KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Borbála Hunyadi, Sabine Van Huffel, Maarten De Vos, (2016), **The power of tensor decompositions in biomedical applications** in Machine Learning in Healthcare Technologies, IET
Archived version	Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher
Published version	Not yet available
Journal homepage	https://eit.europa.eu
Author contact	borbala.hunyadi@esat.kuleuven.be *+ 32 (0)16 32 17 99*
IR	https://lirias.kuleuven.be/handle/123456789/xxxxxx

(article begins on next page)

The power of tensor decompositions in biomedical applications

Borbála Hunyadi, Sabine Van Huffel, Maarten De Vos

June 4, 2016

Contents

1 Introduction to tensors **3**
- 1.1 Tensor representations of multiway data . . . 3
- 1.2 Tensor decomposition techniques . . . 4
  - 1.2.1 Decomposition of matrices . . . 4
  - 1.2.2 Decomposition of tensors . . . 7

2 Construction of tensors in biomedical applications **10**
- 2.1 Naturally occurring tensors . . . 10
  - 2.1.1 Genomic data . . . 10
  - 2.1.2 Repeated multichannel measurements . . . 10
  - 2.1.3 Epoched multichannel measurements . . . 11
- 2.2 Tensor expansion of matrix data . . . 11
  - 2.2.1 Frequency transformation . . . 12
  - 2.2.2 Hankel structure . . . 12
  - 2.2.3 Representation by means of a feature set . . . 13

3 Successful decompositions of biomedical data tensors **14**
- 3.1 Unsupervised tensor decompositions . . . 14
  - 3.1.1 Blind source separation . . . 14
  - 3.1.2 Unsupervised classification . . . 16
- 3.2 Supervised tensor decompositions . . . 16
- 3.3 Coupled tensor decompositions . . . 18
  - 3.3.1 Coupling of multisubject data . . . 19
  - 3.3.2 Temporal coupling . . . 20
  - 3.3.3 Spatial coupling . . . 22

4 Practical considerations **23**
- 4.1 Parameter selection . . . 23
- 4.2 Initialization . . . 24
- 4.3 Tools and algorithms . . . 24

Chapter 1

Introduction to tensors

1.1 Tensor representations of multiway data

Appropriately representating data is crucial for gaining insight and effectively approaching any given problem at hand. We will illustrate in this chapter the use of a particular representation, the tensor or higher order representation, for biomedical data analysis. Tensor algebraic concepts will be introduced by means of easily interpretable examples and visualizations of real world biomedical problems. We also give a formal definition for the most important tensor notions and operations. For a wider number of examples outside biomedical applications, we refer the reader to [15, 30].

A single observation is represented by a scalar value s. For example, the brain potential of a given subject observed at a particular EEG electrode 300ms after a visual stimulus onset can be 1.7uV. However, a single potential value cannot tell much about the overall brain response. We may want to make a series of observations or measure a signal over time, in which case our data is represented by a vector $v \in R^{I_1}$. To continue the previous example, the brain activity at this electrode observed within 0.5s after the stimulus, sampled at 250Hz, results in a vector of length 125. Further, we may want to observe the brain activity at different locations over the scalp. The multichannel EEG signal is now represented in a matrix $B \in \mathbb{R}^{I_1 \times I_2}$ where I_1 is the number of recording electrodes. Repeated experiments can provide information about an even bigger picture, such as the adaptation of the brain to consecutive stimuli. Such data is represented in a third order array or tensor $T \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, where I_3 is the number of consecutive experiments. The concept can be generalized to even higher order tensors $T \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_n}$ by explicitly including more *modes*, for example when performing EEG measurements in different subjects, under different recording conditions, etc.

It is clear that in many situations the observed data naturally takes the form of a tensor. One may be tempted to store and handle such data as a series of matrices. *Matrix unfolding* (see Figure 1.1) may allow easier visualization on a 2D screen, and can be manipulated by means of well-established linear algebra tools. However, persisting with the original tensor representations can be very beneficial for the following two reasons.

First, a tensor representation allows to preserve some crucial information

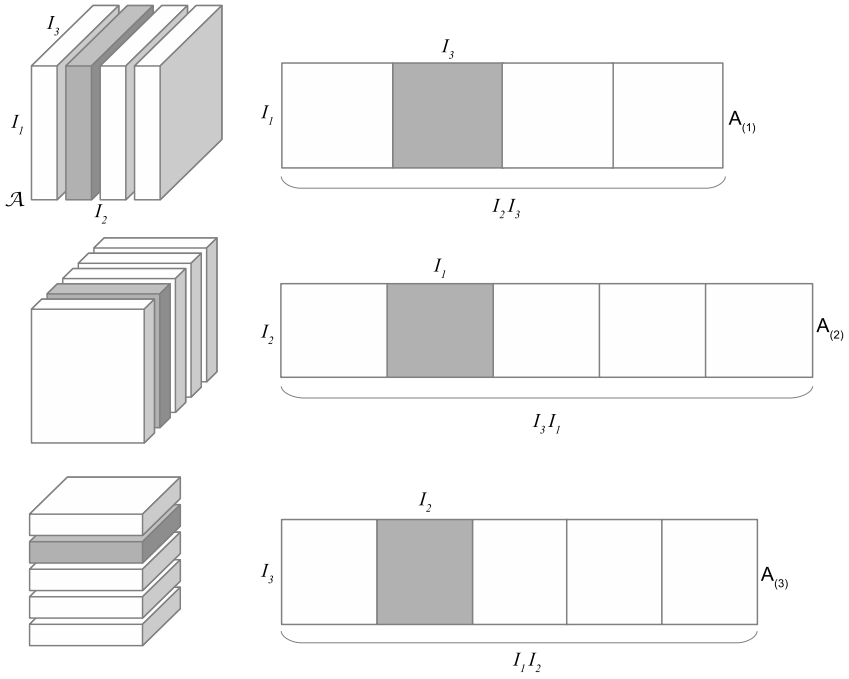


Figure 1.1: The 3 possible unfoldings of a 3D tensor into matrices along the 3 different modes.

residing in the multiway structure. For example, the brain response in a certain cognitive task may be modulated by different levels of difficulty, and may show differences in various pathological conditions. These two effects can be studied separately using different matrix representations, i.e. with a matrix which stores the brain response of each patient in each row, and another matrix with the brain responses at each difficulty level in each row, respectively. However, one should realize that the brain responses may be modulated differently in the different pathological groups. The interaction of these two effects will be hidden in a traditional matrix decomposition while it will become clear when studying the data in the inherent tensor representation, where the brain responses are stored along the *rows*, the patients are organized along the *columns* and the different difficulty levels are organized along the *tubes* (Figure 1.2).

The second motivation for holding on to tensor representation is related to some interesting mathematical properties of tensor manipulation techniques. We are going to demonstrate these in the following section.

1.2 Tensor decomposition techniques

1.2.1 Decomposition of matrices

Some of very common challenges in biomedical data processing include the high dimensionality of the data and low signal to noise ratios, due to the presence of measurement and physiological noise, superimposed on the signals under study.

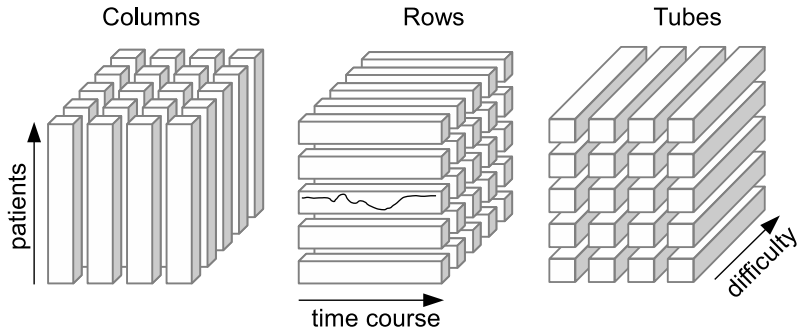


Figure 1.2: The data in a 3D tensor are arranged along the different modes in so-called rows, columns and tubes. In biomedical applications, the variation in different physical quantities are represented in each mode. For example, in a cognitive EEG experiment the time course of the EEG response is stored along the direction of the rows, data from different patients are organized along the columns and the different difficulty levels are organized along the tubes.

A basic concept in linear algebra is the Singular Value Decomposition (SVD), which might offer some benefit when tackling these problems. Formally, the SVD of a matrix $M \in \mathbb{R}^{I_1 \times I_2}$ is the factorization:

$$M = USV^T, \tag{1.1}$$

Where $U \in \mathbb{R}^{I_1 \times I_1}$ and $V \in \mathbb{R}^{I_2 \times I_2}$ are orthogonal matrices, V^T is the transpose of V and S is a non-negative diagonal matrix. The columns of U and V, denoted by $\mathbf{u}_i$ and $\mathbf{u}_i$, are called the left and right singular vectors of M, respectively. The diagonal elements of S, denoted as s_i, appear in decreasing order and are called the singular values of M. Given that the singular values are distinct, the decomposition is unique up to the joint reflection of u_i and v_i. The number of non-zero singular values is equal to the rank of the matrix, i.e. to the number of linearly independent columns/rows of the matrix. M can also be written as the weighted sum of the outer products of the singular vectors:

$$M = \sum \sigma_r \mathbf{u}_r \circ \mathbf{v}_r \tag{1.2}$$

Notice that each term in the summation is rank–1. In fact, the rank R of a matrix M can also be defined as the minimal number of rank-1 terms whose sum equals M. This decomposition is visualized in Figure 1.3.

