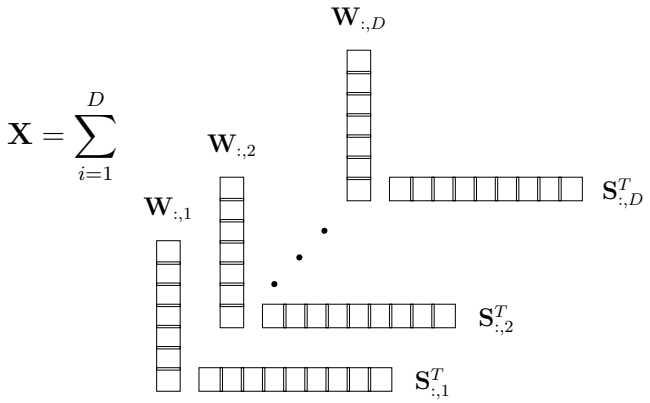


Figure 2.1: Factorization of $\mathbf{X}$ in rank-one component matrices.

The main subject of this chapter is the factorization of a matrix $\mathbf{X} \in \mathbb{R}^{N \times T}$ into a pair of matrices $\mathbf{W} \in \mathbb{R}^{N \times D}$ and $\mathbf{S} \in \mathbb{R}^{T \times D}$, where $D \leq \min\{N, T\}$, such that

$$\mathbf{X} = \mathbf{W}\mathbf{S}^T. \tag{2.1}$$

Matrix factorization (MF) can be employed in many applications, since factors $\mathbf{W}$ and $\mathbf{S}$ may have different interpretations across different frameworks [3]. In a blind source separation problem (BSS), matrix $\mathbf{W}$ plays the role of mixing matrix, while matrix $\mathbf{S}$ expresses the source signals. Specifically, the d-th column of matrix $\mathbf{S}$ expresses the time course of the d-th source signal and d-th column of matrix $\mathbf{W}$ expresses the intensity in

which the d-th source appears to sensors, since

$$\mathbf{X} = \sum_{d=1}^{D} \mathbf{W}_{:,d}\mathbf{S}_{:,d}^{T} = \sum_{d=1}^{D} \mathbf{W}_{:,d} \circ \mathbf{S}_{:,d}, \tag{2.2}$$

where $\circ$ denotes the outer product of two vectors. Thus, each column of $\mathbf{X}$ is reconstructed using a linear combination of the basis elements (columns of $\mathbf{W}$).

In general, MF is more complicated. In many cases of BSS, for example, the number of sources d is unknown. Even if the number of sources D were known, there exist more than one pairs of $\mathbf{W}$ and $\mathbf{S}$ matrices that fulfill equation (2.1). Specifically, for every invertible matrix $\mathbf{\Pi} \in \mathbb{R}^{D\times D}$, if we define $\mathbf{W}' = \mathbf{W\Pi}$ and $\mathbf{S}' = \mathbf{S\Pi}^{-T}$, then

$$\mathbf{W}'\mathbf{S}'^{T} = \mathbf{W\Pi\Pi}^{-1}\mathbf{S}^{T} = \mathbf{WS}^{T} = \mathbf{X}. \tag{2.3}$$

From here on, we will refer to this ambiguity as rotation ambiguity.[1] Also, for every $\alpha \in \mathbb{R}^{*}$, let $\mathbf{W}' = \alpha\mathbf{W}$ and $\mathbf{S}' = \frac{1}{\alpha}\mathbf{S}$. Then

$$\mathbf{W}'\mathbf{S}'^{T} = \alpha\mathbf{W}\frac{1}{a}\mathbf{S}^{T} = \mathbf{WS}^{T} = \mathbf{X}. \tag{2.4}$$

From here on, we will refer to this ambiguity as scalar scaling ambiguity. Therefore, given a matrix $\mathbf{X}$ and the number of sources D, recovering matrices $\mathbf{W}$ and $\mathbf{S}$ is not trivial and further assumptions are required in order to obtain a unique factorization of $\mathbf{X}$.

Consider now a BSS problem where the columns of $\mathbf{X}$ correspond to measurements in different time instances. In this case, we make three implicit assumptions when we select the MF model. The first is that the mixing process is linear, the second is that the mixing matrix $\mathbf{W}$ is time invariant, and the third is that all signals arrive at the sensors at the same time, i.e. without relative delays. In the scope of this report, we study how the MF models can be used for BSS problems under these three assumptions.

2.2 Matrix Factorization Models in fMRI data analysis

In fMRI, the brain of a subject is represented by a finite set of volume elements (voxels). For each voxel, a time series is recorded, during the session, measuring the concentration

[1] A rigorous definition of the rotation ambiguity would require matrix Π to be orthogonal, i.e. $\mathbf{\Pi\Pi}^{T} = \mathbf{\Pi}^{T}\mathbf{\Pi} = \mathbf{I}$.

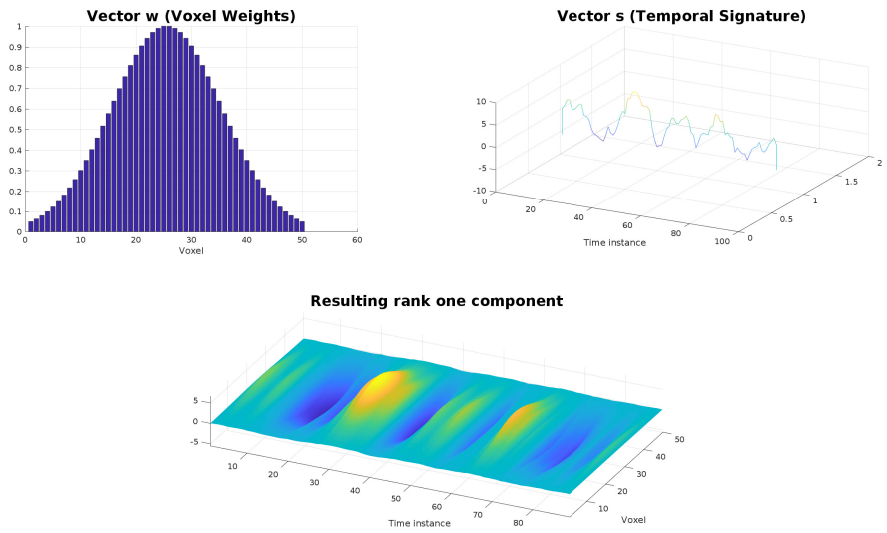


Figure 2.2: Visual example of a rank-one matrix produced by vectors $\mathbf{w}$ and $\mathbf{s}$.

of oxygen in this area over a finite (most often equispaced) set of time points. Hence, fMRI data appear in the form of a four-dimensional array (a fourth-order tensor, as we define in the next chapter), where the first three indices are in correspondence with the spatial coordinates of each voxel, while the fourth index expresses the time domain.

Usually, in fMRI data analysis, the spatial arrangement of the voxels is ignored. Specifically, there exist methods that attempt to define which voxels can be characterized as activated in a voxel-wise fashion, where the presence or the absence of activation are defined by a statistical test. These methods are known as univariate methods. Unlike the univariate methods, multivariate methods provide statistical inference about the whole brain so as to express brain responses in terms of spatial and temporal patterns [5]. Matrix factorization constitutes the basis of many popular multivariate methods, such as Principal Component Analysis (PCA) and Independent Component Analysis (ICA). In MF-based methods, the four-dimensional array is unfolded to the form of a matrix $\mathbf{X} \in \mathbb{R}^{V \times T}$, where V denotes the total number of voxels and T the number of time points. Specifically, by letting vector $\mathbf{x}_j \in \mathbb{R}^T$ denote the time series of the j-th voxel, a matrix $\mathbf{X}$ can be formed as

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_v^T \end{bmatrix}. \tag{2.5}$$

Therefore, by considering each brain voxel as a sensor, a full correspondence with the BSS

framework, as developed in the previous section, emerges, where the MF model can be deployed.

In task-related fMRI experiments, the set of all rank-one components resulting from the factorization of the data matrix can be divided into two categories. The first category contains all rank-one components that are relevant to the task, while the second contains the remaining rank-one components. Thus, a model for this scenario is

$$\mathbf{X} = \mathbf{X}_s + \mathbf{X}_b = \mathbf{W}_s \left[\mathbf{S}_s\right]^T + \mathbf{W}_b \left[\mathbf{S}_b\right]^T = \left[\mathbf{W}_s \ \mathbf{W}_b\right] \left[\mathbf{S}_s \ \mathbf{S}_b\right]^T, \tag{2.6}$$

where term $\mathbf{X}_s$ is used to express the superposition of all task-evoked spatiotemporal responses, while term $\mathbf{X}_b$ expresses the remaining signals that emerge from, irrelevant to the task, background activities. In this setting, one may assume that the different types of stimuli constituting the task are connected with different rank-one components, thereby the number of different types of stimuli is implicitly given by the number of columns of matrix $\mathbf{W}_s$ and by extension of matrix $\mathbf{S}_s$.

Unfortunately, due to the ambiguities concerning that MF model, finding a unique (up to scaling and permutation) set of matrices $\mathbf{W}_s$, $\mathbf{S}_s$, $\mathbf{W}_b$, and $\mathbf{S}_b$ that would explain matrix $\mathbf{X}$ requires prior information to be taken into account, which most of the times is not available. In the next chapter, we discuss how and under which circumstances these difficulties can be circumvented when instead of a matrix $\mathbf{X}$ a set of matrices is available.

Chapter 3

Tensor Factorization Models

3.1 Introduction

In many scientific areas, such as signal processing, neuroimaging, chemometrics, and others, data appear as multidimensional arrays. In this chapter, we study how multidimensional arrays (or tensors) can be processed in order to reveal their hidden structure or patterns across different dimensions. The terms *way* and *order* are used, interchangeably in this report, to express each dimension of an array. In case of arrays in vector form, we have one-*way* array, in case of data in matrix form, we have two-*way* data, etc. In this report, we focus on three-*way* tensors, but the models we discuss can be extended to higher-*way* tensors. An important advantage that comes with multi-*way* factorization models is the existence of tensor decomposition models that do not inherit the rotation ambiguity of the MF model and offer unique decompositions, when some mild conditions hold.

