



HAL
open science

EEG Source Analysis

Marco Congedo

► **To cite this version:**

| Marco Congedo. EEG Source Analysis . Neuroscience. Université de Grenoble, 2013. tel-00880483

HAL Id: tel-00880483

<https://theses.hal.science/tel-00880483>

Submitted on 5 Apr 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Grenoble University

EEG Source Analysis

Habilitation à Diriger des Recherches presented to Doctoral School EDISCE

Marco Congedo
Ph.D., CRI CNRS

Date of Defense: 22/10/2013

Jury:

Prof. Christian Jutten - *President*

University of Grenoble, FRANCE. Senior member of IUF

Prof. Emeritus Fernando Lopes da Silva - *Rapporteur*

Swammerdam Institute for Life Sciences, University of Amsterdam, THE NETHERLANDS

Prof. Juri Kropotov - *Rapporteur*

Institute of the Human Brain, Russian Academy of Sciences, St. Petersburg, RUSSIA

Norwegian University for Science and Technology, Trondheim, NORWAY

Prof. Hichem Snoussi - *Rapporteur*

Université de Technologie de Troyes, FRANCE

Dirk De Ridder - *Examineur*

University of Otago, NEW ZEALAND

Prof. Philippe Kahane - *Examineur*

CHU, Grenoble, FRANCE

Prof. Emeritus Gert Pfurtscheller - *Examineur*

University of Technology, Graz, AUSTRIA,

*To my son Aaron,
with as much love as patience.*

SYNOPSIS

The French “HDR” diploma (Habilitation à Diriger des Recherches) is necessary to be the principal supervisor of PhD theses in French Universities. The candidate must write a dissertation that is defended in front of an international Jury. Typically, a collection of published articles with accessory information may suffice as content of the manuscript. For my HDR I decided to write from scratch a coherent manuscript with a considerable amount of unpublished content. Since my contributions in scientific journals as first author concern mainly methodological works (see chapter I), I decided to compile a manuscript reminding the structure of a small handbook of advanced methods for EEG data analysis. As a matter of fact all methods presented in this manuscript may be understood as a way to study the *latent variables* hidden in EEG recordings, what we name here generically *source analysis*, concept that will be precised as the reading progresses. Such a work is meant to be expanded and enriched in the future. It is addressed to students and peers approaching the field of quantitative EEG data analysis. The aim of the manuscript is to provide a succinct overview of methods that find practical utility, at least in my humble experience. The manuscript focuses on exhaustive descriptions of the algorithms and practical suggestions for their purposeful use. Many of the methods are illustrated by means of extensive simulations and real data, the latter pertaining to clinical, cognitive and Brain-Computer Interface studies. This has given me an opportunity to present many of the studies in which I have participated. The manuscript is organized in *ten* short chapters:

Chapter I contains a CV and other relevant information about my scientific career and achievements. It also contains a brief description of my research interests and main scientific collaborations. This chapter is meant to provide the member of the jury with objective information for evaluating my scientific status. Out of that, it plays no other role in this manuscript. The general reader may skip this chapter altogether.

Chapter II includes a section on notation and nomenclature, which are used consistently throughout the manuscript. Much effort has been spent in defining a simple, yet powerful and consistent notation arranging and harmonizing the different families of methods treated in the manuscript. I must say that this has been actually the stronger involvement in writing the manuscript, since a clear and consistent notation is of paramount importance in the quick grasping of the meaning of equations. In defining the notation I have tried to avoid as much as possible the use of the same symbols for different topics. For

doing so, notation has been simplified considerably and several symbols are used generically to represent objects of the same kind. Furthermore, I have employed both the Latin and Greek alphabet. The reader is invited to study carefully the section *Notation and Nomenclature*, after which the equations in the manuscript should result clear at first glance. If this is not the case then I have not succeeded in my endeavor. The chapter then collects basic known results in linear algebra and statistics, to which the reader is referred throughout the ensuing chapters. All linear algebra results needed, with some more as a bonus, are collected in this section, named *Linear Algebra*. The expert reader may walk through quickly here. Finally, the chapter includes a short introduction to the physiology and physics of EEG, yielding the definition of the *EEG sensor measurement*, which is the starting point of all EEG data analysis, including the source analysis methods that we treat here.

Chapter III introduces the family of *regularized weighted minimum-norm inverse solutions*. It addresses thus the problem of *source localization*. Emphasis is given to the two methods we have been using most, named sLORETA and eLORETA. Both the model driven and data driven version of these methods are presented, uncovering the connection with the well-known family of linearly constrained minimum-variance inverse solutions. Useful suggestions on the use of inverse solutions are provided. The chapter ends with a short overview of my contributions in regional current density estimations, particularly useful for real-time applications, including the use of data-independent filters known as beamformers and data-dependent filters designed to increase the signal to noise ratio, the classification accuracy, or any other sought purpose. This chapter has been included since inverse solutions are heavily employed in chapter VI and VII, hence I felt important to give an account of the mathematical background of these methods.

Chapter IV, V, VI, VII taken together represent a long journey into the wide family of methods based on the *diagonalization of matrices* holding second-order statistics of the data (i.e., covariance matrices and similar). In **Chapter IV** a general framework for the (approximate) joint diagonalization of matrices estimated on multiple data sets is presented, showing that the same optimization in a least-squared framework can be used to solve all problems encountered in these chapters. Two algorithms for solving the general optimization scheme are given. **Chapter V** treats basic spatial filters such as as *principal component analysis, whitening, maximal covariance analysis, canonical correlation analysis* and the *common spatial pattern*. The journey continue in **Chapter VI** with *blind source separation* (BSS) methods based on second-order statistics and associated algorithms, such as AMUSE, SOBI etc. In this chapter we investigate BSS theory and we provide a general conceptual framework and algorithm (AJDC) to deal with all major kinds of EEG data, namely, spontaneous, induced and evoked EEG. Finally, chapter VII treats *group BSS* methods and *Joint BSS* methods, which are extension of

the BSS methods when multiple subjects or multiple data sets are analyzed simultaneously. The different families of methods are illustrated with many real data examples.

Chapter VIII and **IX** are the most original of the manuscript. Chapter **VIII** presents a new *universal framework for BCI classification* based on *Riemann geometry*. We show that the very same signal processing chain with minimal changes can be adapted to all current BCI modalities, including those based on the analysis of event-related (de)synchronizations, evoked-response potentials and steady-state evoked potentials. The framework is well adapted to support a new generation of multi-user BCI functioning without calibration. Our claim is supported with the classification of several data sets for each BCI modality. We believe firmly that the Riemannian framework will receive more and more attention in the BCI community and candidates to become the “standard” very much sought by the field. **Chapter IX** presents, rather exhaustively, current and very recent advances in *differential Riemannian geometry* and *the affine-invariant metric*, that is, it treats the theoretical bases allowing the results presented in chapter VIII. While the Riemannian framework is at first sight mathematically hostile, it turns out to be extremely simple in actual usage, much more simple than most advanced methods presented in chapter III-VII. The reader facing these tools for the first time is invited to tackle this chapter slowly, with an open-minded and challenging attitude. The chapter ends with some theoretical investigations we have started during the very last months, disclosing, among other things, some connections between the Riemannian distance and geometric mean with the material presented in chapter V, VI and VII. Riemann geometry has appeared in EEG literature only five years ago, but is gaining momentum very rapidly. The material presented in chapter VIII and IX is still completely or rather unknown to most EEG specialists. It has been included especially thinking to the members of the Jury, who I hope will find in this chapter new interesting and stimulating ideas, besides sharp results in the field of BCI.

To conclude, **Chapter X** contains an overall discussion of the entire manuscript and some perspectives for future research.

Chapters III to IX form the core of the manuscript. They are self-contained, thus they may be read independently, although several cross-references are included to preserve the unity of the manuscript and to show connections between different research fields; after all, the methods presented in section III and those presented in chapters V-VII are clearly complementary, whereas the connections between the methods presented in chapter VIII-IX and those presented in the others are largely to be uncovered.

PREFACE

"We cannot solve the problems
we have created
with the same thinking that created them."

- Albert Einstein, (1879 - 1955),
1922 Nobel Laureate for Physics

"Atrocities are not less atrocities
when they occur in laboratories
and are called medical research."

- George Bernard Shaw (1856 - 1950),
1925 Nobel Laureate for Literature

Yesterday I came across a post, dated July 3rd 2013, that I found in a blog for video game players¹. The post comments a video describing the advances of the OpenViBE2 project, to which I have participated (2009-2013). OpenViBE2 has assembled several research institutions and several actors of the video game industry with the aim of prospecting the introduction of brain-computer interface (BCI) technology in consumer video games. Here is my English translation of the post:

... Here it is my friends, the stammering of the future! The first steps of a new adventure that we will experience shortly. Because finally, what a few months ago seemed to be just science-fiction, is rising now. Yes, we will experience one day much more than a video game! In our head, we will know, I am sure, the virtual world projected directly in our brain and we will be able to live parallel lives of our choice, where we will be like gods!...

The enthusiasm of the blogger is a prototypical example of the increasing expectation about the possibilities of BCI technology, which is seen by some as a way to surpass natural human capabilities. I must admit that the idea to “live parallel lives of our choice, where we will be like gods!” has left me rather perplexed. Why one may find such a thing desirable? Is this blogger suggesting that BCI may be conceived as a way to edulcorate the human condition? Should our goal be escaping from our natural condition or rather should we try to understand how to live with it? I was still thinking to the implications of the post when, today, I found an article in “Paris Match” titled “*The crazy project of a*

¹ <http://www.team-vips.com/t4201-interface-cerveau-ordinateur>

*Russian billionaire: eternal life*². The goal of the so called “2045 Initiative” is to achieve, by 2045, the embedding of a brain with its consciousness in a chip. This is sometimes referred to as *mind-to-computer uploading*. The billionaire at the origin of the project, which name is Dmitri Itskov, says: « *I am going to get old and then die and all this for what? Life cannot resolve in this sad equation* ». The project aims at an intermediary goal by 2020, which is to create an android avatar completely controlled by a BCI. Far from being just a rumor, it appears that the project has already involved several respectable and authoritative scientists. Again, BCI technology is evoked as a means to achieve what is *naturally precluded to humans*. Moreover, we must admit that the expectation is shared also by some experts. If yesterday I was perplexed, today I am puzzled. Why one may desire to “live” consciously in a chip? Wouldn’t that be a nightmare? What could be human consciousness without human life?

Undoubtedly, the possibility to allow communication without using the natural muscular and peripheral nerves pathways is a unique characteristic of BCI technology. This peculiarity fosters the dreams of the civilized mankind and there is nothing wrong in dreaming. However, we should be aware that BCI technology has risen in medical research - at least this has been its major showcase - as a possibility for those suffering of extremely disabling physical conditions preventing the communication with the external world. Today it is strongly motivated by military aims and is source of inspiration for the whole field of robotics. Put it simply, in my view pretending that BCI technology is meant to empower natural human abilities has nothing to do with science, thus I think it is about time for the BCI scientific community to start discussing seriously the many ethical questions concerning the role and purpose of BCI research in this world.

² <http://www.parismatch.com/Actu/International/Le-projet-fou-d-un-milliardaire-russe-la-vie-eternelle-521150>

AKNOWLEDGEMENTS

The following is a minimal and far from being exhaustive list of people that should find place here.

My esteem goes in the sky for *Prof. Shayle Searle* (Cornell University), who disappeared in 2013 before I could have the pleasure to meet, but whose book *Matrix Algebra Useful for Statistics* (1982) has been determinant for my discovery of the beauty of linear algebra when I read it almost ten years ago. This manuscript can be conceived as the side-effect of the light a single book could shade in the mind of the author, for whom linear algebra at that time was as opaque as a table of logarithms.

My gratitude goes to *Prof. Christian Jutten*, who has constantly guided my career since 2006, when he started working with me to prepare the candidature at CNRS. He has always been an excellent mentor, as such, I would be glad if he is chosen by the members of the jury as the President of the Committee.

It is with honor that I present my thanks to all other members of the jury, in alphabetical order, *Prof. Dirk De Ridder* (Uni. of Otago, New Zealand), *Prof. Fernando Lopes da Silva* (Uni. of Amsterdam, The Netherlands), *Prof. Philippe Kahane* (CHU, Grenoble, France), *Prof. Jury Kropotov* (Russian Academy of Sciences, St. Petersburg, Russia), *Prof. Gert Pfurtscheller* (Uni. of Technology, Graz, Austria) and *Prof. Hichem Snoussi* (Uni. of Technology of Troyes, France). Their participation to the jury has been for me a strong motivation for writing an original manuscript from scratch, rather than contenting myself with the compilation of previously published material.

I am truly indebted to *Dr. Alexandre Barachant*, who has produced the classification results and associated figures of chapter VIII. Alex is at the origin of several ideas presented in chapter VIII. Working with him during his post-doc in the period 2012-2013 has been a pleasure, every single day. Furthermore, with no doubt he has been the best post-doc I have ever seen in action.

My thanks go to *Dr. Alexandre Barachant* and *Jonas Chatel-Goldman*, who will read this manuscript before it will be submitted to the Jury and, surely, will return me useful feedback.

I would like to thank *Prof. Anne Guerin* for supervising all procedures necessary to obtain the HDR.

*Grenoble, 7th of July 2013,
in the day of the Sun.*

INDEX

Synopsis	5
Preface	9
Aknowledgements	11
Chapter I	19
About the Candidate	19
Anagraphic Information	20
Biographical Sketch.....	21
Studies and Positions.....	22
Publication Activity Summary	24
List of Publications.....	25
Grants	32
Teaching at Universities	33
Prizes, Recognitions, Awards	33
Media Coverage	34
Review Consulting	34
Committees.....	35
Invited Lectures	36
Student Supervision.....	37
Web Site	40
Cursus	41
Research Interests and Collaborations	43
BCI, P300, ErrP and MI single-trial classification, SVM, Riemann Geometry.....	44
ICA Neurofeedback.....	50
Normative EEG Database.....	51
Tinnitus.....	52
Software Development	53
Chapter II	57
Background Material	57
Notation and Nomenclature	58
Linear Algebra	62
Invariances	62
The Rank of a Matrix	62
The Trace of Square Matrix $Q \in \mathfrak{R}^{N \times N}$	63
The Symmetric Matrix $S \in \mathfrak{R}^{N \times N}$	63
Symmetric Positive-Definite (SPD) Matrix $C \in \mathfrak{R}^{N \times N}$	64

Orthogonal Matrix $U \in \mathbb{R}^{N \times N}$	64
Eigenvalue-Eigenvector Decomposition (EVD).....	65
Properties of Eigenvalues	65
Power Iterations.....	66
Cholesky Decomposition.....	67
Operators on Symmetric Positive-Definite Matrices	67
Some Results on Matrix Exponential and Logarithm.....	68
Other Results on SPD Matrices	69
Singular Value Decomposition (SVD)	69
Lödwin Orthogonalization.....	69
Moore-Penrose Pseudo-Inverse and Pseudo-Operators	70
Joint Diagonalization of Two Symmetric Matrices	71
Fourier Analysis.....	72
Statistics.....	73
Mean.....	73
Centering Matrix and Common Average Reference.....	73
Sum of Squares and Products	74
Covariance Matrix	74
EEG Basics	75
Advent and Standardization of EEG Recordings.....	75
EEG and Other Neuroimaging Modalities.....	76
The Advent of Quantitative EEG Analysis.....	76
EEG Norms	77
EEG Source Analysis	78
A Short Introduction to the Physiology and Physics of EEG	79
The Sensor Measurement	83
Bipolar and Monopolar Reference.....	83
Common Average Reference.....	84
Chapter III	85
Distributed Inverse Solutions	85
Introduction	86
The Forward Problem.....	87
The Inverse Problem.....	88
Inverse solutions satisfying the sensor measurement	89
Inverse solutions with no localization error for noiseless sensor measurement	90
The Minimum Norm Inverse Solution.....	91
Regularization of Minimum Norm Inverse Solutions.....	91
Weighted Minimum-Norm Inverse Solutions.....	92
Model-Driven sLORETA.....	92
Data-Driven sLORETA.....	93
Model-Driven eLORETA.....	94
Data-Driven eLORETA.....	96

Similarity with the Minimum Variance Beamforming	96
Point Spread Function Simulations.....	97
Conclusions	99
Current Density Estimation in Regions of Interest	100
Data-Independent Filters for Regional Inverse Solutions	102
Data-Dependent Filters for Regional Inverse Solutions	103
Measurement noise suppression	103
Increasing classification accuracy.....	104
Other Filters for Regional Inverse Solutions	108
Co-Registration of Inverse Solutions with MRI	108
Chapter IV.....	109
<i>The Joint Diagonalization Framework</i>	<i>109</i>
Introduction	110
One data set, one matrix	113
One data set, two matrices.....	113
Two data sets, one matrix	114
One data set, several matrices.....	114
Several data sets, several matrices	115
<i>Approximate Joint Diagonalization</i>	<i>116</i>
Least-Squares Functional	117
The Orthogonal Mixing Matrices Case	118
The Invertible Mixing Matrices Case	120
Simulations.....	122
Conclusion.....	125
Chapter V	127
<i>Spatial Filters</i>	<i>127</i>
Introduction	128
Principal Component Analysis (PCA)	128
Whitening	129
Common Spatial Pattern (CSP)	130
Maximum Covariance Analysis (MCA)	131
Canonical Correlation Analysis (CCA)	132
Chapter VI.....	135
<i>Blind Source Separation</i>.....	<i>135</i>
Introduction	136
The BSS Problem for EEG.....	137
A Suitable Class of BSS Solutions	138
BSS Filtering	139
Localization of BSS Components.....	139
Different Approaches for Solving BSS.....	140
BSS Based on the Joint Diagonalization of Two Matrices	142
Closed Form BSS Solutions for Colored Processes	142

Closed Form BSS Solutions for Non-Stationary Processes	143
BSS by Approximate Joint Diagonalization of a Matrix Set	143
The SOBI AJD Methods.....	144
The Fundamental Theorem of AJD-based BSS.....	144
The AJD of Fourier Cospectra (AJDC) Algorithm.....	145
The implementation of AJDC.....	148
Things to know working with AJDC.....	149
Discussion on SOS-based BSS Methods	151
<i>Example Studies Using AJDC.....</i>	154
Spontaneous activity.....	154
Induced activity	155
Evoked activity.....	156
Introduction.....	156
Experimental design	157
AJDC analysis	157
Method	158
Participants.....	158
Trials	158
Data acquisition.....	160
Preprocessing	160
Analysis in the sensor space	161
Analysis in the source space.....	161
Classification of single trials	163
Results	163
Behavioral results.....	163
Sensor space analysis	164
Source analysis.....	165
Source localization	167
Classification of single trials	167
<i>Chapter VII.....</i>	170
<i>Group and Joint Blind Source Separation.....</i>	170
Introduction	171
<i>Group Blind Source Separation (gBSS).....</i>	173
Introduction	173
Method of Our Study.....	174
Databases.....	174
Recording procedures.....	174
Pre-Processing	174
Results	175
Explained variance	177
Deriving group norms.....	178
Experimental Studies with gBSS.....	178
Clinical gBSS studies	178
Cognitive gBSS studies	179

Limitations of the gBSS Approach	181
Joint Blind Source Separation (JBSS).....	187
Introduction	187
The JBSS Framework.....	187
The Extended AJDC Algorithm	190
JBSS Model Order.....	191
Chapter VIII.....	194
Riemann Geometry: a Universal BCI Classification Framework.....	194
Introduction	195
EEG Data Modeling	199
The Classification Framework.....	199
Smart Initialization (Cross-Subject and Cross-Session Generalization).....	201
Adaptation	202
Classification of Motor Imagery.....	202
The form of covariance matrices for motor imagery BCI data.....	202
Analysis of motor imagery BCI data	203
Classification of Event-Related Potentials.....	206
The form of covariance Matrices for of P300 BCI data	206
Analysis of P300 BCI data	208
Classification of Steady-State Evoked Potentials	218
The form of covariance Matrices for Steady-State Evoked Potentials	218
Analysis of Steady-State Visually Evoked Potential BCI data	219
Conclusion and Discussion.....	221
Chapter IX.....	224
Riemann Geometry: a Theoretical Prime	224
Introduction	225
The Riemannian Manifold.....	226
The Exponential and Logarithmic Map	228
The Geodesic	229
The Distance.....	229
The Norm	231
The Geometric Mean of Points on the Manifold	232
Recent Investigations	238
Introduction	238
Connections with Diagonalization Methods.....	238
Blind Source Separation	238
Joint Blind Source Separation.....	239
A Diagonality Function	240
AJD and the geometric mean of a matrix set	241
Standardized Distances to Geometric mean.....	243
The Distance Matrix	243
Wiener Entropy: an Index of Cloud Entropy	244

<i>Chapter X</i>	246
<i>Conclusions and perspectives</i>	246

CHAPTER I

ABOUT THE CANDIDATE

Anagraphic Information



Born in	Bari, Italy
Date of Birth	20 October 1972
Gender	Male
Languages	English, French, Italian (mother tongue)
Citizenship	Italian
Permanent e-mail:	marco.congedo@gmail.com
Website:	http://sites.google.com/site/marcocongedo/home

Biographical Sketch

Marco Congedo obtained the Ph.D. degree in Biological Psychology with a minor in Statistics from the University of Tennessee, Knoxville, in 2003. From 2003 to 2006 he has been a post-doc fellow at the French National Institute for Research in Informatics and Control (INRIA) and at France Telecom R&D, in France. Since 2007 Dr. Congedo is a Research Scientist at the “Centre National de la Recherche Scientifique” (CNRS) in the GIPSA Laboratory, Grenoble, France.

Dr. Congedo has been the recipient of several awards, scholarships and research grants. He is interested in basic human electroencephalography (EEG) and magnetoencephalography (MEG), real-time neuroimaging (neurofeedback and brain-computer interface) and multivariate statistical tools useful for EEG and MEG such as inverse solutions, blind source separation and Riemannian geometry.

Dr. Congedo is a Fellow of the International Society for Neurofeedback and Research and a Consulting Editor for the Journal of Neurotherapy.

Studies and Positions

Legend:  Studies;  Positions

Since 2007  **CNRS (Centre National de la Recherche Scientifique)**, Grenoble, France.



Chargés de Recherche 1ère classe

Research on Neurofeedback, Brain Computer Interface and Digital Signal Processing

2005-2006  **France Telecom R&D**, Grenoble, France.



Post-doctoral fellowship

supervision: Dr. Denis Chêne.

2003-2005  **INRIA** (National Institute for Research in Informatics and Control), Rennes and Grenoble, France.



Post-doctoral fellowship

supervision: Dr. Anatole Lécuyer.

2000-2007  **Nova Tech EEG, Inc.**, Mesa, AZ, USA.

Hardware, Software and Services for Research and Education in Electroencephalography



R&D Director – and Co-Founder.

1999-2003  **University of Tennessee**, Knoxville, USA. Department of Psychology



2003 - Philosophy Doctor major degree (PhD) in *Biological Psychology*.

Supervision: Prof. Joel Lubar

Thesis Title: “*Tomographic Neurofeedback; a new Technique for the Self-Regulation of Brain Electrical Activity*”.



2003 - Philosophy Doctor minor degree (PhD) in *Statistics*.

Supervision: Prof. William Seaver



2001 - Master of Arts degree (M.A.) in *Biological Psychology*

Supervision: Prof. Joel Lubar

Thesis Title: “*On the Comparison to EEG Norms: A new Method and a Simulation Study*”.

1998-1999

Università di Bari, Bari, Italy,

School of Medicine and Hospital, Department of Neurological and Psychiatric Sciences



Internship in *Clinical Psychology and Electroencephalography*.

Supervision: Dr. Rita Carone

1991-1998

Università di Padova, Padova, Italy. Department of Psychology



1998 – Laurea (undergraduate degree & M.A.) in *Experimental Social Psychology*

Supervision: Prof. Dora Capozza

Thesis Title: “Group Distance of the Self and Perception of Homogeneity”.

1996

Université René Descartes (Paris V), Paris, France. Department of Psychology

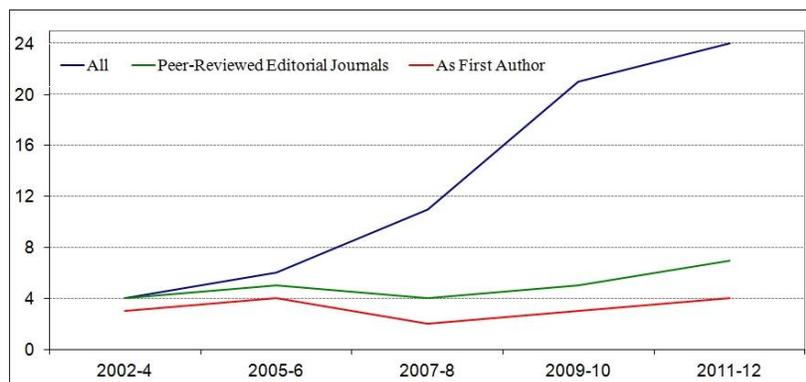


Student exchange Program (ERASMUS) in *Experimental Social Psychology*

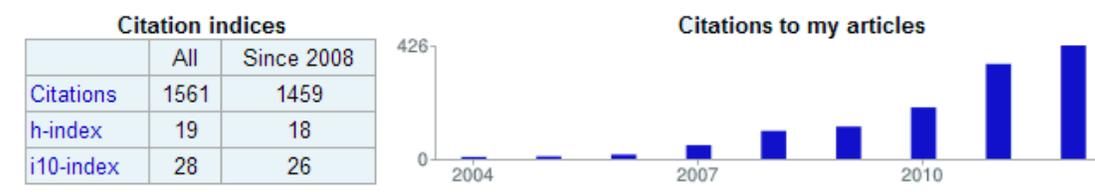
Supervision: Dr. Françoise Askevis.

Publication Activity Summary

Publications per Year



Citations per Year



Source: Google Scholar. *h-index* is the largest number *h* such that *h* publications have at least *h* citations. *i10-index* is the number of publications with at least 10 citations.

Some Co-Authors

Olivier Bertrand	(Lyon)
E. Roy John	(New York)
Christian Jutten	(Grenoble)
Fabien Lotte	(Bordeaux)
Joel Lubar	(Knoxville)
Dinh-Tuan Pham	(Grenoble)
Gert Pfurtscheller	(Graz)
Alain Rakotomamonjy	(Rouen)
Richard Silberstein	(Malbourne)
Dirk De Ridder	(Antwerp)

List of Publications

Legend:  Article in Peer-Reviewed Editorial Journals;  Article in Conference Proceedings;  Book Chapter

2013



Barachant A., Bonnet S., **Congedo M.**, Jutten C. (2013) Classification of covariance matrices using a Riemannian-based kernel for BCI applications, *Neurocomputing*, 112, 172-178.



Barachant A, **Congedo M**, Van Veen G, Jutten C (2013) Classification de potentiels évoqués P300 par géométrie riemannienne, *GRETSI Proceedings*, (in press).



Barachant A, Andreev A, **Congedo M** (2013) The Riemannian Potato: an automatic and adaptive artifact detection method for online experiments using Riemannian geometry, *TOBI Workshop IV*, Sion : Switzerland.



Chatel-Goldman J, Schwartz J-L, Jutten C, **Congedo M** (2013) Nonlocal mind from the perspective of social cognition, *Frontiers in Human Neuroscience* (in press).



Chatel-Goldman J, **Congedo M**, Phlypo R (2013) Joint BSS as a natural analysis framework for EEG-hyperscanning, *ICASSP 2013*, Vancouver: Canada (in press)



Kopřivová J, **Congedo M**, Horáček J, Raszka M, Brunovský M, Praško J (2013) Standardized low-resolution electromagnetic tomography in obsessive-compulsive disorder – a replication study, *Neuroscience Letters*, 548, 185-9.



Kopřivová J, **Congedo M**, Raszka M, Praško J, Brunovský M, Horáček J (2013) Prediction of Treatment Response and the Effect of Independent Component Neurofeedback in Obsessive-Compulsive Disorder: A Randomized, Sham-Controlled, Double-Blind Study, *Neuropsychobiology*, 67:210-223.



Mayaud L, Filipe S, Pétégnief L, Rochecouste O, **Congedo M** (2013) Robust Brain-computer interface for virtual Keyboard (RoBIK): project results, *Ingénierie et Recherche Biomédicale / IRBM BioMedical Engineering and Research* (in press).



Mayaud L, **Congedo M**, Van Laghenhove A, Figère M, Azabou E, Cheliout-Heraut F (2013) A Comparison of Recording Modalities of P300 Event Related Potentials (ERP) for Brain-Computer Interfaces (BCI) Paradigm, *Neurophysiologie Clinique / Clinical Neurophysiology* (in press).



Vanneste S, **Congedo M**, De Ridder D (2013) Pinpointing a highly specific pathological functional connection that turns phantom sound into distress, *Cerebral Cortex* (in press).

2012



Barachant A, Bonnet S, **Congedo M**, Jutten C (2012a) BCI Signal Classification using a Riemannian-based kernel, *20th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, Bruges : Belgique



Barachant A, Bonnet S, **Congedo M**, Jutten C (2012b) Multi-Class Brain Computer Interface Classification by Riemannian Geometry. *IEEE Transactions on Biomedical Engineering* 59(4), 920-928.



Congedo M, Phlypo R, Chatel-Goldman J (2012) Orthogonal and Non-Orthogonal Joint Blind Source Separation in the Least-Squares Sense, *20th European Signal Processing Conference (EUSIPCO)*, Aug 27-31, Bucharest, Romania, 1885-9.



Jrad N, Congedo M (2012) Identification of spatial and temporal features of EEG, *Neurocomputing*, 90, 66-71.



Rousseau S, Jutten M, **Congedo M** (2012a) Designing Spatial Filters Based on Neuroscience Theories to Improve Error-Related Potential Classification, *IEEE International Workshop on Machine Learning for Signal processing*, Santander : Spain.



Rousseau S, Jutten M, **Congedo M** (2012b) Time window selection for improving error-related potential detection, *4th International Conference on Neural Computation Theory and Applications*, Barcelone : Spain



Rousseau S, Jutten C, **Congedo M** (2012c) Closed-looping a P300 BCI using the ErrP, *4th International Conference on Neural Computation Theory and Algorithms*, Barcelone : Spain



Rousseau S, Jutten C, **Congedo M** (2012d) The error-related potential and BCIs, *20th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, Bruges : Belgique



White D, **Congedo M**, Ciorciari J, Silberstein R (2012) Brain oscillatory activity during spatial navigation: Theta and gamma activity link medial temporal and parietal regions. *Journal of Cognitive Neuroscience*, 24(3), 686-697.

2011



Aguilar Herrero M, **Congedo M**, Minguez J (2011) A Data-Driven Process for the Development of an Eyes-closed EEG Normative Database, *Proceedings of the 33rd International IEEE EMBS Conference*, 7306-7309.



Barachant A, Bonnet S, **Congedo M**, Jutten C. (2011a) Réalisation d'un Brain-Switch EEG par Géométrie Riemannienne, *GRETSI*.



Barachant A, Bonnet S, **Congedo M**, Jutten C (2011b) A Brain-Switch Using Riemannian Geometry, *Proceedings of the 5th International BCI Conference*, Graz, Austria, 64-67.



Cecotti H, Rivet B, **Congedo M**, Jutten C, Bertrand O, Mattout J, Maby E (2011) A Robust Sensor Selection Method for P300 Brain-Computer Interfaces, *Journal of Neural Engineering*, 8(1), 016001.



Congedo M, Goyat M, Tarrin N, Varnet L, Rivet B, et al. (2011) "Brain Invaders": a prototype of an open-source P300-based video game working with the OpenViBE platform, *Proceedings of the 5th International BCI Conference, Graz, Austria*, 280-283.



Congedo M, Phlypo R, Pham D-T (2011) Approximate Joint Singular Value Decomposition of an Asymmetric Rectangular Matrix Set, *IEEE Transactions on Signal Processing*, 59(1), 415-424.



De Ridder D, Vanneste S, **Congedo M** (2011) The distressed brain: a group blind source separation analysis on tinnitus, *PLoS One*, 6(10), e24273.



Jrad N, Phlypo R, **Congedo M**. (2011) SVM feature selection for multidimensional EEG data, *International Conference on Acoustic, Speech and Signal Processing, (ICASSP), May 22-27, Praha, Czech Republic*.



Jrad N, **Congedo M** (2011a) Identification of sparse spatio-temporal features in Evoked Response Potentials, *European Symposium on Artificial Neural Networks (ESANN), April 27-29 Bruges, Belgium*.



Jrad N, **Congedo M**. (2011b) Spatio-temporal feature extraction and classification of Event-Related Potentials. *Conférence Francophone d'Apprentissage (CAP), Chambéry, France, May 17-20*.



Jrad N, **Congedo M**, Phlypo R, Rousseau S, Flamary R, Yger F, Rakotomamonjy A (2011) sw-SVM : sensor weighting support vector machines for EEG-based Brain-Computer Interfaces, *Journal of Neural Engineering*, 8(5), 056004.



Kopřivová J, Congedo M, Horáček J, Praško J, Raszka M, Brunovský M, Kohútová B, Höschl C. (2011) EEG source analysis in obsessive-compulsive disorder. *Clinical Neurophysiology*, 122(9), 1735-1743.



Mayaud L, **Congedo M**, Filipe S, Charvet G, Schoettel R, Annane D (2011), Robust Virtual Keyboard for Brain-Computer Interface (ROBIK): An Halfway Update on the Project, *Second IASTED International Conference on Robotics*.



Phlypo R, Jrad N, Rousseau S, **Congedo M** (2011) A Non-Orthogonal SVD-based Decomposition for Phase Invariant Error-Related Potential Estimation, *33rd Annual International IEEE EMBS Conference*.



van der Loo E, **Congedo M**, Vanneste S, Van De Heyning P, De Ridder D (2011) Insular Lateralization in Tinnitus Distress. *Autonomic Neuroscience: Basic and Clinical*, 165(2), 191-194.

2010



Barachant A, Bonnet S, **Congedo M**, Jutten C (2010a) Common Spatial Pattern revisited by Riemannian Geometry, *Proceedings of the IEEE International Workshop on Multimedia Signal Processing*, 472-476.



Barachant A, Bonnet S, **Congedo M**, Jutten C (2010b) Riemannian Geometry Applied to BCI Classification, *Proceedings of Latent Variable Analysis and Signal Separation Conference, Saint Malo, France*, 6365, 629-636.



Cecotti H, Phlypo R, Rivet B, **Congedo M**, Maby E, Mattout J (2010) Impact of the time segment analysis for P300 detection with spatial filtering, *Proceedings of the 3rd International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL), 11 July, Italy*.



Cecotti H, Rivet B, **Congedo M**, Jutten C, Bertrand O, Maby E, Mattout J (2010) Suboptimal Sensor Subset Evaluation in a P300 Brain-Computer Interface, *European Signal Processing Conference (EUSIPCO), August 23-27, Aalborg, Denmark*.



Congedo M, John RE, De Ridder D, Prichep L (2010) Group Independent Component Analysis of Resting-State EEG in Large Normative Samples, *International Journal of Psychophysiology*, 78: 89-99.



Congedo M, John RE, De Ridder D, Prichep L, Isenhardt B (2010) On the "Dependence" of "Independent" Group EEG Sources; an EEG Study on Two Large Databases, *Brain Topography*, 23(2), 134-138.



Congedo M, Sherlin L (2010) EEG Source Analysis: Methods and Clinical Implications. In "Neurofeedback and Neuromodulation Techniques and Applications", (Ed) Coben R., Evans J.R., Academic Press, New York (25-46).



Gouy-Pailler C, **Congedo M**, Brunner C, Jutten C, Pfurtscheller G (2010) Nonstationary brain source separation for multiclass motor imagery. *IEEE Transactions on Biomedical Engineering*. 57(2): 469-78.



Lécuyer A, Congedo M, Gentaz E, Joly O, Coquillart S (2010) Influence of Visual Feedback on Passive Tactile Perception of Speed and Spacing of Rotating Gratings, *Eurohaptics Conference, Amsterdam, The Netherlands, July 8-10 2010: Lecture Note in Computer Science*, 6192, 73-78.



Phlypo R, **Congedo M** (2010) An Extension of the Canonical Correlation Analysis to the Case of Multiple Observations of Two Groups of Variables, *32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Buenos Aires : Argentina*.



Phlypo R, Jrad N, Rivet B, **Congedo M** (2010) Common SpatioTemporal Pattern Analysis, *Proceedings of Latent Variable Analysis and Signal Separation Conference, Saint Malo, France*, 6365, 596-603.



Renard Y, Lotte F, Gibert G, **Congedo M**, Maby E, Delannoy V, Bertrand O, Lécuyer A (2010) OpenVIBE: An Open-Source Software Platform to Design, Test and Use Brain-Computer Interfaces in Real and Virtual Environments, *PRESENCE : Teleoperators and Virtual Environments*, 19(1), 35-53.



Vanneste S, Plazier M, der Loo EV, de Heyning PV, **Congedo M**, De Ridder D (2010) The Neural Correlates of Tinnitus-Related Distress, *Neuroimage*, 52(2), 470-80.

2009



Congedo M, Pham D-T (2009) Least-Squares Joint Diagonalization of a Matrix Set by a Congruence Transformation, *SinFra'09 (Singaporean-French IPAL Symposium)*, Singapore, Feb 18-20.



Gouy-Pailler C, Mattout J, **Congedo M**, Jutten C (2009) Uncued brain-computer interfaces: a variational hidden markov model of mental state dynamics, *Proceedings of European Symposium on Artificial Neural Networks Advances in Computational Intelligence and Learning - ESANN Belgium*.



Gouy-Pailler C, Sameni R, **Congedo M**, Jutten C (2009) Iterative Subspace Decomposition for Ocular Artifact Removal from EEG Recordings, *Independent Component Analysis and Signal Separation - 8th International Conference, ICA 2009, Brasil*.



Kouijzer MEJ, de Moor JMH, Gerrits BJL, **Congedo M**, van Schie HT (2009). Neurofeedback improves executive functioning in children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, 3, 145-162.



Pham D-T, **Congedo M** (2009) Least Square Joint Diagonalization of Matrices Under an Intrinsic Scale Constraint, *ICA 2009 (8th International Conference on Independent Component Analysis and Signal Separation)*, March 15-18, Paraty, Brasil, 298-305.



van der Loo E, Gais S, **Congedo M**, Vanneste S, Plazier M, et al. (2009). Tinnitus Intensity Dependent Gamma Oscillations of the Contralateral Auditory Cortex. *PLoS ONE* 4(10): e7396. doi:10.1371/journal.pone.0007396.

2008



Congedo M, Gouy-Pailler C, Jutten C (2008) On the blind source separation of human electroencephalogram by approximate joint diagonalization of second order statistics. *Clinical Neurophysiology*, 119, 2677-2686.



Congedo M, Jutten C, Sameni R, Gouy-Pailler C, (2008) A new General Weighted Least-Squares Algorithm for Approximate Joint Diagonalization, *Proceedings of the 4th International Brain-Computer Interface Workshop and Training Course, Graz, Austria*, 98-103.



Gouy-Pailler C, **Congedo M**, Brunner C, Jutten C, Pfurtscheller G (2008) Multi-Class Independent Common Spatial Patterns: Exploiting Energy Variations of Brain Sources new General Weighted Least-Squares Algorithm for Approximate Joint Diagonalization. *Proceedings of the 4th International Brain-Computer Interface Workshop and Training Course, Graz, Austria. 20-25.*



Gouy-Pailler C, **Congedo M**, Jutten C, Brunner C, Pfurtscheller G (2008), Model-Based Source Separation for Multi-Class Motor Imagery, *Proceedings of the 16th European Signal Processing Conference (EUSIPCO-2008), EURASIP, Lausanne, Switzerland, August 2008.*

2007



Cannon R, Lubar JF, **Congedo M**, Thornton K, Towler K, Hutchens T (2007) The Effect of Neurofeedback Training in the Cognitive Division of the Anterior Cingulate Gyrus, *International Journal of Neuroscience, 117(3), 337-357.*



Congedo M, Joffe D (2007), Multichannel Tomographic Neurofeedback: Wave of the future?. In *"Handbook of Neurofeedback: Dynamics and Clinical Applications "*, (Ed) Evans J.R., *Haworth Press, New York, (85-107).*



Gouy-Pailler C, Achard S, Rivet B, Jutten C, Maby E, Souloumiac A, **Congedo M** (2007), Topographical dynamics of brain connections for the design of asynchronous brain-computer interfaces, *Proceedings IEEE Eng Med Biol Soc, 1: 2520-2523.*



Gouy-Pailler C, Rivet B, Achard S, Souloumiac A, Jutten C, Maby E, **Congedo M** (2007), Théorie des graphes et dynamique des connexions cérébrales pour la conception d'interfaces cerveau-machines asynchrones, *Proceedings of the 21st Conference GRETSI, Troyes, France.*



Lotte F, **Congedo M**, Lécuyer A, Lamarche F, Arnaldi B (2007) A Review of Classification Algorithms for EEG-based Brain-Computer Interfaces, *Journal of Neural Engineering, 4(2), R1-R13.*



Sherlin L, Budzynski T, Kogan-Budzynski H, **Congedo M**, Fischer ME, Buchwald D (2007) Low-resolution electromagnetic brain tomography (LORETA) of monozygotic twins discordant for chronic fatigue syndrome, *Neuroimage, 34(4), 1438-1442.*



Van der Loo E, **Congedo M**, Plazier M, Van de Heyning P, De Ridder D (2007), Correlation between Independent Components of scalp EEG and intra-cranial EEG (iEEG) time series, *International Journal of Bioelectromagnetism, 9(4), 270-275.*

2006



Congedo M (2006) Subspace Projection Filters for Real-Time Brain Electromagnetic Imaging, *IEEE Transactions on Biomedical Engineering*, 53(8), 1624-34.



Congedo M, Lécuyer A, Gentaz E (2006) The influence of Spatial De-location on Perceptual Integration of Vision and Touch, *PRESENCE Teleoperators and Virtual Environments*, 15(3), 353-357.



Congedo M, Lotte F, Lécuyer A (2006) Classification of Movement Intention by Spatially Filtered Electromagnetic Inverse Solutions, *Physics in Medicine and Biology*, 51, 1971-1989.

2005



Arrouët C, **Congedo M**, Marvie J-E, Lamarche F, Lécuyer A, Arnaldi B (2005) Open-ViBE: a 3D Platform for Real-Time Neuroscience, *Journal of Neurotherapy*, 9(1), 3-25.



Lécuyer A, Burkhardt JM, Le Biller J, **Congedo M (2005)**, A4: A Technique to Improve Perception of Contacts with Uder-Actuated Haptic Devices in Virtual Reality, *Proceedings of the World Haptics Conference, March 18-20, Pisa, Italia*, 316-322.



Sherlin L, **Congedo M (2005)** Obsessive Compulsive Dimension Localized using Low Resolution Electromagnetic Tomography (LORETA), *Neuroscience Letters*, 387(2), 72-74.

2004



Cannon R, Rothove J, Lubar JF, Thornton K, Wilson S, **Congedo M. (2004)**, Limbic Beta Activation and LORETA; can Hippocampal and Related Limbic Activity be Recorded and Changes Visualized using LORETA in an Affective Memory Condition?, *Journal of Neurotherapy*, 8(4), 5- 24.



Congedo M, Lubar JF (2004), Parametric and Non-Parametric Analysis of QEEG: Normative Database Comparisons in Electroencephalography, a Simulation Study on Accuracy. In *"Quantitative Electroencephalographic Analysis (QEEG) Databases for Neurotherapy. Description, Validation, and Application"*, (Ed) Lubar J.F., *Haworth Press, New York*, 1-29.



Congedo M, Lubar JF, Joffe D (2004), Low-Resolution Electromagnetic Tomography neurofeedback, *IEEE Trans. on Neuronal Systems & Rehabilitation Engineering*, 12(4), 387-397.

2003



Congedo M. (2003) Introducing the Logistic Discriminant Function in Electroencephalography, *Journal of Neurotherapy*, 7(2), 5-23.



Congedo M, Lubar JF (2003) Parametric and Non-Parametric Normative Database Comparisons in Electroencephalography: A Simulation Study on Accuracy, *Journal of Neurotherapy*, 7(3/4), 1-29.



Lubar JF **Congedo M, Askew JH (2003)** Low-Resolution Electromagnetic Tomography (LORETA) of Cerebral Activity in Chronic Depressive Disorder, *International Journal of Psychophysiology*, 49, 175-185.

2002



Congedo M, Ozen C, Sherlin L (2002), Notes on EEG Resampling by Natural Cubic Spline interpolation, *Journal of Neurotherapy*, 6(4), 73-80.

Grants

Legend:  *Grant Owner or Principal Investigator (Work Package Leader);*  *Investigator or other Minor Role*

Major Funded Projects



2012-13 RoBIK (Robust Brain-Computer Interface Keyboard)

Founder: Association Française contre les Myopathies – 97K€



2009-12 RoBIK (Robust Brain-Computer Interface Keyboard)

Founder: French National Agency of Research (TecSan) – 156K€



2009-12 Open-ViBE2 (Open Platform for Virtual Brain Environments)

Founder: French National Agency of Research (ContInt 2010-2013) – 122K€



2009-12 Gaze&EEG (Joint synch. EEG signal and ET process. for spatio-temporal analysis of neural activities)

Founder: French National Agency of Research (BLANC 2010-2013)



2006-09 Open-ViBE (Open Platform for Virtual Brain Environments)

Founder: French National Agency of Research (RNTL 2006-2009)

Minor Funded Projects



2011-13 Hyperscanning

Founder: Grenoble INP (Grenoble Institute of technology) – 13K€



2009-10 Independent Component Neurofeedback – 15K€

Founder: Tinnitus Research Initiative (TRI)



2005-08 COST (European Cooperation in the Field of Scientific and Technical Research) B27 Electric Neuronal Oscillations and Cognition (President: Prof. J. Pop-Jordanov),

Founder: European Framework Program 6

Teaching at Universities



Since 2010 **EDISCE** PhD School, Grenoble

Master 2 *Neuropsychologie et Neurosciences Cliniques*, “Real-Time Applications of EEG”, 2h



Since 2008 **EDISCE** PhD School, Grenoble

Master 2 *Sciences Cognitives*, “Real-Time Applications of EEG”, 2h



2011 **EDISCE** PhD School, CNRS, INSERM, Pôle Grenoble Cognition, Grenoble,

Formation en Neuro-Imagerie, “Tests de permutations”, 45m



2003 **University of Tennessee**, Knoxville, USA. Department of Psychology.

“*Statistics in Psychology*”, spring semester.



1999 **University of Bari**, ITALY. Department of Educational Science.

“*Social Psychology*”, spring semester.

Prizes, Recognitions, Awards



2007 Fellow of the International Society of Neurofeedback and Research



2003 Best PhD Thesis of the Department of Psychology, The University of Tennessee, Knoxville.

Media Coverage

Legend:  Press;  Interview;  Video Reportage

-  **2013 Jul 10** Interview on Brain-Computer Interface at Rhône Alpes Regional TV France 3 – [France 3](#), France
-  **2013 Mar 1** The BCI video game *Brain Invaders* is reported in the Journal of CNRS (#271, p. 13) [Journal du CNRS](#), FRANCE
-  **2013 Feb 5** Interview on Brain-Computer Interface at national French radio BFM – [BFM](#), France
-  **2013 Jan 23** Le Cerveau, future manette de jeux video, [Le Nouvel Observateur](#), FRANCE
-  **2013 Jan 22** Congedo M, Bouchet A, Lécuyer A. Neurofeedback : un traitement non invasif du trouble du déficit de l'attention. [Interstice](#), FRANCE
-  **2010 Oct 01** Quand le cerveau parle aux machines. *Doc Sciences (13)*, p. 40-47, France
-  **2008 Mai 26** Interview on BCI at national French radio “BFM” – [L'Atelier Numerique](#), France
-  **2008 Avr 29** Expert Opinion in Web Article - Les jeux vidéo peuvent-ils être prescrits par ordonnance ? [L'Atelier](#), FRANCE

Review Consulting

Funding Agencies



European Commission

- FP7 STREP Project BRAIN (2008-2011) – Interim Reviews
- FP7 CSA Project FUTURE- BNCI (2010-2012) – Interim Reviews
- FP7 STREP Project DECODER (2010-2013) – Interim Reviews
- FP7 STREP Project ABC (2012-2015) – Interim Reviews
- FP7 STREP Project BackHome (2012-2015) – Interim Reviews
- Two calls in ICT within FP7 (2011 and 2013) – Project Selection



French National Research Agency (ANR) – Project Selection



Natural Sciences and Engineering Research Council of Canada (NSERC) – Project Selection



Austrian Science Fund (FWF) – Project Selection



Dutch Ministry of Education, Culture and Science and the Ministry of Economic Affairs

Project BRAINGAIN (2008-2013) – Interim reviews

Scientific Editorial Journals



Clinical Neurophysiology
Computers in Biology and Medicine
Human Brain Mapping
IEEE Transactions on Signal Processing
IEEE Transactions on Biomedical Engineering
IEEE Transactions on Neural System and Rehabilitation Engineering
Journal of Neuronal Engineering
Journal of Neuroscience Methods
Journal of Neurotherapy
Journal of Statistical Planning and Interference
Medical & Biological Engineering & Computing
Neurocomputing
Neuroscience Letters
Nonlinear Biomedical Physics
Statistics in Medicine

Committees

Scientific



2011 **Formation en Neuroimagerie**, organized by EDISCE PhD School, CNRS, INSERM, Pôle Grenoble Cognition, from October 10 to October 14 (30 students).

PhD Theses Jury



2012 Sep 21 Candidate :Hayrettin Gürkök. University : University of Twente, THE NETHERLANDS.
Title : “ Mind the sheep ! User Experience Evaluation & Brain-Computer Interface Games “



2008 Mai 27 Candidate :Vincente Paquette. University : Université de Montreal, CANADA.
Title : “ L’effet de la psychoneurothérapie sur l’act. elect. du cerveau d’individus souffrant du trouble dépressif majeur unipolaire“ President of the Committee: Marc-André Bouchard.

Invited Lectures

At International Scientific Conferences



2009 Group Independent Component Analysis of Brain Resting-State Networks: Nearly Identical Findings on Two EEG Databases, *17th Annual Conference of the International Society for Neuronal Regulation*, Indianapolis, Indiana, USA.



2007 ICoN (Independent Component Neurofeedback), a freeware program for Blind Source Separation of continuous EEG and Extraction of Neurofeedback Weights. , *15th Annual Conference of the International Society for Neurofeedback and Research*, San Diego, CA, USA.



2004 EEG in Real Time: New Perspectives and a Platform for 3D Visualization of Functional Brain Dynamics, *12th Annual Conference of the Int. Society for Neuronal Regulation*, Fort Lauderdale, FL, USA.

Other Lectures



2012 New Developments and Advanced Methods in Neurofeedback, *The Future of Neurofeedback: insights from theory and practice Symposium*, Nijmegen, 28 Nov, THE NETHERLANDS.



2011 Il Cervello, meglio poco ma buono, *Elementary School "Re David"*, Bari, ITALY.



2011 Signal processing: BCI and the analysis of the Joint EEG of two individuals, *2nd Neurotherapy Symposium*, Zurich, SWITZERLAND.



2009 Lagged Connectivity of EEG Resting-State Independent Components, *Opening of the BRAPN Institute (Brain Research center Antwerp for Innovative and Interdisciplinary Neuromodul.)*, Antwerp, BELGIUM.



2007 Enhancing Neurofeedback by means of Multi-Channel Current Source Extraction Methods, *Prague Psychiatric Center (3rd Medical Faculty), Charles University*, Praha, CZECH REPUBLIC.



2006 Multichannel Neurofeedback; a family of methods to improve the specificity of neurotherapy, *Neuroscience Center of Zurich (ZNZ), Department of Psychology*, Zurich, SWITZERLAND.



2006 Linear Decomposition Methods for Real-Time Brain Electromagnetic Imaging, *MEG Center, University of Tuebingen*, Tuebingen, GERMANY.



2005 Recent Trends on Non-Invasive Self-Regulation of Brain Electrical Activity, *14^e Journées Scientifiques du Centre de Recherche en Neuropsychologie Expérimentale et Cognition (CERNEC), Department of Psychology, University of Montreal*, Montreal, CANADA.

Student Supervision

PhD



2012-15 Michael Acquadro, Grenoble University, Grenoble, (co-supervised with Prof. A. Guerin)

Title : TBA

Defense date : To be defended



2010-13 Jonas Chatel-Goldman, Grenoble University, (co-supervised with Prof. C. Jutten and Dr. J.-L. Schwartz)

Title : TBA

Defense date : To be defended



2009-12 Alexandre Barachant, Grenoble University, (co-supervised with Prof. C. Jutten and Dr. S. Bonnet)

Title : " Commande robuste d'un effecteur par une interface cerveau-machine EEG asynchrone "

Defense date : 28 mars 2012



2009-12 Sandra Rousseau, Grenoble University, (co-supervised with Prof. C. Jutten and Dr. J.-L. Schwartz)

Title : " Influence du retour sensoriel dans les ICM EEG: Etude du potentiel d'erreur "

Defense date : 16 Oct 2012



2008-13 David White, Swinburne UT, AUSTRALIA (co-supervised with Dr. J. Ciorciari and Prof. R. Silberstein)

Title : " An exploration of theta oscill. in the human EEG: Modulation via cognitive activity and real-time feedback"

Defense date : 13 Mars 2013 (no actual Defense in this University)



2008-12 Žaneta Kopřivová, Charles University, Prague, CZECH REPUBLIC (co-supervised with Dr. J. Horáček)

Title : " Functional-imaging and EEG correlates of OCD and their potential use in neurofeedback intervention "

Defense date : 03 Dec 2012.



2008-12 Elsa van der Loo, Antwerp Uni., BELGIUM (co-supervised with Prof. P. Van De Heyning and Prof. D. de Ridder)

Title : " The Neurology of Tinnitus Distress "

Defense date : 13 Mars 2012



2007-11 Sven Vanneste, Antwerp Uni., BELGIUM (co-supervised with Prof. P. Van De Heyning and Prof. D. de Ridder)

Title : " The Neural Correlates of Non-Pulsatile Tinnitus "

Defense date : Mars 14 2011



2006-09 Cédric Gouy-Pailler, PhD. Grenoble University, (co-supervised with Prof. C. Jutten)

Title: " Vers une modélisation dynamique de l'activité cérébrale EEG pour la conception d'ICM asynchrones "

Defense date : October 1 2009

Master 2



2011-12 Gijs Van Veen, University of Twente, The Netherlands (Visitor Student, co-supervised with Prof. M. Poel)

Title : " Brain Invaders; a BCI-controlled video-game "

Defense date : 22 Avril 2013.



2011-12 Ayoub Maatallaoui, INPG, Grenoble.

Title : " Watching a movie together: investigation of cospectra in simultaneous electroencephalographic data recording "

Defense date : June 28 2012.



2011-12 Michael Acquadro, INPG, Grenoble

Title : " New frontiers in neuroimaging: a study of several synchronous electroencephalographic recordings "

Defense date : June 28 2012.



2009-10 Soeun Somuny Oudom, INPG, Grenoble (co-supervised with Prof. C. Jutten and Prof. J.-L. Schwartz).

Title : " High Quality of Dual EEG Recording "

Defense date : June 25 2010.



2009-10 Léo Varnet, INPG, Grenoble (co-supervised with Dr. B. Rivet).

Title : " Mise en place d'un prototype de jeux vidéo par une interface cerveau machine pour OrangeLab "

Defense date : Octobre 18 2010.



2008-09 Guillaume Lio, INPG, Grenoble (co-supervised with Prof. C. Jutten).

Title : " Valorisation de l'activité cérébrale par sLORETA & séparation aveugle de sources pour le neurofeedback "

Defense date : June 23 2009.



2007-08 Romain Grandchamps, INPG, Grenoble (co-supervised with Prof. C. Jutten).

Title : " Extraction d'activité neuronale par filtrage spatial en temps réel: application au neurofeedback "

Defense date : June 23 2008.

Master 1



2009-10 Esteve Gallego, (ERASMUS), INPG, Grenoble (co-supervised with Prof. Prof. C. Jutten and Prof. J.-L. Schwartz).

Title : " Synchronous Electroencephalography (EEG) of two subjects: the analysis of their interaction "

Defense date : June 24 2010.



2009-10 Víctor Luis Viña Nogueiras, (ERASMUS), INPG, Grenoble (co-supervised with Prof. C. Jutten and Prof. J.-L. Schwartz).

Title : “ Synchronous Electroencephalography (EEG) of two subjects: the analysis of their interaction ”

Defense date : June 24 2010.

Undergraduate Internships



2008 Simon Rehn, INPG and Universität Karlsruhe (Germany), Grenoble (co-supervised with Prof. C. Jutten).

Title : “ A Normative database for EEG Based in Independent Source Analysis “

Defense date : No defense.

Before Holding a Permanent Position



2005 Fabien Lotte, M.S. at INRIA/INSA-Rennes, “ Classification De Données pour l’Utilisation Des Brain-Computer Interfaces en Réalité Virtuelle “



2004 Cédric Arrouët, M.S. at INRIA/INSA-Rennes, “ Activité Cérébrale et Réalité Virtuelle “.



2002-03 Rex Cannon, M.A. at University of Tennessee, “ The Effect of Neurofeedback Training in the Cognitive Division of the Anterior Cingulate Gyrus “.

WEB SITE

Since 2010 I am maintaining a permanent web page holding relevant information about my scientific activities. The page (<https://sites.google.com/site/marcocongedo/>) contains

- my CV
- a publication list and publication statistics
- a list of funded projects I have been working to
- my national and international collaborations
- a list of PhD students I have supervised
- a list of reviews and grant agency that have been consulting me
- my media interventions
- the tutorials I have written to introduce some of my research topics
- the stand-alone executable software I have made available to the public and the code of the approximate joint diagonalization methods I have proposed.

The site is constantly updated. The aim of this web site is to disseminate my work worldwide through a unique and permanent web space. The figure below (top graph) shows the monthly number of visit to the site in the period February 2010 - Mars 2013. The bottom part reports other statistics for the period, such as the total number of visits (4571), unique visitors (2534) page views (15656), average number of page visited (3.43), average visit duration (2.52 min) and bounce rate (49%). The percentage of unique visitors with respect of the total number of visitors is 55.3%.



Monthly visit to my home page in the period February 2010 - Mars 2013 and other statistics

(source: Google Analytics).

CURSUS

I have entered the CNRS (Centre National de la Recherche Scientifique) in 2007 as a CR1 (Research Scientist). The section of CNRS where I passed the public competition was an inter-disciplinary section. Indeed, merging knowledge from different disciplines and bridging upon them trying to connect researchers coming from different background has always been my natural tendency in science. My studies up until the PhD level have been focusing mainly on *psychology*.

My thesis at Padua University³ was on *experimental social psychology*. I investigated the formation of stereotypes, prejudices, and the bias in the perception and judgment of the external and internal social group. Merging knowledge in psychology and philosophy I elaborated an original theory of the general perception of the psychological objects, the theory of the “distance from the self”. For the experimental part I performed computerized experiments that I programmed myself, having learnt to program in Turbo Pascal during my ERASMUS at University Paris V René Descartes. Then, unsatisfied by the arbitrary method of measurements in social psychology, I switched field of interest completely, turning toward *electrophysiology*. I had discovered the *neurofeedback* technique and I was fascinated by the idea that the human mind may volitionally acquire some form of control over brain functioning.

In order to pursue this research I joined a worldwide renewed expert on neurofeedback: Prof. Joel Lubar, at the University of Tennessee, Knoxville (UTK), who has been a great mentor and to whom I own my entire attitude toward science. My Master thesis at UTK was on a new, non-parametric, way to build normative EEG databases (Congedo and Lubar, 2004). This work has proven very useful for my ensuing research on normative EEG databases based on blind source separation, published in Congedo, John, De Ridder and Prichep (2010). In the meanwhile I decided to focus on *real-time electroencephalography*, which was going to be the pivot of all my research activities since then and still to date. In 2000 I co-funded with another UTK student, Leslie Sherlin, a company providing software and services for research on EEG (Nova Tech EEG., Inc.). I left the company in 2007 as I entered the CNRS.

³ *Italian undergraduate program of Psychology at that time was five years (or more) long and at the end a true dissertation had to be defended.*

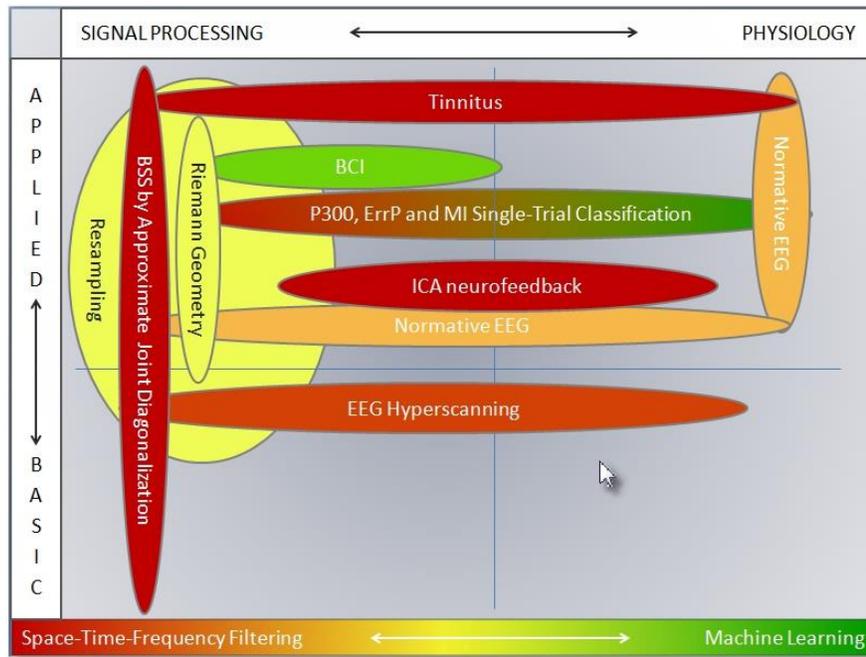
My PhD thesis at UTK was the first neurofeedback study using an EEG inverse solution to estimate on-line an EEG signal with higher spatial specificity (Congedo, Lubar and Joffe, 2004). This research was the first attempt to improve the neurofeedback method itself since its inception in the 60's. It can be considered pioneering, since several independent groups have adopted the method several years later, in Austria (Bauer et al., 2011), Germany (Salari et al., 2012), USA (Choi, 2014), Korea (Im et al., 2007), Switzerland (Liechti et al., 2012) and Turkey (Surmeli and Ertem, 2009).

While doing my PhD studies I started exploring *statistics*, especially *permutation tests* and object-oriented *computer programming* using the Borland Delphi 5 RAD (Rapid Application Development), to which by chance I was exposed during my visit in Zurich to Dr. Pascual-Marqui in 2001. That visit turned out to be fatal for my interest in source analysis methods. I own to Roberto Pascual-Marqui not only the first serious exposure to high-level computer code and mathematical knowledge, but also to an amazing passion for it.

While approaching complex data analysis methods to be used in electroencephalography, I felt the need to understand in depth the methodology I employed. For this reason at UTK I took also a minor PhD in statistics. Then I looked for a post-doc in a *virtual-reality* (VR) laboratory, foreseeing some of the recent trends that are now under the eyes of everybody. To me at that time VR technology was an interesting and powerful tool for the self-exploration of the brain functioning in real-time. This way I arrived for a post-doc at INRIA in Rennes, and then at INRIA in Grenoble, both in France, where I dedicated myself, among other things, to the study of *linear algebra*, a necessary brick to start investigating in depth analysis methods such as linear inverse solution, spatial filtering and blind source separation. I started writing my own linear algebra library in object Pascal, which still today is continuously updated and constitutes the basis of all my original research articles. The encounter with my supervisor in Rennes, Dr. Anatole Lécuyer, would be at the foundation of the first important French national grant on Brain-Computer Interfaces, the OpenViBE ANR project (2006-2009), which has had the merit of drawing much attention to this field in France and has brought several French research groups in this arena.

RESEARCH INTERESTS AND COLLABORATIONS

Because of this multi-disciplinary perspective, I have been collaborating with scientists in the medical and psychological field as well as with statisticians, software engineers and signal processing engineers. Thanks to these collaborations I have developed interest for a wide panel of research fields, certainly disparate, yet all covered under the umbrella of *real-time EEG*. The research fields where I have been active so far are schematically positioned in the figure below along three continua; see the caption for explanations.



Research Interests of the candidate: Each ellipse in the figure represents a research interest. They are schematically arranged on a two-dimensional continuum going from "Basic" to "Applied" (bottom to top axis) and from "Signal Processing" to "Physiology" (left to right axes). The color of each ellipse codes a third continuum depending whether the topic is attacked relying on a "Filtering" or "Machine Learning" approach (red to green). Yellow color codes approaches that do employ neither spatial filtering nor machine learning. Fields of research were both spatial filtering and machine learning are equally important are painted with a red-to-green gradient. High bubbles (e.g., the four on the left of the graph) indicate methods of data analysis; wide bubbles fields of application. Strong links between research interests are represented by overlapping bubbles.

Legend: BCI: Brain-Computer Interface; BSS: Blind Source Separation; EEG: Electroencephalography; ErrP: Error potential; MI: Motor Imagery;

Here below i briefly describe large research perimeters wherein I have been active, giving for each item the associated publications as they can be found in the CV above. Actually this description will serve to present my main scientific collaborations. The research interest are grouped in broad categories, so to highlight the relations among them and the multidisciplinary nature of the work undertaken⁴.

BCI, P300, ErrP and MI single-trial classification, SVM, Riemann Geometry.

Description: Brain-Computer Interface has been my central research interest from 2005 on. It was at the core of the research project presented to the CNRS at the time of my employment. It is also by far the research interest that has resulted in the highest scientific productivity since it has been supported by several research grants. Very recent developments are presented in chapter VIII and IX.

BCI may aim at partially restoring communication capabilities for people affected by severe motor impairment, thus specifically target the clinical population or may aim at enriching the normal communication pathways creating new interfaces, thus specifically targeting healthy people. A recent trend in BCI research is to integrate BCI commands in video-games and more in general to enrich recreational applications via BCI control. In fact, the actual use of BCI by motor-disabled people is problematic for several reasons, including specific cognitive and sensory disabilities jeopardizing the performance of the BCI in clinical population and making the transfer rate achievable by some patients so far non competitive for any practical purposes. Whenever residual motor ability is preserved, albeit minimal, patients usually prefer alternative communication devices such as simple switches and eye-trackers. However, the low transfer rate of a BCI is not a concern for recreational applications; as a consequence patients are usually willing to use BCI technology for recreational purposes such as video-games, painting applications, etc. On the other hand, healthy users can mobilize all available cognitive resources for the BCI therefore they easily achieve satisfactory performance. Among all healthy users, video-gamers are particularly motivated for trying new interface technology. Thus, a BCI video-game is an optimal choice to study the hardcore of a BCI, i.e., the interaction between the

⁴ For the citations in this section that cannot be found in the « Reference » section at the end of the manuscript the reader is kindly directed to the CV here above in this chapter.

interface and the whole signal processing chain, advancing the research toward practical applications that are likely to be used by both healthy users and people suffering from motor impairment.

Collaborations and Publications: This research has started within the ANR (national research) project OpenViBE (2006-2009), where I have investigated the classification of movement intention by a spatially filtered inverse solution (Congedo, Lotte and Lécuyer, 2006) and the use of data-driven and model driven subspace projection filters to increase the special sensitivity and specificity of an inverse solution when used on line (Congedo, 2006). This research is reminded in chapter III. These articles have been among the first to describe the use of EEG inverse solution for classifying BCI data. Project OpenViBE has been instrumental to develop the homonymous software platform for on line EEG analysis and visualization (Arrouët et al., 2005; Renard et al., 2010), which is since the basis of our research on BCI carried out at GIPSA-lab. The platform has been conceived with my arrival at INRIA as a post-doc supervised by Dr. Anatole Lécuyer in 2004. From 2006 to 2011 within the ANR project OpenViBE and OpenViBE2 we have designed and implemented the platform with the chief OpenViBE software engineer Yann Renard. In 2011 Yann left INRIA to create a spinoff company (Mensia Technologies⁵). From that moment OpenViBE has entered in a new phase of its life, becoming a self-sustained open-source platform with tens of developers in the community and with INRIA still continuously redesigning and improving the platform by means of other grants. Our recent developments for OpenViBE at GIPSA-lab are collected in the form of an open-source free add-on package⁶. Mensia technology is today using my free academic software to generate sLORETA (Pascual-Marqui 2002) and eLORETA (Pascual-Marqui, 2007) transformation matrices⁷ in some of their commercial service.

As a successful story, we may tell that OpenViBE has entered the White House in February 2012, when US President Barack Obama has hosted the second annual *White House Science Fair* to celebrate student winners of science, technology, engineering, and math (STEM) competitions from all over across the United States⁸. Anand Srinivasan, 15 y.o., whose “EEG & Prosthetics” project was

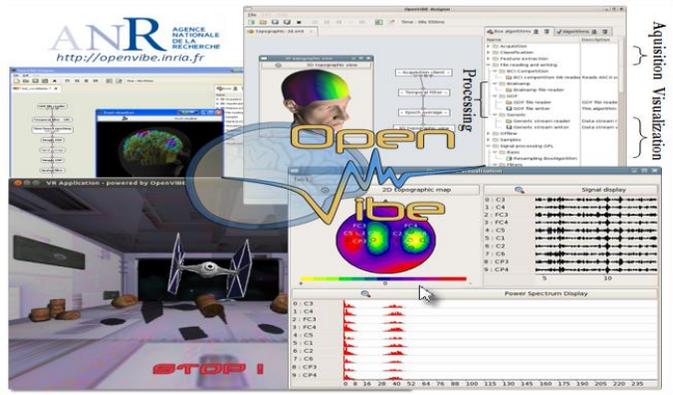
⁵ <http://www.mensiatech.com/>

⁶ <https://code.google.com/p/openvibe-gipsa-extensions/>

⁷ <https://sites.google.com/site/marcocongedo/software/ovtools>

⁸ e.g., <http://neurogadget.com/2012/02/15>

a finalist in the Google Science Fair 2011, impressed the president with a BCI project presented using the Emotiv EPOC headset and OpenViBE.



Open Virtual Brain Environments: OpenViBE is a powerful, flexible and modular software platform for the on-line acquisition, processing and visualization of EEG data. A complete BCI processing chain, from acquisition to visualization, is built as a sequence of “bricks”, very much like in popular software such as LabView and Simulink. While most bricks and the kernel are written in c++ for the sake of efficiency, new bricks can be developed in other languages such as

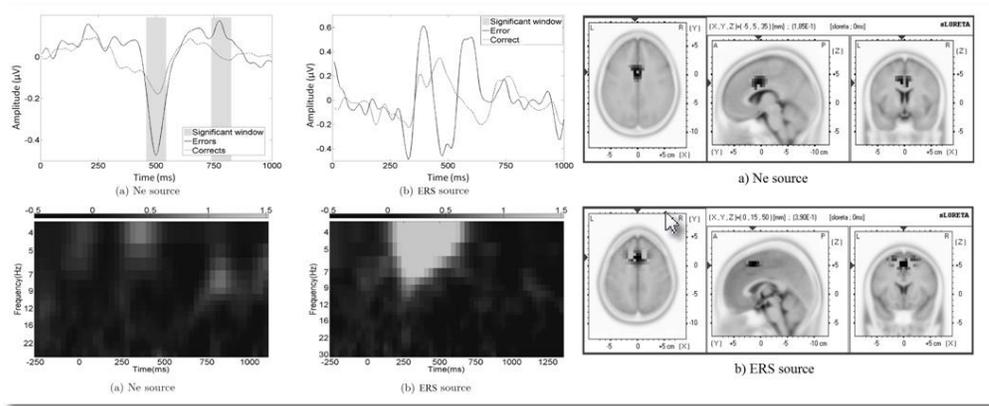
Matlab and Phyton for fast testing and prototyping. OpenViBE is fully compatible with virtual reality technology for advanced visualization. It is open source and free for any kind of use, including academic and commercial (<http://openvibe.inria.fr/>).

During the PhD thesis of Cédric Gouy-Pailler (2006-2009), co-supervised with Prof. Christian Jutten, blind source separation (BSS) strategies previously defined in Congedo, Gouy-Pailler and Jutten (2008) were used to improve single-trial detection of motor imagery over the state of the art common spatial patten method (Gouy-Pailler, Congedo, Jutten, Brunner and Pfurtscheller, 2008; Gouy-Pailler, Congedo, Brunner, Jutten and Pfurtscheller, 2008, 2010). Methods to smooth the BCI output (Gouy-Pailler, Mattout, Congedo and Jutten, 2009) and to remove ocular artifacts (Gouy-Pailler, Sameni, Congedo and Jutten, 2009) were also developed. This research is shown as example of BSS research in chapter VI.

From 2010 the work on BCI has continued within the (second) ANR project Open-ViBE2 (2010-2012), in collaboration with post-doc Nisrine Jrad. In this project it has been developed and tested a general support-vector machine (SVM) classification approach specifically adapted to EEG data. The method seeks the spatial filter for the data that optimizes the SVM cost function (the SVM finds in a high-dimensional space the hyperplane maximizing the margin with the two classes). The method effectively combines the search for the optimal spatial filter and optimal classifier in one algorithm, resulting in very good results that have been extensively documented in the case of P300 and Error-Related Potential data (Jrad and Congedo, 2011a, b; 2012; Jrad et al., 2012; Jrad Phlypo and Congedo, 2011; Phlypo et al., 2011). All this research is not included in this manuscript as we have found in the

meanwhile a more suitable classification framework that does not rely on cross-validation to set parameters.

