TENSOR ANALYSIS OF NEUROIMAGING DATA

by

Esin Karahan Şenvardar

BS, Electrical and Electronics Engineering, Boğaziçi University, 2005

MS, Biomedical Engineering, Boğaziçi University, 2007

Submitted to the Institute of Biomedical Engineering

in partial fulfillment of the requirements

for the degree of

Doctor

of

Philosophy

Boğaziçi University

2015

TENSOR ANALYSIS OF NEUROIMAGING DATA

APPROVED BY:

Prof. Dr. Ahmet Ademoğlu (Thesis Advisor)	....
Prof. Dr. Cengizhan Öztürk	....
Prof. Dr. Tamer Demiralp	....
Doç. Dr. Ali Taylan Cemgil	....
Yrd. Doç. Dr. Adil Deniz Duru	....

DATE OF APPROVAL: 29 September 2015

ACKNOWLEDGMENTS

I would like to express my deepest gratitudes to my advisor Ahmet Ademoğlu for his generous support and guidance during my PhD study. I learned a lot from him about the research and also the topics we discussed during the coffee breaks in the lab. I am grateful to be part of his research team.

I am deeply indebted to my mentor Pedro Valdes-Sosa for his everlasting guidance and support. From the first day of our collaboration, he showed an endless enthusiasm that inspired me all the time. Discussions with him on science and other aspects of life guided me. The work presented in this thesis is a result of his fore vision and our joint work. I am honored to work with him.

I am grateful to Tamer Demiralp for his valuable contributions and collaboration all the time. His vast knowledge on the physiology guided me throughout the PhD study. I would like to thank Cengizhan Öztürk for accepting me to BUMIL at the beginning of my PhD. He was passionate all the time for solving my scientific and financial problems I faced during this study. I would like to express my thanks to Ali Taylan Cemgil and Evrim Acar for introducing the coupled tensor factorization in one of our meetings. I would like to express my sincere thanks to Adil Deniz Duru for his help in EEG. I am thankful to all my jury members for accepting to be part of the thesis committee and review my thesis.

I would like to thank my friends I met during my scientific visits to Cuban Neuroscience Center in Cuba and University of Electronic Sciences and Technology in China. Their kindness and friendship made me feel at home. I would especially thank to Maria Luisa Bringas-Vega and Pedro Ariel Rojas-Lopez for their continuous support and collaboration in research.

I met great people in BME. I would like thank Alper Yaman, Aytaç Durmaz and

Özlem Özmen Okur who were my labmates in my former lab BUMIL for discussions on research and providing a nice and peaceful environment. The time I spent in my second lab NEURAL was enriched with Ali Bayram, Basri Erdoğan, Duygu Şahin, Meltem Sevgi, Moataz Assem, Müge Özker, Seda Dumlu and Sencer Melih Deniz. I am grateful to Moataz for sharing long working hours with me in the lab and reviewing the drafts of the thesis. My friends Ahu Türkoğlu, Bige Vardar, Bora Büyüksaraç, Burcu Tunç, Fatma Şimşek, Filiz Ateş, Gamze Bölükbaşı, Hakan Solmaz, İsmail Devecioğlu, Murat Tümer, Mustafa Ruhi, Özgür Kaya and Uluç Pamuk in BME eased the struggle in PhD works. I want to give my special thanks to Muhammed Avşar for his great help in administrative arrangements. I wish to thank particularly Didar Talat and Nermin Topaloğlu for their friendship that meant a lot to me. I feel deeply sorry for the early loss of our dear colleague Mehmet Susam. It was a privilege to meet him.

I would like to thank the members of the Electrophysiology Lab in Istanbul University Medical School, Physiology Department and especially Elif Kurt, Itır Kaşıkçı and Zübeyir Bayraktaroğlu.

My friends Cihan Çakıcı, Elif Al, Fikret Sancaklı, İlksan Arslan, Neslişah Başaran, Metin Cihan, Öznur Özdamar and Yasemin Ergen shared the joy of life. I was always looking forward for İrem Bozbay's visits to İstanbul. I am indebted to Ebru Aydın to be with me in my most difficult days. My eternal friend Azime Can always supported me even from thousands of kilometers away. I would like to thank Kerem Şenvardar for his love and support during very stressful times of this work.

Lastly, I wish to thank to my family for their endless altruism, patience and support throughout my life. They are the most important thing in my life.

This work was financially supported by the BAP 6675 project. My visit to Cuban Neuroscience Center was funded by the TUBITAK 2214-A Yurtdışı Doktora Sırası Araştırma Bursu Programı and my visit to University of Electronic Sciences and Technology (UESTC) was funded by the DPT Life Sciences Project number 2009K120520 and UESTC.

The numerical calculations reported in this thesis were partly performed on the Future-Oriented Biocomputing System (FOBOS) of UESTC, China.

ACADEMIC ETHICS AND INTEGRITY STATEMENT

I, Esin Karahan Şenvardar, hereby certify that I am aware of the Academic Ethics and Integrity Policy issued by the Council of Higher Education (YÖK) and I fully acknowledge all the consequences due to its violation by plagiarism or any other way.

Name :

Signature:

Date:

ABSTRACT

TENSOR ANALYSIS OF NEUROIMAGING DATA

Acquisition of large amounts of data in neuroimaging research requires development of new methods that can disentangle the underlying information and reveal the features related to cognitive processes. This thesis attempts to propose new methods that favor the multimodality and multidimensionality of the brain data. The main difficulty for the fusion of imaging modalities is the discrepancies in their spatial and temporal resolutions as well as the different physiological processes they reflect. This problem is addressed by decomposing the EEG and fMRI data cast as tensors on both common and discriminant subspaces and computing the common spatial profile from the data on the cortical surface. The Granger causality analysis of brain connectivity is reformulated on tensor space enabling incorporation of tools developed in that area of research. The first approach on this analysis facilitated tensor methods for sparse representation of the connectivity patterns whereas the second method resolved them as atomic structures. General theory and computationally efficient algorithms are presented. The techniques are illustrated on the simultaneous EEG/fMRI recordings for the fusion model and on the fast fMRI data for the connectivity analysis. The proposed approaches may have a wide application area ranging from the early diagnosis of neurological diseases to the brain-computer interface studies.

Keywords: EEG, fMRI, multimodal data fusion, brain connectivity, Granger Causality, autoregressive processes, tensor decomposition, PARAFAC.

ÖZET

NÖROGÖRÜNTÜLEMEDE TENSOR ANALİZİ

Nörogörüntüleme araştırmalarında büyük miktarlarda veri toplanması bilişsel süreçlerle ilgili bilginin ayrıştırılması için yeni yöntemlerin geliştirilmesini gerektirmektedir. Bu tez çalışmasının amacı çok boyutlu ve birden fazla nörogörüntüleme modalitesinden elde edilen beyin verisinin işlenmesine elverişli yöntemler sunmaktadır. Nörogörüntüleme modalitelerinin tümleştirilmesindeki (fusion) en büyük zorluk elde verilerin uzaysal ve zamansal olarak farklı bilgiler taşımasıdır. Bu problem, tensörlerle ifade edilen EEG ve fMRG verisinin hem ortak hem de ayrık altuzaylarda ayrıştırılması ve ortak uzaysal profilin kortikal yüzeyde doğrudan veriden hesaplanması ile aşılmıştır. Aynı şekilde beyin bağlantılılığının Granger nedensellik analizi de tensör tabanlı bir modelle ifade edilmiş ve böylelikle tensör yöntemleri bu problemde kullanılabilmiştir. Bağlantılılık analizi için sunulan ilk yaklaşımda tensör yöntemleri kullanılarak bağlantılılık örüntüsü seyrekleştirilmiştir. İkinci yaklaşımda ise bağlantı örüntüleri atomsal yapılara bölünmüştür. Genel teori ve hesapsal olarak etkin algoritmalar sunulmuştur. Önerilen teknikler tümleştirme modeli için eşzamanlı EEG ve fMRG kayıtlarının üzerinde; bağlantılılık modelleri için hızlı çekim fMRG veri seti üzerinde uygulanmıştır. Önerilen yaklaşımların nörolojik hastalıkların erken teşhisinden beyin-bilgisayar arayüzü gibi uygulamalara kadar geniş bir alanda kullanım imkanı olabilir.

Anahtar Sözcükler:EEG, fMRG, çoklu modalite veri füzyonu, beyin bağlantılılığı, Granger nedenselliği, özbağlanımlı model, tensör ayrıştırması, PARAFAC

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

$\boldsymbol{\mathcal{A}}$	GC connectivity tensor
$\mathbf{A} \otimes \mathbf{B}$	Kronecker product
$\mathbf{A} \odot \mathbf{B}$	Khatri-Rao product
$\mathbf{B}$	fMRI data matrix
$\boldsymbol{\mathcal{B}}$	Time lagged data tensor of GC
$\mathbf{B}_{t-q}$	fMRI time series lagged by q
$\mathrm{circ}(\mathbf{a})$	Circulant matrix
$\mathbf{D}$	Discrete Fourier Transform Matrix
$\mathbf{E}$	Matrix error term
$\boldsymbol{\mathcal{E}}$	Tensor error term
eig	Eigenvalues of a matrix
embed	Circulant embedding operator
$\mathbf{F}_{\mathbf{V}}$	Spectral signature of EEG
fold	Folding operator in the context of t-Operators
$\mathbf{G}$	Primary current density
$\boldsymbol{\mathcal{G}}$	Core tensor of the Tucker decomposition
$\mathbf{H}$	Hemodynamic response matrix
$\mathbf{I}$	Identity matrix
$\boldsymbol{\mathcal{I}}$	Identity tensor
I_{Cx}	Number of sources (EEG or fMRI) on a cortical surface grid
I_E	Number of EEG scalp electrodes
I_{lag}	Number of past time points (time lags) in autoregressive models
I_F	Number of frequency points
$I_{F\delta}$	Number of subsampled frequency points
I_T	Number of time points
$I_{T\delta}$	Number of subsampled time points
$\mathbf{K}$	Lead field matrix
k_A	Kruskal rank

$\mathbf{L}$	Laplacian matrix
$\boldsymbol{\mathcal{L}}$	Lagrangian function
MatVec	Matricization operator in the context of t-Operators
$\mathbf{M_B}$	Discriminant spatial signature of fMRI
$\mathbf{M_C}$	Common spatial signature of EEG and fMRI
$\mathbf{M_V}$	Spatial signature of EEG
$\mathbf{M_G}$	Discriminant source spatial signature of EEG
$\mathbf{M_r}$	Spatial signature for receiver voxels
$\mathbf{M_s}$	Spatial signature for sender voxels
$\boldsymbol{\mathcal{R}}$	Sample covariance tensor of GC
R_B	Number of discriminant atoms of fMRI
R_C	Number of common atoms of EEG and fMRI
R_G	Number of discriminant atoms of EEG
$\boldsymbol{\mathcal{S}}$	EEG data tensor
$\mathbf{T}$	Temporal signature of GC
$\mathbf{T_B}$	Temporal signature of fMRI
$\mathbf{T_V}$	Temporal signature of EEG
$\mathbf{U}$	Factor matrices/Signatures of PARAFAC and Tucker decomposition
$\mathbf{V}$	EEG matrix
$\mathbf{x}$	Vector
$\mathbf{X}$	Matrix
$\boldsymbol{\mathcal{X}}$	Tensor
$\mathbf{X}^H$	Conjugate (Hermitian) transpose
$\mathbf{X}^T$	Transpose
$\mathbf{X}^{-1}$	Matrix inverse
$\boldsymbol{\mathcal{X}}_{(n)}$	Unfolded tensor on the nth dimension
$\mathbf{x}(i)$	ith element of the vector $\mathbf{x}$
$\mathbf{X}(i,j)$	(i,j)th element of the matrix $\mathbf{X}$
$\boldsymbol{\mathcal{X}}(i,j,k)$	(i,j,k)th element of the tensor $\boldsymbol{\mathcal{X}}$
$\lVert\boldsymbol{\mathcal{X}}\rVert_2$	Norm of a tensor
$\lVert\boldsymbol{\mathcal{X}}\rVert_{\circledast}$	Tensor nuclear norm in the context of t-Operators

$\boldsymbol{\mathcal{X}} \bullet \boldsymbol{\mathcal{Y}}$	Tensor contraction
$\boldsymbol{\mathcal{X}}\|_{\{J\|K\}}\boldsymbol{\mathcal{Y}}$	Binary concatenation operator
$[\boldsymbol{\mathcal{X}}_m]_{m=1:M}^{\{J_1\|...\|J_M\}}$	Set concatenation operator
$\boldsymbol{\mathcal{X}} \star \boldsymbol{\mathcal{Y}}$	t-Product
$\mathbf{1}$	Vector or matrix of all ones
β	Regularization parameter for orthogonality constraint
γ	Scale parameter in CMTF
$\mathbf{\Gamma}$	Vasoactive Feed Forward Signal Matrix
λ	Regularization parameters
$\pi(x)$	Prior function of x
σ^2	Error variance

LIST OF ABBREVIATIONS

ADMM	Alternating Direction Method of Multipliers
AIC	Akaike Information Criterion
ALS	Alternating Least Squares
AR	Autoregressive
BEM	Boundary Element Method
BIC	Bayesian Information Criterion
BOLD	Blood Oxygen Level Dependent
CBF	Cerebral Blood Flow
CCA	Canonical Correlation Analysis
CMTF	Coupled Matrix Tensor Factorization
CTF	Coupled Tensor Factorization
DAG	Directed Acyclic Graph
DFT	Discrete Fourier Transform
dof	Degrees of Freedom
DTI	Diffusion Tensor Imaging
EEG	Electroencephalography
EPI	Echo Planar Imaging
FA	Flip Angle
FEM	Finite Element Method
FFT	Fast Fourier Transform
fMRI	functional Magnetic Resonance Imaging
FWHM	Full Width at Half Maximum
GC	Granger Causality
HALS	Hierarchical Alternating Least Squares
HOPLS	Higher Order Partial Least Squares
HRF	Hemodynamic Response Function
ICA	Independent Component Analysis
InI	Inverse Imaging

LFP	Local Field Potential
log	Natural Logarithm
LORETA	Low Resolution Brain Electromagnetic Tomography
M	Motor Cortex
MAR	Multivariate Autoregressive
M-P Diagram	Markov Penrose Diagram
MUA	Multiple Unit Activity
N-PLS	Multiway Partial Least Squares
N-D	N Dimensional
P Diagram	Penrose Diagram
PARAFAC	Parallel Factors
PCA	Principal Component Analysis
PCC	Parietal Cortex
PreM	Pre-motor Cortex
RSS	Residual Sum of Squares
S	Somatosensory Cortex
SPM	Statistical Parametric Mapping
SVD	Singular Value Decomposition
SVM	Support Vector Machines
TAR	Tensor Autoregressive
TCCA	Tensor Canonical Correlation Analysis
TE	Echo Time
TR	Repetition Time
V	Visual Cortex
VFFS	Vasoactive Feed Forward Signal

1. INTRODUCTION

Imaging of the brain function has a wide application area ranging from understanding of the cognitive and perceptual processes in the brain and diagnosis of neurological and mental disorders to the design of neural prostheses and brain-computer interface applications. One of the biggest challenges in this area of research is to develop methods that can handle the complexity of the brain and the type and the size of the datasets acquired from different imaging modalities.

The human brain serves as a highly interconnected network of on the order of 10^{11} neurons and 10^{15} connections between them [1]. This complex network has been studied at different spatial scales from the microscale that investigates the function of the single neurons and their synaptic connections to the macroscale that accounts for the functional and anatomical connectivity between brain regions.

There are several points to consider when addressing the analysis of brain data [2]:

1. Brain data acquired at any level of organization presents a multidimensional nature in space and time.
2. Since each modality is an indirect measurement of the underlying dynamical system, multimodality analysis provides a complementary framework.
3. Each modality is recorded at different spatial and temporal resolutions that should be handled in the multimodal fusion.
4. The analysis should not only be confined to the identification of the brain regions that are specialized on certain functions but also the determination of their interactions.

Among the large variety of acquisition methods available for investigating the brain function, this thesis will focus on two particular modalities: EEG and fMRI. EEG is a non-invasive technique based on the electrical activity produced by the neuronal populations in the milliseconds sampling time. The major drawback of the EEG is the limited spatial resolution which is a result of low number of measurement channels placed on the scalp and the volume conduction effect due to the transmission of electrical currents in the brain tissue. On the other hand fMRI reflects the neuronal activity through metabolic processes whose response evolve slowly compared to the underlying neuronal processes in millimeters spatial scale. The integration of EEG and fMRI on a common space and/or time scale by merging the superiorities of different imaging modalities may reveal the complex dynamics of brain functions and neuronal interactions on a finer spatiotemporal scale [3, 4, 5]. However as stated in [6] and [7] a caution should be taken for the fusion. Under certain circumstances these modalities may not overlap such as oscillatory activity of the neuronal populations represented by EEG may not lead to an increase in fMRI signal. Conversely, asynchronous neural activity will not be detected by EEG but by fMRI.

In this study, a novel EEG-fMRI fusion approach based on tensor methods is presented. This approach has several advantages over the others: 1) The model exploits the inherent multidimensionality of the multimodal data. 2) EEG and fMRI are fused on the common spatial extent by projecting the scalp EEG on the cortical surface with source localization 3) The model takes into account the discrepancies in the neural origins of the two modalities by calculating the common and individual signatures.

The tensor based approach is also adopted for the identification of the brain networks in a causal framework. By using the Granger causality analysis, two problem formulations are presented that may overcome the computational and algorithmic challenges of the analysis of the high dimensional neuroimaging data.

The organization of the thesis is as follows: Chapters are divided on the basis of the presentation of tools and application of those tools on the neuroimaging data. In Chapter 2 the tensor notation and operations that are used throughout the thesis are introduced. Chapter 3 describes tensor decomposition methods and several data fusion techniques based on them. Chapter 4 approaches the EEG/fMRI fusion in the tensor framework and proposes a coupled tensor matrix factorization model. In Chapter 5, brain connectivity is described and Granger causality is reformulated embracing the high dimensionality of neuroimaging data. Finally in Chapter 6 general discussion, conclusion and future work is presented.

2. TENSOR NOTATION AND OPERATIONS

In this chapter tensor notation and operations that will be used throughout the thesis are presented.

2.1 Definitions

Tensors are the generalization of vectors and matrices to higher dimensions. The order or mode of a tensor is the number of its dimensions. In this context vectors are one-dimensional tensors (1-D) denoted by $\mathbf{x} \in \mathbb{R}^{I}$, matrices are two dimensional tensors (2-D) denoted by $\mathbf{X} \in \mathbb{R}^{I \times J}$ and an N-dimensional tensors (N-D) are denoted by $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$. The ith element of a vector is $\mathbf{x}(i)$, (i,j)th element of a matrix is $\mathbf{X}(i,j)$ and (i,j,k)th element of a tensor is $\boldsymbol{\mathcal{X}}(i,j,k)$.

Columns and rows of matrices are replaced with fibers in tensors. Fiber of a tensor is represented by fixing all the indices but one. The column of a matrix denoted by $\mathbf{X}(i,:)$ is the mode-1 fiber and the row of a matrix denoted by $\mathbf{X}(:,j)$ is the mode-2 fiber. Slices are defined as the two-dimensional sections of a tensor by fixing all the indices but two. ith horizontal slice of a three dimensional tensor is denoted by $\boldsymbol{\mathcal{X}}(i,:,:)$. Other types of slices of a 3-D tensor are shown in Figure 2.1.

2.2 Tensor Operations

2.2.1 Mode-n Unfolding

Mode-n unfolding of a tensor is the transformation of the tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ to a matrix denoted by $\boldsymbol{\mathcal{X}}_{(n)} \in \mathbb{R}^{I_n \times I_1 \ldots I_{n-1} I_{n+1} \ldots I_N}$ where mode-n fibers are arranged to be columns of the resulting matrix [8]. (Refer to Figure 2.2). Tensor element

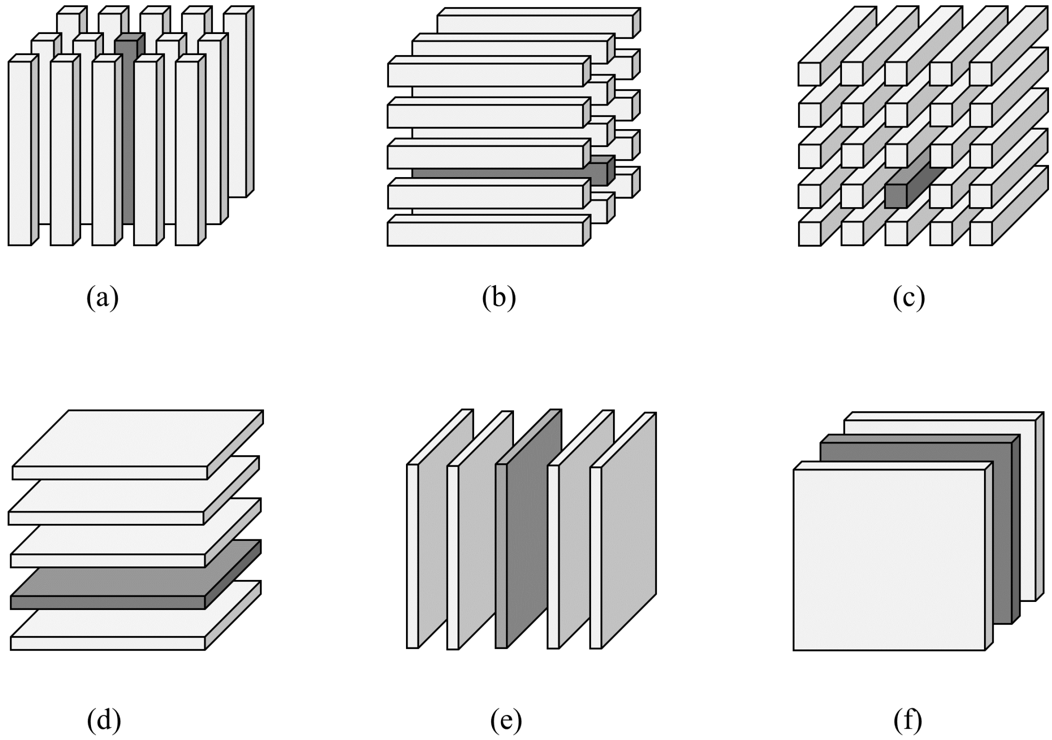


Figure 2.1 Illustration of the fibers and slices of the third order tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{5\times5\times3}$. (a) Mode-1 fibers. The fiber shown in dark color is $\boldsymbol{\mathcal{X}}(:,3,2)$. (b) Mode-2 fibers, $\boldsymbol{\mathcal{X}}(4,:,2)$ is in dark color. (c) Mode-3 fibers, $\boldsymbol{\mathcal{X}}(4,3,:)$ is in dark color. (d) Horizontal slices. The slice shown in dark color is $\boldsymbol{\mathcal{X}}(4,:,:)$. (e) Vertical slices, $\boldsymbol{\mathcal{X}}(:,3,:)$ is in dark color. (f) Frontal slices, $\boldsymbol{\mathcal{X}}(:,:,2)$ is in dark color.

$(i_1, \ldots, i_N)$ corresponds to the matrix element (i_n, j) , where

$$j = 1 + \sum_{\substack{k=1 \\ k\neq n}}^{N} (i_k - 1) J_k \quad \text{with} \quad J_k = \prod_{\substack{k=1 \\ k\neq n}}^{k-1} I_m. \tag{2.1}$$

2.2.2 Kronecker and Khatri-Rao Products

The Kronecker product is a special type of matrix product. Let $\mathbf{A} \in \mathbb{R}^{I\times J}$ and $\mathbf{B} \in \mathbb{R}^{K\times L}$, then the Kronecker product of $\mathbf{A}$ and $\mathbf{B}$ is denoted by $\mathbf{A} \otimes \mathbf{B}$ and is of size $IJ \times KL$. This operation is shown as

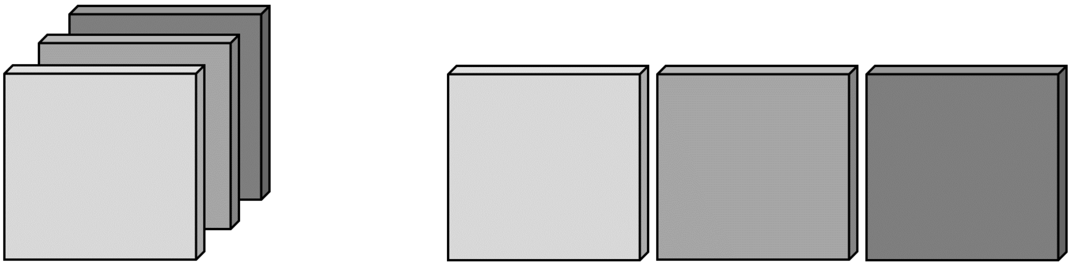


Figure 2.2 Mode-1 unfolding of a tensor. (a) Third order tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I \times J \times 3}$. (b) Mode-1 unfolding of the tensor $\boldsymbol{\mathcal{X}}$ results in a matrix with dimensions $I \times J \cdot 3$.

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} \mathbf{A}(1,1)\mathbf{B} & \mathbf{A}(1,2)\mathbf{B} & \dots & \mathbf{A}(1,J)\mathbf{B} \\ \mathbf{A}(2,1)\mathbf{B} & \mathbf{A}(2,2)\mathbf{B} & \dots & \mathbf{A}(2,J)\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}(I,1)\mathbf{B} & \mathbf{A}(I,2)\mathbf{B} & \dots & \mathbf{A}(I,J)\mathbf{B} \end{bmatrix}. \tag{2.2}$$

The Khatri-Rao product is the columnwise Kronecker product of two matrices. Let $\mathbf{A} \in \mathbb{R}^{I \times J}$ and $\mathbf{B} \in \mathbb{R}^{K \times J}$, then the Khatri-Rao product of $\mathbf{A}$ and $\mathbf{B}$ is denoted by $\mathbf{A} \odot \mathbf{B}$ and is of size $IK \times J$. Khatri-Rao product is formulated as

$$\mathbf{A} \odot \mathbf{B} = \begin{bmatrix} \mathbf{A}(:,1) \otimes \mathbf{B}(:,1) & \mathbf{A}(:,2) \otimes \mathbf{B}(:,2) & \dots & \mathbf{A}(:,J) \otimes \mathbf{B}(:,J) \end{bmatrix}. \tag{2.3}$$

2.2.3 Tensor Contraction

Tensor contraction is the multiplication of two tensors over specified common dimensions. Let $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times \cdots \times I_N \times J_1 \times \cdots \times J_M}$ and $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{J_1 \times \cdots \times J_M \times K_1 \times \cdots \times K_P}$. Multiplication of $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ over common dimensions $J_1, \dots, J_M$ gives the tensor $\boldsymbol{\mathcal{Z}} \in$

$\mathbb{R}^{I_1\times\cdots\times I_N\times K_1\times\cdots\times K_P}$. In scalar notation, this is showed as follows:

$$\left(\boldsymbol{\mathcal{X}} \bullet_{\{j_1,\ldots,j_M\}} \boldsymbol{\mathcal{Y}}\right)(i_1,\ldots,i_N,k_1,\ldots,k_P) \\ = \sum_{j_1,\ldots,j_M=1}^{J_1,\ldots,J_M} \boldsymbol{\mathcal{X}}(i_1,\ldots,i_N,j_1,\ldots,j_M)\boldsymbol{\mathcal{Y}}(j_1,\ldots,j_M,k_1,\ldots,k_P) \quad (2.4)$$

Outer and inner products can also be represented by using tensor contraction notation. For the outer product we make use of the singleton dimensions. Adding singleton dimensions to a tensor does not change the tensor itself: $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1\times\cdots\times I_N\times 1\times 1}$ is the same as $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1\times\cdots\times I_N}$. Outer product of the tensors $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1\times\cdots\times I_N}$ and $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{J_1\times\cdots\times J_M}$ gives a tensor of size $I_1 \times \cdots \times I_N \times J_1 \times \cdots \times J_M$ and is denoted elementwise by

$$(\boldsymbol{\mathcal{X}} \bullet_{\{1\}} \boldsymbol{\mathcal{Y}})(i_1,\ldots,i_N,j_1,\ldots,j_M) = \boldsymbol{\mathcal{X}}(i_1,\ldots,i_N,1)\boldsymbol{\mathcal{Y}}(j_1,\ldots,j_M,1). \quad (2.5)$$

Inner product of the same size tensors $\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}} \in \mathbb{R}^{I_1\times\cdots\times I_N}$ is equal to a scalar and defined by

$$\langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}} \rangle = (\boldsymbol{\mathcal{X}} \bullet_{\{I_1,\ldots,I_N\}} \boldsymbol{\mathcal{Y}}) = \sum_{i_1,\ldots,i_N=1}^{I_1,\ldots,I_N} \boldsymbol{\mathcal{X}}(i_1,\ldots,i_N)\boldsymbol{\mathcal{Y}}(i_1,\ldots,i_N). \quad (2.6)$$

Square of the norm of a tensor is equal to its inner product with itself :

$$\|\boldsymbol{\mathcal{X}}\|_2^2 = (\boldsymbol{\mathcal{X}} \bullet_{\{I_1,\ldots,I_N\}} \boldsymbol{\mathcal{X}}) = \sum_{i_1,\ldots,i_N=1}^{I_1,\ldots,I_N} \boldsymbol{\mathcal{X}}(i_1,\ldots,i_N)^2 \quad (2.7)$$

2.2.4 Tensor Concatenation

Tensor concatenation is the merging of the same order tensors in which tensors are necessarily required to be the same order and have the same dimensions except the concatenation index. We will define the binary and set operators for the tensor concatenation. Let $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1\times\cdots\times I_{n-1}\times J\times I_{n+1}\times\cdots\times I_N}$ and $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{I_1\times\cdots\times I_{n-1}\times K\times I_{n+1}\times\cdots\times I_N}$. Concatenation of $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ on the Jth and Kth dimensions gives the tensor $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{I_1\times\cdots\times I_{n-1}\times J+K\times I_{n+1}\times\cdots\times I_N}$. The binary operator for the concatenation is denoted as follows:

$$\mathcal{Z} = \mathcal{X}|_{\{J|K\}}\mathcal{Y} \tag{2.8}$$

If the concatenation is applied on a set of tensors $\mathcal{X}_m \in \mathbb{R}^{I_1\times\cdots\times I_{n-1}\times J_m\times I_{n+1}\times\cdots\times I_N}$ over the J_m^{th} dimensions for $m = 1,\ldots,M$, the result is of size $I_1 \times \cdots \times I_{n-1} \times (J_1 + \cdots + J_M) \times I_{n+1}$ and denoted as follows:

$$\mathcal{Z} = [\mathcal{X}_m]_{m=1:M}^{\{J_1|\ldots|J_M\}}. \tag{2.9}$$

2.2.5 t-Operators

t-Operators are introduced by Kilmer and her group [9] as an extension of linear algebra tools to tensors. Although there are many t-operators, we will present only the ones that are related to the context of this thesis.

Before giving details, two matrix types that are used extensively in the definitions will be reviewed. If $\mathbf{a} = \begin{bmatrix}\mathbf{a}(1) & \mathbf{a}(2) & \mathbf{a}(3) & \mathbf{a}(4)\end{bmatrix}^T$, then the *circulant matrix* is defined by

$$\mathrm{circ}(\mathbf{a}) = \begin{bmatrix} \mathbf{a}(1) & \mathbf{a}(4) & \mathbf{a}(3) & \mathbf{a}(2) \\ \mathbf{a}(2) & \mathbf{a}(1) & \mathbf{a}(4) & \mathbf{a}(3) \\ \mathbf{a}(3) & \mathbf{a}(2) & \mathbf{a}(1) & \mathbf{a}(4) \\ \mathbf{a}(4) & \mathbf{a}(3) & \mathbf{a}(2) & \mathbf{a}(1) \end{bmatrix}. \tag{2.10}$$

Note that a circulant matrix is completely specified by its first column. An important property of circulant matrices is that they can be diagonalized with the normalized Discrete Fourier Transform (DFT) matrix [10]. Let w be the Mth root of unity, $w = e^{-2\pi i/M}$, the DFT matrix $\mathbf{D} \in \mathbb{R}^{M \times M}$ is defined as

$$\mathbf{D} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & w^1 & \dots & w^{M-1} \\ 1 & w^2 & \dots & w^{2(M-1)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & w^{M-1} & \dots & w^{(M-1)(M-1)} \end{bmatrix}. \tag{2.11}$$

If $\mathbf{a} \in \mathbb{R}^M$ is a column vector and $\mathbf{D} \in \mathbb{R}^{M \times M}$ is the DFT matrix defined in Eq. 2.11, then

$$\mathbf{D}\mathrm{circ}(\mathbf{a})\mathbf{D}^{-1} \tag{2.12}$$

is a diagonal matrix and its diagonal is equal to the DFT of $\mathbf{a}$, and can be calculated by using fast fourier transform (FFT), $\mathrm{fft}(\mathbf{a})$.