Two-way factorization methods, namely MF models, can be applied to multi-way data processing after reshaping data in a two-way form. However, in this case, data-internal relations become corrupted, as the multidimensional structure is flattened. Another way to apply a two-way factorization model is averaging along trials. This choice makes sense only under the assumption that noise in data is uncorrelated, otherwise information would be lost in the aggregation of correlated noise. Nevertheless, making this choice bring back up the uniqueness issues of matrix factorization.

The two most popular factorization models for N-*way* tensors are the Tucker model and the more restricted PARAFAC model. In this report, we focus on the PARAFAC model. The reason for this choice is that the PARAFAC model comes with theoretical background

that guarantees unique factorizations (up to scaling and permutation) of multi-linear arrays when some mild conditions are satisfied. Another reason is related to the interpretability of the resulting decomposition, since complicated multidimensional patterns can be decomposed to collections of one dimensional features/components under the PARAFAC model.

In this chapter, we introduce the PARAFAC and the PARAFAC2 models and show how they can be used in BSS problems when a set of realizations of the same BSS problem is available. As in the previous chapter, we assume that the mixing process is linear, the mixing matrix is time invariant, and that there are no propagation delays between sources and sensors. Additionally, we assume that the source signals and the mixing matrix are realization invariant. As we discuss later, this assumption can be relaxed by passing from the PARAFAC model to the PARAFAC2 model, when some conditions hold. Next, we demonstrate how the PARAFAC and the PARAFAC2 models can be exploited in the context of fMRI data analysis. At last, we propose a *new* tensor decomposition model for task-related multi-subject fMRI data analysis.

3.2 Definitions

Definition 3.2.1 *Let* $\mathbf{a} \in \mathbb{R}^N$, $\mathbf{b} \in \mathbb{R}^P$, *and* $\mathbf{c} \in \mathbb{R}^J$. *The* ***outer product*** *of* $\mathbf{a}$ *and* $\mathbf{b}$ *is defined as the rank-one matrix with elements*

$$[\mathbf{a} \circ \mathbf{b}]_{n,p} = \mathbf{a}_n \mathbf{b}_p, \tag{3.1}$$

for all $n \in \{1, \ldots, N\}$, $p \in \{1, \ldots, P\}$, *and the outer product of* $\mathbf{a}$, $\mathbf{b}$ *and* $\mathbf{c}$ *is defined as the rank-one tensor with elements*

$$[\mathbf{a} \circ \mathbf{b} \circ \mathbf{c}]_{n,p,j} = \mathbf{a}_n \mathbf{b}_p \mathbf{c}_j, \tag{3.2}$$

for all $n \in \{1, \ldots, N\}$, $p \in \{1, \ldots, P\}$, *and* $j \in \{1, \ldots, J\}$.

Definition 3.2.2 *Let* $\mathbf{A} \in \mathbb{R}^{N \times M}$ *and* $\mathbf{B} \in \mathbb{R}^{P \times K}$. *The* ***Kronecker product*** *(or tensor product) of* $\mathbf{A}$ *and* $\mathbf{B}$ *is defined as the matrix*

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} \mathbf{A}_{1,1}\mathbf{B} & \cdots & \mathbf{A}_{1,M}\mathbf{B} \\ \vdots & \ddots & \vdots \\ \mathbf{A}_{N,1}\mathbf{B} & \cdots & \mathbf{A}_{N,M}\mathbf{B} \end{bmatrix} \in \mathbb{R}^{NP \times MK}. \tag{3.3}$$

Definition 3.2.3 *Let* $\mathbf{A} \in \mathbb{R}^{N\times M}$ *and* $\mathbf{B} \in \mathbb{R}^{P\times M}$. *The* ***Khatri-Rao product*** *of* $\mathbf{A}$ *and* $\mathbf{B}$ *is defined as the matrix*

$$\mathbf{A} \circledast \mathbf{B} = \begin{bmatrix} \mathbf{A}_{:,1} \otimes \mathbf{B}_{:,1} & \cdots & \mathbf{A}_{:,M-1} \otimes \mathbf{B}_{:,M-1} & \mathbf{A}_{:,M} \otimes \mathbf{B}_{:,M} \end{bmatrix}. \tag{3.4}$$

Definition 3.2.4 *Let* $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I\times J\times K}$, $\mathbf{A} \in \mathbb{R}^{I\times F}$, $\mathbf{B} \in \mathbb{R}^{J\times F}$, *and* $\mathbf{C} \in \mathbb{R}^{K\times F}$. *We define the rank of* $\boldsymbol{\mathcal{X}}$ *as the minimum integer positive number* F *such that*

$$\boldsymbol{\mathcal{X}} = \sum_{f=1}^{F} \mathbf{A}_{:,f} \circ \mathbf{B}_{:,f} \circ \mathbf{C}_{:,f}. \tag{3.5}$$

3.3 The PARAFAC model

Consider a collection of K realizations of a BSS problem where, for each realization, a matrix of recordings from N sensors and over T time instances, $\mathbf{X}^k \in \mathbb{R}^{N\times T}$, is obtained, for $k = 1, \ldots, K$. This collection of K matrices forms a three-*way* tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{K\times N\times T}$. When the total number of sources D is not greater than $\min(KN, NT, TK)$, the PARAFAC or CP decomposition of tensor $\boldsymbol{\mathcal{X}}$ can be employed in order to investigate the underlying problem, which is given by

$$\boldsymbol{\mathcal{X}}_{k,n,t} = \sum_{d=1}^{D} \mathbf{A}_{k,d}\mathbf{W}_{n,d}\mathbf{S}_{t,d}, \tag{3.6}$$

for all $k \in \{1, \ldots, K\}$, $n \in \{1, \ldots, N\}$, and $t \in \{1, \ldots, T\}$, where $\mathbf{A} \in \mathbb{R}^{K\times D}$ expresses the realization variability, $\mathbf{W} \in \mathbb{R}^{N\times D}$ expresses the mixing matrix, and $\mathbf{S} \in \mathbb{R}^{M\times D}$ expresses the source signals. Relation (3.6) can also be expressed as

$$\boldsymbol{\mathcal{X}} = [\![\mathbf{A}, \mathbf{W}, \mathbf{S}]\!] = \sum_{d=1}^{D} \mathbf{A}_{:,d} \circ \mathbf{W}_{:,d} \circ \mathbf{S}_{:,d}. \tag{3.7}$$

As can be seen, using the PARAFAC model in BSS problems presupposes that the mixing matrix $\mathbf{W}$ and the source matrix $\mathbf{S}$ are invariant across realizations. Then, $\mathbf{A}_{k,d}$ expresses the contribution of the rank-one matrix $\mathbf{W}_{:,d} \circ \mathbf{S}_{:,d}$ in the k^{th} realization. Physical interpretation of matrix $\mathbf{A}$ is related with the framework under which the realizations happen. For instance, in biomedical BSS applications, each realization may correspond to different subjects or different trials.

The PARAFAC model with constrained factors

For completeness, we mention that the factorization problem of a nonnegative tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}_+^{K\times N\times T}$ using the PARAFAC model into a set of nonnegative factor matrices, i.e.

$$\boldsymbol{\mathcal{X}} = [\![\mathbf{A}, \mathbf{W}, \mathbf{S}]\!] = \sum_{d=1}^{D} \mathbf{A}_{:,d} \circ \mathbf{W}_{:,d} \circ \mathbf{S}_{:,d}, \tag{3.8}$$

where $\mathbf{A} \in \mathbb{R}_+^{K\times D}$, $\mathbf{W} \in \mathbb{R}_+^{N\times D}$ and $\mathbf{S} \in \mathbb{R}_+^{T\times D}$ is called nonnegative tensor factorization problem (NTF).

Another interesting case of constrained tensor factorization is the case of the PARAFAC model with unimodal orthogonality constraints, where a tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{K\times N\times T}$ admits the PARAFAC decomposition with factor matrices $\mathbf{A} \in \mathbb{R}^{K\times D}$, $\mathbf{W} \in \mathbb{R}^{N\times D}$, and $\mathbf{S} \in \mathbb{R}^{T\times D}$, i.e. $\boldsymbol{\mathcal{X}} = [\![\mathbf{A}, \mathbf{W}, \mathbf{S}]\!]$, where one of the factors (for example factor $\mathbf{W}$) also satisfies column-wise orthogonality constraints (namely $\mathbf{W}^T\mathbf{W} = \mathbf{I}_D$).

3.3.1 Estimation of the Factor Matrices

Let a tensor $\boldsymbol{\mathcal{X}}^o \in \mathbb{R}^{K\times N\times T}$ admit the PARAFAC factorization form

$$\boldsymbol{\mathcal{X}}^o = [\![\mathbf{A}^o, \mathbf{W}^o, \mathbf{S}^o]\!] = \sum_{d=1}^{D} \mathbf{A}^o_{:,d} \circ \mathbf{W}^o_{:,d} \circ \mathbf{S}^o_{:,d}, \tag{3.9}$$

where matrices $\mathbf{A}^o \in \mathbb{R}^{K\times D}$, $\mathbf{W}^o \in \mathbb{R}^{N\times D}$, and $\mathbf{S}^o \in \mathbb{R}^{T\times D}$. We observe the noisy tensor $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{X}}^o + \boldsymbol{\mathcal{E}}$, where $\boldsymbol{\mathcal{E}}$ is additive noise. Estimates of $\mathbf{A}^o, \mathbf{W}^o$, and $\mathbf{S}^o$ can be obtained by computing matrices $\mathbf{A} \in \mathbb{R}^{K\times D}, \mathbf{W} \in \mathbb{R}^{N\times D}, \mathbf{S} \in \mathbb{R}^{T\times D}$ that solve the optimization problem

$$\underset{\mathbf{A},\,\mathbf{W},\,\mathbf{S}}{\text{minimize}} \quad f_{\boldsymbol{\mathcal{X}}}(\mathbf{A}, \mathbf{W}, \mathbf{S}), \tag{3.10}$$

where $f_{\boldsymbol{\mathcal{X}}}$ is a function measuring the quality of the factorization. A common choice for $f_{\boldsymbol{\mathcal{X}}}$ is

$$f_{\boldsymbol{\mathcal{X}}}(\mathbf{A}, \mathbf{W}, \mathbf{S}) = \frac{1}{2} \|\boldsymbol{\mathcal{X}} - [\![\mathbf{A}, \mathbf{W}, \mathbf{S}]\!]\|_F^2. \tag{3.11}$$