With the PhD thesis of Sandra Rousseau (2010-2012), co-supervised with Prof. Christian Jutten, we have investigated the single-trial detection of Error-Related Potentials (ErrP: Congedo, Rousseau and Jutten, in press) Rousseau, Jutten and Congedo, 2012a, b, c, d). ErrPs are a family of event-related potential (ERP) that can be elicited after the commission of an error. When the feedback is given by the interface the ErrP is characterized both by a negative deflection (an ERP named Ne) and an event-related synchronization (ERS) in the theta band (4-7 Hz). Using the blind source separation framework defined in Congedo, Gouy-Pailler and Jutten (2008) we have been able to estimate simultaneously and separate the source responsible for the Ne and the source responsible for the ERS (Congedo, Rousseau and Jutten, 2014, in press). This research is reported in details in chapter VI as an example of BSS applied to extract ERPs and ERD/ERS.



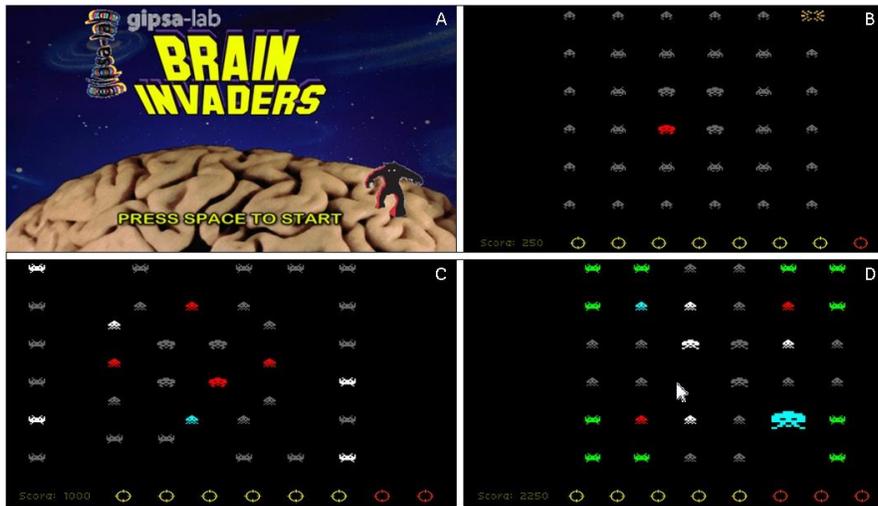
BSS analysis of Error Potentials: Two uncorrelated sources were identified, one responsible for the Ne and one for the ERS. The grand-average ($M=19$) ERP of the Ne source (spatial filter) computed separately for error and correct trials is displayed in the top row (a). The ERP in the top row (b) is obtained using the spatial filter of the ERS source; although differences in amplitude between error and correct trials exist also for this latter source, they are not significant. The ERS generated by the ERS source for error trials is shown in the bottom row (b). The difference as compared to the ERS in the correct trial is significant. The ERS in the bottom row (a) is obtained using the spatial filter of the Ne source on the error trials; the ERS in this case disappears and is no longer significant as compared to the correct trials. On the right part of the figure the three slices (horizontal, sagittal and coronal) through the grand-average ($M=19$) maximal sLORETA current density localizing the Ne source (top) and of the ERS source (bottom) as estimated by BSS. Legend: A=anterior; P=posterior.

In parallel to OpenViBE2, I participated to another ANR project on BCI (RoBiK, 2010-2012) (Mayaud, Congedo, Filipe et al., 2011), seconded by another homonymous grant given by the French Association for Myopathy (2012-2013). Both of the RoBiK grants focus on the clinical applications of P300-based BCIs and on the development of convenient BCI hardware for BCI (small wireless amplifiers and a new headset). The main collaborator for the period 2010-2011 has been post-doc Hubert Cecotti, who has carried out work on optimal sensor selection (Cecotti, Rivet, Congedo et al., 2010; Cecotti et al., 2011) and optimal time segmentation (Cecotti, Phlypo, Rivet et al., 2010) for P300 spatial filtering. This research is not reported in this manuscript since we have found a more suitable classification framework that does not rely on spatial filtering.

Since 2012 the main collaborator has been post-doc Alexandre Barachant, who I had previously co-supervised as a PhD student together with Prof. Christian Jutten and Dr. Stephan Bonnet. The work carried out with Alexandre has represented a major breakthrough in our BCI research group, leading to a universal framework for all kinds of BCI using the Riemannian geometry. The work of Alexandre carried out during the PhD (2009-2011) concerned the definition of the method for motor imagery BCI (Barachant, Bonnet, Congedo and Jutten 2010a, 2010b; 2011a, b; 2012a, b; 2013). The work carried out with us as a post-doc (2012-2013) has extended the Riemannian framework to SSVEP-based and P300-based BCIs (Barachant et al., submitted, in press; Barachant and Congedo, in press; Congedo and Barachant, submitted). The robustness of the Riemannian framework has allowed the initialization of the BCI with generic parameters derived from a database of other users. Then, an adaptive algorithm learns the optimal parameters for the user, effectively *by-passing completely the calibration phase*. In addition to this work, an automated on-line artifact-rejection algorithm has also been developed (Barachant, Andreev, Congedo, 2003). Taken together, this work has completely overridden our previous approaches to BCIs; the Riemannian method is more accurate, it generalizes well across subjects and across sessions, thus making the calibration unnecessary, and is even simpler algorithmically as compared to all our previous attempts based on sharp spatial filtering or machine learning. This research will be presented in details in Chapter VIII.

The two RoBiK projects yielded also the employment of a Anton Andreev as a software engineer (2012-2013). In 2013 Anton has integrated our team thanks to a full permanent position at CNRS. Anton currently maintains our code and technically supports experimentation, constituting a precious resource for the whole team.

In 2011 we have published our prototype of the *Brain Invaders*, a very effective pure-BCI video-game inspired from the vintage game *Space Invaders* (Congedo et al., 2011). During spring and summer 2012 a research project on BCI games has been carried out with a master student, Gijs Van Veen, in collaboration with Prof. Mannes Poel of Twente University (THE NETHERLANDS). With this student during 2012 the gameplay of the *Brain Invaders* has been improved considerably and Anton Andreev has made the code available as an open-source project of our team, along with our own extensions of the Open-ViBE platform that allows using the signal processing chain that has resulted from our research. An extensive experimental study using the *Brain Invaders* BCI is presented in chapter VIII.



The Brain Invaders: As any old-fashioned video game the *Brain Invaders* proceeds by levels. To finish a level the user must destroy a target alien, chosen at random within a grid of 36 aliens and which is indicated by a red circle at the beginning of the level. Aliens may be of different color. The target alien is always red. Aliens move with

patterns that are specific to each level. As in the P300 spellers a repetition of flashes consists in 12 flashes of groups of 6 aliens chosen in such a way that after repetition each alien has flashed two times. After each repetition the system assigns to each alien the probability of being the target according to the signal processing and classification method implemented in the OpenViBE platform and destroys the alien with the highest probability (b). If this alien is the target the level ends, otherwise this alien is eliminated and another repetition of flashes starts. The process is continued until the target alien is destroyed or until 8 non-target aliens have been destroyed, after which another level starts. The current number of attempts per level is indicated by coloring the bullets on the bottom of the screen (a),(b),(c). Between two levels the points obtained in the last level and the cumulative score are shown to the player. The points obtained at each level are inversely proportional to the number of repetitions necessary to destroy the target. Figure a) shows the welcome screen. (b) shows the simplest level, in which the aliens move altogether from the left to the right of the screen as in the original game *Space Invaders*. (c) and (d) shows more complex levels, where aliens move according to elaborated patterns and several aliens are colored green or red, like the target.

ICA Neurofeedback

Description: My PhD thesis was about the first neurofeedback study based on real-time estimation of intra-cranial current density by means of an inverse solution (Congedo, Lubar and Joffe, 2004). This work has been reiterated later on by other research groups (Bauer, Pllana and Sailer, 2011; Liechti et al., 2012; Surmeli and Ertem, 2009). In the meanwhile we have investigated the idea of using blind source separation to design a filter with higher spatial specificity. In fact, current density estimation is focalized on the current flowing in all directions in a particular location. On the other hand a blind source separation filter is optimal for estimating the direction of current flowing, thus it yields a sharper filter.

Collaborations and Publications: This research has been carried out with PhD students Zaneta Koprivova in Praha (CZECH REPUBLIC), co-supervised with Dr. J. Horáček (Kopřivová et al., 2011, 2013) and David White in Melbourne (AUSTRALIA), co-supervise with Prof. Richard Silberstein in Melbourne (White et al., 2012; White, 2012, unpublished PhD dissertation at Swinburne University, Melbourne).

EEG Hyperscanning and Approximate Joint Diagonalization

Description: This is the most recent of our research interests and has acquired a growing importance in our activity since 2010. EEG Hyperscanning is a particular instance of data fusion where two or more individuals are scanned simultaneously and synchronously. We are interested in EEG hyperscanning to study the possible synchronization of the brains of two individuals, that is, whether corresponding areas of the two brains may start working in phase synchrony under special circumstances. This topic has appeared recently in the EEG literature and has soon gained worldwide attention.

Many spatial filtering approaches in EEG work jointly diagonalizing two or more matrices describing relevant aspects of the spatial covariance structure. When more than two matrices are to be diagonalized simultaneously neither a closed form nor an exact solution exists in general. We employ

then approximate joint diagonalization (AJD) iterative algorithms. We are interested in obtaining robust methods obtained by jointly diagonalizing many symmetric matrices. More recently, we have attacked the problem of joint blind source separation, which involves the simultaneous joint diagonalization of several matrix sets, notably with applications in data integration and data fusion modalities such as EEG hyperscanning.

Collaborations and Publications: This research is currently carried out with PhD student Jonas Chatel-Goldman in Grenoble, that I co-supervise with Prof. Christian Jutten and Dr. Jean-Luc Schwartz (Chatel-Goldman, Schwartz, Jutten and Congedo, 2013; Chatel-Goldman, Congedo and Phlypo, 2013) and PhD student Michael Acquadro, co-supervised with Prof. Anne Guerin. Collaborations on the methodological aspects of EEG hyperscanning data analysis have included Prof. Emeritus D-T Pham at CNRS, Grenoble (Congedo and Pham, 2009; Congedo, Phlypo and Pham, 2011; Pham and Congedo, 2009) and Ronald Phlypo (Congedo, Phlypo and Chatel-Goldman, 2012; Phlypo and Congedo (2010), a former post-doc currently working at the MLSP lab of the University of Maryland. During the spring and summer 2012, a research project on hyperscanning has been carried out with two master students, Michael Acquadro and Ayoub Maatallaoui. Previously, in academic year 2009-2010 I have supervised the work of two Spanish ERASMUS students on this subject. The methodological aspect of this research concerns the analysis of the data acquired at the same time on two or more individuals. The innovations proposed stems from our advances on blind source separation (BSS) and approximate joint diagonalization algorithms, which generalization to more than one dataset (Joint Blind Source Separation: JBSS) constitutes the core of the methodology. This research will be presented in chapter IV and VII.

Normative EEG Database

Description: using a large database of healthy individual EEGs it is possible to derive norms for a large number of EEG features such as scalp or current density power, coherence, etc. (Ahn et al., 1980; John et al., 1980a, b, c). These norms serve as an aid in the diagnosis of clinical disorders. We have derived EEG norms based on a group BSS analysis. Such method allows to derive a more compact set of uncorrelated features, with expected improvement in the sensitivity and specificity.

Collaborations and Publications: this research is carried out in collaboration with Dirk de Ridder, formerly in Antwerp (BELGIUM) and now at University of Otago (NEW ZEALAND) and Leslie Prichep in New York (USA). I have been working closely on this subject with, and have been inspired by, senior researcher E. Roy John in New York, who I visited several time at New York University and School of Medicine before he disappeared in 2009. To Roy, a wonderful man, goes all my friendship and esteem. Our main paper describing the method is Congedo et *al.* (2010a). A complement can be found in Congedo et *al.* (2010). This research will be presented in chapter VII.

Tinnitus

Description:Tinnitu is the perception of sound within the human ear in the absence of corresponding external sound. Tinnitus is a complex clinical condition currently under investigation by the TRI (Tinnitus Research Initiative). We are interested in characterizing the neuronal correlates in order to design appropriate treatment strategies, especially those based on neurofeedback and neurostimulation.

Collaborations and Publications: this research is carried with Dirk de Ridder since 2006, formerly located in Antwerp (BELGIUM) and now at University of Otago (NEW ZEALAND). With him i have co-supervised the work of three PhD students in Antwerp (Elsa van Der Loo, Svan Vanneste and Mark Plazier), collaborating on several publications (De Ridder, Vanneste and Congedo, 2011; Vanneste et *al.*, 2010; Van der Loo et *al.*, 2007, 2009; Vanneste, Congedo, De Ridder, 2013). Some of the research carried out on Tinnitus will be presented in chapter VI.

SOFTWARE DEVELOPMENT

I have produced software for EEG data analysis since the beginning of the PhD at University of Tennessee in 2000. At that time I was programming in Borland Turbo Pascal 7. Starting 2001 I have switched to the much more productive Borland Delphi 5 rapid application development, based on the object Pascal language. I have written my own linear algebra library and some graphical objects for EEG data plotting. This software development environment is behind most of my publications. I have written many applications at the usage of my PhD students. Some of them have been released for free to the public.

In the period 2000-2003 I have written software for company Nova Tech EEG, Inc. This software has been freely distributed since 2005. Noteworthy from this period are the two main modules of the NTE pack:

Eureka3!

Continuously recorded (resting-state) EEG data analysis, featuring FFT spectral analysis both in the sensor space and in the sLORETA source space (Pascual-Marqui, 1999; Pascual-Marqui et al., 2002), see also (Bosch-Bayard et al., 2001).

MHyT

Multiple Hypothesis statistical testing for EEG group data (t-tests between, within and to compare the mean to a population value, correlations), adapted for the analysis of data in the frequency domain and both for the sensor space and the source space. The software is very flexible and implements state of the art multiple comparison p -min random permutation tests (Westfall and Young, 2003).

More recent application developments that are freely distributed on my web site are:

NICA

NICA performs group Blind Source Separation (gBSS), build gBSS databases, compares individuals to the databases and performs gBSS statistical tests for two-group designs, both between subjects and within subjects. NICA computes and tests not only source power, but also source (lagged) coherence. This application features very fast gBSS computations, powerful statistical tests (permutation t-max tests) and convenient plots of the results.

Working Memory Trainer

This is a program for training the working memory. The WM Trainer looks and behaves a little bit like a video-game and has been specifically conceived for children attending the primary school. However, it can be used purposefully by people of any age, including adult and elderly. This application features highest graphic quality, a powerful adaptive engine for the difficulty level, a database of users and statistical tools to evaluate the progress. Currently English, French and Italian are supported, but any language can be easily supported. An adaptation of this software has been used by PhD student Sandra Rousseau to carry out research on error potentials.



Screenshots of the WM Trainer

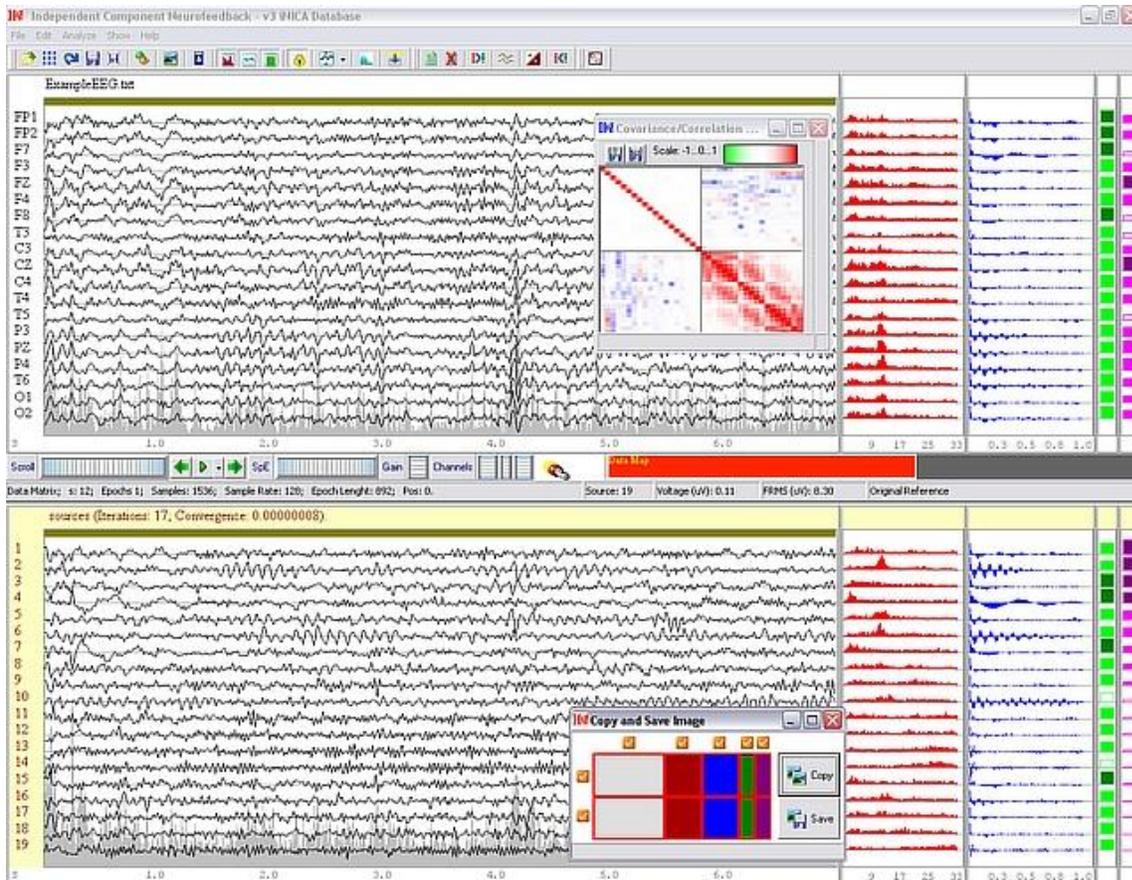
FDR_w False Discovery Rate (weighted)

Simple and efficient, this application performs the Weighted False Discovery Rate procedure of Benjamini and Hochberg (1997) to correct for multiple testing. It allows testing virtually any number of p -values obtained with any test-statistics for any data set. It also allows assigning a-priori weights to

give a better chance to those variables that are deemed important. In practice, this procedure is powerful only with a relatively small number of p -values.

ICoN (Independent Component Neurofeedback)

It is a program for off-line BSS (Blind Source Separation) based on second-order statistics. It has been specifically conceived for EEG data and it is fully automated. ICoN shows on the same screen the original data (top of the figure) and the source components (bottom of the figure), along with their Fourier spectra, Autocorrelation and Hurst exponent. The estimated sources are localized in brain-space using the oldest version of LORETA-Key software, but supports sLORETA and eLORETA as well. ICoN exports the demixing vector to be used as a spatial filter in a program such as Open-ViBE. ICoN implements the AJDC BSS algorithm described in details in chapter VI.



Screenshots of ICoN

CHAPTER II

BACKGROUND MATERIAL

NOTATION AND NOMENCLATURE

A	an <i>Integer</i>
N	<i>Number of Channels</i> in EEG recordings
$P \leq N$	<i>Number of Sources or Components</i> in linear transformations
M	<i>Number of Subjects or Data Sets</i>
Q	<i>Number of Voxels</i> for tomographic inverse solutions
Z	<i>Number of Classes</i> in classification tasks
K	<i>Number of Matrices or Samples</i> in a generic set
F	<i>Number of Frequencies</i> in <i>Fourier Analysis</i>
\mathfrak{R}	The <i>Set of Real Numbers</i>
$a \in \mathfrak{R}$	a real <i>scalar</i> or <i>random variable</i> , but also an <i>index</i>
$a \in \{1, \dots, A\}$	a set of indexes running from 1 to A
r, c	<i>row</i> and <i>column index</i> for matrix entries
$\mathbf{a} \in \mathfrak{R}^N$	an N -dimensional <i>column vector</i>
$\mathbf{a}^T \in \mathfrak{R}^N$	an N -dimensional <i>row vector</i>
a_n	the n^{th} entry of vector \mathbf{a}
$\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_C) \in \mathfrak{R}^{R \times C}$	a <i>matrix</i> of dimension $R \times C$, with $\mathbf{a}_c \in \mathfrak{R}^R$ its c^{th} column vector
$\mathbf{A}^T = (\mathbf{a}_1, \dots, \mathbf{a}_C)^T \in \mathfrak{R}^{C \times R}$	matrix <i>transposition</i>
a_{rc}	the (r, c) entry of matrix \mathbf{A}
$(\mathbf{A}_1, \dots, \mathbf{A}_M) \in \mathfrak{R}^{R \times CM}$	matrix <i>partition</i>

$\begin{pmatrix} \mathbf{A}_{11} & \cdots & \mathbf{A}_{1J} \\ \vdots & \ddots & \vdots \\ \mathbf{A}_{I1} & \cdots & \mathbf{A}_{IJ} \end{pmatrix} \in \mathfrak{R}^{R \times C \times J}$	another notation for matrix <i>partition</i>
$\mathbf{A}_k \in \{\mathbf{A}_1, \dots, \mathbf{A}_K\}$	a <i>set</i> of K matrices indexed by $k \in \{1, \dots, K\}$
$\mathbf{A}^+ \in \mathfrak{R}^{C \times R}$	matrix <i>pseudo-inverse</i>
$\text{rank}(\mathbf{A})$	<i>rank</i> of a matrix
$\text{tr}(\mathbf{A}) = \sum_{r=c} a_{rc}$	<i>trace</i> of a matrix
$\ \mathbf{A}\ _F = \text{tr}^{1/2}(\mathbf{A}^T \mathbf{A}) = \left(\sum_{r=c} a_{rc}^2 \right)^{1/2}$	<i>Frobenius norm</i> of a matrix
$\det(\mathbf{A})$	<i>determinant</i> of a matrix
$\mathbf{1} \in \mathfrak{R}^N$	the <i>unit</i> vector, one in all entries
$\mathbf{0} \in \mathfrak{R}^{R \times C}$	the <i>null</i> matrix, zero in all entries
$\mathbf{I} \in \mathfrak{R}^{N^2}$	the <i>identity</i> matrix (of dimension $N \times N$)
$\mathbf{H} = \mathbf{I} - \frac{1}{N} \mathbf{1} \mathbf{1}^T \in \mathfrak{R}^{N^2}$	the <i>centering</i> matrix
$\mathbf{Q} \in \mathfrak{R}^{N^2}$	a <i>square</i> matrix
$\mathbf{L} \in \mathfrak{R}^{N^2}$	a <i>lower triangular</i> matrix
$\mathbf{S} \in \mathfrak{R}^{N^2}$	a <i>symmetric</i> matrix
$\mathbf{D} \in \mathfrak{R}^{N^2}$	a <i>diagonal</i> matrix
d_n, d_m	two ways to denote the n^{th} <i>diagonal</i> element of \mathbf{D}
$\mathbf{U}, \mathbf{V} \in \mathfrak{R}^{N^2}$	<i>orthogonal</i> matrices, <i>eigenvector</i> and <i>singular vector</i> matrices

A	<i>diagonal matrix of eigenvalues or singular values</i>
$\lambda_n(S)$	n^{th} <i>eigenvalue</i> of matrix S
$\text{diag}(Q)$	matrix Q with off-diagonal elements nullified
$\text{off}(Q)$	matrix Q with diagonal elements nullified
$C \in \mathfrak{R}^{N^2}$	a <i>symmetric positive-definite</i> (SPD), <i>covariance</i> or <i>Cospectra</i>
$C^{-1} \in \mathfrak{R}^{N^2}$	<i>symmetric inverse</i> , such that $C^{-1}C = CC^{-1} = I$
$C^{\frac{1}{2}} \in \mathfrak{R}^{N^2}$	<i>symmetric square root</i> , such that $C^{\frac{1}{2}}C^{\frac{1}{2}} = C$
$C^{-\frac{1}{2}} \in \mathfrak{R}^{N^2}$	<i>symmetric square root inverse</i> , such that $C^{-\frac{1}{2}}C^{\frac{1}{2}} = C^{-\frac{1}{2}}CC^{-\frac{1}{2}} = I$
$W \in \mathfrak{R}^{R \times C}$	a <i>whitening</i> matrix
$P \in \mathfrak{R}^{N^2}$	a <i>permutation</i> matrix, the identity matrix with the rows or column shuffled
$E_n \in \mathfrak{R}^{N^2}$	the <i>elementary diagonal</i> matrix, $e_{nn}=1$, 0 elsewhere
$A \leftarrow B$	affectation: B is written into A
$A_{(\cdot)}$	A matrix depending generically on the argument in parentheses
$a_{n(\cdot)}$	The n^{th} vector of matrix A depending as above
$\delta(C_1 \leftrightarrow C_2)$	A symmetric distance or divergence between matrix C_1 and C_2
$\sum_n a_n$	The <i>sum</i> $a_1 + \dots + a_N$, short for $\sum_{n=1}^N a_n$
$\prod_n a_n$	The <i>product</i> $a_1 \dots a_N$, short for $\prod_{n=1}^N a_n$
$\forall n$	<i>for all</i> $n \in \{1, \dots, N\}$

<i>iff</i>	if and only if
\ll	<i>much smaller than</i>
<i>iid</i>	<i>independently and identically distributed</i>
SNR	<i>signal to noise ratio</i>
Hz	<i>Hertz</i> (cycles per seconds)
SPD	<i>Symmetric Positive-Definite</i> (matrix)
$\mathbb{E}[x]$	statistical <i>expectation</i> of random variable x
$\sim N(\mu, \sigma^2)$	distributed as a <i>Normal</i> (Gaussian) with mean μ and var σ^2
$\sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	distributed as a <i>Multivariate Normal</i> with mean vector $\boldsymbol{\mu}$ and Wishart matrix $\boldsymbol{\Sigma}$
$\sim \ln N(\mu, \sigma^2)$	distributed as a <i>log-Normal</i> , that is, such that its exponential is distributed $\sim N(\mu, \sigma^2)$
a_x, g_x, h_x	<i>Arithmetic, Geometric, Harmonic mean</i> of random variable x

LINEAR ALGEBRA

Invariances

- (1) Two square matrices are *similar* if an invertible matrix \mathbf{B} exists such that $\mathbf{Q}_1 = \mathbf{B}^{-1}\mathbf{Q}_2\mathbf{B}$ (Schott, 1997, p. 152). The product $\mathbf{B}^{-1}\mathbf{Q}_2\mathbf{B}$ is named *similar transformation*. A function or a property of a matrix $f(\mathbf{Q})$ is said *similarity-invariant* if $f(\mathbf{Q}) = f(\mathbf{B}^{-1}\mathbf{Q}\mathbf{B})$.
- (2) Two square matrices are said *congruent* if a matrix \mathbf{A} exists such that $\mathbf{Q}_1 = \mathbf{A}^T\mathbf{Q}_2\mathbf{A}$. The product $\mathbf{A}^T\mathbf{Q}_2\mathbf{A}$ is named *congruent transformation* or *conjugation*. A function or a property of a matrix $f(\mathbf{Q})$ is said *congruence-invariant* if $f(\mathbf{Q}) = f(\mathbf{A}^T\mathbf{Q}\mathbf{A})$.
- (3) Two square matrices are said *orthogonally congruent* or that they are one the *rotation* of the other if an orthogonal matrix \mathbf{U} exists such that $\mathbf{Q}_1 = \mathbf{U}^T\mathbf{Q}_2\mathbf{U}$. The product $\mathbf{U}^T\mathbf{Q}_2\mathbf{U}$ is named an *orthogonal transformation* or *rotation* (see Schott, 1997, p. 60). A function or a property of a matrix $f(\mathbf{Q})$ is said *rotation-invariant* if $f(\mathbf{Q}) = f(\mathbf{U}^T\mathbf{Q}\mathbf{U})$. Congruence invariance (2) implies rotation invariance.

The Rank of a Matrix

- (4) The *rank* of a matrix $\mathbf{A} \in \mathfrak{R}^{R \times C}$ is the number of linearly independent columns or, equivalently, the number of linearly independent rows.
- (5) A square matrix \mathbf{Q} is *invertible* if it is *full rank*, meaning that the rank is equal to the dimension of the matrix. A full-rank symmetric matrix \mathbf{C} has all positive eigenvalues.
- (6) $\text{rank}(\mathbf{A}) \leq \min(R, C)$
- (7) $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}^T) = \text{rank}(\mathbf{A}\mathbf{A}^T) = \text{rank}(\mathbf{A}^T\mathbf{A})$

(8) $\text{rank}(\mathbf{AB}) \leq \min(\text{rank } \mathbf{A}, \text{rank } \mathbf{B})$, for any $\mathbf{B} \in \mathfrak{R}^{C \times Q}$

The Trace of Square Matrix $\mathbf{Q} \in \mathfrak{R}^{N \times N}$

(9) $\text{tr } \mathbf{Q} = \text{tr } \mathbf{Q}^T$

(10) $\text{tr } c\mathbf{Q} = c \cdot \text{tr } \mathbf{Q}$

(11) $\text{tr}(\mathbf{Q}_1 + \mathbf{Q}_2) = \text{tr } \mathbf{Q}_1 + \text{tr } \mathbf{Q}_2$

(12) Trace of the product of two matrices: $\text{tr}(\mathbf{Q}_1\mathbf{Q}_2) = \text{tr}(\mathbf{Q}_1^T\mathbf{Q}_2^T) = \text{tr}(\mathbf{Q}_2\mathbf{Q}_1) = \text{tr}(\mathbf{Q}_2^T\mathbf{Q}_1^T)$

(13) Trace of the product of more than two matrices (cyclic property):

$$\text{tr}(\mathbf{Q}_1\mathbf{Q}_2\mathbf{Q}_3\mathbf{Q}_4\dots) = \text{tr}(\mathbf{Q}_2\mathbf{Q}_3\mathbf{Q}_4\dots\mathbf{Q}_1) = \text{tr}(\mathbf{Q}_3\mathbf{Q}_4\dots\mathbf{Q}_1\mathbf{Q}_2) = \text{tr}(\mathbf{Q}_4\dots\mathbf{Q}_1\mathbf{Q}_2\mathbf{Q}_3)$$

(14) Trace of the product of *three symmetric matrices*: any permutation is allowed

(15) The trace possesses the similarity and rotation invariance, but not the congruence invariance.

The Symmetric Matrix $\mathbf{S} \in \mathfrak{R}^{N \times N}$

(16) $\mathbf{S} = \mathbf{S}^T$

(17) \mathbf{S}^{-1} is symmetric

(18) $\mathbf{SS}^T = \mathbf{S}^T\mathbf{S}$ is symmetric

(19) $\mathbf{S}^{-1}\mathbf{S}^T = \mathbf{I}$

(20) The *sum* and *difference* of symmetric matrices is again symmetric

(21) The *product* of two symmetric matrices is symmetric only if they commute in multiplication.

For instance, for any positive integer P , \mathbf{S}^P is symmetric

(22) Every symmetric matrix is, up to a rotation (3), a *diagonal matrix* (see (37)).

- (23) If two symmetric matrices commute in multiplication, they can be *jointly diagonalized* by orthogonal transformation (3) or by similar transformation (1) of an invertible matrix (Schott, 1997, p. 154-157; Searle, 1982, p. 312).
- (24) For any matrix \mathbf{Q} , its *symmetric part* is given by $\frac{1}{2}(\mathbf{Q} + \mathbf{Q}^T)$ and its *anti-symmetric part* is given by $\frac{1}{2}(\mathbf{Q} - \mathbf{Q}^T)$. The two parts sum up to \mathbf{Q} .
- (25) Symmetry possesses the *congruence-invariance* (hence *rotation-invariance*) but *not the similarity-invariance*.
- (26) A *Symmetric* matrix is *Semi-Positive Definite* iff for any vector \mathbf{y} , $\mathbf{y}^T \mathbf{S} \mathbf{y} \geq 0$
- (27) It is *Positive Definite* if it is also invertible. In this case the equality in (26) holds only for $\mathbf{y}=\mathbf{0}$.

Symmetric Positive-Definite (SPD) Matrix $\mathbf{C} \in \mathfrak{R}^{N \times N}$

- (28) $\det \mathbf{C} > 0$
- (29) $\text{rank } \mathbf{C} = N$
- (30) $\mathbf{C}\mathbf{x} = \mathbf{0}$ holds only for $\mathbf{x} = \mathbf{0}$
- (31) All its eigenvalues are real and positive
- (32) If \mathbf{C}_1 and \mathbf{C}_2 are SPD matrices, so is $\mathbf{C}_1 + s\mathbf{C}_2$, for any $s > 0$.

Orthogonal Matrix $\mathbf{U} \in \mathfrak{R}^{N \times N}$

A square matrix is *Orthogonal* if its transpose is its inverse, that is,

- (33) $\mathbf{U}\mathbf{U}^T = \mathbf{U}^T\mathbf{U} = \mathbf{I}$
- (34) The product of two orthogonal matrices is always orthogonal

Eigenvalue-Eigenvector Decomposition (EVD)

For a symmetric matrix $S \in \mathfrak{R}^{N \times N}$ and for a Positive-Definite matrix $C \in \mathfrak{R}^{N \times N}$ the EVD, also named spectral decomposition is

$$(35) \quad EVD(S) = \mathbf{U} \mathbf{A} \mathbf{U}^T = \sum_{n=1}^N (\lambda_n \mathbf{u}_n \mathbf{u}_n^T),$$

where $\mathbf{U} \in \mathfrak{R}^{N \times N}$ is the orthogonal matrix holding in columns the eigenvectors and $\mathbf{A} \in \mathfrak{R}^{N \times N}$ the diagonal matrix holding the eigenvalues. The eigenvalues are all real for S and also strictly positive for C . We have:

$$(36) \quad \mathbf{S} = \mathbf{U} \mathbf{A} \mathbf{U}^T; \quad \mathbf{S} \mathbf{U} = \mathbf{U} \mathbf{A}; \quad \mathbf{S} \mathbf{u}_n = \lambda_n \mathbf{u}_n, \quad \forall n$$

$$(37) \quad \mathbf{U}^T \mathbf{S} \mathbf{U} = \mathbf{A}$$

The sum of the first $P < N$ terms in the right-end side of (35) yields the matrix $\mathbf{S}_{(P)}$ of rank P closest to \mathbf{S} in the least-square sense, that is (Good, 1969, p. 827)

$$(38) \quad \mathbf{S}_{(P)} = \sum_{n=1}^P (\lambda_n \mathbf{u}_n \mathbf{u}_n^T) \text{ is the solution to the problem } \min_P \left\| \mathbf{S}_{(P)} - \mathbf{S} \right\|_2^F, \text{ with } \text{rank } \mathbf{S}_{(P)} = P.$$

$$(39) \quad \text{The minimum is attained at } \sum_{q=P+1}^N \lambda_q \text{ and is named } \textit{Representation Error}.$$

Properties of Eigenvalues

$$(40) \quad \sum_{n=1}^N \lambda_n = \text{tr } \mathbf{S}; \quad \sum_{n=1}^N \lambda_n^2 = \text{tr } \mathbf{S}^2 = \|\mathbf{S}\|_F^2; \quad \sum_{n=1}^N \lambda_n^k = \text{tr } \mathbf{S}^k$$

$$(41) \quad \prod_{n=1}^N \lambda_n = \det \mathbf{S}$$

$$(42) \quad \lambda_n(\mathbf{Q}) = \lambda_n(\mathbf{Q}^T), \quad \forall n$$

$$(43) \quad \lambda_n(\mathbf{Q} \mathbf{Q}^T) = \lambda_n(\mathbf{Q}^T \mathbf{Q}), \quad \forall n$$

$$(44) \quad \lambda_n(\mathbf{B}^T \mathbf{S} \mathbf{B}) = \lambda_n(\mathbf{S}), \forall n, \text{ for any invertible } \mathbf{B}.$$

For both a diagonal matrix \mathbf{D} and a lower triangular matrix \mathbf{L} the diagonal elements are the eigenvalues:

$$(45) \quad l_n = \lambda_n(\mathbf{L}), \quad d_n = \lambda_n(\mathbf{D})$$

(46) The eigenvalues of an orthogonal matrix \mathbf{U} equal either 1 or -1. With an orthogonal transformation of the data, the eigenvectors with associated eigenvalue 1 determine a *rotation* of the axes, the eigenvectors with associated eigenvalue -1 determine a *rotation and a reflection* of the axes.

The following is known as the basic result of the *extremal properties of eigenvalues* (Schott, 1997, p. 104-128). Let $\lambda_1 \geq \dots \geq \lambda_N$ be the eigenvalues of \mathbf{S} , then

$$(47) \quad \min_{\mathbf{u} \neq 0} \frac{\mathbf{u}^T \mathbf{S} \mathbf{u}}{\mathbf{u}^T \mathbf{u}} = \lambda_N \quad \text{and} \quad \max_{\mathbf{u} \neq 0} \frac{\mathbf{u}^T \mathbf{S} \mathbf{u}}{\mathbf{u}^T \mathbf{u}} = \lambda_1 .$$

This normalized quadratic form is called the *Rayleigh quotient*.

Power Iterations

If the estimation of only one of the eigenvectors is sought one can use the *power method* or the *inverse power method* (Golub and Van Loan, 1996, p. 406; Strang, 2006, p. 359). For instance, the *principal eigenvector* \mathbf{u}_1 of positive-definite matrix \mathbf{C} , i.e., the eigenvector associated with its maximal eigenvalue, is obtained by the following iterative algorithm.

Algorithm (48): Power Iterations

Initialize \mathbf{u}_1 *with a clever guess or with a random unit norm vector*

Repeat

$$\mathbf{u}_1 \leftarrow \mathbf{C} \mathbf{u}_1$$

$$\mathbf{u}_1 \leftarrow \mathbf{u}_1 / \|\mathbf{u}_1\|_F$$

Until the angle between the new estimation and the preceding is smaller than a chosen ε

The convergence speed depends on the ratio between the first and second eigenvalue; the higher the ratio, the faster the convergence. The algorithm fails if the first two eigenvalues are too close to each other.

Cholesky Decomposition

For any *symmetric positive definite matrix* \mathbf{C} it exists a lower triangular matrix \mathbf{L} with all positive entries on the diagonal (positive eigenvalues, see (45)) such that

$$(49) \quad \mathbf{C} = \mathbf{L}\mathbf{L}^T.$$

Such matrix is unique (Schott, 1997, p. 147).

Operators on Symmetric Positive-Definite Matrices

Given \mathbf{C} symmetric positive definite, using $EVD(\mathbf{C}) = \mathbf{U}\mathbf{A}\mathbf{U}^T$ (35) we define the following operators using functions of eigenvalues:

$$(50) \quad \text{Symmetric Inverse} \quad \mathbf{C}^{-1} = \mathbf{U}\mathbf{A}^{-1}\mathbf{U}^T = \sum_n (\lambda_n^{-1} \mathbf{u}_n \mathbf{u}_n^T)$$

$$(51) \quad \text{Symmetric Square Root} \quad \mathbf{C}^{1/2} = \mathbf{U}\mathbf{A}^{1/2}\mathbf{U}^T = \sum_n (\lambda_n^{1/2} \mathbf{u}_n \mathbf{u}_n^T)$$

$$(52) \quad \text{Symmetric Square Root Inverse} \quad \mathbf{C}^{-1/2} = \mathbf{U}\mathbf{A}^{-1/2}\mathbf{U}^T = \sum_n (\lambda_n^{-1/2} \mathbf{u}_n \mathbf{u}_n^T)$$

$$(53) \quad \text{Symmetric Exponential} \quad \exp(\mathbf{C}) = \sum_{i=0}^{+\infty} \frac{\mathbf{C}^i}{i!} = \mathbf{U}e^{\mathbf{A}}\mathbf{U}^T = \sum_n (e^{\lambda_n} \mathbf{u}_n \mathbf{u}_n^T)$$

$$(54) \quad \text{Symmetric Logarithm} \quad \ln(\mathbf{C}) = \sum_{i=0}^{+\infty} \frac{(-1)^i}{i} (\mathbf{C} - \mathbf{I})^i = \mathbf{U} \ln(\mathbf{A}) \mathbf{U}^T = \sum_n (\ln(\lambda_n) \mathbf{u}_n \mathbf{u}_n^T)$$

These functions act in analogy with the algebraic counterpart, for example, all the following are easy to verify:

$$(55) \quad \mathbf{C}^{1/2} \mathbf{C}^{1/2} = \exp(\ln \mathbf{C}) = \ln(\exp \mathbf{C}) = \mathbf{C} ;$$

$$(56) \quad \mathbf{C}^{-\frac{1}{2}}\mathbf{C}^{-\frac{1}{2}} = (\mathbf{C}^{\frac{1}{2}}\mathbf{C}^{\frac{1}{2}})^{-1} = \exp(-\ln \mathbf{C}) = \mathbf{C}^{-1};$$

$$(57) \quad \mathbf{C}^{-\frac{1}{2}}\mathbf{C}\mathbf{C}^{-\frac{1}{2}} = \mathbf{C}^{-\frac{1}{2}}\mathbf{C}^{-\frac{1}{2}}\mathbf{C} = \mathbf{C}\mathbf{C}^{-\frac{1}{2}}\mathbf{C}^{-\frac{1}{2}} = \mathbf{C}^{-\frac{1}{2}}\mathbf{C}^{\frac{1}{2}} = \mathbf{C}\mathbf{C}^{-1} = \mathbf{C}^{-1}\mathbf{C} = \mathbf{I}$$

(58) $\exp(\mathbf{S})$ is always SPD if \mathbf{S} is symmetric, that is, the operator \exp is a projector in the space of SPD matrices.

(59) For a positive definite diagonal matrix the diagonal elements are the eigenvalues, so all the above operators (50) to (54) just apply element-wise to the diagonal elements.

Some Results on Matrix Exponential and Logarithm

If \mathbf{D}_1 and \mathbf{D}_2 are positive-definite diagonal matrices, using properties of the logarithm we have

$$(60) \quad \left\| \ln(\mathbf{D}_2^{-1}\mathbf{D}_1) \right\|_F = \left\| -\ln(\mathbf{D}_1^{-1}\mathbf{D}_2) \right\|_F = \left\| \ln(\mathbf{D}_1^{-1}\mathbf{D}_2) \right\|_F = \left\| -\ln(\mathbf{D}_2^{-1}\mathbf{D}_1) \right\|_F$$

from

$$(61) \quad \sqrt{\sum_n \left(\ln \frac{d_{1n}}{d_{2n}} \right)^2} = \sqrt{\sum_n \left(-\ln \frac{d_{2n}}{d_{1n}} \right)^2} = \sqrt{\sum_n \left(\ln \frac{d_{2n}}{d_{1n}} \right)^2} = \sqrt{\sum_n \left(-\ln \frac{d_{1n}}{d_{2n}} \right)^2} .$$

(62) $\exp(\mathbf{C}_1)\exp(\mathbf{C}_2) = \exp(\mathbf{C}_1 + \mathbf{C}_2)$ and $\ln(\mathbf{C}_1)\ln(\mathbf{C}_2) = \ln(\mathbf{C}_1 + \mathbf{C}_2)$ iff \mathbf{C}_1 and \mathbf{C}_2 commutes in multiplication

$$(63) \quad \ln(\alpha \mathbf{I}) = \ln(\alpha) \mathbf{I}$$

$$(64) \quad \exp(\alpha \mathbf{I}) = \exp(\alpha) \mathbf{I}$$

$$(65) \quad \ln(\mathbf{C}) \text{ commutes with } [(\mathbf{C} - \mathbf{I})\alpha + \mathbf{I}]^{-1}, \alpha > 0$$

$$(66) \quad \det \exp(\mathbf{C}_1 + \mathbf{C}_2) = \det \exp(\mathbf{C}_1) \det \exp(\mathbf{C}_2)$$

$$(67) \quad \ln \det(\mathbf{C}) = \text{tr} \ln(\mathbf{C}); \det \exp(\mathbf{C}) = \exp(\text{tr} \mathbf{C})$$

$$(68) \quad \ln(\mathbf{B}^{-1}\mathbf{C}\mathbf{B}) = \mathbf{B}^{-1}\ln(\mathbf{C})\mathbf{B}; \exp(\mathbf{B}^{-1}\mathbf{C}\mathbf{B}) = \mathbf{B}^{-1}\exp(\mathbf{C})\mathbf{B}, \text{ for any invertible } \mathbf{B}.$$

$$(69) \quad \ln(\mathbf{C}) = -\ln(\mathbf{C}^{-1}); \ln(\mathbf{C}^{-1}) = -\ln(\mathbf{C})$$

$$(70) \quad \exp(\mathbf{C}) = [\exp(-\mathbf{C})]^{-1}$$

$$(71) \quad \mathbf{C}^k = \exp(k \cdot \ln \mathbf{C}), \text{ e.g., } \mathbf{C}^{-1} = \exp(-\ln \mathbf{C})$$

Other Results on SPD Matrices

$$(72) \quad \mathbf{B}^T \mathbf{S} \mathbf{B} = \mathbf{0} \text{ implies } \mathbf{S} \mathbf{B} = \mathbf{0}$$

Singular Value Decomposition (SVD)

There are several possible definitions for the *Singular-Value Decomposition* of a matrix $\mathbf{A} \in \mathfrak{R}^{R \times C}$. We use

$$(73) \quad \text{SVD}(\mathbf{A}) = \mathbf{U} \mathbf{A} \mathbf{V}^T,$$

where $\mathbf{U} \in \mathfrak{R}^{R \times R}$ is the orthogonal matrix holding in columns the left singular vectors, $\mathbf{V} \in \mathfrak{R}^{C \times C}$ the orthogonal matrix holding in columns the right singular vectors, $\mathbf{A} \in \mathfrak{R}^{R \times C}$ the matrix holding at entry $(1, 1), \dots, (P, P)$ the singular vectors and zeros elsewhere and $P = \min(R, C)$. We have

$$(74) \quad \mathbf{A} \mathbf{A}^T = \mathbf{U} \mathbf{A}^2 \mathbf{U}^T \text{ and } \mathbf{A}^T \mathbf{A} = \mathbf{V} \mathbf{A}^2 \mathbf{V}^T,$$

with \mathbf{A} in this case diagonal and of dimension $R \times R$ for the first expression and $C \times C$ for the second expression.

Lödwin Orthogonalization

Given a matrix \mathbf{A} with $\text{SVD}(\mathbf{A}) = \mathbf{U} \mathbf{A} \mathbf{V}^T$ (73), the matrix

$$(75) \quad \mathbf{Z} = \mathbf{U} \mathbf{V}^T = (\mathbf{A} \mathbf{A}^T)^{-\frac{1}{2}} \mathbf{A}$$

is the closest orthogonal matrix in the least-squares sense to \mathbf{A} (Carlson and Keller, 1957), i.e., it is the orthogonal matrix satisfying $\min_{\mathbf{Z}} \|\mathbf{A} - \mathbf{Z}\|_F$, with \mathbf{Z} orthogonal.

Moore-Penrose Pseudo-Inverse and Pseudo-Operators

For a given matrix \mathbf{A} , there is a unique *Pseudo-Inverse* \mathbf{A}^+ , satisfying the following four conditions:

$$(76) \quad \mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}, \quad \mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+, \quad \text{both } \mathbf{A}\mathbf{A}^+ \text{ and } \mathbf{A}^+\mathbf{A} \text{ are symmetric.}$$

The pseudo-inverse satisfies also

$$(77) \quad (\mathbf{A}^+)^+ = \mathbf{A}; \quad (\mathbf{A}^T)^+ = (\mathbf{A}^+)^T; \quad (c\mathbf{A})^+ = \frac{1}{c}\mathbf{A}^+; \quad (\mathbf{A}^T\mathbf{A})^+ = \mathbf{A}^+(\mathbf{A}^T)^+; \quad (\mathbf{A}\mathbf{A}^T)^+ = (\mathbf{A}^T)^+\mathbf{A}^+$$

Using the SVD (73) we have for matrix $\mathbf{A} \in \mathfrak{R}^{R \times C}$ the following *pseudo-operators* (Golub and Reinsch, 1970)

$$(78) \quad \mathbf{A}^k = \mathbf{V}\mathbf{G}^k\mathbf{U}^T$$

where $g_p^k = \lambda_p^k$ if $\lambda_p > 0$, $g_p^k = 0$ otherwise, with $p = \min(R, C)$. For instance, with $k = -1$ we obtain the pseudo-inverse

$$(79) \quad \mathbf{A}^+ = \mathbf{V}\mathbf{G}^{-1}\mathbf{U}^T \quad \text{where } g_p^{-1} = \lambda_p^{-1} \text{ if } \lambda_p > 0, g_p^{-1} = 0 \text{ otherwise.}$$

The pseudo-inverse can also be found as

$$(80) \quad \mathbf{A}^+ = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1} \quad \text{if } \mathbf{A} \text{ is wide } (C > R). \text{ This is also named the right-inverse, in that } \mathbf{A}\mathbf{A}^+ = \mathbf{I},$$

but $\mathbf{A}^+\mathbf{A} \neq \mathbf{I}$.

$$(81) \quad \mathbf{A}^+ = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \quad \text{if } \mathbf{A} \text{ is tall } (R > C). \text{ This is also named the left-inverse in that } \mathbf{A}^+\mathbf{A} = \mathbf{I}, \text{ but}$$

$\mathbf{A}\mathbf{A}^+ \neq \mathbf{I}$.

Joint Diagonalization of Two Symmetric Matrices

Given two positive-definite symmetric matrices C_1 and C_2 , it always exists a matrix B such that

$$(82) \quad \begin{cases} B^T C_2 B = D_2 \\ B^T C_1 B = D_1 \end{cases},$$

where D_1 and D_2 are diagonal and holds all positive elements. This is called the *generalized eigenvalue-eigenvector decomposition (GEVD)*. Matrix B is named the *joint* (or *simultaneous*) *diagonalizer*. The solution is given by the eigenvector matrix of $C_2^{-1}C_1$. Since $C_2^{-1}C_1$ is not symmetric in general, matrix B is not orthogonal in general, thus the *EVD* has not form (35); it is orthogonal *iff* C_1 and C_2 commute in multiplication. Actually the *GEVD* is defined for any two symmetric matrices S_1 and S_2 for which it exists a positive-definite linear combination $aS_1 + bS_2$ (See Schott, 1997, p. 160-165). The *GEVD* can be obtained by *EVD* (35) with a two-step procedure (Fukunaga, 1990, p. 24-33), as it follows.

Algorithm (83): Generalized Eigenvalue-Eigenvector Decomposition (GEVD).

Do $EVD(C_2) = UAU^T$

Do $EVD(A^{-1/2}U^T C_1 U A^{-1/2}) = VDV^T$

The solutions is $V^T A^{-1/2}U^T$ and its inverse $UA^{1/2}V$.

FOURIER ANALYSIS

For real data, if T is the length of the time window and S the sampling rate expressed in Hz, there are $F=T/2+1$ Fourier frequencies with resolution $r=S/T$ equally spacing the range from 0 Hz (DC-level) to the folding frequency (0 Hz, 1r Hz, 2r Hz, ..., $T/2r$ Hz). Typically, we take both S and T as a power of two and spectral estimates may be averaged within arbitrary time intervals by sliding overlapping windows. The latter strategy allows arbitrary time intervals length. The Fourier *cospectra* and *quadrature spectra* are defined as the real and imaginary part of the Fourier *cross-spectra* (Bloomfield, 2000). They are estimations of, respectively, the *in-phase* (or with a half cycle phase shift, i.e., opposite sign) and *out-of-phase* (a quarter cycle in either direction) covariance structure at frequency f . The discrete Fourier transform of sampled time-series $x(t)$ over an epoch of length T is given by

$$(84) \quad d_{f(x)} = \frac{1}{T} \sum_{t=0}^{T-1} x(t) e^{-2\pi i f t} .$$

Let $d_{f(x)}^{\Re}$ and $d_{f(x)}^{\Im}$ be the real and imaginary part of $d_{f(x)}$, respectively⁹. Those coefficients are readily and efficiently estimated by fast Fourier Transform (FFT: Cooley and Tukey, 1965; Frigo and Johnson, 2005). Here below is the formula for computing the 2x2 *cospectral matrix* at frequency f for time-series $x(t)$ and $y(t)$:

$$(85) \quad \mathbf{C}_f = \begin{pmatrix} d_{f(x)}^{\Re} d_{f(x)}^{\Re} + d_{f(x)}^{\Im} d_{f(x)}^{\Im} & d_{f(x)}^{\Re} d_{f(y)}^{\Re} + d_{f(x)}^{\Im} d_{f(y)}^{\Im} \\ d_{f(y)}^{\Re} d_{f(x)}^{\Re} + d_{f(y)}^{\Im} d_{f(x)}^{\Im} & d_{f(y)}^{\Re} d_{f(y)}^{\Re} + d_{f(y)}^{\Im} d_{f(y)}^{\Im} \end{pmatrix} .$$

The formula readily extends to any N -dimensional input time-series to obtain its N -dimensional cospectral matrix. Notice that the cospectral matrix is symmetric and that the diagonal elements are the auto-spectra, better known as *power spectra*.

For an arbitrary long EEG segment we typically obtain an estimate of the cospectral matrix averaging \mathbf{C}_f over overlapping epochs of length T (Welch, 1967). Such estimates may then be summed across

⁹ For the first (0 Hz) and last ($T/2r$ Hz) Fourier frequency the coefficients are real.

adjacent frequencies to obtain estimates within band-pass regions of interest. Summing all of them yields the covariance matrix out of a scaling factor as per Parseval's theorem.

STATISTICS

Let

$$(86) \quad \mathbf{x}(t) = (x_1(t), \dots, x_N(t))^T \in \mathfrak{R}^N$$

be the EEG potential difference data vector for N electrodes unfolding along discrete time samples t . Throughout this manuscript we assume that the data has been band-pass filtered either during acquisition and/or as the unique pre-processing stage required before turning to the methods we treat here. Even if the band-pass region is large, say 0.1 Hz to 70 Hz, each EEG channel has then null expected average, such as

$$(87) \quad \mathbb{E}[\mathbf{x}(t)] = \mathbf{0} \in \mathfrak{R}^N$$

Mean

We have (Searle, 1982, p. 66-68 and p. 349-352)

$$(88) \quad \text{Mean:} \quad \bar{x}(t) = \frac{1}{N}(\mathbf{x}^T(t)\mathbf{1}) \in \mathfrak{R}$$

Centering Matrix and Common Average Reference

$$(89) \quad \text{The data referenced to the common average:} \quad \mathbf{x}_{CAR}(t) = \mathbf{x}(t) - \bar{x}(t)\mathbf{1} = \mathbf{H}\mathbf{x}(t) \in \mathfrak{R}^N$$

$$(90) \quad \mathbf{H} = \mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^T \in \mathfrak{R}^{N^2} \text{ is named the centering matrix.}$$

- (91) The centering matrix has $N-1$ equal eigenvalues and one null eigenvalue corresponding to eigenvector $\mathbf{1}$.
- (92) The centering matrix has the following properties: $\mathbf{H} = \mathbf{H}^T = \mathbf{H}^2$; $\mathbf{H}\mathbf{1} = \mathbf{H}\mathbf{1}\mathbf{1}^T = \mathbf{1}\mathbf{1}^T\mathbf{H} = \mathbf{0}$.

Sum of Squares and Products

Now let $\mathbf{X} \in \mathfrak{R}^{N \times T}$ be a data segment (e.g., a trial for BCI) of T samples. We have (Searle, 1982, p. 349-352)

(93) Sum of squares and products: $\mathbf{X}\mathbf{X}^T \in \mathfrak{R}^{N^2}$

(94) Sum of squares and products in common average reference: $\mathbf{H}\mathbf{X}\mathbf{X}^T\mathbf{H} \in \mathfrak{R}^{N^2}$

Covariance Matrix

Assuming (87) we obtain estimations of the

(95) Sample Covariance Matrix: $\mathbf{C} = \frac{1}{T-1}(\mathbf{X}\mathbf{X}^T) \in \mathfrak{R}^{N^2}$

(96) Sample Covariance Matrix in common average reference: $\mathbf{C} = \frac{1}{T-1}(\mathbf{H}\mathbf{X}\mathbf{X}^T\mathbf{H}) \in \mathfrak{R}^{N^2}$

NB: Hereafter by \mathbf{C} we will indicate *any* form of covariance matrix and not just the sample covariance matrix.

EEG BASICS

Advent and Standardization of EEG Recordings

With approximately 10^{12} neurons in the central nervous system (CNS), 10^{15} synaptic connections releasing and absorbing 10^{18} neuro-transmitters and neuro-modulators per second, the human brain is a net of prodigious complexity. Faugeras *et al.* (1999) compare such an organ with a “computer” capable of processing 10^{12} Gigabits of “information” per second, all in about 1.6 Kg of weight and with a consumption of 10-15 Watts.

The study of the brain “activity” through functional medical imaging devices is named *neuroimaging*. Based on the pioneering work on electricity of animal bodies by Luigi Galvani (1737-1798) and on exposed animal cortex by Richard Caton (1875), Hans Berger (1929) provided the first electroencephalogram (EEG) of a living human. In the whole neuroimaging community, this Berger’s finding is often evoked as a starting point. In order to record EEG, a set of electrodes are applied on the scalp so to establish electrical contact with the skin and in such a way to sample as evenly as possible the available scalp surface. To obtain congruence among different laboratories and different head shapes and sizes, standard placements have been soon proposed, basing the positioning on *proportional distances* along head anatomical landmarks (Jasper, 1958). Such standardization, with the seminal recordings on the exposed human cortex of Penfield and Rasmussen (1950) marks the beginning of modern electroencephalography. The number of electrodes used in research has increased over the years from around 19 of Jasper’s time to as many as 512 today, however the 10-20 system with 19 electrodes is still the dominant standard in clinical settings and most research is carried out with 19 to 64 electrodes (fig. 2.1).

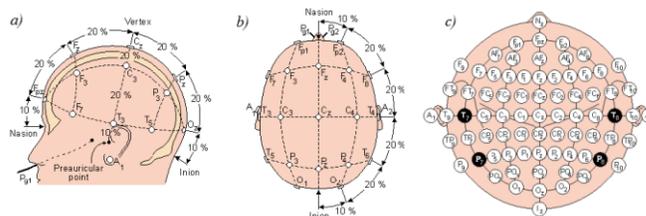


Figure 2.1: (a), (b): the international 10-20 system (Jasper, 1958) seen from left (A) and above (B) the head. Figure labels: A = ear lobe, C=central, Pg=nasopharyngeal, P=parietal, F=frontal, Fp=fronto-polar, T=temporal, O=occipital. (c): location and nomenclature of the intermediate 10-10 system, as standardized by the American Electroencephalographic Society (redrawn from Sharbrough *et al.*, 1991). Figure rearranged from Malmivuo and Plonsey (1995).

EEG and Other Neuroimaging Modalities

In between late 60's and 80's several other neuroimaging techniques have been introduced. The best known are *magnetoencephalography* (MEG), *positron emission tomography* (PET), *functional magnetic resonance imaging* (fMRI) and *functional near-infrared spectroscopy* (fNIRS). Concurrently, studies on brain imaging have literally booming. Yet, the introduction of more sophisticated neuroimaging techniques has not undermined the popularity of EEG. On the contrary, EEG research and practice is still growing while we are writing. There are several reasons that may explain the popularity of electroencephalography. First, EEG research reposes on a long and well established tradition, comforted by intracranial recordings in humans and animals. Second, with the exception of fNIRS, all other modalities require much more bulky and expensive equipment. Third, PET and fMRI may observe brain activity only indirectly, through the metabolic consumption rate (hemodynamic), which changes slowly and which activation peak occurs around 4s after actual cellular activity. Similar arguments apply to fNIRS. On the other hand EEG (and MEG) observes *instantaneous* changes of post-synaptic potential differences, which are directly related to the cell polarization/depolarization, hence to their readiness to discharge. Fourth, EEG (and MEG) is truly non-invasive. There is no limit of time one can safely hold the electrodes on the scalp (EEG) or one may stay in a MEG scanner. Complete safety applies to individuals of any age, including newborn children. Because of this characteristic EEG alone serves in sleep studies and in intensive care units. Fifth, EEG can be recorded in natural settings. Latest EEG equipment may have the size and weight of a handy box, can be powered by common batteries and may wireless store the data on a distant server. EEG can be recorded everywhere there is not excessive electromagnetic interference. Using active electrodes and/or actively shielded electrode cables it can be recorded even on a slowly moving subject, which is a possibility under exploitation in recent times. That is not true for any of the other modalities, including MEG. The ability to be used in natural settings and the absence of acoustic noise (very strong, for example, in fMRI scanners) makes EEG an ideal instrument for research and large public applications, such as video games based on brain-computer interfaces.

The Advent of Quantitative EEG Analysis

First attempts to interpret EEG traces have been based on waveform inspection (morphology). As a matter of fact, current practice of EEG in neurology is still anchored on this kind of analysis. Typically, neurological hospitals perform EEG examinations only for epilepsy, sleep disorder, migraine and a few other pathological conditions for which the waveform bears diagnostic utility, as

for spikes, spindles, generalized slowing, temporal theta, etc. Meanwhile, electrophysiological research took a different path. The introduction of quantitative methods in EEG (qEEG) aided by the (re-)invention of the FFT algorithm (Cooley and Tuckey, 1965) and by the advent of digital equipments, marked the transition into the era of *functional localization* of EEG activity. Spectral analysis (compressed spectral arrays, topographic maps of amplitude, phase, coherence, etc., see for example Duffy et al., 1981) has been used since the 60's in a plethora of cognitive and clinical studies, sustained also by concurrent advances in experimental paradigms. For instance, time-frequency analysis of event-related synchronization and desynchronization (ERS/ERD) has provided means to study brain dynamics in the scale of tens of ms, preserving both spatial and spectral information. This has enlarged the horizon of task-related brain studies beyond evoked response potentials, to allow investigations, for example, of movement-related potentials (for a review see Pfurtscheller and Lopes da Silva, 1999).

EEG Norms

On the clinical front, during the 70's a considerable effort has been injected to establish *normative databases* for a large number of spectral measures (John et al., 1980a,b,c). Comparison of clinical patient to such databases has been shown to offer a powerful tool for aiding the diagnosis of a wide range of disorders including, among others, depression, schizophrenia, learning disabilities, attention deficit disorder with or without hyperactivity and dementia (Hughes and John, 1999; Lopes da Silva, 2005a). The rationale behind normative databases resides in the fact that the normal EEG is largely determined genetically, so that similar space-frequency patterns are observed universally across genders, races and cultures (John et al., 1987). Their age-dependence, reflecting maturational changes during the developmental and aging period, can be taken explicitly into consideration (Ahn et al., 1980).

Another fundamental prerequisite of the validity of normative database comparisons is the intra-subject reliability, which for EEG (especially in a rest condition with the eyes closed) is truly astonishing, even across long periods of time. This is not the case, for example, of fMRI measurements, for which norms still are not available for this reason. Recent extended investigations on vegetative state and minimally conscious patients has shown that EEG is more sensitive than fMRI in detecting residual cognitive abilities and signs of consciousness in non-responsive patients (Monti, 2013, personal communication).

EEG Source Analysis

Solid advances in our understanding of the human brain are achieved linking empirical neuroimaging findings to organic anatomical and physiological knowledge. This way we have reached fairly deep understanding of elementary mental tasks involving primary sensory and primary motor cortical areas. For these areas the positioning of neurons somatotopically maps the sensory and motor organs, hence the anatomical and physiological substratum of the observed findings can be established unambiguously. On the other hand, the spatial resolution of surface EEG is low and does not enable fine spatial discrimination. This effectively limits the precision of the brain dynamics we are able to visualize working in the so-called *sensor space*. As a consequence, much methodological EEG research has been recently devolved to the *improvement of the spatial resolution*. This is achieved working in the so-called *source space*, consisting in the study of linear combinations of the data aiming at extracting latent variables hidden in the EEG. Surface EEG topographies have been first enhanced mapping second spatial derivatives (Laplacian) of the potentials (Hjorth, 1991; Lemos and Fisch, 1991; Nunez and Pilgreen, 1991). These methods actually enhance the spatial resolution only for the radial current component (perpendicular to the scalp).

More recently, EEG and MEG have benefit from several *source localization* and *source extraction (separation)* methods developed in other fields of research. Engineering studies on antenna array reception (beamforming: Van Veen and Buckley, 1988), physics studies on wave propagation in seismology (Backus and Gilbert, 1968) and statistical studies on blind source separation (Jutten and Herault, 1991) have all been adapted to the brain electromagnetic problem for the purpose of extraction of meaningful information and localization in brain space. Those efforts have been successful in providing reasonably accurate *electromagnetic tomographies*, i.e., true EEG or MEG based 3D volumetric functional images of the brain (Bosch-Bayard et al. 2001; Baillet, Mosher and Leahy, 2001; Chen et al., 2006; Greenblatt, Ossadtchi and Pflieger, 2005; Pascual-Marqui, Michel and Lehmann, 1994; Pascual-Marqui, 2002, 2207; Robinson and Vrba, 1999; Sekihara et al., 2004, 2005). In parallel, *spatial filtering* (Cichocki and Amari, 2002), and *blind source separation* (Comon and Jutten, 2010) have yielded means to decompose and optimize the information hidden in the observed data, effectively empowering the analysis in EEG and MEG studies. More recently, tools developed in quantum physics, relativity, elasticity, mechanics, radar, image and diffusion MRI data processing, have been borrowed from the field of *Riemann geometry* to achieve effective classification of mental states (Barachant and Congedo, submitted; Barachant et al., 2012a; Congedo et al., submitted; Li, Wong and De Bruin, 2009, 2012). All these evolutions and many others along the years have turned the study of the human electroencephalogram (EEG) in a strongly *multidisciplinary field*

of research. Besides neuroanatomy, psychology and neurophysiology, knowledge in signal processing, electromagnetism, and multivariate statistics proves nowadays essential prerequisite for high-level electrophysiological research.

A Short Introduction to the Physiology and Physics of EEG

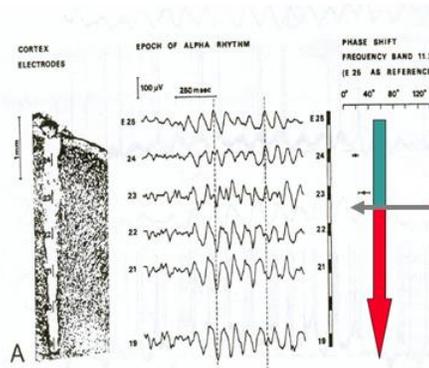
It is well established that the generators of brain electric fields recordable from the scalp are macroscopic post-synaptic potentials created by assemblies of pyramidal cells of the neocortex (Speckmann and Elger, 2005). Pyramidal cells are aligned and oriented perpendicularly to the cortical surface. Their synchrony is possible thanks to a dense net of local horizontal connections (mostly <1mm). At recording distances larger than about three/four times the diameter of the synchronized assemblies the resulting potential behaves as if it were produced by electric dipoles; all higher terms of the multipole expansion vanish and we obtain the often invoked dipole approximation (Lopes da Silva, 2004; Lopes Da Silva and Van Rotterdam, 2005; Nunez, 2005; Nunez and Srinivasan, 2006, Ch. 3, see fig. 2.2 therein).

Three physical phenomena are important for the arguments we advocate in the ensuing chapters. First, unless dipoles are moving there is no appreciable delay in the scalp sensor measurement (Lopes da Silva and Van Rotterdam, 2005). Second, in brain electric fields there is no appreciable electromagnetic coupling (magnetic induction) in the frequencies up to about 1MHz, thus the quasi-static approximation of Maxwell equations holds throughout the spectrum of interest (Nunez and Srinivasan, 2006, p. 535-540). Finally, for source oscillations below 40 Hz it has been verified experimentally that capacitive effects are also negligible, implying that potential difference is in phase with the corresponding generator (Nunez and Srinivasan, 2006, p. 61). These phenomena strongly support the *superposition principle*, according to which the relation between neocortical dipolar fields and scalp potentials may be approximated by a system of linear equations (Sarvas, 1987). We can therefore employ a *linear conduction model*. Because of these properties of volume conduction, scalp EEG potentials describe an *instantaneous mixture* of the fields emitted by several dipoles extending over large cortical areas (fig. 2.3). Whether this is a great simplification, we need to keep in mind that it does not hold true for all cerebral phenomena. Rather, it does at the macroscopic spatial scale concerned by EEG.

**6-Layers
Organization
Of Neocortex**

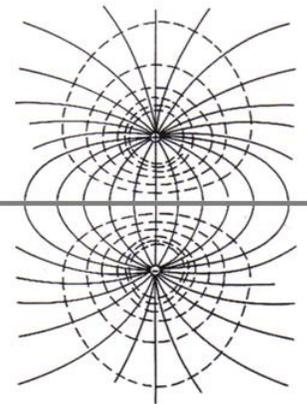


Alpha rhythm from visual cortex of dog
Lopes da Silva and Storm van Leeuwen (1977)



Cortical deep profile shows a polarity reversal at layer IV/V reflecting a **dipolar field**

Electric Dipole
Nunez and Srinivasan (2006)



Current Lines

Isopotentials

Figure 2.2: From left to right, the columnar organization of the mammalian neocortex, organized in six layers, intracranial recordings at the different layers in the visual cortex of the dog (Lopes Da Silva and Storm van Leeuwen, 1977) showing the polarity reversal at layer IV/V and the schematic representation of an electrical dipole (rearranged from Nunez and Srinivasan, 2006).

As a consequence, averaging EEG signals or features extracted from the EEG signal across subjects at the sensor level is not optimal, as different subjects have different spatial patterns and the average is a rather smeared representation of the group activity. We can circumvent this problem in two ways, namely, estimating sources at the individual level and then average such sources or features extracted from such sources or using ad-hoc group source extraction techniques. We will treat both of these approaches.

It is important to realize that the analysis of the scalp EEG signal does not allow per se establishing the position and orientation of sources. In fact, the scalp spatial pattern of the activity of a dipole depends very much on its orientation. This is shown in fig. 2.4 and 2.5.

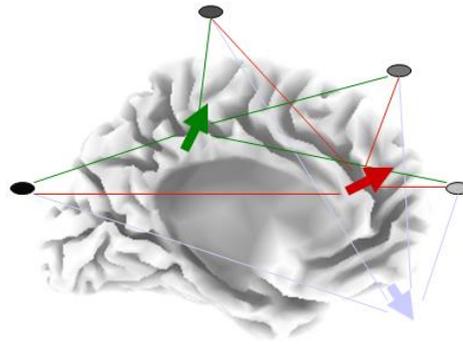


Figure 2.3: Schematic representation of three electrical dipoles (indicated by arrows) of which two are cortical and one is ocular. The disks represent EEG electrodes. The schema indicates the mixing conduction model, wherein each electrode records the activity of all the dipoles.

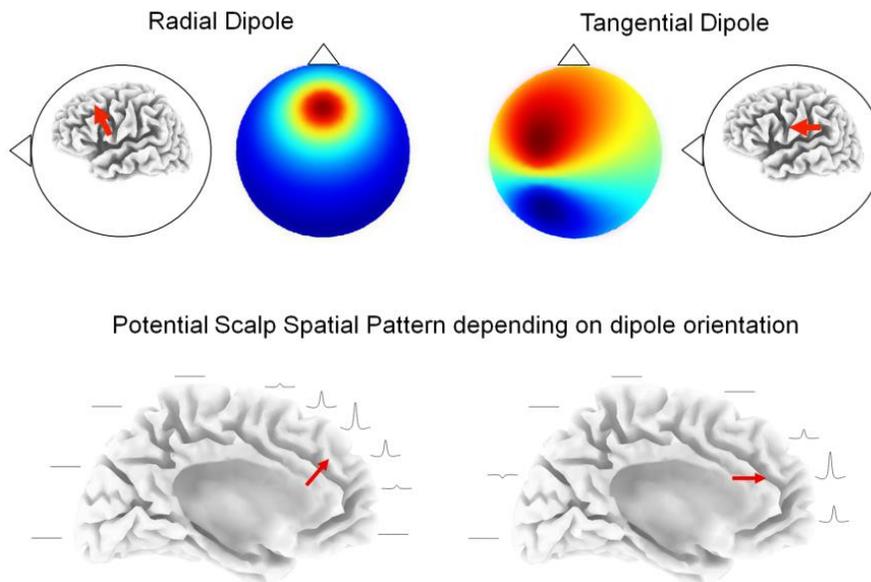


Figure 2.4: On the top left the orientation of a dipole located in the medial cortex is indicated by the arrow and the resulting scalp spatial pattern is shown. Since the dipole is radial to the scalp surface the scalp map is monopolar. On the top right the dipole is lateral and tangential to the scalp surface; the resulting dipole is bipolar. On the bottom two dipoles are located in the same medial position of two brains, but with different orientation. The traces around the brain indicate schematically the amplitude and sign of what we would record from the scalp of the two brains.

THE SENSOR MEASUREMENT

The EEG is, by definition, a measure of *potential difference*. The reference problem is rarely taken into consideration in EEG studies. Arranging the measures at N electrodes in a vector, we may write the potential as a function of time such as

$$\mathbf{x}(t) = \mathbf{x}_s(t) - \mathbf{x}_r(t) \in \mathfrak{R}^N, \quad (97)$$

where subscript s and r denote the scalp and respective reference leads. In practice time is sampled at regular intervals, ranging from hundreds to thousands of samples per seconds, depending on the study. Due to well-known results in A/D (analogue-to-digital) conversion, the sampling rate must be at least twice the maximal frequency contained in the sampled signal (*Nyquist* or *folding* frequency), thus the low-pass filter during data acquisition must be set accordingly.

