Department of Computer Science

# Separation of Synchronous Sources

Miguel Almeida

Aalto University

DOCTORAL DISSERTATIONS

Aalto University publication series
**DOCTORAL DISSERTATIONS** 80/2016

# Separation of Synchronous Sources

**Miguel Almeida**

A doctoral dissertation completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Science, at a public examination held on June 6th 2016 at Instituto Superior Técnico, Lisbon, in room 4.41 of Pavilhão de Civil, at 14h00.

This doctoral thesis is conducted under a convention for the joint supervision of thesis at Aalto University (Finland) and Instituto Superior Técnico (Portugal).

**Aalto University**
**School of Science**
**Department of Computer Science**
**Instituto Superior Técnico**

**Supervising professors**
Erkki Oja, Aalto University, Finland
José Bioucas-Dias, Instituto Superior Técnico, Portugal

**Thesis advisor**
Ricardo Vigário, Universidade Nova de Lisboa, Portugal

**Preliminary examiners**
Ole Jensen, Centre for Cognitive Neuroimaging, Netherlands
Ercan Kuruoglu, Italian National Council of Research, Italy

**Opponent**
Christian Jutten, Joseph Fourier University, France

Aalto University publication series
**DOCTORAL DISSERTATIONS** 80/2016



ISBN 978-952-60-6783-4 (printed)
ISBN 978-952-60-6784-1 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)
http://urn.fi/URN:ISBN:978-952-60-6784-1

Unigrafia Oy
Helsinki 2016

Finland

Publication orders (printed book):
miguelbalmeida@gmail.com

Aalto University

# Abstract

Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

**Author**
Miguel Almeida

**Name of the doctoral dissertation**
Separation of Synchronous Sources

**Publisher** School of Science

**Unit** Department of Computer Science

**Series** Aalto University publication series DOCTORAL DISSERTATIONS 80/2016

**Field of research** Source Separation

**Manuscript submitted** 27 July 2014 **Date of the defence** 6 June 2016

**Permission to publish granted (date)** 24 November 2014 **Language** English

☐ **Monograph** ☒ **Article dissertation** ☐ **Essay dissertation**

**Abstract**

This thesis studies the Separation of Synchronous Sources (SSS) problem, which deals with the separation of signals resulting from a linear mixing of sources whose phases are synchronous. While this study is made in a form independent of the application, a motivation from a neuroscience perspective is presented. Traditional methods for Blind Source Separation, such as Independent Component Analysis (ICA), cannot address this problem because synchronous sources are highly dependent. We provide sufficient conditions for SSS to be an identifiable problem, and quantify the effect of prewhitening on the difficulty of SSS. We also present two algorithms to solve SSS. Extensive studies on simulated data illustrate that these algorithms yield substantially better results when compared with ICA methods. We conclude that these algorithms can successfully perform SSS in varying configurations (number of sources, number of sensors, level of additive noise, phase lag between sources, among others). Theoretical properties of one of these algorithms are also presented. Future work is discussed extensively, showing that this area of study is far from resolved and still presents interesting challenges.

**Keywords**

**ISBN (printed)** 978-952-60-6783-4 **ISBN (pdf)** 978-952-60-6784-1

**ISSN-L** 1799-4934 **ISSN (printed)** 1799-4934 **ISSN (pdf)** 1799-4942

**Location of publisher** Helsinki **Location of printing** Helsinki **Year** 2016

**Pages** 251 **urn** http://urn.fi/URN:ISBN:978-952-60-6784-1

# Contents

# List of Publications

This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.

**I** Miguel Almeida, Ricardo Vigário. Source Separation of Phase-Locked Subspaces. In *Proceedings of the Independent Component Analysis Conference (ICA)*, 203–210, March 2009.

**II** Miguel Almeida, José Bioucas-Dias, Ricardo Vigário. Independent Phase Analysis: Separating Phase-Locked Subspaces. In *Proceedings of the Latent Variable Analysis Conference (LVA)*, 189–196, September 2010.

**III** Miguel Almeida, Jan-Hendrik Schleimer, José Bioucas-Dias, Ricardo Vigário. Source Separation and Clustering of Phase-Locked Subspaces. *IEEE Transactions on Neural Networks*, 22(9):1419–1434, 2011.

**IV** Miguel Almeida, Jan-Hendrik Schleimer, José Bioucas-Dias, Ricardo Vigário. Source Separation and Clustering of Phase-Locked Subspaces: Derivations and Proofs. *arXiv:1106.2474*, 2011.

**V** Miguel Almeida, José Bioucas-Dias, Ricardo Vigário. Separation of Phase-Locked Sources in Pseudo-Real MEG Data. *EURASIP Journal on Advances in Signal Processing*, 2013.1, 2013.

**VI** Miguel Almeida, Ricardo Vigário, José Bioucas-Dias. Phase Locked Matrix Factorization. In *Proceedings of the European Signal Processing Conference (EUSIPCO)*, 1728–1732, August 2011.

**VII** Miguel Almeida, Ricardo Vigário, José Bioucas-Dias. Estimation of the Common Oscillation for Phase Locked Matrix Factorization. In *Proceedings of the International Conference on Pattern Recognition Applications and Methods (ICPRAM)*, 78–85, February 2012.

**VIII** Miguel Almeida, Ricardo Vigário, José Bioucas-Dias. Phase-Locked Matrix Factorization with Estimation of the Common Oscillation. *Mathematical Methodologies in Pattern Recognition and Machine Learning*, 51–66, 2013.

**IX** Miguel Almeida, Ricardo Vigário, José Bioucas-Dias. The Role of Whitening for Separation of Synchronous Sources. In *Proceedings of the Latent Variable Analysis Conference (LVA)*, 139–146, March 2012.

**X** Miguel Almeida, Ricardo Vigário, José Bioucas-Dias. A Comparison of Algorithms for Separation of Synchronous Subspaces. *Bulletin of the Polish Academy of Sciences: Technical Sciences*, 60:455–460, 2012.

**XI** Miguel Almeida, Ricardo Vigário, José Bioucas-Dias. Separation of Synchronous Sources Through Phase Locked Matrix Factorization. *IEEE Transactions on Neural Networks and Learning Systems*, 25(10):1894–1908, 2014.

# Acknowledgements

This thesis would not exist without the influence of a number of people, to whom I am extremely grateful.

The strongest support undoubtedly came from my close family: my girlfriend Katia, my sister Inês, and my parents Cecília and Luís. They were fundamental in keeping me sane at times where a full-time job, writing this thesis, and the usual hurdles in life were proving overwhelming. Thanks for all the laughs in difficult times!

I am extremely thankful to my supervisors, Ricardo Vigário and Erkki Oja from Aalto University, and José Bioucas-Dias from Instituto Superior Técnico. Many of the ideas in this thesis are the product of engaging discussions with them.

A sincere thank you goes to Ms. Emma Holmlund from Aalto University, who made the dealings with Aalto painless and fast. In fact, all the dealings with Aalto were swift and efficient despite being done remotely from Lisbon. All universities should look to Aalto as an example of how to deal with students.

I would also like to thank the Portuguese Government for providing me with a scholarship that allowed me to undertake this degree. The recent crisis has meant that many others with better conditions than I had are not being given the same opportunity, and I hope some solution can be found to allow them to contribute to science.

# Preliminary Note about Double Degree

The goal of this thesis is to obtain a double Doctoral degree from the Instituto Superior Técnico, University of Lisbon, Portugal, and the School of Science, Aalto University, Finland. Due to the formal requirements of both schools,[1] two versions of the thesis will be produced.

You can always find the latest version at the author's homepage.[2] Alternatively, please request the last version from the author by email.[3]

[1] Aalto requires that the thesis is completely finalized before the defense, while IST requires that changes suggested in the defense be incorporated in the final version.

[2] `http://www.it.pt/person_detail_p.asp?id=4124`

[3] miguelbalmeida@gmail.com

# 1. Introduction

## 1.1 Summary

In Blind Source Separation (BSS) problems, the goal is to estimate a set of signals, called *sources*, while having access only to combinations of those sources (the *mixtures*). The most widely studied class of BSS problems is Independent Component Analysis (ICA), which assumes that the sources are statistically independent.

Let there be $N$ sources, which are stacked into a vector $\mathbf{s}(t)$, and $P$ mixed signals, stacked into a vector $\mathbf{y}(t)$. Then, if the mixture process is linear, instantaneous, and noiseless, one can write $\mathbf{y}(t) = \mathbf{M}\mathbf{s}(t)$, where $\mathbf{M}$ is called the *mixing matrix*.

This thesis deals with a specific instance of linear and instantaneous BSS called Separation of Synchronous Sources (SSS). Two complex signals are considered (fully) synchronous if the difference of their arguments, or phases, does not change with time.[1] While SSS itself is not specific to a particular domain, the motivation for this problem comes from neuroscience. Many studies in neuroscience have found that synchrony between brain regions is fundamental for normal information processing in the human brain, such as learning and memory. Furthermore, abnormal synchrony patterns have been associated to pathologies such as schizophrenia and Alzheimer's.

Studies of synchrony in the neuroscience community have, in some cases, been invasive. Electrodes are placed, usually in mice, and the synchrony between brain regions is measured. Extra-cranial signals, such as the

[1] This will be defined formally in section 3.2.2, along with definitions of partial synchrony and some mathematical properties of synchrony. We also discuss there how this can also be applied to real signals instead of complex ones.

ones obtained through an electroencephalogram (EEG) or magnetoencephalogram (MEG), are an alternative and attractive way to study the synchrony of the brain. The non-invasive character of these signals makes them suitable, for example, for diagnostics in human medicine. EEG and MEG signals are the result of a mixing process, and in this thesis we show that this mixing process can destroy synchrony information. Therefore, some researchers have employed BSS techniques, in particular ICA, to extract the original sources from these mixtures, and subsequently analyze the synchrony between the estimated sources. We argue that it is questionable to use ICA algorithms, which assume independent sources, to estimate synchronous sources, because they are highly dependent. Ideally, one would like to use algorithms which directly estimate synchronous sources.

In SSS, a different assumption is made: instead of assuming independence of the sources, as in ICA, it is assumed that the sources are phase-synchronous, which is a particular type of dependency. We shall show two theoretical properties which illustrate a significant parallelism between SSS and ICA:

1. In ICA, if the sources, denoted by the vector $\mathbf{s}$, are statistically independent, any linear combination $\hat{\mathbf{s}} = \mathbf{G}\mathbf{s}$, where $\mathbf{G}$ denotes a square matrix, such that the components of $\hat{\mathbf{s}}$ are independent, must be such that $\hat{\mathbf{s}} = \mathbf{s}$, up to permutation, scaling and sign change, under some mild assumptions. We will show that a similar property exists for SSS, also under mild conditions: if the sources $\mathbf{s}$ have perfect synchrony, any linear combination $\hat{\mathbf{s}} = \mathbf{G}\mathbf{s}$, with square $\mathbf{G}$, in which the synchrony between components of $\hat{\mathbf{s}}$ is perfect must be such that $\hat{\mathbf{s}} = \mathbf{s}$, again up to permutation, scaling and sign change, with the added requirement that $\mathbf{G}$ is non-singular. While this requirement exists in ICA as well, we will see that it introduces an additional difficulty in the design of SSS algorithms which is not present in most ICA algorithms.

2. Whitening has been proven to yield significant advantages as a preprocessing step for ICA; we will show that this is the case for SSS as well, albeit with smaller benefits. In ICA, in the absence of additive noise, whitening reduces the problem to a search for an orthogonal matrix. In SSS, even in the absence of noise, it makes the problem better conditioned, but the search space remains the full space of matrices of

appropriate dimension. This is another property of SSS which makes it fundamentally harder than ICA.

In addition to proving these two properties and comparing them to their ICA counterparts, we will characterize and tackle SSS itself, namely in the following fronts:

- We show that sources which have perfect synchrony can be described in a particular mathematical form, and propose two algorithms to solve the SSS problem.

- We propose an algorithm called Independent Phase Analysis (IPA) which uses property 1 above and directly tries to maximize the synchrony of the estimated sources, $\hat{\mathbf{s}} = \mathbf{W}^T\mathbf{y}$, where $\mathbf{y}$ are the mixtures, as a function of the *unmixing matrix* $\mathbf{W}$. To prevent $\mathbf{W}$ from becoming singular, we use an appropriate term in the objective function that penalizes singular solutions.

- We also propose Phase Locked Matrix Factorization (PLMF), an algorithm which exploits the particular mathematical form that perfectly synchronous sources can be put in. It computes $\hat{\mathbf{M}}$ and $\hat{\mathbf{s}}$, the estimates of the mixing matrix and sources, respectively, which minimize the squared error between the observed data $\mathbf{y}$ and the product $\hat{\mathbf{M}}\hat{\mathbf{s}}$. Unlike IPA, PLMF has theoretical guarantees which prevent the occurrence of singular solutions. In fact, it will be shown that any global optimum of PLMF's cost function corresponds to correct estimations of $\mathbf{M}$ and $\mathbf{s}$, up to permutation, scaling and sign change, under mild assumptions.

Experimental comparisons, using simulated data, show that PLMF generally obtains superior results when compared to IPA, and illustrate the limits where PLMF's performance degrades below a reasonable level, thus pointing directions for future improvements. These algorithms are also compared to ICA, showing that existing ICA algorithms fail to solve the SSS problem. Furthermore, an initial exploration towards real applications is performed with IPA, using pseudo-real MEG data, which are constructed from actual MEG data but in such a way that the true sources are known.

To this author's best knowledge, this thesis presents the first consolidated framework for blind source separation of synchronous signals, contributing with the problem formulation, theoretical properties of the problem, two algorithms for its solution and their theoretical properties, and experimental tests of those algorithms. The following chapters present a summary of this work, on which a total of nine papers were published in peer-reviewed journals and conferences.

## 1.2 Contributions

The contributions of this thesis can be roughly divided into two groups.

### 1.2.1 Problem Formulation and Characterization

This group of contributions concerns the SSS problem itself. The list of contributions is:

1. Characterizing the SSS problem's solutions. In particular, establishing that the usual BSS indeterminacies of permutation, scaling and sign change are present, and that they are the only indeterminacies for non-singular solutions under mild conditions,[2]

2. Establishing that, unlike ICA, SSS can have singular solutions, which are undesirable.

3. Showing that solutions of this problem can be decomposed in a particular way using matrix factorization.

4. Showing that prewhitening results in a bound in the condition number of the equivalent mixing matrix. This can be interpreted as an upper bound on the "difficulty" of the SSS inverse problem.

### 1.2.2 Separation Algorithms

The second group of contributions regards the proposal of algorithms to solve the problem formulated in the previous subsection. The list of contributions is:

[2] We call singular solutions those where the unmixing matrix $\mathbf{W}$ is singular.

1. Proposing Independent Phase Analysis (IPA), a separation algorithm which directly exploits contribution 2 from the previous subsection to find a non-singular solution. This contribution is a significant extension of work that was initially done by Jan-Hendrik Schleimer.

2. Proposing Phase Locked Matrix Factorization (PLMF), another separation algorithm which exploits contribution 3 of the previous subsection to find a suitable matrix factorization of the data which yields the original sources.

3. Implementing both algorithms in MATLAB.

4. Showing, with simulated data, that both algorithms outperform source separation techniques not tailored for SSS, such as ICA methods.

5. Creating pseudo-real MEG data and demonstrating the usefulness of IPA on it. To the author's best knowledge, real-world data where these algorithms could be tested can be collected with current technology, but is not publicly available and their acquisition is non-trivial.

6. Showing that, under certain conditions, all global minima of PLMF's cost function correspond to a solution which recovers the original sources.

## 1.3 Publications

The work in this thesis has resulted in the publication of nine peer-reviewed papers. An extended version of one of these papers was also published, as well as an arXiv supplementary material containing the proofs of some statements. Except where noted, the author of this thesis contributed in the following ways:

- **Algorithms and problem formulation:** In all papers, the design of the algorithms proposed in this thesis and the formulation of the SSS problem were done jointly with the author's supervisors, Ricardo Vigário and José Bioucas-Dias.

- **Theorems:** In all papers, choosing which results should be proven

was done jointly with the supervisors. The proofs were all done by the present author, with feedback from his supervisors.

- **Experiments and code:** In almost all cases, implementation of the algorithms (in MATLAB) and performing the experiments leading to the results shown in the papers was done by the present author. An exception is a re-implementation of the code of PLMF, performed by one of the supervisors, which resulted in a significantly faster code.

- **Papers:** In all cases, the present author initially wrote all papers, and then incorporated suggestions from his supervisors at later stages.

These eleven publications are listed below, alongside a brief description of each. Copies of these papers can be found at the end of the document, in the order listed below. Citations of these publications appear as **[Publication N]**, where **N** is a roman numeral, to distinguish them from publications from other authors, which are not attached to this document, and which appear without bold and with regular numerals, as in [1]. Therefore, **[Publication I]** and [1] refer to different publications.

1. M. Almeida and R. Vigário. Source Separation of Phase-Locked Signals. In *Proceedings of the Independent Component Analysis Conference (ICA)*, pages 203-210, 2009 **[Publication I]**. This paper presented a novel cost function for an algorithm which would, in subsequent publications, become IPA. It yielded very significant improvements relative to work by Jan-Hendrik Schleimer [71], which can be considered the earliest version of IPA.

2. M. Almeida, J. Bioucas-Dias, and R. Vigário. Independent Phase Analysis: Separating Phase-Locked Subspaces. In *Proceedings of the Latent Variable Analysis Conference (LVA)*, pages 189-196, 2010 **[Publication II]**.[3] This paper presents further improvements on the IPA algorithm. The version of the algorithm presented in this paper is the one that is also presented in **[Publication III]**.

3. M. Almeida, J.-H. Schleimer, J. Bioucas-Dias, and R. Vigário. Source Separation and Clustering of Phase-Locked Subspaces. *IEEE Trans-*

[3]The LVA conference is the same as the ICA one; it simply changed name.

*actions on Neural Networks*, 22(9):1419-1434, 2011 **[Publication III]**. This is the main publication about IPA. It contains a description of the algorithm, as well as more extensive experimental results on simulated data. This paper also presented, for the first time, the theorem stating that all non-singular solutions yield the original sources.[4] Some proofs were skipped due to lack of space; they are available in an arXiv paper **[Publication IV]**.

4. M. Almeida, J.-H. Schleimer, J. Bioucas-Dias, and R. Vigário. Source Separation and Clustering of Phase-Locked Subspaces: Derivations and Proofs. *arXiv:1106.2474 [stat.ML]*, available at `http://arxiv.org/abs/1106.2474` **[Publication IV]**. This arXiv paper contains the proofs from **[Publication III]** which were skipped due to lack of space.

5. M. Almeida, J. Bioucas-Dias, and R. Vigário. Separation of Phase-Locked Sources in Pseudo-Real MEG Data. *EURASIP Journal on Advances in Signal Processing*, 32, 2013 **[Publication V]**. In this paper we tested IPA on pseudo-real data from real MEG data and concluded that IPA could separate synchronous sources in such data. In this paper we also presented an optimization strategy for IPA where the regularization to avoid singular solutions is progressively made weaker, so that in the limit one can avoid regularizing and thus be in the conditions of the theorem presented in **[Publication III]**.

6. M. Almeida, R. Vigário, and J. Bioucas-Dias. Phase Locked Matrix Factorization. In *Proceedings of the European Signal Processing Conference (EUSIPCO*, pages 1728-1732, 2011 **[Publication VI]**. This paper discussed the earliest form of the PLMF algorithm. It used an unrealistic assumption: it assumed that the phase of one of the sources was known.

7. M. Almeida, R. Vigário, and J. Bioucas-Dias. Estimation of the Common Oscillation for Phase Locked Matrix Factorization. In *Proceedings*

[4]This paper also presents two other algorithms, called Referenced Phase Analysis (RPA) and Phase Synchronization Cluster Analysis (pSCA), both originally proposed by Jan-Hendrik Schleimer [69, 70]. The author of this thesis reimplemented these two algorithms and performed all the experiments shown in **[Publication III]**. The present author also corrected minor errors in the expressions for the gradients in those two algorithms. However, the two algorithms were left mostly unchanged from their original forms, and they are not considered contributions.

*of the International Conference on Pattern Recognition Applications and Methods (ICPRAM)*, pages 78-85, 2012 **[Publication VII]**. This paper removed that assumption and presented what ended up being called the "1-stage" PLMF, where all variables are estimated simultaneously. Out of roughly 150 papers presented at this conference, 12 were selected to have extended versions published in the Springer Proceedings in Mathematics & Statistics. This paper was among the 12 selected **[Publication VIII]**.

8. M. Almeida, R. Vigário, and J. Bioucas-Dias. Phase-Locked Matrix Factorization with Estimation of the Common Oscillation. In *Mathematical Methodologies in Pattern Recognition and Machine Learning*, pages 51-66. Springer, 2013 **[Publication VIII]**. This is the extended version of the previous paper. It was published in the Springer Proceedings in Mathematics & Statistics. It presents significantly more thorough experimental results than **[Publication VII]**. The previous paper was peer-reviewed, but there was no further peer-review towards this extended version.

9. M. Almeida, R. Vigário, and J. Bioucas-Dias. The Role of Whitening for Separation of Synchronous Sources. In *Proceedings of the Latent Variable Analysis Conference (LVA)*, pages 139-146, 2012 **[Publication IX]**. This paper presented the upper bound on the condition number of the equivalent mixing matrix if prewhitening is performed.

10. M. Almeida, J. Bioucas-Dias, R. Vigário, and E. Oja. A Comparison of Algorithms for Separation of Synchronous Subspaces. *Bulletin of the Polish Academy of Sciences: Technical Sciences*, 60:455-460, 2012 **[Publication X]**. This paper compared the performance of multiple methods to separate subspaces of synchronous sources.

11. M. Almeida, R. Vigário, and J. Bioucas-Dias. Separation of Synchronous Sources Through Phase Locked Matrix Factorization. *IEEE Transactions on Neural Networks and Learning Systems*, 25(10):1894–1908, 2014 **[Publication XI]**. This is the main publication about PLMF. It presents a novel form of PLMF with two subproblems, which we call the "2-stage" approach. In the first subproblem, one of the variables is estimated from a relaxed version of the original problem. In the second subproblem,

this variable is kept fixed and the remaining ones are estimated. This paper also presents theorems stating that all solutions of both subproblems are desirable ones. Finally, the algorithm is extensively studied on simulated data, and comparisons are made between ICA, IPA, and the 1-stage and 2-stage versions of PLMF, concluding that the latter is clearly superior in performance.

While all the above papers are pertinent for this thesis, the author considers the three journal papers **[Publication III]**, **[Publication V]** and **[Publication XI]** as the most important ones. The first and third ones contain the main theoretical results and both algorithms, while the second one presents early work towards application of these algorithms in real situations.

## 1.4 Document Organization

This dissertation is composed of an introductory part plus a list of publications at the end. In accordance with the rules of both Universities involved in this dissertation, the set of two parts needs to be self-sufficient. The introductory part, composed of chapters 1 to 7, contains most of the contributions: only the proofs of the theorems and some experimental results were omitted from the introduction, but they can be found in the publications in appendix.

This thesis is organized as follows. We begin with a brief introduction to Blind Source Separation in Chapter 2. Special focus is given to Independent Component Analysis, since it is the most widely used BSS problem, and because some SSS theoretical results have ICA counterparts.

We then formally introduce phase synchrony in Chapter 3. We show how the phase of a real signal can be computed through the construction of a complex signal. We present a brief motivation from a neuroscience perspective, and mathematically define synchrony to prepare its use in the algorithms that follow.

Chapter 4 contains this thesis' original contributions. First, the SSS problem is formalized and discussed, without considering which algorithm will be used to solve it. Two main results are presented: a theorem stating that, like ICA, SSS is a well-posed problem; and another theorem quantifying the effect of whitening on the "difficulty" of SSS. This chapter also presents the two algorithms that are proposed to solve SSS: IPA and

PLMF. PLMF has some interesting theoretical properties, which are also discussed in this chapter.

Experimental tests with simulated data are presented for both algorithms in chapter 5; also, some results with pseudo-real MEG data, which are considerably more realistic than the simulated data, are shown for IPA.

Chapter 6 discusses future research directions in considerable detail. Conclusions are drawn in chapter 7. Finally, all publications related to this work are presented in Appendix.