A block circulant matrix can also be created from the slices of a tensor. As an example, let $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I \times J \times K}$, then a block circulant of size $IK \times JK$ is

$$\mathrm{circ}(\boldsymbol{\mathcal{X}}) = \begin{bmatrix} \boldsymbol{\mathcal{X}}(:,:,1) & \boldsymbol{\mathcal{X}}(:,:,K) & \dots & \boldsymbol{\mathcal{X}}(:,:,2) \\ \boldsymbol{\mathcal{X}}(:,:,2) & \boldsymbol{\mathcal{X}}(:,:,1) & \dots & \boldsymbol{\mathcal{X}}(:,:,3) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\mathcal{X}}(:,:,K) & \boldsymbol{\mathcal{X}}(:,:,K-1) & \dots & \boldsymbol{\mathcal{X}}(:,:,1) \end{bmatrix}. \tag{2.13}$$

A matrix is called *Toeplitz* if it has constant values along each diagonal. An example of a Toeplitz matrix is given below:

$$\mathbf{T} = \begin{bmatrix} \mathbf{T}(1,1) & \mathbf{T}(1,2) & \mathbf{T}(1,3) & \mathbf{T}(1,4) \\ \mathbf{T}(2,1) & \mathbf{T}(1,1) & \mathbf{T}(1,2) & \mathbf{T}(1,3) \\ \mathbf{T}(3,1) & \mathbf{T}(2,1) & \mathbf{T}(1,1) & \mathbf{T}(1,2) \\ \mathbf{T}(4,1) & \mathbf{T}(3,1) & \mathbf{T}(2,1) & \mathbf{T}(1,1) \end{bmatrix} \tag{2.14}$$

Note that a Toeplitz matrix is completely specified by its first column and row. A Toeplitz matrix can be embedded into a larger size circulant matrix.

We will show the circulant embedding of a block Toeplitz matrix created from the frontal slices of the tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I \times J \times K}$ [11].

$$\operatorname{embed}(\boldsymbol{\mathcal{X}}) = \begin{bmatrix} \boldsymbol{\mathcal{X}}(:,:,1) & \boldsymbol{\mathcal{X}}(:,:,2)^H & \cdots & \boldsymbol{\mathcal{X}}(:,:,2) \\ \boldsymbol{\mathcal{X}}(:,:,2) & \boldsymbol{\mathcal{X}}(:,:,1) & \cdots & \boldsymbol{\mathcal{X}}(:,:,3) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\mathcal{X}}(:,:,K) & \vdots & \ddots & \vdots \\ \Phi & \boldsymbol{\mathcal{X}}(:,:,K) & \ddots & \vdots \\ \boldsymbol{\mathcal{X}}(:,:,K)^H & \Phi & \ddots & \vdots \\ \vdots & \boldsymbol{\mathcal{X}}(:,:,K)^H & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\mathcal{X}}(:,:,2)^H & \boldsymbol{\mathcal{X}}(:,:,3)^H & \cdots & \boldsymbol{\mathcal{X}}(:,:,1) \end{bmatrix} \tag{2.15}$$

where $\Phi = (\boldsymbol{\mathcal{X}}(:,:,K) + \boldsymbol{\mathcal{X}}(:,:,K)^H)/2$. The created matrix is of size $I \cdot (2K) \times J \cdot (2K)$. This type of circulant embedding is especially useful for covariance tensors which will be used in Section 5.4.

The MatVec operator stacks the frontal slices of a tensor to construct a matrix. This matricization operation is slightly different from the mode-n unfolding. Let $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I \times J \times K}$, then $\operatorname{MatVec}(\boldsymbol{\mathcal{X}})$ gives a tensor of size $IK \times J$:

$$\operatorname{MatVec}(\boldsymbol{\mathcal{X}}) = \begin{bmatrix} \boldsymbol{\mathcal{X}}(:,:,1) \\ \boldsymbol{\mathcal{X}}(:,:,2) \\ \vdots \\ \boldsymbol{\mathcal{X}}(:,:,K) \end{bmatrix} \tag{2.16}$$

The fold operation undoes the MatVec operation:

$$\operatorname{fold}(\operatorname{MatVec}(\boldsymbol{\mathcal{X}})) = \boldsymbol{\mathcal{X}} \tag{2.17}$$

2.2.5.1 t-Product. t-product of $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I\times J\times K}$ and $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{J\times L\times K}$ denoted by $\boldsymbol{\mathcal{X}}\star\boldsymbol{\mathcal{Y}}$ is equal to a tensor of size $I \times L \times K$

$$\boldsymbol{\mathcal{X}} \star \boldsymbol{\mathcal{Y}} = \mathrm{fold}(\mathrm{circ}(\boldsymbol{\mathcal{X}}) \cdot \mathrm{MatVec}(\boldsymbol{\mathcal{Y}})). \tag{2.18}$$

If the tensors that are contracted with t-product are sparse, the computation is performed as stated in the definition. However, if the tensors are dense, then diagonalizability property of the circulant matrices can be used and the t-product is calculated by the DFT matrices as follows:

$$(\mathbf{D}^H \otimes \mathbf{I})((\mathbf{D} \otimes \mathbf{I}) \cdot \mathrm{circ}(\boldsymbol{\mathcal{X}}) \cdot (\mathbf{D}^H \otimes \mathbf{I}))(\mathbf{D} \otimes \mathbf{I}) \cdot \mathrm{MatVec}(\boldsymbol{\mathcal{Y}}) \tag{2.19}$$

Transpose of a tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I\times J\times K}$ in t-operator concept is defined as

$$\boldsymbol{\mathcal{X}}^T = \mathrm{fold}\left(\begin{bmatrix} \boldsymbol{\mathcal{X}}(:,:,1)^T \\ \boldsymbol{\mathcal{X}}(:,:,K)^T \\ \vdots \\ \boldsymbol{\mathcal{X}}(:,:,2)^T \end{bmatrix}\right). \tag{2.20}$$

A tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I\times I\times J}$ is t-orthogonal if $\boldsymbol{\mathcal{X}}^T\star\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{X}}\star\boldsymbol{\mathcal{X}}^T = \boldsymbol{\mathcal{I}}$ where $\boldsymbol{\mathcal{I}} \in \mathbb{R}^{I\times I\times J}$ is the t-identity tensor whose frontal slice is the identity matrix and others are zero.

t-inverse of a tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I\times I\times J}$ is $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{I\times I\times J}$ if $\boldsymbol{\mathcal{X}} \star \boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{I}}$ and $\boldsymbol{\mathcal{Y}} \star \boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{I}}$ where $\boldsymbol{\mathcal{I}} \in \mathbb{R}^{I\times I\times J}$ is the t-identity tensor as defined above.

2.2.5.2 t-SVD. **t-SVD** factorizes a real valued tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I\times J\times K}$ as follows

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{U}} \star \boldsymbol{\mathcal{D}} \star \boldsymbol{\mathcal{V}}^T \tag{2.21}$$

where $\boldsymbol{\mathcal{U}} \in \mathbb{R}^{I,I,K}$, $\boldsymbol{\mathcal{V}} \in \mathbb{R}^{J,J,K}$ are t-orthogonal tensors and $\boldsymbol{\mathcal{D}} \in \mathbb{R}^{I\times J\times K}$ is a tensor with diagonal faces. t-SVD allows the tensor $\boldsymbol{\mathcal{X}}$ to be decomposed as

$$\boldsymbol{\mathcal{X}} = \sum_{i=1}^{\min(I,J)} \boldsymbol{\mathcal{U}}(:,i,:) \star \boldsymbol{\mathcal{D}}(i,i,:) \star \boldsymbol{\mathcal{V}}(:,i,:)^T. \tag{2.22}$$

As with the usual matrix SVD, t-SVD provides an optimal approximation of a tensor in the Frobenius norm of the difference. (see Theorem 4.3 in [12]).

2.2.5.3 t-Norm. t-norm is a type of tensor nuclear norm and is defined in [13] as

$$\|\boldsymbol{\mathcal{X}}\|_{\circledast} = \sum_{i=1}^{\min(I,J)} \sum_{k=1}^{K} \tilde{\boldsymbol{\mathcal{D}}}(i,i,k) \tag{2.23}$$

where $\tilde{\boldsymbol{\mathcal{D}}}$ is obtained by taking the Fourier transform of the faces of $\boldsymbol{\mathcal{D}}$.

2.3 Tensor Diagrams

We will use Markov-Penrose Diagrams (M-P Diagram) first introduced in [2] for the visual representation of the tensors and tensor models proposed in this thesis. We will briefly review this concept in this section.

2.3.1 Penrose Diagrams

Penrose Diagrams (P Diagrams) also known as tensor network diagrams have been used for the illustration of tensor objects and operations since Penrose [14]. In P Diagrams tensor objects are the nodes and each line leaving the node is the dimension of the tensor. The order of a tensor is equal to the number of dangling lines. Mathematical expressions of vectors, matrices and tensors and their corresponding P diagrams are shown in Figure 2.3. Nodes representing tensors with random elements are shown as circles and the ones with constant elements are shown as rectangles.

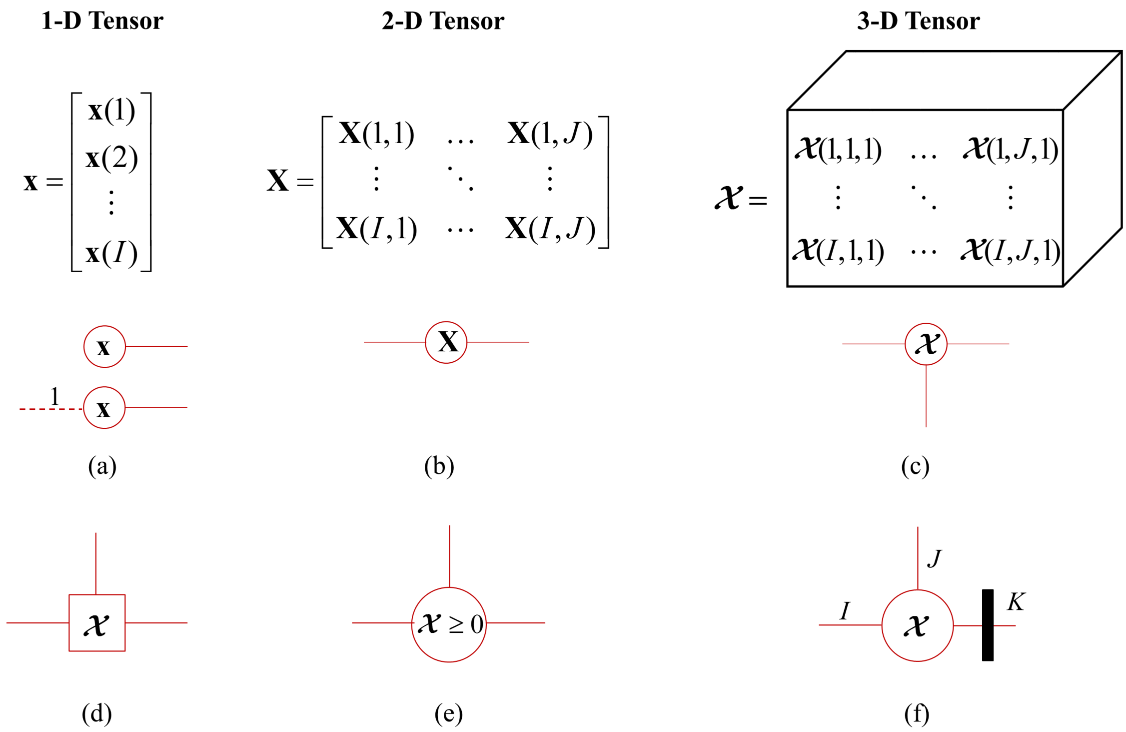


Figure 2.3 P Diagrams of (a) a vector $\mathbf{x} \in \mathbb{R}^{I}$ (b) a matrix $\mathbf{X} \in \mathbb{R}^{I\times J}$ (c) a tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I\times J\times K}$. Note that the number of lines leaving a node is equal to the order of the tensor. (d) Tensor with constant values (e) Nonnegative tensor. Let $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I\times J\times K}$, then $\boldsymbol{\mathcal{X}}$ is a nonnegative tensor if $\boldsymbol{\mathcal{X}}(i,j,k) \geq 0$ for $i = 1,\ldots,I, j = 1,\ldots,J, k = 1,\ldots,K$ (f) Orthogonal tensors are depicted with a square bar on the orthogonal dimension. For the example given, let $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I\times J\times K}$, then $\boldsymbol{\mathcal{X}}_{(3)}^{T}\boldsymbol{\mathcal{X}}_{(3)} = \mathbf{I}$. Adapted from [2].

Some of the tensor operations presented in Section 2.2 are described graphically in Figure 2.4. Note that complicated mathematical expressions in higher dimensions can be easily illustrated with the P Diagrams.

2.3.2 Markov-Penrose Diagrams

P-Diagrams are good representation of higher order arrays and operations between them. However these models do not contain information about the probabilistic dependency between tensors. In [2], a new type of graphical representation called as Markov-Penrose Diagram (M-P Diagram) is proposed to incorporate the Directed Acyclic Graphs (DAGs) [15] with the P Diagrams. In M-P diagrams, undirected links are used for arithmetic operations between tensors whereas directed arrows signify the

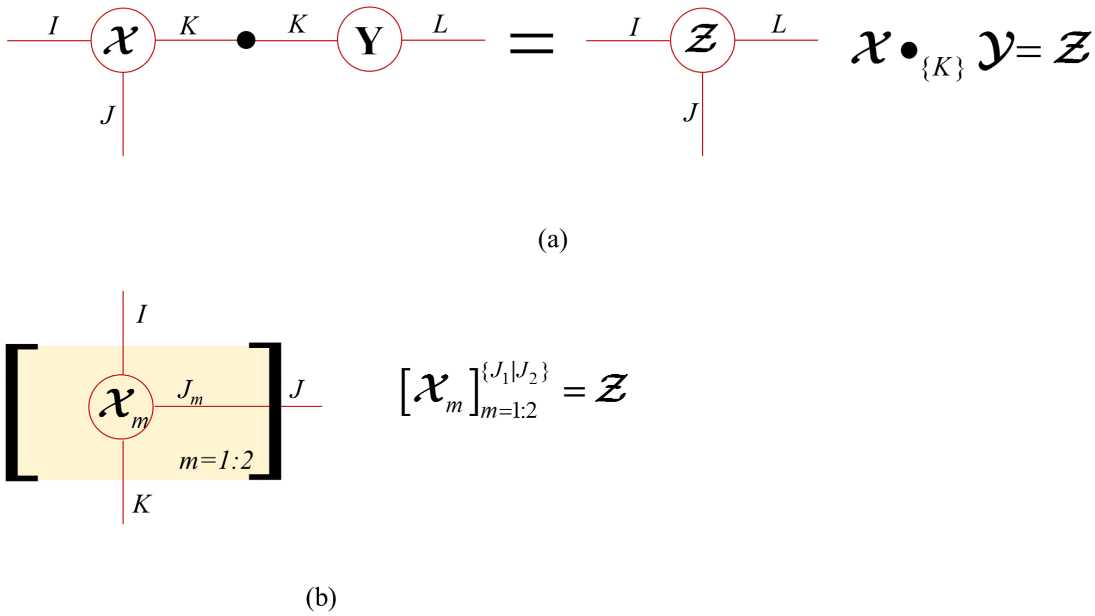


Figure 2.4 P Diagrams for the contraction and concatenation operators. (a) Contraction operation is denoted with a black dot. Contraction of $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I\times J\times K}$ with $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{K\times L}$ on the Kth dimension gives $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{I\times J\times L}$. (b) Concatenation operation defined on a set. Concatenation of $\boldsymbol{\mathcal{X}}_1 \in \mathbb{R}^{I\times J_1\times K}$ with $\boldsymbol{\mathcal{X}}_2 \in \mathbb{R}^{I\times J_2\times K}$ gives $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{I\times J_1+J_2\times K}$. In the diagram the number of tensors m to be concatenated is shown explicitly. The dimension on which concatenation takes place changes outside of the bracket, in this example $J = \sum_{m=1}^{2} J_m$. Adapted from [2].

conditional dependence. By this way, probabilistic models for tensors including imposing a prior distribution on the nodes could be explicitly shown in one diagram. Figure 2.5 shows basic notation in M-P diagrams and two models that compare the DAGs with the M-P notation. These models are selected as the inverse problems of EEG and fMRI which will be explored in detail in Chapter 4.

Cichocki has pointed out the correspondence between certain types of tensor networks and graphical models [16]. In a more detailed analysis, Critch *et al.* showed similarities between a special type of tensor networks - matrix product states and Hidden Markov Models [17]. On the other hand, Yılmaz proposed a novel representation of tensor factorization models that are similar to undirected graphs [18]. M-P Diagrams differ from those models by unifying graphical models with the P Diagrams.

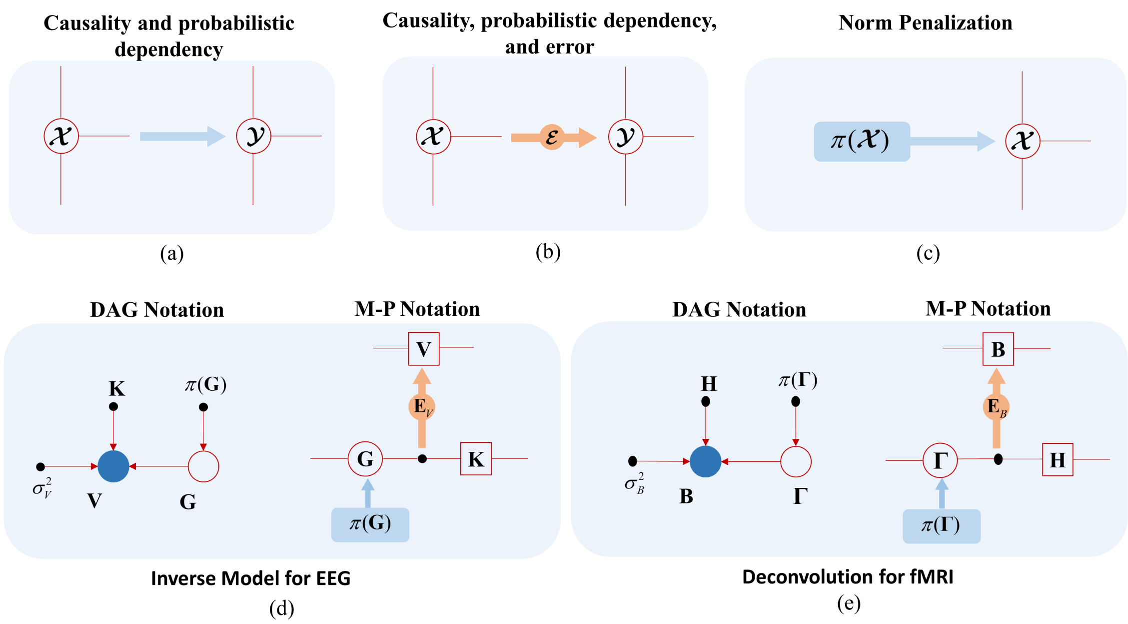


Figure 2.5 M-P Diagrams. (a) An arrow between two tensors indicates a probabilistic dependency between them. (b) Additive error term is added as a circle on the arrow by using $\boldsymbol{\mathcal{E}}$ for tensors and $\mathbf{E}$ for matrices. (c) Prior distribution $\pi(\boldsymbol{\mathcal{X}})$ is denoted by a square with an arrow (d) DAG and M-P graphical notations of the EEG generative model $\mathbf{V} = \mathbf{KG} + \mathbf{E}_V$ is shown. $\mathbf{V}$ is the EEG signal measured on the scalp, $\mathbf{G}$ is the primary current density and $\mathbf{K}$ is the lead field matrix. $\mathbf{G}$ has a prior distribution. (e) fMRI generative model $\mathbf{B} = \mathbf{\Gamma H} + \mathbf{E}_B$ is depicted in DAG and M-P notations. $\mathbf{B}$ is the measured BOLD signal, $\mathbf{\Gamma}$ is the vasoactive feedforward signal and $\mathbf{H}$ is the hemodynamic response function.

3. TENSOR METHODS

Tensor based methods have become a popular tool for handling the high dimensional data in various areas including psychometrics, chemometrics, computer vision and neuroscience. Since the multiway analysis methods have been introduced into the neuroimaging literature, they have attracted great attention. From the first application of multiway analysis on the decomposition of EEG signal [19] and linking the EEG and fMRI activity in time [20], the literature in this field is expanding [21, 22, 23, 24, 25, 26, 27].

The main reason for this interest is that multidimensional nature of neuroimaging data constituted by three dimensional space, time, subjects and even trials can be captured by tensor analysis and underlying structure of data can be represented by a few numbers of components. Decomposition or factorization methods including canonical decomposition [28], Tucker [29], multiway partial least squares [30] are widely used.

Multiway methods are also used for the representation of multivariate functions for the solution of high dimensional integrals, stochastic and parametric partial differential equations, multidimensional convolution in many areas [31]. Recently, new decomposition methods including tensor train [32] and hierarchical Tucker decomposition [33] are introduced for low rank tensor approximations. These methods provide high compression rates for high dimensional data and avoid problems in other multidimensional decomposition methods. These methods are out of the scope of this thesis.

In this chapter, two well-known decomposition methods and data fusion methods based on decomposition will be presented.

3.1 Tensor Decompositions

3.1.1 Parallel Factor Analysis

Parallel factors analysis (PARAFAC) is a decomposition method for higher order arrays which can be considered as a generalization of principal components analysis (PCA). PARAFAC was independently introduced by Harshmann [28] as Parallel Factors and Carroll and Chang [34] as Canonical Decomposition. Möcks independently discovered PARAFAC for event related potentials in which EEG data is organized as a third order tensor of dimensions channel, time and subject and this version of PARAFAC was called as Topographic Component Analysis [35]. In the context of brain imaging, Field and Graupe [36] used PARAFAC to extract consistent ERP components across channels and between subjects and later Miwakeichi *et al.* used PARAFAC for the spectral component extraction [19]. Since then PARAFAC has been used in neuroimaging literature extensively (For a review refer to [2, 37]).

Let $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$, the PARAFAC decomposition of $\boldsymbol{\mathcal{X}}$ is stated as

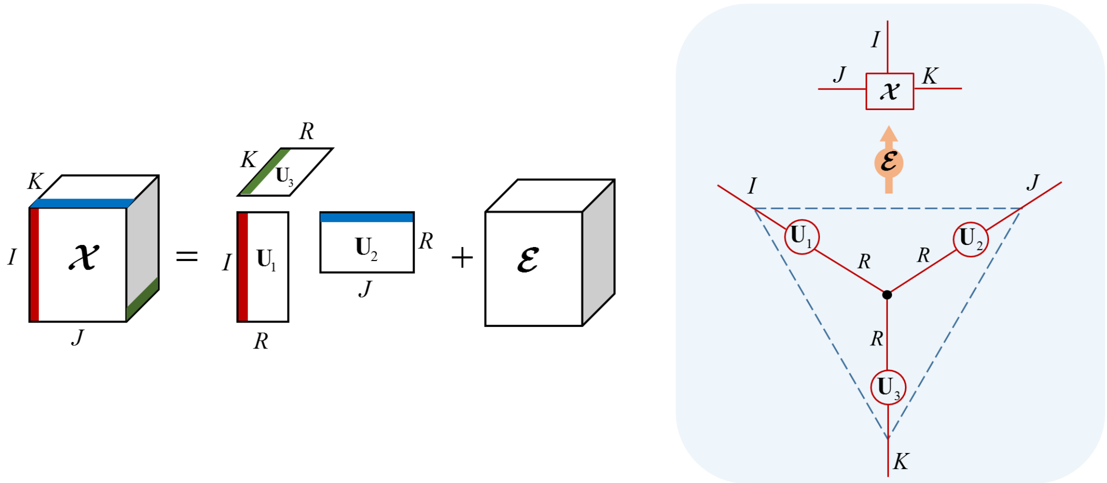
$$\boldsymbol{\mathcal{X}}(i_1, i_2, \ldots, i_N) = \sum_{r=1}^{R} \mathbf{U}_1(i_1, r)\mathbf{U}_2(i_2, r) \ldots \mathbf{U}_N(i_N, r) + \boldsymbol{\mathcal{E}}(i_1, i_2, \ldots, i_N) \tag{3.1}$$

where $\mathbf{U}_1 \in \mathbb{R}^{I_1 \times R}$ to $\mathbf{U}_N \in \mathbb{R}^{I_N \times R}$ are the factor matrices, R is the number of components and $\boldsymbol{\mathcal{E}}$ is the error term. Figure 3.1 shows three-dimensional illustration and M-P Diagram of this model. PARAFAC model can be expressed in the Kruskal notation as [38]

$$\boldsymbol{\mathcal{X}} = [\![\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_N]\!] + \boldsymbol{\mathcal{E}}. \tag{3.2}$$

We can write Eq. 3.1 in matrix format by using mode-n unfolding of the tensor $\boldsymbol{\mathcal{X}}$ and Khatri-Rao product of the factor matrices as follows.

$$\boldsymbol{\mathcal{X}}_{(n)} = \mathbf{U}_n(\mathbf{U}_N \odot \cdots \odot \mathbf{U}_{n-1} \odot \mathbf{U}_{n+1} \odot \cdots \odot \mathbf{U}_1)^T + \boldsymbol{\mathcal{E}}_{(n)} \tag{3.3}$$



(a) (b)

Figure 3.1 Graphical representation of the PARAFAC model for a 3-D tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I\times J\times K}$. $\mathbf{U}_1, \mathbf{U}_2$ and $\mathbf{U}_3$ are the factor matrices. (a) Three dimensional representation (b) M-P Diagram of the same model. Latent variables (components) are denoted by circles and the observed variable (tensor) is denoted by a rectangle.

The PARAFAC model is symmetric and all the factors are treated in the same sense. The most attractive property of the PARAFAC is the uniqueness of the model. We refer the uniqueness in the sense of rotational indeterminacy. It is well-known that in matrix decomposition the factor matrices are not unique. Consider the decomposition of the matrix $\mathbf{Y} \in \mathbb{R}^{I\times J}$

$$\mathbf{Y} = \mathbf{A}\mathbf{B}^T + \mathbf{E}. \tag{3.4}$$

The same model can be obtained by multiplying $\mathbf{A}$ with any non-singular matrix $\mathbf{W}$ and $\mathbf{B}$ with the inverse $(\mathbf{W}^{-1})^T$ as

$$\mathbf{Y} = \mathbf{A}\mathbf{W}\mathbf{W}^{-1}\mathbf{B}^T + \mathbf{E}. \tag{3.5}$$

The uniqueness is provided in PCA by imposing orthogonality on the factors and statistical independency in independent component analysis. However, an R component PARAFAC model gives unique solutions up to scaling and permutation indeterminacies. Scaling ambiguity indicates that the magnitude of the factors is arbitrary whereas due to permutation ambiguity factors can be reordered without changing the model. To avoid this confusion, factors of the PARAFAC are scaled to unit norm and the norm is absorbed in one of the factors and they are ordered according to the ascending of their variance.

The classical and the well-known uniqueness condition is shown by Kruskal [38] based on the k-rank. The k-rank of a matrix $\mathbf{A}$ denoted by k_A is the largest number such that every subset of k_A columns of $\mathbf{A}$ is linearly independent. For the R order PARAFAC model of a third order tensor $\boldsymbol{\mathcal{X}} \approx [\![A, B, C]\!]$ the sufficient condition for the uniqueness is $k_A + k_B + k_C \geq 2R + 2$. This condition is generalized for Nth order tensor in [39] as $\sum_{n=1}^{N} k_{U_n} \geq 2R + N - 1$. For the recent discussions on the uniqueness conditions see [40, 41].

The factor matrices are found by minimizing the sum of squares of the residuals.

$$\{\hat{\mathbf{U}}, \ldots, \hat{\mathbf{U}}_N\} = \underset{\{\mathbf{U}, \ldots, \mathbf{U}_N\}}{\arg\min} \|\boldsymbol{\mathcal{X}} - \hat{\boldsymbol{\mathcal{X}}}\|_2^2 \tag{3.6}$$

PARAFAC model can be solved efficiently by using a sequential algorithm like alternating least squares (ALS). In ALS algorithm, at every step all factors except one are fixed and the problem is solved for that one until all factors are estimated. Since each step of the ALS is a linear regression, penalization methods can be incorporated naturally. A modification of ALS is the hierarchical alternating least squares (HALS) algorithm in which, at each step of ALS, only one of the components of a factor is estimated, fixing other factors of all components [42].

ALS has been improved with line search at each step [43], though it can converge slowly, especially when the components are collinear. Other methods such as gradient-based optimization methods [44] and generalized Schur decomposition [45] have been developed as an alternative to overcome the limitations of ALS. In addition, probabilistic methods for general tensor factorizations are presented in [46, 47].

For an N-D tensor, ALS algorithm is presented in Figure 3.2.

in: $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$

in: model order $R \succeq 0$

Initialization

for $n = 1$ **to** N **do**

  Initialize $\mathbf{U}_n$ randomly or leading eigenvectors of the unfolded tensor

end for

repeat

  for $n = 1$ **to** N **do**

    Fix $\mathbf{U}_1, \ldots, \mathbf{U}_{n-1}, \mathbf{U}_{n+1}, \ldots, \mathbf{U}_N$

    $\mathbf{G} = (\mathbf{U}_N \odot \cdots \odot \mathbf{U}_{n+1} \odot \mathbf{U}_{n-1} \cdots \odot \mathbf{U}_1)$

    $\mathbf{U}_n = \boldsymbol{\mathcal{X}}_{(n)} \mathbf{G} (\mathbf{G}^T \mathbf{G})^{\dagger}$

    if $n \neq N$ **then**

      $\mathbf{U}_n \leftarrow \mathbf{U}_n / ||\mathbf{U}_n||_2$

    end if

  end for

until $||\boldsymbol{\mathcal{X}} - \hat{\boldsymbol{\mathcal{X}}}||_2 / ||\boldsymbol{\mathcal{X}}||_2 < \epsilon$

out: $\mathbf{U}_n \in \mathbb{R}^{I_n \times R}$ **for** $n = 1, \ldots, N$

Figure 3.2 PARAFAC - ALS Algorithm

3.1.2 Tucker Decomposition

Tucker decomposition can also be considered as an extension of PCA to higher dimensions. Tucker decomposition was first introduced by Tucker in 1966 [29] and has been used in various areas under different names related to PCA and SVD such as three-mode PCA, higher order SVD etc.

Let $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$, the Tucker decomposition is defined as

$$\begin{aligned}\boldsymbol{\mathcal{X}}&(i_1, i_2, \ldots, i_N) \\ &= \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \cdots \sum_{r_N=1}^{R_N} \boldsymbol{\mathcal{G}}(r_1, r_2, \ldots, r_N) \mathbf{U}_1(i_1, r_1) \mathbf{U}_2(i_2, r_2) \ldots \mathbf{U}_N(i_N, r_N) \\ &\qquad + \boldsymbol{\mathcal{E}}(i_1, i_2, \ldots, i_N) \qquad (3.7)\end{aligned}$$

where $R_1, \ldots, R_N$ are the number components of the factor matrices $U_1 \in \mathbb{R}^{I_1 \times R_1}$ to $U_N \in \mathbb{R}^{I_N \times R_N}$, respectively. $\boldsymbol{\mathcal{G}} \in \mathbb{R}^{R_1 \times R_2 \times \cdots \times R_N}$ is the core tensor. The core tensor defines the interaction between the factors. In shorthand notation, this model is equal to

$$\boldsymbol{\mathcal{X}} = [\![\boldsymbol{\mathcal{G}}; \mathbf{U}_1, \mathbf{U}_2 \ldots, \mathbf{U}_N]\!] + \boldsymbol{\mathcal{E}}. \qquad (3.8)$$

By using the mode-n unfolding of $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{G}}$, the matricized version of Eq. 3.7 is given as

$$\boldsymbol{\mathcal{X}}_{(n)} = \mathbf{U}_n \boldsymbol{\mathcal{G}}_{(n)} (\mathbf{U}_N \otimes \cdots \otimes \mathbf{U}_{n-1} \otimes \mathbf{U}_{n+1} \otimes \cdots \otimes \mathbf{U}_1)^T + \boldsymbol{\mathcal{E}}_{(n)}. \qquad (3.9)$$

PARAFAC can be considered as the special case of the Tucker decomposition in which the core tensor is super-diagonal and $R_1 = R_2 = \cdots = R_N$. A tensor $\boldsymbol{\mathcal{G}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is called as super-diagonal if $\boldsymbol{\mathcal{G}}(i_1, i_2, \ldots, i_N) \neq 0$ only if $i_1 = i_2 = \cdots = i_N$. Unlike PARAFAC, Tucker decomposition does not ensure uniqueness. The factor matrices can be rotated by multiplying the core tensor with the proper matrix.