If $\boldsymbol{\mathcal{X}} = [\![\mathbf{A}, \mathbf{W}, \mathbf{S}]\!]$, then its matrix unfoldings, with respect to the first, second, and third dimension, are given by [3]

$$\mathbf{X}_{(1)} = \mathbf{A}(\mathbf{S} \circledast \mathbf{W})^T, \quad \mathbf{X}_{(2)} = \mathbf{W}(\mathbf{S} \circledast \mathbf{A})^T, \quad \mathbf{X}_{(3)} = \mathbf{S}(\mathbf{W} \circledast \mathbf{A})^T. \tag{3.12}$$

Thus, $f_{\boldsymbol{\mathcal{X}}}$ can be expressed as

$$\begin{aligned} f_{\boldsymbol{\mathcal{X}}}(\mathbf{A},\mathbf{W},\mathbf{S}) &= \frac{1}{2}\left\|\mathbf{X}_{(1)} - \mathbf{A}\left(\mathbf{S}\circledast\mathbf{W}\right)^T\right\|_F^2 \\ &= \frac{1}{2}\left\|\mathbf{X}_{(2)} - \mathbf{W}\left(\mathbf{S}\circledast\mathbf{A}\right)^T\right\|_F^2 \\ &= \frac{1}{2}\left\|\mathbf{X}_{(3)} - \mathbf{S}\left(\mathbf{W}\circledast\mathbf{A}\right)^T\right\|_F^2. \end{aligned} \tag{3.13}$$

These expressions form the basis for tensor factorization via alternating optimization (AO) in the sense that, if we fix two matrix factors, then we can update the third by solving a least-squares problem of the forms

$$\underset{\mathbf{A}}{\text{minimize}} \quad \frac{1}{2}\left\|\mathbf{X}_{(1)} - \mathbf{A}\left(\mathbf{S}\circledast\mathbf{W}\right)^T\right\|_F^2, \tag{3.14}$$

$$\underset{\mathbf{W}}{\text{minimize}} \quad \frac{1}{2}\left\|\mathbf{X}_{(2)} - \mathbf{W}\left(\mathbf{S}\circledast\mathbf{A}\right)^T\right\|_F^2, \tag{3.15}$$

$$\underset{\mathbf{S}}{\text{minimize}} \quad \frac{1}{2}\left\|\mathbf{X}_{(3)} - \mathbf{S}\left(\mathbf{W}\circledast\mathbf{A}\right)^T\right\|_F^2, \tag{3.16}$$

respectively.

For the case of NTF, the nonnegative factors can be estimated by solving the constrained optimization problem

$$\underset{\mathbf{A}\geq 0,\ \mathbf{W}\geq 0,\ \mathbf{S}\geq 0}{\text{minimize}} \quad f_{\boldsymbol{\mathcal{X}}}(\mathbf{A},\mathbf{W},\mathbf{S}). \tag{3.17}$$

Based on the unfolding equations (3.12) and the expressions of $f_{\boldsymbol{\mathcal{X}}}$ in (3.13), an AO procedure can be employed, in which a sequence of nonnegative least-squares (NNLS) problems

$$\underset{\mathbf{A}\geq 0}{\text{minimize}} \quad \frac{1}{2}\left\|\mathbf{X}_{(1)} - \mathbf{A}\left(\mathbf{S}\circledast\mathbf{W}\right)^T\right\|_F^2, \tag{3.18}$$

$$\underset{\mathbf{W}\geq 0}{\text{minimize}} \quad \frac{1}{2}\left\|\mathbf{X}_{(2)} - \mathbf{W}\left(\mathbf{S}\circledast\mathbf{A}\right)^T\right\|_F^2, \tag{3.19}$$

$$\underset{\mathbf{S}\geq 0}{\text{minimize}} \quad \frac{1}{2}\left\|\mathbf{X}_{(3)} - \mathbf{S}\left(\mathbf{W}\circledast\mathbf{A}\right)^T\right\|_F^2, \tag{3.20}$$

will be solved.

At last, when we are interested in solving the PARAFAC model with unimodal orthogonality constraints (w.l.o.g let matrix $\mathbf{W}$ be that factor), estimates of the parameter matrices $\mathbf{A}$, $\mathbf{W}$ and $\mathbf{S}$ can be obtained by solving the constrained optimization problem

$$\begin{aligned} \underset{\mathbf{A},\ \mathbf{W},\ \mathbf{S}}{\text{minimize}} \quad & f_{\boldsymbol{\mathcal{X}}}(\mathbf{A},\mathbf{W},\mathbf{S}) \\ \text{subject to} \quad & \mathbf{W}^T\mathbf{W} = \mathbf{I}_D. \end{aligned} \tag{3.21}$$

As in the previous cases, an AO procedure can be employed for solving a sequence of least-squares problems

$$\underset{\mathbf{A}}{\text{minimize}} \quad \frac{1}{2}\left\|\mathbf{X}_{(1)} - \mathbf{A}\left(\mathbf{S} \circledast \mathbf{W}\right)^T\right\|_F^2, \tag{3.22}$$

$$\begin{aligned} \underset{\mathbf{W}}{\text{minimize}} \quad & \frac{1}{2}\left\|\mathbf{X}_{(2)} - \mathbf{W}\left(\mathbf{S} \circledast \mathbf{A}\right)^T\right\|_F^2 \\ \text{subject to} \quad & \mathbf{W}^T\mathbf{W} = \mathbf{I}_D, \end{aligned} \tag{3.23}$$

$$\underset{\mathbf{S}}{\text{minimize}} \quad \frac{1}{2}\left\|\mathbf{X}_{(3)} - \mathbf{S}\left(\mathbf{W} \circledast \mathbf{A}\right)^T\right\|_F^2, \tag{3.24}$$

respectively. The constrained least-squares problem in (3.23) is known as the orthogonal Procrustes problem and it has a closed-form solution as we show in the sequel.

The Orthogonal Procrustes problem

Let $\mathbf{Y} \in \mathbb{R}^{N\times T}$, $\mathbf{A} \in \mathbb{R}^{N\times D}$ and $\mathbf{X} \in \mathbb{R}^{T\times D}$. We consider the optimization problem

$$\begin{aligned} \underset{\mathbf{A}}{\text{minimize}} \quad & f_{\mathbf{OP}}\left(\mathbf{A}\right) := \left\|\mathbf{Y} - \mathbf{A}\mathbf{X}^T\right\|_F^2 \\ \text{subject to} \quad & \mathbf{A}^T\mathbf{A} = \mathbf{I}_D. \end{aligned} \tag{3.25}$$

Then, the optimal solution $\hat{\mathbf{A}}$ for this problem is given by setting

$$\hat{\mathbf{A}} = \mathbf{U}\mathbf{V}^T, \tag{3.26}$$

where matrices $\mathbf{U} \in \mathbb{R}^{N\times D}$ and $\mathbf{V} \in \mathbb{R}^{D\times D}$ are given by the singular value decomposition of matrix $\mathbf{M} = \mathbf{Y}\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$.

Proof :

$$
\begin{aligned}
\hat{\mathbf{A}} &= \underset{\mathbf{A}}{\operatorname{argmin}} \left\|\mathbf{Y} - \mathbf{A}\mathbf{X}^T\right\|_F^2 \\
&= \underset{\mathbf{A}}{\operatorname{argmin}} \operatorname{Tr}\left(\left(\mathbf{Y} - \mathbf{A}\mathbf{X}^T\right)^T \left(\mathbf{Y} - \mathbf{A}\mathbf{X}^T\right)\right) \\
&\overset{\mathbf{A}^T\mathbf{A}=\mathbf{I}}{=} \underset{\mathbf{A}}{\operatorname{argmin}} \|\mathbf{X}\|_F^2 + \|\mathbf{Y}\|_F^2 - 2\operatorname{Tr}\left(\mathbf{Y}^T\mathbf{A}\mathbf{X}^T\right) \\
&= \underset{\mathbf{A}}{\operatorname{argmax}} \operatorname{Tr}\left(\mathbf{X}^T\mathbf{Y}^T\mathbf{A}\right) \\
&\overset{\mathbf{Y}\mathbf{X}=\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T}{=} \underset{\mathbf{A}}{\operatorname{argmax}} \operatorname{Tr}\left(\mathbf{V}\mathbf{\Sigma}\mathbf{U}^T\mathbf{A}\right) \\
&= \underset{\mathbf{A}}{\operatorname{argmax}} \operatorname{Tr}\left(\mathbf{\Sigma}\mathbf{U}^T\mathbf{A}\mathbf{V}\right) \\
&\overset{\mathbf{A}=\mathbf{U}\mathbf{L}\mathbf{V}^T}{=} \mathbf{U}\left(\underset{\mathbf{L}}{\operatorname{argmax}} \operatorname{Tr}\left(\mathbf{\Sigma}\mathbf{L}\right)\right)\mathbf{V}^T \\
&= \mathbf{U}\left(\underset{\mathbf{L}}{\operatorname{argmax}} \sum_{i=1}^{D} \mathbf{\Sigma}_{i,i}\mathbf{L}_{i,i}\right)\mathbf{V}^T.
\end{aligned}
\tag{3.27}
$$

Since we require $\mathbf{A}^T\mathbf{A} = \mathbf{I}_D$, this implies that $\mathbf{L}^T\mathbf{L} = \mathbf{I}_D \Rightarrow \sum_{k=1}^{D} \mathbf{L}_{k,i}^2 = 1 \ \forall\, i \in \{1, \ldots, D\}$ $\Rightarrow \mathbf{L}_{i,i} \leq 1 \ \forall\, i \in \{1, \ldots, D\}$. Also, we know that $\Sigma_{i,i} \geq 0 \ \forall\, i \in \{1, \ldots, D\}$. Thus, quantity

$$
\sum_{i=1}^{D} \mathbf{\Sigma}_{i,i}\mathbf{L}_{i,i} \tag{3.28}
$$

achieves the maximum for $\mathbf{L} = \mathbf{I}_D$ and we conclude that the optimal solution is $\hat{\mathbf{A}} = \mathbf{U}\mathbf{V}^T$.

