

TENSOR ANALYSIS OF
ELECTROENCEPHALOGRAPH SIGNAL FOR
LOCALIZATION OF EVENT-RELATED
POTENTIALS

by

SAEED POURYAZDIAN

B.Sc, Ferdowsi University of Mashhad, Mashhad, Iran, 2006
M.Sc, Iran University of Science and Tech., Tehran, Iran, 2009

A dissertation
presented to Ryerson University

in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Program of
Electrical and Computer Engineering

Toronto, Ontario, Canada, 2017
©Saeed Pouryazdian 2017

Declaration

I hereby declare that I am the sole author of this dissertation. This is a true copy of the dissertation, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this dissertation to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this dissertation by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my dissertation may be made electronically available to the public.

Tensor Analysis of Electroencephalogram Signal for Localization of
Event-Related Potentials

Doctor of Philosophy 2017

Saeed Pouryazdian

Electrical and Computer Engineering

Ryerson University

Abstract

Electroencephalogram (EEG) is widely used for monitoring, diagnosis purposes and also for study of brains physiological, mental and functional abnormalities. EEG is known to be a high-dimensional signal in which processing of information by the brain is reflected in dynamical changes of the electrical activity in time, frequency, and space. EEG signal processing tends to describe and quantify these variations into functions with known spatio-temporal-spectral properties or at least easier to characterize. Multi-channel EEG recordings naturally include multiple modes. Matrix analysis, via stacking or concatenating other modes with the retained two modes, has been extensively used to represent and analyze the EEG data. On the other hand, Multi-way (tensor) analysis techniques keep the structure of the data, and by analyzing more dimensions simultaneously, summarize the data into more interpretable components.

This work presents a generalized multi-way array analysis methodology in pattern classification systems as related to source separation and discriminant feature selection in EEG signal processing problems. Analysis of ERPs, as one of the main categories of EEG signals, requires systems that can exploit the variation of the signals in different contextual domains in order to reveal the hidden structures in the data. Temporal, spectral, spatial, and subjects/experimental conditions of multi-channel ERP signals are exploited here to generate three-way and four-way ERP tensors. Two key elements of this framework are the Time-Frequency representation (TFR) and CANDECOMP/PARAFAC model order selection techniques we incorporate for analysis. Here,

we propose a fully data-driven TFR scheme, via combining the Empirical Mode Decomposition and Reassignment method, which yields a high resolution and cross-term free TFR. Furthermore, we develop a robust and effective model order selection scheme that outperforms conventional techniques in mid and low SNRs (i.e. 0 – 10 dB) with a better Probability of Detection (PoD) and almost no extra computational overhead after the CANDECOMP/PARAFAC decomposition.

ERP tensor can be regarded as a mixture that includes different kinds of brain activity, artifacts, interference, and noise. Using this framework, the desired brain activity could be extracted out from the mixture. The extracted signatures are then translated for different applications in brain-computer interface and cognitive neuroscience.

Acknowledgements

First and foremost, this research would not have been possible without the inspiration and wonderful guidance of my supervisors, Professor Sri Krishnan and Professor Soosan Beheshti, throughout all the stages of my PhD studies. Their continued encouragement and support made the whole journey a smooth and pleasant one.

Besides my advisors, I would like to thank the rest of my PhD defence committee: Prof. Karthi Umapathy, Prof. Alagan Anpalagan, Prof. Lawrence Kolasa, and Prof. Dimitrios Hatzinakos, not only for their insightful comments and suggestions that made the dissertation more readable, but also for the hard questions which incited me to widen my research from various perspectives.

I would like to thank my fellow labmates in Signal Analysis Laboratory (SAR) and Signal and Information Processing (SIP) Lab for their support, for the stimulating discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had in the last five years. Also I thank my colleagues in the following institutions, Auditory Development Lab at McMaster University and the Brain-Computer Interface Lab at Universidade Federal do Espirito Santo in Brazil. In particular, I am grateful to Dr. Laurel J. Trainor and Prof. Teodiano F. Bastos Filho for letting me use their ERP database.

I would like to thank my family: my parents, my sisters and my brother for their continued love and encouragement and for supporting me spiritually throughout writing this dissertation and my life in general. This accomplishment would not have been possible without their help.

Last but not least, I would like to gratefully acknowledge the financial support of Canada Research Chairs program, and the Natural Sciences and Engineering Research Council of Canada (NSERC).

Dedication

To my wonderful father and mother, Hossein and Shamsi, for their unconditional love, enormous support and continuous encouragement...

Table of Contents

Abstract	iii
Acknowledgements	v
Dedication	vi
List of Tables	xi
List of Figures	xii
List of Acronyms	xviii
1 Introduction	1
1.1 Electroencephalogram	1
1.1.1 EEG Recordings	3
1.1.2 EEG Rhythms	4
1.2 Fundamentals of EEG Signal Processing	10
1.2.1 Nonlinearity of the medium	11
1.2.2 Nonstationarity	11
1.2.3 Signal Segmentation	12
1.3 EEG Signal Analysis	12
1.3.1 Time Domain Analysis	13
1.3.2 Frequency Domain Analysis	14
1.3.3 Time-Frequency Domain Analysis	15
1.3.4 Spatio-Temporal Domain Analysis	17
1.4 Contributions of The Dissertation	20
1.5 Organization of The Dissertation	24
2 Time-frequency Representation	27
2.1 Classical Time-frequency Distributions	28
2.1.1 Cohens Class Bilinear TFRs	29
2.1.2 Cohen-Posch TFR	29

2.1.3	Spectrogram	30
2.1.4	Wavelet Scalogram	31
2.1.5	Matching Pursuit TFR	32
2.1.6	Dictionary Learning algorithms for TFR	34
2.2	Review and shortcomings of TFR Methods	35
2.3	TFR based on Empirical Mode Decomposition and Reassignment method (RS-EMD)	40
2.3.1	Empirical Mode Decomposition (EMD)	41
2.3.2	EMD-based Detrending of EEG signals	44
2.3.3	Reassignment Method (RM)	45
2.3.4	Quantitative Performance Index	48
2.4	Chapter Summary	49
3	Multi-way Analysis	51
3.1	Multi-way arrays	54
3.1.1	Notation	54
3.1.2	Subarrays, Tubes and Slices	56
3.1.3	Outer, Kronecker, and Khatri-Rao Products	58
3.2	Multi-way Models	60
3.2.1	TUCKER Model	61
3.2.1.1	Algorithms	63
3.2.2	CANDECOMP/PARAFAC MODEL	64
3.2.2.1	Connection between CANDECOMP/PARAFAC Model and TUCKER3 Model	67
3.2.2.2	Algorithms	69
3.2.3	Numerical Complexity	73
3.3	Chapter Summary	76
4	CANDECOMP/PARAFAC Model Order Selection	77
4.1	Problem Formulation	81
4.1.1	Problem Statement: Rank Approximation	83
4.2	Rank Selection Techniques	84
4.2.1	Existing CANDECOMP/PARAFAC model rank selection techniques with white noise	84
4.2.1.1	Core Consistency Diagnostic (CORCONDIA)	85

4.2.1.2	<i>R</i> -D Exponential Fitting Test	87
4.2.1.3	Second ORder sTatistic of the Eigenvalues (SORTE) . .	90
4.2.1.4	Approximate Rank-Detecting Factorization	91
4.3	Proposed Rank Selection approach based on Reconstruction Error	92
4.3.1	Multidimensional Prewhitening	93
4.3.2	Rank Estimation Criterion	94
4.3.2.1	Incorporating CORCONDIA for selection of R_{max}	96
4.3.3	Numerical Complexity	97
4.3.4	Unknown Noise Covariance	99
4.4	Simulations	100
4.4.1	Additive White Noise	102
4.4.2	Additive Colored Noise	109
4.4.2.1	Known Covariance Matrix	109
4.4.2.2	Unknown Covariance Matrix	111
4.5	Chapter Summary	112
5	EEG Tensor Analysis	114
5.1	Multi-way nature of EEG signals	116
5.2	Multiple modes and high-order tensors of EEG data	117
5.3	EEG tensor generation via RS-EMD	118
5.4	Localization of Brain Activities via Multi-way Analysis of EEG Tensor .	120
5.4.1	Methods	120
5.4.2	Three-way EEG Tensor Generation	121
5.4.3	CANDECOMP/PARAFAC Modeling of EEG Tensor	121
5.4.4	Results	123
5.4.4.1	Synthetic EEG data	123
5.4.4.2	Steady-State Visual Evoked Potential	125
5.5	New method for SSVEP Detection via CANDECOMP/ PARAFAC and Canonical Correlation Analysis	129
5.5.1	CANDECOMP/PARAFAC Analysis of EEG tensor	132
5.5.2	Canonical Correlation Analysis (CCA)	132
5.5.3	Methods	133
5.5.3.1	Subjects and EEG preparation	133
5.5.3.2	System architecture, visual stimulus and experimental task	133

5.5.4	Results	134
5.6	Chapter Summary	138
6	New Multi-domain Feature Selection Scheme from EEG tensor	140
6.1	Methods	143
6.1.1	ERP Data Description	143
6.1.2	ERP-tensor Generation via RS-EMD	145
6.1.3	CANDECOMP/PARAFAC Modeling of ERP tensor	146
6.1.3.1	CANDECOMP/PARAFAC model order selection	147
6.1.4	Proposed Multi-domain Feature Selection Scheme	148
6.1.4.1	Template Matching Method	149
6.1.4.2	Discriminability of the Multi-domain Feature	150
6.2	Results	151
6.2.1	Decomposition of Multiple Models	154
6.3	Chapter Summary	157
7	Conclusions and Future Work	158
7.1	Outcome of the proposed work	160
7.2	Limitations and Future Work	164
	Appendix	166
	References	174

List of Tables

1.1	Summary of the proposed solutions and the requirement for efficient EEG signal analysis	24
2.1	Desirable TFR Properties for TF Quantification	39
2.2	Desirable TFR Properties of the proposed technique	50
5.1	Accuracy Results Using WL of 2 s	137
7.1	Summary of the proposed solutions and the requirement for efficient EEG signal analysis	161

List of Figures

1.1	Placement of recording electrodes according the international 10 – 20 system seen from (a) left and above the head. A = Ear lobe, C = central, P = parietal, F = frontal, F_p = frontal polar, O = occipital; (b) indicates a two-dimensional view of the electrode set-up configuration [13].	5
1.2	(a) Bipolar and (b) unipolar measurements. [13].	6
1.3	The EEG signal on the top and four typical dominant brain normal rhythms, from low to high frequencies; the delta wave is observed in infants and sleeping adults, the theta wave in children and sleeping adults, the alpha wave is detected in the occipital brain region when there is no attention, and the beta wave appears frontally and parietally with low amplitude [1].	10
1.4	BSS concept; mixing and blind separation of the EEG signals.	19
1.5	Flowchart of the proposed methodologies highlighting the original contributions with the darker color.	22
1.6	Organization of the dissertation.	25
2.1	Chapter 2- Time-frequency Representation.	27
2.2	A diagram of well-known TF distributions.	28
2.3	A diagram of TF transformation.	29
2.4	Empirical Mode Decomposition of a sample EEG signal. Original signal (Top), extracted IMF1 to IMF7 showing the high to low frequency oscillations, and the residue (Bottom).	43
2.5	Detrending of a EEG signal. left: standardized empirical mean of the fine-to-coarse EMD reconstruction, evidencing $D = 5$ as the change point. Top right: original signal. Middle right: estimated trend obtained from the partial reconstruction with IMFs 6 to 7 + the residual. Bottom right: detrended signal obtained from the partial reconstruction with IMFs 1 to 5.	45

2.6	Time-Frequency representation of a sample 10Hz Steady-State Visual Evoked Potential (SSVEP) signal: (a) Wavelet Transform using Complex Morlet “cmor1-1.5”; (b) RS-EMD, i.e., EMD followed by Reassigned TFR of extracted IMFs.	48
3.1	Chapter 3- Multi-way Analysis.	51
3.2	Illustration of multi-way data: zero-way tensor = scalar, 1-way tensor = row or column vector, 2-way tensor = matrix, N-way tensor = higher-order tensors. The 4-way and 5-way tensors are represented here as a set of the three-way tensors [61].	55
3.3	A three-way array (third-order tensor) $\underline{\mathbf{Y}} \in \mathbb{R}^{7 \times 5 \times 8}$ with elements y_{itq} [61].	56
3.4	(a) Fibers for a third-order tensor (all fibers are treated as column vectors); (b) Slices for a third-order tensor $\underline{\mathbf{Y}} \in \mathbb{R}^{I \times J \times K}$. A colon is used to indicate all elements of a mode [61].	57
3.5	(P, Q, R)-component Tucker3 model, where a three-way array $\underline{\mathbf{Y}} \in \mathbb{R}^{I \times J \times K}$ is modeled with component matrices $\mathbf{A} \in \mathbb{R}^{I \times P}$, $\mathbf{B} \in \mathbb{R}^{J \times Q}$ and $\mathbf{C} \in \mathbb{R}^{K \times R}$ in the first, second and third mode, respectively. $\underline{\mathbf{G}} \in \mathbb{R}^{P \times Q \times R}$ is the core array and $\underline{\mathbf{E}} \in \mathbb{R}^{I \times J \times K}$ contains the error terms.	62
3.6	R-component CANDECOMP/PARAFAC model, where a three-way array $\underline{\mathbf{Y}}$ is expressed as the sum of rank-1 tensors and error terms. $\mathbf{a}_i, \mathbf{b}_i$ and \mathbf{c}_i are the i -th components in the first, second and third mode, respectively. $\underline{\mathbf{E}}$ is a three-way array containing the residuals.	66
3.7	CANDECOMP/PARAFAC model with a superdiagonal core tensor $\underline{\mathbf{G}} = \mathbf{I}^3, R$ for the third-order tensor $\underline{\mathbf{Y}} \cong \mathbf{I}^3, R \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} = \sum_{r=1}^R \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$	68
4.1	Chapter 4- CANDECOMP/PARAFAC Model Order Selection.	77
4.2	Block Diagram of the proposed scheme for model order selection in presence of colored noise which is combination of the multidimensional prewhitening with any model order selection scheme.	80
4.3	Schematic of (a) feed-forward and (b) feed-backward CANDECOMP/PARAFAC model order selection techniques.	85
4.4	The characteristics of the Data Error and Reconstruction Error for a sample rank-five tensor $\underline{\mathbf{Y}}$ while fitting an R-component model ($1 < R < 10$) in CANDECOMP/ PARAFAC analysis.	95

4.5	(a) The CORCONDIA measure for a sample rank-five tensor \underline{Y} along with 12.5% and 80% threshold lines, while fitting an R-component model for $1 < R < 10$; (b) The upperbound of γ_R for different estimation of variances of noise, using available ϵ_R , measured for a sample rank-five tensor $\underline{Y} \in \mathbb{R}^{15 \times 15 \times 20}$ while fitting an R-component CANDECOMP/PARAFAC model for $1 < R < 10$	98
4.6	Comparison of PoD, i.e. $Pr(\hat{R} = R^*)\%$ vs. SNR for different Rank selection methods while fitting a CANDECOMP/PARAFAC model averaged over 1000 Monte Carlo realizations while $\rho_r = 0$ for $r = 1, 2, 3$: (a) for a rank-5 tensor $\underline{Y} \in \mathbb{R}^{15 \times 15 \times 20}$; (b) for a rank-7 $\underline{Y} \in \mathbb{R}^{15 \times 15 \times 20}$	103
4.7	Boxplot of the estimated rank of the observed tensor over 1000 Monte Carlo realizations at $SNR = 0\text{dB}$ using different estimation methods. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points the algorithm considers to be not outliers, and the outliers are plotted individually for: (a) rank-five tensor \underline{Y} ; (b) rank – seven tensor \underline{Y}	104
4.8	Comparison of PoD, i.e. $Pr(\hat{R} = R^*)\%$ vs. SNR for a rank-overlap problem with the true rank of 3 and $\rho_r = 0$ for $r = 1, 2, 3$ for different Rank selection methods averaged over 1000 Monte Carlo realizations of $\underline{Y} \in \mathbb{R}^{15 \times 15 \times 20}$	105
4.9	Comparison of PoD, i.e. $Pr(\hat{R} = R^*)\%$ vs. tensor rank, at $SNR = 5\text{dB}$, $\rho_r = 0$ for $r = 1, 2, 3$ for different rank selection methods averaged over 1000 Monte Carlo realizations of $\underline{Y} \in \mathbb{R}^{15 \times 15 \times 20}$: (a) for an identity tensor in (4.3); (b) for a diagonal tensor in (4.3) with its elements uniformly selected between (0.2, 1).	107
4.10	Estimation accuracy of the factor matrices in terms of MSAE compared to the CRIB for different model orders vs. SNR, $\rho_r = 0$ for $r = 1, 2, 3$ averaged over 100 Monte Carlo realizations of $\underline{Y} \in \mathbb{R}^{15 \times 15 \times 20}$: (a)-(d) for an identity tensor in (4.3); (e)-(h) for a diagonal tensor in (4.3) with its element uniformly selected between (0.2, 1).	108
4.11	Comparison of PoD, i.e. $Pr(\hat{R} = R^*)\%$ vs. Tensor rank at $SNR = 5\text{dB}$ and $\rho_r = 0.5$ for $r = 1, 2, 3$, while fitting a CANDECOMP/PARAFAC model averaged over 1000 Monte Carlo realizations of $\underline{Y} \in \mathbb{R}^{15 \times 15 \times 20}$: (a) for colored noise; (b) for prewhitened noise.	109

4.12	Comparison of PoD, i.e. $Pr(\hat{R} = R^*)\%$, vs. Correlation Coefficient ($0.1 \leq \rho_r \leq 0.9$), at $SNR = 5\text{dB}$, for different Rank selection methods averaged over 1000 Monte Carlo realizations for a rank-7 $\underline{Y} \in \mathbb{R}^{15 \times 15 \times 20}$: (a) without prewhitening of the colored noise; (b) with prewhitening of the colored noise.	111
4.13	PoD vs. SNR of the proposed method with and without having the covariance matrices of the noise, averaged over 1000 Monte Carlo realizations of a rank-7 $\underline{Y} \in \mathbb{R}^{15 \times 15 \times 20}$ contaminated with colored noise of $\rho_r = 0.2, 0.5$ for $r = 1, 2, 3$.	112
5.1	Chapter 5- EEG Tensor Analysis.	114
5.2	third-order tensor of EEG with modes: <i>time</i> \times <i>frequency</i> \times <i>space</i>	118
5.3	Block diagram of the proposed EEG localization algorithm.	122
5.4	two-component CANDECOMP/PARAFAC model, where a three-way array \underline{Y} is expressed as the sum of two rank-1 tensors and error terms . \mathbf{a}_i , \mathbf{b}_i and \mathbf{c}_i are the i th components in the first, second and third mode, respectively. \underline{E} is a three-way array containing the residuals.	123
5.5	A comparison between the localization ratios B , for both the proposed RS-EMD method and the CWT-based TFR, evaluated for a sample EEG signal with 50 Hz oscillation.	124
5.6	3-component CANDECOMP/PARAFAC decomposition of a synthetic three-way EEG tensor $\underline{Y} \in \mathbb{R}^{512 \times 512 \times 32}$. temporal signatures, spectral signatures, and topographic maps of spatial signatures of the extracted factors are depicted: (a) using Wavelet transform for TFR; (b) using the proposed technique for TFR.	126
5.7	CANDECOMP/PARAFAC decomposition of a sample 10Hz SSVEP tensor $\underline{Y} \in \mathbb{R}^{256 \times 512 \times 248}$: (a) Temporal signatures; (b) Spectral signatures; (c) Topographic maps of Spatial signatures of 5 visually selected factors. The extracted factor #17 characterizes 10 Hz SSVEP and some other extracted factors characterizing the EEG background activities are also shown for comparison.	128
5.8	(a) Electrodes location using 10-20 system; (b) LCD screen showing the cue of beginning with the two LEDs coupled.	134

5.9	Signatures of the SSVEP templates for 8 Hz (left panel), and 13 Hz (right panel). The signatures characterise the spectral (top), temporal (middle), and spatial (bottom) dynamics of the corresponding SSVEP signal. . . .	135
5.10	Schematic of the proposed technique for SSVEP detection.	136
5.11	Extracted spectral signatures via CANDECOMP/PARAFAC, (i.e. \mathbf{a}_1 , \mathbf{a}_2 and \mathbf{a}_3) when Subject 1 was stimulated with: (a) 8 Hz and (b) 13 Hz, respectively.	137
6.1	Chapter 6- Multi-domain Feature Selection scheme from EEG tensor. . .	140
6.2	Auditory evoked event-related potential (ERP) waveform averaged across channels located at the mid-frontal area (F1, Fz, F2, FC1, FCz, and FC2) for (a) 10% session;(b) 20% session. Top: standard trial (blue), deviant trial (red), and the difference waveform of deviant minus standard trial (black). The shaded areas indicate the standard error of mean. Bottom: The topography of the MMN shows the typical frontal negativity of the MMN.	144
6.3	Illustration of a four-way ERP tensor of <i>time</i> \times <i>frequency</i> \times <i>channel</i> \times <i>subject/condition</i> which can be decomposed into its constituting factors via CANDECOMP/PARAFAC analysis.	145
6.4	Illustration of two types of error after CANDECOMP/PARAFAC analysis of four-way ERP tensor of <i>time</i> \times <i>frequency</i> \times <i>channel</i> \times <i>subject/condition</i> for a range of possible model orders: (a) Data Error; (b) Reconstruction Error.	149
6.5	(a) Temporal and spectral signatures used in the template matching approach with temporal component peaked at about 100 ms and the spectrum peaked around 5 Hz. (b) Histogram of the calculated angles for both temporal and spectral signatures of all extracted multi-domain features. .	153
6.6	Illustration of three extracted multi-domain features along their temporal, spectral and spatial signatures. The selected feature #21 (top row) matches both temporal and spectral characteristics of the MMN and also differentiate the subjects significantly (p-value < 0.05) for the two experimental conditions. The other two multi-domain features #43, #16 (middle and bottom row) are samples from the total of 52 features that do not satisfy the selection criteria.	154

6.7	CANDECOMP/PARAFAC analysis of four-way ERP tensor of <i>time</i> × <i>frequency</i> × <i>channel</i> × <i>subject/condition</i> for a range of possible model orders among 10 runs under each model: (a) The maximal, averaged and minimal sum of angles of the temporal and spectral signatures for the selected multi-domain feature; (b) The maximal, averaged and minimal p-value of the p-values for the selected features (p-value is to reveal the degree of the difference between two experimental conditions in the selected multi-domain feature in each run of each model).	155
7.1	Flowchart of the proposed contributions.	159

List of Acronyms

AIC	Akaike Information Criterion
ALS	Alternating Least Squares
ANOVA	Analysis Of Variance
AR	Autoregressive
ARMA	Autoregressive Moving Average
ATLD	Alternating TriLinear Decomposition
AWGN	Additive White Gaussian Noise
BIC	Bayesian Information Criterion
BCI	Brain-Computer Interface
BSS	Blind Source Separation
CANDECOMP	Canonical Decomposition
CCA	Canonical Correlation Analysis
CORCONDIA	Core Consistency Diagnostic
CEEMDAN	Complete Ensemble Empirical Mode Decomposition with Adaptive Noise
CRIB	Cramer-Rao Induced Bound
CWD	Choi-Williams Distribution
CWT	Continuous Wavelet Transform
ELS	Enhanced Line Search
EFT	Exponential Fitting Test
EM	Expectation Maximization
EEMD	Ensemble Empirical Mode Decomposition
EMD	Empirical Mode Decomposition
EP	Evoked Potential
ERP	Event Related Potential
EVD	Eigenvalue Decomposition

fMRI	Functional Magnetic Resonance Imaging
ICA	Independent Component Analysis
IMF	Intrinsic Mode Function
ITR	Information Transfer Rate
HHT	Hilbert-Huang Transom
HOSVD	Higher-Order Singular Value Decomposition
MIMO	Multiple Input Multiple Output
MP	Matching Pursuit
MSAE	Mean Square Angular Error
MMN	MisMatch Negativity
MSE	Mean Squared Error
NCG	Non-linear Conjugate Gradient
NMF	Non-negative Matrix Factorization
NTF	Non-negative Tensor Factorization
PARAFAC	Parallel Factor Analysis
PCA	Principle Component Analysis
PET	Positron Emission Tomography
PoD	Probability of Detection
RM	Reassignment Method
SNR	Signal-to-Noise Ratio
SORTE	Second ORder sTatistic of the Eigenvalues
SSVEP	Steady State Visually Evoked Potentials
STFT	Short-Time Fourier Transform
SVD	Singular Value Decomposition
SVM	Support Vector Machine
TFR	Time-Frequency Representation
WT	Wavelet Transform
WVD	Wigner-Ville Distribution

Chapter 1

Introduction

1.1 Electroencephalogram

Electroencephalogram (EEG) was discovered in 1924 by a German scientist, Hans Berger. Since then, this recording technique has become one of the most used tools to study and monitor the brain activity and diagnose various neurological disorders. The intrinsic simplicity and its low cost compared to other recording and brain imaging techniques, such as fMRI and PET, has made EEG a popular and widely accepted modality [1].

EEG is the recording of the electrical fields generated in the brain. Various groups of pyramidal cells of neurons, oriented perpendicularly to the surface of the head, are responsible for generating these electrical fields. Neurons are the source of these electrical fields by generating ionic current flows. Neurons that produce synchronized electric fields are grouped together and can be modeled as a microdipole [2]. EEG essentially records the electrical activity generated by various microdipoles located in the surface cortex of the brain. In order to generate an observable electric field from the scalp, approximately a group of 10^6 neurons orientated in the same direction with synchronized activity is enough [3]. However, EEG is a complex combination of rhythms that reflects the dynamical changes of the activities created in different parts of the brain concurrently. Although brain is the most complex organ in the body, through study of the brain activity,

a huge improvement of knowledge of the brain has perceived in the past decades.

Initially, before computers become ubiquitous, EEG analysis and conclusions were conducted through tedious visual interpretation and via manual measurements of the EEG traces. Hence, the extracted results were unreliable and subject to different interpretation of neurologists. Thanks to progresses in computerized data processing, fast and robust processing of the EEG signal became possible. Nowadays, various parametric and nonparametric methods are incorporated for digital processing of EEG signals to extract some insights from the data for different applications. Progressive developments in electrical engineering and the fascination with the human brain have attracted researchers from different scientific fields to investigate EEG recordings. Some of the main reasons behind the extensive usage of EEG are listed as follows [2]:

- i. The temporal resolution of EEG compared to other neuroimaging modalities is high. EEG can provides a resolution of up to few milliseconds, whereas PET and fMRI are limited to a few seconds.
- ii. The knowledge of the mechanism that generates spontaneous EEG activity has increased. Furthermore, fMRI is inapplicable for registering many types of mental activities, brain disorders and malfunctions of the brain, due to their low impact on the blood oxygenation level.
- iii. EEG recording systems are low cost and way cheaper than the multimillion dollar fMRI and PET scans.
- iv. New advanced signal processing methods such as Blind Source Separation (BSS), or time-frequency analysis like wavelet analysis and multi-way models have emerged and adapted for EEG analysis.

As a result, nowadays EEG has been widely used by many research centres and hospitals across the world in many different applications. Some of the applications in clinical neuroscience are monitoring alertness, coma and brain death [4], Detecting neurodegenerative disease such as Alzheimer's disease [5], Investigating sleep disorders [6]

and epilepsy [7], measuring the depth of anesthesia [8], testing drug effects [9]. Furthermore, EEG has found lots of applications in Brain Computer Interface (BCI) field which is a new communication channel that connects human brain with a device, bypassing the classical neuromuscular communication channels. Controlling video games [10], or controlling a machine [11] are some of BCI systems applications.

1.1.1 EEG Recordings

EEG can be recorded by placing different sensors on the scalp. Sensors are generally constructed from Ag/AgCL (Silver/Silver chloride) and they are typically built in the form of disks with the radius of 5 mm. Different standardizations for recording of EEG in term of placement of the electrodes have been proposed (10 – 20 system, Maudsley system, 10 – 10 system). Nowadays, the 10 – 20 recording system is the most common in the filed. The International 10 – 20 system of electrodes placement presents an uniform coverage of the entire scalp. This system is based on an iterative subdivision of arcs on the scalp starting from craniometric reference points: Nasion (Ns), Inion (In), Left (PAL) and Right (PAR) pre-auricular points [12]. The standard set of electrodes is detailed in Figure 1.1. The 10 and 20 in the International 10 – 20 system are named based on the distance between adjacent electrodes which indicates that distance between adjacent electrodes is either 10% or 20% of the specified distance measured using specific anatomical landmarks, e.g. the total distance between the front and back or left and right of the head. Electrodes are named with a a letter followed by a subscript number. The letter specifies the anatomical area or lobe of the brain where the electrode is places: prefrontal or frontopolar (Fp), frontal (F), central (C), parietal (P), occipital (O), temporal (T) and auricular (A). Depending on where the electrode is located, the subscript takes different values. Letter z as the subscript, indicates zero or midline placement of the electrode, and a numeric subscript indicates lateral placement. Even numbers are used to indicate the electrodes on the right side of the head, while odd numbers correspond to the electrodes on the left side. Besides, the subscript number increase with increasing

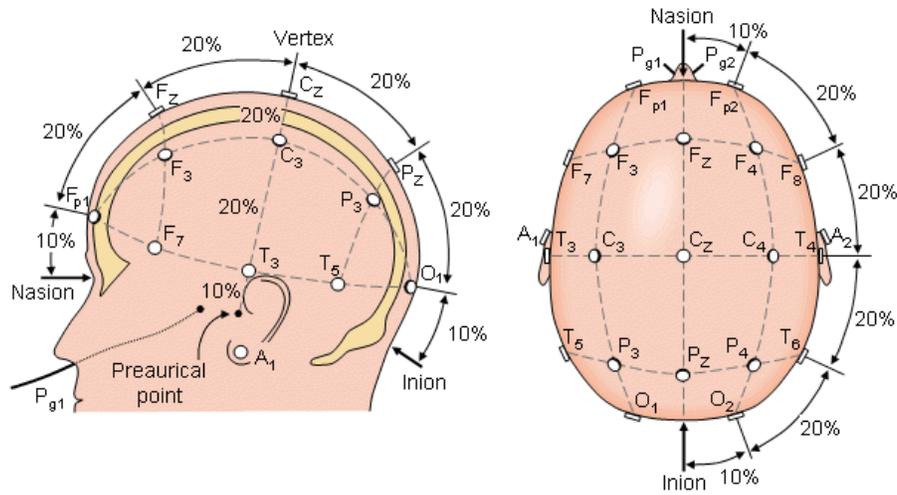
distance from the anterior posterior midline of the head [12].

Having the electrodes placed accordingly, we can use different montages for the recording of brain electrical potentials. Two popular montages for recording the EEG signals are the referential or bipolar montages (Figure 1.2). In referential montages, we have a reference electrode where the voltage differences between all electrodes and this common electrode are recorded. In bipolar montages, however, instead of using a reference electrode, the voltage difference between two designated electrodes is recorded (i.e., each electrode pair is considered as a channel). One of the main disadvantage of the referential montage is related to the fact that no reference is truly inactive, hence we can not find a single reference electrode that is optimal for all situations. As a result, the so-called common average reference, which corresponds to an artificial reference that is obtained by subtracting the time signal averaged over all sensors from the data of each channel, is a new montage that tries to alleviate this shortcoming. Bipolar montages, on the other hand, reduce the effects of common noise/artifacts and eliminate the influence of contaminated references [12].

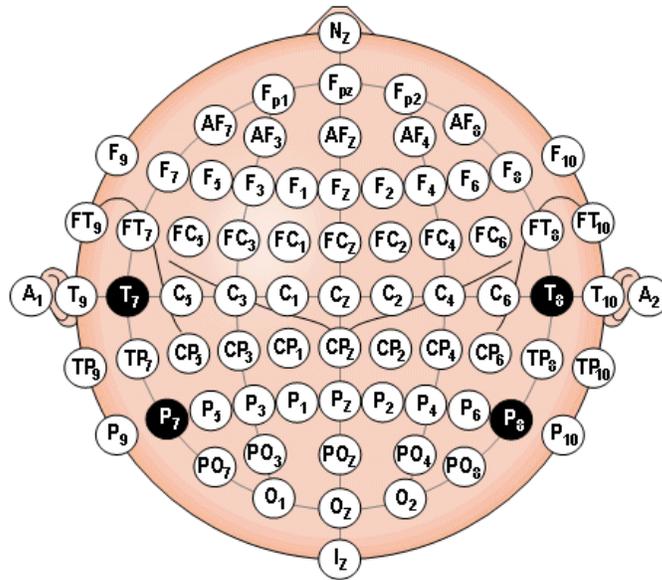
The effective bandwidth for EEG signals is limited to approximately 100 Hz. For many applications the bandwidth may be considered even half of this value. Therefore, in order to satisfy the Nyquist criterion, a minimum frequency of 200 Hz is needed for sampling the EEG signals. However, 256 Hz is the conventional sampling frequency used in most EEG acquisition systems.

1.1.2 EEG Rhythms

EEG signals exhibit several patterns of rhythmic or periodic activity that are associated with various physiological and mental processes. EEG rhythms are highly sensitive to the subject's state and can easily change depending on the subject's task. However, historically there are 5 commonly identified rhythmic activities in EEG the recordings which are divided in different frequency bands. The typical EEG frequency rhythms and their respective frequency bands are [2]:



(a)



(b)

Figure 1.1: Placement of recording electrodes according the international 10 – 20 system seen from (a) left and above the head. A = Ear lobe, C = central, P = parietal, F = frontal, F_p = frontal polar, O = occipital; (b) indicates a two-dimensional view of the electrode set-up configuration [13].

- δ rhythms, found in the frequency band of 1 to 4 Hz (δ band).
- θ rhythms, found in the frequency band of 4 to 8 Hz (θ band).

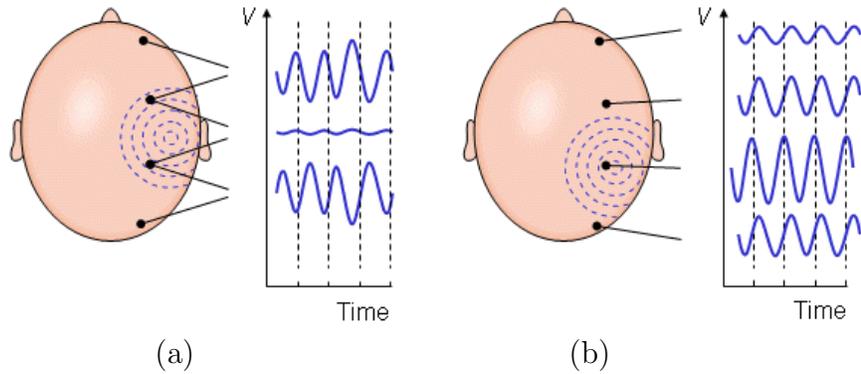


Figure 1.2: (a) Bipolar and (b) unipolar measurements. [13].

- α rhythms, found in the frequency band of 8 to 13 Hz (α band).
- β rhythms, found in the frequency band of 13 to 30 Hz (β band).
- γ rhythms, found in the frequency band of 30 to 100 Hz (γ band).

A brief description of these EEG rhythms are as follows:

δ rhythms

δ rhythms are the slowest of all the existing rhythms. However, they present the higher amplitude of all. These waves are primarily associated with deep sleep and may be present in the waking state. Having these characteristics, it is very likely for the genuine delta response to be mistaken with the artefact signals caused by the large muscles of the neck and jaw.

θ rhythms

θ rhythms are associated with drowsiness, childhood, adolescence and young adulthood. Larger contingents of theta wave activity in the waking adult are abnormal and are caused by various pathological problems. θ rhythms are also found during problem solving, for example mathematical problems such as adding or subtracting. It is located in the prefrontal part of the cortex.

α rhythms

α rhythms are associated with relaxed states. α waves appear in the posterior half of the head, are usually found over the occipital region of the brain, and can be detected in all parts of brain posterior lobes. α rhythms commonly appears as a round or sinusoidal-shaped signal. However, in rare cases it may manifest itself as sharp waves. In such cases, the negative component appears to be sharp and the positive component appears to be rounded. Different α rhythms are found in the human cortex: μ rhythm, α occipital rhythm and α parietal rhythm.

μ rhythms are rhythms that have been widely used in the implementation of BCI systems. They are named based on the similarity of their shape with letter μ (with sharp negative peaks). μ rhythms are identified in the frequency range close to 10 Hz. A known phenomenon related to synchronization of the groups of neurons in the motor cortex area is governed with change of these rhythms. Movement of arms and legs are controlled with these groups of neurons. When limbs are in the idle state and inactive, the μ rhythm presents activity, whereas when a subject moves his limbs the rhythm presents a decrease of amplitude, which is known as desynchronization. It has also been shown that this phenomenon and pattern presents in the brain not only when a subject physicality does an action of movement, but also happens when a subject just thinks about performing the same action without executing that. This has been the motivation for developing BCI systems based on motor imagery [11].

Another example of α rhythms, that are visible at occipital electrodes O1 and O2, is the α occipital rhythms. The amplitude of α occipital rhythms is higher compared to other rhythms. While the subject's eyes are closed, these rhythms enhance their amplitude, and subsequently they reduce the amplitude when the subject keeps his eyes opened. They also decrease in response to visual stimuli.

α parietal rhythm, is the last of the α rhythms that is identified in the parietal area with the dominant response at the electrode Pz. These rhythms can be enhanced when the subject closes his eyes. The functional properties of the parietal rhythm are not fully

understood.

β rhythms

β rhythms are observable at different locations in the cortex and their main feature is having low amplitude with multiple and varying frequencies. A β wave is associated with the usual waking rhythm of the brain that can be identified during active thinking, active attention, focus on the outside world or solving concrete problems, and is found in normal adults. A high level β wave may be acquired when a subject is in a panic state. There are different types of β rhythms, such as β Rolandic rhythms and β frontal rhythms.

β Rolandic rhythms usually present activity around the 20 Hz. Spontaneous activity recorded on electrodes close to the sensorimotor area (C3,Cz and C4), which are usually around the 20 Hz, are identified as β Rolandic rhythms. These rhythms are associated with the intentionality of performing a movement, presenting desynchronization even before than the μ rhythm.

The other type of β rhythms are β frontal rhythms with the dominant activity observable at electrodes F3, Fz and F4. They appear during cognitive tasks related with decision making.

γ rhythms

γ rhythms cover the frequencies above 30 Hz (mainly up to 45 Hz). Sometimes, these rhythms are called fast beta wave. γ rhythms are believed to be associated with higher mental activity, including perception, problem solving, fear, and consciousness. The amplitude of these rhythm are very low and they are rarely seen in EEG recordings. However, detection of these rhythms are important as they can be used as the indicator for certain brain diseases. The frontocentral areas are the region with high EEG frequencies and the highest levels of cerebral blood flow (as well as oxygen and glucose uptake). γ rhythms appear to be involved in higher mental activity, including perception, problem

solving, fear, and consciousness.

Figure 1.3 shows the typical normal brain rhythms with their usual amplitudes levels. These rhythms may last as long as the state of the subject is not changed and therefore they are approximately cyclic in nature.

On the other hand, there are other brain waveforms and some of them are as follows [1]:

- Spiky type signals with a wide frequency range such as k-complexes, vertex waves (which happens during sleep), and some seizure signal.
- Transient signal such as Event Related Potentials (ERP) and contain Positive Occipital Sharp Transient (POST) signals.
- Signals that originate from the defected region of the brain, such as tumoural brain lesions.
- The sleep spindles within the 11 – 15 Hz frequency range.
- Eyelid flutter with closed eyes that originates the frontal artefacts in the alpha band.
- Lambda (λ) waves are most prominent in walking patients, but are not very common. They are characterised with sharp transients occurring over the occipital region of the walking subject during visual explorations.

Usually, detection and understanding of the brain rhythms from scalp EEGs, even with trained eyes, is a difficult job. However, with the help of powerful computers and application of advanced signal processing tools, separation and analysis of the desired waveforms from within the EEG signals has become feasible. We next consider the development in the EEG signal processing context, but before that some fundamentals about it are discussed.

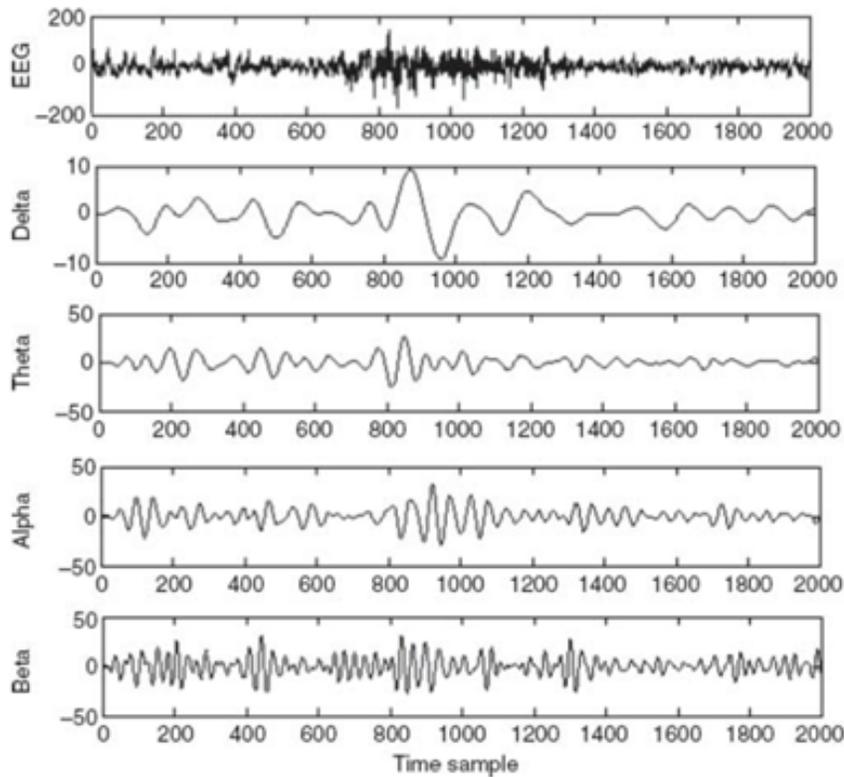


Figure 1.3: The EEG signal on the top and four typical dominant brain normal rhythms, from low to high frequencies; the delta wave is observed in infants and sleeping adults, the theta wave in children and sleeping adults, the alpha wave is detected in the occipital brain region when there is no attention, and the beta wave appears frontally and parietally with low amplitude [1].

1.2 Fundamentals of EEG Signal Processing

EEG signals are signatures of neural activity in the brain that are non-invasively captured from the scalp. These multichannel signals are presented in time domain but is believed to reflect the dynamical changes of the electrical activity in frequency and spatial domains as well. This allows processing of the EEG signals using conventional signal processing algorithms which are developed for analysis of the data in either time, frequency, space, multidimensional or multi-way domains. The fundamental concepts considered in developing EEG signal processing tools and algorithms are as follows [1]:

1.2.1 Nonlinearity of the medium

The head as a mixing medium combines and mixes EEG signals that are locally generated within the brain at the electrode positions. As a system, the output behaviour is directly influenced by its input sources in different situations. Various biological and physiological phenomenon can affect and change the brain metabolism, which may change the mixing process of head and subsequently, these effect can make EEG signal the output of a nonlinear system. Analysis and modelling of such system is very complicated and nobody has yet fully modelled it. However, chaos theory and analysis of dynamics of time series has helped researchers to characterise the nonlinear behaviour of EEG signals.

1.2.2 Nonstationarity

The signals are generally called nonstationary if the statistics of the signals, such as mean, variance, and higher order statistics, change with time. Consequently, we can quantify the nonstationarity of the signals by measuring some statistics of the signals at different time lags. If there is no considerable variation in these statistics, the signals can be considered stationary.

Generally, the multichannel EEG signals can be modeled as a multivariate Gaussian distribution. However, the mean and covariance properties of those models generally change from segment to segment. As a result, EEG signals are considered stationary only within short segments or intervals, that is, quasi-stationary. Furthermore, this Gaussian assumption holds during normal brain condition and it is not valid during mental or physical activities. The changes in the distribution of signal segments can be measured by the deviation of the distribution from Gaussian. Estimates of some higher-order moments, such as skewness and kurtosis can be checked as a measure of the non-Guassianity of the signal. Patterns observed during eye blinking, or during the transitions between various ictal states of seizure, and in the event-related potentials (ERP) and evoked potentials (EP) are some examples of non-stationarity in EEG signals.

1.2.3 Signal Segmentation

Effective assessment of EEG signals by neurophysiologists and clinicians necessitates labelling of the EEG signals into segments with similar characteristics. Within each segment, usually the signals have a similar time and frequency statistics and are considered statistically stationary. As an example, an EEG recorded from an epileptic patient may be divided into three segments of preictal, ictal, and postictal segments. Each may have a different duration.

In segmentation of EEGs time or frequency properties of the signals would be exploited to define a dissimilarity measurement between the adjacent EEG frames. Then, an empirical threshold usually determines whether those frames would belong to the same segment or not. Some of the criteria that are used as dissimilarity measure are based on autocorrelation function, high order statistics, error energy in autoregressive (AR)-modelling of the signals, and spectral error measure of the periodogram.

Although the above criteria can be effectively used for segmentation of EEG signals, In order to detect certain abnormalities better system can be used. By extracting meaningful and discriminative features that describe the behaviour of the signals, the segmentation problem can be translated to a classification problem where different classifiers can be used.

1.3 EEG Signal Analysis

Most traditional methods of signal analysis are based on the assumption of the ergodicity of the time series which requires stationarity of the signals. Physiological time series comply with that assumption only for very limited time intervals and quite often the evolution and variation of the signal in time is of primary interest. This is especially the case for brain electrical activity. According to the present understanding, the processing of information by the brain is reflected in dynamical changes of electrical activity in time, frequency, and space. Therefore, the study of these processes requires methods which

can describe the variation of the signal in time and frequency in a quantitative way.

EEG signal analysis is the art of decomposing the signal into functions with known Spatio-Temporal properties or at least easier to characterize. The conventional methods for EEG analysis can be categorized into the time domain, frequency domain, time-frequency, and spatio-temporal domains. Each of these methods analyse one or at most two of the contextual information embedded in EEG signals which we briefly discuss about in the next section.

1.3.1 Time Domain Analysis

Linear Prediction Models

Finding a set of model parameters that can optimally describe the signal generation system is the main objective of prediction models. These models usually use a noise type input. In autoregressive (AR) modelling of signals each sample of a single channel EEG measurement is defined to be a linear combination of a number of previous samples plus input sample, that is,

$$y[n] = - \sum_{k=1}^p a_k y[n-k] + x[n] \quad (1.1)$$

where $a_k, k = 1, 2, \dots, p$ are the linear parameters, n denotes the discrete sample time, and $x[n]$ is the noise input. In an autoregressive moving average (ARMA) linear predictive model each sample is defined as the linear combination of a number of its previous input and a combination of output sample values, that is,

$$y[n] = - \sum_{k=1}^p a_k y[n-k] + \sum_{k=0}^q b_k x[n-k] \quad (1.2)$$

where $b_k, k = 1, 2, \dots, q$ are the additional linear parameters. The parameters p and q are the model orders. The Akaike Information Criterion (AIC) is traditionally used to determine the optimum order of the model. The parameters of the model can be estimated using normal equations or via some iterative optimization techniques. In [14]

a parametric modelling of a single evoked potential is addressed.

Gaussian Mixture Model

Different parameters defined in the distributions characterises the signals in this approach. The distributions in terms of probability density functions are the sum of a number of weighted Gaussian functions with different means and variances. The overall distribution for K Gaussian components is defined as

$$p(x|\theta_k) = \sum_{k=1}^K \omega_k p(x|\mu_k, \sigma_k), \quad (1.3)$$

where the vector of unknown parameters $\theta_k = [\omega_k, \mu_k, \sigma_k]$ for $k = 1, \dots, K$ subject to $\sum_{k=1}^K \omega_k = 1$. μ_k and σ_k are the mean and variance of the k th Gaussian distribution function denoted by $p(x|\mu_k, \sigma_k)$. Expectation Maximization (EM) algorithm is often used for estimating the parameters of the model. Moreover, AIC or other more efficient model order selection techniques can also be used for selecting the optimum number of components K in the model [15]. In [16], the Gaussian mixture model is used for EEG pattern classification.

1.3.2 Frequency Domain Analysis

Since the invention of the Fourier Transform in early nineteenth century, which was considered a revolution in the signal processing field, along the temporal analysis signals have often been analysed in frequency domain more effectively. EEG signals behave differently for different states of the brain and as said earlier have distinctive frequency contents. So, it is worth analysing the EEG signals in frequency domain. However, the traditional windowed Fourier transform is not sufficient in this case. It is subject to high statistical errors, and it is severely biased as a consequence of the unfulfilled assumption that the signal is either infinite or periodic outside the measurement window. Parametric methods like the autoregressive (AR) model are free from the "windowing" effect since

no assumptions about the signal outside the measurement window are needed, but still, stationarity of the signal is required and signal structures of duration shorter than the measurement window cannot be identified. Moreover, parametric spectrum estimation of EEG signals required that an accurate estimate of the model order to be available, as overestimating or underestimating the model order may result in splitting the true peaks and combining peaks in close proximity in frequency spectrum, respectively.

EEG signals are often statistically nonstationary, particularly when an abnormal event is captured within the signals. As the Fourier transform integrates over the whole observed segment, the temporal information is lost in the process and the characteristics of the signals is not represented accurately. A time-frequency (TF) approach is the solution to the problem.

1.3.3 Time-Frequency Domain Analysis

Signal transformation plays a crucial role in the majority of signal processing applications. The goal of this transformation is to obtain much simpler representation of the signal. For instance, single tune signals are localized in spectral domain and have narrow frequency bands. Hence, they appear sparse in Fourier or cosine transform domains. On the other hand, data such as wideband noise with a flat spectrum data, are localized in temporal domain and have an impulsive behaviour in time. Hence, to more effectively and efficiently analyse the data, transformation becomes the main aspect of signal processing. When the characteristics of the signal is changing over time, we need to capture the non-stationarity with a transformation that considers both temporal and spectral contents. Time-frequency domain information exploits the data characteristics and properties in both time and frequency.

Short-Time Fourier Transform (STFT) is defined as the discrete-time Fourier transform evaluated over a sliding window in an effort to keep track of the temporal and spectral evolution of the analysed signal contemporary. However, due to windowing effect, perfect resolution can not be achieved in both time and frequency-domains.

On the other hand, Wavelet Transform (WT) and related methods such as Matching Pursuit (MP) have brought essential progress in this respect. Unlike the STFT, the time-frequency kernel for the WT-based method can better localize the signal components in timefrequency space.

The use of the wavelet transform in the study of brain signals was first applied by [17], [18], [19] to the evaluation of evoked potentials and also for the analysis of electrocortical activity. Evoked potentials are signals of amplitude an order of magnitude smaller than the ongoing EEG activity. The most commonly applied method of evoked potentials (EP) evaluation is based on averaging single EPs triggered by a repeating stimulus. This method relies on unrealistic assumptions concerning the purely deterministic and repeatable character of the EP, the purely stochastic character of the on-going EEG, and the independence of both signals. Wavelet analysis offers an alternative method of EP evaluation which is based on an important feature of the method-its ability to discriminate between signals lying in the same frequency range but with different temporal localizations. Wavelet analysis has proved to be useful in characterizing changes in brain activity, although in the study of time-locked phenomena such as EPs.

Using Wavelet transform, the time-frequency components characterizing the signals can be calculated and consequently an efficient and compact representation of the signal can be obtained. Usually a small number of wavelet components is sufficient to represent the signal with good fidelity which means often signals are sparse in wavelet domain. Nevertheless, wavelet analysis is itself subject to certain limitations connected with the fact that the bandwidth is inversely proportional to the time scale, which effectively limits the resolution. The bandwidth changes in steps (usually by a factor of two) and therefore we can not get precise frequency estimates especially at high frequencies. There have been different attempts to overcome this limitation. When the orthogonality assumption is relaxed, finer tuning of the time-frequency scale is possible; however, the computational procedures are much more complicated in this case.

Another drawback is that the wavelet representation depends also on the data window

setting and is sensitive to the time shift of the window. This limitation is not critical in the case of time-locked phenomena like EPs. In such cases the wavelet transform offers a universal and convenient parametrization which can be used for comparison of signals, but this kind of representation is much less suitable for transients which occur more or less randomly in the signal (e.g., spindles, K-complexes, etc.).