# 2. Blind Source Separation

## 2.1 Inverse Problems

Consider some physical phenomenon which is taking place, and some sensors that are placed to take measurements about the phenomenon. The *direct problem* is the one of computing what one would measure in the sensors, given the state of the experiment. The *inverse problem* is the problem of computing the state of the experiment given the measurements from the sensors. Blind Source Separation is an inverse problem, as we discuss below.

As an example, we briefly discuss the well-known *cocktail party problem*. In this conceptual problem, several people in a room are talking with one another, and some microphones are scattered throughout the room. The microphones capture sound coming from all the people that are talking, making it difficult to obtain the voice of one person directly from one of the microphones. In this situation, the direct problem would consist of computing the signals measured by each microphone, assuming that we know exactly the sound waves generated by each person, each person's location in the room, the room layout, and so on. While the computations for the direct problem may be non-trivial, conceptually it is a straightforward problem. The inverse problem in this situation involves finding the speech signals produced by each person using the signals measured by the microphones. Conceptually, inverse problems are much harder than their direct counterparts, often being ill-posed without further assumptions about the experimental setup.

Formally, the signals measured at the sensors are known as *mixtures*, *mixed signals*, or *sensors*; we will use these terms interchangeably. The unknown signals which, when mixed, originate the measurements are

usually called *sources*. We adopt this terminology here, and now proceed to define the notation used throughout this work.

In this work, we assume that all sources and sensors are one-dimensional discrete-time signals sampled at times $t = 1, 2, \ldots, T$, that sources are numbered from 1 to $N$, and that sensors are numbered from 1 to $P$. Let $y_i(t)$ denote the discrete-time signal measured at sensor $i$, and $s_j(t)$ the discrete-time signal emitted by source $j$. A rather general BSS problem states that

$$y_i(t) = f(s_1(1), \ldots, s_1(t), \ldots, s_N(1), \ldots, s_N(t)), \tag{2.1}$$

where $f(\cdot)$ is a function which depends on the experimental setup.[1] The measurement at sensor $i$ and time $t$ can depend on the signals of all the sources at all time instants up to the instant considered at the sensor.

Two assumptions can be made which tremendously simplify this problem. The first one is that the information travels instantaneously from the sources to the sensors, making this an *instantaneous* BSS problem. Such problems follow a model of the form

$$y_i(t) = f(s_1(t), \ldots, s_N(t)), \tag{2.2}$$

*i.e.*, the signal at sensor $i$ and time $t$ only depends on the signals of the sources at the same time instant.

The second assumption is linearity: if a problem is instantaneous and **linear**, its model is of the form

$$y_i(t) = \sum_{j=1}^{N} m_{ij} s_j(t), \tag{2.3}$$

where $m_{ij}$ is a mixing coefficient which describes how the signal at sensor $i$ depends on source $j$.

The contributions in this thesis assume that the model in (2.3) holds; this model is further explored in section 2.2. However, it is important to remark that relevant work has been done using other kinds of models. For example, *nonlinear ICA* has been used for image separation using instantaneous but nonlinear models (the exact form is not that of equation (2.2), since the sources and sensors are 2D signals) [42, 4]. Furthermore, *convolutive* BSS problems are an important subclass, where the model is

[1] This is not completely general. BSS problems can deal with signals which are more than one-dimensional, such as images. Also, the function $f$ could depend explicitly on the time $t$, if the mixing process itself varies with time. We do not consider these two possibilities.

of the form

$$y_i(t) = \sum_{j\in\{1,\ldots,N\},\tau\in\{1,\ldots,t\}} m_{ij\tau} s_j(t-\tau). \tag{2.4}$$

Convolutive BSS is linear but non-instantaneous, and has been used in applications such as the cocktail party problem and sound source localization, which in fact is similar to the cocktail-party problem, except that the goal is to locate the sound sources (which can be done from the coefficients $m_{ij\tau}$) and not to estimate the source signals $s_j(t)$. A good overview of convolutive BSS methods is available in [64].

## 2.2 Linear and Instantaneous BSS

Linear and instantaneous BSS is the simplest of all the models presented in the previous section, and is the model used throughout this work. Under this model, the signal measured at time $t$ on sensor $i$, which we denote by $y_i(t)$, is given by equation (2.3). We have one equation of this form for every sensor $i$ and every time instant $t$. We thus have $P \times T$ such equations.

It is common to combine these equations using matrix notation. Define $\mathbf{y}(t) \in \mathbb{R}^P$, $\mathbf{s}(t) \in \mathbb{R}^N$ and $\mathbf{M} \in \mathbb{R}^{P\times N}$ as follows:

$$\mathbf{y}(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_P(t) \end{bmatrix}, \quad \mathbf{s}(t) = \begin{bmatrix} s_1(t) \\ s_2(t) \\ \vdots \\ s_N(t) \end{bmatrix} \quad \text{and} \quad \mathbf{M} = \begin{bmatrix} m_{11} & m_{12} & \ldots & m_{1N} \\ m_{21} & m_{22} & \ldots & m_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ m_{P1} & m_{P2} & \ldots & m_{PN} \end{bmatrix}. \tag{2.5}$$

Matrix $\mathbf{M}$ is called the *mixing matrix*. The $P$ equations of the form (2.3), corresponding to the same time instant $t$, can be compactly expressed as

$$\mathbf{y}(t) = \mathbf{M}\mathbf{s}(t), \tag{2.6}$$

and this can be done for each time instant $t$, thus there are $T$ equations.

These $T$ equations can be further compacted into a single equation. Define $\mathbf{Y} \in \mathbb{R}^{P\times T}$ and $\mathbf{S} \in \mathbb{R}^{N\times T}$ as follows:

$$\mathbf{Y} = \begin{bmatrix} y_1(1) & y_1(2) & \ldots & y_1(T) \\ y_2(1) & y_2(2) & \ldots & y_2(T) \\ \vdots & \vdots & \ddots & \vdots \\ y_P(1) & y_P(2) & \ldots & y_P(T) \end{bmatrix} \quad \text{and} \quad \mathbf{S} = \begin{bmatrix} s_1(1) & s_1(2) & \ldots & s_1(T) \\ s_2(1) & s_2(2) & \ldots & s_2(T) \\ \vdots & \vdots & \ddots & \vdots \\ s_N(1) & s_N(2) & \ldots & s_N(T) \end{bmatrix}. \tag{2.7}$$

We shall use a slight abuse of notation and sometimes call the matrix $\mathbf{Y}$ the *sensor data*, or sometimes simply *sensors*. $\mathbf{S}$ will be called *source data* or *sources*. The $T$ equations of the form (2.6) can be compacted into a single equation:

$$\mathbf{Y} = \mathbf{MS}. \tag{2.8}$$

Usually, the objective of BSS is to find the sources $\mathbf{S}$, using only the data from the sensors $\mathbf{Y}$, although in some cases the goal might be to find the mixing matrix $\mathbf{M}$. In either case, BSS is an ill-posed problem. Equation (2.8) makes the reasons for this clear: in general, there is an infinite number of pairs $(\mathbf{M}, \mathbf{S})$ which, when multiplied, yield the observed data $\mathbf{Y}$. It is, therefore, necessary to make some assumptions on the sources $\mathbf{S}$, on the mixing matrix $\mathbf{M}$, or on both, to make the problem well-posed. Different BSS problem make different assumptions:

- By far, the most well-known BSS problem is Independent Component Analysis (ICA). Its fundamental assumption is that, at each time instant $t$, the value of each source $s_j(t)$ is a realization of a random variable $S_j$, and that the random variables $S_1, S_2, \ldots, S_N$ are statistically independent. ICA will be briefly discussed in Section 2.3; further analysis is deferred to that section.

- A generalization of ICA is Independent Subspace Analysis (ISA). Its fundamental assumption is that there are several sets of sources, which are usually called *subspaces*. Sources in the same subspace can be mutually dependent, but the set of sources in a subspace is independent from the set of all other sources. While ICA can be considered a mature field, ISA is currently being actively researched. It will be briefly discussed in Section 2.4.

- Non-negative Matrix Factorization (NMF) can also be viewed as an instance of BSS, although the literature does not always cast it as such. Its fundamental assumption is that the entries of both the mixing matrix $\mathbf{M}$ and the source data $\mathbf{S}$ are non-negative. Despite being now around ten years old [36, 52], NMF is a very active area of research, with applications in, *e.g.*, acoustic signal processing [44, 45] and hyperspectral unmixing [53]. However, this topic is not central to the work presented in this thesis, and it is not discussed further.

- The topic of this thesis, Separation of Synchronous Sources (SSS), is also an instance of BSS. The fundamental assumption is that the sources have perfect phase synchrony with one another. SSS will be the subject of Section 4.

One can draw several parallelisms between SSS and the well-known case of ICA. For this reason, we now provide a brief overview of ICA.

### 2.3 Independent Component Analysis

The term "Blind Source Separation" began to be used in the early 1990s (see, *e.g.*, [41]), while the term "Independent Component Analysis" became widespread a few years later [17]. However, ICA began to take form in the early 1980s, in France, with works by Hérault, Jutten, and Ans [32, 34, 7, 33]. In [43] (an excellent historical overview, including comments from many pioneer researchers in the field), Jutten places the birth of BSS in 1982. According to [38], despite earlier works presenting solutions to the problem which would become known as ICA (such as [33]), a major turning point in the history of ICA was the publication, in 1995, of an approach based on the infomax principle [10, 9], which drew wider attention to the field. ICA can now be considered a mature field of research, with the ICA conference, created in 1989 and occurring every 18 months until 2009, and then the LVA/ICA conference[2] from 2010 onwards, gathering around 150 researchers from the field.

ICA has seen wide application, even in situations where the independence assumption is not satisfied. Examples of applications are the removal of artifacts from neurophysiological signals [54, 87] and modeling the receptive fields of neurons of the primary visual cortex [11], among many others. Good overviews of ICA include [38, 16, 18].

As was said above, ICA, in its typical form, is a linear and instantaneous BSS problem which assumes that $s_j(t)$, and consequently $y_i(t)$, are realizations of random variables. Specifically, the sources $s_j(t)$ are assumed to be i.i.d. realizations of random variables $S_j$, with $S_1, S_2, \ldots, S_N$ being statistically independent. It turns out that this independence assumption is enough to make the problem "sufficiently well-posed", in a sense that will be rigorously defined below.

For simplicity, throughout this section the number of sources is assumed

[2] LVA stands for the more general designation "Latent Variable Analysis".

to be equal to the number of sensors, *i.e.*, $P = N$, and the mixing matrix $\mathbf{M}$ (which is square for $P = N$) is assumed to be invertible. Most results in this section can be generalized to the case where one has more sensors than sources ($P > N$), which is called the *overdetermined* case. The case where the number of sensors is smaller than the number of sources ($P < N$), which is called *underdetermined*, is considerably harder to tackle.

One of the most important aspects of ICA is that it is still technically ill-posed, in the sense that there are still infinite solutions to equation (2.8), even if the independence assumption is verified. However, the following theorem precisely characterizes the ill-posedness of ICA [17]:

**Theorem 2.3.1.** *Let* $\hat{\mathbf{s}} \equiv \mathbf{G}\mathbf{s}$ *be a set of signals resulting from a linear combination of sources* $\mathbf{s}$*, with square* $\mathbf{G}$*, and let these sources be statistically independent as per the ICA model. Furthermore, assume that at most one of the components of* $\mathbf{s}$ *is Gaussian, and that none of them has a degenerate point-like distribution. The components of* $\hat{\mathbf{s}}$ *are statistically independent if and only if* $\hat{\mathbf{s}} = \mathbf{D}\mathbf{P}\mathbf{s}$ *for some diagonal matrix* $\mathbf{D} \in \mathbb{R}^{N \times N}$ *with nonzero entries in its diagonal, and some permutation matrix* $\mathbf{P} \in \mathbb{R}^{N \times N}$.

This theorem formalizes the previously mentioned statement that ICA is "sufficiently well-posed". While the problem is ill-posed, in the sense that finding independent estimated sources $\hat{\mathbf{s}}$ does not imply that $\hat{\mathbf{s}} = \mathbf{s}$, all those solutions correspond to situations where each component of $\hat{\mathbf{s}}$ depends only on a single source. Specifically, the following indeterminacies exist:

- The *order* of the sources cannot be determined. This happens because the order of the terms in the sum of Equation (2.3) can be changed without affecting the value of the sum. Equivalently, one can permute the rows of matrix $\mathbf{S}$ and apply the same permutation to the columns of $\mathbf{M}$ without affecting the product $\mathbf{MS}$. In the matricial notation of Equation (2.8), it is equivalent to considering a new mixing matrix $\tilde{\mathbf{M}} \equiv \mathbf{M}\mathbf{P}^{-\mathbf{1}}$ and a new source matrix $\tilde{\mathbf{S}} \equiv \mathbf{P}\mathbf{S}$, where $\mathbf{P} \in \mathbb{R}^{N \times N}$ is some permutation matrix. This is called the *permutation indeterminacy*.

- The *scale* of the sources cannot be determined. This happens because one can, in Equation (2.3), apply scaling factors $\alpha_j \neq 0$ to each source $s_j$ and apply the inverse scaling $\frac{1}{\alpha_j}$ to all mixing coefficients involving that source, $m_{1j}, \ldots, m_{Pj}$, without affecting the resulting mixture signals. In the matricial notation of Equation (2.8), it is equivalent to considering

a new mixing matrix $\tilde{\mathbf{M}} \equiv \mathbf{M}\mathbf{D}^{-1}$ and a new source matrix $\tilde{\mathbf{S}} \equiv \mathbf{D}\mathbf{S}$, where $\mathbf{D} \in \mathbb{R}^{N \times N}$ is some diagonal matrix with non-zero entries in its diagonal. This is called the *scaling indeterminacy*.

- The previous indeterminacy can involve negative scaling factors, which result in changes of sign of the estimated sources. While this is already included in the previous case, this is sometimes referred separately as the *sign indeterminacy*.

These indeterminacies are very common in linear and instantaneous BSS. However, note that they may depend on the specific problem. For example, in Non-Negative Matrix Factorization, where the matrices $\mathbf{M}$ and $\mathbf{S}$ are assumed to have non-negative entries, the sign indeterminacy does not exist. Usually, the way to deal with these indeterminacies is that one aims at finding an *unmixing matrix* $\mathbf{W}$ such that the estimated sources, given by $\hat{\mathrm{s}} \equiv \mathbf{W}^T\mathrm{y} = \mathbf{W}^T\mathrm{M}\mathrm{s}$, are a permutation and scaling of the original sources, since the order and scale are impossible to determine. Equivalently, the *gain matrix* $\mathbf{W}^T\mathbf{M}$, should be a permutation of a diagonal matrix with nonzero elements in the diagonal.

If one could simply "maximize the independence" of the estimated sources $\hat{\mathrm{s}}$ as a function of $\mathbf{W}$, then the previous theorem would ensure that all global optima, where the estimated sources are independent, would be good enough as long as the order and scale of the sources were not important. It is important to remark, however, that the notion of "maximizing the independence" of the estimated sources glosses over a lot of the research put into ICA. In fact, it is not easy to measure the independence of a set of random variables when one only has access to a set of realizations. ICA algorithms replace independence with other criteria which approximate independence, such as kurtosis, negentropy, and time-lagged correlations, among others. The exploration of these flavors of ICA is outside the scope of this work; good overviews can be found in [38, 16, 19].

We now describe two very common and very useful preprocessing steps for ICA. Let $\mathbb{E}[\mathrm{y}]$ denote the expected value of the random vector $\mathrm{y}$, and $\mathbb{E}[\mathrm{y}\mathrm{y}^T]$ denote its covariance. In practice, the mean $\mathbb{E}[\mathrm{y}]$ and covariance $\mathbb{E}[\mathrm{y}\mathrm{y}^T]$ are unknown. If a sufficient number of time samples is available, these can be well approximated by their estimators, $< \mathrm{y} >$ and $\frac{T}{T-1} < \mathrm{y}\mathrm{y}^T >$, where $< . >$ denotes the time averaging operator. ICA is usually preceded by **centering** the data, *i.e.*, removing its mean. This

step returns for each time $t$ a new vector given by $\tilde{\mathbf{y}}(t) \equiv \mathbf{y}(t) - \mathbb{E}[\mathbf{y}]$. Equivalently, one can subtract $\mathbb{E}[\mathbf{y}]$ from every column of matrix $\mathbf{Y}$. For simplicity, we shall assume that centering has been applied to the data.

Also, it is normally useful to perform **prewhitening** of the data: this step applies a linear transformation to the data such that their covariance matrix, after the transformation, is the identity matrix. Let

$$\mathbf{C_Y} \equiv \mathbb{E}[\mathbf{y}\mathbf{y}^T] \tag{2.9}$$

denote the covariance matrix of the data.[3] Consider the eigendecomposition of $\mathbf{C_Y}$,

$$\mathbf{C_Y} = \mathbf{V}\mathbf{D}\mathbf{V}^T, \tag{2.10}$$

where $\mathbf{D}$ is a diagonal matrix containing the eigenvalues of $\mathbf{C_Y}$ in some order, and $\mathbf{V}$ is an orthogonal matrix ($\mathbf{V}\mathbf{V}^T = \mathbf{V}^T\mathbf{V} = \mathbf{I}$) containing the eigenvectors of $\mathbf{C_Y}$ in the corresponding order.

Then, prewhitening can be performed by multiplying the data $\mathbf{Y}$ on the left by the matrix

$$\mathbf{B} \equiv \mathbf{D}^{-\frac{1}{2}}\mathbf{V}^T. \tag{2.11}$$

It is easy to see that the covariance of the whitened data $\mathbf{z}$, given by $\mathbf{z} \equiv \mathbf{B}\mathbf{y}$, is the identity matrix:

$$\begin{aligned}
\mathbb{E}[\mathbf{z}\mathbf{z}^T] &= \mathbb{E}[\mathbf{B}\mathbf{y}\mathbf{B}\mathbf{y}^T] \\
&= \mathbf{B}\mathbb{E}[\mathbf{y}\mathbf{y}^T]\mathbf{B}^T \\
&= \mathbf{D}^{-\frac{1}{2}}\underbrace{\mathbf{V}^T\mathbf{V}}_{\equiv\mathbf{I}}\mathbf{D}\underbrace{\mathbf{V}^T\mathbf{V}}_{\equiv\mathbf{I}}\mathbf{D}^{-\frac{1}{2}} \\
&= \mathbf{D}^{-\frac{1}{2}}\mathbf{D}\mathbf{D}^{-\frac{1}{2}} \\
&= \mathbf{I}.
\end{aligned} \tag{2.12}$$

An important consequence of prewhitening is the following result [38].

**Theorem 2.3.2.** *Let* $\mathbf{y} \equiv \mathbf{M}\mathbf{s}$ *be a set of measurements resulting from a linear combination of sources* $\mathbf{s}$*, with invertible* $\mathbf{M}$*, and let these sources be statistically independent as per the ICA model. Furthermore, assume that at most one of the components of* $\mathbf{s}$ *is Gaussian. Let* $\mathbf{z} \equiv \mathbf{B}\mathbf{y} = \mathbf{B}\mathbf{M}\mathbf{s}$ *denote the result of prewhitening the sensor data. Then, there exists an orthogonal matrix* $\mathbf{W} \in \mathbb{O}^{N\times N}$ *such that the components of* $\hat{\mathbf{s}} \equiv \mathbf{W}^T\mathbf{z}$ *are statistically independent.*

[3] Recall that the data have been centered, and therefore their mean is zero. Therefore, we can use the terms “correlation” and “covariance” interchangeably.

Theorem 2.3.2 states that, if prewhitening is performed, we can find a set of independent sources $\hat{\mathbf{s}}$ by searching for an orthogonal matrix. Note that $\hat{\mathbf{s}}$ is a linear combination of the original sources $\mathbf{s}$, because $\hat{\mathbf{s}} \equiv \mathbf{W}^T\mathbf{z} = \mathbf{W}^T\mathbf{B}\mathbf{y} = \mathbf{W}^T\mathbf{B}\mathbf{M}\mathbf{s}$. Therefore, Theorem 2.3.1 ensures that after prewhitening one can solve the ICA problem (in the absence of noise) by searching for an orthogonal matrix. While the literature on ICA usually discusses prewhitening as well [38, 16, 18], it should be emphasized that while prewhitening is a useful preprocessing step for ICA, it is not part of it, nor is it indispensable.

Let $\sigma_{max}(\mathbf{M})$ and $\sigma_{min}(\mathbf{M})$ denote the largest and smallest singular values of matrix $\mathbf{M}$.[4] The **condition number** of $\mathbf{M}$ is defined as

$$\rho(\mathbf{M}) = \frac{\sigma_{max}(\mathbf{M})}{\sigma_{min}(\mathbf{M})}. \tag{2.13}$$

The condition number of a matrix with at least one nonzero element always belongs to the set $[1, +\infty) \cup \{+\infty\}$. In particular, orthogonal matrices have $\rho = 1$, and singular matrices have $\rho = +\infty$.

The condition number of a mixing matrix can be considered an indicator of the difficulty of the corresponding inverse problem.[5] Problems where the mixing matrix has a high condition number are usually harder to solve than problems with a smaller condition number. Theorem 2.3.2 ensures that, if prewhitening is performed, ICA is reduced to a search for an orthogonal matrix, which has $\rho = 1$. This will not be the case for SSS. Thus, in a a certain sense, SSS can be considered a "harder" inverse problem when compared to ICA. We shall return to this interpretation later, when we discuss the effect of whitening on SSS.

## 2.4 Independent Subspace Analysis

The goal of this thesis is to study the problem of separating sources with a particular kind of dependency: they are assumed to be perfectly synchronous. In reality, however, this assumption limits the applicability of the theory and of the methods developed here. In fact, in real-world

[4] For square Hermitian matrices, $\sigma_{max}(\mathbf{M})$ and $\sigma_{min}(\mathbf{M})$ are equal to $|\lambda_{max}(\mathbf{M})|$ and $|\lambda_{min}(\mathbf{M})|$, respectively, where $\lambda_{max}(\mathbf{M})$ and $\lambda_{min}(\mathbf{M})$ are the eigenvalues of $\mathbf{M}$ with the largest and smallest absolute values, respectively.

[5] While essentially all linear inverse problems involve a matrix whose function is similar to the function of our mixing matrix, in many cases it does not correspond to a mixing of signals and therefore is not called "mixing matrix". We still call it "mixing matrix" for brevity.

situations it is unlikely that one only has sources which are perfectly synchronous with one another. It is more realistic to consider situations where (approximately) synchronous sources of interest may be considered independent of all other sources present in the mixing process. A brief motivation from a neuroscience perspective will be presented in section 3.2.1.

In this section we explore a generalization of ICA where, instead of assuming independence between individual sources, one assumes independence between groups of sources, and sources within each group may be dependent. This generalization is called *Independent Subspace Analysis* (ISA). If some of these groups' dependency is strong phase synchrony, as may be the case in the human brain, a two-step procedure where ISA is employed to separate the groups, and SSS is employed to extract synchronous sources from their respective groups, may be adequate.

In ISA one assumes that there are sets of sources, called *subspaces*,[6], where each set is independent from the set of sources not belonging to the subspace. Sources within each such set are not necessarily independent. In mathematical notation, the vector of sources, s, is assumed to have the form

$$\mathrm{s} \equiv \begin{bmatrix} \mathrm{s}^1 \\ \mathrm{s}^2 \\ \vdots \\ \mathrm{s}^K \end{bmatrix}, \text{ with } \mathrm{s}^k \equiv \begin{bmatrix} s_1^k \\ \vdots \\ s_{N_k}^k \end{bmatrix}, \tag{2.14}$$

where $\mathrm{s}^1, \mathrm{s}^2, \ldots, \mathrm{s}^K$ are called the $K$ *subspaces* of s. We use $N_k$ to denote the dimension of each subspace; note that they must obey $N_1 + \ldots + N_K = N$, where $N$ is the number of sources, *i.e.*, the dimension of s. The critical assumption in ISA is that the vectors $\mathrm{s}^1, \mathrm{s}^2, \ldots, \mathrm{s}^K$ are statistically independent. As mentioned above, the components of each subspace vector $\mathrm{s}^k$ need not be independent.

If $N_1 = N_2 = \ldots = N_K = 1$ and $K = N$, all subspaces have dimension 1 and we recover the ICA case. In the other extreme, where $K = 1$, only one subspace is present, and no independence assumption exists. Source separation, in this case, is sometimes called Dependent Component Analysis (DCA), a field with its own active research community (see, *e.g.*, [47] and references therein). SSS is a form of DCA.