3.3.2 Degeneracy and Uniqueness

The problem of finding a best rank-D approximation for third-*order* tensors has no solution, in general. There exists $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ such that

$$
\inf \left\|\boldsymbol{\mathcal{A}} - [\![\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \mathbf{U}^{(3)}]\!]\right\| \tag{3.29}
$$

is not attained by any choice of matrices $\mathbf{U}^{(1)}$, $\mathbf{U}^{(2)}$, $\mathbf{U}^{(3)}$. It is also, in general, not possible to determine a priori if a given $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ will fail to have a best rank-D approximation [6]. Moreover, such failures can occur with positive probability and in some cases with certainty, i.e. where the infimum in (3.29) is never attained. This phenomenon is called PARAFAC degeneracy. Roughly speaking, it refers to solutions in which some component

loadings are highly correlated in all modes and the elements of these components become arbitrarily large [7]. PARAFAC degeneracy makes the estimation unstable, the algorithm slow to converge (or even diverge), and the result difficult to interpret, largely because the model is plagued by strong inter-component cancellations.

Assume that for a tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ there exists a solution of the PARAFAC model, i.e. $\boldsymbol{\mathcal{X}} = \left[\!\left[\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \mathbf{U}^{(3)}\right]\!\right]$ with $\mathbf{U}^{(j)} \in \mathbb{R}^{I_j \times D'}$, for $j = 1, \ldots, 3$. Then, this solution is unique (up to scaling and permutation), when [8]

$$\sum_{j=1}^{3} k_{\mathbf{U}^{(j)}} \geq 2D' + 2, \tag{3.30}$$

where D' is the rank of the tensor and $k_{\mathbf{Y}}$ is the Kruskal rank of a matrix $\mathbf{Y}$, denoting the largest number of columns of $\mathbf{Y}$ that is guaranteed to be linearly independent. Thus, $k_{\mathbf{Y}} \leq rank\,(\mathbf{Y})$.

When one models the data using a low rank approximation ($D < D'$), the above criterion guarantees that the residuals are uniquely defined [9], which leads to essential uniqueness. If $\boldsymbol{\mathcal{A}} = \left[\!\left[\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \mathbf{U}^{(3)}\right]\!\right]$, then essential uniqueness means that $\mathbf{U}^{(1)}, \mathbf{U}^{(2)}$, and $\mathbf{U}^{(3)}$ are unique up to a common permutation and scaling/counter-scaling of columns, i.e. there exists a permutation matrix $\boldsymbol{\Pi}$ and diagonal scaling matrices $\boldsymbol{\Lambda}_1$, $\boldsymbol{\Lambda}_2$, and $\boldsymbol{\Lambda}_3$ such that

$$\hat{\mathbf{U}}^{(1)} = \mathbf{U}^{(1)}\boldsymbol{\Pi}\boldsymbol{\Lambda}_1, \quad \hat{\mathbf{U}}^{(2)} = \mathbf{U}^{(2)}\boldsymbol{\Pi}\boldsymbol{\Lambda}_2, \quad \hat{\mathbf{U}}^{(3)} = \mathbf{U}^{(3)}\boldsymbol{\Pi}\boldsymbol{\Lambda}_3,$$
$$\boldsymbol{\Lambda}_1\boldsymbol{\Lambda}_2\boldsymbol{\Lambda}_3 = \mathbf{I}, \quad \boldsymbol{\mathcal{A}} = \left[\!\left[\hat{\mathbf{U}}^{(1)}, \hat{\mathbf{U}}^{(2)}, \hat{\mathbf{U}}^{(3)}\right]\!\right].$$

The condition (3.10) is sufficient but not necessary for essential uniqueness. This condition does not hold when $D' = 1$. It is also necessary for $D' = 2$ and $D' = 3$ but not for $D' > 3$ [10, 8]. In the presence of noise, if tensor $\boldsymbol{\mathcal{A}}$ for a rank-D does not belong to the degeneration class, all matrices $\mathbf{U}^{(1)}, \mathbf{U}^{(2)}$, and $\mathbf{U}^{(3)}$ will have full rank and uniqueness is guaranteed by proofs given in [11] and [12].

In constrast to the unconstrained PARAFAC model, there exist cases where imposing constraints on the factor matrices can guarantee the existence of an optimal solution, hence degeneration phenomena are fully eliminated. In the case of tensor factorization model (NTF), for any nonnegative tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}_{+}^{d_1 \times d_2 \times d_3}$ and any given $D \in \mathbb{N}$, a best nonnegative rank-D approximation always exists (up to scaling and permutation) in the sense that the infimum in relation (3.9) is attained by some nonnegative tensors $\sum_{d=1}^{D} \mathbf{U}_{:,d}^{(1)} \circ \mathbf{U}_{:,d}^{(2)} \circ \mathbf{U}_{:,d}^{(3)}$

[6]. In [13], Theorem 2 states that if one of the factor matrices $\mathbf{A}, \mathbf{W}, \mathbf{S}$ is constrained to be column-wise orthonormal, then there exists an optimal solution which can be unique under more relaxed conditions than the unconstrained PARAFAC model [10, 14].

Notice that all the above conclusions can be extended, without loss of generality, to higher order tensors (multi-*way* tensors) with the appropriate extension of all above relations.

3.4 The PARAFAC2 model

In the previous section, we introduced the PARAFAC model by considering a collection of K data matrices, $\mathbf{X}^k \in \mathbb{R}^{N \times T}$ for $k = 1, \ldots, K$, that correspond to different realizations of a BSS problem and all together form a three-*way* tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{K \times N \times T}$. By letting tensor $\boldsymbol{\mathcal{X}}$ to admit the PARAFAC model, each matrix $\mathbf{X}^k$ can be expressed as

$$\mathbf{X}^k = \mathbf{W}\mathbf{D}^k\mathbf{S}^T, \tag{3.31}$$

where $\mathbf{D}^k = \mathrm{diag}\left(\mathbf{A}_{k,:}\right)$, for $k = 1, \ldots, K$.

In the PARAFAC2 model, one of the factor matrices is allowed to vary across different slices. For example, choosing factors $\mathbf{A}$ and $\mathbf{W}$ to be invariant across different slices results into expressing matrices $\mathbf{X}^k$ as

$$\mathbf{X}^k = \mathbf{W}\mathbf{D}^k \left[\mathbf{S}^k\right]^T, \tag{3.32}$$

where $\mathbf{D}^k = \mathrm{diag}\left(\mathbf{A}_{k,:}\right)$ and $\mathbf{S}^k \in \mathbb{R}^{T \times D}$, for $k = 1, \ldots, K$. Notice that solving the PARAFAC2 model for tensor $\boldsymbol{\mathcal{X}}$, in the above setting, is equivalent to solving the following MF problem

$$\left[\mathbf{X}^1 \ \ldots \ \mathbf{X}^K\right] = \mathbf{W} \begin{bmatrix} \mathbf{H}^1 \\ \vdots \\ \mathbf{H}^K \end{bmatrix}^T, \tag{3.33}$$

where $\mathbf{H}^k = \mathbf{S}^k\mathbf{D}^k$, for $k = 1, \ldots, K$. Hence, the ambiguities concerning the uniqueness issues of the MF model reappear by passing from the PARAFAC model to the PARAFAC2 model.

In order to maintain the uniqueness properties of the PARAFAC model, it has been proposed to add equality constraints among the cross products $\left[\mathbf{S}^k\right]^T \mathbf{S}^k$. In [15], the authors show that this is equivalent to letting

$$\mathbf{S}^k = \mathbf{B}^k\mathbf{R}, \tag{3.34}$$

where each matrix $\mathbf{B}^k$ is column-wise orthonormal (i.e. $\left[\mathbf{B}^k\right]^T \mathbf{B}^k = \mathbf{I}_D$, for $k = 1, \ldots, K$). Under this set of constraints, relation (3.32) can be rewritten as

$$\mathbf{Y}^k = \mathbf{X}^k \mathbf{B}^k = \mathbf{W}\mathbf{D}^k \mathbf{R}^T, \tag{3.35}$$

which coincides with the PARAFAC model for the tensor that is formed from matrices $\mathbf{Y}^k$.

3.5 Tensor factorization models in fMRI data analysis

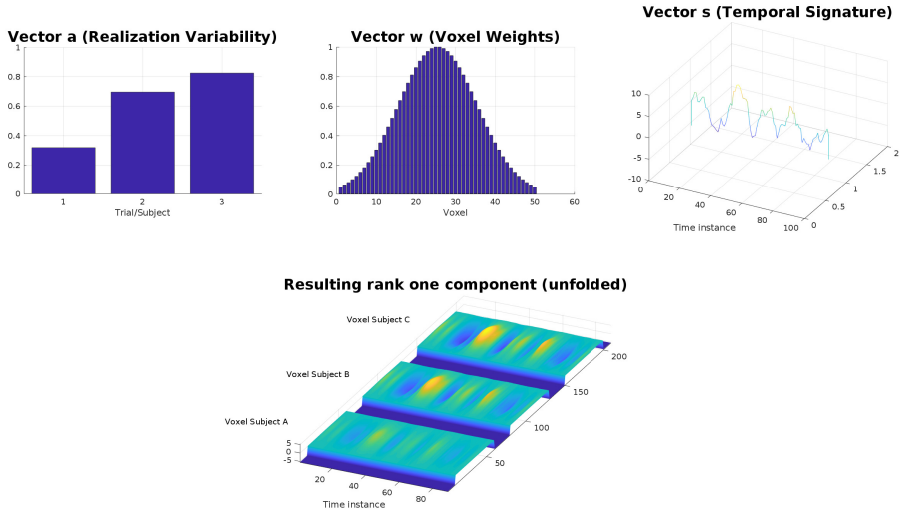


Figure 3.1: Visual example of a rank one tensor produced by vectors $\mathbf{a}$, $\mathbf{w}$ and $\mathbf{s}$.

Tensor factorization models have recently attract the interest of many researchers in the area of fMRI data analysis. In 2004, Andersen and Rayens demonstrated how the PARAFAC (CP) model is useful in the analysis of multi-subject/trial neuroimaging data including task-based fMRI data [16]. Specifically, they proposed to fit the PARAFAC model to the third order tensor $\mathbf{X} \in \mathbb{R}^{K \times N \times T}$ that is formed from the collection of the matricized data (see Chapter 2) of K different subjects/trials. However, degeneration of the unconstrained PARAFAC model is a frequent problem in the analysis of fMRI data.

