부분공간학습을 이용한 동기화 된 신호원의 뇌전자기 영상기법

NEUROELECTROMAGNETIC IMAGING OF CORRELATED SOURCES USING A NOVEL SUBSPACE PENALIZED SPARSE LEARNING

유재 준 (兪在俊 Yoo, JaeJun)
바이오및뇌공학과
Department of Bio and Brain Engineering

KAIST

2013
부분공간학습을 이용한 동기화 된 신호원의 뇌전자기 영상기법

NEUROELECTROMAGNETIC IMAGING OF CORRELATED SOURCES USING A NOVEL SUBSPACE PENALIZED SPARSE LEARNING
NEUROELECTROMAGNETIC IMAGING OF CORRELATED SOURCES USING A NOVEL SUBSPACE PENALIZED SPARSE LEARNING

Advisor : Professor Jong Chul Ye

by

Yoo, JaeJun

Department of Bio and Brain Engineering

KAIST

A thesis submitted to the faculty of KAIST in partial fulfillment of the requirements for the degree of Master of Science in Engineering in the Department of Bio and Brain Engineering. The study was conducted in accordance with Code of Research Ethics¹.


Approved by

Professor Jong Chul Ye

[Advisor]

¹Declaration of Ethical Conduct in Research: I, as a graduate student of KAIST, hereby declare that I have not committed any acts that may damage the credibility of my research. These include, but are not limited to: falsification, thesis written by someone else, distortion of research findings or plagiarism. I affirm that my thesis contains honest conclusions based on my own careful research under the guidance of my thesis advisor.
부분공간학습을 이용한 동기화 된 신호원의 뇌전자기 영상기법

유재준

위 논문은 한국과학기술원 석사학위논문으로 학위논문심사위원회에서 심사 통과하였음.

2012년 12월 6일

심사위원장: 예종철 (인)
심사위원: 정용 (인)
심사위원: 전대종 (인)
심사위원: 임창환 (인)
ABSTRACT

Localization of brain signal sources from EEG/MEG has been an active area of research. Currently, there exists a variety of approaches such as MUSIC, M-SBL, and etc. These algorithms have been applied for various clinical examples and demonstrated excellent performances. However, when the unknown sources are highly correlated, the conventional algorithms often exhibit spurious reconstructions. To address the problem, this paper proposes a new algorithm that generalizes M-SBL by exploiting the fundamental subspace geometry in the multiple measurement problem (MMV). Experimental results using simulation and real epilepsy data show that the proposed algorithm outperforms the existing methods even under a highly correlated source condition.

Keywords: EEG and MEG source imaging, joint sparse recovery, MUSIC, M-SBL
# Contents

Abstract ................................................................................................................... i
Contents .................................................................................................................... ii
List of Tables ............................................................................................................ iv
List of Figures ........................................................................................................... v

Chapter 1. Introduction ......................................................................................... 1

Chapter 2. Background ......................................................................................... 3
  2.1 Physiology of EEG and MEG ............................................................... 3
  2.2 General Approaches to EEG and MEG Source Localization .............. 4
  2.3 Definitions and Problem Statement ....................................................... 7

Chapter 3. Theory ................................................................................................. 8
  3.1 Overview of the Conventional Approaches ......................................... 8
    3.1.1 Least Squares ............................................................................ 8
    3.1.2 Multiple Signal Classification Families ........................................ 9
    3.1.3 Minimum Norm Families .......................................................... 11
  3.2 Bayesian Approaches ............................................................................. 12
    3.2.1 Sparse Bayesian Learning ......................................................... 13
    3.2.2 Multiple Sparse Bayesian Learning .......................................... 15
    3.2.3 Parametric Empirical Bayesian Model using Multiple Priors .......... 16

Chapter 4. Subspace Penalized Sparse Learning - A Review ......................... 18
  4.1 Role of Non-separable Penalty in M-SBL .......................................... 18
  4.2 The SPL Penalty ................................................................................... 19
  4.3 Theoretical Result on Local Minimizers .............................................. 20

Chapter 5. Implementation ................................................................................. 22
  5.1 Preconditioning ..................................................................................... 22
  5.2 Normalization of leadfield matrix .......................................................... 22
  5.3 Extension to Dipole Moment .................................................................... 23

Chapter 6. Experimental Result ......................................................................... 25
  6.1 Simulation Results .................................................................................... 25
6.2 Real Experiments

Chapter 7. Discussion

Chapter 8. Conclusion

References

Summary (in Korean)
List of Tables
List of Figures

2.1 A schematic illustration of the EEG and MEG source generation. ........................................ 3
2.2 Left: Magnetic field generated by a dipole source. Center: Electric potential generated by a dipole source. Right: The magnetic field $B$ at each sensor is calculated with respect to the moment of the dipole and the distance between the center of the dipole and the sensor. 6

6.1 Success rate comparison with varying with SNR(dB) $\in \{25, 30, 25, 40\}$: five uncorrelated signal sources ($k = 5$) given true rank ($r = 5$) ........................................................................................................ 26
6.2 Success rate comparison with varying with SNR(dB) $\in \{25, 30, 25, 40\}$: five correlated signals ($k = 5$) given true rank ($r = 5$) ........................................................................................................ 27
6.3 Success rate comparison with varying SNR(dB) $\in \{25, 30, 25, 40\}$: five uncorrelated signal sources ($k = 5$) given wrong rank ($r = 3, r = 15$). ........................................................................................................ 28
6.4 Success rate comparison with varying SNR(dB) $\in \{25, 30, 25, 40\}$: five correlated signal sources ($k = 5$) given wrong rank ($r = 3, r = 15$). ........................................................................................................ 28
6.5 MUSIC image of five uncorrelated signal sources ($k = 5$) with SNR 40dB: (a) Ground truth, (b) True rank ($r = 5$) (c) Wrong rank ($r = 3$). ........................................................................................................ 29
6.6 SPL image of five uncorrelated signal sources ($k = 5$) with SNR 40dB, (a) Ground truth, (b) True rank ($r = 5$) (c) Wrong rank ($r = 3$). ........................................................................................................ 29
6.7 MUSIC image of five correlated signal sources ($k = 5$) with SNR 40dB: (a) Ground truth, (b) True rank ($r = 5$) (c) Wrong rank ($r = 3$). ........................................................................................................ 29
6.8 SPL image of five correlated signal sources ($k = 5$) with SNR 40dB, (a) Ground truth, (b) True rank ($r = 5$) (c) Wrong rank ($r = 3$). ........................................................................................................ 29
6.9 Success rate comparison with varying with SNR(dB) $\in \{25, 30, 25, 40\}$: four anti correlated signals ($k = 4$) given true rank ($r = 4$) ........................................................................................................ 30
6.10 SPL success rate comparison with and without preconditioning the leadfield matrix $A$, varying SNR(dB) $\in \{25, 30, 25, 40\}$: five uncorrelated signal sources ($k = 5$) given true rank ($r = 5$) ........................................................................................................ 31
6.11 SPL success rate comparison with and without preconditioning the leadfield matrix $A$, varying SNR(dB) $\in \{25, 30, 25, 40\}$: five correlated signal sources ($k = 5$) given true rank ($r = 5$). ........................................................................................................ 31
6.12 Epileptic seizure wave. ........................................................................................................ 32
6.13 ICA results. There are total 22 components and we removed eight components manually; i.e. $(6, 7, 9, 13, 16, 19, 21, 22)$. ........................................................................................................ 32
6.14 Real experimental results using (a) sLORETA, (b) MUSIC, (c) M-SBL (d) SPL (our proposed algorithm). ........................................................................................................ 33
6.15 Real experimental results using SPL (our proposed algorithm) with moving window of size 10. These are the results from 90 to 105 timepoints. ........................................................................................................ 34
Chapter 1. Introduction

Electroencephalography and magnetoencephalography (EEG / MEG) are completely noninvasive techniques using external electromagnetic signal measurements. EEG and MEG have a long history from the end of the 18th century what Galvani called “animal electricity,” today better known as electrophysiology. Recently, brain function imaging and source localization by using EEG and MEG have been an active area of research [1].

EEG and MEG are the only techniques among current methods which uniquely provide a temporal resolution under 100ms. They provide superior temporal resolution over the other imaging technologies such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET). However, they suffer from the poor spatial resolution since their signals are distorted by many physical and physiological factors. As a consequence, it is hard to use EEG and MEG in practical situation for a brain function imaging.

There have been many attempts to solve the problem. Most of the contemporary attempts fall under one of the two methods of equivalent current dipole (ECD) and linearly distributed (LD) approaches [1]. In the ECD approaches, which is also called parametric methods, the signals are assumed to be represented by a relatively small number of focal sources (usually less than five). This class of localization algorithms operates expectation-maximization to estimate the source location where the first guess of sources are initialized from the data. It simulates using the forward model until the residuals between results and the data converge under the criteria. Though the idea is very simple, this works in very intuitive and reasonable way which is very powerful for many problems. The least squares (LS) approach, beamforming approach [2], and multiple signal classification (MUSIC) families [3, 4, 5] are the most popular algorithms in ECD category. Under the favorable and restricted condition, they reach good performances. LS updates the entire parameters simultaneously and tries to minimize the difference between the measurements and the estimated fields. However, since their cost function is highly nonconvex, they have difficulty with numerous local minima. In contrast, beamforming and MUSIC approach scan through the whole possible range of the locations so that they avoid the nonconvexity issue. Still, all these dipole fitting procedures are highly dependent on a priori knowledge of the dipole source numbers.