These limitations can be overcome by the MP technique, offering a maximally adaptive approach and very fine resolution close to the theoretical limit. In MP, from a large dictionary of possible functions, a subset is chosen in such a way as to match optimally the local signal structures. The windowed Fourier transform and wavelet transform (WT) can be considered as particular cases of MP corresponding to restrictions concerning the choice of parameters for the atoms of the dictionary. Each atom is characterized quantitatively by four parameters: time and frequency coordinates, time span, and intensity; in the case of MP all of them can change freely.

The density of the signal in the time-frequency plane can be visualized with MP in a form of Wigner distribution. However, due to the decomposition embedded in MP, TFR achieved by MP provides a clear picture in time-frequency space and does not have interference or cross-terms unlike the Wigner or Cohen class distributions.

Although MP seems to have interesting properties, it still has some problems. Greedy MP algorithms even for a simple combination of dictionary's functions can fail in certain cases, some of them addressed in [20]. One solution with a cost of increased computational requirements could be to use orthogonalized matching pursuit. However, this might also introduce numerical instabilities [21]. Besides, instead of using fixed basis for constructing the dictionary, incorporating the dictionary learning algorithms (DLAs) would enable us to learn the atoms of the dictionary in a data-driven and unsupervised way.

1.3.4 Spatio-Temporal Domain Analysis

Most physiological information, signals, and images are found as mixtures of more than one component. Unmixing and separating these components is then a challenge in this

context. However, if the number of recording channels is large enough and the constituting sources in the mixtures are either uncorrelated, independent, or disjoint in some domain, then we can accurately separate the components. This problem is known as the blind source separation (BSS) problem where some a priori, constraints such as independence of the sources, are considered to solve it. The term blind in BSS refers to the case of having both the sources and the mixing system (or medium) unknown. Principal component analysis (PCA), matrix factorization, independent component analysis (ICA) are different examples of the approaches that have been used in the source separation context.

ICA is the more popular one, as it decomposes the signals into the constituent independent components and plays a crucial role in de-noising and separation of biomedical signals. A perfect separation of the signals requires taking into account the structure of the mixing process. In real applications like EEG, this process is unknown but some assumptions can be made, that is the instantaneous case that the source signals arrive at the sensors at the same time. As a result, the BSS model can be written as

$$\mathbf{X} = \mathbf{H}\mathbf{S} + \mathbf{V} \tag{1.4}$$

where $\mathbf{S} \in \mathbb{R}^{M \times T}$, $\mathbf{X} \in \mathbb{R}^{N \times T}$, $\mathbf{V} \in \mathbb{R}^{N \times T}$ denote the matrix of source signals, observed signals, and noise, respectively. $\mathbf{H} \in \mathbb{R}^{N \times M}$ is the mixing matrix. N is the number of sensors, M is the number of sources, and T is the total number of measured samples. The separation is done using a demixing matrix $\mathbf{W} \in \mathbb{R}^{M \times N}$ that uses only the observed information to recover the independent sources according to $\mathbf{Y} = \mathbf{W}\mathbf{X}$.

The aim of BSS using ICA is to estimate an unmixing matrix \mathbf{W} such that \mathbf{Y} best approximate the independent sources \mathbf{X} . As such, the unmixing matrix can be calculated as the inverse of the mixing matrix, i.e. $\mathbf{W} = \mathbf{H}^{-1}$. This implies that separation is unique up to permutation and scaling ambiguities in the output independent components. BSS has been widely employed in the literature for separating normal EEG rhythms and also artefacts removal [22, 23]. Figure 1.4 illustrates the BSS concept for separation of brain

sources.

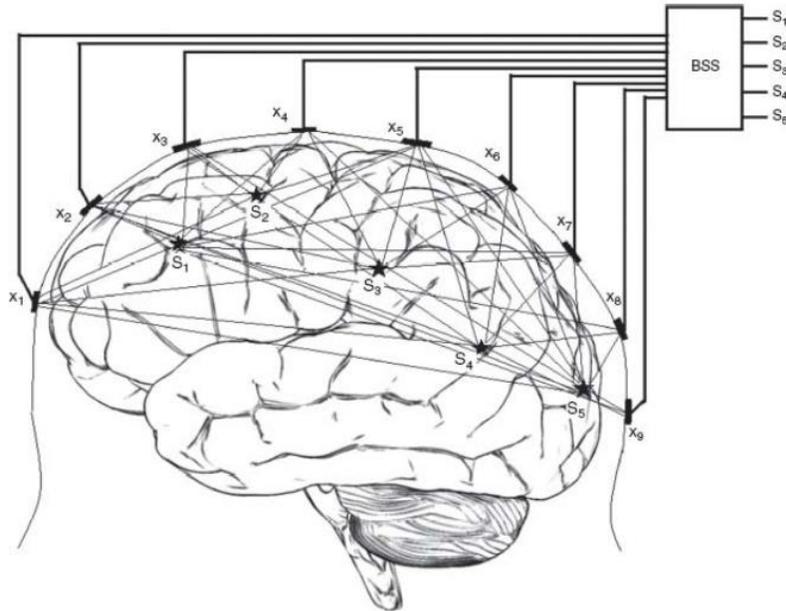


Figure 1.4: BSS concept; mixing and blind separation of the EEG signals.

Furthermore, the optimization underlying the solution of BSS problem may be subject to fulfilment of number of conditions, which could be based on a priori knowledge about the sources or the mixing matrix. These constraints could be statistical as well as geometrical, such as sparsity of sources or the mixing matrix, that can lead to more accurate solutions [24, 25]. In the context of EEG signal processing, although the number of signals mixed at the electrodes seems to be limited, the number of sources corresponding to the neurons firing at a time can be enormous. However, if the objective is to study a certain rhythm in the brain, the problem can be transformed to the time-frequency domain or even to the space-time-frequency domain. In such domains the sources may be considered disjoint and generally sparse. Having said this, our main goal in this dissertation focuses on high-dimensional techniques that achieve representative and discriminant features that potentially improve characterising and quantifying brain activities.

So far in this chapter, we have discussed the importance and challenges of EEG signal processing. Furthermore, we have introduced the conventional techniques, i.e.

temporal domain, spectral domain, TF analysis as a suitable signal representation that provides comprehensive information in signals and their non-stationary behaviour. Also, we have highlighted the popular spatio-temporal techniques. However, these techniques rely on at most two contextual domains in the EEG analysis, although, it is believed that processing of the information in brain is reflected in dynamical changes of the electrical activity in time, frequency, and space. Moreover, two-way analysis is unique subject to some additional constraints which may not be always physiologically meaningful. Study of the brain processes requires methods which can describe these variations of the EEG signal in time, frequency, and space in a quantitative way, which inspires the research of this dissertation. Therefore, the main focus of the present dissertation is to develop EEG signal processing framework that seek to decompose EEG into functions with known spatio-temporal-spectral properties or at least easier to characterize. Furthermore, the extracted components could be more discriminative, easier to interpret, and effectively applied for different applications.

1.4 Contributions of The Dissertation

This work presents a generalized multi-way array analysis methodology in pattern classification systems as related to source separation and discriminant feature selection and classification in EEG signal processing problems. Analysis of ERPs, as one of the main categories of EEG signals, requires systems that can exploit the variation of the signals in different contextual domains in order to reveal the hidden structures in the data. Multi-way analysis framework enables us to consider as much as relevant contextual information in our analysis simultaneously rather than sequentially adding the domains in the analysis which is done conventionally in the two-way analysis. Here in this thesis, the proposed framework via the blessing of the dimensionality of EEG signals, i.e. incorporating temporal, spectral, spatial, subjects/conditions modes in the analysis seeks to separate and localize the EEG sources that might have overlap in some domains. We

investigate three different implications in the proposed system as follows: i) EEG signal decomposition; ii) localizing to the events of interest; and iii) identification and classification of ERP signals. In all three implications, our main objective is to develop techniques that successfully quantify the patterns of interest in EEG signals. The proposed EEG-based framework could also be considered as the brain imaging system for poor people, as it has the good temporal resolution compared to the costly and conventional brain imaging fMRI systems, and still produces a satisfactory and acceptable spatial resolution that could be used for visualizing and analysing the ERP signals.

We use three different EEG datasets in this dissertation as following:

- The multichannel SSVEP dataset from the Swartz Center for Computational Neuroscience (SCCN) at University of California San Diego (UCSD), which consisted of ten healthy adults (8 males and 2 females, mean age: 23 years) with normal or corrected-to-normal vision.
- The multichannel EEG dataset of five healthy male subjects (mean age: 26.2 years old; standard deviation: 2.3) that were recorded in the BCI lab of the Federal University of Espirito Santo in Brazil.
- The multi-channel ERP dataset that was recorded in Auditory Development Lab at McMaster University. 22 participants (17 – 22 years old, mean age 18.93 ± 1.39), neurologically healthy, right-handed, and with normal hearing participated in the experiments.

Every EEG processing system consists of three main parts: Signal Acquisition, Signal Processing, and Translation part. The block diagram in Figure 1.5 shows the overview of the proposed framework. In this block diagram, our contributions' areas are highlighted with the darker color as explained below:

In order to capture the non-stationarity of EEG signals, analysing the temporal and spectral domains concurrently is inevitable which requires TFR techniques that can characterise EEG data accurately. TFR is considered as a pre-processing step in the system

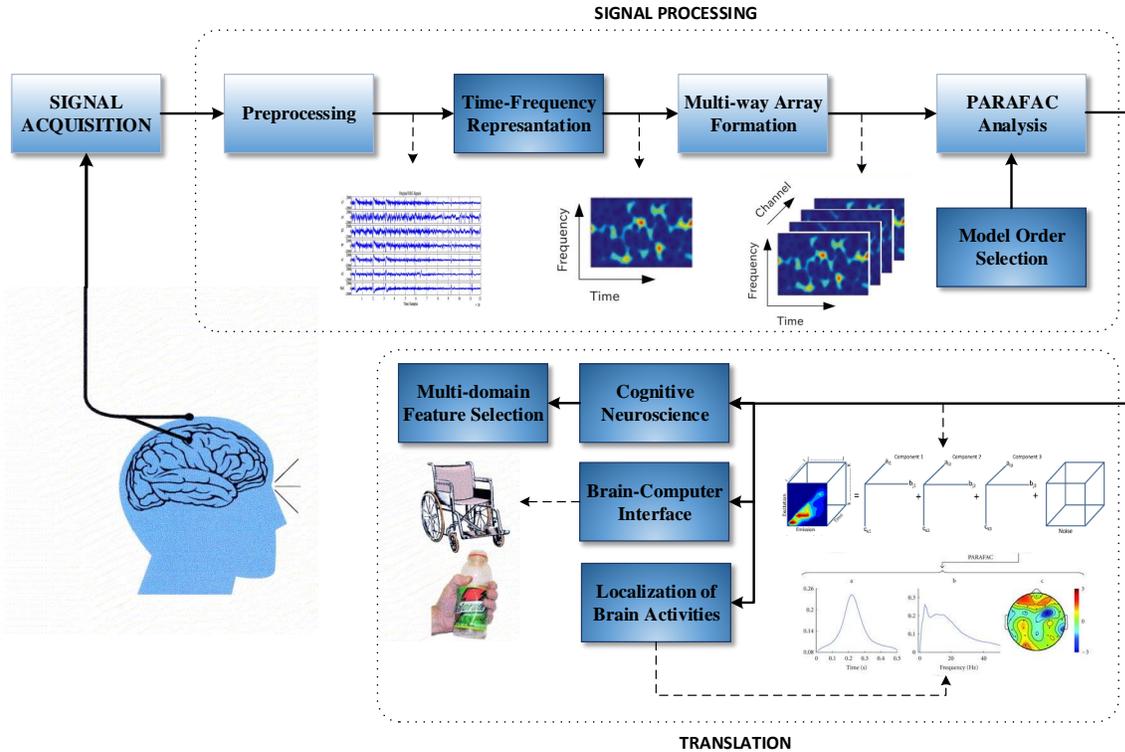


Figure 1.5: Flowchart of the proposed methodologies highlighting the original contributions with the darker color.

to help with better localization/analysis of the events. Having a coarser TFR is okay in some cases but finer TFR helps with ERPs as these are time locked phenomenon that require higher resolution for an effective analysis. Furthermore, as EEG is naturally high-dimensional, the data collected can also be fitted into the multi-way array including multiple modes. Then, the interactions among multiple modes can be revealed via multi-way analysis. Hence, the signal processing part of the system requires methodologies that can characterise and model different interactions among various EEG modes. To fulfill this objective, we exploit the followings:

1. Time-Frequency Representation (TFR): We intend to generate a high-resolution and cross-term free time-frequency representation that increases the effectiveness of localization in EEG signals and captures the non-stationarity and non-linearity within the EEG signal.

2. CANDECOMP/PARAFAC model order selection: Based on the above TFR technique, we form the EEG multi-way array by adding other relevant modes, and our goal is to develop a unique and novel model order selection technique for decomposing the EEG multi-array into its constituting components. The extracted components uniquely characterize the embedded events in the analysed EEG segment and can be used further for various exploratory and classification applications.

In the next step, we intend to translate the extracted meaningful and unique features from the EEG data into different applications. To make this happen, once the multi-way array of EEG signals are formed, based on the contextual modes incorporated in the formation of the EEG tensor, we introduce three applications for multi-modal event-related potentials analysis in a way that they are categorized into individual-level and group-level analysis:

Individual-level Analysis

Three-way array of EEG signals are decomposed and the extracted signatures are used for the following applications:

1. Source localization of brain activities.
2. Brain-Computer Interface.

Additionally, the developed method makes sure that the obtained features are robust to noise and outliers and are effective for classification and detection of the event of interest.

Group-level Analysis

Four-way array of EEG signals, including the mode of subjects, is decomposed for research questions of cognitive neuroscience. The extracted multi-domain feature is used for group-level analysis of ERP signals. We introduce a novel multi-domain feature selection technique in a way that it represents the ERP characteristics and differentiates the two groups of subjects/conditions in the experiments with a proper significance level.

Outcome of the proposed work

Table 1.1, summarizes the various solutions provided by the proposed data-driven signal processing framework in efficiently analyzing non-stationary EEG signals and localizing the events and extracting discriminative features from them. The proposed framework with multi-way analysis as its significant highlight is expected to become a versatile non-stationary signal analysis tool, which has the benefits of localization and discriminant analyses. The outcome of the proposed work as an EEG processing system, including the data acquisition, signal processing, and the translation stages could be summarized as following:

Table 1.1: Summary of the proposed solutions and the requirement for efficient EEG signal analysis

Requirements for efficient EEG signal analysis and feature extraction	Solution provided/ suggested by the proposed work	Chapter Reference
EEG characteristics	Capture non-stationarity with TFR	Chapter 1
TF analysis	Data-driven and high resolution TFR	Chapter 2
Multi-way nature of EEG	Multi-way Analysis	Chapter 3
Capture the true underlying structure of EEG signals	CANDECOMP/PARAFAC Model Order Selection	Chapter 4
EEG tensor formation	Multi-channel Reassigned-EMD	Chapter 5
Localization of Brain Activities	Three-way EEG tensor analysis	Chapter 5
Brain-Computer Interface	PARAFAC-based feature extraction	Chapter 5
Multi-domain feature selection	Four-way EEG tensor analysis	Chapter 6

1.5 Organization of The Dissertation

The dissertation is organized in 7 chapters. The flowchart in Figure 1.6 displays the evolution of this dissertation. We begin with identifying the right TFR technique for the proposed work that best suits the non-stationarity of EEG data and results in a high-resolution and cross-term free TFR. In Chapter 2, the classical TF transformations and the proposed TFR method are explained. Time-Frequency analysis is effective for

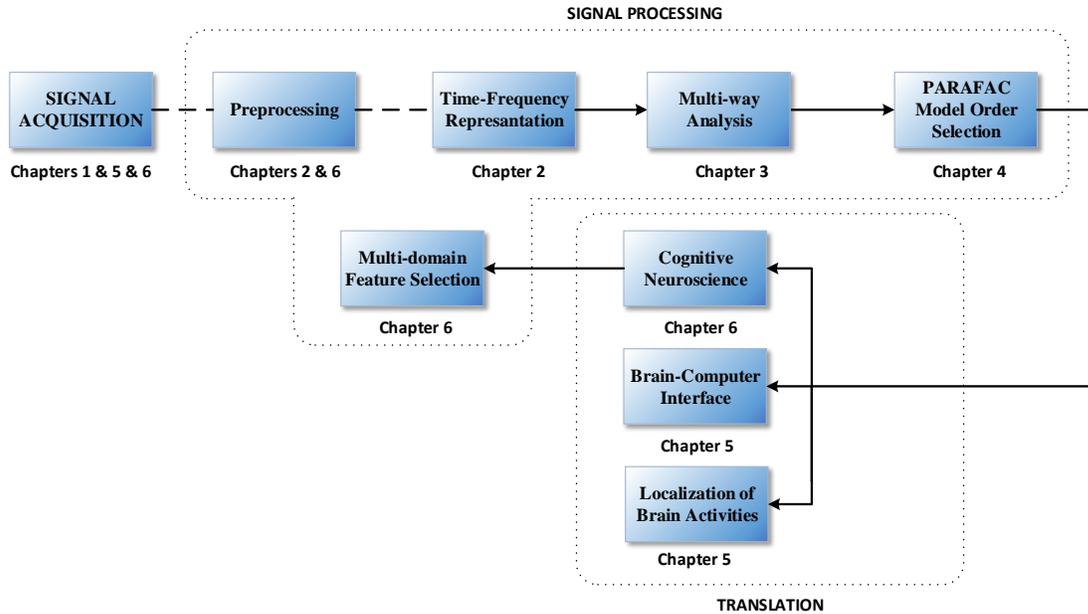


Figure 1.6: Organization of the dissertation.

EEG analysis and has been widely used in the literature, However, it may not be able to extract and separate the overlapping events in EEG. Multi-way analysis techniques, on the other hand, by retaining the additional dimensions, keep the structure of the multi-dimensional data and may reveal the hidden structure that two-way analysis cannot do. The proposed TFR in this chapter is later used for the formation of multi-way EEG tensors. In Chapter 3 multi-way arrays and the corresponding multi-way analysis techniques are introduced. CANDECOMP/PARAFAC and TUCKER are two well-known models in the multi-way analysis context. We select the CANDECOMP/PARAFAC model for our EEG analysis due to its uniqueness property and more interpretable decomposed signatures. Selecting the optimum number of components in CANDECOMP/PARAFAC model is an active challenge in the community. A novel and robust methodology for CANDECOMP/PARAFAC model order selection is developed in Chapter 4. The method is capable of handling both white and colored noises and outperforms the state-of-the-art techniques in the literature. In Chapter 5, two individual-level applications for three-way EEG tensor with modes: $time \times frequency \times channel$ are presented. The first

application is for localizing the brain activities and extracting the event of interest from a mixture of events in multi-channel EEG recordings. The second application is to use the extracted signature from CANDECOMP/PARAFAC analysis of EEG tensor for a brain-computer interface system. The extracted signature can effectively characterise the embedded events in EEG signals and further be used for detection and classification of ERP signals. In Chapter 6, a group-level analysis for four-way EEG tensor with modes: *time × frequency × channel × subjects/conditions* via CANDECOMP/PARAFAC model is presented. Furthermore, in this chapter a novel multi-domain feature selection scheme is developed, that not only characterizes the ERP signal via its coupled temporal and spectral signatures, but also can effectively differentiate the two groups of subjects/conditions with an acceptable significant level. Chapter 7 presents a summary of the complete work with analysis of the achieved results at various stages. The novelty and the multifold benefits of the proposed work are highlighted. The future directions on enhancing this methodology are presented as well.

Chapter 2

Time-frequency Representation

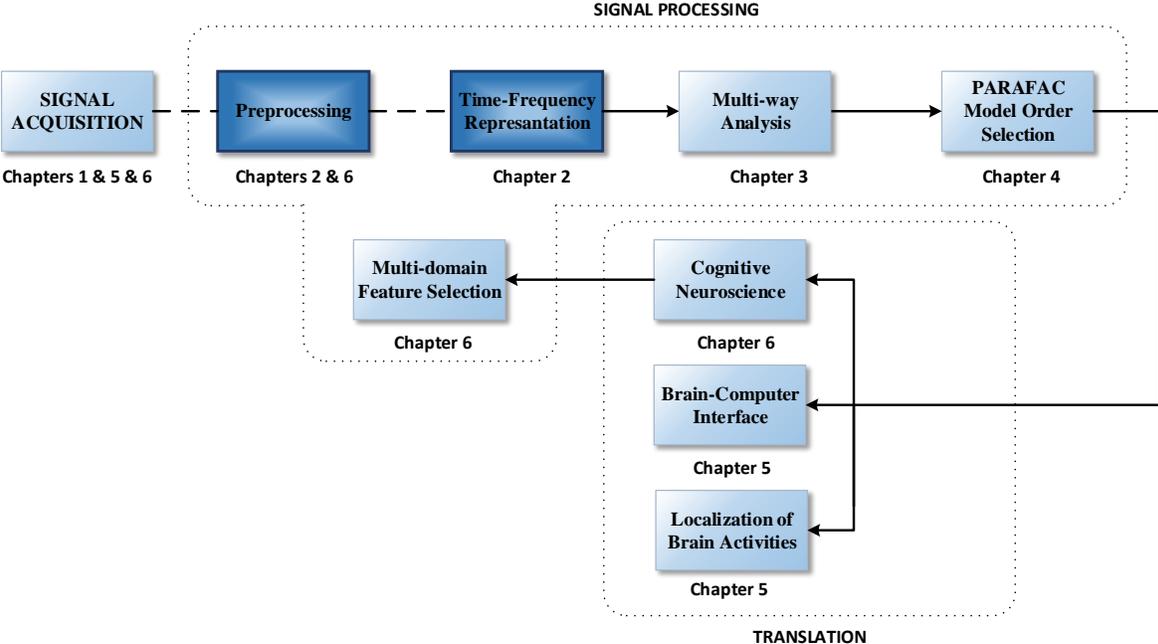


Figure 2.1: Chapter 2- Time-frequency Representation.

EEG signals are non-stationary, and they therefore can not be efficiently represented in the time domain, $x(t)$. Fourier representation, $X(f)$, reveals spectral constitutive features of the signal, but it does not preserve any explicit localization in time. It is well-known that Fourier representation faces limitations when we are looking for non-stationarity

features of the signal. Hence, when it comes to analyze signals with time-varying characteristics, the time-domain and frequency domain analysis tools are not sufficient. To overcome this difficulty and to analyze the non-stationary signals effectively, techniques that provide joint time and frequency information are needed. Joint TF representation (TFR) indicates a two dimensional energy representations of a signal in terms of time and frequency domains. The work in the area of TFR methods is extensive [26–29]. Depending upon the application in hand and the feature extraction strategies, any of the TF approaches could be used. In this chapter, first we review the classical TF representation techniques and highlight their shortcomings and then propose an adaptive TF representation scheme that is suitable for EEG analysis and results in a high-resolution and cross-term free TFR.

2.1 Classical Time-frequency Distributions

Figure 2.2 displays the most well-known TF transformation techniques. Any of these techniques transform a temporal signal, $x(t)$, into a TF distribution denoted with $\mathbf{V}(t, f)$, where t and f locate each sample in the TF plane and $\mathbf{V}(t, f)$ is the TF value at the corresponding location. Such transformation is displayed in Figure 2.3. Depending on the transformation technique, TF distributions with different properties are achieved. For example, some methods construct distributions with non-negative entries while some might result in negative values also. Temporal and Spectral marginals of a TFD are

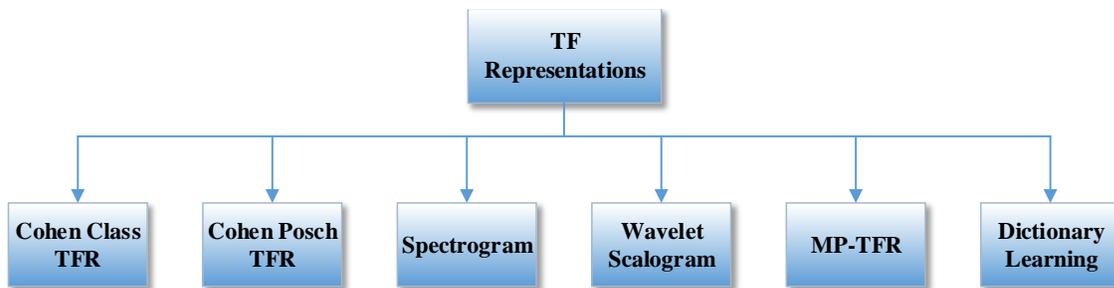


Figure 2.2: A diagram of well-known TF distributions.

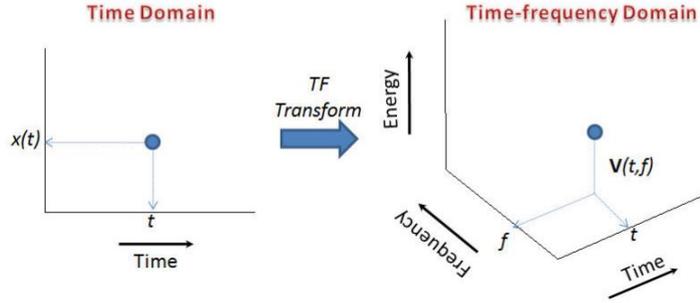


Figure 2.3: A diagram of TF transformation.

calculated along each time and frequency coordinates as shown below:

$$\begin{aligned}
 TM(t) &= \int_{-\infty}^{+\infty} \mathbf{V}(t, f) df, \\
 SM(f) &= \int_{-\infty}^{+\infty} \mathbf{V}(t, f) dt.
 \end{aligned} \tag{2.1}$$

2.1.1 Cohens Class Bilinear TFRs

Due to the windowing embedded in the TFR, perfect resolution can not be achieved in both time and frequency-domains. In an attempt to reduce this effect, quadratic methods of TFR will adapt the analyzed signal as the analysis window, i.e. in order to obtain the 2D representation of the distribution of signal energy over time and frequency, the quadratic TFR incorporates the time varying autocorrelation of the signal as:

$$\mathbf{V}_{WV}(t, f) = \int_{-\infty}^{+\infty} x(t + \tau/2)x^*(t - \tau/2)e^{-i2\pi f\tau} d\tau. \tag{2.2}$$

where, \mathbf{V}_{WV} is Wigner-Ville distribution (WVD) of the signal.

2.1.2 Cohen-Posch TFR

Cohen-Posch TFR, or positive TFR (PTFR), produces a TFR with non-negative entries. Cohen and Posch [30] showed that there is an infinite set of positive TFRs and derived

a formulation via signal dependant kernels to compute the positive TFRs as follows:

$$\mathbf{V}_{PTFR}(t, f) = |x(t)|^2 |X(f)|^2 \{1 + c \cdot \rho(s(t), S(f))\}, \quad (2.3)$$

where

$$s(t) = \int_{-\infty}^t |x(\tau)|^2 d\tau; S(f) = \int_{-\infty}^f |X(\xi)|^2 d\xi, \quad (2.4)$$

and

$$\rho(s(t), S(f)) = h(s(t), S(f)) - h_1(s(t)) - h_2(S(f)) + 1. \quad (2.5)$$

In the above equation, $h(s, S)$ is a positive kernel function of the variables s and S , $0 \leq s, S \leq 1$ and normalized to one. $h_1(s)$ and $h_2(S)$ are the marginals of $h(s, S)$ (defined in (2.1)), and c is a numerical constant in the range of

$$\frac{1}{\max(\rho(s(t), S(f)))} \leq c \leq \frac{1}{\min(\rho(s(t), S(f)))}. \quad (2.6)$$

2.1.3 Spectrogram

Linear TF analysis decompose the signal over a set of basis functions. The simplest linear TF representation is short-time Fourier transform (STFT) of signal, which is based on the stationarity assumption of the signal in short durations and takes the Fourier transform of the windowed segments after multiplying the signal by a window. The basis functions used in Fourier transform are orthonormal cosine functions with varying frequencies. Given a signal $x(t)$ and a window $h(t)$, the STFT of $x(t)$ is defined by

$$\mathbf{V}_{STFT}(t, f) = \int_{\mathbb{R}} x(\tau) h^*(\tau - t) e^{-i2\pi f\tau} d\tau \quad (2.7)$$

where $h^*(t)$ is the complex conjugate of $h(t)$. The spectrogram $\mathbf{S}_{STFT}(t, f)$ is then usually defined as $|\mathbf{V}_{STFT}(t, f)|^2$ and generally used to display the TF energy distribution over the TF plane.

2.1.4 Wavelet Scalogram

Wavelet scalogram is based on wavelet decomposition where orthonormal basis functions with different sizes are used to decompose a signal as given by the following equation:

$$\mathbf{V}_{CWT}(t, s) = \frac{1}{\sqrt{s}} \int x(\tau) g\left(\frac{\tau - t}{s}\right) d\tau, \quad (2.8)$$

where $g(\frac{t}{s})$ is the mother wavelet, and s being the scaling parameter, corresponds to the size of each basis function. The basis function used in wavelets, which are called mother wavelets, are small waves satisfying certain mathematical conditions. Various scaled versions of the mother wavelet are created by expanding and compressing the mother wavelet. These different scaled versions of the mother wavelet are slid over the signals, and captures and models the localized signal structures and characteristics.

We need to find $\mathbf{V}_{CWT}(t, f)$. There is no precise relation between scale and frequency. However, an approximate answer for the relationship between scale and frequency is associated to the centre frequency of the wavelet. Having the centre frequency of the wavelet, F_c , determined, we can relate the scale to frequency as following:

$$F_s = \frac{F_c}{s \cdot \Delta} \quad (2.9)$$

where Δ is the sampling period, F_c is the center frequency of a wavelet in Hz, and F_s is the pseudo-frequency corresponding to the scale s , in Hz. The idea is to associate a purely periodic signal of frequency F_c with a given wavelet which is the frequency F_c , maximizing the Fourier transform of the wavelet modulus.

Wavelet scalogram displays the TF structure obtained from the wavelet transform. In scalogram, each wavelet signal is plotted as a filled rectangle whose its location and size are related to the time interval and the scale range for this wavelet signal. Center frequency of the wavelets are changed by the scaling parameter that expands and compresses the wavelets. Small scale factors corresponds to higher frequencies and larger scale factor corresponds to the lower frequencies. In other words, in order to capture

high-frequency characteristics, wavelets use short time scales and on the contrary, a long time scale is used to capture the low frequency structures in a signal.

2.1.5 Matching Pursuit TFR

Matching pursuit (MP) TF representation is generated based on MP decomposition as proposed in [31]. MP uses an over-complete, redundant, and non-orthogonal basis functions, including all possible translations, modulations and scalings as shown in the following equation:

$$G_{\gamma_i}(t) = \frac{1}{\sqrt{s_i}} g\left(\frac{t - p_i}{s_i}\right) \exp[j(2\pi f_i t + \phi_i)]. \quad (2.10)$$

where $g(t)$ is the primary Gaussian function, and $G_{\gamma_i}(t)$ is a basis function generated from the primary waveform. The scale factor s_i , and the parameter p_i controls the width and the temporal placement of the basis function, respectively. The parameters f_i and ϕ_i are the frequency and phase of the basis function, respectively. A particular combination of the TF decomposition parameters (s_i, p_i, f_i, ϕ_i) is denoted by γ_i . The set of all the possible TF basis functions is called the redundant TF dictionary and each member of this collection is called a TF atom. The term redundant indicates that the number of basis functions in the TF dictionary is much larger than the minimum number of required orthonormal basis functions to completely decompose a given signal. Selection of the dictionary and the basis functions is done based on the specific application in hand. The Gabor dictionary (i.e., sinusoidal oscillations modulated by Gaussian envelope) is especially well suited to describe most of EEG structures and is widely used for the processing and representing of EEG signals in the literature and has shown to offer the best TF localization properties [26], [32], [33].

A real Gabor function can be expressed as:

$$g_\gamma(t) = K(\gamma) e^{-\pi\left(\frac{t-u}{s}\right)^2} \sin\left(2\pi\frac{\omega}{N}(t-u) + \phi\right) \quad (2.11)$$

where N is the size of the signal for which the dictionary is constructed, $K(\gamma)$ is such that $\|g_\gamma\|=1$. $\gamma = \{u, \omega, s, \phi\}$ denotes the parameters of the dictionary's functions (time-frequency atoms). These parameters form a continuous three-dimensional space which yields a dictionary with an infinite number of atoms. Hence, in practice, we usually use a subset of the possible dictionary's functions. The price we pay for this sub-sampling is that the resulted decomposition, due to the particular scheme of sub-sampling the parameter space, is biased. In other words, any fixed sub-sampling scheme of the parameter space in order to form the practical dictionary for MP decomposition, introduces statistical bias in the resulting parametrization. To reduce this effect MP with stochastic dictionaries are used, where the parameters of a dictionary's atoms are randomized before each decomposition.

MP decomposes a signal, $x(t)$, into a linear combination of TF functions $G_{\gamma_i}(t)$ selected from a redundant Gabor dictionary of TF basis functions as given in the following equation:

$$x(t) = \sum_{i=1}^I a_{\gamma_i} G_{\gamma_i}(t) + R_x^I, \quad (2.12)$$

where $x(t)$ being the signal, $a_{\gamma_i} = |\langle R_x^I, G_{\gamma_i}(t) \rangle|$ is the expansion coefficient on $G_{\gamma_i}(t)$, and R_x^I is the residue signal after I iteration.

The signal $x(t)$ in (2.12) is projected over a redundant dictionary of TF functions with all possible combinations of scaling, translations and modulations. At each iteration, a TF function from the Gabor dictionary with the highest correlation is selected. After subtracting the projected part from the signal, the remaining signal called the residue was further projected on the dictionary in the same way at each iteration. After I iterations, signal $x(t)$ could be expressed as the summation in (2.12), where the first part of (2.12) is the summation of selected TF functions until I iterations, and the second part is the residue which can be further decomposed in the subsequent iterations. This process is repeated till all the energy of the signal is decomposed or the energy of the residue is below a certain threshold. At each iteration of the MP, a portion of the signal energy was

modeled with an optimal TF resolution in the TF plane. However, after some iterations, we can see that the coherent and incoherent structures of the signal are captured and separated into the decomposed part and the residue (R_x^I) in (2.12). Consequently, as the residue does not show any TF localization to be decompose after high enough iterations, which depends on the nature and length of the signal, the decomposition residue in (2.12) can be assumed to be due to random noise and be ignored.

After MP selected the collection of TF atoms that accurately model the signal $x(t)$, MP-TFR of the given signal, $\mathbf{V}(t, f)$, is constructed by summing the TFR of each decomposed TF atom as shown below:

$$\mathbf{V}(t, f) = \sum_{i=1}^I |a_{\gamma_i}|^2 \mathbf{WVG}_{\gamma_i}(t, f), \quad (2.13)$$

where $\mathbf{WVG}_{\gamma_i}(t, f)$ is the WVD of the Gabor atom $G_{\gamma_i}(t)$.

Although MP seems to have interesting properties, it has its own problems and drawbacks. In certain cases, the greedy MP may fail to properly decompose a signal containing even a simple combination of dictionary's functions. For some theoretical examples of failure of MP refer to [20]. However, the use of orthogonalized matching pursuit can help us resolve some of the cases, with an increased computational cost and a possibility of introducing numerical instabilities at its side effects [21].

2.1.6 Dictionary Learning algorithms for TFR

Fourier and wavelet dictionaries due to the well-defined mathematical bases, has been widely used for spectral analysis of signals, although they are not flexible enough to represent the shape diversity of EEG patterns. Furthermore, the Gabor dictionary because of its temporal shift-invariance property has also become popular. Nevertheless, it also suffers from a lack of flexibility to represent evoked potentials and EEG bursts. For example, a custom-based dictionary approach is preferred over Fourier or wavelets dictionaries when it comes to representing repeatable and complexly shaped epileptic

activities such as inter-epileptic peaks [34, 35]. In EEG analysis, the spatial modelling has been the focus of various studies by considering the inter-channels links via complex atoms and multichannel decomposition. The EEG temporal modelling is more difficult. Some approaches assume the temporal stationarity and treat only the spatial aspect, which brings about ambiguity. Generic Gabor dictionary is used by other approaches as it is shift-invariant. But it remains difficult to learn an EEG dictionary that integrates both temporal and spatial aspects together.

Recently, dictionary learning algorithms (DLAs) via a data-driven and unsupervised way have enabled us to learn dictionary atoms. It is essentially a set of iterations alternating between sparse approximation and dictionary update that results in learned atoms. The learned atoms are no longer generic but instead are data-driven and adapted to the studied data. Thus, learned dictionaries outperform generic ones for processing of the data that does not comply with the assumptions of the generic dictionaries. A recent one [36], for EEG data via a multivariate and shift-invariant temporal dictionary learning approach has been shown to outperform the Gabor dictionaries in terms of their sparse representative power; i.e. the number of atoms necessary to represent a fixed percentage of the EEG signals.

2.2 Review and shortcomings of TFR Methods

So far in this chapter, we have introduced the classical TFR techniques that have been used for EEG signal analysis in the literature. Now, we highlight the shortcomings of these methods, and then in the next section our proposed technique is presented. A TFR, $\mathbf{V}(t, f)$, that is non-stationary compatible and could be used for extraction of meaningful features should satisfy the following properties [37]:

- A desirable TF transformation provides a high time and frequency resolution. Therefore, the effectiveness of any TFR lies in how well it can transform the signal on to a TF plane with optimal TF resolution. The ideal case would be to have both

time and frequency resolution as high as possible. However, due to the Heisenberg uncertainty principle, high resolutions in both time and frequency domains cannot exist simultaneously. According to the Heisenberg uncertainty principle [38], the TF resolution has to satisfy the condition $\sigma_t \sigma_f \geq \frac{1}{2}$, where σ_t and σ_f are the respective time width and frequency width of the TF structure.

- It is invariant to time shift or amplitude scale in the signal. As in a sensitive TFR to shift, any transformation in the signal would result in a complete change in the structure of the TFR and the extracted TF-based features. Such TFR cannot satisfy the translation invariance property required for the features. Therefore, it is essential for a TFR to follow the same translations as in the signal rather than providing a completely new TF transformation.
- The suitable TF representation provides non-negative TF values:

$$\mathbf{V}(t, f) \geq 0 \tag{2.14}$$

In order to produce meaningful features, the value of the TFR should be positive at each point; otherwise it would be difficult to explain and interpret the extracted features. For example, it is possible to get a negative mean in a negative TFR at a given time, which means that the instantaneous frequency is also negative. In real-world applications, presence of negative energy or negative instantaneous frequency cannot be interpreted [31].

- Satisfies correct time and frequency marginals:

$$\begin{aligned} \int_{-\infty}^{+\infty} \mathbf{V}(t, f) df &= |x(t)|^2, \\ \int_{-\infty}^{+\infty} \mathbf{V}(t, f) dt &= |X(f)|^2. \end{aligned} \tag{2.15}$$

where, $\mathbf{V}(t, f)$ is the TFR of signal $x(t)$ with Fourier transform of $X(f)$. The TFR

which satisfies the non-negativity and marginal criteria is called positive TFR [30]. A high-resolution estimate of joint TF distribution of the signal can be achieved via a positive TFR with correct marginals. A TFR like that is a suitable TF representations for analysis of non-stationary signals and provides a high TF localization of the signal energy.

As mentioned in this chapter, several TFD methods exist; however, not all the methods are nonstationary compatible, or are suitable TF representations for non-stationary feature extraction purposes. WVD offers high TF resolution; however, the WVD contains the so-called interference terms or cross-terms. The presence of cross-terms in the TFR, that do not belong to the signal, would result in incorrect interpretation of the signal properties. Pseudo Wigner-Ville distribution (PWVD), smoothed pseudo Wigner-Ville distribution (SPWVD), Choi-Williams distribution (CWD) and Cohen kernel distribution are examples of other developed TFRs in the literature with the motivation of removing these cross-terms by defining a kernel in ambiguity domain. These distributions belong to a general class called the Cohen's class of bilinear TF representation. These TFRs satisfy time and frequency marginals; however, the distributions do not always satisfy the non negativity constraint.

Cohen-Posch TFR, or positive TFR (PTFR), produces non-negative TFR of a signal that does not contain any cross terms. Even though PTFR successfully constructs a positive and high resolution TFR of a given signal, this method cannot be implemented in most cases. In order to calculate positive kernels, the method requires the signal equation which is not usually known. Therefore, although the existence of PTFRs is proven, their derivation process is too complicated to be considered in most of the applications.

In STFT, due to windowing effect, resolution is limited by the Heisenberg's TF uncertainty and perfect resolution can not be achieved in both time and frequency-domains. It suffers from TF resolution trade off; when time is short, frequency resolution is coarse, and vice-versa. On the other hand, Wavelet Transform (WT) and related methods such as Matching Pursuit (MP) have brought essential progress in this respect. Compared to

the STFT, WT can better localize the signal components in time-frequency space due to the adaptive varying time width defined by the scaling parameter of the mother wavelet in contrary to the fixed time-width window in STFT. Additionally, scalogram provides a positive and cross-term free TF representation. Nevertheless, wavelet analysis is itself subject to certain limitations connected with the fact that the bandwidth is inversely proportional to the time scale, which effectively limits the resolution. The main drawback of the scalogram is its poor temporal resolution at low frequency regions of the TF plane and poor spectral resolution at high frequencies. Therefore, scalogram cannot efficiently display TFR of signals containing components with short durations and low frequencies, or vice versa. We can tackle this drawback and achieve finer tuning of the time-frequency scale by relaxing the orthogonality assumption of the wavelets. There have been different attempts to overcome this limitation. However, the computational cost of this procedure would be increased a lot. Another drawback is that the wavelet representation is not shift-invariant and is sensitive to the time shift of the window. These limitations can be overcome by the MP technique, offering a maximally adaptive approach and very fine resolution close to the theoretical limit. It uses an over-complete dictionary of atoms where each atom is characterized quantitatively by four parameters: time and frequency coordinates, time span, and intensity that can change freely. The windowed Fourier transform and WT can be considered as special cases of MP by restricting the choice of parameters for the atoms of the dictionary.

By virtue of the energy conservation, MP allows us to conveniently visualize its density in the time-frequency plane in the form of a Wigner distribution. Furthermore, unlike the Wigner or Cohen class distributions, MP representation is free of cross-terms and result in a clear picture in time-frequency space. As explained earlier in this chapter, although WVD distribution is a powerful TF representation with high TF resolution, in case of multi-components signals, the TF resolution will be affected by cross-terms. However, when WVD is applied to single components, their summation is a cross-term free TFR. Since practical signals are composed of several components, the cross-term

Table 2.1: Desirable TFR Properties for TF Quantification

Property	Positivity	Marginals	Time-Frequency Resolution	Time-Frequency Localization	Practical Ability
WVD	✗	✓	✓	✗	✓
PTFR	✓	✓	✓	✓	✗
Spectrogram	✓	✗	✗	✗	✓
Scalogram	✓	✗	✗	✗	✓
MP-TFR	✓	✗	✓	✗	✓

free MP-TFR provides a TF representation adaptive to the TF structure of a given signal. Although MP seems to have interesting properties, it still has some problems. The constructed MP-TFR did not satisfy temporal or spectral marginals. Greedy MP algorithm take time for representation and in certain cases can fail to properly decompose signal containing even a simple combination of dictionary's functions. Moreover the signal decomposition is biased by the dictionary selection and sampling exploited in MP-TFR, which means different applications may require different dictionaries. There were efforts to optimize the MP-TFR to construct a positive, high resolution and cross term free TFR that satisfies the marginal criteria, although the computational cost of the approach increased dramatically. These TFR techniques are called adaptive TFR as it is constructed according to the properties of the signal being analyzed [39]. The properties of different TFR techniques are summarized in Table 2.1. Dictionary learning algorithms are good for sparse representation of the signals, however it may not end up to mono-component decomposition and there is no guarantee that the extracted components to be spectrally localized.

Having said the shortcomings of the classical TFR methods, we want to propose a fully data-driven and high-resolution TFR methodology, that satisfies all the TFR requirements and could be effectively used for EEG signal analysis. As EEG is a multi-component signal and non-stationary and we seek to characterize and decompose it into its constituting factors, it is crucial to have a TFR that is cross-term free and highly localized in time and frequency domains concurrently.

2.3 TFR based on Empirical Mode Decomposition and Reassignment method (RS-EMD)

Here, we propose to use the approach that transform the EEG segment into TFR with a high temporal and spectral resolution, which to the authors knowledge is novel in EEG signal processing context. The method is a combination of Empirical Mode Decomposition (EMD) and Reassignment method, which we call it RS-EMD afterwards. EMD is a fully adaptive decomposition technique that is useful for analyzing natural signals, which are most often non-linear and non-stationary and multi-component. EMD has been widely used for EEG analysis; see [40] for a comprehensive review of EMD applications. EMD decomposes the EEG signal into the spectrally localized components, which is desirable here as we could further reduce the interference of different components of EEG in TF representation. Hence, with taking TF transformation of the resulting components after EMD and adding them up, we can calculate the overall TF spectrum of the EEG segment. The idea of sequentially combining the EMD and the TFR of the extracted components is not novel and has been used previously for different applications. For example, combination of EMD and Hilbert transform, known as Hilbert-Huang Transform (HHT), that captures the instantaneous frequency of the extracted IMF via Hilbert transform has been used previously for EEG signal analysis [41]. However, HHT suffers from high variance and usually an average of several segments should be considered for analysis. Besides, combining the EMD with Choi-Williams Distribution was proposed in [42] to suppress the cross-terms interference in the Cohen class TFR. On the other hand, Time-Frequency reassignment [43] can be viewed as a postprocessing technique aimed to overcome the localization and interference trade-off that is usually observed in classical TF representations of multi-component signals. Combining EMD with the reassignment method has been independently used before for assessment of pain expression in infant cry signal [44] and flue gas turbine signal [45]. However, those works are not widely available and combining EMD and Reassignment method is novel in the context

of EEG signal processing and tensor analysis and seems to suit more readily for ERP analysis. Hence, we propose to use the Reassignment method to calculate the TFR of extracted components after empirical mode decomposition of EEG signals. Finally, by adding up the reassigned spectrogram of each of the EMD-based extracted components, via RS-EMD we can get a high resolution and cross-term free TF representation for each of the EEG segments.

In parallel with the Reassignment method another postprocessing technique for enhancing the readability of the TFR was developed and named "synchrosqueezing" which is a special case of reassignment with the additional advantage of allowing for reconstruction and mode retrieval [46]. The SynchroSqueezing Transform (SST) is a post-processing technique applied to the continuous wavelet transform in order to generate localized time-frequency representations of nonstationary signals. In SST the resulting wavelet coefficient that have the same instantaneous frequency are combined via the synchrosqueezing procedure. SST could be resurfaced as an alternative for EMD that lacks solid mathematical foundations due to its data driven and empirical nature [47]. Besides, in situations where the modes in the signal are too close, EMD may fail to separate them, but SST can do the mode separation. However, this condition of having too close modes is not relevant to our applications in this thesis and it is the basis-free nature of EMD that makes it attractive for our analysis. Furthermore, another advantage of EMD is that it could be exploited for completely removing the trend of EEG signals that exists in most recordings. The methods are explained in details in the following sections.

2.3.1 Empirical Mode Decomposition (EMD)

EMD is a fully data-driven technique for decomposing a non-stationary multicomponent time series into a set of finite number of oscillatory components called Intrinsic Mode Functions (IMFs), through a *sifting* process. An IMF satisfies the following two conditions [48]:

1. The number of maxima, which are strictly positive, and the number of minima,

which are strictly negative, for each IMF, are either equal, or differ at most by one.

2. The mean value of the envelope, as defined by the maxima and the minima, for each IMF, is zero.

The EMD process over an arbitrary time series $x[n]$ computes N IMFs $(c_1[n], \dots, c_N[n])$ and a residue signal $r_N[n]$:

$$x[n] = \sum_{i=1}^N c_i[n] + r_N[n] \quad (2.16)$$

The decomposition via the *sifting* process is described as follows:

1. For a given discrete time signal $x[n]$, all the local minima and maxima of $x[n]$ are identified.
2. The upper envelope E_U is calculated by using a cubic spline to connect all the local maxima. Similarly, the lower envelope E_L is calculated from the local minima. The upper and lower envelopes should cover all the data in $x[n]$ between them.
3. The mean $E_{mean} = (E_U + E_L)/2$ of the upper and lower envelopes is calculated, and $x[n]$ is updated by subtracting the mean from it $x[n] \leftarrow x[n] - E_{mean}$.
4. The previous three steps are executed till $x[n]$ is reduced to an IMF $c_1[n]$, which conforms to the properties of IMFs described previously. The first IMF contains the highest oscillation frequencies found in the original data $x[n]$.
5. The first IMF $c_1[n]$ is subtracted from $x[n]$ to get the residue $r_1[n]$.
6. The residue $r_1[n]$ is now taken as the starting point instead of $x[n]$, and the previously mentioned steps are repeated to find all the IMFs $c_i[n]$ so that the final residue r_N either becomes a constant, a monotonic function, or a function with a single maximum and minimum from which no further IMF can be extracted.

Figure 2.4 illustrates a sample EEG signal along its extracted IMFs via EMD. The figure shows that the higher index of IMFs corresponds to lower frequency oscillations. It is

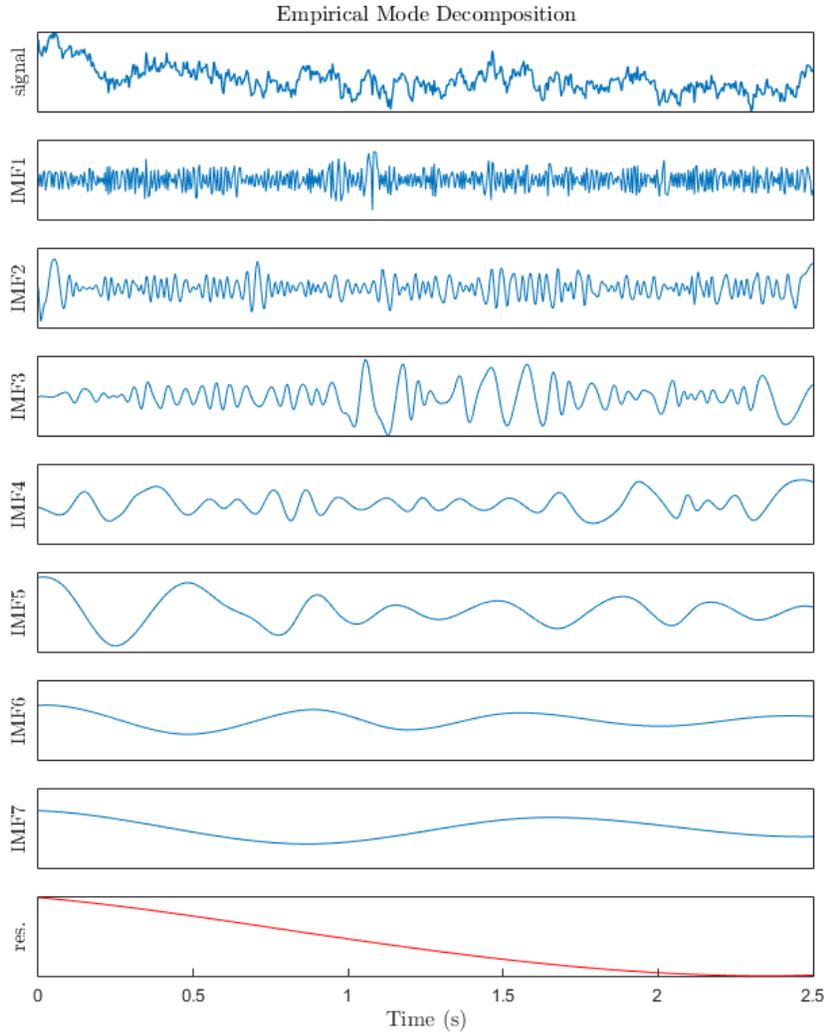


Figure 2.4: Empirical Mode Decomposition of a sample EEG signal. Original signal (Top), extracted IMF1 to IMF7 showing the high to low frequency oscillations, and the residue (Bottom).

known that each of the extracted IMFs has a good spectral localization characteristics. Here in the simulations we use a new variant of EMD, Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) [49], that is based on averaging the modes obtained by EMD applied to several realizations of Gaussian white noise added to the original signal. That would result in better spectral separation of the modes, i.e. to prevent mode mixing which is one of the main drawbacks of ordinary EMD, and a

lesser number of sifting iterations.

Computational Complexity of EMD

It was shown in [50] that the time complexity of EMD/EEMD is $T(n) = 41 \cdot NE \cdot NS \cdot n(\log_2 n)$ which is in the order of $O(n \log n)$, where n is the data length and the parameters NS, NE are the sifting and ensemble numbers respectively. Therefore, the time complexity of the EMD/EEMD is equivalent to the traditional Fourier transform but with a larger factor. This proves that EMD is a computationally efficient method and can be applied to a much larger class of scientific and engineering problems.