The same framework was adopted in later proposals, where additional factor constraints were considered. Such proposals, include the tensor PICA model [17], where the ICA model is extended for tensors, as an attempt to increase the robustness of the PARAFAC model and the PARAFAC model with orthogonality constraints over the spatial mode (matrix factor $\mathbf{W}$), in order to reduce cross-talk (overlap) between different spatial components and

eliminate degeneration phenomena. Also, imposing nonnegativity constraints on factor $\mathbf{A}$ (subject/trial variability) is a common practice that aims to improve the interpretability of the resulting factors.

Tensor factorization models have been also proposed for analyzing multi-subject resting-state fMRI data. Specifically, the authors in [18] propose the fitting of the PARAFAC2 model to the multi-subject data tensor, where the temporal factor $\mathbf{S}^k$ varies across subjects, while the cross products $\left[\mathbf{S}^k\right]^T \mathbf{S}^k$ are constrained to be equal. They conclude that PARAFAC2 model is a viable framework for modeling subject variability while, at the same time, identifying consistent and generalizable functional connectivity networks.

3.6 A new model

While functional connectivity networks are often extracted from resting-state fMRI scans, they have been shown to be active during task performance as well [19, 20, 21]. Thus, considering the PARAFAC model seems inappropriate, since components that are related the functional connectivity networks are not taken into account. On the other hand, the terms expressing the responses to the stimuli in a task experiment are not expected to vary significantly over space and time across subjects.

Therefore, we propose the usage of the model that emerges by combining the PARAFAC and PARAFAC2 models. Specifically, let $\{\mathbf{X}_k\}_{k=1}^K$ be a set of matrices, where $\mathbf{X}_k \in \mathbb{R}^{V \times T}$, where V denotes the number of voxels, T denotes the number of time points, and K the number of subjects/trials. Then, the new model takes the form

$$\mathbf{X}^k = \mathbf{X}^k_{st} + \mathbf{X}^k_{sp} = \mathbf{W}_{st}\mathbf{D}^k_{st}\left[\mathbf{S}_{st}\right]^T + \mathbf{W}_{sp}\mathbf{D}^k_{sp}\left[\mathbf{S}^k_{sp}\right]^T, \quad k = 1, \dots K, \tag{3.36}$$

where $\mathbf{D}^k_c = \text{diag}\left(\left[\mathbf{A}_{st}\right]_{k,:}\right)$ and $\mathbf{D}^k_{sp} = \text{diag}\left(\left[\mathbf{A}_{sp}\right]_{k,:}\right)$. The first term expresses the signals that emerge as responses to the stimuli of the experimental task, while the second term expresses the spontaneous signals that emerge from the activation of the functional connectivity networks.

Of course, the model used in relation (3.36) can be viewed as the PARAFAC2 model with equality constraints, i.e.

$$\mathbf{X}^k = \mathbf{W}_t\mathbf{D}^k_t\left[\mathbf{S}^k_t\right]^T = \left[\mathbf{W}_{st}\ \mathbf{W}_{sp}\right]\mathbf{D}^k_t\left[\mathbf{S}_{st}\ \mathbf{S}^k_{sp}\right]^T, \quad k = 1, \dots K, \tag{3.37}$$

where $\mathbf{D}^k_t = \text{diag}\left(\left[\left[\mathbf{A}_{st}\right]_{k,:}\left[\mathbf{A}_{sp}\right]_{k,:}\right]\right)$.

Chapter 4

Common Component Extraction via gCCA

4.1 Introduction

The measured fMRI signal is corrupted by numerous factors as random noise and various nuisance components related to the hardware, like scanner drift and thermal noise, as well as the subjects themselves, due to subject motion, respiration and heartbeat [22]. On top of that, the signal of interest is relatively weak and appears only to a small percentage of the total number of voxels. As a result, although sophisticated preprocessing methods have been developed, the task of recovering useful spatio-temporal components from a whole brain analysis procedure is, at least, challenging.

In this chapter, we focus on recovering components that are common among different subjects/trials. Specifically, we show how this goal can be stated properly in a mathematical form and achieved effectively, via deploying tools that emerge from the theory of canonical correlation analysis.

4.2 Canonical Correlation Analysis

Canonical correlation analysis (CCA) can be seen as the problem of finding basis vectors for two sets of variables such that the correlation between the projections of the variables onto these basis vectors are mutually maximised [23].

Let $\mathbf{X}_1 \in \mathbb{R}^{V \times t_1}$ and $\mathbf{X}_2 \in \mathbf{R}^{V \times t_2}$ be two known full column rank data matrices. CCA

can be defined as the problem of finding two canonical vectors, $\mathbf{a}_1 \in \mathbf{R}^{t_1}$ and $\mathbf{a}_2 \in \mathbf{R}^{t_2}$, such that the canonical variates $\mathbf{z}_1 = \mathbf{X}_1\mathbf{a}_1 \in \mathbb{R}^V$ and $\mathbf{z}_2 = \mathbf{X}_2\mathbf{a}_2 \in \mathbb{R}^V$ are maximally correlated [24], namely

$$(\mathbf{a}_1, \mathbf{a}_2) = \text{argmax} \frac{\mathbf{z}_1^T\mathbf{z}_2}{\|\mathbf{z}_1\|_2 \|\mathbf{z}_2\|_2} = \text{argmax} \frac{\mathbf{a}_1^T\mathbf{C}_{12}\mathbf{a}_2}{\sqrt{\mathbf{a}_1^T\mathbf{C}_{11}\mathbf{a}_1\mathbf{a}_2^T\mathbf{C}_{22}\mathbf{a}_2}}, \tag{4.1}$$

where $\mathbf{C}_{ij} = \mathbf{X}_i^T\mathbf{X}_j$, for $i, j \in \{1, 2\}$, are estimates of the within-set and between-set covariance matrices, respectively. The optimization problem that appears in equation (4.1) is equivalent to the maximization of $\mathbf{q} = \mathbf{a}_1^T\mathbf{C}_{12}\mathbf{a}_2$ subject to the constraints $\mathbf{a}_1^T\mathbf{C}_{11}\mathbf{a}_1 = \mathbf{a}_2^T\mathbf{C}_{22}\mathbf{a}_2 = 1$. The solution to this problem is given by the eigenvector corresponding to the largest eigenvalue of the following generalized eigenvalue problem [25]

$$\begin{bmatrix} \mathbf{0} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{0} \end{bmatrix} \mathbf{a} = q \begin{bmatrix} \mathbf{C}_{12} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{22} \end{bmatrix} \mathbf{a}, \tag{4.2}$$

where q is the canonical correlation and $\mathbf{a} = \left[\mathbf{a}_1^T, \mathbf{a}_2^T\right]^T$ is the eigenvector.

4.2.1 Generalization of CCA to several sets (gCCA)

Any generalization of the CCA to several sets has to be equivalent to the CCA in the case of two sets. Kettenring [26] studied five versions of generalized CCA (gCCA) in which the pair-wise correlation coefficients are maximized. In this manuscript, we focus on the version that is known as the maximum variance formulation of the gCCA (MAXVAR). Specifically, let $\left\{\mathbf{X}^k\right\}_{k=1}^K$ be a set of K full column rank matrices. Then, the MAXVAR formulation of the gCCA is given by the following optimization problem

$$\begin{aligned} \min_{\{\mathbf{Q}^k\}_{k=1}^K, \mathbf{G}} & \sum_{k=1}^K \left\|\mathbf{X}^k\mathbf{Q}^k - \mathbf{G}\right\|_F^2 \\ \text{s.t. } & \mathbf{G}^T\mathbf{G} = \mathbf{I}. \end{aligned} \tag{4.3}$$

In the next section, we show how the MAXVAR formulation of gCCA can be used for the estimation of common subspaces of a given set of matrices.

4.3 Common Subspace Estimation

Let $\left\{\mathbf{X}^k\right\}_{k=1}^K$ be a set of matrices, where $\mathbf{X}^k \in \mathbb{R}^{N\times T}$, for $k \in \{1,\ldots,K\}$, follow the model in relations (3.36) and (3.37), namely,

$$\mathbf{X}^k = \mathbf{W}_t \mathbf{D}_t^k \left[\mathbf{S}_t^k\right]^T, \quad k = 1,\ldots K. \tag{4.4}$$

In order to simplify the analysis that follows, we assume for now that $\mathbf{D}_t^k = \mathbf{I}$, for all $k = 1,\ldots K$. As we show next, this assumption does not affect the concluding results of this section. Therefore, let the generator model of matrices $\left\{\mathbf{X}^k\right\}_{k=1}^K$ be

$$\mathbf{X}^k = \mathbf{W}_t \left[\mathbf{S}_t^k\right]^T, \quad k = 1,\ldots K, \tag{4.5}$$

where $\mathbf{W}_t \in \mathbb{R}^{N\times D}$ and $\mathbf{S}_t^k \in \mathbb{R}^{T\times D}$.

The key observation here is the fact that *the column spaces of all matrices* $\mathbf{X}^k$ *coincide,* since $\text{span}\left(\mathbf{X}^k\right) = \text{span}\left(\mathbf{W}_t\right)$ for all $k = 1,\ldots K$. Let $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{K\times N\times T}$ be the tensor that emerges from the set $\left\{\mathbf{X}^k\right\}_{k=1}^K$ and $\mathbf{X}_{(2)} \in \mathbb{R}^{N\times KT}$ be the matricization of tensor $\boldsymbol{\mathcal{X}}$ with the respect to the second mode. Then, we obtain

$$\mathbf{X}_{(2)} = \left[\mathbf{X}^1 \ \ldots \ \mathbf{X}^K\right] = \mathbf{W}_t \begin{bmatrix} \mathbf{S}_t^1 \\ \vdots \\ \mathbf{S}_t^K \end{bmatrix}^T, \tag{4.6}$$

which implies that the column space of $\mathbf{X}_{(2)}$ will also coincide the $\text{span}\left(\mathbf{W}_t\right)$, since

$$\text{span}\left(\mathbf{X}_{(2)}\right) = \bigcup_{k=1}^K \text{span}\left(\mathbf{X}^k\right) = \text{span}\left(\mathbf{W}_t\right). \tag{4.7}$$

Therefore, all matrices $\left\{\mathbf{X}^k\right\}_{k=1}^K$ generate a common subspace, determined by matrix $\mathbf{W}_t$.