[6]Technically, "subspaces" is the term for all possible linear combinations of sources in one of these sets. In the literature it is common to also call a "subspaces" to each set of sources. We will employ this slightly abusive terminology.

Assume that it is known that a certain set of observations $\mathbf{y}(t)$ is the result of a mixture of sources $\mathbf{s}(t)$, of dimension $N$, using a mixing matrix $\mathbf{M}$: $\mathbf{y}(t) = \mathbf{M}\mathbf{s}(t)$. Assume also that $\mathbf{s}$ follows the ISA assumption: the $N$ sources can be partitioned into a number of sets $K \in \{1, \ldots, N\}$ and the different sets are independent. ICA can be motivated as the minimization of the *mutual information* between the scalar sources [38, chapter 10]:

$$\min MI(\hat{s}_1, \ldots, \hat{s}_N), \tag{2.15}$$

where $\hat{s}_i$ is the estimate of the $i$-th source. In the presence of subspaces, one natural generalization would be to minimize the mutual information between the various subspaces:

$$\min MI(\hat{\mathbf{s}}^1, \ldots, \hat{\mathbf{s}}^K)$$

This approach has seen some use. However, it presents two problems:

- In general, this approach is a combinatorial optimization problem [13], since one does not know which of the estimated sources should be grouped together [3, 66] when defining the subspaces. One can then test all possible groupings, but they grow very quickly with $N$: the problem rapidly becomes intractable.[7] Alternatively, one could solve a discrete optimization problem by following, *e.g.*, a greedy approach to cluster the estimated sources, an approach that is not guaranteed to yield the optimal solution.

- This approach involves the computation of the entropy of random vectors of dimension $N_k$. Such computation is non-trivial for $N_k \geq 2$ [8], further increasing the complexity of this approach. Nevertheless, this approach has been tackled, *e.g.*, by estimating the entropy of multi-dimensional components using minimum spanning trees [66], or using variational Bayes approaches [3].

We now divide the general ISA problem, of recovering the original sources when subspaces are present, into three successive parts. The first part is called *inter-subspace separation*. The goal of this first part is to obtain a demixing matrix $\mathbf{W}_{inter}$, such that the *gain matrix*, $\mathbf{G} = \mathbf{W}_{inter}^T\mathbf{M}$, is a permutation of a block diagonal matrix with blocks corresponding to the

[7] The number of partitions of a set with $n$ elements is called the $n$-th *Bell number*, $B_n$. One has $B_0 = B_1 = 1$ and $B_{n+1} = \sum_{i=0}^{n} \frac{n!}{i!(n-i)!} B_i$. The first few elements are 1, 1, 2, 5, 15, 52, 203, 877 [2].

subspaces. For example, suppose that there are three subspaces ($K = 3$), the first of which has three components ($N_1 = 3$), while the second and third subspaces have two components each ($N_2 = N_3 = 2$). In this case, the goal is to find a matrix $\mathbf{W}_{inter}$ of the form $\mathbf{W}_{inter}^T \equiv \mathbf{P}\mathbf{B}_{inter}$, where $\mathbf{P}$ is a permutation matrix and $\mathbf{B}_{inter}$ is such that

$$\mathbf{B}_{inter}\mathbf{M} = \begin{bmatrix} \mathbf{U}^1 & \mathbf{0}_{3\times 2} & \mathbf{0}_{3\times 2} \\ \mathbf{0}_{2\times 3} & \mathbf{U}^2 & \mathbf{0}_{2\times 2} \\ \mathbf{0}_{2\times 3} & \mathbf{0}_{2\times 2} & \mathbf{U}^3 \end{bmatrix}.$$

Here, $\mathbf{0}_{m\times n}$ is the $m$-by-$n$ zero matrix, $\mathbf{U}^1$ is a 3-by-3 invertible matrix, and $\mathbf{U}^2$ and $\mathbf{U}^3$ are 2-by-2 invertible matrices. After this step, each entry of the random vector $\mathbf{x}_{inter} \equiv \mathbf{W}_{inter}^T\mathbf{y}$ is a linear combination of sources from one subspace only.

The second step is called *subspace detection*. The goal is to permute the entries of the random vector $\mathbf{x}_{inter}$ so that the first $N_1$ entries of $\mathbf{x}_{inter}$ are linear combinations of sources from the first subspace, the next $N_2$ entries are linear combinations of sources from the second subspace, and so on. Formally, we multiply $\mathbf{x}_{inter}$ by a suitable permutation matrix, $\mathbf{Q}$. Finding $\mathbf{Q}$ is, in general, a combinatorial problem. In the case of SSS, we used a simple heuristic **[Publication III]** to perform this step, with reasonable results.

After the subspace detection is completed, one can define

$$\mathbf{y}^1 \equiv \mathbf{Q}_{(1:N_1\ ,\ :)}\mathbf{x}_{inter} \tag{2.16}$$

$$\mathbf{y}^2 \equiv \mathbf{Q}_{(N_1+1:N_1+N_2\ ,\ :)}\mathbf{x}_{inter} \tag{2.17}$$

$$\vdots \tag{2.18}$$

$$\mathbf{y}^K \equiv \mathbf{Q}_{(N-N_K+1:N\ ,\ :)}\mathbf{x}_{inter}, \tag{2.19}$$

where $\mathbf{Q}_{(a:b\ ,\ :)}$ is a matrix composed of the rows $a$ to $b$ of matrix $\mathbf{Q}$. Thus, $\mathbf{y}^k$ is a $N_k$-dimensional vector containing linear combinations of the sources of the $k$-th subspace.

The third and last step is called *intra-subspace separation*. It involves finding square matrices $\mathbf{W}^k \in \mathbb{R}^{N_k\times N_k}$ such that $\mathbf{s}^k = (\mathbf{W}^k)^T\mathbf{y}^k$, up to permutation, scale and sign change. There are $K$ such matrices to be found, and each can be estimated separately once inter-subspace separation and subspace detection have been performed. This step requires some knowledge about the sources under study or about their interdependency within the subspace. If we know that the interdependency is strong synchrony, SSS algorithms could be employed to perform this step.

Sadly, there is no consensus in the available literature on what the term "Independent Subspace Analysis" means. Some authors (*cf.*, [62, 78]) define ISA as the task of performing all three steps, while others (*cf.*, [39, 73]) define the same term as solving only the first step or the first two steps. To prevent confusion, we shall define *Full ISA* as the task of performing all three steps, and *Partial ISA* as the task of performing only the first step.

Both full and partial ISA have seen increasing interest from the scientific community in recent years. While these problems are usually called "Independent Subspace Analysis" [29, 39, 66, 83], other names have been used, such as "Subspace Independent Component Analysis" [73], "Independent Vector Analysis" [1], and "Multidimensional Independent Component Analysis" [15], among others. ISA was first proposed for fetal electrocardiogram extraction [21]; another important early work is [15]. It has also been applied to capturing inter-subject variability in functional magnetic resonance imaging (fMRI) [1], natural image analysis [37] and analysis of cosmic microwave background radiation [50], among other fields.

Relevant theoretical results have been published about this topic, such as sufficient conditions on the distribution of the sources for full ISA to be achievable through maximization of kurtosis [78] or through minimization of mutual information [62]. A general discussion of contrast functions can be found in [63]. Under the (quite restrictive) conditions stated in these works, then, simple ICA algorithms which maximize kurtosis (such as some variants of FastICA) or minimize mutual information (such as Infomax) can be safely used to perform the first and third steps of full ISA, even though the assumption of independence of the sources is violated. In other words, one can recover the original sources by applying methods which do not consider subspaces at all; one can, if desired, group the recovered sources into subspaces *a posteriori*.

Dedicated algorithms for partial ISA have also been proposed; see, *e.g.*, [39] and [73]. Techniques for subspace detection have also been recently presented [29]. We performed a comparison of ICA and ISA algorithms for partial ISA where the interdependency in each subspace is perfect synchrony **[Publication X]**. The best-performing algorithm was FastICA, which outperformed two other ICA algorithms as well as three ISA ones.

Some researchers have focused on specific types of sources. For example, in [51], second-order statistics are used to perform ISA, and a model is derived for multidimensional Gaussian sources.

The conjecture that full ISA can be solved by using simple ICA and afterwards grouping the sources into subspaces is called *ISA separation principle* [79]. This conjecture has been proven for certain source types (see [79] for an overview), and recent works such as [49] suggest that it may be true for a broad class of sources.

# 3. Phase Synchrony

Unlike ICA, in which we assume that the sources are statistically independent, in Separation of Synchronous Sources (SSS) the underlying assumption is that the sources have a very particular kind of dependence: they are perfectly synchronous. In this section, we provide a brief overview of the concepts of phase and phase synchrony. We begin by discussing how to obtain the phase of a real signal, and then provide precise definitions of synchrony which will be used in the following chapters.

## 3.1 From Real to Complex Signals: The Analytic Signal

In many real-world applications, including the analysis of EEG and MEG signals from the brain, the measurements are real-valued. However, all the methods proposed in this thesis will require us to compute the phase of the signals presented as input. In this section we discuss an important question: how can we define the phase of a real signal $s(t)$?

Typically, this step is performed by obtaining a complex signal $x(t)$ from the given real signal, and defining the phase of the real signal $s(t)$ as the argument, or angle, of the complex signal $x(t)$. This reduces the question of the previous paragraph to a new question: how should we define this complex signal?

Let the original signal be decomposed as $s(t) = a(t)\cos\varphi(t)$, where $a(t)$ is the *amplitude* of the signal and $\varphi(t)$ is its *phase*. Our goal is to know under which conditions we can create a complex signal $x(t)$ equal to $x(t) = a(t)e^{i\varphi(t)}$; we will say that this is a "meaningful" complex signal, and that its phase is "meaningful" as well. Furthermore, the function mapping $s(t)$ to $x(t)$ should be a linear function, such that a linear combination of real signals yields the same linear combination of the corresponding complex signals.

Let $S(\omega)$ denote the Fourier transform (FT) of $s(t)$, given by

$$S(\omega) \equiv \int_{-\infty}^{+\infty} s(t)e^{-i\omega t}dt, \text{ for } \omega \in \mathcal{R}, \tag{3.1}$$

and let $X(\omega)$ be the FT of $x(t)$. Note that $s(t) = \frac{1}{2}a(t)(e^{i\varphi(t)} + e^{-i\varphi(t)}) = \frac{1}{2}(x(t) + x^*(t))$, where $^*$ denotes complex conjugation. Using basic FT properties [61, section 4.6], the FTs of $s(t)$ and $x(t)$ are related through $S(\omega) = \frac{1}{2}(X(\omega) + X^*(-\omega)$.

Suppose now that the support of $X(\omega)$ is contained in $\mathbb{R}^+$. In that case, the support of $X^*(-\omega)$ is contained in $\mathbb{R}^-$, and $X(\omega)$ can be obtained from $S(\omega)$ through

$$X(\omega) = 2U(\omega)S(\omega), \tag{3.2}$$

where $U(\omega)$ is the *step* (or *Heaviside*) function:

$$U(\omega) = \begin{cases} 1 & \text{if } \omega \geq 0 \\ 0 & \text{if } \omega < 0. \end{cases}$$

Therefore, we can construct $X(\omega)$ by computing the FT of $s(t)$ and multiplying it by the step function $U(\omega)$. To obtain $x(t)$, which is called the *analytic signal*, we merely need to compute the inverse Fourier transform (IFT) of $X(\omega)$, given by

$$x(t) \equiv \frac{1}{2\pi}\int_{-\infty}^{+\infty} X(\omega)e^{i\omega t}dt, \text{ for } t \in \mathcal{R}. \tag{3.3}$$

An important question remains to be discussed: for which real signals do the corresponding analytic signals $x(t)$ equal $x(t) = a(t)e^{i\varphi(t)}$? This question is fundamental for this thesis, since apart from the few real-world cases that are intrinsically well represented by complex-valued signals, those will be the signals that our methods can be applied to.

The crucial step in obtaining the analytic signal is the assumption that the support of $X(\omega)$ only contains positive frequencies. For example, consider a signal defined as

$$s(t) = a(t)\cos(\omega_0 t) \tag{3.4}$$

where $a(t)$ is some non-negative signal, and $\omega_0$ is a positive real number. By the shift property of the FT, the FT of $s(t)$ is given by

$$S(\omega) = \frac{A(\omega - \omega_0) + A(\omega + \omega_0)}{2}, \tag{3.5}$$

where $A(\omega)$ is the FT of $a(t)$.

Suppose that the following condition holds:

$$A(\omega) = 0 \text{ for } |\omega| \geq \omega_0. \tag{3.6}$$

In this case, $A(\omega + \omega_0) = 0$ for $\omega > 0$, and the FT of the analytic signal of $s(t)$ (equation (3.2)) yields

$$X(\omega) = \begin{cases} 0 & \text{for } \omega < 0 \\ A(\omega - \omega_0) & \text{for } \omega > 0, \end{cases}$$

and the corresponding analytic signal is

$$x(t) = a(t)e^{i\omega_0 t}.$$

The phase of $x(t)$ is its argument, and since $a(t)$ is real and non-negative, its argument is simply equal to $\omega_0 t$, the argument of the cosine function in the definition of $s(t)$ (equation (3.4)). Therefore, in this case the analytic signal preserves the phase of the real signal in a meaningful way.

In the general case, where $\varphi(t)$ is not constant, the signals for which the analytic signal contains a meaningful phase are those for which the support of $X(\omega)$ is contained in $\mathbb{R}^+$. Note that the analytic signal is given by the product of the amplitude, $a(t)$, and the complex exponential of the phase, $e^{i\varphi(t)}$. In particular, if the support of the FT of $a(t)$ is the interval $[A_{min}, A_{max}]$ and the support of the FT of $e^{i\varphi(t)}$ is the interval $[P_{min}, P_{max}]$, it is sufficient to have

$$A_{min} + P_{min} > 0 \tag{3.7}$$

for the phase of the analytic signal to be meaningful.

An important remark must be made about the extraction of phases of linear combinations of signals. Since this thesis deals with measurements which result from linear mixtures of sources, one must show that the analytic signal of a mixture of those sources is equal to the corresponding mixture of the analytic signals of the sources. In mathematical terms, we must show that, if

$$y(t) = \sum_j s_j(t), \tag{3.8}$$

then the analytic signal of $y(t)$, denoted by $\tilde{y}(t)$, will obey

$$\tilde{y}(t) = \sum_j \tilde{s}_j(t), \tag{3.9}$$

where $\tilde{s}_j(t)$ is the analytic signal of the $j$-th source. The proof is straightforward: one must merely note that the analytic signal $\tilde{x}(t)$ is a linear function of its input signal $x(t)$, since the operations in equations (3.1), (3.2) and (3.3) are all linear.

We can therefore conclude that in any situation where one deals with signals which are the result of linear mixtures of sources, all of which have

non-negative amplitudes and obey condition (3.7), the analytic signal can be employed as a procedure which allows us to extract meaningful phases from real signals.

## 3.2 Phase Synchrony

The original motivation of this work was to develop source separation algorithms appropriate to the study of synchrony in brain electrophysiological signals, such as the electroencephalogram (EEG) and the magnetoencephalogram (MEG). An in-depth overview of neuroscience is completely out of the scope of this thesis. The following subsection intends to provide a starting set of references for a reader unfamiliar with this field, motivate why synchrony is relevant for neuroscience, and why it is an interesting topic in itself. A reader interested in neuroscience fundamentals may find [30] to be a good starting point.

Before discussing synchrony in the neuroscience domain, it is important to remark that synchrony is prevalent in many other physical systems, such as organ pipes, electrical circuits, laser beams, astrophysical objects, some types of fireflies, and even among humans and members of other mammal species. More examples, as well as very good overviews of the mathematical formulation of synchrony, can be found in [46, 65, 77].

### 3.2.1 Neuroscience Motivation

The number of neurons in the human brain was originally grossly estimated to be around $10^{11}$ [89, 6]. Recent, precise estimates place this number around $8.6 \times 10^{10}$ [35]. The firing of a neuron involves the travel of an action potential along its axon. These action potentials trigger the release of neurotransmitters in the connections between neurons (called *synapses*), and these in turn cause electrical potentials in the post-synaptic neurons. While the electrical or magnetic activity of a single post-synaptic potential is not measurable through the skull, if multiple neurons fire simultaneously, and if their post-synaptic receptors (the *dendrites*) are aligned, this activity may be measurable as an EEG or MEG signal. Good overviews of the physical phenomena behind EEG and MEG signals can be found in [31, 59, 72].

The neuroscience community has shown a great deal of interest in synchrony. While researchers had measured synchrony in mammal brains in

| Disorder | Anomalies in neural synchrony |
|---|---|
| Schizophrenia | Reduction of local- and long-range synchronization |
| Epilepsy | Increase in local synchrony; evidence for a reduction in long-range synchronization |
| Autism | Reduced functional connectivity; preliminary evidence for impaired neural synchrony |
| Alzheimer's disease | Reduced neural synchrony during resting state; evidence for reduced functional connectivity |
| Parkinson's disease | Increase in neural synchrony in the basal ganglia, but also between subcortical-cortical structures |

**Table 3.1.** Neurological pathologies associated with anomalous synchrony patterns of the human brain. Adapted from [84, Table 1].

the late 1980s [27], to the author's knowledge, widespread interest from neuroscientists began in the mid 1990s [20], and the second half of that decade saw several experimental verifications that hinted at the role of synchrony in the brain. Two of the most impactful works are, perhaps, [82] and [48]. Of particular relevance to this thesis, as motivating factors, are the verification that the scalp conducts electrical activity and that, therefore, the EEG signals can be considered to be a mixture of sources from inside the scalp [60], and the finding of correlations between several pathologies and anomalous synchrony patterns in the brain [84]. Table 3.1 presents a list of some pathologies and of the corresponding synchrony anomalies.

Today, synchrony continues to be an active area of research in the neuroscience community. Many influential researchers consider phase synchrony a fundamental mechanism for understanding the human brain [74, 23, 24, 90, 25], and it has been found to be involved in many brain functions. Findings include:

- An involvement of synchrony in the processing of learning from mistakes, which appears to be similar between rodents and humans [56].

- Preliminary evidence that the hippocampus, one of the central areas in the brain, "tunes" into different frequencies depending on the need to use path memorization ($\approx$35 Hz) or visual landmarks ($\approx$60 Hz) in orientation [14]. There is now speculation that synchrony plays a role

in prioritizing different sources of information.

- Indications that the memorization of smells, at least in rodents, is connected to synchronized activity between the entorhinal cortex (part of the medial temporal lobe) and the hippocampus, with a frequency around 20 Hz [40].

Also, despite the fact that the association between synchrony and some pathologies has been known for quite a while [84], recent findings keep challenging old knowledge. For example, recent data suggest not only a reduction in local and long-range connectivity in schizophrenia, but also an increase in connectivity in the brain's default mode network [67, 88].[1]

Apart from synchrony, researchers have used other criteria to measure connectivity in the brain. Among these, coherence is a popular choice. For example, [86] applied ICA to MEG recordings and then measured coherence between the resulting sources (in subsequent work [55], a similar approach was presented replacing coherence with synchrony). [57] proposed using the imaginary part of coherence to detect interaction between brain regions, although this was not used to perform source separation. In subsequent work, the same group proposed a BSS technique based on diagonalization of anti-symmetrized cross-correlation matrices [58]. BSS, in particular ICA, has also seen widespread use as a tool for artifact removal in EEG [54] and MEG [87].

A popular measure of synchrony in neuroscience studies is the Phase Locking Factor (PLF), sometimes also called Phase Locking Value (PLV) (see, *e.g.*, [81, 80, 68, 22, 76]). It is introduced in the next section.

### 3.2.2 Phase Locking Factor

This section provides a formal way to measure synchrony between two signals. Unlike in section 3.1, in this section and the remainder of the document, all signals are discrete-time signals.

Given a complex discrete-time signal $s(t)$, we define its *phase* as its argument: $\phi(t) \equiv \arg[s(t)]$. Consider now two signals, $s_j(t)$ and $s_k(t)$, with phases $\phi_j(t)$ and $\phi_k(t)$, respectively, and define their *phase lag* as

[1] The default mode network is part of the rest-state network, a network of regions in the brain that have increased activity when humans are not performing any specific task. The activity in these regions is measurably reduced when humans engage in tasks.

$\Delta\phi_{jk}(t) \equiv \phi_j(t) - \phi_k(t)$. The *Phase Locking Factor* (PLF) between these two signals is defined as

$$\varrho_{jk} \equiv \left| \frac{1}{T} \sum_{t=1}^{T} e^{i\Delta\phi_{jk}(t)} \right| = \left| \left\langle e^{i\Delta\phi_{jk}(t)} \right\rangle \right|, \tag{3.10}$$

where $\langle \cdot \rangle$ is the time average operator. The PLF has the following properties:

1. $0 \leq \varrho_{jk} \leq 1$;

2. $\varrho_{jk} = 1$ if and only if $e^{i\Delta\phi(t)}$ does not depend on $t$. This is equivalent to $\Delta\phi(t)$ being constant modulo $2\pi$. A third equivalent statement is that the phase of each source is equal to a source-dependent and time-independent term plus a source-independent and time-varying term: $\phi_j(t) = \phi_j + \phi(t)$.

3. In particular, the PLF of a signal with itself is 1: $\varrho_{jj} = 1$.

4. $\varrho_{jk}$ is invariant to the scale of the signals: if $s_j$ is multiplied by $\alpha_j$ and $s_k$ is multiplied by $\alpha_k$, where $\alpha_j$ and $\alpha_k$ are two nonzero complex scalars, then the PLF between $\alpha_j s_j$ and $\alpha_k s_k$ is the same as the PLF between $s_j$ and $s_k$.

$\varrho_{jk} = 0$ occurs, by definition, if and only if $\sum_{t=1}^{T} e^{i\Delta\phi_{jk}(t)} = 0$. Unlike the $\varrho_{jk} = 1$ case which is equivalent to a precise characterization of the two signals, the case $\varrho_{jk} = 0$ can correspond to vastly different situations. In the limit where the observation period $T$ tends to $+\infty$, and under ergodicity conditions, the following cases, among many others, yield zero PLF:

- $\phi_j(t) = 0$, and $\phi_k(t)$ uniformly distributed in $[0, 2\pi)$;

- $\phi_j(t)$ and $\phi_k(t)$ are both uniformly distributed in $[0, 2\pi)$ and are statistically independent;

- $\Delta\phi_{jk}(t) = \alpha t$ where $\alpha$ is not a multiple of $2\pi$.

For a finite observation period $T$, even these cases may yield non-zero values of $\varrho_{jk}$, which will tend to become smaller as the observation period grows larger.

We can now provide a precise definition of the term “synchrony”: if $\varrho_{jk} = 1$ we say that signals $s_j(t)$ and $s_k(t)$ are *perfectly synchronized*; if $\varrho_{jk} = 0$ we say that they are *perfectly unsynchronized*; if $0 < \varrho_{jk} < 1$ we say that they are *partially synchronized*.

# 4. Separation of Synchronous Sources

We now have all the elements to formally state the Separation of Synchronous Sources problem: a linear and instantaneous BSS problem, following model (2.3), is called a *Separation of Synchronous Sources (SSS)* problem if all pairs of sources are assumed to have pairwise PLFs of 1. We will later consider perturbed versions of this problem and still call it SSS.