2.3.2 EMD-based Detrending of EEG signals

One of the interferences that is present in most EEG recordings is characterised by a very low frequency oscillation superimposed to the fluctuating process and is known as EEG trend. In general, EMD decompose each EEG signal into several oscillatory modes (IMFs) that can be effectively used for detrending. The IMFs with the lower indices correspond to high frequency oscillations, whereas those with higher indices correspond to the trend in the signal. A simple EMD-based EEG detrending method is described in [51] which is through partial reconstruction of the signal, fine-to-coarse, by adding up lower index IMFs till the mean varies significantly from zero. Denoting the recorded signal as $x[n]$, the detrended signal $y[n]$ can be estimated as follows

$$\hat{y}_D[n] = \sum_{i=1}^D c_i[n], \quad (2.17)$$

where D is the larger IMF index prior contamination by the trend. As the mean of each of the IMFs $\{c_i[n]; i = 1, \dots, D\}$ is zero, an intuitive rule for choosing D is to calculate the empirical mean of $\hat{y}_d[n]$ as a function of a test order d and monitor the evolution of that to identify for which $d = D$ it departs significantly from zero. An example of this approach is given in Figure 2.5, where a 2.5s segment of an EEG signal is considered.

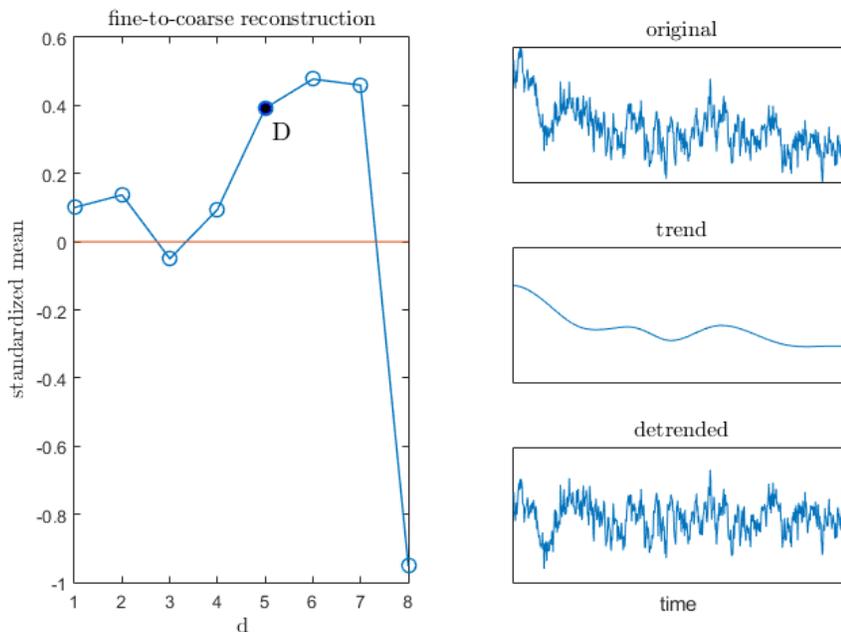


Figure 2.5: Detrending of a EEG signal. left: standardized empirical mean of the fine-to-coarse EMD reconstruction, evidencing $D = 5$ as the change point. Top right: original signal. Middle right: estimated trend obtained from the partial reconstruction with IMFs 6 to 7 + the residual. Bottom right: detrended signal obtained from the partial reconstruction with IMFs 1 to 5.

2.3.3 Reassignment Method (RM)

Discarding phase information in forming TF representation in conventional linear TF methods, including short-time Fourier transform (STFT), has certain consequences including magnitude spreading. Although, there has been research on improving the readability of the TF plane by incorporating the STFT phase information explicitly or implicitly in the process.

Besides, as an alternative path, instead of using the phase information we can also effectively improve the readability of the TF plane with combining STFTs with suitable analysis windows in a way called Reassignment [43]. RM is essentially a postprocessing technique aims to sharpen the TF representation, while keeping the temporal localization, by assigning the local energy to the center of energy distribution. It is particularly well adapted to multicomponent signals [52] and permits to overcome the localization and

interference trade-off that is usually observed in classical TF analysis.

To understand how RM works, it is worth going back to classical TF distributions definitions and properties. Given a signal $x(t)$ and a window $h(t)$, the STFT of $x(t)$ is defined by

$$V_x^h(t, \omega) = \int_{\mathbb{R}} x(\tau)h^*(\tau - t)e^{-i\omega\tau}d\tau \quad (2.18)$$

where $h^*(t)$ is the complex conjugate of $h(t)$. The spectrogram $S_x^h(t, \omega)$ is then usually defined as $|V_x^h(t, \omega)|^2$. Uncertainty principle, which impose that one cannot localize a signal with an arbitrary precision in time and frequency concurrently, is the most significant issue in TF signal analysis. In order to balance between frequency resolution and time localization, TF representation techniques use various tuning parameters. For example, we can tackle that by varying the size of the analysis window in STFT (or spectrogram). The Wigner-Ville distribution, or multi-linear distributions are other examples of many attempts in the literature to optimize this trade-off. However, each of these methods rely on strong assumptions, which makes them suitable only for a specific class of signals.

On the other hand, Reassignment techniques offer an alternative approach with the goal of sharpening the TF representation while keeping the temporal localization. Furthermore, they are particularly well adapted to multicomponent signals. Starting with the definition (2.7) of the STFT, the spectrogram can be written as

$$S_x^h(t, \omega) = \frac{1}{2\pi} \iint_{\mathbb{R}^2} W_h(\tau - t, \nu - \omega)W_x(\tau, \nu)d\tau d\nu, \quad (2.19)$$

where $W_x(t, \omega)$ is the Wigner-Ville distribution (WVD), defined for any function $x(t)$ by

$$W_x(t, \omega) = \int_{\mathbb{R}} x(t + \tau/2)x^*(t - \tau/2)e^{-i\omega\tau}d\tau. \quad (2.20)$$

The question is whether WVD is superior to spectrogram or not ? The answer is yes and no. The spectrogram could be considered as the 2D smoothing of the WVD of the analyzed signal by the WVD of the analyzing window. This alternative formulation

let us compare the the main features of a spectrogram with a WVD. While WVD is favourable as it can sharply localize individual linear chirps in the TF plane, the 2D smoothing involved in the spectrogram computation results in a smearing of their energy distribution. On the other hand, oscillatory interference between individual components (cross-terms) is due to the quadratic nature of the WVD which itself can be removed by the 2D smoothing used in the spectrogram computation.

Reassignment method (RM) seeks to make up for the TF smearing and shifts caused by the 2D smoothing defining the spectrogram. In order to do so, a meaningful TF location need to be determined to assign the local energy given by the spectrogram to it. This location corresponds to the *centroid* of the distribution (2.19), whose coordinates are defined by

$$\begin{aligned}\hat{\omega}_x(t, \omega) &= \frac{1}{S_x^h(t, \omega)} \iint_{\mathbb{R}^2} \nu W_h(\tau - t, \nu - \omega) W_x(\tau, \nu) \frac{d\tau d\nu}{2\pi} \\ \hat{t}_x(t, \omega) &= \frac{1}{S_x^h(t, \omega)} \iint_{\mathbb{R}^2} \tau W_h(\tau - t, \nu - \omega) W_x(\tau, \nu) \frac{d\tau d\nu}{2\pi}.\end{aligned}\quad (2.21)$$

RM then include moving the value of the spectrogram from the point of computation to this so-called centroid:

$$\hat{S}_x^h(t, \omega) = \iint_{\mathbb{R}^2} S_x^h(\tau, \nu) \delta(\omega - \hat{\omega}_x(\tau, \nu)) \delta(t - \hat{t}_x(\tau, \nu)) d\tau d\nu, \quad (2.22)$$

where δ stands for the Dirac distribution [52].

To summarize the TFR technique, after decomposing each EEG channel signal into its constituting IMFs via EMD, the next step is to calculate the RM of each extracted IMF excluding the ones corresponding to the trend of EEG in (2.17), i.e., $c_1(t), \dots, c_D(t)$. Finally, the total Reassigned TFR of a EEG signal is obtained by adding up the RMs of all IMFs, which result in a localized and cross-term free TF representation.

Figure 2.6 shows the TF representation of a sample 10Hz Steady-State Visual Evoked Potential (SSVEP) signal which is characterized by a dominant frequency component at

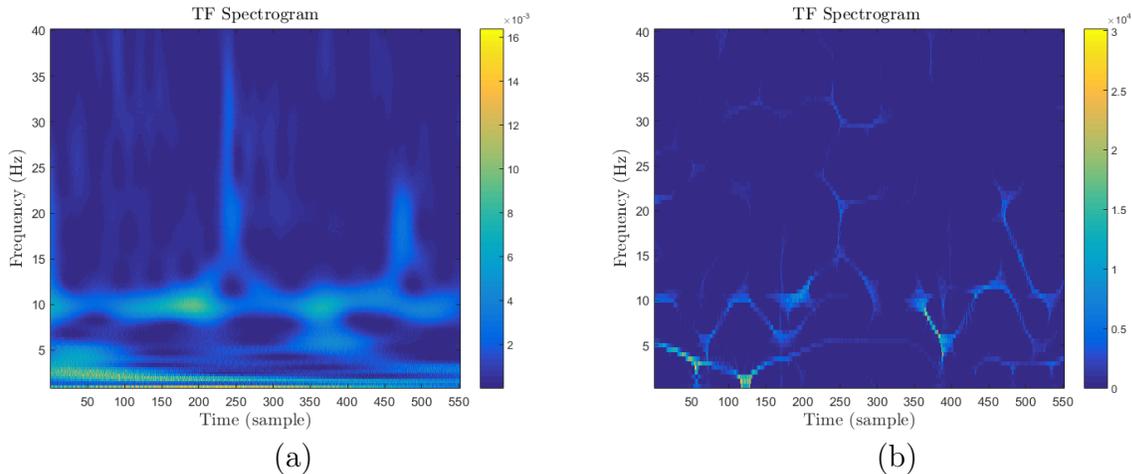


Figure 2.6: Time-Frequency representation of a sample 10Hz Steady-State Visual Evoked Potential (SSVEP) signal: (a) Wavelet Transform using Complex Morlet “cmor1-1.5”; (b) RS-EMD, i.e., EMD followed by Reassigned TFR of extracted IMFs.

10Hz and also higher harmonics at 20 and 30Hz: (a) Continuous Wavelet transform using Complex Morlet “cmor1-1.5”; (b) RS-EMD method which is EMD followed by Reassigned TFR of extracted IMFs. It illustrates that the proposed RS-EMD technique has a good temporal and spectral resolution contemporarily compared to the conventional Wavelet transform mainly used for Multi-way analysis of EEG signal [53–55], which suffers from low temporal resolution at lower frequencies and low spectral resolution at higher frequencies.

The proposed RS-EMD satisfies all the TFR requirements, it is cross-term free due to incorporation of the EMD algorithm as it decomposes the signal into mono-components. Being a variant of spectrogram, it is all positive and due to Reassignment method it satisfies the marginals and have high resolution in both time and frequency domains simultaneously and can be effectively used for practical applications.

2.3.4 Quantitative Performance Index

Parametric identification of signals corrupted with noise is an interesting application of time-frequency analysis. For this purpose it is very important that a time-frequency

distribution highly concentrates signals' energy along the instantaneous frequency. A metric that measures the ratio of energy along the instantaneous frequencies and the energy outside these regions was proposed in [56, 57] as follows

$$B = 10 \log \frac{\iint_{(t,\omega) \in R} \text{TFR}(t, \omega) dt d\omega}{\iint_{(t,\omega) \notin R} \text{TFR}(t, \omega) dt d\omega} \quad (2.23)$$

where the symbol $\text{TFR}(t, \omega)$ denotes the time-frequency representation, and R is the instantaneous frequency path of the desired signal. This ratio quantifies the localization performance of TFR and can be compared for different TFRs, as a function of noise variance. Hence, the TFR that gives greater B for a set of signals with known actual instantaneous frequencies has been more effective in localizing the energy the of signal. In Chapter 5 we compare the RS-EMD with the conventional Wavelet-based TFR method in terms of localization of energy for a synthetic EEG signal with known instantaneous frequency.

2.4 Chapter Summary

This chapter presented the comparison of few well known TF distribution techniques from their TF resolution point of view. The shortcoming of each of the TF methods was explored and a fully data-driven TRF technique with high temporal and spectral resolution was developed. The properties of the proposed TFR are listed in Table 2.2 that satisfies all the TFR requirements. Combination of EMD and Reassigned method, with a high-resolution TF representation characteristic, was exploited for transforming the EEG signals into the time-frequency plane. Furthermore, EMD is favourable as it could be used for EEG detrending. Detrending is simply embedded in the proposed TFR method. Simulation results on real EEG data show that the proposed TFR method could be effectively used for EEG analysis and separation and localization of events in EEG could be explored further via RS-EMD method.

Table 2.2: Desirable TFR Properties of the proposed technique

Property	Positivity	Marginals	Time-Frequency Resolution	Time-Frequency Localization	Practical Ability
RS-EMD	✓	✓	✓	✓	✓

Nevertheless, study of the brain processes requires methods which can describe these variations of the EEG signal in time, frequency, and space in a quantitative way, which inspires the incorporation of more contextual information in the analysis. Hence, we need to consider multi-way arrays (tensor) for the analysis. In the next chapter we introduce the multi-way arrays and the corresponding analysis models.

Chapter 3

Multi-way Analysis

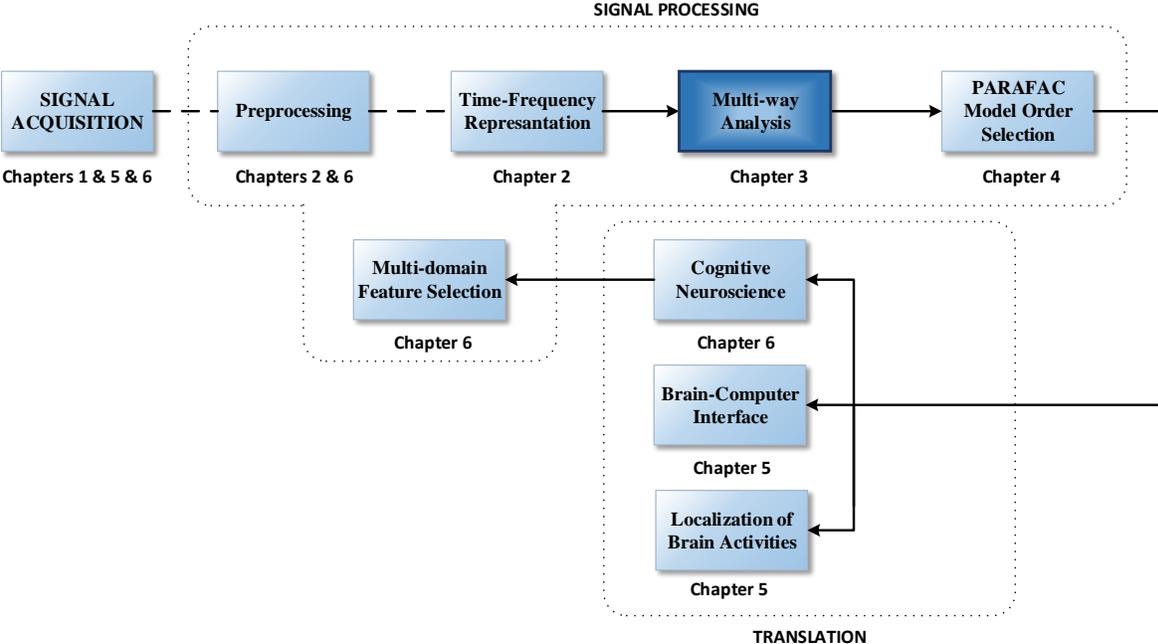


Figure 3.1: Chapter 3- Multi-way Analysis.

EEG signal is characterized as a high-dimensional and noisy data, and there has been numerous attempts to decompose it into functions with known spatio-temporal-spectral properties or at least easier to characterize which is the philosophy behind EEG analysis. In this context, the approaches for EEG analysis can be categorized into time

domain analysis, frequency domain analysis via Fourier Transform, spatio-temporal and time-frequency analysis. In the last two cases, methods have focused on analyzing two dimensions simultaneously.

The spatio-temporal analysis seek to represent the data as a sum of components, each of them having their own temporal and spatial signatures. In order to do so, there is a fundamental assumption about the brain network which states that activity of each brain network has a fixed spatial pattern with the dynamics of it changing in time. However, it is quite known that this kind of two-dimensional decomposition is not unique and in order to obtain meaningful and unique description of the data some additional constraints need to be added to the problem. The two most popular methods used in this type of analysis are the Principal Component (PCA) and the Independent Component Analysis (ICA), which extract the signatures by constraining them to be orthogonal or statistically independent, respectively. These mathematical constraints, although useful in many applications, do not always lead to interpretable and physiologically meaningful decompositions.

On the other hand, time-frequency methods use variants of the Short-Term Fourier Transform (STFT) in an attempt to characterize the non-stationarity of EEG signals and decompose the EEG time series into some basis functions with localized support in time and frequency. The coefficients of this decomposition can be mapped into a time-frequency plane and their square absolute value (energy) is known as the spectrogram. Selecting different basis for decomposition, e.g., Gabor or Wavelet functions, result in different type of spectrograms although they are very similar and statistically equivalent.

Furthurmore, standard matrix factorizations techniques, such as PCA/SVD, ICA, Nonnegative Matrix Factorization (NMF), and their variants, are invaluable analyzing tools for feature selection, dimensionality reduction, noise reduction, and data mining [58]. However, they have only two modes or 2-way representations (say, space and time), and data need to be fitted into a matrix for analysis which make their usage limited. In many applications the data is naturally high-dimensional and in addition to the intrinsic

modes of time, frequency, and space it may also contain higher-order modes (ways) such as trials, subjects, task conditions, and groups. For instance, a sequence of trials in an EEG experiment may lead to a large stream of data with many dimensions: space, time-frequency, subjects, trials, and conditions.

Applying the two-way analysis techniques to such high-dimensional data inevitably requires the data to be unfolded. Clearly the “flat-world view” provided by two-way matrix factorizations (ICA, NMF) may be insufficient and it is natural to use tensor decomposition approaches. Hence, by incorporating the multi-way analysis, all dimensions or modes are retained and because of multilinear models often produce unique and physically meaningful components. It has been shown in numerous research areas [59, 60], that underlying information content of the data may not be captured accurately or identified uniquely by two-way data analysis. Moreover as mentioned earlier two-way analysis is unique only if specific constraints such as statistical independence and orthogonality are enforced on the components. However, these assumptions and constraints are not often necessary for multi-way models and some multi-way analysis techniques are unique under mild conditions.

Multi-way data or tensor analysis is the extension of two-way data analysis to higher-order data sets. Multi-way analysis can be incorporated for revealing the hidden structures and capturing the latent correlation between variables in a multi-way array [61]. For instance, it is very usual in the neuroscience studies to have multiple subjects (people or animals) and trials in the experiments which lead to data structures conveniently represented by multi-way arrays or blocks of three-way data. Analyzing the data of every subject separately, ignoring the multi-way structure of the data, would lose the covariance information among subjects. Hence, it is essential for the analysis tools to keep the multi-dimensional structure of the data, in order to discover hidden components within the data and retain the integrative information. The multi-way analysis (tensor factorizations and decompositions) is a natural choice in this context, for instance, in EEG

studies multi-way analysis of multichannel EEG data enables us to capture the spatial correlation between the channels by representing signals in both time and frequency domains. Furthermore, multi-way analysis framework provides convenient multi-channel and multi-subject time-frequency-space sparse representations, artifacts rejection in the time-frequency domain, feature extraction, multi-way clustering and coherence tracking. Our main objective here is to decompose the time-varying multichannel EEG signals into components with distinct temporal, spectral and spatial signatures which are then discriminative enough across different conditions.

3.1 Multi-way arrays

Multi-way arrays, also referred to as tensors, are higher-order generalizations of vectors and matrices [62]. Higher-order arrays are represented as $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$, where the order of $\underline{\mathbf{X}}$ is N ($N > 2$). A zero-order tensor is a scalar, a first-order tensor is a vector, a second-order tensor is a matrix, and tensors of order three and higher are called higher-order tensors (see Figure 3.2).

Higher-order arrays have a different terminology compared to two-way data sets. Each dimension of a multi-way array is called a *mode* (or a *way*), and the number of variables in each mode is used to indicate the dimensionality of a mode. For instance, $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ is a multi-way array with N modes (called N -way array or N th-order tensor) with I_1, I_2, \dots, I_N dimensions in the first, second, \dots , N th mode, respectively. Each entry of $\underline{\mathbf{X}}$ is denoted by $x_{i_1 i_2 \dots i_N}$. For a special case, where $N = 3$, let $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ be a three-way array that has three modes (or indices or dimensions) as shown in Figure 3.3. Then $x_{i_1 i_2 i_3}$ denotes the entry in the i_1 th row, i_2 th column, and i_3 th tube of $\underline{\mathbf{X}}$.

3.1.1 Notation

We use the notation of Reference [63]. Scalars are denoted by lowercase letters, e.g., a . Vectors are denoted by boldface lowercase letters, e.g., \mathbf{a} . Matrices are denoted by

boldface capital letters, e.g., \mathbf{A} . Higher-order tensors are denoted by underlined boldface letters, e.g., $\underline{\mathbf{A}}$. The i th entry of a vector \mathbf{a} is denoted by a_i , element (i, j) of a matrix \mathbf{A} is denoted by a_{ij} , and element (i, j, k) of a third-order tensor $\underline{\mathbf{X}}$ is denoted by x_{ijk} . The j th column of a matrix \mathbf{A} is denoted by \mathbf{a}_j . Indices typically range from 1 to their capital version, e.g., $i = 1, \dots, I$. The n th element in a sequence is denoted by a superscript in parentheses, e.g., $\mathbf{a}^{(n)}$ denotes the n th vector in a sequence.

The inner product of two same-sized tensors $\underline{\mathbf{X}}, \underline{\mathbf{Y}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ is the sum of the products of their entries, i.e.,

$$\langle \underline{\mathbf{X}}, \underline{\mathbf{Y}} \rangle = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} x_{i_1, i_2, \dots, i_N} y_{i_1, i_2, \dots, i_N}. \quad (3.1)$$

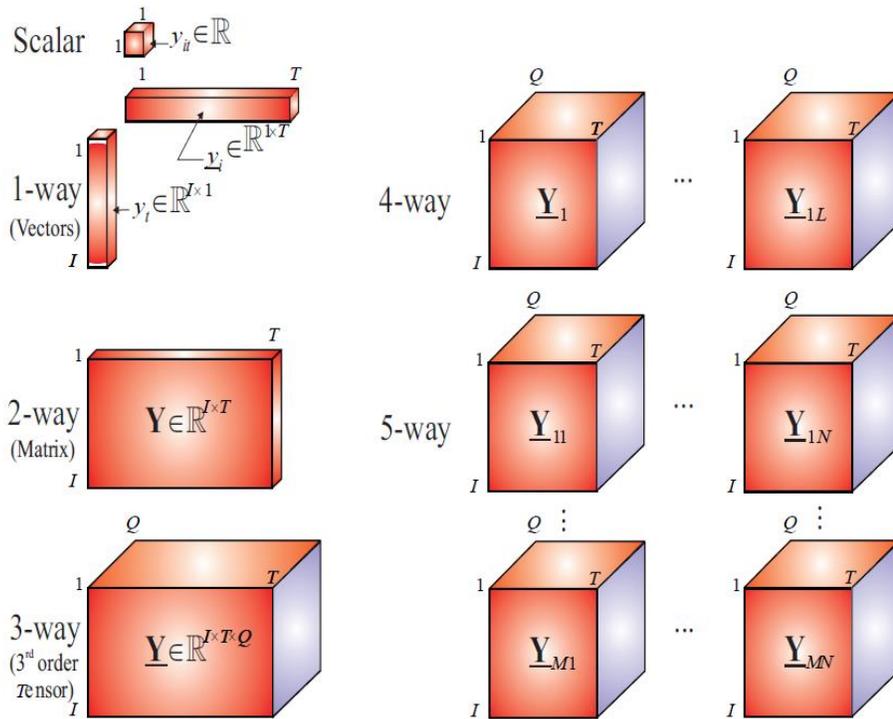


Figure 3.2: Illustration of multi-way data: zero-way tensor = scalar, 1-way tensor = row or column vector, 2-way tensor = matrix, N-way tensor = higher-order tensors. The 4-way and 5-way tensors are represented here as a set of the three-way tensors [61].

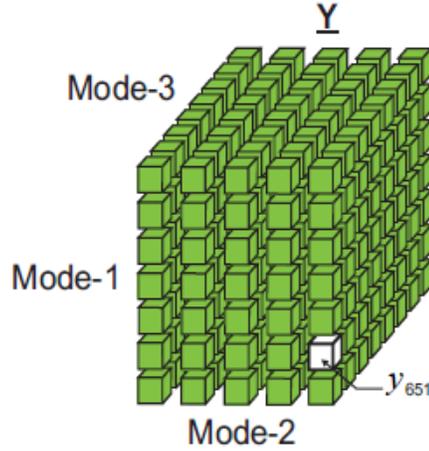


Figure 3.3: A three-way array (third-order tensor) $\underline{\mathbf{Y}} \in \mathbb{R}^{7 \times 5 \times 8}$ with elements y_{itq} [61].

The norm of a tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ is the square root of its inner product with itself, i.e.,

$$\|\underline{\mathbf{X}}\| = \sqrt{\langle \underline{\mathbf{X}}, \underline{\mathbf{X}} \rangle}. \quad (3.2)$$

For matrices (i.e., second-order tensors), $\|\cdot\|$ refers to the analogous Frobenius norm, and, for vectors (i.e., first-order tensors), $\|\cdot\|$ refers to the analogous two-norm.

3.1.2 Subarrays, Tubes and Slices

Subtensors or subarrays are formed when a subset of the indices is fixed. For matrices, these are the rows and columns. A tensor fiber is a one-dimensional fragment of a tensor, obtained by fixing all indices except for one (Figure 3.4(a)). A matrix column is a mode-1 fiber and a matrix row is a mode-2 fiber. Third-order tensors have column, row, and tube fibers. A tensor slice is a two-dimensional section (fragment) of a tensor, obtained by fixing all indices except for two indices. Figure 3.4(b) shows the horizontal, lateral, and frontal slices of a third-order tensor.

It is often very convenient to represent tensors as matrices or to represent multi-way relationships and a tensor decomposition in their matrix forms. *Matricization*, also known as *unfolding* or *flattening*, is the process of reordering the elements of an N -way tensor

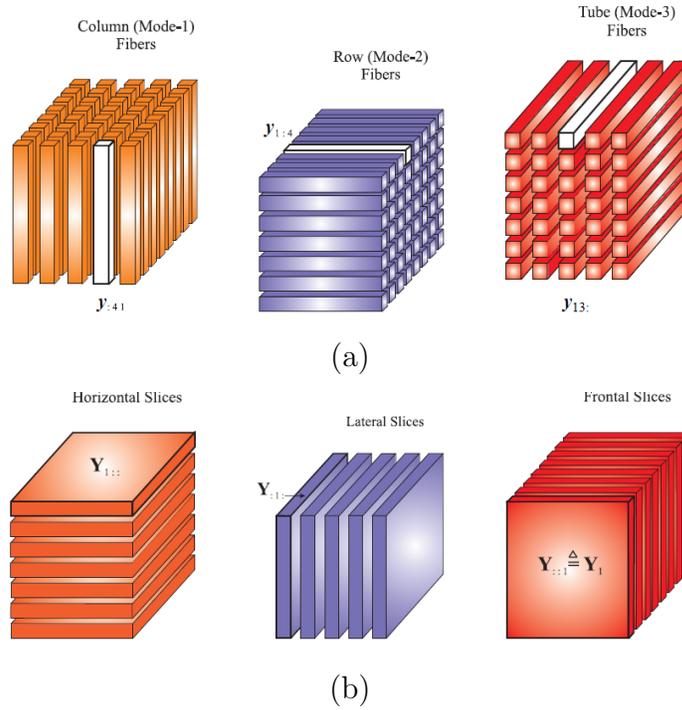


Figure 3.4: (a) Fibers for a third-order tensor (all fibers are treated as column vectors); (b) Slices for a third-order tensor $\underline{\mathbf{Y}} \in \mathbb{R}^{I \times J \times K}$. A colon is used to indicate all elements of a mode [61].

into a matrix. The mode- n matricization of a tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ is denoted by $\mathbf{X}_{(n)}$ and arranges the mode- n one-dimensional fibers to be the columns of the resulting matrix. Specifically, tensor element (i_1, i_2, \dots, i_N) maps to matrix element (i_n, j) where

$$j = 1 + \sum_{\substack{k=1 \\ k \neq n}}^N (i_k - 1) J_k, \quad \text{with} \quad J_k = \prod_{\substack{m=1 \\ m \neq n}}^{k-1} I_m \quad (3.3)$$

Note that $J_k = 1$ if $k = 1$ or if $k = 2$ and $n = 1$. Since matricization is just a rearrangement of the elements, clearly $\|\underline{\mathbf{X}}\| = \|\mathbf{X}_{(n)}\|$ for $n = 1, \dots, N$.

The n -mode (vector) product of a tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ with a vector $\mathbf{v} \in \mathbb{R}^{I_n}$ is denoted by $\underline{\mathbf{X}} \times_n \mathbf{v}$. The result is of order $N - 1$, i.e., the size is $I_1 \times \dots \times I_{n-1} \times I_{n+1} \times \dots \times I_N$.

$\dots \times I_N$. Elementwise, the n -mode product can be formulated as

$$(\underline{\mathbf{X}} \times_n \mathbf{v})_{i_1 \dots i_{n-1} i_{n+1} \dots i_N} = \sum_{i_n=1}^{I_n} x_{i_1 i_2 \dots i_N} v_{i_n}. \quad (3.4)$$

We note that multiplication in all modes results in a scalar and multiplication in every mode except n and p results in a matrix of size $I_n \times I_p$. A tensor may be multiplied by multiple vectors at once. For example, assume $\mathbf{v}^{(n)} \in \mathbb{R}^{I_n}$ for $n = 1, \dots, N$. Then we use the notation $\underline{\mathbf{X}} \times_1 \mathbf{v}^{(1)} \times_2 \mathbf{v}^{(2)} \dots \times_N \mathbf{v}^{(N)}$ to denote multiplication in multiple modes.

3.1.3 Outer, Kronecker, and Khatri-Rao Products

Several special matrix products are important for representation of tensor factorizations and decompositions.

The outer product of the tensors $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2, \dots, \times I_N}$ and $\underline{\mathbf{Y}} \in \mathbb{R}^{J_1 \times J_2, \dots, \times J_M}$ is given by

$$\underline{\mathbf{Z}} = \underline{\mathbf{X}} \circ \underline{\mathbf{Y}} \in \mathbb{R}^{I_1 \times I_2, \dots, \times I_N \times J_1 \times J_2, \dots, \times J_M} \quad (3.5)$$

where $z_{i_1, i_2, \dots, i_N, j_1, j_2, \dots, j_M} = x_{i_1, i_2, \dots, i_N} y_{j_1, j_2, \dots, j_M}$. The symbol \circ represents the vector outer product. Note that, the tensor $\underline{\mathbf{Z}}$ contains all the possible combinations of pair-wise products between the elements of $\underline{\mathbf{X}}$ and $\underline{\mathbf{Y}}$. As special cases, the outer product of two vectors $\mathbf{a} \in \mathbb{R}^I$ and $\mathbf{b} \in \mathbb{R}^J$ yields a rank-one matrix

$$\mathbf{A} = \mathbf{a} \circ \mathbf{b} = \mathbf{a} \mathbf{b}^T \in \mathbb{R}^{I \times J} \quad (3.6)$$

and the outer product of three vectors: $\mathbf{a} \in \mathbb{R}^I$, $\mathbf{b} \in \mathbb{R}^J$ and $\mathbf{c} \in \mathbb{R}^K$ yields a third-order rank-one tensor

$$\underline{\mathbf{Z}} = \mathbf{a} \circ \mathbf{b} \circ \mathbf{c} \in \mathbb{R}^{I \times J \times K} \quad (3.7)$$

where $z_{i,j,k} = a_i b_j c_k$. Generally, an N -way tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2, \dots, \times I_N}$ is rank one if it can

be written as the outer product of N vectors, i.e.,

$$\underline{\mathbf{X}} = \mathbf{a}^{(1)} \circ \mathbf{a}^{(2)} \circ \dots \circ \mathbf{a}^{(n)} \quad (3.8)$$

This means that each element of the tensor is the product of the corresponding vector elements:

$$x_{i_1, i_2, \dots, i_N} = a_{i_1}^{(1)} a_{i_2}^{(2)} \dots a_{i_N}^{(N)} \text{ for all } 1 \leq i_n \leq I_n. \quad (3.9)$$

The Kronecker product of two matrices $\mathbf{A} \in \mathbb{R}^{I \times J}$ and $\mathbf{B} \in \mathbb{R}^{T \times R}$ is a matrix denoted as $\mathbf{A} \otimes \mathbf{B} \in \mathbb{R}^{IT \times JR}$ and defined as

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \dots & a_{1J}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \dots & a_{2J}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{I1}\mathbf{B} & a_{I2}\mathbf{B} & \dots & a_{IJ}\mathbf{B} \end{bmatrix}. \quad (3.10)$$

It should be mentioned that, in general, the outer product of vectors yields a tensor whereas the Kronecker product gives a vector.

The Khatri-Rao product is the matching columnwise Kronecker product. Given matrices $\mathbf{A} \in \mathbb{R}^{I \times K}$ and $\mathbf{B} \in \mathbb{R}^{J \times K}$, their Khatri-Rao product is denoted by $\mathbf{A} \odot \mathbf{B}$. The result is a matrix of size $(IJ) \times K$ and defined by

$$\mathbf{A} \odot \mathbf{B} = [\mathbf{a}_1 \otimes \mathbf{b}_1 \quad \mathbf{a}_2 \otimes \mathbf{b}_2 \quad \dots \quad \mathbf{a}_k \otimes \mathbf{b}_k]. \quad (3.11)$$

Recall that the Kronecker product of two vectors $\mathbf{a} \in \mathbb{R}^I$ and $\mathbf{b} \in \mathbb{R}^J$ is a vector of length IJ defined by

$$\mathbf{a} \otimes \mathbf{b} = \begin{bmatrix} a_1 \mathbf{b} \\ a_2 \mathbf{b} \\ \vdots \\ a_I \mathbf{b} \end{bmatrix}.$$

3.2 Multi-way Models

In general, multi-way data analysis methods are extensions of two-way analysis techniques based on the idea of linear factor models [61, 62]. The factor models in the context of bilinear factor models is defined as follows. Let the matrix $\mathbf{X} \in \mathbb{R}^{M \times N}$ represent the original data set. We can model \mathbf{X} using an R -component bilinear model as

$$x_{ij} = \sum_{r=1}^R a_{ir}b_{jr} + e_{ij}, \quad (3.12)$$

where $\sum_{r=1}^R a_{ir}b_{jr}$ is the part of the model consisting of matrices $\mathbf{A} \in \mathbb{R}^{M \times R}$ and $\mathbf{B} \in \mathbb{R}^{N \times R}$. Each column of \mathbf{B} corresponds to a factor, and each row of \mathbf{A} contains the scores corresponding to R factors. Matrix $\mathbf{E} \in \mathbb{R}^{M \times N}$ contains the residual. We can measure how well a model fits the data by analyzing the residual. Bilinear or multi-linear analysis enables us to extract the factors that are linear combination of variables. These extracted factors summarize the data and reveal the hidden structure and information content of the data and could be used for interpreting the data. Bilinear factor models have a close relation with the definition of rank of a matrix. The rank of \mathbf{X} is the number of its non-zero singular values. Equivalently, it is the minimum r such that \mathbf{X} can be written as

$$\mathbf{X} = \sum_{i=1}^r \mathbf{a}_i \mathbf{b}_i^T, \quad (3.13)$$

with $\mathbf{a}_i \in \mathbb{R}^M, \mathbf{b}_i \in \mathbb{R}^N$. In general, this presentation is unique up to numbering and common scaling of the $\mathbf{a}_i, \mathbf{b}_i$. So, Bilinear factor model decomposes the data into combination of rank-1 matrices. We later extend this into the context of multi-linear factor analysis in order to define the rank of a tensor.

Higher-order tensor decompositions are nowadays frequently used in a variety of fields including psychometrics, chemometrics, image analysis, graph analysis, and signal processing [61]. The most well-known multi-way models in the literature are Tucker family [64] and the PARAFAC family [65, 66], which is also called Canonical Decomposition

(CANDECOMP). CANDECOMP/PARAFAC solution is unique up to a scaling and permutation of the component matrices. It is this uniqueness property that makes CANDECOMP/PARAFAC a popular technique in various fields. Tucker is a more flexible multi-way model compared to CANDECOMP/PARAFAC. The flexibility of a Tucker model is related to the existence of non-restricted higher-order core array which implies that each component in each mode can interact with any component in other modes. On the other hand, in a CANDECOMP/PARAFAC model, due to the super-diagonal core-array in the modeling, a component in a certain mode can be related to a single component in another mode.

These models consist of two parts: a parametric part with multiple factors describing the data structure and a residual part relating to measurement noise. Multi-way models are mainly optimized in the sense of least squares by iterative algorithms, such as Alternating Least Squares (ALS) which estimates one factor at a time keeping the estimates of other components fixed. The formulation of these two models are explained in the following sections.

3.2.1 TUCKER Model

Tucker models are extension of bilinear factor analysis to higher-order data sets, which are also called N-way principal component analysis techniques. Tucker family in consisted of three models, i.e., Tucker1, Tucker2, and Tucker3 models [62]. Tucker1 simply matrixizes the multi-way data and decomposes the resulted unfolded multi-way array via Singular Value Decomposition (SVD). Tucker1 essentially reduce the rank in one mode. However, Tucker2 and Tucker3 models allow rank reduction in more than one mode and are named to reflect the number of modes rank reduction considered in the model. Tucker3 is a more flexible multi-way model due to the core array, which enables the interaction of each component in each mode with all components in other modes. Equation (3.14) shows a commonly-used formulation for a Tucker3 model applied on a three-way array

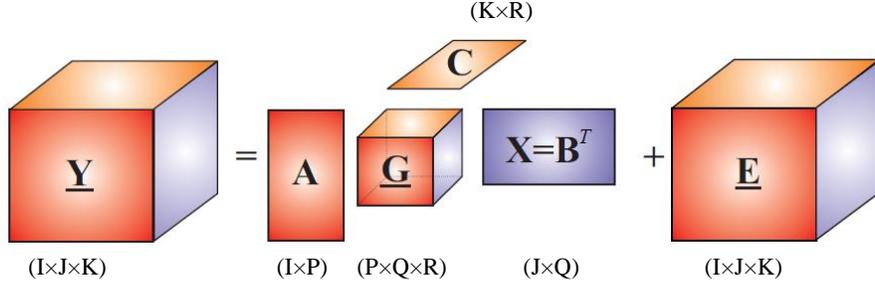


Figure 3.5: (P, Q, R) -component Tucker3 model, where a three-way array $\underline{\mathbf{Y}} \in \mathbb{R}^{I \times J \times K}$ is modeled with component matrices $\underline{\mathbf{A}} \in \mathbb{R}^{I \times P}$, $\underline{\mathbf{B}} \in \mathbb{R}^{J \times Q}$ and $\underline{\mathbf{C}} \in \mathbb{R}^{K \times R}$ in the first, second and third mode, respectively. $\underline{\mathbf{G}} \in \mathbb{R}^{P \times Q \times R}$ is the core array and $\underline{\mathbf{E}} \in \mathbb{R}^{I \times J \times K}$ contains the error terms.

$$\underline{\mathbf{Y}} \in \mathbb{R}^{I \times J \times K}.$$

$$y_{ijk} = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R g_{pqr} a_{ip} b_{jq} c_{kr} + e_{ijk}, \quad (3.14)$$

or equivalently using n -mode multiplication in the form of

$$\underline{\mathbf{Y}} = \underline{\mathbf{G}} \times_1 \underline{\mathbf{A}} \times_2 \underline{\mathbf{B}} \times_3 \underline{\mathbf{C}} + \underline{\mathbf{E}} \quad (3.15)$$

where $\underline{\mathbf{A}} \in \mathbb{R}^{I \times P}$, $\underline{\mathbf{B}} \in \mathbb{R}^{J \times Q}$, and $\underline{\mathbf{C}} \in \mathbb{R}^{K \times R}$ are the component matrices corresponding to the first, second, and third modes, respectively. $\underline{\mathbf{G}} \in \mathbb{R}^{P \times Q \times R}$ is the three-way core array and $\underline{\mathbf{E}} \in \mathbb{R}^{I \times J \times K}$ is a three-way array containing the residuals. Illustration of a Tucker3 model on a three-way array is given in Fig. 3.5.

For a three-way model, the core array is a third-order tensor, $\underline{\mathbf{G}}$ as given in Figure 3.5, where g_{pqr} represents the interaction of the p th factor in the first, q th factor in the second and r th factor in the third mode. Determining the optimum rank in each mode in TUCKER model is a challenge and usually need to be tuned by the user. Practically, the ranks estimated using the SVD on unfolded data in each mode can be used for this purpose. However, there are other systematic rank selection technique, e.g., cross validation and Difference in Fit (DIFFIT), and a Convex-hull based method [67, 68].

Besides, by imposing orthogonality constraints on the extracted components of Tucker model, a model called Higher-Order Singular Value Decomposition (HOSVD) was proposed in [69]. HOSVD can be computed by unfolding the tensor in each mode and estimating the left singular vectors constituting that mode. Those singular vectors which are called n -mode singular vectors then form the component matrices in each mode. Having the components matrices of all modes estimated, a core tensor can be computed consequently.

3.2.1.1 Algorithms

The original algorithm for computing a Tucker3 model was described by Tucker in [64], that for a three-way array $\underline{\mathbf{Y}}$, after unfolding the data in each mode, e.i., $\mathbf{Y}_{(1)}$, $\mathbf{Y}_{(2)}$, and $\mathbf{Y}_{(3)}$, the model calculated the eigenvectors of $\mathbf{Y}_{(i)}\mathbf{Y}_{(i)}^T$, which are essentially the left singular vectors of $\mathbf{Y}_{(i)}$, to form the component matrix for mode i . The final step is to estimate the core tensor using the original data and the component matrices. The number of components extracted from each mode need to be specified priorly. In order to provide the optimum solution in the least square sense another algorithm, called TUCKALS3 [70], was proposed that was the extension of the original one via ALS for three-way arrays. The TUCKALS3 algorithm finds the best approximation for $\underline{\mathbf{Y}} \in \mathbb{R}^{I \times J \times K}$ by minimizing the least squares error function given in

$$\|\underline{\mathbf{Y}} - \tilde{\underline{\mathbf{Y}}}\|^2 = \|\underline{\mathbf{Y}} - \underline{\mathbf{G}} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C}\|_F^2 \quad (3.16)$$

where \mathbf{A} , \mathbf{B} , and \mathbf{C} are columnwise orthogonal component matrices in the first, second, and third modes, respectively, and $\underline{\mathbf{G}}$ is the core tensor. Usually the component matrices obtained in Algorithm 1 are used to initialize the ALS algorithms. Then, these components matrices are alternatively estimated one at a time by keeping the other component matrices fixed. Estimation of component matrices is repeated until convergence is reached, e.g., no change in the model fit. Then, in a final step the core tensor is estimated

via n -mode product (Algorithm 2).

Algorithm 1 TUCKER3 [64]($\underline{\mathbf{Y}} \in \mathbb{R}^{I \times J \times K}$, P, Q, R)

$\underline{\mathbf{Y}} \in \mathbb{R}^{I \times J \times K}$ and P, Q, R as the rank of component matrices. Estimate of \mathbf{A} , \mathbf{B} , \mathbf{C} , and the core tensor $\underline{\mathbf{G}}$.

$\mathbf{A} \leftarrow$ First P left singular vectors of $\mathbf{Y}_{(1)}$. $\mathbf{B} \leftarrow$ First Q left singular vectors of $\mathbf{Y}_{(2)}$.
 $\mathbf{C} \leftarrow$ First R left singular vectors of $\mathbf{Y}_{(3)}$. $\underline{\mathbf{G}} \leftarrow \underline{\mathbf{Y}} \times_1 \mathbf{A}^T \times_2 \mathbf{B}^T \times_3 \mathbf{C}^T$.

Algorithm 2 TUCKER3-ALS [70]($\underline{\mathbf{Y}} \in \mathbb{R}^{I \times J \times K}$, P, Q, R)

$\underline{\mathbf{Y}} \in \mathbb{R}^{I \times J \times K}$ and P, Q, R as the rank of component matrices and a convergence criterion ϵ . Estimate of \mathbf{A} , \mathbf{B} , \mathbf{C} , and the core tensor $\underline{\mathbf{G}}$. Initialize $\mathbf{A} \in \mathbb{R}^{I \times P}$, $\mathbf{B} \in \mathbb{R}^{J \times Q}$, and $\mathbf{C} \in \mathbb{R}^{K \times R}$. *not satisfied* $\underline{\mathbf{X}} \leftarrow \underline{\mathbf{Y}} \times_2 \mathbf{B}^T \times_3 \mathbf{C}^T$. $\mathbf{A} \leftarrow$ the first P left singular vectors of $\mathbf{X}_{(1)}$.
 $\underline{\mathbf{X}} \leftarrow \underline{\mathbf{Y}} \times_1 \mathbf{A}^T \times_3 \mathbf{C}^T$. $\mathbf{B} \leftarrow$ the first Q left singular vectors of $\mathbf{X}_{(2)}$. $\underline{\mathbf{X}} \leftarrow \underline{\mathbf{Y}} \times_1 \mathbf{A}^T \times_2 \mathbf{B}^T$.
 $\mathbf{C} \leftarrow$ the first R left singular vectors of $\mathbf{X}_{(3)}$. $\underline{\mathbf{G}} \leftarrow \underline{\mathbf{Y}} \times_1 \mathbf{A}^T \times_2 \mathbf{B}^T \times_3 \mathbf{C}^T$.

Besides, there are other variants of ALS-based Tucker decomposition that try to improve the performance of the algorithm in terms of speed or accuracy. Another approach is to use the Bayesian framework for Tucker decomposition with automatic order selection. For an overview of different techniques refer to [62].

3.2.2 CANDECAMP/PARAFAC MODEL

One of the most well-known multi-way models in literature is the Parallel Factor Analysis (PARAFAC) model. The model was independently proposed in 1970 by Harshman [71] and by Carroll & Chang [65] which is also called Canonical Decomposition (CANDECAMP). CANDECAMP/PARAFAC model for an N -way array of data $\underline{\mathbf{Y}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ can be formulated as

$$y_{i_1 i_2 \dots i_N} = \sum_{r=1}^R a_{i_1 r}^{(1)} a_{i_2 r}^{(2)} \dots a_{i_N r}^{(N)} + e_{i_1 i_2 \dots i_N}, \quad (3.17)$$

or equivalently, using the vectors outer product as

$$\underline{\mathbf{Y}} = \sum_{r=1}^R \mathbf{a}_r^{(1)} \circ \mathbf{a}_r^{(2)} \circ \dots \circ \mathbf{a}_r^{(N)} + \underline{\mathbf{E}}, \quad (3.18)$$

where $\mathbf{a}_r^{(1)} = [a_{i_1 r}^{(1)}] \in \mathbb{R}^{I_1}$, $\mathbf{a}_r^{(2)} = [a_{i_2 r}^{(2)}] \in \mathbb{R}^{I_2}$, and $\mathbf{a}_r^{(N)} = [a_{i_N r}^{(N)}] \in \mathbb{R}^{I_N}$ are respectively the constituent vectors of different modes of the array and $\underline{\mathbf{E}}$ is an N -way array containing the residuals.

For simplicity and without loss of generality, we stick to three-way array of the data $\underline{\mathbf{Y}} \in \mathbb{R}^{I \times J \times K}$ where CANDECOMP/PARAFAC model according to (3.18) can be written as

$$y_{ijk} = \sum_{r=1}^R a_{ir} b_{jr} c_{kr} + e_{ijk}, \quad (3.19)$$

where $\mathbf{a}_r = [a_{ir}] \in \mathbb{R}^I$, $\mathbf{b}_r = [b_{jr}] \in \mathbb{R}^J$, and $\mathbf{c}_r = [c_{kr}] \in \mathbb{R}^K$ are respectively the constituent vectors of the corresponding factor matrices. This is a special case of (3.14), where for an R -component three-way CANDECOMP/PARAFAC model, the core array is restricted to be a superdiagonal core array, $\underline{\mathbf{G}} \in \mathbb{R}^{R \times R \times R}$ where $g_{ijk} \neq 0$ if $i = j = k$, otherwise $g_{ijk} = 0$. This means that for a superdiagonal core array, i th factor in one mode can only interact with the i th factors of other modes and there is no interaction with other factors. Mathematically, a CANDECOMP/PARAFAC model can be represented as the decomposition of a tensor as a linear combination of rank-one tensors. The coefficient of this decomposition are the elements of the superdiagonal core that constitute a vector of coefficients. (An N th-order rank-one tensor is a tensor that can be written as the outer product of N non-zero vectors. Let $\mathbf{a}, \mathbf{b}, \mathbf{c}$ be column vectors of size $I \times 1, J \times 1$ and $K \times 1$, respectively, and $\underline{\mathbf{Y}}$ be a tensor of size $I \times J \times K$, then $\underline{\mathbf{Y}} = \mathbf{a} \circ \mathbf{b} \circ \mathbf{c}$ is the vector outer product if and only if $y_{ijk} = a_i b_j c_k$. This generalizes $\mathbf{a} \circ \mathbf{b} = \mathbf{a} \mathbf{b}^T$ in the case of matrices). So, equivalently equation (3.19) can be written as

$$\underline{\mathbf{Y}} = \sum_{r=1}^R \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r + \underline{\mathbf{E}}. \quad (3.20)$$

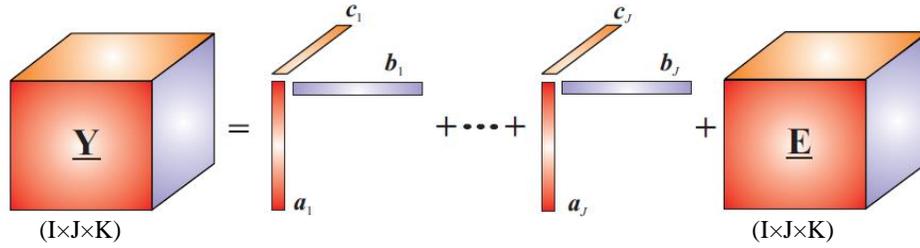


Figure 3.6: R -component CANDECOMP/PARAFAC model, where a three-way array $\underline{\mathbf{Y}}$ is expressed as the sum of rank-1 tensors and error terms. \mathbf{a}_i , \mathbf{b}_i and \mathbf{c}_i are the i -th components in the first, second and third mode, respectively. $\underline{\mathbf{E}}$ is a three-way array containing the residuals.

So, an R -component CANDECOMP/PARAFAC model can be expressed as in (3.20), where \mathbf{a}_i , \mathbf{b}_i , and \mathbf{c}_i indicate the i th column of component matrices $\mathbf{A} \in \mathbb{R}^{I \times R}$, $\mathbf{B} \in \mathbb{R}^{J \times R}$, and $\mathbf{C} \in \mathbb{R}^{K \times R}$, respectively. $\underline{\mathbf{E}} \in \mathbb{R}^{I \times J \times K}$ is a three-way array containing the residuals. Sometimes the “Kruskal operator” shorthand notation is used:

$$[[\mathbf{A}, \mathbf{B}, \mathbf{C}]] = \sum_{r=1}^R \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r \quad (3.21)$$

Illustration of an R -component CANDECOMP/PARAFAC model on a three-way data set is given in Figure 3.6. In an alternative approach, using the matrix notation, a CANDECOMP/PARAFAC model can be written as

$$\mathbf{Y}_k = \mathbf{A} \mathbf{D}_k \mathbf{B}^T + \mathbf{E}_k \quad (3.22)$$

where \mathbf{Y}_k represents the k th frontal slice of a three-way array, and \mathbf{A} and \mathbf{B} are the component matrices in the first and second modes, respectively. \mathbf{D}_k is a diagonal matrix, whose diagonal elements are the k th row of the third component matrix \mathbf{C} . \mathbf{E}_k is the k th frontal slice of the residual.