Let us now take into consideration the presence of additive noise, namely,

$$\mathbf{X}^k = \mathbf{W}_t \left[\mathbf{S}_t^k\right]^T + \mathbf{E}^k, \quad k = 1,\ldots K. \tag{4.8}$$

In this case, determining the relations between $\text{span}\left(\mathbf{X}^k\right)$ and $\text{span}\left(\mathbf{W}_t\right)$ is a difficult problem due to the existence of the perturbation terms $\mathbf{E}^k$. Employing one of the previously discussed matrix or tensor factorization models would only guarantee that the estimate of matrix $\mathbf{W}_t$ will belong to the column space of matrix $\mathbf{X}_{(2)}$, i.e. to the union of spaces that are spanned from the columns of each matrix $\mathbf{X}^k$.

In the sequel, we focus on the problem of estimating the column space of matrix $\mathbf{W}_t$, $\text{span}\,(\mathbf{W}_t)$. Notice that the problem of estimating $\text{span}\,(\mathbf{W})$ is related to the estimation of a D–dimensional orthonormal basis in $\mathbb{R}^N$ that generates the same column space with $\mathbf{W}$, namely $\text{span}\,(\mathbf{G}) = \text{span}\,(\mathbf{W})$. Next, we show how solving the optimization problem (4.24), which arises from the MAXVAR formulation of the gCCA, enables us to calculate estimates of common subspaces, like $\text{span}\,(\mathbf{W}_t)$, from a set of matrices $\left\{\mathbf{X}^k\right\}_{k=1}^K$.

In order to understand how this optimization problem is linked with our goals, let matrix $\mathbf{G} \in \mathbb{R}^{N\times D}$ be a D dimensional orthonormal basis in $\mathbb{R}^N$ ($\mathbf{G}^T\mathbf{G} = \mathbf{I}_D$), where $\text{span}\,(\mathbf{G}) = \text{span}\,(\mathbf{W}_t)$. Then, there exists a matrix $\mathbf{\Lambda} \in \mathbb{R}^{D\times D}$ such that $\mathbf{W}_t = \mathbf{G}\mathbf{\Lambda}$. Now, let

$$\mathbf{Q}^k = \left[\mathbf{S}_t^k\right]^\dagger \mathbf{\Lambda}^{-1} = \mathbf{S}_t^k \left(\left[\mathbf{S}_t^k\right]^T \mathbf{S}_t^k\right)^{-1} \mathbf{\Lambda}^{-1}. \tag{4.9}$$

By multiplying each matrix $\mathbf{X}^k$ by $\mathbf{Q}^k$, respectively, we get

$$\begin{aligned}
\mathbf{X}^k\mathbf{Q}^k &= \left(\mathbf{W}_t \left[\mathbf{S}_t^k\right]^T + \mathbf{E}^k\right)\mathbf{Q}^k \\
&= \mathbf{W}_t \left[\mathbf{S}_t^k\right]^T \mathbf{Q}^k + \mathbf{E}^k\mathbf{Q}^k \\
&= \mathbf{G}\mathbf{\Lambda}\mathbf{S}_k^T\mathbf{S}_k^\dagger\mathbf{\Lambda}^{-1} + \mathbf{E}^k\mathbf{Q}^k \\
&= \mathbf{G} + \mathbf{E}^k\mathbf{Q}^k \\
&= \mathbf{G} + \mathbf{E}^{'k}.
\end{aligned} \tag{4.10}$$

Hence,

$$\mathbf{X}^k\mathbf{Q}^k - \mathbf{G} = \mathbf{E}^{'k}, \text{ for all } k = 1, \dots K. \tag{4.11}$$

After considering the Frobenius norms of the above terms, we can see that finding the matrices $\mathbf{G}$ with $\mathbf{G}^T\mathbf{G} = \mathbf{I}_D$ and $\mathbf{Q}^k$, for all $k \in \{1, \dots, K\}$, that solve the optimization problem

$$\begin{aligned}
\min_{\{\mathbf{Q}^k\}_{k=1}^K, \mathbf{G}} & \sum_{k=1}^K \left\|\mathbf{X}^k\mathbf{Q}^k - \mathbf{G}\right\|_F^2 \\
\text{s.t.} \quad & \mathbf{G}^T\mathbf{G} = \mathbf{I}_D
\end{aligned}$$

coincides with the MAXVAR formulation of gCCA (4.24) under the minimum noise energy assumption.

Also notice that, for all $k \in \{1, \dots, K\}$,

$$\text{span}\left(\mathbf{S}_t^k\right) = \text{span}\left(\mathbf{Q}^k\right) = \text{span}\left(\mathbf{J}^k\right), \tag{4.12}$$

where $\mathbf{Q}^k = \mathbf{J}^k\mathbf{U}^k$, with $\left[\mathbf{J}^k\right]^T\mathbf{J}^k = \mathbf{I}_D$. Hence, by solving the optimization problem in relation (4.24) one can obtain estimates about spaces span $(\mathbf{W}_t)$ and span $\left(\mathbf{S}_t^k\right)$, for all $k \in \{1, \ldots, K\}$, while for the case where $D = 1$ one can obtain estimates about factors $\mathbf{W}_t$ and $\mathbf{S}_t^k$, for all $k \in \{1, \ldots, K\}$.

Denoising

In this section, we focus on the practical aspects of why one should be interested in finding spaces span $(\mathbf{W}_t)$ and span $\left(\mathbf{S}_t^k\right)$. One application that is closely related to is noise reduction, since

$$\begin{aligned}\mathbf{G}\mathbf{G}^T\mathbf{X}^k\mathbf{J}^k\left[\mathbf{J}^k\right]^T &= \mathbf{G}\mathbf{G}^T\mathbf{W}_t\left[\mathbf{S}_t^k\right]^T\mathbf{J}^k\left[\mathbf{J}^k\right]^T + \mathbf{G}\mathbf{G}^T\mathbf{E}^k\mathbf{J}^k\left[\mathbf{J}^k\right]^T \\ &= \mathbf{W}_t\left[\mathbf{S}_t^k\right]^T + \mathbf{G}\mathbf{G}^T\mathbf{E}_k\mathbf{J}^k\left[\mathbf{J}^k\right]^T.\end{aligned} \tag{4.13}$$

Now, we adopt the following definition of signal to noise ratio (SNR)

$$\mathrm{SNR}^k = \frac{\left\|\mathbf{W}_t\left[\mathbf{S}_t^k\right]^T\right\|_F^2}{\|\mathbf{E}^k\|_F^2}. \tag{4.14}$$

Since matrix $\mathbf{G}$ forms a D–dimensional orthonormal basis in $\mathbb{R}^N$, every matrix $\mathbf{P} \in \mathbb{R}^{N\times T}$, can be written as $\mathbf{P} = \mathbf{G}\mathbf{G}^T\mathbf{P} + \mathbf{G}_{com}\mathbf{G}_{com}^T\mathbf{P}$, where $\mathbf{G}_{com}$ is an orthogonal basis for the orthogonal complement of span$(\mathbf{G})$.

Then, the following equality holds

$$\|\mathbf{P}\|_F^2 = \left\|\mathbf{G}\mathbf{G}^T\mathbf{P}\right\|_F^2 + \left\|\mathbf{G}_{com}\mathbf{G}_{com}^T\mathbf{P}\right\|_F^2. \tag{4.15}$$

This relation is also valid for $\mathbf{P} = \mathbf{E}^k$. Hence,

$$\left\|\mathbf{E}^k\right\|_F^2 \geq \left\|\mathbf{G}\mathbf{G}^T\mathbf{E}^k\right\|_F^2 \tag{4.16}$$

for all $k \in \{1, \ldots, K\}$. Analogously,

$$\begin{aligned}\left\|\mathbf{G}\mathbf{G}^T\mathbf{E}^k\right\|_F^2 &= \left\|\mathbf{G}\mathbf{G}^T\mathbf{E}^k\mathbf{J}\left[\mathbf{J}^k\right]^T\right\|_F^2 + \left\|\mathbf{G}\mathbf{G}^T\mathbf{E}^k\mathbf{J}_{com}^k\left[\mathbf{J}_{com}^k\right]^T\right\|_F^2 \\ &\geq \left\|\mathbf{G}\mathbf{G}^T\mathbf{E}^k\mathbf{J}\left[\mathbf{J}^k\right]^T\right\|_F^2.\end{aligned} \tag{4.17}$$

Concluding, we have

$$\left\|\mathbf{E}^k\right\|_F^2 \geq \left\|\mathbf{G}\mathbf{G}^T\mathbf{E}^k\mathbf{J}\left[\mathbf{J}^k\right]^T\right\|_F^2, \tag{4.18}$$

which leads to

$$\mathrm{SNR}^k{}_{\text{before projection}} \leq \mathrm{SNR}^k{}_{\text{after projection}}. \tag{4.19}$$

4.4 Common Component Extraction

In this section, we focus on the case where the task-related fMRI experiment of a session consists of only one type of stimulus. Let $\left\{\mathbf{X}^k\right\}_{k=1}^K$ be a set of data matrices collected from K participating individuals, where $\mathbf{X}^k \in \mathbb{R}^{N\times T}$, for $k \in \{1,\dots,K\}$, with N and T denoting the number of voxels and time instances, respectively. Then, it reasonable to use for all matrices $\mathbf{X}^k$ the model in relation (3.36)

$$\mathbf{X}^k = [\mathbf{a}_{st}]_k \, \mathbf{w}_{st}\mathbf{s}_{st}^T + \mathbf{W}_{sp}\mathbf{D}_{sp}^k \left[\mathbf{S}_{sp}^k\right]^T, \quad k = 1,\dots K, \tag{4.20}$$

where $\mathbf{D}_{sp}^k = \text{diag}\left([\mathbf{A}_{sp}]_{k,:}\right)$. Notice that the prior knowledge about the structure of the task is encoded in our model by constraining the first term to have rank equal to one.