Orthogonality is imposed on the factor matrices for the identifiability of the model [48].

3.2 Tensor Based Data Fusion

We will review some of the tensor based models used for the fusion of data generated from multiple sources. Unlike matrix based data fusion methods, tensor methods can handle heterogeneous datasets i.e. variables with different orders.

3.2.1 Multiway Partial Least Squares

Partial least squares (PLS) is a well-known method that the independent variables cast in a matrix are decomposed into scores and dependent variables are regressed on those scores. Multiway PLS (N-PLS) is an extension of matrix based PLS to higher dimensions [30] in which the independent and dependent variables are decomposed in such a way that the score vectors have maximal covariance.

Let $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I\times J\times K}$ and $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{I\times L\times M}$ be two tensors. Then the decomposition formulations of the N-PLS model based on PARAFAC are given as

$$\begin{aligned} \boldsymbol{\mathcal{X}} &= [\![\mathbf{T}, \mathbf{P}, \mathbf{Q}]\!] + \boldsymbol{\mathcal{E}}_x \\ \boldsymbol{\mathcal{Y}} &= [\![\mathbf{U}, \mathbf{B}, \mathbf{C}]\!] + \boldsymbol{\mathcal{E}}_y \end{aligned} \tag{3.10}$$

where $\mathbf{P} \in \mathbb{R}^{J\times R}$ and $\mathbf{Q} \in \mathbb{R}^{K\times R}$ are the loadings of $\boldsymbol{\mathcal{X}}$ and $\mathbf{B} \in \mathbb{R}^{L\times R}$ and $\mathbf{C} \in \mathbb{R}^{M\times R}$ are the loadings of $\boldsymbol{\mathcal{Y}}$.

The objective of this N-PLS model is to find latent vectors stacked into $\mathbf{T}, \mathbf{U} \in \mathbb{R}^{I\times R}$ matrices that satisfy

$$\mathbf{U} = \mathbf{TD} + \mathbf{E}_U. \tag{3.11}$$

The objective function of the N-PLS can also be formulated as

$$\begin{aligned} f(\mathbf{P},\mathbf{Q},\mathbf{B},\mathbf{C}) = & \underset{\mathbf{P},\mathbf{Q},\mathbf{B},\mathbf{C}}{\arg\min}\{\mathrm{cov}(\mathbf{T},\mathbf{U})\} \\ & \text{s. t. } \boldsymbol{\mathcal{X}} \approx [\![\mathbf{T},\mathbf{P},\mathbf{Q}]\!], \quad \boldsymbol{\mathcal{Y}} \approx [\![\mathbf{U},\mathbf{B},\mathbf{C}]\!] \\ & \|\mathbf{P}\|_2 = \|\mathbf{Q}\|_2 = \|\mathbf{B}\|_2 = \|\mathbf{C}\|_2 = 1. \end{aligned} \tag{3.12}$$

In the multiway PLS, the number of dimensions of dependent and independent variables may change. For instance $\boldsymbol{\mathcal{X}}$ can be a 3-D tensor and $\mathbf{y}$ can be a vector. In that case the covariance between $\mathbf{T}$ and $\mathbf{y}$ is maximized.

In the decomposition step of the matrix PLS, low-rank and subspace approximation are the same. However, in multiway case these two approximations lead to different models [49]. The equivalent of the low-rank approximation in higher dimensions is the PARAFAC and the equivalent of the subspace approximation is the Tucker model. An improved version of N-PLS is suggested in [49] by replacing the PARAFAC with the Tucker decomposition in Eq. 3.10 that gives a better-fitting model.

Another improvement of N-PLS is the higher order PLS (HOPLS). HOPLS is a subspace approximation method in which orthogonal Tucker decompositions are used for the decomposition of both dependent and independent data. Consider two tensors $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ and $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{J_1 \times \cdots \times J_M}$ and assume that $I_1 = J_1$. In HOPLS, $\boldsymbol{\mathcal{X}}$ is approximated by a sum of rank-$(1, L_2, \ldots, L_N)$ decompositions and $\boldsymbol{\mathcal{Y}}$ is approximated by a sum of rank-$(1, K_2, \ldots, K_N)$ decompositions. The HOPLS model is expressed as

$$\begin{aligned} \boldsymbol{\mathcal{X}} &= \sum_{r=1}^{R} [\![\boldsymbol{\mathcal{G}}_r, \mathbf{t}_r, \mathbf{P}_r^{(1)}, \mathbf{P}_r^{(N-1)}]\!] + \boldsymbol{\mathcal{E}}_x \\ \boldsymbol{\mathcal{Y}} &= \sum_{r=1}^{R} [\![\boldsymbol{\mathcal{D}}_r, \mathbf{t}_r, \mathbf{Q}_r^{(1)}, \mathbf{Q}_r^{(M-1)}]\!] + \boldsymbol{\mathcal{E}}_y \end{aligned} \tag{3.13}$$

where R is the number of latent vectors. $\mathbf{P}_r^{(n)} \in \mathbb{R}^{I_{n+1} \times L_{n+1}}$ and $\mathbf{Q}_r^{(m)} \in \mathbb{R}^{J_{n+1} \times K_{n+1}}$ are the loading matrices corresponding latent vector $\mathbf{t}_r$. $\boldsymbol{\mathcal{G}}_r \in \mathbb{R}^{1 \times L_2 \times \cdots \times L_N}$ and $\boldsymbol{\mathcal{D}}_r \in \mathbb{R}^{1 \times K_2 \times \cdots \times K_M}$ are core tensors. To ensure uniqueness, the loading matrices are required to be columnwise orthogonal and core tensors to be all-orthogonal. The loading matrices are estimated by maximizing the covariance on the first mode $\boldsymbol{\mathcal{C}} = \text{cov}(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}}) = \boldsymbol{\mathcal{X}} \bullet_{\{I_1\}} \boldsymbol{\mathcal{Y}}$ with the orthogonality constraints. Details on the estimation and algorithm can be found in [50].

3.2.2 Tensor Canonical Components Analysis

Since Hotelling's formulation [51] in 1936, canonical correlation analysis (CCA) has been known and used for searching linear relations between two variables. CCA finds the best subspaces i.e. transformations of the variables that the two variables or arrays have the maximum correlation.

Consider two data matrices $\mathbf{X} \in \mathbb{R}^{I \times J}$ and $\mathbf{Y} \in \mathbb{R}^{K \times J}$, CCA algorithm finds the transformations $\mathbf{u} \in \mathbb{R}^{I}$ and $\mathbf{v} \in \mathbb{R}^{K}$ that maximize the correlation between $\mathbf{x}' = \mathbf{u}^T\mathbf{X}$ and $\mathbf{y}' = \mathbf{v}^T\mathbf{Y}$ formulated by

$$\max_{\mathbf{u},\mathbf{v}} \frac{\mathrm{E}[\mathbf{x}'\mathbf{y}']}{\sqrt{\mathrm{E}[(\mathbf{x}'^2)(\mathbf{y}')^2]}} = \frac{\mathbf{u}^T\mathbf{X}\mathbf{Y}^T\mathbf{v}}{\sqrt{\mathbf{u}^T\mathbf{X}\mathbf{X}^T\mathbf{u}\mathbf{v}^T\mathbf{Y}\mathbf{Y}^T\mathbf{v}}} \tag{3.14}$$

where E is the empirical expectation. Multiple canonical correlations are found up to $R = \min(\text{rank}(\mathbf{X}, \mathbf{Y}))$ in which new pairs of $\mathbf{u}$ and $\mathbf{v}$ are orthogonal to the previous ones.

Tensor CCA (TCCA) is an extension of standard CCA to higher dimensions that considers two tensors [52]. Since tensors are multidimensional arrays, TCCA offers more possibilities on the number of shared modes. Note that CCA is applied on the unshared dimension and 3-D tensors can share any single or multiple dimensions.

We will present the single shared mode TCCA proposed in [52, 53] for 3-D tensors that finds the canonical transformations for two modes. Let $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I \times J \times K}$ and $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{L \times M \times K}$ share the third dimension. In single shared mode TCCA, pairs of linear transforms $\mathbf{u}_1 \in \mathbb{R}^{I}, \mathbf{u}_2 \in \mathbb{R}^{J}$ and $\mathbf{v}_1 \in \mathbb{R}^{L}, \mathbf{v}_2 \in \mathbb{R}^{M}$ are found that maximize the correlation between projected tensors. This is formulated as

$$\max_{\mathbf{u}_1,\mathbf{u}_2,\mathbf{v}_1,\mathbf{v}_2} = \frac{\mathrm{E}[\mathbf{x}'\mathbf{y}']}{\sqrt{\mathrm{E}[(\mathbf{x}'^2)(\mathbf{y}')^2]}} \tag{3.15}$$

where $\mathbf{x}' = \boldsymbol{\mathcal{X}} \bullet_{\{I\}} \mathbf{u}_1 \bullet_{\{J\}} \mathbf{u}_2$ and $\mathbf{y}' = \boldsymbol{\mathcal{Y}} \bullet_{\{L\}} \mathbf{v}_1 \bullet_{\{M\}} \mathbf{v}_2$. The linear transformations are found by using an alternating algorithm and SVD. For more details refer to [54].

3.2.3 Coupled Tensor Factorization

Coupled tensor factorization (CTF) is the joint decomposition of two or more tensors in which tensors are coupled on single or multiple nodes. First examples are found in [55] as linked PARAFAC and in [56] as multiway multiblock models. Later it is improved by [57] for structured data fusion. Unlike N-PLS or TCCA, CTF finds the factors that are shared between tensors. CTF can also be used for the joint factorization of multiple tensors that are coupled in more than one dimension.

Let $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I\times J\times K}$ and $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{I\times L\times M}$, the CTF model for coupling on the first factor is given as

$$\min \left\{ \|\boldsymbol{\mathcal{X}} - [\![\mathbf{T}, \mathbf{A}_1, \mathbf{A_2}]\!]\|_2^2 + \|\boldsymbol{\mathcal{Y}} - [\![\mathbf{T}, \mathbf{B}_1, \mathbf{B_2}]\!]\|_2^2 \right\} \tag{3.16}$$

where $\mathbf{T} \in \mathbb{R}^{I\times R}$ is the common factor, $\mathbf{A}_1 \in \mathbb{R}^{J\times R}$ and $\mathbf{A}_2 \in \mathbb{R}^{K\times R}$ are the individual factors of $\boldsymbol{\mathcal{X}}$ and $\mathbf{B}_1 \in \mathbb{R}^{L\times R}$ and $\mathbf{B}_2 \in \mathbb{R}^{M\times R}$ are the individual factors of $\boldsymbol{\mathcal{Y}}$. R is the model order of the PARAFAC.

It is important to note that if one of the arrays is a tensor and the other is a matrix, the term Coupled Matrix Tensor Factorization (CMTF) is used instead of CTF. Factors can be estimated by using an alternating algorithm or a gradient based first-order optimization method as described in [57].

In some cases, the common dimension may not be completely coupled but may only share a few components that will allow common and discriminative subspace decompositions of the tensors. If we continue with the model in Eq. 3.16, the shared factor $\mathbf{T}$ is represented as $\mathbf{T} = (\mathbf{U}|\mathbf{V})$ as the concatenation of common ($\mathbf{U}$) and discriminative ($\mathbf{V}$) factors for the tensor $\boldsymbol{\mathcal{X}}$. The coupled factor for the tensor $\boldsymbol{\mathcal{Y}}$ will be $\mathbf{T} = (\mathbf{U}|\mathbf{W})$ as the concatenation of common ($\mathbf{U}$) and discriminative ($\mathbf{W}$) factors. The objective function takes the form

$$\begin{aligned} f(\mathbf{U}, \mathbf{V}, \mathbf{A}_1, \mathbf{A}_2, \mathbf{W}, \mathbf{B}_1, \mathbf{B}_2) = \{ &\| \boldsymbol{\mathcal{X}} - [\![(\mathbf{U}|\mathbf{V}), \mathbf{A}_1, \mathbf{A_2}]\!] \|_2^2 \\ &+ \| \boldsymbol{\mathcal{Y}} - [\![(\mathbf{U}|\mathbf{W}), \mathbf{B}_1, \mathbf{B}_2]\!] \|_2^2 \}. \end{aligned} \tag{3.17}$$

The estimation of factors in this model can be performed with the ALS algorithm. HALS algorithm serves a good option in case of regularization on the common and discriminant factors of the coupled factor. We used this type of CTF for the EEG/fMRI fusion on the cortical surface of the brain which will be explained in detail in Chapter 4.

4. FUSION OF EEG AND FMRI ON THE CORTICAL SURFACE

4.1 Electroencephalography

Scalp EEG signals are a direct measure of the brain electric activity by reflecting the postsynaptic cortical currents generated by the large pyramidal neurons which are located perpendicularly to the cortical surface [58]. The temporal resolution of EEG is high in the 1 to 5 kHz range. However, its spatial resolution is hampered by the small number of measurement sites (electrodes) and the inherent volume conduction effect.

The electrical fields generated by single neurons are too small to be detected at the scalp with the EEG. Thus, the main source of the EEG is the volume currents produced by the temporal and spatial alignment of the pyramidal neurons creating dipoles in macrocolumns. Even though EEG is directly related to the spatially summed bioelectrical activity, precisely localizing the neural activity with EEG is not possible due to the ill-posed inverse problem. Inverse problem is based on the estimation of source configuration that might have caused the potential distribution measured from the surface of an electrically conductive volume, in this case the brain [59]. Ill-posed nature of inverse problem arises from the determination of high number of unknown sources (electrical dipoles) from limited and predetermined number of measurement channels (sensors). For the solution of the inverse problem, the field distribution of a current dipole in the volume conductor that the current propagates is modeled by using the quasi-static approximation of Maxwell equations. The computation of scalp potentials for a known set of neural generators is known as the forward problem. The discretized version of the forward problem is

$$\mathbf{V} = \mathbf{K}\mathbf{G} + \mathbf{E}_{\mathbf{V}} \tag{4.1}$$

where $\mathbf{V} \in \mathbb{R}^{I_E \times I_T}$ is the scalp potential measured by EEG from I_E electrodes at I_T time points. $\mathbf{K} \in \mathbb{R}^{I_E \times I_{Cx}}$ is the lead field or gain matrix that contains the geometric and conductive information about the head volume conductor. Lead field matrix projects I_{Cx} sources on the I_E electrodes. $\mathbf{G} \in \mathbb{R}^{I_{Cx} \times I_T}$ is the primary current density and $\mathbf{E} \in \mathbb{R}^{I_E \times I_T}$ is the noise at the sensors.

Forward model can be analytically solved if the head volume conductor is assumed as three or four concentric spherical shells with different isotropic conductivities. The shells represent the brain, the cerebrospinal fluid, the skull and the scalp tissues [60]. In more realistic head models high resolution anatomical MR image is used to extract the surface boundaries between the brain, the skull and the scalp. Forward fields are calculated by using boundary element method (BEM) assuming isotropy in each tissue compartment [61]. The other method based on the realistic head model is the Finite Element Method (FEM). FEM uses DTI images to model the anisotropy in white matter tracts. Although the numerical methods used in BEM and FEM are computationally demanding algorithms, they offer more accurate solutions to the forward models with respect to analytical methods [59].

Estimation of the $\mathbf{G}$ is known as the inverse problem. Inverse problem approaches are grouped as parametric and imaging methods. In parametric methods sources are modeled as limited number of dipoles - less than the number of sensors - and the strength and orientation of the dipoles are estimated by using nonlinear methods. Imaging methods assume that the sources are located on the brain mesh created by the tessellation of the brain [59]. In this type of method source density is obtained from

$$\arg\min_{G} \|\mathbf{V} - \mathbf{K}\mathbf{G}\|_2^2 + \pi(\mathbf{G}). \tag{4.2}$$

Since the number of sources are much higher than the number of sensors, $I_E \ll I_{Cx}$, the model in Eq. 4.2 is underdetermined. A penalization term $\pi(\mathbf{G})$ is applied based on the anatomical, physiological or mathematical constraints [61]. Figure 4.1 describes the EEG inverse and forward models.

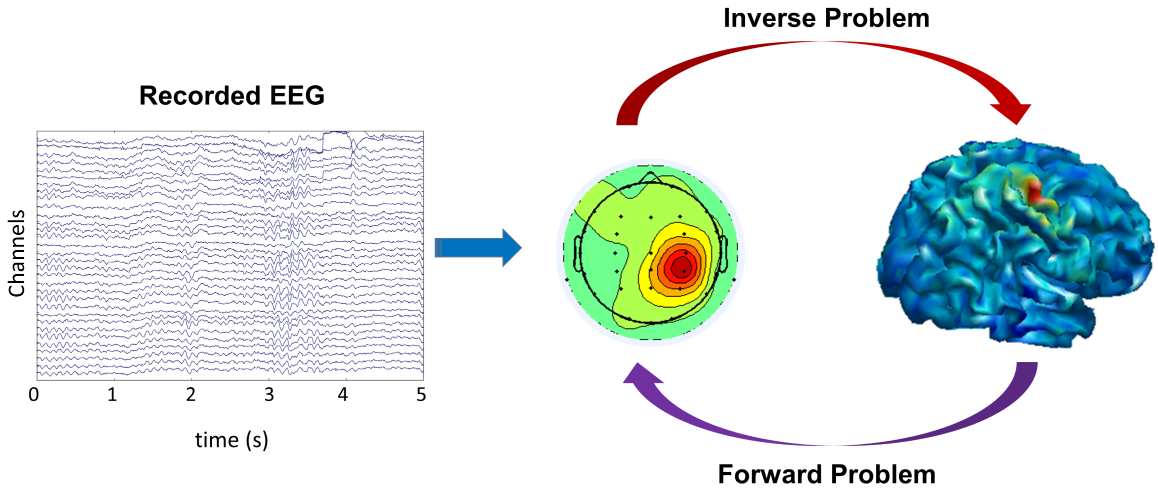


Figure 4.1 Illustration of the EEG inverse and forward problems. Forward problem calculates the distribution of sources on the scalp from a known source configuration. Inverse problem finds the localization of the sources.

4.2 Functional Magnetic Resonance Imaging

By utilizing the spatial power of MR imaging, fMRI indicates indirect neural activity through oxygen metabolism regulated by the oxygen consumption of the neural tissue, the cerebral blood flow and the cerebral blood volume.

fMRI first discovered by Ogawa *et al.* uses an endogenous contrast agent in the blood, deoxyhemoglobin which is a paramagnetic material [62]. Existence of high concentration rates of deoxyhemoglobin in blood distorts the MR signal. When a neural tissue is activated by a stimulus, cerebral blood flow (CBF) increases towards the activated region due to a demand on oxygen and other metabolites. The oxygen supply surpasses the need of the oxygen by the tissues leading an increase in oxyhemoglobin concentration and a decrease in the deoxyhemoglobin concentration. Finally, the relative loss in the concentration difference of deoxyhemoglobin gives rise to BOLD (blood oxygen level dependent) signal (refer to Figure 4.2). It is important to note that BOLD signal is confined to the time course of the slowly evolving hemodynamic activity ($\sim$ 10 s) while exhibiting a high spatial resolution in the order of millimeters.

The relation of the BOLD signal to neural activity has been a controversial topic.

By combining electro-physiological recordings with fMRI Logothetis has shown that the BOLD signal is mostly related to local field potentials (LFPs) rather than multiple unit activities (MUAs) that reflect spiking of neurons [63]. LFPs are generated by the extracellular currents of cell assemblies as a result of the postsynaptic potentials. Due to their low frequency signal content (< 200 Hz), their spatial extent is larger in comparison to action potentials. It was thought that the excitatory postsynaptic potentials were the primary source of the LFPs. However recent studies show that the inhibitory synaptic input may contribute to the LFPs as well [64, 65]. The experiments on monkeys have shown that the temporal course of BOLD signal closely follows of LFPs even the MUA has returned to baseline [63, 66].

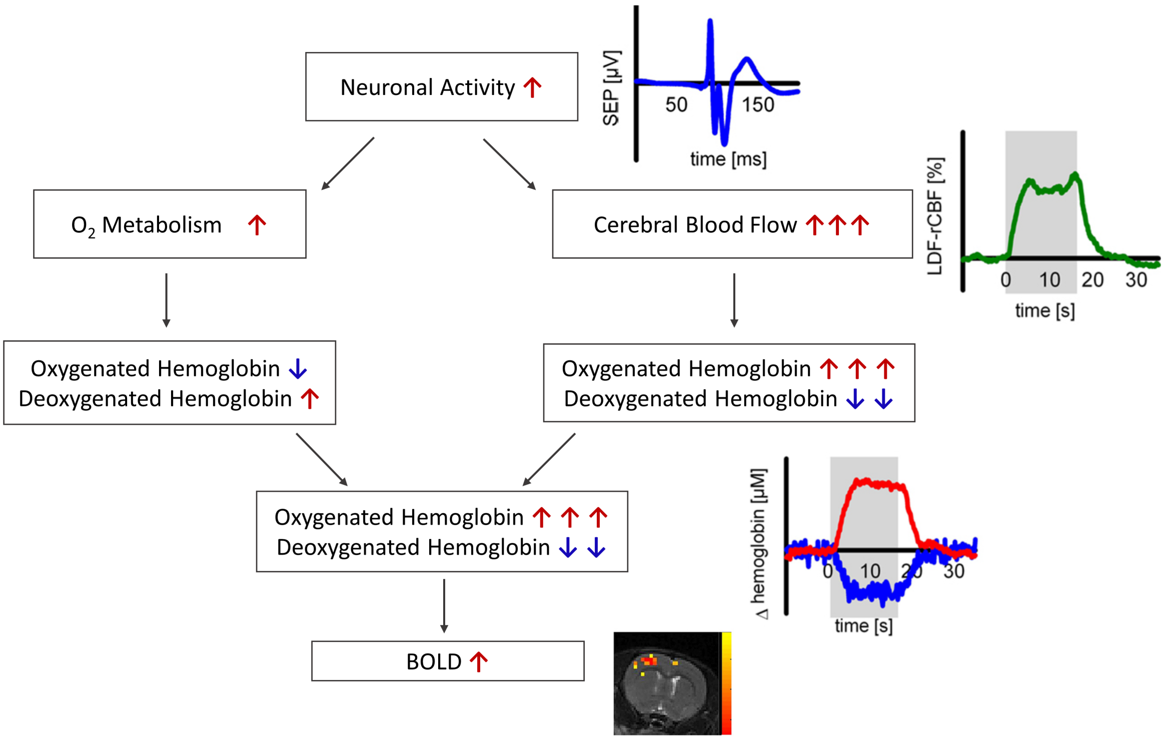


Figure 4.2 Generation of the BOLD response. Adapted from [67].

The forward model for BOLD is nonlinear. The first attempt for modeling the BOLD response was the balloon model that uses cerebral blood volume and deoxyhemoglobin concentration as the state variables, CBF as the input and the BOLD signal as the output. This nonlinear model has extended by adding the dynamic coupling of synaptic activity and flow [68]. More comprehensive models including different cell types and neuronal activity can be found in [69, 70, 71]. For simplicity, we will assume that the forward model of BOLD as linear and use the generative model below [2].

$$\mathbf{B} = \mathbf{\Gamma H} + \mathbf{E_B} \tag{4.3}$$

where $\mathbf{B} \in \mathbb{R}^{I_{Cx} \times I_T}$ is the BOLD signal measured from I_{Cx} voxels at $I_{T\delta}$ time points. Voxel is the smallest volume element of a three-dimensional image. Since the temporal resolution of BOLD is much smaller than that of the EEG $I_{T\delta} \ll I_T$, we will use the symbol δ to emphasize the sampling. $\mathbf{\Gamma} \in \mathbb{R}^{I_{Cx} \times I_T}$ is the vasoactive feed forward signal (VFFS) matrix that links the BOLD signal to neural activity. $\mathbf{H} \in \mathbb{R}^{I_T \times I_{T\delta}}$ is the hemodynamic response matrix whose rows are constituted by shifting the known hemodynamic response function at a finer temporal resolution.

The temporal deconvolution of the fMRI may also be stated as an inverse problem

$$\arg\min_{\mathbf{\Gamma}} \|\mathbf{B} - \mathbf{\Gamma H}\|_2^2 + \pi(\mathbf{\Gamma}) \tag{4.4}$$

where $\pi(\mathbf{\Gamma})$ is the prior for the VFFS. Glover used this deconvolution model with the penalty function $\pi(\mathbf{\Gamma}) = \|\mathbf{\Gamma}\|_2^2$ in the Wiener filter concept [72] and later Valdes-Sosa *et al.* proposed a model based on the $\pi(\mathbf{\Gamma}) = \|\mathbf{L\Gamma}\|_2^2$ where $\mathbf{L}$ is the second order Laplacian operator [3].

4.3 Fusion of EEG and fMRI

The integration of EEG and fMRI on a common space and/or time scale by merging the superiorities of different imaging modalities, to reveal the complex dynamics of brain functions and neuronal interactions, is one of the major current problems of the neuroimaging research. The integration of these two imaging modalities will take the advantage of high temporal resolution of EEG and high spatial resolution of fMRI.

Before giving details in the fusion methods, we want to mention the limitations of the EEG-fMRI fusion that arise from physiological processes. EEG and fMRI sources could be at disparate locations due to distance between neuronal population and vascular tree. Also an increase in BOLD signal does not necessarily mean an increase in neuronal activity. Neurotransmitter synthesis, glial cell metabolism and maintenance of the steady-state transmembrane potential also require oxygen consumption. It is well known that EEG shows the level of synchronization. However hemodynamic activity may also be caused by nonsynchronous activity. In theses cases inconsistency between EEG and fMRI may happen. If the electrophysiological activity is transient, it might not induce any detectable metabolic activity changes [7]. Large EEG amplitudes can be produced by epileptic foci while local metabolic signatures may be reduced due to the reduction in inhibitory activity [73].

Despite limitations of EEG-fMRI fusion, using both modalities gives the possibility of studying finer spatio-temporal structures of neuronal activity. We can elucidate more information at the common substrate of EEG and fMRI which would be harder when using only one modality.

4.4 EEG/fMRI Fusion Methods

There have been fusion studies to comprise the temporal and spatial resolution of EEG and fMRI for exploring the dynamics of brain functions. Fusion methods can be classified according to several criteria.

4.4.1 Asymmetrical Versus Symmetrical Fusion

In asymmetrical fusion one modality is used as a prior for the other modality. To localize the sources of bioelectrical activity measured with EEG, fMRI activation maps are used as priors for the inverse problem. Liu *et al.* used prior anatomical and functional MRI to regularize EEG/MEG inverse problem [74].

EEG to fMRI fusion techniques seek for the fMRI activation regions whose response is temporally correlated with the EEG signal. In this method EEG signals from certain channels which represent expected effect in the experiment, are convolved with the hemodynamic response function and are used as regressors in modeling the BOLD response after downsampling [75, 76]. One of the application of this method is to localize the electrical sources of epileptic discharges by using fMRI. Since time of the epileptic discharges can be easily determined from EEG, regressors are constituted by the convolution of the EEG signal belonging to the discharge time intervals. These studies showed decrease in BOLD during slow wave activity whereas an increase during fast electrical events such as spike and wave discharges [77, 78].

Simultaneous EEG/fMRI studies are also concerned with the hemodynamic correlates of spontaneous activity during resting state. It is well known that during resting state EEG shows a typical posterior rhythm in frequencies between 8-12 Hz named alpha rhythm. In these studies, EEG signals from alpha as well as other frequency bands are integrated into the general linear model to model the fMRI response with voxel based analysis. A positive correlation between thalamic BOLD and occipital alpha oscillations in EEG is observed whereas BOLD activation of occipital-parietal areas are found to be inversely correlated [79, 80, 81]. Inverse correlation originates from an increase in BOLD signal in the absence of marked alpha activity [82].

If there is no common substrate for underlying events of EEG and BOLD, the asymmetrical fusion may lead to serious bias [7]. In contrast symmetrical fusion methods either by using generative models or by maximizing correlation between EEG and fMRI, exploit two modalities for finding common neuronal substrates.

Daunizeau *et al.* [83] established a generative model for the EEG and BOLD signals by assuming common spatial profiles for both modalities. It is assumed that dipoles generating EEG and hemodynamic response function generating BOLD arise from a set of active areas that are characterized by temporal coherence. After parcellating the cortical surface into anatomically and functionally homogeneous areas, temporal dynamics of both signals are modeled. This hierarchical generative model is constituted within a variational Bayesian framework which finds the expectation of the parameter estimates from variational posterior probability distribution functions. Differing from the other methods, the prior probability of the spatial support parameter is assumed to be zero and the coupling between EEG and fMRI is learned from data. The method is also applied to a clinical epilepsy data and it is validated with intra cranial electro-physiological measurements (detailed results can be found in [83]).

Martinez-Montes *et al.* accomplished EEG/fMRI fusion by decomposing EEG and fMRI data as a sum of 'atoms' by using multiway partial least squares algorithm [20]. EEG atoms represent spatial, spectral and temporal signatures whereas fMRI atoms represent spatial and temporal signatures of the data. These atoms are extracted by guaranteeing maximal temporal covariance between temporal signatures of EEG and fMRI. They found similar results with resting state fusion literature: positive correlation in thalamus and negative correlation in occipital-parietal areas between EEG and BOLD alpha atoms.

4.4.2 Data Versus Model Driven Fusion

Data driven fusion establish functional connectivities between observables and seeks temporal or spatial coherence between measured responses of modalities [3]. ICA based fusion methods are in this category [4]. On the other hand, model driven fusion uses a biophysical modeling and attempts to find common neural events for two modalities. EEG and BOLD signals are predicted through the forward models and state space equations [3, 84].

4.5 Coupled Tensor Matrix Factorization for the Fusion of EEG and fMRI

As stated in Section 4.4 symmetrical data fusion approaches use complementary information of both modalities to unveil the common source of neural activity [83, 85]. We propose a new symmetrical data fusion framework based on the joint decomposition of EEG and fMRI on the common and discriminative spatial profile.

It has been shown that time evolving spectrum of the EEG captures oscillations generated by the localized and large scale activity of neuronal populations in spatial domain through volume conduction [58, 86]. Thus, time-frequency decompositions of EEG data matrix $\mathbf{V} \in \mathbb{R}^{I_E \times I_T}$ may show the level of synchronous neural activity. The spectrum of $\mathbf{V}$ over all channels can be constituted as a three-dimensional tensor defined over space, time and frequency dimensions by taking the Wavelet or Gabor transform: $\boldsymbol{\mathcal{S}} \in \mathbb{R}^{I_E \times I_T \times I_F}$, I_F being the number of frequency points. The PARAFAC model of the time-varying EEG spectrum decomposes $\boldsymbol{\mathcal{S}} \in \mathbb{R}^{I_E \times I_T \times I_F}$ into R atoms or components. In scalar notation, this model is

$$\boldsymbol{\mathcal{S}}(i_E, i_T, i_F) = \sum_{r=1}^{R} \mathbf{M_V}(i_E, r)\mathbf{T_V}(i_T, r)\mathbf{F_V}(i_F, r) + \boldsymbol{\mathcal{E}}_{\boldsymbol{\mathcal{S}}} \tag{4.5}$$

where $\mathbf{M_V} \in \mathbb{R}^{I_E \times R}$ is the spatial, $\mathbf{T_V} \in \mathbb{R}^{I_T \times R}$ is the temporal and $\mathbf{F_V} \in \mathbb{R}^{I_F \times R}$ is the spectral factor or signature.