LD approaches are the good alternatives to compensate the limits of ECD approaches. The main appeal of this class is that it does not assume a model for the sources. Moreover this class of algorithm provides an estimate of the current density everywhere in the three dimensional volume of the head. In LD approaches, there are the algorithms such as minimum norm (MN) [6], focal underdetermined system solution (FOCUSS) [7], low resolution electromagnetic tomography (LORETA) [8], and standardized LORETA [9]. However, this suffers from severely underdetermined problem because the number of unknowns is usually larger than the number of measurements. The problem is relaxed by imposing various constraints. The minimum norm solution is one of the approaches that exploits the smoothness constraint. It is selected to have a minimum energy among the infinite solution space. Yet, there is no physiological evidence and moreover, because of its characteristic of constraint, all MN, FOCUSS, LORETA, and sLORETA hardly distinguish the sources near each other. They only reproduce smeared image over the surface.
There are another type of approaches to solve the problem in Bayesian aspect which is different from
the other deterministic algorithms introduced before. Sparse Bayesian learning (SBL) [10] and multiple
sparse Bayesian learning (M-SBL) [11] are the most popular algorithms. M-SBL is the extension
of SBL to the MMV problem. Even though the M-SBL has more computational complexity compared to
greedy algorithms like MUSIC methods, empirical results show that M-SBL is quite robust to noise and
unfavorable restricted isometry property constant (RIC) of the sensing matrix [12]. In spite of their high
performance, there are few researches which have studied the explicit reason except the recent research
done by Wipf et al [13]. They figured out that M-SBL can be converted to a standard compressed
sensing framework with an additional log |·| (log determinant) penalty. Unlike the standard compressed
sensing framework, which is always restricted to factorial priors, the penalty term of M-SBL behave as
a non-separable, sparsity inducing prior which cannot be expressed in the factored form [13].

However, when the unknown sources are highly correlated, many of these algorithms often exhibit
either blurry or spurious reconstructions. To address the problem, in this paper we employ a new algo-
rum that generalizes M-SBL by exploiting the subspace geometry in MMV [14]. The main contribution
of this paper is to demonstrate the effectiveness of the proposed algorithm in EEG and MEG source
localization problem.

More specifically, in [14], the authors showed the prior in M-SBL is indeed a simultaneous sparse
and rank prior that impose the row sparse and low rankness of partial sensing matrix, respectively. The
benefits of additional rank penalty is to significantly reduce the number of local minima compared to the
cases using separable penalties, i.e. factorial priors. Moreover, the authors showed that the log|·| term in
M-SBL should be replaced with a new rank penalty obtained from subspace structures of MMV problem.
Theoretical analysis of the resulting algorithm demonstrates that even though M-SBL is often impossible
to remove all local minimizers, the proposed method can do that under fairly mild conditions, without
affecting the global minimizer. In applying the algorithm for EEG and MEG problem, we performed
extensive simulations and real experiments to validate the proposed algorithm. Our results demonstrate
that the proposed algorithm can provide significantly improved reconstruction in various conditions.

The rest of this thesis is organized as follows. Chapter 2 provides basic physiology and principles in
EEG and MEG source localization. In chapter 3, we will give a brief review of previous approaches to
solve the inverse problem. We discuss the subspace penalized sparse learning algorithm in more detail
at the end of the chapter 3. In chapter 4, implementation issue is discussed and in chapter 5, we show
the experimental results. Chapter 6 and 7 will provide discussion and conclusion, respectively.

Throughout the paper, $x_i$ and $x_j$ correspond to the $i$-th row and the $j$-th column of matrix $X$, respectively. The $(i, j)$ element of $X$ is represented by $x_{ij}$. When $S$ is an index set, $X^S$, $A_S$ corresponds
to a submatrix collecting corresponding rows of $X$ and columns of $A$, respectively. For a matrix $A$,$\text{Tr}(A)$ is the trace of a matrix $A$, $A^*$ is its adjoint, $A^\dagger$ denotes the Penrose-Moor psuedo-inverse, $|A|
refers the determinant, $R(A)$ denotes the range space of $A$, and $P_A$ (or $P_{R(A)}$) and $P_A^\perp$ (or $P_{R(A)}^\perp$) are
the projection on the range space and its orthogonal complement, respectively. The vector $e_i$ denotes
an elementary unit vector whose $i$-th element is 1, and $I$ denotes an identity matrix.

A sensing matrix $A \in \mathbb{R}^{m \times n}$ is said to have a k-restricted isometry property (RIP) if there exist left
and right RIP constants $0 \leq \delta_k^L, \delta_k^R < 1$ such that

$$(1 - \delta_k^L)||x||^2 \leq ||Ax||^2 \leq (1 + \delta_k^R)||x||^2$$

for all $x \in \mathbb{R}^n$ such that $||x||_0 \leq k$. A single RIP constant $\delta_k = \max \{\delta_k^L, \delta_k^R\}$ is often referred to as the RIP constant.
Chapter 2. Background

2.1 Physiology of EEG and MEG

When neuron is excited by the others, excitatory postsynaptic potentials (EPSPs) are generated at the apical dendrite tree. Then, the apical dendrite membrane is transiently depolarized so that it becomes electronegative compared to the cell soma and basal dendrites. The potential difference causes current flows from the unexcited membrane of the soma and basal dendrites to apical dendrite. The current, which takes the short route, flows through the dendritic trunk and this intracellular current is commonly called primary current while extracellular current, secondary or volume current, makes a closed current loop imposed by the conservation law of electric charge [1].

The spatially structured arrangements of cells are crucial for the superposition of neural currents such that they produce measurable fields. Therefore, it is believed that macrocolumns of tens of thousands of synchronously activated large pyramidal cortical neurons are the sources of EEG and MEG generations. These cells are coherently distributed in a parallel or perpendicular orientation to the cortical surface. When they are activated simultaneously, their contributions sum up into a macroscopic current. Both primary and secondary currents of the cell assembly influence the scalp electric potentials and magnetic fields. Note that, however, the electromagnetic fields measured in EEG and MEG are hardly due to action potentials. This is because action potentials generate currents in both opposite directions along the axons so that the sum of their effects is closed to zero in a macro view. They also last too short to be measured when they are compared to the synaptic currents and induced ionic currents stated above. Figure 2.1 shows the schematic illustration of the EEG and MEG source generation.

Because of its signal physiology, EEG and MEG are completely noninvasive techniques to image the source inside the head in real time. Therefore, it has been an active area to study the brain function using EEG and MEG. Recently, fMRI has been used to investigate the brain function abnormalities such as epilepsy [15]. It is possible to detect the blood oxygen level dependent (BOLD) signal changes with fMRI which are caused by the metabolic changes during interictal seizure. However, fMRI often fails to

Figure 2.1: A schematic illustration of the EEG and MEG source generation.
detect the epileptic activity due to its poor temporal resolution limits while abnormalities can be clearly seen on EEG and MEG studies.

Since EEG and MEG directly measure the external electromagnetic signal from brain activities, there are two advantages in brain function study. First, they are able to provide uniquely superior temporal resolutions under 100ms. This allows a potential to detect the millisecond timescale of neural population activities. Second, given the size of the head, their source currents are fast enough to consider that there is no propagation delay, from the neurons to the sensors.

Still, they have a poor spatial resolution compared to the other methods. As a result, it is hard to localize the exact regions of the seizure inducing sources so that it is small enough to operate a surgery. There are several reasons for low spatial resolutions of EEG and MEG source localizations, such as their inherent signal characteristics, relatively few spatial measurements, and ambiguities underlying static electromagnetic inverse problem are the main reasons.

More specifically, EEG measures the electric potential on the scalp by the sources inside the head. In the case of single dipole source, one would observe potential differences on scalp, independent to dipole orientation. Depending on the orientation, however, there will be either one or two of positive and negative patterns. The patterns are very smooth and the dipole activity affects almost all sensors. This means that when several sources are active, what we get from the scalp potential is a blurred image of the activity underneath which is difficult to disentangle and interpret. Moreover, electric potential is mainly the result of volume currents conduction through the head. Therefore, the measured potential will highly depend on the tissue conductivities. This explains why EEG needs a complicated model for the data to account for the inhomogeneity in each of the head tissue. This will be discussed more in detail at the next section.

MEG measures the magnetic field outside the head. Contrary to EEG, MEG does not need a reference point and does not depend upon tissue conductivities and head model. However, it ignores the information of radial sources to the cortex and its signal decays faster than EEG as the source location goes deep. The magnetic fields measured by MEG signal is mainly due to the primary currents because the magnetic fields caused by volume currents seem to cancel out each other. This causes a loss of information. Meanwhile, the electric potential of EEG signal is sensitive to both and mainly volume current that reaches to scalp. As a consequence, to properly image the source of EEG and MEG, one should take into account those anatomical and physical characteristics of the signal in solving the forward and inverse problem [16].

