Bayesian Dictionary Learning for EEG Source Identification

Trine Nyholm Kragh & Laura Nyrup Mogensen
Mathematical Engineering, MATTEK

Master's Thesis

STUDENT REPORT
AALBORG UNIVERSITY

Mathematical Engineering
Aalborg University
http://www.aau.dk

AALBORG UNIVERSITY
STUDENT REPORT

Title:
Bayesian Dictionary Learning for EEG Source Identification

Theme:

Project Period:
Fall Semester 2019

Project Group:
Mattek9

Participant(s):
Trine Nyholm Kragh
Laura Nyrup Mogensen

Supervisor(s):
Jan Østergaard
Rasmus Waagepetersen

Copies: 1

Page Numbers: 19

Date of Completion:
September 24, 2019

Abstract:

Here is the abstract

Matematik-Teknologi
Aalborg Universitet
http://www.aau.dk

AALBORG UNIVERSITET
STUDENTERRAPPORT

Titel:
Bayesian Bibliotek Læring for EEG Kilde Identifikation

Tema:

Projektperiode:
Efterårssemestret 2019

Projektgruppe:
Mattek9

Deltager(e):
Trine Nyholm Kragh
Laura Nyrup Mogensen

Vejleder(e):
Jan Østergaard
Rasmus Waagepetersen

Oplagstal: 1

Sidetal: 19

Afleveringsdato:
24. september 2019

Abstract:

Her er resuméet

Preface

Here is the preface. You should put your signatures at the end of the preface.

Aalborg University, September 24, 2019

Trine Nyholm Kragh
<trijen15@student.aau.dk>

Laura Nyrup Mogensen
<lmogen15@student.aau.dk>

Danish Summary

Dansk resume ?

Contents

Introduction

Introduktion til hele projektet, skal kunne læses som en appetitvækker til resten af rapporten, det vi skriver her skal så uddybes senere. Brug dog stadigvæk kilder.

- kort intro a EEG og den brede anveldelse, anvendelse indenfor høreapperat.
- intro af model, problem med overbestemt system
- Seneste forslag til at løse dette
- vi vil efterviser dette og udvide til realtime tracking
- opbygningen af rapporten

Chapter 1

Problem Analysis

This chapter examines existing literature concerning source localisation from EEG measurements. At first a motivation for the problem is given, considering especially the application within the hearing aid industry. Further, the state of the art methods are presented followed by a description of the contribution proposed in this thesis.

1.1 Motivation

(Hvad er EEG)
EEG recordings or measurements are used within medicine as an imaging technique measuring electric signals on the scalp, caused by brain activity.

The brain consist of an enormous amounts of cells, called neurons. These neurons are mutually connected in neural nets and when a neuron is activated, for instance by some physical stimuli, local current flows are produced[4]. As such the neurons are somehow communicating(?).

The EEG measurements are provided by a varies number of metal electrodes referred to as sensors, placed on the scalp of a human reading electrical signals which are massively amplified and displayed on the computer as a sum of sinusoidal waves relative to time.
It takes a large amount of active neurons to generate an electrical signal that is recordable on the scalp as the current then have to penetrate the skull, skin and several other thin layers.
From this it is clear that the measurements from a single sensor do not correspond to the activity of a single neuron in the brain, but rather a collection of many activities. Here the same neuron activities can be measured by two or more sensors. Furthermore, interfering signals can occur resulting from physical movement of e.g. eyes and jawbone[4].

The waves resulting from EEG have been classified into four groups according to the dominant frequency. The delta wave (0.5 – 4 Hz) is observed from infants and sleeping adults, the Theta wave (4 – 8 Hz) is observed from children and sleeping adults, the alpha wave (8 – 13 Hz) is the most extensively studied brain rhythm, which is induced by an adult laying down with closed eyes. Lastly the beta wave (13 – 30 Hz) is considered the normal brain rhythm for normal adults, associated with active thinking, active attention or solving concrete problems[1, p. 11]. An example of EEG measurement within the four categories are illustrated by figure 1.1.