One of the main motivations behind this work is the empirical verification that mixing destroys synchrony information, a fact which we verified empirically in multiple works. Consider, for example, the sources depicted in the top-left panel of figure 4.1, adapted from **[Publication V]**. These sources have pairwise PLFs of 1 with one another, as depicted in the top-right panel. If these sources are mixed using a $3 \times 3$ matrix whose entries are random and drawn from a Uniform$(-1, 1)$ distribution, a typical result is shown in the bottom-left panel of the figure. As shown on the bottom-right panel, the PLFs of these mixtures are no longer equal to 1, although signals 2 and 3 still exhibit a rather high mutual PLF. One of the main results in this work is a proof that, in the presence of perfectly synchronous sources, mixing will always result in imperfectly synchronous sources (theorem 4.1.1), showing that the example in figure 4.1 is not a coincidence. Another main result of this work is the design, implementation and analysis of two algorithms to solve the SSS problem. The two algorithms are very different in their philosophy:

- In Independent Phase Analysis (IPA), the basic rationale is as follows: Theorem 4.1.1 states that, under certain conditions, maximizing the pairwise PLFs will yield a solution equal to the original sources, apart from the usual indeterminacies. IPA explicitly maximizes an objective function which is the sum of the squares of all pairwise PLFs, plus a

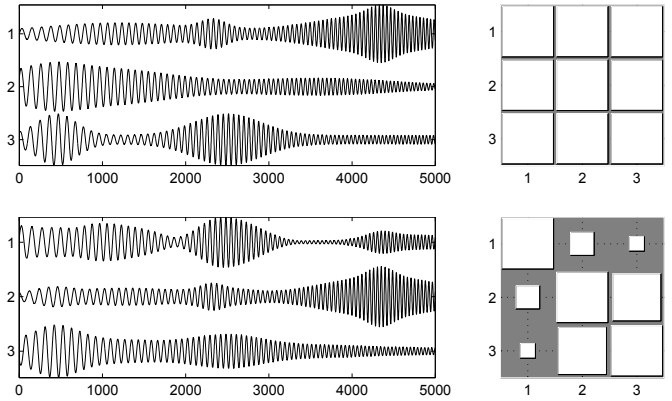


**Figure 4.1.** Top row: The real part of the three original sources (left) and PLFs between them (right). Bottom row: The real part of the three mixed signals (left) and PLFs between them (right). On the right column, the area of the square in position $(i, j)$ is proportional to the PLF between the signals $i$ and $j$. Therefore, large squares represent PLFs close to 1, while small squares represent values close to zero. In this example, the second and third sources have phase lags of $\frac{\pi}{6}$ and $\frac{\pi}{3}$ radians relative to the first source, respectively.

regularization term which penalizes singular solutions.

- In Phase Locked Matrix Factorization (PLMF), the basic rationale is as follows: if the sources are synchronous, then they can be expressed as a product of matrices with specific properties, which are presented below. PLMF explicitly models the sources as such a product and tries to estimate each factor separately. In doing this, singular solutions are automatically avoided.

These algorithms are explained in more detail in the next subsections, along with their theoretical properties and experimental results. Throughout this section, all signals are assumed to be complex-valued.

## 4.1 Problem Definition and Identifiability

A crucial aspect of SSS is that it is "sufficiently well-defined" in a sense similar to the one in which ICA was, but with different assumptions. The following theorem and corollary (both derived in **[Publication III]**[1]) establish that fact.

**Theorem 4.1.1.** *Consider a set of $N$ sources which follow the SSS model,*

[1]The proof of the corollary was somewhat unclear in **[Publication III]**. For that reason, we present it here again in a clearer form.

*such that* $s_k(t) \equiv A_k(t)e^{i(\phi(t)+\phi_k)}$. *Also, consider a linear combination of those sources, given by* $\hat{\mathbf{s}} = \mathbf{G}\mathbf{s}$ *with square* $\mathbf{G}$. *Furthermore, assume that:*

1. *None of the sources and none of the linear combinations are identically zero.*

2. $\phi_1, \phi_2, \ldots$ *are all distinct modulo* $\pi$.

3. *The amplitudes* $A_i(t)$ *are linearly independent (*i.e.*, the matrix* $\mathbf{A}$*, with entry* $(i,t)$ *given by* $A_i(t)$*, has maximum row rank) and positive.*

*Then, if the components of* $\hat{\mathbf{s}}$ *have the form*

$$\hat{s}_j(t) = C_j(t)e^{i(\phi(t)+\alpha_j)}, \tag{4.1}$$

*with positive amplitudes* $C_j$ *one necessarily has, for every* $j$*, that* $\hat{s}_j = Ls_k$ *for some* $k$*, where* $L$ *is a non-zero real number. Equivalently, each row of* $\mathbf{G}$ *has exactly one non-zero element.*

**Corollary 4.1.2.** *In the conditions of theorem 4.1.1, the following two statements are equivalent:*

1. *For all* $j \neq k$, $\hat{s}_j$ *and* $\hat{s}_k$ *are linearly independent.*

2. $\mathbf{G}$ *is non-singular and is thus a permutation of a diagonal matrix with nonzero diagonal elements.*

*In particular, to successfully extract all the original sources up to permutation, scale and sign change,* $\mathbf{G}$ *must be non-singular.*

*Proof.* Proving $1 \implies 2$ is trivial: suppose $\mathbf{G}$ is singular. Since each of its rows has at most one non-zero element, then two of its rows (say, rows $j$ and $k$) have non-zero elements in the same column (say, column $\ell$). This immediately implies that $\hat{s}_j = g_{j\ell}s_\ell$ and $\hat{s}_k = g_{k\ell}s_\ell$, and thus $\hat{s}_j$ and $\hat{s}_k$ are linearly dependent.

We now prove that $2 \implies 1$. This is also straightforward: if for a certain pair $j, k$, we have $\hat{s}_j = L\hat{s}_k$, then by theorem 4.1.1 these two mixtures must both be equal to the same source up to scale and sign: $\hat{s}_j = L\hat{s}_k = Ks_\ell$. This means that rows $j$ and $k$ of $\mathbf{G}$ both have exactly one nonzero element in the $\ell$-th column, thus making $\mathbf{G}$ singular. □

This theorem and corollary state that, if we find a linear combination of the original sources with the form (4.1), and if the assumptions are met, then we have found the original sources up to the typical BSS indeterminacies. Note, however, the specific form of (4.1): the linear combinations $\hat{\mathrm{s}}$ have *exactly the same common oscillation* $\phi(t)$ as the original sources $s$. We will use a theorem presented later (theorem 4.3.3, in section 4.3.4) to show that, under mild assumptions, all linear combinations with pairwise PLFs of 1, *i.e.*, combinations of the form

$$\hat{s}_j(t) = C_j(t)e^{i(\psi(t)+\alpha_j)}, , \tag{4.2}$$

must have $\psi(t) = \phi(t) + \beta$, where $\beta$ is some real number. Since $\beta$ can be absorbed by the $\alpha_j$ phase offsets, one can actually state that one can only find linear combinations of the sources which have pairwise PLFs of 1 if those linear combinations have the form (4.1).

Therefore, this theorem and corollary assert that, similarly to the ICA case, in SSS the sources can be recovered through maximization of the PLF, with the permutation, scale and sign indeterminacies. The requirements on the data are quite different from those of ICA, though. Let us now take a closer look at those requirements:

1. ICA is identifiable only if one has at most one Gaussian source – no such constraint is needed for SSS.

2. On the other hand, SSS's identifiability requires that no pair of sources be either in-phase or in anti-phase.

3. SSS's identifiability also requires that the amplitudes of the sources be linearly independent.

4. SSS's identifiability also requires that the gain matrix G is non-singular. Since the mixing matrix M is assumed non-singular in BSS problems, this requires that the demixing matrix be non-singular. While this is also a requirement for ICA identifiability, we shall see later that fulfilling this requirement is considerably harder in SSS algorithms.

The reader is referred to the proof of theorem 4.1.1 **[Publication III]** for the mathematical reasons for requirements 2–4. We now present simple counterexamples illustrating why each of these requirements is needed.

**Example 4.1.3.** *Let $s_1$ and $s_2$ be two signals exactly in-phase, with a phase lag of zero:*

$$\begin{aligned} s_1(t) &= a_1(t)e^{i\phi(t)} \\ s_2(t) &= a_2(t)e^{i\phi(t)}, \end{aligned} \tag{4.3}$$

*with $a_1(t) = 1$ and $a_2(t) = 2 + t$, with $t = 1, 2, \cdots, T$. Let*

$$\mathbf{G} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}, \tag{4.4}$$

*and thus $\hat{s}_1(t) = b_1(t)e^{i\phi(t)}$ and $\hat{s}_2(t) = b_2(t)e^{i\phi(t)}$ with $b_1(t) = 3 + t$ and $b_2(t) = 1 + t$. Clearly, all requirements are satisfied except that the phase lag is not different from 0 modulo $\pi$.*

*Since $b_1(t), b_2(t) > 0$ for all $t$, the PLF between $\hat{s}_1$ and $\hat{s}_2$ is 1. However, $\hat{s}_1$ and $\hat{s}_2$ are not equal to the original sources $s_1$ and $s_2$, even considering permutation, scaling and sign indeterminacies.*

**Example 4.1.4.** *Let $s_1$ and $s_2$ be two signals with linearly dependent amplitudes and with a phase lag $\Delta\phi \notin \{0, \pi\}$:*

$$\begin{aligned} s_1(t) &= a(t)e^{i\phi(t)} \\ s_2(t) &= 2a(t)e^{i\phi(t)+\Delta\phi}, \end{aligned} \tag{4.5}$$

*with $t = 1, 2, \cdots, T$. Let*

$$\mathbf{G} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}. \tag{4.6}$$

*Clearly, all requirements are satisfied except that the amplitudes of the sources are linearly dependent.*

*Since*

$$\hat{s}_1(t) = a(t)\left[e^{i\phi(t)} + 2e^{i\phi(t)+\Delta\phi}\right] = a(t)e^{i\phi(t)}\left[2e^{i\Delta\phi} + 1\right] \tag{4.7}$$

$$\hat{s}_2(t) = a(t)\left[-e^{i\phi(t)} + 2e^{i\phi(t)+\Delta\phi}\right] = a(t)e^{i\phi(t)}\left[2e^{i\Delta\phi} - 1\right], \tag{4.8}$$

*the PLF between $\hat{s}_1$ and $\hat{s}_2$ is 1. However, $\hat{s}_1$ and $\hat{s}_2$ are not equal to the original sources $s_1$ and $s_2$, even considering permutation, scaling and sign indeterminacies.*

**Example 4.1.5.** *Let $s_1$ and $s_2$ be any two signals, and let*

$$\mathbf{G} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \tag{4.9}$$

*Clearly, all requirements can be made to be satisfied except that the mixing matrix is singular.*

*We have $\hat{s}_1(t) = \hat{s}_2(t) = s_1(t) + s_2(t)$. Since the PLF of a signal with itself is 1 (see equation* (3.10)*), the PLF of signals $\hat{s}_1$ and $\hat{s}_2$ is 1. However, $\hat{s}_1$ and $\hat{s}_2$ are not equal to the original sources $s_1$ and $s_2$, even considering permutation, scaling and sign indeterminacies.*

Requirements 2 and 3, that the amplitudes be linearly independent and that phase lags be different from 0 modulo $\pi$, can be considered mild requirements: they will be met in the vast majority of situations. Note, however, that in practice we will always be dealing with a finite observation period $T$. In that case, as will be empirically shown in section 4.3, the performance of SSS algorithms is stable for phase lags that are far from 0 and $pi$, but degrade as they approach those values.

Requirement 4 is much more profound. It corresponds to a fundamental difference between ICA and SSS:

- In ICA, if the mixing matrix is non-singular, "maximizing independence" poses no risk of leading the algorithm towards singular unmixing matrices, since a signal is not independent of itself.[2]

- In SSS, even with a non-singular mixing matrix, maximizing the PLF poses the risk of leading the algorithm towards singular matrices, since a solution with $\hat{s}_1 = \hat{s}_2 = \ldots = \hat{s}_P$ trivially makes all PLFs equal to 1. Therefore, some means must be employed to prevent this from happening.

We shall see in section 4.2 that this requirement is, to some extent, a drawback of IPA, one of the SSS algorithms proposed in this thesis.

## 4.2 Algorithm: Independent Phase Analysis

Independent Phase Analysis (IPA) explicitly maximizes the PLF of the estimated sources as a function of the demixing matrix $\mathbf{W}$. Let $\hat{\mathbf{s}}(t) \equiv \mathbf{W}^T\mathbf{y}(t)$ denote the vector of estimated sources, let $\hat{s}_j(t)$ denote the $j$-th estimated source, and let $\Delta\hat{\phi}_{jk}(t)$ denote the phase lag between estimated sources $j$ and $k$. These estimated quantities are, naturally, functions of

[2] In fact, ICA algorithms do not maximize independence, but rather optimize some surrogate of it. Even so, the most frequently used ICA algorithms do avoid singular solutions. In some of them, the search is reduced to the space of orthogonal matrices, which immediately avoids singular matrices.

the demixing matrix $\mathbf{W}$ – this dependency will sometimes be omitted for clarity in the following. The PLF between estimated sources $j$ and $k$, denoted by $\hat{\varrho}_{jk}$, is given by:

$$\hat{\varrho}_{jk}(\mathbf{W}) \equiv \left| \frac{1}{T} \sum_{t=1}^{T} e^{i\Delta\hat{\phi}_{jk}(t)} \right| = \left| \left\langle e^{i\Delta\hat{\phi}_{jk}(t)} \right\rangle \right|. \tag{4.10}$$

This estimated PLF is also a function of $\mathbf{W}$, since $\Delta\hat{\phi}_{jk}$ is a function of it.

IPA maximizes the following objective function:

$$J(\mathbf{W}) \equiv (1-\lambda) \sum_{j \neq k;\, j,k=1,\ldots,N} \hat{\varrho}_{jk}^2(\mathbf{W}) + \lambda \log|\det \mathbf{W}|. \tag{4.11}$$

The first term of $J(\mathbf{W})$ is the sum of the squares of all PLFs between pairs of different sources.[3] The second term is a regularization term which we will discuss further below.

IPA constrains the rows of $\mathbf{W}$ to have unit Euclidean norm. The set of $N \times N$ matrices whose rows have unit norm will be denoted as $\mathcal{S}$. Thus, the optimization problem to be solved in IPA is

$$\max_{\mathbf{W} \in \mathcal{S}} J(\mathbf{W}) \tag{4.12}$$

We proved **[Publication IV]** that the gradient of $J$ relative to a column $\mathbf{w}_j$ of the matrix $\mathbf{W}$ is given by

$$\nabla_{\mathbf{w}_j} J = 4\frac{1-\lambda}{N^2} \sum_{k=1}^{N} \hat{\varrho}_{jk} \left\langle \sin\left[\hat{\Psi}_{jk} - \Delta\hat{\phi}_{jk}(t)\right] \frac{\mathbf{\Gamma}(t)}{Y_j(t)^2} \right\rangle \mathbf{w}_j, \tag{4.13}$$

where $S_j \equiv |\hat{s}_j|$ where $\hat{s}_j$ is the $j$-th estimated source, $\hat{\phi}_j \equiv \text{angle}(\hat{s}_j)$ and $\Delta\hat{\phi}_{jk}(t) \equiv \hat{\phi}_j(t) - \hat{\phi}_k(t)$ is the instantaneous phase difference between two estimated sources, $\hat{\Psi}_{jk}(t) \equiv \langle\Delta\hat{\phi}_{jk}(t)\rangle$ is the average phase difference between two estimated sources, and $\mathbf{\Gamma}(t) = \mathbf{y_h}(t)\mathbf{y}^T(t) - \mathbf{y}(t)\mathbf{y_h}^T(t)$, where $\mathbf{y_h} = \text{Imag}(\tilde{\mathbf{y}})$ is the imaginary part of $\mathbf{y}$. $\mathbf{\Gamma}(t)$ is a matrix that can be pre-computed, before the optimization of $J(\mathbf{W})$, because it depends only on the data.

### 4.2.1 Regularization

Let us temporarily focus on the case $\lambda = 0$. For this case, we have $0 \leq J(\mathbf{W}) \leq N(N-1)$. The reason for this is that each $\hat{\varrho}_{jk}^2$ term in the summation is between 0 and 1, and there are $N(N-1)$ such terms.

[3] Note that, since the PLF of a signal with itself is always 1, we could include the case $j = k$ in the summation without changing the solution.

If the assumptions of the SSS problem hold (*i.e.*, if the original sources have PLFs of 1 with one another and the mixing is linear and instantaneous), then one global maximizer of $J(\mathbf{W})$ corresponds to the case where the estimated sources are equal to the original sources. This will yield $\hat{\varrho}_{jk} = \varrho_{jk} = 1$ for all $j, k$, and this is the maximum value each of the $\hat{\varrho}_{jk}$ can take. This case corresponds to making $\mathbf{W}^T = \mathbf{M}^{-1}$, up to permutation, scaling and sign change.

Theorem 4.1.1 ensures that the solution $\mathbf{W}^T = \mathbf{M}^{-1}$ is the only one where $\det(\mathbf{W}) \neq 0$, apart from permutation, scaling and sign change. In more rigorous terms, there are multiple non-singular solutions of the form $\mathbf{W}^T = \mathbf{PDM}^{-1}$, where $\mathbf{P}$ is some permutation matrix, and $\mathbf{D}$ is a diagonal matrix with nonzero entries in its diagonal.[4] We call all solutions of this form *desirable solutions* of IPA. However, there are many other global maximizers of $J(\mathbf{W})$, which have $\det(\mathbf{W}) = 0$. For example, there are $N$ solutions of the form

$$\hat{s}_1 = \hat{s}_2 = \ldots = \hat{s}_N = s_i, \tag{4.14}$$

for some $i \in \{1, \ldots, N\}$. This solution corresponds to making all estimated sources equal to the $i$-th true source. It is equivalent to having a matrix $\mathbf{W}$ such that the gain matrix has all entries in the $i$-th row equal to 1 and all other entries equal to zero. It is simple to see that this is a global maximizer for $\lambda = 0$: all estimated sources are equal to one another, thus their pairwise PLFs are all equal to 1. Apart from these $N$ solutions, there are many other ones, which are not good solutions, in the sense that they are global maximizers of $J(\mathbf{W})$, for $\lambda = 0$, but do not recover the original sources. We call *undesirable solutions* all global maximizers of $J(\mathbf{W})$ that are not desirable solutions. Theorem 4.1.1 ensures that all undesirable solutions have $\det(\mathbf{W}) = 0$.

The fact that the maximizers of the first term of $J(\mathbf{W})$ can correspond to desirable or undesirable solutions is the motivation for the second term, which explicitly penalizes undesirable solutions: these solutions make the second term equal to $-\infty$, whereas desirable solutions make it finite. The parameter $\lambda$ controls the tradeoff between the first term and the regularization one.

The previous reasoning motivates the fact that $0 < \lambda < 1$ is a better choice than $\lambda = 0$. However, there is a significant disadvantage in using regularization: Theorem 4.1.1 characterizes the maximizers of the first

[4] The diagonal entries of $\mathbf{D}$ are not free: they must be such that the rows of $\mathbf{W}$ have unit norm.

term, but the objective function now also contains a regularization term. Theorem 4.1.1 now does not characterize solutions for this case. In general, for $0 < \lambda < 1$, maximizers of $J(\mathbf{W})$ will not exactly correspond to desirable solutions.

The requirement for the columns of $\mathbf{W}$ to be normalized can now be justified: if $\mathbf{W}$ was not constrained, $|\det \mathbf{W}|$ could be made arbitrarily large, without affecting the first term, by scaling the columns of $\mathbf{W}$.

### 4.2.2 Optimization Strategy

The existence of multiple solutions for $\lambda = 0$ shows that IPA is not a concave optimization problem [13] and we have verified, in practice, that optimizing IPA is difficult if a fixed value of $\lambda$ is used. Using small values (or zero) for $\lambda$ tends to yield singular solutions often. Ideally, one would wish to avoid undesirable solutions, which would force us to use relatively large values of $\lambda$, but one would also wish the assurance, through Theorem 4.1.1, that a solution with $\det \mathbf{W} \neq 0$ is a desirable one, which would force us to use $\lambda = 0$. To reconcile these two aspects, we proposed in **[Publication V]** that the value of $\lambda$ should start at a relatively large value, and be decreased throughout the optimization, such that at the end $\lambda$ is zero. The reasoning is that the initial stages, with large $\lambda$, force $\mathbf{W}$ away from undesirable solutions; the later stages, with smaller $\lambda$, corresponds to the optimization of functions which are progressively better approximations of the first term, and are therefore expected to yield results that are progressively closer to desirable solutions, yielding a desirable solution in the final step, with $\lambda = 0$. We have no theoretical guarantee that this method will yield a desirable solution. However, we have verified in practice that it yields much better results than using a fixed value of $\lambda$, as shown in section 5.1.

In the results that follow, each epoch with a fixed $\lambda$ is optimized with 200 iterations of gradient ascent, as shown in table 4.1. After this, MATLAB's implementation of BFGS is run until convergence. The solution found in this manner is used to initialize the optimization with the next value of $\lambda$.

## 4.3 Algorithm: Phase Locked Matrix Factorization

As discussed in the previous section, non-regularized IPA ($\lambda = 0$) suffers from the drawback of allowing undesirable solutions, and regularized IPA

| | INDEPENDENT PHASE ANALYSIS |
|---|---|
| 1: | Input $\mathbf{y}(t)$, $\eta$, $k_{max}$ |
| 2: | $\tilde{\mathbf{y}}(t) \leftarrow$ analytic signal of $\mathbf{y}(t)$ (eqs. (3.1), (3.2), (3.3)) |
| 3: | $\varphi_j(t) \leftarrow \text{angle}(\tilde{y}_j(t))$, $j = 1, \ldots, N$ |
| 4: | $Y_j(t) \leftarrow \lvert\tilde{y}_j(t)\rvert$ |
| 5: | Initialize $\mathbf{W} \sim \mathcal{N}(0, 1)$; set $k = 1$ |
| 6: | **repeat** |
| 7: | $\hat{\tilde{\mathbf{s}}}(t) \leftarrow \mathbf{W}^T \tilde{\mathbf{y}}(t)$ |
| 8: | $\Delta\mathbf{W} \leftarrow$ Eq. (4.13) |
| 9: | $\mathbf{W} \leftarrow \mathbf{W} + \eta\Delta\mathbf{W}$ |
| 10: | $\mathbf{w}_j \leftarrow \mathbf{w}_j / \lVert\mathbf{w}_j\rVert$, $j = 1, \ldots, N$ |
| 11: | $k \leftarrow k + 1$ |
| 12: | **until** $(\lVert\Delta\mathbf{W}\rVert < \epsilon)$ or $(k > k_{max})$ |

**Table 4.1.** Pseudocode for the fixed-$\lambda$ version of IPA using gradient ascent. The varying $\lambda$ version merely runs the fixed version until convergence multiple times with decreasing values of $\lambda$.

($\lambda > 0$) suffers from a bias due to the regularization term. These disadvantages motivated us to design, in a more principled way, an SSS algorithm which would guarantee exact separation. Phase Locked Matrix Factorization (PLMF), the algorithm detailed in this section, accomplishes this, and yields substantially better separation results than IPA.

### 4.3.1 General Approach

Whereas IPA tries to estimate the unmixing matrix $\mathbf{W}$ directly from the data, PLMF is based on a factorized model of synchronous sources. Let us decompose, elementwise, the complex-valued sources into their absolute values and arguments:

$$\mathbf{S} = \mathbf{A} \odot \mathbf{\Phi}, \tag{4.15}$$

where the entries of $\mathbf{A}$ are real-valued and non-negative, those of $\mathbf{\Phi}$ are complex-valued and have unit absolute value, and the symbol $\odot$ denotes elementwise (or Hadamard) product. We call $\mathbf{A}$ and $\mathbf{\Phi}$ the sources' *amplitudes* and *phases*, respectively.

If the sources $\mathbf{S}$ are perfectly synchronous, then the phases $\mathbf{\Phi}$ can be further decomposed as

$$\mathbf{\Phi} = \mathbf{z}\mathbf{f}^T, \tag{4.16}$$

where $\mathbf{z}$ and $\mathbf{f}$ are complex-valued vectors of sizes $N$ and $T$, respectively, and whose entries all have unit absolute value. To see why this decomposition is always possible, recall from section 3.2.2 that, if a set of sources are perfectly synchronous, then the phase of the $j$-th source could be written as $\phi_j(t) = \phi_j + \psi(t)$. Equation (4.16) follows from setting the $j$-th element of $\mathbf{z}$ as $e^{i\phi_j}$, and the $t$-th element of $\mathbf{f}$ as $e^{i\psi(t)}$. Notice that there are more assignments that uphold equation (4.16): one can also set the $j$-th element of $\mathbf{z}$ to $e^{i\phi_j}e^{i\gamma}$ and the $t$-th element of $\mathbf{f}$ to $e^{i\psi(t)}e^{-i\gamma}$, for any real scalar $\gamma$. This is a new indeterminacy, specific to PLMF, which we call the *rotation indeterminacy*. Note that, while these assignments yield different factors $\mathbf{z}$ and $\mathbf{f}$, they do not alter the sources $\mathbf{S}$.