2.2 General Approaches to EEG and MEG Source Localization

Most of source localization approaches in EEG and MEG is characterized by three parts. The first part is related to the definition of the solution spaces and parametric representation of the sources. The second part is all about the physiological, physical, and geometrical information of the head. These two are used in comprising the generative model, also called forward model. Finally, given the forward model, solving the inverse problem with some constraints provides a unique solution. The forward problem computes the relationship between head currents in certain location and the electromagnetic fields outside so that solving this problem means to make a generative model. Conversely, the inverse problem is the process to estimate the location and distribution of the current sources from electromagnetic recordings.

However, it was shown by Helmholtz that even with a complete knowledge of electromagnetic fields outside, we cannot determine internal sources uniquely [17]. This requires solving a highly ill-posed
inverse problem with many possible solutions. Field superposition permits generation of identical external fields for different source configurations so that the selection of particular solution often requires a priori knowledge of overall brain physiology.

To solve the inverse problems, we need to solve the forward problem. Even though the structures of neural cells are simple, their ionic current generation and conduction follow a complex biophysics. If we know the primary source and the surrounding conductivities, the resulting electric potential and magnetic field can be calculated from Maxwell’s equations \[1, 16, 18\]. In neuroelectromagnetism, we generally deal with the frequencies between 0.1 and 100Hz \[1\]. Consequently, we can adopt the quasistatic approximation of Maxwell equations. Under the quasistatic approximation, the current flow \( J(r') \) inside the head can be related rather simply to the magnetic field \( B(r) \) outside as well as scalp potential at location \( r \) though the well-known Biot-Savart law

\[
B(r) = \frac{\mu_0}{4\pi} \int J(r') \times \frac{r - r'}{||r - r'||^3} dv' \tag{2.1}
\]

where \( \mu_0 \) is the permitivity of free space, \( r' \) is the current source location, and \( r \) is the sensor location outside the head. Here, we can partition the total current density, \( J(r') \) into primary and volume current flows:

\[
J(r') = J^p(r') + J^v(r') = \sigma(r')E(r') = J^p(r') - \sigma(r')\nabla V(r') \tag{2.2}
\]

where \( \sigma(r') \) is the conductivity profile of the head tissues and by quasistatic approximation, the electric field \( E(r') \) is the negative gradient of the electric potential \( V(r') \) \[1\].

By combining (2.1) and (2.2), it is obvious that the EEG are dependent to the conductivity profile of the head model. Note that the volume current term, \( \sigma(r')\nabla V(r') \), is dependent to \( \sigma(r') \) while the primary current term does not. Therefore, we need a specific head model for an accurate EEG calculation whose signal is highly from the volume currents. For simplicity, we often assume an isotropic conductivity profile of the head tissues. Even though our heads are anisotropic, the spherical head models works quite well. This simple head model is routinely used in most clinical and research applications to EEG and MEG source localization.

The simplest realistic model for the source in the brain is the current dipole. Obviously, monopolar source cannot produce a potential difference on the scalp while contribution of multipolar source decreases rapidly with distance. Therefore, the dipolar source model seems to be sufficient for the EEG and MEG. For a single dipole at location \( r' \), the magnetic field observed at the sensor location \( r \) is achieved as figure 2.2. This can be extended to the effect of multiple dipoles.

In electromagnetic source localization, we often assume the current dipole as a point source \( J(r') = q\delta(r' - r_q) \) where \( \delta(r) \) is a Dirac delta function with moment \( q \equiv \int J(r')dr' \). Note that this measurement becomes linear in the dipole moment \( q \).

With the source and head models which are used in calculations of the forward problem, we can now provide linear algebraic models which will be taken in the inverse methods. To clarify the different approaches in inverse methods which will be dealt in next section, it is convenient to separate the parameters of dipole moment \( q \) as its magnitude \( q \equiv ||q|| \) and orientation \( \Theta = q/||q|| \). Here, we denoted the dipole moment \( q \) as an unknown \( x \) in the inverse problem point of view. Mathematically, this can be stated as

\[
\begin{align*}
y &= L(r', \Theta)q + \epsilon \\
    &= A(r')\Theta q + \epsilon \\
    &= A(r')x + \epsilon \tag{2.3}
\end{align*}
\]
where $\mathbf{y} \in \mathbb{R}^m$ is a vector containing electromagnetic field measurements at the $m$ sensors and $L(\mathbf{r}', \Theta)$ is a parameterized leadfield matrix with its location and orientation. As the measurement is linear in the dipole moment $\mathbf{q}$ in our model, $L(\mathbf{r}', \Theta)$ can be separated to $A(\mathbf{r}')$ and $\Theta$ where $A \in \mathbb{R}^{m \times 3n}$ is an leadfield matrix whose elements are explicitly specified as below.

$$
\begin{bmatrix}
y_1 \\
n_2 \\
\vdots \\
y_m
\end{bmatrix} =
\begin{bmatrix}
a_{1,x}, a_{1,y}, a_{1,z}, \ldots, a_{3n,z}
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
\vdots \\
x_{3n}
\end{bmatrix} + \epsilon
$$

The columns of $A$ are three times the number of voxels, i.e. $3n$, representing the locations of point dipoles $\mathbf{r}'$ in Cartesian coordinates $(x, y, z)$. Thus, each column $A_{i,j}$ corresponds to $j$th component of a dipole located in the $i$th voxel and each row corresponds to one sensor. Therefore the EEG and MEG at each sensor site is the sum of the weighted lead fields from each source. Then, the goal in EEG and MEG source imaging problems is to estimate unknown $\mathbf{x}$, given $\mathbf{y}$ and $A$.

Now suppose that multiple response vectors have been collected over $\mathcal{N}$ snapshots with an equivalent leadfield matrix $A$. Such a situation is common in EEG and MEG recordings and various applications [3, 11]. Then, the multiple response model becomes

$$
\mathbf{Y} = \mathbf{AX} + \mathbf{E}
$$

where $\mathbf{Y} \in \mathbb{R}^{m \times \mathcal{N}}$, $\mathbf{X} \in \mathbb{R}^{3n \times \mathcal{N}}$, and $\mathbf{E} \in \mathbb{R}^{m \times \mathcal{N}}$.

One of the main technical difficulty is that ensuing inverse problem from $\mathbf{X}$ to $\mathbf{Y}$ becomes highly non-linear. Moreover, the number of unknowns is usually larger than the number of measurements in EEG and MEG. Consequently, estimating the location and distribution of the current sources from electromagnetic recordings requires the solution of highly underdetermined inverse problem.

To deal with such infinite numbers of solutions in the inverse problem, there have been numerous attempts to reduce the range of feasible solution by imposing constraints. One is to make head and source models as accurate as possible. Combining with the physiological information from the other imaging modalities may also give a good constraint to the problem. For example, eliminating the regions such as...
skull from the possible source positions may also exclude large amount of unlike solutions. Using a finite element method (FEM), one can consider the realistic head model with a conductivity information so that the field distortion due to skull is corrected. However, it causes a time-consuming and exhaustive calculation of forward problem.

Other than such heuristic way, there are the algorithmic attempts to solve the problem by exploiting the various penalty functions such as the smoothness and sparsity assumptions. We will discuss these classes of algorithms more in detail at the next chapter of this paper.

2.3 Definitions and Problem Statement

Before moving into the main contents of the paper, let us first define a few concepts such as sparsity and diversity. We first define diversity:

\[ \|x\|_0 \triangleq \sum_{i=1}^{n} I[|x_i| > 0] \]  \hspace{1cm} (2.5)

where \( x \) is a single measurement vector (SMV) with \( i \in \{1, \ldots, n\} \) elements and \( I[\cdot] \) is the indicator function then \( \|x\|_0 \) is diversity. In turn,

\[ \text{sparsity} = n - \text{diversity}. \]  \hspace{1cm} (2.6)

The nonzero elements of \( x \) are usually representing the active sources.

Now, the sparsest solution is obtained by solving

\[ \min_x \|y - Ax\|_F + \lambda \|x\|_0, \quad y = Ax + \epsilon. \]  \hspace{1cm} (2.7)

Unfortunately, it has been shown that even for the SMV case, finding the solution that has the minimum number of nonzero entries is NP-hard nor is it clear how to select the trade-off parameter \( \lambda \) [19].

In MMV framework, we assume that the source locations are relatively stationary up to a certain number of snapshots. Moreover, epileptic sources, for example, are usually sparsely distributed so the problem can be formulated as the following multiple measurement vector problem or joint sparse recovery problem:

\[ \min_X \|Y - AX\|_F + \lambda \|X\|_0, \quad Y = AX + E \]  \hspace{1cm} (2.8)

where \( \|X\|_0 \) denotes the number of non-zero rows.