Assume that the rank of both matrices $\mathbf{W} = [\mathbf{w}_{st}\ \mathbf{W}_{sp}]$ and $\mathbf{S}^k = \left[\mathbf{s}_{st}\ \mathbf{S}_{sp}^k\right]$ is equal to D. Then, solving the MAXVAR problem in relation (4.6), for matrices $\left\{\mathbf{X}^k\right\}_{k=1}^K$ with $\mathbf{G} \in \mathbb{R}^{N\times D}$, would provide an estimate of the column space of matrix $\mathbf{W}$. On the other hand, a set of matrices $\left\{\mathbf{Q}^k\right\}_{k=1}^K$ would be also provided and due to relation (4.12), it turns out that an estimate of the column space of matrix $\mathbf{S}^k$, for all $k = 1,\dots K$, can be also obtained. More specifically, in the noiseless case, we would ideally have that

$$\text{span}\left(\mathbf{S}^k\right) = \text{span}\left(\mathbf{Q}^k\right), \text{ for } k = 1,\dots,K. \tag{4.21}$$

When,

$$\bigcap_{k=1}^K \text{span}\left(\mathbf{S}^k\right) = \text{span}\left(\mathbf{s}_{st}\right), \tag{4.22}$$

also the following relation should hold

$$\bigcap_{k=1}^K \text{span}\left(\mathbf{Q}^k\right) = \text{span}\left(\mathbf{s}_{st}\right). \tag{4.23}$$

Therefore, solving the following MAXVAR problem

$$\begin{aligned} \min_{\{\mathbf{d}^k\}_{k=1}^K, \mathbf{g}} & \sum_{k=1}^K \left\|\mathbf{Q}^k\mathbf{d}^k - \mathbf{g}\right\|_2^2 \\ \text{s.t. } & \mathbf{g}^T\mathbf{g} = 1, \end{aligned} \tag{4.24}$$

should provide a vector $\mathbf{g}$, which would satisfy

$$\mathbf{g} = \frac{\mathbf{s}_{st}}{\|\mathbf{s}_{st}\|_2}. \tag{4.25}$$

4.5 Application to real-world data

In this section, we test the discussed methods from the previous sections in analyzing real-world task-based fMRI data. Specifically, we process two datasets, recorder at the University of Crete, corresponding to two slightly different conditions. Each dataset corresponds to a collection of task-based fMRI data from 25 subjects in slightly different experimental condition. Next, we quote some information regarding the experimental design and the preprocessing pipeline that was applied on the data. At last, we present the results of our gCCA based data analysis for both conditions.

4.5.1 Experimental design

The fMRI block design consists of two action observation conditions. Indicative specifications are presented below. A different video clip illustrating a two-movement action sequence was presented 6 times within each 35 sec block. The stimulus set-up was identical across the total number of blocks, which was 4, consisting of a fixed red spot at the center of the display, presenting a female person sitting behind a table. A white tea cup is positioned on the table and a ceramic bowl 30 cm in diameter is located on a smaller table right next to the person's head. The experimental conditions examining the effects of an action with the same goal but different kinematics are: (i) Fast to cup-Slow to person: Consists of a rapid grasping movement toward the tea cup (700 ms in duration; average velocity = 0.64 m/sec), followed by a much slower movement that brings the cup to the person's mouth (3300 ms in duration; average velocity = 0.14 m/sec). (ii) Slow to cup-Fast to person: Consists of a slow grasping movement toward the tea cup (3300 ms in duration; average velocity = 0.14 m/sec), followed by a much faster movement that brings the cup to the person's mouth (700 ms in duration; average velocity = 0.64 m/sec). The stimulus layout is identical to (i).

4.5.2 Image acquisition and pre-processing

At first, a standard $T2$ weighted $2D$-TSE-FLAIR ($TR/TE/TI$ = 9000/120/2600 ms) sequence with 4 mm transverse slices was performed to ensure absence of congenital anatomic variations or unexpected pathology. For the BOLD-fMRI, a $T2^*$-weighted, fat-saturated $2D$-FID-EPI sequence was used with the following parameters: repetition time

(TR) 3500 ms, echo time (TE) 50 ms, field of view (FOV) $192 \times 192 \times 108$ (x, y, z), acquisition voxel size $3 \times 3 \times 3$ mm. Whole brain scans consisted of 36 transverse slices with 3.0-mm slice thickness and no interslice gap. Additionally, high resolution anatomical images were acquired sagittally, using a $3D$ magnetization-prepared rapid acquisition gradient echo sequence ($3D$-MPRAGE) with the following parameters: TR 9.8 ms, TE 4.6 ms, flip angle 8 deg, inversion time (TI) 922 ms, FOV 180×230 (x, z), with acquisition voxel size of 0.98×0.98 (x, z) and slice thickness of 1 mm.

For both datasets, image preprocessing was performed in SPM8 (Statistical Parametric Mapping software, SPM: Welcome Department of Imaging Neuroscience, London, UK; available at: http://www.fil.ion.ucl.ac.uk/spm/). Initially, EPI scans were spatially realigned to the first image of the first time series using second degree B-spline interpolation algorithms and motion-corrected through rigid body transformations (three translations and three rotations about each axis). Next, images were spatially normalized to a common brain space (MNI template) and smoothed using an isotropic Gaussian filter (FWHM=8 mm). At last, the SPM platform is able to provide a time response component, based on the activation onsets and offsets, which is expected to appear in the activated brain voxels. From now on, we denote this expected response as $\mathbf{s}_{exp}$, which will be the same for both conditions, since as we mentioned above, the stimulus layout, between the two conditions, is identical.

4.5.3 gCCA based data analysis

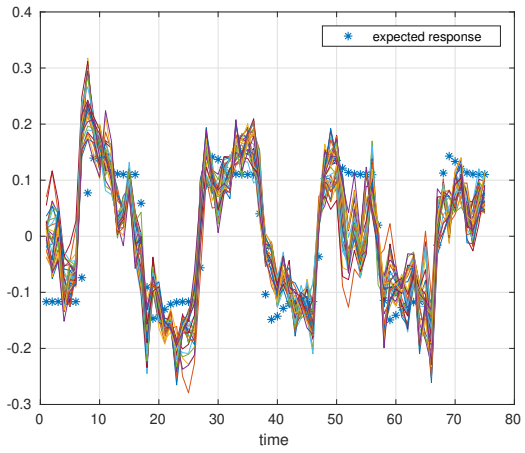
The number of different stimuli, in the experimental design described above, is one, for both conditions. Therefore, we can assume that, in both conditions, the recorded data from the k-th, for $k = 1, \ldots, 25$, should follow the model in relation (4.20), namely

$$\mathbf{X}^k = [\mathbf{a}_{st}]_k \, \mathbf{w}_{st} \mathbf{s}_{st}^T + \mathbf{W}_{sp} \mathbf{D}_{sp}^k \left[\mathbf{S}_{sp}^k\right]^T + \mathbf{E}^k. \tag{4.26}$$

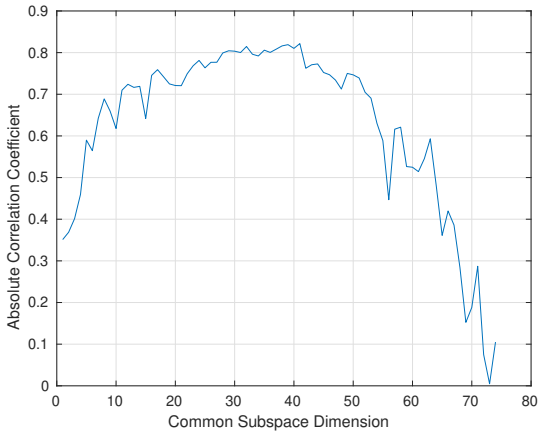
Next, we present the results from our gCCA based analysis on both datasets. In Figures 4.1 and 4.2, we depict

1. (left) the extracted common temporal component $\hat{\mathbf{s}}_{st}$ that emerged for various common subspace dimensions, as well as the expected response $\mathbf{s}_{exp}$.

2. (right) the absolute correlation coefficients between $\mathbf{s}_{exp}$ and the extracted common temporal component $\hat{\mathbf{s}}_{st}$ that emerged for all possible common subspace dimensions.

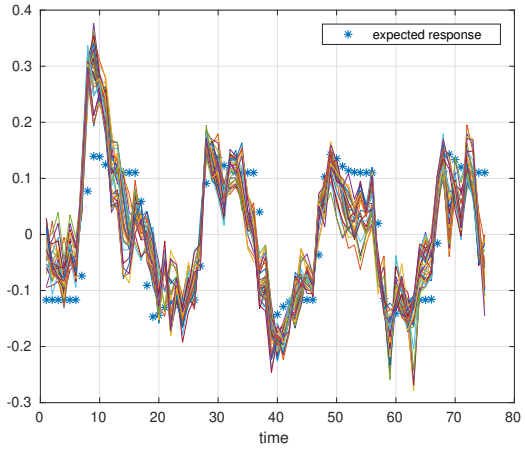


(a) Retrieved $\hat{\mathbf{s}}_{st}$ for varying common subspace dimensions, from 10 to 40, for condition (i). The signal dipictured with blue stars is the $\mathbf{s}_{exp}$.

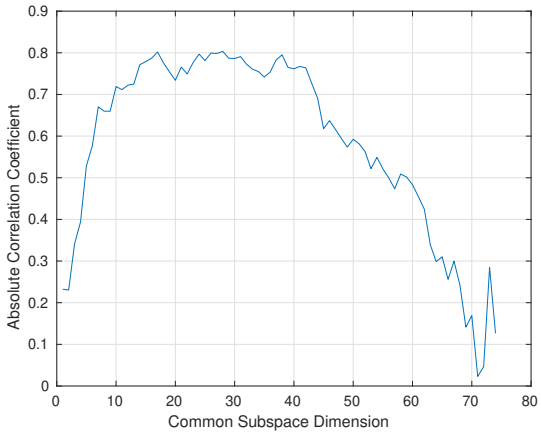


(b) Absolute correlation coefficient between $\mathbf{s}_{exp}$ and $\hat{\mathbf{s}}_{st}$ across different common subspace dimensions.

Figure 4.1: Results from condition (i).



(a) Retrieved $\hat{\mathbf{s}}_{st}$ for varying common subspace dimensions, from 10 to 40, for condition (ii). The signal dipictured with blue stars is the $\mathbf{s}_{exp}$.



(b) Absolute correlation coefficient between $\mathbf{s}_{exp}$ and $\hat{\mathbf{s}}_{st}$ across different common subspace dimensions.

Figure 4.2: Results from condition (ii).

We observe that

1. the estimated common temporal component $\hat{\mathbf{s}}_{st}$ for different common subspace dimensions are very much alike; this implies that our method is *not* sensitive to the common subspace dimension, which is a quantity that is unknown, in general. Thus, we can get useful results over a wide range of values of the dimension of the common subspace.