Truncating the SVD to the first $\rho < R$ terms gives rise to a matrix Y which is the best $rank - \rho$ approximation of the matrix M in the least squares sense, i.e. $||X-Y||^2$ is minimal. As such, SVD is an interesting tool for data compression: a considerable amount of variance in the data can be preserved by storing only the first ρ singular vectors and values. This amounts to the storage of $\rho(1 + I_1 + I_2)$ elements instead of $I_1 I_2$ elements. Moreover, small singular values typically correspond to noise, therefore, truncation at a well-chosen rank has denoising effects.

Now let's consider the problem of processing multidimensional observations which arise as a linear mixture of a number of underlying source signals. For-

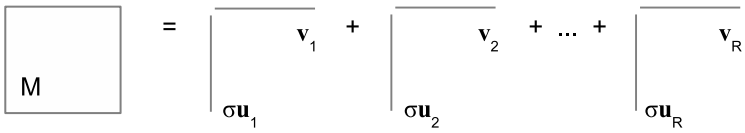


Figure 1.3: SVD of a matrix M in R rank-1 terms

.

mally, let $x(t) = [x_1(t), .., x_p(t))] \in \mathbb{R}^P$ be the observed signal at time instant t and $s(t) = [s_1(t), ..., s_N(t)] \in \mathbb{R}^N$ the underlying sources. Then $x(t)$ an be written as

$$x(t) = As(t) \tag{1.3}$$

where A is an unknown mapping from R^N to R^P. In *blind source separation* the goal is to find the sources $s(t)$ and the mapping or mixing matrix A. The above equation can also be written in a matrix form, in which case we talk about the *factorization of the matrix* X:

$$X = AS \tag{1.4}$$

Note that there are in general an infinite number of solutions for the matrix factorization problem. That is, if AS is a valid solution, then for any invertible matrix M:

$$X = (AM)(M^{-1}S) = \widetilde{A}\widetilde{S} \tag{1.5}$$

However, in blind source separation (BSS), the uniqueness of the obtained solution is crucial. We aim to be able to interpret the results, i.e. match $\mathbf{s}(t)$ (the rows of S) with the true underlying sources, and perhaps remove sources of no interest and reconstruct the clean signal.

In order to find a unique solution for the BSS problem, various decomposition techniques impose different constraints. For example, the so-called principal component analysis (PCA) assumes that the sources underlying the observed signals are mutually uncorrelated. Therefore, it projects the data onto a new, orthonormal basis. Note, still, that there is no unique solution to this problem, as any rotation of the obtained basis is a valid solution as well. This phenomenon is called the rotational invariance property of PCA. One possible solution is to take the right singular vectors, i.e. the columns of V from the SVD of X.

Another popular class of methods, independent component analysis (ICA), imposes a statistical diversity among the underlying sources. In case the observations are non-negative, as well as the sources and the mixing system are presumably non-negative, non-negative matrix factorization (NMF) may provide a solution. However, the success of these approaches strongly depends on the validity of these assumptions, which, unfortunately, are often violated in reality.

1.2.2 Decomposition of tensors

In the current section we will generalize the above matrix concepts to tensors. Focusing on different properties of the SVD, we will obtain two different generalizations for it. The new higher order decompositions will share some powerful properties with SVD. In fact, a possible generalization of the SVD even resolves the ambiguity issue of matrix factorization, as tensor factorizations are unique under mild conditions.

Higher order singular value decomposition

First, we will give a generalization of SVD considering equation 1.1. That is, we are looking for a series of orthogonal projections which will transform the tensor $\mathcal{X}$ into an all-orthogonal and ordered tensor $\mathcal{S}$. Every $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ can be approximated by the product

$$\mathcal{X} = \mathcal{S} \times_1 U_1^{(1)} \times_2 U_1^{(2)} \ldots \times_N U_1^{(N)}, \tag{1.6}$$

where the operator $\times_n$ denotes the mode-n product between a tensor $\mathcal{T} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ and a matrix $U \in \mathbb{R}^{J_n \times I_n}$, defined as

$$(\mathcal{T} \times_n U)_{i_1 i_2 \ldots j_n \ldots i_N} = \sum_{i_n} t_{i_1 i_2 \ldots i_n \ldots i_N} u_{j_n i_n} \tag{1.7}$$

Analogously to the product of two matrices, U makes linear combinations of the columns of $\mathcal{T}$. The product in equation 1.6, termed higher order singular value decomposition (HOSVD) has the following properties:

- $U^{(n)} = [\mathbf{u}_1^{(n)} \mathbf{u}_2^{(n)} \ldots \mathbf{u}_{I_n}^{(2)}]$
- $\mathcal{S} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ is a tensor with subtensors $\mathcal{S}_{i_n=\alpha}$, obtained by fixing the n^{th} index to α, show the following properties:
  - all-orthogonality:
    $< \mathcal{S}_{i_n=\alpha}, \mathcal{S}_{i_n=\beta} >= 0$ when $\alpha \neq \beta$
  - ordering:
    $\|\mathcal{S}_{i_n=1}\| \geq \|\mathcal{S}_{i_n=}2\| \geq \ldots \geq \|\mathcal{S}_{i_n=I_n}\| \geq 0$
    for all n.

The Frobenius-norms of $\|\mathcal{S}_{i_n=i}\|$ are called the n-mode singular values of $\mathcal{X}$ and the vectors $\mathbf{u}_i^n$ are the n-mode singular vectors. The values $I_1, I_2, \ldots, I_N$ correspond to the ranks of the different matrix unfoldings of $\mathcal{X}$ along the different modes. The n-tuple I_n is called the multilinear rank of the tensor $\mathcal{X}$. A graphical representation of the decomposition in case of a third-order tensor is given in Figure 1.4.

By dropping the orthogonality constraints we arrive to the so-called Tucker3 model:

$$\mathcal{X} = \mathcal{G} \times_1 V_1^{(1)} \times_2 V_1^{(2)} \ldots \times_N V_1^{(N)} + \mathcal{E}, \tag{1.8}$$

Choosing a core tensor $\mathcal{G}$ with smaller dimensions than $\mathcal{X}$, but keeping the error $\mathcal{E}$ sufficiently small, one obtains a good compressed estimate of the original dataset. Therefore, HOSVD is often used for dimensionality reduction and feature extraction.

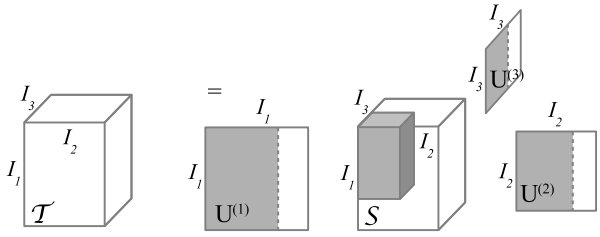


Figure 1.4: Visualization of the higher-order singular value decomposition of a third-order tensor $\mathcal{T}$

.

Canonical polyadic decomposition (CPD)

Another possible generalization of matrix SVD for tensors is considering equation 1.2, i.e. expansion as a sum of rank-1 terms.

As we have seen in the previous sections, the problem of matrix decomposition is ill-posed and additional constraints are needed in order to obtain a unique solution. Interestingly, tensors admit unique decompositions under mild conditions.

Canonical polyadic decompomposition (CPD) approximates a third order tensor $\mathcal{T} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ with a sum of R rank-1 tensors:

$$\mathcal{T} \approx \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r. \tag{1.9}$$

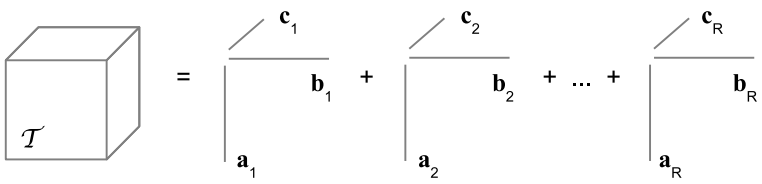


Figure 1.5: CPD of a tensor $\mathcal{T}$ in R rank-1 terms

.

CPD is visualized in Figure 1.5. Note that the definition is formulated for third-order tensors, however, the model can be extended to higher order tensors in a straightforward manner. The rank of the tensor is defined as the smallest R for which (1.9) is exact.

The advantage of the CPD model is its uniqueness up to permutation and scaling under the usually fulfilled conditions [31]. A more general framework for uniqueness has been recently presented in [20, 19].

Block term decomposition (BTD)

The block term decomposition (BTD), introduced in [11, 12, 14] generalizes CPD, as it allows components of low multilinear rank, as opposed to the rank−1

model of CPD. In this chapter we consider one particular case, decomposition into rank-$(L_r, L_r, 1)$ terms.