Consecutively applying equations (2.8), (4.15) and (4.16) yields, for synchronous sources and no observation noise,

$$\mathbf{Y} = \mathbf{M}[\mathbf{A} \odot (\mathbf{z}\mathbf{f}^T)]. \tag{4.17}$$

If one defines $\mathbf{D_z}$ as a $N \times N$ diagonal matrix with diagonal entries equal to the entries of $\mathbf{z}$, and analogously defines $\mathbf{D_f}$ as a $T \times T$ matrix with diagonal entries equal to the entries of $\mathbf{f}$, then equation (4.17) can also be written as

$$\mathbf{Y} = \mathbf{M}\mathbf{D_z}\mathbf{A}\mathbf{D_f}. \tag{4.18}$$

The basic idea behind PLMF is to minimize the squared error of this model relative to the observed data:

$$\min_{\mathbf{M},\mathbf{A},\mathbf{z},\mathbf{f}} \frac{1}{2} \left\| \mathbf{Y} - \mathbf{M}\mathbf{D_z}\mathbf{A}\mathbf{D_f} \right\|_F^2, \tag{4.19}$$

$$\begin{aligned} \text{s.t.: } & 1) \max_{i,j} |m_{ij}| = 1 \\ & 2) |z_j| = 1 \text{ for all } j \\ & 3) |f_t| = 1 \text{ for all } t \end{aligned}$$

where $\|\cdot\|_F$ is the Frobenius norm. The first constraint forces the largest absolute value among all elements of $\mathbf{M}$ to be 1, a constraint that we shall discuss further below. The second and third constraints force $\mathbf{z}$ and $\mathbf{f}$ to have entries with unit absolute value. $\mathbf{M}$ and $\mathbf{A}$ are real, while $\mathbf{z}$ and $\mathbf{f}$ are complex ones.[5]

[5] We also experimented adding the constraint $\mathbf{A} \geq 0$, since our experiments will use amplitudes which obey this constraint. However, adding this constraint makes the algorithm sometimes become "stuck" with many entries of $\mathbf{A}$ equal to zero and unable to depart from that situation, even when the true matrix has only positive entries.

Let the term *desirable solutions* denote all solutions such that the sources $\mathbf{S} = \mathbf{D_z A D_f}$ are equal to the original sources up to permutation, scaling and sign change. Suppose that the data $\mathbf{Y}$ are indeed generated according to the PLMF model (4.17). Then, the true solution (*i.e.*, the values of $\mathrm{M}, \mathbf{A}, \mathbf{z}, \mathbf{f}$ used to generate the data) clearly makes the cost function in equation (4.19) attain its minimum possible value of zero. That solution also obeys constraints 2 and 3, by construction. The true solution will probably not obey constraint 1; however, it is easy to obtain a new solution that obeys it, by multiplying $\mathrm{M}$ by an appropriate scalar and dividing $\mathbf{A}$ by the same scalar. Since the sources are only affected by a scaling factor, this new solution is a desirable one.

There is, therefore, a global minimizer of the objective function of (4.19) which yields a desirable solution. Constraint 1 ensures that there is no scale indeterminacy, but due to the permutation and rotation indeterminacies, there is in fact an infinite number of desirable solutions, all of which minimize the cost function in (4.19).

The PLMF problem is non-convex [13]. There are two reasons for this:

- The cost function involves the product of several variables.

- The feasible sets, *i.e.*, the sets of values of the variables which obey constraints 1, 2, and 3, are not convex sets.

One of the results in this thesis is that, under certain conditions, all global optima of this problem correspond to desirable solutions, as will be shown further below.

The basic PLMF algorithm iteratively solves the minimization problem in (4.19). Each iteration is a sequence of four steps. At each step, three of the four variables are kept fixed, and the minimization problem is solved with respect to the remaining variable. This is known in the literature as a Block Nonlinear Gauss-Seidel (BNGS) method [28], or Block Coordinate Descent, or Alternating Optimization, and it is not guaranteed to converge to the optimal solution in general. It does converge under some assumptions [28], which are not met in our case. We discuss this aspect further in section 6.2.

The first version of PLMF tackled the minimization problem in (4.19) directly, *i.e.*, optimizing on all four variables using the BNGS method. However, we have since shown that PLMF can also be solved by finding a

correct value for $\mathbf{f}$ by solving a relaxed version of (4.19).[6] This value of $\mathbf{f}$ is then kept fixed, and the BNGS method is then applied to optimize on the three remaining variables $\mathbf{M}, \mathbf{A}, \mathbf{z}$. To distinguish these two approaches, when necessary, we will refer to the first one as the "1-stage" PLMF and to the second one as the "2-stage" PLMF. Whenever we do not make a distinction, we will be referring to the second version.

Experimental comparisons have shown that the 2-stage PLMF outperforms the 1-stage version. The reason for this, which we verified experimentally, is that 1-stage PLMF yields local minima frequently; the 2-stage version is not immune to this, but returns local minima less often. Hence, we now proceed to discuss only the 2-stage version; the 1-stage version was presented in **[Publication VI]**.

### 4.3.2 First subproblem

In the first stage of 2-stage PLMF, the goal is to estimate the common oscillation $\mathbf{f}$ up to the rotation indeterminacy. This estimation is performed by solving the subproblem

$$\min_{\mathbf{H},\mathbf{A},\mathbf{f}} \frac{1}{2} \|\mathbf{Y} - \mathbf{H}\mathbf{A}\mathbf{D}_{\mathbf{f}}\|_F^2 , \tag{4.20}$$

$$\begin{aligned} \text{s.t.: } & 1) \max_{i,j} |h_{ij}| = 1 \\ & 2) |f_t| = 1 \text{ for all } t, \end{aligned}$$

where $\mathbf{H}$ can be any complex matrix with the same dimensions as $\mathbf{M}$, as long as the largest absolute value among its entries is 1, to fulfill constraint 1, and $\mathbf{f}$ is complex with entries having unit absolute value, as before. This formulation collapses the product $\mathbf{M}\mathbf{D}_{\mathbf{z}}$ from (4.19) into the matrix $\mathbf{H}$, which is now allowed to be any complex matrix. Note that, despite the fact that we minimize relative to $\mathbf{H}, \mathbf{A}$, and $\mathbf{f}$, the purpose is only to estimate $\mathbf{f}$. The values found for $\mathbf{H}$ and $\mathbf{A}$ are discarded. Since the product of a real matrix $\mathbf{M}$ and a complex diagonal matrix $\mathbf{D}_{\mathbf{z}}$ does not span the space of all complex matrices, the minimization problem (4.20) is a relaxation of (4.19).

If the sources exactly follow the model in equation (4.18), a factorization of the form $\mathbf{Y} = \mathbf{H}\mathbf{A}\mathbf{D}_{\mathbf{f}}$ always exists, since the true factorization is a special case of it. The following theorem, proven in **[Publication XI]**, shows

[6] Recall that, due to the rotation indeterminacy, there are infinite correct values for $\mathbf{f}$. The method finds one of them.

that if $\mathbf{Y} = \mathbf{HAD_f}$ (in particular, if the sources follow the model in equation (4.18)), then finding a solution of (4.20) yields a correctly estimated $\mathbf{f}$, apart from a sign indeterminacy which can be easily compensated, and from the rotation indeterminacy.

**Theorem 4.3.1.** *Let* $\mathbf{Y} = \mathbf{H}_1\mathbf{A}_1\mathbf{D_{f1}}$ *with* $\mathbf{H}_1 \in \mathbb{C}^{P\times N}$, $\mathbf{A}_1 \in \mathbb{R}^{N\times T}$, $\mathbf{D_{f1}} \in \mathbb{D}_1^T$, *where* $\mathbb{D}_1^T$ *is the set of* $T$*-by-*$T$ *diagonal complex matrices whose diagonal entries have unit absolute value, and* $\mathbf{H}_1$ *has full column rank. If there is another factorization of the same form,* $\mathbf{Y} = \mathbf{H}_2\mathbf{A}_2\mathbf{D_{f2}}$, *then one necessarily has* $\mathbf{D_{f2}} = \mathbf{ED_{f1}}$ *where* $\mathbf{E} \in \mathbb{D}_1^T$ *is a diagonal matrix whose diagonal elements belong to the two-element set* $\{-e^{i\gamma}, +e^{i\gamma}\}$, *where* $\gamma$ *is some real number.*

This theorem only ensures a "quasi-identifiability" of $\mathbf{f}$, since $\mathbf{D_f}$ is determined up to multiplication by matrix $\mathbf{E}$. This means that we may not obtain the true $\mathbf{D_f}$, for two reasons (which may occur simultaneously):

1. The first possibility is that all entries of $\mathbf{D_f}$ are multiplied by $e^{i\gamma}$, *i.e.*, all its entries are rotated by an angle $\gamma$. This ambiguity corresponds to the rotation indeterminacy.

2. The second possibility is that some entries of the estimated $\mathbf{D_f}$ are multiplied by $+1$ and some by $-1$. This means that some entries of $\mathbf{D_f}$ are estimated with the wrong sign.

The first issue does not need to be solved at this point, since the rotation indeterminacy does not affect the estimated sources. It will be compensated when we estimate $\mathbf{z}$, in the second subproblem (section 4.3.3).

The second issue can easily be solved if the common oscillation $\mathbf{f}$ is *smooth*, *i.e.*, if it varies slowly with time. In that case, it is natural to expect that $f_{t+1}$ is not very different from $f_t$. Therefore, to correct this sign estimation, we compute, for $t = 1, \ldots, T-1$, the quantity

$$|f_R(t) - f_R(t+1)| + |f_I(t) - f_I(t+1)|, \tag{4.21}$$

where $f_R(t)$ is the real part of the $t$-th entry of $\mathbf{f}$, and $f_I(t)$ is the imaginary part of that entry. It is easy to show that, if $f_{t+1} = -f_t$, then this quantity lies between $\sqrt{2}$ and $2$. For a smoothly varying $\mathbf{f}$, we expect the values of (4.21) to be small if there is no change of sign from time $t$ to time $t+1$, and to be $\gtrapprox\sqrt{2}$ if such a sign change occurs from $t$ to $(t+1)$. In our simulations

we determine that there is a change in sign when

$$|f_R(t) - f_R(t+1)| + |f_I(t) - f_I(t+1)| > 1. \tag{4.22}$$

In the tests presented in section 5.2, this simple procedure successfully captures all sign changes. However, if $\mathbf{f}$ is not sufficiently smooth, better phase unwrapping techniques must be employed [12].

### 4.3.3 Second subproblem

The first subproblem yields an estimate of $\mathbf{f}$. To motivate the second subproblem, let us take the source model (4.18) and multiply both sides of the equation by the inverse of $\mathbf{D_f}$, on the right. Note that the inverse of $\mathbf{D_f}$ is its complex conjugate: $\mathbf{D_f}^* = \mathbf{D_f}^{-1}$. This yields

$$\mathbf{Y}\mathbf{D_f}^* = \mathbf{M}\mathbf{D_z}\mathbf{A}. \tag{4.23}$$

The second subproblem attempts to minimize the squared difference between both sides of this equation:

$$\min_{\mathbf{M},\mathbf{A},\mathbf{z}} \frac{1}{2} \left\| \mathbf{Y}\mathbf{D_f}^* - \mathbf{M}\mathbf{D_z}\mathbf{A} \right\|_F^2, \tag{4.24}$$

$$\text{s.t.: } 1) \max_{i,j} |m_{ij}| = 1$$
$$2) |z_j| = 1 \text{ for all } j.$$

One again has identifiability in this second subproblem, as shown by the following theorem, which was derived in **[Publication XI]**.

**Theorem 4.3.2** (Identifiability of M,A,z)**.** *Let* $\mathbf{Y}\mathbf{D_f}^* = \mathbf{M}\mathbf{D_{z1}}\mathbf{A}_1$ *with* $\mathbf{M}_1 \in \mathbb{R}^{P\times N}$*,* $\mathbf{D_{z1}} \in \mathbb{D}_1^N$*,* $\mathbf{A}_1 \in \mathbb{R}^{N\times T}$*, where* $\mathbb{R}^{N\times T}$ *denotes the set of* $N$*-by-*$T$ *matrices with real entries. Further assume that the phases of all sources are different from one another modulo* $\pi$ *(in other words, that two entries* $e^{i\alpha}$ *and* $e^{i\beta}$ *of the diagonal of* $\mathbf{D_{z1}}$ *never satisfy* $e^{i\alpha} = e^{i\beta}$ *nor* $e^{i\alpha} = -e^{i\beta}$*), and that* $\mathbf{A}_1$ *has maximum row rank. If there is another factorization of the same form,* $\mathbf{Y}\mathbf{D_f}^* = \mathbf{M}_2\mathbf{D_{z2}}\mathbf{A}_2$*, then one necessarily has* $\mathbf{M}_1 = \mathbf{M}_2$*,* $\mathbf{D_{z1}} = \mathbf{D_{z2}}$*, and* $\mathbf{A}_1 = \mathbf{A}_2$*, up to permutation, scaling, and sign change.*

Importantly, note that this theorem does not state that $\mathbf{D_z}$ is estimated up to rotation. This ensures that the rotation indeterminacy, which was a potential issue from the first subproblem, is no longer an issue. At the end of the second subproblem, if a global minimum of problem (4.24) is found, the theorem ensures that one will have a solution $\mathbf{M}, \mathbf{A}, \mathbf{z}, \mathbf{f}$ such

that $\mathbf{Y}\mathbf{D}_{\mathbf{f}}^{*} = \mathbf{M}\mathbf{D}_{\mathbf{z}}\mathbf{A}$, or equivalently, such that $\mathbf{Y} = \mathbf{M}\mathbf{D}_{\mathbf{z}}\mathbf{A}\mathbf{D}_{\mathbf{f}}$ up to permutation, scaling and sign change. While two different solutions may have vectors $\mathbf{z}$ and $\mathbf{f}$ which differ by arbitrary rotations, the two theorems ensure that both pairs $(\mathbf{z}, \mathbf{f})$ yield the same sources.

This theorem assumes that all the arguments of the entries in the diagonal of $\mathbf{D}_{\mathbf{z}}$ are different modulo $\pi$. A similar theorem can be proven for a more general case where $k$ diagonal elements violate this assumption, whereas the remaining $(N-k)$ obey it. In that case, $\mathbf{D}_{\mathbf{z}}$ is still identifiable. However, only $(N-k)$ rows of $\mathbf{A}$ and the corresponding $(N-k)$-by-$(N-k)$ block of $\mathbf{M}$ are identifiable. In other words, only the $(N-k)$ sources with distinct phase values (modulo $\pi$) are identifiable; the remaining sources will, in general, be mixed with one another in the estimated sources. A sketch of this proof was presented in **[Publication XI]** (see footnote 9).

### 4.3.4 Global identifiability

Given that both subproblems of PLMF are identifiable (or quasi-identifiable for the first one), one may naturally ask the question: is the original problem in (4.19) identifiable? In other words, if we find a solution of that problem (not necessarily with PLMF), can we be sure that we found the original sources up to the usual indeterminacies? By combining the two Theorems 4.3.1 and 4.3.2, the answer turns out to be affirmative, with one additional assumption. Along with the assumptions of both theorems, one also needs that the amplitudes $\mathbf{A}$ are non-negative, as shown by the following theorem, which was also derived in **[Publication XI]**.

**Theorem 4.3.3.** *Let* $\mathbf{Y}$ *be data generated according to the model in equation* (4.18) *with all the elements of* $\mathbf{A}$ *non-negative, and let* $\mathbf{Y} = \mathbf{M}_1\mathbf{D}_{\mathbf{z1}}\mathbf{A}_1\mathbf{D}_{\mathbf{f1}}$ *be a factorization of the data such that the entries of* $\mathbf{A}_1$ *are non-negative, the constraints of problem* (4.19) *are satisfied,* $\mathbf{M}_1$ *has full column rank, the phases of the entries of* $\mathbf{z}_1$ *are different modulo* $\pi$*, and* $\mathbf{A}_1$ *has maximum row rank. Let* $\mathbf{Y} = \mathbf{M}_2\mathbf{D}_{\mathbf{z2}}\mathbf{A}_2\mathbf{D}_{\mathbf{f2}}$ *be another such factorization. Then, the two factorizations are equal up to permutation, scaling, and rotation.*

This theorem also shows that there is no loss of generality in the assumption, in theorem 4.1.1, that the common oscillation of the mixed signals is equal to the common oscillation of the sources, if the sources' amplitudes are positive. Recall that we proved that, for the case where the number of sources is equal to the number of mixed signals, if the sources

were given by

$$s_k(t) \equiv A_k(t)e^{i(\phi(t)+\phi_k)}, \tag{4.25}$$

then, under the assumptions of theorem 4.1.1, the only way to construct a linear combination $\mathbf{y} = \mathbf{M}\mathbf{s}$ such that

$$y_j(t) = C_j(t)e^{i(\phi(t)+\alpha_j)} \tag{4.26}$$

was to have $y_i = Ks_j$ for some $i$ and $j$, where $K$ is a non-zero real number. At the time, we noted that (4.26) enforces a very particular form for the mixtures $\mathbf{y}$, since their common oscillation, $\phi(t)$, must be equal to that of the sources $\mathbf{s}$.

Using Theorem 4.3.3, it is straightforward to show that this must be the case, and therefore that no generality was lost in 4.1.1. One merely needs to decompose the sources according to 4.18:

$$\mathbf{S} = \mathbf{D}_{\mathbf{z}_s}\mathbf{A}_s\mathbf{D}_{\mathbf{f}_s}, \tag{4.27}$$

and similarly decompose the data as

$$\mathbf{Y} = \mathbf{D}_{\mathbf{z}_y}\mathbf{A}_y\mathbf{D}_{\mathbf{f}_y}. \tag{4.28}$$

Note that Theorem 4.1.1 assumes that the mixtures $\mathbf{y}$ have a PLF of 1, therefore the factorization (4.28) must exist.

We then use the equality $\mathbf{Y} = \mathbf{MS}$ and plug in equation (4.27) to obtain

$$\mathbf{Y} = \mathbf{M}\mathbf{D}_{\mathbf{z}_s}\mathbf{A}_s\mathbf{D}_{\mathbf{f}_s}. \tag{4.29}$$

Finally, we use the fact that the left-hand sides of equations (4.28) and (4.29) are the same, to obtain

$$\mathbf{D}_{\mathbf{z}_y}\mathbf{A}_y\mathbf{D}_{\mathbf{f}_y} = \mathbf{M}\mathbf{D}_{\mathbf{z}_s}\mathbf{A}_s\mathbf{D}_{\mathbf{f}_s}. \tag{4.30}$$

One can now use theorem 4.3.3 for the two factorizations in (4.30) to show that the common oscillation for the linear mixtures, $\mathbf{D}_{\mathbf{f}_y}$, must be the same as that of the sources, $\mathbf{D}_{\mathbf{f}_s}$. This implies that no generality is lost in assuming that any linear combinations of the sources that have a PLF of 1 must be of the form (4.26).

### 4.3.5 Optimization Strategy

The PLMF algorithm is presented in Table 4.2. We now explain in further detail how each of the two subproblems is tackled. We employ the BNGS method in both optimizations; in the first subproblem, we randomly initialize the variables $\hat{\mathbf{H}}$, $\hat{\mathbf{A}}$ and $\hat{\mathbf{f}}$, and iteratively optimize relative

to each of them, while keeping all others fixed (lines 4-8 of Table 4.2). Similarly, for the second subproblem, we randomly initialize $\hat{\mathbf{M}}$, $\hat{\mathbf{A}}$ and $\hat{\mathbf{z}}$ and optimize each of them while keeping all others fixed (lines 15-19 of Table 4.2). The use of BNGS has a great advantage: problems (4.20) and (4.24), which are hard to directly solve, in particular due to the presence of products of variables, are solved through an iteration of constrained least-squares problems, which we optimize in a simple, albeit suboptimal, procedure. There is a downside: BNGS is not guaranteed to converge. We discuss this aspect further in section 6.2.

The two subproblems (4.20) and (4.24) are convex in some variables and non-convex in other variables. Instead of trying to find the global minimum for a certain variable at each iteration, we chose to always solve for each variable without enforcing any constraints, then projecting that solution onto the feasible set; this projection is an approximation of the true solution. Our choice is motivated for three reasons: simplicity, because like this all variables are optimized in a similar way; speed, which allowed us to run the extensive experiments shown in section 5.2; and the quality of the results in those experiments. Note that, while this is a sub-optimal procedure, the fact that the two subproblems are non-convex in some variables would prevent us from having a guaranteed optimal solution.

Each iteration of the Gauss-Seidel method simply involves solving an unconstrained least squares problem, which we solve using the Moore-Penrose pseudoinverse. After finding the solution of the unconstrained problem, that solution is "projected" into the space of feasible solutions. For example, in the first subproblem, solving for $\mathbf{H}$ (line 5) is done without taking the first constraint of (4.20) into account. After the unconstrained solution is found, $\mathbf{H}$ is multiplied by a scalar so that the largest absolute value of its elements becomes exactly 1. All variables, in both subproblems, are handled in a similar manner.

We use the values of the cost functions of problems (4.20) and (4.24) as imperfect indicators of the goodness of a solution. For this reason, each subproblem is solved multiple times for given data $\mathbf{Y}$; we then keep only the solution which yielded the lowest cost value for that subproblem (lines 10 and 21 of table 4.2), to partially cope with the possible existence of non-absolute minima.

| | **PHASE LOCKED MATRIX FACTORIZATION** |
|---|---|
| 1: | Given: data $\mathbf{Y}$, $\text{MaxRuns}_{\mathbf{f}}$, $\text{MaxIter}_{\mathbf{f}}$, $\text{MaxRuns}_{\mathbf{M,A,z}}$, $\text{MaxIter}_{\mathbf{M,A,z}}$ |
| **I**: | ESTIMATION OF $\mathbf{f}$ |
| 2: | **for** run $\in$ {1,2,. . .,$\text{MaxRuns}_{\mathbf{f}}$}, **do** |
| 3: | Randomly initialize $\hat{\mathbf{H}}$, $\hat{\mathbf{A}}$, $\hat{\mathbf{f}}$ |
| 4: | **for** iter $\in$ {1,2,. . .,$\text{MaxIter}_{\mathbf{f}}$}, **do** |
| 5: | Solve minimization (4.20) for $\mathbf{H}$ |
| 6: | Solve minimization (4.20) for $\mathbf{A}$ |
| 7: | Solve minimization (4.20) for $\mathbf{f}$ |
| 8: | **end for** |
| 9: | **end for** |
| 10: | From the $\text{MaxRuns}_{\mathbf{f}}$ solutions, choose the one which yields the lowest value of the function being minimized in (4.20) |
| 11: | Store $\mathbf{f}$ and discard $\mathbf{H}$ and $\mathbf{A}$ |
| 12: | Correct sign of $\mathbf{f}$ by detecting values of (4.21) greater than 1 |
| **II**: | ESTIMATION OF $\mathbf{M}$, $\mathbf{A}$, $\mathbf{z}$ |
| 13: | **for** run $\in$ {1,2,. . .,$\text{MaxRuns}_{\mathbf{M,A,z}}$}, **do** |
| 14: | Randomly initialize $\hat{\mathbf{M}}$, $\hat{\mathbf{A}}$, $\hat{\mathbf{z}}$ |
| 15: | **for** iter $\in$ {1,2,. . .,$\text{MaxIter}_{\mathbf{M,A,z}}$}, **do** |
| 16: | Solve problem (4.24) for $\mathbf{M}$ |
| 17: | Solve problem (4.24) for $\mathbf{A}$ |
| 18: | Solve problem (4.24) for $\mathbf{z}$ |
| 19: | **end for** |
| 20: | **end for** |
| 21: | From the $\text{MaxRuns}_{\mathbf{M,A,z}}$ solutions, choose the one which yields the lowest value of the function being minimized in eq. (4.24) |
| 22: | **return** $\mathbf{M}$,$\mathbf{A}$, $\mathbf{z}$, $\mathbf{f}$ |

**Table 4.2.** The Phase Locked Matrix Factorization algorithm.

## 4.4 The Effect of Whitening in SSS

Theorem 4.1.1 can be considered the SSS version of Theorem 2.3.1, which establishes identifiability conditions for ICA. In this section we present an SSS version of Theorem 2.3.2, showing that prewhitening yields some advantages when performing SSS. This result was originally presented in **[Publication IX]**.