Bipolar and Monopolar Reference

The choice of the *reference* is arbitrary. In clinical EEG sequential (transversal and longitudinal) *bipolar* recordings are preferred because of their higher signal-to-noise ratio. This is due to the fact that low spatial frequency noise cancels out when the difference is computed between two closely spaced leads. In research, *monopolar* recordings with a common reference for all electrodes are used because having a common reference for all leads allows treating the measurement vector with linear algebra tools (fig 2.6). Monopolar recordings have form

$$\mathbf{x}(t) = \mathbf{x}_s(t) - \kappa(t)\mathbf{I}, \quad (98)$$

where κ is the electrical reference, which is now *common to all leads* and \mathbf{I} the N -dimensional vector of 1's. Changing the reference changes dramatically the observed potentials. This arbitrariness has been partially resolved with the advent of *reference-free* methods such as the Laplacian and the inverse solutions treated in chapter III. Blind source separation methods treated in chapter V, VI and VII are not reference-free. Riemannian operations (chapter VIII and IX) are completely invariant with respect to any reference one can construct by premultiplying the data with an invertible matrix. This is not the case, for example, of the *common average reference* (CAR), which is particularly relevant in chapter III.

Common Average Reference

Given an arbitrary common reference $\kappa(t)$, the CAR potentials is given by

$$\mathbf{H}(\mathbf{x}_s(t) - \kappa(t)\mathbf{I}) = \mathbf{H}\mathbf{x}_s(t), \quad (99)$$

where \mathbf{H} is the centering matrix (90) and the equality stems from (92). In CAR data the sum of the potentials across electrodes at each time instant is null and *the centering matrix plays the role played by the identity matrix in raw potentials*. For instance, regularizing the CAR covariance matrix (96) will be obtained adding $\alpha\mathbf{H}$ to it, with $\alpha > 0$, whereas regularizing the covariance matrix in the original reference (95) is obtained adding $\alpha\mathbf{I}$. Note that the covariance matrix of CAR data has at most rank $N-1$, because one of its eigenvalues is null (91), thus the CAR is not a suitable reference for Riemannian operations, which are defined only for positive definite matrices.

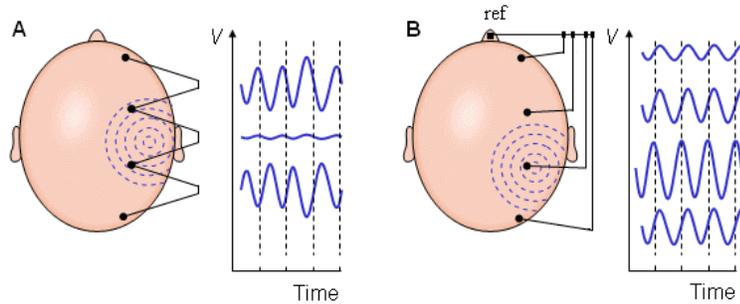


Figure 2.6: Longitudinal bipolar (A) and monopolar (referenced to the nose) (B) measurement. Notice the phase reversal between the first and third trace in (A). Figure edited and rearranged from Malmivuo and Plonsey (1995).

NB: Throughout this work the data are considered in common average reference when dealing with inverse solutions (chapter III) and in the arbitrary original reference when dealing with spatial filters, blind source separation and Riemannian methods (chapters IV to IX).

CHAPTER III

DISTRIBUTED INVERSE SOLUTIONS

Introduction

Research with quantitative methods in EEG has received a strong impulse since the introduction of *EEG inverse solutions* (Baillet, Mosher and Leahy, 2001). Hämäläinen and Ilmoniemi (1984) proposed the first inverse solution for EEG. Sarvas (1987) has first formalized the problem for magnetoencephalography, which has been then readily transposed to EEG. According to the linear instantaneous model of EEG generation discussed in chapter II, the electrical potential as measured on the scalp may be approximated by an *instantaneous weighted sum* of dipolar activity. The aim of an inverse solution is estimating the activity of the dipoles responsible of the observable measurement, given the observed measurement and a model of the conduction medium.

There are two types of inverse solutions: they may attempt to estimate a pre-defined number of dipoles (e.g., Mosher, Lewis and Leahy, 1992) or to estimate the activity in all the cortical volume without fixing the number of active dipoles. This latter approach comprises the so-called *distributed* inverse solutions and is the object of this chapter. Also, inverse solutions differ depending on whether they aim at estimating current flowing in all directions within the whole cortical grey matter, or whether they limits themselves to the estimation of current flowing in one or two directions through the scalp. In the former case, known as the *vector type*, we obtain 3D volumetric images and we need to estimate current flowing in three orthogonal directions. In the latter, known as *scalar type*, we obtain cortical surface maps and typically we estimate current flowing in the radial direction only (EEG) or in the two tangential directions only (MEG). We treat here the more involving vector type.

In a number of clinical, cognitive and methodological EEG papers we have used with success the well known LORETA method (Congedo Lubar and Joffe, 2004; Lubar, Congedo and Askew, 2003; Sherlin and Congedo, 2005; Sherlin et al., 2007) and, more recently, the sLORETA method (Congedo, 2006; Congedo et al., 2006, 2010; De Ridder et al., 2011; Kopřivová et al., 2011, 2013; van der Loo et al., 2007, 2009; Vanneste et al., 2010; White et al., 2012) and eLORETA method (van der Loo et al., 2011).

In this chapter we summarize useful knowledge to use these methods giving emphasis on practical issues and to material that cannot be found readily and compactly elsewhere. We also provide full but succinct algorithmic explanations and we report on how turning these *model driven* inverse solutions into *data driven* inverse solutions, establishing the relation between the minimum norm family of methods (Pascual-Marqui, 1999, 2007) and the minimum variance beamforming methods (Sekihara et

al., 2004; van Veen van Drongelen and Suzuki, 1997), which are the two most prominent families in current MEG and EEG literature.

The Forward Problem

Indicating by $\Omega(t)$ the set of active dipoles at each instant time, the EEG model discussed in chapter II takes the simple linear form

$$\mathbf{x}(t) = \sum_{i \in \Omega(t)} \mathbf{K}_i \mathbf{j}_i(t) - \kappa(t) \mathbf{I} \quad (100)$$

where $\mathbf{j}_i \in \mathfrak{R}^3$ holds the x , y , and z component of the dipolar current at space location i , $\mathbf{x}(t)$ is the sensor measurement and $\kappa(t) \mathbf{I}$ is a common reference (98) used for recording. Given Q voxels¹⁰ covering the entire *solution space*, typically restricted to the cortical grey matter, the matrices

$$\mathbf{K}_q = \left(\mathbf{k}_{q(x)}, \mathbf{k}_{q(y)}, \mathbf{k}_{q(z)} \right) \in \mathfrak{R}^{N \times 3} \quad (101)$$

are the Q partitions of

$$\mathbf{K} = \left(\mathbf{K}_1, \dots, \mathbf{K}_Q \right) \in \mathfrak{R}^{N \times 3Q}, \quad (102)$$

which is referred to as the *leadfield* matrix. The leadfield embeds the physical properties of the volume conduction model, i.e., the conduction of the current in the head. More precisely, each column of the leadfield is the *scalp field for unit-length dipole* located at the q^{th} position and pointing in one of three orthogonal directions, indicated in Cartesian coordinates by x , y and z (fig. 2.5). Notice that three orthogonal coordinates suffice to explain current flowing in *any direction*. Given an accurate head model and corresponding accurate leadfield¹¹, the forward problem fully describes the sensor measurement, under the assumption that the current is generated within the solution space. The *distributed forward problem* is conveniently expressed by linear equation

¹⁰ *Voxels* stands short for “volume elements”.

¹¹ A research field is dedicated to leadfield modeling; see for example Fuchs *et al.* (2002) and Wolters *et al.* (2006).

$$\mathbf{x}(t) = \mathbf{K}\mathbf{j}(t) - \kappa(t)\mathbf{I} \quad (103)$$

where

$$\mathbf{j} = (\mathbf{j}_1^T, \dots, \mathbf{j}_Q^T)^T \in \mathfrak{R}^{3Q}; \mathbf{j}_q = (j_{q(x)}, j_{q(y)}, j_{q(z)})^T \in \mathfrak{R}^3, \forall q \in \{1, \dots, Q\} \quad (104)$$

now accounts for current originating anywhere with any direction within the solution space. First of all, we get rid of the reference issue (Pascual-Marqui, 2007). For any instant time we seek

$$\min_{\kappa(t)} \|\mathbf{x}(t) - \mathbf{K}\mathbf{j}(t) - \kappa(t)\mathbf{I}\|_F^2 \quad (105)$$

The solution is

$$\kappa(t) = \frac{\mathbf{I}^T}{\mathbf{I}^T \mathbf{I}} (\mathbf{x}(t) - \mathbf{K}\mathbf{j}(t)). \quad (106)$$

Plugging (106) into (103) we obtain

$$\mathbf{H}\mathbf{x}(t) = (\mathbf{H}\mathbf{K})\mathbf{j}(t), \quad (107)$$

where \mathbf{H} is the centering matrix (90). Therefore, hereafter in this chapter the sensor measurement and the leadfield will always be assumed in the *common average reference* (89). The forward problem is then written simply as

$$\mathbf{x}(t) = \mathbf{K}\mathbf{j}(t). \quad (108)$$

The Inverse Problem

We do observe scalp potentials and we wish to estimate the current source location and orientation. This is named the *inverse problem*. The *distributed inverse problem* consists in estimating the whole current vector $\mathbf{j}(t)$, given head model \mathbf{K} and noisy scalp potentials $\mathbf{x}(t)$, both in common average reference, as we have seen. Typically $Q \gg N$, that is, we divide the cortical space in thousands of voxels, but we have only tens of electrodes. Thus the inverse problem is strongly *underdetermined* (Cichocki and Amari, 2002) and has infinite solutions with form

$$\mathbf{j}(t) = \mathbf{T}^T \mathbf{x}(t), \quad (109)$$

where

$$\mathbf{T}^T = (\mathbf{T}_1, \dots, \mathbf{T}_Q)^T \in \mathfrak{R}^{3Q \times N} \quad (110)$$

is a generalized inverse of \mathbf{K} termed the *Transfer* matrix. Note that the current estimation for each voxel is given by

$$\mathbf{j}_q(t) = \mathbf{T}_q^T \mathbf{x}(t) \in \mathfrak{R}^3, \quad (111)$$

where matrices $\mathbf{T}_q^T \in \mathfrak{R}^{3 \times N}$ are the Q partitions of (110). Once obtained the current vector we usually compute the *current density* at each voxel, effectively discarding the information about the orientation of the current. The current density at each voxel is given by the sum of the squares of the current in the three orthogonal directions, that is, the square of the length of the vector formed by the three coordinates x , y and z at each location (104):

$$\gamma_q(t) = j_{q(x)}^2(t) + j_{q(y)}^2(t) + j_{q(z)}^2(t). \quad (112)$$

Notice that the current flowing in each one of the three orthogonal directions is analogous to the scalp voltage (it oscillates around zero), while the current density is analogous to the scalp power¹².

There are infinitely many possible definition of a transfer matrix with some desirable properties (Chen et al., 2006; Pascual-Marqui et al., 1994; Pascual-Marqui, 2002; Robinson and Vrba, 1999). Here we consider two of such properties.

Inverse solutions satisfying the sensor measurement

One may wish a transfer matrix satisfying the measurement, that is, substituting the right-end side of (109) into (108), \mathbf{T} should satisfy then

$$\mathbf{x}(t) = \mathbf{K}\mathbf{T}^T \mathbf{x}(t), \text{ implying } \mathbf{K}\mathbf{T}^T = \mathbf{H}, \quad (113)$$

¹² Like current density, the EEG scalp power is a non-negative sum of squares. In the time domain it is the square of the amplitude. In the frequency domain, e.g., in Fourier analysis, it is the sum of the squares of the sine and cosine (orthogonal) components at each discrete frequency.

where because of the common average reference the centering matrix plays the role of the identity matrix. Such a \mathbf{T} is named the right-inverse of \mathbf{K} (80).

Inverse solutions with no localization error for noiseless sensor measurement

A relevant property of a transfer matrix is the ability to always localize correctly single dipoles, regardless their location and orientation, at least in noiseless sensor measurement. To check this property one uses *point spread functions* (Pascual-Marqui, 1999, 2002). As we have seen the scalp field for unit-length dipole located at the q^{th} position and pointing in one of three orthogonal directions is one of the columns of leadfield \mathbf{K} (102). We require that feeding the inverse solution with these columns, one by one, results in a current density vector having maximum at the location corresponding to the leadfield vector. That is, if

$$\hat{\mathbf{j}} = \mathbf{T} \mathbf{k}_{q(\cdot)} \quad (114)$$

is the current estimation for the leadfield vector representing the unit-length dipole located at the q^{th} position and pointing in any of three orthogonal directions (the dot in parenthesis stands for orientation in x , y , or z direction) and $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_Q)^T$ its current density (112) estimated for the entire volume, we require that the maximum in $\boldsymbol{\gamma}$ is at the q^{th} position. All point spread functions can be done at once and all relevant point spread function properties can be found in the *Resolution Matrix* of Backus and Gilbert (1968), which holds in its $3Q$ columns the collection of all $3Q$ point spread functions (114), such as

$$\mathbf{T}^T \mathbf{K} \in \mathfrak{R}^{3Q \times 3Q}. \quad (115)$$

Note that the no-localization error property, as seen by point spread functions, is valid only for noiseless sensor measurement in the case of a single active dipole. Of course, this is a hypothetical situation with no practical utility; however it can be considered a minimal requirement for any good candidate as an EEG inverse solution. How the inverse solution behaves with noise and in case of multiple active dipoles can be analyzed theoretically (Greenblatt, Ossadtchi and Pflieger, 2005; Pascual-Marqui, 2007; Sekihara, Sahani and Nagarajan, 2005), by simulations like in Wagner, Fuchs and Kastner (2004) and empirically by crossing information obtained with multiple neuroimaging modalities and other investigation means (Pascual-Marqui *et al.*, 2002).

The Minimum Norm Inverse Solution

The first solution proposed has been to define \mathbf{T} as the Moore-Penrose pseudo-inverse (79), with a Tikhonov regularization (Hämäläinen and Ilmoniemi, 1984). The minimum norm solution seeks the current vector \mathbf{j} satisfying (113) with minimal norm $\|\mathbf{j}\|_F^2$ in the least-squares sense (Cichocki and Amari, 2002, p. 57-58). It is given by

$$\mathbf{T}^T = (\mathbf{K}\mathbf{K}^T + \alpha\mathbf{H})^{-1} \mathbf{K}^T, \text{ for } \alpha \geq 0, \quad (116)$$

that is, for $\alpha = 0$ (no regularization) (116) reduces to (80). For $\alpha > 0$ the regularized solution becomes more robust to measurement noise imposing a weak smoothness constraint on the set of admissible solutions. All solutions belonging to the minimum norm family provide smooth and blurred current density reconstruction. Even if there is only one active dipole delimited within a single voxel, the reconstructed current density will cover a more or less large region, fading out in all directions with the distance from the location of the maximum. For vector type inverse solutions the minimum-norm (116) has large localization errors, that is, the maximum current density reconstructed for single dipoles is generally not in the dipole location. Particularly, it tends to localize dipoles always on the most superficial part of the volume. To alleviate this problem several weighting of the leadfield matrix have been proposed, but all of these first attempts have displayed large localization errors for the distributed case (Pascual-Marqui, 1999). The LORETA (low-resolution electromagnetic tomography) inverse solution (Pascual-Marqui, Michel and Lehmann, 1994) also displays localization errors, even if much smaller as compared to the previous attempts, thus we do not consider here it further. The first minimum-norm kind of solution with no localization error throughout the volume has been sLORETA (standardized LORETA: Pascual-Marqui, 2002), followed by eLORETA (exact LORETA: Pascual-Marqui, 2007). Both of them are regularized, hence, before turning to their description, let us discuss how to deal with the α parameter.

Regularization of Minimum Norm Inverse Solutions

Whether the best value of α is exactly zero for noise-free measurements, for real data a positive value is necessary to prevent spurious reconstruction due to measurement noise. The optimal value of α grows exponentially with noise and the number of electrodes. We have been using values as small as 0 for 19 electrodes and high SNR data and as large as 10^4 for 64-electrode data. It is better to

overestimate then to underestimate α . In fact, the underestimation of α in the presence of noise yields spurious current density distributions, very far from the expected ones. This can be easily checked by means of simulations. Yet, regularization engenders further smoothing, proportionally to α , which lowers the spatial resolution, that is, the ability to resolve two dipoles close to each other in space. Thus, we should always try to set α as close as possible to the optimal value. The optimal choice of α can be estimated by cross-validation. A safe strategy with real data is to check different solutions with decreasing value of α and stop just before the solution deviates significantly in term of maxima location and spatial distribution of current sources. This is true for all minimum-norm kind of inverse solutions we consider in this chapter.

Weighted Minimum-Norm Inverse Solutions

The family of (symmetricly) weighted minimum-norm inverse solutions has general form

$$\mathbf{T}^T = \boldsymbol{\Theta}^{-1} \mathbf{K}^T (\mathbf{K} \boldsymbol{\Theta}^{-1} \mathbf{K}^T + \alpha \mathbf{H})^+, \text{ with } \boldsymbol{\Theta} \in \mathfrak{R}^{3Q \times 3Q} \text{ symmetric and invertible.} \quad (117)$$

For $\boldsymbol{\Theta} = \mathbf{I}$ we have the minimum norm solution (116). With other choices of $\boldsymbol{\Theta}$ we obtain other weighted minimum-norm solutions, which, for a given $\boldsymbol{\Theta}$ are solutions to the problem

$$\min_j (\|\mathbf{x} - \mathbf{K}\mathbf{j}\|_F^2 + \alpha \mathbf{j}^T \boldsymbol{\Theta} \mathbf{j}). \quad (118)$$

eLORETA is obtained choosing $\boldsymbol{\Theta}$ block diagonal, with Q diagonal blocks $\boldsymbol{\Theta}_m \in \mathfrak{R}^{3 \times 3}$, that is

$$\boldsymbol{\Theta} = \begin{pmatrix} \boldsymbol{\Theta}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \boldsymbol{\Theta}_M \end{pmatrix}. \quad (119)$$

Model-Driven sLORETA

The general form of the transfer matrix (110) for the sLORETA solutions is

$$\mathbf{T}_q^T = \left[(\mathbf{K}_q^T \mathbf{Z} \mathbf{K}_q)^{-\frac{1}{2}} \mathbf{K}_q^T \mathbf{Z} \right], \forall q \in \{1, \dots, Q\}, \mathbf{Z} \in \mathfrak{R}^{N \times N} \text{ symmetric,} \quad (120)$$

where $\mathbf{K}_q^T \mathbf{Z} \mathbf{K}_q \in \mathfrak{R}^{3 \times 3}$ and superscript $-1/2$ indicates the *unique* symmetric square root inverse (52). The original (model-driven) standardized low-resolution electromagnetic tomography (sLORETA) solution proposed by Pascual-Marqui (2007) is given by (120) with

$$\mathbf{Z} = (\mathbf{K} \mathbf{K}^T + \alpha \mathbf{H})^+ . \quad (121)$$

Equivalently, one can show that sLORETA is obtained after minimum norm current estimation (109) (obtained using \mathbf{T} in (116)) by the following voxel-by-voxel weighting (Pascual-Marqui, 2002)

$$\gamma_q(t) = \mathbf{j}_q^T(t) [\mathbf{T}^T \mathbf{K}]_q^{-1} \mathbf{j}_q(t), \text{ for all } q \in \{1, \dots, Q\}, \quad (122)$$

where $[\mathbf{T}^T \mathbf{K}]_q \in \mathfrak{R}^{3 \times 3}$ is the 3x3 diagonal block of the resolution matrix that we have already encountered in (115) and \mathbf{T}^T is again the minimum-norm transfer matrix. We see that sLORETA outputs a standardized estimate of current density. Eq. (122) has the form of the Mahalanobis distance of the current vector from the origin, i.e., the actual length of the vector, taking into account the covariance structure of its three components. As a consequence, standardized current density estimations are expressed on the same (dimensionless) metric all across the volume, regardless the norm of the leadfield columns. sLORETA is an unbiased estimator of source location in noiseless measurements, meaning that it is able to correctly estimate the location of a single active source regardless of location and orientation. This result has been demonstrated both by point spread functions (Pascual-Marqui, 2002), and theoretically (Greenblatt, Ossadtchi and Pflieger, 2005; Pascual-Marqui, 2007; Sekihara, Sahani and Nagarajan, 2005).

sLORETA is also capable of separating simultaneously active sources given that their energy is comparable and that their distance exceeds the spatial resolution attained, which depends on head model and number of sensors (Wagner, Fuchs and Kastner, 2004). In general, the resolution with multiple sources increases as the orientation of dipoles diverges and superficial sources tend to dominate deeper sources.

Data-Driven sLORETA

Let us consider a ‘‘Bayesian’’ interpretation of (120) and (121); The resolution matrix is the estimated source covariance matrix assuming the identity matrix as its prior, $\alpha \mathbf{H}$ as the noise covariance matrix prior and $\mathbf{K} \mathbf{K}^T + \alpha \mathbf{H}$ as the sensor measurement covariance prior (Pascual-Marqui, 2002). Assuming

$\mathbf{K}\mathbf{K}^T + \alpha\mathbf{H}$ as the sensor measurement covariance prior amount to the *maximal incertitude* about the active dipole location and orientation. In fact, by definition (108) and (115) notice that $\mathbf{K}\mathbf{K}^T$ is the *sensor measurement covariance matrix* that one would obtain if all Q dipoles throughout the whole volume are active with current flowing in all directions. Such a solution is completely “uninformed”, that is, it assumes no knowledge whatsoever about the number of active dipoles, nor about their location and orientation. This choice of \mathbf{Z} can be used in any situation; however it will produce the maximally smooth solution. The maximally smooth solution is the one with minimal spatial resolution, i.e., minimal ability to resolve spatially close dipoles. On the opposite, for a single active dipole with leadfield k^{I3} , the maximally sharp solution is given by $\mathbf{Z} = (\mathbf{k}\mathbf{k}^T + \alpha\mathbf{H})^+$. Similarly, if I dipoles are active in the time period under analysis, with $i \in \{1, \dots, I\}$, the maximally sharp solution is given by $\mathbf{Z} = \left(\sum_i \sigma_i \mathbf{k}_i \mathbf{k}_i^T + \alpha\mathbf{H}\right)^+$, where σ_i is the strength of the i^{th} dipole. In practice, we rarely have precise a-priori knowledge about the number, location and orientation of active dipoles for the time interval under analysis, nonetheless we can use its empirical estimations based on the observed sensor measurement covariance matrix. This yields the data-driven version of sLORETA. It is given by (120) with

$$\mathbf{Z} = (\mathbf{C} + \alpha\mathbf{H})^+ \quad (123)$$

and \mathbf{C} the CAR data covariance matrix (96). Notice that we may use other estimations of \mathbf{Z} , for example using instead of \mathbf{C} in (123) the outer product of some columns of the mixing matrix estimated by blind source separation as we will see in chapter VI and VII.

Model-Driven eLORETA

eLORETA is a weighted minimum norm solution (117) and makes use of a block-diagonal form for the weighting matrix as per (119). The model driven eLORETA is defined choosing for (117)

$$\boldsymbol{\Theta}_q^{-1} = \left[\mathbf{K}_q^T (\mathbf{K}\boldsymbol{\Theta}^{-1}\mathbf{K}^T + \alpha\mathbf{H})^+ \mathbf{K}_q^T \right]^{\frac{1}{2}}, \text{ for all } q \in \{1, \dots, Q\} \quad (124)$$

¹³ We do not introduce further notation here. For a dipole located at the q^{th} voxel, vector \mathbf{k} here actually stands for the linear combination of the three orthogonal vectors in \mathbf{K}_q describing the actual direction of the dipole.

Notice that the computation of $\boldsymbol{\theta}$ is nested into itself. This prevents any closed-form solution, however the solution can be found by an efficient iterative algorithm. The eLORETA problem (124) is the optimization of (Pascual-Marqui, 2007)

$$\min_{\boldsymbol{\theta}} \left\| \mathbf{I} - \left[\boldsymbol{\theta}^{-1} \mathbf{K}^T (\mathbf{K} \boldsymbol{\theta}^{-1} \mathbf{K}^T + \alpha \mathbf{H})^+ \mathbf{K} \boldsymbol{\theta}^{-1} \right] \right\|_F^2, \quad (125)$$

which satisfies the set of matrix equations

$$\boldsymbol{\theta}_q^2 = \mathbf{K}_q^T (\mathbf{K} \boldsymbol{\theta}^{-1} \mathbf{K}^T + \alpha \mathbf{H})^+ \mathbf{K}_q, \text{ for all } q \in \{1, \dots, Q\}. \quad (126)$$

The algorithm is

Algorithm (127) eLORETA

Initialize $\boldsymbol{\theta} = \mathbf{I}$

Repeat

$$\boldsymbol{\Pi} \leftarrow (\mathbf{K} \boldsymbol{\theta}^{-1} \mathbf{K}^T + \alpha \mathbf{H})^+ \quad (128)$$

$$\text{for } q=1 \text{ to } Q \text{ do } \boldsymbol{\theta}_q^{-1} \leftarrow (\mathbf{K}_q^T \boldsymbol{\Pi} \mathbf{K}_q)^{-1/2}$$

Until Convergence

(the Frobenius norm of the difference between two successive update of $\boldsymbol{\theta}$ is $< \varepsilon$)

Finally compute $\mathbf{T}_q^T = \boldsymbol{\theta}_q^{-1} \mathbf{K}_q^T (\mathbf{K} \boldsymbol{\theta}^{-1} \mathbf{K}^T + \alpha \mathbf{H})^+$ for all $q \in \{1, \dots, Q\}$

Note that, being $\boldsymbol{\theta}^{-1} \in \mathfrak{R}^{3Q \times 3Q}$ block diagonal as per (119), we do not need to compute the full matrix multiplication $\mathbf{K} \boldsymbol{\theta}^{-1} \mathbf{K}^T$ as this matrix reduces in this case to $\sum_q (\mathbf{K}_q \boldsymbol{\theta}_q^{-1} \mathbf{K}_q^T)$. Also note that we do not need to invert the full matrix $\boldsymbol{\theta}$ since, because of its block-diagonal structure, it holds

$$\boldsymbol{\theta}^{-1} = \begin{pmatrix} \boldsymbol{\theta}_1^{-1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \boldsymbol{\theta}_Q^{-1} \end{pmatrix}. \quad (129)$$

We may notice that eLORETA assumes that there is no correlation between the current in different voxels and seeks the solution with least correlation among directions within each voxel.

Data-Driven eLORETA

The data-driven version of eLORETA is obtained with the same algorithm (127), but with (128) replaced by

$$\mathbf{\Pi} = (\mathbf{C} + \alpha\mathbf{H})^+, \quad (130)$$

which is analogous to what we have done for the data-driven sLORETA (123).

Similarity with the Minimum Variance Beamforming.

The linearly constrained minimum variance beamforming is very popular in the MEG literature. (LCMV: Van Veen, Drongele and Suzuki, 1997). This is a data-driven method with solution

$$\mathbf{T}_q^T = \left[(\mathbf{K}_q^T \mathbf{C}^{-1} \mathbf{K}_q)^{-1} \mathbf{K}_q^T \mathbf{C}^{-1} \right], \forall q \in \{1, \dots, Q\}, \quad (131)$$

where \mathbf{C} is the sample covariance matrix and the current density obtained by this transfer matrix is somehow normalized (Sekihara et al., 2004; Sekihara, Sahani and Nagarajan, 2005). As we have seen the sLORETA data-driven solution is given instead by using (123) in (120), yielding

$$\mathbf{T}_q^T = \left[(\mathbf{K}_q^T \mathbf{C}^{-1} \mathbf{K}_q)^{-\frac{1}{2}} \mathbf{K}_q^T \mathbf{C}^{-1} \right], \forall q \in \{1, \dots, Q\}. \quad (132)$$

If the dipole orientation at each voxel is constrained in the radial direction or the direction is estimated by a data-driven method, so that the leadfield is now a collection of Q vectors such as

$\mathbf{K} = (\mathbf{k}_1, \dots, \mathbf{k}_Q)$, we have the scalar type LCMV solution

$$\mathbf{t}_q^T = \left[\frac{\mathbf{k}_q^T (\mathbf{C} + \alpha\mathbf{H})^{-1}}{\sqrt{\mathbf{k}_q^T (\mathbf{C} + \alpha\mathbf{H})^{-1} \mathbf{k}_q}} \right], \forall q \in \{1, \dots, Q\} \quad (133)$$

and the sLORETA scalar type solution

$$\mathbf{t}_q^T = \left[\frac{\mathbf{k}_q^T (\mathbf{C} + \alpha\mathbf{H})^{-1}}{\mathbf{k}_q^T (\mathbf{C} + \alpha\mathbf{H})^{-\frac{1}{2}} \mathbf{k}_q} \right], \forall q \in \{1, \dots, Q\}. \quad (134)$$

The difference between the two is just a different voxel-wise normalization. As a matter of fact sLORETA does not need further normalization, whereas the LCMV beamformer does.

Point Spread Function Simulations

Both sLORETA and eLORETA features zero localization error as seen by point spread functions, however they are quiet different solution mathematically. They also differ in the theoretical ability to give exact results in different theoretical noisy conditions (Pascual-Marqui, 2007). In order to compare them we have performed point spread function simulations using the three shell spherical head model implemented in the LORETA-Key Software (Pascual-Marqui, 1999). We have studied the behavior of these two inverse solutions plus the behavior of the sLORETA method informed with the *position* of the test dipole and the sLORETA method informed with both its *position and orientation*. To inform about the *position* we define matrix \mathbf{Z} in (123), to be used in (120) to obtain the transfer matrix, as

$$\mathbf{Z} = \left(\mathbf{K}_q \mathbf{K}_q^T \right)^+ \quad (135)$$

where for each one of the $3Q$ simulations, \mathbf{K}_q corresponds to the true test location q . To inform about both the position and orientation we define matrix \mathbf{Z} as

$$\mathbf{Z} = \left(\mathbf{k}_{q(.)} \mathbf{k}_{q(.)}^T \right)^+, \quad (136)$$

where for each one of the $3Q$ simulations, $\mathbf{k}_{q(.)}$ is the exact leadfield vector used in the simulation. Notice that this latter case corresponds to the ideal situation where the maximum a-priori information is given to sLORETA, that is, the prior used is the exact prior describing just the simulated activity. Furthermore, the results presented herein concern simulations with only one active dipole and no noise whatsoever, thus they represent the upper limit of performance attainable by the inverse solution. Insomuch they are particularly interesting. We have performed point spread function simulations using 19, 32, 64 and 90 electrodes. For each method and number of electrodes we have analyzed two kinds of errors:

- The *Spread Error*, defined as the sum of the energy in the other locations divided by the energy in the test location. For each of the $3Q$ point spread functions the current density is computed all over the volume. No localization error is achieved if the maximum of the current density is found in the location where the

dipole was simulated. However, we wish that the current density in the other locations be as small as possible. Hence, the definition of the spread error; the lower the spread error the more focal the current density reconstruction around the maximum. In table 3.1 we report the spread error averaged across the $3Q$ point spread function simulations.

- The *Equalization Error*, defined as the variance of energy across test locations and orientations. A desirable property of an inverse solution is that for a unit-length dipole simulation the energy in the test locations be equal wherever we place the dipole and whatever its orientation. For instance, we desire that deep and superficial sources with the same energy are reconstructed with similar energy. Therefore, we compute the current density at the test location for all $3Q$ simulations and we compute the variance of these values; the lower the variance the more homogeneous is the current density across test locations. In table 3.2 we report this variance multiplied by 10^4 .

Table 3.1: Average Spread Error across all $3Q$ point spread function simulation for eLORETA, sLORETA, sLORETA with exact dipole position information (Pos) and sLORETA with exact information of dipole position and orientation (Pos+Orient), for 19, 32, 64 and 90 electrodes.

Electrodes	eLORETA	sLORETA	Pos	Pos+Orient
19	429.06	483.69	0.48	0.32
32	281.26	350.14	0.28	0.19
64	158.56	219.66	0.11	0.09
90	109.28	151.53	0.03	0.03

Table 3.2: Equalization Error for the $3Q$ point spread function simulation for eLORETA, sLORETA, sLORETA with exact dipole position information (Pos) and sLORETA with exact information of dipole position and orientation (Pos+Orient), for 19, 32, 64 and 90 electrodes.

Electrodes	eLORETA	sLORETA	Pos	Pos+Orient
19	0.0881	23.3484	0.3151	0.2256
32	0.5328	54.5325	0.1824	0.1283
64	2.1251	117.7894	0.1068	0.0722
90	3.9541	160.6805	0.0194	0.019

About the spread error we can notice that the spatial resolution (inverse of the spread) increases with the number of electrodes for all methods, as expected. eLORETA features a lower spread error for all number of electrodes considered. We can also notice that informing the solution with position only engenders a dramatic reduction of the spread error, which is only slightly lowered further by adding information on the orientation as well. For what it concerns the equalization error we notice that it is about 265 times larger for sLORETA as compared to eLORETA with 19 electrodes and about 40 times larger with 90 electrodes. Also for the equalization error, the reduction is important when adding information of the location of the dipole, while the further reduction obtained adding the information on the dipole orientation is not very strong.

Conclusions

We conclude that the eLORETA method clearly displays a favorable behavior as compared to sLORETA in noiseless point spread function simulations. Based on our simulations, we also conclude that the use of data-driven methods is clearly to be preferred whenever possible. Several things however should be considered before opting for a data-driven method: the transfer matrix must be computed for each data segment to be analyzed, whereas for the model-driven methods it is computed only once. Also, with real data the optimal regularization amount may need to be re-estimated for each data segment to be analyzed. Finally, the data segments used to estimate the prior (e.g., the covariance matrix) to be used in (123) for sLORETA and in (130) for eLORETA should reflect the activity of the smallest possible number of dipoles; if the activity of many dipoles is contained in the covariance matrix, the advantage of the data-driven computation of the transfer matrix becomes irrelevant. An interesting option is to use as prior the outer product of the columns of the mixing matrix estimated by blind source separation (see chapter VI and VII), taken one at a time, instead of the whole covariance matrix. Such a prior is optimal to localize the sources estimated by BSS. In fact the columns of the mixing matrix are a decomposition of the EEG activity in a number of independent dipoles or dipole clusters, thus each column of the mixing matrix hold information about the activity of a much smaller number of dipoles as compared to the total number of active dipoles in the data. However, we have not investigated the effect of plugging into sLORETA and eLORETA data-driven equations non positive definite matrices.

One should also consider that using a model-driven method is actually preferable if the covariance matrix estimation, or whatever is the prior used, is biased; informing the data-driven inverse solution with a biased prior may actually result in a worse reconstruction as compared to what is obtained with

a uninformed (model-driven) method. We have observed this in many simulations (data not shown). In practice, our opinion is that data-driven methods may result useful for very high SNR data (for example, for averaged evoked response potentials) or when the data is decomposed in a number of uncorrelated components and a data-driven transfer matrix can be computed for each of them separately.

Current Density Estimation in Regions of Interest

Besides our simulations, so far in this chapter we have presented material published by others and, particularly, the work of R.D. Pascual-Marqui. Hereafter we report some of our contributions. In many situations we are interested in the estimation of the current density in a region of interest (ROI), defined as a cluster of connected voxels covering the anatomical area of interest. This is the case for example in real-time applications such as neurofeedback based on EEG inverse solutions (Bauer, Pllana and Sailer, 2011; Choi, 2014; Congedo, Lubar and Joffe, 2004; Kopřivová et al., 2013; Liechti et al., 2012; Salari et al., 2012; Surmeli and Ertem, 2009) or real-time monitoring in general (Im et al., 2007). To compute the total current density in the ROI we do not need to compute the current in all voxels belonging to the ROI via (109) and then summing the current density obtained via (112). Instead, we here show a faster method that also opens the way to model-based filters to attenuate the interference coming from other regions (beamformers). In fact in real-time the use of data-driven inverse solutions is impractical.

The following developments apply to whatever linear inverse solution for which a model-driven transfer matrix \mathbf{T}^T has been computed (Congedo, 2006), as for example the sLORETA or eLORETA model and data driven transfer matrix. Let us indicate the ROI simply by a set of voxels Ω . First notice that for the voxel at location i the current density estimation (112) can be written such as

$$\gamma_i(t) = \mathbf{j}_i^T(t) \mathbf{j}_i(t). \quad (137)$$

Substituting in (137) the right-end side of (111) yields

$$\gamma_i(t) = \mathbf{x}^T(t) \mathbf{\Xi}_i \mathbf{x}(t), \quad (138)$$

where we name

$$\mathbf{\Xi}_i = \mathbf{T}_i \mathbf{T}_i^T, \quad (139)$$

the (quadratic) *inverse operator*. For estimating the total current density in ROI Ω we compute simply

$$\gamma_{\Omega}(t) = \sum_{i \in \Omega} (\mathbf{x}^T(t) \mathbf{\Xi}_i \mathbf{x}(t)) = \mathbf{x}^T(t) \mathbf{\Xi}_{\Omega} \mathbf{x}(t), \quad (140)$$

where

$$\mathbf{\Xi}_{\Omega} = \sum_{i \in \Omega} \mathbf{\Xi}_i. \quad (141)$$

Notice that the inverse operator is always symmetric. For a point region, i.e., a single voxel at location i , $\text{rank}(\mathbf{\Xi}_i)=3$. For an extended ROI, $\text{rank}(\mathbf{\Xi}_{\Omega}) \geq 3$, due to the fact that leadfield vectors corresponding to distant points in solution space progressively diverge.

An even faster expression for the regional current density can be obtained using the factorization of inverse operator $\mathbf{\Xi}_{\Omega}$, given by

$$\mathbf{\Xi}_{\Omega} = \mathbf{G}_{\Omega} \mathbf{G}_{\Omega}^T, \quad (142)$$

where

$$\mathbf{G}_{\Omega} = (\lambda_1^{1/2} \mathbf{u}_1, \dots, \lambda_p^{1/2} \mathbf{u}_p), \in \mathfrak{R}^{N \times P} \quad (143)$$

$P < N-1$ ¹⁴ and \mathbf{u}_p and λ_p are the eigenvectors and associated eigenvalues of $\mathbf{\Xi}_{\Omega}$ arranged in descending order of eigenvalues (35). The current density in the ROI is then given by

$$\gamma_{\Omega}(t) = \left\| \mathbf{G}_{\Omega}^T \mathbf{x}(t) \right\|_F^2. \quad (144)$$

Equation (144) derives directly from (140) using factorization (142). We write first

$$\mathbf{x}^T(t) \mathbf{\Xi}_{\Omega} \mathbf{x}(t) = \text{tr}(\mathbf{x}^T(t) \mathbf{\Xi}_{\Omega} \mathbf{x}(t)) = \text{tr}(\mathbf{x}^T(t) \mathbf{G}_{\Omega} \mathbf{G}_{\Omega}^T \mathbf{x}(t)),$$

and then (144) follows from the properties of the trace. Typically, one wishes to compute the current density in the ROI for a time interval (e.g., a BCI trial or sliding overlapping windows in neurofeedback), given its covariance matrix (or Fourier cospectral matrix) \mathbf{C} . In this case the current density estimation is given by

¹⁴ It is $N-1$ and not N because one dimension is already lost by the common average reference.

$$\gamma_{\Omega} = \text{tr}(\mathbf{G}_{\Omega}^T \mathbf{C} \mathbf{G}_{\Omega}) = \sum_p (\mathbf{g}_p^T \mathbf{C} \mathbf{g}_p), \quad (145)$$

where \mathbf{g}_p is the p^{th} column vector of \mathbf{G}_{Ω} and \mathbf{g}_p^T its transpose.

Data-Independent Filters for Regional Inverse Solutions

Actually, using (144) or (145) and choosing P in such a way that the representation error (39) is small (say, <0.01), one obtain a *beamformer* effectively suppressing the interference of current flowing outside the ROI (Congedo, 2006). Beamforming refers to the use of spatial filters in order to enhance the receptivity of the sensors to sources emitting from a chosen region. It has been widely applied to other emission/reception systems, like sonar, radar and satellite/antennas (Van Veen and Buckley, 1988). The data is projected on a reduced space, called the beamspace, with dimension $P < N-1$. The aim of the beamforming filter considered here is to reduce the interference emitted by uninteresting sources, both cranial and extra-cranial. The method is illustrated here using the three-shell spherical head model available with the LORETA-Key software (Pascual-Marqui, 1999). The solution space in this head model includes 2394 voxel of dimension 7mm^3 each. We define a deep ROI composed of 36 voxels, roughly covering the anterior cingulate cognitive division, which, for instance, is of practical therapeutic utility in the treatment of the attention deficit disorder via neurofeedback (Chabot et al., 2005; Congedo, Lubar and Joffe, 2004). The ROI covers only $36/2394=1.5\%$ of the total solution space. We analyze the eigenvalue spectrum of the inverse operator defined on the ROI (141) using 6, 12, 19, 32, 64 and 90 electrodes evenly spaced on the scalp (fig. 3.1). As it can be seen, for all electrode montages with the exception of 6 electrodes only, there is a large gap in between the third and fourth eigenvalue, with the eigenvalues dropping by two orders of energy. Thus the representation error setting $P=3$ in (143) is small and by using (144) with $P=3$ one can reduce the interference coming from other regions.

Since data-independent filters exploit the model, but not the data, in general they are more effective on data-driven inverse solutions. For more information on beamformers and their application to inverse solutions see Bolton et al. (1999), Chen at al. (2006), Congedo (2006), Gross and Ioannides (1999) and Rodríguez-Rivera et al. (2006).

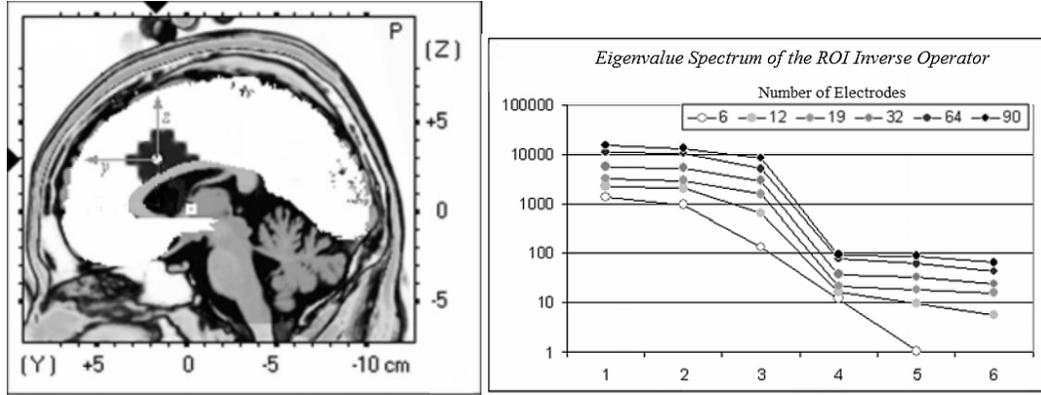


Figure 3.1: Left: medial view of the brain. Left of picture is front of the brain. The ROI is indicated by the dark shaded area and corresponds roughly to the cognitive division of the anterior cingulate. Right, eigenvalue spectrum including the first 6 eigenvalues arranged in descending order for the ROI's inverse operator (141) obtained with different number of electrodes.

Data-Dependent Filters for Regional Inverse Solutions

In order to derive filters for inverse solution we can also exploit the data. Contrary to what happens for data-independent filters, in general data-dependent filters will be more effective on model-driven inverse solutions.

Measurement noise suppression

Consider first the case when we want to suppress measurement noise, which is useful when using a large number of electrodes (say, >30). For doing so one filter the data $x(t)$ before applying (144) or (145) with a *principal component analysis*, which we will encounter in the chapter V. The data is *projected* in the signal subspace such as

$$\hat{x}(t) = UU^T x(t), \quad (146)$$

for time points and

$$\hat{C} = UU^T C U U^T \quad (147)$$

for time intervals given covariance matrix C , where

$$U = (u_1, \dots, u_p) \in \mathbb{R}^{N \times p} \quad (148)$$

and \mathbf{u}_p are the eigenvectors of the data covariance matrix sorted in descending order of corresponding eigenvalues. As for data-independent filters, $P < N-1$, with P chosen in such a way that the remaining $N-P$ eigenvalues of the covariance matrix explain a small proportion of the total variance, that is, that the representation error is small. One then feed (144) and (145) with filtered data (146) and (147), respectively, depending whether time points or time segments are analyzed.

Increasing classification accuracy

One may also apply data-dependent spatial filters with specific properties. For example in Congedo, Lotte and Lécuyer (2006) we have used sLORETA to classify motor intention data in a BCI experiment. Here the filter aims at better separating the classes. Data comprised one subject and constituted the BCI competition 2003, dataset IV, provided by the Berlin BCI group, Berlin Institute of Technology (Blankertz et al., 2004). The task of the subject was to press with the index and little fingers keys using either the left or right hand, in a self-paced timing and self-chosen order. Epochs of 500 ms were extracted ending 130 ms before the key press, thus only *movement intention* can be used for classification. The epochs were divided in a *training* set and a *test* set (316 and 100 trials, respectively). EEG data were acquired using 28 electrodes and sampled at 1000 Hz. Since there were two-classes, left and right hand motor intention, we have applied to sLORETA a *common spatial pattern* filter (see chapter V) for eliminating current not relevant for classification purposes. The common spatial pattern (CSP) filter is a matrix \mathbf{F} diagonalizing simultaneously both the grand-average covariance matrix of left and right training trials. Once we find this matrix we keep a small numbers of vectors maximizing the ratio of the variance between the two classes, which we name \mathbf{F}_R for the right sensorimotor motor cortex (desynchronized during left-hand motor intention) and \mathbf{F}_L for the left sensorimotor cortex (desynchronized during right hand motor intention). Using training data we also defined two single-voxel ROIs, Ω_R for the right sensorimotor cortex and Ω_L for the left sensorimotor cortex, to which regional inverse operator \mathbf{E}_R and \mathbf{E}_L (141) and their factorization \mathbf{G}_R and \mathbf{G}_L (143) corresponded. Since the ROIs comprise one single voxel, the rank of the inverse operator is exactly three and no data-independent beamforming can be applied¹⁵. Figure 3.2 shows the sLORETA source localization of the common spatial patter (CSP) spatial filters. This is obtained by feeding sLORETA

¹⁵ Note that in this case, since the ROI is composed of one voxel only, \mathbf{G}_R (\mathbf{G}_L) is equal to the partition of the transfer matrix \mathbf{T}_q corresponding to the voxel forming the Ω_R (Ω_L) region.

with the columns of the pseudo-inverse of \mathbf{F}^T . Since data were re-referenced to the common average, a total of $N-1=27$ filters were estimated by the CSP. By construction of the CSP the first $P \ll N/2$ vectors of the filter are optimal for one class and the last P for the other class (see (201)). Based on the source localization of the filters (fig. 3.2) we define *data-dependent filters*

$$\mathbf{F}_L = (\mathbf{f}_1, \dots, \mathbf{f}_P) \in \mathfrak{R}^{N \times P} \quad (149)$$

and

$$\mathbf{F}_R = (\mathbf{f}_{N-P-1}, \dots, \mathbf{f}_{N-1}) \in \mathfrak{R}^{N \times P} \quad (150)$$

as the first and last P columns of the CSP filter, respectively. In this study we have fixed $P=2$ based on results in fig. 3.2.

For the test trials with covariance matrix \mathbf{C} , the filtered data is then given by

$$\hat{\mathbf{C}}_L = \left[\mathbf{F}_L (\mathbf{F}_L^T \mathbf{F}_L)^{-1} \mathbf{F}_L^T \right] \mathbf{C} \left[\mathbf{F}_L (\mathbf{F}_L^T \mathbf{F}_L)^{-1} \mathbf{F}_L^T \right] \quad (151)$$

and

$$\hat{\mathbf{C}}_R = \left[\mathbf{F}_R (\mathbf{F}_R^T \mathbf{F}_R)^{-1} \mathbf{F}_R^T \right] \mathbf{C} \left[\mathbf{F}_R (\mathbf{F}_R^T \mathbf{F}_R)^{-1} \mathbf{F}_R^T \right] \quad (152)$$

for the left and right sensorimotor cortex, respectively. The final current density estimations in the right and left sensorimotor cortex are then

$$\gamma_R = \text{tr}(\mathbf{G}_R^T \hat{\mathbf{C}}_R \mathbf{G}_R) \quad (153)$$

and

$$\gamma_L = \text{tr}(\mathbf{G}_L^T \hat{\mathbf{C}}_L \mathbf{G}_L). \quad (154)$$

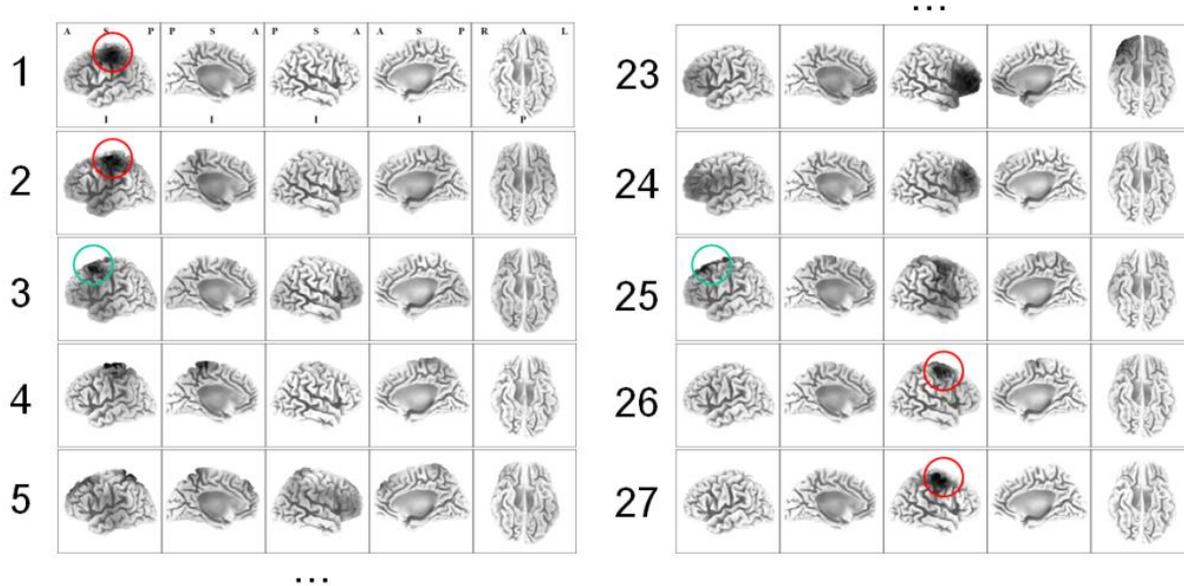


Figure 3.2: sLORETA cortical images of the spatial patterns associated with vector 1-5 and 23-27 of the spatial filter. For each image, from left to right, are shown the left lateral and medial view, the right lateral and medial view and the bottom view. Each image is scaled to its own maximum. The activity is color-coded with black representing the maximum current density and white representing zero. Note that filter 1, 2, are localized in the left sensorimotor cortex and filter 26 and 27 in the right sensorimotor cortex. Based on these results vector 1 and 2 have been used to form the spatial filter F_L for estimating current density in the left sensorimotor cortex, while vector 26 and 27 have been used to form the spatial filter F_R for estimating current density in the right sensorimotor cortex. Filter 3 and 25 are localized in pre-motor areas, however for these data they proved little useful for classification and were not used. Legend: A=Anterior; P=Posterior; S=Superior; I=Inferior; L=Left; R=Right;

Given current density estimation γ_L and γ_R for a given trial, the classification is obtained simply by looking at the difference in current density; if $\gamma_R < \gamma_L$ the synchronization in the right hemisphere has been stronger, therefore the trial is assigned to the “left-hand” class, otherwise it is assigned to the “right hand” class. This procedure yields a very simple classifier with no parameters to be set. The plots of γ_R vs. γ_L are shown in fig. 3.3 for the unfiltered sLORETA method and the CSP-filtered sLORETA method here exposed. Results are presented for the training data (316 trials) and for the test data (100 trials). Using the filtered sLORETA the classification accuracy improves from 73.72% to 83.65% for the training data and from 73% to 83% for the test data. The effect of the CSP filtering can be appreciated as mitigation of the scatter of the plots. These results are consistent and in line with the winner of that BCI competition, who used more features and more complex classification (Wang et al., 2004).

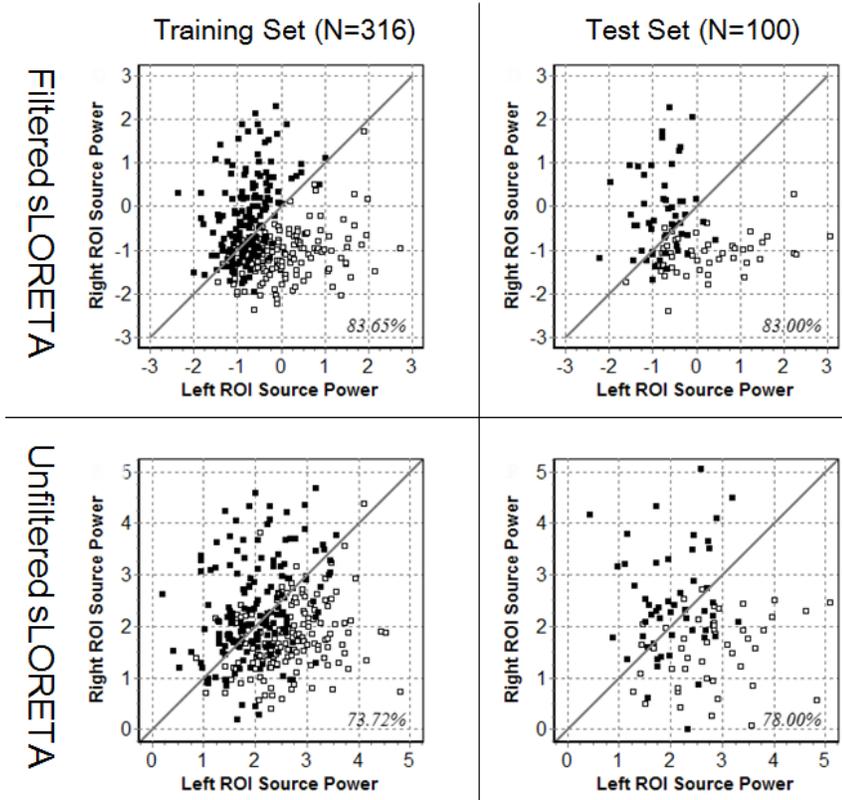


Figure 3.3: In each plot on the x-axis and y-axis is the source power in the left and right region of interest (ROI), respectively. Left column: results on the training set (159 Left + 157 Right trials). Right column: results on the test set (49 Left + 51 Right trials). Top row: results obtained using the CSP filter. Bottom row: results obtained with no filter (raw sLORETA). The untrained classifier is represented by the thick grey line, which has equation $y=x$. Right fingers movement intention trials (black squares) are correctly classified if they fall above the line, while left fingers movement intention trials (white squares) are correctly classified if they fall below the line. The classification accuracy is printed as percentage of correctly classified trials near the bottom-right corner of each plot.

Notice that in (151) and (152) we have employed projectors with general form

$$\mathbf{F} (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T \quad (155)$$

whereas in (146) and (147) the projector has general form

$$\mathbf{U} (\mathbf{U}^T \mathbf{U})^{-1} \mathbf{U}^T = \mathbf{U} \mathbf{U}^T, \quad (156)$$

with the equality due to the orthogonality of the columns of U , which does not hold for the columns of non-orthogonal filters like the CSP¹⁶.

Other Filters for Regional Inverse Solutions

One may design other data-dependent filters with general form (155). For instance, they may derive from a blind source separation or any other decomposition method that we treat in details in chapter IV to VII (see eq. (218)). One may also use *both* a data-dependent and data-independent filter. For data vectors this is achieved feeding equation (144) with filtered data such as (146). For data covariance matrices this is achieved feeding equation (145) with filtered covariance matrices such as (147) or (151)-(152).

Co-Registration of Inverse Solutions with MRI

Throughout this manuscript we present inverse solution images obtained using the free software LORETA-Key (Pascual-Marqui, 2001). A first version was presented briefly in Pascual-Marqui (1999). The newer version of the software makes use of revisited realistic electrode coordinates (Jurcak, Tsuzuki and Dan, 2007) and the head model (and corresponding leadfield matrix) produced by Fuchs *et al.* (2002), applying the boundary element method on the MNI-152 (Montreal neurological institute, Canada) template of Mazziotta *et al.* (2001). This sLORETA-key anatomical template divides and labels the neocortical (including hippocampus and anterior cingulate cortex) MNI-152 volume in 6239 voxels of dimension 5 mm³, based on probabilities returned by the Demon Atlas (Lancaster *et al.*, 2000). The co-registration makes use of the correct translation from the MNI-152 space into the Talairach and Tournoux (1988) space (Brett *et al.*, 2002). The cortical anatomical images are based on the CARET software (van Essen, 2005).

¹⁶ Note that in Congedo, Lotte and Lécuyer (2006) we have used by mistake as projectors $F_L F_L^T$ and $F_R F_R^T$, thus our published results possibly underestimate the classification accuracy of the method.

CHAPTER IV

THE JOINT DIAGONALIZATION FRAMEWORK

Introduction

This and the next three chapters are a long journey into *diagonalization procedures* encompassing a very large range of methods used ubiquitously over the last 20 years in EEG research and practice. This journey will give us an opportunity to summarize the knowledge in this domain, presented here in compact form, but also to present some of our own algorithmic contributions and the studies using them in which we have participated. This chapter gives an ensemble view of *all* diagonalization methods presented in chapter V, VI and VII from the point of view of *optimization theory* and presents a general algorithm for solving the most general case. The reader who is not familiar with optimization methods and blind source separation may want to skip this chapter and maybe come back to it after reading chapters V-VII.

Let $\mathbf{x}(t)$ be the observed EEG data. A linear transformation of EEG data

$$\mathbf{y}(t) = \mathbf{B}^T \mathbf{x}(t) \tag{157}$$

is designed to diagonalize one or more matrices holding statistics of the data. For example, principal component analysis (PCA) is obtained computing the covariance matrix (95) of the data \mathbf{C} and choosing the orthogonal matrix \mathbf{B} in such a way that $\mathbf{B}^T \mathbf{C} \mathbf{B}$ is diagonal. Such a choice outputs transformed time-series (157) with *uncorrelated components*. The simplest diagonalization procedure is indeed the PCA. The most involved is the joint blind source separation (JBSS), which is achieved by diagonalizing several matrices at the same time in each of several data sets. In between them we find many well-known methods such as whitening, the common spatial pattern (CSP), maximal covariance analysis (MCA), canonical correlation analysis (CCA) with its several extensions, extensions of the singular value decomposition to handle several matrices, blind source separation (BSS) methods such as AMUSE, FOBI, SOBI, JADE, etc. The point we want to make here is that all these methods can be conceived as a way to solve the *same general optimization problem*. The general problem involves the diagonalization of $K \geq 1$ different forms of covariance matrices for $M \geq 1$ data sets, where M typically, but not necessarily, refers to the number of subjects analyzed simultaneously¹⁷. The case $M=1$, that is, the single-subject/single data set analysis scenario, is by far the most common.

¹⁷ For instance M may refer to different data modalities or data filtered in different frequency band-pass regions for the same individuals.

However, it can be seen as a special case of the more general setting $M \geq 1$, for which the data takes form

$$\mathbf{x}(t) = \begin{pmatrix} \mathbf{x}_1(t) \\ \dots \\ \mathbf{x}_M(t) \end{pmatrix} \in \mathfrak{R}^{NM} \quad (158)$$

and the M joint linear transformations are given by

$$\begin{pmatrix} \mathbf{y}_1(t) \\ \dots \\ \mathbf{y}_M(t) \end{pmatrix} = \begin{pmatrix} \mathbf{B}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{B}_M \end{pmatrix}^T \begin{pmatrix} \mathbf{x}_1(t) \\ \dots \\ \mathbf{x}_M(t) \end{pmatrix}. \quad (159)$$

Notice that for the sake of simplicity and in order to highlight the modularity of the diagonalization approach presented in this manuscript, setting

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{B}_M \end{pmatrix} \in \mathfrak{R}^{NM \times NM} \quad (160)$$

equation (159) is written as (157) regardless whether $M=1$ or $M>1$. However, there is an important differences in between the two cases: the K covariance matrices that can be derived from the multiple data set (158) have a block structure with M^2 blocks, such as.

$$\begin{pmatrix} \mathbf{C}_{11,k} & \dots & \mathbf{C}_{1M,k} \\ \vdots & \ddots & \vdots \\ \mathbf{C}_{M1,k} & \dots & \mathbf{C}_{MM,k} \end{pmatrix} \in \mathfrak{R}^{NM \times NM}, k \in \{1, \dots, K\}. \quad (161)$$

The $N \times N$ matrices on the diagonal (the diagonal blocks) hold the auto-statistics of the subject (or data sets in general), while the matrices on the off-diagonal (off-diagonal blocks), whenever they are available, hold the cross-statistics between subjects. The reader should pay attention to this composite form of covariance matrices as it will be found over and over again throughout the reminder of this manuscript. Working within setting $M>1$ is useful when we can assume that *data is correlated between-subjects* (or data sets), otherwise, as we will see precisely, the problem reduces to a collection of M independent diagonalization procedures for the case $M=1$. Each diagonalization procedure is defined by a specific choice of M , K and the kind of covariance information contained in the matrices to be diagonalized. However the resulting set of statistics have always general form (161). Whatever is

the diagonalization method, we end up with the task of diagonalizing all available matrices forming the $N \times N$ blocks in (161) by congruent transformation $\mathbf{B}_i^T \mathbf{C}_{ij,k} \mathbf{B}_j$, with $i,j \in \{1, \dots, M\}$. In linear algebra form this is written as

$$\begin{pmatrix} \mathbf{B}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{B}_M \end{pmatrix}^T \begin{pmatrix} \mathbf{C}_{11,k} & \cdots & \mathbf{C}_{1M,k} \\ \vdots & \ddots & \vdots \\ \mathbf{C}_{M1,k} & \cdots & \mathbf{C}_{MM,k} \end{pmatrix} \begin{pmatrix} \mathbf{B}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{B}_M \end{pmatrix} = \begin{pmatrix} \mathbf{Q}_{11,k} & \cdots & \mathbf{Q}_{1M,k} \\ \vdots & \ddots & \vdots \\ \mathbf{Q}_{M1,k} & \cdots & \mathbf{Q}_{MM,k} \end{pmatrix}. \quad (162)$$

As we will see in general the strip-diagonal form (162) cannot be obtained exactly. In a *least-squares optimization framework* the task is to find M matrices $\mathbf{B}_1, \dots, \mathbf{B}_M$ minimizing the sum of squares of the off-diagonal elements of all products $\mathbf{Q}_{ij,k} = \mathbf{B}_i^T \mathbf{C}_{ij,k} \mathbf{B}_j$, that is, making them all at the same time (jointly) as diagonal as possible. The general optimization problem is then written such as

$$\min_{\mathbf{B}_1, \dots, \mathbf{B}_M} \sum_{i,j=1}^M \sum_{k=1}^K \left\| \text{off} \left(\mathbf{B}_i^T \mathbf{C}_{ij,k} \mathbf{B}_j \right) \right\|_F^2. \quad (163)$$

Some structural constraints have to be imposed on matrices $\mathbf{B}_1, \dots, \mathbf{B}_M$ in order to avoid the trivial solution obtained setting them equal to $\mathbf{0}$. We will discuss them later. Table (4.1) lists some of the methods that may be solved by optimization (163), classified depending on the number of observations (K) and data sets (M) involved. Note that very diverse families of methods can be seen as a particular instance of the general optimization (163). In chapter V, VI and VII we will analyze in necessary details all these methods. For each of them we provide the “solutions” as matrices $\mathbf{B}_1^T, \dots, \mathbf{B}_M^T$ and their “inverse”. The solutions extract time series as per (159), reducing to (157) for the case $M=1$, while their inverse are the matrices holding the scalp spatial patterns associated to each extracted component. Chapter V treats several *spatial filters*. Chapter VI treats the family of *blind source separation (BSS)* methods. Chapter VII treats *group BSS* methods. All these families of methods provide linear transformation of the data possessing some statistical property. However, there is a fundamental difference between spatial filtering and blind source separation approaches: filters $\mathbf{B}_1, \dots, \mathbf{B}_M$ and the corresponding scalp spatial patterns can be interpreted physiologically only in the case of blind sources separation; for generic spatial filters they have no physiological meaning.

For all families the general optimization (163) is our unifying framework. The reader will realize that it will be optimized over and over again throughout the next three chapters, reducing to particular cases for each method. Here below is a short overview of the different cases of table 4.1.

Table 4.1: Taxonomy of several signal processing methods depending on the number of observations and data sets involved. All these methods are diagonalization procedures and can be solved by optimization (163). Legend: PCA=principal component analysis; CSP=common spatial pattern; AMUSE=algorithm for multiple source extraction; FOBI=fourth-order blind identification; SOBI=Second-order blind identification; JADE=Joint diagonalization of eigenmatrices; MCA=maximum covariance analysis; CCA=canonical correlation analysis; AJSVD=approximate joint singular value decomposition; JSSS=joint blind source separation.

		Observations (K)		
		1	2	>2
Data Sets (M)	1	PCA, Whitening	CSP, AMUSE, FOBI, ...	BSS: SOBI, JADE...
	2	MCA, CCA	Extended CCA	AJSVD,
	>2	Extended CCA	Extended CCA	JBSS

One data set, one matrix

Suppose we have only one data set ($M=1$), that is, one subject, and one observation ($K=1$), for example, the only matrix to be diagonalized is the sample covariance matrix. Optimization (163) reduces in this case to

$$\min_B \left\| \text{off} \left(\mathbf{B}^T \mathbf{C} \mathbf{B} \right) \right\|_F^2. \quad (164)$$

Constraining the solution \mathbf{B} to be orthogonal the problem is the PCA and the solution is given by the eigenvector matrix of \mathbf{C} . The minimum attainable is zero. With a specific constraint on the norm of the vectors of \mathbf{B} the problem becomes the whitening problem.

One data set, two matrices

Suppose we have only one data set ($M=1$), that is, one subject, but two observations ($K=2$), that is, we require the diagonalization of two forms of covariance matrices. Optimization (163) reduces in this case to

$$\min_{\mathbf{B}} \sum_{k=1}^2 \left\| \text{off} \left(\mathbf{B}^T \mathbf{C}_k \mathbf{B} \right) \right\|_F^2. \quad (165)$$

Depending on the constraints on the norm of the vectors of \mathbf{B} and on the choice of \mathbf{C}_1 and \mathbf{C}_2 the problem is the CSP, AMUSE, FOBI,... and the solution is the generalized eigenvector matrix of some linear combination of \mathbf{C}_1 and \mathbf{C}_2 . The minimum attainable is still zero, but \mathbf{B} will be no more orthogonal, unless \mathbf{C}_1 and \mathbf{C}_2 commute in multiplication.

Two data sets, one matrix

Suppose we have two data sets ($M=2$), for example, two subjects, and only one observation ($K=1$), for example, we require the diagonalization of the covariance matrix of the two data sets stacked as in (158) having composite form (161). Optimization (163) reduces in this case to

$$\operatorname{argmin}_{\mathbf{B}_1, \mathbf{B}_2} \sum_{i,j=1}^2 \left\| \text{off} \left(\mathbf{B}_i^T \mathbf{C}_{ij} \mathbf{B}_j \right) \right\|_F^2. \quad (166)$$

Depending on the constraints on the norm of the vectors of \mathbf{B}_1 and \mathbf{B}_2 the problem is the MCA or CCA problem and the solution is achieved by SVD. The minimum attainable is still zero. \mathbf{B}_1 and \mathbf{B}_2 will be orthogonal for the MCA and non-orthogonal for the CCA.

One data set, several matrices

Suppose we have one data set ($M=1$), for example, one subjects, and many observation ($K>2$), for example, we require the diagonalization of several covariance matrices. Optimization (163) reduces in this case to

$$\operatorname{argmin}_{\mathbf{B}} \sum_k \left\| \text{off} \left(\mathbf{B}^T \mathbf{C}_k \mathbf{B} \right) \right\|_F^2. \quad (167)$$

Depending on the constraints on the norm of the vectors of \mathbf{B} and on the choice of the matrices \mathbf{C}_k the problem yields a wide family of BSS problems (SOBI, JADE, etc.) and the solution is achieved by *approximate joint diagonalization* iterative algorithms. The minimum attainable is no more zero. \mathbf{B} will be non-orthogonal, unless all matrices in the set pair-wise commute in multiplication.

Several data sets, several matrices

Suppose finally we have more than two data sets ($M > 2$), for example many subjects, and many observations ($K > 2$), for example, we require the diagonalization of several covariance matrices with composite form (161). Optimization (163) applies as it is in this case, since we have reached the most general form. The problem is the general JBSS problem. Let us now see how to solve this most general problem.

APPROXIMATE JOINT DIAGONALIZATION

The optimization problems (163)-(167) may be solved by a panel of numerical methods. The cases ($M=1, K=2$) and ($M=2, K=1$) have closed form algebraic solutions. The others can be solved only by iterative algorithms. Here we tackle the most general problem when $M \geq 2$ and/or $K \geq 2$ (163). We present two algorithms for finding matrices $\mathbf{B}_1, \dots, \mathbf{B}_M$, one constraining the matrices to be orthogonal and the other constraining them to be just invertible. These two algorithms apply to simpler problems as well, particularly to the useful problem (167), yielding the many blind source separation solutions by *approximate joint diagonalization (AJD)* that we will treat in chapter VI and VII. In general, the simultaneous diagonalizer of more than two matrices has no closed form solution. There exist many *AJD* algorithms for the single subject case ($M=1, K>2$) (Afsari, 2008; Cardoso and Souloumiac, 1993; Congedo and Pham, 2009; Degerine and Kane, 2007; Fadaili, Moreau and Moreau, 2007; Flury and Gautschi, 1986; Iferroudjene, Abed-Meraim and Belouchrani, 2009; Li and Zhang, 2007; Mesloub, Abeb-Meraim and Belouchrani, 2013; Pham, 2001b; Pham and Congedo, 2009; Souloumiac, 2009, 2011; Tichavsky and Yeredor, 2009; Wang, Liu and Zhang, 2007; Wax and Sheinvald, 1997; Vollgraf and Obermayer, 2006; Yeredor, 2002; Ziehe et al., 2004; Zhou et al., 2008). Recently, extensions to the multisubject case ($M>1, K>2$) that we treat here, have appeared as well (Anderson, Adali and Li, 2012; Vía et al. 2011; Li, Adali and Anderson, 2011; Li et al., 2009). As we are contending, these may be seen as extension of the single-subject *AJD* case.

The *AJD* algorithms proposed so far differ according to whether they estimate directly the demixing matrix or its inverse (the mixing matrix), the restrictions imposed on the matrices that can be diagonalized (Hermitian/symmetric, positive semi-definite, normal), the restrictions imposed on the joint diagonalizer sought (unitary/orthogonal or just invertible), their convergence rate and computational complexity per iteration. More importantly, they differ in terms of the cost function to be optimized. For the sake of efficiency some algorithms rely on heuristics (e.g., Souloumiac, 2009; Tichavsky and Yeredor, 2009; Ziehe et al., 2004), whereas others have focused on the more intuitive formulation based on the general Frobenius norm off-diagonal minimization that we have already introduced in (167) (Congedo and Pham, 2009; Degerine and Kane, 2007; Fadaili, Moreau and Moreau, 2007; Pham and Congedo, 2009; Vollgraf and Obermayer, 2006). The *AJD* algorithms may perform poorly when the true mixing matrix to be estimated or the matrices in the diagonalization set are ill-conditioned. They may be more or less robust to noise, more or less prone to be trapped in local minima, more or less stable, etc. Therefore, the goodness of the *AJD* solution greatly influence the

goodness of the estimation of the (de)mixing matrix and the choice of the *AJD* algorithm is instrumental for the source separation *tout court*. It is out of the scope of this work the throughout analysis and review of *AJD* algorithms. Rather, we here present our contributions in *AJD* algorithm developments. The two algorithms we present here have been published in Congedo, Phlypo and Chatel-Goldman (2012). They are extensions of previous algorithms published in Congedo, Phlypo and Pham (2011), Congedo and Pham (2009) and Pham and Congedo (2009), which are restricted cases of the more general form we present here.