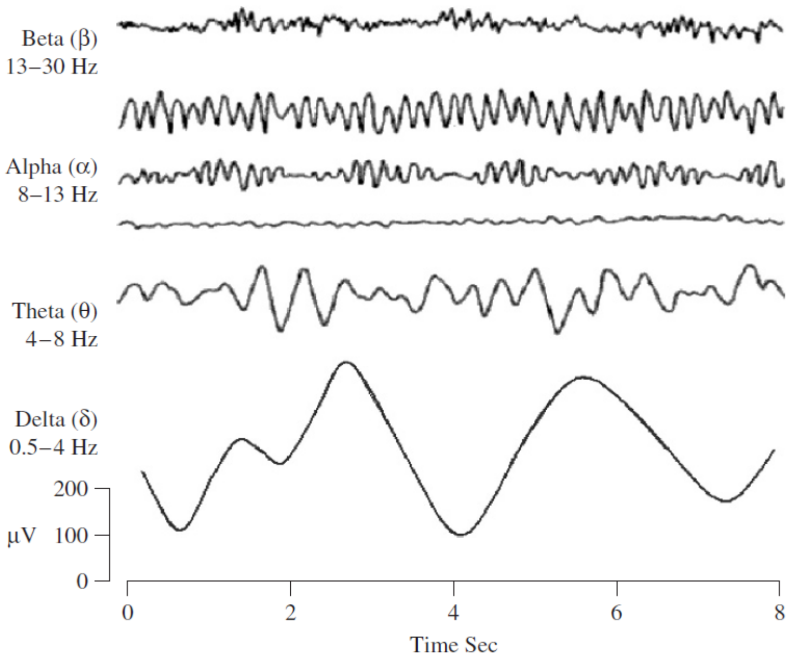

Figure 1.1: Example of time dependent EEG measurements within the four defined categories alpha, beta, theta and delta. Image source: [1]

(Hvad bruges det til)
EEG is widely used within medicine and research of the cognitive processes in the brain. Diagnosis and Management of neurological disorders such as epilepsy is one example.
A great advantage of EEG is the speed. Neural activity can be measured within fractions of a second after a stimuli has been provided[4, p. 3]. When a person is exposed to certain stimuli, e.g. visual or audible, the measured activity is said to result from evoked potential.

slå de to sammen..
(Hvad er problemet, localisation)

Due to EEG being non-invasive and fast it is widely used to study of the dynamical behaviour of the brain. Over the past two decades especially functional integration has become an area of interest, that is the interplay between functionally segregate brain areas[3](evt. friston 2011).
This concerns the localization of the single cortical sources causing the united signal measured by a EEG sensor.

(Modellering)
When considering the issue of identifying and localizing the activated sources from the EEG measurements, one known option is to model the data by the following linear system

$$\mathbf{Y} = \mathbf{A}\mathbf{X},$$

where $\mathbf{Y} \in \mathbb{R}^{M} \times N_d$ is the EEG measurements from N sensors at N_d data points, $\mathbf{A} \in \mathbb{R}^{M \times N}$ is an unknown mixing matrix and $\mathbf{X} \in \mathbb{R}^{N} \times N_d$ is the actual activation of sources within the brain. The i^{th} column of A represent the relative projection weights from the i^{th} source to every sensor or channel[**phd2015**]. From this model the aim is to identify both A and X given the measurements Y. This set up is in general referred to as the inverse EEG problem.

To solve the problem the concept of compressive sensing makes a solid foundation including sparse signal recovery and dictionary learning. Independent Component Analysis (ICA) is a common applied method to solve the inverse problem
problemet med source og sensors.. low density
(Anvendelse i praksis)
The hearing aid industry is one example of the use of this research. At Eriksholm research center which is a part of the hearing aid manufacture Oticon is cognitive hearing science an important research area. The main goal is to make it possible to the hearing aid to identify the attendeted speaker by reading the signal from the brain, which is where the EEG measurements are used(`https://www.eriksholm.com/research/cognitive-hearing-science`).

evt. lave subsections her..

1.2 Related Work and Our Contribution

nb. husk at dette kapitel skal vise et helt system og hvorhenne i det system vi kigger nærmere og kommer ind med vores bidrag. Det skal gøres klar hvilke områder vi vælger at ligge vores kræfter i.

Chapter 2

Problem Statement

forslag:
How can sources of activation within the brain be localized from EEG measurement, in the cases of less sensors than sources and how can this be extented to a realtime application within the hearing aid developmental

Underspørgsmål:

From EEG measurements how do one localize activated sources within the brain and can this be done in real time with the possibility to include feedback

Vi skal have styr på hvad det er vi vil dem vores realtime implementering: - måle om der er støj, så vi kan skrue ned for den støj,jeg tror det var det Jan snakked om i første omgang - kæde støjen sammen med locationerne for de aktive sources, måske det man gør i forhold til at retningsbestemme støj? - er første del blot at localisere sources

Chapter 3

Sparse Signal Recovery

Through this chapter an introduction to the concept compressive sensing is described with associated theory which later on will be used in the development of the algorithm with used methods known from compressive sensing to estimate the mixing matrix A and the sparse source matrix X.