2. the estimated common temporal components $\hat{\mathbf{s}}_{st}$ are quite similar to the expected signal $\mathbf{s}_{exp}$, since their correlation coefficient takes values at about 0.8, and even higher in some cases.

Consequently, we conclude that our method effectively estimates the common temporal component.

In the sequel, we will use the estimates to derive correlation coefficient maps. Since projecting the data of each subject onto the subspace that is spanned from matrix $\mathbf{G}$ is able to reduce the noise effect (when $\mathbf{G}$ spans that common subspace of interest), in Figures 4.3 and 4.4, we present spatial maps that emerge from calculating the voxelwise correlation coefficient between the denoised data and signal $\hat{\mathbf{s}}_{st}$, for both conditions. Specifically, a correlation coefficient map has been obtained from the data of each subject and then, we created a binary map, which indicates the voxels that have a positive correlation coefficient across all subjects.

For both conditions, we can deduce that voxels from the occipital and parietal lobes present a positive correlation with the estimated common temporal component $\hat{\mathbf{s}}_{st}$, across all subject. Furthermore, we can observe that, for the case of condition (i), voxels from the premotor cortex (PMv area) and inferior frontal gyrus also present a positive correlation with the estimated common temporal component $\hat{\mathbf{s}}_{st}$, across all subject. Both observations are consistent with the nature

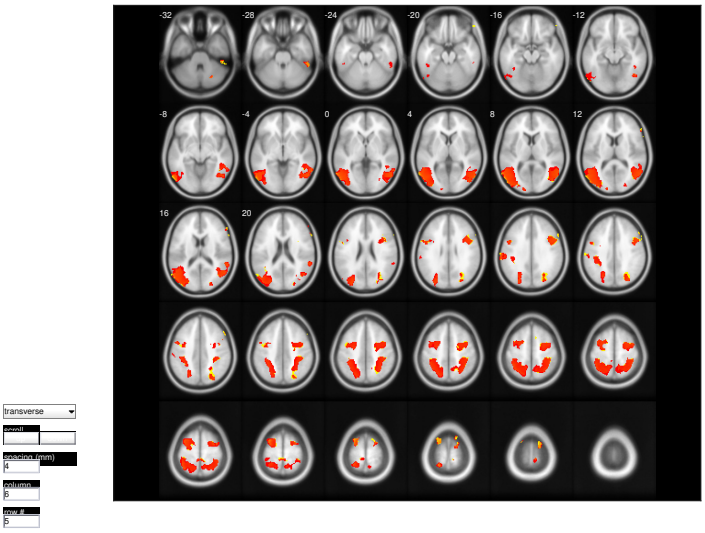


Figure 4.3: Correlation coefficient maps calculated for common subspaces with dimension 19 (red) and 20 (yellow) for condition (i).

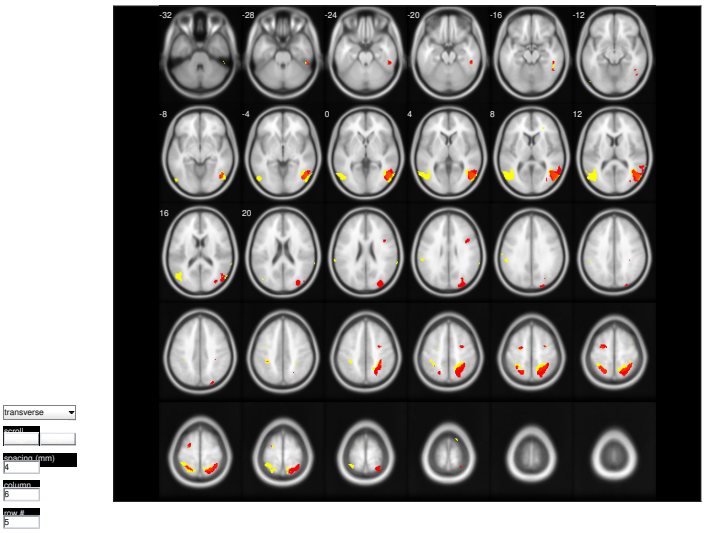


Figure 4.4: Correlation coefficient maps calculated for common subspaces with dimension 14 (red) and 44 (yellow) for condition (ii)

.

Chapter 5

Conclusion

In this report, we studied how tensor decomposition models can be used in analyzing multi-subject/trial fMRI data. We saw that the PARAFAC model is inadequate, since it does not take into account activations that are related with functional connectivity networks. Then, we proposed a constrained version of PARAFAC2 model that alleviates the discussed shortcomings.

In addition, we emphasized the fact that signals of interest are usually weak and appear only in a small percentage of the total brain voxels. Hence, there is the imperative need in devising denoising methods.

Exploiting the fact that the signals of interest are common across subjects, Canonical Correlation Analysis can be employed as a basis for developing efficient algorithms for interesting problems as signal denoising and common component retrieval.

Bibliography

[1] A. Rayshubskiy, T. J. Wojtasiewicz, C. B. Mikell, M. B. Bouchard, D. Timerman, B. E. Youngerman, R. A. McGovern, M. L. Otten, P. Canoll, G. M. McKhann *et al.*, "Direct, intraoperative observation of˜ 0.1 hz hemodynamic oscillations in awake human cortex: implications for fmri," *Neuroimage*, vol. 87, pp. 323–331, 2014.

[2] P. M. Kroonenberg, *Applied multiway data analysis*. John Wiley & Sons, 2008, vol. 702.

[3] A. Cichocki, R. Zdunek, A. H. Phan, and S.-i. Amari, *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*. John Wiley & Sons, 2009.

[4] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.

[5] V. Calhoun, T. Adali, G. Pearlson, and J. Pekar, "Spatial and temporal independent component analysis of functional mri data containing a pair of task-related waveforms," *Human brain mapping*, vol. 13, no. 1, pp. 43–53, 2001.

[6] L.-H. Lim and P. Comon, "Nonnegative approximations of nonnegative tensors," *Journal of chemometrics*, vol. 23, no. 7-8, pp. 432–441, 2009.

[7] A. Stegeman, "Degeneracy in candecomp/parafac and indscal explained for several three-sliced arrays with a two-valued typical rank," *Psychometrika*, vol. 72, no. 4, pp. 601–619, 2007.

[8] N. D. Sidiropoulos and R. Bro, "On the uniqueness of multilinear decomposition of n-way arrays," *Journal of chemometrics*, vol. 14, no. 3, pp. 229–239, 2000.

[9] M. Mørup, L. K. Hansen, S. M. Arnfred, L.-H. Lim, and K. H. Madsen, "Shift-invariant multilinear decomposition of neuroimaging data," *NeuroImage*, vol. 42, no. 4, pp. 1439–1450, 2008.

[10] G. Favier and A. L. de Almeida, "Overview of constrained parafac models," *EURASIP Journal on Advances in Signal Processing*, vol. 2014, no. 1, p. 142, 2014.

[11] R. A. Harshman, "Determination and proof of minimum uniqueness conditions for parafac1," *UCLA Working Papers in phonetics*, vol. 22, no. 111-117, p. 3, 1972.

[12] J. Mocks, "Topographic components model for event-related potentials and some biophysical considerations," *IEEE transactions on biomedical engineering*, vol. 6, no. 35, pp. 482–484, 1988.

[13] W. P. Krijnen, T. K. Dijkstra, and A. Stegeman, "On the non-existence of optimal solutions and the occurrence of "degeneracy" in the candecomp/parafac model," *Psychometrika*, vol. 73, no. 3, pp. 431–439, 2008.

[14] M. Sørensen, L. D. Lathauwer, P. Comon, S. Icart, and L. Deneire, "Canonical polyadic decomposition with a columnwise orthonormal factor matrix," *SIAM Journal on Matrix Analysis and Applications*, vol. 33, no. 4, pp. 1190–1213, 2012.

[15] H. A. Kiers, J. M. Ten Berge, and R. Bro, "Parafac2—part i. a direct fitting algorithm for the parafac2 model," *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 13, no. 3-4, pp. 275–294, 1999.

[16] A. H. Andersen and W. S. Rayens, "Structure-seeking multilinear methods for the analysis of fmri data," *NeuroImage*, vol. 22, no. 2, pp. 728–739, 2004.

[17] C. F. Beckmann and S. M. Smith, "Tensorial extensions of independent component analysis for multisubject fmri analysis," *Neuroimage*, vol. 25, no. 1, pp. 294–311, 2005.

[18] K. H. Madsen, N. W. Churchill, and M. Mørup, "Quantifying functional connectivity in multi-subject fmri data using component models," *Human brain mapping*, vol. 38, no. 2, pp. 882–899, 2017.

[19] Q. R. Razlighi, "Task-evoked negative bold response in the default mode network does not alter its functional connectivity," *Frontiers in computational neuroscience*, vol. 12, 2018.

[20] D. A. Fair, B. L. Schlaggar, A. L. Cohen, F. M. Miezin, N. U. Dosenbach, K. K. Wenger, M. D. Fox, A. Z. Snyder, M. E. Raichle, and S. E. Petersen, "A method for using blocked and event-related fmri data to study "resting state" functional connectivity," *Neuroimage*, vol. 35, no. 1, pp. 396–405, 2007.

[21] M. D. Fox and M. E. Raichle, "Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging," *Nature reviews neuroscience*, vol. 8, no. 9, p. 700, 2007.

[22] M. A. Lindquist *et al.*, "The statistical analysis of fmri data," *Statistical science*, vol. 23, no. 4, pp. 439–464, 2008.

[23] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor, "Canonical correlation analysis: An overview with application to learning methods," *Neural computation*, vol. 16, no. 12, pp. 2639–2664, 2004.

[24] B. Afshin-Pour, G.-A. Hossein-Zadeh, S. C. Strother, and H. Soltanian-Zadeh, "Enhancing reproducibility of fmri statistical maps using generalized canonical correlation analysis in npairs framework," *NeuroImage*, vol. 60, no. 4, pp. 1970–1981, 2012.

[25] K. Mardia, J. Kent, and J. Bibby, "Multivariate analysis. 1979," *Probability and mathematical statistics. Academic Press Inc.*

[26] J. R. Kettenring, "Canonical analysis of several sets of variables," *Biometrika*, vol. 58, no. 3, pp. 433–451, 1971.