Least-Squares Functional

According to what we have said the optimization for the most general problem (163) leads to cost function

$$\psi^{OFF}(\mathbf{B}) = \sum_{i,j=1}^M \sum_k \left\| \text{Off}(\mathbf{B}_i^T \mathbf{C}_{ij,k} \mathbf{B}_j) \right\|_F^2. \quad (168)$$

Our task is to find the M matrices in (160) making the products

$$\mathbf{Q}_{ij,k} = \mathbf{B}_i^T \mathbf{C}_{ij,k} \mathbf{B}_j \quad (169)$$

as diagonal as possible. The overall strategy is to sequentially update each matrix \mathbf{B}_i , for $i \in \{1, \dots, M\}$ and iterate such sequential search until convergence. In the sequel, let us define the functional of interest for any given i as

$$\psi_{\mathbf{B}_i | \mathbf{B}}^{off} = \sum_k \left\| \text{off}(\mathbf{Q}_{ii,k}) \right\|_F^2 + 2 \sum_{i \neq j=1}^M \sum_k \left\| \text{off}(\mathbf{Q}_{ij,k}) \right\|_F^2 \quad (170)$$

wherein we have separated the products of (169) for $i=j$ (first Frobenius norm) and for $i \neq j$ (second Frobenius norm), corresponding to the diagonal and off-diagonal blocks of matrices in (162). Such partition of the total diagonalization functional is very useful for finding the gradient, but also illustrates exhaustively the fundamental difference between the case $M=1$ and $M>1$. One thing appears suddenly: if we do not consider the $i \neq j$ portion of (170), that is, if we do not assume that *sources are correlated between subjects or data sets*, the functional of interest reduces to a collection of M problem to be solved independently each with $M=1$. In fact, using (170) notice that the optimization of the $i=j$ portion involves for each m only the \mathbf{B}_m block of \mathbf{B} (160). Thus the sequential update strategy

amounts to run M AJD algorithms independently; the update of the \mathbf{B}_m block does not influence the other blocks. On the other hand, using the whole functional in (170) the update of each \mathbf{B}_m block depends on the whole matrix \mathbf{B} . So, each update influence and constraint the others. This dependencies is highlighted by notation $\mathbf{B}_i|\mathbf{B}$ (read: “ \mathbf{B}_i given \mathbf{B} ”). Now, noticing that the functional (168) can be partitioned also as

$$\psi^{OFF}(\mathbf{B}) = \sum_{i,j=1}^M \sum_k \left\| \left(\mathbf{B}_i^T \mathbf{C}_{ij,k} \mathbf{B}_j \right) \right\|_F^2 - \sum_{i,j=1}^M \sum_k \left\| \text{Diag} \left(\mathbf{B}_i^T \mathbf{C}_{ij,k} \mathbf{B}_j \right) \right\|_F^2, \quad (171)$$

we can rewrite it as

$$\psi_{\mathbf{B}_i|\mathbf{B}}^{off} = \psi_{\mathbf{B}_i|\mathbf{B}}^{tot-diag} = \psi_{\mathbf{B}_i|\mathbf{B}}^{tot} - \psi_{\mathbf{B}_i|\mathbf{B}}^{diag}, \quad (172)$$

where the *total* and *diagonal* parts are

$$\psi_{\mathbf{B}_i|\mathbf{B}}^{tot} = \sum_k \left[\text{tr}(\mathbf{Q}_{ii,k}^2) + 2 \sum_{i \neq j} \text{tr}(\mathbf{Q}_{ij,k} \mathbf{Q}_{ij,k}^T) \right] \quad (173)$$

and

$$\psi_{\mathbf{B}_i|\mathbf{B}}^{diag} = \sum_k \left[\sum_{n=1}^N \left(\mathbf{b}_{n(i)}^T \mathbf{C}_{ii,k} \mathbf{b}_{n(i)} \right) + 2 \sum_{i \neq j} \sum_{n=1}^N \left(\mathbf{b}_{n(i)}^T \mathbf{C}_{ij,k} \mathbf{b}_{n(j)} \right) \right], \quad (174)$$

respectively. In Eq. (174), $\mathbf{b}_{(n)i}$ is the n^{th} column vector of \mathbf{B}_i , and $\mathbf{b}_{(n)i}^T$ its transpose.

The Orthogonal Mixing Matrices Case

If we constraint the matrices \mathbf{B}_m to be orthogonal as in Cardoso and Souloumiac (1993) the first expression in the right-end side of the cost function (171) is

$$\sum_{i,j} \sum_k \left\| \mathbf{B}_i^T \mathbf{C}_{ij,k} \mathbf{B}_j \right\|_F^2 = \text{tr} \left[\sum_{i,j=1}^M \mathbf{B}_i^T \sum_k \left(\mathbf{C}_{ij,k} \mathbf{B}_j \mathbf{B}_j^T \mathbf{C}_{ij,k}^T \right) \mathbf{B}_i \right], \quad (175)$$

which for the orthogonality of all matrices \mathbf{B}_m and for the rotation-invariance property of the trace simplifies to

$$\text{tr} \left[\sum_{i,j=1}^M \sum_k (\mathbf{C}_k \mathbf{C}_k^T) \right]. \quad (176)$$

Hence, this term does not depend on \mathbf{B} and can be dropped. Hence, the cost function (171) reduces to

$$\psi^{OFF}(\mathbf{B}) = - \sum_{i,j=1}^M \sum_k \left\| \text{Diag}(\mathbf{B}_i^T \mathbf{C}_{ij,k} \mathbf{B}_j) \right\|_F^2 \quad (177)$$

and we are left with the problem of maximizing iteratively the *diag* functional in (174), for $i \in \{1, \dots, M\}$. Let us rewrite the objective function (174) as

$$\psi_{\mathbf{B}_i | \mathbf{B}}^{diag} = \sum_k \left(\left\| \text{diag}(\mathbf{Q}_{ii,k}) \right\|_F^2 + 2 \sum_{i \neq j} \left\| \text{diag}(\mathbf{Q}_{ij,k}) \right\|_F^2 \right), \quad (178)$$

and then as

$$\psi_{\mathbf{B}_i | \mathbf{B}}^{diag} = \sum_k \left[\text{tr} \sum_n (\mathbf{E}_n \mathbf{Q}_{ii,k} \mathbf{E}_n \mathbf{E}_n \mathbf{Q}_{ii,k} \mathbf{E}_n) + 2 \sum_{i \neq j} \left(\text{tr} \sum_n (\mathbf{E}_n \mathbf{Q}_{ij,k} \mathbf{E}_n \mathbf{E}_n \mathbf{Q}_{ji,k} \mathbf{E}_n) \right) \right], \quad (179)$$

where matrix \mathbf{E}_n is the elementary matrix filled with entry 1 at position (n,n) and 0 elsewhere. The above functional is a matrix polynomial of second degree in \mathbf{B}_i . The derivative is of third degree in \mathbf{B}_i for the first trace and of first degree in \mathbf{B}_i for the second trace. However, using the symmetry of matrices $\mathbf{C}_{ii,k}$, the gradient simplifies to

$$\frac{\partial \psi_{\mathbf{B}_i | \mathbf{B}}^{diag}}{\partial \mathbf{B}_i} = \sum_k \left[4 \sum_n (\mathbf{C}_{ii,k} \mathbf{B}_i \mathbf{E}_n \mathbf{E}_n \mathbf{Q}_{ii,k} \mathbf{E}_n) + 4 \sum_{i \neq j} \sum_n (\mathbf{C}_{ij,k} \mathbf{B}_j \mathbf{E}_n \mathbf{E}_n \mathbf{Q}_{ji,k} \mathbf{E}_n) \right]. \quad (180)$$

Thus

$$\frac{\partial \psi_{\mathbf{B}_i | \mathbf{B}}^{diag}}{\partial \mathbf{B}_i} = 4 \left(\mathbf{R}_{(i)(n)} \mathbf{b}_{n(i)}, \dots, \mathbf{R}_{(i)(N)} \mathbf{b}_{N(i)} \right), \quad (181)$$

where

$$\mathbf{R}_{(i)(n)} = \sum_k \sum_{j=1}^M (\mathbf{C}_{ij,k} \mathbf{b}_{n(j)} \mathbf{b}_{n(j)}^T \mathbf{C}_{ij,k}^T), \quad (182)$$

In words, the gradient of the N vectors of \mathbf{B}_i should be taken as the eigenvectors of corresponding matrices $\mathbf{R}_{(i)(n)}$ associated with their largest eigenvalue, for all $n \in \{1, \dots, N\}$. In order to update these

vectors we limit ourselves to a single pass of power iterations (Congedo and Pham, 2009; Congedo, Phlypo and Pham, 2011; Pham and Congedo, 2009). After updating all vectors of \mathbf{B}_i we need to orthogonalize \mathbf{B}_i so as to ensure that at each step \mathbf{B}_i stays in the orthogonal group. Therefore, we have the following simple updating rule:

$$\text{for all } i=1 \text{ to } M \text{ do } \left\{ \begin{array}{l} \mathbf{B}_i \leftarrow \left(\mathbf{R}_{(i)(n)} \mathbf{b}_{n(i)}, \dots, \mathbf{R}_{(i)(N)} \mathbf{b}_{N(i)} \right) \\ \text{orthogonalize } \mathbf{B}_i \end{array} \right\} \quad (183)$$

The iterative algorithm is summarized here below:

Algorithm(184): Orthogonal Joint Least-Squares Diagonalization (OJLSD).

Optimization (167).

Initialize $\mathbf{B}_1, \dots, \mathbf{B}_M$ by orthogonal clever guesses or by \mathbf{I} if no guess is available.

Repeat

For $i=1$ to M do

Obtain $\mathbf{R}_{(i)(1)}, \dots, \mathbf{R}_{(i)(N)}$ by (182)

For $n=1$ to N do one pass of power iterations $\mathbf{b}_{n(i)} \leftarrow \mathbf{R}_{(i)(n)} \mathbf{b}_{n(i)}$

Make \mathbf{B}_i orthogonal by Lödwin orthogonalization (75)

End For i

Until Convergence

(The sum of difference of $\sum_i \|\mathbf{B}_i\|_F$ in two successive iterations is smaller than ε)

Note that in practice the orthogonalization is computed as $\mathbf{B}_i \leftarrow \mathbf{U}\mathbf{V}^T$, where $\mathbf{U}\mathbf{A}\mathbf{V}^T$ is the SVD of \mathbf{B}_i .

The Invertible Mixing Matrices Case

In this case the total function is not invariant to \mathbf{B} , hence we need to explicitly minimize the whole off functional in (168). Furthermore we need to avoid the trivial solution $\mathbf{B}_m^T = \mathbf{0}$, for any $m \in \{1, \dots, M\}$.

Several constraints on \mathbf{B} may serve this purpose. For example orthogonality of \mathbf{B}_m , such that $\mathbf{B}_m^T \mathbf{B}_m = \mathbf{I}$, which we have just considered, or $\text{diag}(\mathbf{B}_m^T \mathbf{C}_0 \mathbf{B}_m) = \mathbf{I}$, where \mathbf{C}_0 is an arbitrary symmetric positive-definite matrix (Degerine and Kane, 2007; Vollgraf and Obermayer, 2006). In Congedo and Pham (2009) we have tackled the non-orthogonal (invertible) solution and introduced for the first time an *intrinsic constraint*, that is, a constraint on the norms of the vectors of \mathbf{B}_m that does not depend on any matrices external to the diagonalization set. We minimize the off functional with a constraint (w.c.) on the norm of the column vectors of \mathbf{B}_m , such as

$$\Psi_{\mathbf{B}_i|\mathbf{B}}^{\text{tot-dia}}, \text{ w.c. } \mathbf{b}_{n(i)}^T \mathbf{R}_{(i)(n)} \mathbf{b}_{n(i)} = 1, \forall n \in \{1, \dots, N\}, \quad (185)$$

where matrices $\mathbf{R}_{(i)(n)}$ are given in (182). We apply the method of Lagrange multipliers, leading us to minimize

$$\begin{aligned} & \sum_k \left[2 \left(\sum_{i \neq j} \text{tr}(\mathbf{Q}_{ij,k} \mathbf{Q}_{ji,k}) \right) + \text{tr}(\mathbf{Q}_{ij,k}^2) \right] \\ & - \sum_k \left[4 \left(\sum_{i \neq j} \sum_n v_n (\mathbf{C}_{ij,k} \mathbf{B}_j \mathbf{E}_n^2 \mathbf{Q}_{ji,k} \mathbf{E}_n) \right) + 4 \left(\sum_n v_n (\mathbf{C}_{ii,k} \mathbf{B}_i \mathbf{E}_n^2 \mathbf{Q}_{ii,k} \mathbf{E}_n) \right) \right] \end{aligned} \quad (186)$$

where the multipliers v_n are adjusted in order to satisfy constraint (185). Using the symmetry of the matrices $\mathbf{C}_{ii,k}$ and exploiting the previous gradient results in (181) for the *diag* part, the gradient of the Lagrangian reads

$$\frac{\partial L(\Psi_{\mathbf{B}_i|\mathbf{B}}^{\text{tot-dia}})}{\partial \mathbf{B}_i} = \sum_k \left[4 \sum_{j \neq i} (\mathbf{C}_{ij,k} \mathbf{B}_j \mathbf{Q}_{ij,k}) + 4 (\mathbf{C}_{ii,k} \mathbf{B}_i \mathbf{Q}_{ii,k}) \right] - \frac{\partial \Psi_{\mathbf{B}_i|\mathbf{B}}^{\text{diag}}}{\partial \mathbf{B}_i}, \quad (187)$$

which is

$$\frac{\partial L(\Psi_{\mathbf{B}_i|\mathbf{B}}^{\text{tot-dia}})}{\partial \mathbf{B}_i} = 4 \mathbf{\Gamma}_{(i)} \mathbf{B}_i - 4 \left(\mathbf{R}_{(i)(1)} \mathbf{b}_{1(i)}, \dots, \mathbf{R}_{(i)(N)} \mathbf{b}_{N(i)} \right). \quad (188)$$

Notice that matrices $\mathbf{\Gamma}_{(i)}$ are the sum of the corresponding N $\mathbf{R}_{(i)(n)}$ matrices found in (182), that is

$$\mathbf{\Gamma}_{(i)} = \sum_k \sum_{j=1}^M (\mathbf{C}_{ij,k} \mathbf{B}_j \mathbf{B}_j^T \mathbf{C}_{ij,k}^T) = \sum_n \mathbf{R}_{(i)(n)}. \quad (189)$$

Setting the gradient to zero gives us a stationary point for our optimization, which for each \mathbf{B}_i is given by

$$\Gamma_{(i)} \mathbf{B}_i = \left(\mathbf{R}_{(i)(1)} \mathbf{b}_{1(i)}, \dots, \mathbf{R}_{(i)(N)} \mathbf{b}_{N(i)} \right). \quad (190)$$

This is a set of N generalized eigenvalue-eigenvector problems for matrix pencils $(\Gamma_{(i)}, \mathbf{R}_{(i)(1)}), \dots, (\Gamma_{(i)}, \mathbf{R}_{(i)(N)})$. The N vector of \mathbf{B}_i are the principal eigenvector of the N matrices $\Gamma_{(i)}^{-1} \mathbf{R}_{(i)(1)}, \dots, \Gamma_{(i)}^{-1} \mathbf{R}_{(i)(N)}$. In practice, in order to find the principal eigenvectors avoiding the inversion of $\Gamma_{(i)}$, we can update the vector $\mathbf{b}_{n(i)}$ by a single-pass of a special power iteration. We compute the Cholesky decomposition (49) $\Gamma_{(i)} = \mathbf{L}\mathbf{L}^T$, with \mathbf{L} lower triangular and then solve two systems of linear equations by forward and backward substitution (Golub and van Loan, 1996, p. 88-89). Limiting ourselves to one single power iteration per update step yields algorithm:

Algorithm (191): Invertible Joint Least-Squares Diagonalization (IJLSD). Optimization (167).

Initialize $\mathbf{B}_1, \dots, \mathbf{B}_M$ by non-singular clever guesses or by \mathbf{I} if no guess is available.

Repeat

For $i=1$ to M do

Obtain $\mathbf{R}_{(i)(1)}, \dots, \mathbf{R}_{(i)(N)}$ by (182) and heir sum (189) $\sum_n \mathbf{R}_{(i)(n)} = \Gamma_{(i)}$

Obtain Cholesky decomposition (49) $\Gamma_{(i)} = \mathbf{L}\mathbf{L}^T$

For $n=1$ to N do solve $\mathbf{L}\mathbf{v} = \mathbf{R}_{(i)(n)} \mathbf{b}_{n(i)}$ for \mathbf{v} , $\mathbf{L}^T \mathbf{z} = \mathbf{v}$ for \mathbf{z} and update $\mathbf{b}_{n(i)} \leftarrow \mathbf{z} \left(\mathbf{z}^T \mathbf{R}_{(i)(n)} \mathbf{z} \right)^{-1/2}$

End For m

Until Convergence (The difference of $\sum_m \|\mathbf{B}_m\|_F$ in two successive iterations is smaller than ε).

Simulations

In this section, taken from Congedo, Phlypo and Chatel-Goldman (2012), the behavior of the proposed OJLSD and IJLSD algorithms is assessed by means of simulations¹⁸. For simulations input matrices $\mathbf{C}_{ij,k}$ are generated according to the model

¹⁸ Notice that in that paper the two algorithms were given a different acronym and we have used a different notation.

$$\begin{pmatrix} \mathbf{C}_{11,k} & \cdots & \mathbf{C}_{1M,k} \\ \vdots & \ddots & \vdots \\ \mathbf{C}_{M1,k} & \cdots & \mathbf{C}_{MM,k} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{A}_M \end{pmatrix} \begin{pmatrix} \mathbf{D}_{11,k} & \cdots & \mathbf{D}_{1M,k} \\ \vdots & \ddots & \vdots \\ \mathbf{D}_{M1,k} & \cdots & \mathbf{D}_{MM,k} \end{pmatrix} \begin{pmatrix} \mathbf{A}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{A}_M \end{pmatrix}^T + \begin{pmatrix} \mathbf{N}_{11,k} & \cdots & \mathbf{N}_{1M,k} \\ \vdots & \ddots & \vdots \\ \mathbf{N}_{M1,k} & \cdots & \mathbf{N}_{MM,k} \end{pmatrix} \quad (192)$$

where the matrices $\mathbf{A}_1, \dots, \mathbf{A}_M$ (mixing matrices, see chapter VI) are the inverse of $\mathbf{B}_1^T, \dots, \mathbf{B}_M^T$. Matrices $\mathbf{D}_{ij,k}$ are generated as square diagonal matrices with each diagonal entry randomly distributed as a chi-squared with M degrees of freedom and divided by M . Noise is added to the generated matrices $\mathbf{C}_{ij,k}$ as additive matrices $\mathbf{N}_{ij,k}$ symmetric and with entries randomly Gaussian distributed with zero mean and σ standard deviation (sd). The parameter σ controls the signal to noise ratio of the input matrices. Several different values of σ are considered in the simulations. The data generating matrices \mathbf{A}_m , $m \in \{1, \dots, M\}$, are generated as orthogonal or non-orthogonal invertible. Orthogonal matrices are generated by first generating a matrix with entries randomly drawn from a Gaussian distribution with zero mean and $\text{sd}=1$ and then taking its left singular vector matrix. In this case the conditioning of the mixing matrices does not jeopardize the performance of the algorithms and we can evaluate their robustness with respect to noise. In order to generate non-orthogonal matrices, the matrices generated as above are perturbed by adding to each entry a number randomly drawn from a Gaussian distribution with zero mean and $\text{sd}=1/2$. In this case the mixing matrices have variable conditioning and we can evaluate the behavior of the algorithms with respect to the conditioning of the mixing matrices.

The algorithms estimate the matrices $\mathbf{B}_1^T, \dots, \mathbf{B}_M^T$ which, according to BSS theory (see chapter VI), should approximate the pseudo-inverse of actual mixing matrices $\mathbf{A}_1, \dots, \mathbf{A}_M$ up to row scaling (including sign) and global permutation. Then, matrices $\mathbf{\Phi}_m = \mathbf{B}_m^T \mathbf{A}_m$ should approximate as much as possible a scaled permutation matrix. For each estimated demixing matrix we consider the Amari-like performance index (Comon and Jutten, 2010), which is computed as

$$\pi_m = \left[\sum_r \left(\frac{\sum_c |\phi_{m,rc}|}{\max_c |\phi_{m,rc}|} - 1 \right) + \sum_c \left(\frac{\sum_r |\phi_{m,rc}|}{\max_r |\phi_{m,rc}|} - 1 \right) \right] / 2N(N-1) \quad (193)$$

where indexes $r, c \in \{1, \dots, N\}$ are the rows and columns of matrices $\mathbf{\Phi}_m$, $\phi_{m,rc}$ is the (r, c) entry of matrix $\mathbf{\Phi}_m$, and $|\cdot|$ denotes the absolute value of the argument. We define the composite performance as a function of the geometric mean of the performance indexes obtained over the M matrices, such as

$$\pi = -\log_{10} \left(1 - \sqrt[M]{\prod_m (1 - \pi_m)} \right) \quad (194)$$

Values of π above two indicate a very good performance. The higher the value of the composite index (194), the higher the performance. Likewise, the composite conditioning with respect to matrix inverse of the mixing matrices is defined as

$$\chi = \log_{10} \left(\prod_m \left| \frac{\max \text{eig}(\mathbf{A}_m)}{\min \text{eig}(\mathbf{A}_m)} \right| \right), \quad (195)$$

where $\max \text{eig}$ and $\min \text{eig}$ are the largest and smallest eigenvalue of the argument, respectively. Figure 4.1 shows the performance (194) obtained by the OJLSD and IJLSD algorithms with orthogonal mixing (input) matrices, $N=P=3$ sensors/sources and several combinations of M (number of datasets), K (number of observations), and σ (noise level). One hundred simulations have been performed for each algorithm and for each combination of M , K and σ . Each dot represent the intersection of the performance obtained in one simulation when the algorithm is initialized with identity matrices (x-axis), that is, possibly far from the optimal solution, and with the exact solutions (y-axis). Dots lying on the 45° line indicate that the algorithms have a stable attractor, despite the added noise. Dots lying above the 45° line indicate that the algorithm gets far from the exact solution. These results show that for both algorithms the degradation engendered by noise is mitigated by increasing either K or M . Also, when either M or K are much larger than N no divergence of the algorithms is noticed. Overall, IJLSD appears more stable than OJLSD.

Figure 4.2 shows the performance of the IJLSD algorithm (y-axis) vs. the mixing matrices conditioning (195) with non-orthogonal mixing (input) matrices, $N=P=3$ and several combinations of M , K , and σ . Results show that the more noise there is in the system the more the conditioning affects the performance. Furthermore, the degradation is not mitigated by increasing K and only moderately mitigated by increasing M .

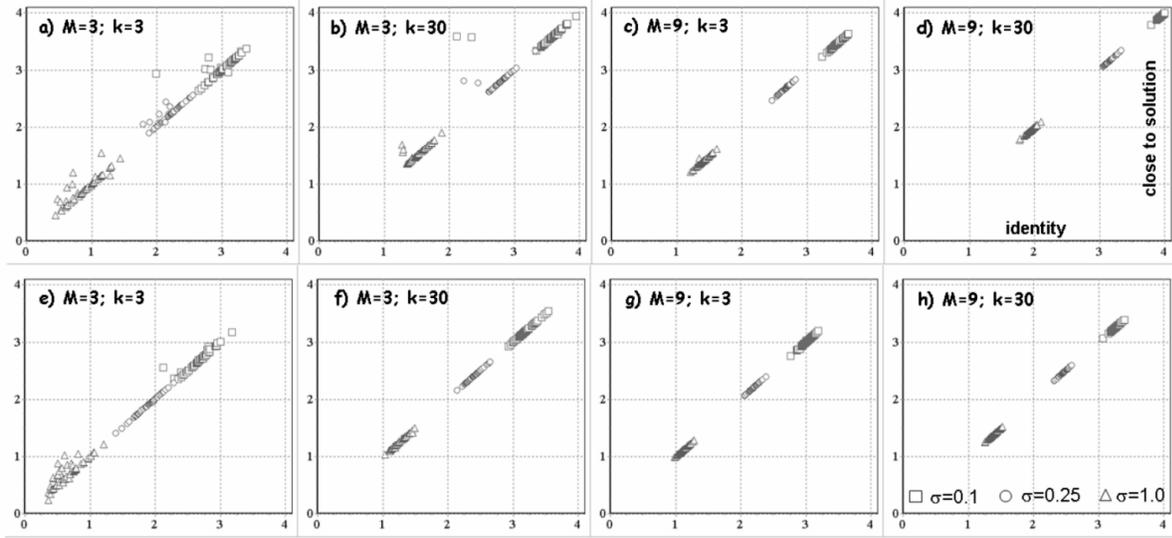


Figure 4.1: Composite performance of the OJLSD (top row: a to d) and IJLSD (bottom row: e to h) algorithms (194) when initialized with the identity matrices (x -axis) vs. when initialized with the inverse of the actual mixing matrices (y -axis), for $N=P=3$, three noise levels (σ) and several combinations of M and K .

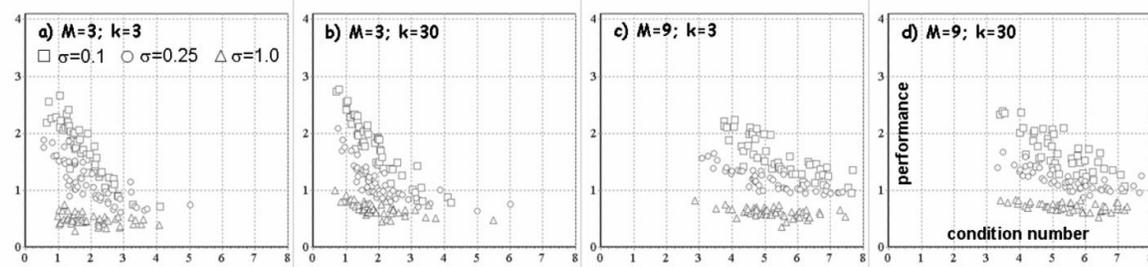


Figure 4.2: Composite performance of the ILSD algorithm (y -axis) vs. composite condition number (195) of mixing matrices. Same parameters as fig. 4.1.

Conclusion

We have presented two algorithms for solving in a least-squares framework the most general joint diagonalization problem (163). These algorithms will be considered to solve the joint blind source separation problem discussed in more details in chapter VII, where a further extension is also presented. These algorithms are very simple and can be implemented rapidly. Their computational complexity per iteration is low, however their convergence is only linear. Furthermore our simulations show that both OJLSD and IJLSD may be prone to be trapped in local minima in the presence of noise. As such, for single-subject ($M=1$) AJD problem we keep using state of the art AJD algorithms

such as the one by Tichavsky and Yeredor (2009). On the other hand for the case $M>1$ these algorithms are competitive against the state of the art. Future research will establish how to optimize the relevant functionals for the optimal weighted sum of input matrices in order to improve the convergence toward the optimal solution.

CHAPTER V

SPATIAL FILTERS

Introduction

Because of the linear instantaneous conduction of dipolar activity discussed in chapter II, each electrode on the scalp records a weighted sum of the underlying sources. As a consequence the covariance matrix of EEG data is highly non-diagonal, meaning that the EEG channels are highly correlated. The analysis of the data in the sensor space is problematic for this reason. Since the EEG acquired with multiple electrodes covering the whole scalp contains a considerable amount of spatial information, we can design spatial filters in order to represent the data in a different space possessing some desirable statistical property. Typically, such space is obtained enforcing uncorrelation, that is, diagonalization of some matrices holding statistics. What property the transformed data should possess is the rationale of each method and depends on the application. Referring to the arguments advanced in chapter IV, the spatial filters presented in this chapter belongs all to the case $K=1$ or $K=2$ and $M=1$ or $M=2$, where at least one between M or K equals 1.

Principal Component Analysis (PCA)

PCA was proposed by Karl Pearson (1901) and Harold Hotelling (1933). Let $\mathbf{X} \in \mathfrak{R}^{N \times T}$ be a data segment composed of T samples drawn from process $\mathbf{x}(t) \in \mathfrak{R}^N$ and let \mathbf{C} its covariance matrix (95). Being \mathbf{C} a positive-definite symmetric matrix, eq. (37) states that its eigenvector matrix \mathbf{U} diagonalizes \mathbf{C} by rotation as $\mathbf{U}^T \mathbf{C} \mathbf{U} = \mathbf{A}$. The eigenvalues in \mathbf{A} after such rotation are all positive. The linear transformation $\mathbf{Y} = \mathbf{U}^T \mathbf{X}$ yields uncorrelated data with variance of the n^{th} components equal to the respective eigenvalue λ_n , that is, $\frac{1}{T-1} \mathbf{Y} \mathbf{Y}^T = \mathbf{A}$. We always arrange the eigenvectors in such a way that their respective eigenvalues are sorted in descending order of variance, such as $\lambda_1 \geq \dots \geq \lambda_N$. Then because of (47) the first component (time-series) of \mathbf{Y} holds the linear combination of \mathbf{X} with maximal variance, the second the linear combination with maximal residual variance and so on, subject to $\mathbf{U}^T \mathbf{U} = \mathbf{I}$. This is known as *principal component analysis* (PCA).

Algorithm (196): Principal Component Analysis (PCA) . Optimization (164).

Do $\text{EVD}(\mathbf{C}) = \mathbf{U}^T \mathbf{A} \mathbf{U}$

The solutions is \mathbf{U}^T and its inverse \mathbf{U} .

The time course and related spatial pattern of the PCA are not informative for interpreting EEG data in general, nonetheless the PCA remains possibly the most common and useful pre-processing step, with applications in data compression (dimensionality reduction), classification, eye-blinks extraction and much more.

Whitening

As before let C be the covariance matrix of a data segment. From (37) it follows that the matrix

$$\mathbf{W}^T = \mathbf{A}^{-\frac{1}{2}} \mathbf{U}^T \quad (197)$$

diagonalizes and standardizes C by rotation and axes stretching, yielding $\mathbf{W}^T \mathbf{C} \mathbf{W} = \mathbf{I}$. Linear transformation $\mathbf{Y} = \mathbf{W}^T \mathbf{X}$ then yields uncorrelated data with unit variance of all components. This is called *whitening* or *sphering*. Whitened data stays whitened after whatever further rotation. That is, for any orthogonal matrix \mathbf{V} it holds

$$\mathbf{V}^T \left[\frac{1}{T-1} (\mathbf{Y} \mathbf{Y}^T) \right] \mathbf{V} = \mathbf{V}^T \mathbf{I} \mathbf{V} = \mathbf{I}. \quad (198)$$

Hence there exist an infinite number of possible whitening matrices with general form $\mathbf{V}^T \mathbf{A}^{-\frac{1}{2}} \mathbf{U}^T$. Particularly important is the only symmetric one, which is the symmetric square root inverse (SSRI) $\mathbf{C}^{-\frac{1}{2}} = \mathbf{U} \mathbf{A}^{-\frac{1}{2}} \mathbf{U}^T$ (52) (Hyvärinen, Karhunen and Oja, 2001). It has been shown that if the mixing matrix is symmetric the SSRI whitening is the solution to the BSS problem (Cichocki and Georgiev, 2003, theorem 1), but the mixing matrix is never symmetric with EEG data. Nonetheless whitening plays a fundamental role as a first step in many of the blind source separation algorithms we will encounter in chapter VI and VII and in countless other algorithms and situations.

Algorithm (199): Whitening . Optimization (164).

Do $EVD(\mathbf{C}) = \mathbf{U}^T \mathbf{A} \mathbf{U}$

The solutions is $\mathbf{W}^T = \mathbf{V}^T \mathbf{A}^{-\frac{1}{2}} \mathbf{U}^T$ and its inverse is $\mathbf{U} \mathbf{A}^{\frac{1}{2}} \mathbf{V}$, with \mathbf{V} any orthogonal matrix.

Typically \mathbf{V}^T is taken as the identity or as \mathbf{U} , yielding in this latter case the SSRI (52).

Common Spatial Pattern (CSP)

A useful instance of the generalized eigenvalue-eigenvector decomposition (82) is the simultaneous diagonalization of two covariance matrices C_1, C_2 computed on data segments X_1 and X_2 drawn from two processes $\mathbf{x}_1(t), \mathbf{x}_2(t) \in \mathfrak{R}^N$ and of their sum $C = C_1 + C_2$ (Fukunaga, 1990, p. 31-33), yielding a spatial filter known as *common spatial pattern*. A solution in this case is given by the matrix holding the eigenvectors of $C^{-1}C_1$ or, equivalently, of $C^{-1}C_2$. The desired properties, however, are obtained with a two-step procedure:

Algorithm (200): Common Spatial Pattern. Optimization (165)

Given matrices C_1 and C_2 compute their sum $C = C_1 + C_2$, its SSR $C^{1/2}$ (51) and SSRI (52) $C^{-1/2}$
do $EVD(C^{-1/2}C_1C^{-1/2}) = U_1 A_1 U_1^T$
The solution is $B^T = U_1^T C^{-1/2}$ and its inverse is $A = C^{1/2} U_1$.

This solution verifies

$$\begin{cases} B^T C B = I \\ B^T C_1 B = A_1 \\ B^T C_2 B = A_2 = I - A_1 \end{cases} \quad (201)$$

That is to say, after congruent transformation, $\lambda_1 \geq \dots \geq \lambda_N$ are the eigenvalues of C_1 and $1 - \lambda_1 \leq \dots \leq 1 - \lambda_N$ are the corresponding eigenvalues of C_2 . We see that this particular *GEVD* maximizes the ratio between the variance of the corresponding components of transformed processes $Y_1 = B^T X_1$ and $Y_2 = B^T X_2$, ratio which is given naturally ordered by descending order as

$$\lambda_1 / (1 - \lambda_1) \geq \dots \geq \lambda_N / (1 - \lambda_N). \quad (202)$$

If C_1 and C_2 are covariance matrices of two classes, the CSP is the joint diagonalizer maximizing the ratio between them, insomuch such decomposition is used for separating two classes, for example, in BCI based on motor imagery. An efficient spatial filter is obtained forming matrix $F = (F_1, F_2) \in \mathfrak{R}^{N \times 2P}$ keeping in matrix $F_1 = (\mathbf{b}_1, \dots, \mathbf{b}_P)$ only $P \ll N$ among the first vectors of B and in

$\mathbf{F}_2 = (\mathbf{b}_{N-P+1}, \dots, \mathbf{b}_N)$ the corresponding last P vectors of \mathbf{B} (typically $P=2$ or $P=3$ in BCI). The filter \mathbf{F} then explains much of the variance ratio between the two classes. Given the covariance matrix of an unknown trial, filtering with \mathbf{F} effectively removes components that are not useful for the classification. Since its inception in EEG (Koles, 1991; Koles and Song, 1998) and then in BCI research (Guger, Ramoser and Pfurtscheller (2000); Ramoser, Muller-Gerking, Pfurtscheller, 2000), the CSP with its adaptations and extensions has become the most popular spatial filter for improving class separability in BCI (Lotte et al., 2007; Lotte and Guan, 2011).

Maximum Covariance Analysis (MCA)

When two EEG data sets are recorded simultaneously, i.e., there is a one-to-one correspondence between the samples of the two data sets, we may be interested in analyzing their covariance. The extension of the covariance bivariate measure for multivariate measurements is the *maximum covariance analysis* (MCA). It can be conceived as the extension of a PCA on two data sets simultaneously ($M=2$). Given synchronized processes $\mathbf{x}_1(t)$, $\mathbf{x}_2(t)$ from which we draw data segments of T samples, let us stack vertically the data such as

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \in \mathfrak{R}^{2N \times T}, \quad (203)$$

From which we have covariance matrix with form

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_1 & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_2 \end{pmatrix} \in \mathfrak{R}^{2N \times 2N}, \quad (204)$$

the MCA finds two orthogonal bases, to be applied one to \mathbf{X}_1 and one to \mathbf{X}_2 , maximizing their covariance. Note that by construction $\mathbf{C}_{12} = \mathbf{C}_{21}^T$. We are looking for two orthogonal matrices \mathbf{U}_1 and \mathbf{U}_2 yielding

$$\mathbf{U}_1^T \mathbf{C}_{12} \mathbf{U}_2 = \mathbf{U}_2^T \mathbf{C}_{21} \mathbf{U}_1 = \mathbf{A}, \quad (205)$$

where sorted singular values $\lambda_1 \geq \dots \geq \lambda_N$ of the third expression hold the maximal covariances between the transformed data $\mathbf{Y}_1 = \mathbf{U}_1^T \mathbf{X}_1$ and $\mathbf{Y}_2 = \mathbf{U}_2^T \mathbf{X}_2$ under constraint of orthogonality of \mathbf{U}_1 and \mathbf{U}_2 . The set of equations (205) can be written in linear algebra form as

$$\begin{pmatrix} \mathbf{U}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_2 \end{pmatrix}^T \begin{pmatrix} \mathbf{C}_1 & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_2 \end{pmatrix} \begin{pmatrix} \mathbf{U}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{A} \\ \mathbf{A} & \mathbf{A}_2 \end{pmatrix} \quad (206)$$

The new bases are found as per

Algorithm (207): Maximal Covariance Analysis (MCA). Optimization (166).

Do $SVD(\mathbf{C}_{12}) = \mathbf{U}_1 \mathbf{A} \mathbf{U}_2^T$

The solutions are \mathbf{U}_1^T and \mathbf{U}_2^T and their inverses are \mathbf{U}_1 and \mathbf{U}_2 , respectively.

Notice that the two data sets may be defined in any way, for example, they may be the recordings on two individuals, the electrodes on the right and left hemisphere on one individual, EEG and EMG electrodes, etc. As such the MCA method is pretty general. It should be kept in mind that MCA is sensitive to the amplitude of each process since the covariance is maximized and not the correlation. Thus, the covariance will be driven by the process with highest amplitude.

Canonical Correlation Analysis (CCA)

Much more common in these situations is the use of *canonical correlation analysis* (CCA). If the MCA is the multivariate extension of the bivariate covariance, the CCA is the multivariate extension of the bivariate (Pearson product-moment) correlation. It should better be called *maximum correlation analysis*, but the name has been kept for historical reasons (Hotelling, 1936). We work in the same framework as the previous paragraph on MCA. CCA finds matrices \mathbf{B}_1 and \mathbf{B}_2 such that

$$\begin{cases} \mathbf{B}_1^T \mathbf{C}_1 \mathbf{B}_1 = \mathbf{I} \\ \mathbf{B}_2^T \mathbf{C}_2 \mathbf{B}_2 = \mathbf{I} \\ \mathbf{B}_1^T \mathbf{C}_{12} \mathbf{B}_2 = \mathbf{B}_2^T \mathbf{C}_{21} \mathbf{B}_1 = \mathbf{A} \end{cases}, \quad (208)$$

with \mathbf{A} holding in diagonal the correlation sorted by descending order $\lambda_1 \geq \dots \geq \lambda_N$ of the transformed data $\mathbf{Y}_1 = \mathbf{B}_1^T \mathbf{X}_1$ and $\mathbf{Y}_2 = \mathbf{B}_2^T \mathbf{X}_2$. In linear algebra form the set of equation (208) reads

$$\begin{pmatrix} \mathbf{B}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_2 \end{pmatrix}^T \begin{pmatrix} \mathbf{C}_1 & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_2 \end{pmatrix} \begin{pmatrix} \mathbf{B}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{A} \\ \mathbf{A} & \mathbf{I} \end{pmatrix}. \quad (209)$$

Notice that the difference with the MCA is only a scaling of the eigenvalues of the transformed data, however matrices \mathbf{B}_1 and \mathbf{B}_2 are no longer orthogonal. The solutions to the CCA \mathbf{B}_1 and \mathbf{B}_2 are given by the eigenvector matrix of (non-symmetric) $\mathbf{C}_1^{-1}\mathbf{C}_{12}\mathbf{C}_2^{-1}\mathbf{C}_{21}$ and $\mathbf{C}_2^{-1}\mathbf{C}_{21}\mathbf{C}_1^{-1}\mathbf{C}_{12}$, respectively. A numerically preferable solution, accommodating also the case where the dimension of $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$ are different, is the following two-step procedure:

Algorithm (210): Canonical Correlation Analysis. Optimization (166).

For matrices \mathbf{C}_1 and \mathbf{C}_2 compute their SSRI (52) $\mathbf{C}_1^{-1/2}, \mathbf{C}_2^{-1/2}$ and their SSR (51) $\mathbf{C}_1^{1/2}, \mathbf{C}_2^{1/2}$.

Do $\text{SVD}(\mathbf{C}_1^{-1/2}\mathbf{C}_{12}\mathbf{C}_2^{-1/2}) = \mathbf{U}\mathbf{A}\mathbf{V}^T$

The solutions are $\mathbf{B}_1^T = \mathbf{U}^T \mathbf{C}_1^{-1/2}$ and $\mathbf{B}_2^T = \mathbf{V}^T \mathbf{C}_2^{-1/2}$ and their inverse are $\mathbf{A}_1 = \mathbf{C}_1^{1/2}\mathbf{U}$ and $\mathbf{A}_2 = \mathbf{C}_2^{1/2}\mathbf{V}$, respectively.

It holds

$$\begin{cases} \mathbf{A}_1 \mathbf{A}_1^T = \mathbf{C}_1 \\ \mathbf{A}_2 \mathbf{A}_2^T = \mathbf{C}_2 \\ \mathbf{A}_1 \mathbf{A}_2^T = \mathbf{C}_{12}; \mathbf{A}_2 \mathbf{A}_1^T = \mathbf{C}_{21} \end{cases}, \quad (211)$$

that is, CCA is a special kind of joint full-rank factorization of \mathbf{C}_1 and \mathbf{C}_2 respecting the third equalities here above. Since it maximizes the correlation, differently from the MCA, the CCA is not sensitive to the amplitude of the two processes. This should be kept in mind, as the correlation of two signals may be high even if the amplitude of one of the two signal is very low, say, at the noise level; in such a case the resulting correlation is spurious and meaningless¹⁹. In general, methods such as

¹⁹ Similar considerations apply to the frequency-domain version of the (squared) correlation coefficient, the well-known coherence, which is a meaningless measure if one of the signals or both are of very low amplitude (Bloomfield, 2000). This is rarely checked in papers dealing with massive coherence (and similar synchronization measures) estimations.

MCA and CCA are meaningful when the amplitude of the two processes is comparable. The same applies to their extension, the joint blind source separation, which we will treat in chapter VII.

CHAPTER VI

BLIND SOURCE SEPARATION

Introduction

So far we have considered *arbitrary transformations of the data* possessing some specific properties. However, the resulting components (time course) of the transformed signal $\mathbf{y}(t) = \mathbf{B}^T \mathbf{x}(t)$ and their corresponding spatial patterns given in the columns of the inverse of \mathbf{B}^T *do not have any physiological meaning*²⁰. Here and hereafter we treat the much more involved task of finding transformations of EEG signals possessing some properties, as before, but also estimating the *waveform* and the *spatial pattern* of *physiological sources* generating the signal. BSS has enjoyed considerable interest worldwide only starting a decade after the pioneering works carried out in our laboratory in Grenoble (Ans et al., 1985; Héroult and Jutten, 1986), inspired by the seminal papers of Jutten and Héroult (1991), Comon (1994) and Bell and Sejnowski (1995). Thanks to its flexibility and power BSS has today greatly expanded encompassing a wide range of applications such as speech enhancement, image processing, geophysical data analysis, wireless communication and biological signal analysis (Comon and Jutten, 2010). In EEG BSS has enjoyed an amazing popularity starting at the end of the last millennium, being used, just to name a few examples, for improving brain computer interfaces (Kachenoura et al., 2008; Serby et al., 2005; Wang and James, 2007;), for increasing the SNR of single-trial time-locked responses (Cao et al., 2002; Guimaraes et al., 2007; Lemm et al., 2006; Sander et al., 2005; Tang et al., 2006; Zeman et al., 2007) and for denoising/artifact rejection (Crespo-Garcia et al., 2008; Fitzgibbon et al., 2007; Frank and Frishkoff, 2007; Halder et al., 2007; Ille, Berg and Scherg, 2002; Iriarte et al., 2003; Joyce et al., 2004; Jung et al., 2000; Kierkels et al., 2006; Phlypo et al., 2007; Romero et al., 2008; Vigário, 1997; Vorobyov and Cichocki, 2002).

This chapter begins with the consideration of early single-subject BSS attempts based on the simultaneous diagonalization of two matrices (*GEVD*), that is, referring to the arguments introduced in chapter IV, the cases $M=1$ and $K=2$. Then we illustrate a more flexible and accurate BSS scheme requiring the estimation of second-order statistics (SOS) only and based on the approximate joint diagonalization of several matrices ($K>2$). A general algorithm is described. The method operates in the same way in the time or frequency domain (or both at the same time) and is capable of modeling explicitly physiological and experimental source of variations with remarkable flexibility. At this stage we provide several examples illustrating the analysis of continuous recording EEG (spontaneous

²⁰ In some circumstances the CSP and the CCA may output physiologically meaningful time course of components and associated spatial patterns. For CSP for example, this is when the data is chosen respecting the non-stationarity BSS framework presented in this section. We will comment further on this point.

activity), event-related de/synchronizations (induced activity) and event-related potentials (evoked activity).

The BSS Problem for EEG

For N scalp sensors and $P \leq N$ EEG dipolar fields with fixed location and orientation in the analyzed time interval, the linear BSS model simply states the superposition principle discussed in chapter II, i.e.,

$$\mathbf{x}(t) - \kappa(t)\mathbf{I} = \mathbf{A}\mathbf{s}(t) + \boldsymbol{\eta}(t), \quad (212)$$

where $\mathbf{x}(t) \in \mathfrak{R}^N$ is the *sensor measurement vector* with common reference κ , $\mathbf{A} \in \mathfrak{R}^{N \times P}$ is assumed a time-invariant full column rank *mixing matrix*, $\mathbf{s}(t) \in \mathfrak{R}^P$ holds the unknown time-course of the source components and $\boldsymbol{\eta}(t) \in \mathfrak{R}^N$ is unknown additive noise, assumed temporally white, uncorrelated to $\mathbf{s}(t)$ and with spatially uncorrelated components. We treat here the problem $P \leq N$, that is, we try to estimate at most as many sources as available sensors. Equation (212) states in the language of linear algebra what has been said in words in chapter II, i.e., that the EEG sensor measurement $\mathbf{x}(t)$ is a linear combination (mixing) of sources $\mathbf{s}(t)$, given by the coefficients in the corresponding column of matrix \mathbf{A} . In contrast to inverse solutions (chapter III) neither $\mathbf{s}(t)$ nor \mathbf{A} are supposed to be known, that is why the problem is said to be *blind*. Although this is the classical BSS model we need a few clarifications for the EEG case: by $\boldsymbol{\eta}(t)$ we model *instrumental* noise only. In the following we drop the noise term because the instrumental (and quantization) noise of modern EEG equipment is typically low ($< 1\mu\text{V}$). *Biological* noise (extra-cerebral artifacts such as eye movements and facial muscle contractions) and *environmental* noise (external electromagnetic interference) may obey a mixing process as well, thus they are generally modeled as components of $\mathbf{s}(t)$, along with cerebral ones. Notice that while biological and environmental noise can be identified as separated components of $\mathbf{s}(t)$, hence removed, source estimation will be affected by the underlying cerebral *background noise* propagating with the same coefficients as the signal (Belouchrani and Amin, 1998). In contrast to the approach for inverse solutions seen in chapter III, where data is referenced to the common average in order to eliminate the reference (see (105) to (107)), we here leaves the estimation of sources up to the reference in order to preserve the full-rank of the data. We then obtain the simplified model

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t). \quad (213)$$

Our source estimation is given by inverting this equation such as

$$\hat{\mathbf{s}}(t) = \mathbf{B}^T \mathbf{x}(t), \quad (214)$$

where $\mathbf{B}^T \in \mathfrak{R}^{P \times N}$ is called the *demixing* or *separating matrix*. This is what we want to estimate in order to recover the sources from EEG²¹.

A Suitable Class of BSS Solutions

Tackling problem (214) assuming knowledge of sensor measurement only appears a little pretentious. The great achievement of BSS theory is to demonstrate that it is possible with some assumptions on the statistical properties of the sources, effectively reducing the number of admissible solutions. Particularly, we are interested in weak restrictions converging toward condition

$$\hat{\mathbf{s}}(t) = \mathbf{\Phi}\mathbf{s}(t) \quad (215)$$

where $\mathbf{s}(t)$ holds the time-course of the true (unknown) source processes, $\hat{\mathbf{s}}(t)$ our estimation and the *system matrix*

$$\mathbf{\Phi} = \mathbf{B}^T \mathbf{A} \approx \mathbf{D}\mathbf{P} \quad (216)$$

approximates a signed scaling (a diagonal matrix \mathbf{D}) and row permutation (\mathbf{P}). Equation (215) is obtained substituting (213) in (214). Whether condition (215) may be satisfied is a problem of *identifiability*, which establishes the theoretical ground of BSS theory (Tong, Inouye and Liu, 1993; Cardoso, 1998; Pham and Cardoso, 2001; Pham, 2002). The identification capability makes of BSS a much more useful method as compared to the linear transformations we have considered in chapter V²². Matching condition (215) implies that we can recover faithfully the source *waveform*, but only out of a *scale* (including sign) and *permutation* (order) indeterminacy, that is, the admissible solutions

²¹ We do not use anymore the symbol $\mathbf{y}(t)$ for transformed data to stress that BSS estimates the waveform of the actual sources.

²² We will talk later on about how *identifiability* is sought in practice with the proposed BSS approaches.

cannot be constrained any further and we are left with a permutation matrix and a diagonalization matrix that are arbitrary. This limitation is not constraining for EEG, since it is indeed the waveform that bears meaningful physiological and clinical information. Notice the correspondence between the p^{th} source, its *separating vector* (p^{th} row of \mathbf{B}^T) and its *scalp spatial pattern* (mixing vector), given by the p^{th} column of

$$\mathbf{A} = (\mathbf{B}^T)^+, \quad (217)$$

where superscript $+$ indicates the Moore-Penrose pseudo-inverse (79). The mono-dimensionality of those vectors and their sign/energy indeterminacy implies the explicit modeling of the orientation and localization parameters of the p^{th} source, but not its energy. This is also the case of scalar-type inverse solutions we have encountered in chapter III, but not of the vector-type inverse solutions, which estimates current in three directions (111), with consequent lower spatial resolution. Notice that we are not placing a “hat” on \mathbf{A} and on \mathbf{B}^T albeit they are both estimated from the data in order to keep notation clean.

BSS Filtering

Linearity allows switching back from the source space into the sensor space. Substituting (214) into (213) yields *BSS filtering*

$$\hat{\mathbf{x}}(t) = \mathbf{AZ}\hat{\mathbf{s}}(t) = \mathbf{AZB}^T\mathbf{x}(t), \quad (218)$$

where we have interposed a diagonal matrix \mathbf{Z} with p^{th} diagonal element equal to 1 if the p^{th} component is to be retained and equal to 0 if it is to be removed. BSS filtering is common practice to remove artifacts from the EEG data (e.g., Vigário, 1997; Jung et al., 2000).

Localization of BSS Components

The vectors \mathbf{a}_p of the mixing matrix \mathbf{A} are the spatial patterns associated to the p^{th} source. Taken one by one these vectors may be used to create topographic maps or may be fed to inverse solutions, as if they were single EEG samples. Note that for inverse solutions treated in chapter III input data must be in

common average reference (107), therefore we actually need to feed the inverse solution of chapter III with $\mathbf{H}\mathbf{a}_p$, where \mathbf{H} is the centering matrix (89).

Different Approaches for Solving BSS

It has been known for a long time that in general the BSS problem cannot be solved for sources that are Gaussian, independent and identically distributed (*iid*) (Darmois, 1953). The *iid* condition implies that each sample of the source components is statistically independent from the others and that they all follow the same probability distribution. Therefore, in order to solve the BSS problem the sources must be either (a) possibly *iid*, but non-Gaussian or (b) possibly Gaussian but *not iid*.

In case (a), one assumes that at most one source is Gaussian and that they are all mutually statistically independent. The mutual independence assumption (spatial independence of all pair-wise sources) should not be confused with the *iid* condition (temporal independence of successive samples within each source process). Actually, the *iid* condition implies that no temporal information is used, thus the method is efficient regardless the temporal dependence of sources and outputs the same mixing and demixing matrix estimation on the data as on *any shuffled version of the data*. Those methods are known as *independent component analysis* (ICA) (Cardoso, 1989; Comon, 1994; Hyvärinen, 1999; Jutten and Héroult, 1991). ICA requires higher order statistics (HOS), explaining why it may succeed only if at most one source has Gaussian distribution: in fact Gaussian distributions are fully defined by their statistics up to the second order (SOS). The idea of (b) is to break the non-Gaussianity assumption. This can be done using SOS statistics only by assuming that source components are all pair-wise uncorrelated and that they are not *iid*, that is, that they possess a *temporal structure*.

If these assumptions are fulfilled the separating matrix can be identified uniquely, thus source can be recovered *regardless the true mixing process* (uniform performance property: see for example Cardoso, 1998) and regardless the distribution of sources, which is a remarkable theoretical advantage. The fundamental question is therefore whether or not the above statistical assumptions fit the EEG data. In Congedo, Jutten and Gouy-Pailler (2008) we have reported statistical and neurophysiological considerations about the choice of SOS vs. HOS statistics considering the most common kinds of EEG data, namely *spontaneous*, *induced* and *evoked* EEG. We have contended that several EEG phenomena

are strongly colored²³ and/or, since they are episodic or appears in more or less short bursts, their energy over time is never constant²⁴ (Lopes da Silva, 2005b; Niedermeyer, 2005 a,b,c; Steriade, 2005; Buzsáki, 2006, Ch. 6, 7). These phenomena include, among others,

- several kinds of EEG artefacts such as eye-blinks and facial muscular electromyography,
- several episodic spontaneous EEG phenomena such as sleep spindles (7-14 Hz) (Niedermeyer, 2005 b; Steriade, 2005), frontal Theta (4-7 Hz) and Beta (13-35 Hz) waves (Niedermeyer, 2005 a),
- several sustained spontaneous EEG phenomena such as slow Delta (1-2 Hz) waves during deep sleep stages III and IV (Niedermeyer, 2005 b), the Rolandic Mu rhythms (around 10 Hz and 20 Hz) and the posterior dominant rhythms (in the Alpha range: 8-12 Hz) (Niedermeyer, 2005 a),
- induced EEG such as ERD/ERS (Pfurtscheller and Lopes da Silva, 2004; Steriade, 2005)
- evoked activity (ERP: Lopes Da Silva, 2005 b).

Therefore we have concluded that SOS statistics are appropriate for capturing the relevant information contained in most observable EEG phenomena²⁵. Based on these conclusions in the reminder of this and next chapter we will focus exclusively on SOS time-frequency approaches, which are well established in other technical fields (Belouchrani and Amin, 1998; Pham 2002; Choi et al., 2002; Bousbia-Salah et al., 2003). In order to work with SOS statistics we proceed assuming that source components are *all pair-wise uncorrelated* and that either

SOS BSS Assumption 1 (Coloration)

Within each source component the successive samples are temporally correlated, (Féty and Uffelen, 1988; Tong et al., 1990, 1991 b; Molgedey and Schuster, 1994; Belouchrani et al., 1997)

and/or

²³ “Colored” signals, as opposed to “white” signals, display a non-flat power spectrum.

²⁴ Signals which energy changes over time are said “non-stationary”

²⁵ On the other hand HOS are more adapted to capture spikes, sharp waves and spike-wave complexes in epileptic disorder (Niedermeyer, 2005 c), vertex waves during sleep (Niedermeyer, 2005 b), and similar transient activity with abrupt potential changes.

SOS Assumption 2 (Non-stationarity)

Samples in successive time intervals do not have the same energy (Matsuoka et al., 1995; Souloumiac, 1995; Choi and Cichocki, 2000; Pham and Cardoso, 2001; Choi et al., 2002).

Identifiability (215) according to these assumption is achieved when *all sources taken pair-wise have non-proportional power spectrum* under assumption 1 (coloration) or when the *changes in energy over data windows for all sources taken pair-wise is different* under assumption 2 (non-stationarity).

BSS Based on the Joint Diagonalization of Two Matrices.

If sources are uncorrelated and assumption 1 or assumption 2 is verified the model is identifiable in closed form. Algebraic solutions to the BSS problem share the same conceptual scheme (Parra and Saida, 2003): they are all achieved by a generalized eigenvalue-eigenvector decomposition (*GEVD*) of two square matrices (82), of which one is always the covariance matrix of the sensor measurement (95) and the other is a covariance matrix embedding information about the chosen assumption. The covariance matrix of the sensor measurement is common to all methods, while the other makes every method different.

Closed Form BSS Solutions for Colored Processes

Molgedey and Shuster (1994) specified the model given by SOS assumption 1. Their algorithm consists in jointly diagonalizing the covariance matrix and a lagged covariance matrix with lag τ . Alternatively, the same result is achieved by a two-step process analogous to the one used in (83): first we find a whitening matrix \mathbf{W} (197) such that, for the whitened data $\mathbf{Y}=\mathbf{W}^T\mathbf{X}$ it holds true $\mathbf{C}_y=\mathbf{I}$. Then, by singular value decomposition (SVD) we diagonalize the lagged covariance matrix of the whitened data $\mathbf{C}_{y(t-\tau)}$ (Chicochi and Amari, 2002, p. 146), or by *EVD* its symmetric part (24) (Thong et al., 1991b), or, more, the covariance matrix $\mathbf{C}_{\tilde{y}(t)}$, where $\tilde{y}(t)=y(t)+y(t-\tau)$ (Chicochi and Amari, 2002, p. 120), all leadings to results practically identical under assumption 1. This family of two-step algorithms is known as AMUSE (algorithm for multiple source extraction).

Closed Form BSS Solutions for Non-Stationary Processes

To comply with SOS assumption 2, the data is split in two non-overlapping windows \mathbf{X}_1 and \mathbf{X}_2 . Then we compute their covariance matrices, denoted \mathbf{C}_1 and \mathbf{C}_2 and the BSS problem is solved as above by the *GEVD* of \mathbf{C}_1 and \mathbf{C}_2 . Importantly, if the source is active in one interval and completely inactive in the other, the obtained filter is optimal (Souloumiac, 1995). Along these lines see the discussion on *super-efficiency* in Pham and Cardoso (2001). In the same fashion, the time intervals \mathbf{X}_1 and \mathbf{X}_2 may pertain to two different experimental conditions or two different classes in BCI. In this latter case this procedure is equivalent to the CSP. It should be noted however that in general the CSP is not a BSS (waveform-preserving) decomposition. In order to extract physiological sources the appropriate choice of \mathbf{X}_1 and \mathbf{X}_2 is crucial; EEG data within each window must be stationary, which for EEG implies the choice of short windows. For the CSP usually a large amount of data is included in \mathbf{X}_1 and \mathbf{X}_2 in order to obtain good generalization to unseen data. This ensures that the ratio of the variance between the classes is maximized in general, but does not necessarily succeed in recovering the sources.

BSS by Approximate Joint Diagonalization of a Matrix Set

More recent research has generalized and extended the *GEVD* methods by *Approximate Joint Diagonalization (AJD)* of several matrices (Cardoso and Souloumiac, 1993; Pham, 2001b; Pham and Congedo, 2009; Ziehe et al., 2004; Vollgraf and Obermayer, 2006; Dégerine and Kane, 2007; Theis and Inouye (2006); Tichavsky and Yeredor, 2009), which we have considered in chapter IV. In this chapter we treat the reduced problem (167). Given a set of covariance matrices $\mathbf{C}_k \in \{\mathbf{C}_1, \dots, \mathbf{C}_K\}$, the *AJD* seeks a matrix \mathbf{B}^T such that the products $\mathbf{B}^T \mathbf{C}_k \mathbf{B}$ are as diagonal as possible $\forall k \in \{1, \dots, K\}$. Given an appropriate choice of the *diagonalization set* \mathbf{C}_k , such matrix \mathbf{B} is indeed an estimation of the demixing matrix in (214) and one obtain an estimate of the mixing matrix as per (217). The joint diagonalization is applied on matrices that *change* according to the assumptions about the source, exactly as in the case of the diagonalization of two matrices. However adopting *AJD* the changes are more likely to be detected. These extensions have better performance because they make use of more available sample statistics of the data. Hence, the estimation of the mixing process is more robust. Also, the *AJD* extensions are more robust to noise as none of the matrices in the diagonalization set is diagonalized exactly, so different noise structures in different matrices do not distort as much the

solution. Nonetheless, the additional assumption remains essentially the same and each method in turn fails whenever its peculiar additional assumption is not matched.

The SOBI AJD Methods

The SOBI method of Belouchrani et co-workers (1997) is an extension of the AMUSE algorithm obtained by *AJD* of several autocorrelation matrices with several τ estimated on whitened data. Very similar, the method of Ziehe and Müller (1998) seeks a non-orthogonal *AJD* of several autocorrelation matrices with several τ estimated on raw data.

The Fundamental Theorem of AJD-based BSS

In order to make purposeful use of SOS-based AJD methods let us consider precisely the necessary and sufficient conditions for source identifiability (215). They are described by the fundamental *AJD*-based BSS theorem (Afsari, 2008; see also Aïssa-El-Bey et *al.*, 2008): let matrix set $\{S_1, \dots, S_k\}$ hold the K (unknown) covariance matrices of sources corresponding to the covariance matrices included in the diagonalization set $\{C_1, \dots, C_k\}$. Denote by $s_{rc,k}$ the (r,c) element of S_k (entry at the r^{th} row and c^{th} column). The diagonal elements of these matrices $s_{rc,k}, r=c$, hold the source variance. The off-diagonal elements $s_{rc,k}, r \neq c$, are null as sources are assumed to be uncorrelated. Let

$$\Psi = (\psi_1, \dots, \psi_P)^T = \begin{pmatrix} s_{11,1} & \dots & s_{11,K} \\ \vdots & \ddots & \vdots \\ s_{PP,1} & \dots & s_{PP,K} \end{pmatrix} \quad (219)$$

be the matrix formed by stacking one below the other the P row vectors $\psi_1^T, \dots, \psi_P^T$ constructed as shown in fig. 6.1. Each row vector of Ψ is

$$\psi_p^T = (s_{pp,1}, \dots, s_{pp,K}), \quad (220)$$

thus it holds the *profile along the diagonalization set* for each source, with $p \in \{1, \dots, P\}$ and P the number of estimated sources. For instance, if the matrix in the diagonalization set are lagged covariances the profile is the autocorrelation of the sources. If the matrices are Fourier cospectra the

profile is the source power spectrum, etc. *The fundamental AJD-based BSS theorem says that the p^{th} source can be separated as long as its profile vector ψ_p is not collinear with any other vector in Ψ* ²⁶. Said differently, the wider the angle between ψ_p and any other vector in Ψ , the greater the chance to separate the p^{th} source. Importantly, Even if two vectors are collinear, the other sources can still be identified. The theorem says that sources will be identified if their profile is non-proportional to the profile of any other source. Therefore, to succeed with AJD-based method we have to create diagonalization set with this characteristic. In order to do so we first specify a general framework combining the two SOS assumptions.

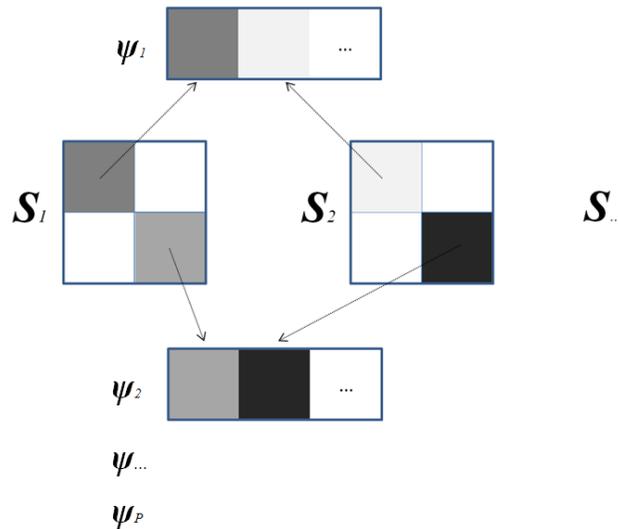


Figure 6.1: Graphical illustration of the construction of the source energy profile vectors ψ_p .

The AJD of Fourier Cospectra (AJDC) Algorithm

The model imposed by the statistical assumptions often is too restrictive for practical purposes. For instance, whether several sources may well have different power spectra, often they do not have *all* different power spectra. This is the case for example of the several occipital dominant rhythms that can be found, which have very close spectral profile (Nunez, Wingeier and Silberstein, 2001). Similarly, non-stationarity may be present or not and sometimes it is difficult, if not arbitrary, to select

²⁶ Two vectors are collinear if they are equal out of a scaling factor, that is, if the profile is proportional.

time windows for which the energy is different. To overcome these difficulties we may want to combine the two basic theoretical frameworks for working in a SOS framework, the *coloration* and the *non-stationary* in such a way that *identifiability is reached whenever either assumption is respected* for a given source to be extracted. Actually, the *AJD* can be easily and conveniently transposed in the frequency or time-frequency domain, whether we perform the frequency expansion for several time segments. In fact, applying to (213) any invertible and linearity-preserving transform \mathcal{T} leads to (Congedo, Gouy-Pailler and Jutten, 2008)

$$\mathcal{T}[\mathbf{x}(t)] = \mathbf{A}\mathcal{T}[\mathbf{s}(t)], \quad (221)$$

which preserves the mixing model. The basic idea is to estimate several Fourier cospectral matrices (85) on several time windows, classes or experimental conditions in order to capture both the source spectra and non-stationarity profile. We will use Fourier cospectral matrices for a range $f \in \{1, \dots, F\}$ of discrete frequencies and for a range $i \in \{1, \dots, I\}$ of temporal windows, classes and/or experimental conditions. The epochs on which the cospectra are estimated should be short enough to capture the energy variations over time and wide enough to allow satisfactory estimations for each of them separately.

Notice that working in the frequency domain is advantageous for several reasons: first, covariance statistical estimations in the time domain are distorted for temporally correlated processes like EEG (Beran, 1994). Second, estimating cospectral matrices in the frequency domain is computationally more efficient than estimating delayed covariance matrices in the time domain²⁷. Finally, the *AJD* of cospectra has been connected to the *Gaussian mutual information* criterion (Pham 2001a, 2002). This places the ensuing method at the heart of the BSS theory and steers toward the Cramér-Rao bound (Pham, 2001a; Pham and Cardoso, 2001) when the sources are Gaussian, a *working assumption* that we will make also in chapter VIII in the framework of Riemann geometry. An example of applying the *AJD* of several cospectral matrices to the continuous stream of EEG is shown in fig. 6.2.

²⁷ *Fourier cospectra estimations may take advantage of efficient split-radix fast Fourier transform (FFT: Cooley and Tukey, 1965) algorithms such as FFTW3 (Frigo and Johnson, 2005); in typical situations we may expect the computation complexity of Fourier cospectral matrices be 20 to 100 times smaller as compared to lagged covariance matrices.*

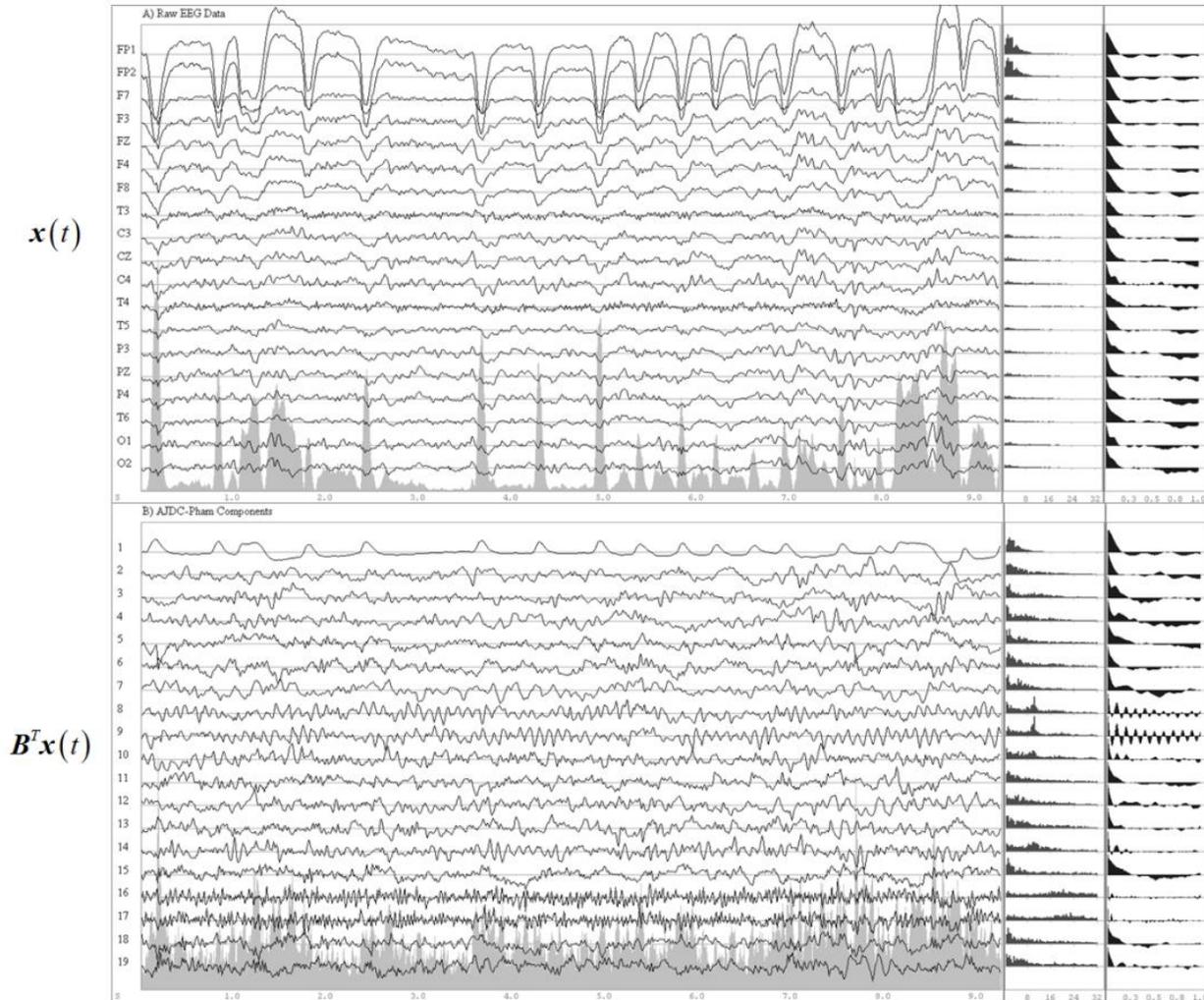


Figure 6.2: On the top about 10s of EEG continuous recording obtained on a 26 y.o. male healthy individual with 19 electrodes placed according to the international 10/20 system (Jasper, 1958). On the bottom the sources obtained by AJDC. Matrix B is found as the approximate joint diagonalizer of all 1-Hz spaced cospectral matrices in the frequency range 1-28 Hz. With this setting the AJDC is exploiting only the coloration assumption. Cospectra are estimated with a 50% overlapping sliding window according to the classical method attributed to Welch (1967). Next to each traces there is the corresponding power spectrum in the range 1-32 Hz and the autocorrelation function with lags 0 to 1s. One may notice the first source, capturing the eye blinks (notice that the amplitude and the sign of the BSS solution is arbitrary, see (215)), sources 8 and 9 holding alpha rhythms and sources 16 and 17 holding EMG activity visible in the EEG recording at electrodes T3 and T4.

The implementation of AJDC

Let us see how to implement the AJDC algorithm (Approximate Joint Diagonalization of Cospectra) in practice. One can obtain a solution by simply finding the *AJD* matrix of the cospectral matrix set $\mathbf{C}_k \in \{\mathbf{C}_1, \dots, \mathbf{C}_K\}$, as

$$\mathbf{B}^T = \text{AJD}(\mathbf{C}_k) \quad (222)$$

Such a one-step procedure performs well when the number of electrodes is not large (say, < 20). Otherwise the covariance matrix being ill-conditioned, that is, with several eigenvalues close to zero, the *AJD* algorithm may not converge appropriately. For this reason we always adopt a two-step procedure, as in general it is done in BSS methods:

Algorithm (223): AJDC (AJD of Cospectral Matrices). Optimization (167).

Given diagonalization set $\mathbf{C}_k \in \{\mathbf{C}_1, \dots, \mathbf{C}_K\}$ indexed by $k \in \{1, \dots, K\}$, let

$$\mathbf{C}_{TOT} = \sum_k \mathbf{C}_k \quad (224)$$

And

$$\mathbf{W}^T = \mathbf{C}_{TOT}^{-\frac{1}{2}}$$

its whitening matrix (197). Now partition the whitening matrix such as

$$\mathbf{W}^T = [\mathbf{F} \ \mathbf{N}]^T, \quad (225)$$

where $\mathbf{F}^T \in \mathfrak{R}^{P \times N}$ holds the first P rows of \mathbf{W}^T (signal subspace) and \mathbf{N}^T the remaining rows (noise subspace). The AJD problem on the reduced and whitened data is now

$$\mathbf{E}^T = \text{AJD}(\mathbf{F}^T \mathbf{C}_k \mathbf{F}) \quad (226)$$

$$\text{The solution (demixing matrix) is } \mathbf{B}^T = \mathbf{F}^T \mathbf{E}^T \in \mathfrak{R}^{P \times N} \quad (227)$$

$$\text{and its inverse (mixing matrix) is } \mathbf{A} = (\mathbf{E}^T)^{-1} (\mathbf{F}^T)^+ \in \mathfrak{R}^{N \times P} \quad (228)$$

Things to know working with AJDC

The choice of the AJD algorithm

Note that C_{TOT} (224) is the sum of cospectral matrices at several frequencies, time intervals, classes, experimental conditions, etc., thus W^T it is a “global” whitening matrix. One thing should be noticed at this time; it is known that pre-whitening and then constraining the ensuing AJD solution E^T (226) to be orthogonal jeopardizes the separation performance due to the estimation error of the data covariance matrix and noise (Cardoso, 1994; Yeredor, 2000; Pham, 2001a). Exact diagonalization of the covariance matrix is required for the orthogonal constraint on E^T to be valid; however it implies the diagonalization of the estimation errors as well, which distorts the solution. The problem is solved simply by *not constraining the ensuing AJD solution (226) to be orthogonal*. Whereas orthogonal AJD solutions are favorable in term of convergence and stability, non-orthogonal AJD algorithms performs well when the covariance matrices are well conditioned and when the solution is close to the orthogonal form, which is the case after whitening. This is the reason why *this two-step procedure is always preferable*.