3.2.2.1 Connection between CANDECOMP/PARAFAC Model and TUCKER3 Model

Multi-linear models, capture the multi-linear structure in data. Multi-linearity of the model denotes that the model is linear in each mode, and factors extracted from each mode are linear combinations of the variables in that mode. The extracted factors form the columns of a component matrix that summarize the structure in each mode.

As mentioned earlier, the CANDECOMP/PARAFAC summarization due to the uniqueness property makes it a popular technique in various fields. Uniqueness is achieved by imposing some restrictions on the model, that factors in different modes can only interact with the corresponding factors. A core array in multi-way models represent the interaction between factors in different modes, as given in (3.14). By assuming that each component matrix in CANDECOMP/PARAFAC decomposition has full rank, we can write (3.20) as following

$$\underline{\mathbf{Y}} = \underline{\mathbf{I}}_{3,R} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} + \underline{\mathbf{E}}, \quad (3.23)$$

where $\underline{\mathbf{I}}_{3,R}$ represents a three-way super-identity tensor of size $R \times R \times R$, the elements of which are equal to 1 when all indices are equal and 0 otherwise (Figure 3.7).

Due to the presence unrestricted core array in a Tucker3 model, $\underline{\mathbf{G}}$ in (3.14), this model is more flexible compared CANDECOMP/PARAFAC model. This flexibility is originated from the fact that in Tucker3 model a factor in a mode can interact with any factor in the other modes. On the other hand, in a PARAFAC model, a component in a certain mode can be related to a single component in another mode. While the core array enables us to explore the underlying structure of a multi-way dataset much better than a restricted CANDECOMP/PARAFAC model, full-core array structure in Tucker3 has some drawbacks. First, this property is the reason for rotational indeterminacy in Tucker3 models. Unlike CANDECOMP/PARAFAC, a Tucker3 model cannot determine component matrices uniquely. Consider rotating a component matrix by a rotation matrix, then we can

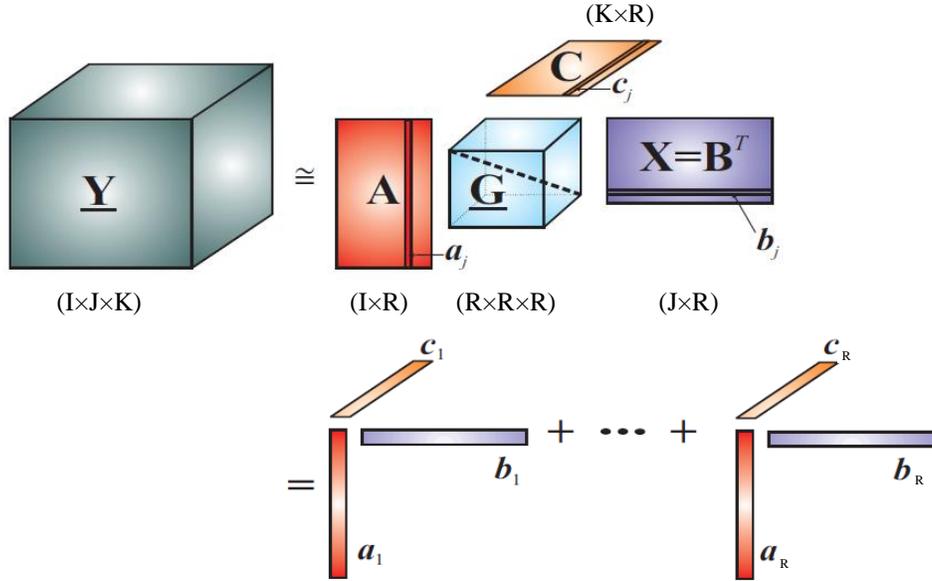


Figure 3.7: CANDECOMP/PARAFAC model with a superdiagonal core tensor $\underline{\mathbf{G}} = \mathbf{I}_{3,R}$ for the third-order tensor $\underline{\mathbf{Y}} \cong \mathbf{I}_{3,R} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} = \sum_{r=1}^R \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$.

easily cancel the rotation effect by applying the inverse of the rotation to the core array and still get the same model fit. Therefore, a Tucker3 model can determine component matrices only up to a rotation. Second, interpretation of Tucker3 models is much more difficult compared to CANDECOMP/PARAFAC models.

In a CANDECOMP/PARAFAC model, we extract the same number of components in each mode. When one component is identified as an artifact, for instance in the analysis of an EEG tensor, that particular component shows the signature of an artifact in the time domain in the first mode, in the frequency domain in the second mode and in the spatial domain in the third mode. Therefore, we actually identify an artifact using the rank-1 tensor corresponding to it. By using CANDECOMP/PARAFAC model for analysis of three-way tensor of multichannel EEG with modes of time, frequency and electrodes (space), we assume that each brain activity is characterised by certain modalities in time and frequency domains and a certain mixture of brain activities is recorded at each electrode.

3.2.2.2 Algorithms

The problem of computing CANDECOMP/PARAFAC model given a three-way tensor $\underline{\mathbf{Y}}$ and a choice for R , is to find the factor matrices $(\mathbf{A}, \mathbf{B}, \mathbf{C})$. Using the “Kruskal operator” notation defined in (3.21), the problem could be formulated as a least-squares optimization problem:

$$\min f(\mathbf{A}, \mathbf{B}, \mathbf{C}) \equiv \frac{1}{2} \|\underline{\mathbf{Y}} - \llbracket \mathbf{A}, \mathbf{B}, \mathbf{C} \rrbracket\|^2. \quad (3.24)$$

The original algorithms proposed for computing a CANDECOMP/PARAFAC model are based on the ALS approach [65]. The component matrices are estimated one at time as follow by rewriting the CANDECOMP/PARAFAC model on a three-way array $\underline{\mathbf{Y}} \in \mathbb{R}^{I \times J \times K}$ given in (3.19) and (3.20) using the Khatri-Rao product as

$$\mathbf{Y}_{(1)} = \mathbf{A}(\mathbf{C} \odot \mathbf{B})^T + \mathbf{E}_{(1)} \quad (3.25)$$

where $\mathbf{A} \in \mathbb{R}^{I \times R}$, $\mathbf{B} \in \mathbb{R}^{J \times R}$, and $\mathbf{C} \in \mathbb{R}^{K \times R}$ are the component matrices corresponding to the first, second, and third modes, respectively and the three-way arrays $\underline{\mathbf{Y}}$ and $\underline{\mathbf{E}}$ are unfolded in the first mode, and matrices of size $I \times JK$, denoted by $\mathbf{Y}_{(1)}$ and $\mathbf{E}_{(1)}$, are formed. The objective is then to minimize the least squares objective function written for a CANDECOMP/PARAFAC model as

$$\|\underline{\mathbf{Y}} - \tilde{\underline{\mathbf{X}}}\|^2 = \|\mathbf{Y}_{(1)} - \mathbf{A}(\mathbf{C} \odot \mathbf{B})^T\|_F^2. \quad (3.26)$$

where $\tilde{\underline{\mathbf{X}}}$ is the reconstructed multi-way array from the estimated factor matrices. Minimizing this function with respect to \mathbf{A} , is equivalent to finding the least squares estimate for \mathbf{A} as $\mathbf{Y}_{(1)}((\mathbf{C} \odot \mathbf{B})^T)^\dagger$ where \dagger is the Moore-Penrose pseudoinverse operator. Similarly, we can also calculate the least square estimates of other factors. Hence, the ALS-based algorithm iteratively updates these estimates until a convergence criterion, e.g., no change in model fit, is satisfied.

Algorithm 3 PARAFAC-ALS [71]($\underline{\mathbf{Y}} \in \mathbb{R}^{I \times J \times K}$, R)

$\underline{\mathbf{Y}} \in \mathbb{R}^{I \times J \times K}$, R rank of tensor, and a convergence criterion ϵ . Estimate of \mathbf{A} , \mathbf{B} , \mathbf{C} . Initialize $\mathbf{A} \in \mathbb{R}^{I \times R}$, $\mathbf{B} \in \mathbb{R}^{J \times R}$, and $\mathbf{C} \in \mathbb{R}^{K \times R}$. ϵ not satisfied $\mathbf{Z} \leftarrow \mathbf{C} \odot \mathbf{B}$. $\mathbf{A} \leftarrow \mathbf{X}_{(1)} \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1}$. $\mathbf{Z} \leftarrow \mathbf{C} \odot \mathbf{A}$. $\mathbf{B} \leftarrow \mathbf{X}_{(2)} \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1}$. $\mathbf{Z} \leftarrow \mathbf{B} \odot \mathbf{A}$. $\mathbf{C} \leftarrow \mathbf{X}_{(3)} \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1}$.

Although Algorithm 3 is the most commonly used approach to fit a CANDECOMP/PARAFAC model, it is prone to local minima and convergence to the global optimum is not guaranteed. Besides, like any other iterative approach, it is sensitive to the initialization of the factor matrices and different initializations of the algorithm may converge to different local optima. Random initialization or using generalized eigenvalue decompositions are common approaches for initializing a CANDECOMP/PARAFAC model. Furthermore, ALS suffers from slow convergence rate. Alternative algorithms have been proposed for fitting a CANDECOMP/PARAFAC model, in particular for three-way tensors, with the objective to improve ALS algorithm in terms of convergence rate and robustness to overfactoring. Acar *et al.* [72] proposed an approach using a first-order gradient-based optimization method to solve the CANDECOMP/PARAFAC optimization problem in (3.26); the performance of the method that is called OPT using a nonlinear conjugate gradient method is showed to outperform the ALS algorithm and leads to increased accuracy in the case of overfactoring. Furthermore, ALS methods need to unfold the tensor frequently and often converge slowly. It could also suffer from the so called degeneracy problem [73] where some factors diverge and at the same time tend to cancel each other as the goodness of fit progresses. CANDECOMP/PARAFAC-degeneracies occur when one attempts to approximate a tensor by another of lower rank, causing two or more factors to tend to infinity, and at the same time to almost cancel each other, giving birth to a tensor of higher rank. If cancellation between diverging rank-1 tensors cannot occur, then neither CANDECOMP/PARAFAC-degeneracy can. This is in fact what happens with tensors with positive entries. Thus, imposing the non-negativity during the execution of successive iterations is one way to prevent degeneracy in CAN-

DECOMP/PARAFAC decomposition.

ALS algorithm was compared with a number of competing algorithms in [77] including, direct trilinear decomposition (DTLD), alternating trilinear decomposition (ATLD), alternating coupled vectors resolution (ACOVER), selfweighted alternating trilinear decomposition (SWATLD), pseudo-alternating least-squares (PALS), alternating slice-wise diagonalization (ASD) and alternating coupled matrices resolution (ACOMAR). Although, ALS outperformed all of the algorithms in terms of the quality of solution, but when the computation time is a priority, ASD could be considered as an alternative.

Nearly all existing algorithms are based on carefully adapting existing optimization algorithms, see [62] for an overview of the literature. More recent developments for general tensors include work on increasing the efficiency and robustness of gradient-based and Newton-like methods [72], modifying and improving ALS [74], studying the convergence of ALS [75] and reducing the cost of the unfolding operations required during the approximation [76].

Optimization approach for fitting CANDECOMP/PARAFAC

As an alternative to the ALS approach, which tries to estimate a factor matrix one at a time by fixing the other factor matrices, all the factor matrices could be solved simultaneously using a gradient-based optimization approach (OPT). Although f in (3.24) is written as a function of matrices, it could be thought as a scalar-valued function by vectorizing and stacking the component matrices in a column vector and then it would be straight forward to drive the gradient of it. Having the derivatives determined, we can use any first order optimization technique. for example in [72] the numerical performance of a generic nonlinear conjugate gradient (NCG) method was explored. Compared to ALS, instead of alternatively estimating the factor matrices, OPT solves for all factor matrices simultaneously. This has the benefit of increased accuracy in the case of overfactoring.

On the other hand, ALS can be remarkably fast, but it is not always accurate. OPT is more accurate, but maybe not as fast as ALS.

PARAFAC decomposition based on Single Mode blind Source separation

The objective of most ALS algorithms is to minimize the fitting error exploiting some algebraic properties of the factors. However, in addition to the algebraic properties, we often have access to some a priori knowledge about the components in specific modes. For instance, components in a specific mode may have clear physical meaning such as independency, sparsity, nonnegativity or have temporal structures. For example, in brain signal processing spectral components are nonnegative and usually sparse, while temporal components are often statistically independent. By incorporating this type of knowledge and keeping in mind the special structure of PARAFAC model as in (3.25), where the matricized tensor in each mode could be expressed as the product of the corresponding components matrix by the Khatri-Rao product of other component matrices, we can extract the latent components from their arbitrary linear mixtures by using suitable constrained matrix factorization methods such as independent component analysis (ICA), sparse component analysis, and nonnegative matrix factorization (NMF).

By assuming that some a priori knowledge about the components at least in one mode is available, the components in this mode can be estimated first by incorporating standard Blind Source Separation (BSS) algorithms. Looking at the matricized expression of PARAFAC model help us to understand it better. As $\mathbf{Y}_{(1)} = \mathbf{A}(\mathbf{C} \odot \mathbf{B})^T$, the columns of $\mathbf{Y}_{(1)}$ are just linear mixtures of the columns of \mathbf{A} , hence this is the motivation of extracting \mathbf{A} first by using a proper BSS algorithm Ψ such that

$$\hat{\mathbf{A}} = \Psi(\mathbf{Y}_{(1)}) = \Psi(\mathbf{A}(\mathbf{C} \odot \mathbf{B})^T) = \mathbf{A} \mathbf{P} \mathbf{D} \quad (3.27)$$

where \mathbf{P} and \mathbf{D} are $R \times R$ permutation matrix and invertible diagonal matrix, respectively and denote the unavoidable ambiguity of permutation and scaling. Having the factor

matrix \mathbf{A} estimated, exploiting the Khatri-Rao product structure using a series of rank-1 singular value decompositions the remaining component matrices (\mathbf{B} and \mathbf{C}) are then computed. The success of this technique relies on proper priori knowledge of one mode. The method fails if such knowledge is not available or we have used incorrect priori knowledge, since (3.27) may not hold anymore. However, once such priori knowledge is available, using this method can help us achieve faster convergence speed. This method is extremely efficient and overcomes the bottleneck issue of CANDECOMP/PARAFAC decomposition which occurs when two or more components in at least one mode are highly collinear. Particularly, when the ordinary CANDECOMP/PARAFAC decompositions are not unique, more interpretable results can be achieved via this method even with physical meaning [78].

3.2.3 Numerical Complexity

In this section our goal is to give an idea of the cost per iteration of various multi-way analysis algorithms. In fact, just considering total number of iterations for comparison of iterative algorithms is not very meaningful, as the costs per iteration for each algorithm might be significantly different. Furthermore, we can reduce the complexity of the product between two matrices or other matrix operations by taking into account any possible structure. Kronecker product of matrices create some structure that can be incorporated to reduce the overall computational complexity. However, in this section we are not taking advantage of such structure and the calculations are for the general matrices.

Matrix inversion is one of the main parts of most tensor decomposition algorithms, which is implemented via singular value decomposition (SVD). First, we calculate the complexity of SVD and then derive the complexity of other algorithms based on that.

Singular Value Decomposition (SVD)

Let the so-called reduced SVD of a $I \times J$ matrix \mathbf{A} of rank r , $I \geq J \geq r$, be written as:

$$\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T = \mathbf{U}_r \mathbf{\Sigma}_r \mathbf{V}_r^T, \quad (3.28)$$

where \mathbf{U} and \mathbf{V} are orthogonal matrices, $\mathbf{\Sigma}$ is $I \times J$ diagonal, \mathbf{U}_r is a $I \times r$ submatrix of \mathbf{U} , $\mathbf{U}_r^T \mathbf{U}_r = \mathbf{I}_r$, $\mathbf{\Sigma}_r$ is $r \times r$ diagonal and \mathbf{V}_r is a $J \times r$, $\mathbf{V}_r^T \mathbf{V}_r = \mathbf{I}_r$. The calculation of the diagonal matrix $\mathbf{\Sigma}$ needs $2IJ^2 - 2J^3/3$ multiplications, if matrices \mathbf{U} and \mathbf{U} are not explicitly formed, but kept in the form of a product of Householder symmetries and Givens rotations [79]. In order to calculate matrices \mathbf{U}_r and \mathbf{V}_r , additional $5Ir^2 - r^3/3$ and $5Jr^2 - r^3/3$ multiplications are required, respectively.

For instance if $r = J$, the total complexity for computing $\mathbf{\Sigma}$, \mathbf{U}_J and \mathbf{V} is of order $7IJ^2 + 11J^3/3$. If $I \gg J$, this complexity can be decreased to $O(3IJ^2)$ by resorting to Chans algorithm [80].

High-Order SVD (HOSVD)

For a three-way array $\underline{\mathbf{Y}} \in \mathbb{R}^{I \times J \times K}$, calculating the HOSVD involves three SVD of dimensions $I \times JK$, $J \times KI$ and $K \times IJ$ respectively, where only the left singular matrix is required explicitly. Consider computing the reduced SDV of rank R_i in mode i , then the overall computational complexity for the three SVDs can be summarized as

$$2IJK(I + J + K) + 5(R_1^2 JK + IR_2^2 K + IJR_3^2) - 2(I^3 + J^3 + K^3)/3 - (R_1^3 + R_2^3 + R_3^3)/3$$

Furthermore, estimating the core tensor has extra burden and for that singular matrices also need to be contracted on the original tensor. As we need to contract it on all the modes, the complexity depends on the order in which the contractions are executed. Assuming $I \geq J \geq K \geq R_i$, it is better to contract the third mode first and the first mode last. This computation requires $IR_1 JK + R_1 JR_2 K + R_1 R_2 KR_3$ additional multiplications.

Alternating Least Square (ALS)

Considering the complexity of one iteration of ALS for estimating the factor matrix in the first mode. For that we need to solve the linear system below for \mathbf{A} in the LS sense according to (3.26):

$$\mathbf{Z}\mathbf{A}^T = \mathbf{T} \quad (3.29)$$

where the dimensions of \mathbf{Z} , \mathbf{A} , and \mathbf{T} are $m \times r$, $q \times r$ and $m \times q$ respectively. First, we need to compute matrix $\mathbf{Z} = \mathbf{B} \odot \mathbf{C}$ which is the Khatri-Rao product two rank- R matrices and requires JKR multiplications. Then the SVD of matrix \mathbf{Z} of size $JK \times R$ needs $7JKR^2 + 11R^3/3$ multiplications, according to the previous part. Finally the product $\mathbf{V}_R \Sigma_R^{-1} \mathbf{U}_R^T \mathbf{T}$ represents $RJKI + RI + R^2I$ multiplications.

The total cumulated operations for one iteration of ALS algorithm for a three-way array is as follows,

$$(JK + KI + IJ)(7R^2 + R) + 3RIJK + (I + J + K)(R^2 + R) + 11R^3$$

where the two last terms are often negligible.

The convergence of the ALS algorithm sometimes may take up a very large number of iterations. Good initialization of the algorithm, in some cases, help to reach the global minimum very quickly. However, it is not always true and sometimes even trying any starting point does not help us to reach a global minimum quickly via ALS. Some reasons for this type of behaviour could be related to the solution that is embedded in a deep swamp, or is in fact unreachable at the solution rank, and can only be approached through an infinite series of diverging sets of loadings. When the convergence is slow we can use a variant of ALS, called line search with the idea of limiting the number of iterations of a given cycle in the convergence path. This is done by predicting the value of the loading factors a certain number of iterations ahead by computing a sort of linear regression and optimizing the coefficients. The problem then involves solving a system of three polynomials with three unknowns, which leads to a high numerical complexity that

for the simplest case leads to solving a polynomial of degree 6 in R for a three-way array. According to [81], the Enhanced Line Search (ELS) adds $(8R + 10)IJK$ multiplications to each iteration of ALS. However, ELS makes the number of iterations decrease, hence allowing a decrease of the overall complexity.

3.3 Chapter Summary

In many applications the data structures often contain higher-order contextual information, and representing the data via multi-way arrays enables us to retain the structure of the data and analyze all the dimensions concurrently using multi-way analysis techniques. Moreover, it is known that the underlying information content of the data may not be captured accurately or identified uniquely by two-way data analysis. In this chapter the multi-way array and the corresponding analysis explored for extracting hidden structures and capturing underlying correlations between variables and summarize the data into unique and physically meaningful components. CANDECOMP/PARAFAC and TUCKER are two common multi-way models that are used in various applications. We choose CANDECOMP/PARAFAC for EEG analysis as it is more favourable due to its uniqueness property and better interpretability of the resulting components. Furthermore, various decomposition techniques for CANDECOMP/PARAFAC model are presented. ALS is the most popular algorithm due to its simplicity and satisfactory performance. The computational complexity of ALS algorithm for CANDECOMP/PARAFAC decomposition is also explored. Most of the decomposition methods require that the rank (model order) of the tensor to be known a priori, which is not available in practical application. Hence, optimum rank selection is an active challenge in this context and various techniques have been proposed for that in the literature. We cover the state-of-the-art techniques for CANDECOMP/PARAFAC order selection and also our proposed approach in the next chapter.

Chapter 4

CANDECOMP/PARAFAC Model Order Selection

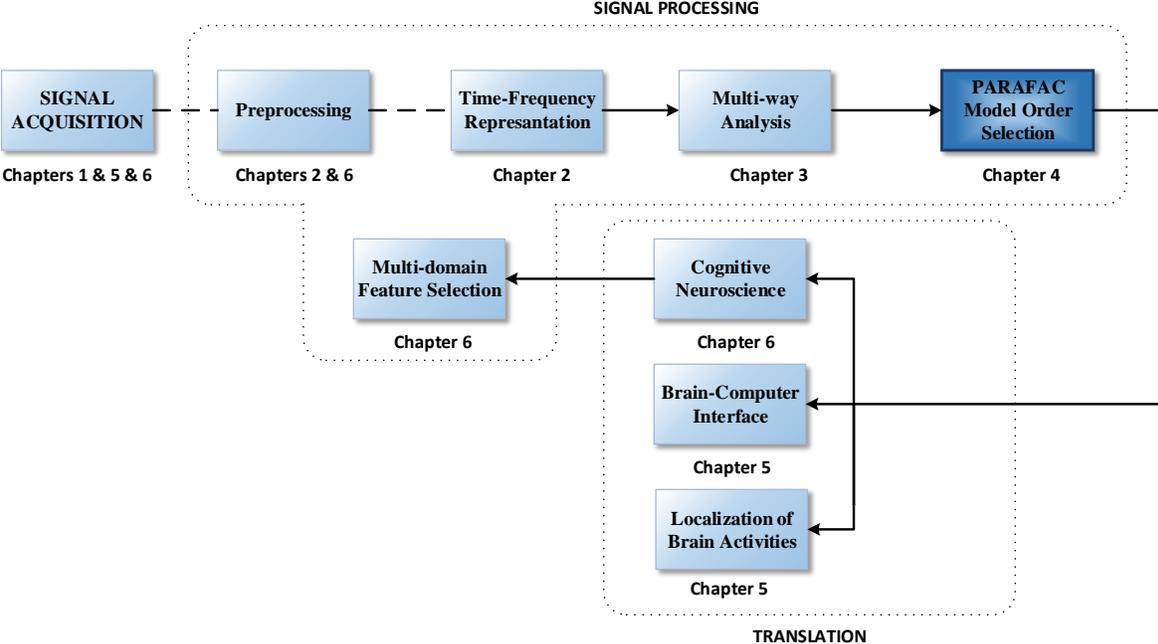


Figure 4.1: Chapter 4- CANDECOMP/PARAFAC Model Order Selection.

One of the main reasons behind the rapid increase in popularity of Multi-way (Tensor) analysis, as evidenced by numerous research areas, is that underlying latent informa-

tion content of the high-dimensional data may not be captured accurately or identified uniquely by two-way data analysis. The multi-way data analysis techniques are generalization of two-way (matrix) data analysis techniques based on the idea of linear factor models [60, 62]. Tensor analysis seeks to extract a more interpretable data summarization by exploring the multi-linear relationships between the variables used to represent the data. CANDECOMP/PARAFAC [71], also known as Canonical Polyadic Decomposition (CPD), is a well-known and common tensor factorization model. This approach has been applied in many areas such as chemometrics [82], wireless communication [83, 84], analysis of fMRI data [85], Electroencephalogram (EEG) signal processing [53–55], data mining and web analysis [86], classification, and clustering [87, 88].

In a CANDECOMP/PARAFAC model, a tensor can be decomposed as a linear combination of rank-one tensors. An N th-order rank-one tensor is a tensor that can be written as the outer product of N non-zero vectors. Each of these extracted rank-one tensors corresponds to a component or a factor of the data that itself could be uniquely characterized by its associated vectors. For example, in the context of EEG signal processing, when we generate a three-way Time-Frequency-Space EEG tensor, the observed EEG tensor could be decomposed using CANDECOMP/PARAFAC into rank-one EEG tensors. Each of the extracted components technically representing an event in EEG that could be uniquely characterized by its time, frequency and spatial signatures [53].

In order to capture the true underlying latent structure of observed data, it is crucial to extract the right number of signal components in a CANDECOMP/PARAFAC model. Knowing that underestimating or overestimating the number of signal components may mislead us in interpreting the data, estimation of the tensor rank or in other words, the optimum number of rank-one tensors has been addressed in several CANDECOMP/PARAFAC modeling problems [67, 88–90]. The essence of effective model order selection and the situations in which overestimation or underestimation of model order is preferable is discussed in [91]. In some applications like radar imagery a detection method biased toward overestimation is preferable as the cost of missing a target is much higher

than a false alarm. However, in most applications underestimation is favourable in low SNRs as using an overestimated source number yields inaccurate parameter estimates which necessitate overwhelming costs.

Moreover, in many applications this problem is more challenging in the presence of colored noise or interference. As the signal information is extracted based on its correlation structure assuming the noise samples to be uncorrelated, the presence of colored noise with correlated samples, seriously affects the signal parameter estimation performance. Prewhitening is an effective way to tackle this problem, which decorrelates the noise samples, therefore minimizing the disturbance between the noise and signal subspaces, and consequently avoiding the performance degradation.

Prewhitening requires that an estimate of the noise covariance matrix to be available. This covariance matrix is characterized by numerous parameters in the case of a non-parametric noise model. Hence, for a high-dimensional data, a large number of samples are required to estimate the noise covariance matrix properly. However, by assuming a parametric structure for the noise, we can reduce the number of required parameters and also the complexity of the noise characterizing method [92, 93]. Here we assume that the noise has a Kronecker structure. Estimating the noise parameters with Kronecker structure has been addressed in many works such as [94–96]. In EEG experiments, the measured signal usually has a very low signal-to-noise ratio (SNR). This could be due to low signal strength that is coupled with noise from different sources, such as the EEG acquisition system, the brain, or external environment. Brain noise, from ongoing or non-stimulus correlated activity, is often the dominating noise in EEG recordings and has a complicated structure, with correlations in both time and space. These correlations are usually reflected in temporal and spatial covariance matrices. A model that combines these two covariance matrices using the Kronecker product has been shown to be satisfactory and effective in the EEG noise modeling context [97–99]. Moreover, the Kronecker structured covariance matrix has practical applications in other fields like sensor array processing, Multiple-Input Multiple-Output (MIMO) wireless systems [100],

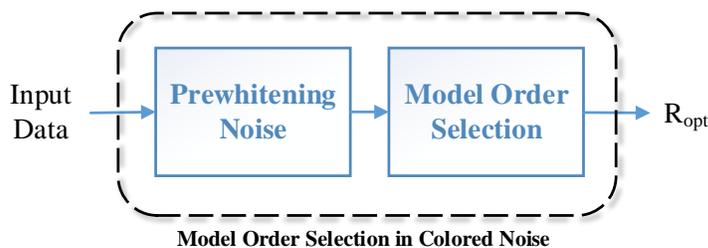


Figure 4.2: Block Diagram of the proposed scheme for model order selection in presence of colored noise which is combination of the multidimensional prewhitening with any model order selection scheme.

and genomics [101], due to its efficient characterization.

After Prewhitening, all the parameter estimation and model order selection methods, developed with the assumption of white noise, could be exploited for analysis. Then, if needed, through a dewhitening step, the original signal subspace could be retrieved. The combination of model order selection schemes with prewhitening schemes was first used in [102]. Here, we are extending it for the multidimensional case (Figure 4.2).

There are several techniques for determining the number of signal components of the CANDECOMP/PARAFAC model in the literature, for example, residual analysis [67], convex-hull-based method [68], CORE CONSistency DIAGNOSTIC (CORCONDIA) [89, 103, 104] that are mostly based on Alternating Least Squares (ALS) algorithm. Although, there are other CANDECOMP/PARAFAC decomposition methods that may be in some situations more efficient than ALS, (e.g. [78, 90, 105–110]), some of them even with the capability of automatic rank determination or even handling the Kronecker structure colored noise, i.e. Closed-Form PARAFAC based Model Order Selection (CFP-MOS) method [111]. Nevertheless, those methods are not in the scope of this work and we are mainly focusing on the ALS-based CANDECOMP/PARAFAC approaches due to its programming simplicity and the satisfactory performance [112]. Furthermore, it is often suggested to use several diagnostic tools together rather than a single method, as there is no rule of thumb in determining the optimal number of signal components (optimal in terms of interpretation) for real data [91].

In this Chapter, our goal is to propose a novel, theoretical, robust, and consistent method, based on the available data, for the ALS-based CANDECOMP/PARAFAC model order selection that can handle additive colored noise with the Kronecker structure. The rest of the chapter is organized as following: First the rank selection problem formulation and statement are explained followed by a brief overview of the rank selection techniques. Then, our proposed technique is explained.

4.1 Problem Formulation

An N -way tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ is of rank one if it can be written as the outer product of N nonzero vectors, i.e.,

$$\underline{\mathbf{X}} = \mathbf{a}^{(1)} \circ \mathbf{a}^{(2)} \circ \dots \circ \mathbf{a}^{(N)} \quad (4.1)$$

where the symbol \circ denotes the vector outer product and each element of the tensor $\underline{\mathbf{X}}$ is the product of the corresponding vector elements. In other words, $\underline{\mathbf{X}}$ is of rank one, if and only if $x_{i_1 i_2 \dots i_N} = a_{i_1}^{(1)} a_{i_2}^{(2)} \dots a_{i_N}^{(N)}$.

In a CANDECOMP/PARAFAC model, a tensor is decomposed into the sum of rank-one tensors. For simplicity and without loss of generality here we only consider a three-way tensor. Let $\mathbf{a}, \mathbf{b}, \mathbf{c}$ be column vectors of size $I \times 1, J \times 1$ and $K \times 1$, respectively, and $\underline{\mathbf{X}}$ be a tensor of size $I \times J \times K$, therefore, an R^* -component CANDECOMP/PARAFAC model for a three-way array of noiseless data $\underline{\mathbf{X}} \in \mathbb{R}^{I \times J \times K}$ can be written as

$$\underline{\mathbf{X}} = \sum_{r=1}^{R^*} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r, \quad (4.2)$$

where R^* is the true number of signal components in the model (True Rank), $\mathbf{a}_r, \mathbf{b}_r,$ and \mathbf{c}_r indicate the r th column of component matrices $\mathbf{A} \in \mathbb{R}^{I \times R^*}, \mathbf{B} \in \mathbb{R}^{J \times R^*},$ and $\mathbf{C} \in \mathbb{R}^{K \times R^*},$ respectively. By assuming that each component matrix in CANDECOMP/PARAFAC decomposition has full rank, we can also write (4.2) in terms of

n -mode products as follows:

$$\underline{\mathbf{X}} = \mathbf{I}_{3,R^*} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C}. \quad (4.3)$$

In practice the observed data is a contaminated version of $\underline{\mathbf{X}}$ and can be represented as

$$\underline{\mathbf{Y}} = \underline{\mathbf{X}} + \underline{\mathbf{E}}^{(c)}, \quad (4.4)$$

where $\underline{\mathbf{E}}^{(c)} \in \mathbb{R}^{I \times J \times K}$ is a three-way array containing the additive colored noise samples, which are typically correlated with each other in several modes and has the white noise as its special case.

Here we focus on the multidimensional colored noise with a Kronecker structure, that is characterized by the following correlation structure:

$$\mathbf{Cov} = \mathbf{Cov}_3 \otimes \mathbf{Cov}_2 \otimes \mathbf{Cov}_1, \quad (4.5)$$

where $\mathbf{Cov} \in \mathbb{R}^{M \times M}$ with $M = I \times J \times K$ is the noise covariance matrix in all dimensions and $\mathbf{Cov}_1 \in \mathbb{R}^{I \times I}$, $\mathbf{Cov}_2 \in \mathbb{R}^{J \times J}$ and $\mathbf{Cov}_3 \in \mathbb{R}^{K \times K}$ are the covariance matrices in mode-1, mode-2 and mode-3 respectively which are defined as [93]:

$$\mathbf{Cov} \triangleq \frac{1}{\sigma_n^2} \mathbf{E}\{\text{vec}(\underline{\mathbf{E}}^{(c)}) \cdot \text{vec}(\underline{\mathbf{E}}^{(c)})^T\}, \quad (4.6)$$

$$\mathbf{Cov}_r \triangleq \frac{1}{\alpha_r} \mathbf{E}\{\mathbf{E}_{(r)}^{(c)} \cdot (\mathbf{E}_{(r)}^{(c)})^T\}, \quad (4.7)$$

where $r = 1, 2, 3$, and α_r is the scaling coefficient and σ_n^2 is the noise power after prewhitening. In the context of EEG signal processing, the noise free EEG events could be well characterized by Time-Frequency-Space EEG tensors [53]. However, EEG is typically contaminated with strong background noise that is correlated in both time and across channels. It has been found that the joint spatio-temporal covariance ma-

trix can be modelled as the Kronecker product of the spatial and temporal covariance matrices [97].

Applying the Eigenvalue decomposition (EVD) or the Cholesky decomposition to \mathbf{Cov} and \mathbf{Cov}_r , $r = 1, 2, 3$, yields [93]

$$\mathbf{Cov} = \mathbf{L}\mathbf{L}^T, \quad \mathbf{Cov}_r = \mathbf{L}_r\mathbf{L}_r^T, \quad r = 1, 2, 3, \quad (4.8)$$

where $\mathbf{L} \in \mathbb{R}^{M \times M}$ is the noise correlation factor in the joint dimensions and $\mathbf{L}_1 \in \mathbb{R}^{I \times I}$, $\mathbf{L}_2 \in \mathbb{R}^{J \times J}$ and $\mathbf{L}_3 \in \mathbb{R}^{K \times K}$ are the correlation factors in mode-1, mode-2 and mode-3 respectively. It can be verified [93] that $\mathbf{L} = \mathbf{L}_3 \otimes \mathbf{L}_2 \otimes \mathbf{L}_1$ and the multidimensional colored noise $\underline{\mathbf{E}}^{(c)}$, can be related to the i.i.d white Gaussian noise tensor according to

$$\underline{\mathbf{E}}^{(c)} = \underline{\mathbf{E}} \times_1 \mathbf{L}_1 \times_2 \mathbf{L}_2 \times_3 \mathbf{L}_3, \quad (4.9)$$

where $\underline{\mathbf{E}} \in \mathbb{R}^{I \times J \times K}$ is a tensor with samples from zero-mean i.i.d Gaussian distribution with variance of σ_n^2 .

4.1.1 Problem Statement: Rank Approximation

The rank approximation of a noisy tensor $\underline{\mathbf{Y}}$, is defined according to (4.10) as, when the R^* is unknown, the R for which a criterion $f(\underline{\mathbf{Y}}, R)$ is minimized, such that $\underline{\mathbf{Y}}$ is expressed as a sum of R rank-one tensors,

$$\hat{R} := \underset{R}{\operatorname{argmin}} \left\{ f(\underline{\mathbf{Y}}, R) \mid \underline{\mathbf{Y}} \cong \sum_{r=1}^R \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r \right\}. \quad (4.10)$$

Finding the best R -component CANDECOMP/PARAFAC model that approximates the observed noisy tensor $\underline{\mathbf{Y}}$ well is a challenge in this regard. While most of the criteria developed in this context are ad-hoc, here our goal is to provide a robust theoretically based criterion for rank approximation in the presence of colored noise, which relies on characteristics of available data.

4.2 Rank Selection Techniques

Model order selection techniques can be categorized into two methods: feed-forward and feed-backward. While in the former category, the parameters of the model and the model order itself are estimated simultaneously, (e.g. [78, 113]), or the model order is estimated first and then parameters of the model for that order are estimated separately, (e.g. [103, 111]). In the later category, first the parameters of the model are estimated for a range of possible model orders, and then in a backward direction, suitability of the considered model orders according to a criterion is evaluated and the optimum model order is selected [104]. Figure 4.3 shows the schematic of these two categories.

While each of the above mentioned categories has its own advantages and disadvantages, the feed-forward methods are often computationally more efficient than the feed-backward methods as decomposing and estimating the factor matrices of the corresponding CANDECOMP/PARAFAC model, (i.e. $\mathbf{A}, \mathbf{B}, \mathbf{C}$) is done just for the priorly estimated model order. However, feed-forward methods do not have post evaluation capability; In other words, the estimated factor matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$ have no control over the order selection. In addition, most of them cannot validate and compare estimated factors for different orders. On the other hand, the feed-backward model order selection methods, use the complete resulted estimates of CANDECOMP/PARAFAC, $\mathbf{A}, \mathbf{B}, \mathbf{C}$, in validation of their criteria. Furthermore, the selecting criterion in most of the cases has a meaningful interpretation (e.g. Mean Squared Error (MSE)), which is good from the validation point of view and the noteworthy advantage of the feed-backward methods over the feed-forward ones.

4.2.1 Existing CANDECOMP/PARAFAC model rank selection techniques with white noise

Most of the CANDECOMP/PARAFAC model order selection methods start by computing a set of candidate model orders, which is assumed to contain the true model

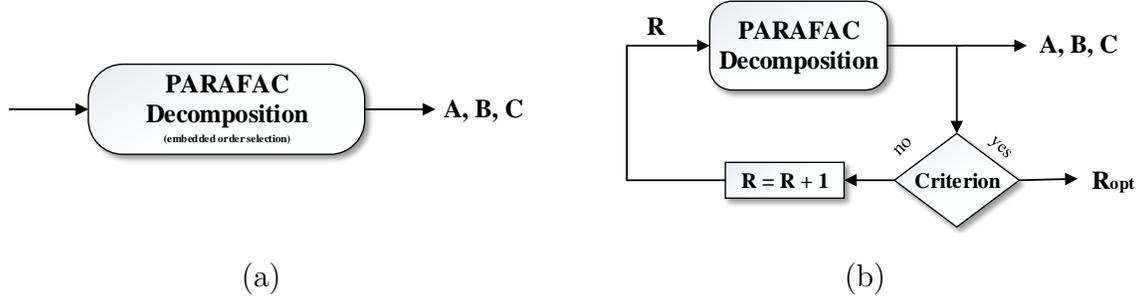


Figure 4.3: Schematic of (a) feed-forward and (b) feed-backward CANDECOMP/PARAFAC model order selection techniques.

order, and then based on an appropriate model selection criteria, the optimum number of signal components is determined. Some of these methods include DIFFIT [67], CORCONDIA [89], Threshold-CORCONDIA [103], and a Convex-hull based method [68].

4.2.1.1 Core Consistency Diagnostic (CORCONDIA)

CANDECOMP/PARAFAC can be written as a special Tucker3 model where the core is superdiagonal with ones on the diagonal and zeros elsewhere. This special way of writing the model can be used to check the adequacy of a CANDECOMP/PARAFAC model by estimating what Tucker3 core is found if estimated unconstrained from the CANDECOMP/PARAFAC loadings. The core consistency [89] is given as the percentage of variation in this core array consistent with the theoretical superdiagonal array. The expression of CORCONDIA (CC) is as follows

$$CC(R) = 100 \left(1 - \frac{\sum_{i=1}^R \sum_{j=1}^R \sum_{k=1}^R (g_{ijk} - t_{ijk})^2}{R} \right) \quad (4.11)$$

where g_{ijk} and t_{ijk} are elements of the estimated Tucker3 core and the superdiagonal core array, respectively according to (3.14).

The maximum core consistency (100) is thus found according to (4.11). Consistencies found well below 70 to 90 indicate that either too many components are used or the model is otherwise mis-specified. The consistency can also become negative which means

that the model is not reasonable. Note that core consistency is an ad-hoc method. It often works well on real data, but not as well with simulated data. CORCONDIA does not provide proof of dimensionality, but it can give a good indication. The estimate of CORCONDIA, denoted as \hat{R}_{CC} , is given by

$$\hat{R}_{CC} = \max R \text{ subject to } CC(R) \geq \eta \quad (4.12)$$

where $0 < \eta < 1.0$ is the threshold coefficient (typically $70\% < \eta < 90\%$ is used). The CORCONDIA measure is almost 100% for both true and underestimated model orders, and then starts to drop for overestimation of model order. So, having the estimate of the component matrices for a given model order using CANDECOMP/PARAFAC, we estimate the core array based on the Tucker3 model, and then compare it with a same size super-diagonal core array to see how good the model is fitted. If the Frobenius norm of the difference is less than a threshold the model fit is good. We do the same process by increasing the model order up to a point that the distance between those two core arrays are increased further than the threshold [89]. The CORCONDIA has been commonly applied in the literature and is known as a benchmark for CANDECOMP/PARAFAC model order selection. The problem with this method is that the effectiveness of the CORCONDIA is highly determined with a subjective threshold that should be increased with Signal-to-Noise ratio (SNR) to keep a consistent performance. In Threshold-CORCONDIA [103] the authors have tried to tackle this problem by considering the difference of the core consistency for two adjacent candidate ranks as a means of rank estimation. This method also requires a thresholding, and the optimal threshold coefficients are selected using the probability of detection, but the problem is that, still a common threshold is used for a wide range of noise levels which does not improve the performance in all SNRs and furthermore, the computation of the threshold adds to the complexity of the method.

The above mentioned methods are all based on the idea of misfit or the goodness-of-fit. From another perspective, several model selection methods based on the r-mode

unfoldings are also developed. RD-RMT [114] tackles the CANDECOMP/PARAFAC model rank selection by extending the classical r-mode matrix unfolding to multi-mode matrix unfolding and through a sequence of nested hypothesis tests, in the framework of random matrix theory, the true model rank is enumerated. Furthermore, some model selection methods based on Higher Order Singular Value Decomposition (HOSVD) were proposed in [111]: R-Dimensional Exponential Fitting Test (R-D EFT), R-D AIC, and R-D MDL.

4.2.1.2 R-D Exponential Fitting Test

The eigenvalues of the sample covariance matrix play a major role in the model order estimation. It is showed that the noise eigenvalues exhibiting a Wishart profile can have their profile approximated by an exponential curve [115]. Let $\mathbf{X} \in \mathbb{R}^{M \times N}$ be the observation matrix and λ_i be the i -th eigenvalue of the sample covariance matrix. The exponential model may be expressed as

$$E\{\lambda_i\} = E\{\lambda_1\} \cdot q(\alpha, \beta)^{i-1}, \quad (4.13)$$

where $E\{\cdot\}$ is the expectation operator and we assume that the eigenvalues are sorted so that λ_1 is the largest and we at least have one noise eigenvalue. The term $q(\alpha, \beta)$ is heuristically defined as

$$q(\alpha, \beta) = \exp \left\{ -\sqrt{\frac{30}{\alpha^2 + 2} - \sqrt{\frac{900}{(\alpha^2 + 2)^2} - \frac{720\alpha}{\beta(\alpha^4 + \alpha^2 - 2)}}} \right\} \quad (4.14)$$

so that $0 < q(\alpha, \beta) < 1$ and $\alpha = \min\{M, N\}$ and $\beta = \max\{M, N\}$. Then an analytic expression of the ordered eigenvalues profile, obtained under noise only hypothesis, is used to fit the eigenvalues of the noise [115].

Basically the method is based on the observation that, in noise-only case, the profile of ordered eigenvalues can be well approximated by a decaying exponential. The strategy

for model order selection is simple: it consists in looking for a break in profile by comparing observed profile and the theoretical noise-only one. Assume d non-coherent sources corrupted by additive noise and let P denote the number of candidate noise eigenvalues. The basic idea behind the EFT method is to choose the highest P for which the candidate noise eigenvalues can be well fitted to the theoretical decaying exponential. Moreover, the method assumes that there is at least one noise eigenvalue, i.e., $d < M$. For each value of P in the range $1 < P < M - 1$ we find the parameters of the decaying exponential which best fit the observations and we compute the prediction for the $(M - P)$ -th eigenvalue, which we denote by $\hat{\lambda}_{M-P}$. Thereby, it is possible to obtain $M - 1$ predicted eigenvalues, namely $\hat{\lambda}_{M-P}$ for $1 < P < M - 1$. To decide whether the $(M - P)$ -th eigenvalue λ_{M-P} fits to the exponential profile we measure its relative distance to the predicted eigenvalue $\hat{\lambda}_{M-P}$. By setting a threshold η_P , which could be obtained by Monte Carlo simulations carried out in the noise-only case, the following hypotheses is formulated:

$$\begin{aligned}
H_{P+1} & : \lambda_{M-P} \text{ is a noise EV, } \frac{\lambda_{M-P} - \hat{\lambda}_{M-P}}{\hat{\lambda}_{M-P}} \leq \eta_P \\
\bar{H}_{P+1} & : \lambda_{M-P} \text{ is a signal EV, } \frac{\lambda_{M-P} - \hat{\lambda}_{M-P}}{\hat{\lambda}_{M-P}} > \eta_P
\end{aligned} \tag{4.15}$$

The largest P for which the test H_{P+1} , fails, denoted by P_{opt} determines the estimated model order $d = M - P_{opt}$.

R-D EFT method is the extension of EFT for high-dimensional data. By having different r -mode unfolding, we can apply the EFT to each of the R observation matrices obtained after unfolding in the associated mode. Then the link to R-D EFT is in the way that eigenvalues in different modes are combined, which is done by defining the Global Eigenvalues as follows: Let us define $M = \prod_{r=1}^R I_r$. For the r -mode unfolding the sample covariance matrix is defined as

$$\hat{\mathbf{R}}_{\mathbf{xx}}^{(r)} = \frac{I_r}{M} \mathbf{X}_{(r)} \cdot \mathbf{X}_{(r)}^T \in \mathbb{R}^{I_r \times I_r}. \tag{4.16}$$

The global eigenvalues are based on the r -mode eigenvalues represented by $\lambda_i^{(r)}$ for $r = 1, \dots, N$ and for $i = 1, \dots, I_r$. For the sake of simplicity, let us first assume that $I_1 = I_2 = \dots = I_N$. Then we can define *global eigenvalues* as being

$$\lambda_i^{(G)} = \lambda_i^{(1)} \cdot \lambda_i^{(2)} \dots \cdot \lambda_i^{(N)} \quad (4.17)$$

Therefore, based on (4.13), it is straight forward that the noise global eigenvalues also follow an exponential profile, since

$$E\{\lambda_i^{(G)}\} = E\{\lambda_1^{(G)}\} \cdot (q(\alpha_1, \beta_1) \cdot \dots \cdot q(\alpha_N, \beta_N))^{i-1}. \quad (4.18)$$

Using the product of eigenvalues in the definition of global eigenvalues has another benefit which increases the gap between the noise and actual eigenvalues. Therefore, the break in the profile is easier to detect via global eigenvalues than using only one mode eigenvalues. To cope with the fact that in many applications the size of the N dimensions might differ, without loss of generality, let us consider the case in which $I_1 > I_2 > \dots > I_N$. We start by estimating d with the modified EFT method considering the first unfolding only. If $d < I_2$, we could have taken advantage of the second mode as well. Therefore, we compute the global eigenvalues $\lambda_i^{(G)}$ as (4.18) for $1 < i < I_2$, thus discarding the $I_1 - I_2$ last eigenvalues of the first mode. We can obtain a new estimate d . If $d < I_3$ we could continue in the same fashion, by computing the global eigenvalues considering the first 3 modes. Clearly, the full potential of the proposed method can be achieved when all modes are used to compute the global eigenvalues. This happens when $d_{opt} < I_N$, so that $\lambda_i^{(G)}$ can be computed for $1 \leq i \leq I_N$.

Having the thresholds calculated, the computational complexity of the R-D EFT method is comparable to R-D AIC and R-D MDL. The method is also limited to additive white Gaussian noise, however, it is reported to be more robust and consistent compared to the state-of-the-art model order selection techniques [111].

4.2.1.3 Second Order sTatistic of the Eigenvalues (SORTE)

SORTE [88], is based on the profile of eigenvalues of the sample covariance matrix of the unfolded tensor in one mode. SORTe is based on the second order statistics of difference of the consecutive eigenvalues, which makes it free of any required threshold. By computing the sample covariance matrix of mode- n unfolded tensor and by extracting its eigenvalues, compute the differences of eigenvalues $\nabla \hat{\lambda}_t = \hat{\lambda}_t - \hat{\lambda}_{t+1}$, $t = 1, \dots, I_n - 1$. Then, for a correct number of components R , we have

$$\nabla \hat{\lambda}_{R+1} = \nabla \hat{\lambda}_{R+2} = \dots = \nabla \hat{\lambda}_{I_n-1} = 0. \quad (4.19)$$

To compute the gap between eigenvalue, compute the variance of the sequence $\{\nabla \hat{\lambda}_i\}_{i=k}^{I_n-1}$ as

$$\hat{\sigma}_k^2 = \frac{1}{I_n - k} \sum_{i=k}^{I_n-1} \left(\nabla \hat{\lambda}_i - \frac{1}{I_n - k} \sum_{i=k}^{I_n-1} \nabla \hat{\lambda}_i \right)^2, \quad (4.20)$$

where $k = 1, \dots, I_n - 1$. From (4.19) and (4.20), it is easy to check that

$$\begin{cases} \hat{\sigma}_k^2 > 0, & k = 1, \dots, R, \\ \hat{\sigma}_k^2 = 0, & k = R + 1, \dots, I_n - 1. \end{cases} \quad (4.21)$$

Further, define a Second Order sTatistic of the Eigenvalues (SORTE) as follows:

$$SORTE(k) = \begin{cases} \frac{\hat{\sigma}_{k+1}^2}{\hat{\sigma}_k^2}, & \hat{\sigma}_k^2 > 0, \\ +\infty, & \hat{\sigma}_k^2 = 0, \end{cases} \quad (4.22)$$

where $k = 1, \dots, I_n - 2$. Then, from the definition (4.22), we have

$$\begin{cases} SORTE(k) > 0, & k = 1, \dots, R - 1, \\ SORTE(k) = 0, & k = R, \\ SORTE(k) = +\infty, & k = R + 1, \dots, I_n - 3, \\ SORTE(k) = 0, & k = I_n - 2, \end{cases} \quad (4.23)$$

According to (4.23), we can perform the model selection by the following criterion:

$$\hat{R} = \underset{k=1, \dots, I_n-3}{\operatorname{argmin}} \operatorname{SORTE}(k). \quad (4.24)$$

SORTE is computationally more efficient than R-D EFT. However, unlike R-D EFT that uses the tensor structure to combine the eigenvalues, SORTTE suffers from the lack of exploiting the information of different modes and needs a further step to somehow combine the extracted model orders from different modes of the observed tensor. Besides, SORTTE tends to overestimate the model in low SNRs and is only applicable when the White Gaussian Noise is present.

4.2.1.4 Approximate Rank-Detecting Factorization

Finding common linear subspaces in data is a classical and well-studied problem in Machine Learning and the applied sciences. A standard way to address these problems algorithmically is jointly diagonalizing the matrices. However, in some scenarios a complete diagonalization may not be necessary, and the problem is finding a single common rank one constituents of the matrices. A new framework called AROFAC2 [113] can be used for CANDECOMP/PARAFAC decomposition and rank selection. AROFAC2 algorithm uses the intrinsic algebraic structure of a low-rank degree tensor in the calculations, and reduces determination of rank to a clustering problem. The method avoids spurious components, and is stable with respect to outliers and non-Gaussian noise, however its performance degrades with increasing the rank of the tensor.

Most of the CANDECOMP/PARAFAC feed-forward model order selection techniques use the well known ALS as the core of decomposition algorithm which suffers from its high computational cost and the so called degeneracy problem when the collinearity of the factors in at least one mode of observed tensor is high. Although, there exists some feed-forward methods like CP-SMBSS and deflation method [78, 107] and even CFP-

MOS [111], that are not based on ALS and avoids the estimation of spurious components when the rank is overestimated, feed-backward methods based on variants of ALS, aiming to mitigate the degeneracy problem by adding some constraints on the extracted factors, are still widely used and therefore requires attention.