Let $\mathrm{Cov}[\mathbf{y}\mathbf{y}^T]$ denote the covariance matrix of the mixtures. Prewhitening [38] involves multiplying the data $\mathbf{Y}$ on the left by the matrix

$$\mathbf{B} \equiv \mathbf{D}^{-\frac{1}{2}}\mathbf{V}^H, \tag{4.31}$$

where $\mathbf{D}$ is a $N \times N$ diagonal matrix containing only the nonzero eigenvalues of $\mathrm{Cov}[\mathbf{y}\mathbf{y}^T]$ in its diagonal, $\mathbf{V}$ is a $P \times N$ matrix with the corresponding eigenvectors in its columns, and $(\cdot)^H$ denotes the conjugate transpose of a matrix. Then, the equation $\mathbf{BY} = \mathbf{BMS}$ defines a new BSS problem, with new data $\mathbf{BY}$ which now has $N$ rows.[7] The new mixing matrix $\mathbf{BM}$ is called the *equivalent mixing matrix*, and is now square.

In SSS, the mixing matrix $\mathbf{M}$ is real but the data $\mathbf{Y}$ are complex. Therefore, if $\mathbf{B}$ is defined as in (4.31), the equivalent mixing matrix $\mathbf{BM}$ is, in general, complex. Thus, without whitening, one is searching for a real $P \times N$ mixing matrix (or equivalently, a real $N \times P$ unmixing matrix); with whitening one has to search for a complex $N \times N$ mixing matrix (or a complex $N \times N$ unmixing matrix). We now show how one can transform this into a search for a real $N \times N$ mixing (or unmixing) matrix.

We split the data matrix $\mathbf{Y}$ into its real part $\mathbf{Y}_R \equiv \mathrm{real}(\mathbf{Y})$ and its imaginary part $\mathbf{Y}_I \equiv \mathrm{imag}(\mathbf{Y})$, and define $\mathbf{S}_R$ and $\mathbf{S}_I$ in a similar way for the source matrix $\mathbf{S}$. Since $\mathbf{M}$ is real, the initial complex problem $\mathbf{Y} = \mathbf{MS}$ can be turned into an equivalent real problem in two different ways:

$$\begin{bmatrix}\mathbf{Y}_R \\ \mathbf{Y}_I\end{bmatrix} = \begin{bmatrix}\mathbf{M} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}\end{bmatrix}\begin{bmatrix}\mathbf{S}_R \\ \mathbf{S}_I\end{bmatrix} \quad \text{or} \quad [\mathbf{Y}_R\ \mathbf{Y}_I] = \mathbf{M}\,[\mathbf{S}_R\ \mathbf{S}_I]\,. \tag{4.32}$$

We call the first formulation the "vertically stacked form" (VS form) and the second one the "horizontally stacked form" (HS form). Clearly, any of these two formulations is equivalent to the original one, in the sense that a solution for either of them is transformable into a solution for the original problem.

[7] In the presence of additive noise, all eigenvalues of $\mathrm{Cov}[\mathbf{y}\mathbf{y}^T]$ will be nonzero. Even in that case, this reasoning remains valid if noise levels are low and only the $N$ largest eigenvalues are used to construct $\mathbf{D}$.

Recall from section 2.3 that the condition number can be used as an indicator of the difficulty of an inverse problem. The following theorem states that, in SSS, this difficulty is bounded above if prewhitening is performed. One can apply the whitening procedure to the left-hand side of either the VS form or the HS form, both of which are real. Both of these methods would yield the same upper bound for the condition number of the equivalent mixing matrix in the theorem that follows. We have empirically found, however, that the condition number of the equivalent mixing matrix is, on average, farther from the upper bound presented ahead (and thus, that the matrix is better conditioned) if the HS form is used. Therefore, we focus on that formulation only.

The upper bound for the condition number of the mixing matrix after whitening is given by the following theorem, derived in **[Publication IX]**.

**Theorem 4.4.1.** *Let* $\mathbf{S}_{RI} \equiv [\mathbf{S}_R\ \mathbf{S}_I]$ *and* $\mathbf{Y}_{RI} \equiv [\mathbf{Y}_R\ \mathbf{Y}_I]$. *Let* $\mathbf{B}$ *be the result of applying the procedure from equation* (4.31) *to* $\mathbf{Y}_{RI}$. *Let* $a_j(t) = |s_j(t)|$ *and* $\phi_j(t) = \text{angle}(s_j(t))$. *Furthermore, suppose that the following assumptions hold:*

- $\mathbf{M}$ *and* $\mathbf{S}$ *both have maximum rank.*
- *There is no additive noise; thus,* $\mathbf{Y} = \mathbf{MS}$ *holds.*
- $a_j(t)$*, are i.i.d. realizations of a random variable which we denote by* $A_j$*;*
- $A_j$ *is independent of* $A_k$ *for* $j \neq k$*;*
- $\phi_j(t)$*, are i.i.d. realizations of a random variable which we denote by* $\Phi_j$*;*
- $A_j$ *is independent of* $\Phi_k$ *for any* $j$ *and* $k$*, including* $j = k$*;*
- *All* $A_j$ *have the same distribution (we denote by* $A$ *a generic random variable with that distribution);*
- $s_j$ *and* $s_k$ *have maximum PLF,* i.e.*, they have a constant phase lag; this implies that there exists* $\phi(t)$*, independent of* $j$*, such that* $\phi_j(t) = \phi_j + \phi(t)$ *for all* $j$ *and* $t$*;*
- *For each value of* $t$*,* $\phi(t)$ *is random, and uniformly distributed in* $[0, 2\pi)$*;*

*note, however, that $\phi(t)$ does not need to be i.i.d..*

*Then, the condition number of the equivalent mixing matrix, denoted by $\rho(\mathbf{BM})$, obeys*

$$\rho(\mathbf{BM}) \leq \sqrt{1 + N\frac{\mathbb{E}[A]^2}{\mathrm{Var}[A]}}, \tag{4.33}$$

*where $N$ is the number of sources, $\mathbb{E}[\cdot]$ is the expected value operator and* $\mathrm{Var}[\cdot]$ *is the variance operator. Furthermore, this upper bound is tight, meaning that in some cases equation* (4.33) *holds with equality.*

In the ICA case, prewhitening ensures that we can restrict the search to orthogonal unmixing matrices. Equivalently, the equivalent mixing matrix after prewhitening is guaranteed to have a condition number of 1. In SSS, the condition number of the equivalent mixing matrix can be larger than 1, but it is bounded above by a value which depends on properties of the amplitudes of the sources.

In **[Publication IX]**, we presented experimental results confirming the validity of this bound, by randomly generating mixing matrices and sources which obeyed the assumptions of the theorem, for a few types of sub- and super-gaussian distributions. We then applied prewhitening and computed the condition number of the equivalent mixing matrix, verifying that the upper bound is correct and that it is tight.

The assumptions of theorem 4.4.1 are quite restrictive, and will probably not be obeyed in most practical situations. Nevertheless, we have empirically found that even in such situations, prewhitening improves the quality of the results of SSS, and improves the convergence time of the separation methods. This was shown for PLMF in **[Publication XI]**. In the results that follow, prewhitening was always applied to the data before any separation algorithm was used.

# 5. Experimental Results

## 5.1 IPA Results

Results obtained with various versions of IPA have been presented in four papers **[Publication I]**, **[Publication II]**, **[Publication III]**, **[Publication V]**. For conciseness, we present only the two most significant results, and omit some details; the reader can consult the above references for more information.

*Comparison of IPA with other BSS techniques on simulated data*

In **[Publication III]** we tackled the problem of full ISA where the dependency within each subspace was perfect synchrony. Our goal was to show that IPA could successfully be used as the third step (intra-subspace separation). For this, we generated data that obeys the ISA model. We then used TDSEP[1] [91] for the first step (inter-subspace separation) and a simple heuristic for subspace detection. Since perfectly synchronous sources are strongly dependent, we did not expect TDSEP to be appropriate for the third step. Our goal was to study whether applying TDSEP to separate subspaces, and then using IPA to separate within each subspace, was effective.[2]

In order to study this, we randomly generated 300 sets of 12 sources,

[1]TDSEP is an algorithm which separates sources based on the principle that independent sources should have $\mathbb{E}[s_i(t)s_j(t+\tau)] = \mu_\tau \delta_{ij}$, where $\mu_\tau$ is a real number that depends on the time lag $\tau$, and $\delta_{ij}$ is Kronecker's delta. TDSEP constructs multiple time-lagged correlation matrices $\mathbf{C}_\tau$ whose $(i,j)$ element is $\mathbb{E}[s_i(t)s_j(t+\tau)]$ and performs joint diagonalization on them to estimate the original sources.

[2]These results were obtained with a constant $\lambda$ optimization strategy. The varying $\lambda$ strategy detailed in section 4.2.2 was only proposed later, in **[Publication V]**.

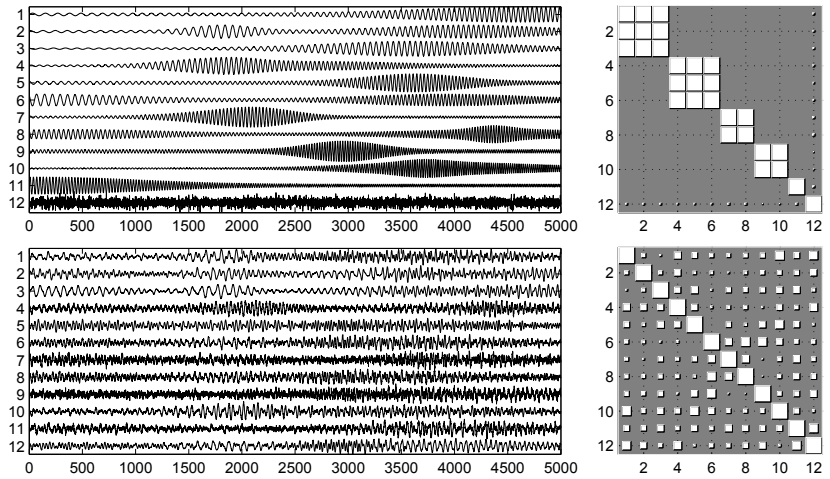


**Figure 5.1.** Example dataset with subspaces of dimensions 3,3,2,2,1,1. *First row*: Original sources *(left)* and PLFs between them *(right)*. *Second row*: Mixed signals and PLFs between them. In the second column, the numbers denote the indexes of the sources, and the area of each square is proportional to their pairwise PLF.

grouped in 6 subspaces. The subspaces have sizes 3, 3, 2, 2, 1, 1, and the sources were such that different subspaces could not simply be separated through a bandpass filter (see the right panel of figure 5 of the paper). We also generated corresponding mixing matrices randomly. An example of a set of sources generated this way is depicted in Figure 5.1.[3]

As an illustration of the need for specific techniques for separating synchronous sources, figure 5.2 shows the results of applying FastICA to the set of signals from figure 5.1. The results are quite poor, since the sources are very strongly dependent. This kind of result was consistently obtained throughout our experiments with synchronous sources.

We applied TDSEP to each of the 300 sets of mixed signals. Then, a simple heuristic procedure was applied to estimate the subspaces present in the data.[4] If subspaces could be detected, IPA was applied on each of these estimated subspaces. Otherwise, the result of TDSEP was returned with no further processing.

The above procedure, which we denote TDSEP+IPA, was compared with simply applying TDSEP to the sets of mixed signals with no further pro-

[3]The specific way in which the sources and mixing matrices were generated can be found in **[Publication III]**.

[4]We used a hard threshold on the matrix containing the pairwise PLFs. Several values of the threshold are swept; if any of them returns a block-diagonal structure, the subspace structure corresponding to those blocks is considered correct. Otherwise, we consider that the subspaces cannot be detected.

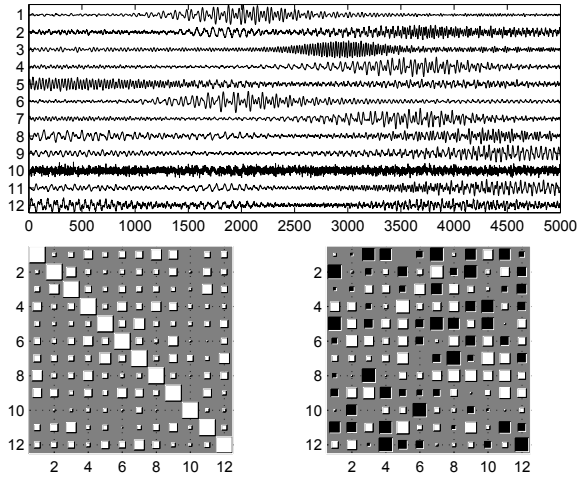


**Figure 5.2.** Result of FastICA applied to the dataset of figure 5.1. *Top*: Sources estimated by FastICA. *Bottom left*: PLFs between the estimated sources. *Bottom right*: Estimated gain matrix. It is clear that FastICA is not adequate for the problem.

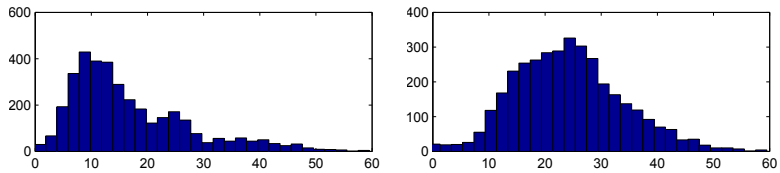


**Figure 5.3.** *(Left)* Histogram of the signal-to-noise ratio (SNR) between the sources found by TDSEP and the original sources. *(Right)* Similar histogram for the sources found by IPA.

cessing (*i.e.* no subspace detection and no application of IPA to the detected subspaces). To measure the quality of the output, we measured the signal-to-noise ratio (SNR) of the estimated sources relative to the corresponding true sources. We compensate the permutation indeterminacy using a simple heuristic, and then define the SNR of source $i$ as $10 \log_{10} \frac{E[(\alpha \hat{s}_i)^2]}{E[(\alpha \hat{s}_i \pm s_i^2]}$, where the real scalar $\alpha$ and the $\pm$ sign are chosen to maximize the SNR value. This ensures that these SNR values are independent of permutation, scaling, and sign, as is common in source separation contexts. We then compute the average SNR over all estimated sources. Histograms of the average SNRs of the 300 runs are shown in figure 5.3. The average output SNR is 16.78 dB for TDSEP and 24.18 dB for TDSEP+IPA. These results show that IPA was able to perform an effective separation within the subspaces separated by TDSEP.

The result of these two procedures (TDSEP+IPA and TDSEP only) for the set of sources and mixtures in figure 5.1 is shown in figure 5.4. These

two results correspond to the modes of the histograms from figure 5.3, and thus they can be considered typical results. It can be seen that TDSEP is quite successful in separating the subspaces, but that it does not correctly estimate the original sources – this observation is what led us to use TDSEP as a subspace-identification method. It can also be concluded that IPA performs a correct separation within the subspaces.

We also tested the sensitivity of both approaches to additive noise. To that effect, we added random Gaussian white noise to each of the 300 sources with SNRs of 60, 50, 40, 30, and 20 dB (to avoid confusion with the SNR as a quality measure, we use the term "output SNR" for that measure and "input SNR" for the measure of the noise added to the mixture). We then repeated the two procedures above for each of these sets of 300 signals. Figure 5.5 shows the average output SNR of the 300 runs, for each input SNR level. It can be seen that both TDSEP+IPA and TDSEP alone yielded similar results, and rather poor ones (output SNR around 11 dB), when the input SNR was 30 dB. For SNR values below 30 dB, TDSEP+IPA actually yields worse results than TDSEP by itself, whereas for an input SNR above 30 dB, TDSEP+IPA performed better than TDSEP alone.

This set of results shows that TDSEP+IPA is capable of separating synchronous sources with significantly better results than other BSS methods in low-noise situations. They also show that even moderate levels of noise hinder TDSEP+IPA's ability to successfully separate this type of sources.

*Application of IPA to pseudo-real MEG data*

The second set of results illustrates a more realistic test of IPA; it was presented in **[Publication V]**. Ideally, one should validate an SSS algorithm on real-life data. For this, one would need a set of real-world signals which were the result of a mixture of synchronous sources; one would also need direct measurements of those sources, or knowledge of the mixing matrix (or both), to be able to assess the quality of the separation results. For example, for EEG or MEG, we would need not only the EEG/MEG recordings from outside the scalp, but also simultaneous acquisitions of the sources from inside the scalp. We are not aware of any dataset where these acquisitions were simultaneously performed. On the other hand, simulated data, such as the data used in the previous set of results, can only go a certain length in showing the usefulness of an algorithm in real

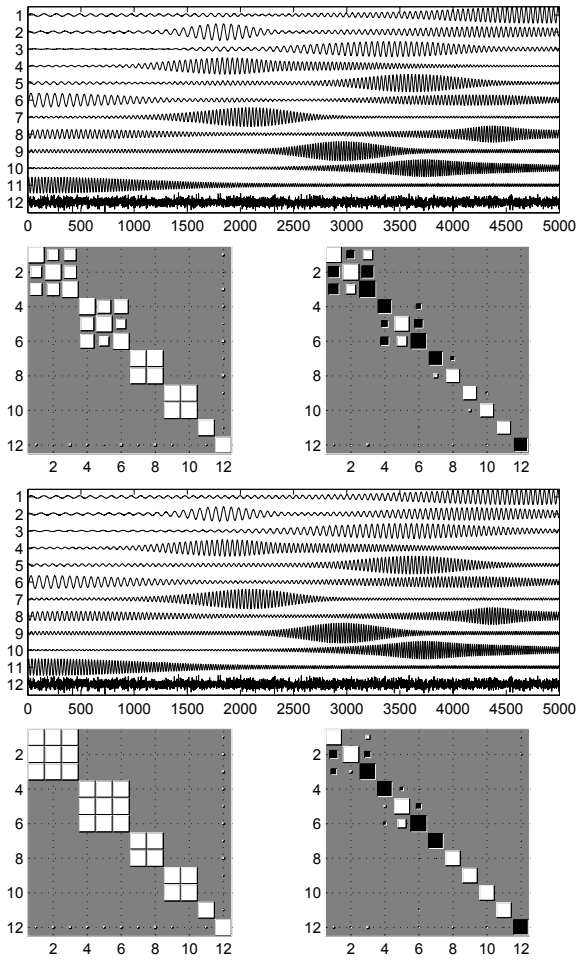


**Figure 5.4.** *First row*: Sources resulting from TDSEP. Note that the inter-subspace PLFs *(second row, left)* are very close to zero, but the intra-subspace PLFs are not all close to 1. Furthermore, the intra-subspace separation is poor, as can be seen from inspection of the gain matrix estimated by TDSEP *(second row, right)*. *Third row*: Results found after applying IPA to each subspace. The estimated sources are very similar to the original ones. This is corroborated by the PLFs between the estimated sources *(fourth row, left)* and the final gain matrix *(fourth row, right)*. The permutation of the sources was manually corrected. White squares represent positive values and black squares represent negative values.

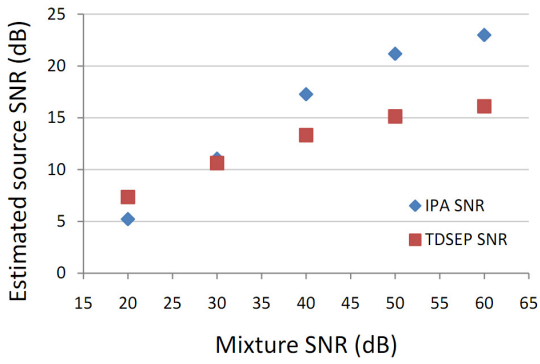


**Figure 5.5.** Effect of noise on separation quality, for TDSEP+IPA and for TDSEP alone.

situations.

In an attempt to obtain the best of both worlds, we generated a set of pseudo-real data from actual MEG recordings. By doing this, we were able to generate a set of sources on which we knew the true sources and the true mixing matrix, while still using sources that were of a nature similar to that of the signals one observes in real-world MEG. We begin by summarily describing the process that we used to generate a data set with perfectly synchronous sources. We then explain how we modified these data to analyze non-perfect cases as well.

We began by obtaining a realistic mixing matrix, using the EEGIFT software package and a real-world EEG dataset[5] it includes to obtain a $64 \times 20$ mixing matrix. In each run, we then selected $N$ random rows and $N$ random columns, and formed the $N \times N$ mixing matrix by taking the corresponding submatrix.

The second step involved obtaining a set of physiologically plausible sources which obey the SSS model. For this, we used the MEG dataset previously studied in [87], and selected $N$ sources at random from its 122 channels. These $N$ sources were bandpass filtered using a filter with zero phase. The passband was 18-24 Hz, which is of a width similar to filters used in typical MEG studies [85]. We then computed the Hilbert transform of each of these bandpass-filtered sources and extracted their amplitudes and phases.

To generate sources which were phase-locked, we generated new *pseudo-*

[5]This is not a typo. We did use an EEG dataset to obtain the mixing matrix and an MEG dataset to obtain the sources. This was intended only as a proof-of-concept with more realistic data, therefore we do not believe this to be a serious issue.

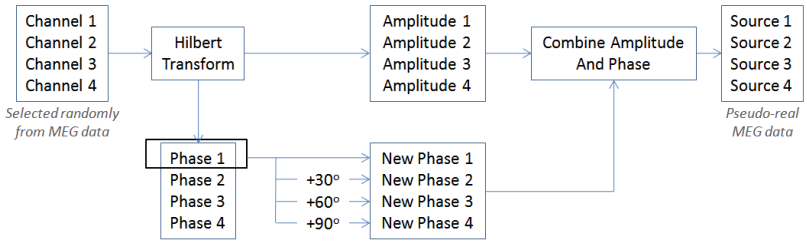


**Figure 5.6.** The process used to generate the pseudo-real MEG sources.

*real* sources which used the amplitudes of the original sources, but whose phases were phase-lagged versions of the first source's phase. As an example, suppose that $N = 4$. The first pseudo-real source was equal to the first original channel. We replaced the phase of the second of these channels with the phase of the first channel with a constant phase lag of $\frac{\pi}{6}$ radians. The phase of the third channel was replaced with the phase of the first channel with a constant phase lag of $\frac{\pi}{3}$ radians, and that of the fourth channel with the phase of the first channel with a lag of $\frac{\pi}{2}$ radians. The amplitudes of the four sources were kept as the original amplitudes of the four random channels themselves. The process is illustrated in figure 5.6.

The above procedure yielded sources which exactly obeyed the SSS model. To study situations which deviated from the model, we multiplied each each sample $t$ of each source $j$ by $e^{i\delta_j(t)}$, where the phase jitter $\delta_j(t)$ was drawn from a random Gaussian distribution with zero mean and standard deviation $\sigma$. We tested IPA for $\sigma$ from 0 to 20 degrees, in 5 degrees steps. One example with $\sigma = 5$ degrees is shown in figure 5.7, and one with $\sigma = 20$ degrees is shown in figure 5.8. 100 different datasets were generated for each of these phase jitter values, by selecting at random different rows and different columns from the original $64 \times 20$ mixing matrix, and by selecting at random different channels from the original 122 channels of the MEG data.

Figure 5.9 shows the output SNR and Amari Performance Index[6] as a function of the phase jitter for $N = 4$. It can be seen that for phase jitter values from 0 (jitterless case) to 5 degrees (mild jitter) there was virtually no loss of performance. The performance at a jitter of 10 degrees was already deteriorated but still acceptable with an average output SNR of

[6]The Amari Performance Index is another quality measure for source separation. It measures how different the gain matrix $\mathbf{W}^T\mathbf{M}$ is from a diagonal matrix. Please see [5] for a formal definition.

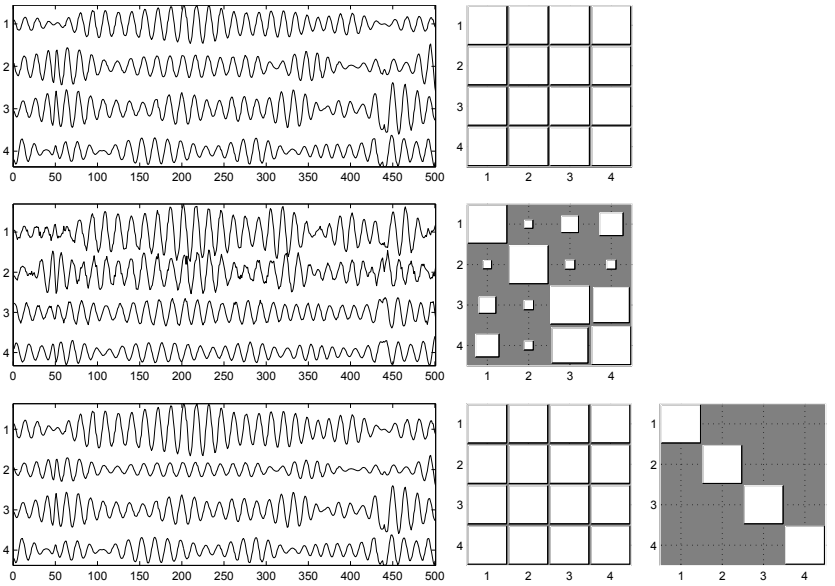


**Figure 5.7.** Example of a dataset where $\sigma = 5$ degrees. Only a short segment of the signals is shown, for clarity. *Top row*: original sources *(left)* and PLFs between them *(right)*. *Middle row*: mixed signals *(left)* and PLFs between them *(right)*. *Bottom row*: estimated sources, after manual compensation of permutation, scaling, and sign *(left)*; PLFs between them *(middle)*; and the gain matrix $\mathbf{W}^T\mathbf{A}$ *(right)*. The gain matrix is virtually equal to the identity matrix, indicating a correct separation.