The rank-$(L_r, L_r, 1)$ block term decomposition of a third order tensor $\mathcal{T} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ into a sum of rank-$(L_r, L_r, 1)$ terms $(1 \leq r \leq R)$ is given as

$$\mathcal{T} \approx \sum_{r=1}^{R} \left(\mathbf{A}_r \cdot \mathbf{B}_r^T \right) \circ \mathbf{c}_r, \tag{1.10}$$

in which the matrix $\mathbf{D}_r = \mathbf{A}_r \cdot \mathbf{B}_r^T \in \mathbb{R}^{I_1 \times I_2}$ has rank L_r and the vector $\mathbf{c}_r$ is nonzero. In addition to permutation and scaling, inherited from the CPD, the factors $\mathbf{A}_r$ may be postmultiplied by any nonsingular matrix $\mathbf{F}_r \in \mathbb{R}^{L_r \times L_r}$, provided that $\mathbf{B}_r^T$ is premultiplied by the inverse of $\mathbf{F}_r$. When the matrices $[\mathbf{A}_1 \cdots \mathbf{A}_R]$ and $[\mathbf{B}_1 \cdots \mathbf{B}_R]$ are full column rank and the matrix $[\mathbf{c}_1 \cdots \mathbf{c}_R]$ does not contain collinear columns, the decomposition is guaranteed to be unique up to the above indeterminacies.

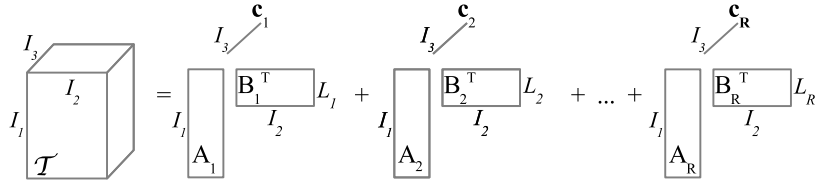


Figure 1.6: BTD-$(L_r, L_r, 1)$ of a tensor $\mathcal{T}$.

Figure 1.6 visualizes the decomposition of a tensor in rank-$(L_r, L_r, 1)$ terms.

The uniqueness of the decomposition is paramount for blind source separation, as it allows to give physical interpretation to the results and match the resulting components to true underlying processes. In the matrix case uniqueness is ensured by various constrains such as orthogonality or independence, which often has no physical meaning or is a too strong assumptions. However, the weaker uniqueness conditions of CPD and BTD are met for a wide range of parameters. This makes these tensor decompositions very interesting tools for various blind source separation applications in biomedical analysis problems.

Chapter 2

Construction of tensors in biomedical applications

There are several ways to organize biomedical data in a tensor. Often this organization comes naturally from the way the data was collected. At other times a specific tensorization method is applied in order to convey additional information about the data, which is thought to be of interest in the given problem. We will provide several examples of both approaches, some of which are also visualized in Figure 2.1.

2.1 Naturally occurring tensors

2.1.1 Genomic data

Genome-scale signals, such as mRNA expression levels, protein's DNA-binding occupancy levels or copy number variations can be recorded using DNA microarrays. These signals provide information about cellular processes, which may characterize normal or pathological regulatory mechanisms. A single sample of DNA microarray probes the signal of a certain number of genes. Samples may be analysed from multiple patients. Repeated probing under different experimental conditions can give additional insight, such as collecting samples at different time points during different oxidative stress conditions [38]. A tensorial framework allows to integrate these experimental conditions and analyse them simultaneously. Depending on the number of different experimental conditions we wish to vary, we might represent the data as a higher-order tensor with dimensions $\mathit{patients} \times \mathit{genes} \times \mathit{experimental\ condition}\ 1 \times \cdots \times \mathit{experimental\ condition}\ N$.

2.1.2 Repeated multichannel measurements

As already mentioned, biomedical data comes naturally in a multiway form in case of repeated multichannel measurements. For example the brain activity in response to the task paradigm can be studied using functional magnetic resonance imaging (fMRI). The fMRI signals are measured in a large number of voxels (points on a high resolution 3D grid defined over the volume of the

brain). In order to study the modulation of brain responses to a particular stimulation sequence, [6] organized continuous multisession and multisubject fMRI data in a $voxel \times time \times session$ and $voxel \times time \times subject$ tensor. One can assume that the same task elicits a similar response in the same brain regions across subjects and with similar timing in repeated experiments. However, the strength of the activation may vary in different subjects, or in consecutive sessions performed by the same subjects. In other words, we expect that each source in the brain has the same temporal and spatial signature, and these signatures are scaled over the different subjects or sessions, giving rise to a $rank-1$ structure. As such, the CPD model in equation 1.9 is appropriate to analyse data. Choosing the appropriate number of components R, each term in the CPD decomposition corresponds to a distinct brain source. Their spatial and temporal characteristics, as well as the modulation over the repetitions can be studied using the signatures $\mathbf{a}_r$, $\mathbf{b}_r$ and $\mathbf{c}_r$.

2.1.3 Epoched multichannel measurements

Relevant information sometimes resides in well-defined epochs within the continuous data, rather than throughout the whole measurement. This is the case for example in event-related potential (ERP) data, where the brain response in each trial, observed on EEG (or MEG), follows a specific waveform within a few hundred milliseconds time-locked to the stimulus onset. During a typical experiment a few hundred stimuli are presented to a subject. In order to analyse the EEG, the continuous data is broken down in epochs with a predefined window length around the stimuli. Such data is naturally organized in a $channel \times time \times trial$ tensor. The decomposition of such a tensor can help to find patterns which are representative for one type of stimulus, but not for the other. This information can be utilized in order to recognize users' intention based on their brain responses to different stimuli in a BCI setting [51, 48] or extract the localisation of repeated spikes during a neonatal seizure [17, 18]. Similarly, in case of ECG measurements, it may be important to study the variations in the ECG waveform of the consecutive heart beats. For example, an alternating pattern in the amplitude of consecutive T waves, called T wave alternans, is a possible indicator for risk of sudden cardiac death. A convenient way to do so is segmenting the ECG into single beats. After appropriate alignment of the T waves, the multi-lead ECG data is represented in a $lead \times time \times beats$ tensor. In case the patient has T wave alternans, this will be indicated by the presence of an ABABAB pattern in the trial mode signature of an $R = 1$ CPD model [23].

2.2 Tensor expansion of matrix data

We have seen that multichannel time series naturally take the form of a $channel \times time$ matrix. There are several different approaches to extend this to a tensorial representation by expanding the time course into an extra dimension, with the aim of conveying additional information about the signal.

2.2.1 Frequency transformation

The frequency content of biomedical signals often carries crucial information. This information can be conveyed by expanding the time series by means of a time-frequency transformation. This has been exploited in various ways. In a study aiming at classifying different pathological heart beats, the ECG signal was analysed using a short-time Fourier transformation to construct a $channel \times time \times frequency$ tensor [25]. Alternatively, wavelet transformation [2, 16] or Wigner-Ville distribution [49] is often used to expand the EEG matrix into a tensor. A particularly elegant example is the extraction of stereotypical oscillatory brain sources in the alpha and theta bands, related to resting and mental arithmetic, as revealed by the CPD of wavelet transformed EEG [36]. Studying wavelet transformed event-related potential data in various subjects and sessions, represented in a 5-way tensor, helped to reveal a quantitative difference in occipital gamma band response between different conditions of a visual paradigm [37]. It could also facilitate the classification of different event related potentials in a brain computer interface [33].

A frequency transformation can also be applied over space. Local spatial Fourier transform computed on the EEG matrix gives rise to a $space \times time \times wave \times vector$ tensor. This formulation will allow to separate sources with correlated but shortly delayed activities, which is the case when interictal epileptic activity spreads between two regions [5].

2.2.2 Hankel structure

A less intuitive but also powerful way of deriving a tensor representation is by means of the Hankel decomposition. Biomedical signals may be modelled as the sum of exponentially damped sinusoids [27, 10, 22]. Such signal model allows unique blind source separation in rank-$(L_r, L_r, 1)$ terms [13]. To exploit the desired structure, each channel signal $\mathbf{a}_{ch} = [a_{ch}(1)\ a_{ch}(2)\ \cdots\ a_{ch}(S)]$, $ch = 1, ..., I_1$ is mapped to a Hankel matrix as follows:

$$\begin{bmatrix} a_{ch}(1) & a_{ch}(2) & a_{ch}(3) & \cdots & a_{ch}(I_3) \\ a_{ch}(2) & a_{ch}(3) & \cdots & a_{ch}(I_3) & a_{ch}(I_3+1) \\ a_{ch}(3) & \cdots & a_{ch}(I_3) & a_{ch}(I_3+1) & a_{ch}(I_3+2) \\ \vdots & \iddots & \iddots & \iddots & \iddots \\ a_{ch}(I_2) & a_{ch}(I_2+1) & \cdots & a_{ch}(S-1) & a_{ch}(S) \end{bmatrix}$$

One can show that the Hankel matrix associated with a signal generated by L_r distinct poles is $rank - L_r$. For example, the Hankel matrix of a pure exponential is rank 1, while the one of a pure sinusoid or an exponentially damped sinusoid is rank−2. Noisy or nonstationary signals such as chirps give rise to Hankel matrices of higher rank. Since the Hankel mapping is linear, and assuming that the channel signals are linear combinations of the underlying sources, the above matrix is the linear combination of the Hankel matrices associated with the sources. It follows that the multichannel EEG data, represented by a tensor in the form of $Hankel\ matrix \times channels$, can be decomposed in block terms of $(L_r, L_r, 1)$ (equation 1.2.2) in order to retrieve the original sources. Indeed, it was shown that BTD combined with Hankel tensor representation of EEG successfully extracts and localizes epileptic seizure sources [26], and

can also reliably estimate arterial activity from surface ECG for the purpose of analysing arterial fibrillation in cardiology patients [40].