Size and Content of the diagonalization set

The key for succeeding with BSS by AJD is the definition of an adequate size and content of the diagonalization set; it should include matrices estimated on data as homogeneous as possible for each matrix, with enough samples to allow a proper estimation, in frequency regions and time blocks when the signal-to-noise ratio is high and with an high probability to uncover unique source profiles. Table 6.1 reports useful information to define an appropriate diagonalization set so as to ensure identifiability of sources.

It is also important to consider that the number of matrices in the diagonalization set should be high enough to help non-collinearity of source profiles (219). One may want to have at least as many matrices in the diagonalization set as sources to be estimated, but this is not strictly necessary. On the other hand one should not try to increase the number of matrices indefinitely to the detriment of the goodness of their estimation, i.e., selecting too many discrete frequencies or blocks of data that are too shorts.

Table 6.1: Criteria to achieve identifiability of sources in BSS methods based on AJD of SOS.

Assumption on the sources	Covariance Matrices (CM) Estimation	What is the source profile	Sufficient condition for Identifiability	Examples of data
Coloration	<ul style="list-style-type: none"> i. Lagged Covariance matrices, ii. Fourier Cospectral Matrices, iii. CM estimated with a filter bank 	<ul style="list-style-type: none"> i. The source <i>autocorrelation</i>, ii. The source <i>power spectrum</i> iii. as in ii. 	The power spectrum of the source is non-proportional to the power spectrum of any other sources	Spontaneous oscillation with characteristic power spectrum such as posterior dominant rhythms (Alpha), Somatosensory Mu rhythms, frontal midline Theta, Beta bursts, etc.
Non-Stationarity	CM estimated on <ul style="list-style-type: none"> j. Different <i>blocks</i> of data jj. Different <i>experimental conditions</i> jjj. Different <i>classes</i> 	The variation of the source energy along the <ul style="list-style-type: none"> j. <i>Blocks</i> or jj. <i>Experimental conditions</i> or jjj. <i>Classes</i> 	The variation of the source energy along the <ul style="list-style-type: none"> j. <i>Blocks</i> or jj. <i>Experimental conditions</i> or jjj. <i>classes</i> do not correlate with the same variation of any other sources 	<ul style="list-style-type: none"> - Blocks of data according to physiological reactivity of EEG oscillations (e.g., eyes-close vs. eyes-open) - CM estimated before and after the event in ERD/ERS - CM estimated on different peaks in ERP (after averaging the ERP) - Active vs. Control condition, ...

Weighting matrices in the diagonalization set

Particularly unproductive is the inclusion in the set of matrices with small signal-to-noise ratio. Such matrices are nearly diagonal, as in general the noise is little spatially correlated. For continuous recording EEG, matrices at high frequencies are nearly diagonal. Above 20 Hz the EMG becomes predominant (Whitham et al. 2007) and EMG is little spatially correlated. However the higher informative frequency depends very much on the data and instrumentation at hand. To make sure that we diagonalize matrices holding relevant information we normalize all cospectra in the diagonalization set to unit trace and we weight them by a non-diagonality function (Congedo, Jutten and Gouy-Pailler, 2008) such as

$$\varpi(\mathbf{C}_k) = \frac{1}{N-1} \frac{\sum_{r \neq c} c_{rc,k}^2}{\sum_{r=c} c_{rc,k}^2} \quad (229)$$

where $c_{rc,k}$ is the entry of matrix \mathbf{C}_k at row r and column c and N is the size of the matrix (number of channels). For a positive definite matrix, non-diagonality measure (229) is bounded inferiorly by zero,

for a diagonal matrix, and superiorly by 1.0, for a uniform matrix. Thus the higher the non-diagonality function the higher the weight. Noise-suppression may be promoted by zeroing the weights above a cut-off frequency. According to our experience, such a weighting function generally allows satisfactory source estimation with EEG. We have observed that the non-diagonality function (229) is highly correlated with the overall energy (trace of the cospectral matrices), but is not as much influenced by the dominant occipital rhythms (8-12 Hz). Using a non-diagonality weighting function is in line with previous works in time-frequency BSS where the diagonalization effort is concentrated on high-energy time-frequency regions (Belouchrani and Amin, 1998).

Dimensionality reduction

Finally, notice that dimensionality reduction by truncated pre-whitening (226) loses some of the variance of the data. If $\mathbf{W}^T = \mathbf{C}_{TOT}^{-1/2} = \mathbf{U}\mathbf{A}^{-1/2}\mathbf{U}^T$ is the complete whitening matrix and $\lambda_{TOT} = \sum_n \lambda_n$ the total variance of the data, the truncated whitening matrix \mathbf{F}^T (226) retains exactly $\sum_p \lambda_p / \lambda_{TOT}$ of the variance and the *representation error* is

$$g_{(p)} = \sum_{q=p+1}^N \lambda_q / \lambda_{TOT} . \quad (230)$$

This is the amount of variance (with respect to the total) that will be lost by truncated pre-whitening. In general, losing up to 1% of the variance is a safe strategy, as only noise is removed, whereas the reduced dimension ensures that the reduced covariance matrix is well-conditioned.

Discussion on SOS-based BSS Methods

The possibility of recovering both the waveform and the spatial pattern of EEG unknown source in a blind fashion is with no doubt a seducing ability of BSS. Nonetheless, besides the correct modeling of the fundamental assumptions and the correct algorithmic implementation, which is never completely automatic (e.g., the definition of proper weighting for continuously recorded EEG), the linear BSS instantaneous model (212) makes a number of restrictive general assumptions that are rarely checked or investigated. We here discuss them:

- One general assumption is that the number of sources is not greater than the number of sensors. When this is not the case (undetermined case) it is not possible to solve the BSS problem unless other constrains on the sources are introduced (e.g., sparsity: Gribonval and Lesage, 2006). In

practice, with truncated pre-whitening chosen so as to yield a small representation error (230) we limit the estimation to P sources, to which the remaining sources, all at noise level, will be mixed.

- A related assumption for the exactly determined model is that the mixing matrix A in model (213) is full-column rank. With truncated pre-whitening we assume that it has rank P . The columns of A are scalp spatial pattern vectors of the source components and the more the electrodes are close to each other, the more those vectors will be collinear. Consequently, it is always better to space the electrodes as much as possible on the scalp²⁸.

- One may also wonder if during the analyzed time interval the number of active dipoles is stable (Li et al., 2006). In practice, brain electrical “source components” are macroscopic electric dipole with relatively high SNR formed by the synchronous activity of pyramidal cells over large cortical areas (Nunez and Silberstein, 2000; Nunez and Srinivasan, 2006). For sufficiently small time intervals one may assume that such high- SNR layers are limited in number. Other concurrently active cortical columns may be ignored if their current is comparatively negligible and it does not matter if the dipoles are active throughout the time interval or intermittently (actually such non stationarity signature can be explicitly exploited). Henceforth, assuming at least as many sensors as relevant sources does not appear problematic if we consider *a sufficiently small time interval*.

- No definitive solution exists to the problem of estimating the number of source components in the overdetermined case (more sensor than source components). Whereas correct dimensionality reduction by pre-whitening (226) allows exact determination, the amount of dimensionality reduction is somehow arbitrary and over-reduction must be avoided since in this case identifiability is lost and several generators are extracted mixed in one component. A safe strategy is to identify a few meaningful components and keep reducing the dimension until those components are not distorted (step-down).

- Several restrictive assumptions are made by model (213) also on the nature of brain electric fields. One may ask whether it is reasonable to assume that dipoles keep fixed orientation and location in the analyzed time interval. For a fixed spatial sensor configuration with respect to the brain, which

²⁸ This suggests that placing many electrodes closely spaced above the brain region of interest, as it is sometimes done, is not a convenient strategy if multivariate statistical methods are to be employed. This is true for all source analysis methods considered in this manuscript.

is the case of a single EEG recording session, the orientation and location of electric dipoles are fixed by the anatomy and physiology of the cortical convolution forming the dipole. However, the dipole approximation becomes untenable for sources distributed over large areas (Malmivuo and Plonsey, 1995; Nunez and Srinivasan 2006). Also, there is evidence of traveling waves phenomena in the brain; long wavelength waves originating in a region and propagating via cortico-cortical connections to other regions (Lopes da Silva and Van Rotterdam, 2005; Srinivasan et al., 2006; Thorpe et al., 2007).

- Also, the longer the time interval under analysis the less tenable is the stationarity assumption, which is basic to SOS estimations (Hyvärinen et al., 2001, p. 49). At the same time one must take care to retain enough data points for analysis in order to avoid *overfitting* (Müller et al., 2004). Särelä and Vigário (2003) reported that using small time intervals the output may contain artefacts that are not present in the data. For HOS methods artifacts takes the form of artificial spikes and bumps, whereas for SOS methods they take the form of artificial sinusoid waves. This should be kept in mind while checking and validating the BSS output. Meinecke et al. (2002) and Müller et al. (2004) addressed the problem of obtaining robust and reliable source estimates. They proposed a resampling-based methods consisting in running the algorithms on different time intervals and retain only the source processes that can be found consistently.

In conclusion, although statistical estimations improve with the number of samples we advocate the use of multiple time intervals as short as possible (enough to avoid overfitting while justifying the BSS method assumptions, say, 8 to 40 seconds, depending on the data), modeling appropriately the stationarity within intervals while exploiting explicitly the non stationarity between intervals. In this sense an efficient time-frequency approach appears a precious option.

EXAMPLE STUDIES USING AJDC

We have presented a general framework to address a multitude of experimental EEG data by means of an unique algorithm: AJDC. In this section we show several examples of the use of AJDC for EEG data analysis and classification. Aim of this section is on one hand to show the power and flexibility of the method, on the other to present some of the research that we have been carried out.

Spontaneous activity

Typically, SOS BSS methods are used on continuously recorded EEG (*spontaneous activity*). As an example in Van der Loo *et al.* (2007) we have checked the validity of source extraction thanks to a joint EEG and ECoG (Electrocorticogram) recording. A 27 y.o. female patient suffering from right unilateral white noise tinnitus was implanted with two arrays of eight extra-dural ECoG electrodes posed on the secondary auditory cortex. The exact location was established by means of fMRI BOLD signal change associated with the tinnitus pitch. A neuronavigation system crossing fMRI and MRI information was used during surgery. The standard 19 EEG electrodes according to the 10/20 system (Jaspers, 1958) were recorded synchronously. Importantly, all leads, both EEG and ECoG, were referenced at the same location (vertex).

The AJDC BSS algorithm was applied on 184 seconds artifact free EEG data²⁹. Since no information was available about the presence and absence of the target rhythm during recording, only assumption 1 (coloration) is exploited. As suggested by the non-diagonality function (229), the non-diagonality of Fourier cospectra for scalp EEG data dropped significantly after 28 Hz. Accordingly, AJDC consisted in the *AJD* of all 1Hz-spaces cospectral matrices in the range 1-28 Hz. As many sources as sensors were extracted ($P=N=19$).

Figure 6.3 show the main results. This result has recently been replicated on six more Tinnitus patients with ECoG electrodes placed in various locations. The bigger study with these six further subjects is currently under submission (Van Der Loo *et al.*, submitted).

²⁹ Typically 20 to 40 seconds of data suffice to AJDC.

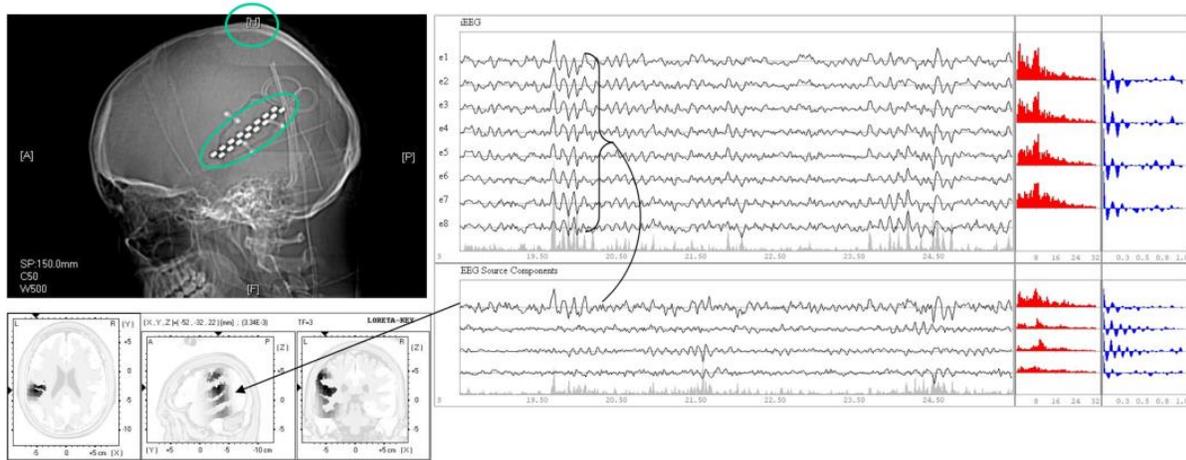


Figure 6.3: On the left the computerized tomography (CT) scan showing the emplacement of the ECoG electrode arrays and the reference on the vertex. On the right, about seven seconds of ECoG recording at the top 8-electrode array (e1-e8) and below the synchronous estimations of four of the sources obtained by AJDC. The first source clearly correlates very strongly with all 8 ECoG recordings. Source localization by means of model-driven sLORETA (121) pointed to the left temporal lobe as the location of this source. Results using the bottom ECoG array were very similar.

Induced activity

Since it can model both coloration and non-stationarity, the AJDC method is particularly suited to analyze induced activity like event-related de/synchronization (ERD/ERS). We have used it for motor imagery-based brain-computer interfaces (BCI). Motor imagery engenders frequency-specific ERD of the mu/beta rhythm over delimited areas of the sensory-motor cortex, corresponding to the body part interested by movement imagination, followed by a beta ERS, named “beta rebound” (Pfurtscheller and Neuper, 2001; Pfurtscheller et al., 2006). The spatial specificity of the ERD and ERS for different body parts is usually exploited by a common spatial pattern filter (200). This is still today considered a state of the art approach since it performs pretty well against competitors and is relatively simple (Lotte et al., 2007). In order to capture the source energy diversity in ERD and ERS (non-stationarity framework) in Gouy-Pailler et al. (2010) we have partition the 1s trials of motor imagery in four 250ms intervals estimating cospectral matrices in the mu and beta frequency range separately for these intervals. The *AJD* of these eight matrices was performed to extract $P=N$ sources. A mutual-information criterion between the source and the classes (Grosse-Wentrup and Buss, 2008) was used to select eight sources related to the ERD/ERS and the log-power of these sources was used in a logistic regression classifier. The method was applied to dataset 2a of BCI Competition IV (2008), provided by our collaborators working at the Institute for Knowledge Discovery (Laboratory of Brain-

Computer Interfaces), Graz University of Technology. In the data set nine subjects were involved in a four-class two-session motor imagery-based BCI experiment. The four classes were right and left hand, feet, tongue. EEG data was acquired at 22 electrodes concentrated on and around the sensory-motor areas. The performance of the method was tested against the CSP using a cross-validation procedure on the two sessions separately. No difference between sessions was found, suggesting that the data were homogeneous in the two sessions. However, the *AJD* method proved superior to the CSP aggregating the data of the two sessions ($t_{(17)} = 2.87$; $p = 0.027$). We then analyzed the session-to-session transfer ability of the filters. Training was performed on each one of the session and the test on the other session. In this case also the *AJD* method proved superior to the CSP ($t_{(17)} = 2.98$, $p = 0.022$). Importantly, we have shown that randomizing the indices of the data intervals before computing the averages, effectively disrupting the estimation of the changes in source energy, the advantage of the BSS method disappeared, proving that the superiority we have found springs indeed from the correct exploitation of ERD/ERS non-stationarity, which the CSP cannot do.

Evoked activity

We here report an unpublished study recently submitted for publication as a book chapter (Congedo, Jutten and Rousseau, in press). The study demonstrates the use of AJDC for the simultaneous extraction and analysis of an event-related potential and an event-related synchronization related to error detection. This study is provided as a practical example of BSS analysis for the general AJD-based BSS framework. This study is reported in several details, so as to illustrate the actual involvement in this kind of experimentation and data analysis. Also, by reporting in addition the analysis in the sensor space we can illustrate the advantage of EEG source analysis.

Introduction

ErrPs are a family of event-related potential (ERP) that can be elicited after the commission of an error, firstly reported in Miltner, Braun and Coles (1997) as associated to receiving external negative feedback after error commission. This feedback error-related potential (ErrPf) is characterized by a negative deflection peaking between 250 and 400 ms with a fronto-central scalp distribution. The authors named it the feedback-related negativity (FRN) and put it in relation with the response error related negativity (ERN) that had been previously reported (Felkenstein et al., 1991; Gehring et al., 1993), also characterized by a negative deflection. Initially the ErrPf has been studied prevalently in

the case of gambling tasks with monetary gain and loss. More recently it has attracted much attention in the brain-computer interface (BCI) community because its on-line detection provides a unique opportunity to automatically correct erroneous BCI operations, effectively increasing the consistency and transfer rate of a BCI system (Farquhar and Hill, 2013). In order to do so accurate on-line single-trial ErrP detection is necessary.

Experimental design

We study the feedback related potential in the case of a memory task, with no monetary gain or loss. The feedback is returned when the subject gives the answer and no reward is given to the subject except a score, thus our participants have no other interest besides their own performance. Such an experimental protocol allows to study the ErrP in a real "error versus correct" condition. The protocol we use is a memory task inducing a high cognitive load. The subject is continuously engaged in a demanding task (and not only on the feedback presentation), mimicking the actual conditions of a BCI use, where focus, concentration and attention are essential requisite for successful BCI operation. Then, in this study the feedback corresponds to the actual performance achieved in the task, again approximating the actual operation of a BCI. Finally, the *memory task continuously adapts* to the ability of the participants during the whole experiment. This ensures that the cognitive load is approximately constant across the duration of the experiment, that it is comparable across individuals regardless their memory span and that the error rate across subjects is approximately equal. This latter point is particularly important in ErrP studies since it is known that the error rate affects the ErrP ([8]). In this study the adaptive algorithm is tuned to engender an error rate of about 20%, which amount approximately to the reasonable accuracy of a reactive BCI operation in real-world situations.

AJDC analysis

Some of the previous studies on single trial ErrP classification (correct vs. error) have reached encouraging results (around 70% of overall accuracy) using only little a-priori knowledge on this potential. As usual, electrophysiological knowledge about the investigated phenomena can be used to select more relevant and robust features for the purpose of single-trial on line detection. Previous studies showed that the ErrP can be characterized both in the temporal domain as an ERP (time and phase-locked event) and as an event-related synchronization, or ERS (time but non-phase-locked event). The ERP is characterized by a negative deflection, named Ne, sometimes followed by a positive one named Pe (Gentsch, Ullsperger and Ullsperger, 2009; Steinhauser and Kiesel, 2011). The

ERS is characterized by an increased oscillatory activity in the theta frequency (4-7.5 Hz) occurring approximately in the same time window and spatial location as the Ne (Trujillo and Allen, 2007).

Source localization of the FRN using dipole analysis has suggested generators in the anterior cingulate cortex (ACC) and the supplementary motor area (Gehring and Willoughby, 2002; Miltner, Braun and Coles, 1997). Similar results have been obtained for the ErrPr.

Hereby we apply a sharp BSS approach with the aim to disentangling the sources responsible for the ERP and the ERS; if this proves feasible, then the ERP and ERS components will yield independent features to feed the classifier, hence potentially increasing the on-line accuracy.

As a first objective we identify the different components of the ErrP along dimensions time, space and frequency by means of a multivariate analysis both in the sensor space and in the source space. We jointly estimate the brain sources at the origin of the ERP and ERS components and assess their different role in error reaction.

Finally, we look at how these results impact on ErrP single-trial classification, which is the essential step in integrating ErrPs in BCI systems.

Method

Participants

22 healthy volunteers participated to this experiment. All subjects were BCI-naive at the time of the experiment and none of them reported neurological or psychiatric disorders in their lifetime. Due to the presence of excessive artifacts in the EEG data, three subjects were subsequently excluded from all analyses, leaving $M=19$ participants, of which 9 female and 10 male, with age ranging from 20 to 30 with a mean and a standard deviation of 24 and 2.52, respectively. All data was acquired in our laboratory in Grenoble.

Trials

The experiment involved two sessions lasting altogether approximately half an hour. Each session consisted of six blocks of six trials, for a total of $6 \times 6 \times 2 = 72$ trials. Participants seated comfortably 80cm in front of a 21-inch computer screen. Nine square boxes were arranged in a circle on the screen. Each trial consisted of the same memory retrieval task: the trial started with the display of the current

score for 3000ms (initialized at zero) followed by a fixation cross, also displayed for 3000 ms (fig 6.4a). Then the memorization sequence started; each memorization comprised a random sequence of two to nine digits appearing sequentially in random positions, with each digit of the sequence randomly assigned to a different box for each sequence (fig 6.4b). Subjects were instructed to retain positions of all digits. At the end of the sequence the target digit (always contained in the previous sequence) was displayed (fig 6.4c) and subjects had to click with the aid of a mouse on the box where it had appeared. Once the subject had answered, the interface waited for 1500 ms in order to avoid any contamination of ErrP by beta rebound motor phenomena linked to mouse clicking (Pfurtscheller and Lopes da Silva, 1999). Then, if the answer was correct, the chosen box background color turned into green ("correct" feedback), otherwise it turned into red ("error" feedback). Subjects were then asked to report if the feedback (error/correct) matched their expectation by a mouse click ("yes"/ "no") (fig 6.4d). Following this answer a random break of 1000 to 1500ms preceded the beginning of the new trial.

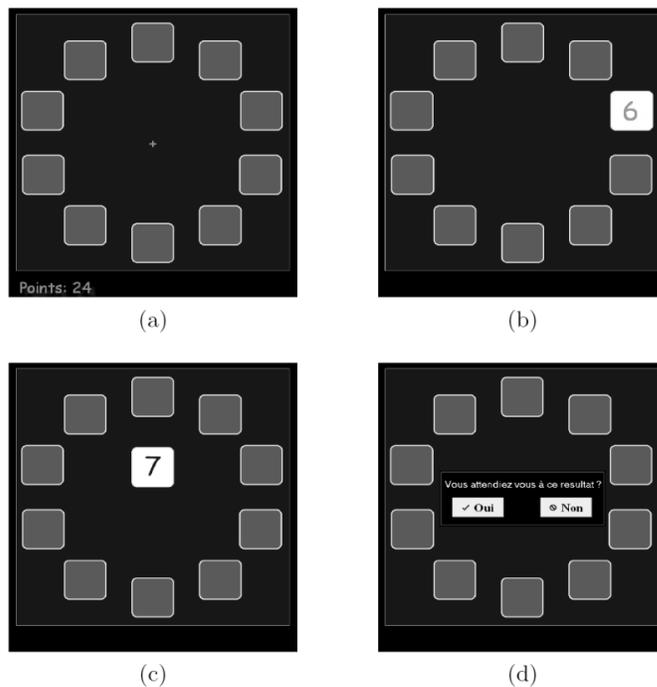


Figure 6.4: Screen shots from the experiment representing different steps of the trials. a): Fixation cross. b): One digit appearing in the memorization sequence. c): Target digit appearing. d): Feedback report question: 'Vous attendiez-vous à ce resultat' = 'Did you expect this result?', 'Oui'='Yes' and 'Non'='No'.

In order to keep the subjects motivated throughout the experiment, the accumulated score was computed at the beginning of each trial. When subjects localized correctly the target digits their score increased, otherwise, it remained unchanged. The number of digits in the sequence was always between two and nine, fixed within blocks and updated, at the beginning of each block, according to the change in performance from the block just finished and the previous one, as assessed on-line by statistical t-tests. The first block started always with four digits for all subjects. The parameters of the adaptation were set thanks to a pilot study and a computer simulation and were chosen to yield about 20% of errors, regardless the working memory ability. Moreover, our learning approach is capable of adapting to fatigue as well as other possible nuisance intervening during the experiment. A random rest break was allowed between blocks, during which the boxes performed a colorful animation chosen each time at random among four preset animations. Between the two sessions the screen was shut down to allow a rest break of 2 - 3 minutes.

Data acquisition

EEG recordings were acquired from $N=31$ silver/chloride electrodes positioned according to the extended 10/20 system (FP1, FPz, FP2, F7, F3, Fz, F4, F8, FT7, FC3, FCz, FC4, FT8, T7, C3, Cz, C4, T8, TP7, CP3, CPz, CP4, TP8, P7, P3, Pz, P4, P8, O1, Oz, O2) with the aid of a standard elastic cap. Both earlobes, digitally linked, were used as electrical reference. The ground sensor was positioned on the forehead. The impedance of each sensor was kept below 5k. The EEG was band-pass filtered in the range 0.1-70 Hz and digitized at 500 Hz using the Mitsar 202 DC EEG acquisition system (Mitsar Co. Ltd., Saint Petersburg, Russia). During recording, the stimulation program continuously sent to the Mitsar system triggers to track precisely all event onsets for each trial.

Preprocessing

Data were filtered in the 1-40 Hz band-pass region using an order four Butterworth FIR filter with linear phase response in the band-pass region. Ocular artifacts were extracted using the SOBI algorithm (Belouchrani et al., 1993) available in the EEGLAB toolbox (Delorme and Makeig, 2004). One EOG source corresponding to eye-blinks was suppressed for each subject. It was manually selected using both the temporal shape of the source and its topography. All other artifacts were left into the signal, so as to approximate the conditions of on-line analysis of EEG data acquired during BCI operation.

Analysis in the sensor space

ERPs were analyzed contrasting the average potential obtained from each subject at each electrode and time-sample. ERS were analyzed contrasting the average time-frequency map obtained on each trial from each subject at each electrode. In order to compute ERS we employed a multi-tapering Hanning sliding window (frequency dependent, with the taper equal to 4 cycles for each frequency) over the 2-32 Hz band-pass region using a 1-Hz step, as implemented in the Fieldtrip software (Oostenveld et al., 2011). ERS were computed on time window [-0,5s 1,2s] using a time step of 0,03s and a baseline defined as [-1s 0s] pre-stimulus.

The statistical analysis in the sensor space for contrasting “error” vs. “correct” trials needs to be performed for each electrode, discrete frequency and time segment in the case of ERS and for each electrode and time segment for ERP data. In order to account for the extreme multiple-comparison nature of the test we employed a permutation strategy. The test chosen is a slight modification of the supra-threshold cluster size permutation test originally proposed for neuroimaging data by Holmes et al. (1996). Here the statistic is not the supra-threshold cluster size, but the supra-threshold cluster intensity, defined as the sum of the t-values within the supra-threshold clusters. As compared to the test described by Holmes et al. (1996) such a statistic is influenced not only by the spatial extent of the clusters, but also by the strength of the effect. The test is sensitive to effects that are contiguous in space (adjacent electrodes), frequency and time, in line with physiological considerations. The family-wise error rate for multiple comparisons was set to 0.05, meaning that the probability of falsely rejecting even only one hypothesis is less than 0.05. All permutation tests were approximated by the use of 5000 random permutations.

Analysis in the source space

As we have seen BSS computes a weighted sum (linear combination) of the signal obtained at each electrode, isolating delimited dipolar sources from each other. We apply here the AJDC method introduced in this chapter adapting it to ERP data. Our goal is to separate the source of the Ne (ERP) and the source for the theta ERS. We need to separate them one from the other, but also from background EEG activity. For our purpose we need to include in the diagonalization set matrices holding

- a) the spatial structure of the ERP component,
- b) the spatial structure of the ERS component, as well as

c) the spatial structure of the spontaneous EEG oscillations and persistent artifacts such as lateral and horizontal eye movements, jaw muscle contractions, etc.

We then put into practice the guidelines provided in table 6.1.

For (a) and (b) we compute the relevant covariance matrices both on error trials and correct trials so to exploit variations of source energy between the two conditions. For the ERP components (a) we estimate the covariance matrix of the average ERP in the three time windows where the ERP analysis in the sensor space revealed significant results (see “result” section). Covariance matrices were separately computed for error and correct conditions, providing $3 \times 2 = 6$ matrices. *These six matrices provide unique source energy profile about ERP that have different potential in error vs. correct trials.*

For the ERS component (b) we estimate the averaged covariance matrix in the time-frequency region where the sensor space analysis revealed significant results (see “result” section). These matrices were computed as the covariance matrices of the EEG filtered in the frequency band of interest. Again, matrices were computed separately for error and correct conditions, providing two additional matrices. *These two matrices provide unique source energy profile about ERS that display different power in the theta band in error vs. correct trials³⁰.*

To separate possible sources of ERP and ERS from spontaneous EEG oscillations and artifacts (c) we include in the set all co-spectral matrices (Bloomfield, 2000) of the signal during the fixation cross sequence in the frequency range 2-20 Hz using a frequency step of 2 Hz, providing 10 additional matrices. *These latter 10 matrices provide unique source energy profile to separate all spontaneous sources having non-proportional power spectrum.*

In summary, our BSS algorithm jointly diagonalizes a total of 18 matrices. We define an exactly determined BSS model, that is to say, we estimate as many sources as electrodes ($N=P=31$). For solving the approximate joint diagonalization we employ the iterative algorithm proposed by Tichavsky and Yeredor (2009), which is fast and in our long-lasting practice has proven robust.

³⁰ Notice that matrices for the ERP and the ERS components are substantially different: for the ERP components EEG trials are averaged before computing the covariance matrix (thus only both time-locked and phase-locked signals are preserved), while for the ERS components trials are averaged only after computing covariance matrices on single-trial data (thus non-phase-locked nor time-locked signal are preserved).

Once estimated the 31 sources, they were inspected analyzing their ERP, ERS, topographies and the mutual information criterion between the source and the error class (Grosse-Wentrup and Buss, 2008). Meaningful sources were localized in a standard brain using the model driven sLORETA inverse solution (120). Source localization was conducted on each participant separately, normalized to unit global current density (the input of the inverse solution is a vector estimated by BSS up to a scale indeterminacy) and summed up over participants in the brain space.

Classification of single trials

For classifying single trials, data were band-pass filtered using an order four Butterworth FIR filter with linear phase response between 1-10 Hz for the ERP component and 4-8 Hz for the ERS component. Data were then spatially filtered using the results of the BSS analysis. Only samples corresponding to 250-750ms were kept. For the ERP component we used the temporal signal down-sampled at 32 Hz, providing 16 samples (features) for the classification. For the ERS component we used the square of the temporal signal (power) dawn-sampled at 32 Hz, providing 16 samples (features) for the classification as well. This procedure assigns to each component equal chance for classification. As a classifier we employed a LDA (linear discriminant analysis). One hundred random cross-validations were performed with the classifier trained on a randomly selected set containing 80% of the data (both errors and corrects) and then tested on the remaining data.

Results

Behavioral results

All subjects performed the task with a convenient error-rate, with mean (sd) = 22.2 (4)% and a quasi-equal repartition of expected and unexpected errors, with mean (sd) = 10.4 (4.3)% and 11.8(3)%, respectively. Reaction time was higher for error trials as compared to correct trials in 80% of the subjects (all t-tests with $p < 0.05$). The maximum number of digits to memorize for each subject was highly variable, ranging from 4 to 10, with mean (sd) = 6.5 (1.37). These results demonstrate that our presentation software succeeded in equalizing the cognitive load across subjects, despite the great inter-subject variability of digit memory span.

Sensor space analysis

The ERP in the error trials differed from the correct trials in three time windows with different timing and/or electrode location (fig. 6.5). A significant positivity for errors was found at time window [320ms 400ms] at electrode Cz ($p < 0.01$), a significant negativity for errors at time window [450ms 550ms] at clustered electrodes Fz, FCz, Cz ($p < 0.01$) and a significant positivity for errors at time [650ms 775ms] at clustered electrodes Fz, FCz ($p = 0.025$).

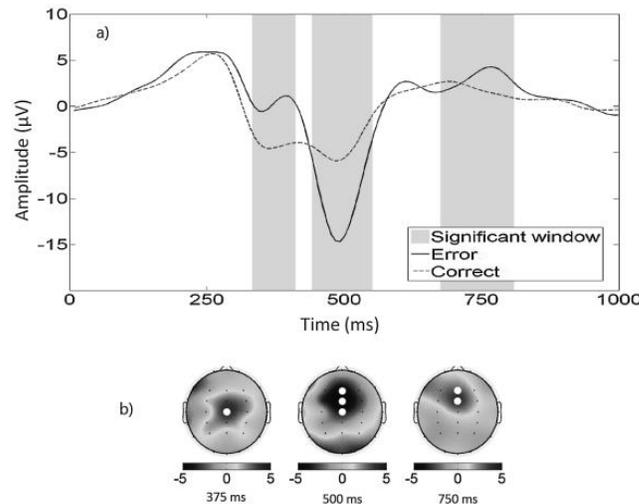


Figure 6.5: a): grand average ($M=19$) ERP for correct (pointed line) and error (solid line) trials. Time windows where the difference in amplitude between the two conditions is significant (grey panels) and (b) scalp topographies of t-values computed within the three significant windows. White disks show the significant clustered electrodes.

An ERS (power increase as compared to baseline) could be seen in the theta band in both correct and error feedback at fronto-midline locations. This synchronization unfolds from around 250ms to 600ms post-stimulus. In some subject it goes up to more than 200% of power increase for error trials. Albeit present in both conditions, this ERS is significantly more intense for error trials as compared to correct ones (fig. 6.6) in the frequency band pass region 5-8 Hz and time window [350ms 600ms] post-stimulus over the clustered electrodes Fz and FCz ($p = 0.015$).

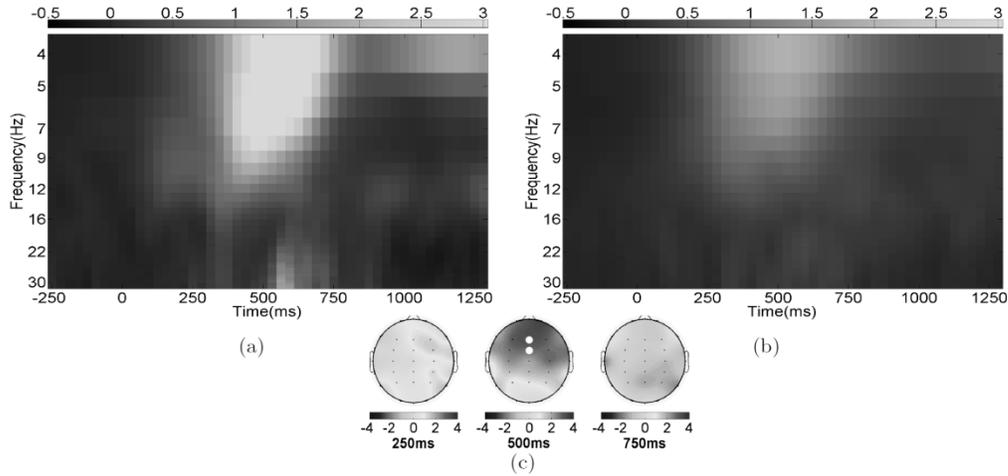


Figure 6.6: Grand average ($M=19$) ERS averaged at electrodes (Fz, FCz, Cz, CPz) for error (a) and correct (b) trials. (c): topographic maps of t -values averaged over the theta band and time window [350ms 600ms]. White disks show the significant clustered electrodes.

Source analysis

BSS analysis revealed two uncorrelated sources with variable sensitivity and specificity, however clearly responsible one for the ERP findings and one for the ERS findings. The source responsible for the ERP differences between error and correct trials, to which hereafter we will refer to as the “Ne source”, was significantly different in error vs. correct trials in two time windows, with a first negative peak at time window [460ms 540ms] ($p < 0.01$) and a positive peak at time [750ms 830ms] ($p = 0.015$). The grand-average ERP of this source computed separately for error and correct trials is displayed in fig. 6.7a. In fig. 6.7b it is displayed the same grand average ERP when computed using the spatial filter of the source responsible of the ERS differences between error and correct trials, to which hereafter we will refer to as the “theta source”; although differences in amplitude exist also for this latter source, they are not significant.

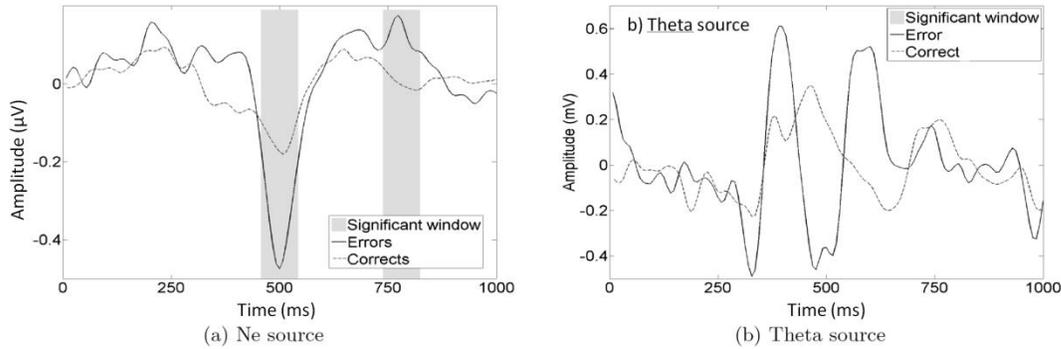


Figure 6.7: Grand averaged ($M=19$) of the ERP generated by the Ne source (a) and by the theta source (b) for error (solid line) and correct (pointed line) trials. Time windows where the difference in amplitude between the two conditions is significant are highlighted by grey panels.

On the other hand the theta source power increase was significant in frequency band-pass region 5-8 Hz for time window [300ms 600ms] ($p < 0.01$). The ERS generated by this source is shown in fig. 6.8b. In fig. 6.8a it is displayed the same ERS when computed using the spatial filter of the Ne source instead; the ERS in this case disappears. These results suggest that the Ne source and the theta source correspond to separate phenomena generated by different brain structures with different dynamics. The source responsible for the ERS (theta source) appears more specific.

We can now illustrate the advantage brought upon from the BSS analysis with these data. Compare fig. 6.7 to fig. 6.5 and fig. 6.8 to fig. 6.6. *Although in both cases results in the sensor space are computed for the optimal cluster of electrodes, it is clear that working in the source space allows a better sensitivity and specificity: in both cases the difference between the error and correct trials is highlighted.*

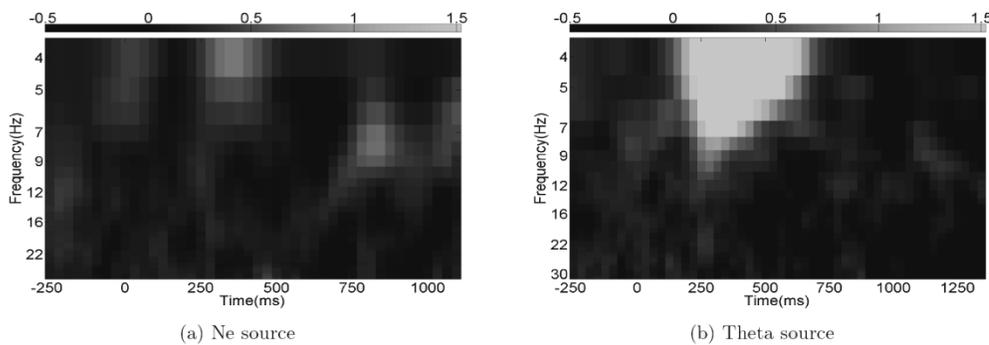


Figure 6.8: Grand average ($N=19$) of the ERS generated by the Ne source (a) and by the theta source (b) for error trials.

Source localization

The BSS source responsible for the ERP (Ne source) difference between correct vs. error trials was localized by sLORETA in the anterior cingulate gyrus (BA 24). The BSS source responsible for the ERS (theta source) was localized close to the supplementary motor area (BA 6) (fig. 6.9). Keeping in mind the approximation of a source localization method applied on a standard head model, these anatomical results are perfectly in line with results reported by previous studies (Gehring and Willoughby, 2002; Herrmann *et al.*, 2004; Nieuwenhuis *et al.*, 2003).

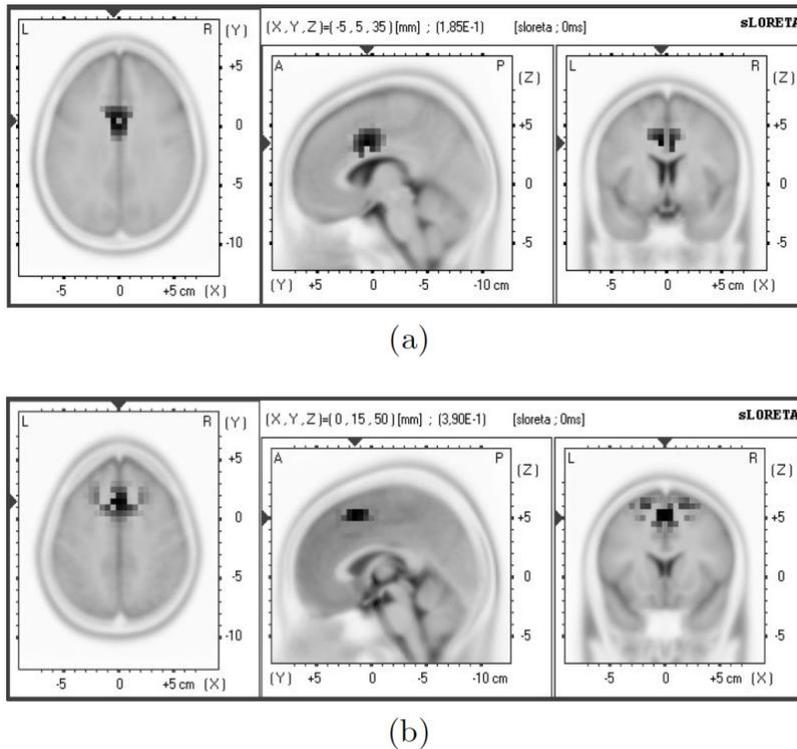


Figure 6.9: (a) Ne source sLORETA localization. The source is localized in BA 32. (b) Theta source sLORETA localization. The source is localized in BA 6. For each image, from left to right are the axial, sagittal and coronal views across the maximum. The images (a) and (b) are scaled to their own maximum. The activity is color-coded with black representing the maximum and transparent representing zero. Legend: A=Anterior; P=Posterior; S=Superior; I=Inferior; L=Left; R=Right.

Classification of single trials

The Ne source alone leads to better accuracy in classifying error trials as compared to the theta source alone ($p < 0.01$). The theta source leads to better accuracy for classifying correct trials ($p = 0.028$). These corroborate the conclusion that the ERP and ERS represent different phenomena of the ErrP.

When looking at the average classification rate $(T_e+T_c)/2$, with T_e the classification rate of error trials and T_c the classification rate of correct trials, one see that the use of both components leads to better results for 14 subjects out of 19. The use of both components increases the mean classification rate on the 19 subjects from 67% up to 71%. We performed a repeated measure two-way ANOVA with factor “type” (2 levels: error vs. correct) and “feature” (3 levels: Ne source ERP, theta source ERS, both). It revealed a main effect on the “type” factor ($p < 0.01$) with correct trials being better classified than error trials and a “type” x “feature” interaction ($p = 0.013$), demonstrating that the use of both the ERP and the ERS feature in the source space improves the performance of single trial classification. It should be noticed that with a total of 72 trials per subject, training set included only a mean of 17 single trials for the error condition, thus the classification task for this data set is hard since the training sets include very few examples of error trials. Better results are expected applying the method adopted in this study to larger data sets.

CHAPTER VII

GROUP AND JOINT BLIND SOURCE SEPARATION

Introduction

The BSS methods described in chapter VI target single individuals. The analysis of group data is necessary to generalize findings to the population of interest, may it be clinical or healthy, but also to construct EEG norms on large healthy populations so as to assess the deviance of single individuals (Ahn et al., 1980; John et al., 1980a,b,c). In neuroimaging, interest toward group BSS analysis has started with fMRI data (e.g., Calhoun et al., 2001) and has appeared in EEG literature only later (Congedo et al., 2010; Eiciele et al., 2011; Mueller et al., 2011; Ponomarev et al., 2013). There are many ways in which it is possible to extend BSS methods to group data. One major difficulty is that BSS methods estimate sources with arbitrary sign, energy and order. Therefore performing BSS for each individual and average the results is problematic. We do not consider here methods consisting in the clustering of results obtained at the individual level, as this is not a true group BSS approach.

Among true group BSS approaches, early attempts have aimed at estimating a single demixing matrix for all individuals (e.g., Calhoun, Liu et Adali, 2009). If there are M individuals in the sample, the extension of the linear instantaneous model (213) is then

$$\mathbf{x}_m(t) = \mathbf{A} \mathbf{s}_m(t), m \in \mathcal{R}\{1, \dots, M\}. \quad (231)$$

Since the mixing matrix, hence the demixing matrix, may be substantially different from one individual to another (fig. 2.4) such model in practice is useful only when large groups are analyzed; an average demixing matrix in this case represents gross directions and locations of dipolar current and the corresponding spatial patterns (mixing matrix) are very smeared and smoothed. Nonetheless, useful results can be found when the database is *very* large (Kropotov, personal communication). We will refer to the approaches tackling model (231) as *gBSS* (group BSS).

More recently, approaches inferring group sources capable of finding a specific demixing matrix for each individual in the group has appeared. They include extension of the CCA (208), *independent vector analysis* (Anderson et al., 2013), which we do not treat here and *joint blind source separation* (JBSS). Here we consider the JBSS approach, which is a natural extension of the *AJD*-based BSS methods considered so far for single-individuals. The underlying model is in this case

$$\mathbf{x}_m(t) = \mathbf{A}_m \mathbf{s}_m(t), m \in \mathcal{R}\{1, \dots, M\}. \quad (232)$$

In addition to the individual BSS assumptions, we assume that sources are correlated between subjects, that is

$$\mathbb{E}\left(\mathbf{s}_i^T(t)\mathbf{s}_j(t)\right)=\mathbf{D}_{ij}, \forall i,j \in \{1, \dots, M\}, \quad (233)$$

where \mathbf{D}_{ij} is diagonal and different from $\mathbf{0}$. As we have seen in chapter IV, given (233) true, model (232) is very different from a simple collection of M BSS, to which it reduces only *iff* $\mathbf{D}_{ij} = \mathbf{0}$ for all i,j . JBSS modeling is a very powerful tool, however we may use it only for data for which sample-by-sample correlations between subjects can be estimated. This is possible, for instance, in *hyperscanning studies*, a new paradigm for the neuroscience of social interaction (Adolphs, 2006, 2010; De Jaegher, Di Paolo and Gallagher, 2010; Frith and Frith, 1999, 2010; Hari and Kujala, 2009) where EEG is acquired synchronously and simultaneously on several subjects or for time and phase locked data, where a time “zero” can be established and sample-by-sample correspondence for different individuals (or different classes, experimental conditions, etc) is established. When this is not possible, the only available group BSS approach is the gBSS approach. We now turn to the presentation of our contribution to both approaches.

GROUP BLIND SOURCE SEPARATION (GBSS)

Introduction

We are concerned here with group model (231). In Congedo *et al.* (2010) we have presented a straightforward extension of the AJDC method to treat group EEG data (cognitive studies) and normative EEG of resting-states (Raichle and Snyder, 2007) for clinical studies. Many research groups have tackled model (231) by concatenating the group data in one single data matrix (Calhoun, Liu et Adali, 2009; Eichele *et al.*, 2011; Mueller *et al.*, 2011; Ponomarev *et al.*, 2013; Schmithorst and Holland, 2004). Given finite EEG observations for the M subjects $\mathbf{X}_1, \dots, \mathbf{X}_M$, where the observations may be spontaneous, induced or evoked activities, this entails constructing a big data matrix concatenating all observations such as (horizontal stacking)

$$\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_M) \quad (234)$$

and performing BSS with the chosen method on these data. Algorithmically, there is no difference from the single-subject BSS methods. In order to extend the AJDC framework to group data in Congedo *et al.* (2010) we simply diagonalize by *AJD* the grand-average cospectral matrices computed as the average of the cospectra obtained on the M \mathbf{X}_m data observations. Having averaged the cospectra, we can use the two-step algorithm described in (223) as it is. The method is analogous to the gBSS method used for fMRI and presented in Schmithorst and Holland (2004). Furthermore, it is evidently equivalent to the methods performing BSS on concatenated data if the EEG observations \mathbf{X}_m are of equal length. If they are not of equal length, averaging the cospectra is a preferable procedure in order to weight each subject equally. Actually, since the EEG is affected by an individual scaling factor, due to scalp thickness and other individual variables (Hernández *et al.* 1994; Goncalves *et al.* 2006), we also normalize the cospectra of each subject at each frequency to unit trace before computing the grand average. After normalization the contribution of each subject at each frequency is weighted exactly in the same way. Then the cospectra are weighted according to non-diagonality function (229). We now present our study on gBSS.

Method of Our Study

Databases

We applied the method on continuously recording eyes-closed resting state EEG employing a test-retest strategy in two independent large-sample normative databases ($M=57$ and $M=84$). One was a subset of the normative database of the Brain Research Laboratory (BRL), New York University School of Medicine ($M=57$; age range 17-30) and the other the normative database of Nova Tech EEG (NTE), Inc., Mesa, AZ ($M=84$; age range 18-30), which has been built by the author and Dr. Leslie Sherlin while at University of Tennessee. Exclusion criteria for the BRL database were known psychiatric or neurological illness, history of drug/alcohol abuse, current psychotropic/CNS active medications, history of head injury (with loss of consciousness) or seizure disorder. Exclusion criteria for the NTE database were a psychiatric history in any relative and participant of drug/alcohol abuse, head injury (at any age, even very mild), headache, physical disability and epilepsy.

Recording procedures

Recording procedures and settings were very similar for the two databases. In both cases 3 to 20 minutes of EEG data was continuously recorded while the participant sat with the eye-closed on a comfortable chair in a quiet and dimly lit room. EEG data were acquired from the 19 standard locations prescribed by the 10-20 international system (Jasper, 1958) using linked ear reference and enabling a 60 Hz notch filter to suppress power line contamination. The impedance of all electrodes was kept below 5K Ohms. Data of the NTE database were acquired using the 12-bit A/D NeuroSearch-24 acquisition system (Lexicor Medical technology, Inc., Boulder, CO) and sampled at 128 Hz, whereas data of the BRL database were acquired using the 12-bit A/D BSA acquisition system (Neurometrics, Inc., New York, NY) and sampled at 100 Hz. For consistency, we subsequently up-sampled the BRL database to 128 Hz using a natural cubic spline interpolation routine (Congedo, Ozen and Sherlin, 2002).

Pre-Processing

In order to minimize inter-subject variability we removed from all data any biological, instrumental and environmental artifacts, paying particular attention to biological artifacts generated by the eyes, the hearth and the muscles of the neck, face and jaw. All recordings included in this study were

artifact-free and featured high overall SNR. *This is important for gBSS analysis as the inter-individual variability of the artifacts and noise may influence negatively the solution.* The mean length and standard deviation of artifact-free data in the BRL and NTE database were 102.9 (27.5) and 92.5 (29.79) seconds, respectively.

Results

Using the proposed gBSS method and the sLORETA (model-driven) inverse solution (123) we could closely replicate in the two databases both the spatial distribution and spectral pattern of *seven source components*. That is to say, for the truncated pre-whitening in (225) we fixed the dimension of F to $P=7$. The source localizations of the seven components along with their absolute and normal power spectrum for the two databases are shown in fig. 7.1. Table 6.2 reports the main cortical structure involved in each component. For the physiological interpretation of these components, awaiting confirmation by experimental and clinical data, we refer the reader to Congedo et al. (2010).

Table 6.2: Anatomical structures and Brodmann areas (BAs) where high-power (more than 50% of the maximum) current source is located for the seven gBSS components in the two databases (leftmost column). BAs with top 10% current density power are highlighted in bold. When the side is not specified labels and corresponding BAs apply bilaterally.

Anatomical Structures and Brodmann Areas	
1	Anterior Cingulate (BA 23/ 24 /32/33/ 25), Insula (BA 13), Middle/Superior Frontal Gyrus and Paracentral Lobule (BA 4/5/6), Parahippocampal/Subcallosal Gyrus (BA 28/34/35/36)
2	Cuneus/Precuneus/ (BA 7/31 /18/19/), Post-central gyrus (BA 3/4/5), Superior Parietal and Paracentral Lobule (BA 5/7), Posterior Cingulate Gyrus (BA 23/31)
3	Cuneus/Precuneus/ (BA 30/31/7), Right superior parietal lobule (BA 7), Posterior Cingulate (BA 30), Lingual/Parahippocampal Gyrus (BA 18/19/30), Right Fusiform Gyrus (BA 19)
4	Cuneus/Precuneus/Posterior Cingulate (BA 23/30/31), Lingual Gyrus/Fusiform Gyrus/Middle and Inferior Occipital Gyrus (Occipital Pole) (BA 17/18/19)
5	Anterior Cingulate (BA 24/25 /32), Medial Frontal Gyrus (BA 32/9/10/11), Rectal/Orbital Gyrus (BA 11/47), Inferior Frontal Gyrus (BA 47), Parahippocampal Gyrus (BA 28/34)
6	Medial Frontal/Rectal Gyrus/Anterior Cingulate (BA 11, 25), Middle Frontal Gyrus (BA 11), Inferior Frontal Gyrus (BA 47), Parahippocampal Gyrus (BA 28/34), Insula (BA 13)
7	Post-central Gyrus (BA 1/2/3), Supramarginal Gyrus/Inferior Parietal Lobule (BA 40), Precentral Gyrus (BA 6), Cuneus/ Precuneus (BA 17/18/19/31), Middle Occipital Gyrus (BA 18), Superior and Middle temporal Gyrus (BA 21/22/39/41), Insula (BA 13), Angular Gyrus (BA 39)

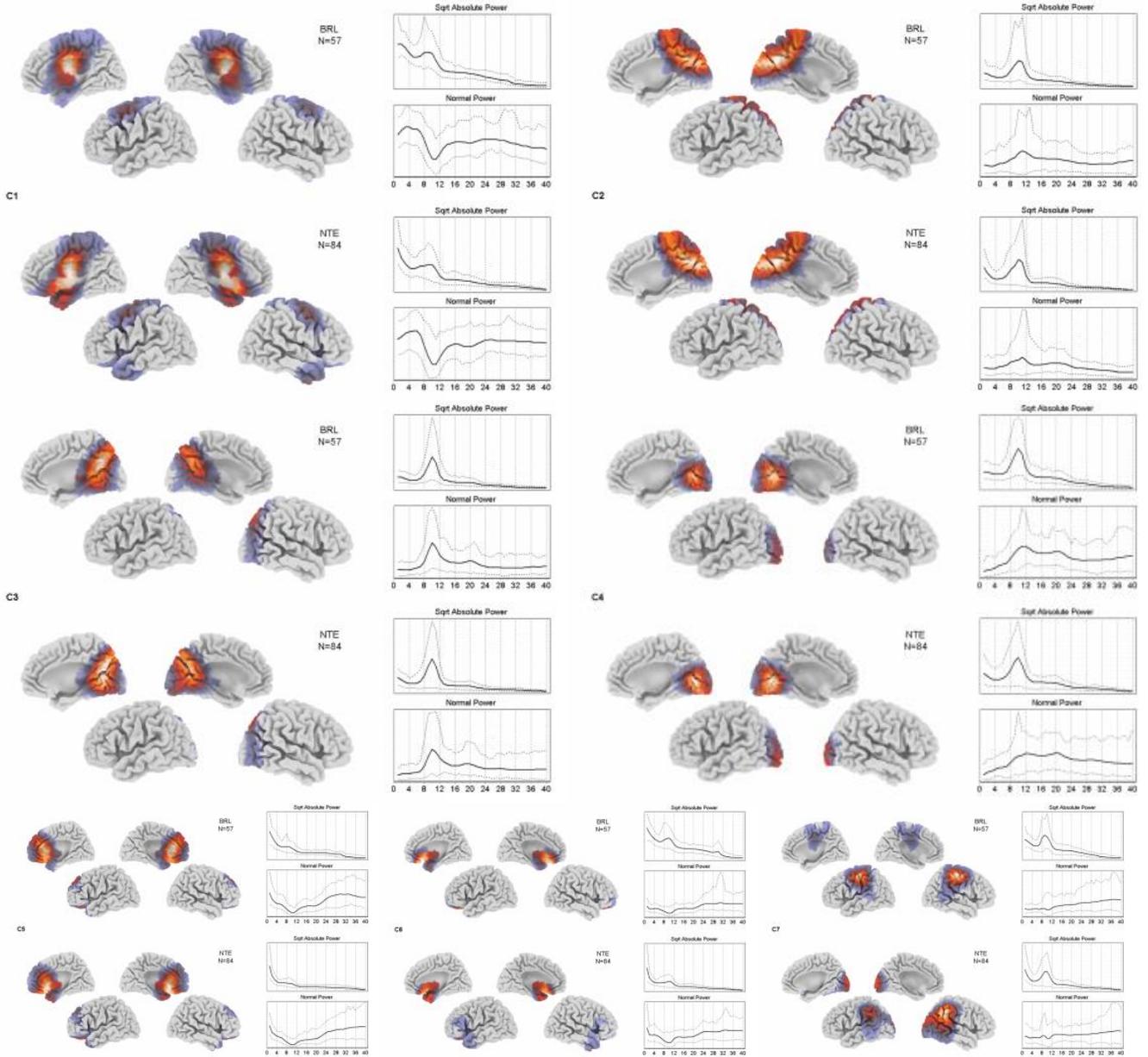


Figure 7.1: sLORETA cortical current density images and associated frequency spectrum of the seven independent components for the BRL (top) and NTE (bottom) database. From left to right:

- the sLORETA cortical image medial and lateral views of the left and right hemisphere. The current density is thresholded at half the maximum.

- the mean (solid line) and 95% confidence interval (dotted line) of the grand-average frequency spectrum in the range 0.5-40 Hz for absolute and normal power. The vertical axis is adjusted individually in each plot. The absolute power for the m^{th} individual and f^{h} frequency is given by $\mathbf{b}_p^T \mathbf{C}_{mf} \mathbf{b}_p$. The normal power is obtained by normalizing each \mathbf{C}_{mf} to unit trace (energy) before computing the power. Such normalization eliminates the dominance of frequencies in the range 8-12 Hz and shows the relative involvement of each frequency with respect to the others. Cortical images have been produced by the sLORETA-Key software (see chapter III).

Explained variance

The reader may wonder why we have found seven replicable components and not a higher number. We have used an empirical approach; starting with a small number of components we have increased the number until the output components had close spatial localization and power spectrum in the two databases; fixing the dimensionality reduction to more than seven resulted in components no more replicable in the two databases. This is understandable if we study the explained variance of the components, introduced in Congedo *et al.* (2008). Figure 7.2 shows the proportion of variance explained by each component with respect to the total as obtained on the two databases. Note that not only the output components are sorted identically for the two databases, but also that each component individually contributes a similar increase of explained variance in the two databases. Note also that component 7, which spatial distribution in the two databases matches only roughly (fig 7.1), explains less than 5% of the variance. This suggests that increasing P over seven would require the estimation of components with very low SNR , which explains why it is hard with these data to closely replicate more than seven components.

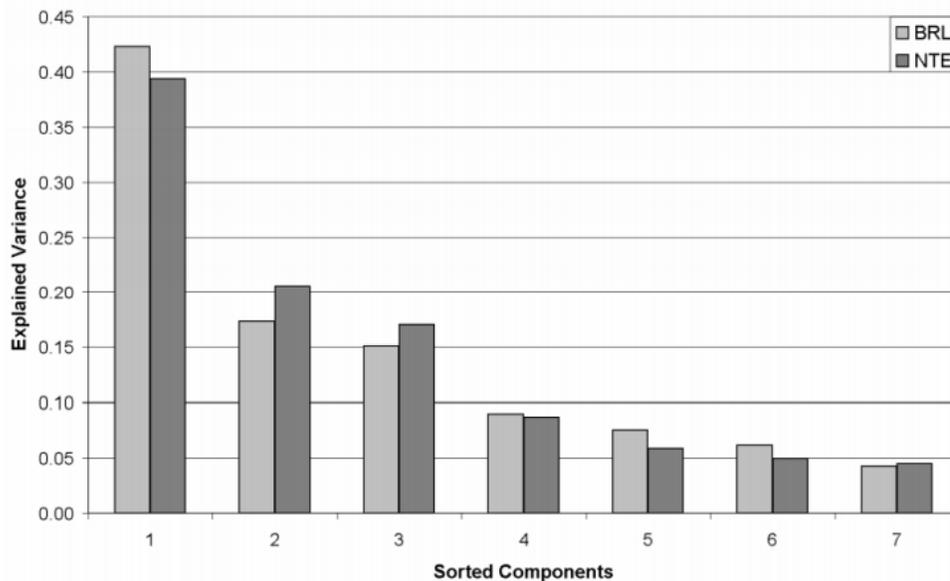


Figure 7.2: Proportion of the grand-average explained variance (Congedo *et al.*, 2008) for the seven gBSS components found independently on the NTE and BRL databases.

Deriving group norms

Norms were then constructed for the spectral power of the seven components so as to allow testing patients against the norms. This work update the seminal work of Ahn *et al.* (1980) and John *et al.* (1980a,b,c; 1987) from the sensor space to the source space. Norms are defined as the empirical distribution of the power across the normative database (Congedo and Lubar, 2004). When an individual is to be tested against the norm its source power spectra are compared to these distributions; for each feature to be tested the probability to deviate from the norms is given by a function of the position of the individual within the sorted normative empirical distribution (percentiles). The reader is referred to Congedo and Lubar (2004) and to Congedo *et al.* (2010) for details on how to derive norms and test the deviance of individuals against the norms.

Note that as compared to existing normative databases based on scalp spectral features, the gBSS tool defines a smaller number of features with very little inter-correlation. Furthermore, these features may be physiological meaningful as they relate the activity of several brain regions, forming a total of seven patterns, each with a peculiar spatial distribution and spectral profile. On the other hand, having reduced considerably the number of extracted sources, very likely several sources are mixed together in each component. In order to obtain more specific and replicable components, the method should be applied on *very large* databases ($M > 1000$).

Experimental Studies with gBSS

In this section we illustrate the gBSS methods at work with real data. We present a cognitive study and two clinical studies, carried out in collaboration with some of my PhD students and their principal supervisors.

Clinical gBSS studies

In a well-controlled clinical study on obsessive-compulsive disorder (OCD) by Kopřivová *et al.* (2011) we have replicated one more time the seven normative components on a third independent database ($M=50$). Only one component was found to display differences of power in the OCD sample ($M=50$) as compared to matched controls ($M=50$). The source localization and power spectrum of this normative component is shown in fig. 7.3. The component corresponds very closely to component 1 in fig. 7.1. Notice that both the spatial distribution of current density and the power spectrum found by

Kopřivová *et al.* (2011) are very similar to those found by Congedo *et al.* (2010). The relative and normal power of the OCD sample were significantly higher as compared to the controls in the frequency range 3-6 Hz (multiple-comparison permutation test (Holmes *et al.*, 2006; Westfall and Young, 1993), $p < 0.05$ corrected). The difference remained similar and significant at the 0.1 level when SSRI-medicated ($M=30$) and medication-free ($M=20$) patients were tested separately against the controls. This finding has been used in Kopřivová *et al.* (2013) to train selectively this component yielding the first gBSS-based spatially filtered neurofeedback study.

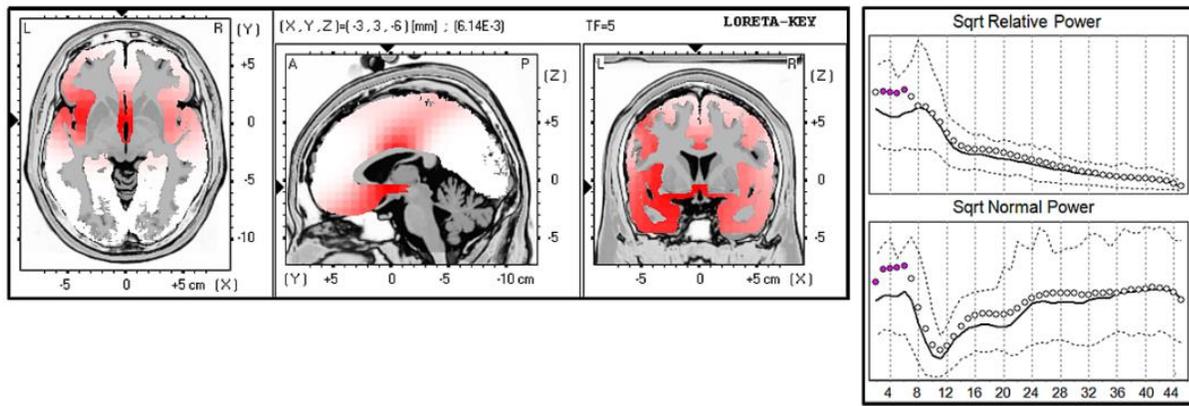


Figure 7.3: Left: axial (top of the picture is front of the head), sagittal (left of the picture is front of the head) and coronal (top of the picture is top of the head) view of the normative component displaying significant power differences between OCD patients and controls. The intensity of the red color indicates the intensity of the signal contributing to the component. This component is the same we have previously found in Congedo *et al.* (2010), see component 1 in fig. 7.1. Right: The solid (dashed) line(s) is the mean (5% and 95% percentile) power spectrum of the component for the control group ($M=50$). The disks indicate the mean for the OCD sample ($M=50$). Filled disks indicate a significant difference between OCD and controls ($p < 0.05$, corrected for multiple comparison).

Cognitive gBSS studies

The gBSS approach is not limited to clinical studies. For example, we have used it in a cognitive study on spatial navigation (White *et al.*, 2012). Using 64 EEG electrodes, 26 subjects performed a spatial navigation task in a 3D virtual city. The gBSS method was applied to the EEG data acquired during the spatial navigation so as to extract source components related to the task. For the AJDC algorithm a subspace reduction by truncated whitening limited the estimation to 36 components, allowing more than 99% of explained variance to be preserved in the solution (Representation Error = 0.0093, see

(230)). Once estimated the group demixing matrix, it was used to extract source power in both the spatial navigation and a control condition. Power for all 36 sources extracted was estimated in all 1-Hz spaced discrete frequencies in the range 1-44 Hz in all subjects and compared between the baseline and navigation condition. Permutation t-max tests (Holmes et al., 2006; Westfall and Young, 1993) were employed to correct for multiplicity of frequencies, whilst a Bonferroni adjusted significance threshold of $p=0.05/36=0.00139$ was used to control for multiple comparisons across the 36 components.

Two sources were found to exhibit significant spectral power differences during navigation with respect to the control condition and were subject to source localization using model-driven sLORETA (120). These two sources were localized as a right parietal component with gamma activation and a *right medial-temporal-parietal component* with activation in theta and gamma bandwidths. The power in the theta band for the latter source was significantly higher in the navigation condition as compared to the baseline condition. The source localization and its grand average power spectrum in the baseline and navigation condition are shown in fig. 7.4.

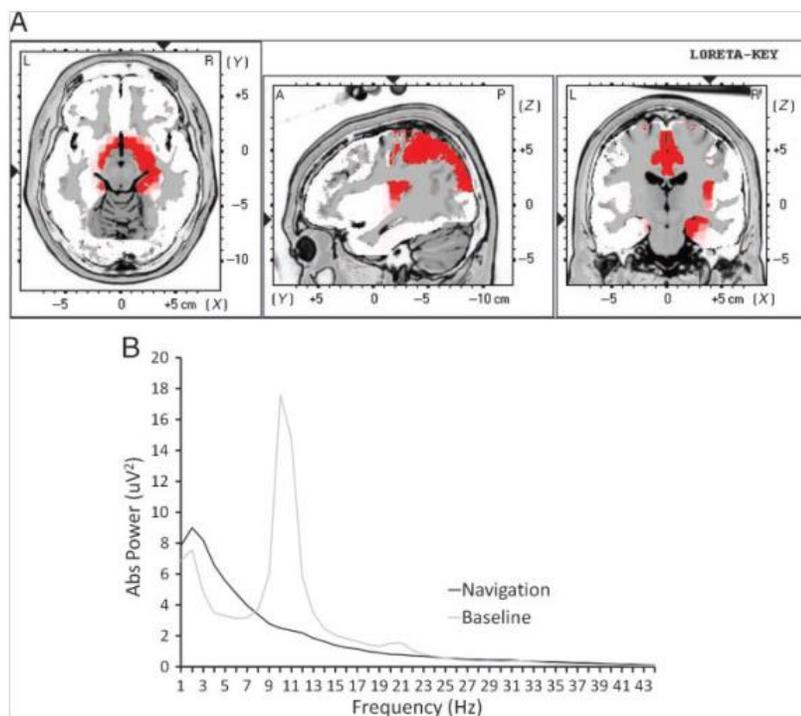


Figure 7.4: (A) sLORETA source maps for the right medial-temporal-parietal component. From left to right: axial, sagittal, and coronal sections. A = posterior, P = posterior, L = left, R = right. (B) Grand average ($M=64$) power spectra for the component during navigation and baseline conditions.

Furthermore, the theta activity on the medial-temporal/parietal source was positively correlated with more efficient navigation performance (fig. 7.5), measured with the navigation latency, in seconds, where a shorter latency indicates higher efficiency in reaching the landmarks.

These findings are intriguing, as it is usually very difficult to extract deep medio-temporal activity by means of EEG. Nonetheless, gamma and theta oscillations have been linked with numerous aspects of human spatial navigation using intracranial EEG (Caplan *et al.*, 2001; Ekstrom *et al.*, 2005; Jacobs *et al.*, 2009) and MEG (Cornwell *et al.*, 2008, 2010). Our study suggests that the gBSS approach successfully visualized medio-temporal (possibly parahippocampal) theta oscillations related with spatial navigation and that this activity is related to parietal activation, which is in line with current knowledge. However, a replication of this study is needed to support this claim.

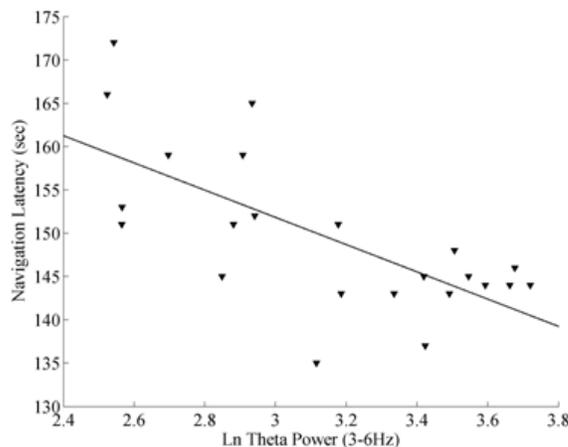


Figure 7.5: Scatter plot of navigation latency against the right medial-temporal–parietal component theta (3–6 Hz) power during navigation ($M=24$, two outliers were removed from the analysis; $r=-.659$, $p=.001$).

Limitations of the gBSS Approach

To end up the gBSS section we here present unpublished data inspecting the extent to which group components fit appropriately individual data. For more details on the gBSS method we refer the reader to Congedo *et al.* (2010). For a similar approach on ERP data see Mueller *et al.* (2011) and Ponomarev *et al.* (2013).

- An important question is whether the gBSS method is a valid instrument, that is, how precisely the components extracted by gBSS on each individual of the database match the corresponding group components. We have studied the similarity between the individual and group component power spectra. We have done that for each individual and for each one of the seven components. We computed the group spectra in between 2 Hz and 40 Hz for each gBSS component and then all individual spectra in between 2 Hz and 40 Hz for the same gBSS components. We then tested the hypothesis that the individual spectra are the same as the group spectra, repeating the test for all components and for all individuals³¹. The p -values for each component are shown as box-plots in fig. 7.6. As it can be seen, the minimum p -value is larger than 0.5, thus for no individual and for no component the spectra significantly differs from the corresponding group spectra. This result demonstrates that the gBSS spatial filters extract activity with similar power spectrum in all individuals. Particularly consistent appear the results for component 1.

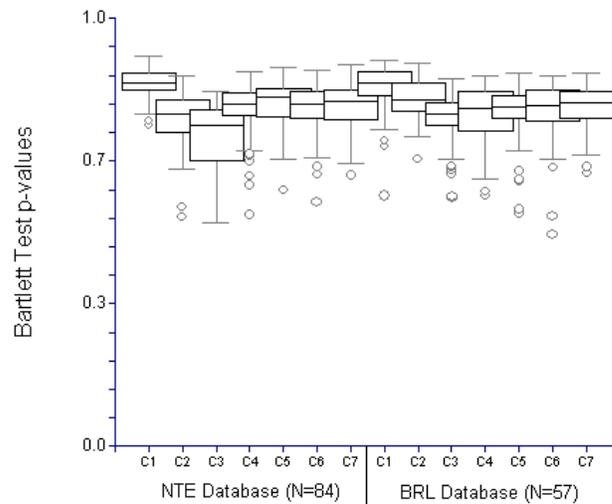


Figure 7.6: Box plots of the p -values of the Kolmogorov-Smirnov test for equality of the power spectrum of the gBSS components and of the individual components derived with the gBSS demixing matrix, for all components (C1 to C7) and the two databases analyzed in Congedo et al. (2010).

³¹ The test we used is the Kolmogorov-Smirnov test for uniformity of the absolute difference between the two spectra. That is, after normalizing the spectra so as to have total power equal to 1, we compute the difference between the individual and group spectra and take the absolute value of this difference. If the resulting spectrum is uniform (white), then there is no difference between the two spectra.