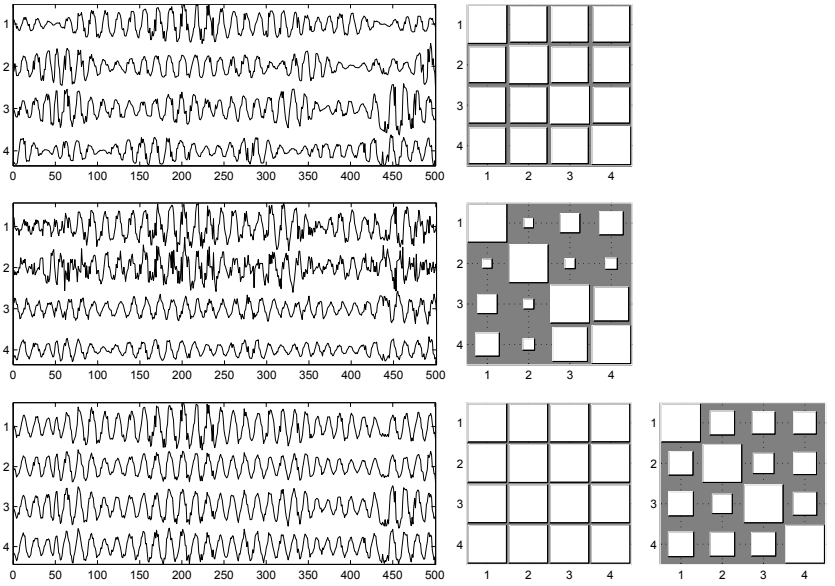


**Figure 5.8.** Similar to figure 5.7, but with $\sigma = 20$ degrees. The gain matrix has significant values outside the diagonal, indicating that a complete separation was not achieved. Nevertheless, the largest values are in the diagonal, corresponding to a partial separation.

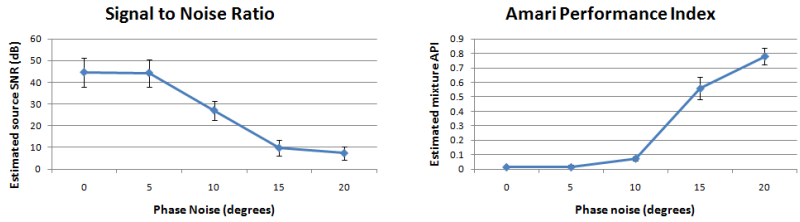


**Figure 5.9.** Result of applying IPA to pseudo-real MEG data with $N = 4$, with varying phase jitter: Signal to Noise Ratio (left) and Amari Performance Index (right).

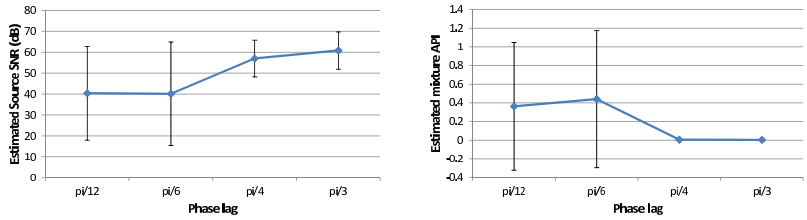


**Figure 5.10.** Effect of applying IPA to pseudo-real MEG data with varying phase lags between the sources, with $N = 2$: Signal to Noise Ratio (left) and Amari Performance Index (right).

27 dB. The quality of the separation then gradually deteriorated until 15 degrees (strong jitter), after which performance remained at a low level. We added the noise after the bandpass filtering; we expect that results would be better if the noise had been added before that filtering step.

We also studied the effect of the phase lag between the sources. For this, we used $N = 2$, and generated 100 data sets at random as above, with phase lags of $\frac{\pi}{12}$, $\frac{2\pi}{12}$, $\frac{3\pi}{12}$, $\frac{4\pi}{12}$ (*i.e.*, multiples of 15 degrees). The results are shown in figure 5.10. The results show that phase lags of $\frac{\pi}{6}$ and below yield significant variability in performance, with the error bars including SNR values from around 60 dB to under 20 dB. In contrast, values of $\frac{\pi}{4}$ and above yield consistently good results.

We also studied how results varied with the choice of $N$, by generating 100 datasets for each of $N = 2, 3, 4, 5$. Phase lags were multiples of $\frac{\pi}{6}$.[7] The results, shown in figure 5.11 are quite surprising: while performance was consistently good for $N = 3, 4, 5$, with only slight deterioration as $N$ increased, there was a significant decrease in performance for $N = 2$. We do not have a solid explanation for this fact. Our conjecture, which remains open, is that the presence of some pairs of sources with larger phase lags (for example, for $N = 4$, the first and third sources had a phase

[7] In other words, for $N = 2$, sources had phase lags of 0 and $\frac{\pi}{6}$. For $N = 3$ they had phase lags of 0, $\frac{\pi}{6}$ and $\frac{\pi}{3}$, and so on.

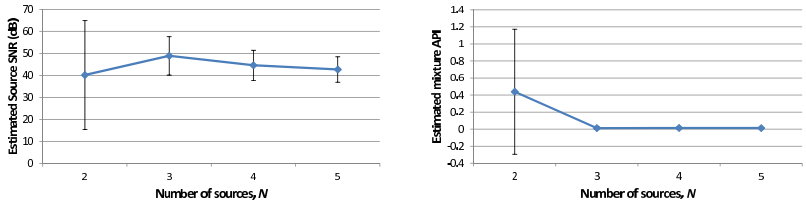


**Figure 5.11.** Effect of applying IPA to pseudo-real MEG data with varying values of $N$: Signal to Noise Ratio (left) and Amari Performance Index (right).

| | $\lambda$ | 0.025 | 0.05 | 0.1 | 0.2 | 0.4 |
|---|---|---|---|---|---|---|
| SNR | fixed | $17.5 \pm 21.2$ | $27.5 \pm 18.0$ | $34.4 \pm 4.3$ | $27.2 \pm 3.6$ | $13.5 \pm 5.5$ |
| | varying | | | $48.9 \pm 8.7$ | | |
| API | fixed | $0.795 \pm 0.570$ | $0.369 \pm 0465$ | $0.048 \pm 0.057$ | $0.079 \pm 0.027$ | $0.327 \pm 0.097$ |
| | varying | | | $0.013 \pm 0.015$ | | |

**Table 5.1.** Values of Signal to Noise Ratio (SNR) and Amari Performance Index (API) for jitterless data with $N = 3$, for various fixed values of $\lambda$, as well as for the varying-lambda strategy detailed in the text. While the best fixed value, $\lambda = 0.1$, yields decent results, the results using a varying value of $\lambda$ are consistently better, with a large margin.

lag of $\frac{\pi}{3}$ and the first and fourth sources had a phase lag of $\frac{\pi}{2}$) aids in the separation of all the sources, including the ones with small phase lags.

Finally, we compared the varying-$\lambda$ strategy (which was used for all the above results) with a fixed-$\lambda$ strategy. Results are shown in table 5.1 for $N = 3$. While the best fixed $\lambda$ yielded a decent separation quality, with an average output SNR close to 35 dB, using a varying $\lambda$ yielded much better results, with an average output SNR of almost 50 dB.

Globally, these results illustrate that IPA can handle realistically simulated signals and that it is robust to a variation in the number of sources up to, at least, $N = 5$. They also illustrate that it can handle phase lags of $\frac{\pi}{4}$ and above. Results also show that IPA can handle mild deviations from the true SSS model, by exhibiting some robustness to phase jitter. Even in cases with strong phase jitter, the imperfectly separated sources were normally closer to the true sources than the original mixed signals. Finally, the results show that the varying $\lambda$ strategy yielded big improvements in separation quality, when compared to any strategy with a fixed $\lambda$.

## 5.2 PLMF Results

Results obtained using PLMF with simulated data were reported in **[Publication VI]**, **[Publication VII]**, **[Publication VIII]**, **[Publication XI]**. PLMF was most extensively studied in **[Publication XI]**. In

that paper, we studied the effect of several variables (number of sources, number of sensors, amount of additive noise, amount of phase jitter, number of time samples, and more) by starting from a "central case" where PLMF yielded good results, and changing one of these variables at a time, to find how the algorithm's performance varied with each of them. We report the most important experimental results from that paper in what follows.

We used a noisy variant of the source model in expression (4.18) to generate the data. This variant accomodated two deviations from the noiseless case: the presence of additive noise and of phase jitter. The model used to generate the data was

$$\mathbf{Y} \equiv \mathbf{M}(\mathbf{A} \odot (\mathbf{z}\mathbf{f}^T) \odot \mathbf{J}) + \mathbf{N}, \tag{5.1}$$

where $\mathbf{J}$ is a $N \times T$ matrix of complex values with unit absolute value, representing phase jitter, and $\mathbf{N}$ is a $P \times T$ matrix of complex values representing additive sensor noise. If all entries of $\mathbf{J}$ are equal to 1 and all entries of $\mathbf{N}$ are equal to zero, we recover the noiseless and jitterless model of Equation (4.18).

We generated 1000 datasets for each set of parameters studied. For each dataset, a mixing matrix $\mathbf{M}$ was randomly generated, with each entry uniformly distributed between -1 and 1, the vector of phase lags $\mathbf{z}$ was generated as $[0, \Delta\phi, \ldots, (N-1)\Delta\phi]^T$ ($\Delta\phi$ is determined below), and the common oscillation $\mathbf{f}$ was generated as a sinusoid: $\mathbf{f} = [0, \exp(i\Delta t), \exp(i2\Delta t), \ldots, \exp(i(T-1)\Delta t)]^T$, with $T = 100$ and $\Delta t = 0.1$. While this was a very specific choice (a phase which grows linearly with time), it is representative of the smoothly-varying $\mathbf{f}$ case which is treated in that paper. We have empirically verified that PLMF worked well with other choices for $\mathbf{f}$ as long as they were smoothly-varying (otherwise, the correction of phase jumps, mentioned at the end of section 4.3.2, became unreliable).

The amplitude $\mathbf{A}$ was generated as the result of lowpass filtering a Gaussian white noise signal. Specifically, we began by generating random Gaussian white noise of length $T$. We then took the discrete cosine transform (DCT) of that signal, kept only the 10% of coefficients corresponding to the lowest frequencies, and took the inverse DCT of the result. We then added a constant to this filtered signal to ensure that it was non-negative[8], and the result became $\mathbf{a}_1(t)$, the first row of $\mathbf{A}$. The pro-

[8]While the algorithm presented in this work does not require positive amplitudes, we compared it to other algorithms which do require this assumption.

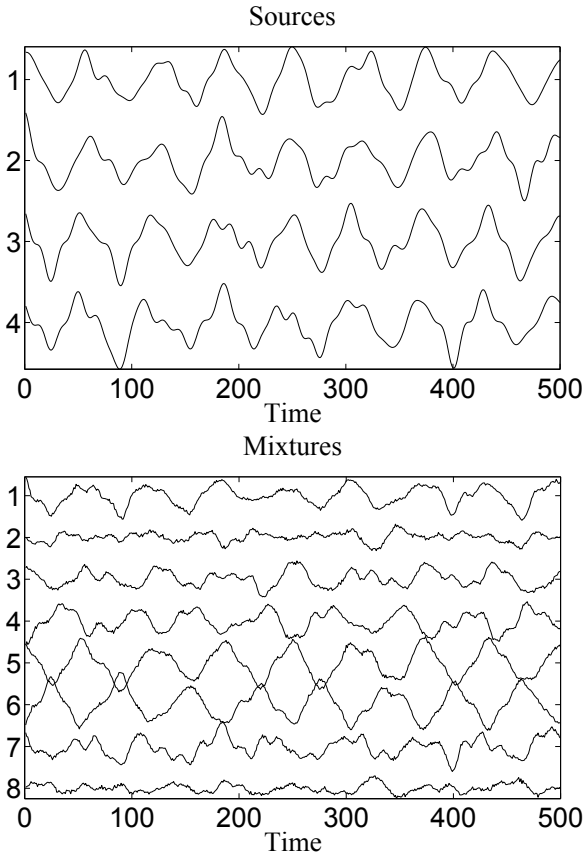


**Figure 5.12.** *Top*: the real part of a typical set of four sources generated as described in the text, with no phase jitter. *Bottom*: the real part of a corresponding set of eight mixtures, with an input signal-to-noise ratio (SNR) of 20 dB. The horizontal axis measures time in samples. Note that in most of the following experiments, only 100 points were used.

cess was repeated, with different random initializations, for each of the remaining rows of $\mathbf{A}$.

One example of a set of signals generated in this manner is depicted in figure 5.12, where we present an extended time period ($T = 500$) to better illustrate the structure of the signals.

The paper **[Publication XI]** studied the effect of the following variables:

- Additive noise N, as measured by the signal-to-noise ratio (SNR) of each mixture. The energy of the noise in each channel was selected so that all channels had the same SNR, which is called the input SNR. We studied the cases of an SNR of 80, 60, 40, 20, and 0 dB.

- Phase jitter $\mathbf{J}$. We studied two types of jitter:
  - The first case was jitter where each entry of $\mathbf{J}$ was of the form $e^{i\delta}$, where $\delta$ was independently drawn from a Gaussian distribution with zero mean and standard deviation $\sigma_{iid}$. We studied the cases of $\sigma_{iid} = 0, 0.02, 0.04, \ldots, 0.1$. We name this *i.i.d. jitter*, since the jitter for time $t$ and for source $k$ was independent from the jitter in any other entry of $\mathbf{J}$.
  - The second case is called *correlated jitter*. We generate a matrix $\mathbf{Q}$ in a similar manner to the way in which we generated the amplitude $\mathbf{A}$, except that positivity was not enforced, and that we kept the lowest 2% of coefficients of the DCT, instead of the lowest 10%. This yielded a very slowly varying signal. We then generated the jitter $\mathbf{J}$ as $e^{i\mathbf{Q}}$, where the exponential is taken elementwise. This resulted in a jitter which was slow-varying. While in a statistical sense the sources are uncorrelated, due to the finite observation time $T$, this jitter is correlated from one source to another. In the context of this correlated jitter, we will use the symbol $\sigma_{corr}$ to denote the standard deviation of the Gaussian white noise used in the generation of the jitter.
- Phase lag $\Delta\phi$. We studied the cases of $\Delta\phi = \pi/50, 2\pi/50, \ldots, 12\pi/50$.
- Number of sources $N$ and number of sensors $P$. We studied the cases $N = 2, 4, \ldots, 10$, with $P = N$ and with $P = 2N$.
- Number of time samples $T$. We studied the values $T = 100, 200, 400, 800$.

It would have been extremely cumbersome to compute and show results for all possible combinations of the above variables. To avoid this while still studying all variables, we studied a "central case" where PLMF performed very well, and then changed the above variables, one at a time. In total, we studied 64 different cases. The central case had $N = 4$ sources, $P = 8$ sensors, $T = 100$ time samples, an input SNR of 80 dB, no jitter, and a phase lag of $\Delta\phi = \pi/10$.

We first applied both IPA and PLMF to 1000 datasets in each of four situations, all of which had $N = P = 2$ sources and sensors, and no phase jitter: low noise and large phase lag (input SNR of 80 dB, $\Delta\phi = \pi/3$), low

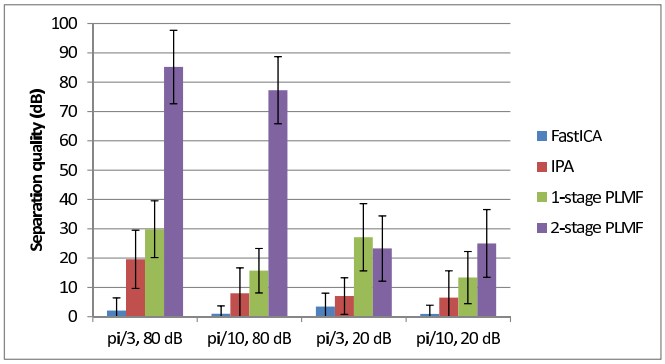


**Figure 5.13.** Comparison of FastICA, IPA, one-stage PLMF, and two-stage PLMF. Error bars correspond to plus or minus one standard deviation. The two-stage PLMF algorithm clearly dominates the other two algorithms, except for one situation ($\Delta\phi = \pi/3$, input SNR of 20 dB) where it is essentially tied for first place with one-stage PLMF.

noise and small phase lag (input SNR of 80 dB, $\Delta\phi = \pi/10$), moderate noise and large phase lag (input SNR of 20 dB, $\Delta\phi = \pi/3$), and moderate noise and small phase lag (input SNR of 20 dB, $\Delta\phi = \pi/10$). We also compared with the first variant of PLMF **[Publication VII]**, in which all four variables are estimated simultaneously, and with FastICA [38].

The results are shown in figure 5.13. Apart from one situation where both versions of PLMF are tied, these results show a clear superiority of 2-stage PLMF when compared to the other two SSS algorithms. FastICA performed poorly, as expected, given that the sources were strongly interdependent.[9]

In other experiments **[Publication XI]**, we showed how PLMF performed when each of the above variables was varied. We concluded that, unlike IPA, the performance of PLMF decreased gracefully with the input SNR (see figure 5.14), since the separation quality was only 2-3 dB below the input SNR, except for low-noise cases (input SNR of 80 and 60 dB), in which the separation quality was nevertheless very good (65 and 55 dB, respectively).

PLMF can handle very small phase lags. Figure 5.15 shows how the separation quality varied with the phase lag $\Delta\phi$. For most values of this parameter, the separation quality was very high. However, it became progressively lower when $\Delta\phi$ approached zero, where the hypothesis of The-

[9]We used the MATLAB FastICA implementation available from http://research.ics.aalto.fi/ica/fastica/code/dlcode.shtml. All parameters were left at their default values, except for the nonlinearity option where we tried all possibilities. All such options yield very similar results; the results reported here use the default option.

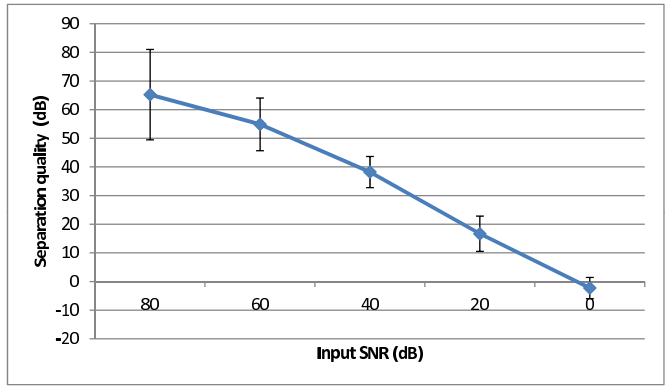


**Figure 5.14.** Separation quality versus input SNR. Under heavy noise, PLMF can recover the sources with about as much noise as they had in the input.

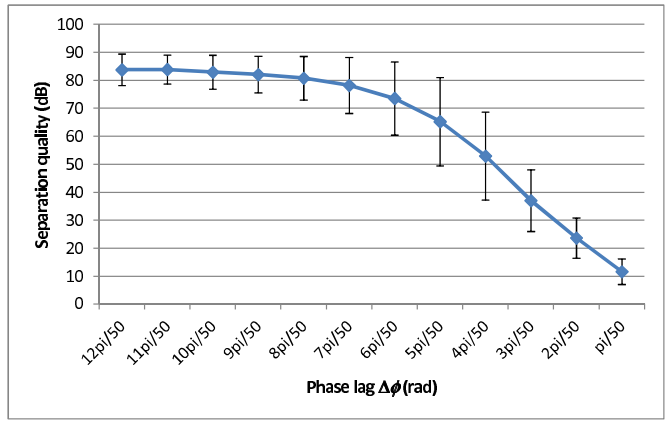


**Figure 5.15.** Separation quality versus phase lag. PLMF's results are, in general, good, but they deteriorate progressively as one approaches the case where $\Delta\phi = 0$, where theorem 4.3.2 fails to hold.

orem 4.3.2 fails to hold. Nevertheless, this deterioration in performance was gradual, and was only relevant for very small phase lags (smaller than $\frac{2\pi}{50}$, or 7.2 degrees, which yielded a separation quality of 23.7 dB).

We also concluded that, in low-noise situations, having as many sensors as sources ($P = N$) was enough to obtain a good separation: little benefit is brought by having $P > N$. However, that benefit became significant in the presence of noise. Figure 5.16 shows the effect of varying the number of sources $N$ and the number of sensors $P$. Generally, the quality of the results decreased with increasing $N$, which is expected since the size of the problem variables $\mathbf{M}$, $\mathbf{A}$ and $\mathbf{z}$ increases. When there was very little noise (input SNR of 80 dB), there was little benefit in doubling the number of sensors from $P = N$ to $P = 2N$. However, when there was considerable noise (input SNR of 20 dB), that benefit became significant, especially for

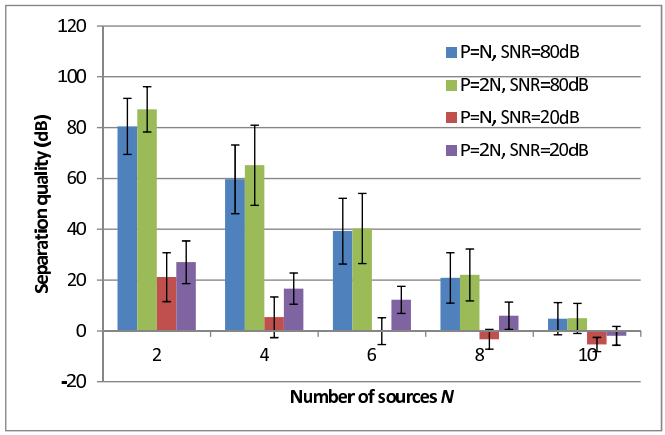


**Figure 5.16.** Separation quality versus number of sources ($N$), number of sensors ($P$), and input SNR.

$P = 4, 6, 8$ where the improvement exceeded 10 dB.

Finally, we observed that PLMF did not handle the presence of phase jitter well. Even small amounts of jitter brought the performance from around 65 dB to 30 dB, and larger amounts lowered it further. The paper presents a more involved discussion explaining that the effect of i.i.d. jitter can probably be mitigated using lowpass filtering of f as post-processing, after f is estimated, whereas correlated jitter cannot be mitigated that way.

# 6. Future Work

In this chapter we present a few directions for future work based on the research presented above. Some of these directions were already studied in a limited depth, and we present suggestions to deepen them.

## 6.1 Tests on Real-Life Data

We begin this section by briefly discussing an important aspect which was not dealt with in this thesis: validation using real-world data. Indeed, perhaps the most important future development would be the acquisition of non-simulated data, and the testing of the algorithms proposed here on those data. The study with pseudo-real data, presented in **[Publication V]**, can be considered a step in this direction; however, those pseudo-real data were generated in such a way that they exactly followed the SSS model, or only deviated from it in controlled ways. Real-life data may deviate from the model in other ways, and only tests on those data can tell how the proposed algorithms perform in such situations.

Acquisition of real-life data in an EEG or MEG context, in a form that allows the assessment of the performance of the proposed algorithms, is not easy, for several reasons:

- Expensive equipment is required.

- Mixture data can be acquired from the outside of the skull, but the sources, which are needed to assess the algorithms' performance, must be acquired from inside the skull. Invasive procedures are, therefore, necessary.

- Neuroscience experts are required, to ensure that data are acquired

from areas which are expected to exhibit synchrony, and to help in evaluating the results.