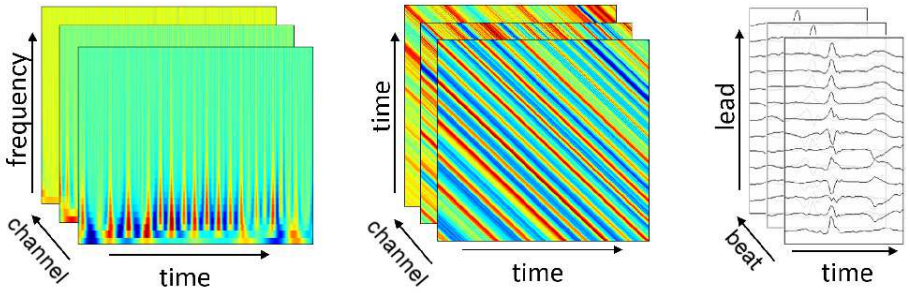


Figure 2.1: Biomedical data are often represented in a tensor. A multichannel EEG measurement may be expanded into the frequency dimension using wavelet transform, resulting in a $channel \times time \times frequency$ tensor (left). A $channel \times Hankel\ matrix$ representation may be used to model the multichannel EEG as a sum of exponentially damped sinusoids (middle). Epoched multichannel measurements naturally take the form of a tensor. For example, a 12 lead ECG recording segmented around each heartbeat forms a $lead \times time \times beat$ tensor (right).

2.2.3 Representation by means of a feature set

Sometimes very specific knowledge is available about which properties of the signal are interesting for a given problem. For example, it has been shown that multiscale entropy (MSE), which can characterize the complexity of a signal at different time scales, shows differences between the electrical brain activity of patients with Alzheimer's disease and healthy controls. Therefore, multichannel MEG measurements of various patients and controls can be analysed where the MSE values are organized in a $subject \times channel \times temporal\ scale$ tensor. A subsequent tensor decomposition reveals a characteristic filter, defined by the combination of the spatial and temporal factors. By projecting the data from a new subject onto this subspace, the resulting weight value can give an indication about the class membership of the subject [21]. In different problems, various different signal characteristics or features may be of interest. For example, various time and frequency domain features may be extracted from consecutive EEG windows in order to characterize normal versus epileptic seizure patterns [1]. This way, the multichannel EEG matrix is expanded to a $channel \times epoch \times feature$ tensor.

Chapter 3

Successful decompositions of biomedical data tensors

Once we have an appropriate tensor representation, a suitable tensor decomposition method must be chosen. The optimal choice depends on three main considerations: the purpose of the data analysis, the a-priori information available, and the structure of the data. In case of exploratory data analysis, aiming at understanding hidden factors in the data, or extracting the sources underlying an observed signal, fully unsupervised tensor decompositions would be the method of choice. If some a-priori knowledge is available, such as non-negativity of the sources, constraints on the factor matrices can be imposed. For example, nonnegative CPD and additional l_1-regularization successfully differentiates tumour tissue types using MR spectroscopic imaging [7]. Sometimes complementary observations, e.g. recordings of different signal modalities are available. Knowledge may be transferred between these modalities using coupled tensors decompositions. Labelled data represents even stronger a-priori knowledge, which will allow us to use supervised tensor decomposition when the goal is to differentiate classes in grouped data. Finally, the parameters of the tensor decomposition must be set carefully. We will dedicate a separate section to this topic in Chapter 4. In the following sections we illustrate the use of these different decompositions with several examples.

3.1 Unsupervised tensor decompositions

3.1.1 Blind source separation

Electroencephalography (EEG) measures the changes in brain's electrical activity over time using electrodes placed over the scalp. The electrical potentials, propagating in all directions from their sources, travel through different tissues before reaching the scalp. Therefore, the signals measured at the electrodes are a mixture of the attenuated electrical activity of various brain sources. Besides, EEG also picks up other physiological sources such as muscle (EMG) or eye (EOG) activity. It is also very sensitive to non-physiological artefacts, such as

power line noise or electrode movement. Therefore, the interpretation of the EEG is often difficult and requires careful preprocessing.

One of the most important applications of EEG is recording and studying epileptic seizure activity. The voltage distribution over the electrodes can give an indication about the location of the seizure source. However, due to involuntary movements and discomfort of the patient, these recordings are often contaminated by artefacts. Below we illustrate how unsupervised tensor decompositions can help to separate the distinct sources underlying the noisy EEG and characterize the seizure source.

The epileptic seizure activity is known as an oscillatory phenomenon, consisting of rhythmical waves in a frequency band below 30Hz which evolve in amplitude, frequency and location [45]. Consider for example the seizure segment depicted in Figure 3.2a. The seizure pattern, most prominent on the T1 channel, begins with a few distinct sharp waves. Between $2-4s$ after the start of the segment the waves occur more rhythmically 4 times every second. Later on, from 7s, the waves become sharper and shorter, repeating more rapidly, at a rate of 8Hz. Despite of the clear frequency evolution, a short seizure segment, such as the pattern between $2-3s$, can be considered stationary. The continuous wavelet transform (CWT) of this stationary seizure segment, visualized in Figure 3.2b, results in an approximately rank–1 $time \times frequency$ matrix where large coefficients are present only at certain frequency scales, corresponding to the rhythm of the seizure pattern. These large coefficients are distributed along the whole length of the segment, following the actual phase of the oscillation. Moreover, we can assume that within this short segment the seizure does not spread yet from its source to other brain regions. Therefore, the seizure pattern will be the most prominent on the channel which is nearby the true source, and will also be visible on adjacent channels, although with moderate amplitude. The voltage distribution of the seizure pattern over the channels thus gives an indication of the location of the seizure source. Notice that if these assumptions are correct, then the seizure pattern can be represented by a rank-1 third order tensor, which is the outer product of a frequency signature, a temporal signature and a spatial signature.

Based on these considerations, epileptic seizures were successfully localized using the spatial signature of the CP decomposition of the wavelet transformed multichannel EEG segment at seizure onset [16, 2]. The CPD of the seizure in the previous example is shown on Figure 3.2c. Observing the time course in comparison with the raw EEG segment and the spatial signature with frontal dominance, it is clear that the first component captures eye blinks. The second component represents the seizure source. An oscillatory pattern is observed on the temporal signature, which is similar to the T1 channel of the raw EEG. Moreover, the spatial signature shows an anterior temporal onset, which is in agreement with the pathology of the patient. However, note that the frequency signature, showing a wide peak around 6Hz, is not very informative about the spectral content of the seizure. This is because the rank-1 CP components can not capture the evolving nature of the pattern. A BTD offers more flexibility, as it allows components of higher rank. In Figure 3.2d we show the decomposition of the same EEG segment into block terms. The first rank-1 term captures the eye blinks, as in CPD. However, choosing the second term as rank–$(2, 2, 1)$, a more detailed characterization of the seizure pattern is possible. The temporal signature on the left contains early slow activity, while the one on the right

captures the late fast oscillatory pattern of the seizure. The frequency characteristics can be directly seen from the frequency signatures, namely the 4Hz peak on the left and the 8Hz peak in the right frequency signature.

3.1.2 Unsupervised classification

Tensor decompositions can be used for unsupervised classification as well. Let us consider a classification problem, where the datapoints to be classified are represented as an $N-$dimensional feature set. Then, the entire dataset can be organized in an $(N+1)-$dimensional tensor, where the indices of the datapoints are along the last mode. After applying an appropriate tensor decomposition method, the signature of the last mode can give an indication about the class membership. This approach has been successful to classify attended and non-attended trials in an auditory BCI dataset [50]. Different data representations were explored, including a $channel \times time \times trial$, and a $channel \times frequency \times trial$ tensor, obtained by applying fast-fourier transform on each ERP time course. In Figure 3.1 we illustrate the ERP classification, achieved by a rank-1 CPD on the former representation. The spatial mode and temporal mode factors show a stereotypical P300 scalp topography and time course. Target (attended) and non-target (non-attended) trials are classified with over 70% accuracy based on the trial mode factor.