The MMV problem further complicates the problem due to the nature of joint sparsity. Regularization methods are the most popular alternatives which relax the problem based upon minimizing the cost function with \( l_p \) norm. This \( l_p \) norm surrogates the \( l_0 \) norm which is a sparsity penalty, where \( 0 \leq p < 1 \).

\[ \min_X \|Y - AX\|_F + \lambda \|X\|_p, \quad Y = AX + E \]  \hspace{1cm} (2.9)

Finally, with the matrix \( A \), spark is defined as the size of smallest linearly dependent subset of \( a_i \), column vector of \( A \). For example, if \( A \) has full column rank, \( \text{spark}(A) = n + 1 \). We say the matrix \( A \in \mathbb{R}^{m \times n} \) is overcomplete when \( \text{rank}(A) = m \) and \( m < n \).
Chapter 3. Theory

3.1 Overview of the Conventional Approaches

Using the algebraic model, there are two general approaches to localize multiple dipole sources within the brain, namely equivalent current dipole (ECD) and linear distributed (LD) methods. Many advanced reconstruction methods are currently available in either approach. For ECD approaches, they includes least squares (LS) and multiple signal classification (MUSIC) \[3\] while the LD approaches have minimum norm (MN)[6], focal underdetermined system solution (FOCUSS) \[7\], low resolution electromagnetic tomography (LORETA) \[8\], and standardized LORETA[8].

Compared to the other methods, there are also somewhat different approaches which solve the problem by using Bayesian aspect. Among these, such as sparse Bayesian learning and multiple-SBL (M-SBL) are known best \[11\]. They have a different root with the others while sharing similar properties with LD approaches.

However, there are several key problems which impair the performances for each approach. For ECD cases, they have difficulties in estimating the accurate source numbers in advance. It is also hard to apply extended source models for their localization. Furthermore, they exhibit some intensity changes depending on the depth of source location.

Meanwhile, LD and Bayesian approaches suffer from the highly underdetermined and ill-posed inverse problem. Without constraints, this class of algorithms takes a tremendous amount of information and computational loads to get a 5mm resolution \[16\]. The inversion can be very unstable even in the case that the unique solution is possible under the constraints \[20\]. As a result, the conventional algorithms often exhibit either blurry or spurious reconstructions. In this case, proper regularization methods are inevitable to reduce the range of feasible solutions.

In this section, we briefly review the conventional algorithms and study the M-SBL as one of the state-of-art algorithms.

3.1.1 Least Squares

Let us assume that the forward model can be represented with a nonlinear parametric estimation model

\[ Y = A(\mathbf{r}, \Theta)X + E \]  \hspace{1cm} \text{(3.1)}

where \( Y \) is an \( m \times N \) spatio-temporal matrix with \( k \) dipole sources. Here, \( A \in \mathbb{R}^{m \times n} \) is an leadfield matrix which is parameterized with location \( \mathbf{r} \) and orientation \( \Theta \) of \( n \) dipoles. The least squares approach attempts to minimize the differences between the measured data and the fields computed from the estimated sources using a forward model. The goal of the inverse problem is to estimate the parameters, \( \{\mathbf{r}, \Theta, X\} \) given the measurements, \( Y = \{y_1, y_2, \cdots, y_N\} \). This search continues until the following cost function is minimized under various constraints \[21\]:

\[ L_{LS} = \|Y - A(\mathbf{r}, \Theta)X\|_F^2. \]  \hspace{1cm} \text{(3.2)}
We can easily relax the problem by adjusting the cost function into a projection minimization problem as
\[
\mathcal{L}_{LS} = \|Y - P_A Y\|_F^2 = \|P_A^\perp Y\|_F^2
\]
(3.3)
where \(P_A^\perp \in \mathbb{R}^{m \times m}\) projects \(Y\) on to the orthogonal complement of \(A(r, \Theta)\) [22]. However, this kind of mapping suffers from the computational cost because a huge number of parameters are simultaneously optimized.

The performance of the LS method, however, is still substantially subject to the dipole model. It is crucial to decide the right number of sources a priori. If too few dipoles are selected, the resulting parameters are impaired by the missing dipoles and if too many dipoles are selected, then the redundant dipoles will decrease the accuracy. In practice, this is a significant drawback because there is no way to know the exact number of sources in advance. Furthermore, as the number of sources increase, the possibility of trapping in the local minima increases due to the nonconvexity of its cost function.

### 3.1.2 Multiple Signal Classification Families

**MUSIC**

Multiple signal classification (MUSIC) approach was developed in the array signal processing community [3]. As before, let the problem be \(Y = A(r, \Theta)X + E\) and assume that there are \(k\) dipole sources. Here, \(A \in \mathbb{R}^{m \times n}\) is parametric leadfield matrix. Then, its goal is to find the appropriate parameter set, \(\{r, \Theta, X\}\) for \(k\) sources. For convenience, we drop the notation of parameters inside \(A\) in following contents.

MUSIC tries to find the source by calculating the subspace correlation between the signal and the estimated subspace. In the presence of noise, the sample correlation matrix is
\[
R_Y = E\{YY^*\} = E\{AXX^*A + AXE^* + EX^*A + EE^*\}
\]
(3.4)
where \(E\{\cdot\}\) means the expectation of the argument. If we assume that our noise model follows the white Gaussian distribution with variance \(\sigma^2 I\) and the snapshot \(N \to \infty\), we may think that the (3.4) converges to:
\[
\lim_{N \to \infty} R_Y = \lim_{N \to \infty} E\{YY^*\} = E\{AXX^*A + EE^*\} = AR_XA + \sigma^2 I .
\]
(3.5)

If we compute the singular value decomposition (SVD) of \(R_Y\), we will find
\[
R_Y = \begin{bmatrix} U_s & U_n \end{bmatrix} \begin{bmatrix}
\sigma_1^2 + \sigma^2 & & \\
\sigma_2^2 + \sigma^2 & \ddots & \\
& \ddots & \ddots \\
& & \sigma_k^2 + \sigma^2 \\
& & \ddots & \sigma^2
\end{bmatrix} \begin{bmatrix} U_s^* \\ U_n^* \end{bmatrix} .
\]
(3.6)

It looks that the noise only shifts the eigenvalue with value \(\sigma^2\) and there is no problem in searching the exact signal numbers. However, this regular shifting property may not hold when the snapshot, \(N\), is finite.
By directly looking at the singular value decomposition (SVD) of $Y$, $\text{svd}(Y) = U\Sigma V^*$, we can get a similar result with (3.6). We can select the rank by seeing the eigenvalue profile, $\Sigma$. After the number of source, $k$, is selected, provided that $A$ has full rank and $m > k$ without noise, the corresponding first $k$ columns of $U$ spans the exact signal subspace, $\text{span}(U_s)$, while the other columns will span the exact noise subspace, $\text{span}(U_n)$.

Then, finding the true source location and orientation becomes finding the parameters maximizing a cost function of MUSIC

$$L_{\text{MUSIC}}(i) = \frac{1}{\bar{a}_i^* U_n U_n^* \bar{a}_i}, \quad i \in \{1, \cdots, n\}$$

which results in a huge peak at the true position as the denominator of (3.7) becomes zero in the ideal case. Therefore, MUSIC estimation also has a problem in determining the exact number of sources as LS approach does. If we overspecify the true rank, the probability increases that we unintentionally include a noise-only subspace which correlates with the forward models. If we underspecify the rank, the performance can dramatically be reduced so some caution is called for in rank selection.

Compared to LS methods, however, MUSIC can find each source by scanning through the possible set of locations and orientations whereas LS methods search the parameters for all sources simultaneously [1].

Two problems arise with the use of MUSIC in EEG and MEG signals. MUSIC suffers if the number of snapshot is not sufficient to identify accurate signal subspace estimation. Moreover, the basic regularity condition for the success of MUSIC is that the sources should not be correlated. In EEG and MEG, the sources of interest usually have transient time courses. This limit the number of time samples. It is well-known that MUSIC cannot detect the sources when the measurement $Y$ is rank deficient, i.e. the number of targets is greater than the rank of signal subspace.

Secondly, the independency assumption fails in many clinical or experimental situations. Under the epilepsy seizure, EEG and MEG signal often show an abnormal and hypersynchronous neuronal activity in brain. If two dipole sources have synchronous activation, then the two-dimensional signal subspace collapses to a one-dimensional subspace and MUSIC will fail to localize either of the sources.

**Recursive MUSIC**

There are several families of MUSIC approaches which have tried to overcome such limits and improved the performance. Recursive MUSIC (R-MUSIC) [4] and recursive applied and projected MUSIC (RAP-MUSIC) [5] are one of the popular examples in neuroelectromagnetic source localization research.

In R-MUSIC, for example, Mosher et al [4] argued that correlated signal can be corrected by adjusting the concept of source models. They proposed a spatio-temporal independent topographies (IT) model instead of single dipole model. They tried to build up a new source model for data as the sum of fixed number of IT’s finding with a recursive manner. R-MUSIC takes advantage of the recursive procedure. During each step of the procedure, it exploits the support set, which is found previously, to explain the signal subspace. As a result, R-MUSIC only needs to separate the global solution at the time. Although there is an issue of selecting a stopping criterion, its source detecting performance is better than the original MUSIC.

However, the complexity of the search increases in combinatorial sense to the increase of correlated source number and possible source locations. Furthermore, because of its greedy searching property, if any of the supports are wrongly selected during the iteration, this will affect the following selection of supports.
3.1.3 Minimum Norm Families

LD approaches have several advantages over ECD approaches. When dealing with more than one sources, ECD approaches have been reported to be less reliable than LD approaches even under the favorable conditions [23]. Moreover, unlike ECD, LD approaches considers all possible source locations as a solution with the orientation constrained to equal the local surface normal. Since the only unknowns are source magnitudes, the inverse problem is linear while the fixed orientation assumption remains as an inherent drawback of the methods leaving a possibility of wrong solution [1]. Contrast to ECD, this class of algorithms does not assume a model for the sources so that it can be used without the prior knowledge. However, the problem is severely underdetermined and ill-posed because the number of unknowns is usually larger than the number of measurements. This problem is relaxed by imposing some constraints on the inverse problem such as the smoothness of the inverse solution.