Our contribution here is to provide a consistent feed-backward method for CANDECOMP/PARAFAC model order selection with additional advantages. The proposed method not only has the ability of handling colored noise, but also, the criterion for different orders is a quantitative measure that can evaluate the models for each specific order. For example, if the true order is 10, our algorithm not only finds the value 10, but also provides the trade-off to settle for a lower dimension. This is a practical issue where in some modeling problem, some error could be tolerated for the purpose of using a lower dimensional model. For example, if tolerance ϵ is acceptable, our reconstruction error can show what is the lowest order within the tolerance of ϵ error, which in this case may be a model with an order of less than 10.

4.3 Proposed Rank Selection approach based on Reconstruction Error

In applications where we encounter colored noise, the rank estimation problem could be tackled, first, by whitening the noise and then using the conventional rank estimation approaches. Having the noise whitened, any of the previously mentioned methods, developed with the additive white noise assumption, could be exploited for the rank selection.

A multidimensional whitening technique, developed in [93], is used here for the prewhitening step. The prewhitening method in [93] assumes that the true CANDECOMP/PARAFAC model rank is available. However, using a similar approach, here our goal is to propose an algorithm that simultaneously does the prewhitening and rank selection, which is novel in this context.

4.3.1 Multidimensional Prewhitening

Having either prior knowledge about the true noise covariances or by using the available noise only measurement tensor $\underline{\mathbf{E}}^{(c)}$, the sample covariance matrix (SCM) in each mode could be calculated according to (4.7)

$$\hat{\mathbf{Cov}}_{\mathbf{r}} = \frac{1}{\beta_r} \mathbf{E}_{(r)}^{(c)} \cdot (\mathbf{E}_{(r)}^{(c)})^T, \quad r = 1, 2, 3 \quad (4.25)$$

where β_r is chosen such that $tr(\hat{\mathbf{Cov}}_{\mathbf{r}}) = N_r$ where N_r is the size of the corresponding mode, i.e. I, J, K for mode-1, mode-2 and mode-3 respectively. There are other approaches for estimating the covariance matrices with Kronecker structure including the maximum likelihood estimator of the factor matrices that leads to an iterative alternating algorithm called the flip-flop algorithm [99], or the recent algorithm developed for estimation of rank deficient covariance matrices [116].

Having the SCMs estimated, we can obtain an estimate of the correlation factors $\hat{\mathbf{L}}_{\mathbf{r}}$ by applying an EVD, i.e., $\hat{\mathbf{Cov}}_{\mathbf{r}} = \mathbf{Q}_{\mathbf{r}} \cdot \mathbf{\Lambda}_{\mathbf{r}} \cdot \mathbf{Q}_{\mathbf{r}}^T$ which yields $\hat{\mathbf{L}}_{\mathbf{r}} = \mathbf{Q}_{\mathbf{r}} \cdot \mathbf{\Lambda}_{\mathbf{r}}^{1/2}$.

The prewhitening is then performed by

$$\underline{\mathbf{Y}}' = \underline{\mathbf{Y}} \times_1 \hat{\mathbf{L}}_1^{-1} \times_2 \hat{\mathbf{L}}_2^{-1} \times_3 \hat{\mathbf{L}}_3^{-1}, \quad (4.26)$$

where the inverse of $\hat{\mathbf{L}}_{\mathbf{r}}$ can be obtained as $\hat{\mathbf{L}}_{\mathbf{r}}^{-1} = \mathbf{\Lambda}_{\mathbf{r}}^{-1/2} \cdot \mathbf{Q}_{\mathbf{r}}^T$. In case that any of the correlation factors are rank deficient, their pseudo-inverse can be used. Using (4.4),(4.9) and the n-mode product properties, we can write (4.26) as

$$\begin{aligned} \underline{\mathbf{Y}}' &= \underline{\mathbf{X}} \times_1 \hat{\mathbf{L}}_1^{-1} \times_2 \hat{\mathbf{L}}_2^{-1} \times_3 \hat{\mathbf{L}}_3^{-1} \\ &+ \underline{\mathbf{E}} \times_1 (\mathbf{L}_1 \cdot \hat{\mathbf{L}}_1^{-1}) \times_2 (\mathbf{L}_2 \cdot \hat{\mathbf{L}}_2^{-1}) \times_3 (\mathbf{L}_3 \cdot \hat{\mathbf{L}}_3^{-1}), \end{aligned} \quad (4.27)$$

where the noise tensor becomes approximately white, and by having full-rank correlation factor matrices, the rank of prewhitened signal subspace $\underline{\mathbf{X}}' = \underline{\mathbf{X}} \times_1 \hat{\mathbf{L}}_1^{-1} \times_2 \hat{\mathbf{L}}_2^{-1} \times_3 \hat{\mathbf{L}}_3^{-1}$ remains unaffected and the same as the rank of $\underline{\mathbf{X}}$. Therefore, we can do the rank

selection using the prewhitened signal $\underline{\mathbf{Y}}'$ instead of $\underline{\mathbf{Y}}$. From now on, we assume that the noise is whitened and for simplicity of notations we use $\underline{\mathbf{Y}}$ as the prewhitened signal.

Having the colored noise whitened, first we develop a method based on Reconstruction error for rank selection of the whitened tensor. Then, in the next section, for the cases that no prior information about the covariance of noise is available, we develop an iterative algorithm for simultaneously whitening and rank estimation.

4.3.2 Rank Estimation Criterion

Like most of the CANDECOMP/PARAFAC-decomposition based methods, for each candidate rank, the factor matrices of the corresponding CANDECOMP/PARAFAC model, (i.e. $\mathbf{A}, \mathbf{B}, \mathbf{C}$), are first estimated via ALS algorithm [71] and then according to (4.3) an estimate of the data with that candidate rank is calculated. Traditionally, one of the criterion that is widely used as the stopping criteria of the ALS algorithm, as well as in the backward direction for optimum order selection is the Data error ϵ_R defined as following:

$$\epsilon_R = \frac{1}{M} \|\underline{\mathbf{Y}} - \hat{\underline{\mathbf{X}}}_R\|^2 \quad (4.28)$$

where $\underline{\mathbf{Y}}$ is the noisy observed data tensor and $M = I \times J \times K$ is the total number of samples and $\hat{\underline{\mathbf{X}}}_R$ is the estimation of noiseless data via (4.3) by using an R -component CANDECOMP/PARAFAC model which factor matrices are estimated using the ALS algorithm.

The Data error ϵ_R , based on the maximum likelihood theorem, is known to be a monotonically decreasing function of the model order and cannot be solely used for model order selection problems. The characteristics of the Data error for a sample rank-five tensor while fitting an R -component CANDECOMP/PARAFAC model is depicted in Figure 4.4(a). Increasing the model order would result in a smaller Data error. In order to overcome this limitation, the classical model order selection methods, like AIC and BIC, have tried to add a penalty term, which is an increasing function of the model order,

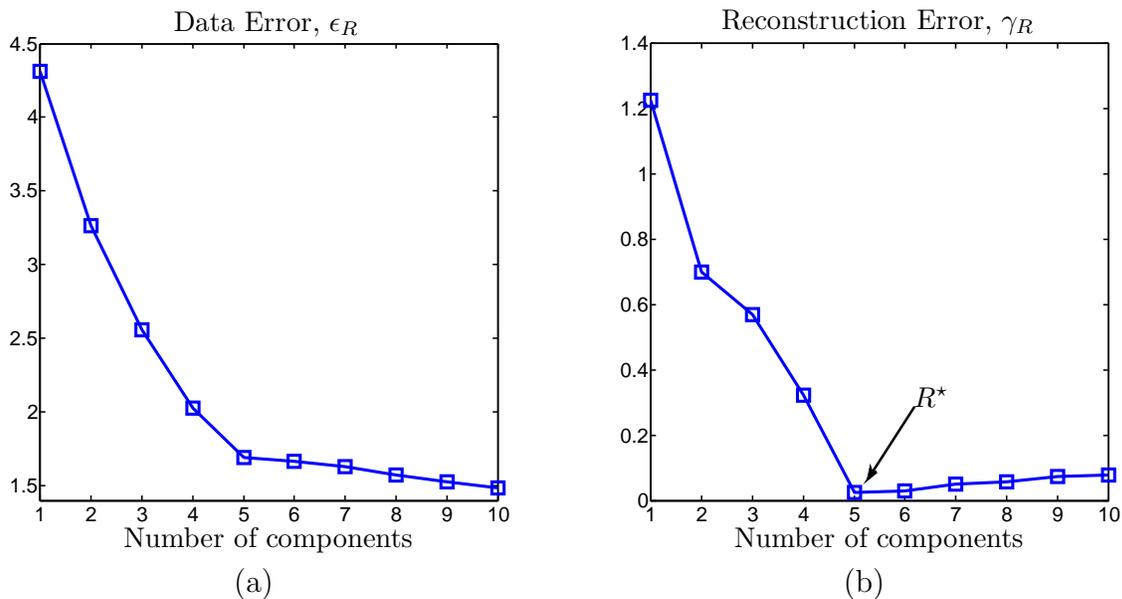


Figure 4.4: The characteristics of the Data Error and Reconstruction Error for a sample rank-five tensor $\underline{\mathbf{Y}}$ while fitting an R -component model ($1 < R < 10$) in CANDECOMP/ PARAFAC analysis.

to compensate that behaviour. Besides, in the context of CANDECOMP/PARAFAC model order selection, the Data error in the form of DIFFIT [67] or in combination with CORCONDIA [104] have been used for model order enumeration in order to overcome its downside.

On the other hand, by looking at the characteristics of Reconstruction error γ_R (see Figure 4.4(b)) defined as:

$$\gamma_R = \frac{1}{M} \|\underline{\mathbf{X}} - \hat{\underline{\mathbf{X}}}_R\|^2 \quad (4.29)$$

where $\underline{\mathbf{X}}$ is the noiseless data tensor associated with the observed noisy tensor according to (4.4), the true rank of the tensor ($R^* = 5$ in this case) is revealed at the minimum of γ_R by its intrinsic characteristics. Therefore, the Reconstruction error seems to be a potential metric for detecting the true rank of a tensor. However, the challenge here is that in real applications we do not have access to the noiseless data tensor $\underline{\mathbf{X}}$ and the only thing that we observe is $\underline{\mathbf{Y}}$ which leads to the Data error.

It can be shown that by observing the available Data error, ϵ_R , and having prior

knowledge about the variance of the noise, the Reconstruction Error could be modelled as a sample of a Chi-Squared distribution. Using that, we can find a probabilistic upperbound with a confidence region for the Reconstruction Error which has the AIC and BIC as its special cases [117]. The estimate of the Reconstruction Error then could be used for the optimum model order selection. The upperbound of γ_R is computed as follows:

$$\overline{\gamma_R(Q(\beta), \underline{\mathbf{Y}}, Q(\alpha))} = \frac{RI}{M}\sigma_n^2 + U_R(\underline{\mathbf{Y}}, Q(\alpha)) + \beta\frac{\sqrt{2RI}}{M}\sigma_n^2, \quad (4.30)$$

where σ_n^2 is the noise variance after prewhitening and R is the number of signal components used in the CANDECOMP/PARAFAC model. $Q(\alpha)$ and $Q(\beta)$ are the validation probability and the confidence probability used to provide the probabilistic upper bound for γ_R . Furthermore, U_R is the upperbound corresponding to unmodeled or discarded possible components (A full derivation of (4.30) is addressed in Appendix A).

Therefore, the optimum model order for CANDECOMP/PARAFAC in the context of Reconstruction Error, according to (4.10), would be the smallest R that would minimize the upperbound of γ_R among a set of possible candidates $1 \leq R \leq R_{max}$:

$$\hat{R} = R_{opt} = \underset{R}{\operatorname{argmin}} \overline{\gamma_R(Q(\beta), \underline{\mathbf{Y}}, Q(\alpha))}. \quad (4.31)$$

The maximum candidate rank, R_{max} , either can be available as prior knowledge or can be selected based on the CORCONDIA measure explained in the following section. The proposed rank selection method is summarized in Algorithm 4.

4.3.2.1 Incorporating CORCONDIA for selection of R_{max}

The performance of CORCONDIA is highly dependent on a threshold which could lead to overestimation or underestimation of the number of signal components in different scenarios [104]. The CORCONDIA (CC) measure is almost 100% for both true and underestimated model ranks and then starts to drop for overestimation of model rank (Figure 4.5(a)), where with the fixed threshold of 80%, the model is overestimated, and

Algorithm 4 Rank Selection with known Covariance

$\underline{\mathbf{Y}} \in \mathbb{R}^{I \times J \times K}$ and a maximum rank R_{max} and \mathbf{Cov}_r for $r = 1, 2, 3$. Optimum Rank \hat{R} . obtain the correlation factor matrices \mathbf{L}_r . obtain the prewhitened tensor $\underline{\mathbf{Y}}'$ via (4.26). $R \leftarrow 1$ R_{max} using $\underline{\mathbf{Y}}'$, estimate the factor matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$ for an R -component CANDECOMP/PARAFAC model using ALS. find an estimate of noiseless data $\hat{\underline{\mathbf{X}}}'_R$ via (4.3). $\epsilon_R \leftarrow \frac{1}{M} \|\underline{\mathbf{Y}}' - \hat{\underline{\mathbf{X}}}'_R\|^2$. obtain the upperbound of the Reconstruction Error $\overline{\gamma}_R$ via (4.30).
 $\hat{R} \leftarrow \underset{R}{\operatorname{argmin}} \overline{\gamma}_R$.

in order to have the best performance the threshold should be around 90%). The gap between the CC measures of true order and the overestimated one varies depending on the SNR. When the SNR is high, the drop of CC for the overestimated orders is very small, although for low SNRs the drop is significant. Therefore, the ideal threshold should increase with SNR, and to overcome this drawback in [104], the authors suggested the use of the difference of fit in consecutive model orders as an aid to increase the performance of CORCONDIA, and to remove the subjective interpretation of the threshold.

Regardless of the CORCONDIA plot trajectory, the CC measure is shown to be small and robust for the high overestimates of the model order. Therefore, a lower bound threshold is robust in providing an upper limit for the model rank, (i.e. R_{max}):

$$R_{max} = \max R \text{ subject to } CC(R) \geq \eta_{lb}, \quad (4.32)$$

where η_{lb} is the user-defined lowerbound of the threshold coefficient. Empirically range of $5\% \leq \eta_{lb} \leq 20\%$ works well.

4.3.3 Numerical Complexity

The goal of this section is to give the overall complexity of the proposed rank selection technique. As we saw in the previous Chapter, the number of required multiplications per iteration was calculated. Although, usually there is a stopping criterion using the relative error rather than a fixed number of iterations for convergence, In fact, to calculate the

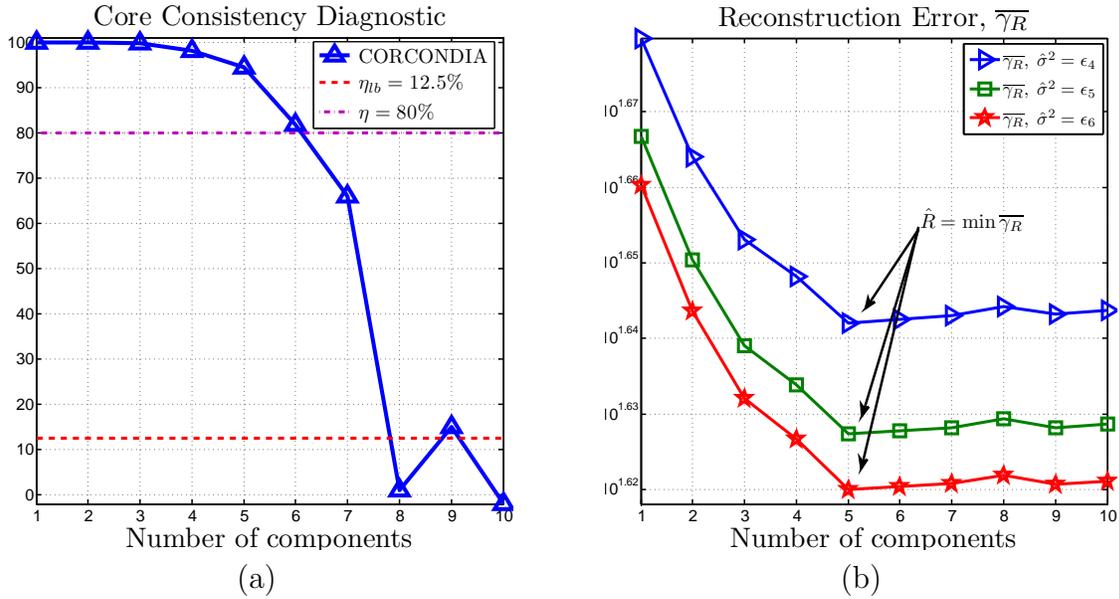


Figure 4.5: (a) The CORCONDIA measure for a sample rank-five tensor $\underline{\mathbf{Y}}$ along with 12.5% and 80% threshold lines, while fitting an R -component model for $1 < R < 10$; (b) The upperbound of γ_R for different estimation of variances of noise, using available ϵ_R , measured for a sample rank-five tensor $\underline{\mathbf{Y}} \in \mathbb{R}^{15 \times 15 \times 20}$ while fitting an R -component CANDECOMP/PARAFAC model for $1 < R < 10$.

cost of iterative algorithms a fixed number of iterations is considered. Computational complexity of ALS for a three-way array is $O((JK + KI + IJ)(7R^2 + R)l + 3RIJKl + (I + J + K)(R^2 + R)l + 11R^3l)$, where I, J, K are dimension in the corresponding modes, R the rank of the array, and l is the fixed number of iterations in the optimization stage. Computational complexity of Reconstruction Error based model order selection is $O(R_{max}) \times O((JK + KI + IJ)(7R^2 + R)l + 3RIJKl + (I + J + K)(R^2 + R)l + 11R^3l)$. This is obtained based on R_{max} required runs of ALS for estimating the minimum of γ_R , while almost no extra computation overhead after CANDECOMP/PARAFAC decomposition is required.

4.3.4 Unknown Noise Covariance

In situations in which we have no prior information about noise covariances and neither, the signal free measurements are available, (e.g. for EEG signals), we propose an algorithm that simultaneously estimates the covariance matrices and the rank selection. In Algorithm 5, first, through an iterative process, the covariance matrices for each mode are estimated, then the obtained noise covariance matrices are used for prewhitening and rank selection afterwards.

The algorithm estimates the noise covariances as following: it calculates the pre/de-whitened tensor $\check{\mathbf{Y}}'$ in three steps: prewhitening of \mathbf{Y} via (4.26), truncated HOSVD via (4.33), and dewhitening via (4.34).

$$\check{\mathbf{Y}} = \check{\mathbf{S}}^{[t]} \times_1 \mathbf{U}_1^{[t]} \times_2 \mathbf{U}_2^{[t]} \times_3 \mathbf{U}_3^{[t]} \quad (4.33)$$

where $\check{\mathbf{Y}} \in \mathbb{R}^{I \times J \times K}$ is the truncated HOSVD, $\check{\mathbf{S}}^{[t]} \in \mathbb{R}^{p_1 \times p_2 \times p_3}$, and $\mathbf{U}_r^{[t]} \in \mathbb{R}^{M_r \times p_r}$ with $p_r = \min(M_r, R)$. Exploiting the HOSVD-based low rank approximation we obtain the denoised prewhitened tensor $\check{\mathbf{Y}}$. Finally, for parameter estimation, the original signal subspace should be recovered through the dewhitening stage as following

$$\check{\mathbf{Y}}' = \check{\mathbf{Y}} \times_1 \hat{\mathbf{L}}_1 \times_2 \hat{\mathbf{L}}_2 \times_3 \hat{\mathbf{L}}_3. \quad (4.34)$$

Consequently, the estimates of the factor matrices of an R -component CANDECOMP/PARAFAC model are calculated and the noiseless tensor is reconstructed using (4.3), then the residual tensor, $\hat{\mathbf{E}}_R^{(c)} = \mathbf{Y} - \hat{\mathbf{X}}_R$, is used for estimating the covariance matrices of the noise. We found in most of the cases, less than 5 iterations is required to converge to the true parameters. A stopping criteria is developed in [93] using a similar approach for cases that we may end up in non-convergence parameter estimates. Note that for each model rank R , ($1 \leq R \leq R_{max}$), we get a different estimate of the noise tensor and consequently a different pre/de-whitened tensor $\check{\mathbf{Y}}'_R$, (i.e. the estimate of noise tensor, $\hat{\mathbf{E}}_R^{(c)}$, is a function of number of signal components R). Hence, we would use the

Algorithm 5 Simultaneous Noise estimation and Rank Selection

$\underline{\mathbf{Y}} \in \mathbb{R}^{I \times J \times K}$ and a maximum rank R_{max} and a stopping iteration T . Optimum Rank \hat{R} . $it \leftarrow 0$ $\hat{\mathbf{L}}_{\mathbf{r}}^R \leftarrow \mathbf{I}_{M_r}$ $R \leftarrow 1$ R_{max} $it \leq T$ obtain the pre/de-whitened tensor $\check{\underline{\mathbf{Y}}}'_R$ in 3 steps: prewhitened via (4.26), truncated HOSVD via (4.33) and dewhitened via (4.34). using $\check{\underline{\mathbf{Y}}}'_R$, estimate the factor matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$ of an R -component CANDECOMP/PARAFAC model using ALS. find an estimate of noiseless data $\hat{\underline{\mathbf{X}}}'_R$ via (4.3). $\hat{\underline{\mathbf{E}}}_R^{(c)} \leftarrow \underline{\mathbf{Y}} - \hat{\underline{\mathbf{X}}}'_R$. given $\hat{\underline{\mathbf{E}}}_R^{(c)}$, estimate the correlation factor matrices $\hat{\mathbf{L}}_{\mathbf{r}}^R$ via (4.25). $it \leftarrow it + 1$
 using $\hat{\mathbf{L}}_{\mathbf{r}}^R$, obtain the prewhitened tensor $\underline{\mathbf{Y}}'_R$ via (4.26).
 estimate the factor matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$ of $\underline{\mathbf{Y}}'_R$ for an R -component CANDECOMP/PARAFAC model using ALS. estimate the noiseless data $\hat{\underline{\mathbf{X}}}'_R$ via (4.3). $\epsilon_R \leftarrow \frac{1}{M} \|\underline{\mathbf{Y}}'_R - \hat{\underline{\mathbf{X}}}'_R\|^2$. obtain the upperbound of the Reconstruction Error $\overline{\gamma}_R$ via (4.30). $\hat{R} \leftarrow \underset{R}{\operatorname{argmin}} \overline{\gamma}_R$.

corresponding whitened tensor, $\underline{\mathbf{Y}}'_R$, for estimating the factor matrices of an R -component CANDECOMP/PARAFAC model and finding the optimum model rank, unlike the fixed whitened tensor $\underline{\mathbf{Y}}'$ that we used in Algorithm 4 for a known noise covariance.

4.4 Simulations

We are considering real-valued data in all the simulations. In the CANDECOMP/PARAFAC model, the factor matrices in (4.3) contain i.i.d Zero-Mean Gaussian entries with unit variance. The colored noise is generated according to (4.9) using zero-mean white noise samples with unit variance. Along each mode the colored noise is modeled as a first-order autoregressive process

$$e_{i+1}^{(c)} = \rho_r \cdot e_i^{(c)} + \sqrt{1 - |\rho_r|^2} \cdot e_{i+1}, \quad (4.35)$$

such that both the correlation factor $\mathbf{L}_{\mathbf{r}}$ and covariance matrix $\mathbf{Cov}_{\mathbf{r}}$, are functions of a single correlation coefficient ρ_r . For example, $\mathbf{Cov}_{\mathbf{r}}$ for $M_r = 3$ has the following

structure

$$\mathbf{Cov}_r = \begin{bmatrix} 1 & \rho_r & \rho_r^2 \\ \rho_r & 1 & \rho_r \\ \rho_r^2 & \rho_r & 1 \end{bmatrix}.$$

A scaling parameter η is used to obtain different SNRs. Let $\underline{\mathbf{E}}^{(c)}$ be the colored noise tensor, and let $\underline{\mathbf{X}}$ be the noise-free tensor. Then

$$\underline{\mathbf{Y}} = \underline{\mathbf{X}} + \eta \frac{\|\underline{\mathbf{X}}\|}{\|\underline{\mathbf{E}}^{(c)}\|} \underline{\mathbf{E}}^{(c)} \quad (4.36)$$

is the noisy version of the tensor where η is the percentage of noise to add. In the simulations we use the Tensor Toolbox ¹ for CANDECOMP/PARAFAC decomposition with the following configurations: convergence criterion set as $10e^{-6}$, no constraints on the modes, line-search acceleration scheme initialized, and using direct trilinear decomposition for initialization. For each parameter setting, 1000 independent Monte Carlo runs have been conducted. The performance measure is the PoD, (i.e., $Pr(\hat{R} = R^*)$), the probability that the estimated rank is equal to the true rank, averaged over noisy realizations of all Monte Carlo runs.

In the simulations we assume that the true rank of the observed tensor is smaller than the size of its corresponding modes. The simulations are divided into two parts. First, by setting the correlation coefficients $\rho_r = 0$ for $r = 1, 2, 3$, we evaluate the performance of the proposed method compared to other methods. In this scenario, no prewhitening is needed and evaluation is based solely on the merit of effectiveness of different methods in true rank selection considering white noise. Second, we evaluate the effect of prewhitening on the rank selection of the proposed method in the colored noise scenario. In the following sections we discuss the results of each scenario.

¹MATLAB Tensor Toolbox, Available online, <http://www.sandia.gov/~tgkolda/TensorToolbox/>.

4.4.1 Additive White Noise

The correlation coefficients in this case, are set as $\rho_r = 0$ for $r = 1, 2, 3$ which means that the noise is white and the covariance matrices are identity. We compare our method (rank selection based on Reconstruction Error) in this scenario with the following schemes: CORCONDIA with a fixed typical threshold setting of 80%, aided-CORCONDIA [104], and also rank selection based on the AROFAC2 [113], the SORTe [88], the Convex-hull method [68], the RADOI method [118], and R-D AIC, R-D MDL and R-D EFT methods [111]. In the proposed scheme, we set $\alpha = 100$ and $\beta = 200$ (refer to [117] for the discussion on selection of these parameters). The lower bound threshold in the CORCONDIA is also chosen as $\eta_{lb} = 12.5\%$.

The Data error ϵ_R at the true order, according to its definition (4.28), could be a good estimate of the noise variance if we assume that the CANDECOMP/PARAFAC model fits the measurements well. Although we do not have prior knowledge about the true order, we found through our simulations that $\overline{\gamma_R}$ is not that sensitive to the variance of noise, σ_n^2 , as long as it is selected from the vicinity of true model order in ϵ_R . Figure 4.5(b) shows the estimated Reconstruction Error $\overline{\gamma_R}$ of a sample realization three-way CANDECOMP/PARAFAC model of size $I = 15, J = 15, K = 20$ and $R^* = 5$ for $1 \leq R \leq R_{max}$. $\overline{\gamma_R}$ is plotted using 3-different noise variances, selected as $\epsilon_4, \epsilon_5, \epsilon_6$ from the Data error, which are all minimized at the true order $\hat{R} = 5$. Therefore, we used the average of Data error, ϵ_R for $1 \leq R \leq R_{max}$, as an estimate of the noise variance in the simulations. The other approach is to use the majority voting technique to choose the mode of the selected orders via different noise variances as the optimum order.

Next, the PoD versus SNR for $R^* = 5$ and $R^* = 7$ components in a three-way CANDECOMP/PARAFAC model of size $I = 15, J = 15, K = 20$ using different selection methods are depicted in Figure 4.6(a) and Figure 4.6(b), respectively. Comparing the simulation results we can see that the CORCONDIA with a fixed threshold 80%, which is highly dependent on the threshold value and SNR, has a low PoD at high SNRs due to frequent over-enumeration. 80% threshold only seem to be a proper one at $SNR = -5dB$

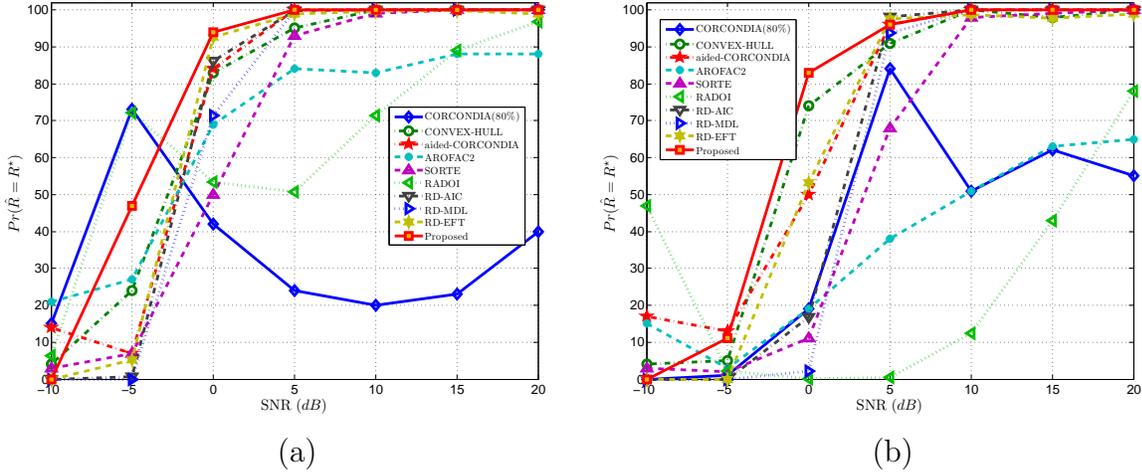


Figure 4.6: Comparison of PoD, i.e. $Pr(\hat{R} = R^*)\%$ vs. SNR for different Rank selection methods while fitting a CANDECOMP/PARAFAC model averaged over 1000 Monte Carlo realizations while $\rho_r = 0$ for $r = 1, 2, 3$: (a) for a rank-5 tensor $\underline{\mathbf{Y}} \in \mathbb{R}^{15 \times 15 \times 20}$; (b) for a rank-7 $\underline{\mathbf{Y}} \in \mathbb{R}^{15 \times 15 \times 20}$.

and $SNR = 5dB$ for $R^* = 5$ and $R^* = 7$ respectively, where the PoD is almost high. Thus making it clear that this method is not consistent for different settings and model orders. The performance of AROFAC2, which is developed based on structure of low-rank tensors, degrades with increasing tensor-rank. SORTE tends to underestimate in low SNRs and has an inferior performance compared to aided-CORCONDIA, which uses the discriminative power of Data error in the neighbourhood of true model order. Although aided-CORCONDIA is an efficient method and almost comparable with the proposed method, it still requires an extra computational load to calculate its threshold. The performance of RADOI method, which is an empirical method, is inferior to other methods and degrades with increasing the model order and requires higher SNRs to perform better. The performance of R-D AIC, R-D MDL are close and consistent with the results of [111] which are deemed inferior compared to R-D EFT. The convex-hull based method has a comparable performance with the proposed method, however, the proposed method outperforms other methods for both considered model orders in all SNRs, especially in mid and low SNR scenarios, (i.e. 0 – 10 dB), by having a higher PoD.

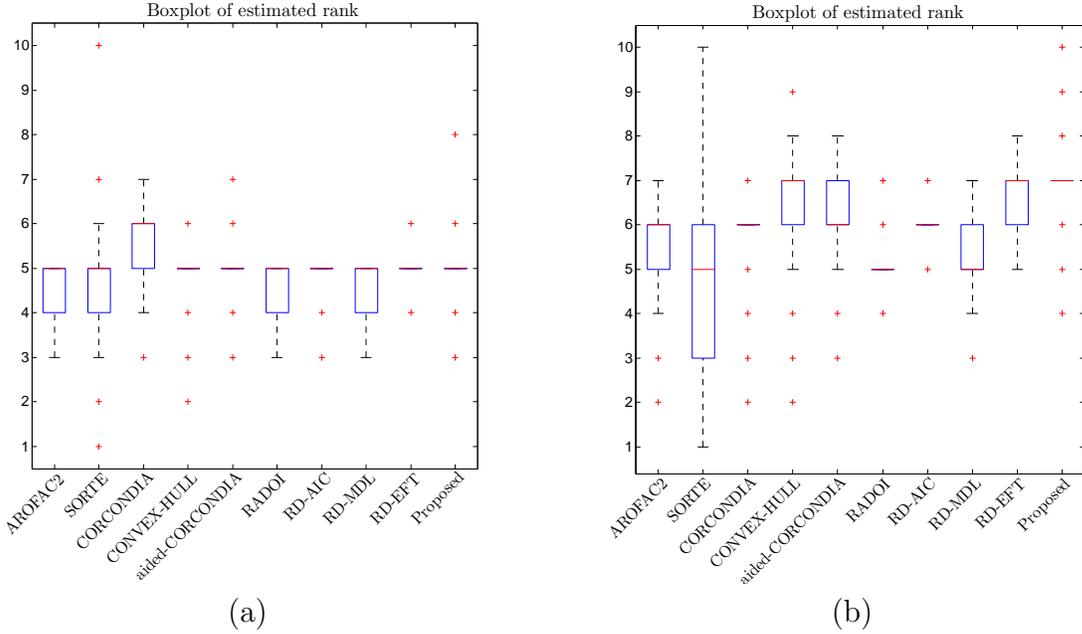


Figure 4.7: Boxplot of the estimated rank of the observed tensor over 1000 Monte Carlo realizations at $SNR = 0\text{dB}$ using different estimation methods. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points the algorithm considers to be not outliers, and the outliers are plotted individually for: (a) rank-five tensor $\underline{\mathbf{Y}}$; (b) rank-seven tensor $\underline{\mathbf{Y}}$.

Furthermore, to evaluate the statistics of estimated ranks using different methods, the boxplots of the estimated rank at $SNR = 0\text{dB}$ for $R^* = 5$ and $R^* = 7$ are illustrated in Figure 4.7(a) and Figure 4.7(b), respectively. The boxplot shows the median, 25th and 75th percentiles, and outliers of the estimated rank. Figure 4.7(a) shows that, for a true rank of 5, CORCONDIA with a fixed 80% threshold overestimates the rank in most cases, the median at $\hat{R} = 6$, while AROFAC, SORTE, RADOI, and R-D MDL tend to underestimate when they couldn't estimate the true rank (25th percentile is at $\hat{R} = 4$).

Fig. 4.7(b) shows for a true rank of 7, the proposed method detects the correct rank at $\hat{R} = 7$ consistently, and the only methods that have a close performance, but still tend to underestimate, are R-D EFT and Convex-hull. Other methods tend to underestimate the rank more by having the median at $\hat{R} = 6$ for AROFAC, CORCONDIA, aided-CORCONDIA, R-D AIC and the median at $\hat{R} = 5$ for R-D MDL, RADOI and SORTE. Also, SORTE has the highest inconsistency in model order selection by having a widely

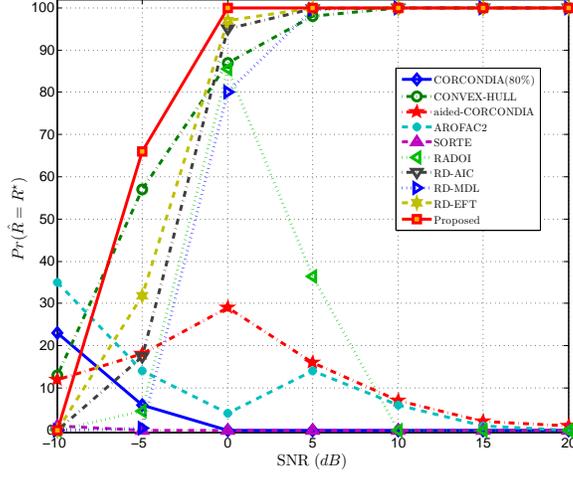


Figure 4.8: Comparison of PoD, i.e. $Pr(\hat{R} = R^*)\%$ vs. SNR for a rank-overlap problem with the true rank of 3 and $\rho_r = 0$ for $r = 1, 2, 3$ for different Rank selection methods averaged over 1000 Monte Carlo realizations of $\underline{\mathbf{Y}} \in \mathbb{R}^{15 \times 15 \times 20}$.

distributed boxplot.

In order to evaluate the performance of the proposed method when we have linear dependence in constituting factor matrices of a tensor or especially the rank-overlap problem, (i.e. some loading components of the factor matrices are collinear, e.g. bottlenecks and swamps), an other simulation is performed. A simple example for the rank-overlap problem with a rank-3 tensor is as follows

$$\underline{\mathbf{X}} = a_1 \circ b_1 \circ c_1 + a_1 \circ b_2 \circ c_2 + a_2 \circ b_2 \circ c_3 . \quad (4.37)$$

Figure 4.8 depicts the PoD versus SNR for $R^* = 3$ when we have the rank-overlap problem, averaged over 1000 Monte Carlo realizations. As the figure shows the SORTE method that is based on the mode-1 unfolding completely fails, as the number of eigenvalues in mode-1 unfolding is 2. RADOI method which also uses the mode-1 unfolding fails in high SNRs, however, in lower SNRs it has a high PoD that might be due to the fact that RADOI empirically tries to model the signal and noise subspaces with two different discriminant functions and in lower SNRs one of the noise eigenvalues might be

misclassified as the signal subspace and hence the correct order is selected. The performance of CORCONDIA, AROFAC, and the aided-CORCONDIA degrades drastically in the rank-overlap problem. RD-based methods which are based on global eigenvalues have a good performance. The Convex-hull method which is based on the goodness of fit also performs good. However, the proposed scheme outperforms the other state-of-the-art methods in the rank-overlap problem as it is based on reconstruction error in which correlation of factors does not affect its performance.

Furthermore, we compared the performance of the methods for different model ranks at a fixed SNR. Figure 4.9(a) depicts the PoD versus tensor rank for $R^* = 1, 2, \dots, 14$ at $SNR = 5\text{dB}$ averaged over 1000 Monte Carlo realizations when all the factors in the model have the same magnitude, (i.e. in (4.3) we use identity tensor). The results show that the CORCONDIA with a fixed threshold has the highest inconsistency among the compared methods and once more confirms the downside of this method. The performance of RADOI degrades drastically as the model order increases. The proposed method, on the other hand, has remarkably superior and consistent performance for different model orders compared to other methods where the drop in the PoD of the proposed method is small even for high model orders. In Figure 4.9(b), we compare the performance of the methods, where unlike the previous one, instead of having the same magnitude we have different magnitudes for each factor by setting a diagonal tensor in (4.3) with its elements being uniformly distributed between $(0.2, 1)$. We chose this range to prevent having any component with a very small and negligible contribution. The results show that the performance of all methods degrades for higher model orders, but the proposed method still has the best performance.

In addition, Figure 4.9 shows that PoDs are relatively low, and might fall even lower than 50% in some methods when the true rank is high and approaches the tensor dimensions. It is nonsense to do rank selection when the desired components are not recovered. So, we need to figure out what is the cause of degradation for high model orders, whether it is due to shortcoming of ALS algorithm or the model order selection scheme itself.

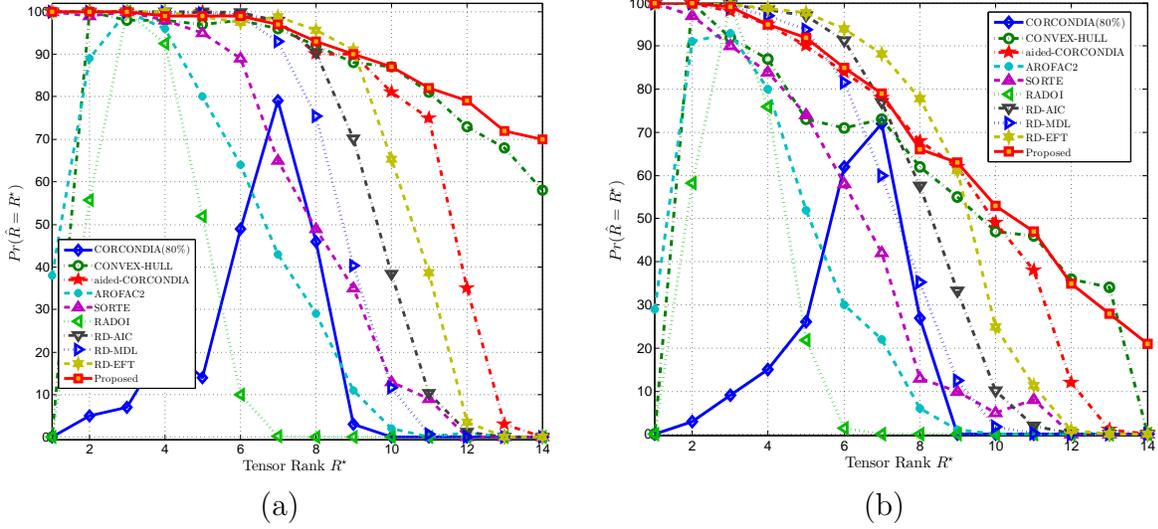


Figure 4.9: Comparison of PoD, i.e. $Pr(\hat{R} = R^*)\%$ vs. tensor rank, at $SNR = 5\text{dB}$, $\rho_r = 0$ for $r = 1, 2, 3$ for different rank selection methods averaged over 1000 Monte Carlo realizations of $\mathbf{Y} \in \mathbb{R}^{15 \times 15 \times 20}$: (a) for an identity tensor in (4.3); (b) for a diagonal tensor in (4.3) with its elements uniformly selected between $(0.2, 1)$.

The fit of the model is generally used to evaluate performance of a tensor decomposition, although this measure does not imply quality of the estimated components. In order to verify theoretically the quality and accuracy of the estimated components via ALS and validate the efficiency of the proposed algorithm an appropriate measure is an essential prerequisite. The squared angular error between the estimated component and its original one is such a measure [119]. In Figure 4.10 the estimation accuracy of the factor matrices in terms of Mean Square Angular Error (MSAE) compared to the theoretically achievable Cramer-Rao Induced Bound (CRIB) for different model orders and a range of SNR values is depicted. When the noise is additive Gaussian noise, the CRIB on estimation of factor matrices in CANDECOMP/PARAFAC model can be calculated [119]. The CRIB may serve a gauge of achievable accuracy of estimation/CP decomposition. It is an asymptotically (in the sense of variance of the noise going to zero) tight bound on the angular error between an estimated and true factor. For example, it is known that when CRIB on the squared angular error (in dB) is lower than 20dB, which corre-

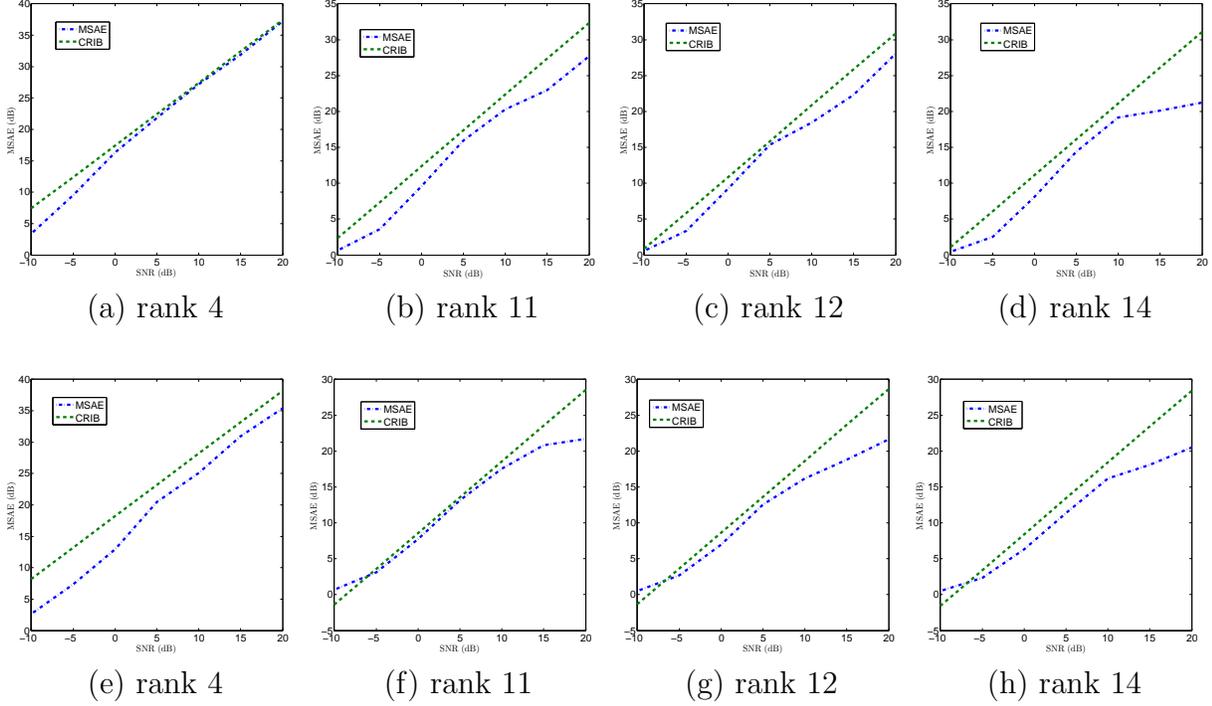


Figure 4.10: Estimation accuracy of the factor matrices in terms of MSAE compared to the CRIB for different model orders vs. SNR, $\rho_r = 0$ for $r = 1, 2, 3$ averaged over 100 Monte Carlo realizations of $\underline{\mathbf{Y}} \in \mathbb{R}^{15 \times 15 \times 20}$: (a)-(d) for an identity tensor in (4.3); (e)-(h) for a diagonal tensor in (4.3) with its element uniformly selected between (0.2, 1).

sponds to the standard deviation of 5.7° , it is hard to retrieve the components. CRIB can explain performance of simulations when the tensor rank is high and approaches the tensor dimension. We can see in Figure 4.10 that the CRIB at SNR = 5 dB for high model orders is not within the reliable range (i.e. MSAE < 20 dB). Moreover, CRIB is lower for a diagonal tensor (lower row) in (4.3), compared to identity tensor (upper row), which reflects the difficulty the ALS algorithm faces in estimating the factor matrices by having components with different magnitudes. Hence, we can say that the degradation of PoD when the tensor rank is high could be related to the ALS algorithm, which can not estimate the factor matrices correctly, and not to the model order selection schemes.

Now, acknowledging the effectiveness of the proposed method with additive white noise, in the next section we evaluate the effect of adding correlation between noise

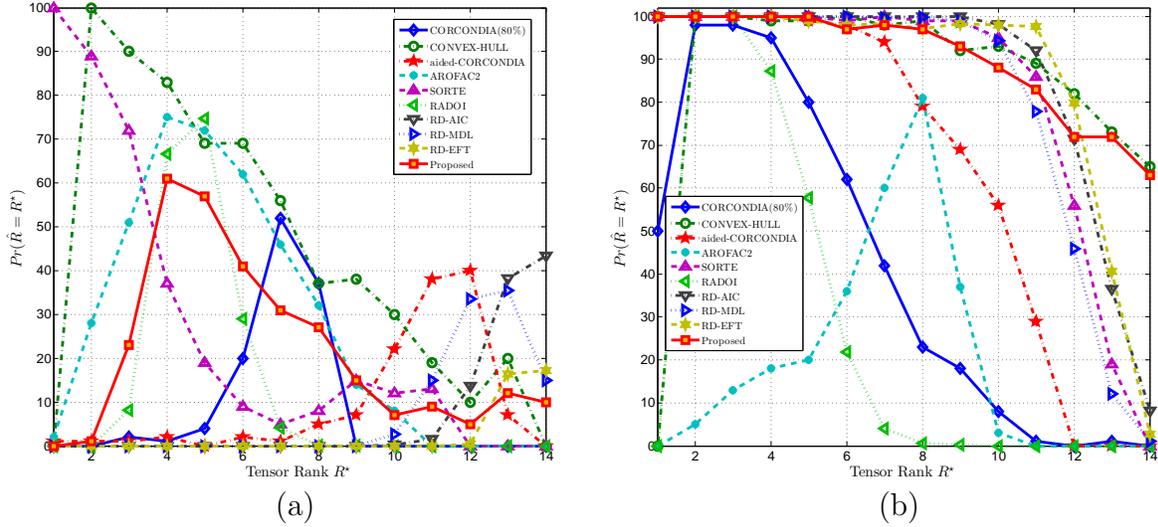


Figure 4.11: Comparison of PoD, i.e. $P_r(\hat{R} = R^*)\%$ vs. Tensor rank at $SNR = 5\text{dB}$ and $\rho_r = 0.5$ for $r = 1, 2, 3$, while fitting a CANDECOMP/PARAFAC model averaged over 1000 Monte Carlo realizations of $\underline{\mathbf{Y}} \in \mathbb{R}^{15 \times 15 \times 20}$: (a) for colored noise; (b) for prewhitened noise.

samples on the rank selection performance and see how the proposed method can tackle the colored noise problem.

4.4.2 Additive Colored Noise

In this scenario different correlation coefficients, (i.e. ρ_r for $r = 1, 2, 3$), are considered to evaluate the effect of additive colored noise on degradation of the rank selection performance and how prewhitening can help us to prevent that.

4.4.2.1 Known Covariance Matrix

Here, we assume that we have access to the signal free measurements and the true covariance matrices of the colored noise in each mode are also available. Algorithm 4 is used in this part and for all other methods considered for the comparison, first the prewhitening is done and then rank selection is performed. Figure 4.11(a) shows the PoD of the true rank versus tensor rank, averaged over 1000 Monte Carlo realizations, for $R^* = 1, 2, \dots, 14$ at $SNR = 5\text{dB}$ with $\rho_r = 0.5$ for $r = 1, 2, 3$. The only parametric

change in this setting, compared to Figure 4.9, is the consideration of the additive colored noise. The results show that even for a moderate correlation coefficient $\rho_r = 0.5$, all of the rank selection methods degrades drastically. Although, Convex-hull method still does a relatively good job when the true rank is low. The performance of the RADOI method which is believed to be able to handle colored noise also degrades and the RD-based methods that are based on white noise assumption fail completely in this scenario.

In Figure 4.11(b) the effect of prewhitening of the colored noise on the true rank selection is depicted. We can see that by prewhitening we can effectively prevent performance degradation in all methods except the AROFAC2 method, where not much improvement in the rank selection is achieved for the low rank orders range. It could be due to the fact that AROFAC2 is based on clustering and prewhitening, which is like a rotation in the signal subspace, might affect the number of detected clusters.

Next, we compare the effect of having colored noise samples with different correlation coefficients for a range of $0.1 \leq \rho_r \leq 0.9$ on PoD at $SNR = 5\text{dB}$ for a rank-7 $\mathbf{Y} \in \mathbb{R}^{15 \times 15 \times 20}$. Figure 4.12(a) shows the performance degradation effect of having colored noise with different correlation coefficients, while Figure 4.12(b) shows that prewhitening can consistently prevent performance degradation for all range of correlation coefficients. It shows that for $\rho_r \geq 0.4$ the performance of proposed method, without prewhitening of the colored noise, degrades and the probability of detection of the true rank is almost zero for any $\rho_r \geq 0.6$. We can see that the Aided-CORCONDIA, SORTe, RD-based methods and the core of the proposed method, which are developed based on the white noise assumption, have a considerable improvement in their PoD after prewhitening. However, CORCONDIA(%80) and AROFAC2 are unresponsive to prewhitening, which is due to the fact that AROFAC2 is said to be independent of the noise structure and the correlation between the noise samples does not influence its performance that much. CORCONDIA, on the other hand, although is the most inconsistent method among the compared ones, has almost no performance degradation in the presence of colored noise which might be due to the fact that, CORCONDIA measures the

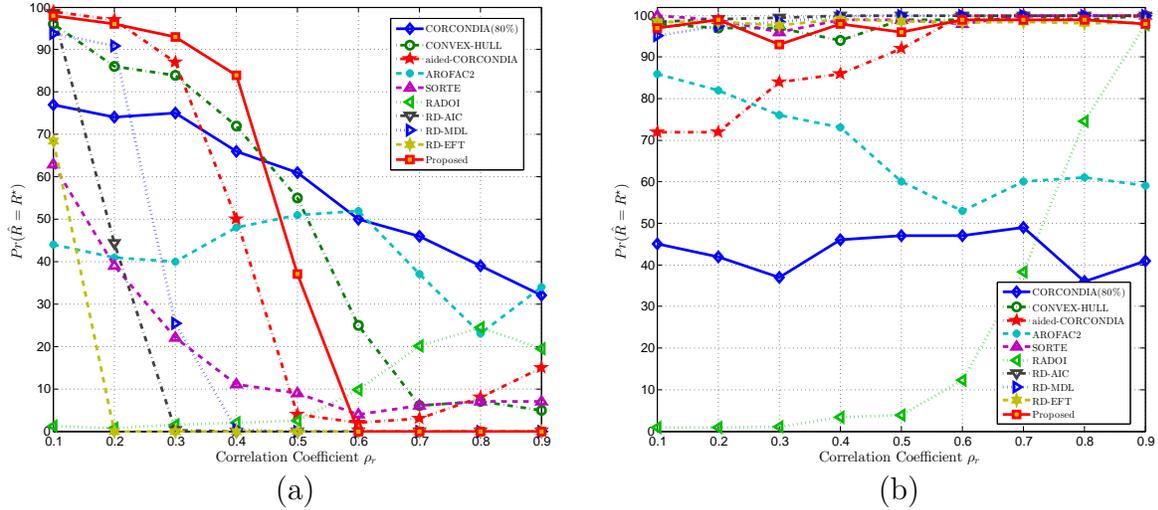


Figure 4.12: Comparison of PoD, i.e. $Pr(\hat{R} = R^*)\%$, vs. Correlation Coefficient ($0.1 \leq \rho_r \leq 0.9$), at $SNR = 5\text{dB}$, for different Rank selection methods averaged over 1000 Monte Carlo realizations for a rank-7 $\mathbf{Y} \in \mathbb{R}^{15 \times 15 \times 20}$: (a) without prewhitening of the colored noise; (b) with prewhitening of the colored noise.

similarity of the estimated core tensor with the theoretical super diagonal one and presence of additive colored noise samples may not influence the trend of the CORCONDIA measure. On the other hand, prewhitening does not improve the RADOI method, as the true rank-7 is fairly high for this method and it completely fails.