It was not possible to acquire such data in this thesis' work. Nevertheless, it is certainly important to test SSS algorithms in real-world data. A non-invasive, but also less conclusive approach is to collect only data from the outside of the skull, apply SSS to those data, and see whether the resulting sources "make sense". This eliminates the invasiveness requirement, but makes neuroscientists' input even more important. This approach has been followed in [86], which applied ICA to MEG recordings and then measured coherence between the resulting sources. A similar approach was followed in [55], which measured synchrony instead of coherence. Other works, such as [58], which used BSS with non-synchrony criteria to extract sources, also follow this approach.

Another possibility is to validade these algorithms on another domain where SSS's assumptions are (approximately) valid. One such domain might be music. Musical instruments which are playing the same tone will have the same fundamental frequency, and should therefore have perfect phase synchrony. It is possible that the algorithms presented here can be applied to the separation of instruments playing the same tone, and subspace versions of these algorithms may be used to separate sets of instruments playing different tones. If the music being played is known, this knowledge can be incorporated in a variant of the PLF measure known as $n : m$ synchrony [82]. Finally, the domain of multipath communication systems [26] may find the techniques proposed here, or adapted versions of them, useful.

## 6.2 Improvements on PLMF

One future direction that was briefly discussed in **[Publication XI]** consists in turning each of PLMF's subproblems into a sequence of convex problems. More specifically, the goal is to have an equivalent, or near-equivalent, formulation, such that the optimization in each variable is a convex problem. In its present state, the first subproblem in PLMF (problem (4.20)) is not solved through a sequence of convex problems: it would be necessary that $\mathbf{H}$, $\mathbf{A}$ and $\mathbf{f}$ each lie in convex sets, and that is not true for $\mathbf{H}$ and $\mathbf{f}$. The second subproblem (problem (4.24)) is also not solved through a sequence of convex problems for a similar reason involving $\mathbf{M}$

and z.

Let us discuss why one should be interested in doing this. There is considerable theoretical work on BNGS methods. In particular, [28] gives sufficient conditions for the following property: if a BNGS method converges to some limit solution, then that limit solution is a critical point of the problem. Turning each of PLMF's subproblems into a sequence of convex problems (or more specifically, ensuring that the feasible set for each variable is a closed, non-empty, convex set) would allow direct application of this theorem, ensuring that if the algorithms for each subproblem converged, we would have found critical points of the subproblems.

### 6.3 Generalizing PLMF for subspaces

The algorithms presented in this thesis could be adapted to work with subspaces. One possibility would be to incorporate them into a full ISA framework, using a first step which is agnostic to synchrony (*i.e.*, it does not use synchrony information at all) to separate subspaces, followed by a second step which uses synchrony to separate within a subspace. In **[Publication III]**, we used an approach of this kind, where the synchrony-agnostic first step was TDSEP, and the second step was IPA. The references presented at the end of section 2.4 illustrate that in some situations this is possible using ICA methods as the first step.

However, if it is known that the dependency within each subspace is synchrony, and that sources from different subspaces are not synchronous, it makes sense to develop algorithms which exploit this information. One possible advantage is that these algorithms may not require full independence between sources in different subspaces.

### 6.4 Partial Synchrony

While sources with partial synchrony were tackled with the algorithms presented in this work, the model assumes that the sources have full synchrony. Yet another improvement would involve devising a model which does not make this assumption. It is likely that this would lead to techniques quite different from the ones presented here.

One possibility is to consider a Bayesian framework under which the phase of source $j$, at time $t$, is no longer given by the product of $z_j$ and

$f(t)$, but rather by the product of $z_j$ and $f_j(t)$, where $f_j(t)$ is, for example, a Gaussian random variable centered at $f(t)$, with some variance. As the variance tends to zero we recover the perfectly-synchronous model of SSS.

## 6.5 Partial ISA

Some work was developed in the study of the ISA separation principle (defined at the end of section 2.4) for a particular type of sources. Since some details of the proof are not finished, it is omitted, and the result is here presented as a conjecture.

The conjecture is the following: by minimizing the sum of the entropies of the individual estimated sources, one can solve Partial ISA, for sources which are a sum of Gaussian components with zero mean. Formally, consider the usual BSS model, $\mathbf{y} = \mathbf{M}\mathbf{s}$, where sets of sources are independent as in (2.14), repeated here for convenience:

$$\mathbf{s} \equiv \begin{bmatrix} \mathbf{s}^1 \\ \mathbf{s}^2 \\ \vdots \\ \mathbf{s}^K \end{bmatrix}, \text{ where } \mathbf{s}^k \equiv \begin{bmatrix} s_1^k \\ \vdots \\ s_{N_k}^k \end{bmatrix}. \tag{6.1}$$

We consider only sources of a particular type: Gaussian mixture densities where each Gaussian component has zero mean. Concretely, the probability density functions (PDFs) of the subspaces are assumed to be given by

$$p(\mathbf{s}^k) = \sum_i \alpha_{ik} N(\mathbf{s}^k | 0, \mathbf{A}_{ik}). \tag{6.2}$$

Furthermore, let $\hat{\mathbf{s}} = \mathbf{W}\mathbf{y}$ denote the estimated sources, and assume that the data $\mathbf{y}$ have been pre-whitened. Our conjecture is that any solution of

$$\min_{\mathbf{W}} \sum_{i=1}^{N} H(\hat{s}_i) \tag{6.3}$$
$$\text{s.t. } \mathbf{W}^T\mathbf{W} = \mathbf{I}$$

yields a gain matrix $\mathbf{W}^T\mathbf{M}$ which is a permutation of a block diagonal matrix, with blocks corresponding to each of the subspaces, as exemplified in section 2.4. Consequently, each estimated source will be a linear combination of true sources from a single subspace.

If this conjecture turns out to be true, its applicability is considerable. If each Gaussian component could have an arbitrary mean, sources of the

form (6.2) could approximate any density in a broad class of densities arbitrarily well as the number of Gaussian components increases, a result which has been known for over forty years [75]. Future work in this direction would involve determining the veracity of this conjecture, by presenting a fully correct proof or a counter-example. Furthermore, it would be relevant to characterize which densities can be well approximated by densities of the form (6.2), and investigating what is the intersection between those densities and densities which describe sources with perfect synchrony.

# 7. Conclusions

In this thesis we have studied the problem of Separation of Synchronous Sources, an instance of blind source separation where the sources exhibit perfect synchrony. Since synchrony is present in many topics in neuroscience, and EEG/MEG signals can be considered mixtures of underlying sources, SSS is a relevant problem for this field. SSS had not been studied before; this thesis presents the first formalization of the problem. Previously, to separate synchronous sources, researchers were forced to use algorithms which make inadequate assumptions about the sources (such as ICA, which assumes independence of the sources).

It was shown that SSS is sufficiently well-defined, like in the case of ICA. Unlike ICA, singular solutions are an issue in SSS, and they should be taken into consideration when designing algorithms. We showed that prewhitening results in a bound on the condition number of the equivalent mixing matrix, ensuring a relatively well-conditioned numerical problem. We presented two algorithms to solve SSS problems: IPA, which penalizes singular solutions through regularization, and PLMF, which factorizes the sources using a matrix factorization model which automatically avoids singular solutions.

Experimental tests with simulated data showed that both approaches yield very significant improvements in separation quality compared to ICA algorithms. Within the SSS algorithms, PLMF yielded significantly better results than IPA in experimental comparisons. PLMF also possesses better theoretical properties.

The contributions in this thesis allow, for the first time, the separation of synchronous sources within a grounded theoretical framework, with appropriate algorithms which yield good separation quality and avoid singular solutions, and in the case of PLMF, with identifiability guarantees. Some directions for further work were presented, which illustrate that

the topic is far from being solved and present interesting research topics for the future.

# References


[1] T: Adali, M. Anderson, and G.-S. Fu. Diversity in independent component and vector analyses. *IEEE Signal Processing Magazine*, pages 18–33, 2014.

[2] M. Aigner. A characterization of the Bell numbers. *Discrete Mathematics*, 205:207–210, 1999.

[3] E. Alhoniemi, A. Honkela, K. Lagus, and J. Seppä. Compact modeling of data using independent variable group analysis. *IEEE Transactions on Neural Networks*, 18:1762–1776, 2007.

[4] L. B. Almeida. Synthesis lectures on signal processing. In *Nonlinear Source Separation*. Morgan & Claypool, 2006.

[5] S. Amari, A. Cichocki, and H. H. Yang. A new learning algorithm for blind signal separation. In *Advances in NIPS*, volume 8, pages 757–763, 1996.

[6] B. B. Andersen, L. Korbo, and B. Pakkenberg. A quantitative study of the human cerebellum with unbiased stereological techniques. *Journal of Computational Neurology*, 326:549–560, 1992.

[7] B. Ans, J. Herault, and C. Jutten. Adaptive neural architectures: detection of primitives. In *Proceedings of COGNITIVA'85*, 1985.

[8] J. Beirlant, E. Dudewicz, L. Gyorfi, and E. van der Meulen. Nonparametric entropy estimation: an overview. *International Journal of Mathematical and Statistical Sciences*, 6:17–39, 1997.

[9] A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–1159, 1995.

[10] A. J. Bell and T. J. Sejnowski. A non-linear information maximization algorithm that performs blind separation. In *Advances in Neural Information Processing Systems*, 1995.

[11] A. J. Bell and T. J. Sejnowski. Edges are the" independent components" of natural scenes. In *Advances in Neural Information Processing Systems*, pages 831–837, 1996.

[12] J. Bioucas-Dias and G. Valadão. Phase unwrapping via graph cuts. *IEEE Transactions on Image Processing*, 16:698–709, 2007.

[13] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[14] H. O. Cabral, M. Vinck, C. Fouquet, C. M. A. Pennartz, L. Rondi-Reig, and F. P. Battaglia. Oscillatory dynamics and place field maps reflect sequence and place memory processing in hippocampal ensembles under nmda receptor control. *Neuron*, 81:402–415, 2014.

[15] J.-F. Cardoso. Multidimensional independent component analysis. In *Proc. of Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP '98)*, 1998.

[16] A. Cichocki and S. Amari. *Adaptive blind signal and image processing - Learning algorithms and applications*. John Wiley & Sons, 2002.

[17] P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36:287–314, 1994.

[18] P. Comon and C. Jutten. *Handbook of Blind Source Separation*. Academic Press, 2010.

[19] P. Comon and C. Jutten, editors. *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Academic Press, 2010.

[20] B. A. Conway, D. M. Halliday, S. F. Farmer, U. Shahani, P. Maas, A. I. Weir, and J. R. Rosenberg. Synchronization between motor cortex and spinal motoneuronal pool during the performance of a maintained motor task in man. *Journal of Physiology*, 489:917–924, 1995.

[21] L. De Lathauwer, B. De Moor, and J. Vandewalle. Fetal electrocardiogram extraction by source subspace separation. In *Proc. IEEE SP/ATHOS Workshop HOS*, 1995.

[22] S. M. Doesburg, A. B. Roggeveen, K. Kitajo, and L. M. Ward. Large-scale gamma-band phase synchronization and selective attention. *Cerebral Cortex*, 18:386–396, 2008.

[23] A. K. Engel, P. Fries, and W. Singer. Dynamic predictions: oscillations and synchrony in top–down processing. *Nature Reviews Neuroscience*, 2:704–716, 2001.

[24] P. Fries. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in Cognitive Sciences*, 9:474–480, 2005.

[25] P. Fries. Neuronal gamma-band synchronization as a fundamental process in cortical computation. *Annual Review of Neuroscience*, 32:209–224, 2009.

[26] A. Goldsmith. *Wireless Communicatnions*. Cambri, 2005.

[27] C. Gray, P. König, A. Engel, and W. Singer. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, 338:334–337, 1989.

[28] L. Grippo and M. Sciandrone. On the convergence of the block nonlinear Gauss-Seidel method under convex constraints. *Operations Research Letters*, 26:127–136, 2000.

[29] H. W. Gutch, J. Krumsiek, and F. J. Theis. An ISA algorithm with unknown group sizes identifies meaningful clusters in metabolomics data. In *Proceedings of the European Signal Processing Conference (EUSIPCO)*, 2011.

[30] A. Guyton and J. Hall. *Textbook of Medical Physiology*. Saunders, 12th edition, 2011.

[31] M. Hämäläinen, R. Hari, R. J. Ilmoniemi, J. Knuutila, and O. V. Lounasmaa. Magnetoencephalography theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev. Mod. Phys.*, 65:413–497, 1993.

[32] J. Herault and B. Ans. Circuits neuronaux à synapses modifiables: décodage de messages composites par apprentissage non supervisé. In *Comptes rendus de l'Academie des sciences. Serie III, Sciences de la vie*, volume 299, pages 525–528, 1983.

[33] J. Herault and C. Jutten. Space or time adaptive signal processing by neural network models. *Neural networks for computing*, 151:206–211, 1986.

[34] J. Herault, C. Jutten, and B. Ans. Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé. In *10é Colloque sur le traitement du signal et des images*, pages 1017–1022, 1985.

[35] S. Herculano-Houzel. The remarkable, yet not extraordinary, human brain as a scaled-up primate brain and its associated cost. *Proceedings of the National Academy of Sciences*, 109:10661–10668, 2012.

[36] P. Hoyer. Non-negative matrix factorization with sparseness constraints. *Journal of Machine Learning Research*, 5:1457–1469, 2004.

[37] A. Hyvärinen and P. Hoyer. Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces. *Neural Computation*, 12:1705–1720, 2000.

[38] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Wiley & Sons, 2001.

[39] A. Hyvärinen and U. Köster. FastISA: A fast fixed-point algorithm for independent subspace analysis. In *Proceedings of the European Symposium on Artificial Neural Networks (ESANN)*, 2006.

[40] K. M. Igarashi, L. Lu, L. L. Colgin, M.-B. Moser, and E. I. Moser. Coordination of entorhinal-hippocampal ensemble activity during associative learning. *Nature*, 2014.

[41] C. Jutten and J. Herault. Blind separation of sources: an adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24:1–10, 1991.

[42] C. Jutten and J. Karhunen. Advances in nonlinear blind source separation. In *Proc. of the 4th Int. Symp. on Independent Component Analysis and Blind Signal Separation (ICA2003)*, 2003.

[43] C. Jutten and A. Taleb. Source separation: From dusk till dawn. In *Proceedings of the 2nd International Workshop on Independent Component Analysis and Blind Source Separation*, pages 15–26, 2000.

[44] H. Kameoka, N. Ono, K. Kashino, and S. Sagayama. Complex NMF: a new sparse representation for acoustic signals. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2009.

[45] B. J. King and L. Atlas. Single-channel source separation using complex matrix factorization. *IEEE Transactions on Audio, Speech, and Language Processing*, 19:2591–2597, 2011.

[46] Y. Kuramoto. *Chemical Oscillations, Waves and Turbulences*. Springer Berlin, 1984.

[47] E. Kuruoglu and F. Theis. Dependent component analysis. *EURASIP Journal on Advances in Signal Processing*, 2013:185, 2013.

[48] J.-P. Lachaux, E. Rodriguez, J. Martinerie, and F. J. Varela. Measuring phase synchrony in brain signals. *Human Brain Mapping*, 8:194–208, 1999.

[49] D. Lahat, J. Cardoso, and H. Messer. Second-order multidimensional ICA: Performance analysis. *IEEE Transactions on Signal Processing*, 60:4598–4610, 2012.

[50] D. Lahat, J.-F. Cardoso, M. Le Jeune, and H. Messer. Multidimensional ICA and its performance analysis applied to CMB observations. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, pages 3724–3727, 2011.

[51] D. Lahat, J.-F. Cardoso, and H. Messer. Second-order multidimensional ICA: Performance analysis. *IEEE Transactions on Signal Processing*, 60(9):4598–4610, 2012.

[52] D. Lee and H. Seung. Algorithms for non-negative matrix factorization. In *Advances in Neural Information Processing Systems*, volume 13, pages 556–562, 2001.

[53] W. Ma, J. Bioucas-Dias, T. Chan, N. Gillis, P. Gader, A. Plaza, A. Ambikapathi, and C. Chi. A signal processing perspective on hyperspectral unmixing: Insights from remote sensing. *IEEE Signal Processing Magazine*, 31:67–81, 2014.

[54] S. Makeig, A. J. Bell, T. P. Jung, and T. J. Sejnowski. Independent component analysis of electroencephalographic data. In *Advances in Neural Information Processing Systems*, pages 145–151, 1996.

[55] C. Meinecke, A. Ziehe, J. Kurths, and K.-R. Müller. Measuring phase synchronization of superimposed signals. *Physical Review Letters*, 94, Mar 2005.

[56] N. S. Narayanan, J. F. Cavanagh, M. J. Frank, and M. Laubach. Common medial frontal mechanisms of adaptive control in humans and rodents. *Nature Neuroscience*, 16:1888–1895, 2013.

[57] G. Nolte, O. Bai, L. Wheaton, Z. Mari, S. Vorbach, and M. Hallett. Identifying true brain interaction from EEG data using the imaginary part of coherency. *Clin. Neuroph.*, 115:2292–2307, 2004.

[58] G. Nolte, A. Ziehe, F. Meinecke, and K.-R. Müller. Analyzing coupled brain sources: Distinguishing true from spurious interaction. In *Advances in Neural Information Processing Systems (NIPS)*, 18, pages 1027–1034, 2006.

[59] P. L. Nunez and R. Srinivasan. *Electric Fields of the Brain: the neurophysics of EEG*. Oxford University Press, 2006.

[60] P. L. Nunez, R. Srinivasan, A. F. Westdorp, R. S. Wijesinghe, D. M. Tucker, R. B. Silberstein, and P. J. Cadusch. EEG coherency I: statistics, reference electrode, volume conduction, laplacians, cortical imaging, and interpretation at multiple scales. *Electroencephalography and clinical Neurophysiology*, 103:499–515, 1997.

[61] A. V. Oppenheim, A. S. Willsky, and S. Hamid. *Signals and Systems*. Signal Processing Series. Prentice-Hall, 1996.

[62] J. A. Palmer and S. Makeig. Blind separation of dependent sources and subspaces by minimum mutual information. Technical report, University of California, San Diego, 2010.

[63] J. A. Palmer and S. Makeig. Contrast functions for independent subspace analysis. In *Latent Variable Analysis and Signal Separation*, pages 115–122, 2012.

[64] M. S. Pedersen, J. Larsen, U. Kjems, and L. C. Parra. *Springer Handbook of Speech Processing*, chapter A Survey of Convolutive Blind Source Separation methods, pages 1065–1084. Springer Press, 2008.

[65] A. Pikovsky, M. Rosenblum, and J. Kurths. *Synchronization: A universal concept in nonlinear sciences*. Cambridge Nonlinear Science Series. Cambridge University Press, 2001.

[66] B. Póczos and A. Lörincz. Independent subspace analysis using geodesic spanning trees. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 673–680, 2005.

[67] E. Pomarol-Clotet, R. Salvador, S. Sarró, J. Gomar, F. Vila, Á Martínez, A. Guerrero, J. Ortiz-Gil, B. Sans-Sansa, A. Capdevila, J. M. Cebamanos, and P. J. McKenna. Failure to deactivate in the prefrontal cortex in schizophrenia: dysfunction of the default mode network? *Psychological Medicine*, 38:1185–1194, 2008.

[68] B. Roach and D. Mathalon. Event-related eeg time-frequency analysis: An overview of measures and an analysis of early gamma band phase locking in schizophrenia. *Schizophrenia Bulletin*, 34:907–926, 2008.

[69] J.-H. Schleimer and R. Vigário. Reference-based extraction of phase synchronous components. In *Proceedings of the International Conference on Artificial Neural Networks*, 2006.

[70] J.-H. Schleimer and R. Vigário. Clustering limit cycle oscillators by spectral analysis of the synchronisation matrix with an additional phase sensitive rotation. In *Proc. of the Int. Conf. on Artificial Neural Networks*, 2007.

[71] J.-H. Schleimer and R. Vigário. Order in complex systems of nonlinear oscillators: Phase locked subspaces. In *Proc. of the Eur. Symp. on Neural Networks*, 2007.

[72] D. L. Schomer and F. L. da Silva. *Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Wolters Kluwer Health, 2012.

[73] A. Sharma and K. K. Paliwal. Subspace independent component analysis using vector kurtosis. *Pattern Recognition*, 39:2227–2232, 2006.

[74] W. Singer. Neuronal synchrony: A versatile code for the definition of relations? *Neuron*, 24:49–65, Sep 1999.

[75] H. W. Sorenson and D. L. Alspach. Recursive bayesian estimation using gaussian sums. *Automatica*, 7:465–479, 1971.

[76] K. Spencer, M. Niznikiewicz, M. Shenton, and R. McCarley. Sensory-evoked gamma oscillations in chronic schizophrenia. *Biological Psychiatry*, 63:744–747, 2008.

[77] S. Strogatz. From kuramoto to crawford: exploring the onset of synchronization in populations of coupled oscillators. *Physica D*, 143:1–20, 2000.

[78] Z. Szabó, B. Póczos, and A. Lörincz. Undercomplete blind subspace deconvolution. *Journal of Machine Learning Research*, 8:1063–1095, 2007.

[79] Z. Szabó, B. Póczos, and A. Lörincz. Separation theorem for independent subspace analysis and its consequences. *Pattern Recognition*, 45:1782–1791, 2012.

[80] C. Tallon-Baudry, O. Bertand, C. Delpuech, and J. Pernier. Oscillatory $\gamma$-band (30-70 hz) activity induced by a visual search task in humans. *The Journal of Neuroscience*, 17:722–734, 1997.

[81] C. Tallon-Baudry, O. Bertrand, C. Delpuech, and J. Pernier. Stimulus speci?city of phase-locked and non-phase-locked 40 hz visual responses in human. *The Journal of Neuroscience*, 16:4240–4249, 1996.

[82] P. Tass, M. G. Rosenblum, J. Weule, J. Kurths, A. Pikovsky, J. Volkmann, A. Schnitzler, and H.-J. Freund. Detection of n:m phase locking from noisy data: Application to magnetoencephalography. *Physical Review Letters*, 81:3291–3294, 1998.

[83] F. J. Theis. Towards a general independent subspace analysis. In *Advances in Neural Information Processing Systems*, pages 1361–1368, 2007.

[84] P. J. Uhlhaas and W. Singer. Neural synchrony in brain disorders: Relevance for cognitive dysfunctions and pathophysiology. *Neuron*, 52:155–168, Oct 2006.

[85] F. Varela, J.-P. Lachaux, E. Rodriguez, and J. Martinerie. The brainweb: phase synchronization and large-scale integration. *Nature Reviews Neuroscience*, 2:229–239, 2001.

[86] R. Vigário and O. Jensen. Identifying cortical sources of corticomuscle coherence during bimanual muscle contraction by temporal decorrelation. In *Proceedings of IEEE International Symposium on Signal Processing and Its Applications*, 2003.

[87] R. Vigário, V. Jousmäki, M. Hämäläinen, R. Hari, and E. Oja. Independent component analysis for identification of artifacts in magnetoencephalographic recordings. In *Advances in NIPS*, 1997.

[88] S. Whitfield-Gabrieli, H. W. Thermenos, S. Milanovic, M. T. Tsuang, S. V. Faraone, R. W. McCarley, M. E. Shenton, A. I. Green, A. Nieto-Castanon, P. LaViolette, J. Wojcik, J. D. E. Gabrieli, and L. J. Seidman. Hyperactivity

and hyperconnectivity of the default network in schizophrenia and in first-degree relatives of persons with schizophrenia. *Proceedings of the National Academy of Sciences*, 4:1279–1284, 2009.

[89] R. W. Williams and K. Herrup. The control of the neuron number. *Annual Review of Neuroscience*, 11:423–453, 1988.

[90] T. Womelsdorf, J.-M. Schoffelen, R. Oostenveld, W. Singer, R. Desimone, A. K. Engel, and P. Fries. Modulation of neuronal interactions through neuronal synchronization. *Science*, 316:1609–1612, 2007.

[91] A. Ziehe and K.-R. Müller. TDSEP - an efficient algorithm for blind separation using time structure. In *International Conference on Artificial Neural Networks*, pages 675–680, 1998.

Aalto-DD 80/2016


ISBN 978-952-60-6783-4 (printed)
ISBN 978-952-60-6784-1 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

**Aalto University**
**School of Science**
**Department of Computer Science**
**www.aalto.fi**

**BUSINESS + ECONOMY**

**ART + DESIGN + ARCHITECTURE**

**SCIENCE + TECHNOLOGY**

**CROSSOVER**

**DOCTORAL DISSERTATIONS**