MN

The minimum norm (MN) solution is one of the first approaches to linear inverse solutions in LD methods [6]. It chooses the solution that has minimum energy which imposes a smoothness constraint on the inverse problem. As all LD methods do, its only unknown is the magnitudes of sources so that the problem becomes \( Y = AX + E \) where \( A \in \mathbb{R}^{m \times 3n} \) and \( n \) is the every possible location in head.

Here, the minimum energy means the solution \( X \) which has the least Euclidean norm value

\[
\min_X \|X\|_2^2 \quad \text{subject to} \quad Y = AX. \tag{3.8}
\]

It is shown that the solution can be found by multiplying the pseudo-inverse of \( A \) to \( Y \), \( X = A^+Y \) [21].

However, there is no physiological evidence that the solution with minimum energy is the most appropriate one. There have been studies that the MN solution indeed tends to retrieve only weak and localized activation patterns [8]. It results in a blurred image due to its smoothness constraint and it becomes hard to disentangle the sources nearby. Besides, the solution of (3.8) are biased toward the scalp due to the increasing signal power of the source which is near the sensor on the scalp [21].

WMN and FOCUSS

The weighted minimum norm (WMN) and focal underdetermined system solution (FOCUSS) algorithms are introduced to compensate this problem by giving a weight inversely to the source distance from the scalp [7]. Hence, the equation (3.8) is changed to

\[
\min_X \|WX\|_2^2 \quad \text{subject to} \quad Y = AX \tag{3.9}
\]

with a solution \( X = W^{-1}A^+(AW^{-1}A^*)X \). Here, \( W \in \mathbb{R}^{3n \times 3n} \) is a diagonal matrix giving a weight to the sources as:

\[
W = \begin{bmatrix}
\frac{1}{\|a_1\|_2} & 0 & 0 \\
0 & \ddots & 0 \\
0 & 0 & \frac{1}{\|a_{3n}\|_2}
\end{bmatrix}.
\tag{3.10}
\]

By exploiting the weight to the depth, unlike the MN, weighted MN can truly use the three dimensional solution space.

FOCUSS furthermore manipulates the cost function of weighted minimum norm by iteratively updating the \( W \) using the information gathered from the previous results in (3.9). The iteration stops
when there is no significant change in the estimation. These two algorithms show some improvements over the MN solution in blurred source image.

Still, there is no evidence that the mathematical penalty has a relationship with some physiological factors. Moreover, FOCUSS has a problem with the local minimum highly dependent on the initialization of the algorithm.

**LORETA and sLORETA**

Suggesting some physiological assumptions in mathematical process, Pascual-Marqui *et al* proposed low resolution electromagnetic tomography algorithm (LORETA) [8] and standardized LORETA (sLORETA) [9].

They assume that: (1) neighboring neurons are synchronously activated with only the negligible changes in their orientations and (2) the signals recorded on the scalp is mainly due to the superficial areas of the cortex. LORETA uses a Laplacian operator producing a smooth and superficial solution

\[
\min_X \| BW X \|_2^2 \text{ subject to } Y = AX
\]

where \( B \) is the Laplacian operator. There exist some controversy about the assumptions and stability of LORETA [24]. Nevertheless, it is one of the most widely used algorithms in the field. Substantial consistency between LORETA and other traditional neuroimaging techniques has been reported [25].

Meanwhile, sLORETA choose the solution which minimizes the different cost function:

\[
\min_X \| Y - AX \| + \lambda \| X \|_2^2
\]

where it uses a zero-order Tikhonov-Phillips regularization [9]. Hence, this provides a unique solution to the inverse problem:

\[
\hat{X} = A'(AA' + \lambda I)^{-1} Y = RX
\]

where \( \hat{X} \) is the candidate sources and \( X \) are the actual sources. It is obvious that the resolution matrix \( R \) becomes:

\[
R = A'(AA' + \lambda I)^{-1} A.
\]

sLORETA is reported to have superior performance over LORETA having zero-localization error [9]. However, every LD method demonstrated until now fail to detect the deep source in the presence of superficial source. Under the condition of closely located sources, they will only reproduce a single source in-between the original sources with a broadly blurred image. Thus, even sLORETA fails to successfully localize two simultaneously active sources if they do not have distinct fields with a similar strength.

### 3.2 Bayesian Approaches

Unlike the other methods illustrated before, there exists a very different class of powerful algorithms that are based on Bayesian approaches while sharing some similar properties with LD approaches. The LD approaches represents a highly undetermined but linear system as mentioned before. This linear model can be treated in a Bayesian way, using priors to achieve a unique solution.

This Bayesian approaches with priors inducing sparseness on the solution, for example, provide the powerful automatic relevance determination (ARD) cost function which are popular in the machine
learning community [26]. Their cost function automatically switches off the irrelevant parameters during the learning procedure which provides the powerful engine to obtain a sparse solution.

Bayesian approaches with this sparsity approximation problem are typically divided into two categories: (1) maximum a posteriori (MAP) estimation using a fixed, computationally tractable priors and (2) empirical Bayesian approach that uses a parameterized prior that is iteratively updated by learning from the data.

Note that the classical regularization approaches can be seen as MAP estimation context with single prior. They are characterized by the priors inside their cost function. Interestingly, we can interpret the previous EEG and MEG source localization methods by using a variety of implicit Bayesian priors under some proper assumptions (e.g. Gaussian assumption)[27].

Therefore, in the following subsections, we will concentrate in the empirical Bayesian approaches and among these, so-called sparse Bayesian learning (SBL) and its extension to MMV problem are known best [28].

3.2.1 Sparse Bayesian Learning

Suppose that we have the measurement signal with the form of sparse weight representation from overcomplete dictionaries

\[ y = Ax + \epsilon \]  

(3.14)

where \( A \in \mathbb{R}^{m \times n} \) is an overcomplete matrix, \( x \) is a weight vector to be learned, \( y \) is an observed signal, and \( \epsilon \) is Gaussian noise.

As the LD methods, the inverse problem is a highly underdetermined so that its estimation and optimization tasks are uneasy. In Bayesian framework, it provides a remedy by assuming that the unknown weight \( x \) has drawn from some distribution \( p(x) \). This allows us to narrow the solution space with application-specific constraints. There are two approaches in this framework, one is maximum a posteriori (MAP) estimation and the other is empirical Bayesian approach.

Because directly solving (2.7) needs a difficult optimization challenge, the MAP strategy of estimation is to replace the troublesome prior

\[ p(x|\theta) \sim \exp \left( -\| x \|_0 \right) \]  

(3.15)

with a distribution that is manageable. Therefore, in conventional MAP based algorithms, prior probability is fixed and the weight prior is given by

\[ p(x) \sim \exp \left( -\sum_{i=1}^{n} |x_i|^p \right) \]  

(3.16)

where \( p \in [0, 1] \).

Given that fixed prior, MAP solution is

\[
x_{MAP} = \max_{x} p(x|y) \\
= \max_{x} p(x, y)p(y) \\
= \max_{x} p(y|x)p(x) \\
= \max_{x} -\log p(y|x) - \log p(x) \\
= \min_{x} ||y - Ax||^2 + \lambda \sum_{i=1}^{n} |x_i|^p.
\]  

(3.17)
When $p = 1$ in (3.16), often referred to as Basis Pursuit (BP) algorithm [29] and the LASSO [30], they solve the minimum $l_1$ norm regularization solution. With Basis Pursuit, cost function is now a convenient convex function so that it always converges to global solution. In practice, however, its solution fails to achieve a sufficient sparseness with a large number of small nonzero elements. This is called structural error [10].

When $0 \leq p < 1$, on the other hand, the cost function is used in FOCUSS algorithm. As $p \to 0$, this prior becomes similar with (3.15). In this case, however, it suffers from a lot of local minima. Then the solution frequently converges to suboptimal solution which is referred to as convergence error [10].

Instead, sparse Bayesian learning (SBL) settle the problem with an empirical prior on the weights which encourages sparsity [10]. SBL assumes the Gaussian likelihood model:

$$p(y|x; \lambda) = \frac{1}{(2\pi)^{m/2} \lambda} e^{-\frac{1}{2\lambda} ||y - Ax||^2} \quad (3.18)$$

and a parameterized prior from data:

$$p(x; \gamma) = \prod_{i=1}^{n} \frac{1}{(2\pi \gamma_i)^{1/2}} e^{-\frac{1}{2\gamma_i} x_i^2} \quad (3.19)$$

where $\gamma = [\gamma_1, \cdots, \gamma_n]^*$ is a hyperparameter vector which controls the prior variances inducing a sparse solution. Note that, if the prior variance $\gamma$ becomes sparse, so does the $x$ by its mean zero distribution, (3.19).