- Another limitation of the gBSS approach is that the seven gBSS filters are “grand-average” filters and the resulting individual sources (the sources extracted by applying such filter to individual data) cannot be very well decorrelated. As a consequence the gBSS demixing matrix is a sub-optimal spatial filter to derive individual components. This is due to the fact that the grand-average filters cannot take into consideration the individual physical head model and individual source distribution. In fig. 7.7 we report the correlations between individual sources in the range 2-40 Hz. As it can be seen, for the majority of source pairs and frequencies the correlation is low ($<abs(0.2)$), whereas only for some of them the correlation is high ($>abs(0.4)$).

This result is not surprising as approximate decorrelation for all source pairs and all frequencies can be obtained only if we apply blind source separation (BSS) to individual data or by JBSS; in other terms, if we want to work at the individual level the grand-average demixing matrix can be regarded only as a rough approximation. Whenever individual component issued from a gBSS analysis are needed with precision, as for example to train selectively gBSS components by means of neurofeedback, we propose here a solution:

Algorithm (235): Individually Refining gBSS Components

Referring to model (231), let $\mathbf{x}_m(t)$ be the individual data and let $\bar{\mathbf{A}}, \bar{\mathbf{B}}^T$ be the gBSS (grand-average) mixing and demixing matrix, respectively.

First project the individual data in the gBSS space as

$$\mathbf{y}_m(t) = \bar{\mathbf{B}}^T \mathbf{x}_m(t), \quad (236)$$

Second extract individual sources as

$$\mathbf{s}_m(t) = \mathbf{F}_m^T \mathbf{y}_m(t), \quad (237)$$

where \mathbf{F}_m^T is the demixing matrix obtained running again the BSS algorithm on individual projected data.

The solution (individual demixing matrix) is $\mathbf{F}_m^T \bar{\mathbf{B}}^T$ and its inverse $\bar{\mathbf{A}} (\mathbf{F}_m^T)^+$.

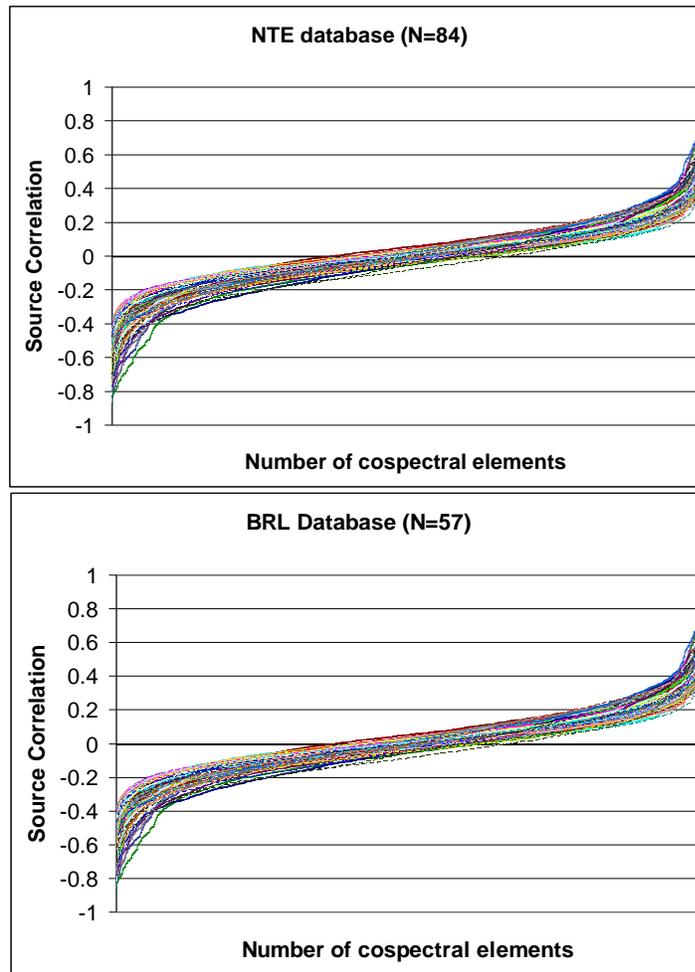


Figure 7.7: Correlations between the gBSS sources as extracted on all individuals in the data bases studied in Congedo *et al.* (2010). Each line in the plot corresponds to one subject in the database. Each line traces the correlation between all pair-wise sources (21 pairs for seven sources) times 77 frequencies (from 2 to 40 Hz in 0.5-Hz increment), that is, a total of $21 \times 77 = 1617$ points, which are sorted in ascending order for all subjects in order to make the plot readable.

The rationale for this double BSS procedure is this: the first projection filters out data that is far away from the gBSS components. This will result in a set of seven components that are still somehow correlated. The second refines the source estimation applying a BSS on the individual projected data. The resulting sources will be well decorrelated and most probably well correlated to the gBSS sources. The process is illustrated in fig. 7.8. In the figure:

- a) shows 10 seconds of EEG data of one subject of the NTE database. Next to the EEG traces there is the corresponding power spectrum in between 1 and 32 Hz (red plot), the autocorrelation function (blue plot) and the Hurst exponent³² (green bar).
- b) shows the corresponding (time locked) seven time-courses of the gBSS components, that is $y_m(t)$ of (236). These are the gBSS projected data.
- c) shows the sources found by applying the BSS algorithm on the projected data in b), that is, $s_m(t)$ in (237).
- d) shows the 7x7 correlation matrix (all correlations are computed in the time domain and using the whole recording, not just the visible data) of b) (top-left square), the 7x7 correlation matrix of the data in c) (bottom right square) and the cross-correlation between data in b) and data in c) (top-right square). In the bottom-left square you see its transpose, that is, the cross-correlation between data in c) and in b). In all this plots the correlation is color coded, with red coding positive correlation and blue coding negative correlation. In BSS models the sign of the source cannot be recovered, thus only the absolute magnitude of the correlation is relevant. Note that the diagonal elements in the figure are the autocorrelations, thus corresponding to a correlation=1.0. As it can be seen the gBSS sources in b) are only approximately decorrelated (top-left 7x7 square), whilst the individual sources are decorrelated very well (bottom-right 7x7 square). Now, suppose we are interested in training by neurofeedback gBSS component 3. From the top-right 7x7 matrix one can see that the gBSS component 3 in b) correlates very highly ($r= 0.98$) with the first refined individual BSS component in c). Thus, we can use the filter of the refined BSS and be sure that we have found optimal coefficients to train indeed gBSS component 3. If component 4 was the component of interest, then we would use the fifth spatial filter of the refined BSS, which correlates very well. Note the similarity of the time courses, power spectrum and autocorrelation function for the gBSS “group” sources 3 and 4 in b) with the “individually refined” BSS sources 1 and 5 in c), respectively. In other circumstances the process does not issue conclusive results. For example component 2 correlates moderately with at least four of the individually refined BSS sources, so the group source cannot be found uniquely on this subject.

³² The Hurst Exponent (HE) is an index of the “memory” of a time series related to the fractal dimension and it is comprised between 0 and 1. Roughly speaking an $HE < 0.6$ is characteristic of EEG signals contaminated by EMG, $0.5 < HE < 0.82$ is characteristic of clean normal EEG and $0.8 < HE < 1.0$ is characteristic of EEG signals contaminated by eye-movements. It is sometimes used to select independent components for artifact removal (e.g., Vorobyov and Cichocki, 2002).

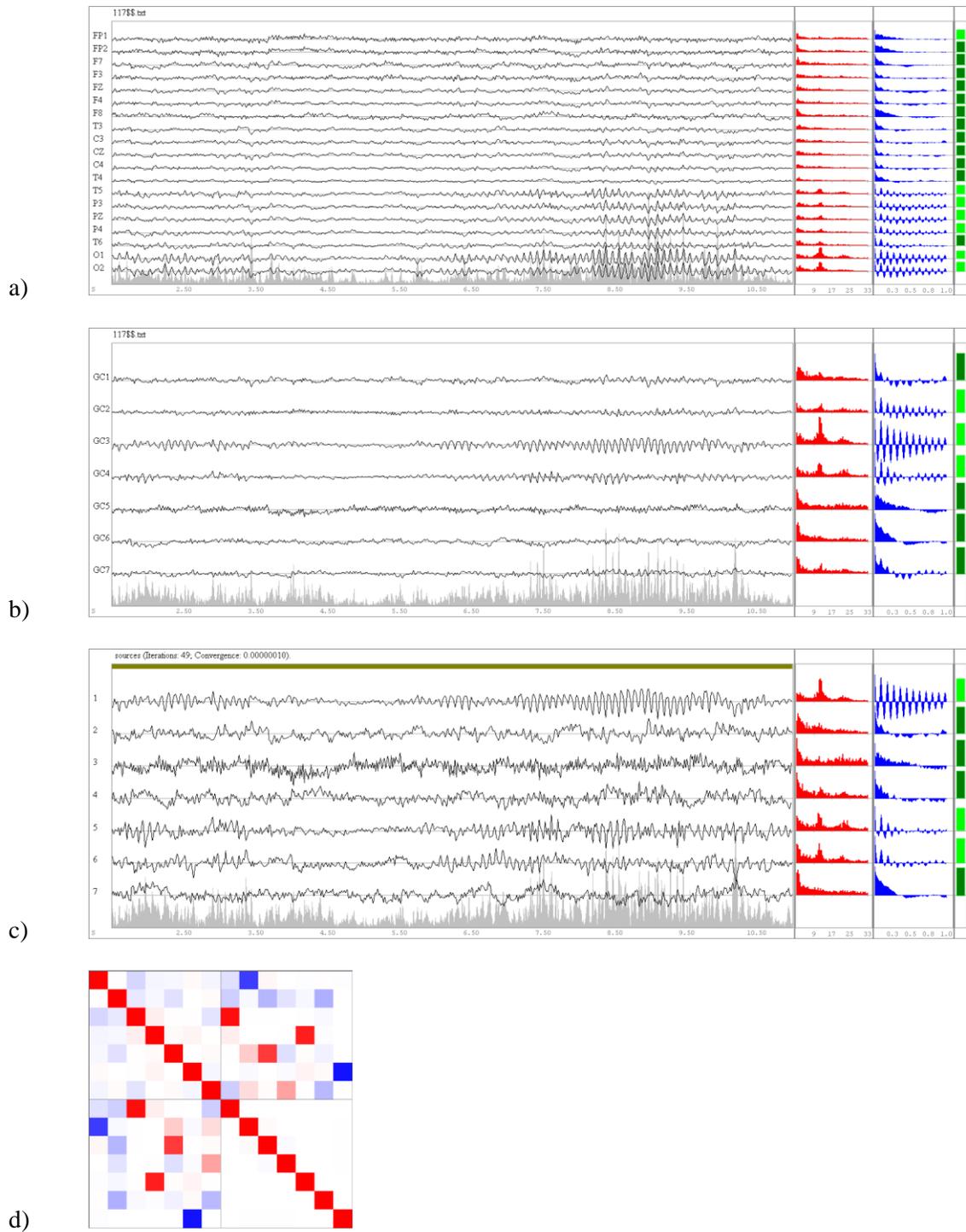


Figure 7.8: Illustration of the individualization of gBSS components (235) obtained by refining gBSS components by an individual BSS. See text for details.

JOINT BLIND SOURCE SEPARATION (JBSS)

Introduction

Extensions of the *AJD* BSS methods to the analysis of simultaneous multiple-subject EEG data is currently an active line of research in our laboratory (Chatel-Goldman, Congedo and Phlypo, 2013; Congedo, Phlypo and Pham, 2011; Congedo, Phlypo and Chatel-Goldman, 2012). In chapter IV we have presented two algorithms that can solve the joint blind source separation (JBSS) problem. We here precise the JBSS model in light of the theory of BSS. We also present unpublished material on an extension of the OJLSD and IJLSD algorithms presented in chapter IV allowing more flexibilities on the JBSS modeling.

The JBSS Framework

Suppose we are given M datasets, e.g., M subjects, with $m \in \{1, \dots, M\}$. As usual, we suppose that each data set is multidimensional, such as $\mathbf{x}_m(t) = [x_m^{(1)}(t), \dots, x_m^{(N)}(t)]^T$, wherein N random variables as before indexed by $n \in \{1, \dots, N\}$, unfold over the discrete dimension $t \in \{1, \dots, T\}$. Note that n and t may refer to, for instance, space (EEG channels) and time, respectively, as it is the case for EEG, but this is not important for the sequel. As for the single subject BSS framework, suppose further that K observations are available for each dataset, indexed by $k \in \{1, \dots, K\}$, yielding KM groups of N variables. As in the SOS framework presented for single-subject BSS in chapter VI, the K observations may refer to K experimental conditions, to recordings in K different times (or trials), or to an expansion of the original data in K discrete frequencies or K time-frequency regions. We are concerned here with model (232) $\mathbf{x}_m(t) = \mathbf{A}_m \mathbf{s}_m(t)$, where $\mathbf{A}_m \in \mathbb{R}^{N \times P}$ is a (t, k) -invariant full column rank mixing matrix specific to each set and $\mathbf{s}_{m,k}(t) \in \mathbb{R}^P$ (with $P \leq N$) holds the source components over the t dimension. So, contrary to the gBSS approach, each one of the M EEG observations has its own mixing matrix and sources, however as usual the mixing matrix is the same for each dataset along the K observations. The M set of sources are assumed decorrelated within sets as in single-subject BSS, but also correlated between data sets as per (233). In symbols, while for single subject BSS model $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$ we assume that the matrix set of statistics $\mathbf{C}_k \in \{\mathbf{C}_1, \dots, \mathbf{C}_K\}$ is generated theoretically by process

$$\mathbf{C}_k = \mathbf{A}\mathbf{S}_k\mathbf{A}^T, \quad (238)$$

where the \mathbf{S}_k matrices are (unknown) source covariance matrices and are assumed diagonal because of the fundamental decorrelation assumption, for the JBSS model we assume that the matrix set of statistics $\mathbf{C}_{ij,k}$, with $k \in \{1, \dots, K\}$ as before and $i, j \in \{1, \dots, M\}$, is generated theoretically by process

$$\mathbf{C}_{ij,k} = \mathbf{A}_i \mathbf{S}_{ij,k} \mathbf{A}_j^T. \quad (239)$$

In addition to the single-subject BSS framework we assume that the sources are correlated between data sets, that is, that each matrix $\mathbf{S}_{ij,k}$, $i \neq j$, is diagonal and different from $\mathbf{0} \forall i, j \in \{1, \dots, M\}$. Important for exploiting cross data sets SOS statistics, as we do in this framework, the data samples over dimension t must reflect the same metric for all data sets. For example, if t is time, as per EEG, the samples must be taken synchronously in all data sets or a reference sample must be identified and relative intervals between samples across data sets must be equal (for instance, relative to a stimulus presentation). This is essential to estimate correctly the cross-statistics between data sets, which is the core of the JBSS approach. This model has been proposed already several times also by others (Anderson, Adali and Li, 2012; Vía et al. 2011; Li, Adali and Anderson, 2011; Li et al., 2009) and is currently attracting much attention because of its generality and its ability to exploit dependencies between data set. The framework is similar to the one presented for MCA and CCA (203), however there are substantial advantages working in the JBSS framework, namely, we may allow $M > 2$, we may allow $K > 1$ and the JBSS is a waveform-preserving (source estimation) framework.

In JBSS we require to find the M matrices $\mathbf{B}_m \in \mathbb{R}^{N \times P}$, $m \in \{1, \dots, M\}$, yielding source estimates

$$\hat{\mathbf{s}}_m(t) = \mathbf{B}_m^T \mathbf{x}_m(t), \quad (240)$$

where the demixing matrices can be estimated up to a sign, scale and permutation indeterminacy, as in the BSS case. However in JBSS we require the permutation be the same for all M datasets, otherwise the analysis of the corresponding sources in the M datasets becomes difficult. We say that the output sources are *aligned* across data sets. This key advantage of the JBSS approach is made possible thanks to additional assumption on the diagonality and non-vanishing cross-statistics. Notice that for the sake of simplicity we suppose hereafter that $N=P$, this quantity being the same across datasets, i.e., we consider the *exactly determined* case, however the OJLSD and IJLSD algorithms presented in chapter IV apply equally well for the case $P < N$. The JBSS model generation (239) can be written in matrix form as

$$\begin{pmatrix} \mathbf{C}_{11,k} & \cdots & \mathbf{C}_{1M,k} \\ \vdots & \ddots & \vdots \\ \mathbf{C}_{M1,k} & \cdots & \mathbf{C}_{MM,k} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{A}_M \end{pmatrix} \begin{pmatrix} \mathbf{S}_{11,k} & \cdots & \mathbf{S}_{1M,k} \\ \vdots & \ddots & \vdots \\ \mathbf{S}_{M1,k} & \cdots & \mathbf{S}_{MM,k} \end{pmatrix} \begin{pmatrix} \mathbf{A}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{A}_M \end{pmatrix}^T \quad (241)$$

and its inverse equation is

$$\begin{pmatrix} \mathbf{S}_{11,k} & \cdots & \mathbf{S}_{1M,k} \\ \vdots & \ddots & \vdots \\ \mathbf{S}_{M1,k} & \cdots & \mathbf{S}_{MM,k} \end{pmatrix} = \begin{pmatrix} \mathbf{B}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{B}_M \end{pmatrix}^T \begin{pmatrix} \mathbf{C}_{11,k} & \cdots & \mathbf{C}_{1M,k} \\ \vdots & \ddots & \vdots \\ \mathbf{C}_{M1,k} & \cdots & \mathbf{C}_{MM,k} \end{pmatrix} \begin{pmatrix} \mathbf{B}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{B}_M \end{pmatrix}. \quad (242)$$

The JBSS framework is illustrated graphically in fig. 7.9.

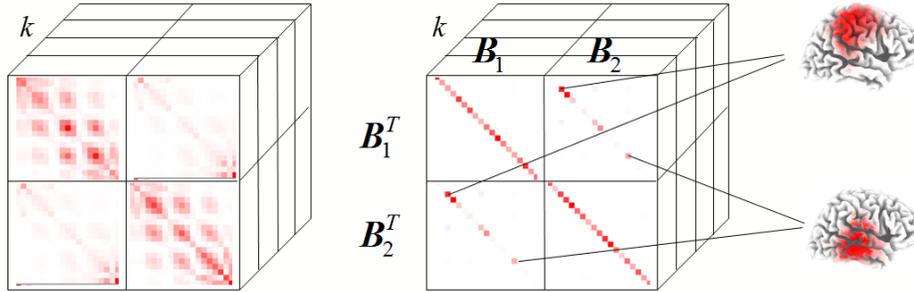


Figure 7.9: Illustration of the JBSS framework for the case $M=2$. On the left are the K slices of observed statistics (241) and on the right the estimated source statistics (242). The diagonal blocks on the right are the source auto-statistics

$$\mathbf{S}_{11,k} = \mathbf{B}_1^T \mathbf{C}_{11,k} \mathbf{B}_1 \text{ and } \mathbf{S}_{22,k} = \mathbf{B}_2^T \mathbf{C}_{22,k} \mathbf{B}_2$$

and on the off-diagonal blocks the source cross-statistics

$$\mathbf{S}_{12,k} = \mathbf{B}_1^T \mathbf{C}_{12,k} \mathbf{B}_2 = \mathbf{S}_{21,k} = \mathbf{B}_2^T \mathbf{C}_{21,k} \mathbf{B}_1,$$

with the last two equalities due to the diagonality of these matrices. All blocks are assumed diagonal, yielding for the whole source covariance structure (242) what we name a strip-diagonal form. This implies that the source statistics within datasets are diagonalized (for $i=j$), as in the BSS framework. In addition, the output cross-statistics between datasets are also diagonalized, thus corresponding sources across datasets may be correlated. The sources in the two data sets are extracted with the same permutation, meaning that the n^{th} source for the first data set correspond to the n^{th} source for the second data set. This is illustrated with putative source localizations shown on the right of the figure; we expect that each source extracted corresponds to a dipole or dipole cluster with corresponding localization in the brain of the M individuals. Nonetheless, the mixing and demixing matrix are allowed to vary across individuals, so that the optimal spatial filter to recover the source can be estimated in each individual, contrary to gBSS.

The Extended AJDC Algorithm

Since $\mathbf{B}_i^T \mathbf{C}_{ij,k} \mathbf{B}_j$ is the transpose of $\mathbf{B}_j^T \mathbf{C}_{ji,k} \mathbf{B}_i$, in order to apply JBSS we do not need to estimate all the KM^2 blocks of (241) from the data; it suffice to estimate the $K[M(M+1)/2]$ blocks on whatever triangular part. Once we have estimated these statistics we may apply the IJLSD algorithm (191). As for the BSS case, we adopt a two-step procedure, which is particularly useful when N is large. It helps also the algorithm because the ensuing AJD matrix to be found is close to orthogonal form, hence well-conditioned (see results in fig. 4.2). The two-step algorithm is:

Algorithm (243): JAJDC (Joint AJD of Cospectral Matrices). Optimization (163).

Given diagonalization set $\mathbf{C}_{ij,k} \in \{\mathbf{C}_{ij,1}, \dots, \mathbf{C}_{ij,K}\}$ indexed by $i, j \in \{1, \dots, M\}$ and $k \in \{1, \dots, K\}$, do

For $m=1$ to M do

$$\mathbf{C}_m^{TOT} = \sum_{i,k} \mathbf{C}_{ii,k} \quad (\text{auto-statistics only}) \quad (244)$$

$$\mathbf{W}_m^T = (\mathbf{C}_m^{TOT})^{-\frac{1}{2}}$$

Partition the whitening matrix such as

$$\mathbf{W}_m^T = (\mathbf{F}_m, \mathbf{N}_m)^T, \quad (245)$$

where $\mathbf{F}_m^T \in \mathfrak{R}^{P \times N}$ holds the first P rows of \mathbf{W}_m^T (signal subspace) and \mathbf{N}_m^T the remaining rows (noise subspace).

End For m

The joint AJD problem on the reduced and whitened data is now solved by IJLSD (191) as

$$\mathbf{E}_1^T, \dots, \mathbf{E}_M^T = \text{JAJD}(\mathbf{F}_i^T \mathbf{C}_{ij,k} \mathbf{F}_j) \quad (246)$$

$$\text{The solutions (demixing matrices) are } \mathbf{B}_m^T = \mathbf{F}_m^T \mathbf{E}_m^T \in \mathfrak{R}^{P \times N} \quad (247)$$

$$\text{and their inverses (mixing matrices) are } \mathbf{A}_m = (\mathbf{E}_m^T)^{-1} (\mathbf{F}_m^T)^+ \in \mathfrak{R}^{N \times P} \quad (248)$$

JBSS Model Order

So far in the JBSS section we have summarized our published contributions. This section contains a recent development, still unpublished. We have moved from the realization that in many real-world situations only a small number of sources can be assumed dependent among data sets. For instance, in an EEG experiment where two or more subjects watch together the same movie we may expect that one or a few sources in the visual cortex are synchronized across individuals as a consequence of the simultaneous processing of the same visual stream. Maybe we may expect some other sources to be synchronized, for example those involved in the analysis of the emotional content of the movie. In all cases we may expect there exist some sources that are not synchronized across individuals. In this case, which is realistic, neither a full BSS nor a full JBSS model is optimal. We hereby introduce *incomplete JBSS models*, a general model under which only partial dependence between the datasets is accounted for, and which has full BSS (no source is correlated among datasets) or full JBSS model (all sources are correlated among data sets) as particular cases. Hence the framework presented here is even more general than the JBSS approach. The incomplete approach follows the same least-squares framework presented in chapter IV. More precisely, given N electrodes and P sources for each data set, we assume that there are $L \leq P \leq N$ correlated sources among data sets. Consider the partition of the functional of interest in (170) and let us write it for the sake of clarity as

$$\psi_{\mathbf{B}_i | \mathbf{B}}^{off} = \psi_{\mathbf{B}_i | \mathbf{B}_i}^{BSS\ off} + \psi_{\mathbf{B}_i | \mathbf{B}_{j \neq i}}^{JBSS\ off}, \quad (249)$$

where

$$\psi_{\mathbf{B}_i | \mathbf{B}_i}^{BSS\ off} = \sum_{i=1}^M \sum_k \left\| \text{off}(\mathbf{q}_{ii,k}) \right\|_F^2 \quad (250)$$

and

$$\psi_{\mathbf{B}_i | \mathbf{B}_{j \neq i}}^{JBSS\ off} = 2 \sum_{i \neq j=1}^M \sum_k \left\| \text{off}(\mathbf{q}_{ij,k}) \right\|_F^2. \quad (251)$$

The JBSS partition (for $i \neq j$) corresponds to the pure JBSS cost function and the BSS partition (for $i=j$) the pure BSS cost function. Typical works in the JBSS literature have considered the “equally mixed” model given by the non-weighted sum of the two terms as in (249). In all previous work so far the same combination of the two terms applies to every source to be extracted. However, *we may optimize whatever linear combination of the two terms*. Furthermore, *the linear combination may be different*

for each source. This is what we do here; given that $L \leq N$ sources may be assumed correlated among data sets, we intend to solve for the first L sources according to the JBSS separation criterion only and for the remaining $P-L$ sources according to the BSS separation criterion only. When $L=P$ we obtain a full JBSS model and when $L=0$ we obtain a full BSS model. Diversifying the separation criteria in the whole search is straightforward in a least-squares framework, wherein the total cost function of the JBSS algorithms can be easily decomposed as the sum of a “pure” BSS part and a “pure” JBSS part as per (249) and moreover the search for each demixing matrix can be carried out independently for each of its vectors. In order to do so, let us partition further the cost function in (249), defining the variance explained by each column vector of the \mathbf{B}_i matrices as taken separately. We have the following

Proposition (252).

We assume that among the P sources to be extracted there are L of them that are dependent between data sets and independent within data sets, and the remaining $P-L$ of them that are independent within data sets, but not dependent between data sets, with $0 \leq L \leq P$. Then, for any \mathbf{B}_i matrix to be estimated, with $i \in \{1, \dots, M\}$, we search separately for each $\mathbf{b}_{n(i)}$ vectors of \mathbf{B}_i , with $n \in \{1, \dots, N\}$. Any incomplete JBSS model can be solved by using the JBSS term only (251) when $n \leq L$ and the BSS term only (250) when $n > L$.

The OJLSD algorithm makes use at each iterations of matrices (182). The IJLSD algorithm makes use of these matrices and also of their sum (189). To comply with our proposition all we need is to define the matrices $\mathbf{R}_{(i)(n)}$ differently, such as

$$\mathbf{R}_{(i)(n)} = \begin{cases} \mathbf{R}_{(i)(n)}^{JBSS} = \sum_k \sum_{i \neq j} \left(\mathbf{C}_{ij,k} \mathbf{b}_{n(j)} \mathbf{b}_{n(j)}^T \mathbf{C}_{ij,k} \right) & \text{if } n \leq L \\ \mathbf{R}_{(i)(n)}^{BSS} = \sum_k \left(\mathbf{C}_{ii,k} \mathbf{b}_{n(j)} \mathbf{b}_{n(j)}^T \mathbf{C}_{ii,k} \right) & \text{if } n > L \end{cases}, \quad (253)$$

After which algorithms OJLSD and IJLSD can be applied with the exact same algorithms (184) and (191) for any $L \in \{0, \dots, N\}$ using (253) to compute matrices $\mathbf{R}_{(i)(n)}$ and from there obtaining their sum $\mathbf{\Gamma}_{(i)}$. We name the incomplete version of the algorithms iOJLSD and iIJLSD, where “i” stands for incomplete.

CHAPTER VIII

RIEMANN GEOMETRY: A UNIVERSAL BCI

CLASSIFICATION FRAMEWORK

Introduction

Over the past 25 years the field of brain-computer interfaces (BCI) has grown considerably and has become the most prominent applied research area for electroencephalography (EEG). Thanks to substantial granting by the EC in Europe and by the NIH and NSF in the USA, among others, recently there has been a striking acceleration of BCI research and applications, both for healthy users and for clinical populations (Allison *et al.*, 2012; Kübler *et al.*, 2001; Tan and Nijholt, 2010; Wolpaw and Wolpaw, 2012). Yet, still today it has to be admitted that “efforts to commercialize research findings have been tiepid, hampered by a general lack of robustness when translating technologies to uncontrolled environments beyond the research laboratory” (Obeid and Picone, 2013). While trying to combine advances from different projects it has become evident that efforts toward the standardization of EEG data format, BCI interfaces and processing tools are of paramount importance (see the roadmap of the coordination “Future BNCF” project, 2012). Since the inception about 20 years ago of inverse solutions (see Chapter III) and of diagonalization methods (see chapter IV to VII) such as the common spatial pattern, canonical correlation analysis, independent component analysis, with the many variants of each and possible combinations, we may say there has been no further major innovation in the signal processing of BCI data; new methods based on these tools effectively bring only moderate improvement and do not in general increase reliability in a significant way.

Each of the three main BCI paradigms, namely, motor-imagery (MI), steady-state evoked potentials (SSEP) and P300 are currently treated with dedicated pre-processing, signal processing and classification tools. The number of papers presenting improvements for each one of these steps is very large and contributes to the fragmentation of the field. Traditionally and still today we can divide existing BCI in two categories: those that follow a “hard machine learning” approach, and those that use “spatial filtering” to increase the signal to noise ratio followed by a simple classification algorithm. The “hard machine-learning” kind generalizes fairly well across sessions and across subjects, but require a substantial amount of training data. Furthermore, it is often computationally intensive. The opposite happens for the “spatial filtering” kind, where bad generalization capabilities are compensated by a fast training and lower computational cost. In light of this situation it has been stated that “the field would benefit from a new paradigm in research development that focuses on robust algorithm development” (Obeid and Picone, 2013).

In this chapter we focus on the standardization of the *core of a BCI*, that is, the processing and classification algorithm. We propose a new paradigm for signal processing and classification in BCI capable of supporting a completely new BCI mode of operation. To proceed in this direction we have

moved from the trends currently followed by the BCI community and by the specification of the characteristics a BCI should possess. Among recent trends in BCI research we find:

1. The conception, analysis and testing of *generic model classifiers*, allowing the so-called *transfer learning*, whereas data from other sessions and/or other subjects is used to increase the performance of low-performance users and to initialize a BCI so as to start using it without calibration (Colwell et al., 2013; Herweg, Kaufmann and Kübler, 2013; Kindermans and Schrauwen, 2013; Jin et al., 2012). In this direction is also relevant the use of unsupervised classifiers (Kindermans, Verstraeten and Schrauwen, 2012).

2. The conception, development and maintenance of world-wide massive *databases* (e.g., Obeid and Picone, 2013). Such a resource is necessary to boost research by allowing massive testing of algorithms. It also enables the systematic study of the source of variation in individual EEG pattern and their relation to BCI capabilities and individual attainable performances. Finally, it yields a smart initialization of a BCI (by specific transfer learning), which is necessary to use effectively a BCI without calibration.

3. The *continuous on-line adaptation of the classifier*, which combines the smart initialization mentioned above, in that the adaptation ensures that optimal performance is achieved regardless how good the initialization is. It also allows keeping optimal performance by adapting to mental and environmental changes during the session, achieving the sought reliability (Barachant and Congedo, submitted; Panicker, Puthusserypady and Sun, 2010; Schettini et al., 2013).

Other current lines of research that should be taken into consideration in designing a new generation of BCIs include:

I) The improvement of the performance by *dynamic stopping*, that is, minimizing the amount of data necessary to send a command (the duration of a BCI trial, e.g., the number of repetitions in the P300 speller) while keeping the same performance (Mainsah et al., 2013; Kindermans and Schrauwen, 2013; Schettini et al, 2013).

II) The improvement of the *BCI interface*, e.g., the introduction of *language models* in BCI spellers (for example, letter and word prediction: Mainsah et al., 2013; Kaufmann et al., 2012; Kindermans and Schrauwen, 2013).

III) The improvement of the *BCI paradigm* itself, e.g., for P300-based BCI, the use of faces for flashing to improve the accuracy (Kaufmann et al., 2012), the use of random or pseudo-random flashing instead of row-column flashing (Congedo et al., 2011; Jin et al., 2011; Townsend et al., 2010), the use of inter-stimulus intervals randomly drawn from an exponential distribution and not constant (Congedo et al., 2011), etc. For SSVEP-based BCI improvements of the paradigm include the use of precise tagging of the flickering so as to use phase information (e.g., Jia et al., 2011), the use of random flickering sequences (code modulation: e.g., Bin et al., 2011), etc.

IV) *Multi-subject BCIs*, that is, BCI systems controlled by several users, in proximity one to the other or remotely connected (Bonnet, Lotte and Lécuyer, 2013; Yuan et al., 2013; Schultze-Kraft et al., 2013). Besides allowing remote social operation, this functioning has potential to achieve perfect accuracy on single trials, by combining the data of several users.

V) *BCI Hybridization*, that is, the combination of several BCI paradigms on the same interface to increase the bit rate, the ergonomy, usability and the accuracy (e.g., Lee and Park, 2013).

A proposition for a BCI realizing characteristics 1-3, while keeping in mind points I-V, is presented schematically in fig. 8.1. See the caption of the figure for a generic description of such a BCI. In order to achieve this BCI functioning, the BCI processing and classification core should possess the following characteristics:

- a) It should be *accurate* in general, as compared to existing approaches.
- b) It should be *reliable*, that is, it should maintain as much constant as possible its functions and accuracy in routine circumstances, as well as in hostile or unexpected circumstances.
- c) It should perform generally well as *initialized with generic parameter*, even for a naïve user, that is, it should possess good generalization abilities both cross-subject and cross-session.
- d) It should *learn fast* the individual characteristics and then maintains optimality, *adapting* fast to the mental state of the user and to environmental changes.
- e) It should be *universal*, that is, applicable to all BCI paradigms (hence to hybrid systems).
- f) It should be *algorithmically simple*, so as to be robust and usable in unsupervised on-line operation.
- g) It should be *computationally efficient*, so as to work on small portable devices in line with the current trend in portability of micro-electronic devices.

In this chapter we present a BCI signal processing and classification framework possessing *all* characteristics a)-g), that is, a candidate to the sought standardization effort for the processing and classification core of a BCI. The framework, based on recent advances in *Riemann geometry*, allows the conception of a simple BCI core that can be applied with minor changes to all the main three BCI paradigms. We show by means of real offline and online data that an effective “traditional” BCI can be obtained with a simple classification algorithm and very little pre-processing, regardless the chosen paradigm. Then we briefly delineate strategies to evolve the framework so as to work as in the concept illustrated in fig. 8.1. These claims are supported with experimental data. The material here presented is today still largely unknown to colleagues involved in EEG analysis and BCI, yet, we predict that Riemann geometry will play shortly a prominent role in BCI classification, forming the core methodology for new BCIs. The theory of Riemann geometry will be presented in next chapter (chapter IX).

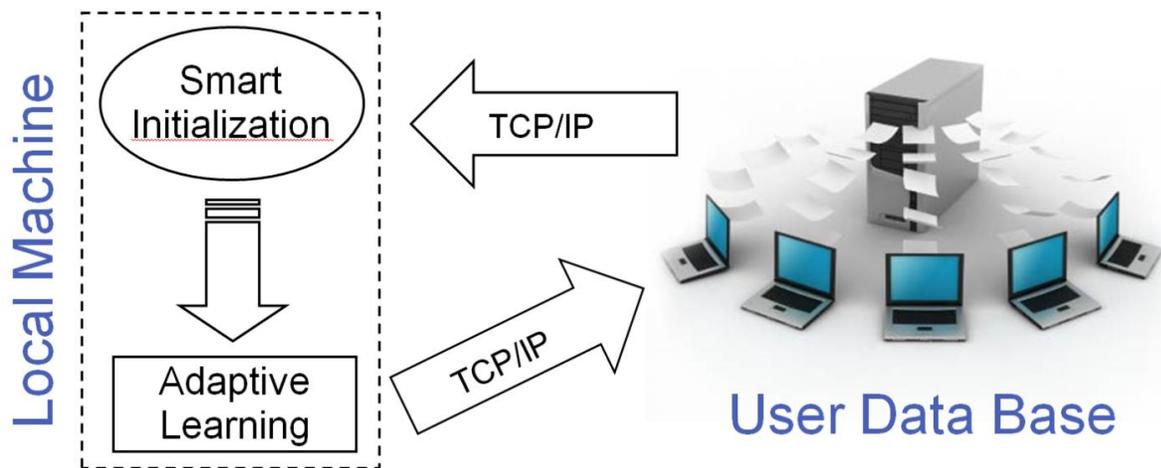


Figure 8.1: Concept for a “new generation” of BCIs. At start-up a BCI queries a database to obtain an initialization, possibly sending minimal EEG data of the user so that the database can elaborate a smart initialization that fits appropriately even a naive user. The BCI is operational straightaway, albeit suboptimal at the very beginning. While being used the BCI adapts to the user and send back the data to the database, along user information, so as to enrich the database and to allow smarter initializations in future sessions of the same user. Multiple subjects may use at the same time the same BCI, in which case the core of the BCI may be located on the server, so to exploit the multitude of data to increase performance.

EEG Data Modeling

Let $\mathbf{x}(t) \in \mathfrak{R}^N$ be the zero-mean EEG data vector for N electrodes at discrete time sample t and let $\mathbf{X}_z \in \mathfrak{R}^{N \times T}$ be a finite time-interval realization (a trial for BCI) comprised of T samples belonging to class $z \in \{1, \dots, Z\}$. The classification method we propose in this section applies to the multi-class situations, that is, it applies in the same way it does not matter how big is Z . Note also that in this chapter most of the time we do not need to index the trial as well, so usually the trial has the class index when it is labeled (training data: \mathbf{X}_z) and no index at all when it is unlabeled (test trial: \mathbf{X}). The data is always assumed having zero mean, therefore the *sample covariance matrix* of a given trial \mathbf{X}_z in a wide-sense stationary (Yeredor, 2000, 2010) belonging to class z as in (95) is given by

$$\mathbf{C}_z = 1/(T-1) (\mathbf{X}_z \mathbf{X}_z^T). \quad (254)$$

Assuming a multivariate Gaussian distribution the *Wishart matrix* $\boldsymbol{\Sigma}_z$ is the unique parameter of the data distribution, assumed unique for each class, such as

$$\mathbf{x}_z(t) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_z), \quad (255)$$

and (254) is a sample covariance matrix estimation of the Wishart matrix.

The Classification Framework

First of all we shall describe a classification algorithm that can be applied *universally* with an appropriate definition of the “covariance matrix”, so as to capture the relevant information of the trials. The relevant information, hence the band-pass filtering and the form of the covariance matrix, depends on the paradigm at hand, however we require the rest of the signal processing chain be the same for all paradigms. Note that in this chapter we employ the term “covariance matrix” referring to a generic structured covariance matrix, depending on the paradigm and of which the sample covariance matrix (254) typically is just a block. We assume data modeling (255) for an extended definition of data, allowing well separated associated Wishart matrices and covariance matrices estimated on single-trials.

In BCI we face the problem of classifying single trials. It does not matter if we deal with motor imagery (MI) trials, steady-state evoked potentials (SSEP) trials, or event-related potentials (ERP)

trials, we have a number of training trials \mathbf{X}_z for each class $z \in \{1, \dots, Z\}$. The classification task consists in assigning an unlabeled trial \mathbf{X} , from which a special form of covariance matrix \mathbf{C} is computed, to one of the Z classes. Using training data we may compute a “mean” for each covariance matrix of the Z classes $\mathbf{M}_1, \dots, \mathbf{M}_Z$ and then simply assign the unlabeled trials to the class which mean is the closest. In order to do so we need an appropriate *metric* to estimate the class *means* and to assess the *distance* between the unlabeled trials and the means $\delta_R(\mathbf{C} \leftrightarrow \mathbf{M}_z)$. This is what we obtain thanks to the Riemann framework. Our universal classification algorithm is summarized such as:

Universal MDM BCI Classifier (256)

- Given a number of training trials \mathbf{X}_z for each class $z \in \{1, \dots, Z\}$ do appropriate preprocessing, estimate an appropriate form of covariance matrix \mathbf{C}_z and estimate their Z class means $\mathbf{M}_1, \dots, \mathbf{M}_Z$.
- For unknown trial \mathbf{X} do the same preprocessing, estimate the same form of covariance matrix \mathbf{C} and assign to class k as per

$$\underset{z}{\operatorname{argmin}} \delta_R(\mathbf{C} \leftrightarrow \mathbf{M}_z)$$

that is, to the class which mean is the closest to the covariance matrix, according to distance δ_R .

This is the simplest classification method one can think of and is known as *minimum distance classifier* (MDM). The classification algorithm is illustrated in fig. 8.2 for the case of a two-class BCI ($Z=2$). It works exactly in the same way for whatever number of classes. As it is well known, defining the mean as the arithmetic mean and the distance as the Euclidean distance gives very poor classification results (see for example Li and Wong, 2013). However, the message we want to convey here is that we do not need to complexify the classification algorithm or to apply sophisticated preprocessing and sophisticated spatial filtering or machine learning. It turns out that an appropriate definition of the covariance matrix, of the mean and of the distances performs as well as most sophisticated methods presented in chapter III, IV, V, VI and VII. To make a metaphor, it appears that we have started a long time ago measuring distances with a biased ruler. Then we have developed complex instruments in order to replace the malfunctioning ruler. Finally, we have found a valid ruler, so that the complex instruments are no more necessary. Providing the valid ruler to measure distances is the main achievement of the Riemann framework. Defining appropriate covariance matrices embedding relevant information depending on the data is our job. We will treat the appropriate

mathematical definition of mean and distance later in chapter IX. Let us now consider in details how the universal MDM framework applies to different BCI paradigm and let us evaluate its performance.

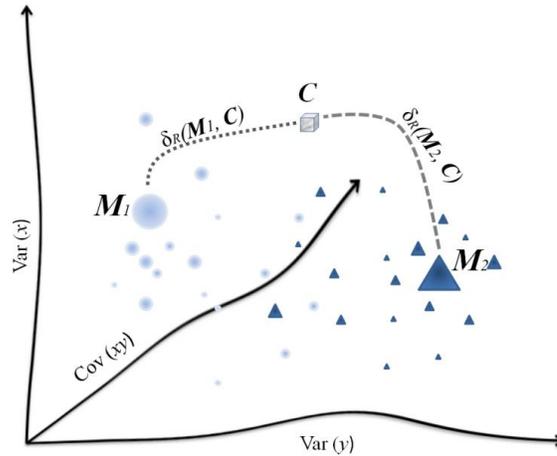


Figure 8.2: *The Minimum Distance to Classifier Algorithm. For covariance matrices of dimension 2×2 , given two geometric means M_1 and M_2 on and unlabeled trial with covariance matrix C , the algorithm assigns the trial to the class given by the closer mean according to an appropriate distance measure δ_R . The distance measure is not linear, as we will see in chapter IX. This is represented by curved lines in the figure.*

Smart Initialization (Cross-Subject and Cross-Session Generalization)

We have said that a BCI processing chain should possess *both* fast adaptation abilities and good generalization across-session and across-subject. This is where the Riemann framework proves advantageous as compared to the state of the art methods, which in general possess one but not the other. The initialization of the classifier state by previous data, either coming from other individuals (cross-subject) or from previous sessions (cross-session) is sometimes named in the literature as *transfer learning*. Cross-subject transfer learning is the only option for a naïve user. From the second use of the same BCI on we can use cross-session transfer learning as well. When an optimal subset of available data is used to specifically initialize the system for a given user we say the system is *smartly initialized*. For instance, how to optimally blend the cross-subject and cross-session initialization after the first session and what part of the database should be used to initialize the classifier for a given user are largely unexplored topics (see for example Schettini et al., 2013).

Adaptation

Given an initialization we want to learn individual classification parameters to achieve optimal performance and adapt to environmental changes, mental state changes and any other intervening condition that may affect the classification performance. We also want to do this as fast as possible. In order to do so we actually set up two parallel classification MDM algorithms, a *generic* one and an *individual* one, the latter being supervised. The classifier output will be given by a weighted sum of the two classifiers, say, with the two weights summing up to 1. The generic classifier will have weight 1 at the beginning of the session and smaller and smaller weight as the session progresses. The individual classifier will have weight zero at the beginning of the session for a naïve user. For a user for which data from previous sessions is available the initial weight can be raised proportionally to the amount of training data. In any case the weight of the individual classifier will rise along the session and will approach 1 by the end of the session. How these two parallel classifiers should evolve over time without supervision is an intriguing and non-trivial research topic.

Classification of Motor Imagery

The form of covariance matrices for motor imagery BCI data

The sample covariance matrix as defined in (254) contains only *spatial information*. The diagonal elements hold the variance of the signal at each electrode and the off-diagonal elements hold the covariance between all electrode pairs. As such it suffices for classifying motor imagery (MI) trials because MI trials for different classes do indeed produce a different scalp pattern, but not necessarily a different frequency pattern or temporal pattern (Pfurtscheller and Lopes da Silva, 1999). We then set

$$\mathbf{X}_z^{MI} = \mathbf{X}_z . \quad (257)$$

In case of MI-based BCIs there are as many classes (with associated training trials) as motor imagery tasks. A no motor imagery class can be added if sought. The only pre-processing step requested is filtering the data in the frequency band pass regions involved in motor imagery (e.g., 8-30 Hz). Then algorithm (256) applies using the regular form of covariance matrix as per (254). Extensive testing has proven that the MDM method is reliable and accurate for motor imagery data (Barachant, Bonnet, Congedo and Jutten 2010a, 2010b; 2011a, b; 2012a, b; Barachant et al., 2012a).

Analysis of motor imagery BCI data

Method

We have applied the MDM method to dataset 2a of BCI Competition IV (2008), provided by the Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology. The data set includes nine subjects involved in a four-class ($Z=4$) two-session motor imagery-based BCI experiment. The four classes were right hand, left hand, feet, tongue. EEG data was acquired by means of 22 electrodes concentrated on and around the sensorymotor areas. The trials were band-pass filtered in the range 8-30 Hz. Two seconds of data in each trial were used for the analysis. We consider both binary classification of each class against the others - this is what the CSP (common spatial pattern, see (200)) does better – and the true multiclass case, where the four classes are treated altogether. The MDM handles equally well and in the same way both the binary and the multiclass case.

We present results of the offline analysis to compare MDM to state of the art competitors. For binary classification we compare the MDM against CSP + LDA (linear discriminant analysis) classification algorithm (Lotte and Guan, 2011)³³. Three pairs of CSP filters were retained. This is the unique parameter to be set with this approach. For multiclass classification we compare the MDM with the BSS approach proposed by Grosse-Wentrup and Buss (2008). Their method consists in the approximate joint diagonalization (AJD) of the four class covariance matrices, selection of the eight best filters using a mutual information criterion and a sparse logistic regression classifier³⁴. For this approach also, the number of filters must be set. On the other hand, the MDM approach is fully automatic. The results concern the *cross-session performance*, that is, the algorithms are trained on one session and tested on the other. This is a more difficult test-bed as compared to cross-validation within the same session.

Results and Conclusions

Results in term of accuracy (percent of correctly classified trials) for all subjects, the two sessions, one against the other, and for all methods are given in table 8.1. The chance level is 25%. We compared

³³ For the CSP + LDA we used the code of Dr. Lotte, available under request by e-mail.

³⁴ For the BSS + LG we have used the code available at <http://people.kyb.tuebingen.mpg.de/moritzgw/MulticlassCSP.zip>.

Notice that this AJD approach is a specific instance of the AJDC framework we have described in (223).

statistically the average performance in term of percent correct classified trial (accuracy) of the MDM and the state of the art competitors. For the multiclass classification the MDM proved marginally superior on the average of the 18 sessions as compared to the BSS + LG method (paired t -test₍₁₇₎=1.9, $p=0.074$, two-tailed). For the average of all the binary classification there was no difference between the MDM and the CSP+LDA method.

In Gouy-Pailler *et al.* (2010) we have developed a BSS method for motor imagery classification exploiting the non-stationarity of the ERD/ERS during the trial. The method proposed by Grosse-Wentrup and Buss (2008) was the starting point of our investigation. We have shown (see chapter VI) that the non-stationarity BSS method performs better as compared to the BSS method of Grosse-Wentrup and Buss (2008). The basic BSS + LG method implemented in Gouy-Pailler *et al.* (2010) was virtually identical to the BSS method of Grosse-Wentrup and Buss (2008). The only difference was the *AJD* algorithm employed; we used the algorithm by Pham (2001b) and they used the algorithm of Ziehe *et al.* (2004). The non-stationarity extension developed in Gouy-Pailler *et al.* (2010) is obtained estimating several covariance matrices in successive time intervals within the trial and diagonalizing all these matrices for all classes simultaneously. We therein also implemented a CSP + LDA “one class vs. all” for comparison. The means (sd) obtained by the three methods implemented in Gouy-Pailler *et al.* (2010) for the cross-session accuracy were, in the order, 63.3 (13.48), 63.8 (12.28) and 60.5 (11.09). These means are directly comparable to the means reported in table 8.1. None of these methods are on the average significantly superior to the MDM multiclass. We conclude that the MDM for MI data performs as well as the most sophisticated method found in the literature, but is fully automatic (no parameter to be set) and respect the requirements of a next generation of BCI.

To take a closer look at the results we plotted the accuracy of the MDM against its competitor (fig. 8.3). It appears then the result we have observed consistently when comparing the Riemann MDM approach with state of the art approaches: the performance of the MDM approach is more or less equivalent for subjects performing well, while it is better for subjects performing poorly (see the position of the dots in the lower-bottom corner of the plots). This behavior springs from the robustness of the Riemann distance (chapter IX).

Table 8.1: Accuracy results for each subject using as training data session 1 and test data session 2 (rows “Session 1->Session 2”) and vice versa (rows “Session 2->Session 1”), for the binary classification of all pairs of classes (numbered 1 to 4), the average of all binary classifications (columns “Ave”) and for the multiclass classification (columns “4-class”). The MDM method applies to both the binary and multiclass classification. As state of the art competitors we used the CSP+LDA for binary classification and the BSS+LG for multiclass classification.

Subject	MDM								CSP + LDA						BSS + LR		
	1 / 2	1 / 3	1 / 4	2 / 3	2 / 4	3 / 4	Ave	4-class	1 / 2	1 / 3	1 / 4	2 / 3	2 / 4	3 / 4	Ave	4-class	
1	93.75	96.53	98.61	97.22	99.31	70.14	92.59	78.82	93.06	98.61	98.61	97.22	100.00	69.44	92.82	76.74	Session 1 -> Session 2
2	63.19	78.47	68.06	74.31	72.22	74.31	71.76	46.88	50.69	68.75	67.36	81.25	63.89	69.44	66.90	43.40	
3	94.44	89.58	86.81	95.14	97.92	66.67	88.43	70.83	96.53	94.44	94.44	93.06	96.53	69.44	90.74	76.04	
4	75.00	88.89	88.19	92.36	84.72	62.50	81.94	61.81	70.14	78.47	86.81	88.89	85.42	56.94	77.78	55.21	
5	63.19	73.61	72.22	72.22	76.39	70.14	71.30	50.00	59.03	63.19	68.75	68.75	65.28	70.83	65.97	35.07	
6	71.53	71.53	64.58	65.28	63.89	72.92	68.29	47.57	68.06	59.03	71.53	63.19	65.97	67.36	65.86	44.44	
7	72.92	91.67	88.19	90.97	85.42	78.47	84.61	66.32	79.86	97.92	95.14	99.31	97.22	81.25	91.78	63.19	
8	96.53	85.42	82.64	90.28	76.39	70.83	83.68	72.57	93.75	87.50	90.97	86.81	91.67	82.64	88.89	69.44	
9	91.67	93.06	97.22	72.22	81.94	90.28	87.73	74.31	92.36	95.14	95.14	84.72	81.94	88.89	89.70	79.17	
1	74.31	94.44	95.14	92.36	98.61	79.86	89.12	71.18	77.78	95.14	95.83	91.67	99.31	80.56	90.05	73.61	Session 2 -> Session 1
2	50.00	76.39	53.47	77.08	74.31	77.78	68.17	50.00	50.00	74.31	59.03	59.72	60.42	77.78	63.54	29.51	
3	88.89	82.64	92.36	85.42	94.44	81.25	87.50	74.65	91.67	86.81	96.53	90.97	93.75	85.42	90.86	78.47	
4	65.28	72.22	78.47	70.83	72.92	68.75	71.41	49.65	66.67	72.92	81.25	79.17	71.53	71.53	73.84	42.01	
5	63.89	65.28	72.22	66.67	72.22	59.03	66.55	37.85	61.11	51.39	70.83	54.17	67.36	53.47	59.72	26.39	
6	61.81	70.14	59.72	68.06	61.11	63.19	64.00	42.71	70.83	76.39	61.11	66.67	65.97	68.06	68.17	38.19	
7	73.61	87.50	85.42	87.50	87.50	77.78	83.22	65.97	66.67	91.67	95.83	98.61	97.22	86.11	89.35	67.71	
8	94.44	79.17	87.50	88.89	87.50	79.86	86.23	71.53	97.92	83.33	96.53	93.06	92.36	86.11	91.55	75.35	
9	81.25	88.19	98.61	83.33	89.58	88.89	88.31	72.92	91.67	93.75	100.00	81.25	92.36	84.03	90.51	74.65	
mean	76.43	82.48	81.64	81.67	82.02	74.04	79.71	61.42	76.54	81.60	84.76	82.14	82.68	74.96	80.45	58.26	
sd	14.06	9.43	13.75	10.75	11.58	8.58	9.44	13.15	16.09	14.33	14.27	14.03	14.67	10.24	12.30	18.80	

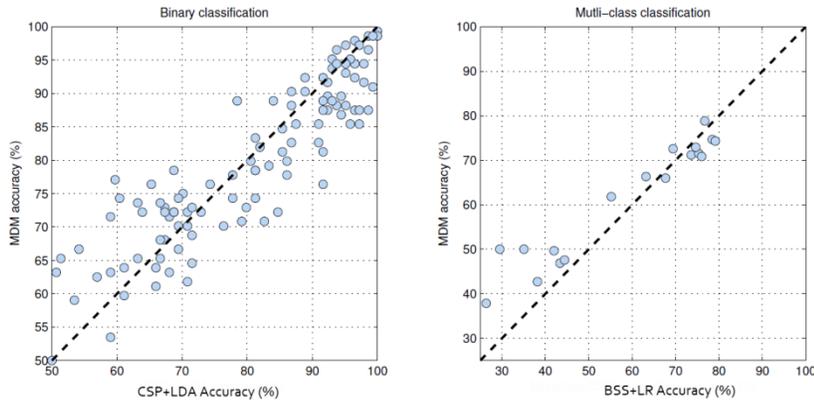


Figure 8.3: Performance in terms of percent correctly classified trials (accuracy), for the binary classification and for the multiclass classification. Each dot represents a subject and a session and has coordinates given by the accuracy of the MDM method against the state of the art competitor.

Classification of Event-Related Potentials

The form of covariance Matrices for of P300 BCI data

For ERP-based BCI the standard covariance matrices (254) are not efficient as the ERP features amplitude much smaller as compared to the background EEG, thus the spatial structure contained in the covariance matrix of a single trial does not hold sufficient information for classification. As a matter of fact the covariance matrix does not contain *any temporal information at all*, which is easily seen if we consider that shuffling at random the samples of trial \mathbf{X}_z the covariance matrix (254) is unchanged. However ERPs have a specific time signature; it is this signature that differentiates an ERP from another or an ERP from the absence of the ERP, so this is the information we need to extract and embed in a “covariance matrix”. In order to do so let us consider again a number of training trials \mathbf{X}_z , for $z \in \{1, \dots, Z\}$ classes. In this case each class corresponds to a different ERP and a no-ERP class is usually added. For example, in P300-based BCI, one class is the *target* class, containing a P300, and the other is the *non-target* class ($Z=2$). Let us now construct the *super-trial*

$$\mathbf{X}_z^{ERP} = \begin{pmatrix} \bar{\mathbf{X}}_{(1)} \\ \dots \\ \bar{\mathbf{X}}_{(Z)} \\ \mathbf{X}_z \end{pmatrix} \in \mathfrak{R}^{N(Z+1) \times T}, \quad (258)$$

where $\bar{\mathbf{X}}_{(1)}^T, \dots, \bar{\mathbf{X}}_{(Z)}^T$ are the grand average ERPs obtained on the training data, previous sessions of the user or even on a database of other users (transfer learning) for each class; we call these grand-average ERP the *temporal prototype*. We specify a prototype for each class. Note that we have introduced index (z) in parenthesis to highlight the difference with the z^{th} training class of the trial. Now, for a training trial \mathbf{X}_z the covariance matrix of the super-trial has the following block structure:

$$\mathbf{C}_z = \frac{1}{(T-1)} \left(\mathbf{X}_z^{ERP} \left(\mathbf{X}_z^{ERP} \right)^T \right) = \frac{1}{(T-1)} \begin{pmatrix} \bar{\mathbf{X}} \cdot \bar{\mathbf{X}}^T & \left(\mathbf{X}_z \bar{\mathbf{X}}^T \right)^T \\ \mathbf{X}_z \bar{\mathbf{X}}^T & \mathbf{X}_z \mathbf{X}_z^T \end{pmatrix} \in \mathfrak{R}^{N(Z+1) \times N(Z+1)}, \quad (259)$$

$$\text{where } \bar{\mathbf{X}} \cdot \bar{\mathbf{X}}^T = \begin{pmatrix} \bar{\mathbf{X}}_{(1)} \bar{\mathbf{X}}_{(1)}^T & \dots & \bar{\mathbf{X}}_{(1)} \bar{\mathbf{X}}_{(Z)}^T \\ \vdots & \ddots & \dots \\ \bar{\mathbf{X}}_{(Z)} \bar{\mathbf{X}}_{(1)}^T & \dots & \bar{\mathbf{X}}_{(Z)} \bar{\mathbf{X}}_{(Z)}^T \end{pmatrix} \in \mathfrak{R}^{NZ \times NZ} \quad (260)$$

$$\text{and } \mathbf{X}_z \bar{\mathbf{X}}^T = \left(\mathbf{X}_z \bar{\mathbf{X}}_{(1)}^T, \dots, \mathbf{X}_z \bar{\mathbf{X}}_{(Z)}^T \right) \in \mathfrak{R}^{N \times NZ}. \quad (261)$$

Let us take a close look to the structure of this covariance matrix:

The $N \times N$ diagonal blocks of $\bar{\mathbf{X}} \cdot \bar{\mathbf{X}}^T$ hold the covariance matrix of the Z temporal prototypes and its $N \times N$ off-diagonal blocks their cross covariance. All these blocks are not useful for classification as they do not change from trial to trial.

The $N \times N$ block $\mathbf{X}_z \mathbf{X}_z^T$ holds the covariance matrix (254) of the trial \mathbf{X}_z , which contains the *spatial information* of the trial and will be little useful for classification, as we have said.

The $N \times N$ blocks of $\mathbf{X}_z \bar{\mathbf{X}}^T$ contains the cross-covariances between the trial and the Z prototypes, that is, these blocks contain the *temporal covariances*. Notice that shuffling the samples of the trial now *does disrupt* these covariances. These blocks contain the relevant information for classification as the cross-covariance will be large only in the blocks where the class of the trial coincides with the class of the prototype. The means of the “super” covariance matrices \mathbf{C}_z constructed as per (259) on training data, which we denote by $\mathbf{M}_1, \dots, \mathbf{M}_z$ for the Z classes, have the same structure as \mathbf{C}_z . With an unlabeled trial \mathbf{X} we then construct the super-trial as per (258), where \mathbf{X} replaces \mathbf{X}_z , and the corresponding covariance matrix \mathbf{C} as per (259). Then, the classification is obtained as before using MDM (256). The only pre-processing required is to filter the data in the frequency band pass region containing the ERPs, typically 1-16 Hz. The exact choice of the band-pass region is not crucial for ERP classification. Extensive testing has shown that the method is reliable and robust, generalizes better than state of the art methods both across-session and across-subject and is prone to fast adaptation (Barachant et al., 2013; Barachant and Congedo, submitted). We report here below new results corroborating this conclusion.

Notice that the temporal prototypes may be defined in any plausible way, that is, they may be given by models, expectations as we do here, guesses, etc. This way to construct covariance matrices embedding both spatial and temporal information is very flexible thus we believe it will be useful in other domains of research. Notice also that as compared to the classification chain proposed for the MI, only the band-pass region and the form of the covariance matrix is different; anything else stays the same. Finally, it is worth mentioning that often we deal only with the presence and absence of an ERP, as it is the case of P300-bases BCIs, where there are only two classes, a *target* (P300 present) and *non-target* (P300 non-present) class. In this case one can use a simplified version of super-trial (258) given by

$$\mathbf{X}_z^{P300} = \begin{pmatrix} \bar{\mathbf{X}}_{(+)} \\ \mathbf{X}_z \end{pmatrix} \in \mathfrak{R}^{2N \times T}, \quad (262)$$

where $\bar{\mathbf{X}}_{(+)}^T$ is the temporal prototype of the P300 (target class) and $z \in \{+, -\}$, with “+” indicating the target class and “-” indicating the non-target class. For a training trial \mathbf{X}_z the covariance matrix of the super-trial has now the simpler block structure:

$$\mathbf{C}_z = \frac{1}{(T-1)} \left[\mathbf{X}_z^{P300} \left(\mathbf{X}_z^{P300} \right)^T \right] = \frac{1}{(T-1)} \begin{pmatrix} \bar{\mathbf{X}}_{(+)} \bar{\mathbf{X}}_{(+)}^T & \bar{\mathbf{X}}_{(+)} \mathbf{X}_z^T \\ \mathbf{X}_z \bar{\mathbf{X}}_{(+)}^T & \mathbf{X}_z \mathbf{X}_z^T \end{pmatrix} \in \mathfrak{R}^{2N \times 2N}. \quad (263)$$

As in (259) the covariance of the prototype $\bar{\mathbf{X}}_{(+)} \bar{\mathbf{X}}_{(+)}^T$ does not change from trial to trial and is useless for classification.

The covariance of the trial $\mathbf{X}_z \mathbf{X}_z^T$ will be little useful for classification, as we have seen.

The temporal covariance between the prototype and the trial $\bar{\mathbf{X}}_{(+)} \mathbf{X}_z^T$ will be large if the trial pertains to a target and small if it does not, so (263) suffices to classify efficiently target and non-target trials, as we will show.

Equation (263) is the super-trial we have been using in Barachant et al. (2013) and Barachant and Congedo (submitted) and we have found it equivalent to the more general form (258) for a two-class P300-based BCI. This is also the super-trial we have used for the results presented here below.

Analysis of P300 BCI data

Method

We present several results issued from an extensive experiment performed in our laboratory in Grenoble. 24 subjects performed one session of the P300-based BCI video-game *Brain Invaders* (Congedo et al., 2011, see chapter I). Seven of these subjects performed seven more sessions, twice a week, for a total of eight sessions. Each session consisted of two runs of the Brain Invaders, one using the typical training-test procedure (non-adaptive mode) and the other without any training using an initialization and an adaptation scheme (adaptive mode), as discussed. The two runs looked exactly identical to the subjects, in that in both cases a training session preceded a test session. However, the

BCI classification was different in the two runs. For the non-adaptive mode the universal MDM classification algorithm (256) was trained in the training run and applied in the test run. For the adaptive mode the classification was initialized using a database for the first session and then with the cumulated data of previous sessions of the user, so that training data was simply discarded for the adaptive mode. The order of the two runs was randomized and the design was double-blinded; at any time neither the subject nor the experimenter could know in what mode the BCI was running. For the MDM algorithm we use as definition of the trial (262) and the corresponding covariance matrices obtained by equation (263).

Data was acquired with a g.USBamp amplifier (g.Tec, Graz, Austria) using 16 active wet electrodes positioned at Fp1, Fp2, Afz, F5, F6, T7, Cz, T8, P7, P3, Pz, P4, P8, O1, Oz, O2, referenced at the right earlobe with a cephalic ground and sampled at 512 Hz. In online operation and for offline analysis EEG data were band-pass filtered in the range 1-16 Hz and downsampled to 128 Hz.

We present both online results and offline results, the latter in order to compare the MDM algorithm with two popular state of the art algorithms: XDawn (Rivet et al., 2009, 2011) and the stepwise linear discriminant analysis (SWLDA, Farwell and Donchin, 1988). For XDawn the two most discriminant spatial filters were retained. EEG data was then spatially filtered, decimated to 32 Hz and vectorized so as to classify the obtained 32x2 features with a regularized linear discriminant analysis (LDA), using an automatic setting of the regularization parameter (Ledoit and Wolf, 2004; Vidaurre et al., 2009). For the SWLDA EEG data were decimated to 32 Hz and vectorized so as to feed the classifier with the obtained 32x16 features.

Results and Conclusions

We present several offline results of the performance pertaining to the non-adaptive mode, including the classic training-test setting and the cross-subject and cross-session initialization comparing several classifiers. We also present the online results obtained in the adaptive and non-adaptive mode. These latter results are the most relevant as they report the actual performance achieved by the universal MDM algorithm in real operation. All performance results for this experiment are reported in terms of AUC (area under the curve). Before we detail the performance results, let us visualize the structure of the covariance matrix (263) for one subject. This illustrates well the rationale behind the choice of this form of the covariance matrix for ERP-based BCI. See Fig. 8.4 and its caption for details.

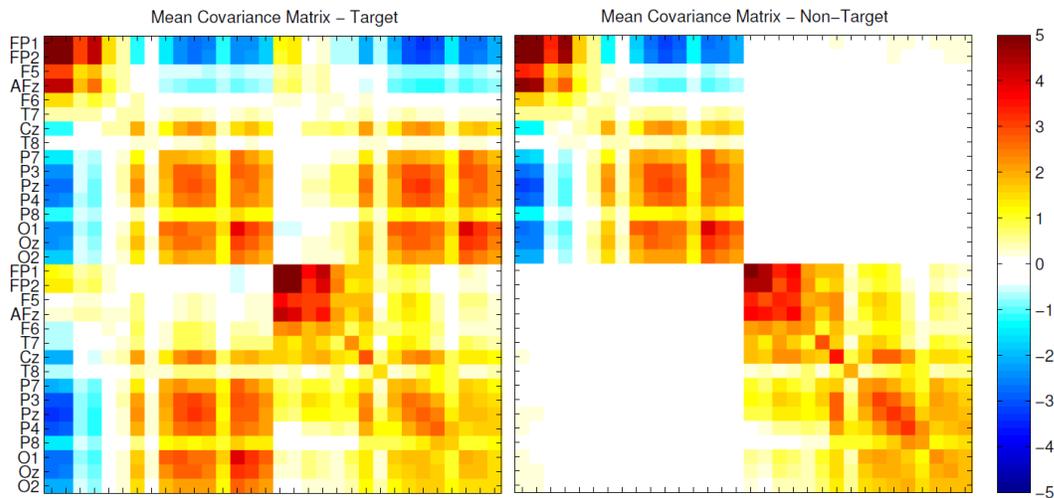


Figure 8.4: Covariance matrix with form (263) computed on one subject using the Brain Invaders P300-based BCI. The matrix is divided in four 16x16 blocks. The upper-left diagonal block is the grand-average sample covariance matrix of the target prototype computed on the other six subjects (cross-subject initialization). This block is the same on the left (target) and right (non-target) part of the figure. The lower-right diagonal block is the sample covariance matrix of the average ERPs obtained on the subject with 106 flashes for target (left) and 530 for non-target (right). The off-diagonal blocks are the temporal covariances between the prototype and the average ERPs; this covariance is high only for the target class since only in this case the signal produced by target flashes correlates with the prototype. All covariance matrices are computed on the ERPs recorded 1s after the flash. The diagonal blocks are scaled so as to make the plot readable.

Offline results: the “classic” training-test mode.

Fig. 8.5 shows the grand average (7 subjects x 8 sessions) AUC accuracy criterion for the three classification methods, obtained training the classifiers on the training run and testing on the test run (“Classic” column). Table 8.2 reports the detailed results for each subject and session. Paired t-tests revealed that the mean AUC obtained by the MDM is significantly superior to the mean AUC obtained by the SWLDA method ($t_{(55)}= 3.377$, $p=0.001$), and equivalent to the mean AUC obtained by XDawn.

Offline results: the cross-subject initialization.

These results are obtained using a leave-one-out method. Fig. 8.5 shows the grand average (7 subjects x 8 sessions) AUC accuracy criterion for the three classification methods obtained training the classifiers on the test data of all subjects excluding the one on which the performance are computed

(“Cross-subject” column). Table 8.2 reports the detailed cross-subject results for each subject and session. As compared to the classic mode the average AUC with cross-subject transfer learning is significantly lower for all classification methods ($p < 0.002$ for all of them). This is an expected result as no information at all about the subject actually using the BCI is provided to the classifiers. Paired t-tests comparing the average performance of the three classification methods in the cross-subject mode reveal that the average AUC obtained by the MDM is marginally superior to the average AUC obtained by the SWLDA ($t_{(55)} = 1.676, p = 0.099$) and by XDAWN ($t_{(55)} = 1.755, p = 0.085$).

Offline results: the cross-session initialization.

These results are also shown in Fig. 8.5 (“Cross-session” column). The mean AUC is obtained initializing the classifier with any possible combination of S number of sessions among the eight available sessions and testing on the remaining $8-S$ sessions. The results are given for S in the range $1, \dots, 7$ and correspond to the average of all subjects and all combinations (which number depends on S). The MDM algorithm proves superior both in the rapidity of learning from previous subject’s data and in the performance attained for all values of S . Note that XDAWN, which is a spatial filter approach, performs fairly well even when only one session is available for training, but its performance grows slowly as more data is available for training. This is because the spatial filter is influenced negatively by the difference in electrode placements across sessions and, in general, by all factors that may change from one session to the other. On the other hand the SWLDA classifier performs poorly when only one session is available for training, however it learn fast as the number of available sessions increase. This is because the SWLDA, being an “hard machine learning” approach, tends to perform well only when a lot of training data is available. So, XDAWN possesses fast learning capabilities, but lacks good transfer learning, whilst the opposite holds for SWLDA. The MDM algorithm possesses both desirable properties.

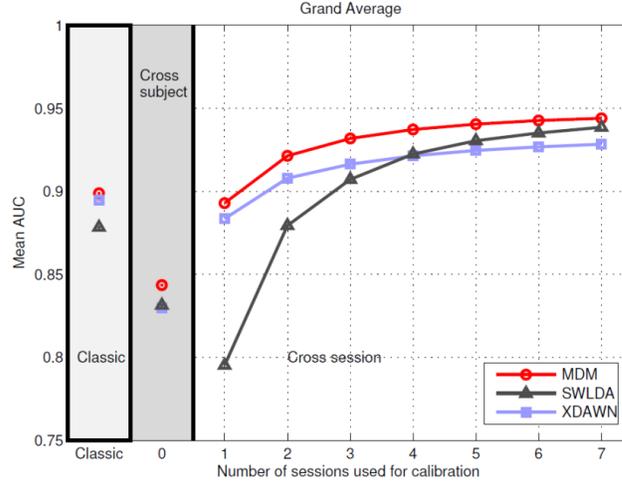


Figure 8.5: Classic (training-test), cross-subject and cross-session offline AUC performance for the P300-based Brain Invaders BCI experiment. Results are the grand average of 7 subjects playing 8 sessions of the Brain Invaders. See text for details.

Table 8.2: Classic (training-test) and cross-subject AUC performance for the P300-based BCI experiment. AUC performance is given separately for each one of the seven subjects (SS) and of the eight sessions (Sess) of the Brain Invaders. See text for details.

Sess	SS	Classic			Cross-Subject			SS	Classic			Cross-Subject		
		MDM	SWLDA	XDAWN	MDM	SWLDA	XDAWN		MDM	SWLDA	XDAWN	MDM	SWLDA	XDAWN
1	1	0.96	0.93	0.94	0.94	0.94	0.94	5	0.98	0.96	0.99	0.85	0.88	0.89
2	1	0.91	0.88	0.91	0.89	0.90	0.94	5	0.97	0.96	0.95	0.90	0.92	0.89
3	1	0.79	0.74	0.75	0.86	0.90	0.88	5	0.87	0.85	0.84	0.85	0.87	0.85
4	1	0.90	0.94	0.89	0.89	0.75	0.69	5	0.94	0.94	0.97	0.88	0.91	0.88
5	1	0.94	0.85	0.91	0.91	0.95	0.94	5	0.73	0.83	0.84	0.83	0.93	0.88
6	1	0.96	0.93	0.95	0.95	0.98	0.98	5	0.91	0.83	0.77	0.88	0.89	0.87
7	1	0.90	0.87	0.89	0.87	0.95	0.94	5	0.87	0.84	0.87	0.88	0.91	0.88
8	1	0.90	0.96	0.90	0.97	0.98	0.97	5	0.92	0.89	0.87	0.81	0.83	0.84
1	2	0.87	0.89	0.94	0.77	0.82	0.85	6	0.99	0.93	0.96	0.89	0.86	0.92
2	2	0.85	0.74	0.80	0.79	0.85	0.83	6	0.87	0.83	0.85	0.91	0.93	0.91
3	2	0.87	0.84	0.87	0.75	0.69	0.69	6	0.80	0.91	0.95	0.97	0.94	0.95
4	2	0.94	0.91	0.93	0.85	0.84	0.85	6	0.96	0.91	0.96	0.92	0.94	0.97
5	2	0.99	0.98	0.99	0.74	0.80	0.78	6	0.95	0.91	0.92	0.82	0.87	0.91
6	2	0.86	0.79	0.84	0.77	0.76	0.74	6	0.85	0.76	0.83	0.91	0.92	0.92
7	2	0.86	0.90	0.88	0.77	0.82	0.80	6	0.94	0.96	0.97	0.77	0.70	0.76
8	2	0.93	0.95	0.98	0.68	0.68	0.75	6	0.99	0.97	0.96	0.89	0.85	0.89
1	3	0.85	0.78	0.89	0.68	0.60	0.59	7	0.84	0.74	0.77	0.85	0.86	0.86
2	3	0.90	0.85	0.88	0.80	0.73	0.72	7	0.96	0.97	0.97	0.87	0.78	0.83
3	3	0.92	0.91	0.94	0.80	0.70	0.66	7	0.88	0.90	0.94	0.83	0.80	0.76
4	3	0.97	0.96	0.93	0.87	0.82	0.79	7	0.83	0.78	0.75	0.77	0.80	0.80
5	3	0.82	0.73	0.86	0.79	0.65	0.68	7	0.86	0.88	0.75	0.88	0.84	0.80
6	3	0.89	0.92	0.96	0.82	0.69	0.71	7	0.91	0.91	0.87	0.82	0.76	0.82
7	3	0.99	0.97	0.99	0.88	0.77	0.77	7	0.98	0.97	0.98	0.79	0.78	0.76
8	3	0.94	0.83	0.86	0.77	0.73	0.75	7	0.73	0.75	0.78	0.82	0.80	0.84
1	4	0.87	0.83	0.87	0.94	0.88	0.83	M	0.90	0.88	0.89	0.84	0.83	0.83
2	4	0.95	0.96	0.94	0.86	0.85	0.85	SD	0.06	0.07	0.07	0.06	0.09	0.09
3	4	0.94	0.94	0.90	0.87	0.86	0.82							
4	4	0.82	0.79	0.84	0.86	0.81	0.78							
5	4	0.94	0.91	0.99	0.83	0.84	0.81							
6	4	0.90	0.88	0.92	0.86	0.84	0.82							
7	4	0.83	0.78	0.81	0.79	0.79	0.82							
8	4	0.85	0.85	0.85	0.80	0.80	0.81							

Online results: adaptation.