4.4.2.2 Unknown Covariance Matrix

In this part, we assume that no information about the covariance of the colored noise is available and we use the developed Algorithm 5 for estimating the covariance matrices of the noise and simultaneously obtaining the whitened tensor for model rank selection. We compare the PoD of the true rank versus SNR for the proposed method, having a Rank-7 three-way CANDECOMP/PARAFAC model of size $I = 15, J = 15, K = 20$ contaminated with additive colored noise with two different correlation coefficients of $\rho_r = 0.2, 0.5$ for $r = 1, 2, 3$. Figure 4.13 shows that the proposed approach is effective for the moderate correlation factors for almost all the considered SNRs, and the performance is comparable to the scenario that we have access to the true covariance matrices of the

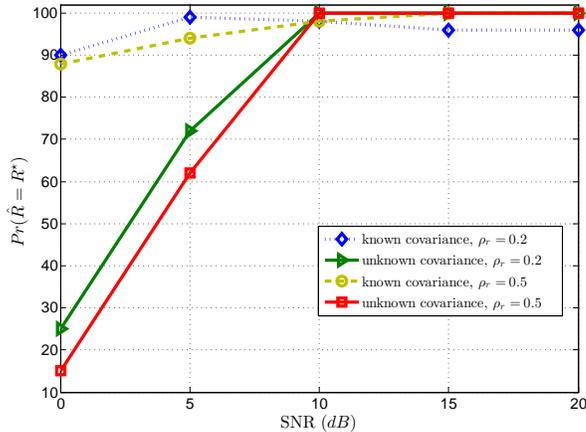


Figure 4.13: PoD vs. SNR of the proposed method with and without having the covariance matrices of the noise, averaged over 1000 Monte Carlo realizations of a rank-7 $\underline{\mathbf{Y}} \in \mathbb{R}^{15 \times 15 \times 20}$ contaminated with colored noise of $\rho_r = 0.2, 0.5$ for $r = 1, 2, 3$.

colored noise. However, for the low SNR of 0dB, the performance is slightly inferior. This proves the essence of availability of a good estimate for the noise covariance matrix in order to detect the true rank efficiently. Perhaps incorporating more robust noise covariance estimation techniques in the algorithm could further improve the results, especially when we have a limited number of measurements and the sample covariance matrix is not reliable.

4.5 Chapter Summary

CANDECOMP/PARAFAC decomposes a tensor into a summation of rank-one tensors. The ALS algorithm used for this decomposition is attractive for its simplicity and satisfactory performance for well-defined problems. In this chapter we reviewed various ALS-based approaches that have been proposed in the literature to extract the true rank of the observed tensor, most of them based on the so called model fit or Data error. Although, the available Data error is informative, it partially models the noise and known to be a monotonically decreasing function of the model order, that makes it inapplicable to rank selection. Furthermore, in this chapter we developed a novel methodology for

CANDECOMP/PARAFAC model order selection. we propose using the Data error to approximate the unknown Reconstruction error. The Reconstruction error, as the rank selection criterion, intrinsically reveals the true rank of the observed tensor. In addition the range of possible ranks is determined exploiting the CORCONDIA measure. The proposed algorithms handle the additive colored noise with the Kronecker structure through an embedded multidimensional prewhitening scheme. Furthermore, when we have no information about the covariance of the noise priorly, the proposed algorithm can also estimate the noise covariance through an iterative method while simultaneously determines the rank. The results show the advantage of the proposed method in both robustness and accuracy, in terms of PoD, over the conventional CANDECOMP/PARAFAC model order selection methods such as different variants of CORCONDIA. Consequently, the proposed method could be effectively exploited for CANDECOMP/PARAFAC model order selection, while almost no extra computation overhead after CANDECOMP/PARAFAC decomposition is required.

In the following Chapters we present some practical applications of EEG tensor analysis using CANDECOMP/PARAFAC model where we use the TFR technique proposed in Chapter 2 for EEG tensor formation and also estimate the model order by our developed technique based on the Reconstruction Error.

Chapter 5

EEG Tensor Analysis

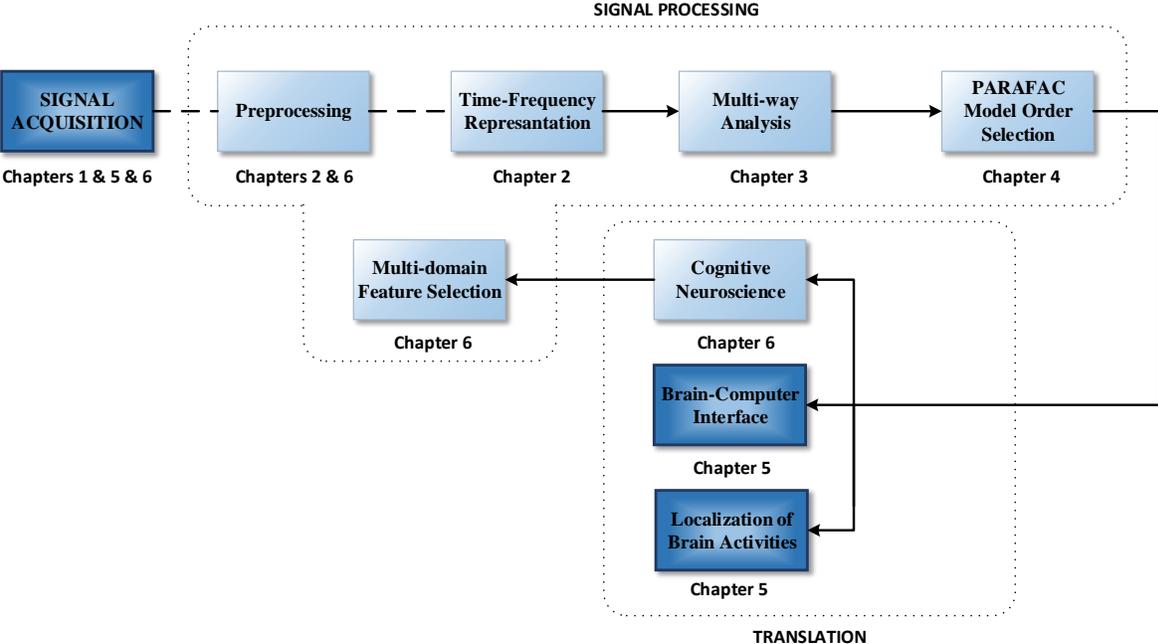


Figure 5.1: Chapter 5- EEG Tensor Analysis.

Data processing and analysis has a key role in study of the brain and brain research using brain imaging tools. The recorded brain imaging data has a structural nature and can be fitted and represented by a one-way series (a vector), a two-way array (a matrix), and a multi-way array (a tensor). Hence, depending on how the data is represented

various signal processing and analysis methods are applied.

EEG is known to be a high-dimensional signal in which processing of the information by the brain is reflected in dynamical changes of the electrical activity in time, frequency, and space. Study of these processes requires methods which can describe these variations of the signal in time, frequency, and space in a quantitative way. In general, EEG signal processing seek to decompose EEG into functions with known Spatio-Temporal-Spectral properties or at least easier to characterize. For example, in the early stage of EEG studies, when the temporal analysis techniques were popular, EEG data were represented by a time series and all data samples were carried by a 1D vector. Since then, after invention of Fourier transform, power spectrum analysis of the time series has been often applied for exploring EEG characteristics in spectral domain. These methods only use one contextual dimension which is either temporal or spectral domain.

Conventional methods use two-way decomposition techniques. By exploiting the popular time-frequency analysis (TFA) techniques, the time series of single-trial EEG were mapped to 2D images. Furthermore, it is very common that multiple electrodes, spanning different locations of the scalp, to be used in the experiments to collect EEG data. Hence, EEG recordings naturally include two modes of time and space, at least. These EEG data representations require a two-way array or matrix and subsequently two-way signal processing methods has been extensively used for EEG data analysis. Two of the most famous techniques in this context are PCA and ICA that along other matrix factorization techniques have been performed on the observed EEG matrix to remove artifacts and to extract brain activities of interest [61]. These techniques consider just two contextual dimensions, e.g. time and frequency or time and spatial domains. However, two-way analysis is unique subject to some additional constraints which may not be always physiologically meaningful. On the other hand, Multi-way analysis techniques, by retaining the additional dimensions, keep the structure of the multidimensional data and by analyzing more dimensions simultaneously, summarize the data into more interpretable components.

5.1 Multi-way nature of EEG signals

Indeed, in EEG experiments, along the two modes of time and space that are naturally inherited in the recordings, usually other modes are also present. For instance, different subject groups may have participated in the experiments and analysis of EEG signals may compare responses between groups (e.g. comparison of responses in a healthy control group and a clinical group). Thus, at least one more mode could be considered in the analysis and it is the subject. Furthermore, in an experiment to elicit event-related potentials (ERPs), usually the stimuli is presented several times to the subject and EEG is recorded for each of the trials. There might be also various stimulus presentation conditions. Hence, there are modes of EEG trial and stimulus condition. This means the brain data collected by EEG techniques can be naturally fit into a multi-way array including multiple modes.

However, most of the analysis tools for brain research are aligned for one-way or two-way data, which means that in order to use the two-way signal processing methods the recorded data needs to be fitted in a matrix. This is often called unfolding a multi-way array into a matrix. To do so, the extra modes besides the two modes of time and space are often concatenated or stacked with the time or the space mode to form a matrix. The difference between concatenation and stacking is in the way we connect the data in a plane. In the former the data are horizontally connected, while in the later the data are vertically connected in a plane. For EEG data, such unfolding inevitably loses some structure and potentially the existing interactions between / among the folded modes, such as time, frequency and space modes. However, the interactions between various modes of data can be of research interest and may reveal hidden structure of data. Consequently, multi-way array processing tools, by keeping the structure of the data, are considered promising tools for investigating the latent structure and interactions among multiple modes [120].

5.2 Multiple modes and high-order tensors of EEG data

There are three different type of EEG data, including spontaneous EEG, ERPs, and ongoing EEG that regardless of which type we collect, the raw recording are continuous and the length of the recording in time can be dozens of minutes or even a few hours or days. For off-line analysis the continuous EEG data are usually segmented in terms of stimulus types. These segments, which can also be called epochs or trials in case of ERPs, naturally constitute the additional mode segment beside to the two modes of time and space. Besides, if the EEG data of one segment is transformed into the time-frequency domain, another mode called as frequency is produced. Moreover, particularly in ERPs, usually two or more experimental conditions are tested. These variations could be a change in the stimulus properties, e.g. change of frequency or length of a sound in the experiment for mismatch negativity (MMN), or any environmental change during the EEG data recording. Furthermore, often several participants are involved in the experiments to facilitate the within-subject analysis. This mean that modes of condition and subject also exist. Also, regarding the between-subject analysis in an ERP experiment, there are two or more groups of subjects. Therefore, another mode is the group.

As a result, in an EEG experiment, potentially, there could be even 7 modes including time, frequency, space, trial, condition, subject and group. Although, in the ERP study, the mode of trials disappears due to the synchronized averaging usually applied on EEG data over single trials to improve the SNR. Consequently, the high-order tensors including some of the seven modes do naturally exist in any EEG experiment. Obviously, such a tensor may be very big in sizes.

It is not straightforward to visualize the tensors with more than three modes. For demonstration, Figure 5.2 shows a third-order tensor including the three modes of time, frequency and space. The third-order tensor consists of the time-frequency representation (TFR) of the multi-channels EEG data.

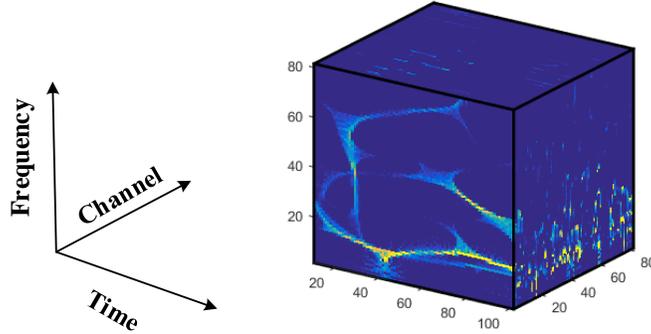


Figure 5.2: third-order tensor of EEG with modes: $time \times frequency \times space$

5.3 EEG tensor generation via RS-EMD

In order to form a high-order EEG tensor that includes the mode of frequency in it, we need to transform EEG data of one segment into the time-frequency domain. As EEG is a multi-component signal and tensor analysis seeks to decompose it into its constituting factors, it is crucial to have a TF representation that is cross-term free and highly localized in time and frequency domains concurrently.

Existing method and its main problem

Continuous Wavelet Transform (CWT) is previously used for generating EEG tensor [53–55, 121]. The continuous wavelet transform $C(a, \tau)$ at scale a of a signal $x(t)$ is defined as

$$C(a, \tau) = \int_{-\infty}^{+\infty} x(t)\phi(a, t, \tau)dt, \quad (5.1)$$

where ϕ is the chosen wavelet. Common choices include the class of biorthogonal wavelets, Debauchy wavelets, and the Morlet wavelets [122]. The connection between the scale a and the frequency f is given by

$$f \approx \frac{f_c}{a \cdot \Delta t}, \quad (5.2)$$

where f_c is the center frequency of the wavelet and Δt is the sampling interval for $x(t)$. The disadvantage of CWT-based time-frequency is the limited resolution, especially in the low-frequency region, which is very important in EEG signal analysis. CWT suffers from the lack of ability to exactly localize a single oscillatory event in time and frequency contemporarily. Indeed, every wavelet that is well-defined in frequency cannot be localized in a well-defined time window and vice versa.

Here, we use the approach that was proposed in Chapter 2 to transform the EEG segment into TFR with a high temporal and spectral resolution, which to the authors knowledge is novel in this context. The RS-EMD method is a combination of Empirical Mode Decomposition (EMD) and Reassignment method.

Multi-way analysis of EEG signals has been incorporated for various application, the purpose of them is often source localization of brain activity [123], research questions of cognitive neuroscience or clinical neuroscience [60], or related to brain-computer interface [124]. Regardless of the specific application, two issues should be addressed when it comes to exploiting multi-way analysis. The first is what type of model to choose for analysis, whether the CANDECOMP/PARAFAC model or the Tucker model would be used for analysis and also how to select the number of components for that model which is regarding the model order selection. The second issue is whether the tensor is for individual-level analysis or for group-level analysis. Regarding the former analysis, the tensor is generated using the EEG data of one subject and usually includes the data of one segment (or called as epoch, trial) of that subject in an EEG experiment. As for the latter analysis, the tensor has more modes and contain the data of multiple trials of one subject or multiple subjects. We categorize the applications of tensor decomposition on EEG based on the above-mentioned classes. In this Chapter, we cover two individual-level applications and then in the next Chapter a group-level analysis is presented. RS-EMD and Reconstruction Error based CANDECOMP/PARAFAC model order selection are the core of these applications.

5.4 Localization of Brain Activities via Multi-way Analysis of EEG Tensor

EEG signal, which measures the electrical activity of neurons, is now widely used to study brain’s physiological, mental and functional abnormalities, due to its non-invasiveness, mobility and its high temporal resolution [1]. The EEG recordings can also be used for monitoring of spontaneous and evoked brain activity that allow for spatio-temporal localization of neuronal activity. Localization of the electrical activities of brain has gained considerable attention over the last decades [125, 126], because of its potential diagnostic value for epilepsy [127], stroke [128], traumatic brain injury [129], and even in brain computer interface applications [130].

Multichannel EEG data in [55, 121] has been constructed as a third-order tensor, with modes: $time \times frequency \times channel$, for seizure analysis. The idea is to stack the Time-Frequency (TF) representations of different channels of EEG as the third mode of the tensor. Effective localizing of EEG events using tensor analysis necessitate exploiting a TF representation technique with high spectral and temporal localization.

Here, our goal is to localize the EEG events using multi-way analysis of $frequency \times time \times space$ EEG tensor, which is generated using RS-EMD method. Then, tensor analysis is exploited for extracting localized signatures that uniquely characterize the underlying EEG events.

5.4.1 Methods

The proposed technique could be divided into two main parts: EEG tensor generation, and EEG tensor analysis. $time \times frequency \times space$ EEG tensor is first generated by calculating the TF representation of all EEG channels, which is done here via RS-EMD that is the combination of EMD followed by calculating the reassigned TF representation of the extracted intrinsic mode functions. Then in the second step, CANDECOMP/PARAFAC decomposition of EEG tensor enables us to localize the spatial and spectral signatures of

the analysed EEG events. CANDECOMP/PARAFAC is more favourable here compared to TUCKER models, as it provides a unique decomposition and the same number of components in each mode that also makes it easier to interpret.

A block diagram of the proposed localization technique is illustrated in Figure 5.3.

5.4.2 Three-way EEG Tensor Generation

One of the key elements of *time* \times *frequency* \times *space* EEG tensor generation is the TF representation techniques we choose, and localization of the extracted signatures are highly dependent on the temporal and spectral resolution of the resulting TF images. Here, we use the proposed RS-EMD method for TFR that exploits the benefits of EMD in decomposing the EEG into spectrally localized components called Intrinsic Mode Functions (IMF), and then take advantage of the reassignment technique that gives us a near optimal resolution in TF plane [43].

5.4.3 CANDECOMP/PARAFAC Modeling of EEG Tensor

Mathematically, a PARAFAC model for a three-way tensor of noiseless data $\underline{\mathbf{X}} \in \mathbb{R}^{I \times J \times K}$ can be written as

$$\underline{\mathbf{X}} = \sum_{r=1}^R \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r, \quad (5.3)$$

where R is the number of components in the model and symbol \circ denotes the vector outer product. In a CANDECOMP/PARAFAC model, a tensor is decomposed into the sum of rank-one tensors, which are defined as the outer product of vectors. In practice the observed data is a contaminated version of $\underline{\mathbf{X}}$ and can be represented as

$$\underline{\mathbf{Y}} = \underline{\mathbf{X}} + \underline{\mathbf{E}}, \quad (5.4)$$

where $\underline{\mathbf{E}} \in \mathbb{R}^{I \times J \times K}$ is a three-way array containing the White Gaussian noise elements. Illustration of a two-component CANDECOMP/PARAFAC model on a three-way EEG

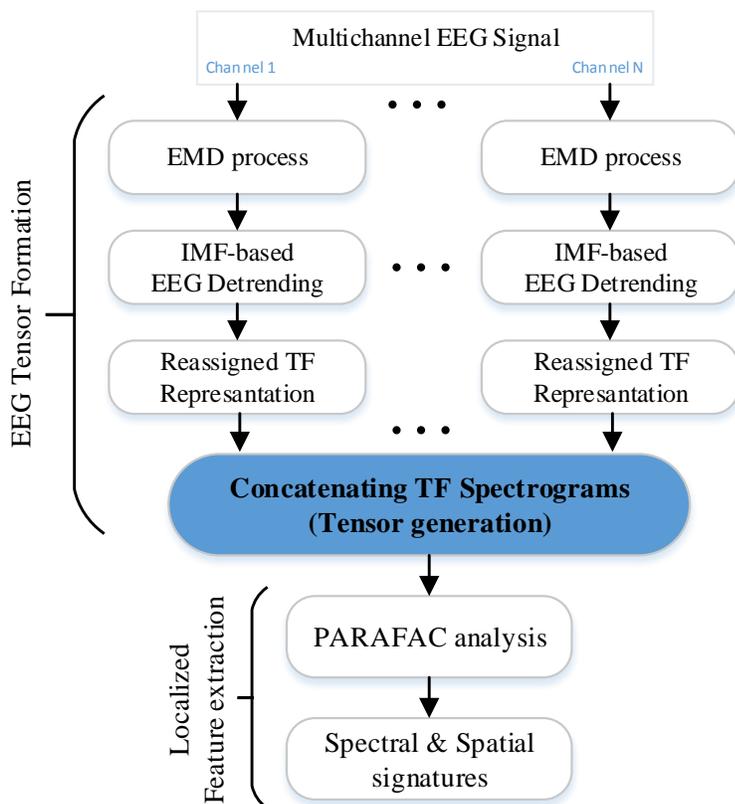


Figure 5.3: Block diagram of the proposed EEG localization algorithm.

data set is given in Figure 5.4.

CANDECOMP/PARAFAC analysis of EEG tensor assumes that every event in the EEG could be uniquely characterized and represented by a temporal and a spectral signature, and that event is elicited in different channel locations with different variations, making the elements of the space signature. TF representations of multichannel EEG data using CWT have been used previously in [121] to construct a three-way EEG tensor which is then decomposed into time-frequency-space components using CANDECOMP/PARAFAC analysis for seizure analysis and source localization. This allows the spectral, spatial, and temporal signatures of EEG events to be found and extracted.

Besides, in order to capture the true underlying structure in data, it is important to extract the right number of components R in the CANDECOMP/PARAFAC model

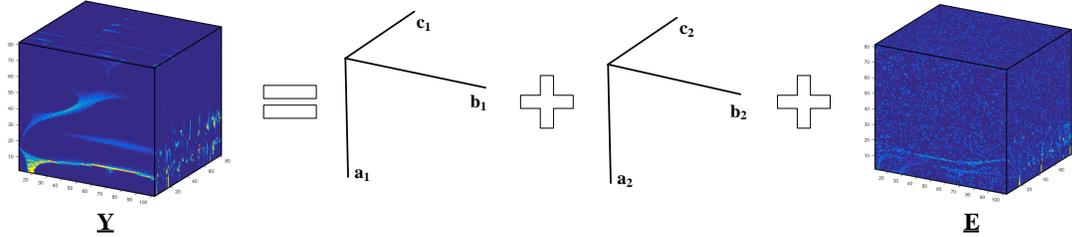


Figure 5.4: two-component CANDECOMP/PARAFAC model, where a three-way array $\underline{\mathbf{Y}}$ is expressed as the sum of two rank-1 tensors and error terms. \mathbf{a}_i , \mathbf{b}_i and \mathbf{c}_i are the i th components in the first, second and third mode, respectively. $\underline{\mathbf{E}}$ is a three-way array containing the residuals.

(5.3), which is the number of existing events in EEG tensor analysis. Selecting the optimum number of components in CANDECOMP/PARAFAC model is still an active challenge. Here, we use our developed method based on Reconstruction Error in the previous Chapter for selecting the optimum model order.

5.4.4 Results

First, we examine the localization capability of the proposed method on a simulated multichannel EEG data and then we use a real EEG dataset to evaluate our algorithm.

5.4.4.1 Synthetic EEG data

In this section, we simulate 2 seconds of 32-channel EEG data sampled at 256 Hz while embedding 3 different events with distinct characteristics in it. These activities are 50 Hz oscillations of amplitude of 0.8 on all channels over the whole 2-second data mimicking the electronic noise, two burst of 35 Hz sinusoidal oscillations with an amplitude of 1.0 and duration of 0.5 second were placed in Channels 30, 31 and 32 at the posterior areas resembling occipital gamma activity, and one burst of 25 Hz oscillation with amplitude of 1.5 and duration of 0.4 second, contemporary happening with previous event, were generated at each ear at Channels 11 and 15. Finally, normal distributed random noise of power 1.0 was added to all channels.

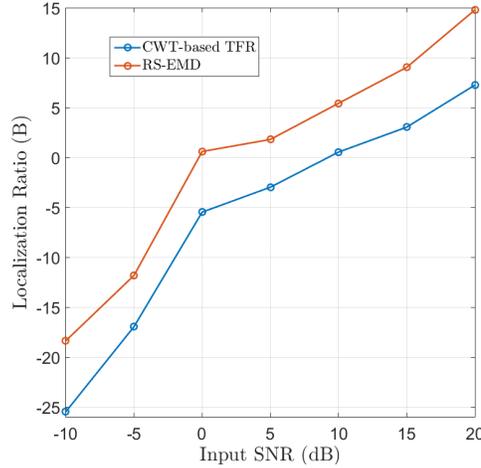


Figure 5.5: A comparison between the localization ratios B , for both the proposed RS-EMD method and the CWT-based TFR, evaluated for a sample EEG signal with 50 Hz oscillation.

Localization Performance Index

In order to quantify the localization performance of the RS-EMD compared to conventional wavelet-based TFR we use the performance index introduced in Chapter 2. The B ratio in (2.23) for one of the EEG channels with 50Hz oscillation and one burst of 25 Hz was compared for RS-EMD and Complex Wavelet transform. The reason for comparing our results with CWT is that it was the only TFR that was previously used for EEG tensor analysis and we used it in our analysis for benchmarking. The resulting mean instantaneous frequency is given as the mean of frequencies when more than one component is present in the signal. However, as we apply EMD to decompose the multi-component signal into mono-components then the instantaneous frequencies are 50 Hz and 25 Hz for each component. Having the instantaneous frequency path, we can calculate the B ratio for a range of different SNR for the input signal. The results are shown in Figure 5.5 where reassigned method gives B for 5.5 dB greater than the Complex Wavelet transform for EEG signal, particularly when the SNR of the input signal is relatively high. This index shows that the proposed RS-EMD techniques can generate a more localized TFR compared to CWT-based TFR. We later in this section explore the effect of this

advantage on the extracted signatures after EEG tensor decomposition.

Synthetic EEG tensor Analysis

The corresponding three-way EEG tensor was generated where we used the window length of 127 and 512 frequency bins in RM calculations. Besides, we also generated another three-way EEG tensor where Wavelet transform with Complex Morlet mother wavelet was used for TFR. We used linear frequency bins within 1 – 80 Hz with the resolution of 0.25 Hz in CWT-based TFR. As we knew priorly that there were 3 components embedded in the EEG data, we approximated the constructed EEG tensor of $\underline{\mathbf{Y}} \in \mathbb{R}_+^{512 \times 512 \times 32}$ with a 3-component CANDECOMP/PARAFAC model. We used the well-known Alternating Least Squares (ALS) algorithm, with non-negative constraint, for estimating the factor matrices. The resulting factor matrices after RS-EMD based EEG tensor analysis are illustrated in Figure 5.6(b). The extracted factor matrices reveals the temporal, spectral, and spatial signatures of all activities where each of the extracted components correspond to one of the embedded events. The results of CWT-based EEG tensor analysis are depicted in 5.6(a). Considering the WT-based results and by particularity comparing the extracted signatures in the spectral domains of both techniques, it interestingly aligns with the previous B ratio results where the extracted signatures using RS-EMD are more localized and confirms the advantage of RS-EMD over CWT for EEG tensor analysis and localizing embedded events. The extracted spectral signatures using CWT for tensor generation show spectral overlap between components in the frequency domain, although it is not true in this case. Therefore, the proposed method is able to effectively localize and separate the ongoing activities in the brain even for the temporally overlapping events.

5.4.4.2 Steady-State Visual Evoked Potential

Steady-State Visual Evoked Potential (SSVEP) is a periodic response elicited in the brain by the repetitive representation of visual stimulus, at frequencies greater than 6

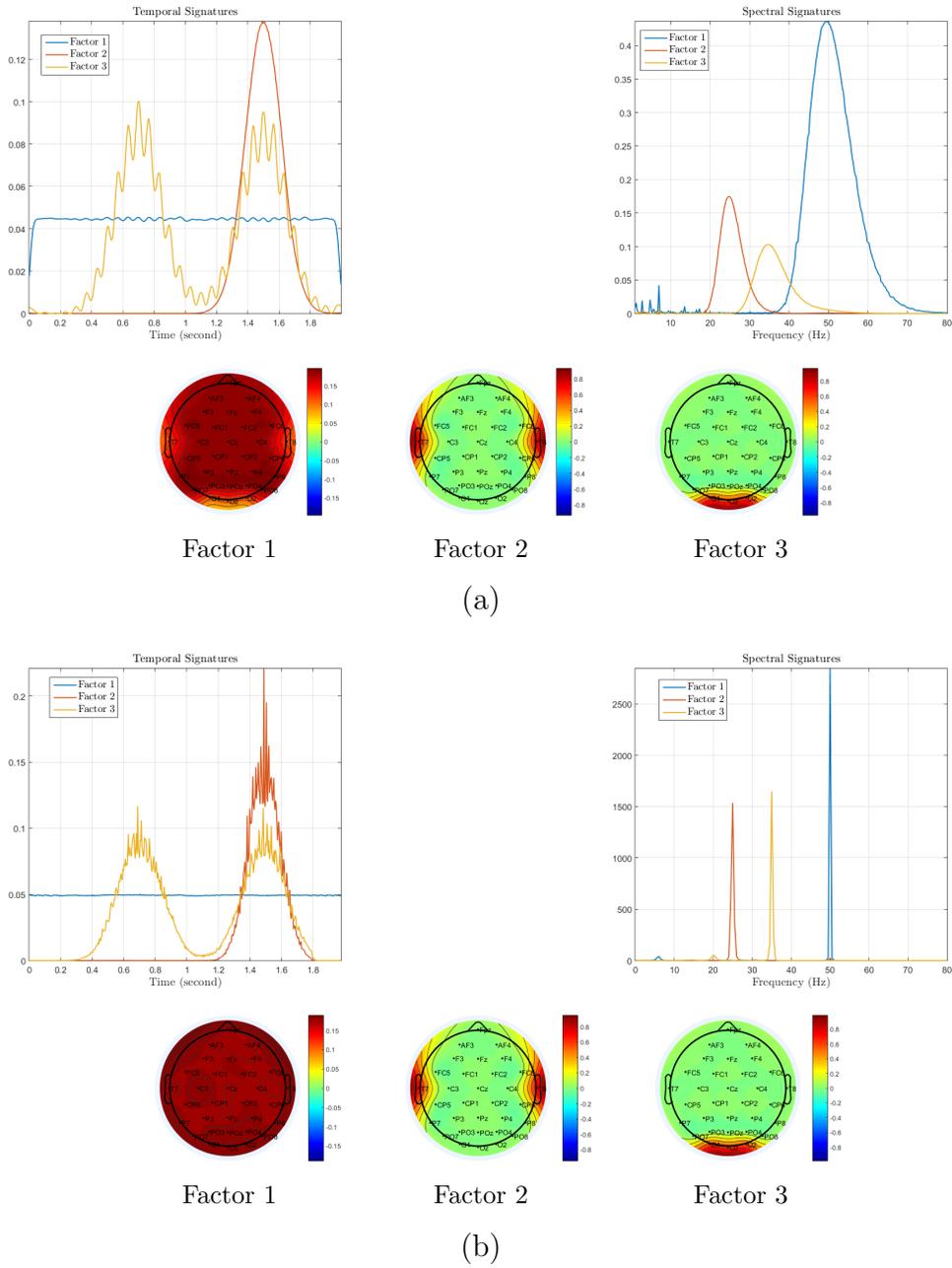


Figure 5.6: 3-component CANDECOMP/PARAFAC decomposition of a synthetic three-way EEG tensor $\underline{\mathbf{Y}} \in \mathbb{R}^{512 \times 512 \times 32}$. temporal signatures, spectral signatures, and topographic maps of spatial signatures of the extracted factors are depicted: (a) using Wavelet transform for TFR; (b) using the proposed technique for TFR.

Hz. SSVEP is characterised with localized spectral and spatial signatures, having a sharp spectral component at the harmonics of the stimulus frequency spread mainly over

the occipital region of the brain [131]. Knowing these characteristics, we now evaluate our method to see how successful it is in localizing SSVEPs and separating it from the ongoing activities of the brain.

SSVEP dataset

The multichannel SSVEP dataset used in this part is from Swartz Center for Computational Neuroscience (SCCN) at University of California San Diego (UCSD) [132], which consisted of ten healthy adults (8 males and 2 females, mean age: 23 years) with normal or corrected-to-normal vision. The 248-channel EEG data were recorded according to 10-20 international system and down sampled to 256 Hz.

Following the onset of SSVEP, a set of EEG channels develop rhythmic activity that is typically composed of multiple frequency harmonics of the visual stimulator. What makes the SSVEP distinctive from ongoing background activities of the brain is the presence of specific frequency components localized on certain channels over the occipital region. Hence, SSVEP would be an EEG event that has localized spectral and spatial characteristics.

Keeping that in mind, we expect that after CANDECOMP/PARAFAC analysis of a SSVEP tensor, one of the extracted factors would correspond to the SSVEP activity characterized by having a frequency component localized over a certain region of topographic map of the brain, while a non SSVEP factors would have several frequency components distributed over topographic map of brain. Due to nonstationarity nature of EEG and as most of the EEG activity is limited up to 50 Hz, a two second long epoch is used and we consider 256 frequency bins between 0 – 50 Hz and a window length of 127 in RS calculations. Therefore for a 249-channel EEG, the constructed *frequency* \times *time* \times *space* EEG tensor is $\underline{\mathbf{Y}} \in \mathbb{R}_+^{256 \times 512 \times 248}$. We approximated the EEG tensor with a 40-component CANDECOMP/PARAFAC model, selected based on our proposed method in the previous Chapter, using ALS algorithm with non-negativity constraint on extracted factors. Nevertheless, it is often suggested to use several diagnostic tools together rather

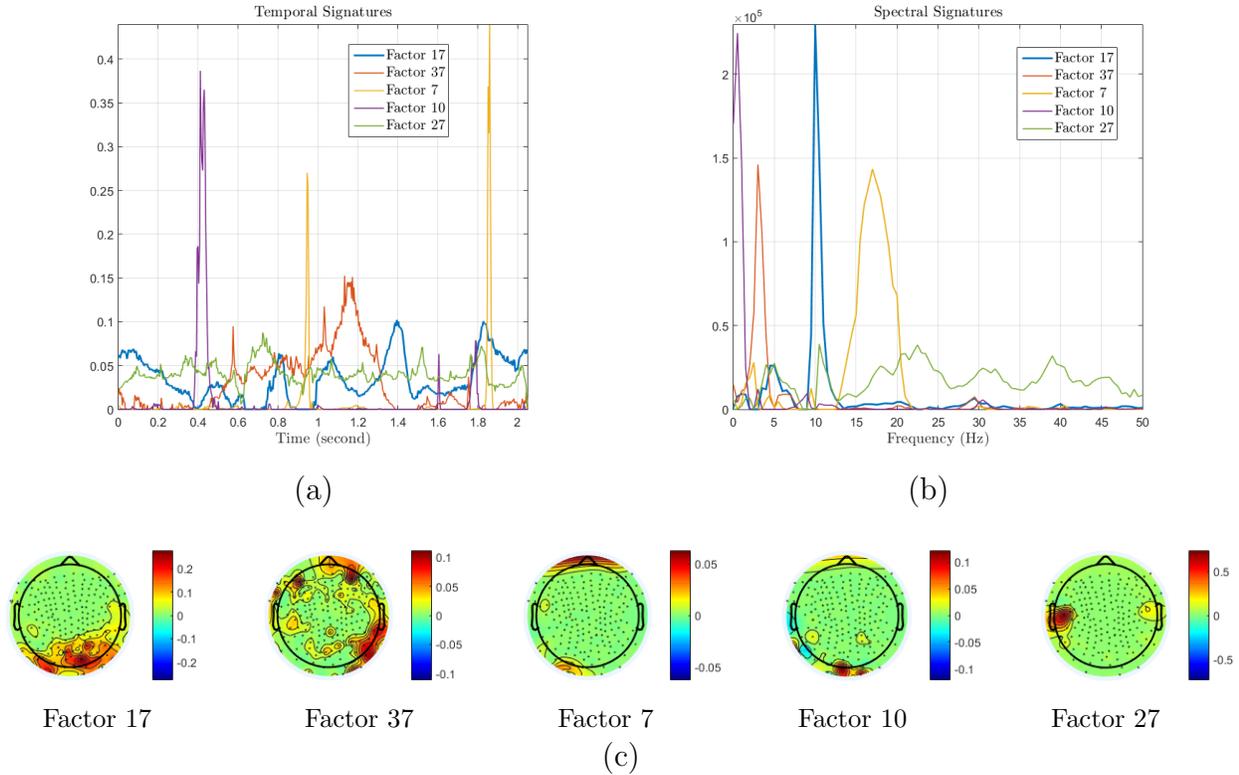


Figure 5.7: CANDECOMP/PARAFAC decomposition of a sample 10Hz SSVEP tensor $\underline{\mathbf{Y}} \in \mathbb{R}^{256 \times 512 \times 248}$: (a) Temporal signatures; (b) Spectral signatures; (c) Topographic maps of Spatial signatures of 5 visually selected factors. The extracted factor #17 characterizes 10 Hz SSVEP and some other extracted factors characterizing the EEG background activities are also shown for comparison.

than a single method, as there is no rule of thumb in determining the optimal number of components (optimal in terms of interpretation) for real data and usually a range of model orders are considered. In our experiments we were able to assign a component to SSVEP, i.e. find a one that satisfies the SSVEP characteristics, for the range of 10 – 50 model orders.

Figure 5.7 shows the temporal, spectral and spatial signatures of the 5 extracted factors out of 40, after CANDECOMP/PARAFAC analysis of a 2-second segment of a 10 Hz SSVEP signal from subject 1 in the database. These 5 factors are visually selected from the total of 40 extracted factors. The first component (Factor #17) is characterized by having a spectral signature localized over 10 Hz, temporal activity over

the whole 2s segment, and with a spatial signature distributed over the occipital region which matches with SSVEP properties. Other selected factors are representing the EEG background activities in different frequency bands and various lobes of the brains. For example, factor #7 could be representing an eye blink artifact, as it has a sharp temporal signature and a spatial signature localized over the frontal region. We got similar and consistent results from all 10 subjects in our simulations. These examples prove that if an EEG event has intrinsically localized spectral and spatial characteristics, multi-way analysis of the EEG tensor, in the form of the proposed framework, can be used for localizing the brain activities, with a set of signatures that uniquely characterize that event. Extracted localized features, then, could be effectively used for various purposes, e.g. Brain Computer Interface for classification of different brain states or for seizure localization and determining its onset-zones.

5.5 New method for SSVEP Detection via CANDECOMP/ PARAFAC and Canonical Correlation Analysis

The decomposition of EEG signals into interpretable building components has been of great interest along the years. However, traditionally two-way decomposition techniques, considering just two contextual dimensions or “signatures”, e.g. time and frequency or time and space, have been used in the literature, which needs exploiting some additional constraints in the decomposition for the sake of uniqueness. In order to obtain a more natural representation of an original multi-dimensional data structure, the use of tensor decomposition approaches are necessary, since additional dimensions or modes can be retained only in multi-linear models to produce structures that are unique and which admit interpretations that are neurophysiologically meaningful. Advances in developing high-spatial density array EEG have called to multi-dimensional signal processing tech-

niques as multi-way analysis, multi-way array (tensor) factorization/decomposition or dynamic tensor analysis [61].

Tensor-based methods are a more natural approach to handle signals that vary in more than two dimensions and seek to summarize the data into extracted factors, which are linear combination of the constituting variables. The well-known CANDECOMP/PARAFAC decomposition is a powerful approach to decompose a tensor into building components.

In the last years, several works have been performed in applying CANDECOMP/PARAFAC for EEG signal analysis, e.g., for estimating the sources of cognitive processing using a Wavelet decomposition [53], Event-Related Potential (ERP) analysis [54], artifact removal [55] and epileptic seizure localization [121]. It was also used in a BCI based on motor imagery [61].

On the other hand, several approaches to detect the SSVEPs have been developed in the literature. SSVEP occurs as evoked potential during intermittent photic stimulation (IPS). IPS is widely used in clinical practice (e.g. to detect abnormal epileptogenic sensitivity to flickering light) and in neuroscience research. In basic research IPS can be used to obtain a better understanding of the alpha rhythm and it can be utilized for BCI applications. If the stimulation frequency is > 5 Hz, then individual VEP responses overlap and result a quasi-sinusoid oscillation with the same frequency as the stimulus (SSVEP). “SSVEP-BCI” uses the excitation of the retina of eye by a stimulus at a certain frequency, making the brain generating an electrical activity of the same frequency with its multiples or harmonics. The objective in SSVEP-based BCI applications is to detect this frequency reliably with high accuracy and furthermore to detect when the frequency is not present, thus when the person does not look at the stimulus. The detection of the SSVEP is challenging if the stimulus frequency is in the near of the natural alpha rhythm because SSVEP and alpha rhythms can merge into one because the alpha changes its frequency after flicker onset towards the flicker frequency (entrainment effect). Therefore appropriate time-frequency techniques must be used in order to realize a reliable SSVEP detection with an optimal time-frequency resolution, i.e. for each electrode N (space)

a time-frequency representation of the dimension sampling points (time) \times frequency bins (frequency) result. The results of topographic time-frequency analysis can be seen as three-way tensor. With the application of tensor decomposition the EEG tensor can be summarized into components which can improve the interpretation and visualization of such data. The resulting factors (a number of factors for the modes space, time, and frequency) can be used as input data for classification algorithms (e.g. k-means, Support Vector Machine (SVM)). There are several studies which use this strategy for VEP analysis, detection and classification [133–136]. In [135], three classes of geometric figures were evaluated, flickering at 15 Hz, and a SVM classifier was used to discriminate classes from feature vectors, achieving the highest classification accuracy (80%). [136] used two techniques (“maximum method” and “sharpness method”) for recognition of SSVEPs and all possible 2, 4 and 12 stimuli combinations are compared for a BCI system.

Here, we modify this procedure and introduces a combination of tensor decomposition and canonical correlation analysis (CCA) in order to detect SSVEP, which is novel in this context as CCA have been used for SSVEP detection in the original signal domain and using CCA in the transformed domain is novel. The automatic detection of SSVEPs is done through correlation analysis between tensor models (using three-way: time \times frequency \times space) and simulated tensor model (“template”). The CCA procedure maximizes the correlation between the linear combination of variables. Being a simple correlation in the end, CCA is often thought of as a descriptive procedure to investigate the relationship between (two random) variables. Using CCA as the detection algorithm we compare each of the extracted signatures of the CANDECOMP/PARAFAC model with the corresponding simulated signatures of a target SSVEP signal. We only consider the spectral and spatial signatures, considering that the SSVEP is present in the temporal signature. A maximum criterion decides which target is intended every 1 second, which makes it suitable for real-time BCI applications.

5.5.1 CANDECOMP/PARAFAC Analysis of EEG tensor

Three-way EEG tensor of *time* \times *frequency* \times *space* was generated and decomposed according to (5.3). The multi-way array shows the variation of EEG signals in temporal, spectral and spatial domain (Figure 5.2). TF representation of EEG signals was calculated via the novel RS-EMD method developed in the previous section exploiting EMD and RM techniques. We used ALS algorithm, with non-negative constraint, for estimating the factor matrices. Each of the extracted components via CANDECOMP/PARAFAC analysis uniquely represents one of the embedded events in the analyzed EEG segment which is characterized with three signatures in time, frequency and spatial domain.

5.5.2 Canonical Correlation Analysis (CCA)

As a multivariate statistical method, canonical correlation analysis (CCA) explores the underlying correlation between two sets of data. Given two sets of random variables $\mathbf{X} \in \mathbb{R}^{I_1 \times J}$ and $\mathbf{Y} \in \mathbb{R}^{I_2 \times J}$, which are normalized to have zero mean and unit variance, CCA is to seek a pair of linear transforms $\mathbf{w}_x \in \mathbb{R}^{I_1}$ and $\mathbf{w}_y \in \mathbb{R}^{I_2}$, called canonical variables, such that the correlation between linear combinations $\tilde{\mathbf{x}} = \mathbf{w}_x^T \mathbf{X}$ and $\tilde{\mathbf{y}} = \mathbf{w}_y^T \mathbf{Y}$ is maximized as

$$\begin{aligned} \max_{\mathbf{w}_x, \mathbf{w}_y} \rho &= \frac{\mathbb{E} [\tilde{\mathbf{x}} \tilde{\mathbf{y}}^T]}{\sqrt{\mathbb{E} [\tilde{\mathbf{x}} \tilde{\mathbf{x}}^T] \mathbb{E} [\tilde{\mathbf{y}} \tilde{\mathbf{y}}^T]}} \\ &= \frac{\mathbf{w}_x^T \mathbf{X} \mathbf{Y}^T \mathbf{w}_y}{\sqrt{\mathbf{w}_x^T \mathbf{X} \mathbf{X}^T \mathbf{w}_x \mathbf{w}_y^T \mathbf{Y} \mathbf{Y}^T \mathbf{w}_y}}. \end{aligned} \quad (5.5)$$

The maximum of correlation coefficient ρ with respect to \mathbf{w}_x and \mathbf{w}_y is the maximum canonical correlation.

A set of template signals is usually generated for each of the targets in SSVEP-BCI and after calculating the correlation between recorded SSVEP and all the targets, the one that has the highest correlation ρ is selected as the intended target by the subject. CCA

is a technique widely used in the processing of multi-channel EEG signals for detection of SSVEP components [137, 138]. In most of them CCA is exploited in the original domain, meaning that template signals are in time domain resembling sine and cosine functions with different frequencies. However, the novelty of the proposed technique is that we calculate the correlation in the signature domain, meaning that extracted spectral and spatial signatures after CANDECOMP/PARAFAC analysis are used to calculate the correlation with a set of defined templates for each of the targets.

5.5.3 Methods

5.5.3.1 Subjects and EEG preparation

Five healthy male subjects (mean age: 26.2 years old; standard deviation: 2.3) were recruited to participate in this study. The experiments were performed in the BCI lab of the Federal University of Espirito Santo in Brazil according to the rules of the ethics committee of UFES/Brazil, under registration number CEP-048/08. Volunteers were screened and selected having no problem in the topics related to precautions as visual problems, headaches, family history with epilepsy and brain damage.

5.5.3.2 System architecture, visual stimulus and experimental task

The equipment used for EEG signal recording was the BrainNet-36, manufactured by Lynx Tecnologia Ltd. Twelve channels of EEG signal with the reference electrode at the left ear lobe were recorded at 600 samples/s, with 1 to 100 Hz pass-band filter. The ground (GND) electrode was placed on the forehead. Using the extended international 10-20 system, the electrode positions chosen were P7, PO7, PO5, PO3, POz, PO4, PO6, PO8, P8, O1, O2 and Oz (see Figure 5.8(a)). The volunteers sat on a comfortable chair, in front of a 17-Inch LCD screen, 70 cm far from this.

A coupling structure of small boxes (4cm x 4cm x 4cm) containing Light-Emitting Diodes (LEDs) was mounted in two sides of the LCD and away from each other in

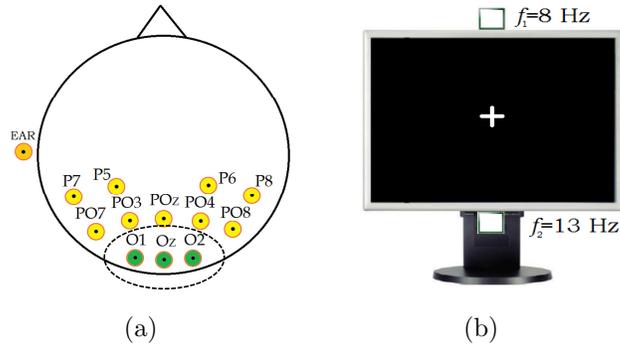


Figure 5.8: (a) Electrodes location using 10-20 system; (b) LCD screen showing the cue of beginning with the two LEDs coupled.

30cm (see Figure 5.8(b)). The participants were asked to watch the visual stimuli (two LED flickers), the timing was controlled by a microcontroller (PIC18F4550, Microchip Technology Inc., USA) with 50/50% on-off duties). Two LEDs (luminous intensity of 10.000 mcd) of white color, covered with thin white papers diffusers were used. The flickering frequencies were 8.0 Hz (top) and 13.0 Hz (bottom), which were presented simultaneously to the user. The frequencies were selected from outside of 9 – 11Hz band to avoid the interference of the alpha rhythm.

The experiments were performed off-line. During the first five seconds a cross on the screen is shown to the volunteers. Then, a beep sound is played indicating the onset of trial. The volunteer was instructed to fix the attention on the flicker located on the top. After fifteen seconds, the volunteer rested for five seconds. Finally, in the last fifteen seconds, the volunteer had to fix the attention to the bottom flicker.

5.5.4 Results

Initially, two three-way *time* \times *frequency* \times *space* template tensors for two target SSVEP signals of 8 Hz and 13 Hz were generated. The template SSVEP signals constituted of the summation of sine and cosine signals at the target SSVEP's frequencies. A 2s signal sampled at 256Hz for each target frequency is generated and its TFR is calculated afterwards, considering 80 frequency bins. Based on the hypothesis that the visual evoked

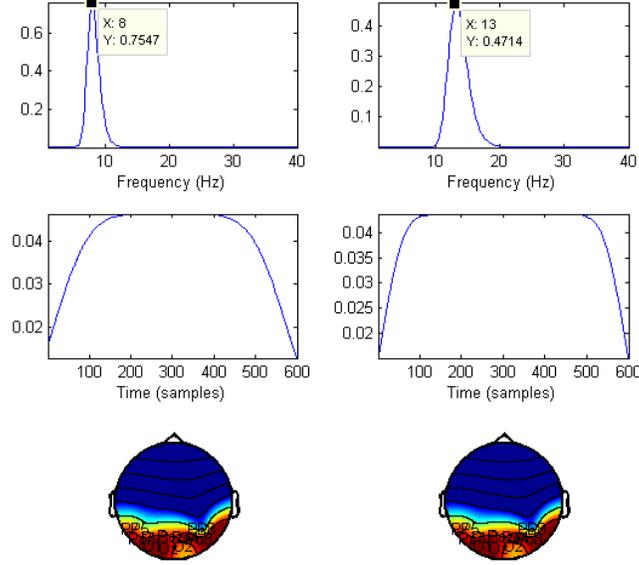


Figure 5.9: Signatures of the SSVEP templates for 8 Hz (left panel), and 13 Hz (right panel). The signatures characterise the spectral (top), temporal (middle), and spatial (bottom) dynamics of the corresponding SSVEP signal.

potentials contain more energy in the occipital region, the spatial signature was considered just for O1, O2 and Oz channels, and zero elsewhere. Hence, a three-way template tensor of $\underline{\mathbf{X}} \in \mathbb{R}_+^{80 \times 512 \times 12}$ was formed for each target frequency. Then a 1-component CANDECOMP/PARAFAC model was fitted to extract the signatures of the templates. Figure 5.9 shows the signatures of SSVEP templates for 8 Hz and 13 Hz, reflecting an activity during the whole segment with a dominant corresponding frequency component localized over the occipital region of the brain.

On the other hand, every 2-second of the 12-channel input EEG signals without overlapping was used to generate the $\underline{\mathbf{Y}} \in \mathbb{R}_+^{80 \times 512 \times 12}$ EEG tensor. Then a CANDECOMP/PARAFAC model was fitted where the model order R was selected via our proposed approach in Chapter 4. The output of each CANDECOMP/PARAFAC model for EEG signal has R atoms or signature (spectral (\mathbf{a}_r), temporal (\mathbf{b}_r), and spatial (\mathbf{c}_r), where $r = 1, 2, \dots, R$), in each mode according to (5.3).

Based on the proposed approach, the estimation of correlations between the extracted signatures of EEG tensor and the corresponding template signatures should be calculated.