Therefore, the resulting posterior density is,

$$p(x|y; \gamma) = p(x, y; \gamma) / p(y; \gamma) = p(y|x; \lambda)p(x; \gamma) / \int p(x, y; \gamma) dx = N(\mu, \Sigma) \quad (3.20)$$

with mean and covariance given by

$$\mu = \Gamma A^* \Sigma_y^{-1} y$$
$$\Sigma = \Gamma - \Gamma A^* \Sigma_y^{-1} A \Gamma \quad (3.21)$$

It is quite intuitive and reasonable to select the posterior mean $\mu$ as a estimator for $x$. Here, marginal likelihood distribution becomes

$$p(y; \gamma, \lambda) = \int p(y|x; \lambda)p(x; \gamma) dx = (2\pi)^{-m/2} |\Sigma_y|^{-1/2} \exp \left[-\frac{1}{2} y^* \Sigma_y^{-1} y \right] \quad (3.22)$$

where $\Gamma = \text{diag}(\gamma)$ and $\Sigma_y = \lambda I + A \Gamma A^*$.

Now, the problem has changed from finding a sparse solution to estimating a hyperparameter vector. Then, to obtain the right $x$ is tantamount to achieve the right $\gamma$ which maximizes the marginal likelihood. For simplicity, $-2 \log(\cdot)$ transformation is added and the $\gamma$ which is maximizing the marginal likelihood is equal to the one minimizing the cost function:

$$\mathcal{L}(\gamma) = -2 \log \int p(y|x)p(x; \gamma) dx = -2 \log p(y; \gamma) = \log |\Sigma_y| + y^* \Sigma_y^{-1} y. \quad (3.23)$$

There are two ways of minimizing the cost function $\mathcal{L}(\gamma)$ with respect to $\gamma$, EM and fixed point update rule. However, original EM shows a very slow convergence so that we prefer to replace the M step of original EM to the fixed point approach at the expense of proven convergence [28].

- 14 -
Thus, a fixed point equation is formed by taking derivative of (3.23) with respect to \( \gamma \) and equating to zero. Then, \( \gamma \) can be estimated:

\[
\gamma^{\text{new}}_i = \frac{\mu^2}{1 - \gamma^{-1}_i \Sigma_{ii}}, \quad i = 1, \ldots, n.
\]  

(3.24)

Here, \( \mu \) is updated as \( \hat{x} \), a \( \chi \) estimator. The iteration between \( \mu \) and \( \gamma \) is repeated until the convergence.

### 3.2.2 Multiple Sparse Bayesian Learning

We can extend SBL to MMV problem and this is called M-SBL. Under the appropriate assumptions of noise and signal Gaussian statistics, one can show that M-SBL minimizes the following cost function in a so-called \( \gamma \) space [31]:

\[
L^\gamma(\gamma) = \text{Tr}(\Sigma^{-1}_y YY^*) + N \log |\Sigma_y|
\]  

(3.25)

where

\[
\Sigma_y = \lambda I + A\Gamma A^* , \quad \Gamma = \text{diag}(\gamma).
\]  

(3.26)

With an estimate of \( \Gamma \), a commonly accepted point estimate for \( X \) is given by

\[
X = \Gamma A^*(\lambda I + A\Gamma A^*)^{-1}Y.
\]  

(3.27)

Note that if \( \Gamma \) is sparse, the corresponding coefficient estimate \( X \) will be sparse as well.

Wipf et al [13] demonstrated the fundamental duality between both M-SBL and standard sparse recovery framework. In particular, the cost functions associated with both approaches can be expressed either in \( \chi \) space or in \( \gamma \) space.

Then, the minimization problem of the cost function (3.25) in \( \gamma \) space, which is somewhat unclear, can be equivalently represented as the following standard sparse recovery framework in \( \chi \) space which is more transparent in optimizing process [13]:

\[
\min_X L^\chi(X), \quad L^\chi(X) = \|Y - AX\|^2_F + \lambda g_{\text{msbl}}(X)
\]  

(3.28)

where \( g_{\text{msbl}}(X) \) is a penalty given by

\[
g_{\text{msbl}}(X) \equiv \min_{\gamma \geq 0} \text{Tr}(X^*\gamma^{-1}X) + N \log |\lambda I + A\Gamma A^*|.
\]  

(3.29)

Here, \( X \) ia a global minimum of (3.28) if and only if \( \gamma \) is a global minimum of (3.25) which is also the sparsest solution. This duality enables us to suggest alternative means of constructing cost functions and algorithms for promoting sparsity and offers an avenue for improvement. Wipf et al [13] proposed the following alternating minimization approach to solve the minimization problem (3.28).

**Step 1: Minimization with respect to \( X \)**

For a given estimate \( \gamma^{(t)} \) at the \( t \)-th iteration, we can find a close form solution for \( X \) in (3.28):

\[
X^{(t)} = \Gamma^{(t)} A^*(\lambda I + A\Gamma^{(t)} A^*)^{-1}Y, \quad \Gamma^{(t)} = \text{diag}(\gamma^{(t)}).
\]  

(3.28)
Step 2: Minimization with respect to $\gamma$

In this step, for a given $X(t)$ we need to solve the following minimization problem:

$$
\gamma^{(t+1)} = \arg \min_{\gamma \geq 0} G^{(t)}(\gamma)
$$

where

$$
G^{(t)}(\gamma) \equiv \text{Tr} \left( X(t)^* \Gamma^{-1} X(t) \right) + N \log |\Sigma_y| .
$$

(3.30)

Wipf et al. [13] have found the coordinate-wise minimizer of (3.30) by calculating the derivative. Note that the derivative with respect to each component is given by

$$
\frac{\partial G^{(t)}}{\partial \gamma_i} = - \sum_j |x_{ij}^{(t)}|^2 / \gamma_i^2 + Na^*_i (\lambda I + A \Gamma A^*)^{-1} a_i
$$

since

$$
\frac{\partial |\Sigma_y|}{\partial \gamma_i} = |\Sigma_y| \text{Tr} \left( \Sigma_y^{-1} \frac{\partial \Sigma_y}{\partial \gamma_i} \right) .
$$

Setting the derivative to zero after fixing $\Gamma := \Gamma^{(t)}$, this observation leads us the following fixed point update of $\gamma$:

$$
\gamma_i^{(t+1)} = \left( \frac{1}{N} \sum_j |x_{ij}^{(t)}|^2 / a^*_i (\lambda I + A \Gamma A^*)^{-1} a_i \right)^{\frac{1}{2}} .
$$

(3.32)

These kinds of parametric empirical Bayesian (PEB) models are reported to be very powerful in EEG and MEG inverse problem. As mentioned before, this is a special instance of Bayesian analysis in which spatial priors are estimated from the data. PEB models are hierarchical linear models. This hierarchical form enables, for example, hyperparameters to constrain the parameters and therefore act as empirical priors.

Here, M-SBL exploits the inverse gamma ($1/\gamma_i$) hyperpriors in maximization scheme. This approach gives better results in source localization compared to the previous minimum norm estimators. However, they suffer from the increasing computational loads depending upon the increase in the number of sources, (i.e. the number of hyperparameters).

3.2.3 Parametric Empirical Bayesian Model using Multiple Priors

As mentioned in the start of this section, the classical regularization approaches can be seen as MAP estimation in Bayesian context with a single prior. Under the Gaussian assumptions, solving the cost function of WMN directly relates to the Bayesian estimate of the source posterior density $p(X|Y)$. In this aspect, the constraints can be easily expressed in terms of prior source covariance such that $C_p = (AW_p^*W_p)^{-1}$ where $X \sim N(0, C_p)$ [27]. Note that if some locations of the $C_p$ becomes small, then its dipole amplitude is more likely to shrink. For MN, the prior covariance is an identity matrix, $C_p = I_p$. Such smoothness prior considers that the sources be equally likely expressed everywhere. Therefore, it is easily seen that MN constraint is highly improbable in practice.

SBL and M-SBL are the examples of the class exploiting the multiple inverse gamma hyperpriors. Each hyperparameter represents the precision of the source locations which act as ARD priors[26]. This is one of the reason that SBL and MSBL give the better results compared to the others.

Recently, there has been more advanced approach to exploit this empirical Bayesian by automatically selecting the multiple priors with restricted maximization likelihood (ReML) estimates [32]. Instead of
using a small number of carefully specified prior covariance components, Friston et al suggested to use an empirically determined mixture of prior covariance components. In contrast to SBL and M-SBL, this approach uses covariance rather than precision hyperparameters.

Using prior covariance gives two advantages. First, the fixed-form variational process, which is used in SBL and M-SBL estimation, changes to very simple and efficient classical covariance component estimation based on ReML. Second, this will avoid the improper densities associated with noninformative ARD priors.

Let us briefly introduce the parametric empirical Bayesian (PEB) model scheme to explain how this approach runs. Here, the linear matrix equation is given:

\[ Y = AX + E \]  \hspace{1cm} (3.33)

where \( Y \in \mathbb{R}^{m \times N} \), \( A \in \mathbb{R}^{m \times p} \), \( X \in \mathbb{R}^{p \times N} \) and \( p \) dipoles with fixed position and orientation.