An equivalent representation of Eq. 4.5 with Kruskal notation which will simplify the equation by making the indices implicit is

$$\boldsymbol{\mathcal{S}} = [\![\mathbf{M_V}, \mathbf{T_V}, \mathbf{F_V}]\!] + \boldsymbol{\mathcal{E}}_{\boldsymbol{\mathcal{S}}} \tag{4.6}$$

where the factors are normalized and the scale is absorbed by one of the factors. This model is first applied on an EEG dataset acquired from subjects during the resting state and during mental arithmetic. Spectral signatures showed an elevated level in alpha atom for the resting state and in theta atom for the mental arithmetic [19]. In the same study, source localization that is performed after extracting the spatial signature $\mathbf{M_V}$ showed activity in occipital areas for the resting state alpha atom and in frontal areas for the mental arithmetic task theta atom.

Instead of applying source localization on the identified spatial signatures as in [19], the decomposition can be directly performed on the source space. The PARAFAC model will be

$$\boldsymbol{\mathcal{S}} = [\![\mathbf{K}\mathbf{M}_{\mathbf{G}}, \mathbf{T}_{\mathbf{V}}, \mathbf{F}_{\mathbf{V}}]\!] + \boldsymbol{\mathcal{E}}_{\boldsymbol{\mathcal{S}}} \tag{4.7}$$

where $\mathbf{M}_{\mathbf{G}} \in \mathbb{R}^{I_{Cx} \times R}$ is the source spatial signatures.

For the fusion of the EEG and fMRI, we propose a joint decomposition approach based on the coupled matrix tensor factorization (CMTF). In this method, as described in the model in Eq. 4.7 EEG is considered as a three dimensional tensor composed of spatial, temporal and spectral signatures. Furthermore, by incorporating the EEG inverse model into the decomposition, EEG spatial signatures are found in the source space. fMRI is considered as a spatio-temporal two dimensional tensor. Spatial signatures are coupled during decomposition. Unlike conventional CMTF algorithms where a single dimension is considered to be fully coupled between two datasets, we prefer to project part of the datasets on a common and discriminative subspace [87]. This enables us to deal with the cases in which EEG and fMRI sources may differ [7, 73]. Coupled and uncoupled spatial profiles are obtained for each modality.

Assume that $\mathbf{M}_{eeg}$ is the source spatial factor of the EEG tensor $\boldsymbol{\mathcal{S}}$ and $\mathbf{M}_{fmri}$ is the spatial factor of the fMRI matrix $\mathbf{B}$, in the proposed framework these factors will be; $\mathbf{M}_{eeg} = (\mathbf{M}_{\mathbf{C}}|_{\{R_C|R_G\}}\mathbf{M}_{\mathbf{G}})$ and $\mathbf{M}_{fmri} = (\mathbf{M}_{\mathbf{C}}|_{\{R_C|R_B\}}\mathbf{M}_{\mathbf{B}})$ where subscript $\mathbf{C}$ is for the common part and subscript $\mathbf{G}$ ($\mathbf{B}$) is for the discriminant factor of EEG (fMRI). R_C is the number of common atoms, R_B is the number of discriminative atoms of fMRI, and R_G is the number of discriminative atoms of EEG. In this way different model orders can be assigned to the decomposition of $\boldsymbol{\mathcal{S}}$ and $\mathbf{B}$ as long as the number of common components are kept the same i.e. the column number of $\mathbf{M}_{\mathbf{C}}$.

In order to match the spatial resolution of EEG and fMRI, the inverse problem of EEG is included in the tensor decomposition. By using the transformation matrix - defined as the lead field matrix, $\mathbf{K}$ the spatial factor of EEG is transformed from sensor space to source space. We assume that EEG and fMRI share the spatial grid on the cortical surface defined by I_{Cx} voxels. The EEG spectrum is sampled at $I_{F\delta}$

frequency points and $I_{T\delta}$ same as the temporal points of the fMRI. Table 4.1 describes the dimensions of the variables.

Table 4.1
Symbols for EEG and fMRI

Symbol	**Definition**	**Dimension**
$\boldsymbol{\mathcal{S}}$	EEG tensor	$I_E \times I_{T\delta} \times I_{F\delta}$
$\mathbf{B}$	fMRI matrix	$I_{Cx} \times I_{T\delta}$
$\mathbf{K}$	Lead field matrix	$I_E \times I_{Cx}$
$\mathbf{L}$	Laplacian matrix	$I_{Cx} \times I_{Cx}$
$\mathbf{F_V}$	Spectral signature of EEG	$I_{F\delta} \times R_1$
$\mathbf{M_C}$	Common spatial signature of EEG and fMRI	$I_E \times R_C$
$\mathbf{M_G}$	Discriminant source spatial signature of EEG	$I_{Cx} \times R_G$
$\mathbf{M_B}$	Discriminant spatial signature of fMRI	$I_{Cx} \times R_B$
$\mathbf{T_V}$	Temporal signature of EEG	$I_{T\delta} \times R_1$
$\mathbf{T_B}$	Temporal signature of fMRI	$I_{T\delta} \times R_2$

General decomposition formulation for $\boldsymbol{\mathcal{S}}$ and $\mathbf{B}$ is

$$\min_{\substack{\mathbf{M_C},\mathbf{M_G},\mathbf{M_B},\\ \mathbf{F_V},\mathbf{T_V},\mathbf{T_B}}} \left\{ \frac{1}{2}\left\| \boldsymbol{\mathcal{S}} - [\![\mathbf{K}(\mathbf{M_C}|\mathbf{M_G}), \mathbf{T_V}, \mathbf{F_V}]\!] \right\|_2^2 + \gamma\frac{1}{2}\left\| \mathbf{B} - [\![(\mathbf{M_C}|\mathbf{M_B}), \mathbf{T_B}]\!] \right\|_2^2 \right\}. \tag{4.8}$$

Note that we dropped the concatenation indices on the spatial signatures to simplify the notation.

Furthermore, we impose non-negativity, orthogonality, smoothness and sparsity constraints on the spatial factors to ensure uniqueness. The corresponding M-P diagram is shown in Figure 4.3 and the problem is stated as:

$$\begin{aligned}
\min_{\substack{\mathbf{M_C},\mathbf{M_G},\mathbf{M_B},\\ \mathbf{F_V},\mathbf{T_V},\mathbf{T_B}}} & \left\{ \frac{1}{2}\left\| \boldsymbol{\mathcal{S}} - [\![\mathbf{K}(\mathbf{M_C}|\mathbf{M_G}),\mathbf{T_V},\mathbf{F_V}]\!] \right\|_2^2 + \gamma\frac{1}{2}\left\| \mathbf{B} - [\![(\mathbf{M_C}|\mathbf{M_B}),\mathbf{T_B}]\!] \right\|_2^2 \right. \\
& + \lambda_1\left\|\mathbf{M_C}\right\|_1 + \frac{1}{2}\lambda_2\left\|\mathbf{LM_C}\right\|^2 + \lambda_3\left\|\mathbf{M_G}\right\|_1 + \frac{1}{2}\lambda_4\left\|\mathbf{LM_G}\right\|^2 \\
& \left. + \lambda_5\left\|\mathbf{M_B}\right\|_1 + \frac{1}{2}\lambda_6\left\|\mathbf{LM_B}\right\|^2 \right\} \\
& \text{s.t.}(\mathbf{M_C}|\mathbf{M_G})^T(\mathbf{M_C}|\mathbf{M_G}) = \mathbf{I}, \quad (\mathbf{M_C}|\mathbf{M_B})^T(\mathbf{M_C}|\mathbf{M_B}) = \mathbf{I} \\
& \mathbf{M_C} \geq 0, \quad \mathbf{M_G} \geq 0, \quad \mathbf{M_B} \geq 0, \quad \mathbf{F_V} \geq 0.
\end{aligned} \tag{4.9}$$

The model in Eq. 4.9 can also be interpreted as the estimation of neuronal activity through two sources of information with multiple priors. The γ parameter takes into account the scale difference between EEG and fMRI.

Determination of the number of common and discriminant components is a very important task for the interpretation of the model. Since Eq. 4.9 is a modified PARAFAC decomposition, methods used for the selection of model order of PARAFAC can be used. The Core Consistency Diagnostic (Corcondia) is used to verify whether the core array of PARAFAC is a superidentity tensor. Superidentity property of the core tensor is a sign for the trilinearity of the tensor and the validity of PARAFAC model. The maximum number of components providing high Corcondia is taken as the model order [39]. Although Corcondia can be used to determine the model orders of EEG and fMRI decomposition model, selection of the number of common components, R_C needs further research. Separate decomposition of two datasets and observation of spatial factors may give a first estimate for R_C and the algorithm may be run for several R_C's.

4.5.1 Estimation of the Signatures of the CMTF

HALS algorithm combined with orthogonality [88] and other penalties is used for the estimation of the spatial signatures. Remaining factors are estimated in the

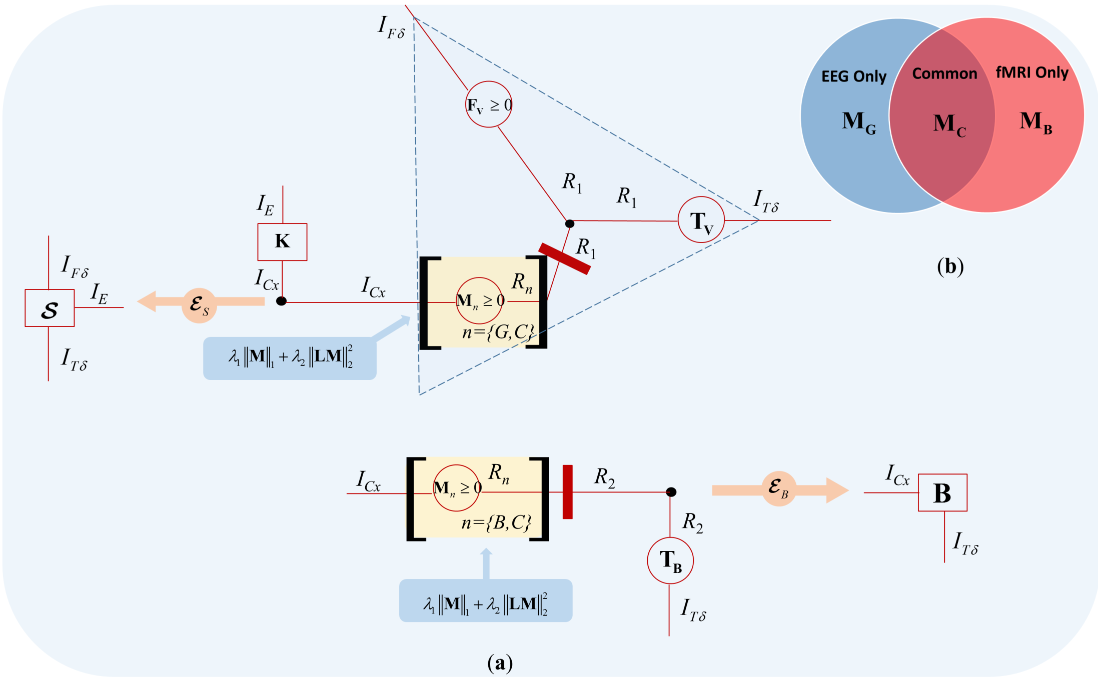


Figure 4.3 Coupled matrix tensor factorization. (a) M-P diagram for the coupled matrix tensor factorization for EEG/fMRI fusion. The EEG tensor $\boldsymbol{\mathcal{S}}$ and the fMRI matrix $\mathbf{B}$ are decomposed simultaneously on common and discriminant spatial subspaces to encompass different physiological sources. The spatial signature $\mathbf{M}$ involves common component $\mathbf{M_C}$ and two uncommon $\mathbf{M_G}, \mathbf{M_B}$ components. The fMRI spatial signature is $(\mathbf{M_C}|\mathbf{M_B})$ and the temporal signature is $\mathbf{T_B}$. For EEG, the spatial signature of the generators is $(\mathbf{M_C}|\mathbf{M_G})$, the temporal signature is $\mathbf{T_V}$, and the spectral signature is $\mathbf{F_V}$. By incorporating the lead field matrix $\mathbf{K}$, the model extends the decomposition of EEG to source space. M-P diagrams of EEG and fMRI are separated for a better visualization. (b) Explicit representation for the common and discriminative subspaces. Note that the common subspace is represented with $\mathbf{M_C}$.

alternating least squares algorithm.

4.5.1.1 Estimation of the Spatial Signatures. Common spatial signature, $\mathbf{M_C}$, individual spatial signature of EEG, $\mathbf{M_G}$ and individual spatial signature of fMRI, $\mathbf{M_B}$ are estimated by matricizing the Eq. 4.9 as follows:

$$\min_{\substack{\mathbf{M_C},\mathbf{M_G},\mathbf{M_B},\\ \mathbf{F_V},\mathbf{T_V},\mathbf{T_B}}} \left\{ \frac{1}{2}\left\| \boldsymbol{\mathcal{S}}_{(1)} - \mathbf{K}(\mathbf{M_C}|\mathbf{M_G})(\mathbf{F_V} \odot \mathbf{T_V})^T \right\|_2^2 + \gamma\frac{1}{2}\left\| \mathbf{B} - (\mathbf{M_C}|\mathbf{M_B})\mathbf{T_B}^T \right\|_2^2 \right.$$
$$+ \lambda_1\left\|\mathbf{M_C}\right\|_1 + \frac{1}{2}\lambda_2\left\|\mathbf{LM_C}\right\|^2 + \lambda_3\left\|\mathbf{M_G}\right\|_1 + \frac{1}{2}\lambda_4\left\|\mathbf{LM_G}\right\|^2$$
$$\left. + \lambda_5\left\|\mathbf{M_B}\right\|_1 + \frac{1}{2}\lambda_6\left\|\mathbf{LM_B}\right\|^2 \right\}$$
$$\text{s.t.}(\mathbf{M_C}|\mathbf{M_G})^T(\mathbf{M_C}|\mathbf{M_G}) = \mathbf{I}, \quad (\mathbf{M_C}|\mathbf{M_B})^T(\mathbf{M_C}|\mathbf{M_B}) = \mathbf{I}$$
$$\mathbf{M_C} \geq 0, \quad \mathbf{M_G} \geq 0, \quad \mathbf{M_B} \geq 0, \quad \mathbf{F_V} \geq 0. \tag{4.10}$$

HALS algorithm fits very well into the coupled factorization since the spatial signature matrices are divided into common and discriminative atoms in a columnwise manner. Call $\mathbf{P} = (\mathbf{F_V} \odot \mathbf{T_V})$ and similarly represent $\mathbf{P}$ in two subspaces as follows $\mathbf{P} = (\mathbf{P_C}|_{\{R_C,R_G\}}\mathbf{P_G})$. It is clear that

$$\mathbf{P_C} = (\mathbf{F_V}(:, 1 : R_C) \odot \mathbf{T_V}(:, 1 : R_C)) \tag{4.11}$$
$$\mathbf{P_G} = (\mathbf{F_V}(:, R_C + 1 : R_C + R_G) \odot \mathbf{T_V}(:, R_C + 1 : R_C + R_G)) \tag{4.12}$$

We do the same formulation for fMRI: $\mathbf{Q} = (\mathbf{Q_C}|_{\{R_C,R_B\}}\mathbf{Q_B})$ where

$$\mathbf{Q_C} = \mathbf{T_B}(:, 1 : R_C) \tag{4.13}$$
$$\mathbf{Q_B} = \mathbf{T_B}(:, R_C + 1 : R_C + R_B). \tag{4.14}$$

Orthogonality constraint on the nonnegative spatial signatures can be imposed column-wise [88]. The reason for this is that for a nonnegative matrix $\mathbf{X} \in \mathbb{R}^{I\times J}$, orthogonality condition $\mathbf{X}^T\mathbf{X} = \mathbf{I}$ can be replaced by $2J$ column-wise coefficients:

$$\mathbf{X}^T\mathbf{X} = \mathbf{I} \;\Rightarrow \begin{cases} \mathbf{X}(:, j)^T\mathbf{X}(:, j) = 1, & j = 1, \ldots, J \;\wedge \\ \sum\limits_{k\neq j}^{J} \mathbf{X}(:, k)^T\mathbf{X}(:, j) = 0, & j = 1, \ldots, J \end{cases} \tag{4.15}$$

where the symbol $\wedge$ is used for the *and* operator.

For our case, the orthogonality condition is expressed as follows:

$$(\mathbf{M_C}|\mathbf{M_G})^T(\mathbf{M_C}|\mathbf{M_G}) = \mathbf{I} \;\Rightarrow \begin{cases} \mathbf{M_C}(:,j)^T\mathbf{M_C}(:,j) = 1,\; j = 1,\ldots,R_C \;\wedge \\ \mathbf{M_G}(:,j)^T\mathbf{M_G}(:,j) = 1,\; j = 1,\ldots,R_G \;\wedge \\ \sum_{k\neq j}^{R_C} \mathbf{M_C}(:,k)^T\mathbf{M_C}(:,j) = 0,\; j = 1,\ldots,R_C \;\wedge \\ \sum_{k\neq j}^{R_G} \mathbf{M_G}(:,k)^T\mathbf{M_G}(:,j) = 0,\; j = 1,\ldots,R_G \;\wedge \\ \sum_{k=1}^{R_G} \mathbf{M_G}(:,k)^T\mathbf{M_C}(:,j) = 0, j = 1,\ldots,R_C \;\wedge \\ \sum_{k=1}^{R_C} \mathbf{M_C}(:,k)^T\mathbf{M_G}(:,j) = 0, j = 1,\ldots,R_G. \end{cases} \tag{4.16}$$

$$(\mathbf{M_C}|\mathbf{M_B})^T(\mathbf{M_C}|\mathbf{M_B}) = \mathbf{I} \;\Rightarrow \begin{cases} \mathbf{M_C}(:,j)^T\mathbf{M_C}(:,j) = 1,\; j = 1,\ldots,R_C \;\wedge \\ \mathbf{M_B}(:,j)^T\mathbf{M_B}(:,j) = 1,\; j = 1,\ldots,R_B \;\wedge \\ \sum_{k\neq j}^{R_C} \mathbf{M_C}(:,k)^T\mathbf{M_C}(:,j) = 0,\; j = 1,\ldots,R_C \;\wedge \\ \sum_{k\neq j}^{R_B} \mathbf{M_B}(:,k)^T\mathbf{M_B}(:,j) = 0,\; j = 1,\ldots,R_B \;\wedge \\ \sum_{k=1}^{R_B} \mathbf{M_B}(:,k)^T\mathbf{M_C}(:,j) = 0,\; j = 1,\ldots,R_C \;\wedge \\ \sum_{k=1}^{R_C} \mathbf{M_C}(:,k)^T\mathbf{M_B}(:,j) = 0,\; j = 1,\ldots,R_B. \end{cases} \tag{4.17}$$

Eq. 4.16 and Eq. 4.17 are unified for $\mathbf{M_C}$ as

$$\mathbf{W}^{(j)} = \sum_{k\neq j}^{R_C} \mathbf{M_C}(:,k) + \sum_{k=1}^{R_G} \mathbf{M_G}(:,k) + \sum_{k=1}^{R_B} \mathbf{M_B}(:,k) \tag{4.18}$$

And the orthogonality constraint is formulated as

$$(\mathbf{W}^{(j)})^T\mathbf{M_C}(:,j) = 0, \quad j = 1,\ldots,R_C. \tag{4.19}$$

First, we will present the estimation of the common spatial signature $\mathbf{M_C}$. Estimation of the others will follow. The objective function for the estimation of the jth column of $\mathbf{M_C}$ with the orthogonality constraint can be formulated in Lagrangian as

$$\mathcal{L}(\mathbf{M_C}(:,j),\beta_1(j)) = \left\{\frac{1}{2}\left\|\tilde{\boldsymbol{\mathcal{S}}}_{(1)} - \mathbf{K}\mathbf{M_C}(:,j)\,\mathbf{P_C}(:,j)^T\right\|_2^2 + \gamma\frac{1}{2}\left\|\tilde{\mathbf{B}} - \mathbf{M_C}(:,j)\mathbf{Q_C}(:,j)^T\right\|_2^2 \right.$$
$$\left. + \lambda_1\|\mathbf{M_C}(:,j)\|_1 + \frac{1}{2}\lambda_2\|\mathbf{L}\,\mathbf{M_C}(:,j)\|^2 + \beta_1(j)(\mathbf{W}^{(j)})^T\mathbf{M_C}(:,j)\right\} \tag{4.20a}$$

where

$$\tilde{\boldsymbol{\mathcal{S}}}_{(1)} = \boldsymbol{\mathcal{S}}_{(1)} - \mathbf{K}\sum_{k\neq j}^{R_C}\mathbf{M_C}(:,k)\mathbf{P_C}(:,k)^T - \mathbf{K}\,\mathbf{M_G}\mathbf{P_G}^T \tag{4.20b}$$

$$\tilde{\mathbf{B}} = \mathbf{B} - \sum_{k\neq j}^{R_C}\mathbf{M_C}(:,k)\mathbf{Q_C}(:,k)^T - \mathbf{M_B}\mathbf{Q_B}^T \tag{4.20c}$$

$\beta_1(j)$ is the weighting parameter for the orthogonality constraint on the jth column of $\mathbf{M_C}$.

Gradient of the objective function in Eq. 4.20a is found as

$$\frac{\partial\mathcal{L}}{\partial\mathbf{M_C}(:,j)} = \left\{-\mathbf{K}^T\tilde{\boldsymbol{\mathcal{S}}}_{(1)}\mathbf{P_C}(:,j) + \mathbf{K}^T\mathbf{K}\mathbf{M_C}(:,j)\;\mathbf{P_C}(:,j)^T\mathbf{P_C}(:,j) - \gamma\tilde{\mathbf{B}}\,\mathbf{Q_C}(:,j)\right.$$
$$\left. + \gamma\mathbf{M_C}(:,j)\mathbf{Q_C}(:,j)^T\mathbf{Q_C}(:,j) + \lambda_1\mathbf{M_C}(:,j) + \lambda_2\mathbf{L}^T\mathbf{L}\,\mathbf{M_C}(:,j) + \beta_1(j)\mathbf{W}^{(j)}\right\} \tag{4.21}$$

Since factors are normalized $\mathbf{P_C}(:,j)^T\mathbf{P_C}(:,j) = 1$ and $\mathbf{Q_C}(:,j)^T\mathbf{Q_C}(:,j) = 1$. Then $\mathbf{M_C}(:,j)$ is estimated by setting Eq. 4.21 to zero. The estimate is found as follows:

$$\hat{\mathbf{M}}_{\mathbf{C}}(:,j) = \Big[(\mathbf{K}^T\mathbf{K} + \lambda_2\mathbf{L}^T\mathbf{L} + \gamma\mathbf{I})^{-1}$$
$$(\mathbf{K}^T\tilde{\boldsymbol{\mathcal{S}}}_{(1)}\mathbf{P_C}(:,j) + \gamma\tilde{\mathbf{B}}\mathbf{Q_C}(:,j) - \lambda_1\mathbf{1} - \beta_1(j)\mathbf{W}^{(j)})\Big]_+ \tag{4.22}$$

where the nonnegativity condition is satisfied through the function

$$[x]_+ = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \tag{4.23}$$

We set the regularization parameter for orthogonality constraint as described in [88]. Multiplication of Eq. 4.21 by $\mathbf{W}^{(j)^T}(\mathbf{K}^T\mathbf{K} + \lambda_2\mathbf{L}^T\mathbf{L} + \gamma\mathbf{I})^{-1}$ from the left and noting $\mathbf{W}^{(j)T}\mathbf{M_C}(:,j) = 0$, the regularization parameter $\beta_1(j)$ is found as follows:

$$\beta_1(j) = \frac{\mathbf{W}^{(j)^T}(\mathbf{K}^T\mathbf{K} + \lambda_2\mathbf{L}^T\mathbf{L} + \gamma\mathbf{I})^{-1}(\mathbf{K}^T\tilde{\boldsymbol{\mathcal{S}}}_{(1)}\mathbf{P_C}(:,j) + \gamma\tilde{\mathbf{B}}\mathbf{Q_C}(:,j) - \lambda_1\mathbf{1})}{\mathbf{W}^{(j)^T}(\mathbf{K}^T\mathbf{K} + \lambda_2\mathbf{L}^T\mathbf{L} + \gamma\mathbf{I})^{-1}\mathbf{W}^{(j)}} \tag{4.24}$$

Note that in Eq. 4.22, the size of the matrix to be inverted is $I_{Cx} \times I_{Cx}$, which can be very large in real problems. So we use the inversion formula in Chapter 3 of [89] for the reformulation.

Call $(\frac{\lambda_2}{\gamma}\mathbf{L}^T\mathbf{L} + \mathbf{I}) = \mathbf{R}^T\mathbf{R}$ and $\mathbf{H} = (\tilde{\mathbf{B}}\mathbf{Q_C}(:,j) - \frac{\lambda_1}{\gamma}\mathbf{1} - \frac{\beta_1(j)}{\gamma}\mathbf{W}^{(j)})$.

$\mathbf{R}$ can be found from Cholesky decomposition. Eq. 4.22 will be:

$$\hat{\mathbf{M}}_\mathbf{C}(:,j) = (\mathbf{K}^T\mathbf{K} + \gamma\mathbf{R}^T\mathbf{R})^{-1}(\mathbf{K}^T\tilde{\boldsymbol{\mathcal{S}}}_{(1)}\mathbf{P_C}(:,j)^T + \gamma\mathbf{H}) \tag{4.25a}$$

$$= \mathbf{R}^{-1}(\tilde{\mathbf{K}}^T\tilde{\mathbf{K}} + \gamma\mathbf{I})^{-1}(\tilde{\mathbf{K}}^T\tilde{\boldsymbol{\mathcal{S}}}_{(1)}\mathbf{P_C}(:,j)^T + \gamma\mathbf{R}^{-T}\mathbf{H}) \tag{4.25b}$$

$$= \mathbf{R}^{-1}\left\{\tilde{\mathbf{K}}^T(\tilde{\mathbf{K}}\tilde{\mathbf{K}}^T + \gamma\mathbf{I})^{-1}(\tilde{\boldsymbol{\mathcal{S}}}_{(1)}\mathbf{P_C}(:,j)^T - \tilde{\mathbf{K}}\mathbf{R}^{-T}\mathbf{H}) + \mathbf{R}^{-T}\mathbf{H}\right\} \tag{4.25c}$$

where $\tilde{\mathbf{K}} = \mathbf{K}\mathbf{R}^{-1}$.

The same matrix manipulation can be used for the computation of the orthogonality parameter in Eq. 4.24:

$$\beta_1(j) = \frac{(\mathbf{W}^{(j)})^T\mathbf{R}^{-1}\left\{\tilde{\mathbf{K}}^T(\tilde{\mathbf{K}}\tilde{\mathbf{K}}^T + \gamma\mathbf{I})^{-1}(\tilde{\boldsymbol{\mathcal{S}}}_{(1)}\mathbf{P_C}(:,j)^T - \tilde{\mathbf{K}}\mathbf{R}^{-T}\mathbf{H}) + \mathbf{R}^{-T}\mathbf{H}\right\}}{\frac{1}{\gamma}(\mathbf{W}^{(j)})^T\mathbf{R}^{-T}(\mathbf{I} - \tilde{\mathbf{K}}^T(\tilde{\mathbf{K}}\tilde{\mathbf{K}}^T + \gamma\mathbf{I})^{-1}\tilde{\mathbf{K}})\mathbf{R}^{-1}\mathbf{W}^{(j)}} \tag{4.26}$$

where $\mathbf{H} = (\tilde{\mathbf{B}}\mathbf{Q_C}(:,j) - \frac{\lambda_1}{\gamma}\mathbf{1})$.

We skip the derivations of the discriminative signatures since formulation is very similar to the common one. We present the final results.

Discriminative signature of EEG is estimated as:

$$\hat{\mathbf{M}}_{\mathbf{G}}(:,j) = \left[\mathbf{L}^{-1}\left\{\tilde{\mathbf{K}}^T\left(\tilde{\mathbf{K}}\tilde{\mathbf{K}}^T + \mathbf{I}\right)^{-1}\left(\tilde{\boldsymbol{\mathcal{S}}}_{(1)}\mathbf{P}_{\mathbf{G}}(:,j)^T - \tilde{\mathbf{K}}\mathbf{L}^{-T}\mathbf{H}\right)\right\}\right]_+ \tag{4.27a}$$

where

$$\tilde{\boldsymbol{\mathcal{S}}}_{(1)} = \boldsymbol{\mathcal{S}}_{(1)} - \mathbf{K}\sum_{k\neq j}^{R_G}\mathbf{M}_{\mathbf{G}}(:,k)\mathbf{P}_{\mathbf{G}}(:,k)^T - \mathbf{K}\,\mathbf{M}_{\mathbf{C}}\mathbf{P}_{\mathbf{C}}{}^T \tag{4.27b}$$

$$\mathbf{W}^{(j)} = \sum_{k=1}^{R_C}\mathbf{M}_{\mathbf{C}}(:,k) + \sum_{k\neq j}^{R_G}\mathbf{M}_{\mathbf{V}}(:,k) \tag{4.27c}$$

$$\mathbf{H} = -\beta_2(j)\mathbf{W}^{(j)} - \lambda_3\mathbf{1} \tag{4.27d}$$

$$\tilde{\mathbf{K}} = \mathbf{K}\mathbf{L}^{-1} \tag{4.27e}$$

Regularization parameter for the orthogonality constraint of the discriminative signature of EEG is found as:

$$\beta_2(j) = \frac{(\mathbf{W}^{(j)})^T\mathbf{L}^{-1}\left\{\tilde{\mathbf{K}}^T\left(\tilde{\mathbf{K}}\tilde{\mathbf{K}}^T + \mathbf{I}\right)^{-1}\left(\tilde{\boldsymbol{\mathcal{S}}}_{(1)}\mathbf{P}_{\mathbf{G}}(:,j)^T - \tilde{\mathbf{K}}\mathbf{L}^{-T}\mathbf{H}\right)\right\}}{(\mathbf{W}^{(j)})^T\mathbf{L}^{-1}\left(\mathbf{I}_{I_{Cx}} - \tilde{\mathbf{K}}^T\left(\tilde{\mathbf{K}}\tilde{\mathbf{K}}^T + \mathbf{I}\right)^{-1}\tilde{\mathbf{K}}\right)\mathbf{L}^{-T}\mathbf{W}^{(j)}} \tag{4.28}$$

Discriminative signature of the fMRI is estimated as:

$$\hat{\mathbf{M}}_{\mathbf{B}}(:,j) \;= \left[\left(\mathbf{I} + \lambda_6\mathbf{L}^T\mathbf{L}\right)^{-1}\left(\gamma\tilde{\mathbf{B}}\mathbf{Q}_{\mathbf{B}}(:,j) - \lambda_5\mathbf{1} - \beta_3(j)\mathbf{W}^{(j)}\right)\right]_+ \tag{4.29a}$$

where

$$\tilde{\mathbf{B}} = \mathbf{B} - \sum_{k\neq j}^{R_B}\mathbf{M}_{\mathbf{B}}(:,k)\mathbf{Q}_{\mathbf{B}}(:,k)^T - \;\mathbf{M}_{\mathbf{C}}\mathbf{Q}_{\mathbf{C}}{}^T \tag{4.29b}$$

$$\mathbf{W}^{(j)} = \sum_{k=1}^{R_C}\mathbf{M}_{\mathbf{C}}(:,k) + \sum_{k\neq j}^{R_{DB}}\mathbf{M}_{\mathbf{B}}(:,k) \tag{4.29c}$$

Orthogonality regularization parameter is found as:

$$\beta_3(j) = \frac{(\mathbf{W}^{(j)})^T\left(\mathbf{I} + \lambda_6\mathbf{L}^T\mathbf{L}\right)^{-1}\left(\gamma\tilde{\mathbf{B}}\mathbf{Q}_{\mathbf{B}}(:,j) - \lambda_1\mathbf{1}\right)}{(\mathbf{W}^{(j)})^T\left(\mathbf{I} + \lambda_6\mathbf{L}^T\mathbf{L}\right)^{-1}\mathbf{W}^{(j)}} \tag{4.30}$$

4.5.1.2 Estimation of Other Signatures. Other signatures are estimated from ALS as follows:

$$\mathbf{T_V} = \boldsymbol{\mathcal{S}}_{(2)}(\mathbf{F_V} \odot \mathbf{K}(\mathbf{M_C}|\mathbf{M_G}))^{\dagger} \tag{4.31}$$

$$\mathbf{F_V} = \boldsymbol{\mathcal{S}}_{(3)}(\mathbf{T_V} \odot \mathbf{K}(\mathbf{M_C}|\mathbf{M_G}))^{\dagger} \tag{4.32}$$

$$\mathbf{T_B} = \mathbf{B}^T(\mathbf{M_C}|\mathbf{M_B})^{\dagger} \tag{4.33}$$

The CMTF algorithm is described in Figure 4.4.