3.1 Compressive Sensing

Compressive sensing is the theory of efficient recover/reconstruct a signal from minimal measurements. This recovery is often described as a linear model/system

$$\mathbf{y} = \mathbf{A}\mathbf{x},$$

which consist of observed data $\mathbf{y} \in \mathbb{R}^M$, a matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$ which models the linear measurements and signal $\mathbf{x} \in \mathbb{R}^N$. In compressive sensing terminology, $\mathbf{y}$ is the signal of interest that is wish recovered from minimal measurements meaning that the signal $\mathbf{x}$ must be sparse.
A signal is said to be k-sparse if the signal has at most k non-zeros coefficient

$$\|\mathbf{x}\|_0 = \text{card}(\text{supp}(\mathbf{x})) \leq k,$$

where the ℓ_0-norm is used. The function card is the cardinality of the support of $\mathbf{x}$. The support of $\mathbf{x}$ is giving as

$$\text{supp}(\mathbf{x}) = \{j \in [N] \ : \ x_j \neq 0\},$$

where $[N]$ a set [2, p. 41]. The set of all k-sparse signals is denoted as

$$\Sigma_k = \{\mathbf{x} \ : \ \|\mathbf{x}\|_0 \leq k\}.$$

It is of interest to recover the signal $\mathbf{y}$ when $M << N$ and $k < M$ [5, p. 8]. This lead to that the matrix $\mathbf{A}$ becomes rank-deficient and therefore have a non-empty null-space [5, p. ix].

The linear model and finding the sparse signal $\mathbf{x}$ can be written as an optimisation problem

$$\min_{\mathbf{z}} \|\mathbf{z}\|_0 \quad \text{s.t} \quad \mathbf{y} = \mathbf{A}\mathbf{z},$$

where $\mathbf{z}$ is all possible candidates to an k-sparse signal $\mathbf{x}$.
Unfortunately, this optimisation problem is non-convex because of ℓ_0-norm and is therefore difficult to solve – it is a NP-hard problem. Instead by replacing the ℓ_0-norm with its convex approximation, the ℓ_1-norm, the optimisation problem become computational feasible [5, p. 27]

$$\min_{\mathbf{z}} \|\mathbf{z}\|_1 \quad \text{s.t} \quad \mathbf{y} = \mathbf{A}\mathbf{z}, \tag{3.1}$$

and instead we find the best k-term approximation of the signal $\mathbf{x}$.

3.1.1 Conditions on the Mixing Matrix

To ensure an exact or an approximate reconstruction of the sparse signal $\mathbf{x}$ some conditions associated on the matrix $\mathbf{A}$ must be satisfied.

Null Space Conditions

The null space property (NSP) is some necessary and sufficient condition for exact recovery. The null space of the matrix A is defined as

$$\mathcal{N}(A) = \{z \ : \ Az = 0\}.$$

Restricted Isometry Conditions

NSP do not take account for noise and we must therefore look at some stronger conditions which incoperate noise, the following restricted isometry property (RIP)

Definition 3.1 (Restricted Isometry Property)
A matrix A satisfies the RIP of order k if there exists a $\delta_k \in (0,1)$ such that

$$(1-\delta_k)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta_k)\|x\|_2^2,$$

holds for all $x \in \Sigma_k$

If a matrix A satisfy RIP then it will also satisfy the NSP as RIP is strictly stronger than NSP.

Theorem 3.1.1
If A satisfies the RIP of order $2k$ with the constant $\delta_{2k} < \sqrt{2} - 1$. Then

$$C = \frac{2}{1 - (1 + \sqrt{2})\delta_{2k}}$$

Coherence

The NSP provide a unique solution to the optimisation problem, (3.1), but is unfortunately complicated to investigate. Instead an alternative measure used for sparsity is presented.
Coherence is a measure of quality and determine if the matrix A is a good choice for the optimisation problem (3.1). A small coherence describe the performance of a recovery algorithm as good with that choice of $\mathbf{A}$.

Definition 3.2 (Coherence)
Coherence of the matrix $A \in \mathbb{R}^{M \times N}$, denoted as $\mu(A)$, with columns $\mathbf{a}_1, \ldots, \mathbf{a}_N$ for all $i \in [N]$ is given as

$$\mu(A) = \max_{1 \leq i < j \leq n} \frac{|\langle a_i, a_j \rangle|}{\|a_i\|_2 \|a_j\|_2}.$$

3.2 ICA

Through this section the mathematical concepts of Independent Component Analysis (ICA) will be explained and defined.
Lets set up an situations. We have some measurements that has been affect by some surrounding noise or "**sideløbende**" measurements such as different conversations in a room. The measurements can be described by a vector $\mathbf{y}$ if we look at the one-dimensional case. $\mathbf{y}$ consist of the measurement from the original signal, a vector $\mathbf{x}$ and surrounding measurements, a matrix $\mathbf{A}$. This situation can be described as the linear model

$$\mathbf{y} = \mathbf{A}\mathbf{x} = \sum_{i=1}^{n} \mathbf{a}_i x_i$$

We known the measurements $\mathbf{y}$ but if also knew the mixing parameter in $\mathbf{A}$ then by inverting the linear model we could solve the system and find the original signal. But this is not the case as the mixing matrix also is unknown.