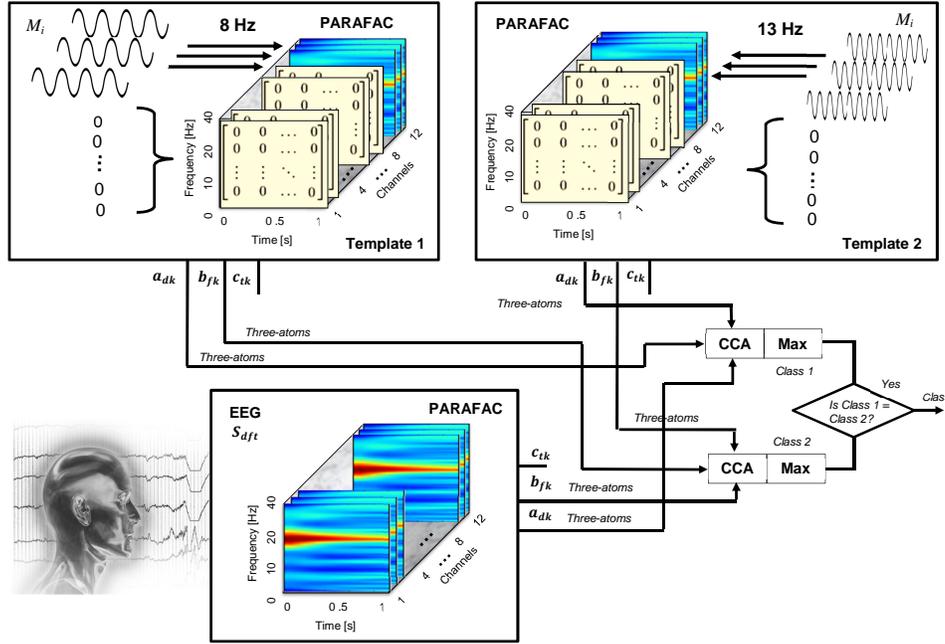


Figure 5.10: Schematic of the proposed technique for SSVEP detection.

After decomposition only the spectral and spatial signature are used here for calculating the correlations, and the temporal signature is ignored, as we assume that there is a SSVEP signal in the EEG segment under analysis. Therefore, these signatures were analyzed with the aim of ensuring the existence of evoked potentials in that region of the brain and also locating its respective frequency component. Finally, the frequency that has the highest correlation is selected as the intended target. A graphical explanation of the process is illustrated in Figure 5.10.

The spectral signature of all extracted components for subject 1 is depicted in Figure 5.11. We can see clear peaks at 8 and 13 Hz when the subject was stimulated with 8 and 13 Hz, respectively. The figure also shows baseline frequency activity extracted in some components. Exploiting a detrending technique could improve the results. Accuracy and Information Transfer Rate (ITR) [bits/min] for all the subjects participated in the experiments are shown in Table 5.1.

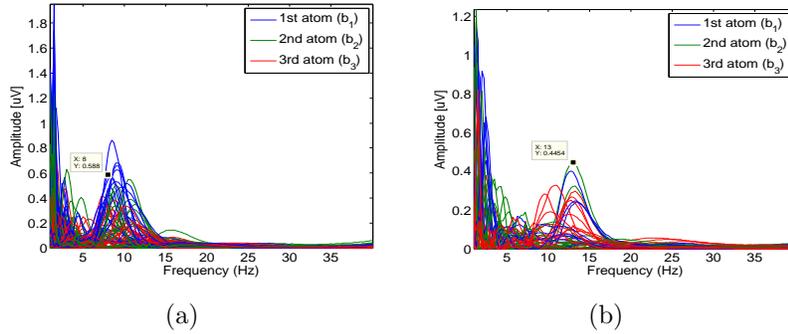


Figure 5.11: Extracted spectral signatures via CANDECOMP/PARAFAC, (i.e. \mathbf{a}_1 , \mathbf{a}_2 and \mathbf{a}_3) when Subject 1 was stimulated with: (a) 8 Hz and (b) 13 Hz, respectively.

ITR can be calculated as follows

$$ITR = \log_2 N + P \log_2 P + (1 - P) \log_2 \frac{1 - P}{N - 1}, \quad (5.6)$$

where N is the number of targets in BCI and P the accuracy of classification [139].

The results are quite promising: the subjects 3 and 5 achieved the highest mean accuracy of 83.3% considering a window length of 2 second. Indeed, increasing the window length would improve the results further as there is a trade-off between accuracy and speed. On the other hand, SSVEP-based BCIs are preferable due to achievable high information transfer rates and no need of training for the subjects. Besides, the extracted components via CANDECOMP/PARAFAC can separate overlapped events in the brain and localize the SSVEP signals. If there exists an SSVEP in the EEG

Table 5.1: Accuracy Results Using WL of 2 s

Acc. [%]	Frequency (class)		Results	
	8 Hz	13 Hz	Average	ITR [bits/min]
Subject 1	80.0%	60.0%	70.0%	7.13
Subject 2	73.3%	60.0%	66.7%	4.91
Subject 3	86.7%	80.0%	83.3%	21.01
Subject 4	73.3%	80.0%	76.7%	12.98
Subject 5	80.0%	86.7%	83.3%	21.01

segment, CANDECOMP/PARAFAC enables us to extract and separate it from ongoing activities of the brain. Then, the extracted signatures can be effectively used via CCA for SSEVP detection. The proposed methodology was effectively used in the BCI system and detected events can be further translated into commands for controlling a wheelchair or any device control.

5.6 Chapter Summary

In this Chapter we talked about Multi-way nature of EEG signals and various applications including source localization of brain activities, research questions of cognitive neuroscience, or brain-computer interface. Two individual-level analysis was presented in this Chapter. the first one, is for localization of brain events. Various oscillatory activities within EEG may overlap, However, tensor decomposition by means of Time-Frequency-Space transformation makes it possible in many cases to isolate each oscillatory behaviour well, even when these activities are not well-separated in two-way domain. Temporal, spatial and spectral information of the multichannel EEG are used here to generate a three-way EEG tensor. RS-EMD that is combination of EMD and Reassigned TFR, with a high-resolution TF representation characteristic, is exploited here for EEG tensor generation. Simulation results on synthetic and real three-way EEG data show that the proposed method could be effectively used for separation and localization of events in EEG. Besides, the clinically meaningful extracted signatures could be used for various purposes. In the second application, an SSVEP-BCI system was designed that uses the extracted signatures via CANDECOMP/PARAFAC in order to detect SSVEP signals using a canonical correlation analysis approach. The results are quite promising and proves the virtue of the proposed framework for EEG analysis. The proposed framework let us map the multi-channel EEG signals into higher-dimensions and help us separate and localize the events that have overlap in some domains.

In the next Chapter, a group-level analysis of EEG signals using a four-way EEG

tensor, including the mode of subjects/conditions, is presented. We also develop a methodology to select the desired component characterising the ERP after CANDECOMP/PARAFAC analysis.

Chapter 6

New Multi-domain Feature Selection Scheme from EEG tensor

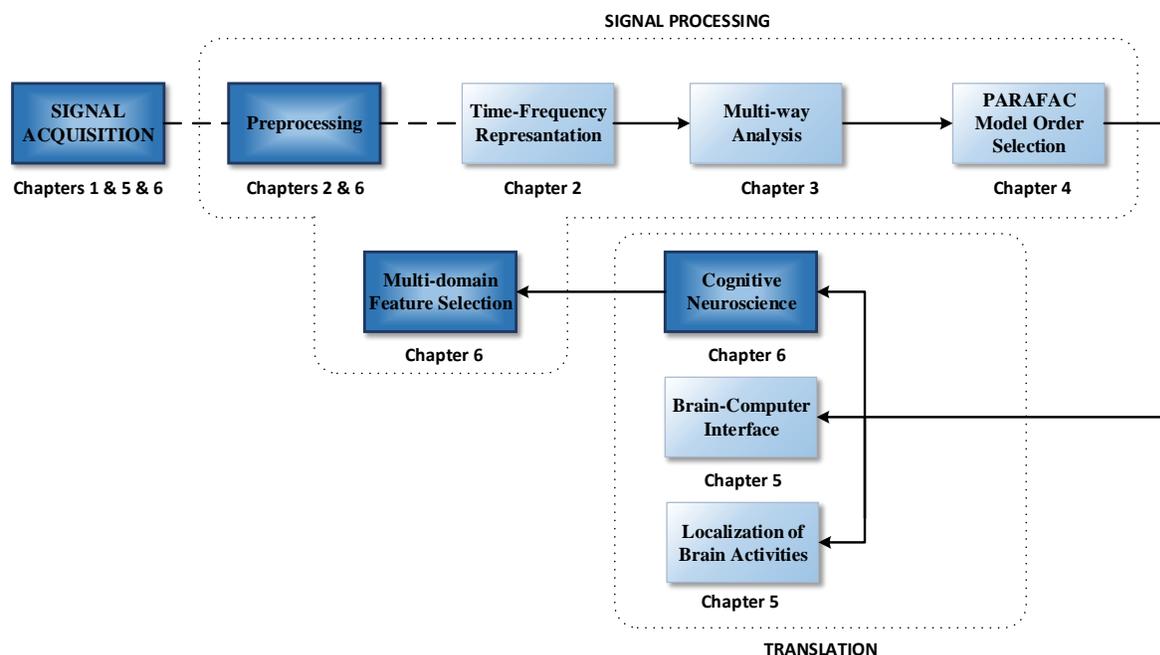


Figure 6.1: Chapter 6- Multi-domain Feature Selection scheme from EEG tensor.

EEG reflects the electrical activity of the brain neurons and it has been widely used for different clinical and diagnostic purposes. These neuronal activities are not only the

spontaneous brain activities, but also include those elicited by external stimuli which are referred to as event-related potentials (ERPs) [140]. The main sensory modalities used for eliciting the response are auditory, somatosensory, and visual modalities. One of the small ERPs that is evoked with small changes in the acoustic environment is the mismatch negativity (MMN). MMN is elicited via an oddball paradigm which is the presentation of an *oddball* or *deviant* event, embedded in a stream of repeated or familiar events, the *standards*. It is a negative component in the response obtained by subtracting the response to the *standard* event from the response to the *deviant* event. MMN is characterised with peak amplitude of 0.5 - 5 μV with the latency of about 100 - 250 *ms* from the deviant onset, and exhibits the strongest intensity in the temporal and frontal areas of topographic scalp maps [141]. The MMN reflects the brain's ability to perform automatic comparisons between consecutive stimuli and provides an electrophysiological index of sensory learning and perceptual accuracy, and it has been extensively used in the research of cognitive processing, clinical neuroscience, and neuro-pharmacology [141, 142].

As the signal to noise ratio (SNR) is very low in EEG recordings and the MMN is a relatively small ERP, signal processing becomes an important part of the MMN research. The most common feature that is used to represent MMN, like other ERPs, is the peak amplitude. This feature is not robust as it only reveals the information of MMN in the time-domain for one electrode and is subject to high sensitivity to many experimental conditions. Other contextual information, e.g., spectrum in the frequency domain, the region of interest in the time-frequency representation (TFR), and even combination of one of these features with the information in the spatial domain to produce the topography map of the ERP, are also used to characterise the ERPs in the literature [140]. So, the information in different domains is exploited sequentially and the features are measured across multiple channels and participants. In order to detect differences at the group or condition level, some statistical analysis on these data are carried out. Group level analyses play an important role in paradigms where SNR is very low [143]. Indeed, the individual features that take into account one domain, may be less sensitive

to identify the difference between groups of subjects than an analysis that can exploit more information of brain responses in more domains simultaneously.

Recently, the so-called multi-domain features were used for group-level analysis of ERPs [144, 145]. In this method the natural structure of the EEG signals is exploited in the form of multi-way array (tensor) by representing the ERP of multi-participants in the time, frequency and spatial domains concurrently. The extracted multi-domain feature of an ERP using Nonnegative Tensor Factorization (NTF) [61], reflect the multi-linear structure of the associated brain activity across multiple subjects and can reveal the properties of the ERP in different domains simultaneously.

CANDECOMP/PARAFAC is one of the popular NTF models that decomposes the observed tensor into summation of rank-one tensors [62]. Each of these extracted rank-one components represent a multi-domain feature along its constituting signatures. We are facing two challenges for representing MMN with a multi-domain feature: First, what is the optimum number of features (rank-one components) that we need to extract from the ERP-tensor, and second, how to select the multi-domain feature extracted by NTF that really corresponds to the desired MMN. CANDECOMP/PARAFAC analysis enables us to decompose the observed ERP tensor into its constituting brain activities which are characterised by a set of signatures. The challenge is to select the component that corresponds to the desired activity. Here in this Chapter, we address the multi-way feature selection problem which is usually done by visual inspection of all the extracted temporal and spectral signatures and finding the one that matches our prior knowledge about the MMN characteristics in time and frequency domain. We propose a new scheme based on a template matching approach which is novel in this context. Furthermore, we use the Reconstruction Error CANDECOMP/PARAFAC model order selection and the RS-EMD method for TFR of the ERP segments in the analysis.

6.1 Methods

The proposed technique could be divided into two main parts: ERP tensor generation, and multi-domain feature selection via ERP tensor analysis. Having a multi-channel dataset across two group of subjects/conditions, a 4-way ERP tensor of $time \times frequency \times channel \times subjects/condition$ is first generated by calculating the Time-Frequency (TF) representation of all EEG channels, which here is done by the combination of Empirical Mode Decomposition (EMD) followed by calculating the reassigned TF spectrogram of the extracted intrinsic mode functions. Then in the second step, CANDECOMP/PARAFAC analysis of ERP tensor enables us to decompose the tensor into the constituting components and select the desired multi-way feature via a template matching approach.

6.1.1 ERP Data Description

The multi-channel ERP dataset was recorded in Auditory Development Lab at McMaster University. 22 participants (17 – 22 years old, mean age 18.93 ± 1.39), neurologically healthy, right-handed, and with normal hearing, were recruited from the McMaster University community. The McMaster University Research Ethics Board approved all procedures. The 128-channel EEG data were recorded using Biosemi Active Two amplifier according to 10 – 20 international system and later on reduced to 81-channels and down sampled to 256 Hz for analysis. The experiments were conducted in a sound-attenuated room. Two recorded piano tones with frequencies of 262 Hz and 494 Hz were used as the standard and the deviant stimuli, respectively. Each participant was presented with a continuous sequence of tones in two sessions, each lasting 30 minutes, while they watched a silent movie on a computer screen. The tones were presented in an oddball sequence with the inter-onset interval (IOI) of 500 ms and a constraint that two deviant tones could not be presented sequentially. Total of 3600 tones were totally presented in each session to the subjects with the deviance occurrence rate of 10% in one session and 20%

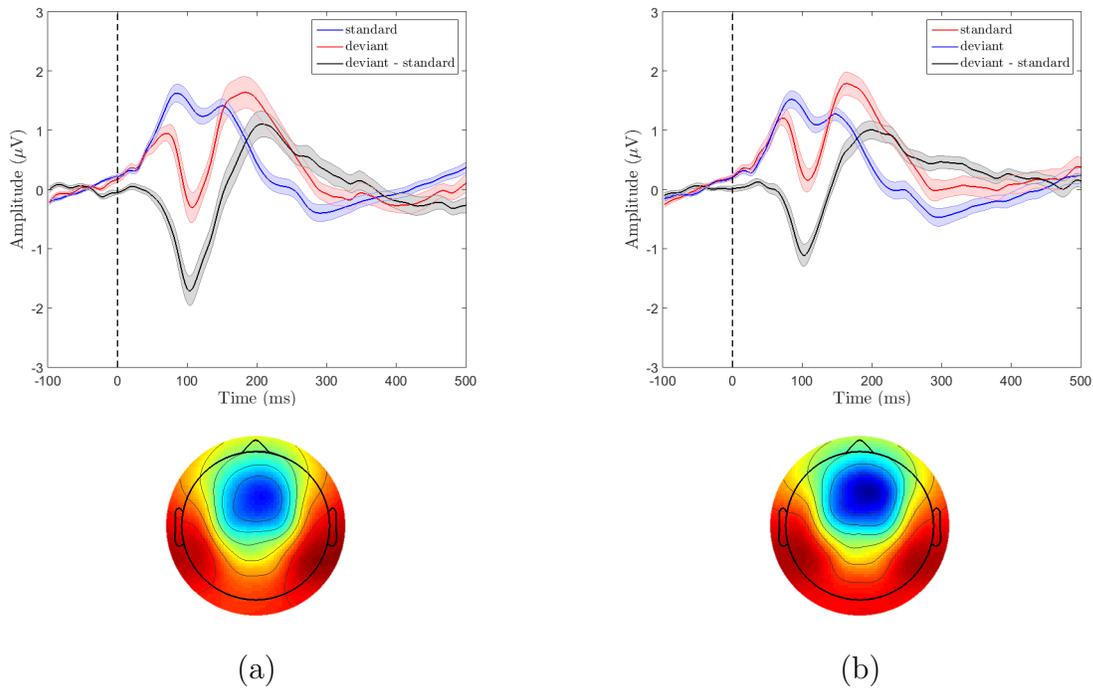


Figure 6.2: Auditory evoked event-related potential (ERP) waveform averaged across channels located at the mid-frontal area (F1, Fz, F2, FC1, FCz, and FC2) for (a) 10% session;(b) 20% session. Top: standard trial (blue), deviant trial (red), and the difference waveform of deviant minus standard trial (black). The shaded areas indicate the standard error of mean. Bottom: The topography of the MMN shows the typical frontal negativity of the MMN.

in the other session. For more details about the dataset and experimental settings please refer to [146].

The continuous EEG data were band-pass filtered between 0.4 – 25 Hz and segmented into single trials including 100 ms pre-stimulus period and 500 ms after the stimulus onset. The baseline was corrected based on the average amplitude of the 100 ms pre-stimulus period. Trials with signal amplitudes beyond the range between -100 and $100 \mu V$ in any recording channel were rejected. The remaining artifact-free single trials were then averaged to obtain the ERP data for the standard and deviant stimuli and subsequently the difference ERP. Figure 6.2 shows the resulted waveforms for 10% session and 20% session, including the MMN negative peak around 100 ms and the topography of the MMN with the typical frontal negativity. The negative peak in 10% session has the

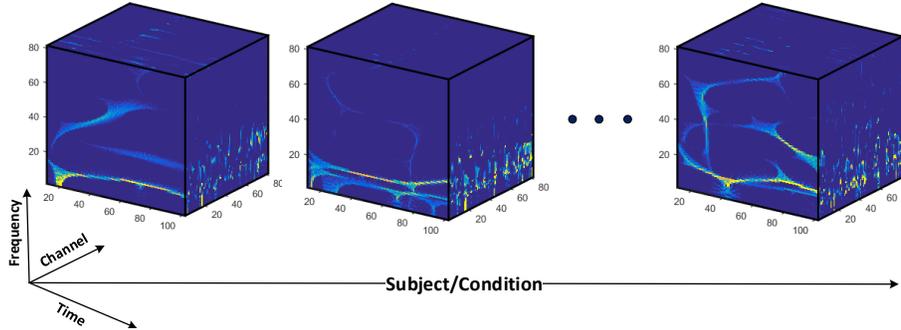


Figure 6.3: Illustration of a four-way ERP tensor of $time \times frequency \times channel \times subject/condition$ which can be decomposed into its constituting factors via CANDECOMP/PARAFAC analysis.

tendency to be higher than the negative peak in 20% session, however this peak analysis may not be significantly different across two group of subjects. Hence, considering more contextual information in the form of multi-way array could be beneficiary in a sense that the extracted multi-domain feature can significantly discriminate two groups/conditions.

6.1.2 ERP-tensor Generation via RS-EMD

Multichannel EEG data are naturally high-dimensional with variation of the signal in temporal, spectral and spatial domains. Furthermore, for a group-level analysis of ERP, the subjects/condition mode can also be considered for formulating the 4-way ERP-tensor. The observed ERP-tensor can be regarded as a mixture that includes different kinds of brain activity, artifacts, interference, and noise. Using tensor decomposition, the desired brain activity could be extracted out from the mixture.

A 4-way ERP tensor of $time \times frequency \times channel \times subjects/condition$ is generated by calculating the Time-Frequency (TF) representation of all EEG channels. Figure 6.3 illustrates the 4-way tensor used for group-level ERP analysis. The key element of this procedure is the TF representation techniques we incorporate, as the localization of the extracted signatures via NTF is highly dependent on the temporal and spectral resolution of the resulting TF images. Here, we use the RS-EMD technique that we proposed earlier

in Chapter 2 and also in [147] which exploits the combination of EMD of ERP signal and the Reassignment method (RM) calculation of the resulting Intrinsic Mode Functions (IMF) to yield a high resolution and cross-term free TF representation for better localizing of brain activities.

6.1.3 CANDECOMP/PARAFAC Modeling of ERP tensor

Mathematically, a CANDECOMP/PARAFAC model for a 4-way tensor of noiseless data $\underline{\mathbf{X}} \in \mathbb{R}^{I \times J \times K \times L}$ can be formulated according to (3.18) as

$$\underline{\mathbf{X}} = \sum_{r=1}^R \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r \circ \mathbf{d}_r, \quad (6.1)$$

where R is the number of components in the model and \mathbf{a}_r , \mathbf{b}_r , \mathbf{c}_r , and \mathbf{d}_r indicate the r th column of component matrices $\mathbf{A} \in \mathbb{R}^{I \times R}$, $\mathbf{B} \in \mathbb{R}^{J \times R}$, $\mathbf{C} \in \mathbb{R}^{K \times R}$, and $\mathbf{D} \in \mathbb{R}^{L \times R}$, respectively. The symbol \circ denotes the vector outer product. In a CANDECOMP/PARAFAC model, a tensor is decomposed into the sum of rank-one tensors, which are defined as the outer product of vectors. Vector outer product for a four-way tensor is defined as follows. Let \mathbf{a} , \mathbf{b} , \mathbf{c} , and \mathbf{d} be column vectors of size $I \times 1$, $J \times 1$, $K \times 1$ and $L \times 1$ respectively, and $\underline{\mathbf{X}}$ be a tensor of size $I \times J \times K \times L$, then $\underline{\mathbf{X}} = \mathbf{a} \circ \mathbf{b} \circ \mathbf{c} \circ \mathbf{d}$ if and only if $x_{ijkl} = a_i b_j c_k d_l$ where x_{ijkl} is the element (i, j, k, l) of a forth-order tensor and the i th entry of a vector \mathbf{a} is denoted by a_i . CANDECOMP/PARAFAC model of a four-way tensor can also be written in the form of n-mode tensor products as

$$\underline{\mathbf{X}} = \mathbf{I}_{4,R} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} \times_4 \mathbf{D}. \quad (6.2)$$

In practice the observed data is a contaminated version of $\underline{\mathbf{X}}$ and can be represented as

$$\underline{\mathbf{Y}} = \underline{\mathbf{X}} + \underline{\mathbf{E}}, \quad (6.3)$$

where $\underline{\mathbf{E}} \in \mathbb{R}^{I \times J \times K \times L}$ is a four-way array containing the white Gaussian noise elements. Illustration of a four-way ERP tensor data is given in Figure 6.3.

CANDECOMP/PARAFAC algorithms try to estimate the component matrices through minimization of some distance, usually the Frobenius norm, between the observed tensor data and the used model as follows

$$\mathcal{D}(\underline{\mathbf{Y}}|\{\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}, \hat{\mathbf{D}}\}) = \frac{1}{2} \|\underline{\mathbf{Y}} - \underline{\mathbf{I}}_{4,R} \times_1 \hat{\mathbf{A}} \times_2 \hat{\mathbf{B}} \times_3 \hat{\mathbf{C}} \times_4 \hat{\mathbf{D}}\|_F^2. \quad (6.4)$$

where $\hat{\cdot}$ indicates the estimated factor matrices. Alternating Least Squares (ALS) is the conventional method for estimating the factor matrices for CANDECOMP/PARAFAC analysis which takes the number of components R as the input parameter and estimates the component matrices. Hence, the four-way ERP tensor of *time* \times *frequency* \times *channel* \times *subject/condition* can be decomposed using PARAFAC into spectral, temporal, spatial and the subject/condition (feature) factor matrices including R components in each one. This allows the multi-domain features along their spectral, spatial, and temporal signatures of ERP to be found and extracted. As we are interested in group-level analysis of the subjects, the extracted factor in the subject/condition mode would be the feature to consider. Indeed, each column of the subject factor is a multi-domain feature for group-level analysis that spans the same spectral, temporal and spatial components [144]. Generating ERP-tensor for group-level analysis assumes that every multi-domain feature in the ERP could be uniquely characterized and represented by a temporal, a spectral, and a spatial signature; and the variation of this feature in different subjects/conditions is reflected in the subject/condition signature.

6.1.3.1 CANDECOMP/PARAFAC model order selection

In order to capture the true underlying structure in data, it is important to extract the right number of components R in the CANDECOMP/PARAFAC model (6.1), which is the number of multi-domain features in ERP tensor analysis. Selecting the optimum

number of components in CANDECOMP/PARAFAC model is still an active challenge in the community. The existing ALS-based methods include DIFFIT, core consistency diagnostic (CORCONDIA), Convex-hull based approach, and our proposed method based on Reconstruction Error [148] that was explained in Chapter 4. Most of them are based on the so called model fit or Data Error in (6.4). The available Data Error is a monotonically decreasing function of the model order, which makes it inapplicable to rank selection. In this study, we use the Reconstruction Error based technique developed by the authors for selecting the optimum model order in CANDECOMP/PARAFAC. The Reconstruction Error (6.5), which is the error between the reconstructed data and the unavailable noiseless data, intrinsically reveals the true rank of the model at its minimum and has a superior performance compared to other state-of-the-art methods specially in low SNR scenarios [148].

$$\gamma_R = \frac{1}{M} \|\underline{\mathbf{X}} - \underline{\mathbf{I}}_{4,R} \times_1 \hat{\mathbf{A}} \times_2 \hat{\mathbf{B}} \times_3 \hat{\mathbf{C}} \times_4 \hat{\mathbf{D}}\|_F^2. \quad (6.5)$$

Figure 6.4 illustrates the behaviour of the Data Error and Reconstruction Error for CANDECOMP/PARAFAC model order selection of the four-way ERP tensor.

6.1.4 Proposed Multi-domain Feature Selection Scheme

According to (6.1), the number of multi-domain features extracted using CANDECOMP/PARAFAC analysis is R , which would be in the order of 50–70 features for ERP-tensors. Visual inspection of the extracted features is usually done to select the desired one that matches with our prior knowledge about the MMN characteristics. Instead, here we propose a template matching approach which is novel in this context. Two concerns should be addressed when it comes to a proper multi-domain feature selection for group-level analysis. First, all the temporal, spectral and spatial information of MMN can be used as a prior knowledge, to reduce the number of features to the ones that adhere to the MMN characteristics. Second, as we want to obtain a multi-domain feature of

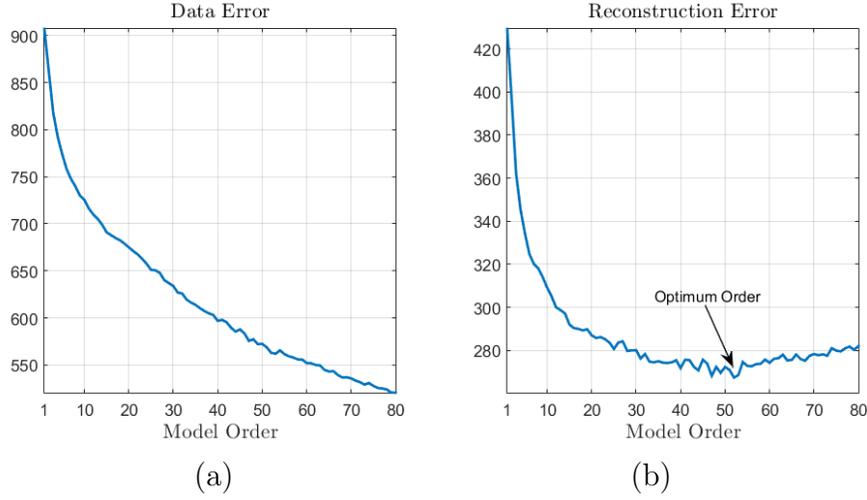


Figure 6.4: Illustration of two types of error after CANDECOMP/PARAFAC analysis of four-way ERP tensor of *time* \times *frequency* \times *channel* \times *subject/condition* for a range of possible model orders: (a) Data Error; (b) Reconstruction Error.

MMN that can better discriminate the two groups of subjects/conditions, consequently, from the short-listed set of multi-domain features, the one that shows a significant level of difference, proved by some statistical tests, between two group of subjects/conditions would be selected.

6.1.4.1 Template Matching Method

ERPs are time-locked, so the temporal signature of MMN could be used in a template matching approach to find the desired multi-domain feature. Moreover, the spectrum of MMN is in a range of 2 to 8.5 Hz and the peak of the spectrum is below 5 Hz. In order to use the temporal and spectral signatures of MMN for template matching, we need to define a similarity or distance metric which measures the distance between the associated signature of decomposed multi-domain feature and the template. We can use the normalized cross-correlation coefficient, the angle between the two vectors as the similarity measure. In our experiment, we use the angle between the template and the extracted signatures as the distance measure. Suppose that the template is \mathbf{a}^{temp} , the extracted column of the factor matrix being $\{\mathbf{a}_r, r = 1, \dots, R\}$, R being the number of

multi-domain features, we calculate the angle of \mathbf{a}_r with the template \mathbf{a}^{temp} :

$$c_r = \cos^{-1} \left(\frac{\text{abs}(\mathbf{a}_r^T \mathbf{a}^{temp})}{\|\mathbf{a}_r\| \cdot \|\mathbf{a}^{temp}\|} \right), \quad r = 1, \dots, R \quad (6.6)$$

where the superscript ' T ' denotes transpose and $\|\cdot\|$ denotes the vector norm. Here we use the absolute value $\text{abs}(\mathbf{a}_r^T \mathbf{a}^{temp})$ to cope with the scale ambiguity problem in CANDECOMP/ PARAFAC decomposition. Then we calculate the minimum distance

$$c_i = \min(c_r), \quad r = 1, \dots, R \quad (6.7)$$

If c_i is less than a predefined threshold ξ for both temporal and spectral signature, we can select the i th subject/condition component as a potential multi-domain feature.

6.1.4.2 Discriminability of the Multi-domain Feature

The significance of the group difference in the multi-domain feature of MMN could be evaluated by a statistical analysis. ANalysis Of VAriance (ANOVA) test was conducted to select the multi-domain feature that yields a significant difference level between the two group of subjects/conditions by having the lower p-value. Indeed, if the means of multi-domain feature for one group/condition is μ_1 and for the other group/condition is μ_2 , then the Null hypothesis H_0 and the Alternative hypothesis H_1 can be formulated as following

$$\begin{aligned} H_0 &: \mu_1 = \mu_2 \text{ or } \mu_1 - \mu_2 = 0, \\ H_1 &: \mu_1 \neq \mu_2 \text{ or } \mu_1 - \mu_2 \neq 0. \end{aligned} \quad (6.8)$$

Hence, the parameter of interest is the difference in means $\mu_1 - \mu_2$, and the test statistics, assuming an equal variance for both groups/conditions, can be written as [149]

$$t_0 = \frac{\bar{\mu}_1 - \bar{\mu}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (6.9)$$

where $s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$ and $\bar{\mu}_1, \bar{\mu}_2$ are sample means, s_1^2, s_2^2 are sample variances, and n_1 and n_2 are the samples sizes of the two groups/conditions considered in the t-test. We reject the Null hypothesis if $t_0 < -t_{\alpha/2, n_1+n_2-2}$ or $t_0 > t_{\alpha/2, n_1+n_2-2}$, where the $t_{\alpha, n}$ is the t-distribution with n degrees of freedom and significance level α which is usually set in practice as $\alpha = 0.05$. Hence, if we reject the Null hypothesis and go in favor of Alternative hypothesis, that means there is enough evidence that the two groups/conditions are statistically different.

6.2 Results

After preprocessing and synchronized averaging of the segmented multichannel EEG signals, the final 81-channel difference MMN signals are obtained by subtracting the standard trials from the deviant trials. The resulting 81-channel evoked potentials have 155 samples in time domain. Transforming each of the channels into TF plane, via the proposed technique [147] and retaining only the 0 – 20 Hz band with the resolution of 0.25 Hz, would yield 80 frequency bins in a 80×150 TF image. We used the window length of 127 and 1024 frequency bins in RM calculations. Besides, considering all the 81-channels and the 21 subjects, each of them with two conditions of deviance occurrence rate of 10% and 20%, we have 81 channels and $(22 \times 2) = 44$ subjects/conditions. Hence, the four-way ERP tensor of $\underline{\mathbf{Y}} \in \mathbb{R}_+^{80 \times 155 \times 81 \times 44}$ was generated which is a non-negative tensor due to the TF transformations.

CANDECOMP/PARAFAC decomposition of the ERP tensor for a range of possible

orders was obtained and the factor matrices were estimated with the following configurations: convergence criterion set as $10e^{-6}$, non-negativity constraint on the modes, line-search acceleration scheme initialized, and using direct trilinear decomposition for initialization.

The maximum possible order was set to $R_{max} = 80$ and the optimum order was selected using the Reconstruction Error based criterion [148] using $\alpha = 100$ and $\beta = 1900000$, which turned out to be $R = 52$ (Figure 6.4). Hence, we have a spectral factor $\mathbf{A} \in \mathbb{R}_+^{80 \times 52}$, a temporal factor $\mathbf{B} \in \mathbb{R}_+^{155 \times 52}$, a spatial factor $\mathbf{C} \in \mathbb{R}_+^{81 \times 52}$, and a subject/condition factor $\mathbf{D} \in \mathbb{R}_+^{44 \times 52}$. It's worth noting that there is no rule of thumb when it comes to the optimum CANDECOMP/PARAFAC model order selection and usually a range of possible orders would be suggested for practical applications.

The extracted 52 multi-domain features are then evaluated and the angle between their corresponding temporal and spectral factors and the MMN's temporal and spectral templates are calculated via (6.6). Fig. 6.5(a) shows the temporal and spectral templates used for calculating the angles, that aligns with our prior knowledge about the MMN. We used a Gabor function to generate the 155 and 80-sample vectors as the temporal and spectral templates, respectively. The width and center of each Gabor function was set according to the MMN characteristics. Fig. 6.5(b) depicts the histogram of resulting angles for the extracted features. The histogram shows that we can discriminate the desired multi-domain features from the undesired ones by measuring their angles with the given templates. Most of the angles are above 1 radian and only a few multi-domain features are close to the desired templates. By setting a threshold around $\xi = 0.5$ radian, all the multi-domain features that their angles are higher than the threshold are discarded. The remaining ones are the desired multi-domain features that matched with the MMN templates. The subject/condition signature of the remaining multi-domain features are tested for group-level statistical significance. 2 multi-domain features of #21 and #49 matched the MMN temporal and spectral templates, and we run the t-test on their two groups of subject/condition signatures and got the p-value = 0.001, 0.4, respectively,

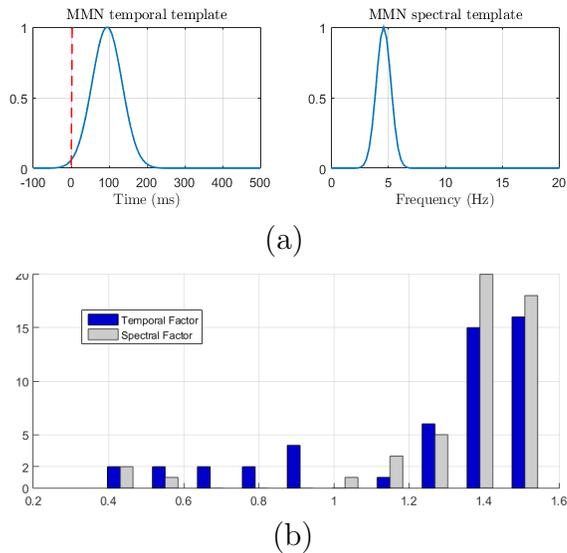


Figure 6.5: (a) Temporal and spectral signatures used in the template matching approach with temporal component peaked at about 100 ms and the spectrum peaked around 5 Hz. (b) Histogram of the calculated angles for both temporal and spectral signatures of all extracted multi-domain features.

which indicates that there is only a significant difference and a main effect in MMN multi-domain feature #21 for the two rates of deviant occurrence between subjects.

PARAFAC analysis of ERP-tensor can effectively localize and separate the ongoing activities in the brain even if they are temporally or spectrally overlapped. Each of the 52 extracted components in PARAFAC modeling, is made of 4 factors of temporal, spectral, spatial and subject/condition that are coupled together and corresponds to one of the embedded events in the ERP segment. Fig. 6.6 shows three multi-domain features (subject/condition signature) #21, #43 and #16 along their constituting temporal, spectral and spatial factors. These factors reveal the temporal, spectral, and spatial signatures of the multi-domain feature that are common for all the subjects. Besides, the strength and variation of the multi-domain feature itself, indicates the difference of the brain activity among the subjects or conditions. The selected multi-domain feature #21, with the significant difference between two conditions (p -value = 0.001), has the smallest angle according to (6.6) and simultaneously possess the properties of MMN in the time and frequency domains with its temporal component peaked at about 100 ms and the

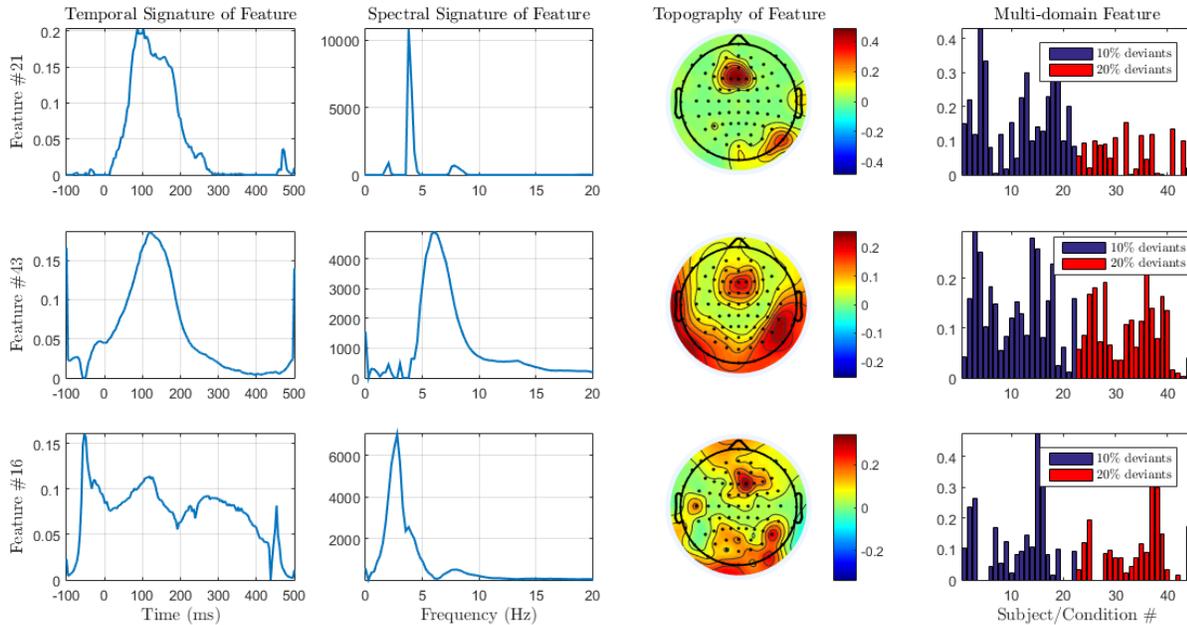


Figure 6.6: Illustration of three extracted multi-domain features along their temporal, spectral and spatial signatures. The selected feature #21 (top row) matches both temporal and spectral characteristics of the MMN and also differentiate the subjects significantly (p -value < 0.05) for the two experimental conditions. The other two multi-domain features #43, #16 (middle and bottom row) are samples from the total of 52 features that do not satisfy the selection criteria.

spectrum peaked around 5 Hz. The spatial component of the feature indicates that the difference in MMN between two deviants appeared in the frontal area. The two other depicted multi-domain features either do not statistically satisfy the t-test (#43) or the template matching criteria (#16).

6.2.1 Decomposition of Multiple Models

In order to verify the consistency of the results, for each of the model orders, the PARAFAC decomposition was run 10 times and the results of each decomposition analyzed with the template signatures to select the most similar one to the templates. For each run the minimum angle was determined and the statistics of the angles for 10 run were obtained. Figure 6.7(a) shows the maximal, averaged and minimal of the sum of the

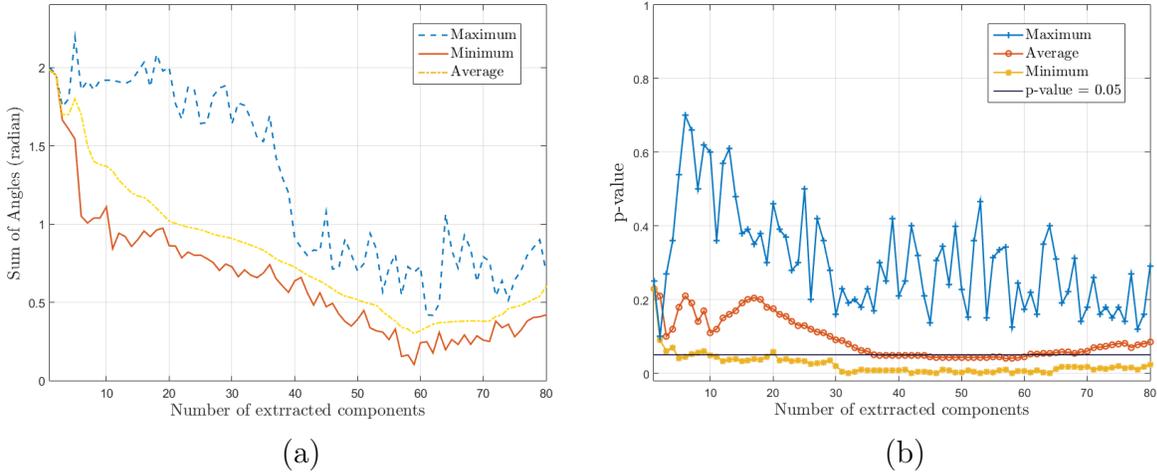


Figure 6.7: CANDECOMP/PARAFAC analysis of four-way ERP tensor of *time* \times *frequency* \times *channel* \times *subject/condition* for a range of possible model orders among 10 runs under each model: (a) The maximal, averaged and minimal sum of angles of the temporal and spectral signatures for the selected multi-domain feature; (b) The maximal, averaged and minimal p-value of the p-values for the selected features (p-value is to reveal the degree of the difference between two experimental conditions in the selected multi-domain feature in each run of each model).

temporal and spectral signatures' angles for the selected multi-domain feature (according to (6.6)) among 10 runs under each model. The figure shows that the minimal angle and the averaged angle are very close to each other when the model order is larger than 40, which indicates that the selected feature in this range is similar to the MMN template. This is interesting, as based on our model order selection criterion, the optimum model order is $R = 52$, however we are still able to find a feature similar to the MMN templates for any model order between 40 – 70 which supports why in practice we usually specify a range of possible orders.

Besides, figure 6.7(b) shows the maximal, averaged and minimal p-value of the p-values for the selected multi-domain features (according to (6.6)) among 10 runs under each model. We have two experimental conditions in the EEG experiment and we compare the mean of selected multi-domain features for the two experimental conditions. P-value reflects the degree of the difference between two experimental conditions in the selected multi-domain feature in each run of each model. If p-value is lower that a

threshold, then there is enough evidence that the mean of multi-domain feature for two experimental conditions are significantly different. In particular, when the number of components is larger than 40, the minimal p-value and the averaged p-value are very close to each other and both are smaller than the conventional threshold ($p = 0.05$). Hence, we can tell that when the number of components was larger than 40, the difference between two groups of subjects in the selected multi-domain feature of MMN was significant and there was a main effect in most runs.

Furthermore, figure 6.7(b) shows that even for some models with a smaller number of components than 40, p-value is smaller than the threshold and subsequently there is a main effect, which suggest that the mean of the selected multi-domain feature is significantly different between two experimental conditions. However, Figure 6.7(a) demonstrates that the corresponding rank-1 tensor associated with the multi-domain feature is not closely correlated with the MMN template signatures (the sum of angles is a large number). This means the selected multi-domain feature is representing other brain activities, rather than MMN.

The results show that although the conventional peak analysis for ERPs might not be discriminative enough for group analysis, the multi-way analysis framework help us to extract more discriminative features. Due to presence of different source of noise and interference in EEG recordings, EEG datasets might be heterogeneous in time domain and the peak amplitude could be easily contaminated. However, we can reduce the heterogeneous effect by mapping the EEG dataset to multi-way domain as the multi-domain pattern of desired brain activity can not simply be contaminated. As a result, the multi-domain feature can outperform the feature in one or two domains in discriminating different groups/conditions.

6.3 Chapter Summary

In this Chapter a group-level analysis of EEG signals via tensor decomposition by exploiting the mode of subjects/conditions in the analysis presented. Adding the mode of subject enables us to compare different group of subjects in the experiments. With CANDECOMP/PARAFAC analysis of 4-way ERP tensor of *time* \times *frequency* \times *channel* \times *subject/condition* of MisMatch Negativity (MMN), a multi-domain feature is extracted that is more discriminative than the peak amplitude of MMN for group-level analysis. In addition, in this Chapter we developed a methodology to select the desired feature. We showed that through a template matching approach the desired multi-domain feature could be selected which effectively discriminate the subjects for the two experimental conditions with an acceptable significance level. Furthermore, the temporal and spectral signatures of the selected multi-domain feature of MMN do match the temporal, spectral characteristics of the MMN in our study.

Chapter 7

Conclusions and Future Work

In this dissertation, we presented a robust signal processing framework for efficiently analyzing multi-channel EEG signals. The proposed method was evaluated using synthetic and real world signals in different stages, and desirable results were achieved. Figure 7.1 displays the contribution flowchart as evolved throughout this dissertation.

Chapter 1 presented a detailed introduction on EEG signal and conventional analysis techniques, including temporal, spectral, time-frequency, and Spatio-temporal analysis. In Chapter 2, popular time-frequency representation techniques that can be exploited to capture the non-stationarity of EEG signals and their shortcomings were reviewed. In addition in this Chapter the RS-EMD method, a fully data driven and cross-term free TFR with high-resolution in temporal and spectral domains, was explored and evaluated for characterizing the EEG properties. Knowing that two-way analysis may not capture the underlying content of the EEG signals, in Chapter 3 multi-way arrays and the corresponding analysis models were introduced that by retaining the structure of the high-dimensional EEG data tries to summarize the information into physiologically meaningful factors. In this chapter, we studied CANDECOMP/PARAFAC model that decomposes a tensor into a summation of rank-one tensors. One of the key parameters in this model is the number of components. Various ALS-based approaches have been proposed in the literature to extract the true rank of the observed tensor, most of them based

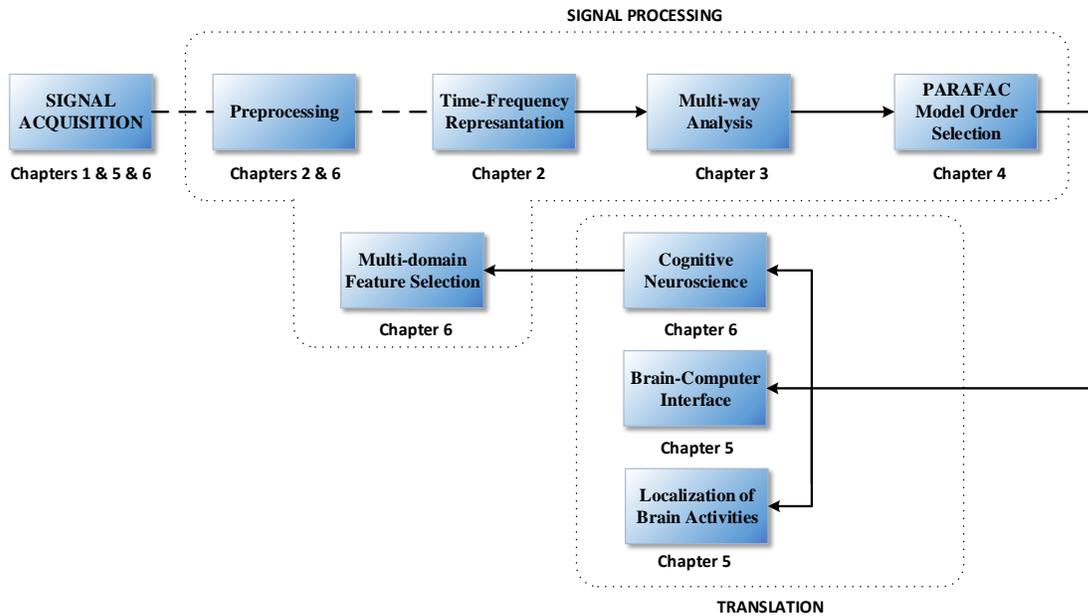


Figure 7.1: Flowchart of the proposed contributions.

on the so called model fit or Data error. Although, the available Data error is informative, it partially models the noise and known to be a monotonically decreasing function of the model order, that makes it inapplicable to rank selection. In Chapter 4, we proposed using the Data error to approximate the unknown Reconstruction error. The Reconstruction error, as the rank selection criterion, intrinsically reveals the true rank of the observed tensor. The results show the advantage of the proposed method in both robustness and accuracy, in terms of PoD, over the conventional CANDECOMP/PARAFAC model order selection methods.

The brain data collected by EEG techniques can be naturally fit into a multi-way array including multiple modes. In Chapters 5 and 6, we consider two classes of applications for multi-modal ERP tensor analysis, whether the tensor is for individual-level analysis or for group-level analysis and considered three different EEG datasets for those applications. Regarding the former analysis, the tensor usually includes the data of one segment (or called as epoch, trial) of one subject in an EEG experiment. As for the latter analysis, the tensor tends to contain the data of multiple trials of one subject

or multiple subjects. In Chapter 5, individual-level analysis are explored with applications in source localization of brain activity, and brain-computer interface. Various oscillatory activities within EEG may overlap, However, tensor decomposition by means of time-frequency-space transformation makes it possible in many cases to isolate each oscillatory behaviour well, even when these activities are not well-separated in two-way domain. Temporal, spatial and spectral information of the multichannel EEG are used here to generate a three-way EEG tensors. Reassigned-EMD with a high-resolution TF representation characteristic, is exploited for EEG tensor formation. Besides, the clinically meaningful extracted signatures are used in a Brain-Computer Interface system for detecting and identifying the desired event.

In Chapter 6, a group-level analysis of four-way EEG data with an application in research questions of cognitive neuroscience is explored. Adding the mode of subjects enables us to compare different group of subjects in the experiments. With CANDECOMP/PARAFAC analysis of 4-way ERP tensor of $time \times frequency \times channel \times subject/condition$ of MisMatch Negativity (MMN), a multi-domain feature is extracted that is more discriminative than the peak amplitude of MMN for group-level analysis. In addition, in this Chapter we developed a methodology to select the desired multi-domain feature. We showed that through a template matching approach the desired multi-domain feature could be selected that not only could effectively discriminate the subjects for the two experimental conditions with an acceptable significance level, but also the temporal and spectral signatures of the selected multi-domain feature of MMN do match the temporal, spectral characteristics of the MMN in our study that is another advantage of the proposed methodology in characterizing the ERP signals.

7.1 Outcome of the proposed work

Table 7.1, summarizes the various solutions provided by the proposed data-driven signal processing framework in efficiently analyzing non-stationary EEG signals and localizing

Table 7.1: Summary of the proposed solutions and the requirement for efficient EEG signal analysis

Requirements for efficient EEG signal analysis and feature extraction	Solution provided/ suggested by the proposed work	Chapter Reference
EEG characteristics	Capture non-stationarity with TFR	Chapter 1
TF analysis	Data-driven and high resolution TFR	Chapter 2
Multi-way nature of EEG	Multi-way Analysis	Chapter 3
Capture the true underlying structure of EEG signals	CANDECOMP/PARAFAC Model Order Selection	Chapter 4
EEG tensor formation	Multi-channel Reassigned-EMD	Chapter 5
Localization of Brain Activities	Three-way EEG tensor analysis	Chapter 5
Brain-Computer Interface	PARAFAC-based feature extraction	Chapter 5
Multi-domain feature selection	Four-way EEG tensor analysis	Chapter 6

the events and extracting discriminative features from them. The proposed framework with multi-way analysis as its significant highlight is expected to become a versatile non-stationary signal analysis tool, which has the benefits of localization and discriminant analyses. The outcome of the proposed work as an EEG processing system, including the data acquisition, signal processing, and the translation stages could be summarized as following:

Our main contribution in signal processing stage focuses attention on developing a framework for multi-way discriminative analysis of EEG signals. Analysis of ERPs, as one of the main categories of EEG signals, requires systems that can exploit the variation of the signals in different contextual domains. Multi-way analysis framework enables us to consider as much as relevant contextual information in our analysis, without ruining the natural structure of the data, in order to reveal the hidden multi-linear structures in the data. To fulfill this objective, in the first point, a fully data-driven time-frequency representation methodology is selected to capture the non-stationarity in the EEG signals which is robust to noise and artifacts. TFR is considered as a pre-processing step in this systems approach to help with better localization/analysis of events. Having a coarser TFR is okay in some cases but finer TFR helps with time-locked ERPs that require

higher resolution.