4.5.2 Selection of the Model Parameters

Regularization parameters are selected by using Bayesian Information Criterion (BIC) [90]. For the BIC calculation, the following formula is used

$$\mathrm{BIC} = \log(\hat{\sigma}^2) + \mathrm{dof}\,\frac{\log(N)}{N} \tag{4.34}$$

where N is the number of observations, dof is the degrees of freedom and $\hat{\sigma}^2$ is the error variance estimated from the residual sum of squares (RSS) as: $\hat{\sigma}^2 = \mathrm{RSS}/N$.

BIC formulations for coupled and uncoupled components of the spatial factor are given as:

$$\begin{aligned}\mathrm{BIC}(\mathbf{M_C}) = \log\Bigg(&\Big\|\boldsymbol{\mathcal{S}} - [\![\mathbf{K}(\mathbf{M_C}|\mathbf{M_G}), \mathbf{T_V}, \mathbf{F_V}]\!]\Big\|_2^2\Big/(n_V + n_B)\\ &+\gamma\Big\|\mathbf{B} - [\![(\mathbf{M_C}|\mathbf{M_B}), \mathbf{T_B}]\!]\Big\|_2^2\Big/(n_V + n_B)\Bigg)\\ &+ \mathrm{dof}(\mathbf{M_C})\log(n_V + n_B)\big/(n_V + n_B)\end{aligned} \tag{4.35}$$

$$\mathrm{BIC}(\mathbf{M_V}) = \log\left(\Big\|\boldsymbol{\mathcal{S}} - [\![\mathbf{K}(\mathbf{M_C}|\mathbf{M_G}), \mathbf{T_V}, \mathbf{F_V}]\!]\Big\|_2^2/n_V\right) + \mathrm{dof}(\mathbf{M_V})\log(n_V)/n_V \tag{4.36}$$

in: $\boldsymbol{\mathcal{S}}, \mathbf{B}, \mathbf{K}, \mathbf{L}$

in: $R_C, R_V, R_B, \gamma, \{\lambda_j\}_{j=1}^{6}$

Initialize: $\mathbf{M_C}, \mathbf{M_G}, \mathbf{T_V}, \mathbf{F_V}, \mathbf{M_B}, \mathbf{T_B}$

repeat

  Set $\mathbf{R}, \mathbf{P_C}, \mathbf{P_G}, \mathbf{Q_C}, \mathbf{Q_B}$

  for $j = 1$ **to** $\max(R_C + R_V, R_C + R_B)$ **do**

    if $j \leq R_C$ **then**

      Estimate $\beta(j)$ from Eq. 4.24

      Estimate $\mathbf{M_C}(:, j)$ from Eq. 4.22

      $\mathbf{M_C}(:, j) \leftarrow \mathbf{M_C}(:, j)/||\mathbf{M_C}(:, j)||_2$

    end if

    if $(j > R_C)$ & $(j \leq R_C + R_V)$ **then**

      Estimate $\beta(j)$ from Eq. 4.28

      Estimate $\mathbf{M_G}(:, j)$ from Eq. 4.27a

      $\mathbf{M_G}(:, j) \leftarrow \mathbf{M_G}(:, j)/||\mathbf{M_G}(:, j)||_2$

    end if

    if $(j > R_C)$ & $(j \leq R_C + R_B)$ **then**

      Estimate $\beta(j)$ from Eq. 4.30

      Estimate $\mathbf{M_B}(:, j)$ from Eq. 4.29a

      $\mathbf{M_B}(:, j) \leftarrow \mathbf{M_B}(:, j)/||\mathbf{M_B}(:, j)||_2$

    end if

  end for

  Estimate $\mathbf{T_V}$ from Eq. 4.31

  $\mathbf{T_V}(:, j) \leftarrow \mathbf{T_V}(:, j)/||\mathbf{T_V}(:, j)||_2$

  Estimate $\mathbf{F_V}$ from Eq. 4.32

  Estimate $\mathbf{T_B}$ from Eq. 4.33

until $||\boldsymbol{\mathcal{S}} - \hat{\boldsymbol{\mathcal{S}}}||_2/||\boldsymbol{\mathcal{X}}||_2 < \epsilon$ & $||\mathbf{B} - \hat{\mathbf{B}}||_2/||\mathbf{B}||_2 < \epsilon$

out: $\mathbf{M_C}, \mathbf{M_G}, \mathbf{T_V}, \mathbf{F_V}, \mathbf{M_B}, \mathbf{T_B}$

Figure 4.4 CMTF HALS Algorithm

$$\mathrm{BIC}(\mathbf{M_B}) = \log\left(\left\|\mathbf{B} - [\![(\mathbf{M_C}|\mathbf{M_B}), \mathbf{T_B}]\!]\right\|_2^2 / n_B\right) + \mathrm{dof}(\mathbf{M_B})\log(n_B)/n_B \qquad (4.37)$$

where n_V and n_B are the number of elements in $\boldsymbol{\mathcal{S}}$ and $\mathbf{B}$,respectively and dof is the degrees of freedom computed as in [91]. Hyper-parameters λ_1 to λ_6 and γ in Eq. 4.9 are found as the minimum of the BIC multidimensional arrays given above.

4.5.3 Optimization of the PARAFAC

Since estimation of the factor matrices of the PARAFAC decomposition is a nonconvex optimization problem, an algorithm may reach different solutions with different starting points. In the ALS method, the objective function decreases at each step of the algorithm, but there is no guarantee that global minimum may be reached. We address the problem heuristically by means of the following approaches:

- running the algorithm with multiple random initial values;
- alternatively using, as a starting point, the eigenvectors of the unfolded tensor to be fitted [147];
- using a combination of the previous two schemes;
- estimating all runs and retaining the one with the best fit;
- in the case of models with nonnegative factor matrices, using a modified nonnegative double singular value decomposition proposed in [92] for the initial eigenanalysis;

However, these techniques do not guarantee convergence to the global minima; this is an area of increased current research [93, 94]. A definitive solution might be obtained by approximating the models here with alternative convex ones.

4.6 Real Data Analysis

We applied the proposed algorithm on a simultaneously recorded EEG - fMRI data [95]. In this experiment, flashing light stimuli in thirteen frequencies in the range of 6 Hz to 42 Hz were presented in a block design paradigm. For this analysis, data from 6 Hz stimulation session of one subject is used.

A Philips 1.5 T MR system was used to acquire T2* weighted images (TR/TE/FA = 2981ms/50ms/90, matrix size = $64 \times 64 \times 32$ axial slices, voxel size = $3.59 \times 3.59 \times 4$) with a gradient echo EPI sequence. EEG was recorded simultaneously by using an MR compatible EEG amplifier (BrainAmp MR+, Brain Products, Germany) with 30 channels EEG and 1 channel ECG. The EEG signal was filtered between 0.01 and 250 Hz and digitized with a sampling rate of 5 kHz. Gradient and ballistocardiographic artifacts in the EEG were removed by using average artifact subtraction technique implemented in the Brain Analyzer software [96]. All preprocessing fMRI analyses were performed with SPM5 software [97]. Motion correction, spatial smoothing with a Gaussian kernel of FWHM 8 mm and a high-pass temporal filtering were applied. Images were spatially normalized into a standard space (MNI152, 2mm) [98].

EEG was down-sampled to 250 Hz and further filtered with a high-pass filter with a cutoff frequency at 60 Hz. Afterwards, EEG was segmented in 2981 ms duration segments and Thomson multitaper method is used to calculate the power spectrum of each segment [99]. In Thomson multitaper method, DFT is applied on the time domain signal windowed by the orthogonal Slepian tapers or discrete prolate spheroidal sequences to decrease the spectral leakage between adjacent frequencies. We extracted the resting periods of the whole experiment and used for further analysis. At the end we had an EEG tensor with dimensions 31(channels) $\times$ 38(time points) $\times$ 58(frequency points). Lead field was computed using a realistic head model with three homogeneous isotropic conductor boundaries based on the MNI brain atlas [98].

After preprocessing of fMRI, voxels inside the cortical mesh of EEG source space were found. Grand mean scaling over the session for the voxels inside the mesh was performed and BOLD values were normalized to obtain a percentage change. At the end, we had an fMRI tensor with dimensions 5124(voxels) $\times$ 38(time points).

CMTF was initialized with the PARAFAC atoms found from separate decompositions of EEG and fMRI. Model orders for two datasets were selected as 2 based on explained variance and Corcondia measure. Since the power spectrum of EEG was used, nonnegative tensor decomposition was performed for EEG. We set the number of common factor to be 1 by examining the initial atoms coming from independent analysis of two datasets. BIC was employed for the selection of smoothness and sparsity regularization parameters and weighting parameter γ. Weighting parameter for orthogonality constraints were optimized inside the algorithm as suggested in [88]. We used the same regularization parameters for common and discriminant spatial signatures to decrease the computational load due to exhaustive search over 7 parameters. Since the number of parameters to be optimized was equal to 3, we searched over a three-dimensional parameter grid space and best parameters were selected as the minimum of the BIC volume.

Figure 4.5 shows the spatial, temporal and spectral signatures of the common atom. Since the two datasets were coupled only in spatial dimension, two temporal signatures for each modality were obtained. Common spatial signature, shows a clear activation in occipital areas. This activation is characterized by the alpha activity peaked at 10 Hz in the corresponding EEG spectral signature (Refer to Figure 4.5(e)) in line with the findings in the literature [20, 79, 100]. Pearson's correlation coefficient between spectral signatures of EEG and fMRI of this common factor is found to be -0.3346 with a p-value of 0.04 showing an inverse relation. However, since this value is not obtained from population statistics, further analysis should be pursued.

Discriminant fMRI atom is shown in Figure 4.6. Spatial signature shows activation mostly in inferior frontal areas of left and right hemispheres, inferior parietal and middle temporal areas of right hemisphere, precuneus and caudate. When the model order of the fMRI is increased these regions are distributed on separate atoms

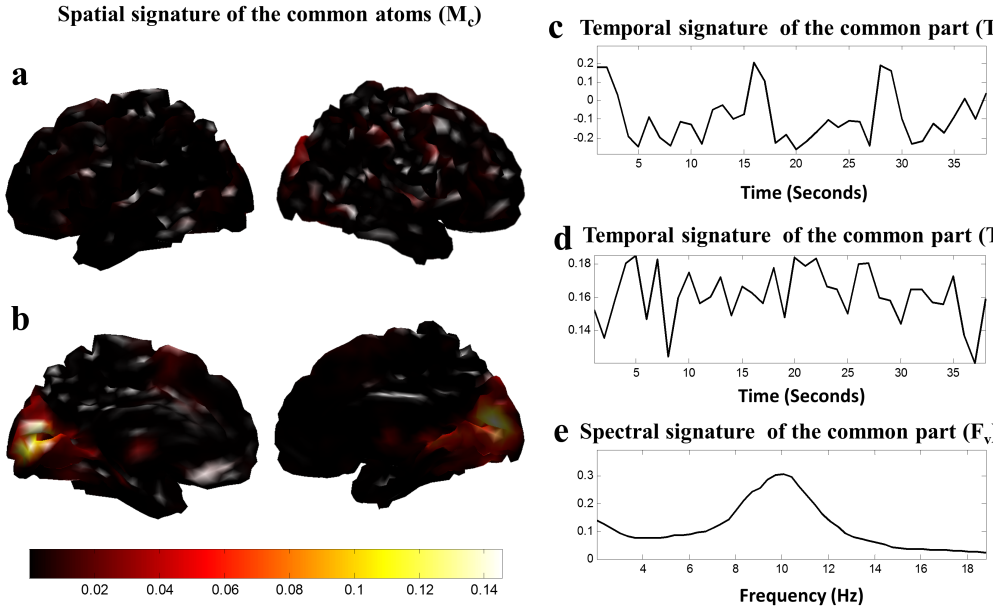


Figure 4.5 Common atom extracted from CMTF. (a) and (b) The spatial signatures $\mathbf{M_C}$ shows the distribution of activation of the spatial signature on the lateral and medial views of left and right hemispheres. Activity is localized in the occipital cortex. (c) The fMRI temporal signature of the common atom $\mathbf{T_B}(:,1)$. (d) The EEG temporal signature of the common atom $\mathbf{T_V}(:,1)$. (e) The EEG spectral signature of the common atom $\mathbf{F_V}(:,1)$. The 10-Hz peak in the EEG spectral signature indicates an alpha band activity.

(results not shown). It can be said that discriminant atom of fMRI shows a unification of resting state networks.

Discriminant EEG atom shows diffused activations in inferior and middle frontal areas, temporal areas of both hemispheres (Refer to Figure 4.7 (a,b)). Temporal signature has an intermittent activity as shown in Figure 4.7(c). Spectral signature is characterized by the well-known (1/frequency) pattern of the resting state EEG, the energy of the factor decreases towards higher frequency values (Refer to Figure 4.7 (d)). In [100], this pattern is called as ξ process and proposed to reflect the neural activity of diffuse and correlated generators.

In resting state data of one subject, our algorithm showed promising results. Common atom is found to be related alpha activity in EEG and spatial signature showed increased activation in occipital regions. This method still needs validation

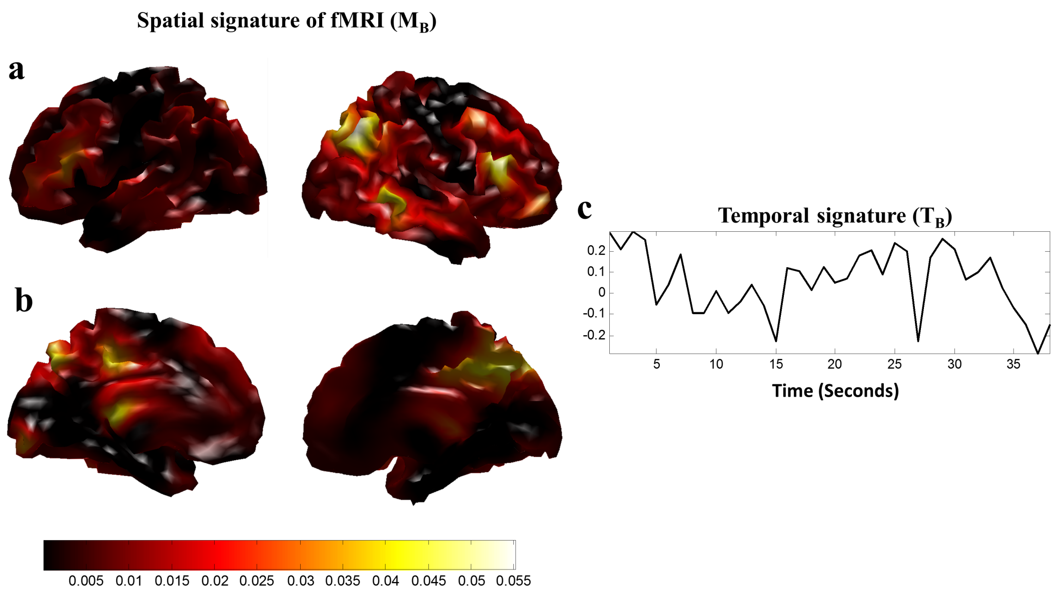


Figure 4.6 Discriminant fMRI atom. (a) and (b) The spatial signature of the discriminant fMRI atom $\mathbf{M_B}$ projected on the lateral and medial views of the left and right hemispheres. (c) The temporal course of the discriminant fMRI atom $\mathbf{T_B}(:,2)$. fMRI activity is diffused mostly in the frontal and temporal regions.

with more subjects and detailed statistical analysis on the relationship between temporal signatures of EEG and fMRI. In this dataset, we used the resting periods of a stimulation paradigm, which may be reason for absence of EEG rhythms in other frequency bands. Automatic selection of model orders instead of using heuristic methods still requires effort.

Incompatibility between temporal and spatial resolutions of EEG and fMRI complicates the simultaneous analysis of the two even in the case when there is discrepancy in the neural origins. CMTF of EEG and fMRI on spatial domain improves this difficulty and also presents the flexibility for searching discriminant sources of neural activation. We suggest that CMTF analysis can reveal more information about brain function by exploiting the complementary properties of two modalities.

CMTF is not limited to PARAFAC decomposition and can be modified by using other decomposition methods, e.g., the Tucker method, to account for the interactions between the signatures [25, 101, 102]. CMTF differs from the linked ICA [103] in the

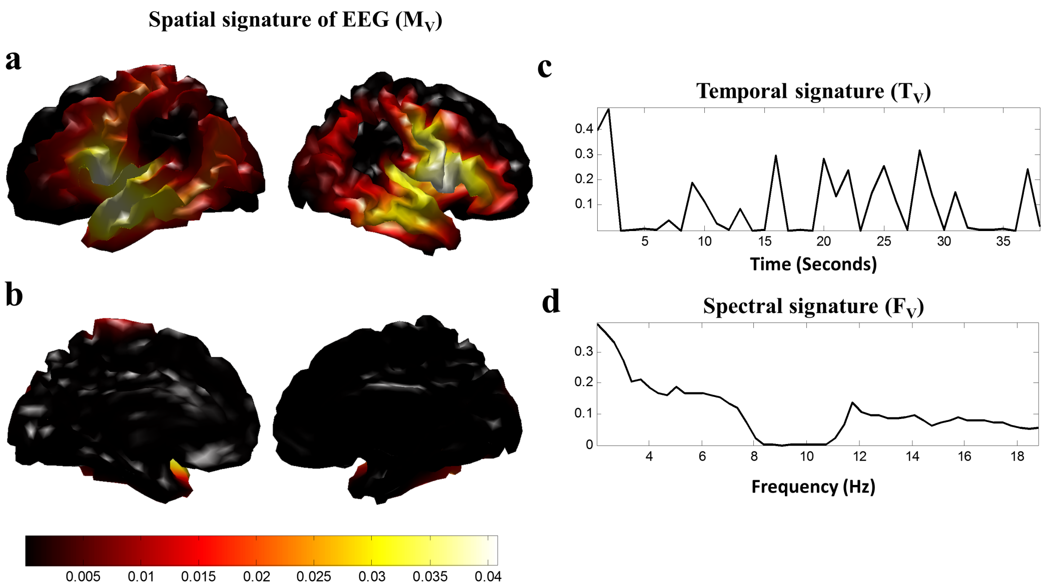


Figure 4.7 Discriminant EEG atom. (a) and (b) The spatial signature of the discriminant EEG atomprojected on the lateral and medial views of the left and right hemispheres $\mathbf{M_G}$. A diffused activity is revealed. (c) The temporal signature $\mathbf{T_V}(:,2)$. (d) The spectral signature $\mathbf{F_V}(:,2)$. Energy of the spectral signature decreases toward higher frequencies showing the ξ process. Spatial distribution is diffused over temporal and inferior frontal areas. All of the signatures are normalized to the unit norm.

sense that statistical independence of the spatial signatures is not required and common profiles can be divided into two subspaces. Recently, scalable and fast algorithms for CMTF have been developed and applied on the decomposition of fMRI and behavioral data [104].

5. TENSORIAL ANALYSIS OF BRAIN CONNECTIVITY

5.1 Brain Connectivity

Understanding the way how brain processes information requires knowledge about the functional organization of the brain. There have been two approaches regarding the organization. Functional specialization approach is built on the idea that a brain function can be localised in a cortical area and in the same sense a cortical area is specialised for some aspects of perceptual or motor processing [105].

Since fMRI gives a full coverage of the brain in a finer spatial resolution, it is possible to design experiments to find the functional localization of certain brain tasks. Statistical Parametric Mapping (SPM) is one of the best tools to find the brain areas that are engaged in the task through statistical tests [106]. However identification of brain regions that are activated in response to experimental manipulation does not answer the question that how these regions are related.

Functional integration on the other hand explores how large scale neural networks and brain regions interact with each other. Functional integration and specialization are complementary processes in the formation of the brain function. For example in a study of face processing one may find activations in the dorsolateral prefrontal cortex, superior parietal cortex and fusiform gyrus. However the coactivation of these brain regions do not explain how these regions are functionally connected. One explanation might be top-down connections from the prefrontal cortex to other regions or bottom-up connections from visual processing in temporal and parietal lobes to prefrontral cortex. An influence from a third region to these regions may also be possible [107]. In order to answer these questions we need to know how these regions are functionally *integrated*.

Functional integration studies give rise to two different connectivity measures in the brain: functional and effective connectivity. Functional connectivity explores the quantification of the operational interactions of multiple spatially distinct brain regions that are engaged simultaneously. Analysis techniques for functional connectivity use the correlation or covariance of activities derived from the BOLD data [108, 109]. The techniques that are used to reveal networks include ICA [110] and support vector machines (SVM) [111].

Effective connectivity is on the other hand defined as "the influence that one neural system exerts over another either directly or indirectly" [84]. The key concept in effective connectivity analysis is the causality which could be studied in the view of temporal precedence or physical influences [112]. Wiener-Akaike-Granger-Schweder (WAGS) influence usually termed as Granger Causality (GC) is based on the former understanding of the causality. If the event A (variable or time series) decreases the uncertainty in the predictability of the event B, then it is said that A G-causes B.

The second approach of causality is best exemplified with the Dynamical Causal Modeling (DCM). DCM combines neural model with an empirically validated biophysical forward model of the transformation from neuronal activity into a BOLD response [84]. In this type of modeling, brain is treated as a nonlinear deterministic system with known inputs and hidden states. By perturbating the system with a known input, hidden state variables are estimated. In DCM the rate of change in neuronal state is modeled via a neuronal model. Hemodynamic state equations utilizes Balloon - Windkessel model [113, 114] to constitute observed BOLD response from flow-inducing signal, cerebral blood flow, cerebral blood volume and deoxyhemoglobin content.

In this thesis we will propose a tensor based autoregressive model for the analysis of Granger Causality.

5.2 Granger Causality

In 1956 Wiener introduced a concept of causality based on the temporal precedence of causes to their effects [115]. However Granger was the first to show an implementation of this idea in econometrics by using linear autoregressive models. Also Akaike and Schweder had similar works at the time of Granger [116, 117].

Although the first practice of GC was on linear vector AR (VAR) processes, it can also be applied on infinite order VAR and vector AR moving average models [112]. Nonlinear GC has also been proposed and applied on the neuroimaging data [118].

The application of GC on the neuroimaging data, namely EEG and fMRI brings several challenges. For EEG, the volume conduction effect obscures the real dynamics of neural activity. Thus for the GC analysis on EEG, the inverse problem stated in Eq. 4.2 should be taken into account [119]. Vinck *et al.* describes the complexities of EEG GC and how to overcome them [120].

For fMRI, the challenge in GC application arises from several factors [121]. First is the variability of the hemodynamic response function(HRF) across brain regions and subjects. This can be a described in a simple example: Consider two brain regions X and Y such that in neuronal level X G-causes Y which can be represented as: $X \rightarrow Y$. However if the latency of the hemodynamic function of X is higher than that of region Y, GC analysis may find a spurious relation as $Y \rightarrow X$ in the fMRI level. In [122], simulated BOLD responses are generated by convolving the standard HRF with the LFPs acquired from the macaque cortex. GC is tested for various hemodynamic and neuronal delays, sampling time and the signal to noise ratio. It is shown that the detection of the underlying network with GC is robust to delays and the sensitivity of GC analysis increases with high sampling rate and high signal to ratio. A similar finding is reported in [123] stating when the variability is unlikely to be systematic, GC can reflect neural influences. It is suggested that higher sampling rates and application of statistical criteria improve the GC results. Also in [124] it was shown that GC is invariant to HRF variability on both theoretical and simulation models.

The second issue is the low pass behavior of the hemodynamic response that may lead to misidentification of the underlying fast neuronal influences. The third difficulty in fMRI is the low temporal resolution of the data which is typically in the range $1-3$ s. However this issue can be overcome by using ultrafast imaging. Extensive simulations by Rodrigues and Andrade suggest that the optimal sampling frequency is around 100 ms [125]. This is also argued experimentally by Lin *et al.*, who employed a fast fMRI sequence of 100 ms [126] sampling rate. We used this dataset for the GC analyses described in this chapter.

Now we will turn to the definition of the GC. Granger causality can be formulated in the linear regression context. Consider two random values $\mathbf{Y}_1 \in \mathbb{R}^T$ and $\mathbf{Y}_2 \in \mathbb{R}^T$ that are generated by stochastic processes and their values at time t are dependent on their own and the other's past values. The bivariate linear autoregressive model can be written as

$$\begin{aligned}
\mathbf{Y}_1(t) &= \sum_{q=1}^{p} \mathbf{A}_q(1,1)\mathbf{Y}_1(t-q) + \sum_{q=1}^{p} \mathbf{A}_q(1,2)\mathbf{Y}_2(t-q) + \varepsilon_1(t) \\
\mathbf{Y}_2(t) &= \sum_{q=1}^{p} \mathbf{A}_q(2,1)\mathbf{Y}_1(t-q) + \sum_{q=1}^{p} \mathbf{A}_q(2,2)\mathbf{Y}_2(t-q) + \varepsilon_2(t)
\end{aligned} \tag{5.1}$$

where $\varepsilon_1(t)$ and $\varepsilon_2(t)$ are uncorrelated Guassian white noise. p is the number of lags or model order, $\mathbf{A}_q \in \mathbb{R}^{2\times 2}$ is defined for each q time lag that quantifies the contribution of lagged observations to the predicted values. If the coefficients in $\mathbf{A}_q(1,2)$ is not equal to 0, then it can be said that $\mathbf{Y}_2$ Granger causes $\mathbf{Y}_1$. If two time series are independent then $\mathbf{A}_q(1,2) = 0$ and $\mathbf{A}_q(2,1) = 0$. Statistical significance of the Granger causality can be determined by using F-statistic of the null hypothesis $\mathbf{A}_q(1,2)$ [121]. The model order p can be determined by using Akaike Information Criterion (AIC) [127] or BIC.

The bivariate model in Eq. 5.1 can be extended to a multivariate autoregressive model (MAR) by using N time series. Let $\mathbf{Y} \in \mathbb{R}^{N\times T}$ is constructed by concatenating

$\mathbf{Y}_1$ to $\mathbf{Y}_N$ as follows

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1(p+1) & \dots & \mathbf{Y}_1(T+p) \\ \vdots & \ddots & \vdots \\ \mathbf{Y}_N(p+1) & \dots & \mathbf{Y}_N(T+p) \end{bmatrix}. \tag{5.2}$$

Similarly the lagged time series are concatenated to form $\mathbf{X} \in \mathbb{R}^{p \cdot N \times T}$

$$\mathbf{X} = \begin{bmatrix} \mathbf{Y}_1(p) & \dots & \mathbf{Y}_1(T+p-1) \\ \vdots & \ddots & \vdots \\ \mathbf{Y}_N(p) & \dots & \mathbf{Y}_N(T+p-1) \\ \vdots & \ddots & \vdots \\ \mathbf{Y}_1(1) & \dots & \mathbf{Y}_1(T) \\ \vdots & \ddots & \vdots \\ \mathbf{Y}_N(1) & \dots & \mathbf{Y}_N(T) \end{bmatrix}. \tag{5.3}$$

New coefficients matrix $\mathbf{W} \in \mathbb{R}^{N \times p \cdot N}$ is formulated explicitly as follows

$$\mathbf{W} = \begin{bmatrix} \mathbf{A}_1(1,1) & \dots & \mathbf{A}_1(1,N) & \dots & \mathbf{A}_p(1,1) & \dots & \mathbf{A}_p(1,N) \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{A}_1(N,1) & \dots & \mathbf{A}_1(N,N) & \dots & \mathbf{A}_p(1,1) & \dots & \mathbf{A}_p(N,N) \end{bmatrix}. \tag{5.4}$$

After defining the variables, we can write the MAR model as

$$\mathbf{Y} = \mathbf{W}\mathbf{X} + \mathbf{E} \tag{5.5}$$

Note that this model is a multivariate linear regression. The coefficient matrix $\mathbf{W}$ can be estimated by using maximum likelihood estimation

$$\hat{\mathbf{W}} = \arg\min_{\hat{\mathbf{W}}} ||\mathbf{Y} - \mathbf{W}\mathbf{X}||^2. \tag{5.6}$$

and the ordinary least squares estimate is found as

$$\hat{\mathbf{W}} = \mathbf{Y}\mathbf{X}^T(\mathbf{X}^T\mathbf{X})^{\dagger}. \tag{5.7}$$

However in case of high dimensional datasets e.g. brain data, the number of observations N is much higher than the number of time points T. In this case, the number of parameters to be estimated is $p \cdot N^2 + \frac{N^2+N}{2}$.

A solution to this problem may be reducing N. For fMRI datasets, a small set of region of interests (ROI) in the brain can be selected based on anatomical or functional constraints and the BOLD signal of the voxels inside the ROIs can be averaged. Then bivariate Granger causality analysis or parameter estimation techniques of multivariate regression can be performed.

Another approach could be application of regression methods based on selection of variables to extract the sparse networks. Penalization methods combined with statistical tests are shown to deal with high dimensional fMRI data in [128].

In this work, we will first reformulate GC analysis as a tensor problem and then propose two tensor based regression methods that will impose sparsity on the tensor and decomposition level. The first method uses the t-norm defined in Section 2.2.5 in the Levinson-Durbin context whereas second method incorporates tensor regression with the PARAFAC.

5.3 Granger Causality as a Tensor Regression

The fMRI data measured from I_{Cx} brain sites or voxels at $I_{T\delta}$ discrete time points can be casted as a matrix $\mathbf{B} \in \mathbb{R}^{I_{Cx} \times I_{T\delta}}$ as similar in Chapter 4. The MAR model of the brain data considers the brain as a complex network with voxels/ROIs to be the nodes. We aim to estimate the direction and weight of the connections where time series of voxels are modulated by their own past values and past values of other voxels. We will show that this problem is inherently a tensor problem due to its multidimensionality in space and time.

To extend the matrix model in Eq. 5.5 to tensor space, we will introduce two tensors constructed by concatenating the coefficients and lagged time series matrices along the time lag dimension, I_{lag}. By referring the definition of lagged time courses in

Eq. 5.3, the data tensor $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{I_{lag} \times I_{Cx} \times I_{T\delta}}$ is constructed by concatenating the lagged time series along I_{lag}:

$$\boldsymbol{\mathcal{B}} = [\mathbf{B}_{t-q}]_{q=1:I_{lag}}^{\{1|\ldots|1\}} \tag{5.8a}$$

where

$$\mathbf{B}_{t-q} = \begin{bmatrix} \mathbf{B}(1, I_{lag}+1-q) & \ldots & \mathbf{B}(1, I_T + I_{lag} - q) \\ \vdots & \ddots & \vdots \\ \mathbf{B}(I_{Cx}, I_{lag}+1-q) & \ldots & \mathbf{B}(I_{Cx}, I_T + I_{lag} - q) \end{bmatrix}. \tag{5.8b}$$

The autoregressive coefficients are also essentially a tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{I_{Cx} \times I_{Cx} \times I_{lag}}$ that is obtained by concatenating $\mathbf{A}_q$ matrices along the I_{lag} dimension.This operation is stated explicitly as follows

$$\boldsymbol{\mathcal{A}} = [\mathbf{A}_q]_{q=1:I_{lag}}^{\{1|\ldots|1\}} \tag{5.9a}$$

where

$$\mathbf{A}_q = \begin{bmatrix} \mathbf{A}_q(1,1) & \ldots & \mathbf{A}_q(1,N) \\ \vdots & \ddots & \vdots \\ \mathbf{A}_q(N,1) & \ldots & \mathbf{A}_q(N,N) \end{bmatrix}. \tag{5.9b}$$

Note that in both concatenation operations in Eq. 5.8a and Eq. 5.9a we make use of the property of tensors that adding singleton dimensions to a tensor will not change the tensor itself.

Now the tensor AR model can be stated by means of the contraction operator defined in Section 2.2.3

$$\mathbf{B} = \boldsymbol{\mathcal{A}} \bullet_{\{I_{Cx}, I_{lag}\}} \boldsymbol{\mathcal{B}} + \mathbf{E}. \tag{5.10}$$

An illustration of the MAR and tensor AR (TAR) is shown in Figure 5.1.