If we used the statistical properties of $\mathbf{x}$ then it would be possible to estimate both the mixing matrix and then the original signal. What ICA do is to assume statistical independence

Lets define the ICA model which is a generative model meaning that the observed data is generated by a process of mixing components which are latent component. Let n be the observed random variables such that $y_1, \dots, y_n$ are model as a linear combination of the random variables $x_1, \dots, x_n$:

$$y_i = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n, \quad i = 1, \dots, n$$
$$\mathbf{y} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \mathbf{x}$$

where $\mathbf{y} = \{y_i\}_{i\in[1,n]}$ and $\mathbf{x} = \{x_t\}_{t\in[1,n]}$. Furthermore, $\mathbf{x}$ is statistically mutually independent.

3.2.1 Estimation of Independent Components

Notes: Estimation with maximization of nongaussianity (see section 7.5 for nonguassianity)

Kurtosis

When estimation ICA with maximization of nongaussianity a measure of the nongaussianity is needed. Kurtosis is quantitative measure used for nongaussianity of random variables. Kurtosis of a random variable y is defined as

$$\text{kurt}(y) = \mathbb{E}[y^4] - 3(\mathbb{E}[y^2])^2,$$

which is he fourth-order cumulant of the random variable y. By assume that the random variable y have been normalised such that its variance $\mathbb{E}[y^2] = 1$, the kurtosis is rewritten as

$$\text{kurt}(y) = \mathbb{E}[y^4] - 3.$$

Because of this definition the kurtosis of gaussian random variables will then be zero and for nongaussian random variables the kurtosis will almost always be non-zero [**ICA**].
By using the absolute value of the kurtosis gaussian random variables are still zero but the nongaussian random variables will be greater than zero. In this case the random variables are called supergaussian.

For ICA the wish is to maximise the nongaussianity and therefore maximise the absolute value of kurtosis. One way to do this is to you a gradient algorithm.

One complication with the use of kurtosis as measure is the used of measured samples as the kurtosis is sensitive to outliers in the measured data set [**ICA**].
Notes: A measure of nongaussianity for the vector b which estimate 1 IC Have some outliners so we introduce negentropy

Negentropy

Another measure of nongaussianity is the negentropy which based of the differential entropy known from information theory.
The differential entropy H of a random variable $\mathbf{y}$ with density $p_y(\boldsymbol{\theta})$ is defined as

$$H(\mathbf{y}) = -\int p_y(\boldsymbol{\theta}) \log(p_y(\boldsymbol{\theta})\ d\boldsymbol{\theta}$$

Gaussian random variable has a high entropy.
The negentropy is defined as

$$J(\mathbf{y}) = H(\mathbf{y}_{\text{gaus}}) - H(\mathbf{y}),$$

which is also can be seen as a normalised differential entropy. $\mathbf{y}_{\text{gaus}}$ is a gaussian random variable.

Approximation of Kurtosis and Negentropy

Algorithm 1 Gradient Algorithm

1. Center the observed data $\mathbf{y}$. $\Delta\mathbf{w} \propto \text{sign}(\text{kurt}(\mathbf{w}^T\mathbf{z}))\mathbb{E}[\mathbf{z}(\mathbf{w}^T\mathbf{z})^3]$
2. $\mathbf{w} \leftarrow \frac{\mathbf{w}}{\|\mathbf{w}\|}$

Notes: ICA can be used on Gaussian variables as little is done in addition to decorrelate for Gaussian variable
Whiting is useful to be done beore ICA

3.3 Cov-DL

3.4 MSB

Bibliography

[1] Saeid Sanei, J.A. Chambers. *EEG Signal Processing*. Johm Wiley and sons, Ltd, 2007.

[2] Simon Foucart, Hoger Rauhut. *A Mathematical Introduction to Compressive Sensing*. Springer Science+Business Media New York, 2013.

[3] Steen Luca Faes, Esin Karahan Jitkomut Songsiri Pedro A. Valdes-Sosa Daniele Marinazzo1 Frederik Van de. "Critical Comments on EEG Sensor Space Dynamical Connectivity Analysis". In: *Brain Topography* 32 (2019).

[4] Teplan, M. "Fundamentals of EEG". In: *Measurement science review* 2 (2002).

[5] Yonina C. Eldar, Gitta Kutyniok. *Compressed Sensing: Theory and Application.* Cambridge University Presse, New York, 2012.

Appendix A

Appendix A