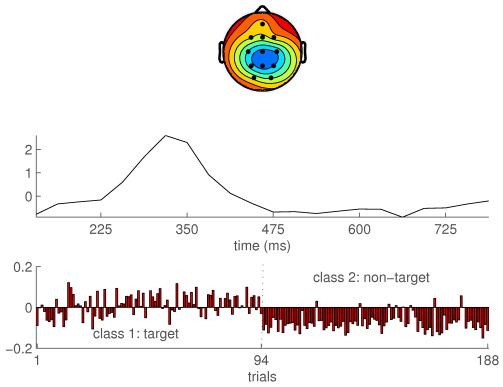
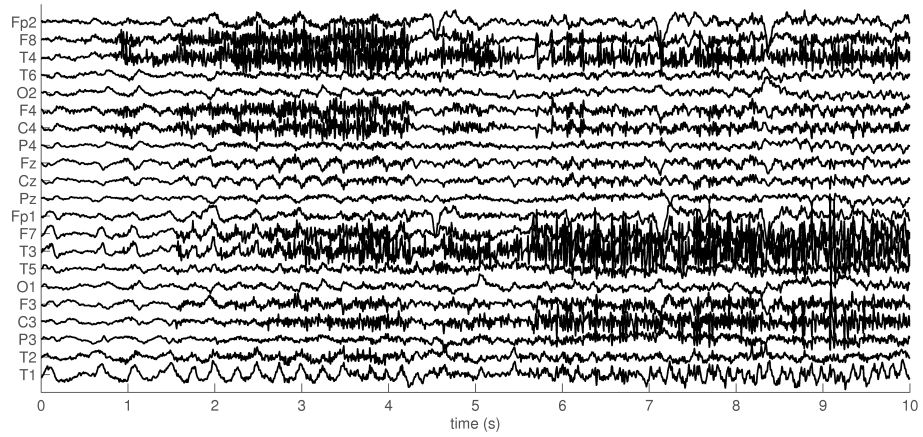


Figure 3.1: Rank-1 CPD of a auditory ERP dataset. The spatial mode and temporal mode factors show a stereotypical P300 scalp topography and time course. Target (attended) and non-target (non-attended) trials are classified with 73.4% accuracy based on the trial mode factor.

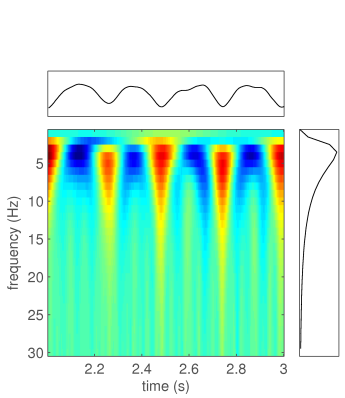
.

3.2 Supervised tensor decompositions

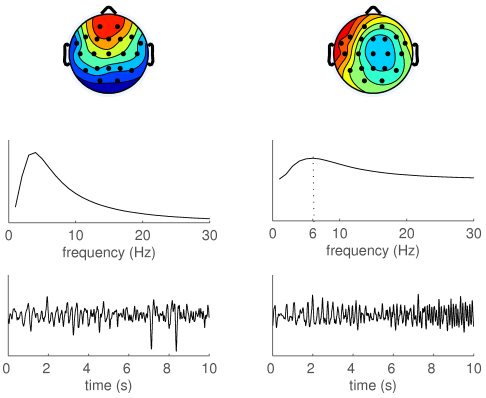
The tensor decompositions discussed before are extremely powerful as they provide a concise view on the underlying, intrinsic structure of the data. This is particularly useful for exploratory data analysis, independent of the type of data, as it will reveal the underlying natural low-dimensional representation by means of a Tucker, CPD or BTD decomposition in a fully unsupervised way.



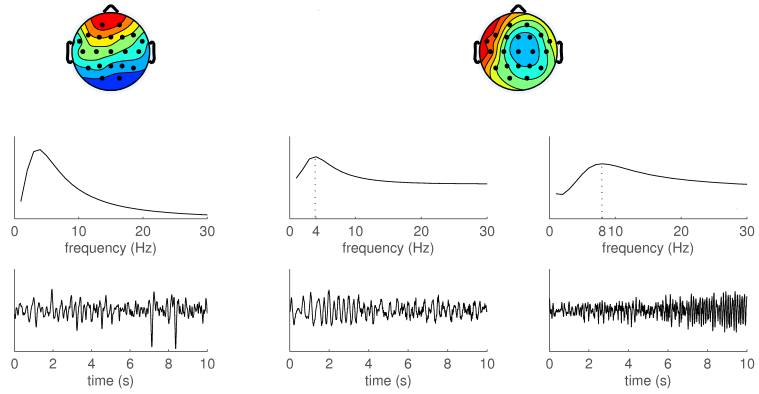
(a) Raw EEG



(b) CWT of a short segment



(c) CPD with $R = 2$



(d) BTD with $R = 2$, $L_1 = 1$ and $L_2 = 2$

Figure 3.2: (a) An EEG segment showing epileptic seizure activity. (b) Continuous wavelet transform of a stationary segment is an essentially rank−1 matrix. (c) CPD extracts eye and seizure activity with 2 components. (d) BTD extracts very similar eye activity, but captures more detailed features of the frequency content of the seizure pattern.

When considering supervised learning methods, one thinks of a traditional classification problem where the goal is to learn the boundary between classes based on given labels in the dataset. Such a boundary is characterized by the decision function, derived from the labeled datasets. Hence, traditional classifiers learn from labeled data to classify new data points into the corresponding classes. However, the properties of higher-order models might also be exploited in a supervised way, and might in certain cases outperform traditional supervised methods. The goal is to fuse known class labels and the intrinsic structure of the data in order to provide class labels of unseen data in a more robust way than when machine learning techniques that do not exploit structure would have been used. Different groups have proposed some individual approaches to combine tensorial frameworks within learning problems, e.g. [47, 24]. Certain types of structured learning can also be formalized in a general framework. When the assumed underlying structure relates to a CPD model, higher dimensional learning tasks on multidimensional arrays might be reformulated to problems where the classification solution is a solution of an optimization problem constrained by the structure [42, 27]. Alternatively, the Tucker decomposition can also incorporate discriminative constraints in its generalization known as the Higher Order Discriminant Analysis (HODA). Rather than only aiming to model the underlying variance as good as possible, HODA estimates a multidimensional decomposition where the subspaces in the different modes encode for optimal separability of the different classes. This can be seen as a generalization of classical Linear Discriminant Analysis (LDA) where every datasample is represented in a higher dimensional way. Such a method has been successfully used to discriminate between different neonatal EEG states [35]. A flow chart of the classification is visualized in Figure 3.3, where a Tucker decomposition is computed on the tensor constructed by concatenating different higher dimensional training samples. After decomposition, an additional classifier can be used on the most discriminative features from the core tensor. Reducing the dimensionality and preserving class-discriminative features can lead to a particularly robust method for classification in high dimensional spaces.

3.3 Coupled tensor decompositions

In order to understand a complex biomedical system, multimodal measurements might be beneficial, which are able to capture complementary aspects of the same system. For instance, simultaneous EEG and fMRI measurements are highly beneficial for studying brain function, as the former has good temporal resolution, and the latter good spatial resolution. In addition, brain anatomy or structure may be studied using MRI or DTI images. Several strategies exist to fuse the different modalities. For diverse datasets, a parallel processing of the individual modalities takes place, followed by a decision making step. Alternatively, in an integrative approach knowledge from one modality is used as a constraint in the analysis of the other. Finally, a data fusion approach allows a symmetrical interaction between the modalities [32]. Integration and fusion may be achieved through the coupled decomposition of the different modalities. This approach requires that the datasets are linked through a common dimension. In the following sections we discuss a few examples, illustrating the benefit of this common link in time, in space, or variability among multiple subjects.

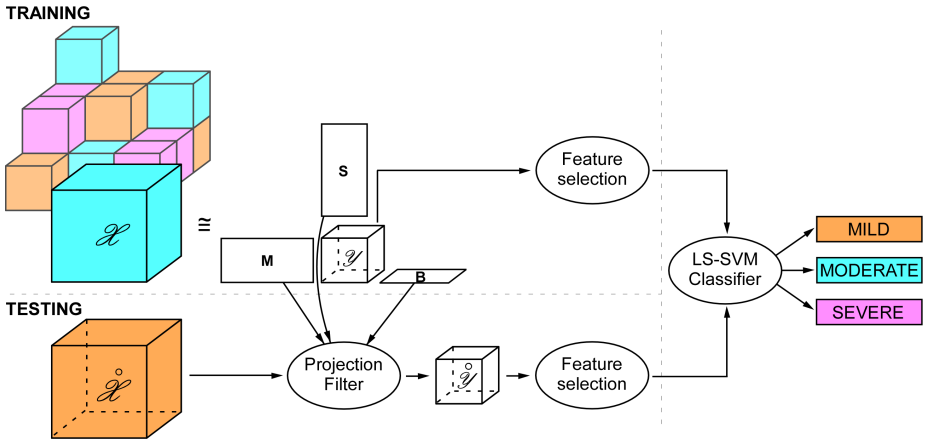


Figure 3.3: An example of how tensor decompositions can be used for a classification problem. 3 types of tensors, represented in 3 colors are used as training examples to discriminate between 3 classes. These training examples can be decomposed in a tensor decomposition that reduces the dimensionality and in the same time aims to maximise class differences. This leads to projection matrices that can be used to evaluate the class of an unseen example.

3.3.1 Coupling of multisubject data

Coupling in a multisubject database may be achieved by exploiting the subject-by-subject variability. More specifically, it is plausible that the different mental tasks involve more intense neural processing in one subject than in another. Moreover, one can assume that the relative level of involvement appears both in the strength of the EEG as well as in the strength of the fMRI response. Therefore, in case one arranges the measurements in a $subject \times EEG\ response$ and a $subject \times fMRI\ response$ matrix, each matrix is generated as a mixture of the same underlying neural sources by the same mixing matrix. This is the principle behind the jointICA approach, which fuses fMRI and event-related potential data into spatiotemporal snapshots to describe the dynamic relationship between hemodynamic and electomagnetic brain sources [9].