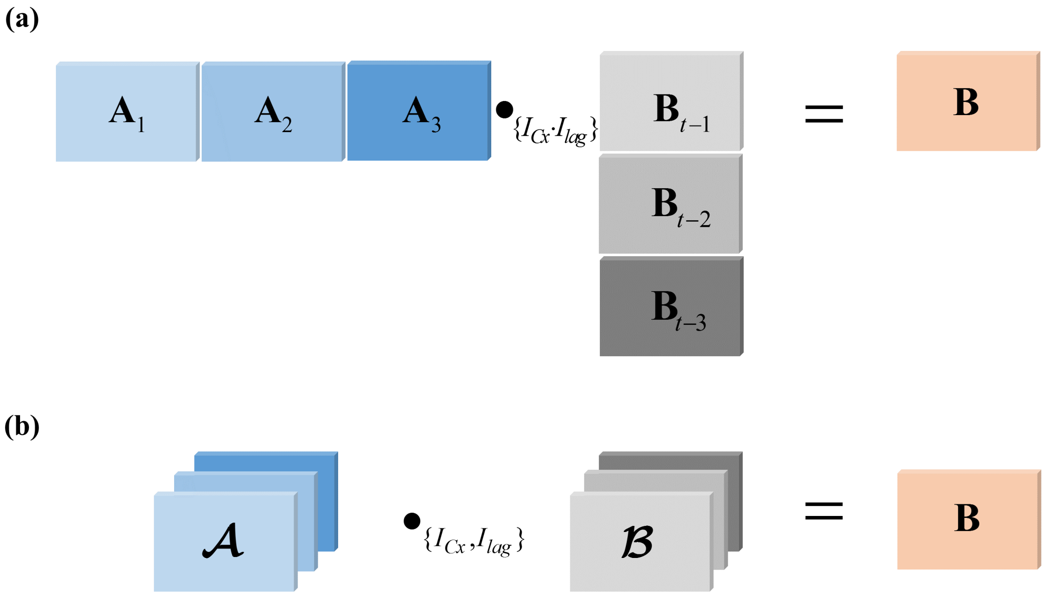


Figure 5.1 Illustration of matrix and tensor AR models for $I_{lag} = 3$. (a) In matrix AR contraction is performed on $I_{Cx} \cdot I_{lag}$ (b) TAR is formulated by contraction on the I_{Cx} and I_{lag} dimensions. Concatenation is made explicit by using the same color blocks.

The M-P diagram of this model is given in Figure 5.2. The tensor $\boldsymbol{\mathcal{A}}$ can be found by imposing priors as

$$\hat{\boldsymbol{\mathcal{A}}} = \underset{\boldsymbol{\mathcal{A}}}{\arg\min} \left\{ \left\| \mathbf{B} - \boldsymbol{\mathcal{A}} \bullet_{\{I_{Cx}, I_{lag}\}} \boldsymbol{\mathcal{B}} \right\|_2^2 + \pi(\boldsymbol{\mathcal{A}}) \right\} \tag{5.11}$$

We will propose two estimation methods one based on t-Product and the other on PARAFAC decomposition in the subsequent sections. The symbols used throughout the chapter are listed in Table 5.1.

5.4 Granger Causality with t-Products

We use t-Product operators defined in Section 2.2.5 for the estimation of AR coefficients. For this, we first define the sample covariance tensor $\boldsymbol{\mathcal{R}} \in \mathbb{R}^{I_{Cx} \times I_{Cx} \times (I_{lag}+1)}$

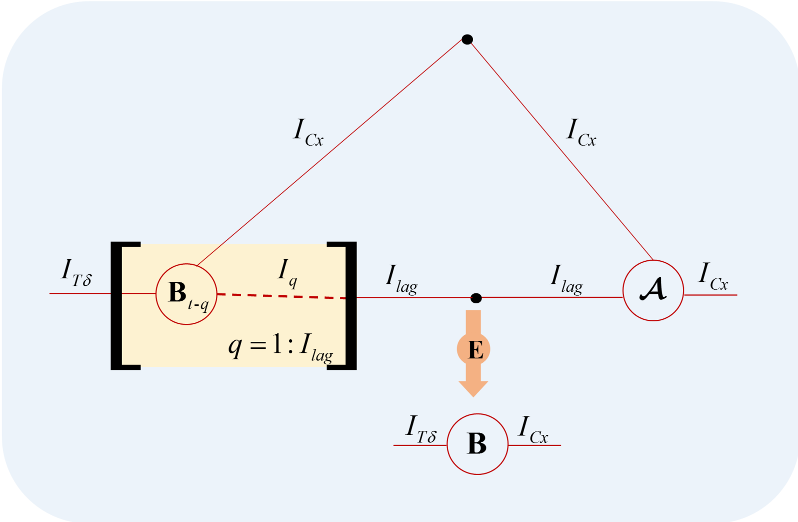


Figure 5.2 M-P Diagram of the TAR

as

$$\boldsymbol{\mathcal{R}}(i_{Cx}, i_{Cx}, q) = \begin{cases} \frac{1}{I_{T\delta}} \sum_{i_{T\delta}=1}^{I_{T\delta}} \mathbf{B}(i_{Cx}, i_{T\delta})\mathbf{B}(i_{Cx}, i_{T\delta}) & \text{if } q = 0 \\ \frac{1}{I_{T\delta}} \sum_{i_{T\delta}=1}^{I_{T\delta}} \boldsymbol{\mathcal{B}}(q, i_{Cx}, i_{T\delta})\mathbf{B}(i_{Cx}, i_{T\delta}) & \text{if } 0 < q \leq I_{lag} \end{cases} \tag{5.12}$$

where each horizontal slice of $\boldsymbol{\mathcal{R}}$ denoted by $\boldsymbol{\mathcal{R}}(:,:,q)$ is an $I_{Cx} \times I_{Cx}$ cross-covariance matrix.

The Levinson-Durbin equations [129] of the MAR model are presented in tensorial framework as

$$\begin{bmatrix} \boldsymbol{\mathcal{R}}(:,:,0) & \boldsymbol{\mathcal{R}}(:,:,1)^H & \cdots & \boldsymbol{\mathcal{R}}(:,:,I_{lag}-1)^H \\ \boldsymbol{\mathcal{R}}(:,:,1) & \boldsymbol{\mathcal{R}}(:,:,0) & \cdots & \boldsymbol{\mathcal{R}}(:,:,I_{lag}-2)^H \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\mathcal{R}}(:,:,I_{lag}-1) & \boldsymbol{\mathcal{R}}(:,:,I_{lag}-2)^H & \cdots & \boldsymbol{\mathcal{R}}(:,:,0) \end{bmatrix} \begin{bmatrix} \boldsymbol{\mathcal{A}}(:,:,1) \\ \boldsymbol{\mathcal{A}}(:,:,2) \\ \vdots \\ \boldsymbol{\mathcal{A}}(:,:,I_{lag}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mathcal{R}}(:,:,1) \\ \boldsymbol{\mathcal{R}}(:,:,2) \\ \vdots \\ \boldsymbol{\mathcal{R}}(:,:,I_{lag}) \end{bmatrix} \tag{5.13}$$

Note that the first term in Eq. 5.13 is a block Toeplitz matrix. By using the circulant embedding of the Toeplitz matrix, Eq. 5.13 may be written in t-operator

Table 5.1
Symbols for Granger causality formulation

Symbol	**Definition**	**Dimension**
$\mathbf{B}$	fMRI matrix	$I_{Cx} \times I_{T\delta}$
$\mathbf{B}_{t-q}$	fMRI time series lagged by q	$I_{Cx} \times I_{T\delta}$
$\boldsymbol{\mathcal{B}}$	Time lagged data tensor of GC	$I_{lag} \times I_{Cx} \times I_{T\delta}$
$\boldsymbol{\mathcal{A}}$	GC connectivity tensor	$I_{Cx} \times I_{Cx} \times I_{lag}$
$\boldsymbol{\mathcal{R}}$	Sample covariance tensor of GC	$I_{Cx} \times I_{Cx} \times I_{lag}$
$\mathbf{M_r}$	Spatial signature for receiver voxels	$I_{Cx} \times R$
$\mathbf{M_s}$	Spatial signature for sender voxels	$I_{Cx} \times R$
$\mathbf{T}$	Temporal signature for GC	$I_{lag} \times R$
$\mathbf{L}$	Laplacian matrix	$I_{Cx} \times I_{Cx}$

notation as follows:

$$\text{embed}(\boldsymbol{\mathcal{R}}_1)\text{MatVec}(\boldsymbol{\mathcal{A}}) = \text{MatVec}(\boldsymbol{\mathcal{R}}_2) \tag{5.14}$$

$$\boldsymbol{\mathcal{R}}_1 \star \boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{R}}_2 \tag{5.15}$$

where $\boldsymbol{\mathcal{R}}_1 = \boldsymbol{\mathcal{R}}(:,:,0 : I_{lag} - 1)$ and $\boldsymbol{\mathcal{R}}_2 = \boldsymbol{\mathcal{R}}(:,:,1 : I_{lag})$. The naïve solution of $\boldsymbol{\mathcal{A}}$ is

$$\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{R}}_1^{-1} \star \boldsymbol{\mathcal{R}}_2 \tag{5.16}$$

where inverse is actually a t-inverse. However, since $\boldsymbol{\mathcal{R}}_1$ is calculated from sample covariance, this type of solution is not numerically stable. One approach is to regularize the estimate of $\boldsymbol{\mathcal{R}}_1$ with t-norm. This is formulated as

$$\hat{\boldsymbol{\mathcal{R}}}_1 = \arg\min_{\Lambda} \left\{ \|\boldsymbol{\mathcal{R}}_1 - \Lambda\|_2^2 + \lambda \|\Lambda\|_{\circledast} \right\}. \tag{5.17}$$

The estimate of $\boldsymbol{\mathcal{R}}_1$ is found by using a proper shrinking function applied on the t-singular values of t-SVD($\boldsymbol{\mathcal{R}}_1$). We preferred to use the function defined in [130] since it was defined specifically for shrinking sample covariance estimator towards a stable target. The estimate is found as

$$\hat{\boldsymbol{\mathcal{R}}}_1 = \boldsymbol{\mathcal{U}} \star \rho(\boldsymbol{\mathcal{D}}) \star \boldsymbol{\mathcal{V}}^T \tag{5.18}$$

where ρ operates on the singular values of each face of $\boldsymbol{\mathcal{D}} \in \mathbb{R}^{I_{Cx} \times I_{Cx} \times I_{lag}}$ extracted as $\mathbf{d}(i) = \boldsymbol{\mathcal{D}}(i,i,k)$. The shrinking function is defined as

$$\rho(\mathbf{d}(i)^2) = \frac{-I_{Cx} + \sqrt{I_{Cx}^2 + 4\lambda\alpha(I_{Cx}\mathbf{d}(i)^2 + \lambda(1-\alpha))}}{2\lambda\alpha} \tag{5.19}$$

where λ and α are estimated from the data.

Finally, the estimate of $\boldsymbol{\mathcal{A}}$ is obtained by

$$\hat{\boldsymbol{\mathcal{A}}} = \boldsymbol{\mathcal{V}} \star (\rho(\boldsymbol{\mathcal{D}}))^{-1} \star \boldsymbol{\mathcal{U}}^T \star \boldsymbol{\mathcal{R}}_2 \tag{5.20}$$

Figure 5.3 describes the estimation steps explicitly.

5.5 Granger Causality with PARAFAC

As noted previously, one of the challenges in the MAR or TAR modeling of the fMRI data is the high number of nodes of the network which may lead to spurious connections. Structured sparsity of $\boldsymbol{\mathcal{A}}$ is a key concept that can be achieved by imposing a PARAFAC structure on the connectivity tensor. Recall that PARAFAC decomposition reveals the high-variance information as factor matrices or signatures.

We shall denote a node as a sender if it influences another set of nodes, and as a receiver if its activity is caused by other nodes. We propose a TAR model in which connectivity tensor is decomposed into spatial signatures denoting sender and receiver

in: $\mathbf{B}, I_{lag}, \lambda, \alpha$

Normalize $\mathbf{B}$: $\mathbf{B} = \frac{\mathbf{B}-\text{mean}(\mathbf{B})}{\text{var}(\mathbf{B})}$

Calculate the sample covariance tensor from Eq. 5.12

function embed **input:** $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I\times I\times K}$, **output:** $\boldsymbol{\mathcal{S}} \in \mathbb{R}^{I\times I\times 2K}$

$\boldsymbol{\mathcal{S}}(:,:,1:K) = \boldsymbol{\mathcal{X}}$

$\boldsymbol{\mathcal{S}}(:,:,K+1) = 1/2(\boldsymbol{\mathcal{X}}(:,:,K) + \boldsymbol{\mathcal{X}}(:,:,K)^H)$

$\boldsymbol{\mathcal{S}}(:,:,K+2:2K) = \boldsymbol{\mathcal{X}}(:,:,K:-1:2)^H$

end function

Define $\boldsymbol{\mathcal{R}}_1 = \boldsymbol{\mathcal{R}}(:,:,0:I_{lag}-1)$ and $\boldsymbol{\mathcal{R}}_2 = \boldsymbol{\mathcal{R}}(:,:,1:I_{lag})$.

Circulant embedding of $\boldsymbol{\mathcal{R}}_1$: $\boldsymbol{\mathcal{R}}_1 = \text{embed}(\boldsymbol{\mathcal{R}}_1)$

Circulant embedding of $\boldsymbol{\mathcal{R}}_2$: $\boldsymbol{\mathcal{R}}_2 = \text{embed}(\boldsymbol{\mathcal{R}}_2)$

$\tilde{\boldsymbol{\mathcal{R}}}_1 = \text{fft}(\boldsymbol{\mathcal{R}}_1, [\,], 3)$ {Fourier transform is applied on the third dimension}

$\tilde{\boldsymbol{\mathcal{R}}}_2 = \text{fft}(\boldsymbol{\mathcal{R}}_2, [\,], 3)$

for $k = 1$ **to** $2I_{lag}$ **do**

$[\mathbf{U}, \mathbf{D}, \mathbf{V}] = \text{svd}(\tilde{\boldsymbol{\mathcal{R}}}_1(:,:,k))$

for $i = 1$ **to** I_{Cx} **do**

$\mathbf{d}(i) = \mathbf{D}(i,i)$

$\mathbf{y}(i) = \rho(\mathbf{d}(i)^2)$ {Use the definition in Eq. 5.19}

$\mathbf{D}(i,i) = \sqrt{\mathbf{y}(i)}$

end for

$\tilde{\boldsymbol{\mathcal{A}}}(:,:,k) = \mathbf{V}\mathbf{D}^{-1}\mathbf{U}^T\tilde{\boldsymbol{\mathcal{R}}}_2(:,:,k)$

end for

$\boldsymbol{\mathcal{A}} = \text{ifft}(\tilde{\boldsymbol{\mathcal{A}}}(:,:,1:I_{lag}), [\,], 3)$ {Inverse Fourier transform is applied on the third dimension}

out: $\boldsymbol{\mathcal{A}}$

Figure 5.3 Granger Causality t-Product Algorithm nodes and a temporal signature for the lag. This model may be written as

$$\begin{aligned}
\min_{\mathbf{M_s},\mathbf{M_r},\mathbf{M_t}} & \left\{ \frac{1}{2}\left\|\mathbf{B} - \boldsymbol{\mathcal{A}} \bullet_{\{I_{Cx},I_{lag}\}} \boldsymbol{\mathcal{B}}\right\|_2^2 + \lambda_1\left\|\mathbf{M_s}\right\|_1 + \frac{1}{2}\lambda_2\left\|\mathbf{L}\mathbf{M_s}\right\|^2 \right. \\
& \left. + \lambda_3\left\|\mathbf{M_r}\right\|_1 + \frac{1}{2}\lambda_4\left\|\mathbf{L}\mathbf{M_r}\right\|_2^2 + \lambda_5\left\|\mathbf{T}\right\|_1 + \frac{1}{2}\lambda_6\left\|\mathbf{L}\mathbf{T}\right\|_2^2 \right\} \qquad (5.21) \\
& \text{s.t. } \boldsymbol{\mathcal{A}} = [\![\mathbf{M_s}, \mathbf{M_r}, \mathbf{T}]\!], \\
& \quad \mathbf{M_s} \geq 0, \quad \mathbf{M_r} \geq 0, \quad \mathbf{M_s}^T\mathbf{M_s} = \mathbf{I}, \quad \mathbf{M_r}^T\mathbf{M_r} = \mathbf{I}
\end{aligned}$$

where $\mathbf{M_s} \in \mathbb{R}^{I_{Cx}\times R}$ is the spatial signature for the sender nodes, $\mathbf{M_r} \in \mathbb{R}^{I_{Cx}\times R}$ is the spatial signature for the receiver nodes and $\mathbf{T} \in \mathbb{R}^{I_{lag}\times R}$ is the temporal signature for causal lags. R is the model order of the PARAFAC model.

In this model, the identifiability is enhanced by enforcing nonnegativity, orthogonality, smoothness, and sparseness for the spatial signatures and a smooth Lasso-type constraint for the lag signature. In other words, these constraints tend to estimate smooth patches of voxels on the cortex. Orthogonality and nonnegativity constraints guarantee that spatial factors can have only one nonnegative element in each row which can be interpreted as the cluster centroids [131, 132]. In this way, the connected spatial regions are confined to be nonoverlapping patches. This model is the generalization of clustering in which connectivity tensor is decomposed into sum of rank one triclusters [133].

The atomic decomposition of the 3-D connectivity tensor for the model of Eq. 5.21 favors a parsimonious model where the number of parameters to be estimated is $(2I_{Cx} + I_{lag})R$. The M-P diagram is shown in Figure 5.4. We will describe the estimation of the factors and the algorithm implemented.

5.5.1 Estimation of the Signatures

For the estimation of the signatures, we will introduce the PARAFAC constraint in Eq. 5.21 as a quadratic penalty term and the objective is written as

$$\min_{\boldsymbol{\mathcal{A}},\mathbf{M_s},\mathbf{M_r},\mathbf{T}} \left\{ \frac{1}{2} \left\| \mathbf{B} - \boldsymbol{\mathcal{A}} \bullet_{\{I_{Cx}, I_{lag}\}} \boldsymbol{\mathcal{B}} \right\|_2^2 + \frac{\lambda_p}{2} \left\| \boldsymbol{\mathcal{A}} - [\![\mathbf{M_s}, \mathbf{M_r}, \mathbf{T}]\!] \right\|_2^2 + P(\lambda, \mathbf{M_s}, \mathbf{M_r}, \mathbf{T}) \right\}$$
$$\text{s.t.}\ \ \mathbf{M_s} \geq 0, \quad \mathbf{M_r} \geq 0, \quad \mathbf{M_s}^T \mathbf{M_s} = \mathbf{I}, \quad \mathbf{M_r}^T \mathbf{M_r} = \mathbf{I} \tag{5.22}$$

where we put all the penalization terms of signatures into function $P(\lambda, \mathbf{M_s}, \mathbf{M_r}, \mathbf{T})$. λ_p is the weight of the penalization for $\boldsymbol{\mathcal{A}}$ to have a PARAFAC decomposition. We used Alternating Direction Method of Multipliers (ADMM) for the estimation of the signatures.

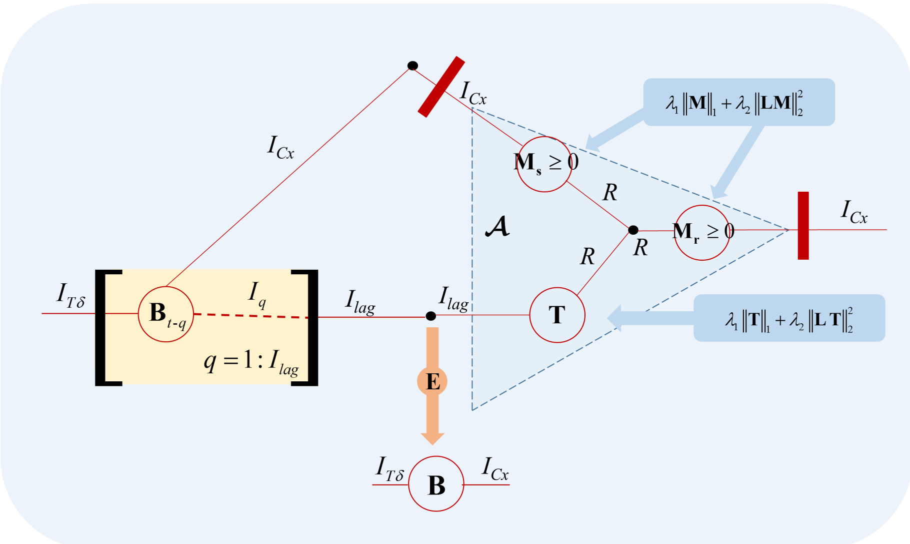


Figure 5.4 M-P Diagram of the GC-PARAFAC

5.5.1.1 Alternating Direction Method of Multipliers. The ADMM algorithm was first proposed by Gabay *et al.* and Glowinski *et al.* in the 1970s [134, 135]. However, it has attracted attention recently due to successful applications on large scale tensor completion problems with multiple nonsmooth penalization terms [136, 137].

The ADMM algorithm solves the problem

$$\begin{aligned} &\min_{\mathbf{x},\mathbf{y}} f(\mathbf{x}) + g(\mathbf{y}) \\ &\text{s.t. } \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y} = \mathbf{c} \end{aligned} \tag{5.23}$$

where the variables are $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$, the matrices are $\mathbf{A} \in \mathbb{R}^{p\times n}$, $\mathbf{B} \in \mathbb{R}^{p\times m}$ and $\mathbf{c} \in \mathbb{R}^p$. The functions f and g are assumed to be convex, though it has been shown that ADMM also works well with non-convex functions [138]. In this problem there are two sets of variables with separable objectives. By introducing a Lagrange multiplier $\mathbf{w} \in \mathbb{R}^p$ for the equality constraint, the augmented Lagrangian function could be written as

$$\boldsymbol{\mathcal{L}}(\mathbf{x},\mathbf{y},\mathbf{w}) = f(\mathbf{x}) + g(\mathbf{y}) + \mathbf{w}^T(\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y} - \mathbf{c}) + \frac{\nu}{2}\left\|\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y} - \mathbf{c}\right\|_2^2. \tag{5.24}$$

The ADMM algorithm is given as:

$$\begin{aligned} \mathbf{x}^{k+1} &= \arg\min_{\mathbf{x}} \boldsymbol{\mathcal{L}}(\mathbf{x},\mathbf{y}^k,\mathbf{w}^k) \\ \mathbf{y}^{k+1} &= \arg\min_{\mathbf{y}} \boldsymbol{\mathcal{L}}(\mathbf{x}^{k+1},\mathbf{y},\mathbf{w}^k) \\ \mathbf{w}^{k+1} &= \mathbf{w}^k + \nu(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{y}^{k+1} - \mathbf{c}) \end{aligned} \tag{5.25}$$

where k is the step number. The parameter ν is set inside the algorithm. Note that by applying sequential optimization, the parameters are decoupled.

5.5.1.2 ADMM Algorithm for GC-PARAFAC. It can be observed that the first two terms in Eq. 5.22 are coupled in the variable $\boldsymbol{\mathcal{A}}$. To decouple these terms, we introduce a new variable $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{I_{Cx}\times I_{Cx}\times I_{lag}}$ as follows:

$$\begin{aligned} \min_{\substack{\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{Z}},\\ \mathbf{M_s},\mathbf{M_r},\mathbf{M_t}}} &\left\{\frac{1}{2}\left\|\mathbf{B} - \boldsymbol{\mathcal{A}} \bullet_{\{I_{Cx},I_{lag}\}} \boldsymbol{\mathcal{B}}\right\|_2^2 + \frac{\lambda_p}{2}\left\|\boldsymbol{\mathcal{Z}} - [\![\mathbf{M_s},\mathbf{M_r},\mathbf{T}]\!]\right\|_2^2 + P(\lambda,\mathbf{M_s},\mathbf{M_r},\mathbf{T})\right\} \\ &\text{s.t. } \boldsymbol{\mathcal{A}} - \boldsymbol{\mathcal{Z}} = 0 \\ &\qquad \mathbf{M_s} \geq 0, \quad \mathbf{M_r} \geq 0, \quad \mathbf{M_s}^T\mathbf{M_s} = \mathbf{I}, \quad \mathbf{M_r}^T\mathbf{M_r} = \mathbf{I} \end{aligned} \tag{5.26}$$

Now, the parameters can be estimated with ADMM. The augmented Lagrangian function is given as

$$\begin{aligned}\boldsymbol{\mathcal{L}}(\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{Z}},\mathbf{M_s},\mathbf{M_r},\mathbf{T}) = \Bigg\{ & \frac{1}{2}\Big\|\mathbf{B}-\boldsymbol{\mathcal{A}}\bullet_{\{I_{Cx},I_{lag}\}}\boldsymbol{\mathcal{B}}\Big\|_2^2 + \frac{\lambda_p}{2}\big\|\boldsymbol{\mathcal{Z}}-[\![\mathbf{M_s},\mathbf{M_r},\mathbf{T}]\!]\big\|_2^2 \\ & + P(\lambda,\mathbf{M_s},\mathbf{M_r},\mathbf{T}) + \frac{\nu}{2}\Big\|\boldsymbol{\mathcal{A}}-\boldsymbol{\mathcal{Z}}\Big\|_2^2 \\ & + \boldsymbol{\mathcal{W}}\bullet_{\{I_{Cx},I_{Cx},I_{lag}\}}(\boldsymbol{\mathcal{A}}-\boldsymbol{\mathcal{Z}})\Bigg\}\end{aligned} \tag{5.27}$$

where $\boldsymbol{\mathcal{W}} \in \mathbb{R}^{I_{Cx}\times I_{Cx}\times I_{lag}}$ is the Lagrange multiplier. ADMM algorithm is given as follows:

$$\begin{aligned}\boldsymbol{\mathcal{A}}^{k+1} &= \arg\min_{\boldsymbol{\mathcal{A}}} \boldsymbol{\mathcal{L}}(\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{Z}}^k,\boldsymbol{\mathcal{W}}^k,\mathbf{M_s}^k,\mathbf{M_r}^k,\mathbf{T}^k) \\ \boldsymbol{\mathcal{Z}}^{k+1} &= \arg\min_{\boldsymbol{\mathcal{Z}}} \boldsymbol{\mathcal{L}}(\boldsymbol{\mathcal{A}}^{k+1},\boldsymbol{\mathcal{Z}},\boldsymbol{\mathcal{W}}^k,\mathbf{M_s}^k,\mathbf{M_r}^k,\mathbf{T}^k) \\ \mathbf{M_s}^{k+1} &= \arg\min_{\mathbf{M_s}} \boldsymbol{\mathcal{L}}(\boldsymbol{\mathcal{A}}^{k+1},\boldsymbol{\mathcal{Z}}^{k+1},\boldsymbol{\mathcal{W}}^k,\mathbf{M_s},\mathbf{M_r}^k,\mathbf{T}^k) \\ \mathbf{M_r}^{k+1} &= \arg\min_{\mathbf{M_r}} \boldsymbol{\mathcal{L}}(\boldsymbol{\mathcal{A}}^{k+1},\boldsymbol{\mathcal{Z}}^{k+1},\boldsymbol{\mathcal{W}}^k,\mathbf{M_s}^{k+1},\mathbf{M_r},\mathbf{T}^k) \\ \mathbf{T}^{k+1} &= \arg\min_{\mathbf{T}} \boldsymbol{\mathcal{L}}(\boldsymbol{\mathcal{A}}^{k+1},\boldsymbol{\mathcal{Z}}^{k+1},\boldsymbol{\mathcal{W}}^k,\mathbf{M_s}^{k+1},\mathbf{M_r}^{k+1},\mathbf{T}) \\ \boldsymbol{\mathcal{W}}^{k+1} &= \boldsymbol{\mathcal{W}}^k + \nu(\boldsymbol{\mathcal{A}}^{k+1}-\boldsymbol{\mathcal{Z}}^{k+1})\end{aligned} \tag{5.28}$$

We will give the ADMM updates in terms of variables as listed in Eq. 5.28.

(i) ADMM update step for $\boldsymbol{\mathcal{A}}$

The gradient of Eq. 5.27 with respect to $\boldsymbol{\mathcal{A}}$ is found as

$$\frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{\mathcal{A}}} = -(\mathbf{B}-\boldsymbol{\mathcal{A}}\bullet\boldsymbol{\mathcal{B}})\bullet\boldsymbol{\mathcal{B}}^H + \nu(\boldsymbol{\mathcal{A}}-\boldsymbol{\mathcal{Z}}) + \boldsymbol{\mathcal{W}}. \tag{5.29}$$

By setting the gradient to zero, $\boldsymbol{\mathcal{A}}$ is found as

$$\hat{\boldsymbol{\mathcal{A}}} = (\boldsymbol{\mathcal{B}}\bullet\boldsymbol{\mathcal{B}}^H + \nu\boldsymbol{\mathcal{I}})^{-1}\bullet(\mathrm{B}\bullet\boldsymbol{\mathcal{B}}^H + \nu\boldsymbol{\mathcal{Z}} - \boldsymbol{\mathcal{W}}). \tag{5.30}$$

(ii) ADMM update step for $\boldsymbol{\mathcal{Z}}$

ADMM update of $\boldsymbol{\mathcal{Z}}$ is found by finding the gradient of Eq. 5.27 with respect to $\boldsymbol{\mathcal{Z}}$

$$\frac{\partial \boldsymbol{\mathcal{L}}}{\partial \boldsymbol{\mathcal{Z}}} = \lambda_p(\boldsymbol{\mathcal{Z}} - \boldsymbol{\mathcal{Q}}) - \nu(\boldsymbol{\mathcal{A}} - \boldsymbol{\mathcal{Z}}) - \boldsymbol{\mathcal{W}}. \tag{5.31}$$

where $\boldsymbol{\mathcal{Q}} = [\![\mathbf{M_s}, \mathbf{M_r}, \mathbf{T}]\!]$. By setting the gradient to zero, $\boldsymbol{\mathcal{Z}}$ is found as

$$\hat{\boldsymbol{\mathcal{Z}}} = \frac{1}{\lambda_p + \nu}(\lambda_p \boldsymbol{\mathcal{Q}} + \nu \boldsymbol{\mathcal{A}} + \boldsymbol{\mathcal{W}}). \tag{5.32}$$

(iii) ADMM update step for $\mathbf{M_s}$

The signature matrices of the PARAFAC decomposition are estimated from the functional

$$\begin{aligned} \boldsymbol{\mathcal{L}}(\mathbf{M_s}, \mathbf{M_r}, \mathbf{M_t}) = \Bigg\{ & \frac{\lambda_p}{2} \|\boldsymbol{\mathcal{Z}} - [\![\mathbf{M_s}, \mathbf{M_r}, \mathbf{T}]\!]\|_2^2 + +\lambda_1 \|\mathbf{M_s}\|_1 + \frac{1}{2}\lambda_2 \|\mathbf{L}\mathbf{M_s}\|^2 \\ & + \lambda_3 \|\mathbf{M_r}\|_1 + \frac{1}{2}\lambda_4 \|\mathbf{L}\mathbf{M_r}\|_2^2 + \lambda_5 \|\mathbf{T}\|_1 + \frac{1}{2}\lambda_6 \|\mathbf{L}\mathbf{T}\|_2^2 \Bigg\} \\ & \text{s.t. } \mathbf{M_s} \geq 0, \quad \mathbf{M_r} \geq 0, \quad \mathbf{M_s}^T \mathbf{M_s} = \mathbf{I}, \quad \mathbf{M_r}^T \mathbf{M_r} = \mathbf{I} \end{aligned} \tag{5.33}$$

At this step any type of solver for tensor decompositions can be used. We prefer to use HALS algorithm as similar to Section 4.5.1, since the constraints are the same for spatial signatures: nonnegativity, orthogonality, sparsity and smoothness. For the estimation of temporal signature we will use proximal maps which will be explained afterwards.