In PEB, this can be changed in two-level hierarchical model:

\[ Y = AX + E_1 \]  \hspace{1cm} (3.34)
\[ X = 0 + E_2 \]  \hspace{1cm} (3.35)

where \( E_1 \sim N(0, C_{\text{noise}}) \), \( E_2 \sim N(0, C_p) \). The matrices \( C_{\text{noise}} \) and \( C_p \) are modeled as the linear combination of covariance components:

\[ C_{\text{noise}} = \mu_1 Q_{\text{noise}}^1 + \mu_2 Q_{\text{noise}}^2 + \cdots \]
\[ C_p = \lambda_1 Q_p^1 + \lambda_2 Q_p^2 + \cdots \]  \hspace{1cm} (3.36)

where \( \mu \in \{\mu_1, \mu_2, \cdots\} \) and \( \lambda \in \{\lambda_1, \lambda_2, \cdots\} \) are the unknown hyperparameters and \( C_p \) behaves as a prior covariance.

The basic idea of this approach is that any combinations of this various prior covariance components can be optimized and compared by their evidence using ReML estimates not a standard ML [32]. By achieving the model in this way, it obviates the needs to use priors in specific form and the approach automatically selects the model which explains the measurements at best.

Furthermore, this model selection takes several advantage of ReML. In PEB, the hyperparameters are estimated by using EM. After \( C_{\text{noise}} \) and \( C_p \) is estimated, the solution \( X \) is given by its matching MAP estimator. In this process, ReML is used and accounts for the loss of degrees of freedom caused by conditional uncertainty about the parameters (see the Appendix B in [27] for its objective function). Moreover, an efficient optimization is available by using ReML because the size of covariance of the data used \((m, \text{channel space})\) does not increase with the number of sources \((p, \text{source space})\).

Applying this concept may give a new area to research for improvements in our studies. However, this is not the main in this paper and we will postpone and continue the study about this subject in our future works.
Chapter 4. Subspace Penalized Sparse Learning - A Review

In this section, we first seek for the role of non-separable penalty in M-SBL. Secondly, we discuss a novel subspace sparse learning (SPL) algorithm that generalizes M-SBL by exploiting the subspace geometry in MMV. The main contribution of this paper is that we have adopted the proposed algorithm in EEG and MEG source localization problem and compared the proposed algorithm to the conventional ones under various conditions.

4.1 Role of Non-separable Penalty in M-SBL

When compared to the standard sparse recovery framework, Wipf et al [13] showed that M-SBL cost function in $x$-space has an additional log $|·|$ term which is a nondecreasing, concave function in $\gamma$. It is well-known that the concave, nondecreasing functions of the coefficient magnitudes favor sparse solutions so that it works as a sparsity inducing term [33]. This non-separable penalty is so powerful that the M-SBL cost function provides a tighter approximation than $||x||_1$ while producing much fewer local minima than when using $||x||_0$ directly. They also showed that, in certain settings, this class of penalty is always superior than the separable regularization term [13].

In order to develop a new joint sparse recovery algorithm that improves on M-SBL, we need to understand the roles of the regularization term in M-SBL. We revealed that there is another important role for this additional regularization term. It turns out that log $|\lambda I + A^\Gamma A^\ast|$ works as a good (non-convex) proxy for rank$(A^\Gamma A^\ast)$ [34]. This can be shown using the following argument originally developed by Mohan and Fazel [34]. Define the smooth Schatten-$p$ function for a matrix $Z \in \mathbb{R}^{m \times n}$ as

$$f_p(Z) = \text{Tr}(Z^\ast Z + \lambda I)^{p/2},$$

where $\sigma_i(Z)$ denotes the $i$-th largest singular values of $Z$. Note that $f_p(Z)$ is differentiable for $p > 0$ and convex for $p \geq 1$; and if $p = 1$ and $\lambda = 0$, then $f_1(Z) = \|Z\|_*$, which is also known as the nuclear norm of the matrix $Z$. Note that for $\lambda = 0$ and $p \to 0$, $f_p(Z) \to \text{rank}(Z)$ [34]. Furthermore, for any positive scalar $x$, as $\lim_{p \to 0} \frac{1}{p} (x^p - 1) = \log x$, we have

$$\lim_{p \to 0} \frac{f_p(Z) - n}{p} = \sum_{i=1}^{n} \frac{1}{2} \log(\sigma_i^2(Z) + \lambda)$$

$$= \frac{1}{2} \log |Z^\ast Z + \lambda I|.$$

It follows that for any matrix $Z$ and for sufficiently small $\lambda$, the quantity log $|Z^\ast Z + \lambda I|$ is a proxy for $\text{rank}(Z)$. To use this observation, we rewrite

$$\log |\lambda I + A^\Gamma A^\ast| = \log |\lambda I + \Gamma^\frac{1}{2} A^\ast A \Gamma^\frac{1}{2}| + (m - n) \log \lambda,$$

where we use the determinant property $|C + DE| = |C||I + EC^{-1}D|$ for an invertible matrix $C$. Therefore, up to the constant of $(m - n) \log \lambda$, we know that log $|\lambda I + A^\Gamma A^\ast|$ is a proxy for the rank$(A^\Gamma A^\ast)$. While Wipf
et al [13] showed that the superior performance of M-SBL is owing to the non-separability of the penalty $\log |A + A^\dagger A^*|$ in terms of $\gamma$, which can avoid many local minimizers, we now add another interpretation that the reason of M-SBL good performance over the conventional sparse recovery algorithms is due to the use of rank penalty.

However, one of the interesting observations is that in M-SBL the rank penalty term does not change its form regardless of the number of snapshots. This appears mysterious, and this paper shows that the penalty term can be indeed modified using the spark reduction property of MMV problem [35].

### 4.2 The SPL Penalty

In order to develop a new joint sparse recovery algorithm that improves M-SBL, we need to understand the roles of the regularization term in M-SBL. It turns out that $\log |A + A^\dagger A^*|$ works a good (non-convex) proxy for $\text{rank}(A^\dagger A)$, and the penalty imposes a sparsity in the selected sub-matrix. However, the following theorem shows that rather than minimizing the $\text{rank}(A^\dagger A)$, there exists another rank penalty that is more suitable for minimizing in the context of MMV.

**Theorem 4.2.1** Suppose that we are given a noiseless observation matrix $Y \in \mathbb{R}^{m \times N}$, and let $R(Q) = R(Q)$. Assume that $A \in \mathbb{R}^{m \times n}$, $X_\star \in \mathbb{R}^{n \times r}$, $Y \in \mathbb{R}^{m \times r}$ satisfy $AX_\star = Y$ where $\|X_\star\|_0 = k$ and the columns of $Y$ are linearly independent and $r = \text{rank}(Y)$. If $A$ satisfies a restricted isometry property (RIP) condition $0 \leq \delta_{2k-r+1}(A) < 1$, then for noiseless measurement we have

$$k - r = \min_{I \subseteq \mathbb{N}} \text{rank}(Q^\dagger A_I),$$

and the corresponding support set $I_\star$ is equivalent to $\text{supp}X_\star$.

Theorem 4.2.1 indicates that by minimizing a new rank prior, we can obtain the true sparse support. Hence, we propose the following penalty to replace $g_{\text{msbl}}(X)$ in (3.28):

$$g_{\text{SPL,} \epsilon}(X) \equiv \min_{\gamma \geq 0} G_{\text{SPL,} \epsilon}(\gamma, X)$$

where

$$G_{\text{SPL,} \epsilon}(\gamma, X) = \text{Tr} \left( X^\dagger \Gamma^{-1} X \right) + N \cdot \text{Rprox} \left( Q^\dagger A^\dagger A^\dagger Q + \epsilon I \right)$$

where $\text{Rprox} \left( Q^\dagger A^\dagger A^\dagger \right)$ denotes a proxy for $\text{rank}(Q^\dagger A^\dagger A^\dagger)$. Note that minimizing $\text{Rprox} \left( Q^\dagger A^\dagger A^\dagger \right)$ with respect to $\gamma$ is equivalent to find the index set $I$ that minimizes $\text{Rprox} \left( Q^\dagger A_I \right)$.

Therefore, if $|\cdot|$ proxy is used, we have

$$G_{\text{SPL,} \epsilon}(\gamma, X) = \text{Tr} \left( X^\dagger \Gamma^{-1} X \right) + N \log |Q^\dagger A^\dagger A^\dagger Q + \epsilon I|$$

Using the proposed SPL penalty, we formulate the following noiseless SPL minimization problem:

$$\min_X g_{\text{SPL,} \epsilon}(X), \quad \text{subject to } Y = AX.$$  \hspace{1cm} (4.5)

For noisy measurement, we have

$$\min_X \|Y - AX\|_F^2 + \lambda g_{\text{SPL,} \epsilon}(X).$$  \hspace{1cm} (4.6)

Using (4.6), a noisy SPL formulation can be written as

$$\min_{X, \gamma \geq 0, \Psi \in S_+^{m \times r}} C(X, \gamma, \Psi)$$  \hspace{1cm} (4.7)
where the augmented cost function is given by
\[
C(X, \gamma, \Psi) = \|Y - AX\|^2_F + \text{Tr} \left( X^*\Gamma^{-1}X \right) + N\text{Tr} \left( (Q^*A\Gamma A^*Q + \epsilon I)\Psi \right) - N \log |\Psi| .
\] (4.8)