Time-frequency Representation

Reassigned-EMD, a fully data-driven TFR technique is proposed that satisfies the general TFR requirements. It has a high-resolution in both temporal and spectral domains and also a cross-term free TFR due to its inherited properties from EMD that decomposes the EEG into mono-component factors. Moreover, EMD can be effectively used for capturing the non-linearity in the signal and detrending the EEG signals that is simply embedded in the proposed EEG tensor formation.

In the second point, by adding other contextual modes, i.e. spatial and subject/condition mode to the analysis, we form the high-dimensional ERP tensor in order to extract the meaningful and unique features from EEG tensor. Then, for an effective analysis of EEG data, an optimum multi-way model order selection scheme is developed.

CANDECOMP/PARAFAC Model Order Selection

CANDECOMP/PARAFAC decomposes a tensor into a summation of rank-one tensors. The ALS algorithm used for this decomposition is attractive for its simplicity and satisfactory performance. Estimating the optimum model order is a crucial part of this analysis. Various ALS-based approaches have been proposed in the literature to extract the true rank of the observed tensor, most of them based on the so called model fit or Data error. Although, the available Data error is informative, it partially models the noise and known to be a monotonically decreasing function of the model order, that makes it inapplicable to rank selection. We developed a novel methodology for CANDECOMP/PARAFAC model order selection, that propose using the Data error to approximate the unknown Reconstruction error. The Reconstruction error, as the rank selection criterion, intrinsically reveals the true rank of the observed tensor.

Finally, in the translation part of the system, several applications are employed to evaluate the proposed work for multi-modal ERP analysis. These application are divided

into individual-level and group-level analysis and are as follows:

Individual-level Analysis

In this analysis the tensor usually includes the data of one segment (or called as epoch, trial) of one subject in an EEG experiment. Two applications are considered in this context. First application is for localization of brain events. Various oscillatory activities within EEG may overlap, However, tensor decomposition by means of Time-Frequency-Space transformation makes it possible in many cases to isolate each oscillatory behaviour well, even when these activities are not well-separated in two-way domain. Temporal, spatial and spectral information of the multichannel EEG are used here to generate a three-way EEG tensor. Reassigned-EMD, with a high-resolution TF representation characteristic, is exploited here for EEG tensor generation. Simulation results on synthetic and real three-way EEG data show that the proposed method could be effectively used for separation and localization of events in EEG. Besides, the clinically meaningful extracted signatures could be used for various purposes. In the second application, an SSVEP-BCI system was designed that uses the extracted signatures via CANDECOMP/PARAFAC in order to detect SSVEP signals using a canonical correlation analysis approach. The results are quite promising and proves the virtue of the proposed technique for EEG analysis.

Group-level Analysis

In this analysis the tensor tends to contain the data of multiple trials of one subject or multiple subjects. A group-level analysis of four-way EEG data with an application in research questions of cognitive neuroscience is explored. Adding the mode of subject enables us to compare different group of subjects in the experiments. With CANDECOMP/PARAFAC analysis of 4-way ERP tensor of $time \times frequency \times channel \times subject/condition$ of MisMatch Negativity (MMN), a multi-domain feature is extracted that is more discriminative than the peak amplitude of MMN for group-level analysis.

In addition a methodology to select the desired feature is developed. We showed that through a template matching approach the desired multi-domain feature could be selected which effectively discriminate the subjects for the two experimental conditions with an acceptable significance level. Furthermore, the temporal and spectral signatures of the selected multi-domain feature of MMN do match the temporal, spectral characteristics of the MMN in our study.

7.2 Limitations and Future Work

While the ERP datasets used in this work were almost limited in terms of number of subjects participated in the experiments, the analysis results might not reflect a general outcome. Hence, accessing to more subjects can help us make less biased and more statistically significant conclusions about the data.

The following could be the directions for future work in applying and enhancing the proposed work with more intelligence and accuracy.

- Bayesian Inference has proved its effectiveness in model selection and parameter estimation in various applications. Using the framework of Bayesian Learning it is also possible to consider correlation on factor matrices (i.e. temporal or spatial domains) in CANDECOMP/PARAFAC model for EEG analysis. However, considering the correlation would make the problem more complex and challenging to solve.
- It is known that EEG sources are sparse in some domains. Exploiting sparsity in the constituting domains for EEG analysis, could be beneficiary in source localization and may result in more interpretable factor analysis.

Potential trend: Tensor decomposition for data of two modalities simultaneously

Recently, research of the simultaneous EEG and functional magnetic resonance imaging (fMRI) has become very attractive in order to overcome the inherent shortcoming of EEG (low spatial resolution) and fMRI (low temporal resolution). Therefore, how to process the data of the two modalities becomes a very significant research question. Currently, most of the data processing methods for the data of the two modalities are in terms of matrix decomposition. Nevertheless, when EEG data and fMRI data are represented by two tensors, the tensor decomposition methods for the data of two modalities can be applied. Indeed, as long as data of one modality is denoted by a tensor (data of the other mode can be represented by a vector, a matrix or a tensor) and they have a common mode in both dataset, the data of two modalities can be factorized by the coupled tensor decomposition methods.

Appendix

Derivation of the probabilistic upperbound in (4.30)

The Reconstruction error γ_R in (4.29) is the sum of squared Gaussians and therefore is a member of non-central Chi-squared distributions with the mean calculated as follows:

$$\mathbb{E}(\Gamma_R) = \frac{1}{M} \mathbb{E} \left(\|\underline{\mathbf{X}} - \hat{\underline{\mathbf{X}}}_R\|_{\text{F}}^2 \right). \quad (.1)$$

Assuming the true rank of the tensor to be within the range of $(1, R_{max})$, we can rewrite (2) as following

$$\begin{aligned} \underline{\mathbf{X}} &= \sum_{r=1}^R \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r + \sum_{r=1}^{\Delta_R} \lambda_r \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r, \\ &= \underline{\mathbf{I}}_{3,R} \times_1 \mathbf{A}^R \times_2 \mathbf{B}^R \times_3 \mathbf{C}^R + \underline{\mathbf{\Lambda}}_{3,\Delta_R} \times_1 \mathbf{A}^{\Delta_R} \times_2 \mathbf{B}^{\Delta_R} \times_3 \mathbf{C}^{\Delta_R}. \end{aligned} \quad (.2)$$

where R is the considered model order and $\Delta_R = R_{max} - R$ corresponds to the unmodeled components. $\underline{\mathbf{\Lambda}}_{3,\Delta_R}$ is a diagonal tensor with its diagonal elements λ_r , including $(R^* - R)$ ones and $(R_{max} - R^*)$ zeros.

By assuming i.i.d Gaussian noise and the observed noisy tensor in the form of $\underline{\mathbf{Y}} = \underline{\mathbf{X}} + \underline{\mathbf{E}}$, the estimate of an R -component CANDECOMP/PARAFAC model for a three-way tensor using ALS can be written as follows

$$\hat{\underline{\mathbf{X}}}_R = \sum_{r=1}^R \hat{\mathbf{a}}_r \circ \hat{\mathbf{b}}_r \circ \hat{\mathbf{c}}_r, \quad (.3)$$

where the factor matrices are estimated using the least squares. For example, for estimating matrix \mathbf{A} we have

$$\mathbf{Y}_{(1)} = \mathbf{A}(\mathbf{C} \odot \mathbf{B})^T + \mathbf{E}_{(1)} \quad (.4)$$

$$\hat{\mathbf{A}} = \mathbf{A}^R + \mathbf{E}_{(1)} [(\mathbf{C}^R \odot \mathbf{B}^R)^T]^\dagger \quad (.5)$$

where \odot is the Khatri Rao product and \dagger is the pseudo-inverse of a matrix. Therefore, we can write the estimate of the $\hat{\underline{\mathbf{X}}}_R$ in each sequence of ALS for \mathbf{A} as following

$$\begin{aligned} \hat{\underline{\mathbf{X}}}_R &= \mathbf{I}_{3,R} \times_1 \hat{\mathbf{A}} \times_2 \mathbf{B}^R \times_3 \mathbf{C}^R \\ &= \mathbf{I}_{3,R} \times_1 \left(\mathbf{A}^R + \mathbf{E}_{(1)} [(\mathbf{C}^R \odot \mathbf{B}^R)^T]^\dagger \right) \times_2 \mathbf{B}^R \times_3 \mathbf{C}^R \\ &= \mathbf{I}_{3,R} \times_1 \mathbf{A}^R \times_2 \mathbf{B}^R \times_3 \mathbf{C}^R + \mathbf{I}_{3,R} \times_1 \mathbf{E}_{(1)} [(\mathbf{C}^R \odot \mathbf{B}^R)^T]^\dagger \times_2 \mathbf{B}^R \times_3 \mathbf{C}^R. \end{aligned} \quad (.6)$$

Accordingly, using (.2) and (.6), we can write (.1) as

$$\begin{aligned} \mathbb{E}(\Gamma_R) &= \frac{1}{M} \mathbb{E} \left(\left\| \underline{\mathbf{A}}_{3,\Delta_R} \times_1 \mathbf{A}^{\Delta_R} \times_2 \mathbf{B}^{\Delta_R} \times_3 \mathbf{C}^{\Delta_R} \right. \right. \\ &\quad \left. \left. - \mathbf{I}_{3,R} \times_1 \mathbf{E}_{(1)} [(\mathbf{C}^R \odot \mathbf{B}^R)^T]^\dagger \times_2 \mathbf{B}^R \times_3 \mathbf{C}^R \right\|_F^2 \right). \end{aligned} \quad (.7)$$

Expanding the norm and knowing that the expectation of noise is zero, we can simplify and write (.7) as follows

$$\begin{aligned} \mathbb{E}(\Gamma_R) &= \frac{1}{M} \mathbb{E} \left(\left\| \underline{\mathbf{A}}_{3,\Delta_R} \times_1 \mathbf{A}^{\Delta_R} \times_2 \mathbf{B}^{\Delta_R} \times_3 \mathbf{C}^{\Delta_R} \right\|_F^2 \right) \\ &\quad + \frac{1}{M} \mathbb{E} \left(\left\| \mathbf{I}_{3,R} \times_1 \mathbf{E}_{(1)} [(\mathbf{C}^R \odot \mathbf{B}^R)^T]^\dagger \times_2 \mathbf{B}^R \times_3 \mathbf{C}^R \right\|_F^2 \right). \end{aligned} \quad (.8)$$

where the first term, which we call it Δ_R , is the norm of the unmodeled or discarded possible components and is independent of noise and a decreasing function of model order, whereas the second term is noise dependant and an increasing function of model

order R . So, we can write (.8) as

$$\mathbb{E}(\Gamma_R) = \frac{1}{M} \mathbb{E}(\|\Delta_R\|_F^2) + \frac{1}{M} \mathbb{E}\left(\left\|\sum_{r=1}^R \hat{\mathbf{e}}_{(1)r} \circ \mathbf{b}_r \circ \mathbf{c}_r\right\|_F^2\right). \quad (.9)$$

where $\hat{\mathbf{e}}_{(1)r} \in \mathbb{R}^{I \times 1}$ is r th column of the rotated matrix $\mathbf{E}_{(1)} \in \mathbb{R}^{I \times JK}$ with rotation $[(\mathbf{C}^R \odot \mathbf{B}^R)^T]^\dagger \in \mathbb{R}^{JK \times R}$.

Furthermore, every element of the second term in (.9) could be written as

$$\hat{e}_{ijk} = \sum_{r=1}^R \hat{e}_{(1)ir} b_{jr} c_{kr}, \quad (.10)$$

where according to the definition of Frobenius norm we can fix and factor out the noise dependent term $\hat{e}_{(1)ir}$, $i \in (1, I)$ from the summation leading to RI number of independent noise samples. Therefore, we can write (.9) as follows

$$\mathbb{E}(\Gamma_R) = \frac{1}{M} \|\Delta_R\|_F^2 + \mathbb{E}\left(\sum_{i=1}^{RI} |u_i|^2\right). \quad (.11)$$

where the u_i s are independent, zero mean, white Gaussian noises with variance σ_n^2/M . Hence, Reconstruction error γ_R is a sample of random variable χ_{RI}^2 and for this random variable we have

$$\frac{M}{\sigma_n^2} \left(\Gamma_R - \frac{1}{M} \|\Delta_R\|_F^2 \right) \sim \chi_{RI}^2 \quad (.12)$$

where Γ_R is a Chi-square random variable of order RI . The expected value and the variance of Γ_R are

$$\mathbb{E}(\Gamma_R) = \frac{RI}{M} \sigma_n^2 + \frac{1}{M} \|\Delta_R\|_F^2, \quad (.13)$$

$$\text{var}(\Gamma_R) = \frac{2RI}{M^2} (\sigma_n^2)^2. \quad (.14)$$

The MSE that is the expected value of Γ_R has two terms. The first term is the noise

dependent part and is a monotonically increasing function of R . The second term is a function of the unmodeled or discarded possible components Δ_R . The norm of the unmodeled components is a decreasing function of R [117].

Using the pdf of the random variable Γ_R , for a given confidence probability p_1 , there exists a D_R for which the reconstruction error γ_R is bounded as follows [117]

$$Pr\{|\gamma_R - E(\Gamma_R)| \leq D_R\} = p_1. \quad (.15)$$

The value of D_R is a function of p_1 and the variance of Γ_R , and can be calculated using the Chi-square CDF table. Therefore, with probability p_1 the reconstruction error is bounded with

$$\underline{\gamma_R(p_1)} \leq \gamma_R \leq \overline{\gamma_R(p_1)} \quad (.16)$$

where

$$\overline{\gamma_R(p_1)} = E(\Gamma_R) + D_R = \frac{1}{M} \|\Delta_R\|_F^2 + \frac{RI}{M} \sigma_n^2 + D_R. \quad (.17)$$

The terms of these values are functions of R , M , σ_n , and the confidence probability, except one term which is a function of the unmodeled components. The structure of the available Data error ϵ_R is such that it can be used to provide probabilistic bounds on $\|\Delta_R\|_F^2$. It is shown that the Data error is a sample of a Chi-square random variable \mathcal{E}_R [117]

$$\frac{M}{\sigma_n^2} (\mathcal{E}_R) \sim \chi_{M-RI}^2 \quad (.18)$$

where χ_{M-RI}^2 is a Chi-square random variable of order $M - RI$. The random variable \mathcal{E}_R similarly has the following expected value and variance [117]

$$E(\mathcal{E}_R) = \left(1 - \frac{RI}{M}\right) \sigma_n^2 + \frac{1}{M} \|\Delta_R\|_F^2 \quad (.19)$$

$$\text{var}(\mathcal{E}_R) = \frac{2}{M} \left(1 - \frac{RI}{M}\right) (\sigma_n^2)^2 + \frac{4\sigma_n^2}{M^2} \|\Delta_R\|_F^2 \quad (.20)$$

Given the noisy data ϵ_R , one sample of the random variable \mathcal{E}_R is available. The variance of this random variable is of order $\frac{1}{M}$ of its expected value. If the data length is long enough, the variance of this random variable is close to zero. In this case, one method of estimating $\|\Delta_R\|_F^2$ is to assume that the available sample ϵ_R from (4.28) is a good estimate of its expected value in (.19). Therefore, by assuming that $E(\mathcal{E}_R) \approx \epsilon_R$, it yields [117]:

$$\frac{1}{M}\|\Delta_R\|_F^2 \approx \epsilon_R - \left(1 - \frac{RI}{M}\right) \sigma_n^2. \quad (.21)$$

However, for a finite-length data, the validity of this estimation depends on the exact behaviour of the variance of \mathcal{E}_R which is completely ignored in this estimation. It is proved in [117] that using the observed data, with validation probability p_2 , we can provide probabilistic bounds on the unavailable unmodeled components effects,

$$L_R(\underline{\mathbf{Y}}, \sigma_n, p_2) \leq \frac{1}{M}\|\Delta_R\|_F^2 \leq U_R(\underline{\mathbf{Y}}, \sigma_n, p_2) \quad (.22)$$

where the lower and upper bounds are functions of only the observed data and the validation probability. With validation probability p_2 and confidence probability p_1 , the Reconstruction error is bounded as follows [117]:

$$\underline{\gamma_R(p_1, \underline{\mathbf{Y}}, p_2)} \leq \gamma_R \leq \overline{\gamma_R(p_1, \underline{\mathbf{Y}}, p_2)} \quad (.23)$$

where

$$\overline{\gamma_R(p_1, \underline{\mathbf{Y}}, p_2)} = U_R(\underline{\mathbf{Y}}, p_2) + \frac{RI}{M}\sigma_n^2 + D_R(p_1, \sigma_n, M) \quad (.24)$$

The bounds on the Reconstruction error are only functions of the Chi-square CDF table, σ_n^2 , R , M , probabilities p_1 and p_2 , and the observed data tensor $\underline{\mathbf{Y}}$.

If R is large enough, we can estimate the Chi-square distribution of Γ_R with a Gaussian distribution. In this case, there is no need to check the Chi-square CDF table and with $p_1 = Q(\beta)$, the probabilistic event in (.15) can be written in the form

$$\Pr \left\{ |\Gamma_R - \mathbb{E}(\Gamma_R)| \leq \beta \sqrt{\text{var}(\Gamma_R)} \right\} = Q(\beta) \quad (.25)$$

where $Q(\beta) = \int_{-\beta}^{\beta} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$. Estimating this random variable with a Gaussian implies that D_R in (.24) is simply

$$D_R(Q(\beta), \sigma_n, M) = \beta \sqrt{\text{var}(\Gamma_R)} = \beta \frac{\sqrt{2RI}}{M} \sigma_n^2. \quad (.26)$$

On the other hand, if $M - RI$ is large enough, we can estimate the Chi-square distribution of \mathcal{E}_R with a Gaussian distribution. In this case, there is no need to use the Chi-square CDF table to calculate the bounds of the unmodeled coefficients effect. The values of these bounds with validation probability $p_2 = Q(\alpha)$ are provided in Appendix D of [117] as

$$\Pr \left\{ |\mathcal{E}_R - \mathbb{E}(\mathcal{E}_R)| \leq \alpha \sqrt{\text{var}(\mathcal{E}_R)} \right\} = Q(\alpha). \quad (.27)$$

Using this to find the upperbound for $\|\Delta_R\|_2^2$ we should solve the following inequality:

$$\mathbb{E}(\mathcal{E}_R) - \alpha \sqrt{\text{var}(\mathcal{E}_R)} \leq \epsilon_R. \quad (.28)$$

This inequality provides the upperbound as long as p_2 , or equivalently, $Q(\alpha)$ has been chosen sufficiently large such that [117]

$$\alpha \geq \frac{M}{\sqrt{2(M - RI)}} \left(1 - \frac{RI}{M} - \frac{\epsilon_R}{\sigma_n^2} \right). \quad (.29)$$

Finally, the upperbound of γ_R in (.24) can be written as follows:

$$\overline{\gamma_R(Q(\beta), \underline{\mathbf{Y}}, Q(\alpha))} = \frac{RI}{M} \sigma_n^2 + U_R(\underline{\mathbf{Y}}, Q(\alpha)) + \beta \frac{\sqrt{2RI}}{M} \sigma_n^2, \quad (.30)$$

where σ_n^2 is the noise variance after prewhitening and R is the number of signal compo-

nents used in the CANDECOMP/PARAFAC model. $Q(\alpha)$ and $Q(\beta)$ are the validation probability and the confidence probability used to provide the probabilistic upper bound for γ_R . Furthermore, U_R is the upperbound corresponding to unmodeled or discarded possible components which could be obtained as [117]

$$U_R(\underline{\mathbf{Y}}, Q(\alpha)) = \epsilon_R - m_n + \frac{2\alpha^2\sigma_n^2}{M} + K_R(\alpha), \quad (.31)$$

where m_n and $K_R(\alpha)$ are defined as:

$$m_n = \left(1 - \frac{RI}{M}\right)\sigma_n^2, \quad (.32)$$

$$K_R(\alpha) = 2\alpha \frac{\sigma_n}{\sqrt{M}} \sqrt{\frac{\sigma_n^2}{M} + \epsilon_R - \frac{1}{2}m_n}. \quad (.33)$$

□

References

- [1] S. Sanei and J. A. Chambers, *EEG signal processing*. John Wiley & Sons, 2013.
- [2] J. Kropotov, *Quantitative EEG, event-related potentials and neurotherapy*. Academic Press, 2010.
- [3] P. L. Nunez and R. Srinivasan, *Electric fields of the brain: the neurophysics of EEG*. Oxford university press, 2nd ed., 2006.
- [4] E. F. Wijdicks, P. N. Varelas, G. S. Gronseth, and D. M. Greer, “Evidence-based guideline update: Determining brain death in adults Report of the Quality Standards Subcommittee of the American Academy of Neurology,” *Neurology*, vol. 74, no. 23, pp. 1911–1918, 2010.
- [5] J. Dauwels, F. Vialatte, and A. Cichocki, “Diagnosis of Alzheimer’s disease from EEG signals: where are we standing?,” *Current Alzheimer Research*, vol. 7, no. 6, pp. 487–505, 2010.
- [6] R. M. Benca, W. H. Obermeyer, C. L. Larson, B. Yun, I. Dolski, K. D. Kleist, S. M. Weber, and R. J. Davidson, “EEG alpha power and alpha power asymmetry in sleep and wakefulness,” *Psychophysiology*, vol. 36, no. 04, pp. 430–436, 1999.
- [7] M. Le Van Quyen, J. Martinerie, V. Navarro, P. Boon, M. D’Havé, C. Adam, B. Renault, F. Varela, and M. Baulac, “Anticipation of epileptic seizures from standard EEG recordings,” *The Lancet*, vol. 357, no. 9251, pp. 183–188, 2001.

- [8] I. J. Rampil, "A primer for EEG signal processing in anesthesia," *The Journal of the American Society of Anesthesiologists*, vol. 89, no. 4, pp. 980–1002, 1998.
- [9] W. T. Blume, "Drug effects on EEG," *Journal of Clinical Neurophysiology*, vol. 23, no. 4, pp. 306–311, 2006.
- [10] S. G. Mason, R. Bohringer, J. F. Borisoff, and G. E. Birch, "Real-time control of a video game with a direct brain–computer interface," *Journal of Clinical Neurophysiology*, vol. 21, no. 6, pp. 404–408, 2004.
- [11] B. Graimann, B. Z. Allison, and G. Pfurtscheller, *Brain-computer interfaces: Revolutionizing human-computer interaction*. Springer Science & Business Media, 2010.
- [12] B. J. Fisch and R. Spehlmann, *Fisch and Spehlmann's EEG primer: basic principles of digital and analog EEG*. Elsevier Health Sciences, 1999.
- [13] J. Malmivuo and R. Plonsey, *Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields*. Oxford university press, 1995.
- [14] M. Von Spreckelsen and B. Bromm, "Estimation of single-evoked cerebral potentials by means of parametric modeling and Kalman filtering," *Biomedical Engineering, IEEE Transactions on*, vol. 35, no. 9, pp. 691–700, 1988.
- [15] S. Beheshti, M. Dahleh, *et al.*, "A new information-theoretic approach to signal denoising and best basis selection," *Signal Processing, IEEE Transactions on*, vol. 53, no. 10, pp. 3613–3624, 2005.
- [16] T. Tsuji, O. Fukuda, H. Ichinobe, and M. Kaneko, "A log-linearized Gaussian mixture network and its application to EEG pattern classification," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 29, no. 1, pp. 60–72, 1999.

- [17] E. Bartnik and K. Blinowska, “Waveletsnew method of evoked potential analysis,” *Medical and Biological Engineering and Computing*, vol. 30, no. 1, pp. 125–126, 1992.
- [18] N. V. Thakor, G. Xin-Rong, S. Yi-Chun, and D. F. Hanley, “Multiresolution wavelet analysis of evoked potentials,” *Biomedical Engineering, IEEE Transactions on*, vol. 40, no. 11, pp. 1085–1094, 1993.
- [19] M. Akay, Y. M. Akay, P. Cheng, and H. H. Szeto, “Time-frequency analysis of the electrocortical activity during maturation using wavelet transform,” *Biological cybernetics*, vol. 71, no. 2, pp. 169–176, 1994.
- [20] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM journal on scientific computing*, vol. 20, no. 1, pp. 33–61, 1998.
- [21] G. M. Davis, S. G. Mallat, and Z. Zhang, “Adaptive time-frequency decompositions,” *Optical Engineering*, vol. 33, no. 7, pp. 2183–2191, 1994.
- [22] S. Makeig, A. J. Bell, T.-P. Jung, T. J. Sejnowski, *et al.*, “Independent component analysis of electroencephalographic data,” *Advances in neural information processing systems*, pp. 145–151, 1996.
- [23] T.-P. Jung, S. Makeig, C. Humphries, T.-W. Lee, M. J. Mckeown, V. Iragui, and T. J. Sejnowski, “Removing electroencephalographic artifacts by blind source separation,” *Psychophysiology*, vol. 37, no. 02, pp. 163–178, 2000.
- [24] Y. Li, S.-I. Amari, A. Cichocki, D. Ho, and S. Xie, “Underdetermined blind source separation based on sparse representation,” *Signal Processing, IEEE Transactions on*, vol. 54, pp. 423–437, Feb 2006.
- [25] A. Hyvärinen, P. Hoyer, and M. Inki, “Topographic independent component analysis,” *Neural computation*, vol. 13, no. 7, pp. 1527–1558, 2001.

- [26] P. J. Durka, “Time-frequency analyses of EEG,” *Praca doktorska. Uniwersytet Warszawski, Warszawa*, 1996.
- [27] B. Boashash, *Time-frequency signal analysis and processing: a comprehensive reference*. Academic Press, 2015.
- [28] S. Qian and D. Chen, “Joint time-frequency analysis,” *Signal Processing Magazine, IEEE*, vol. 16, no. 2, pp. 52–67, 1999.
- [29] S. Blanco, R. Q. Quiroga, O. Rosso, and S. Kochen, “Time-frequency analysis of electroencephalogram series,” *Physical Review E*, vol. 51, no. 3, p. 2624, 1995.
- [30] L. Cohen and T. E. Posch, “Positive time-frequency distribution functions,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 1, pp. 31–38, 1985.
- [31] S. G. Mallat and Z. Zhang, “Matching pursuits with time-frequency dictionaries,” *Signal Processing, IEEE Transactions on*, vol. 41, no. 12, pp. 3397–3415, 1993.
- [32] P. J. Durka, A. Matysiak, E. M. Montes, P. V. Sosa, and K. J. Blinowska, “Multichannel matching pursuit and EEG inverse solutions,” *Journal of neuroscience methods*, vol. 148, no. 1, pp. 49–59, 2005.
- [33] P. J. Durka, “Adaptive time-frequency parametrization of epileptic spikes,” *Physical Review E*, vol. 69, no. 5, p. 051914, 2004.
- [34] I. Tošić and P. Frossard, “Dictionary learning,” *Signal Processing Magazine, IEEE*, vol. 28, no. 2, pp. 27–38, 2011.
- [35] B. Hamner, R. Chavarriaga, and J. d. R. Millán, “Learning dictionaries of spatial and temporal EEG primitives for brain-computer interfaces,” in *Workshop on Structured Sparsity: Learning and Inference, ICML 2011*, no. EPFL-CONF-166740, 2011.

- [36] Q. Barthélemy, C. Gouy-Pailler, Y. Isaac, A. Souloumiac, A. Larue, and J. I. Mars, “Multivariate temporal dictionary learning for EEG,” *Journal of neuroscience methods*, vol. 215, no. 1, pp. 19–28, 2013.
- [37] F. Auger and P. Flandrin, “Improving the readability of time-frequency and time-scale representations by the reassignment method,” *Signal Processing, IEEE Transactions on*, vol. 43, no. 5, pp. 1068–1089, 1995.
- [38] S. Mallat, *A wavelet tour of signal processing*. Academic press, 1999.
- [39] S. Krishnan, R. M. Rangayyan, G. D. Bell, and C. B. Frank, “Adaptive time-frequency analysis of knee joint vibroarthrographic signals for noninvasive screening of articular cartilage pathology,” *Biomedical Engineering, IEEE Transactions on*, vol. 47, no. 6, pp. 773–783, 2000.
- [40] N. E. Huang, *Hilbert-Huang transform and its applications*, vol. 16. World Scientific, 2014.
- [41] A. Pigorini, A. G. Casali, S. Casarotto, F. Ferrarelli, G. Baselli, M. Mariotti, and M. Rosanova, “Time–frequency spectral analysis of TMS-evoked EEG oscillations by means of hilbert–huang transform,” *Journal of neuroscience methods*, vol. 198, pp. 236–245, 2011.
- [42] W. Liu and X. Guo, “Detection of transient power quality disturbances based emd combined with choi-williams distribution,” in *2012 IEEE International Conference on Automation and Logistics*, pp. 588–591, IEEE, 2012.
- [43] P. Flandrin, F. Auger, E. Chassande-Mottin, *et al.*, “Time-frequency reassignment: from principles to algorithms,” *Applications in Time-Frequency Signal Processing*, vol. 5, pp. 179–203.
- [44] B. Mijovic, M. Silva, B. Van den Bergh, K. Allegaert, J.-M. Aerts, D. Berckmans, S. Van Huffel, *et al.*, “Assessment of pain expression in infant cry signals using

- empirical mode decomposition,” *Methods of information in medicine*, vol. 49, no. 5, pp. 448–452, 2010.
- [45] W. Hao, Z. Laibin, W. Zhaohui, D. Lixiang, and L. Wei, “Application of ISVD denoising and reassigned spectrogram in HHT time-frequency spectrum analysis of flue gas turbine signal [j],” *Chinese Journal of Scientific Instrument*, vol. 3, p. 030, 2009.
- [46] I. Daubechies and S. Maes, “A nonlinear squeezing of the continuous wavelet transform based on auditory nerve models,” *Wavelets in medicine and biology*, pp. 527–546, 1996.
- [47] I. Daubechies, J. Lu, and H.-T. Wu, “Synchrosqueezed wavelet transforms: an empirical mode decomposition-like tool,” *Applied and computational harmonic analysis*, vol. 30, no. 2, pp. 243–261, 2011.
- [48] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, “The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis,” in *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 454, pp. 903–995, The Royal Society, 1998.
- [49] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin, “A complete ensemble empirical mode decomposition with adaptive noise,” in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pp. 4144–4147, May 2011.
- [50] Y.-H. Wang, C.-H. Yeh, H.-W. V. Young, K. Hu, and M.-T. Lo, “On the computational complexity of the empirical mode decomposition algorithm,” *Physica A: Statistical Mechanics and its Applications*, vol. 400, pp. 159–167, 2014.

- [51] P. Flandrin, P. Goncalvs, and G. Rilling, “Detrending and denoising with empirical mode decompositions,” in *Signal Processing Conference, 2004 12th European*, pp. 1581–1584, Sept 2004.
- [52] F. Auger, P. Flandrin, Y.-T. Lin, S. McLaughlin, S. Meignen, T. Oberlin, and H.-T. Wu, “Time-frequency reassignment and synchrosqueezing: An overview,” *IEEE Signal Processing Magazine*, vol. 30, no. 6, pp. 32–41, 2013.
- [53] F. Miwakeichi, E. Martinez-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, no. 3, pp. 1035–1045, 2004.
- [54] M. Mørup, 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, no. 3, pp. 938–947, 2006.
- [55] E. Acar, C. Aykut-Bingol, H. Bingol, R. Bro, and B. Yener, “Multiway analysis of epilepsy tensors,” *Bioinformatics*, vol. 23, no. 13, pp. i10–i18, 2007.
- [56] I. Djurović and L. Stanković, “Time–frequency representation based on the re-assigned S-method,” *Signal Processing*, vol. 77, no. 1, pp. 115–120, 1999.
- [57] A. Ahrabian, D. Looney, L. Stanković, and D. P. Mandic, “Synchrosqueezing-based time-frequency analysis of multivariate data,” *Signal Processing*, vol. 106, pp. 331–341, 2015.
- [58] A. Cichocki and S.-i. Amari, *Adaptive blind signal and image processing: learning algorithms and applications*, vol. 1. John Wiley & Sons, 2002.
- [59] F. Estienne, N. Matthijs, D. Massart, P. Ricoux, and D. Leibovici, “Multi-way modelling of high-dimensionality electroencephalographic data,” *Chemometrics and Intelligent Laboratory Systems*, vol. 58, no. 1, pp. 59–72, 2001.

- [60] E. Acar and B. Yener, “Unsupervised multiway data analysis: A literature survey,” *Knowledge and Data Engineering, IEEE Transactions on*, vol. 21, no. 1, pp. 6–20, 2009.
- [61] A. Cichocki, R. Zdunek, A. H. Phan, and S.-i. Amari, *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*. John Wiley & Sons, 2009.
- [62] T. Kolda and B. Bader, “Tensor decompositions and applications,” *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.
- [63] H. A. Kiers, “Towards a standardized notation and terminology in multiway analysis,” *Journal of chemometrics*, vol. 14, no. 3, pp. 105–122, 2000.
- [64] L. Tucker, “Some mathematical notes on three-mode factor analysis,” *Psychometrika*, vol. 31, no. 3, pp. 279–311, 1966.
- [65] J. Carroll and J.-J. Chang, “Analysis of individual differences in multidimensional scaling via an n-way generalization of eckart-young decomposition,” *Psychometrika*, vol. 35, no. 3, pp. 283–319, 1970.
- [66] R. A. Harshman, “PARAFAC2: Mathematical and technical notes,” *UCLA Working Papers in Phonetics*, vol. 22, pp. 30–44, 1972b.
- [67] M. E. Timmerman and H. A. L. Kiers, “Three-mode principal components analysis: Choosing the numbers of components and sensitivity to local optima,” *British Journal of Mathematical and Statistical Psychology*, vol. 53, no. 1, pp. 1–16, 2000.
- [68] E. Ceulemans and H. A. L. Kiers, “Selecting among three-mode principal component models of different types and complexities: A numerical convex hull based method,” *British Journal of Mathematical and Statistical Psychology*, vol. 59, pp. 133–150, 2006.

- [69] L. D. Lathauwer, B. D. Moor, and J. Vandewalle, “A multilinear singular value decomposition,” *SIAM J. Matrix Anal. Appl.*, vol. 21, pp. 1253–1278, 2000.
- [70] P. Kroonenberg and J. Leeuw, “Principal component analysis of three-mode data by means of alternating least squares algorithms,” *Psychometrika*, vol. 45, no. 1, pp. 69–97, 1980.
- [71] R. A. Harshman, “Foundations of the parafac procedure: models and conditions for an” explanatory” multimodal factor analysis,” 1970.
- [72] E. Acar, D. M. Dunlavy, and T. G. Kolda, “A scalable optimization approach for fitting canonical tensor decompositions,” *Journal of Chemometrics*, vol. 25, no. 2, pp. 67–86, 2011.
- [73] P. Comon, X. Luciani, and A. L. De Almeida, “Tensor decompositions, alternating least squares and other tales,” *Journal of Chemometrics*, vol. 23, no. 7-8, pp. 393–405, 2009.
- [74] Y. Chen, D. Han, and L. Qi, “New ALS methods with extrapolating search directions and optimal step size for complex-valued tensor decompositions,” *Signal Processing, IEEE Transactions on*, vol. 59, no. 12, pp. 5888–5898, 2011.
- [75] A. Uschmajew, “Local convergence of the alternating least squares algorithm for canonical tensor approximation,” *SIAM Journal on Matrix Analysis and Applications*, vol. 33, no. 2, pp. 639–652, 2012.
- [76] A. Cichocki and P. Anh-Huy, “Fast local algorithms for large scale nonnegative matrix and tensor factorizations,” *IEICE transactions on fundamentals of electronics, communications and computer sciences*, vol. 92, no. 3, pp. 708–721, 2009.
- [77] N. K. M. Faber, R. Bro, and P. K. Hopke, “Recent developments in CANDECOMP/PARAFAC algorithms: a critical review,” *Chemometrics and Intelligent Laboratory Systems*, vol. 65, no. 1, pp. 119 – 137, 2003.

- [78] G. Zhou and A. Cichocki, “Canonical polyadic decomposition based on a single mode blind source separation,” *Signal Processing Letters, IEEE*, vol. 19, no. 8, pp. 523–526, 2012.
- [79] G. H. Golub and C. F. Van Loan, *Matrix computations*, vol. 3. JHU Press, 2012.
- [80] S. Van Huffel, J. Vandewalle, and A. Haegemans, “An efficient and reliable algorithm for computing the singular subspace of a matrix, associated with its smallest singular values,” *Journal of computational and applied mathematics*, vol. 19, no. 3, pp. 313–330, 1987.
- [81] M. Rajih, P. Comon, and R. A. Harshman, “Enhanced line search: A novel method to accelerate PARAFAC,” *SIAM journal on matrix analysis and applications*, vol. 30, no. 3, pp. 1128–1147, 2008.
- [82] C. M. Andersen and R. Bro, “Practical aspects of PARAFAC modeling of fluorescence excitation-emission data,” *Journal of Chemometrics*, vol. 17, no. 4, pp. 200–215, 2003.
- [83] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, “Parallel factor analysis in sensor array processing,” *Signal Processing, IEEE Transactions on*, vol. 48, no. 8, pp. 2377–2388, 2000.
- [84] L. De Lathauwer and J. Castaing, “Tensor-based techniques for the blind separation of DS-CDMA signals,” *Signal Processing*, vol. 87, no. 2, pp. 322–336, 2007.
- [85] A. H. Andersen and W. S. Rayens, “Structure-seeking multilinear methods for the analysis of fMRI data,” *NeuroImage*, vol. 22, no. 2, pp. 728–739, 2004.
- [86] E. Acar, S. A. Camtepe, and B. Yener, “Collective sampling and analysis of high order tensors for chatroom communications,” in *Intelligence and security informatics*, pp. 213–224, Springer, 2006.

- [87] B. W. Bader, M. W. Berry, and M. Browne, “Discussion tracking in enron email using PARAFAC,” in *Survey of Text Mining II*, pp. 147–163, Springer, 2008.
- [88] Z. He, A. Cichocki, S. Xie, and K. Choi, “Detecting the number of clusters in n-way probabilistic clustering,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 11, pp. 2006–2021, 2010.
- [89] R. Bro and H. A. Kiers, “A new efficient method for determining the number of components in PARAFAC models,” *Journal of Chemometrics*, vol. 17, no. 5, pp. 274–286, 2003.
- [90] M. Mørup and L. K. Hansen, “Automatic relevance determination for multi-way models,” *Journal of Chemometrics*, vol. 23, no. 7-8, pp. 352–363, 2009.
- [91] K. Liu, H. C. So, J. P. C. L. da Costa, F. Roemer, and L. Huang, “Efficient source enumeration for accurate direction-of-arrival estimation in threshold region,” *Digital Signal Processing*, vol. 23, no. 5, pp. 1668 – 1677, 2013.
- [92] J. P. C. L. Da Costa, F. Roemer, and M. Haardt, “Deterministic prewhitening to improve subspace based parameter estimation techniques in severely colored noise environments,” *Proc. 54th Int. Scientif. Coll.(IWK) Ilmenau, Ilmenau, Germany*, 2009.
- [93] J. P. C. L. Da Costa, K. Liu, H. C. So, S. Schwarz, M. Haardt, and F. Römer, “Multidimensional prewhitening for enhanced signal reconstruction and parameter estimation in colored noise with Kronecker correlation structure,” *Signal Processing*, vol. 93, no. 11, pp. 3209–3226, 2013.
- [94] P. Wirfalt and M. Jansson, “On Toeplitz and Kronecker structured covariance matrix estimation,” in *Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pp. 185–188, IEEE, 2010.

- [95] K. Greenewald, T. Tsiligkaridis, and A. O. Hero, “Kronecker sum decompositions of space-time data,” in *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 5th International Workshop on*, pp. 65–68, IEEE, 2013.
- [96] T. Tsiligkaridis and A. O. Hero, “Covariance estimation in high dimensions via Kronecker product expansions,” *Signal Processing, IEEE Transactions on*, vol. 61, pp. 5347–5360, 2013.
- [97] H. M. Huizenga, J. C. De Munck, L. J. Waldorp, and R. P. Grasman, “Spatiotemporal EEG/MEG source analysis based on a parametric noise covariance model,” *Biomedical Engineering, IEEE Transactions on*, vol. 49, no. 6, pp. 533–539, 2002.
- [98] L. Beltrachini, N. von Ellenrieder, and C. Muravchik, “Shrinkage approach for spatiotemporal EEG covariance matrix estimation,” *Signal Processing, IEEE Transactions on*, vol. 61, pp. 1797–1808, April 2013.
- [99] F. Bijma, J. C. De Munck, and R. M. Heethaar, “The spatiotemporal MEG covariance matrix modeled as a sum of Kronecker products,” *NeuroImage*, vol. 27, no. 2, pp. 402–415, 2005.
- [100] K. Werner, M. Jansson, and P. Stoica, “On estimation of covariance matrices with Kronecker product structure,” *Signal Processing, IEEE Transactions on*, vol. 56, no. 2, pp. 478–491, 2008.
- [101] J. Yin and H. Li, “Model selection and estimation in the matrix normal graphical model,” *Journal of multivariate analysis*, vol. 107, pp. 119–140, 2012.
- [102] J. P. C. L. da Costa, F. Roemer, and R. T. de Sousa Jr, “Enhanced model order estimation in colored noise scenarios via noise prewhitening,” in *XXIX Simpósio Brasileiro de Telecomunicações - (SBrT11), Curitiba, Brazil*, 2011.
- [103] J. P. C. L. da Costa, M. Haardt, and F. Romer, “Robust methods based on the HOSVD for estimating the model order in PARAFAC models,” in *Sensor Array and*

- Multichannel Signal Processing Workshop, 2008. SAM 2008. 5th IEEE*, pp. 510–514, IEEE, 2008.
- [104] K. Liu, H. C. So, J. P. C. L. da Costa, and H. Lei, “Core consistency diagnostic aided by reconstruction error for accurate enumeration of the number of components in PARAFAC models,” in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vancouver, 2013.
- [105] Q. Zhao, L. Zhang, and A. Cichocki, “Bayesian CP factorization of incomplete tensors with automatic rank determination,” *arXiv preprint arXiv:1401.6497*, 2014.
- [106] A.-H. Phan, P. Tichavsky, and A. Cichocki, “Low complexity damped Gauss–Newton algorithms for CANDECOMP/PARAFAC,” *SIAM Journal on Matrix Analysis and Applications*, vol. 34, no. 1, pp. 126–147, 2013.
- [107] A.-H. Phan, P. Tichavsky, and A. Cichocki, “Deflation method for CANDECOMP/PARAFAC tensor decomposition,” in *Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on*, pp. 6736–6740, 2014.
- [108] A. Cichocki, D. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. Caiafa, and H. Phan, “Tensor decompositions for signal processing applications: From two-way to multiway component analysis,” *Signal Processing Magazine, IEEE*, vol. 32, pp. 145–163, March 2015.
- [109] A.-H. Phan, P. Tichavsky, and A. Cichocki, “CANDECOMP/PARAFAC decomposition of high-order tensors through tensor reshaping,” *Signal Processing, IEEE Transactions on*, vol. 61, no. 19, pp. 4847–4860, 2013.
- [110] J. d. M. Goulart and G. Favier, “An algebraic solution for the CANDECOMP/PARAFAC decomposition with circulant factors,” *SIAM Journal on Matrix Analysis and Applications*, vol. 35, no. 4, pp. 1543–1562, 2014.

- [111] J. P. C. L. da Costa, F. Roemer, M. Haardt, and R. T. de Sousa Jr, “Multi-dimensional model order selection,” *EURASIP Journal on Advances in Signal Processing*, vol. 2011, no. 1, pp. 1–13, 2011.
- [112] A. Karfoul, L. Albera, and L. De Lathauwer, “Iterative methods for the canonical decomposition of multi-way arrays: Application to blind underdetermined mixture identification,” *Signal Processing*, vol. 91, no. 8, pp. 1789–1802, 2011.
- [113] F. Kiraly and A. Ziehe, “Approximate rank-detecting factorization of low-rank tensors,” in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vancouver, 2013.
- [114] K. Liu, H. C. So, and L. Huang, “A multi-dimensional model order selection criterion with improved identifiability,” in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pp. 2441–2444, IEEE, 2012.
- [115] A. Quinlan, J.-P. Barbot, P. Larzabal, and M. Haardt, “Model order selection for short data: An exponential fitting test (eft),” *EURASIP Journal on Advances in Signal Processing*, vol. 2007, no. 1, p. 071953, 2007.
- [116] M. Castaneda and J. Nossék, “Estimation of rank deficient covariance matrices with Kronecker structure,” in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pp. 394–398, May 2014.
- [117] S. Beheshti and M. A. Dahleh, “Noisy data and impulse response estimation,” *Signal Processing, IEEE Transactions on*, vol. 58, no. 2, pp. 510–521, 2010.
- [118] E. Radoi and A. Quinquis, “A new method for estimating the number of harmonic components in noise with application in high resolution radar,” *EURASIP Journal on Advances in Signal Processing*, vol. 2004, no. 8, pp. 1177–1188, 2004.
- [119] P. Tichavsky, A. Phan, and Z. Koldovsky, “Cramér-Rao-Induced Bounds for CAN-

- DECOMP/PARAFAC tensor decomposition,” *Signal Processing, IEEE Transactions on*, vol. 61, pp. 1986–1997, April 2013.
- [120] F. Cong, 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, 2015.
- [121] M. De Vos, 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, p. 11, 2007.
- [122] C. Torrence and G. P. Compo, “A practical guide to wavelet analysis,” *Bulletin of the American Meteorological society*, vol. 79, no. 1, pp. 61–78, 1998.
- [123] H. Becker, L. Albera, P. Comon, M. Haardt, G. Birot, F. Wendling, M. Gavaret, C.-G. Bénar, and I. Merlet, “EEG extended source localization: tensor-based vs. conventional methods,” *NeuroImage*, vol. 96, pp. 143–157, 2014.
- [124] A. Cichocki, Y. Washizawa, T. Rutkowski, H. Bakardjian, A.-H. Phan, S. Choi, H. Lee, Q. Zhao, L. Zhang, and Y. Li, “Noninvasive BCIs: Multiway signal-processing array decompositions,” *Computer*, no. 10, pp. 34–42, 2008.
- [125] K. Wendel, O. Väisänen, J. Malmivuo, N. G. Gencer, B. Vanrumste, P. Durka, R. Magjarevic, S. Supek, M. L. Pascu, H. Fontenelle, *et al.*, “EEG/MEG source imaging: methods, challenges, and open issues,” *Computational intelligence and neuroscience*, vol. 2009, p. 13, 2009.
- [126] M. A. Jatoi, N. Kamel, A. S. Malik, I. Faye, and T. Begum, “A survey of methods used for source localization using EEG signals,” *Biomedical Signal Processing and Control*, vol. 11, no. 0, pp. 42 – 52, 2014.

- [127] C. Plummer, A. S. Harvey, and M. Cook, “EEG source localization in focal epilepsy: where are we now?,” *Epilepsia*, vol. 49, no. 2, pp. 201–218, 2008.
- [128] S. Finnigan and M. J. van Putten, “EEG in ischaemic stroke: quantitative EEG can uniquely inform (sub-) acute prognoses and clinical management,” *Clinical neurophysiology*, vol. 124, pp. 10–19, 2013.
- [129] P. W. Kaplan and A. O. Rossetti, “EEG patterns and imaging correlations in encephalopathy: encephalopathy part ii,” *Journal of Clinical Neurophysiology*, vol. 28, no. 3, pp. 233–251, 2011.
- [130] M. G. Wentrup, K. Gramann, E. Wascher, and M. Buss, “EEG source localization for brain-computer-interfaces,” in *Neural Engineering, 2005. Conference Proceedings. 2nd International IEEE EMBS Conference on*, pp. 128–131, IEEE, 2005.
- [131] G. R. Müller-Putz, R. Scherer, C. Brauneis, and G. Pfurtscheller, “Steady-state visual evoked potential (ssvep)-based communication: impact of harmonic frequency components,” *Journal of neural engineering*, vol. 2, no. 4, p. 123, 2005.
- [132] M. Nakanishi, Y. Wang, Y.-T. Wang, Y. Mitsukura, and T.-P. Jung, “Generating visual flickers for eliciting robust steady-state visual evoked potentials at flexible frequencies using monitor refresh rate,” *PloS one*, vol. 9, no. 6, p. e99235, 2014.
- [133] Y. Zhang, G. Zhou, J. Jin, M. Wang, X. Wang, and A. Cichocki, “L1-regularized multiway canonical correlation analysis for SSVEP-based BCI,” *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 21, no. 6, pp. 887–896, 2013.
- [134] “Multiway canonical correlation analysis for frequency components recognition in SSVEP-based BCI—s, author=Zhang, Yu and Zhou, Guoxu and Zhao, Qibin and Onishi, Akinari and Jin, Jing and Wang, Xingyu and Cichocki, Andrzej, booktitle=Neural Information Processing, pages=287–295, year=2011, organization=Springer,”

- [135] J. Li, L. Zhang, and Q. Zhao, “Pattern classification of visual evoked potentials based on parallel factor analysis,” in *Advances in Cognitive Neurodynamics ICCN 2007*, pp. 571–575, Springer, 2008.
- [136] N. V. Manyakov, N. Chumerin, A. Combaz, A. Robben, M. van Vliet, and M. M. Van Hulle, “Decoding SSVEP responses based on parafac decomposition.,” in *BIOSIGNALS*, pp. 443–447, 2012.
- [137] Z. Lin, C. Zhang, W. Wu, and X. Gao, “Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs,” *Biomedical Engineering, IEEE Transactions on*, vol. 53, no. 12, pp. 2610–2614, 2006.
- [138] G. Bin, X. Gao, Z. Yan, B. Hong, and S. Gao, “An online multi-channel SSVEP-based brain–computer interface using a canonical correlation analysis method,” *Journal of neural engineering*, vol. 6, no. 4, p. 046002, 2009.
- [139] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, “Brain–computer interfaces for communication and control,” *Clinical neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.
- [140] S. J. Luck, *An introduction to the event-related potential technique*. MIT Press, Cambridge, MA, 2nd ed., 2014.
- [141] M. I. Garrido, J. M. Kilner, K. E. Stephan, and K. J. Friston, “The mismatch negativity: a review of underlying mechanisms,” *Clinical neurophysiology*, vol. 120, no. 3, pp. 453–463, 2009.
- [142] R. Näätänen, P. Paavilainen, T. Rinne, and K. Alho, “The mismatch negativity (MMN) in basic research of central auditory processing: A review,” *Clinical Neurophysiology*, vol. 118, no. 12, pp. 2544 – 2590, 2007.
- [143] R. Näätänen, T. Kujala, C. Escera, T. Baldeweg, K. Kreegipuu, S. Carlson, and C. Ponton, “The mismatch negativity (MMN)—a unique window to disturbed cen-

- tral auditory processing in ageing and different clinical conditions,” *Clinical Neurophysiology*, vol. 123, no. 3, pp. 424–458, 2012.
- [144] F. Cong, A. H. Phan, Q. Zhao, T. Huttunen-Scott, J. Kaartinen, T. Ristaniemi, H. Lyytinen, and A. Cichocki, “Benefits of multi-domain feature of mismatch negativity extracted by non-negative tensor factorization from EEG collected by low-density array,” *International journal of neural systems*, vol. 22, no. 06, p. 1250025, 2012.
- [145] F. Cong, G. Zhou, P. Astikainen, Q. Zhao, Q. Wu, A. K. Nandi, J. K. Hietaanen, T. Ristaniemi, and A. Cichocki, “Low-rank approximation based non-negative multi-way array decomposition on event-related potentials,” *International journal of neural systems*, vol. 24, no. 08, p. 1440005, 2014.
- [146] A. Chang, D. J. Bosnyak, and L. J. Trainor, “Unpredicted pitch modulates beta oscillatory power during rhythmic entrainment to a tone sequence,” *Frontiers in Psychology*, vol. 7, no. 327, 2016.
- [147] S. Pouryazdian, S. Beheshti, and S. Krishnan, “Localization of brain activities using multiway analysis of EEG tensor via EMD and reassigned TF representation,” in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pp. 113–116, Aug 2015.
- [148] S. Pouryazdian, S. Beheshti, and S. Krishnan, “CANDECOMP/PARAFAC model order selection based on reconstruction error in the presence of kronecker structured colored noise,” *Digital Signal Processing*, vol. 48, pp. 12 – 26, 2016.
- [149] D. C. Montgomery and G. C. Runger, *Applied Statistics and Probability for Engineers*. Wiley, 6th ed., 2013.