For the estimation of spatial signature $\mathbf{M_s}$, we will use the matricized notation of PARAFAC. Call $\mathbf{G} = (\mathbf{T} \odot \mathbf{M_r})$, Eq. 5.33 may be written as

$$\begin{aligned} \boldsymbol{\mathcal{L}}(\mathbf{M_s}) = & \left\{ \frac{\lambda_p}{2} \|\boldsymbol{\mathcal{Z}}_{(1)} - \mathbf{M_s}\mathbf{G}^T\|_2^2 + \lambda_1 \|\mathbf{M_s}\|_1 + \frac{1}{2}\lambda_2 \|\mathbf{L}\mathbf{M_s}\|^2 \right\} \\ & \text{s.t. } \mathbf{M_s} \geq 0, \quad \mathbf{M_s}^T \mathbf{M_s} = \mathbf{I} \end{aligned} \tag{5.34}$$

The orthogonality constraints can be imposed column-wise in HALS as stated in [88]. Orthogonality constraint on $\mathbf{M_s}$ is expressed as

$$\mathbf{M_s}^T\mathbf{M_s} = \mathbf{I} \;\Rightarrow \begin{cases} \mathbf{M_s}(:,j)^T\mathbf{M_s}(:,j) = 1, \qquad j = 1,\ldots,R \;\wedge \\ \sum\limits_{k\neq j}^{R} \mathbf{M_s}(:,k)^T\mathbf{M_s}(:,j) = 0, \; j = 1,\ldots,R \end{cases} \tag{5.35}$$

Denote $\mathbf{W}^{(j)} = \sum\limits_{k\neq j}^{R} \mathbf{M_s}(:,k)$ then the orthogonality constraint is equal to

$$(\mathbf{W}^{(j)})^T\mathbf{M_s}(:,j) = 0 \text{ for } j = 1,\ldots,R. \tag{5.36}$$

By incorporating orthogonality constraint in Eq. 5.34 and fixing all columns except j, we get

$$\begin{aligned} \boldsymbol{\mathcal{L}}(\mathbf{M_s}(:,j)) = \Bigg\{ & \frac{\lambda_p}{2}\big\|\tilde{\boldsymbol{\mathcal{Z}}}_{(1)} - \mathbf{M_s}(:,j)\mathbf{G}(:,j)^T\|_2^2 + \lambda_1\big\|\mathbf{M_s}(:,j)\big\|_1 \\ & + \frac{1}{2}\lambda_2\big\|\mathbf{L}\mathbf{M_s}(:,j)\big\|^2 + \lambda_{orth}(\mathbf{W}^{(j)})^T\mathbf{M_s}(:,j)\Bigg\} \\ & \text{s.t. } \mathbf{M_s}(:,j) \geq 0 \end{aligned} \tag{5.37}$$

where $\tilde{\boldsymbol{\mathcal{Z}}}_{(1)} = \boldsymbol{\mathcal{Z}}_{(1)} - \sum\limits_{k\neq j}^{R} \mathbf{M_s}(:,k)\mathbf{G}(:,k)^T$ and λ_{orth} is the regularization parameter for orthogonality. Note that we did not apply the same subtraction on sparsity and smoothness inducing penalty functions since they are already operating on the columns.

The gradient of $\boldsymbol{\mathcal{L}}(\mathbf{M_s}(:,j))$ with respect to $\mathbf{M_s}(:,j)$ is found as

$$\begin{aligned} \frac{\partial\boldsymbol{\mathcal{L}}}{\mathbf{M_s}(:,j)} = \Big\{ & -\lambda_p\tilde{\boldsymbol{\mathcal{Z}}}_{(1)}\mathbf{G}(:,j) + \lambda_p\mathbf{M_s}(:,j)\mathbf{G}(:,j)^T\mathbf{G}(:,j) + \lambda_1\mathbf{1} \\ & + \lambda_2\mathbf{L}^T\mathbf{L}\mathbf{M_s}(:,j) + \lambda_{orth}\mathbf{W}^{(j)}\Big\} \end{aligned} \tag{5.38}$$

By setting the gradient to zero, the estimate is found as

$$\hat{\mathbf{M}}_\mathbf{s}(:,j) = \Big[\left(\lambda_p\mathbf{I} + \lambda_2\mathbf{L}^T\mathbf{L}\right)^{-1}\left(\lambda_p\tilde{\boldsymbol{\mathcal{Z}}}_{(1)}\mathbf{G}(:,j) - \lambda_1\mathbf{1} - \lambda_{orth}\mathbf{W}^{(j)}\right)\Big]_+ \tag{5.39}$$

Inside the algorithm the other factors are normalized to ensure $\mathbf{G}(:,j)^T\mathbf{G}(:,j) = 1$.

In order to set the λ_{orth}, we multiply Eq. 5.38 with $(\mathbf{W}^j)^T\left(\lambda_p\mathbf{I}+\lambda_2\mathbf{L}^T\mathbf{L}\right)^{-1}$ from the left and use the expression $(\mathbf{W}^{(j)})^T\mathbf{M_s}(:,j)=0$ to obtain

$$\lambda_{orth}=\frac{(\mathbf{W}^j)^T+\left(\lambda_p\mathbf{I}+\lambda_2\mathbf{L}^T\mathbf{L}\right)^{-1}\left(\lambda_p\tilde{\boldsymbol{\mathcal{Z}}}_{(1)}\mathbf{G}(:,j)-\lambda_1\mathbf{1}\right)}{(\mathbf{W}^j)^T+\left(\lambda_p\mathbf{I}+\lambda_2\mathbf{L}^T\mathbf{L}\right)^{-1}\mathbf{W}^j} \tag{5.40}$$

(iv) ADMM update step for $\mathbf{M_r}$

The estimation of the other spatial signature $\mathbf{M_r}$ follows the same procedure. So we skip the derivations and give the final result:

$$\hat{\mathbf{M}}_{\mathbf{r}}(:,j)=\left[\left(\lambda_p\mathbf{I}+\lambda_4\mathbf{L}^T\mathbf{L}\right)^{-1}\left(\lambda_p\tilde{\boldsymbol{\mathcal{Z}}}_{(2)}\mathbf{H}(:,j)-\lambda_3\mathbf{1}-\lambda_{orth}\mathbf{W}^{(j)}\right)\right]_+ \tag{5.41}$$

where $\mathbf{W}^{(j)}=\sum_{k\neq j}^{R}\mathbf{M_r}(:,k)$, $\mathbf{H}=(\mathbf{T}\odot\mathbf{M_s})$ and $\tilde{\boldsymbol{\mathcal{Z}}}_{(2)}=\boldsymbol{\mathcal{Z}}_{(2)}-\sum_{k\neq j}^{R}\mathbf{M_r}(:,k)\mathbf{H}(:,k)^T$

The orthogonality parameter is calculated as

$$\lambda_{orth}=\frac{(\mathbf{W}^j)^T+\left(\lambda_p\mathbf{I}+\lambda_4\mathbf{L}^T\mathbf{L}\right)^{-1}\left(\lambda_p\tilde{\boldsymbol{\mathcal{Z}}}_{(2)}\mathbf{H}(:,j)-\lambda_3\mathbf{1}\right)}{(\mathbf{W}^j)^T+\left(\lambda_p\mathbf{I}+\lambda_4\mathbf{L}^T\mathbf{L}\right)^{-1}\mathbf{W}^j} \tag{5.42}$$

(v) ADMM update step for $\mathbf{T}$

Since orthogonality and nonnegativity are not imposed on the temporal lag signature $\mathbf{T}$, the estimation procedure is different. The minimizer for $\mathbf{T}$ is written as

$$\arg\min_{\mathbf{T}}\boldsymbol{\mathcal{L}}(\mathbf{T})=\arg\min_{\mathbf{T}}\left\{\frac{\lambda_p}{2}\|\boldsymbol{\mathcal{Z}}_{(3)}-\mathbf{T}\mathbf{F}^T\|_2^2+\lambda_5\|\mathbf{T}\|_1+\frac{1}{2}\lambda_6\|\mathbf{L}\mathbf{T}\|^2\right\} \tag{5.43}$$

where a matricized notation for PARAFAC is used and $\mathbf{F}=(\mathbf{M_s}\odot\mathbf{M_r})$. Since $\mathbf{T}$ is multiplied by $\mathbf{F}$ from the right and by $\mathbf{L}$ from the left in Eq. 5.43, estimation of $\mathbf{T}$ is not easy. One method might be the vectorization of Eq. 5.43 and using Kronecker products which is not favorable since the scale of the problem will be high. Instead, we prefer to reformulate the problem by introducing a new variable to split the quadratic and penalization functions as follows:

$$\arg\min_{\mathbf{T}} \boldsymbol{\mathcal{L}}(\mathbf{T}) = \arg\min_{\mathbf{T}} \left\{ \frac{\lambda_p}{2} \|\boldsymbol{\mathcal{Z}}_{(3)} - \mathbf{T}\mathbf{F}^T\|_2^2 + \lambda_5 \|\mathbf{V}\|_1 + \frac{1}{2}\lambda_6 \|\mathbf{L}\mathbf{V}\|^2 \right\} \quad \textbf{s.t. } \mathbf{T} - \mathbf{V} = 0 \tag{5.44}$$

where $\mathbf{V} \in \mathbb{R}^{I_{lag} \times R}$ is an auxiliary matrix. Note that first term of Eq. 5.44 is decoupled from the rest, thus this problem can be solved with ADMM.

The augmented Lagrangian of the problem in Eq. 5.44 is

$$\boldsymbol{\mathcal{F}}(\mathbf{T}) = \left\{ \frac{\lambda_p}{2} \|\boldsymbol{\mathcal{Z}}_{(3)} - \mathbf{T}\mathbf{F}^T\|_2^2 + \lambda_5 \|\mathbf{V}\|_1 + \frac{1}{2}\lambda_6 \|\mathbf{L}\mathbf{V}\|^2 + \frac{\tau}{2} \|\mathbf{T} - \mathbf{V}\|_2^2 + \mathbf{Y} \bullet (\mathbf{T} - \mathbf{V}) \right\} \tag{5.45}$$

where $\mathbf{Y} \in \mathbb{R}^{I_{lag} \times R}$ is the Lagrange multiplier. We will use ADMM-in to emphasize that this ADMM algorithm is the update step of the outer ADMM.

ADMM-in update step for T

$\mathbf{T}$ is estimated from the functional

$$\arg\min_{\mathbf{T}} \boldsymbol{\mathcal{F}}(\mathbf{T}) = \arg\min_{\mathbf{T}} \left\{ \frac{\lambda_p}{2} \|\boldsymbol{\mathcal{Z}}_{(3)} - \mathbf{T}\mathbf{F}^T\|_2^2 + \frac{\tau}{2} \|\mathbf{T} - \mathbf{V}\|_2^2 + \mathbf{Y} \bullet (\mathbf{T} - \mathbf{V}) \right\} \tag{5.46}$$

The gradient of the functional with respect to $\mathbf{T}$ is

$$\frac{\partial \boldsymbol{\mathcal{F}}}{\partial \mathbf{T}} = \left\{ -\lambda_p \boldsymbol{\mathcal{Z}}_{(3)} \mathbf{F} + \lambda_p \mathbf{T}\mathbf{F}^T\mathbf{F} + \tau(\mathbf{T} - \mathbf{V}) + \mathbf{Y} \right\} \tag{5.47}$$

The estimate of $\mathbf{T}$ is found by setting the gradient to zero and solving for $\mathbf{T}$

$$\hat{\mathbf{T}} = \left(\lambda_p \tilde{\boldsymbol{\mathcal{Z}}}_{(3)} \mathbf{F} + \tau\mathbf{V} - \mathbf{Y}\right)\left(\lambda_p \mathbf{F}^T\mathbf{F} + \tau\mathbf{I}\right)^{-1} \tag{5.48}$$

Note that since the dimensions of $\mathbf{F}^T\mathbf{F}$ is $R \times R$, the inverse operation is not computationally demanding.

ADMM-in update step for V

$\mathbf{V}$ is estimated from the functional

$$\underset{\mathbf{V}}{\arg\min}\, \boldsymbol{\mathcal{F}}(\mathbf{V}) = \underset{\mathbf{V}}{\arg\min} \left\{ \frac{\tau}{2} \left\| \mathbf{T} - \mathbf{V} \right\|_2^2 + \mathbf{Y} \bullet (\mathbf{T} - \mathbf{V}) + \lambda_5 \left\| \mathbf{V} \right\|_1 + \frac{1}{2} \lambda_6 \left\| \mathbf{L}\mathbf{V} \right\|^2 \right\} \tag{5.49}$$

The problem in Eq. 5.49 is convex but not differentiable due to the L1 norm. We used proximal gradient ascent estimation of the $\mathbf{V}$ as described in [139]. Proximal map of a function $P(x)$ is defined as

$$\mathrm{prox}_p(y, \lambda) = \underset{x}{\arg\min} \left\{ \frac{1}{2} \| x - y \|_2^2 \right\} + \lambda P(x) \tag{5.50}$$

In our problem $P(x) = \|x\|_1$. The proximal map of the L1 norm is defined as follows

$$\mathrm{prox}_p(y, \lambda) = \mathrm{sign}(x) \left[|x| - \lambda \right]_+ \tag{5.51}$$

where sign is a function that is equal to 1 if $x > 0$ and -1 if $x < 0$.

By using the quadratic approximation of Eq. 5.49, we found $\mathbf{V}$ as

$$\hat{\mathbf{V}} = \mathrm{prox} \left(\mathbf{V} + \frac{1}{L} \left(\mathbf{Y} + \tau \mathbf{T} - (\tau \mathbf{I} + \lambda_6 \mathbf{L}^T \mathbf{L}) \mathbf{V} \right), \frac{\lambda_6}{L} \right) \tag{5.52}$$

where L is the Lipschitz constant found as [139]

$$L = \max \mathrm{eig} \left(\tau \mathbf{I} + \lambda_6 \mathbf{L}^T \mathbf{L} \right) \tag{5.53}$$

The algorithm of the GC-PARAFAC is summarized in Figure 5.5. It is important to note that GC analysis with PARAFAC requires a decomposition at each step of the ADMM algorithm; in this analysis, a "warm start" is used with initial values for the decomposition set as the factor estimates from the previous step.

in: $\mathbf{B}, \boldsymbol{\mathcal{B}}, \mathbf{L}, R, \{\lambda_j\}_{j=1}^{6}$

initialize: $\mathbf{M_r}, \mathbf{M_s}, \mathbf{T}$

for $k = 1$ **to** K **do**

  Estimate $\boldsymbol{\mathcal{A}}^{(k+1)}$ from Eq. 5.30

  Estimate $\boldsymbol{\mathcal{Z}}^{(k+1)}$ from Eq. 5.32

  for $j = 1$ **to** R **do**

    Estimate λ_{orth} of $\mathbf{M_s}$ from Eq. 5.40

    Estimate $\mathbf{M_s}^{(k+1)}$ from Eq. 5.39

    Estimate λ_{orth} of $\mathbf{M_r}$ from Eq. 5.42

    Estimate $\mathbf{M_r}^{(k+1)}$ from Eq. 5.41

  end for

  $L = \max \ \operatorname{eig}\left(\tau\mathbf{I} + \lambda_6 \mathbf{L}^T\mathbf{L}\right)$

  repeat

    Estimate $\mathbf{T}^{k+1}$ from Eq. 5.48

    Estimate $\mathbf{V}^{k+1}$ from Eq. 5.52

    $\mathbf{Y}^{(k+1)} = \mathbf{Y} + \tau(\mathbf{T}^{(k+1)} - \mathbf{V}^{(k+1)})$

  until (Primal Residual + Dual Residual) $< \epsilon$

  $\boldsymbol{\mathcal{Q}}^{k+1} = [\![\mathbf{M_s}, \mathbf{M_r}, \mathbf{T}]\!]$

  $\boldsymbol{\mathcal{W}}^{k+1} = \boldsymbol{\mathcal{W}} + \nu(\boldsymbol{\mathcal{A}}^{k+1} - \boldsymbol{\mathcal{Z}}^{k+1})$

end for

out: $\boldsymbol{\mathcal{A}}, \mathbf{M_s}, \mathbf{M_r}, \mathbf{T}$

Figure 5.5 Granger Causality - PARAFAC ADMM Algorithm

5.6 Real Data Analysis

We applied the GC t-Product and GC-PARAFAC analyses on the fMRI data of one subject reported in [126]. In this study, a reversing checkerboard at 8 Hz was presented that captures the left or right visual hemifield. Subjects were required to press the button with their corresponding hands to right or left visual hemifield stimuli. 500 ms duration stimulus was presented at a randomized onset with a uniform distribution of inter stimulus intervals in the $3 - 16$ s range.

fMRI images were acquired using inverse imaging (InI) providing a high temporal resolution (10 Hz) with whole brain coverage. After collecting the reference scan, functional scans were acquired by TR$=$ 100 ms, TE$=$ 30 ms and FA$=$ 30. The k-space InI reconstruction algorithm was used for the estimation of spatial encodings along the anterior-posterior axis [126].

InI time series of each subject was registered to their cortical surface and then to a spherical brain. After applying general linear model analysis, five functional ROIs were determined according to the BOLD activation: visual cortex (V), parietal cortex (PCC), pre-motor cortex (PreM), somatosensory cortex (S) and motor cortex (M). Each ROI consists of different number of activated voxels. The ROIs are shown in Figure 5.6. We did not average the time series of the voxels within each ROI for the analysis contrary to [126]. However, for reporting the results we took the sum of the ROIs which will be explained below. The mean value and the linear drift of the time series were removed. At the end we had a fMRI data matrix of size 299 time points and 1100 voxels.

For the t-Product analysis, we normalized the data as described in Figure 5.3. Then we calculated the covariance tensor and estimated the connectivity tensor. The results are reported by taking the sum of the connections between ROIs. The connectivity pattern is depicted in Figure 5.7(a). The method was able to deal with high-dimensional data having more than 1000 nodes and 20 lags with stable numerical results. It is also interesting to note that this estimate of connectivity seems to be

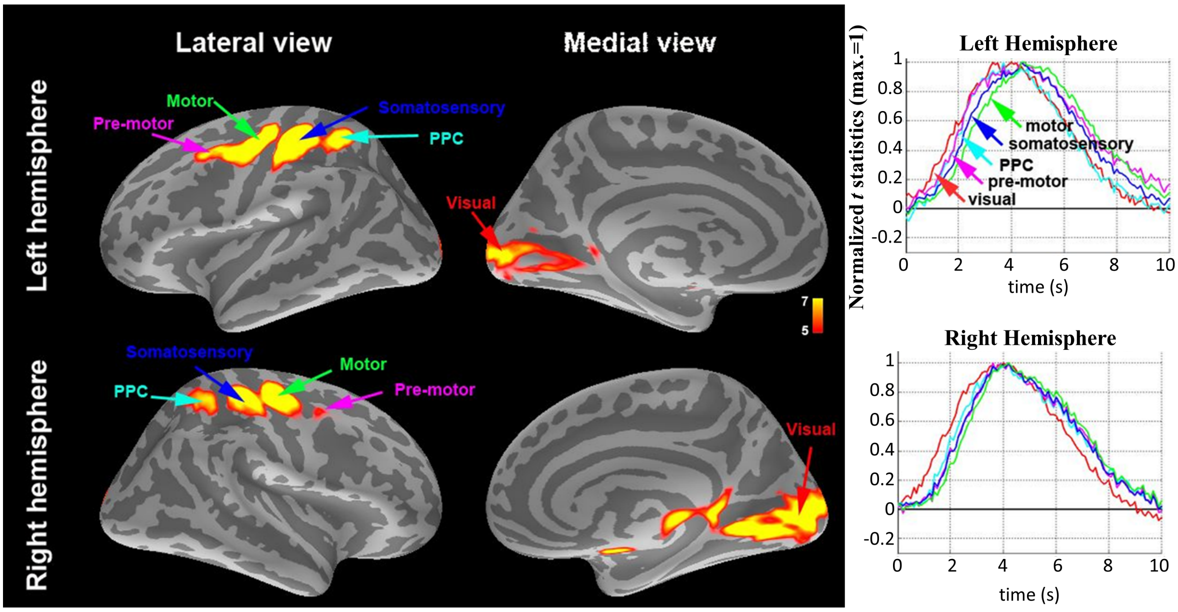


Figure 5.6 Locations of the functional ROIs are depicted on the cortical surfaces of the left and right hemisperes. The ROIs are selected according the t-values of the mean of the BOLD signal between 4 and 7s after the visual onset. Time courses of the BOLD responses and the estimated neuronal activity calculated from the deconvolution are shown on the right. Adapted from [126].

much more sensitive than the simple bivariate approach.

For the GC-PARAFAC analysis, a time period of 500 ms corresponding to five time lags was selected as the temporal factor. A graph Laplacian matrix is used as the smoother matrix $\mathbf{L}$ in Eq. 5.21. The model order of PARAFAC was set to 3. Each atom of receiver $\mathbf{M_r}$ and sender $\mathbf{M_s}$ signatures extracted from the PARAFAC are grouped according to ROIs and the sum of each ROI is taken. Figure 5.7(c) shows the existence of strong bottom-up and weak top-down connections between VC, PCC, M, and S. There is also lateral information flow from left to right visual areas. The temporal atoms, encoded in matrix $\mathbf{T}$, showed an ascending connectivity influence, peaking at the first lag (100 ms) and slowly decaying afterwards. Results of all analysis methods show that there is a predominance in causal directionality emerging from the V and PCC cortex to the rest of the brain areas.

In this section, we showed that the GC analysis for the brain data is inherently a tensor problem due to its multidimensionality in space and time. We proposed two

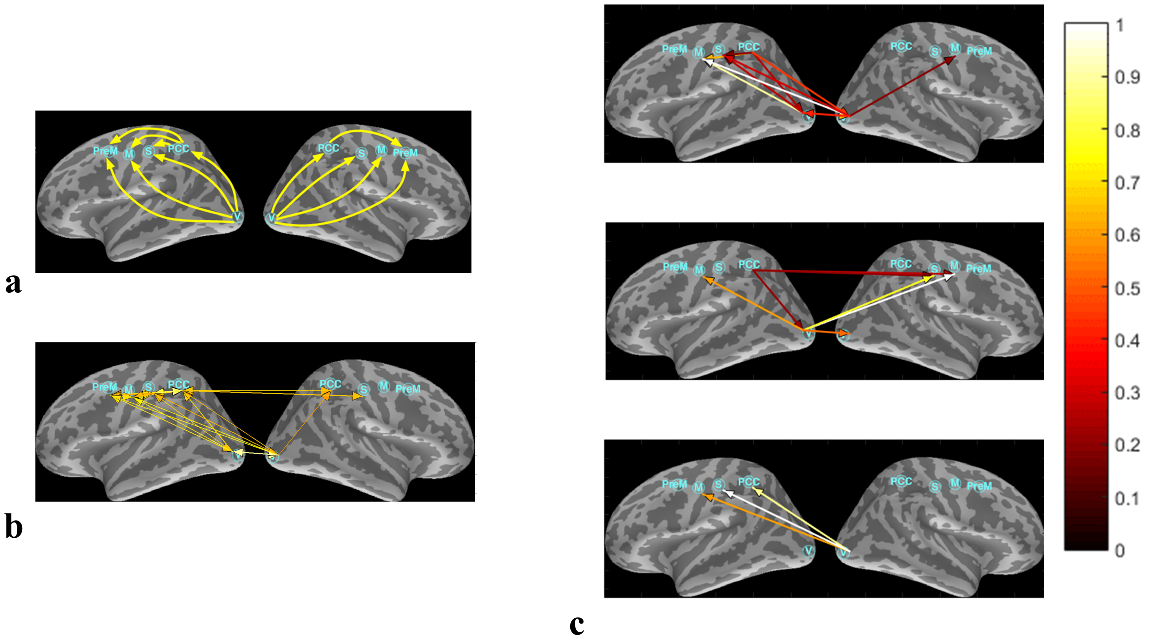


Figure 5.7 Granger causality in real data. The arrows denote directional dominant flows of Granger causality between the visual V, parietal PPC, premotor PreM, somatosensory S, and motor M cortical regions. (a) The original results were published in [126] and extracted from the Figure 2 of that reference. This is the dominant information flow calculated from the difference between two unidirectional Granger estimates among the ROIs. Only connections that have a p-value ≤ 0.05 are shown. (b) Results using the t-product (c) The resulting three spatial atoms of the connectivity tensor retrieved by the GC analysis with PARAFAC decomposition. Connectivity maps are generated for each atom by using directed arrows that are pointed from the cortical regions of senders which have a value greater than zero to positively active regions of the corresponding receiver signature. Magnitude of the connectivity is symbolized by the color bar on the right of the figure.

analysis methods, one based on the regularization of the covariance tensor and the other on the PARAFAC decomposition of the connectivity tensor. We suggest that the connection between GC and tensor analysis may lead to the use of other tensor based methods which are tailored for high-dimensional data.

6. DISCUSSION, CONCLUSION AND FUTURE WORK

6.1 Discussion

In Chapter 4, a new symmetric multimodal fusion technique is presented and applied on the simultaneous EEG/fMRI data. In this technique, time varying EEG spectrum is represented as a 3D tensor with spatial, temporal and spectral dimensions. fMRI data matrix is formed by using the time courses of the voxels on the cortical surface. Data coming from EEG and fMRI are decomposed simultaneously by coupling on the spatial dimension. This technique is summarized as follows:

- Fusion of EEG and fMRI is performed on the cortical surface which requires the source localization in EEG. This differs from other ICA/PARAFAC methods in which decomposition on the sensor space is followed by the localization in the source space [19, 140].
- Both common and uncommon spatial sources are identified for two modalities which enables to assign different model orders on the decomposition models of both modalities. This approach differs from joint-ICA and N-PLS based models in which model orders are kept the same.
- The identifiability of the fusion model is enhanced with the sparsity, smoothness, non-negativity and orthogonality constraints on the spatial factors. The interpretation of these constraints is to find a few, smooth and non-overlapping spatial sources. CMTF inherits the uniqueness of the PARAFAC however under some circumstances such as low signal-to-noise ratio and correlated factors, it may fail to identify the real underlying factors [141]. Constrained CMTF may avoid these problems. The orthogonality constraint may be physiologically demanding in some cases and can be relaxed [142, 143].

In this thesis, heuristic methods are used for the determination of the model orders of the decompositions of EEG and fMRI data tensors as well as the number of common components. This should be improved by using automatic selection of the model orders [144, 145]. The selection of the weight parameter γ in Eq. 4.8 is important since it determines the effects of the modalities on the identification of the common spatial factor $\mathbf{M_C}$. We used a rough estimate of this parameter through the BIC formulation of $\mathbf{M_C}$ by testing for various values. However, this parameter can be directly estimated in the alternating algorithm by using probabilistic approaches [146].

An improvement of the fusion model might be modification of the spatial definition of fMRI. We confined fMRI sources on the cortical surface. For the estimation of the discriminant components of fMRI whole-brain can be used. Furthermore, the EEG data is modeled as a spectral tensor which diminishes the phase information. As a future work, complex space-time decomposition can be performed.

In Chapter 5, a tensor AR model is proposed for modeling the causal brain networks. It is shown that the Granger causality analysis can be formulated within the tensor framework. Two methods are proposed:

1. Levinson-Durbin equations are reformulated by using t-products. Tensor nuclear norm is used for the estimation of the inverse of the covariance tensor calculated from the fMRI data matrix for several temporal lags.

2. Connectivity tensor is represented with the sender and receiver spatial signatures and a temporal lag signature by using PARAFAC. Both the connectivity tensor and the signatures are estimated by the ADMM algorithm.

Both of the algorithms could handle the high dimensionality of the data and find a sparse representation of the connectivity patterns. For the first method, we used tensor nuclear norm to find a stable estimate of the inverse of the covariance tensor. An alternative method might be the application of the tensor nuclear norm directly on the connectivity tensor which will give a lower rank estimate [13].

6.2 Conclusion

EEG and fMRI are mediated by different physiological processes from neural activation that lead to differences in their spatial and temporal resolutions. Due to the indirect nature of these signals, inverse problems for each modality should be solved to cover the interactions between modalities which are intrinsically ill-posed in their nature.

In this thesis, a general framework for the tensor analysis of multimodal data fusion and brain connectivity is presented. Detailed descriptions of the models and algorithms for the proposed approaches are presented. M-P diagrams that unify the graphical tensor notations with the directed acyclic graphs description of Bayesian statistical models are used for the illustration of the models.

All of the algorithms developed for this thesis are available at http://neurosignal.boun.edu.tr/software/tensor.

As the amount of neuroimaging data increase tremendously, methods dealing with this problem should be developed. Statistical methods based on tensors embraces the high dimensionality of the multimodal data.

6.3 Future Work

The application area of the proposed fusion method is not limited to the EEG and fMRI. Other types of data fusion such as DTI with fMRI, electrocortiography with EEG, or DTI with EEG may be used. As a future work, we will validate our model on datasets from multiple subjects and other modalities.

All of the proposed methods investigate the brain function on the macro-scale by using linear models. The generative models proposed are limited and do not include biophysical models. It is known that the forward model of fMRI is nonlinear. The methods can be improved by incorporating nonlinear models as suggested in [2].

For both the fusion and connectivity models, statistical methods for testing the significance of the results are needed. As a future work, statistical inference in higher dimensions should be developed.

APPENDIX A. LIST OF PUBLICATIONS RELATED TO THE THESIS

1. Tensor Analysis and Fusion of Multimodal Brain Images, **E. Karahan**, P.A. Rojas-Lopez, M.L. Bringas-Vega, P.A. Valdes-Hernandez, P.A. Valdes-Sosa, *Proceedings of the IEEE*, Vol. 103, pp: 1531-1559, 2015.

2. fMRI Responses of Alzheimer's Disease and Mild Cognitive Impairment Patients during Target Detection, M. Assem, M. H. Alpsan, **E. Karahan**, A. Bayram, B. Bilgiç, H. Gürvit, A. Ademoglu, T. Demiralp, *20th Annual Meeting of the Organization for Human Brain Mapping*, Hamburg, Germany, 2014.

3. Temporal Frequency Responses of Human Geniculate Nucleus and Primary Visual Cortex in fMRI, A. Bayram, **E. Karahan**, B. Bilgiç, A. Ademoğlu, T. Demiralp, *20th Annual Meeting of the Organization for Human Brain Mapping*, Hamburg, Germany, 2014.

4. EEG-fMRI fusion on the cortical surface using Coupled Tensor-Matrix Factorization: A simulation study, **E. Karahan**, A. D. Deniz Duru, P. A. Valdes-Sosa, A. Ademoğlu, *INCF Neuroinformatics Conference*, Stockholm, Sweden, 2013.

5. Simultaneous EEG/fMRI analysis of the resonance phenomena in steady-state visual evoked responses, A. Bayram, Z. Bayraktaroğlu, **E. Karahan**, B. Erdoğan, B. Bilgiç, M. Özker, I. Kaşıkçı, A.D. Duru, A. Ademoğlu, C. Öztürk, K. Arıkan, N. Tarhan, T. Demiralp, *Clinical EEG and Neuroscience*, Vol. 42, pp: 98-106, 2011.

6. Simultaneous EEG/fMRI Analysis of Steady-State Visual Evoked Responses, **E. Karahan**, M. Özker, A. Bayram, Z. Bayraktaroglu, B. Erdoğan, I. Kaşıkçı, C. Öztürk, A. Ademoğlu, T. Demiralp, *17th Annual Meeting of the Organization for Human Brain Mapping*, Quebec City, Canada, 2011.

7. A Group Study on BOLD Change to the Steady State Visual Stimuli with Bayesian Inference, M. Sevgi, **E. Karahan**, A. Bayram, A.D. Duru, C. Öztürk,

A. Ademoğlu, T. Demiralp, *17th Annual Meeting of the Organization for Human Brain Mapping*, Quebec City, Canada, 2011.

8. Frequency response characteristics of lateral geniculate nucleus and primary visual cortex, T. Demiralp, A. Bayram, **E. Karahan**, B. Bilgiç, N. Tarhan and A. Ademoğlu, *Front. Hum. Neurosci. Conf. Abs.: 11th International Conference on Cognitive Neuroscience*, 2011.

9. Visual Stimulation Frequency Dependent Changes in BOLD Transients, A. Bayram, A. Ademoğlu, **E. Karahan**, B. Bilgiç, AD Duru, N. Tarhan, T. Demiralp, *Front. Hum. Neurosci. Conf. Abs.: 11th International Conference on Cognitive Neuroscience*, 2011.

10. Hemodynamic correlates of brain electrical oscillations related with working memory, I. Kaşıkçı, A. Bayram, **E. Karahan**, B. Bilgiç, A. Ademoğlu, T. Demiralp, *Front. Hum. Neurosci. Conf. Abs.: 11th International Conference on Cognitive Neuroscience*, 2011.

11. Steady State Visual Evoked Potential Informed fMRI Analysis for Alpha, Beta and Gamma Bands, **E. Karahan**, M. Özker, B. Erdoğan, A. Bayram, Z. Bayraktaroglu, C. Öztürk, A. Ademoğlu, T. Demiralp, *16th Annual Meeting of the Organization for Human Brain Mapping*, Barcelona, Spain, 2010.

12. Comparison of Feature Selection Methods for Classification of Temporal fMRI Volumes Using SVM, A.E. Ercan, **E. Karahan**, O. Özyurt, C. Öztürk, *Proceedings of the 18th Annual Meeting of ISMRM*, Stockholm, Sweden, 2010.