Finally, we show the actual online results for the adaptive and non-adaptive mode of functioning. Let us remind that the adaptive and non-adaptive runs were performed in a double-blinded fashion and randomized order. At the beginning of each of its 12 levels the game Brain Invaders shows to the subject a target alien, chosen randomly among 36 aliens. After each repetition of random flashing of each alien, in such a way that each alien is flashed two times (Congedo *et al.*, 2011), the classification algorithm destroys the alien with the highest probability of being the target based on the MDM output. If the destroyed alien is the target the subject wins the level and goes to the next level, otherwise another repetition of flashes is done. Starting from the second repetition the MDM used the cumulated distance of all repetitions to select the alien with the highest probability. Hence, the number of repetitions to destroy the target (NRD) is a direct measure of performance: the lower the NRD the higher the performance.

Figure 8.6 shows the mean and standard deviation NRD as a function of levels for the first session performed by all 24 subjects. As we can see, the non-adaptive MDM features a non-significant negative slope ($p=0.142$), meaning constant performance across levels, whereas the adaptive MDM features a significantly negative slope ($p=0.02$), meaning that the performance increases as the algorithm learns from the data of the subject. On the other hand, the slope of difference of the means between adaptive and non-adaptive mode is not significant. This result shows that the adaptation is effective in leading the user toward good performances already at the first session.

Figure 8.7 shows the histogram and percent cumulative distribution of the NRD for all 24 subjects and all 12 levels of the Brain Invaders game. The cumulative distribution at the third level is 94.44% for the non-adaptive mode and 95.49% for the adaptive mode, that is to say, on the average of all levels and subjects about 95% of the times three or less repetitions suffice to destroy the target. These results demonstrate that our adaptive system *without calibration* yields performances equivalent to the traditional system with calibration, already at the first session.

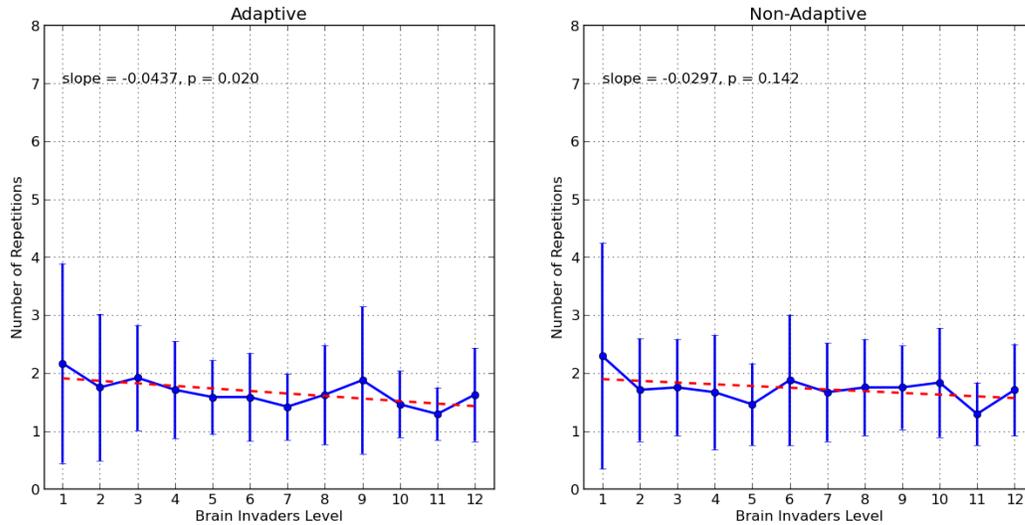


Figure 8.6: Mean (disks) and standard deviation (bars) number of repetitions necessary for destroying the target (NRD) for the 24 subjects across the 12 levels of the first session of Brain Invaders, for the adaptive run (left) and the non-adaptive run (right). On top of the plots is printed the slope of the means and its p -value for the two-tailed test of the slope being significantly different from zero.

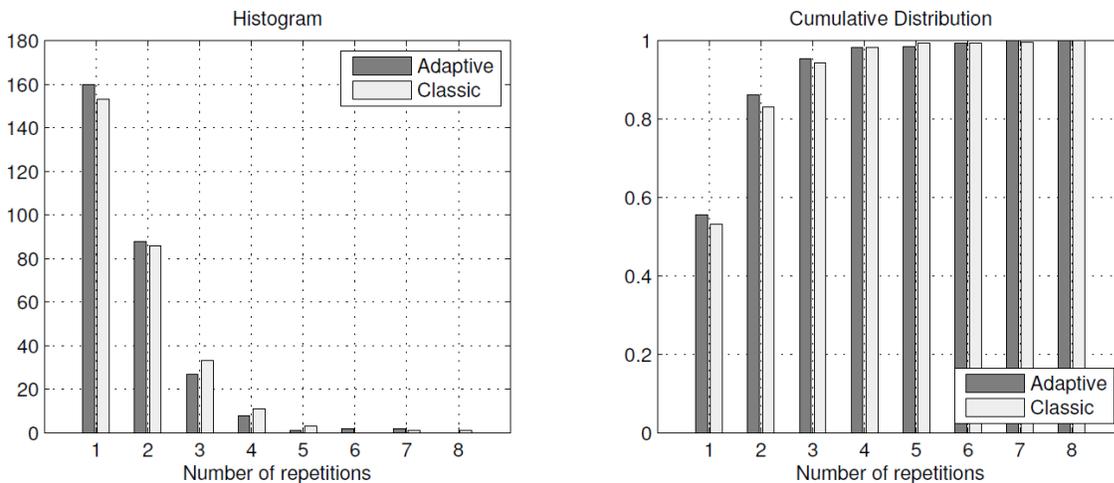


Figure 8.7: Raw histogram (left) and percent cumulative distribution (right) of the number of repetitions necessary to destroy the target (NRD) for all 24 subjects and all 12 levels of the first session of the Brain Invaders game.

Figure 8.8 shows the means and standard deviations of the NRD for the 7 subjects across the 8 sessions of the Brain Invaders, for the adaptive runs and the non-adaptive runs. Neither slope is

significantly different from zero, however the slope of the difference of the means between adaptive and non-adaptive mode is significantly smaller than zero ($slope=-0.0304$; $p=0.047$, one-tailed), demonstrating that over session the performances in the adaptive mode becomes better as compared to the non-adaptive mode. We can also appreciate the smaller standard deviation of the NRD in the adaptive mode, in all sessions. This result is striking since in non-adaptive mode the system is calibrated with data recorded just before the test.

Figure 8.9a and 8.9b shows the histogram and percent cumulative distribution of the NRD for all 12 levels of the Brain Invaders game for the seven subjects performing eight sessions. These figures show the data separately for each session. We see that while the distribution for the non-adaptive classification algorithm is constant across sessions, for the adaptive classification algorithm the performance increases sharply starting at the third session.

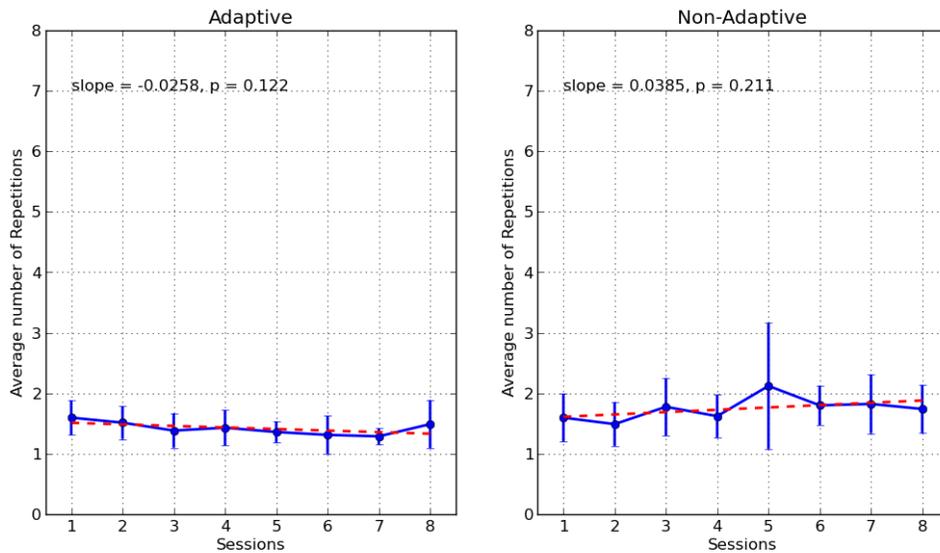


Figure 8.8: Means (disks) and standard deviations (bars) of the number of repetitions necessary for destroying the target (NRD) for the 7 subjects across the 8 sessions of the Brain Invaders, for the adaptive runs (left) and the non-adaptive runs (right). On top of the plots is printed the slope of the means and its p-value for the two-tailed test of the slope being significantly different from zero.

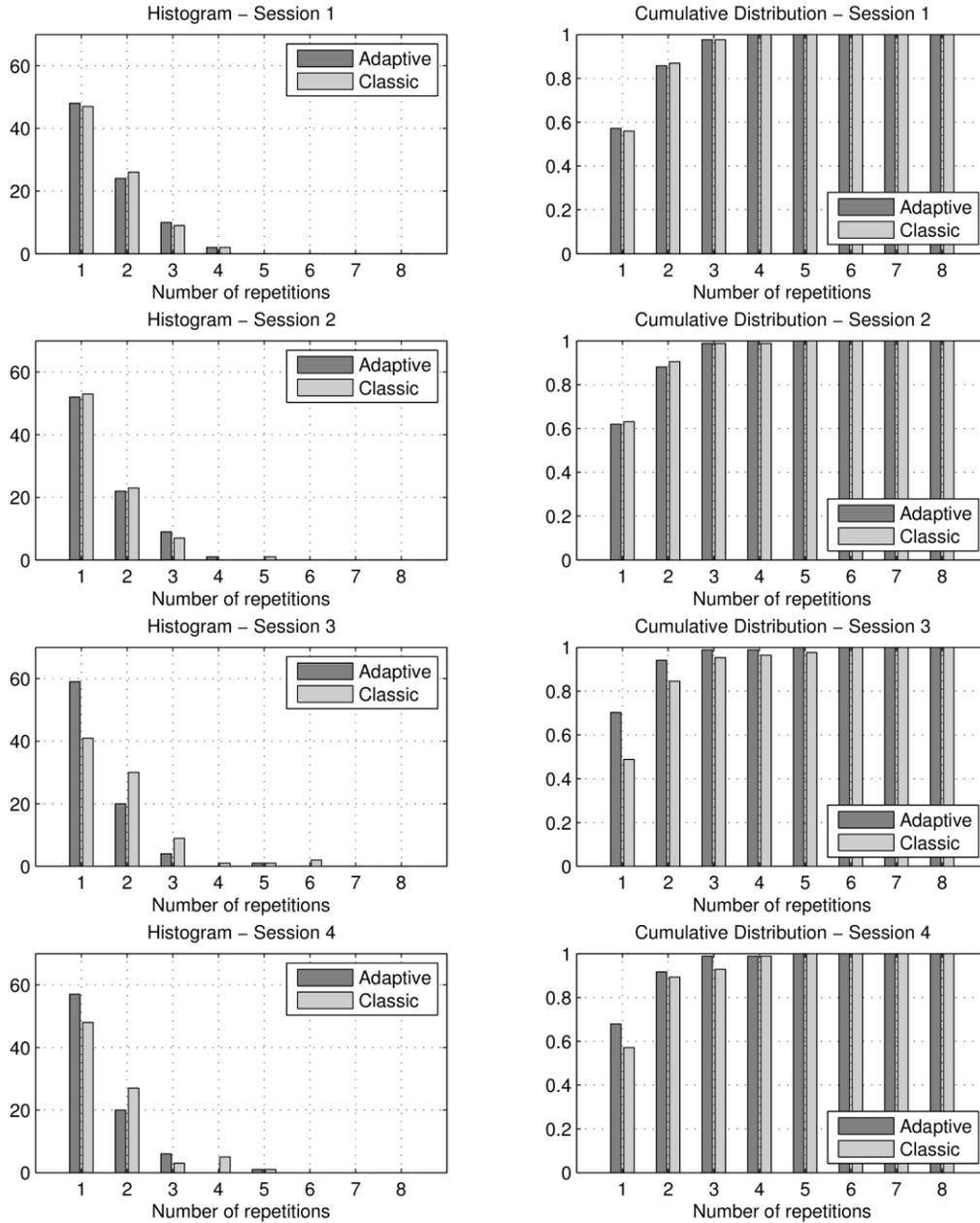


Figure 8.9a: Raw histogram (left) and percent cumulative distribution (right) of the number of repetitions necessary to destroy the target (NRD) for all 12 levels of the Brain Invaders game and for the seven subjects performing eight sessions. Each row is the data of one session. Session 1 to 4.

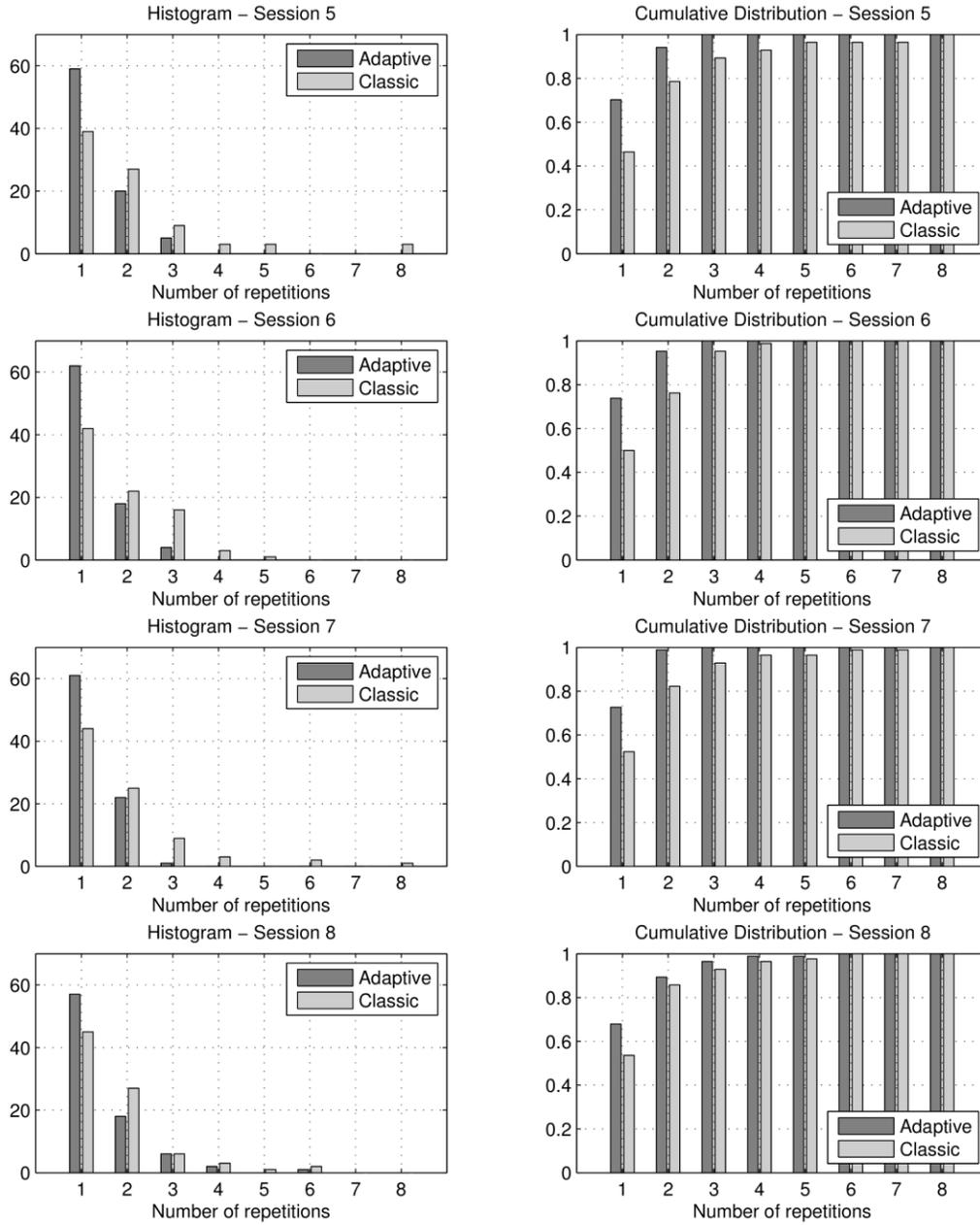


Figure 8.9b: as in figure 8.9a, but for session 5 to 8.

Classification of Steady-State Evoked Potentials

The form of covariance Matrices for Steady-State Evoked Potentials

The same MDM method can be used for SSEP data classification as well. We make here the example of steady-state visually evoked potentials (SSVEP). The Z classes here represent F different flickering frequencies and a no-flickering class can be added as well if sought. In this case the relevant information is the diversity of the frequencies engendering oscillations in the visual cortex, while the spatial pattern may be the same for different frequencies. In order to exploit the frequency diversity we construct super trial

$$\mathbf{X}_z^{SSEP} = \begin{pmatrix} \mathbf{X}_{(1)} \\ \dots \\ \mathbf{X}_{(F)} \end{pmatrix} \in \mathfrak{R}^{NF \times T} \quad (264)$$

where $\mathbf{X}_{(f)}^T$ is the trial filtered in the band-pass region for flickering frequency $f \in \{1, \dots, F\}$. More simply, one may use the Fourier cospectra (85) for the exact flickering frequencies. The covariance matrix of super-trial (264) has the following block structure:

$$\mathbf{C}_z = \frac{1}{(T-1)} \left[\mathbf{X}_z^{SSEP} \left(\mathbf{X}_z^{SSEP} \right)^T \right] = \frac{1}{(T-1)} \begin{pmatrix} \mathbf{X}_{(1)} \mathbf{X}_{(1)}^T & \dots & \mathbf{X}_{(1)} \mathbf{X}_{(F)}^T \\ \vdots & \ddots & \vdots \\ \mathbf{X}_{(F)} \mathbf{X}_{(1)}^T & \dots & \mathbf{X}_{(F)} \mathbf{X}_{(F)}^T \end{pmatrix} \in \mathfrak{R}^{NF \times NF}. \quad (265)$$

The $N \times N$ diagonal blocks holds the covariance matrices of the F frequencies. When comparing an unlabeled trial with the mean of the different classes, only the mean with the block indexing the frequency corresponding to the frequency of the trial will have large values. Thus the diagonal blocks will be useful for classification. On the other hand the off-diagonal blocks hold the cross-covariance between frequencies, thus are not very meaningful. We can put them to zero since the resulting matrix

$$\mathbf{C}_z = \frac{1}{(T-1)} \begin{pmatrix} \mathbf{X}_{(1)} \mathbf{X}_{(1)}^T & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{X}_{(F)} \mathbf{X}_{(F)}^T \end{pmatrix} \in \mathfrak{R}^{NF \times NF} \quad (266)$$

is still symmetric positive definite. Given training data we estimate the class means $\mathbf{M}_1, \dots, \mathbf{M}_z$. For an unlabeled trail \mathbf{X} we compute the super trial with (264), where \mathbf{X} replaces \mathbf{X}_z , then covariance matrix

C using (266). Finally, we use again algorithm (256) to assign the unlabeled trial. The only pre-processing required is to filter the data in the frequencies corresponding to the SSVEP flickering frequencies or, equivalently, estimating the Fourier cospectra at the F flickering frequencies. Note that if the phase of the SSVEP is known thanks to precise data tagging, as it is done in Jia et al. (2011), or code modulation is used (Bin et al., 2011) one can exploit both the frequential and the temporal information, constructing a super trial mixing the strategy used here for ERP (258) and for SSVEP (264).

Analysis of Steady-State Visually Evoked Potential BCI data

Method

Just as an example on the use of MDM for classifying SSEP data, we have applied the MDM algorithm using form of covariance matrix given by (266) to a steady-state visually evoked potential (SSVEP) dataset distributed with the OpenViBE software (Renard et al., 2010). The dataset is from one subject performing 32 SSVEP trials lasting six seconds. There were four classes ($Z=4$); no SSVEP (rest), 12, 15 and 20 Hz. Data was acquired by a g.tec amplifier at 512 Hz with six electrodes (CPz, O1, Oz, O2, POz, Iz). Preprocessing consisted in a 5th order Butterworth 2-Hz large band-pass filter centered at the three flickering frequencies.

Results and Conclusions

Let us visualize the structure of the covariance matrix (266) for one subject. This illustrates well the rationale behind the choice of this form of the covariance matrix for SSEP-based BCI. See fig. 8.10 and its caption for details.

We applied the MDM using a 8-fold cross-validation procedure and using as data segment duration 1s, 2s, 3s, 4s, 5s and 6s. Accuracy results in term of average percent correctly classified trials are shown in table 8.3.

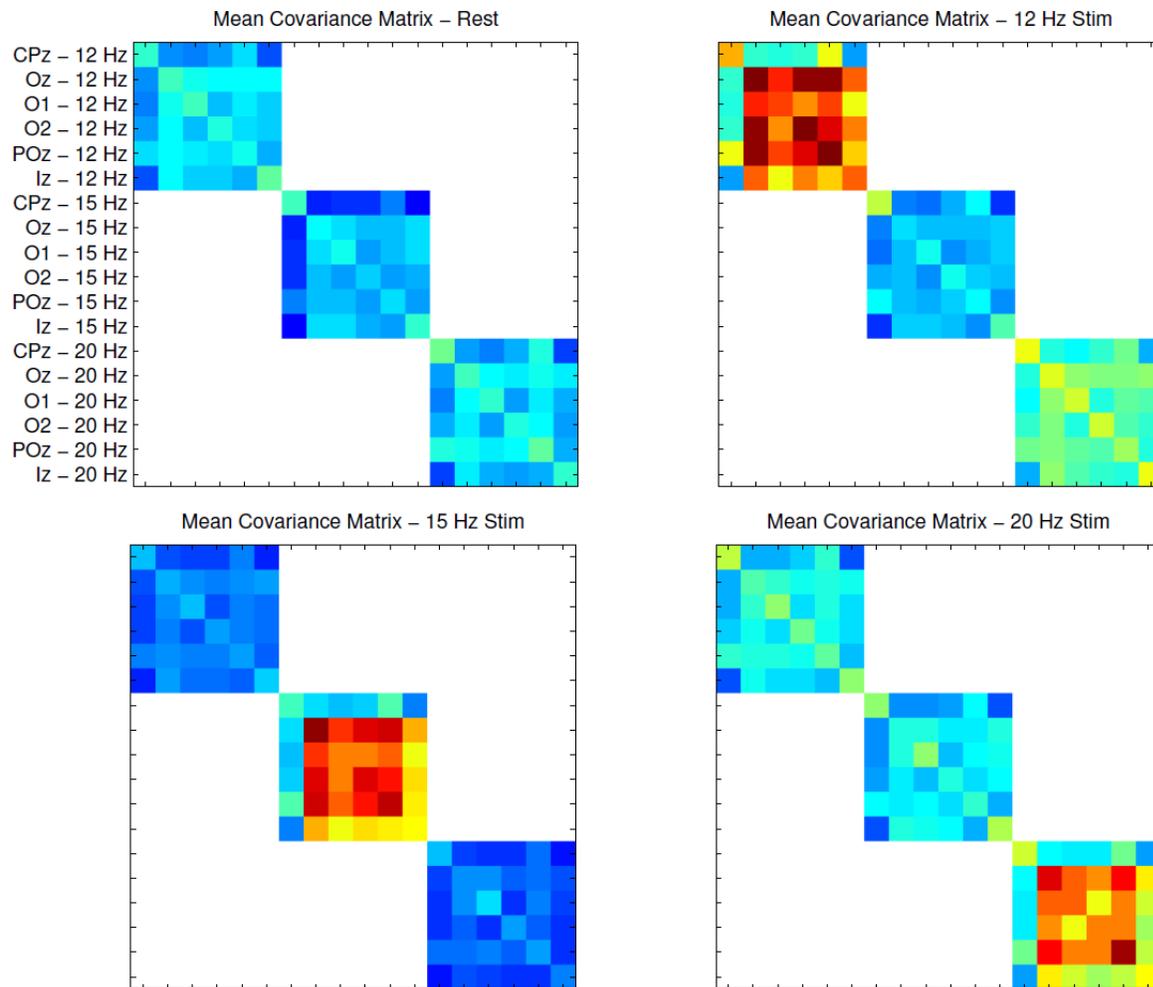


Figure 8.10: Covariance matrix of form (266) computed on the subject performing the SSVEP experiment. Data were acquired at six electrodes. Each diagonal block of the matrices is the 6x6 covariance matrix of the data sharply band-pass filtered around the three flickering frequencies used for stimulation: 12, 15 and 20 Hz. The four matrices represented in the figure are the grand-average obtained for the “rest” class (no flickering) and for the trials with the three flickering frequencies. Notice that for each class only the block corresponding to the actual flickering frequency has high values. For the no-SSVEP (rest) data none of the blocks features high values; this is sufficient to classify well trials belonging to any of the four classes.

Table 8.3: Eight-fold cross-validation offline accuracy performance for the SSVEP data using a window size ranging from one to six seconds. There were four classes, so the chance level for classification is 25%.

Windows Size (s)	1	2	3	4	5	6
Accuracy (%)	53.125	75.000	87.500	93.750	100.000	100.000

Conclusion and Discussion

Based on the presented results we conclude that the MDM classification algorithm do indeed possess fast learning capabilities and that it is apt to exploit transfer learning. The adaptive classification scheme we have tested has proved effective; once the subject arrives at the twelfth level of the Brain Invaders less than two repetitions on the average suffice to destroy the target (fig. 8.6), which is a very good result as compared to the state of the art.

It does not matter how the covariance matrices are defined, the classification algorithms we have proposed using the Riemannian framework remains the same for all the three BCI paradigms we have considered. Furthermore, it remains astonishing simple. Note that at no point there is a parameter to be tuned; it is all deterministic and completely parameter-free. This is in contrast with sophisticated machine learning approach such as SVM, where one or more parameter must be learned, for example, by cross-validation. For this reason we claim that the strategy we have delineated is truly universal. In fact, taken together the simplicity of the MDM classification, its ability to learn rapidly (with little training data) and its good across-subject and across-session generalization, make of this strategy a very good candidate for building a new generation of BCIs. Such BCIs will be smartly initialized using remote massive databases and will adapt to the user fast and effectively in the first minute of use. They will be reliable, robust and will maintain good performances. Having analyzed and tried several among the strategies that can be found in the literature, we believe that the Riemannian framework is the ideal candidate, in that it is the only one possessing all necessary properties a)-g) we have listed in the introduction.

In Barachant et al. (2012a) we have shown that motor imagery classification can be improved significantly over the results shown in table 8.1 in a Riemannian framework mapping the covariance matrices in the tangent space and applying a feature selection + LDA in the tangent space (see chapter IX). In Barachant et al. (2012b) a support vector machine embedded with a Riemann kernel was used. These two methods outperform the state of the art but they require tuning parameters. In Barachant et al. (2010b) we have mapped the covariance matrices in the tangent space, applied a supervised projection of the points (regularized LDA with automatic regularization) in order to increase the class separability and then remapped the data in the Riemann manifold where the MDM applies. This method does not require tuning parameters, but it still is more involving as compared to the simple MDM. Thus, using more sophisticated classification methods in the Riemannian framework one may find a way to outperform the state of the art, but only at the expenses of the ergonomic requirements of the BCI. Similar results apply to other BCI modalities. In our view the simple MDM method is a good

trade-off between accuracy, robustness and ergonomics, therefore could be considered as a starting point for the sake of a new standard suiting a large spectrum of BCI applications. Further research will find the optimal trade-off between sophistication of classification methods based on Riemann geometry and the effectiveness/usability of the method in actual online operation. The method candidating to become a standard for BCI data should work without tuning parameters as MDM does and should keep the fast learning and good transfer learning capabilities.

We will see in next chapter that, given N sensors, classifying sensor covariance matrices in the Riemannian framework is equivalent to classifying in the optimal N -dimensional source space, establishing a connection between the MDM method and spatial filtering/blind source separation approaches. We obtain this result without having to estimate the sources, which is cumbersome, prone to errors and does not generalize well across subjects. This is a very strong result, making the Riemannian framework simple as the most simple classification methods based on sensor data (e.g., Laplacian log-power in motor imagery), but as powerful as the most sophisticated spatial filtering/blind source separation. It is in this property that resides the flexibility and usability of the Riemann framework.

CHAPTER IX

RIEMANN GEOMETRY: A THEORETICAL PRIME

Introduction

The Riemann framework establishes appropriate tools to manipulate the covariance matrices we have defined in chapter VIII for motor imagery, ERP-based and SSEP-based BCI. Covariance matrices are symmetric positive definite (SPD). The study of operators for SPD matrices such as distance and geometric mean has recently grown very fast, driven by practical problems in radar data processing, image processing, medical imaging (especially diffusion MRI), elasticity, mechanics and machine learning. Interestingly, in this endeavor disparate perspectives from matrix analysis, operator theory, differential geometry, probability and numerical analysis have converged to the same results. However, Riemann geometry as a tool in EEG analysis has appeared only very recently, essentially with the parallel work of a group in Canada focusing on sleep stage classification (Li, Wong and Debruin, 2009, 2012; Li and Wong, 2013) and our own group focusing on BCI (Barachant et al., 2010a,b; 2011a,b; 2012a,b; 2013; Barachant and Congedo, 2013, 2014; Congedo, Barachant and Andreev, submitted) and automatic artifact detection (Barachant, Andreev and Congedo, 2013). In this chapter we provide a clear account of useful knowledge accumulated in this field. We skip proof and derivations, referring the reader to the appropriate references. Instead we focus on concepts and useful results. We also present some results on preliminary theoretical investigations we have started recently.

Covariance matrices are symmetric positive definite (SPD) matrices. SPD matrices of dimension N live in a $\frac{1}{2}N(N+1)$ -dimensional hyper cone. Their topology in the Euclidian space is shown in fig. 9.1. Up until recently in EEG analysis we have treated SPD matrices in the normal vector (Euclidian) space of general symmetric matrices created by the metric

$$tr(\mathbf{Q}_1^T \mathbf{Q}_2) \tag{267}$$

and associated Frobenius norm $\|\mathbf{Q}_1\|_F$, however, the native space of SPD is not the Euclidean vector space. We introduce here the use of an exponential map for symmetric matrices, which is always SPD and induces on the space of SPD matrices an *affine-invariant metric* in a *Riemann manifold*. Such operation replaces the convex pointed cone in the vector space of fig. 9.1 with a regular manifold of constant curvature without boundaries, developing instead infinitely in all of its $\frac{1}{2}N(N+1)$ dimensions. This is a curved space, but appropriate operations allow treating it as a vector space, ensuring to remain in the space of SPD matrices. While at first the Riemann framework appears unusual and odd, we provide here a comprehensible account. For our exposition we follow the account given by Bathia (2013), Moakher (2005) and Pennec et al. (2004) integrating the exposition with other articles and

books. It is interesting to notice that we can arrive at exactly the same mathematical formalism and we end up with the very same Riemann metric from a pure *differential geometry*³⁵ point of view (Bhatia, 2003, 2013; Bhatia and Holbrook, 2006; Fillard et al., 2005; Förstner and Moonen, 1999; Moakher and Batchelor, 2006; Pennec et al., 2004) as we do here, or from a statistical point of view, assuming the multivariate Normal distribution of the data (255) and adopting the *Fisher Information metric*³⁶ (Goh and Vidal, 2008; Skovgaard, 1984), tracing back to the seminal works of Rao (1945) and Amari (1985). We prefer the former approach because we believe it is more intuitive.

The Riemannian Manifold

In differential geometry, a differentiable manifold is a topological space that is locally similar to the Euclidean space and has a globally defined differential structure. One can endow a manifold \mathcal{M} with a *Riemannian Metric*. A (smooth) *Riemannian manifold* or (smooth) Riemannian space \mathcal{M}^{37} is a real smooth manifold equipped with an inner product on the tangent space $\mathcal{T}_{\mathbf{Q}}\mathcal{M}$ defined at each point \mathbf{Q} that varies smoothly from point to point. In the SPD manifold, for any two points \mathbf{v}_1 and \mathbf{v}_2 in the tangent space the inner product through point \mathbf{Q} is given by

$$\text{tr}(\mathbf{Q}^{-1}\mathbf{v}_1\mathbf{Q}^{-1}\mathbf{v}_2). \tag{268}$$

Notice that in this chapter we use the term point, covariance matrix or simply matrix to designate a point on the manifold, whereas we use the term vector (and associated lower-case italic and bold letter notation as \mathbf{v}) to designate a point on the tangent space. In all cases, all these quantities in reality are matrices.

³⁵ *The theory of plane and space curves and of surfaces in the three-dimensional Euclidean space formed the basis for development of differential geometry during the 18th century and the 19th century. Since the late 19th century, differential geometry has grown into a field concerned more generally with the geometric structures on differentiable manifolds.*

³⁶ *In information geometry, the Fisher information metric is a particular Riemannian metric which can be defined on a smooth statistical manifold. It can be used to calculate the informational difference between measurements. The metric is interesting in several respects. It can be understood to be the infinitesimal form of the relative entropy or Kullback–Leibler divergence; specifically, it is the Hessian of this divergence.*

³⁷ *So named after the German mathematician Bernhard Riemann.*

If we consider a curve on the manifold, we can compute at each point its instantaneous velocity (derivative) and its norm. The tangent space $\mathcal{T}_{\Omega}\mathcal{M}$ at point Ω is the Euclidean vector space containing the tangent vectors to all curves on \mathcal{M} passing through Ω . A Riemannian metric makes it possible to define various useful geometric notions on a Riemannian manifold, such as distances, means, deviations from the mean, lengths of curves, areas (or volumes), curvature, gradients of functions, etc. As usual for geometric mathematical concepts, it is useful to visualize these concepts with the aid of figures (fig. 9.2).

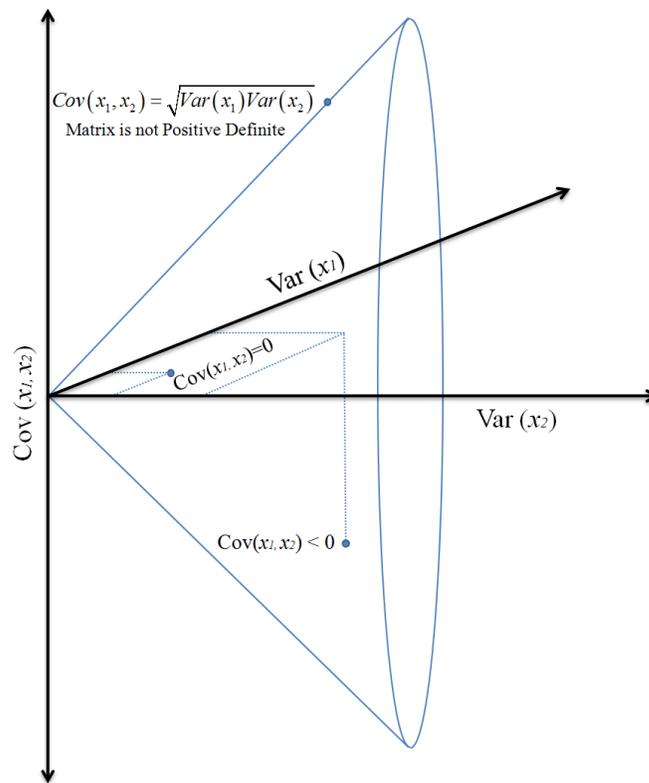


Figure 9.1: Covariance matrices are constrained by their symmetry, the strict positivity of the diagonal elements (variance) and the Cauchy-Schwarz inequalities bounding the absolute value of the off-diagonal elements: $|\text{Cov}(x_i x_j)| \leq (\text{Var}(x_i) \text{Var}(x_j))^{1/2}$, for all $i, j \in \{1, \dots, N\}$. This topology is easily visualized in case of 2x2 matrices; any 2x2 covariance matrix can be seen as point in 3D Euclidean space, with two coordinates given by the two variances (diagonal elements) and the third coordinate given by the covariance (either one of the off-diagonal element). By construction a covariance matrix must stay within the cone boundaries. As soon as the point touches the boundary of the cone, the matrix is no more positive definite; at the boundary anywhere it has a null eigenvalue, while only at the vertex it has two null eigenvalues. Notice that a matrix inside the cone may be very close in Euclidian space to the boundary. This is in contrast with the Riemannian manifold that we introduce in this chapter. Note also that if the point lies on the horizontal plane with $\text{Cov}(x_i x_j) = 0$, the covariance matrix is diagonal.

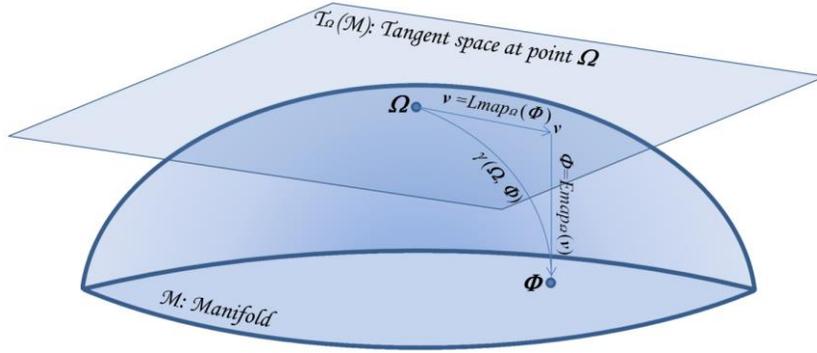


Figure 9.2: The Manifold and the tangent space at a point. Consider a point Ω on \mathcal{M} and construct the tangent space $\mathcal{T}_\Omega\mathcal{M}$ on it. Now take a tangent vector \mathbf{v} departing from Ω , which is our reference point. There exists one and only one geodesic on the manifold starting at Ω that corresponds to \mathbf{v} ; think at rolling the plane (tangent space) on the surface (manifold) in such a way that the vector always touches the surface. The end point on \mathcal{M} is Φ . We see that the geodesics on \mathcal{M} through Ω are transformed into straight lines and the distances along all geodesics are preserved (this is true in the neighborhood of Ω). (Rearranged from Goh and Vidal, 2008).

The *geodesic* between two points of \mathcal{M} is the curve joining the two points with minimum length. Such curve, indicated by $\gamma(\Omega \rightarrow \Phi)$ in the figure, is unique for a given metric. The length of the geodesic between these two points is their *distance*. Since these points are SPD matrices, the half-point on this curve, according to the chosen metric, is the *mean* of the two matrices. Here and hereafter the *mean* should be understood as a *geometric* concept, not as an *arithmetic* concept.

The Exponential and Logarithmic Map

The exponential and logarithmic maps are shown graphically in fig. 9.2. The function that maps a vector $\mathbf{v} \in \mathcal{T}_\Omega\mathcal{M}$ onto the point Φ of the manifold \mathcal{M} following the geodesic starting at Ω , is named the *exponential map* and denoted $\Phi = \text{Emap}_\Omega(\mathbf{v})$. This is the map from the tangent space to the manifold, $\mathcal{T}_\Omega\mathcal{M} \rightarrow \mathcal{M}$. The map is defined in the whole tangent space, but is generally one-to-one only locally around the point $\mathbf{0}$ in $\mathcal{T}_\Omega\mathcal{M}$, which corresponds to Ω in the manifold \mathcal{M} . The exponential map is defined such as

$$\Phi = \text{Emap}_\Omega(\mathbf{v}) = \Omega^{1/2} \exp\left(\Omega^{-1/2} \mathbf{v} \Omega^{-1/2}\right) \Omega^{1/2}. \quad (269)$$

The inverse operation is the function mapping the geodesic starting at the reference point Ω and going to a point Φ of the manifold \mathcal{M} onto the shortest tangent vector $\mathbf{v} \in \mathcal{T}_\Omega \mathcal{M}$. It is named the *logarithmic map* and denoted $\mathbf{v} = Lmap_\Omega(\Phi)$. The logarithmic map is defined such as

$$\mathbf{v} = Lmap_\Omega(\Phi) = \Omega^{1/2} \ln\left(\Omega^{-1/2} \Phi \Omega^{-1/2}\right) \Omega^{1/2}. \quad (270)$$

See the definition of matrix symmetric square root (51) and its inverse (52), the matrix exponential (53) and its inverse the matrix logarithm (54).

We can already see a crucial difference between working in a Riemannian manifold and working in an Euclidean vector space: *every operation on the manifold is made with respect to a reference point*. The same operation with respect to another point yields a different result. The operations in the Riemannian manifold are always *local*. Said differently, the quantities we compute can be conceived as proportions with respect to a reference point.

The Geodesic

Given two points Ω and Φ on their native space \mathcal{M} , the unique *Riemannian geodesic* is the curve with minimum length joining them, given by

$$\gamma_R(\beta, \Omega \rightarrow \Phi) = \Omega^{1/2} \left(\Omega^{-1/2} \Phi \Omega^{-1/2}\right)^\beta \Omega^{1/2}, \quad \beta \in [0, 1], \quad (271)$$

where beta is the step size. With $\beta=1/2$ we obtain the mean of the two points.

The Distance

Given two points Ω and Φ on their native space \mathcal{M} , the *Riemannian distance* between them is the length of the geodesic. It is given by (Bhatia, 2013; Moakher and Batchelor, 2006; Pennec et al., 2004)

$$\delta_R(\Omega \leftrightarrow \Phi) = \left\| \ln\left(\Phi^{-1/2} \Omega \Phi^{-1/2}\right) \right\|_F = \sqrt{\text{tr} \ln^2(\mathbf{A})} = \sqrt{\sum_n \ln^2 \lambda_n}, \quad (272)$$

where $\ln^2 \lambda_n = \left[\ln(\lambda_n) \right]^2$ and \mathbf{A} holds in diagonal the eigenvalues of any of the following four expressions,

$$\Phi^{-1}\Omega; \Omega^{-1}\Phi; \Phi^{-\frac{1}{2}}\Omega\Phi^{-\frac{1}{2}}; \Omega^{-\frac{1}{2}}\Phi\Omega^{-\frac{1}{2}}, \quad (273)$$

showing that such distance is symmetric. This distance has a remarkable number of properties, some of which are reported in table 9.1 (Förstener and Moonen, 1999; Sra, 2012; Moakher, 2005). For more inequalities see Sra (2012) and Bhatia (2007).

These properties have a number of interesting consequences, to which we will come back later on. Notice that other possible Riemannian distances can be used as well (for example see Li and Wong, 2013). Notice also that eq. (272) is used in the universal MDM BCI classifier (256) throughout chapter VIII.

Table 9.1: Some important properties and inequalities of the Riemannian Affine-Invariant Distance. * Products $\mathbf{B}^{-1}\Omega\mathbf{B}$ and $\mathbf{B}^{-1}\Phi\mathbf{B}$ are not necessarily symmetric.

Fundamental Properties of the Riemannian metric (distance)		
(274) Postivity	$\delta_R(\Omega \leftrightarrow \Phi) \geq 0.$	with equality iff $\Omega = \Phi$
(275) Symmetry	$\delta_R(\Omega \leftrightarrow \Phi) = \delta_R(\Phi \leftrightarrow \Omega)$	
(276) Congruence-Invariance	$\delta_R(\Omega \leftrightarrow \Phi) = \delta_R(\mathbf{B}^T\Omega\mathbf{B} \leftrightarrow \mathbf{B}^T\Phi\mathbf{B}),$	for any invertible \mathbf{B}
(277) Invariance under Inversion	$\delta_R(\Omega \leftrightarrow \Phi) = \delta_R(\Omega^{-1} \leftrightarrow \Phi^{-1})$	
(278) Proportionality	$\delta_R(\Omega \leftrightarrow \gamma_R(\alpha, \Omega \rightarrow \Phi)) = \alpha\delta_R(\Omega \leftrightarrow \Phi)$	
Some inequalities of the Riemannian metric (distance)		
(279)	$\delta_R(\gamma_R(\alpha, \Omega \rightarrow \Phi) \leftrightarrow \gamma_R(\alpha, \Omega \rightarrow \Xi)) \leq \alpha\delta_R(\Phi \leftrightarrow \Xi)$	
(280)	$\delta_R(\Omega \leftrightarrow \Phi) \geq \ \log\Omega - \log\Phi\ _F,$	with equality iff Ω and Φ commute

The Norm

Since the distance is congruence-invariant (see (276) in Table 9.1), we can equate the distance between two matrices $\mathbf{\Omega}$ and $\mathbf{\Phi}$ with the distance from a matrix to the identity matrix; after congruence transformation by matrix $\mathbf{\Omega}^{-1/2}$ and since $\mathbf{\Omega}^{-1/2}\mathbf{\Omega}\mathbf{\Omega}^{-1/2} = \mathbf{I}$ we have

$$\delta_R(\mathbf{\Omega} \leftrightarrow \mathbf{\Phi}) = \delta_R(\mathbf{I} \leftrightarrow \mathbf{\Omega}^{-1/2}\mathbf{\Phi}\mathbf{\Omega}^{-1/2}) = \sum_n \ln^2 \lambda_n(\mathbf{\Omega}^{-1/2}\mathbf{\Phi}\mathbf{\Omega}^{-1/2}), \quad (281)$$

where $\lambda_n(\mathbf{\Omega}^{-1/2}\mathbf{\Phi}\mathbf{\Omega}^{-1/2})$ are the eigenvalues of $\mathbf{\Omega}^{-1/2}\mathbf{\Phi}\mathbf{\Omega}^{-1/2}$.

We define then the Riemannian *norm* as the Riemannian distance from the matrix to the identity, hence as the distance of its logarithm to the zero point:

$$\|\mathbf{\Omega}^{-1/2}\mathbf{\Phi}\mathbf{\Omega}^{-1/2}\|_R = \delta_R(\mathbf{I} \leftrightarrow \mathbf{\Omega}^{-1/2}\mathbf{\Phi}\mathbf{\Omega}^{-1/2}) = \|\ln(\mathbf{\Omega}^{-1/2}\mathbf{\Phi}\mathbf{\Omega}^{-1/2})\|_F. \quad (282)$$

Notice that, if \mathbf{M}_z is the Riemannian mean of given trials $\mathbf{C}_{kz} \in \{\mathbf{C}_{1z}, \dots, \mathbf{C}_{Kz}\}$ pertaining to class z , the distance of each trial to the mean is $\sum_n \ln^2 \lambda_n(\mathbf{M}_k^{-1/2}\mathbf{C}_{kz}\mathbf{M}_k^{-1/2}) = \sum_n \ln^2 \lambda_n(\mathbf{M}_k^{-1}\mathbf{C}_{kz})$, i.e., it is a non-linear function of the eigenvalues of the whitened trials.

The Riemannian norm is zero only for the identity matrix (while the Frobenius norm is zero only for the null matrix). Either eigenvalues smaller and greater than 1 increase the norm and the norm goes to infinity as any eigenvalue goes to either infinity or zero. Importantly, because of the square of the log, an eigenvalue λ increases the norm as much as an eigenvalue $1/\lambda$ does (from which the invariance under inversion (277)). The way this metric works is illustrated in fig. 9.3.

We have also the following result:

$$\sum_n \ln^2 \lambda_n(\mathbf{\Omega}^{-1/2}\mathbf{\Phi}\mathbf{\Omega}^{-1/2}) = \sum_n \ln^2 \lambda_n(\mathbf{\Omega}^{-1}\mathbf{\Phi}), \quad (283)$$

that is, the Riemannian norms of $\mathbf{\Omega}^{-1/2}\mathbf{\Phi}\mathbf{\Omega}^{-1/2}$ and $\mathbf{\Omega}^{-1}\mathbf{\Phi}$ are equal, hence the fact that any of the expressions in (273) can be used to compute the distance. Notice that $\mathbf{\Omega}^{-1/2}\mathbf{\Phi}\mathbf{\Omega}^{-1/2} = \mathbf{\Omega}^{-1}\mathbf{\Phi}$ only if $\mathbf{\Omega}^{-1/2}$ and $\mathbf{\Phi}$ commute in multiplication, which is not true in general. If they commute $\mathbf{\Omega}^{-1}\mathbf{\Phi}$ is symmetric. Hence, neither the trace (sum of the eigenvalues), nor the determinant (product of the

eigenvalues), nor the Frobenius norm of these two quantities are equal, but the sum of the squares of the log of the eigenvalues is the same. Also, since $\Phi^{-1}\Omega$ and $\Omega^{-1}\Phi$ are not necessarily symmetric, their EVD has a form different from (35). Note that one has to use the appropriate eigenvalue-eigenvector decomposition algorithms if these two expressions are used to compute the distance or the norm, or, just use either $\Phi^{-1/2}\Omega\Phi^{-1/2}$ or $\Omega^{-1/2}\Phi\Omega^{-1/2}$, which are always symmetric.

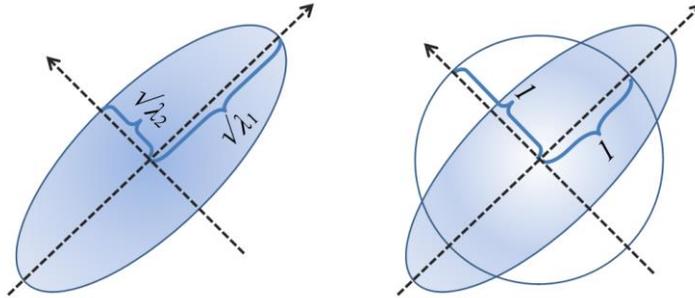


Figure 9.3: The ellipsoids in the figure are isolines of constant density of bivariate Gaussian distributions. The semiaxes are proportional to the square root of the eigenvalues of the covariance matrix. If we ask how far the ellipsoid is from the circle, which is the definition of the Riemannian norm (282), we see that an eigenvalue = 2 contribute to the distance from the identity as much as an eigenvalue = 1/2, as one would expect, since the eigenvalues are squared quantities. Neither the sum nor the sum of the logarithm of the eigenvalue has this property.

The Geometric Mean of Points on the Manifold

It is very useful to compute *means* of two or more points in the manifold. For instance, such a mean may be used to represent a class in BCI experiment, as we have seen in (256), or more in general to obtain a prototype for a cloud of points (a matrix set). Among the many means proposed by the ancient Greeks, three of them, the *arithmetic, harmonic and geometric mean* are widely used still today. For K samples of a univariate random variable x the arithmetic mean is

$$a_x = 1/K \sum_k x_k, \quad (284)$$

the harmonic mean is

$$h_x = \left(1/K \sum_k x_k^{-1}\right)^{-1} \quad (285)$$

and the geometric mean is

$$g_x = \sqrt[K]{\prod_k x_k} = \exp\left(\frac{1}{K} \sum_k \ln(x_k)\right). \quad (286)$$

When working with matrices, a straightforward definition of the matrix arithmetic mean in analogy with (284) is

$$\frac{1}{K} \sum_k \mathbf{C}_k \quad (287)$$

and a straightforward definition of the harmonic mean in analogy with (285) is

$$\left(\frac{1}{K} \sum_k \mathbf{C}_k^{-1}\right)^{-1}, \quad (288)$$

however a straightforward definition of geometric mean does not exist. To see this, it suffices to realize that the matrices in the cloud do not necessarily commute in multiplication, so in analogy with (286) one would obtain different means depending on the permutation of the index. Hence, one may define the geometric mean as the K^{th} square root of the matrix product, in analogy with (286), only if the matrices all pair-wise commute.

It turns out that there are infinite ways to define a geometric mean for the general case $K > 2$ (Bhatia, 2013). Researchers have begun by listing a number of desirable properties a mean should possess. Ten such properties are known in the literature as the ALM properties, from the seminal paper of Ando, Li and Mathias (2004). When doing so, one finds out that the arithmetic and harmonic mean do not possess all desirable properties of a mean. For example, we require that the inverse of the mean of the inverse of several matrices be the mean of these matrices, but this is not true for the arithmetic, nor for the harmonic mean. Such a property, as we will see, is possessed by the geometric mean. Another simple example will provide motivation to investigate the geometric mean of SPD matrices; ask yourself, what should be the “mean” of \mathbf{C} and \mathbf{C}^{-1} ? Their arithmetic mean is

$$\frac{1}{2}(\mathbf{C} + \mathbf{C}^{-1}) \text{ and their harmonic mean is } \left[\frac{1}{2}(\mathbf{C} + \mathbf{C}^{-1})\right]^{-1}.$$

Only their geometric mean gives us the intuitive answer: the “mean” of \mathbf{C} and \mathbf{C}^{-1} is the identity matrix. Another reason to praise the geometric mean is that it considers proportional (relative) variations, which is meaningful when working with variances and covariances, while the arithmetic mean considers absolute variations. Nonetheless, the main reason to discard the arithmetic mean of SPD matrices as representation of a matrix cloud is that it performs poorly as a representation of several points. For example, in a study of Li and Wong (2013) performances in classification of sleep

stages using an Euclidean distance proved way inferior as compared to a Riemannian distance. For all these reasons in the last 25 years there has been an intense effort to define and estimate a geometric mean of several PSD matrices. For $K=2$ the problem has been solved soon. For the case $K > 2$ it has proven elusive for a quarter of century (see Bhatia, 2013, for an historical perspective).

In the Riemann framework it is convenient to use the definition of Fréchet means and the ensuing variational approach: in a univariate context, while the *arithmetic mean* (284) minimizes the sum of the squared Euclidean distances to K given positive scalars, such as

$$a_x = \underset{a_x}{\operatorname{argmin}} \sum_k (a_x - x_k)^2, \quad (289)$$

the *geometric mean* (286) minimizes the sum of the squared hyperbolic distances to K given positive numbers, such as

$$g_x = \underset{g_x}{\operatorname{argmin}} \sum_k (\ln g_x - \ln x_k)^2. \quad (290)$$

In analogy, we define the (least-squares) mean \mathbf{M} of K SPD matrices $\mathbf{\Omega}_k$ such as (Bhatia and Holbrook, 2006; Moakher, 2005)

$$\mathbf{M} = \underset{\mathbf{M}}{\operatorname{argmin}} \sum_k \delta_R^2(\mathbf{M} \leftrightarrow \mathbf{\Omega}_k). \quad (291)$$

In words, the geometric mean is the matrix minimizing the sum of the squared Riemannian distances of all elements from itself. \mathbf{M} is the *unique SPD geometric mean*³⁸ satisfying non-linear matrix equation (Moakher, 2005)

$$\sum_k \ln(\mathbf{\Omega}_k^{-1} \mathbf{M}) = \mathbf{0} \text{ or, equivalently, } \sum_k \ln(\mathbf{M}^{-\frac{1}{2}} \mathbf{\Omega}_k \mathbf{M}^{-\frac{1}{2}}) = \mathbf{0}. \quad (292)$$

Notice that if the matrices are covariance matrices of trials belonging to a given class, the second expression here above says that the sum of the log of the whitened trials equals the null matrix. We will come back to this later on.

³⁸ This least-squares geometric mean is also referred to as the barycenter, the center of mass or the Karcher mean.

Given two points Ω_1 and Ω_2 on their native space \mathcal{M} , the *Geometric Mean* between them, indicated in the literature by $\Omega_1 \# \Omega_2$, is given by (Bhatia, 2003, 2007; Bathia and Karandikar, 2011; Bini and Iannazzo, 2011; Moakher, 2005)

$$\Omega_1 \# \Omega_2 = \Omega_1^{1/2} \left(\Omega_1^{-1/2} \Omega_2 \Omega_1^{-1/2} \right)^{1/2} \Omega_1^{1/2} = \Omega_1 \left(\Omega_1^{-1} \Omega_2 \right)^{1/2} = \left(\Omega_2 \Omega_1^{-1} \right)^{1/2} \Omega_1, \quad (293)$$

or

$$\Omega_1 \# \Omega_2 = \Omega_1^{1/2} \exp \left(\frac{1}{2} \log \left(\Omega_1^{-1/2} \Omega_2 \Omega_1^{-1/2} \right) \right) \Omega_1^{1/2} = \Omega_1 \exp \left(\frac{1}{2} \log \left(\Omega_1^{-1} \Omega_2 \right) \right), \quad (294)$$

which, as one should expect, is the midpoint of the geodesic in (271). In the above the indexes 1 and 2 can be switched to obtain as many more expressions. This geometric mean of two matrices turns out to be the unique solution of a quadratic Ricatti equation (Bhatia, 2007; Nakamura, 2009), yielding

$$(\Omega_1 \# \Omega_2) \Omega_2^{-1} (\Omega_1 \# \Omega_2) = \Omega_1; \quad (\Omega_1 \# \Omega_2) \Omega_1^{-1} (\Omega_1 \# \Omega_2) = \Omega_2 \quad (295)$$

Given a set $\Omega : \{\Omega_1, \dots, \Omega_K\}$ of $K > 2$ points (e.g., covariance matrices), there is no closed form solution for computing their geometric mean. Several iterative algorithms have been proposed. We use the following (Pennec et al., 2004; Manton, 2004; see also Bini and Iannazzo, 2011) algorithm.

Algorithm (296): Geometric Mean M of K SPD matrices $\Omega_k \in \{\Omega_1, \dots, \Omega_K\}$.

Initialize M by a smart guess (e.g., the cheap mean of Bini and Iannazzo, 2011) or with the arithmetic mean.

Repeat

$$M \leftarrow M^{1/2} \exp \left[\frac{1}{K} \sum_k \ln \left(M^{-1/2} \Omega_k M^{-1/2} \right) \right] M^{1/2},$$

Until Convergence

$$\left(\left\| \sum_k \ln \left(M^{-1/2} \Omega_k M^{-1/2} \right) \right\|_F < \varepsilon, \text{ according to (292)}. \right)$$

These iterations have linear convergence. In simulations with high-dimensional matrices it has been found that they do not converge if the matrices are very distant one from the others (Bini and

Iannazzo, 2011). With real EEG data and up to 32-electrode data we have always seen this algorithm converging reliably, thus the problem thus not apply for real BCI applications. For EEG in general, a principal component analysis (196) can be applied to reduce the dimension if many electrodes are available. Numerical simulations and an analysis of the complexity reveal that this algorithm, compared to others, is also convenient in term of speed of computations (Jeuris et al., 2012). Notice that, in essence, this algorithm iteratively maps the points in the tangent space through the current estimation of the mean, computes the algebraic mean in the tangent space (where the arithmetic mean makes sense) and maps back the updated mean estimation on the manifold, until convergence (fig. 9.4).

The research of algorithms for estimating the geometric mean of three or more matrices is currently a very active field (Bhatia, 2013; Jeuris et al., 2012; Nakamura, 2009; Poloni, 2010). For recent developments see Bini and Iannazzo (2011, 2013) and Moakher (2012). For a matrix cloud (set) Ω : $\{\Omega_1, \dots, \Omega_k\}$ of $K > 2$ points with geometric mean $M(\Omega)$, we have listed important properties of the geometric mean in table 9.2 (Moakher, 2005; Nakamura, 2009).

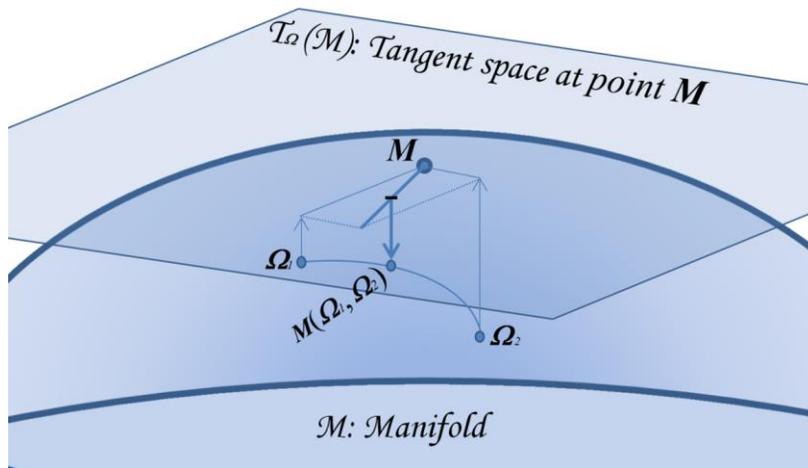


Figure 9.4: Zoom on the manifold as it is represented in fig. 9.2. Consider two points Ω_1 and Ω_2 on \mathcal{M} and construct the tangent space $\mathcal{T}_{\Omega}\mathcal{M}$ through their current estimation of the mean M , initialized as the arithmetic mean. At each iteration, the algorithm maps the points on the tangent space, computes the mean vector and maps back the point on the manifold. At each iteration the estimation of the mean is updated, thus the point of transition into the tangent space changes, until convergence, that is, until this transition point will not change anymore, coinciding with the geometric mean, that is, satisfying (292).

Table 9.2: Some important properties of the geometric mean

Properties of the Geometric Mean	
(297) Invariance by reordering	The GM of K SPD matrices is the same in any order
(298) Invariance under congruence transformation	$\mathbf{B}^T \mathbf{M}(\Omega) \mathbf{B} = \mathbf{M} \{ \mathbf{B}^T \Omega_1 \mathbf{B}, \dots, \mathbf{B}^T \Omega_K \mathbf{B} \}$
(299) Self-Duality	$\mathbf{M}(\Omega) = \left[\mathbf{M} \{ \Omega_1^{-1}, \dots, \Omega_K^{-1} \} \right]^{-1}$
(300) Joint Homogeneity	$\mathbf{M} \{ \alpha_1 \Omega_1, \dots, \alpha_K \Omega_K \} = \left(\prod_k \alpha_k \right)^{1/K} \mathbf{M}(\Omega), \alpha \geq 0$
(301) Determinant Identity	$\det \mathbf{M}(\Omega) = \left(\prod_k \det \Omega_k \right)^{1/K}$
(302) for any PSD Φ	$\delta_R^2(\mathbf{M}(\Omega) \leftrightarrow \Phi) \leq \sum_k 1/K \left[\delta_R^2(\Omega_k, \mathbf{M}(\Omega)) - \delta_R^2(\Omega_k, \Phi) \right]$
(303) iff all matrices Ω_k pair-wise commute	$\mathbf{M}(\Omega) = \left(\prod_k \Omega_k \right)^{1/K}$

RECENT INVESTIGATIONS

Introduction

Some of the properties of the geometric mean, particularly (298) (table 9.2), which is analogous to (276) (table 9.1) for the distance between two SPD matrices, have important consequences for the use of these tools with EEG data. In this section we present some of our recent investigations in this sense. In particular, we have very recently inquired about the connection between these tools and the diagonalization methods presented in chapter IV to VII. We also propose a standardized measure of distance, a distance matrix quantity and an index of cloud entropy.

Connections with Diagonalization Methods

Blind Source Separation

Consider the typical linear instantaneous mixing model for EEG we have used over and over again in chapter VI:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t), \quad (304)$$

where \mathbf{A} is the unknown mixing matrix and $\mathbf{s}(t)$ the unknown source process. Model (304) is the base of practically all single-subject blind source separation signal processing approaches used in EEG, including independent component analysis. Take the covariance matrix of trial \mathbf{X} , which using (304) can be expressed as a function of the source sample covariance matrix \mathbf{S} such as

$$\mathbf{A}\mathbf{S}\mathbf{A}^T = \mathbf{C}. \quad (305)$$

From the congruence invariance property of the distance (276) we see that, given two realizations \mathbf{C}_1 and \mathbf{C}_2 , with associated source covariance matrix \mathbf{S}_1 and \mathbf{S}_2 , the distance between \mathbf{C}_1 and \mathbf{C}_2 equals the distance between \mathbf{S}_1 and \mathbf{S}_2 . That is to say, *working with the Riemannian distance with sensor covariance matrices is equivalent to working in the source space with an optimal estimation of the sources*. This is true whenever \mathbf{A} is invertible, that is, in the $\frac{1}{2}N(N+1)$ dimensional Riemannian space.

Joint Blind Source Separation

Consider multi-subject BCIs, in which multiple users interact at the same time with the same interface (Bonnet, Lotte and Lécuyer; Yuan et al., 2013; Schultze-Kraft et al., 2013). The interesting point of such set up is that we can in theory obtain a BCI with *100% accuracy on single-trial given a sufficient number of subjects*; instead of averaging data or classification scores across trials, we can do the same across subjects on a single-trial³⁹ We can actually go a little further and exploit not only the multitude of trials, but also their theoretical synchronization between subjects. Let us make the example of P300-based BCIs. For other paradigms the development is similar. As we have seen, for single-subject using a P300-based BCI the super-trials have form (262). Referring to notation introduced in chapter VIII, for the multi-subject case (MS) the trial for class $z \in \{1, \dots, Z\}$ is given by

$$\mathbf{X}_z^{MS \text{ P300}} = \begin{pmatrix} \bar{\mathbf{X}}_{(+)}^T \\ \mathbf{X}_{1z}^T \\ \dots \\ \mathbf{X}_{Mz}^T \end{pmatrix} \in \mathfrak{R}^{(N(M+1))N}, \quad (306)$$

where M is the number of subjects and N the number of sensors. Notice that the temporal prototype is still just one, as for the single-subject case in (262), as it applies to all subjects. The covariance matrix of (306) for the example case $M=2$ has form

$$\mathbf{C}_z = \frac{1}{(T-1)} \left[\mathbf{X}_z^{MS \text{ P300}} \left(\mathbf{X}_z^{MS \text{ P300}} \right)^T \right] = \frac{1}{(T-1)} \begin{pmatrix} \bar{\mathbf{X}}_{(+)} \bar{\mathbf{X}}_{(+)}^T & \bar{\mathbf{X}}_{(+)} \mathbf{X}_{1z}^T & \bar{\mathbf{X}}_{(+)} \mathbf{X}_{2z}^T \\ \mathbf{X}_{1z} \bar{\mathbf{X}}_{(+)}^T & \mathbf{X}_{1z} \mathbf{X}_{1z}^T & \mathbf{X}_{1z} \mathbf{X}_{2z}^T \\ \mathbf{X}_{2z} \bar{\mathbf{X}}_{(+)}^T & \mathbf{X}_{2z} \mathbf{X}_{1z}^T & \mathbf{X}_{2z} \mathbf{X}_{2z}^T \end{pmatrix}. \quad (307)$$

Block $\bar{\mathbf{X}}_{(+)} \bar{\mathbf{X}}_{(+)}^T \in \mathfrak{R}^{N \times N}$ is again the covariance matrix of the temporal prototype. This does not change across trials and has no value for classification.

The off-diagonal blocks $\mathbf{X}_{1z} \bar{\mathbf{X}}_{(+)}^T$ and $\mathbf{X}_{2z} \bar{\mathbf{X}}_{(+)}^T \in \mathfrak{R}^{N \times N}$ (or their transpose $\bar{\mathbf{X}}_{(+)} \mathbf{X}_{1z}^T$ and $\bar{\mathbf{X}}_{(+)} \mathbf{X}_{2z}^T$) hold the covariances between the trial of the subjects and the prototype and are relevant for

³⁹ !

classification just as in (259); the only difference is that instead of having only one of such covariance, now we have two of them, increasing accordingly the classification power.

Moreover, consider now off-diagonal block $\mathbf{X}_{2z}\mathbf{X}_{1z}^T$. This (or its transpose $\mathbf{X}_{1z}\mathbf{X}_{2z}^T$) holds the covariance between the trial of the two subjects; since the P300 response is synchronized, this covariance will be large for the target trials only, thus it hold some information useful for classification.

Notice that classifying based on (307) does not amount to simply add the data of each individual. This is what we would obtain putting to zero the off-diagonal blocks $\mathbf{X}_{2z}\mathbf{X}_{1z}^T$ and $\mathbf{X}_{1z}\mathbf{X}_{2z}^T$ of (307). Here we are actually exploiting also the synchronization of the ERP response and the covariance of the synchronized response of the two subjects. The same goes with more than two subjects, wherein all pair-wise cross-subject covariances may be exploited. This is the same endeavor of the joint blind source separation (JBSS) approach encountered in chapter VII. The JBSS aims at finding M demixing matrices simultaneously, maximizing the covariances among individuals. The sample covariance matrix in this case has form (extending (305))

$$\begin{pmatrix} \mathbf{A}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{A}_M \end{pmatrix} \begin{pmatrix} \mathbf{S}_1 & \cdots & \mathbf{S}_{1M} \\ \vdots & \ddots & \vdots \\ \mathbf{S}_{M1} & \cdots & \mathbf{S}_M \end{pmatrix} \begin{pmatrix} \mathbf{A}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{A}_M \end{pmatrix}^T = \mathbf{C}, \quad (308)$$

where \mathbf{A}_m is the mixing matrix for the m^{th} subject, \mathbf{S}_m is the source covariance matrix of the m^{th} subject and \mathbf{S}_{ij} is the cross-covariance between the sources of the i^{th} and j^{th} subject, with $i, j \in \{1, \dots, M\}$; thus, once again, due to the congruence-invariance of the mean and the distance, using the Riemann framework we are classifying in the sensor space as if we were estimating the optimal JBSS filtering.

A Diagonality Function

Consider the Riemannian distance of a matrix to its diagonal part $\delta_R(\mathbf{C} \leftrightarrow \text{diag}(\mathbf{C}))$, where $\text{diag}(\cdot)$ nullifies the off-diagonal elements of the argument. Using the congruence invariance of the Riemannian distance we have that

$$\delta_R(\mathbf{C}, \text{diag}(\mathbf{C})) = \left\| \ln \left(\text{diag}(\mathbf{C})^{-\frac{1}{2}} \mathbf{C} \text{diag}(\mathbf{C})^{-\frac{1}{2}} \right) \right\|_F, \quad (309)$$

that is, the Riemannian distance of a covariance matrix from its diagonal part is the Riemannian norm (282) of the correlation matrix. We can then use this definition to obtain a diagonality function to weight matrices in AJDC and replace the Euclidean measure (229).

AJD and the geometric mean of a matrix set

The following proposition establishes the relation between the Riemannian geometric mean and the approximate joint diagonalization (AJD) of a matrix set.

Proposition (310)

Let matrix set $C = \{C_1, \dots, C_K\}$ be composed by SOS statistics of data generated under instantaneous linear mixing model (213) $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$, where \mathbf{A} is the mixing matrix and its inverse \mathbf{B}^T is the demixing matrix, and let \mathbf{M} be the least-squares geometric mean of set C . Then

$$\mathbf{M} \approx \mathbf{A} \left[\prod_k \left(\text{diag}(\mathbf{B}^T \mathbf{C}_k \mathbf{B}) \right) \right]^{\frac{1}{K}} \mathbf{A}^T, \quad (311)$$

with equality iff $\mathbf{B}^T \mathbf{C}_k \mathbf{B}$ is diagonal for all $k \in \{1, \dots, K\}$.

Proof:

Since diagonal matrices commutes in multiplication, it follows directly from the fact that the geometric mean of a set $\{\mathbf{\Omega}_1, \dots, \mathbf{\Omega}_K\}$ of matrices that all pair-wise commute in multiplication after congruent transformation $\mathbf{F}^T \mathbf{\Omega}_k \mathbf{F}$ is $\mathbf{F}^{-T} \left[\prod_k (\mathbf{F}^T \mathbf{\Omega}_k \mathbf{F}) \right]^{\frac{1}{K}} \mathbf{F}^{-1}$ (Bhatia and Holbrook, 2006, Proposition 18).

If the set can be diagonalized exactly, that is if $\text{diag}(\mathbf{B}^T \mathbf{C}_k \mathbf{B}) = \mathbf{B}^T \mathbf{C}_k \mathbf{B}$, for all C_k , equality (311) holds exactly.