A limitation of the jointICA technique is the fact that it uses a single EEG channel, overlooking spatial information from the EEG. Multichannel EEG information can be incorporated via horizontal or vertical channel concatenation [46], or, formulating the problem as a Coupled Matrix-Tensor Factorization (CMTF) [3] .

Below we propose a CMTF scheme to characterize various spatiotemporal brain sources during interictal epileptic discharges (IED) measured by EEG-fMRI. Traditionally, epileptic EEG-fMRI is analysed within the general linear model (GLM) framework, where a regressor is defined based on the timing of the interictal spikes observed in the EEG. This regressor is then used to find voxels showing similar blood oxygen level dependent (BOLD) fMRI signal fluctuations. The GLM results often show widespread activations in the brain, which is partly explained by the mismatch between the temporal dynamics of EEG and fMRI. As the BOLD signal in response to a transient neural event peaks after several seconds, fMRI cannot differentiate the nuances of all under-

lying neural processes which are reflected in the millisecond resolution EEG. In fact, an IED often starts with a sharp spike followed by a slow wave. Source localization studies have shown that the spike propagates within a few tens of milliseconds. Moreover, slow waves are considered to be related to inhibitory activity. Therefore, we argue that both the IED and the GLM maps capture a mixture of underlying neural activity. In order to disentangle these sources, similar consideration are made as in jointICA, explained above. That is, we work with a group of 10 temporal lobe epilepsy cases, assuming that the same neural processes are reflected in the EEG and fMRI, and the strength of the neural processes vary in each patient. Average IEDs from each patient are organized in a $channel \times time \times patients$ tensor. The GLM-based activations maps of each patient are masked using a thresholded average GLM-based map, the images are vectorized and stored in a $voxels \times patients$ matrix. A CMTF is performed, where the two modalities share the same factor in the *patient* mode. A rank of $R = 2$ was chosen based on the core consistency diagnostic [8] of the EEG. The patient mode factors were fixed according to the patient-by-patient amplitudes of the spike and the slow wave.

The results of the decomposition are shown in Figure 3.4. Comparing the EEG sources with the grand average IED, one can observe that the first source captures the spike, while the second source captures the slow wave activity. Note the close resemblance of the spatial signatures (top left) and the scalp distributions of the spike and the slow wave in the grand average IED (top right). The average GLM-based activation maps is shown on the bottom left. Widespread activations are present in the right temporal lobe and in the occipital lobe. The first fMRI source, corresponding to the spike, shows predominantly temporal lobe activation. Interestingly, the second fMRI source, corresponding to the slow wave, captures the activation in the occipital lobe. Previous tractography studies have shown strong structural connection between the temporal lobe and the occipital lobe, as well as occipital activations in TLE. However, to our knowledge, a relationship between slow waves and occipital lobe activations has not been established. Further analysis is needed to confirm and interpret our results. Nevertheless, we believe that the joint factorization of EEG and fMRI can lead to new insights in the characterization of epileptic network activity.

3.3.2 Temporal coupling

Continuous EEG and fMRI data may be integrated based on the assumption that they capture the same changes in brain activity over time. As the sampling rate and the dynamics of the EEG and fMRI signals are different, some preprocessing is necessary. The following procedure was proposed in [34], one of the first studies to fuse multimodal neuroimaging data. The EEG signal recorded during a single fMRI image was defined as a segment. Then, a time-varying EEG spectrum was computed over the consecutive EEG segments, forming a $channel \times frequency \times time$ EEG tensor. The fMRI images were vectorized and the consecutive images were organized in a $voxel \times time$ matrix. As such, the time mode of the two datasets are aligned. Finally, the coupled decomposition was formulated mathematically as a multiway partial least squares (N-PLS) problem, i.e. the simultaneous factorization of the fMRI matrix and a CPD of the EEG tensor, with a constraint that maximizes the covariance between the temporal signatures of the EEG (independent variable) and the fMRI (depen-

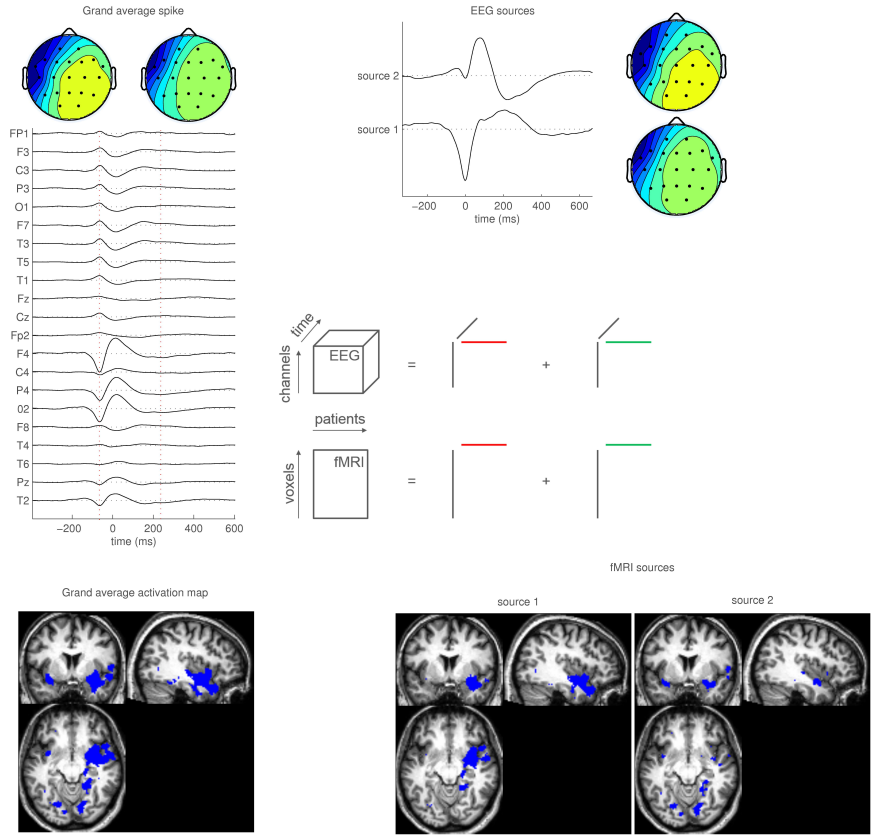


Figure 3.4: Coupled tensor-matrix factorization (CMTF) was performed on an EEG-fMRI dataset recorded in temporal lobe epilepsy patients. The schematic representation of the factorization into two components, where the modalities share the same factors in patient mode, is shown in the middle. The multichannel average interictal epileptic discharges (IEDs) and GLM-based fMRI activation maps of each patient form the input tensor and matrix, respectively. The grand average IED and grand average activation map are shown on the left. A spike and a slow wave, with different scalp potential distributions, are observed in the EEG, while the fMRI map shows widespread activations. The resulting EEG and fMRI sources are shown on the top and bottom right, respectively. Comparing the sources with the original EEG, it is clear that the first source captures the spike, while the second source captures the slow wave activity. The first fMRI source, corresponding to the spike, shows predominantly temporal lobe activation. The second fMRI source, corresponding to the slow wave, captures the activation in the occipital lobe.

dent variable). The study identified possible brain regions which participate in generating or controlling spontaneous brain rhythms such as alpha activity.

Ocular artefacts often obscure the EEG data. There are many approaches to remove such artefacts, among which one of the most robust ways is estimating the eye movements based on simultaneously recorded EOG. The estimation can be formulated as a CMTF problem. As the effect of eye movement is not exactly the same on the EEG and EOG signals, [41] proposed a relaxed CMTF solution, which estimates correlated shared factors, instead of equivalent ones. Additionally, their formulation also allows that the first or second derivatives of the factors are correlated rather than the original factors themselves. It was shown in both synthetic and real signals that refining coupling based on such similarities rather than equivalence have improved the estimation the factors.

3.3.3 Spatial coupling

In the previous example the EEG-fMRI integration was carried out using the common temporal dimension as a link between the two datasets. Alternatively, the decomposition may be coupled along the common spatial dimension, as proposed in [28]. In order to account for the different spatial resolution of the EEG and fMRI, fMRI data at voxels on the cortical grid of the EEG source space were extracted. Then, the integration is solved as a CMTF problem. The authors have applied a special formulation, which takes into account not only coincidence but also diversity among the modalities, by allowing one common component, one individual EEG and one individual fMRI component.

Chapter 4

Practical considerations

The previous sections clearly illustrated the power of using tensor decompositions in a variety of cases. In all examples, the final result for the optimal models was shown. However, it is important to highlight a few practical issues that are crucial for obtaining these results, and that might not be clear when one is not familiar with this class of methods.

4.1 Parameter selection

The appropriate choice of model parameters is crucial for obtaining an interpretable and useful result. Whereas in supervised classification the optimal parameters may be determined using cross-validation, choosing the model parameters is a more difficult question in unsupervised problems.