13. Nonlinear Modeling of BOLD Signal with Particle Filters, **E. Karahan**, C. Öztürk, *26th Annual Scientific Meeting of the ESMRMB* Antalya, Turkey, 2009.

14. Multivariate Classification of fMRI Images, **E. Karahan**, C. Öztürk, *13th National Biomedical Engineering Conference*, İzmir, Turkey, 2009.

15. Mapping of the Visual Cortex: A FreesurferTM-based Approach, M. Yorulmaz, **E. Karahan**, A. Hamamcı, C. Öztürk, *13th National Biomedical Engineering Conference*, İzmir, Turkey, 2009.

16. Studying Familiarity of Different Stimulus Types, **E. Karahan**, Ö. Özmen-Okur, Ö. Alkan, T. Yıldırım, and C. Öztürk, *Proceedings of the 17th Annual Meeting of ISMRM*, Honolulu, USA, 2009.

REFERENCES

1. Sporns, O., G. Tononi, and R. Kötter, "The human connectome: A structural description of the human brain," *PLoS Computational Biology*, Vol. 1, p. e42, Sept. 2005.

2. Karahan, E., P. A. Rojas-Lopez, M. L. Bringas-Vega, P. A. Valdes-Hernandez, and P. A. Valdes-Sosa, "Tensor analysis and fusion of multimodal brain images," *Proceedings of the IEEE*, Vol. 103, pp. 1531–1559, Sept. 2015.

3. Valdés-Sosa, P. A., J. M. Sanchez-Bornot, R. C. Sotero, Y. Iturria-Medina, Y. Aleman-Gomez, J. Bosch-Bayard, F. Carbonell, and T. Ozaki, "Model driven EEG/fMRI fusion of brain oscillations," *Human Brain Mapping*, Vol. 30, pp. 2701–21, Sept. 2009.

4. Sui, J., T. Adalı, Q. Yu, J. Chen, and V. D. Calhoun, "A review of multivariate methods for multimodal fusion of brain imaging data," *Journal of Neuroscience Methods*, Vol. 204, pp. 68–81, Feb. 2012.

5. Uludağ, K., and A. Roebroeck, "General overview on the merits of multimodal neuroimaging data fusion," *NeuroImage*, Vol. 102, pp. 3–10, May 2014.

6. Bayram, A., *Investigation of oscillatory mechanisms and thalamo-cortical circuitry of the visual system by simultaneous EEG-fMRI*. PhD thesis, Bogazici University, Istanbul, Turkey, 2014.

7. Rosa, M. J., J. Daunizeau, and K. J. Friston, "EEG-fMRI integration: A critical review of biophysical modeling and data analysis approaches," *Journal of Integrative Neuroscience*, Vol. 09, no. 04, p. 453, 2010.

8. Kolda, T. G., and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, Vol. 51, no. 3, p. 455, 2009.

9. Kilmer, M. E., K. Braman, N. Hao, and R. C. Hoover, "Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging," *SIAM Journal on Matrix Analysis and Applications*, Vol. 34, no. 1, pp. 148–172, 2013.

10. Golub, G. H., and C. F. V. Loan, *Matrix Computations*, Baltimore, Maryland: Johns Hopkins University Press, 3rd ed., 1996.

11. Dietrich, C. R., and G. N. Newsam, "Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix," *SIAM Journal on Scientific Computing*, Vol. 18, no. 4, pp. 1088–1107, 1997.

12. Kilmer, M. E., and C. D. Martin, "Factorization strategies for third-order tensors," *Linear Algebra and Its Applications*, Vol. 435, pp. 641–658, Aug. 2011.

13. Semerci, O., N. Hao, M. E. Kilmer, and E. L. Miller, "Tensor-based formulation and nuclear norm regularization for multienergy computed tomography," *IEEE Transactions on Image Processing*, Vol. 23, no. 4, pp. 1678–1693, 2014.

14. Penrose, R., "Applications of negative dimensional tensors," in *Combinatorial Mathematics and its Applications*, pp. 221–244, London: Academic Press, 1971.

15. Bishop, C. M., *Pattern Recognition and Machine Learning*, New York: Springer-Verlag, 1st ed., 2006.

16. Cichocki, A., "Tensor networks for big data analytics and large-scale optimization problems," *arXiv preprint arXiv:1407.3124*, pp. 1–36, 2014.

17. Critch, A., and J. Morton, "Algebraic geometry of matrix product states," *Symmetry, Integrability and Geometry: Methods and Applications*, Vol. 10, Sept. 2014.

18. Yılmaz, Y. K., *Generalized Tensor Factorization*. PhD thesis, Bogazici University, Istanbul, Turkey, 2012.

19. Miwakeichi, F., E. Martínez-Montes, P. A. Valdés-Sosa, N. Nishiyama, H. Mizuhara, and Y. Yamaguchi, "Decomposing EEG data into space-time-frequency components using parallel factor analysis," *NeuroImage*, Vol. 22, pp. 1035–45, July 2004.

20. Martínez-Montes, E., P. A. Valdés-Sosa, F. Miwakeichi, R. I. Goldman, and M. S. Cohen, "Concurrent EEG/fMRI analysis by multiway partial least squares," *NeuroImage*, Vol. 22, pp. 1023–34, July 2004.

21. Mørup, M., L. K. Hansen, C. S. Herrmann, J. Parnas, and S. M. Arnfred, "Parallel Factor Analysis as an exploratory tool for wavelet transformed event-related EEG," *NeuroImage*, Vol. 29, pp. 938–47, Feb. 2006.

22. Acar, E., C. Aykut Bingol, H. Bingol, R. Bro, and B. Yener, "Multiway analysis of epilepsy tensors," *Bioinformatics*, Vol. 23, no. 13, pp. i10–18, 2007.

23. De Vos, M., L. De Lathauwer, B. Vanrumste, S. Van Huffel, and W. Van Paesschen, "Canonical decomposition of ictal scalp EEG and accurate source localisation: principles and simulation study," *Computational Intelligence and Neuroscience*, Vol. 2007, pp. 1–10, 2007. Article ID: 58253.

24. Deburchgraeve, W., P. Cherian, M. De Vos, R. Swarte, J. Blok, G. Visser, P. Govaert, and S. Van Huffel, "Neonatal seizure localization using PARAFAC decomposition," *Clinical Neurophysiology*, Vol. 120, pp. 1787–1796, Oct. 2009.

25. Cichocki, A., "Tensor decompositions: A new concept in brain data analysis?," *arXiv preprint arXiv:1305.0395*, May 2013.

26. Eliseyev, A., and T. Aksenova, "Recursive N-Way partial least squares for brain-computer interface," *PLoS ONE*, Vol. 8, no. 7, 2013.

27. Pester, B., C. Ligges, L. Leistritz, H. Witte, and K. Schiecke, "Advanced insights into functional brain connectivity by combining tensor decomposition and partial directed coherence," *PLoS ONE*, Vol. 10, no. 6, p. e0129293, 2015.

28. Harshman, R. A., "Foundations of the PARAFAC procedure: Models and conditions for an explanatory multimodal factor analysis," *UCLA Working Papers in Phonetics*, Vol. 16, pp. 1–84, 1970.

29. Tucker, L. R., "Some mathematical notes on three-mode factor analysis," *Psychometrika*, Vol. 31, pp. 279–311, Sept. 1966.

30. Bro, R., "Multiway calibration. multilinear PLS," *Journal of Chemometrics*, Vol. 10, pp. 47–61, Jan. 1996.

31. Grasedyck, L., D. Kressner, and C. Tobler, "A literature survey of low-rank tensor approximation techniques," *GAMM Mitteilungen*, Vol. 36, pp. 53–78, Aug. 2013.

32. Oseledets, I. V., "Tensor-train decomposition," *SIAM Journal on Scientific Computing*, Vol. 33, pp. 2295–2317, Jan. 2011.

33. Tobler, C., *Low-rank Tensor Methods for Linear Systems and Eigenvalue Problems.* PhD thesis, EPFL, Switzerland, 2012.

34. Carroll, J. D., and J.-J. Chang, "Analysis of individual differences in multidimensional scaling via an n-way generalization of Eckart-Young decomposition," *Psychometrika*, Vol. 35, pp. 283–319, Sept. 1970.

35. Möcks, J., "Topographic components model for event-related potentials and some biophysical considerations," *IEEE Transactions on Biomedical Engineering*, Vol. 35, pp. 482–484, June 1988.

36. Field, A. S., and D. Graupe, "Topographic component (Parallel Factor) analysis of multichannel evoked potentials: Practical issues in trilinear spatiotemporal decomposition," *Brain Topography*, Vol. 3, no. 4, pp. 407–423, 1991.

37. Cong, F., Q.-H. Lin, L.-D. Kuang, X.-F. Gong, P. Astikainen, and T. Ristaniemi, "Tensor decomposition of EEG signals: A brief review," *Journal of Neuroscience Methods*, Vol. 248, pp. 59 – 69, June 2015.

38. Kruskal, J. B., "Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear Algebra and its Applications*, Vol. 18, no. 2, pp. 95–138, 1977.

39. Bro, R., and H. A. L. Kiers, "A new efficient method for determining the number of components in PARAFAC models," *Journal of Chemometrics*, Vol. 17, pp. 274–286, June 2003.

40. Domanov, I., and L. De Lathauwer, "On the uniqueness of the canonical polyadic decomposition of third-order tensors—part ii: Uniqueness of the overall decomposition," *SIAM Journal on Matrix Analysis and Applications*, Vol. 34, pp. 876–903, July 2013.

41. Stegeman, A., and T. T. T. Lam, "Improved uniqueness conditions for canonical tensor decompositions with linearly dependent loadings," *SIAM Journal on Matrix Analysis and Applications*, Vol. 33, pp. 1250–1271, Nov 2012.

42. Cichocki, A., R. Zdunek, and S.-i. Amari, "Hierarchical ALS algorithms for nonnegative matrix and 3D tensor factorization," in *Independent Component Analysis, ICA07*, pp. 169–176, 2007.

43. Comon, P., X. Luciani, and A. L. F. de Almeida, "Tensor decompositions, alternating least squares and other tales," *Journal of Chemometrics*, Vol. 23, pp. 393–405, July 2009.

44. Acar, E., D. M. Dunlavy, and T. G. Kolda, "A scalable optimization approach for fitting canonical tensor decompositions," *Journal of Chemometrics*, Vol. 25, pp. 67–86, Feb. 2011.

45. De Lathauwer, L., B. De Moor, and J. Vandewalle, "Computation of the canonical decomposition by means of a simultaneous generalized Schur decomposition," *SIAM Journal on Matrix Analysis and Applications*, Vol. 26, pp. 295–327, Jan. 2004.

46. Yılmaz, Y. K., and A. T. Cemgil, "Algorithms for probabilistic latent tensor factorization," *Signal Processing*, Vol. 92, pp. 1853–1863, Aug. 2012.

47. Schmidt, M., and S. Mohamed, “Probabilistic non-negative tensor factorisation using Markov Chain Monte Carlo,” in *Proceedings of the 17th European Signal Processing Conference*, pp. 1918–1922, 2009.

48. De Lathauwer, L., B. De Moor, and J. Vandewalle, “A multilinear singular value decomposition,” *SIAM Journal on Matrix Analysis and Applications*, Vol. 21, pp. 1253–1278, Jan. 2000.

49. Bro, R., A. K. Smilde, and S. de Jong, “On the difference between low-rank and subspace approximation: improved model for multi-linear PLS regression,” *Chemometrics and Intelligent Laboratory Systems*, Vol. 58, pp. 3–13, Sept. 2001.

50. Zhao, Q., C. F. Caiafa, D. P. Mandic, Z. C. Chao, Y. Nagasaka, N. Fujii, L. Zhang, and A. Cichocki, “Higher order partial least squares (HOPLS): a generalized multilinear regression method.,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, pp. 1660–73, July 2013.

51. Hotelling, H., “Relations between two sets of variates,” *Biometrika*, Vol. 28, pp. 321–377, Dec. 1936.

52. Kim, T.-K., S.-F. Wong, and R. Cipolla, “Tensor canonical correlation analysis for action classification,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, June 2007.

53. Harshman, R. A., “Generalization of canonical correlation to N-way arrays,” in *Thirty-fourth Annual Meeting of the Statistical Society of Canada*, 2006.

54. Kim, T.-K., and R. Cipolla, “Canonical correlation analysis of video volume tensors for action categorization and detection.,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 31, pp. 1415–28, Aug. 2009.

55. Harshman, R. A., and M. E. Lundy, “PARAFAC: Parallel factor analysis,” *Computational Statistics & Data Analysis*, Vol. 18, pp. 39–72, Aug. 1994.

56. Smilde, A. K., J. A. Westerhuis, and R. Boque, “Multiway multiblock component and covariates regression models,” *Journal of Chemometrics*, Vol. 14, pp. 301–331, May 2000.

57. Acar, E., T. G. Kolda, and D. M. Dunlavy, “All-at-once optimization for coupled matrix and tensor factorizations,” in *KDD Workshop on Mining and Learning with Graphs*, 2011.

58. Nunez, P. L., and R. Srinivasan, *Electric Fields of the Brain: The Neurophysics of EEG*, Oxford University Press, 2nd ed., December 2005.

59. Baillet, S., J. C. Mosher, and R. Leahy, “Electromagnetic brain mapping,” *IEEE Signal Processing Magazine*, Vol. 18, no. 6, pp. 14–30, 2001.

60. Srinivasan, R., P. L. Nunez, D. M. Tucker, R. B. Silberstein, and P. J. Cadusch, “Spatial sampling and filtering of EEG with spline laplacians to estimate cortical potentials.,” *Brain Topography*, Vol. 8, no. 4, pp. 355–366, 1996.

61. Phillips, C. L. M., *Source estimation in EEG*. PhD thesis, Université de Liège, Liège, Belgium, 2000.

62. Ogawa, S., T. M. Lee, A. R. Kay, and D. W. Tank, “Brain magnetic resonance imaging with contrast dependent on blood oxygenation,” *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 87, no. 24, pp. 9868–9872, 1990.

63. Logothetis, N. K., J. Pauls, M. Augath, T. Trinath, and A. Oeltermann, "Neurophysiological investigation of the basis of the fMRI signal," *Nature*, Vol. 412, pp. 150–7, July 2001.

64. Trevelyan, A. J., "The direct relationship between inhibitory currents and local field potentials," *Journal of Neuroscience*, Vol. 29, pp. 15299–15307, Dec. 2009.

65. Berens, P., N. K. Logothetis, and A. S. Tolias, "Local field potentials, BOLD and spiking activity: Relationships and physiological mechanisms," in *Visual population codes - toward a common multivariate framework for cell recording and functional imaging* (Kriegeskorte, N., and G. Kreiman, eds.), pp. 599–624, MIT Press, Jan 2012.

66. Goense, J. B., and N. K. Logothetis, "Neurophysiology of the BOLD fMRI signal in awake monkeys," *Current Biology*, Vol. 18, pp. 631–640, May 2008.

67. Lindauer, U., U. Dirnagl, M. Füchtemeier, C. Böttiger, N. Offenhauser, C. Leithner, and G. Royl, "Pathophysiological interference with neurovascular coupling - when imaging based on hemoglobin might go blind," *Frontiers in Neuroenergetics*, Vol. 2, Oct 2010. Available: http://journal.frontiersin.org/article/10.3389/fnene.2010.00025/full.

68. Friston, K. J., a. Mechelli, R. Turner, and C. J. Price, "Nonlinear responses in fMRI: the balloon model, volterra kernels, and other hemodynamics.," *NeuroImage*, Vol. 12, no. 4, pp. 466–477, 2000.

69. Riera, J. J., and A. Sumiyoshi, "Brain oscillations: ideal scenery to understand the neurovascular coupling," *Current Opinion in Neurology*, Vol. 23, pp. 374–381, July 2010.

70. Sotero, R. C., and N. J. Trujillo-Barreto, "Modelling the role of excitatory and inhibitory neuronal activity in the generation of the BOLD signal," *NeuroImage*, Vol. 35, pp. 149–165, Mar. 2007.

71. Babajani, A., and H. Soltanian-Zadeh, "Integrated MEG/EEG and fMRI model based on neural masses," *IEEE Transactions on Biomedical Engineering*, Vol. 53, pp. 1794–1801, Sept. 2006.

72. Glover, G. H., "Deconvolution of impulse response in event-related BOLD fMRI," *NeuroImage*, Vol. 9, pp. 416–29, Apr. 1999.

73. Nunez, P. L., and R. B. Silberstein, "On the relationship of synaptic activity to macroscopic measurements: does co-registration of EEG with fMRI make sense?," *Brain Topography*, Vol. 13, pp. 79–96, Jan. 2000.

74. Liu, A. K., J. W. Belliveau, and a. M. Dale, "Spatiotemporal imaging of human brain activity using functional MRI constrained magnetoencephalography data: Monte Carlo simulations," *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 95, pp. 8945–50, July 1998.

75. Laufs, H., a. Kleinschmidt, a. Beyerle, E. Eger, a. Salek-Haddadi, C. Preibisch, and K. Krakow, "EEG-correlated fMRI of human alpha activity," *NeuroImage*, Vol. 19, pp. 1463–1476, Aug. 2003.

76. de Munck, J., S. Gonçalves, R. Mammoliti, R. Heethaar, and F. Lopes da Silva, "Interactions between different EEG frequency bands and their effect on alpha-fMRI correlations," *NeuroImage*, Vol. 47, pp. 69–76, Aug 2009.

77. Archer, J., D. F. Abbott, A. B. Waites, and G. D. Jackson, “fMRI deactivation of the posterior cingulate during generalized spike and wave,” *NeuroImage*, Vol. 20, pp. 1915–1922, Dec. 2003.

78. Hamandi, K., A. Salek-Haddadi, D. R. Fish, and L. Lemieux, “EEG/functional MRI in epilepsy: The queen square experience,” *Journal of Clinical Neurophysiology*, Vol. 21, no. 4, pp. 241–248, 2004.

79. Goldman, R. I., J. M. Stern, J. Engel, and M. S. Cohen, “Simultaneous EEG and fMRI of the alpha rhythm,” *Neuroreport*, Vol. 13, pp. 2487–92, Dec. 2002.

80. Moosmann, M., “Correlates of alpha rhythm in functional magnetic resonance imaging and near infrared spectroscopy,” *NeuroImage*, Vol. 20, pp. 145–158, Sept. 2003.

81. Laufs, H., J. L. Holt, R. Elfont, M. Krams, J. S. Paul, K. Krakow, and a. Kleinschmidt, “Where the BOLD signal goes when alpha EEG leaves.,” *NeuroImage*, Vol. 31, pp. 1408–18, July 2006.

82. Laufs, H., “Endogenous brain oscillations and related networks detected by surface EEG-combined fMRI.,” *Human Brain Mapping*, Vol. 29, pp. 762–9, July 2008.

83. Daunizeau, J., C. Grova, G. Marrelec, J. Mattout, S. Jbabdi, M. Pélégrini-Issac, J.-M. Lina, and M. H. Benali, “Symmetrical event-related EEG/fMRI information fusion in a variational bayesian framework.,” *NeuroImage*, Vol. 36, pp. 69–87, May 2007.

84. Friston, K. J., L. M. Harrison, and W. D. Penny, “Dynamic causal modelling,” *NeuroImage*, Vol. 19, pp. 1273–1302, Aug. 2003.

85. Trujillo-Barreto, N. J., E. Martínez-Montes, L. Melie-García, and P. A. Valdés-Sosa, “A symmetrical bayesian model for fMRI and EEG/MEG neuroimage fusion,” *International Journal of Bioelectromagnetism*, Vol. 3, no. 1, 2001.

86. Basar, E., T. Demiralp, and M. Schürmann, “Oscillatory brain dynamics, wavelet analysis, and cognition,” *Brain and Language*, Vol. 183, pp. 146–183, January 1999.

87. Liu, W., J. Chan, J. Bailey, C. Leckie, and K. Ramamohanarao, “Mining labelled tensors by discovering both their common and discriminative subspaces,” in *Proceedings of the 2013 SIAM Conference on Data Mining*, 2013.

88. Kimura, K., Y. Tanaka, and M. Kudo, “A fast hierarchical alternating least squares algorithm for orthogonal nonnegative matrix factorization,” in *JMLR: Workshop and Conference Proceedings*, pp. 129–141, 2014.

89. Tarantola, A., *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM, 2005.

90. Schwarz, G., “Estimating the dimension of a model,” *The Annals of Statistics*, Vol. 6, no. 2, pp. 461–464, 1978.

91. Hebiri, M., “Regularization with the smooth-lasso procedure,” *arXiv preprint arXiv: 0803.0668*, 2008.

92. Boutsidis, C., and E. Gallopoulos, “SVD based initialization: A head start for nonnegative matrix factorization,” *Pattern Recognition*, Vol. 41, pp. 1350–1362, Apr. 2008.

93. Uschmajew, A., "Local convergence of the alternating least squares algorithm for canonical tensor approximation," *SIAM Journal on Matrix Analysis and Applications*, Vol. 33, pp. 639–652, Jan. 2012.

94. Wang, L., M. T. Chu, and B. Yu, "Orthogonal low rank tensor approximation: Alternating least squares method and its global convergence," *SIAM Journal on Matrix Analysis and Applications*, Vol. 36, pp. 1–19, Jan. 2015.

95. Bayram, A., Z. Bayraktaroglu, E. Karahan, B. Erdogan, B. Bilgic, M. Ozker, I. Kasikci, A. D. Duru, A. Ademoglu, C. Oztürk, K. Arikan, N. Tarhan, and T. Demiralp, "Simultaneous EEG/fMRI analysis of the resonance phenomena in steady-state visual evoked responses.," *Clinical EEG and Neuroscience*, Vol. 42, pp. 98–106, Apr. 2011.

96. Allen, P. J., O. Josephs, and R. Turner, "A method for removing imaging artifact from continuous EEG recorded during functional MRI," *NeuroImage*, Vol. 12, pp. 230–239, Aug. 2000.

97. Wellcome Trust Centre for Neuroimaging, "Statistical parametric mapping software." Available: http://www.fil.ion.ucl.ac.uk/spm.

98. Duru, A. D., *Neuroimaging of Brain Activity using Spatio-temporal Signal Modelling*. PhD thesis, Bogazici University, Istanbul, Turkey, 2012.

99. Thomson, D., "Spectrum estimation and harmonic analysis," *Proceedings of the IEEE*, Vol. 70, no. 9, pp. 1055–1096, 1982.

100. Valdés-Sosa, P. A., J. Bosch-Bayard, R. Grave, J. Hernandez, J. Riera, R. Pascual, R. Biscay, J. Bosch, R. Grave, J. Hernandez, J. Riera, R. Pascual, R. Biscay, J. Bosch-Bayard, R. Grave, J. Hernandez, J. Riera, R. Pascual, and R. Biscay, "Frequency domain models of the EEG," *Brain Topography*, Vol. 4, no. 4, pp. 309–319, 1992.

101. Yılmaz, Y. K., A. T. Cemgil, and U. Simsekli, "Generalised coupled tensor factorisation," in *NIPS*, pp. 1–9, 2011.

102. Sorber, L., M. Van Barel, and L. De Lathauwer, "Structured data fusion," *IEEE Journal of Selected Topics in Signal Processing*, Vol. 9, pp. 586–600, June 2015.

103. Groves, A. R., C. F. Beckmann, S. M. Smith, and M. W. Woolrich, "Linked independent component analysis for multimodal data fusion," *NeuroImage*, Vol. 54, pp. 2198–217, Feb. 2011.

104. Papalexakis, E. E., C. Faloutsos, T. M. Mitchell, and N. D. Sidiropoulos, "Turbo-SMT: Accelerating coupled sparse matrix-tensor factorizations by 200x," in *SIAM International Conference on Data Mining*, SIAM, 2014.

105. Friston, K., "Functional integration in the brain," in *Human Brain Function* (Friston, K. J., C. D. Frith, R. Dolan, C. J. Price, S. Zeki, J. T. Ashburner, and W. D. Penny, eds.), pp. 971–997, London: Academic Press, 2nd ed., 2003.

106. Friston, K. J., P. Jezzard, and R. Turner, "Analysis of functional MRI time-series," *Human Brain Mapping*, Vol. 1, pp. 153–171, October 1994.

107. Huettel, S. A., A. W. Song, and G. McCarthy, *Functional Magnetic Resonance Imaging*, Sunderland, Massachusetts: Sinauer Associates, Inc, 2nd ed., 2009.

108. Rogers, B. P., V. L. Morgan, A. T. Newton, and J. C. Gore, "Assessing functional connectivity in the human brain by fMRI," *Magnetic Resonance Imaging*, Vol. 25, pp. 1347–57, Dec. 2007.

109. Horwitz, B., "The elusive concept of brain connectivity," *NeuroImage*, Vol. 19, pp. 466–470, June 2003.

110. Calhoun, V. D., T. Adalı, G. D. Pearlson, and J. J. Pekar, "Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms," *Human Brain Mapping*, Vol. 13, pp. 43–53, May 2001.

111. Mourão Miranda, J., A. L. Bokde, C. Born, H. Hampel, and M. Stetter, "Classifying brain states and determining the discriminating activation patterns: Support vector machine on functional MRI data," *NeuroImage*, Vol. 28, pp. 980–995, Dec. 2005.

112. Valdés-Sosa, P. A., A. Roebroeck, J. Daunizeau, and K. J. Friston, "Effective connectivity: Influence, causality and biophysical modeling," *NeuroImage*, Vol. 58, pp. 339–361, Apr. 2011.

113. Buxton, R. B., E. C. Wong, and L. R. Frank, "Dynamics of blood flow and oxygenation changes during brain activation: The balloon model," *Magnetic Resonance in Medicine*, Vol. 39, pp. 855–864, June 1998.

114. Mandeville, J. B., J. J. A. Marota, C. Ayata, G. Zaharchuk, M. A. Moskowitz, B. R. Rosen, and R. M. Weisskoff, "Evidence of a cerebrovascular postarteriole Windkessel with delayed compliance," *Journal of Cerebral Blood Flow & Metabolism*, Vol. 19, pp. 679–689, June 1999.

115. Wiener, N., "The theory of prediction," in *Modern Mathematics for Engineers* (Beckenbach, E. F., ed.), New York: McGraw-Hill, 1956.

116. Akaike, H., "On the use of a linear model for the identification of feedback systems," *Annals of the Institute of Statistical Mathematics*, Vol. 20, no. 1, pp. 425–439, 1968.

117. Schweder, T., "Composable markov processes," *Journal of Applied Probability*, Vol. 7, no. 2, pp. 400–410, 1970.

118. Marinazzo, D., W. Liao, H. Chen, and S. Stramaglia, "Nonlinear connectivity by Granger causality," *NeuroImage*, Vol. 58, pp. 330 – 338, February 2011.

119. Haufe, S., R. Tomioka, G. Nolte, K.-R. Müller, and M. Kawanabe, "Modeling sparse connectivity between underlying brain sources for EEG/MEG," *IEEE Transactions on Biomedical Engineering*, Vol. 57, pp. 1954–1963, Aug 2010.

120. Vinck, M., L. Huurdeman, C. A. Bosman, P. Fries, F. P. Battaglia, C. M. Pennartz, and P. H. Tiesinga, "How to detect the Granger-causal flow direction in the presence of additive noise?," *NeuroImage*, Vol. 108, pp. 301–318, Mar 2015.

121. Bressler, S. L., and A. K. Seth, "Wiener-Granger causality: A well established methodology," *NeuroImage*, Vol. 58, no. 2, pp. 323–329, 2011.

122. Deshpande, G., K. Sathian, and X. Hu, "Effect of hemodynamic variability on Granger causality analysis of fMRI," *NeuroImage*, Vol. 52, pp. 884–896, Sep 2010.

123. Wen, X., G. Rangarajan, and M. Ding, "Is Granger causality a viable technique for analyzing fMRI data?," *PLoS ONE*, Vol. 8, p. e67428, July 2013.

124. Seth, A. K., P. Chorley, and L. C. Barnett, "Granger causality analysis of fMRI BOLD signals is invariant to hemodynamic convolution but not downsampling," *NeuroImage*, Vol. 65, pp. 540–555, Jan. 2013.

125. Rodrigues, J. a., and A. Andrade, "Lag-based effective connectivity applied to fMRI: A simulation study highlighting dependence on experimental parameters and formulation," *NeuroImage*, Vol. 89, pp. 358–377, Apr. 2014.

126. Lin, F. H., J. Ahveninen, T. Raij, T. Witzel, Y. H. Chu, I. P. Jääskeläinen, K. W. K. Tsai, W. J. Kuo, and J. W. Belliveau, "Increasing fMRI sampling rate improves Granger causality estimates," *PLoS ONE*, Vol. 9, no. 6, pp. 26–29, 2014.

127. Akaike, H., "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, Vol. 19, pp. 716–723, Dec. 1974.

128. Valdés-Sosa, P. A., J. M. Sanchez-Bornot, A. Lage-Castellanos, M. Vega-Hernández, J. Bosch-Bayard, L. Melie-García, and E. Canales-Rodríguez, "Estimating brain functional connectivity with sparse multivariate autoregression.," *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, Vol. 360, pp. 969–81, May 2005.

129. Hayes, M., *Statistical Digital Signal Processing and Modeling*, John Wiley & Sons, Inc., 1st ed., 1996.

130. Chi, E. C., and K. Lange, "Stable estimation of a covariance matrix guided by nuclear norm penalties," *Computational Statistics and Data Analysis*, Vol. 80, pp. 117–128, May 2014.

131. Huang, H., C. Ding, D. Luo, and T. Li, "Simultaneous tensor subspace selection and clustering: The equivalence of high order SVD and k-means clustering," in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (New York, USA), p. 327, ACM Press, 2008.

132. Ding, C., T. Li, and M. I. Jordan, "Convex and semi-nonnegative matrix factorizations.," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, pp. 45–55, Jan. 2010.

133. Papalexakis, E. E., N. D. Sidiropoulos, and R. Bro, "From k-means to higher-way co-clustering: Multilinear decomposition with sparse latent factors," *IEEE Transactions on Signal Processing*, Vol. 61, pp. 493–506, Jan. 2013.

134. Gabay, D., and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Computers & Mathematics with Applications*, Vol. 2, no. 1, pp. 17–40, 1976.

135. Glowinski, R., and A. Marrocco, "Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéares," *Revue Française d'Automatique, Informatique, et Recherche Opérationelle*, Vol. 9, no. 2, pp. 41–76, 1975.

136. Gandy, S., B. Recht, and I. Yamada, "Tensor completion and low-n-rank tensor recovery via convex optimization," *Inverse Problems*, Vol. 27, p. 025010, Feb. 2011.

137. Signoretto, M., L. De Lathauwer, and J. A. K. Suykens, "Nuclear norms for tensors and their use for convex multilinear estimation," tech. rep., ESAT-SISTA, K.U. Leuven, Leuven, 2010.

138. Boyd, S., N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, Vol. 3, no. 1, pp. 1–122, 2010.

139. Beck, A., and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, Vol. 2, pp. 183–202, Jan. 2009.

140. Makeig, S., S. Debener, J. Onton, and A. Delorme, "Mining event-related brain dynamics," *Trends in Cognitive Sciences*, Vol. 8, pp. 204–210, May 2004.

141. Acar, E., R. Bro, and A. K. Smilde, "Data fusion in metabolomics using coupled matrix and tensor factorizations," *Proceedings of the IEEE*, Vol. 103, pp. 1602–1620, Sept 2015.

142. Acar, E., M. Nilsson, and M. Saunders, "A flexible modeling framework for coupled matrix and tensor factorizations," in *Proceedings of the 22nd European Signal Processing Conference*, pp. 111–115, 2014.

143. Li, B., G. Zhou, and A. Cichocki, "Two efficient algorithms for approximately orthogonal nonnegative matrix factorization," *IEEE Signal Processing Letters*, Vol. 22, no. 7, pp. 843–846, 2011.

144. Zhao, Q., L. Zhang, and A. Cichocki, "Bayesian CP factorization of incomplete tensors with automatic rank determination," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 37, pp. 1751–1763, Sept. 2015.

145. Mørup, M., and L. K. Hansen, "Automatic relevance determination for multi-way models," *Journal of Chemometrics*, Vol. 23, pp. 352–363, July 2009.

146. Simsekli, U., B. Ermis, A. Cemgil, and E. Acar, "Optimal weight learning for coupled tensor factorization with mixed divergences," in *Proceedings of the 21st European Signal Processing Conference*, pp. 1–5, Sept 2013.