Corollary (312)

Since the scaling of the columns of \mathbf{B} is arbitrary, let us scale \mathbf{B} such that

$$\left[\prod_k \left(\text{diag}(\dot{\mathbf{B}}^T \mathbf{C}_k \dot{\mathbf{B}}) \right) \right]^{\frac{1}{K}} = \mathbf{I}. \quad (313)$$

After such scaling, using (311) one can estimate approximately the geometric mean with the scaled approximate joint diagonalizer or its inverse as

$$\mathbf{M} \approx (\mathbf{B}\mathbf{B}^T)^{-1} \approx \mathbf{A}\mathbf{A}^T. \quad (314)$$

Again, this estimate is the true geometric mean \mathbf{M} iff $\mathbf{B}^T \mathbf{C}_k \mathbf{B}$ is diagonal for all $k \in \{1, \dots, K\}$. Equation (314) also says that the mixing matrix is equal to $\mathbf{M}\mathbf{B}$ and the demixing matrix equal to $\mathbf{A}^T \mathbf{M}^{-1}$ after such scaling.

Corollary (315)

Using the properties of matrix exponential and logarithm and assuming that the off-diagonal elements of the sum $\sum_k \ln(\mathbf{B}^T \mathbf{C}_k \mathbf{B})$ vanish as K goes to infinity we have asymptotic result

$$\lim_{k \rightarrow \infty} \mathbf{M} = \mathbf{A} \left[\exp \left(\frac{1}{K} \sum_k \ln(\mathbf{B}^T \mathbf{C}_k \mathbf{B}) \right) \right] \mathbf{A}^T \quad (316)$$

or use instead approximation

$$\mathbf{M} \approx \mathbf{A} \text{diag} \left[\exp \left(\frac{1}{K} \sum_k \ln(\mathbf{B}^T \mathbf{C}_k \mathbf{B}) \right) \right] \mathbf{A}^T \quad (317)$$

Corollary (318)

For sets composed of two matrices only both $\mathbf{B}^T \mathbf{C}_1 \mathbf{B}$ and $\mathbf{B}^T \mathbf{C}_2 \mathbf{B}$ are exactly diagonal, so in this case

$$\mathbf{M} = \mathbf{A} (\mathbf{B}^T \mathbf{C}_1 \mathbf{B} \mathbf{B}^T \mathbf{C}_2 \mathbf{B})^{\frac{1}{2}} \mathbf{A}^T = \sum_n (\sqrt{d_{1n} d_{2n}} \mathbf{a}_n \mathbf{a}_n^T), \quad (319)$$

where d_{1n} and d_{2n} are the N diagonal elements of $\mathbf{B}^T \mathbf{C}_1 \mathbf{B}$ and $\mathbf{B}^T \mathbf{C}_2 \mathbf{B}$, respectively.

Corollary (318) is the strongest result as it shows the *exact* relation between the geometric mean of two matrices to be diagonalized and the joint diagonalizer. This result applies to all spatial filtering and BSS methods based on the diagonalization of two matrices presented in chapter V and VI.

In case of more than two matrices, note that these approximations may be used to initialize iterative algorithms for the geometric mean or for estimating a “pseudo” geometric mean of non positive-

definite symmetric matrices whether the matrices in the set share a common eigen-structure (i.e., they can be approximately jointly diagonalized).

Standardized Distances to Geometric mean.

Riemann distances to geometric mean do not have a symmetric distribution. To make their distribution symmetric we may use the *standardized distances to the geometric mean*, which we here define using the definition of geometric mean and standard deviation of random variables. Given training points $\mathbf{\Omega}_{zk}$ of K covariance matrices (trials) with $k \in \{1, \dots, K\}$ for class z , we use the distances of the $\mathbf{\Omega}_{zk}$ from their class geometric mean \mathbf{M}_z given by $\delta_{zk} (\mathbf{\Omega}_{zk} \leftrightarrow \mathbf{M}_z)$. Then we have for the class trials:

$$\text{the geometric mean of the } \delta_{zk} \text{ distances:} \quad \mu_z = \exp\left(\frac{1}{K} \sum_k \ln(\delta_{zk})\right) \quad (320)$$

$$\text{the geometric standard deviation of the } \delta_{zk} \text{ distances:} \quad \sigma_z = \exp\left(\sqrt{\frac{1}{K} \sum_k \left(\ln\left(\frac{\delta_{zk}}{\mu_z}\right)\right)^2}\right) \quad (321)$$

$$\text{the geometric z-score (standardized) of the deviations:} \quad z(\delta_{zk}) = \ln\left(\frac{\delta_{zk}}{\mu_z}\right) / \ln(\sigma_z) \quad (322)$$

The z-score of the deviation can be used instead of the raw distance in MDM (see chapter VIII). It can also be used to improve a method we have developed to detect automatically artifacts (Barachant Andreev and Congedo, 2013).

The Distance Matrix

For any given point $\mathbf{\Omega}_k$ in the set $\mathbf{\Omega}_k \in \{\mathbf{\Omega}_1, \dots, \mathbf{\Omega}_K\}$ we define the *distance matrix* from the geometric mean matrix \mathbf{M} as

$$\mathbf{A}(\mathbf{\Omega}_k \leftrightarrow \mathbf{M}) = \mathbf{M}^{-1/2} \mathbf{\Omega}_k \mathbf{M}^{-1/2}. \quad (323)$$

Contrary to the Riemannian distance, this matrix holds *directional information* in the Manifold. Moreover, it is easy to show that the distance between the point $\mathbf{\Omega}_k$ and the geometric mean is the same as the *distance between the distance matrix and the identity matrix*, therefore it is comparable for different geometric means. Note that the distance matrix is the *whitened trial*, since $\mathbf{M}^{-1/2}$ is the

whitening matrix, that is, it holds true $\mathbf{M}^{-1/2}\mathbf{M}\mathbf{M}^{-1/2} = \mathbf{I}$. Since the distance matrices hold directional information, the geometric mean of *iid* distance matrices approaches the identity very fast, *regardless the magnitude of the deviations*. This fact can be used for classification: for instance in the P300-based BCI at each repetition one observes an unknown trial \mathbf{C} . So far we have been summing along repetitions the distances of each \mathbf{C} to the target and non-target geometric means \mathbf{M}_+ (target) and \mathbf{M}_- (non-target) and applied the MDM. Instead, one can compute the cumulating distance matrices

$$\Delta(\mathbf{C} \leftrightarrow \mathbf{M}_+) \text{ and } \Delta(\mathbf{C} \leftrightarrow \mathbf{M}_-) \quad (324)$$

obtained along repetitions such as

$$\exp\left(\frac{1}{\#rep} \sum_i^{\#rep} \ln\left(\mathbf{M}_z^{-1/2} \mathbf{C}_i \mathbf{M}_z^{-1/2}\right)\right) \quad (325)$$

for class $z \in \{+, -\}$ and then apply the MDM algorithm taking the minimum distance from these cumulated matrices to the identity matrix. The cumulated distance (325) provides us with physiological information to interpret the data. In fact the expected form of the cumulated distance matrix is the identity; any deviation from this form can be mapped to understand source of variations in the actual sample.

Wiener Entropy: an Index of Cloud Entropy

For a given cloud of points, besides its central location, we may want to characterize its spread. For instance we may want to know if the training data for two classes have similar spread or not. For doing this we borrow the little-known *Wiener Entropy* index. For a power spectrum at $f \in \{1, \dots, F\}$ discrete frequencies, the *spectral flatness* or *Wiener entropy* is the ratio between the geometric mean of the power spectra and their arithmetic mean. It is a dimensionless measure bounded in between 0 (a pure sine wave) and 1 (white noise, flat spectrum). The more colored is the signal, the lower the Wiener entropy. Usually it is expressed in decibels to increase the dynamic range.

When used on Riemann distances the Wiener entropy expresses the *cloud compactness*: the Wiener entropy is 1 when the geometric mean equals the arithmetic mean, that is, when all points are at the same distance from the geometric mean. We have thus the following

Conjecture (326)

In a set Ω of N -dimensional SPD matrices the cloud compactness may equal 1 iff there are at most $\frac{1}{2}N(N-1)$ points, otherwise it will smaller than 1.

Since the Riemann distance is a *multiplicative operation of strictly positive quantities*, it is reasonable to model its distribution as a *log-normal distribution*. Furthermore, location and scale parameter for data distributed log-normally are usually treated in term of geometric mean and geometric standard deviation, which suits perfectly the Riemannian framework. The geometric mean and standard deviation of a log-normal distribution are e^μ and e^σ , respectively, where μ and σ are the mean and standard deviation of a normal distribution.

For a log-normal distribution the harmonic (h), geometric (g), and arithmetic (a) means are related such as

$$h = g^2/a, a = g^2/h \text{ and } g = \sqrt{ah} \quad (327)$$

Using this and starting from the definition of Wiener entropy as the ratio between the geometric and arithmetic mean we obtain

$$\text{Wiener entropy} = h/g = g/a. \quad (328)$$

We see that the three means, which are always in relation $a \geq g \geq h$ (Bhatia, 2007), for a log-normal distribution are in logarithmic relation with *stretch factor* $a/g = g/h$. Therefore, given a set of K points $\Omega_k, k \in \{1, \dots, K\}$, if we express the Wiener entropy in minus decibel we obtain a *dimensionless additive measure of cloud entropy*, equal to zero when $a = g = h$ and going to $+\infty$ as the stretch factor grows:

$$e_\Omega = -10 \log_{10} [g_\Omega/a_\Omega]. \quad (329)$$

We predict that the higher the cloud entropy the more valuable the contribution of treating covariance matrices in their native space (Riemannian) instead of approximating with Euclidian geometry.

CHAPTER X

CONCLUSIONS AND PERSPECTIVES

The leitmotiv of this manuscript is the *EEG source analysis*, that is, the study of latent variables hidden in EEG scalp recordings. We have described several families of methods. They follow different paths and have different aims, but all allow to disclose a considerable amount of information to which we do not have access in the sensor space.

In chapter III we have presented distributed linear inverse solutions. The eLORETA method is a true weighted linear inverse solution respecting the sensor measurement and achieving zero-localization error in point spread function. Its use with real data is rather straightforward; the only free parameter to be set is the regularization parameter. We have provided a solution to this choice that is effective in practice. One can hardly imagine a significant improvement over eLORETA in the search of a distributed linear inverse solution, respecting the constraints of linearity, fit to sensor measurement and no-localization error in point spread function. Nonetheless other solutions with specific properties may be devised dropping one or more of these constraints. All inverse solutions make use of the leadfield matrix. The goodness of the head model chosen for a given subject is crucial for the accuracy of the results. Therefore, it is important to spend effort in the precision of head modeling. This is in contrast with diagonalization methods that rely only on statistics of the data. As for EEG source analysis tools in general, it is hard to check the validity of inverse solutions. For this purpose one may want to use different neuroimaging modalities and replicate the experiments several times.

In chapter IV-VII we have presented a wide family of statistical methods based on diagonalization of matrices holding second-order statistics of the data. These methods serve many purposes. The use of spatial filters (chapter V) is straightforward and well established in EEG and many other research fields. They do not output data to be interpreted physiologically, so their use in practice does not pose critical questions. The use of blind source separation (BSS) is more cumbersome, since BSS is meant to provide physiologically plausible information, but much more arbitrariness is left to the researcher; for example in AJDC a different definition of the diagonalization set may lead to sources that are significantly different. Different normalization and weighting of the matrices in the diagonalization set also may give results that are quite different. So, the actual solution retained tends to be biased by the expectation of the researcher. As for inverse solutions, it is difficult to check the validity of the BSS output; the BSS will always output *something*, even if the assumptions are not respected at all. Still, when used carefully striving to respecting the assumptions of the BSS method and exploiting physiological knowledge about the data submitted to analysis, BSS is a very powerful tool allowing the study of spatial, temporal and frequential EEG dynamics with high precision. The good news in BSS methods based on approximate joint diagonalization (AJD) is that research on AJD algorithms is now mature and several algorithms with good performance are available. As compared to single-

subject BSS, group BSS and joint BSS makes use of even stronger assumptions and inherit all the difficulties of the BSS method, hence, while representing possibly the most powerful multivariate and multi-subject signal processing tool we know to date, even more care should be applied.

In chapter VIII, accompanied by the theoretical chapter IX, we have delineated a new universal strategy for brain-computer interface (BCI) classification based on Riemann geometry. We have proposed a very simple and fully automated classification framework, leaving to the researcher only the definition of the right form of the “covariance matrix” to be extracted from the data. Everything else is automatic and parameter-free. This is in contrast with the methods presented in previous chapters and represents a decisive advantage. We have shown that the framework is flexible and powerful, despite its simplicity. Riemann geometry has an important advantage over the other source analysis methods we have presented in this manuscript; both inverse solutions and diagonalization methods are very sensitive to noise and are useful only for high SNR data, whereas Riemann geometry is robust to noise. The drawback of Riemannian tools is that the data is difficult to interpret. We can visualize the geometric mean of covariance matrices and make scalp topographical maps (or source localization by inverse solutions) with that, but in this space (sensor space) the data is still mixed, so such analysis is not very useful; *while inverse solutions and BSS disclose the hidden variable, Riemann geometry keeps them hidden*. We have started to address this limitation with the introduction of the distance matrix in chapter IX. It should be kept in mind that Riemann geometry is proposed here as a tool for classification, but we are very far from being able to analyze explicitly the hidden variables. If the purpose of a study is the analysis of the source dynamics and localization, Riemann geometry is not, at least to date, well adapted. On the other hand Riemannian geometry is our method of choice for *classification* and *brain decoding*. Also, while inverse solutions and diagonalization methods are very sensitive to noise and are useful only on high SNR data, Riemann geometry is robust to noise and can be used purposefully also on low SNR data (e.g., real conditions).

In the next years we will continue our investigation in multi-variate multi-subject EEG analysis. While we have given in this manuscript several developments in this direction, much is left to do. We believe that the analysis of data recorded on several subject simultaneously, either in close contact or at a distance, is a topic that will acquire more and more importance in the years to come, both in neuroimaging research and in brain-computer interface. This will call for the construction, exploitation and maintenance of massive databases and multiple-client/server operations, paving the way for a new conception of neuroimaging and brain-computer interfaces. After all, the interaction with other individuals is a fundamental ingredient of our existence not only in our ontogeny, but also in our phylogeny. The exponential technological progress in mobility and communication of our society is

intensifying the occurrence of both face to face and distant interactions. Today the expressions of humanity in a natural or technological environment is conceived more and more as *social*, that is, embedded in a situated framework in which in order to achieve efficient communication the reciprocal understanding of the emotional and cognitive state is considered as of paramount importance. While neuroimaging studies have traditionally focused on the study of one individual while performing simple, often non-ecological, tasks in isolation, there is growing interest in analyzing groups of individuals simultaneously in ecological experimental settings. Starting with the new millennium, the field of social neuroscience has been defined, aiming at discovering those brain mechanisms supporting close coupling and attunement between the self and other (Hari and Kujala, 2009). There is no doubt that such understanding will be capital for the well being of human society. The analysis of the data gathered in these realistic situations will be the object of our further investigations.

REFERENCES

- Adolphs R (2006) How do we know the minds of others? Domain-specificity simulation, and enactive social cognition. *Brain Research*, 1079, 25-35.
- Adolphs R (2010) Conceptual Challenges and Directions for Social Neuroscience. *Neuron*, 65, 752-767.
- Ahn H, Prichep LS, John ER, Baird H, Trepetin M, Kaye H. (1980) Developmental Equations reflect Brain Dysfunctions, *Science*, 210, 1259-1262.
- Afsari B (2008) Sensitivity Analysis for the Problem of Matrix Joint Diagonalization. *SIAM J Matrix Anal & Appl*, 30(3), 1148–1171.
- Aissa-El-Bey A, Linh-Trung N, Abed-Meraim K, Belouchrani A, Grenier Y. (2007) Underdetermined Blind Separation of Nondisjoint Sources in the Time-Frequency Domain. *IEEE Trans Signal Process*, 55(3), 897-907.
- Allison BZ, Dunne S, Leeb R, Millán JdR, Nijholt A (2012) *Toward Practical Brain-Computer Interfaces* (Eds.), Springer, London.
- Amari S (1985) *Differential-Geometrical methods in Statistics*, Springer, Heidelberg.
- Anderson M, Adali T, Li X (2012) Joint Blind Source Separation With Multivariate Gaussian Model: Algorithms and Performance Analysis, *IEEE Trans Signal Process*, 60(4), 1672-1683.
- Anderson M, Fu G-S, Phlypo R, Adali T (2013) Independent Vector Analysis: Identification Conditions and Performance Bounds, [arXiv:1303.7474](https://arxiv.org/abs/1303.7474)
- Ando T, Li CK, Mathias R (2004) Geometric Means, *Linear Algebra Appl*, 385, 305-334.
- Ans B, Héroult J, Jutten C (1985) Adaptive Neural Architectures: Detection of Primitives. In : *Proc. COGNITIVA*, 593-597.
- Arrouët C, Congedo M, Marvie J-E, Lamarche F, Lécuyer A, Arnaldi B (2005) Open-ViBE: a 3D Platform for Real-Time Neuroscience, *Journal of Neurotherapy*, 9(1), 3-25 .
- Backus G, Gilbert F (1968) The resolving power of gross earth data, *Geophysical Journal of Royal Astronomic Society*, 16, 169-205.
- Baillet S, Mosher JC, Leahy RM (2001) Electromagnetic brain mapping, *IEEE Signal Processing Magazine*, 18(6), 14-30.
- Barachant A, Andreev A, Congedo M (2013) The Riemannian Potato: an automatic and adaptive artifact detection method for online experiments using Riemannian geometry, *TOBI Workshop IV*, Sion : Switzerland.
- Barachant A, Bonnet S, Congedo M, Jutten C (2012a) Multi-Class Brain Computer Interface Classification by Riemannian Geometry, *IEEE Transactions on Biomedical Engineering*, 59(4), 920-928.
- Barachant A, Bonnet S, Congedo M, Jutten C (2012b) BCI Signal Classification using a Riemannian-based kernel, *ESANN Conference*, Bruges, Belgium..
- Barachant A, Bonnet S, Congedo M, Jutten C (2011a) Réalisation d'un Brain-Switch EEG par Géométrie Riemannienne, *GRETSI conference*, Bordeaux, France.
- Barachant A., Bonnet S., Congedo M., Jutten C (2011b) A Brain-Switch Using Riemannian Geometry, *Proceedings of the 5th International BCI Conference*, Graz, Austria, 64-67.
- Barachant A., Bonnet S., Congedo M, Jutten C (2010a) Common Spatial Pattern revisited by Riemannian Geometry, *Proceedings of the IEEE International Workshop on Multimedia Signal Processing*, 472-476.

- Barachant A, Bonnet S, Congedo M, Jutten C. (2010b) Riemannian Geometry Applied to BCI Classification. Proceedings of Latent Variable Analysis and Signal Separation Conference, St. Malo, France, 6365, 629-636
- Barachant A, Congedo M (2013) A Universal BCI Classification Framework based on Riemannian Geometry, submitted.
- Barachant A, Congedo M (2014) A Robust and Efficient Method for Single Trial Classification of ERP Using Information Geometry. IEEE Trans Biom Eng, in press.
- Barachant A, Congedo M, Van Veen G, Jutten C (2013) Classification de potentiels évoqués P300 par géométrie riemannienne, GRETSI Proceedings, in press.
- Bauer H, Pllana A, Sailer U (2011) The EEG-based local brain activity (LBA-) feedback training. *Activitas Nervosa Superior Rediviva*, 53(3), 107-113.
- Bell AJ, Sejnowski TJ. (1995) An Information-Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Comput*, 7, 1129-1159.
- Belouchrani A, Abed-Meraim K, Cardoso J-F, Moulines E (1997) A blind source separation technique using second-order statistics. *IEEE Trans Signal Process* 1997, 45(2), 434-444.
- Belouchrani A, Amin MG (1998) Blind Source Separation Based on Time-Frequency Signal Representations. *IEEE Trans Signal Process*, 46(11), 2888-2897.
- Beran J. (1994) *Statistics for Long-Memory processes*. London: Chapman & Hall.
- Berger H (1929) Über das Elektroenkephalogram des Menschen. *Archives of Psychiatry*, 87, 527-70.
- Bhatia R (2003) On the exponential metric intrinsic property. *Linear Algebra and its Applications*, 375, 211-220.
- Bhatia R (2007) *Positive Definite Matrices*. Princeton University Press, New Jersey.
- Bhatia R (2013) The Riemannian Mean of Positive Matrices. Ch 2 in Nielsen F. and Bhatia R. (Eds.) *Matrix Information Geometry*, Springer, London.
- Bhatia R, Holbrook J (2006) Riemannian geometry and matrix geometric mean. *Linear Algebra and its applications*, 413, 594-618.
- Bhatia R, Karandikar RL (2011) The matrix geometric mean. Research Report isid/ms/2011/02. Indian Statistical Institute.
- Bin G, Gao X, Wang Y, Li Y, Hong B, Gao S (2011) A high-speed BCI based on code modulation VEP. *Journal of neural engineering*, 8(2), 025015.
- Bini DA, Iannazzo B (2011) A note on computing matrix geometric means. *Adv. Comput. Math.* 35(2-4), 175-192.
- Bini DA, Iannazzo B (2013) Computing the Karcher mean of symmetric positive definite matrices, *Linear Algebra Appl*, 438-4, 1700-1710.
- Blankertz B, Müller KR, Curio G, Vaughan TM, Schalk G, Wolpaw JR, Schlgl A, Neuper C, Pfurtscheller G, Hinterberger T, Schrder M, Birbaumer N (2004) The BCI Competition 2003: Progress and Perspectives in Detection and Discrimination of EEG Single Trials. *IEEE Trans on Biom Eng*, 51(6), 1044-1051.
- Bloomfield P (2000) *Fourier Analysis of Time Series*. New York: John Wiley & Sons.
- Bolton JP, Gross J, Liu CL, Ioannides AA (1999) SOFIA: spatially optimal fast initial analysis of biomagnetic signals, *Phys Med Biol*, 44(1), 87-103.

- Bonnet L, Lotte F, Lécuyer A (2013) Two Brains, One Game: Design and Evaluation of a Multi-User BCI Video Game Based on Motor Imagery, *IEEE Trans on Computational Intelligence and AI in Games*.
- Bosch-Bayard J, Valdés-Sosa P, Virues-Alba T, Aubert-Vázquez E, John ER, Harmony T et al. (2001) 3-D statistical parametric mapping of EEG source spectra by means of variable resolution electromagnetic tomography (VARETA), *Clin Electroencephalogr*, 32, 47–61.
- Bousbia-Salah A, Belouchrani A, Bousbia-Salah H (2003) A one step time-frequency blind identification. *7th Int Symp Sig Process Applications*, 1(1), 581- 584.
- Brett M, Anton J-L, Valabregue R, Poline J-B (2002) Region of interest analysis using an SPM toolbox [abstract] presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2–6, 2002, Sendai, Japan. Available on CD-ROM in *NeuroImage* 16(2).
- Buzsáki G (2006) *Rhythms of the Brain*. New York: Oxford Univ Press, 2006.
- Calhoun VD, Liu J, Adali T. (2009) A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data, *Neuroimage*, 45(1 Suppl), S163-72.
- Calhoun VD, Adali T, Pearlson GD, Pekar JJ (2001) A method for making group inferences from functional MRI data using independent component analysis. *Hum Brain Mapp*, 14, 140-151.
- Cao J, Murata N, Amari S-I, Cichocki A, Takeda T (2002) Independent component analysis for an averaged single-trial MEG data decomposition and single-dipole localization, *Neurocomputing*, 49, 255-277.
- Caplan JB, Madsen JR, Raghavachari S, Kahana MJ (2001). Distinct patterns of brain oscillations underlie to basic parameters of human maze learning. *Journal of Neurophysiology*, 86, 368-380.
- Cardoso J-F (1989) Source separation using higher order moments. In *Proc. IEEE ICASSP*, 4, 2109-2112.
- Cardoso J-F (1994) On the performance of orthogonal source separation algorithms. *Proc EUSIPCO*, Edinburg (UK) 1994, 776-779.
- Cardoso J-F (1998) Blind Signal Separation: Statistical Principles. *IEEE Proc*, 9(10), 2009-2025.
- Cardoso J-F (1999) High-Order Contrasts for Independent Component Analysis. *Neural Comput*, 11(1), 157-192.
- Cardoso J-F, Souloumiac A. (1993) Blind beamforming for non-Gaussian signals. *IEE Proc-F (Radar and Signal Process)*, 140(6), 362-370.
- Carlson BC and Keller M (1957) Orthogonalization Procedures and the Localization of Wannier Functions, *Physical review*, 105(1), 102-103.
- Caton R (1875) The electrical currents of the brain. *British Medical Journal*, 2, 278.
- Chabot RJ, di Michele F, Prichep L (2005) The role of quantitative electroencephalography in child and adolescent psychiatric disorders, *Child and adolescent psychiatric clinics of North America*, 14(1), 21-53.
- Chatel-Goldman J, Congedo M, Phlypo R (2013) Joint BSS as a natural analysis framework for EEG-hyperscanning *ICASSP*, Vancouver, Canada.
- Chebbi Z, Moakher M (2012) Means of Hermitian positive-definite matrices based on the log-determinant α -divergence function. *Linear Algebra and its Application*, 436, 1872-1889.
- Chen YS, Cheng CY, Hsieh JC, Chen LF (2006) Maximum contrast beamformer for electromagnetic mapping of brain activity, *IEEE Transactions on Biomedical Engineering*, 53(9), 1765-74.

- Choi K (2014) Electroencephalography (EEG)-based neurofeedback training for brain-computer interface (BCI). *Experimental Brain research*, in press.
- Choi S, Cichocki A (2000) Blind Separation of nonstationary sources in noisy mixtures. *Electron Lett*, 36, 848-849.
- Choi S, Cichocki A, Belouchrani (2002) Second Order Nonstationary Source Separation. *J VLSI Sig. Process*, 32(1-2), 93-104.
- Cichocki A, Amari S-I (2002) *Adaptive Blind Signal and Image Processing. Learning Algorithms and Applications*, John Wiley & Sons, New-York.
- Cichocki A, Georgiev P (2003). Blind Source Separation Algorithms with Matrix Constraints. *IEICE Transactions Fundamentals*, E86-A(1), 1-9, 2003.
- Colwell K, Throckmorton C, Collins L, Morton K (2013) Transfer Learning for Accelerated P300 Speller Classifier Training, *Proc. Fifth Int. BCI Meeting*, June 3-7, Pacific grove, California. ID: 004.
- Comon P (1994) Independent component analysis, A new concept? *Signal Processing* 36, 287-314.
- Comon P, Jutten C (2010) *Handbook of Blind Source Separation: Independent Component Analysis and Applications*, Academic Press, Oxford.
- Congedo M (2006) Subspace Projection Filters for Real-Time Brain Electromagnetic Imaging, *IEEE Transactions on Biomedical Engineering*, 53(8), 1624-34.
- Congedo M, Barachant A, Andreev A (2014) Riemannian Geometry Useful for Machine Learning: a Theoretical Prime, in press.
- Congedo M, Goyat M, Tarrin N, Varnet L, Rivet B, Ionescu G, et al. (2011) "Brain Invaders": a prototype of an open-source P300-based video game working with the OpenViBE platform. *Proc of the 5th Int BCI Conference*, Graz, Austria, 280-283
- Congedo M, Gouy-Pailler C, Jutten C (1998) On the blind source separation of human electroencephalogram by approximate joint diagonalization of second order statistics, *Clinical Neurophysiology* 119, 2677-2686.
- Congedo M, John ER, De Ridder D, Prichep L (2010) Group Independent Component Analysis of Resting-State EEG in Large Normative Samples, *International Journal of Psychophysiology* 78, 89-99.
- Congedo, Jutten, Rousseau (in press) An Introduction to EEG Source Analysis with an Illustration of a Study on Error-Related Potentials. In: *Guide to Brain-Computer Music Interfacing*, Miranda ER Castet J, Knapp B (Eds), Springer, London
- Congedo M, Lotte F, Lécuyer A. (2006) Classification of Movement Intention by Spatially Filtered Electromagnetic Inverse Solutions, *Physics in Medicine and Biology*, 51, 1971-1989.
- Congedo M, Lubar JF (2004) Parametric and Non-Parametric Analysis of QEEG: Normative Database Comparisons in Electroencephalography, a Simulation Study on Accuracy. In "Quantitative Electroencephalographic Analysis (QEEG) Databases for Neurotherapy. Description, Validation, and Application", Lubar JF (Ed), Haworth Press, New York
- Congedo M, Lubar JF, Joffe D (2004) Low-Resolution Electromagnetic Tomography neurofeedback, *IEEE Transactions on Neuronal Systems & Rehabilitation Engineering*, 12(4), 387-397.
- Congedo M, Ozen C, Sherlin L (2002). Notes on EEG Resampling by Natural Cubic Spline Interpolation, *Journal of Neurotherapy*, 6(4), 73-80.

- Congedo M, Pham D-T (2009) Least-squares joint diagonalization of a matrix set by a congruence transformation, Proceedings of the Singaporean-French IPAL Symposium.
- Congedo M, Phlypo R, Chatel-Goldman J (2012) Orthogonal and Non-Orthogonal Joint Blind Source Separation in the Least-Squares Sense, 20th European Signal Processing Conference (EUSIPCO), 1885-9.
- Congedo M, Phlypo R, Pham D-T (2011), Approximate joint singular value decomposition of an asymmetric rectangular matrix set, *IEEE Trans Signal Process*, 59(1), 415-424.
- Cooley TW, Tukey JW (1965) An algorithm for the machine computation of the complex Fourier series. *Mathematics of Computations*, 19, 297-301.
- Cornwell BR, Johnson LL, Holroyd T, Carver FW, Grillon C. (2008) Human hippocampal and parahippocampal theta during goal-directed spatial navigation predicts performance on a virtual Morris water maze. *Journal of Neuroscience*, 28, 5983-5990.
- Cornwell BR, Salvatore G, Colon-Rosario V, Latov DR, Holroyd T, Carver FW, et al. (2010) Abnormal hippocampal functioning and impaired spatial navigation in depressed individuals: Evidence from whole-head magnetoencephalography. *American Journal of Psychiatry*, 167, 836-844.
- Crespo-Garcia M, Atienza M, Cantero JL. (2008) Muscle artifact removal from human sleep EEG by using independent component analysis. *Ann Biomed Eng*, 36(3), 467-75.
- Darmois G (1953) Analyse générale des liaisons stochastiques. *Rev Inst Inter Stat*, 21, 2-8.
- Dégerine S, Kane E (2007) A Comparative Study of Approximate Joint Diagonalization Algorithms for Blind Source Separation in Presence of Additive Noise. *IEEE Trans Signal Process*, 55 (6-2), 3022-3031.
- De Jaegher H, Di Paolo E, Gallagher S (2010) Can social interaction constitute social cognition?, *Trends in Cognitive Sciences*, 14(1), 441-447.
- Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods*, 134(1), 9-21.
- De Ridder D, Vanneste S, Congedo M (2011) The distressed brain: a group blind source separation analysis on tinnitus. *PLoS One*, 6(10), e24273.
- Duffy FH, Bartels PH, Burchfiel JL (1981) Significance Probability Mapping: An Aid in the Topographic Analysis of Brain Electrical Activity, *Electroencephalography and clinical Neurophysiology*, 51, 455-462.
- Eichele T, Rachakonda S, Brakedal B, Eikeland R, Calhoun VD (2011) EEGIFT: Group Independent Component Analysis for Event-Related EEG Data. *Comput Intell Neurosci*, 129365.
- Ekstrom AD, Caplan JB, Ho E, Shattuck K, Fried I, Kahana MJ (2005). Human hippocampal theta activity during virtual navigation. *Hippocampus*, 15, 881-889.
- Fadaili EM, Moreau NT, Moreau E (2007) Nonorthogonal Joint Diagonalization/Zero Diagonalization for Source Separation Based on Time-Frequency Distributions. *IEEE Trans Signal Process*, 55(5-1), 1673-1687.
- Falkenstein M, Hoormann J, Christ S, Hohnsbein J (2000) ERP components on reaction errors and their functional significance: a tutorial. *Biol Psychol*, 51(2-3), 87-107.
- Farquhar J, Hill NJ (2013) Interactions between pre-processing and classification methods for event-related-potential classification : best-practice guidelines for brain-computer interfacing. *Neuroinformatics*, 11(2), 175-92.
- Farwell LA, Donchin E (1988) Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials *Electroenceph. Clin. Neurophysiol.*, 70, 510-23

- Faugeras O, Clément F, Deriche R, Keriven R, Papadopoulo T, Roberts J, et al. (1999) The inverse EEG and MEG problems: The adjoint state approach I: The continuous case. Research report # 3673, National Institute for Research in Informatics and Control (INRIA), Sophia-Antipolis, France.
- Féty L, Uffelen J-P (1988) New Methods for Signal Separation. Proc. of the 14th conf. on HF Radio System and Techniques, London, 226-230.
- Fillard P, Arsigny V, Ayache N, Pennec X (2005) A Riemannian Framework for the Processing of Tensor-Valued Images. DSSCV, 112-123
- Fitzgibbon SP, Powers DM, Pope KJ, Clark CR (2007) Removal of EEG noise and artifact using blind source separation. *J Clin Neurophysiol*, 24(3), 232-243.
- Flury BN, Gautschi WG (1986) An algorithm for orthogonal transformation of several positive definite symmetric matrices to nearly diagonal form, *SIAM J Sci Stat Comp*, 7(1), 169-184.
- Förstner W, Moonen B (1999) A metric for covariance matrices. In Krumm K and Schwarze VS eds. *Qho vadis geodesia...?*, number 1999.6 in tech. report of the Dep. Of Geodesy and Geoinformatics, p. 113-128, Stuttgart University.
- Frank RM, Frishkoff GA (2007) Automated protocol for evaluation of electromagnetic component separation (APECS): Application of a framework for evaluating statistical methods of blink extraction from multichannel EEG. *Clin Neurophysiol*. 118(1), 80-97.
- Frigo M, Johnson SG (2005) The Design and Implementation of FFTW3. *Proc IEEE*, 93(2), 216-231.
- Frith CD, Frith U (1999) Interacting Minds-A Biological Basis, *Science*, 286(5445),1692-1695.
- Frith U, Frith CD (2010) The social brain: allowing humans to boldly go where no other species has been, *Phil. Trans. R. Soc. B*, 365,165-176
- Fuchs M, Kastner J, Wagner M, Hawes S, Ebersole JS (2002) A standardized boundary element method volume conductor model. *Clin Neurophysiol*, 113, 702-12.
- Fukunaga K (1990) Introduction to Statistical Pattern Recognition (2nd Ed.), Academic Press, London.
- Gehring WJ, Goss B, Coles MGH, Meyer DE, Donchin E (1993) A neural system for error detection and compensation. *Psychol. Sci.*, 4(Suppl 6), 385-390.
- Gehring WJ, Willoughby AR (2002) The medial frontal cortex and the rapid processing of monetary gains and losses. *Science*, 295(5563), 2279- 2282.
- Gentsch A, Ullsperger P, Ullsperger M (2009) Dissociable medial frontal negativities from a common monitoring system for self-and externally caused failure of goal achievement. *Neuroimage*, 47(4), 2023-2030.
- Goh A, and Vidal R (2008) Unsupervised Riemannian Clustering of probability density functions. In W. Daelemans et al. (Eds.): *ECML PKDD 2008, Part I, LNAI 5211*, 377-392.
- Golub GH, Reinsch C, (1970) Singular Value Decomposition and Least Squares Solutions, *Numer. Math*, 14, 403-420.
- Golub GH, van Loan CF (1996) *Matrix Computations*, The Johns Hopkins University Press (3rd Ed.), Baltimore.
- Goncalves SI, de Munck JC, Pouwels PJW, Schoonhoven R, Kujer JPA, Maurits NM, et al. (2006) Correlating the alpha rhythm to BOLD using simultaneous EEG/fMRI: inter-subject variability. *Neuroimage*, 30, 203-213.
- Good IJ (1969) Some Applications of the Singular Decomposition of a Matrix. *Technometrics*, 11(4), 823-831.

- Gouy-Pailler C, Congedo M, Brunner C, Jutten C, Pfurtscheller G (2010) Nonstationary brain source separation for multiclass motor imagery, *IEEE Transactions on Biomedical Engineering* 57(2), 469-78.
- Greenblatt RE, Ossadtchi A, Pflieger ME (2005) Local Linear Estimators for the Bioelectromagnetic Inverse Problem. *IEEE Trans Signal Process*, 53(9), 3403-3412.
- Gribonval R, Lesage S, (2006) A survey of Sparse Component Analysis for blind source separation: principles, perspectives, and new challenges. *Proc. of Eur Symp Artif Neural Netw (ESANN 2006)*, 323-330.
- Gross J, Ioannides AA (1999) Linear Transformation of data space in MEG, *Phys Med Biol*, 44(1), 87-103.
- Grosse-Wentrup M, Buss M (2008) Multiclass Common Spatial Patterns and Information Theoretic Feature Extraction, *IEEE Transactions on Biomedical Engineering*, 55(8), 1991-2000.
- Guger C, Ramoser H, Pfurtscheller G (2000) Real-Time EEG Analysis with Subject-Specific Spatial Patterns for a Brain-Computer Interface. *IEEE Transactions on Rehabilitation Engineering*, 8(4), 447-456.
- Guimaraes MP, Wong DK, Uy ET, Grosenick L, Suppes P. (2007) Single-trial classification of MEG recordings. *IEEE Trans Biomed Eng*, 54(3), 436-443.
- Halder S, Bensch M, Mellinger J, Bogdan M, Kübler A, Birbaumer N, et al. (2007) Online artifact removal for brain-computer interfaces using support vector machines and blind source separation. *Comput Intell Neurosci*: 82069.
- Hämäläinen MS, Ilmoniemi RJ (1984) Interpreting measured magnetic fields of the brain: estimates of current distributions. *Tech. Rep. TKK-F-A559*, Helsinki University of Technology, Espoo.
- Hari R, Kujala MV (2009) Brain Basis of Human Social Interaction: From Concepts to Brain Imaging. *Physiol Rev*, 89, 453-479.
- Hérault J, Jutten C (1986) Space or time adaptive signal processing by neural network models. *Proc Int Conf Neural Netw Computing, Snowbird (Utah)*, 151, 206-211.
- Herrmann MJ, Rommler J, Ehlis AC, Heidrich A, Fallgatter AJ (2004) Source localization (LORETA) of the error-related-negativity (ERN/Ne) and positivity (Pe). *Cognitive Brain Research*, 20(2), 294-299.
- Hernández JL, Valdés P, Biscay R, Virues T, Szava S, Bosch J, et al. (1994) A global scale factor in brain. *Int J Neurosci* 76, 267-278.
- Herweg A, Kaufmann T, Kübler A (2013) Using Generic Models to Improve Tactile ERO-BCI performance of Low Aptitude Users. *Proc. Fifth Int. BCI Meeting, June 3-7, Pacific grove, California*. ID: 097.
- Hjorth B (1991) Principles for transformation of scalp EEG from potential field into source distribution, *Journal of Clinical Neurophysiology*, 8(4), 391-6.
- Holmes AP, Blair RC, Watson JDG, Ford I (1996) Non-Parametric Analysis of Statistic Images From Functional Mapping Experiments. *Journal of Cerebral Blood Flow and Metabolism*, 16, 7-22.
- Hotelling H (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24, 417-441, and 498-520.
- Hotelling H (1936). Relations between two sets of variates. *Biometrika*, 27, 321-77.
- Hughes, M.D., John, E.R. (1999) Conventional and Quantitative Electroencephalography in Psychiatry, *Journal of Neuropsychiatry and Clinical Neuroscience*, 11, 190-208.
- Hyvärinen A (1999) Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans Neural Netw* 1999, 10(3), 626-634.

- Hyvärinen A, Karhunen J, Oja E (2001) *Independent Component Analysis*. New York: John Wiley & Sons.
- Iferroudjene R, Abed-Meraim K, Belouchrani A (2009) A new jabobi-like method for joint diagonalization of arbitrary non-defective matrices, *Applied Mathematics and Computation*, 211, 363–373.
- Ille N, Berg P, Scherg M (2002) Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies. *J Clin Neurophysiol*, 19(2), 113-124.
- Im CH, Hwang HJ, Che H, Lee S (2007) An EEG-based real-time cortical rhythmic activity monitoring system, *Physiol Meas*, 28(9),1101-13.
- Iriarte J, Urrestarazu E, Valencia M, Alegre M, Malanda A, Viteri C, Artieda J (2003) Independent component analysis as a tool to eliminate artifacts in EEG: a quantitative study. *J Clin Neurophysiol*, 20(4): 249-257.
- Jacobs J, Korolev IO, Caplan JB, Ekstrom AD, Litt B, Baltuch G, et al. (2010) Right-lateralized brain oscillations in human spatial navigation. *Journal of Cognitive Neuroscience*, 22, 824-836.
- Jasper HH (1958) Report of the Committee on Methods of Clinical Examination in Electroencephalography, *Electroencephalography and Clinical Neurophysiology*, 10, 370-1.
- Jeuris B, Vanderbril R, Vandereycken B (2012) A survey and comparison of contemporary algorithms for computing the matrix geometric mean. *Electronic Transactions on Numerical Analysis*, 39, 379-402.
- Jia C, Gao X, Hong B, Gao S (2011) Frequency and phase mixed coding in SSVEP-based brain-computer interface. *IEEE Trans Biomed Eng*, 58(1),200-206.
- Jin J, Allison BZ, Sellers EW, Brunner C, Horki P, Wang X, Neuper C. (2011) Optimized stimulus presentation patterns for an event-related potential EEG based brain-computer interface. *Med Biol Eng Comput*, 49(2),181-91.
- Jin J, Sellers EW, Zhang Y, Daly I, Wang X, Cichocki A (2012) Whether generic model works for rapid ERP-based BCI calibration. *J Neurosci Meth*, 212, 94-99.
- John ER, Ahn H, Prichep LS, Trepetin M, Brown D, Kaye H. (1980a). Developmental equations for the electroencephalogram. *Science* 210. 1255-1258.
- John ER, Karmel BZ, Corning WC, Easton P, Brown D, Ahn, H, et al. (1980b) *Neurometrics*, *Science*, 196, 1393-1409.
- John ER, Karmel BZ, Corning WC, Easton P, Brown D, Ahn H, et al. (1980c). *Neurometrics*. *Science* 196, 1393-1409.
- John ER, Prichep LS, Easton P (1987) Normative Data Banks and Neurometrics. Basic Concepts, Method and Results of Norm Constructions.in *Method of Analysis of brain Electrical and Magnetic Signals. EEG Handbook* (revised series. Vol. 1). (Gevins, A. S., and Remond, A. Ed.). Elsevier Science Publishers B.V. (Biomedical Division).
- Joyce CA, Gorodnitsky IF, Kutas M (2004) Automatic removal of eye movement and blink artifacts from EEG data using blind component separation. *Psychophysiology*, 41(2), 313-25.
- Jung TP, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski TJ (2000) Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clin Neurophysiol*, 111(10), 1745-58.
- Jurcak V, Tsuzuki D, Dan I (2007) 10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems. *Neuroimage*, 34(4), 1600-1611.

- Jutten C, Herault J (1991) Blind separation of sources, Part 1: an adaptive algorithm based on neuromimetic architecture. *Signal Process*, 24(1), 1-10.
- Kachenoura A, Albera L, Senhadji L, Comon P (2008) ICA: a potential tool for BCI systems. *IEEE signal Process Mag*, 25(1), 57-68.
- Kaufmann T, Völker S, Gunesch L, Kübler A (2012) Spelling is just a click away – a user-centered brain-computer interface including auto-calibration and predictive text entry. *Frontiers in Neuroscience*, 6(72), 1-10
- Kierkels JJM, van Boxtel GJM, Vogten LLM (2006) A model-based objective evaluation of eye movement correction in EEG recordings. *IEEE Trans Biomed Eng*, 53(2), 246-253.
- Kindermans P-J, Schrauwen B (2013) Dynamic Stopping in a Calibration-less P300 Speller (2013) Proc. Fifth Int. BCI Meeting, June 3-7, Pacific grove, California. ID: 075.
- Kindermans P-J, Verstraeten D, Schrauwen B (2012) A Bayesian Model for Exploiting Application Constraints to Enable Unsupervised Training of a P300-based BCI. *PLoS ONE*, 7(4), e33758.
- Koles ZJ (1991) The Quantitative extraction and Topographic Mapping of the Abnormal Components in the Clinical EEG. *Electroencephalography and Clinical Neurophysiology*, 79, 440-447.
- Koles ZJ, Soong A (1998) EEG Source Localization: Implementing the Spatio-Temporal Decomposition Approach. *Electroencephalography and Clinical Neurophysiology*, 107, 343-352.
- Kopřivová J, Congedo M, Raszka M, Praško J, Brunovský M, Horáček J (2013) Prediction of Treatment Response and the Effect of Independent Component Neurofeedback in Obsessive-Compulsive Disorder: A Randomized, Sham-Controlled, Double-Blind Study, *Neuropsychobiology*, 67:210-223.
- Kopřivová J, Congedo M, Horáček J, Praško J, Raszka M, Brunovský M, et al. (2011) EEG source analysis in obsessive-compulsive disorder, *Clinical Neurophysiology* 122(9), 1735-1743.
- Kübler A, Kotchoubey B, Kaiser J, Wolpaw JR, Birbaumer N (2001) Brain-computer communication: unlocking the locked in. *Psycholog Bull*, 127(3), 358-75.
- Lancaster JL, Woldor MG, Parsons LM, Liotti M, Freitas CS, Rainey L, et al. (2000) Automated talairach atlas labels for functional brain mapping. *Human brain mapping*, 10(3), 120-131.
- Ledoit O, Wolf M (2004) A well-conditioned estimator for large-dimensional covariance matrices. *J Multivar Anal*, 88, 365–411.
- Lee JS, Park KS (2013) A New Stimulation Method of Virtual Speller for Simultaneous P300 and SSVEP Responses. Proc. Fifth Int. BCI Meeting, June 3-7, Pacific grove, California. ID: 162
- Lemm S, Curio G, Hlushchuk Y, Müller K-R (2006) Enhancing the Signal-to-Noise Ratio of ICA-Based Extracted ERPs. *IEEE Trans Biomed Eng*, 53(4), 601-607.
- Lemos MS, Fisch BJ (1991) The weighted average reference montage, *Electroencephalography and Clinical Neurophysiology*, 79(5), 361-70.
- Li X-L, Adali T, Anderson M (2011), Joint blind source separation by generalized joint diagonalization of cumulant matrices, *Signal Process*, 91(10), 2314-2322.
- Li X-L, Zhang X-D (2007) Nonorthogonal Joint Diagonalization Free of Degenerate Solution. *IEEE Trans Sig Proc*, 55(5), 1803-1814.
- Li Y, Cichocki A, Amari S-I (2006) Blind estimation of channel parameters and source components for EEG signals: a sparse factorization approach. *IEEE Trans Neural Netw*, 17(2), 419-431.

- Li Y, Wong KM, De Bruin H (2009) EEG signal classification based on a Riemannian distance measure, *Proc TIC-STH*, 268 – 273.
- Li Y, Wong KM, De Bruin H (2012) EEG signals classification for sleep-state decision – A Riemannian geometry approach, *IET Signal Processing*, 6(4), 288–299.
- Li Y-O, Adali T, Wang W, Calhoun VD (2009), Joint blind source separation by multi-set canonical correlation analysis, *IEEE Trans. Signal Process*, 57(10), 3918-29.
- Liechti MD, Maurizio S, Heinrich H, Jäncke L, Meier L, Steinhausen H-C, et al. (2012) First clinical trial of tomographic neurofeedback in attention-deficit/hyperactivity disorder: Evaluation of voluntary cortical control, *Clinical Neurophysiology*, 123(10), 1989–2005.
- Lopes da Silva FH (2004) Functional Localization of Brain Sources using EEG and/or MEG data: Volume Conductor and Source Models. *Magn Res Img*, 22, 1533-1538.
- Lopes da Silva FH (2005a) Computer-Assisted EEG diagnosis: Pattern Recognition and Brain Mapping. In: *Electroencephalography. Basic Principles, Clinical Applications, and Related Fields*. Niedermeyer E and Lopes da Silva FH (Eds), 5th ed., New York: Lippincott Williams & Wilkins, 1233-1263.
- Lopes da Silva FH (2005b) Event Related Potentials: Methodology and Quantification. In: *Electroencephalography. Basic Principles, Clinical Applications, and Related Fields*. Niedermeyer E and Lopes da Silva FH (Eds), 5th ed., New York: Lippincott Williams & Wilkins, 991-1001.
- Lopes da Silva FH, Storm van Leeuwen W (1977) The cortical source of the alpha rhythm. *Neurosci Lett* 6, 237-41.
- Lopes da Silva FH, Van Rotterdam A (2005), Biophysical Aspects of EEG and Magnetoencephalogram Generation. In: *Electroencephalography. Basic Principles, Clinical Applications, and Related Fields*. Niedermeyer E and Lopes da Silva FH (Eds), 5th ed., New York: Lippincott Williams & Wilkins, 107-125.
- Lotte F, Congedo M, Lécuyer A, Lamarche F, Arnaldi B (2007), A review of classification algorithms for EEG-based brain-computer interfaces, *J Neural Eng*, 4(2), pp. R1–R13.
- Lotte F, Guan CT (2011) Regularizing Common Spatial Patterns to Improve BCI Designs: Unified Theory and New Algorithms, *IEEE Transactions on Biomedical Engineering*, 58(2), 355-362.
- Lubar JF, Congedo M, Askew JH (2003) Low-Resolution Electromagnetic Tomography (LORETA) of Cerebral Activity in Chronic Depressive Disorder, *International Journal of Psychophysiology*, 49, 175-185.
- Mainsah BO, Collins LM, Colwell K, Throckmorton CS (2013) Improving Dynamic Data Collection in P300 Spellers With a Language Model. *Proc. Fifth Int. BCI Meeting*, June 3-7, Pacific grove, California. ID: 107.
- Malmivuo J, Plonsey R (1995) *Bioelectromagnetism. Principles and Applications of Bioelectric and Biomagnetic Fields*. New York: Oxford Univ Press, New York.
- Manton JH (2004) A globally convergent numerical algorithm for computing the centre of mass on compact Lie groups. *ICARCV Conference proceeding*, 2211-2216.
- Matsuoka K, Ohya M, Kawamoto M (1995) A neural net for blind separation of nonstationary signals. *Neural Netw*, 8(3), 411-419.
- Mazziotta J, Toga A, Evans A, Fox P, Lancaster J, Zilles K, et al. (2001) A probabilistic atlas and reference system for the human brain: International consortium for brain mapping (icbm). *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 356(1412), 1293-1322.
- Meinecke F, Ziehe A, Kawanabe M, Müller KR. (2002) A resampling approach to estimate the stability of one- or multidimensional independent components. *IEEE Trans Biomed Eng*, 49, 1514-1525.

- Mesloub A, Abeb-Meraim K, Belouchrani A (2013) A new algorithm for complex non orthogonal joint diagonalization based on Shear and Givens rotations, arXiv:1306.0331.
- Miltner WHR., Braun CH, Coles MGH (1997) Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a generic neural system for error detection. *Journal of Cognitive Neuroscience*, 9(6),788-798.
- Moakher M (2005) A differential geometric approach to the arithmetic and geometric means of operators in some symmetric spaces. *SIAM. J. Matrix Anal. Appl.*, 26 (3), 735-747.
- Moakher M (2012) Divergence measures and means of symmetric positive-definite matrices, Ch. 16 in: *New Developments in the Visualization and Processing of Tensor Fields*, D. Laidlaw and A. Vilanova, eds., Springer, 307–321.
- Moakher M, Batchelor PG (2006) Symmetric positive-definite matrices: From geometry to applications and visualization. *Visualization and Processing of Tensor Fields*, 285-298
- Molgedey L, Schuster HG (1994) Separation of a Mixture of Independent Signals using Time Delayed Correlations. *Phys Rev Lett*, 72, 3634-3636.
- Mosher JC, Lewis PS, Leahy RM (1992) Multiple dipole modeling and localization from spatio-temporal MEGdata. *IEEE Trans Biomed Eng*, 39(6), 541-557.
- Mueller A, Candrian G, Grane VA, Kropotov JD, Ponomarev VA, Baschera G-M (2011) Discriminating between ADHD adults and controls using independent ERP components and a support vector machine: a validation study. *Nonlinear Biomed Phys*, 5, 5.
- Müller KR, Vigario R, Meinecke F, and Ziehe A (2004) Blind source separation techniques for decomposing event-related brain signals. *Int J Bifurcat Chaos*, 14(2), 773-791.
- Nakamura N (2009) Geometric means of Positive Operators, *KYUGPOOK Math J*, 167-181.
- Niedermeyer E (2005a) The Normal EEG of the waking Adult. In: *Electroencephalography. Basic Principles, Clinical Applications, and Related Fields*. Niedermeyer E and Lopes da Silva FH (Eds), 5th ed., New York: Lippincott Williams & Wilkins, 167-191.
- Niedermeyer E (2005b) Sleep and EEG. In: *Electroencephalography. Basic Principles, Clinical Applications, and Related Fields*. Niedermeyer E and Lopes da Silva FH (Eds), 5th ed., New York: Lippincott Williams & Wilkins, 193-207.
- Niedermeyer E (2005c) Epileptic Seizure Disorders. In: *Electroencephalography. Basic Principles, Clinical Applications, and Related Fields*. Niedermeyer E and Lopes da Silva FH (Eds), 5th ed., New York: Lippincott Williams & Wilkins, 505-619.
- Nieuwenhuis S, Yeung N, Van Den Wildenberg W, Ridderinkhof KR (2003) Electrophysiological correlates of anterior cingulate function in a go/no-go task: Effects of response convict and trial type frequency. *Cognitive, Affective, & Behavioral Neuroscience*, 3(1), 17-26..
- Nunez PL (1995) *Neocortical Dynamics and Human EEG Rhythms*. Oxford University Press.
- Nunez PL, Srinivasan R (2006) *Electric Field of the Brain*, 2nd ed., New York: Oxford Univ Press.
- Nunez PL, Pilgreen KL (1991) The spline-Laplacian in clinical neurophysiology: a method to improve EEG spatial resolution, *Journal of Clinical Neurophysiology*, 8(4), 397-413.
- Nunez PL, Silberstein RB (2000) On the relationship of Synaptic Activity to Macroscopic Measuremets: Does Co-Registration of EEG with fMRI Make sense?, *Brain Topography*, 13(2), 79-96.

- Nunez PL, Wingeier BM, Silberstein RB (2001) Spatial-temporal structures of human alpha rhythms: theory, microcurrent sources, multiscale measurements, and global binding of local networks, *Human Brain Mapping*, 13(3), 125-64.
- Obeid I, Picone J. (2013) Bringing Big Data to neural Interfaces. Proc. Fifth Int. BCI Meeting, June 3-7, Pacific grove, California. ID: 180
- Oostenveld R, Fries P, Maris E, Schoffelen J-M (2011) Fieldtrip: open source software for advanced analysis of meg, eeg, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, ID 156869.
- Panicker RC, Puthusserypady S , Sun Y (2010) Adaptation in P300 Brain-Computer Interfaces: A Two-Classifer Co-Training Approach”, *IEEE Tran Biomed Eng*, 57(12), 2927-35.
- Parra L, Sajda P (2003) Blind Source Separation via Generalized Eigenvalue Decomposition, *J Mach Learn Res*, 4, 1261-1269.
- Pascual-Marqui RD (1999) Review of methods for solving the EEG inverse problem, *Int J Bioelectromagn*, 1(1), 75–86.
- Pascual-Marqui RD (2001) LORETA-Key software package, freeware of the Key Institute for Brain-Mind research, Zurich, Switzerland. Available at <http://www.unizh.ch/keyinst/NewLORETA/LORETA01.htm>.
- Pascual-Marqui RD (2002) Standardized Low Resolution brain electromagnetic Tomography (sLORETA): technical details, *Methods Findings in Experimental Clinical Pharmacology*, 24(D), 5-12.
- Pascual-Marqui RD (2007) Discrete, 3D distributed, linear imaging methods of electric neuronal activity. Part 1: exact, zero error localization, *arXiv:0710.3341v2*.
- Pascual-Marqui RD, Esslen M, Kochi K, Lehmann D (2002) Functional imaging with low resolution brain electromagnetic tomography (LORETA): A review, *Meth. Findings Exp Clin Pharmacol*, 24C, 91–95.
- Pascual-Marqui RD, Michel CM, Lehmann D (1994) Low Resolution Electromagnetic Tomography: a New Method for Localizing Electrical Activity in the Brain, *International Journal of Psychophysiology*, 18, 49-65.
- Pearson K. (1901) On Lines and Planes of Closest Fit to Systems of Points in Space. *Philosophical Magazine* 2 (11), 559–572.
- Penfield W, Rasmussen T (1950) *The Cerebral Cortex of Man: A Clinical Study of Localization of Function*, Macmillan, New York.
- Pennec X, Fillard P, Ayache N (2004) A Riemannian Framework for Tensor Computing. Research Report #5255, INRIA, Sophie-Antipolis, France.
- Pfurtscheller G, Brunner C, Schlögl A, Lopes da Silva FH (2006), Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks. *Neuroimage*, 31, (1), 153–159.
- Pfurtscheller G, Lopes da Silva FH (1999) Event-related EEG/MEG synchronization and desynchronization: basic principles, *Clinical Neurophysiology*, 110(11), 1842-57.
- Pfurtscheller G, Neuper C (2001) Motor imagery and direct braincomputer communication, *Proc. IEEE*, 89(7), 1123–1134.
- Pham D-T (2001a) Blind Separation of Instantaneous Mixture of Sources via the Gaussian Mutual Information Criterion. *Signal Process*, 81, 855-870.
- Pham D-T (2001b) Joint Approximate Diagonalization of Positive Definite Matrices. *SIAM J. on Matrix Anal and Appl*, 22(4), 1136-1152.

- Pham D-T (2002) Exploiting source non stationary and coloration in blind source separation. Proceedings of the 14th conference on Digital Signal Processing, 1, 151-154.
- Pham D-T, Cardoso J-F. (2001) Blind Separation of Instantaneous Mixtures of Non Stationary Sources. IEEE Trans Signal Process, 49(9), 1837-1848.
- Pham D-T, Congedo M (2009) Least square joint diagonalization of matrices under an intrinsic scale constraint, Proceedings of the 8th ICA Int. Conf., 298–305.
- Phlypo R, Boon P, D'Asseler Y, Lemahieu I (2007) Removing ocular movement artefacts by a joint smoothed subspace estimator. Comput Intell Neurosci, ID 75079.
- Poloni F (2010) Constructing matrix geometric means, Electron. J. Linear Algebra, 20, 419–435.
- Ponomarev VA, Mueller A, Candrian G, Grin-Yatsenko VA, Kropotov JD (2013), Group Independent Component Analysis (gICA) and Current Source Density (CSD) in the study of EEG in ADHD adults, Clinical neurophysiology, in press.
- Raichle ME, Snyder AZ (2007) A default mode of brain function: A brief history of an evolving idea. Neuroimage 37, 1083-1090.
- Ramoser H, Muller-Gerking J, Pfurtscheller G (2000). Optimal Spatial Filtering of single trial EEG during Imagined Hand Movement. IEEE Transactions on Rehabilitation Engineering, 8(4), 441-446.
- Rao CR (1945) Information and accuracy attainable in the estimation of statistical parameters, Bull Calcutta Math Soc, 37, 81-89.
- Renard Y, Lotte F, Gibert G, Congedo M, Maby E, Delannoy V, et al. (2010) OpenViBE: An Open-Source Software Platform to Design, Test and Use Brain-Computer Interfaces in Real and Virtual Environments. Presence : teleoperators and virtual environments, 19(1), 35-53.
- Rivet B, Souloumiac A, Attina V, Gibert G. (2009) xDAWN algorithm to enhance evoked potentials: application to brain-computer interface. IEEE Transactions on Biomedical Engineering, 56, 8, 2035-43.
- Rivet B, Cecotti H, Souloumiac A, Maby E, Mattout J (2011) Theoretical analysis of xDAWN algorithm: application to an efficient sensor selection in a P300 BCI, EUSIPCO proceedings, 8-29
- Robinson SE, Vrba J (1999) Functional neuroimaging by Synthetic Aperture Magnetometry (SAM), in Recent Advances in Biomagnetism, T.Yoshimoto et al. (eds), Tohoku Univ. Press, Sendai, Japan, 302-305.
- Rodríguez-Rivera A, Baryshnikov BV, Van Veen BD, Wakai RT (2006) MEG and EEG Source Localization in Beamspace. IEEE Trans Biomed Eng, 53(3), 430-441.
- Romero S, Mañanas MA, Barbanoj MJ (2008) A comparative study of automatic techniques for ocular artifact reduction in spontaneous EEG signals based on clinical target variables: A simulation case. Comput Biol Med, 38(3), 348-360.
- Salari N, Büchel C, Rose M (2012) Functional Dissociation of Ongoing Oscillatory Brain States. PLoS ONE 7(5), e38090.
- Sander TH, Burghoff M, Curio G, Trahms L (2005) Single Evoked Somatosensory MEG Responses Extracted by Time Delayed Decorrelation. IEEE Trans Signal process, 53(9), 3384-3392.
- Särelä J, Vigário R (2003) Overlearning in Marginal Distribution-Based ICA: Analysis and Solutions. JMach Learn Res, 4, 1447-1469.
- Sarvas J (1987) Basic Mathematical and Electromagnetic Concepts of the Biomagnetic Inverse Problem. Phys Med Biol, 32(1), 11-22.

- Schettini F, Aloise F, Aricò P, Salinari S, Di Mattia D, Cincotti F (2013) Self-Calibration in an Asynchronous P300-Based BCI. Proc. Fifth Int. BCI Meeting, June 3-7, Pacific grove, California. ID: 124.
- Schmithorst VJ, Holland SK (2004) Comparison of three methods for generating group statistical inferences from independent component analysis of functional magnetic resonance imaging data. *Magn Reson Imaging* 19(3), 365-368.
- Schott JR (1997) *Matrix Analysis for statistics*, John Wiley & Sons, New York.
- Schultze-Kraft, Görgen K, Wenzel M, Haynes J-D, Blankertz B (2013) Cooperating Brains: Joint Control of a Dual-BCI. Proc. Fifth Int. BCI Meeting, June 3-7, Pacific grove, California. ID: 046.
- Searle SR (1982) *Matrix Algebra Useful for Statistics*, John Wiley & Sons, New York.
- Sekihara K, Sahani M, Nagarajan SS (2005) Localization Bias and Spatial Resolution of Adaptive and non-Adaptive Spatial Filters for MEG Source Reconstruction, *Neuroimage*, 25(4), 1056-67.
- Sekihara K, Nagarajan SS, Poeppel D, Marantz A (2004), Asymptotic SNR of Scalar and Vector Minimum-Variance Beamformers for Neuromagnetic Source Reconstruction, *IEEE Trans Biomed Eng*, 51(10), 1726-1734.
- Serby H, Yom-Tov E, Inbar GF, (2005) An improved P300-Brain-Computer Interface. *IEEE Trans Neural Syst Rehabil Eng*, 13(1), 89-98.
- Sharbrough F, Chatrian G-E, Lesser RP, Lüders H, Nuwer M, Picton TW (1991) American Electroencephalographic Society Guidelines for Standard Electrode Position Nomenclature, *Journal of Clinical Neurophysiology*, 8, 200-2.
- Sherlin L, Budzynski T, Kogan-Budzynski H, Congedo M, Fischer ME, Buchwald D (2007) Low-resolution electromagnetic brain tomography (LORETA) of monozygotic twins discordant for chronic fatigue syndrome, *Neuroimage*, 34(4), 1438-1442.
- Sherlin L, Congedo M (2005) Obsessive Compulsive Dimension Localized using Low Resolution Electromagnetic Tomography (LORETA), *Neuroscience Letters*, 387(2), 72-74.
- Skovgaard L (1984) A Riemannian geometry of the multivariate normal model. *Scand J Statistics*, 11, 211-223.
- Souloumiac A (1995) Blind Source Detection and separation using second order nonstationarity. In Proc ICASSP, 1912-1915.
- Souloumiac A (2009), Nonorthogonal joint diagonalization by combining givens and hyperbolic rotations, *IEEE Trans. Signal Process.*, 57(6), 2222–2231.
- Souloumiac A (2011) A Stable and Efficient Algorithm for Difficult Non-Orthogonal Joint Diagonalization Problems, *EUSIPCO proceedings*.
- Speckmann E-J, Elger CE (2005) Introduction to the Neurophysiological Basis of the EEG and DC Potentials. In: *Electroencephalography. Basic Principles, Clinical Applications, and Related Fields*. Niedermeyer E and Lopes da Silva FH (Eds), 5th ed., New York: Lippincott Williams & Wilkins, 17-29.
- Sra S (2012) A new Metric on the manifold of kernel matrices with application to matrix geometric means, *NIPS conference*, 1-9.
- Steinhauser M, Kiesel A (2011) Performance monitoring and the causal attribution of errors. *Cognitive, Affective, & Behavioral Neuroscience*, 1-12.
- Steriade M (2005) Cellular Substrates of Brain Rhythms. In: *Electroencephalography. Basic Principles, Clinical Applications, and Related Fields*. Niedermeyer E and Lopes da Silva FH (Eds), 5th ed., New York: Lippincott Williams & Wilkins, 31-83.

- Strang (2006) *Linear Algebra and its Applications*, 4th Ed., Thomson Brooks/Cole, New York.
- Surmeli T, Ertem A (2009) QEEG guided neurofeedback therapy in personality disorders: 13 case studies. *Clin EEG Neurosci*, 40(1), 5-10.
- Talairach J, Tournoux P (1988) *Co-planar stereotaxic atlas of the Human Brain*, Thieme, New York.
- Tan DS, Nijholt A (2012) *Brain-Computer Interfaces (Eds.)*, Springer, London.
- Tang AC, Sutherland MT, Wang Y (2006) Contrasting single-trial ERPs between experimental manipulations: Improving differentiability by blind source separation. *Neuroimage* 2006, 29: 335-346.
- Theis FJ, Inouye Y (2006) On the use of joint diagonalization in blind signal processing. In *Proc. ISCAS*, Kos, Greece, 2006.
- Thorpe SG, Nunez PL, Srinivasan R (2007) Identification of wave-like spatial structure in the SSVEP: Comparison of simultaneous EEG and MEG. *Stat Med*, 26, 3911-3926.
- Tang AC, Liu J-Y, Sutherland MT (2005) Recovery of correlated neuronal sources from EEG: The good and bad ways of using SOBI. *Neuroimage*, 28, 507-519.
- Tichavsky P, Yeredor A, Nielsen J (2008) A Fast Approximate Joint Diagonalization Algorithm using a Criterion with a Block Diagonal Matrix. *Proc ICASSP*, Las Vegas, USA.
- Tichavsky, P. Yeredor, A. 2009. Fast Approximate Joint Diagonalization Incorporating Weight Matrices. *IEEE Trans Sig Process* 57(3), 878-891.
- Tong L, Inouye Y, Liu RW (1993) Waveform-Preserving Blind Estimation of Multiple Independent Sources *IEEE Trans Signal Process*, 41(7), 2461-2470.
- Tong L, Liu RW, Huang Y-F (1990) Blind Estimation of correlated source signals. *Sig Syst Computers*, 1: 258-262.
- Tong L, Liu RW, Soon VC, Huang Y-F (1991a) Indeterminacy and Identifiability of Blind Identification. *IEEE Trans Circuits Syst*, 38(5), 499-509.
- Tong L, Soon V, Huang Y. Liu RW (1991b). A necessary and sufficient condition Waveform-Preserving Blind Estimation of Multiple Independent Sources. *IEEE Trans Signal Process*, 41(7), 2461-2470.
- Townsend G, LaPallo BK, Boulay CB, Krusienski DJ, Frye GE, Hauser CK, et al. (2010) A novel P300-based brain-computer interface stimulus presentation paradigm: moving beyond rows and columns. *Clinical Neurophysiology*, 121(7), 1109-20.
- Trujillo LT, Allen JJB (2007). Theta EEG dynamics of the error-related negativity. *Clinical Neurophysiology*, 118(3), 645-668.
- van der Loo E, Congedo M, Vanneste S, Van De Heyning P, De Ridder D (2011) Insular Lateralization in Tinnitus Distress. *Autonomic Neuroscience: Basic and Clinical*, 165(2), 191-194.
- van Der Loo E, Congedo M, Plazier M, Van De Heyning P, De Ridder D (2007) Correlation between Independent Components of scalp EEG and intra-cranial EEG (iEEG) time series *Int J Bioelectromagnetism*, 9(4), 270-275.
- van der Loo E, Gais S, Congedo M, Vanneste S, Plazier M, et al. (2009) Tinnitus Intensity Dependent Gamma Oscillations of the Contralateral Auditory Cortex, *PLoS ONE* 4(10), e7396.
- van Essen DC (2005) A Population-Average, Landmark- and Surface-based (PALS) atlas of human cerebral cortex. *Neuroimage* 28(3), 635-62.

- Vanneste S, Plazier M, der Loo EV, de Heyning PV, Congedo M, De Ridder D (2010) The Neural Correlates of Tinnitus-Related Distress, *Neuroimage* 52(2), 470-480.
- van Veen BD, Buckley M (1988) Beamforming: A Versatile Approach to Spatial Filtering, *IEEE ASSP Magazine*, 5, 4-24.
- van Veen BD, van Drongelen W, Suzuki A (1997). Localization of Brain Electrical Activity via Linearly Constrained Minimum Variance Spatial Filter, *IEEE Transactions on Biomedical Engineering*, 44(9), 867-880.
- Vía J, Anderson M, Li X-L, Adali T (2011), Joint blind source separation from second-order statistics: Necessary and sufficient identifiability conditions, *ICASSP 2011*, 2520-23.
- Vidaurre C, Krämer N, Blankertz B, Schlögl A (2009) Time domain parameters as a feature for eeg-based brain computer interfaces. *Neural Networks* 22, 1313–1319.
- Vigário RN (1997) Extraction of ocular artifacts from EEG using independent component analysis. *Electroenceph Clin Neurophysiol*, 103, 395-404.
- Vollgraf R, Obermayer K (2006) Quadratic Optimization for Simultaneous Matrix Diagonalization. *IEEE Trans Sig Process*, 54(9), 3270- 3278.
- Vorobyov S, Cichocki A (2002) Blind Noise Reduction for Multisensory Signals using ICA and Subspace Filtering, with Applications to EEG Analysis. *Biol Cybern*, 86, 293-303.
- Yeredor A (2000) Blind Separation of Gaussian Sources via Second-Order Statistics with Asymptotically Optimal Weighting, *IEEE Signal Process Lett*, 7(7), 197-200.
- Yeredor A (2002) Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation. *IEEE Trans Signal Process*, 50 (7), 1545-1553.
- Yeredor A (2010) Second-order methods based on color. In: *Handbook of Blind Source Separation: Independent Component Analysis and Applications*, Comon P and Jutten C Eds, Academic Press, Oxford.
- Wagner M, Fuchs M, Kastner J (2004) Evaluation of sLORETA in the presence of noise and multiple sources, *Brain Topography*, 16(4), 277-80.
- Wang F, Liu F, Zhang J (2007) Nonorthogonal joint diagonalization algorithm based on trigonometric parameterization, *IEEE Trans. Signal Process.*, 55(11), 5299–5308.
- Wang S, James CJ (2007) Extracting Rhythmic Brain Activity for Brain-Computer Interfacing through Constrained Independent Component Analysis. *Comput Intell Neurosci*, ID 41468.
- Wang Y, Zhang Z, Li Y, Gao X and Gao S (2004) BCI Competition 2003-Data set IV: an Algorithm Based on CSSD and FDA for classifying Single-Trial EEG *IEEE Trans. on Reh. Eng.* 51(6), 1081-1086.
- Wax M, Sheinvald J (1997) A Least-Squares Approach to Joint Diagonalization. *IEEE Signal Process Lett*; 4(2) 52-53.
- Welch PD (1967) The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms. *IEEE Trans Audio Electroacoustics*, 15(2), 70-74.
- White D, Congedo M, Ciorciari J, Silberstein R (2010) Brain oscillatory activity during spatial navigation: Theta and gamma activity link medial temporal and parietal regions, *Journal of Cognitive Neuroscience* 24(3), 686-697.
- Whitham EM, Pope KJ, Fitzgibbon SP, Lewis T, Clark CR, Loveless S, et al. (2007) Scalp electrical recording during paralysis: quantitative evidence that EEG frequencies above 20 Hz are contaminated by EMG. *Clin Neurophysiol*, 118(8), 1877-88.

Wolpaw J, Wolpaw EW (2012) *Brain-Computer Interfaces: Principles and Practice*, Oxford University Press, Oxford.

Wolters CH, Anwander A, Tricoche X, Weinstein D, Koch MA, MacLeod RS (2006) Influence of tissue conductivity anisotropy on EEG/MEG field and return current computation in a realistic head model: a simulation and visualization study using high-resolution finite element modeling, *Neuroimage*, 30(3), 813-26.

Zeman PM, Till BC, Livingston NJ, Tanaka JW, Driessen PF (2007) Independent component analysis and clustering improve signal-to-noise ratio for statistical analysis of event-related potentials. *Clin Neurophysiol*, 118(12), 2591-2604.

Zhou G, Yang Z, Wu Z, Zhang J (2008) Non-orthogonal joint diagonalization with diagonal constraints, *Progress in Natural Science*, 18(6), 735–739.

Ziehe A, Laskov P, Nolte G, Müller R-K (2004) A Fast Algorithm for Joint Diagonalization with Non Orthogonal Transformations and its Application to Blind Source Separation. *J Mach Learn Res*, 5, 777-800.

Ziehe A, Müller K-R (1998) TDSEP—an efficient algorithm for blind separation using time structure. *Proc Int Conf Artif Neural Netw (ICANN'98)*, 675–680.