Different representations of the same dataset have different algebraic properties, therefore, the chosen tensorization will influence the optimal number of components and ranks. For example, oscillatory sources represented in a Hankel matrix will certainly be different from rank−1, therefore, a BTD must be chosen. Besides mathematical considerations, background knowledge from the application field may help to estimate the number of terms as the expected number of underlying sources. For instance, in case of a seizure localization problem, one may expect that only a few distinct sources exist, including a seizure, an artefact and a background activity source. Apart from utilizing such heuristics, several automated techniques exist to estimate the tensor ranks. For a brief overview of different techniques, we refer the reader to [26] and references therein. Below we will illustrate the effect of different model parameter selection in a practical example.

Let us consider the P300 classification problem discussed previously in section 3.1.2. A CPD was performed for different ranks between 1 and 5. The relative fit of the model and the core consistency [8] was computed for each model. We assume that one component captures the P300 source while other components model noise and other brain sources. It is not known a-priori which component captures the P300, therefore, the classification of the trials was attempted using each trial signature separately. We report the best classification for each model, i.e. we assume that in a real application we will find a way to automatically select the relevant component. The results are shown in Figure

4.1a.

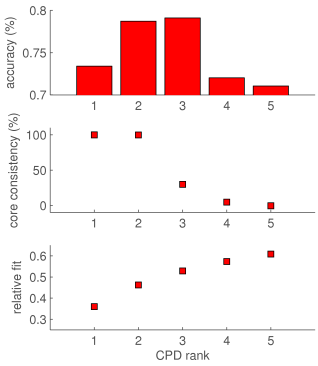
Recall that a classification accuracy of 73.4% was achieved already using a rank−1 CPD. When fitting a rank−2 CPD model on the data, the relative fit increases from 0.36 to 0.46 and the accuracy reaches 79%. It seems that the second component models some significant effects in the data, and this leaves more room for the first component to capture additional P300-specific variability in the data compared to the rank−1 model. The very high core consistency values indicate that both the rank−1 and rank−2 solutions follow a CPD model, i.e. a trilinear interaction between the signatures sufficiently describes the data. However, the core consistency drops to 30% for a rank−3 model, suggesting that a considerable amount of non-trilinear variability is present in the data. Nevertheless, the P300 characteristics are still captured reliably, yielding a 79% accuracy. Finally, although model fit increases slightly, core consistency values around zero indicate invalid models. Indeed, classification accuracy drops as well. Figure 4.1b depicts the components obtained with a rank−5 model. It is easily observed that the spatial and the temporal signatures of the first 3 components are highly correlated. When correlations in multiple factors are observed, one should check whether 2 or more terms nearly cancel each other. This phenomenon is called degeneracy and may indicate that the CPD rank is set too high.

4.2 Initialization

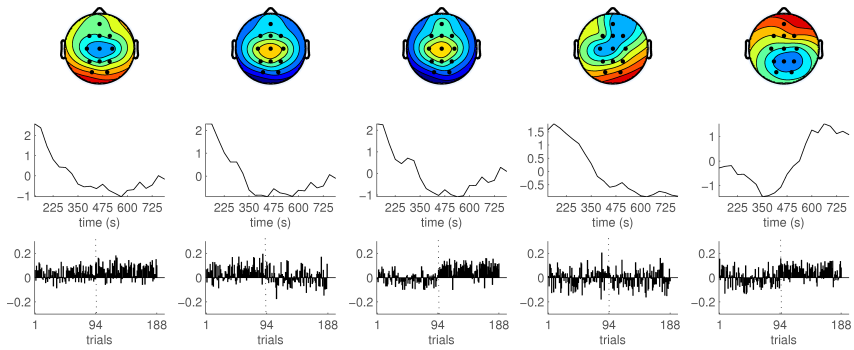
Tensor decompositions are computed using optimization algorithms: an initial guess is updated iteratively in a well-chosen direction, in order to minimize the objective function value, given as the fit of the model. The iterative procedure stops when the stepsize or the relative change of the objective function value is smaller than a predefined value. A tensor decomposition is a nonlinear and non-convex problem. As such, even though theoretically unique, there is no guarantee that the optimization algorithm will find the unique solution. In fact, the algorithm may converge to a local minimum. A good initialization is important to make sure that the algorithm converges fast to a good optimum. In general, it is recommended to run the decomposition algorithm multiple times from different initializations. Then, the best solution can be selected as the one with the smallest objective function value. Good initializations include generating pseudorandom factor matrices drawn for uniform or standard normal distributions, orthogonalizing such factors using QR factorization [43], or computing the initial factors based on HOSVD or using generalized eigenvalue decomposition. For more details and references, we refer the reader to [29].

4.3 Tools and algorithms

Below we list a few useful Matlab toolboxes, which implement the tensor decomposition methods mentioned in this bookchapter. The Tensorlab toolbox for Matlab [44] offers various different optimization algorithms to compute a CPD or a BTD, including the popular alternating least squares method, or the nonlinear least squares algorithm [43]. Furthermore, its structured data fusion module allows to implement coupled matrix and tensor decompositions, as well



(a)



(b)

Figure 4.1: The same P300 classification problem is considered as in section 3.1.2. 4.1a Classification accuracy, core consistency and fit of CPD models with increasing rank. 4.1b Rank−5 CPD of the same auditory ERP dataset as shown on Figure 3.1.

as to incorporate various constraints. Besides the built-in options, such as non-negativity, orthogonal, polynomial, Hankel, etc., its domain specific language allows the users to implement their own desired factor structures.

The Matlab Tensor Toolbox [4] extends Matlab's built-in capabilites to manipulate multidimensional arrays. Different classes are implemented in order to efficiently handle dense, sparse and factored tensors either as a Tucker-type or CPD-type approach.

Finally, the Tensor Toolbox for Feature Extraction and Applications [39] implements efficient tensor decompositions for multilinear discriminative feature extraction based on Constrained Tucker/CP models.

Acknowledgements

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Advanced Grant: BIOTENSORS (n^o 339804). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information.

Bibliography

[1] Evrim Acar, Canan Aykut-Bingol, Haluk Bingol, Rasmus Bro, and B ulent Yener Seizure recognition on epilepsy feature tensor. In *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pages 4273–4276, Aug 2007.

[2] Evrim Acar, Canan Aykut-Bingol, Haluk Bingol, Rasmus Bro, and B ulent Yener. Multiway analysis of epilepsy tensors. *Bioinformatics*, 23(13):i10–i18, 2007.

[3] Evrim Acar, Tamara G Kolda, and Daniel M Dunlavy. All-at-once optimization for coupled matrix and tensor factorizations. *arXiv preprint arXiv:1105.3422*, 2011.

[4] Brett W. Bader, Tamara G. Kolda, et al. Matlab tensor toolbox version 2.6. Available online, February 2015.

[5] Hanna Becker, Laurent Albera, Pierre Comon, Martin Haardt, Gwénaël Birot, Fabrice Wendling, Martine Gavaret, Christian-George Bénar, and Isabelle Merlet. Eeg extended source localization: tensor-based vs. conventional methods. *NeuroImage*, 96:143–157, 2014.

[6] Christian F. Beckmann and Stephen M. Smith. Tensorial extensions of independent component analysis for multisubject {FMRI} analysis. *NeuroImage*, 25(1):294 – 311, 2005.

[7] H.N. Bharath, Diana M. Sima., Nicolas Sauwen, Uwe Himmelreich, Lieven De Lathauwer, Sabine Van Huffel. Tensor Based Tumor Tissue Type Differentiation Using Magnetic Resonance Spectroscopic Imaging. in *Proc. of the Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE (EMBC), Milan, Italy, Aug. 2015*, pages 7003–7006, 2015.

[8] Rasmus Bro and Henk A. L. Kiers. A new efficient method for determining the number of components in PARAFAC models. *Journal of Chemometrics*, 17(5):274–286, 2003.

[9] Vince D. Calhoun, Tülay Adalı, G. D. Pearlson, and K. A. Kiehl. Neuronal chronometry of target detection: Fusion of hemodynamic and event–related potential data. *NeuroImage*, 30(2):544 – 553, 2006.

[10] Wim De Clercq, Bart Vanrumste, Jean-Marie Papy, Wim Van Paesschen, and Sabine Van Huffel. Modeling common dynamics in multichannel signals

with applications to artifact and background removal in eeg recordings. *Biomedical Engineering, IEEE Transactions on*, 52(12):2006–2015, 2005.

[11] Lieven De Lathauwer. Decompositions of a higher-order tensor in block terms - Part I: Lemmas for partitioned matrices. *Siam J. Matrix Anal. Appl.*, 30(3):1022–1033, 2008.

[12] Lieven De Lathauwer. Decompositions of a higher-order tensor in block terms - Part II: Definitions and Uniqueness. *Siam J. Matrix Anal. Appl.*, 30(3):1033–1066, 2008.

[13] Lieven De Lathauwer. Blind separation of exponential polynomials and the decomposition of a tensor in rank-(L_r, L_r, 1) terms. *Siam J. Matrix Anal. Appl.*, 32(4):1451–1474, 2011.

[14] Lieven De Lathauwer and Dimitri Nion. Decompositions of a higher-order tensor in block terms - Part III: Alternating Least Squares Algorithms. *Siam J. Matrix Anal. Appl.*, 30(3):1067–1083, 2008.

[15] Lieven De Lathauwer and Bart De Moor. From matrix to tensor: Multilinear algebra and signal processing. In *Institute of Mathematics and its Applications Conference Series*, volume 67, pages 1–16. Citeseer, 1998.

[16] Maarten De Vos, Anneleen Vergult, Lieven. De Lathauwer, Wim De Clercq, Sabine Van Huffel, Patrick Dupont, Andre Palmini, and Wim Van Paesschen. Canonical decomposition of ictal scalp EEG reliably detects the seizure onset zone. *NeuroImage*, 37:844–854, 2007.

[17] Wouter Deburchgraeve, Perumpillichira J. Cherian, Maarten De Vos, Renate M Swarte, Joleen H Blok, Gerhard Henk Visser, Paul Govaert, and Sabine Van Huffel. Neonatal seizure localization using parafac decomposition. *Clinical Neurophysiology*, 120(10):1787–1796, 2009.

[18] Ivana Despotovic, Perumpillichira J Cherian, Maarten Vos, Hans Hallez, Wouter Deburchgraeve, Paul Govaert, Maarten Lequin, Gerhard H Visser, Renate M Swarte, Ewout Vansteenkiste, et al. Relationship of eeg sources of neonatal seizures to acute perinatal brain lesions seen on mri: a pilot study. *Human brain mapping*, 34(10):2402–2417, 2013.

[19] Ignat Domanov and Lieven De Lathauwer. On the uniqueness of the canonical polyadic decomposition-part i: Basic results and uniqueness of one factor matrix. *SIAM Journal on Matrix Analysis and Applications*, 34(3):855–875, 2013.

[20] Ignat Domanov and Lieven De Lathauwer. On the uniqueness of the canonical polyadic decomposition-part ii: Overall uniqueness. *SIAM Journal on Matrix Analysis and Applications*, 34(3):876–903, 2013.

[21] Javier Escudero, Evrim Acar, Alberto Fernandez, and Rasmus Bro. Multiscale entropy analysis of resting-state magnetoencephalogram with tensor factorisations in alzheimer's disease. *Brain Research Bulletin*, pages –, 2015.

[22] Piotr J. Franaszczuk and Katarzyna J. Blinowska. Linear model of brain electrical activity - eeg as a superposition of damped oscillatory modes. *Biological Cybernetics*, 53(1):19–25, 1985.

[23] Griet Goovaerts, Carolina Varon, Bert Vandenberk, Rik Willems, and Sabine Van Huffel. Tensor-based detection of t wave alternans in multi-lead ecg signals. In *Computing in Cardiology Conference (CinC), 2014*, pages 185–188. IEEE, 2014.

[24] David R. Hardoon and John Shawe-Taylor. Decomposing the tensor kernel support vector machine for neuroscience data with structured labels. *Machine Learning*, 79(1-2):29–46, 2010.

[25] Kai Huang and Liqing Zhang. Cardiology knowledge free ecg feature extraction using generalized tensor rank one discriminant analysis. *EURASIP Journal on Advances in Signal Processing*, 2014(1), 2014.

[26] Borbála Hunyadi, Daan Camps, Laurent Sorber, Wim Van Paesschen, Maarten De Vos, Sabine Van Huffel, and Lieven De Lathauwer. Block term decomposition for modelling epileptic seizures. *EURASIP Journal on Advances in Signal Processing*, 2014(1):1–19, 2014.

[27] Borbála Hunyadi, Marco Signoretto, Stefan Debener, Sabine Van Huffel, and Maarten De Vos. Classification of structured eeg tensors using nuclear norm regularization: Improving p300 classification. In *Pattern Recognition in Neuroimaging (PRNI), 2013 International Workshop on*, pages 98–101, 2013.

[28] Esin Karahan, Pedro A. Rojas-Lopez, Maria L. Bringas-Vega, Pedro A. Valdes-Hernandez, and Pedro A. Valdes-Sosa. Tensor analysis and fusion of multimodal brain images. *Proceedings of the IEEE*, 103(9):1531–1559, Sept 2015.

[29] Stefan Kindermann and Carmeliza Navasca. New algorithms for tensor decomposition based on a reduced functional. *Numerical Linear Algebra with Applications*, 21(3):340–374, 2014.

[30] Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.

[31] Joseph B Kruskal. Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear algebra and its applications*, 18(2):95–138, 1977.

[32] Dana Lahat, Tülay Adalı, and Christian Jutten. Multimodal data fusion: An overview of methods, challenges, and prospects. *Proceedings of the IEEE*, 103(9):1449–1477, Sept 2015.

[33] Hyekyoung Lee, Young-Deok Kim, Andrzej Cichocki, and Seungjin Choi. Nonnegative tensor factorization for continuous eeg classification. *International Journal of Neural Systems*, 17(04):305–317, 2007. PMID: 17696294.

[34] Eduardo Martínez-Montes, Pedro A. Valdás-Sosa, Fumikazu Miwakeichi, Robin I. Goldman, and Mark S. Cohen. Concurrent eeg/fmri analysis by multiway partial least squares. *NeuroImage*, 22(3):1023 – 1034, 2004.

[35] Vladimir Matic, Perumpillichira J Cherian, Ninah Koolen, Gunnar Naulaers, Renate M Swarte, Paul Govaert, Sabine Van Huffel, and Maarten De Vos. Holistic approach for automated background eeg assessment in asphyxiated full-term infants. *Journal of neural engineering*, 11(6):066007, 2014.

[36] Fumikazu Miwakeichi, Eduardo Martinez-Montes, Pedro A. Valdes-Sosa, Nobuaki Nishiyama, Hiroaki Mizuhara, and Yoko Yamaguchi. Decomposing EEG data into space–time–frequency components using parallel factor analysis. *NeuroImage*, 22(3):1035 – 1045, 2004.

[37] Morten Mørup, Lars K. Hansen, Christoph S. Hermann, Josef Parnas, and Sidse M. Arnfred. Parallel Factor Analysis as an exploratory tool for wavelet transformed event-related EEG. *NeuroImage*, 29(3):938–947, 2006.

[38] Larsson Omberg, Gene H. Golub, and Orly Alter. A tensor higher-order singular value decomposition for integrative analysis of dna microarray data from different studies. *Proceedings of the National Academy of Sciences*, 104(47):18371–18376, 2007.

[39] Anh Huy Phan. Nfea: Tensor toolbox for feature extraction and application. Available online, February 2011.

[40] Lucas N. Ribeiro, Antonio R. Hidalgo-Munoz, and Vicente Zarzoso. Atrial signal extraction in atrial fibrillation electrocardiograms using a tensor decomposition approach. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pages 6987–6990, Aug 2015.

[41] Bertrand Rivet, Marc Duda, Anne Guerin-Dugue, Christian Jutten, and Pierre Comon. Multimodal approach to estimate the ocular movements during eeg recordings: A coupled tensor factorization method. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pages 6983–6986, Aug 2015.

[42] Marco Signoretto, Quoc Tran Dinh, Lieven De Lathauwer, and Johan A.K. Suykens. Learning with tensors : a framework based on convex optimization and spectral regularization. *Machine Learning*, accepted, 2013.

[43] Laurent Sorber, Marc Van Barel, and Lieven De Lathauwer. Optimization-based algorithms for tensor decompositions: canonical polyadic decomposition, decomposition in rank-(L_r, L_r, 1) terms and a new generalization. *SIAM Journal on Optimization*, 23 (2):695–720, 2013.

[44] Laurent Sorber, Marc Van Barel, and Lieven De Lathauwer. Tensorlab v2.0, Available online, URL: http://esat.kuleuven.be/sista/tensorlab/, January 2014.

[45] John M. Stern and Jerome Engel. *An Atlas of EEG Patterns*. Lippincott Williams & Wilkins, 2005.

[46] Wout Swinnen, Borbála Hunyadi, Esra Acar, Sabine Van Huffe, and Maarten De Vos. Incorporating higher dimensionality in joint decomposition of eeg and fmri. In *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European*, pages 121–125. IEEE, 2014.

[47] Dacheng Tao, Xuelong Li, Xindong Wu, Weiming Hu, and Stephen J. Maybank. Supervised tensor learning. *Knowl Inf Syst*, 13:1–42, 2007.

[48] Katrien Vanderperren, Bogdan Mijović, Nikolai Novitskiy, Bart Vanrumste, Peter Stiers, Bea R Van den Bergh, Lieven Lagae, Stefan Sunaert, Johan Wagemans, Sabine Van Huffel, and M.aarten De Vos. Single trial erp reading based on parallel factor analysis. *Psychophysiology*, 50(1):97–110, 2013.

[49] Martin Weis, Florian R omer, Martin Haardt, Dunja Jannek, and Peter Husar. Multi-dimensional space-time-frequency component analysis of event related eeg data using closed-form parafac. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pages 349–352, April 2009.

[50] Rob Zink, Borbála Hunyadi, Sabine Van Huffel, and Maarten De Vos. Exploring cpd based unsupervised classification for auditory bci with mobile eeg. In *Neural Engineering (NER), 2015 7th International IEEE/EMBS Conference on*, pages 53–56, April 2015.

[51] Rob Zink, Borbála Hunyadi, Sabine Van Huffel, and Maarten De Vos. Classifying the auditory p300 using mobile eeg recordings without calibration phase. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pages 1777–1780. IEEE, 2015.