Due to the non-negativity constraint for \(\gamma\), a critical solution should satisfy the following first order Karush-Kuhn-Tucker (KKT) necessary conditions [36]:
\[
\frac{\partial C(X, \gamma, \Psi)}{\partial X} = -2A^*(Y - AX) + \Gamma^{-1}X = 0 \quad (4.9)
\]
\[
\frac{\partial C(X, \gamma, \Psi)}{\partial \Psi} = N(Q^*A\Gamma A^*Q + \epsilon I) - N\Psi^{-1} = 0 \quad (4.10)
\]
\[
\frac{\partial C(X, \gamma, \Psi)}{\partial \gamma_i} - \lambda_i = -\sum_j |x_{ij}^{(t)}|^2 \gamma_i^2 + N\gamma_i^*Q\Psi^{(t)}a_i - \lambda_i = 0, \quad \forall i \quad (4.11)
\]
\[
\lambda_i \gamma_i = 0, \quad \lambda_i \geq 0, \quad \gamma_i \geq 0, \quad \forall i \quad (4.12)
\]

This leads us to the following fixed point iterations:

**Minimization with respect to \(X\)**

For a given estimate \(\gamma^{(t)}\), (4.9) leads us to a close form solution for \(X^{(t+1)}\):
\[
X^{(t+1)} = \Gamma^{(t)}A^* (\lambda I + A\Gamma^{(t)} A^*)^{-1}Y, \quad \Gamma^{(t)} = \text{diag}(\gamma^{(t)}).
\]

**Determination of \(\Psi\)**

For a given estimate \(\gamma^{(t)}\), using (4.10), we can find a close form solution for \(\Psi^{(t)}\):
\[
\Psi^{(t)} = (Q^*A\Gamma^{(t)} A^*Q + \epsilon I)^{-1}.
\]

**Estimation of \(\gamma\)**

For a given \(\Psi^{(t)}\) and \(X^{(t)}\), using Eqs. (4.11) and (4.12), we have
\[
\gamma_i \frac{\partial C(X^{(t)}, \gamma, \Psi^{(t)})}{\partial \gamma_i} = \gamma_i \left( -\sum_j |x_{ij}^{(t)}|^2 \gamma_i^2 + N\gamma_i^*Q\Psi^{(t)}a_i \right) = 0 .
\] (4.13)

Here, if \(a_i^*Q\Psi^{(t)}Q^*a_i \neq 0\) and \(\sum_j |x_{ij}^{(t)}|^2 = 0\), (4.13) gives us \(\gamma_i = 0\). In general, \(a_i^*Q\Psi^{(t)}Q^*a_i \neq 0\), we have the following update equation:
\[
\gamma_i^{(t)} = \left( \frac{\frac{1}{N} \sum_j |x_{ij}^{(t)}|^2}{a_i^*Q\Psi^{(t)} Q^*a_i} \right)^\frac{1}{2} = \left( \frac{\frac{1}{N} \sum_j |x_{ij}^{(t)}|^2}{a_i^*Q(Q^*A\Gamma^{(t)} A^*Q + \epsilon I)^{-1}Q^*a_i} \right)^\frac{1}{2} ,
\] (4.14)

The new algorithm was named *subspace-penalized sparse learning (SPL)* by excluding term “Bayesian” because the resulting new algorithm is no more Bayesian due to the deterministic penalty based on geometric argument.

### 4.3 Theoretical Result on Local Minimizers

In [13], Wipf provided a condition to remove all local minimizers for the case of SBL. The authors in [14] have extended the results and showed that the conditions for removing local minimizers are more favorable for the proposed method than for M-SBL.
Theorem 4.3.1 Let $X_*$ denote a maximally sparse solution to be $Y = AX_*$ with $\|X_*\|_0 = k$. Let $A$ satisfy the RIP condition with $0 \leq \delta_{2k-r+1}(A) < 1$, where $r = \text{rank}(Y)$. Suppose $X$ represent a coefficient such that $S = \text{supp}X$ and $k < |S| = p \leq m$, and $X^S = A^S_Y$. Let $\Psi = (Q^*A^*A + \epsilon I)^{-1}$ and the subscript $\setminus i$ denotes the corresponding matrix by removing the $i$-th columns of the matrix $A$. Then, the following statements are true:

1. For some $j \notin S$, if we have
   \[ v^*_{S,j} (\bar{W}_{S,j} \bar{R}_S \bar{W}_{S,j}) v_{S,j} > 1, \]  
   (4.15)
   where $v_{S,j} = A^S a_j$, $\bar{R}_S = [\bar{r}_{i'i'}]_{i,i' \in S}$ and $\bar{W}_{S,j} = \text{diag}([\bar{w}_i]_{i \in S})$ such that
   \[ \bar{r}_{i'i'} = \frac{x^i(x_i')^*}{\|x^i\| \|x_i'\|}, \quad \bar{w}_i = \frac{a_i^* Q \Psi \setminus Q^* a_i}{a_j^* Q \Psi Q^* a_j}, \]  
   (4.16)
   for $i,i' \in S$ and $j \notin S$, then $X$ is not a local minimizer.

2. In particular, if $\text{rank}(Q^*A_S) \leq m-r$ and $|S \cap \text{supp}X_*| \geq k-r$, then there always exists $\epsilon_1 > 0$ such that for any $0 < \epsilon < \epsilon_1$, $X$ is not a local minimizer.

The new local minimizer analysis clearly shows why the proposed method is better than M-SBL. More specifically, as long as $k-r$ correct supports are included in a local minimizer, the proposed method can escape from the local minimizer. In addition, the proposed method is more robust to the condition number of the unknown signal $X$ compared to M-SBL. More specifically, in (4.15), $\text{rank}(\bar{R}_S) \leq k$, so there always exist null spaces when the rank of $\bar{R}_S$ is not full. In particular, in noisy cases, the corresponding numerical rank can be reduced when the condition number of $\bar{R}_S$ is bad, which can happen when the rows of $X^S$ are highly correlated. In this case, there are more chances that $\bar{W}_{S,j} v_{S,j}$ can fall into the null space so that the condition (4.15) cannot be met. However, in the proposed method, we can still prevent such situation as long as $k-r$ correct support are included in $X^S$. 
Chapter 5. Implementation

5.1 Preconditioning

To compare the performance with and without preconditioning process, under both situations we calculated the success rates of our proposed method only.

In EEG and MEG, the RIP condition of the leadfield matrix $A$ are bad. Since the performance of many algorithms depend on the RIP of the dictionary, high correlation between the column vectors of leadfield matrix significantly reduces the recovery performance [37, 38, 39]. To relax the problem, we let a preconditioning weighting matrix $P \in \mathbb{R}^{m \times m}$ produce the following sensing equation [40]:

$$PY = PAX + PE.$$  \hfill (5.1)

The RIP constant is known to be affected by preconditioning so that it is possible to precondition the dictionary $PA$ having a better RIP constant. However, it is combinatorial optimization problem to optimize $P$ having the best one. Thus, we approximate the optimization problem as the minimization of Frobenius norm resulting in the optimal preconditioner [40]

$$P = (AA^*)^{-1/2}.$$  \hfill (5.2)

Consequently, the correlation of the signal after the optimal preconditioning becomes

$$\begin{align*}
(PA)^*PY &= A^*P^*PY = A^*(AA^*)^{-1}Y = A^TY.
\end{align*}$$  \hfill (5.3)

Due to ill-posedness of leadfield matrix $A$, the inversion in (5.2) may become very unstable. We can treat this problem by including a regularization parameter $\epsilon > 0$ in the inversion

$$P = (AA^* + \epsilon I)^{-1/2}.$$  \hfill (5.4)

In our experiments, however, $\epsilon = 0$ seems to be sufficient, so all the results from preconditioned SPL algorithm are conducted under the condition.

5.2 Normalization of leadfield matrix

Before employing the algorithms, entire columns of $A$ are normalized in Euclidean norm sense, $\|a_i\|_2 = 1$ where $i \in \{1, \cdots, n\}$. However, preconditioned version of SPL especially has one more step before the normalization so that its normalization is done on the preconditioned matrix $PA$

$$\begin{align*}
Y &= (AW)W^{-1}X + E \\
    &= \tilde{A}\tilde{X} + E.
\end{align*}$$  \hfill (5.5)

where $W$ is the diagonal matrix for normalization, $\tilde{A}$ is $AW$ and $\tilde{X}$ is $W^{-1}X$. After $\tilde{X}$ is recovered, we denormalize it to $X$ by multiplying $W$ to get a final result.
5.3 Extension to Dipole Moment

For easier understanding, so far we have explained the proposed algorithm under a simplified assumption that the lead field matrix is $m \times n$ assuming that the source is a scalar. However, equivalent derivation can be easily obtained for $m \times 3n$ cases with considering block structure for dipole moments. In neuroelectromagnetic imaging, a single dipole source is represented as a three-dimensional vector in Cartesian coordinates $(x, y, z)$ rather than a scalar. Thus, the problem formulation can be made with a block structure so that the leadfield matrix $A$ and signal matrix $X$ can be interpreted as

$$A = \begin{bmatrix} A_1, A_2, \cdots, A_n \end{bmatrix}, \quad X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$$ (5.6)

where $A_i \in \mathbb{R}^{m \times 3}$, $X_i \in \mathbb{R}^{3 \times N}$, and $i \in \{1, \cdots, n\}$. Therefore, it is reasonable to have same $\gamma$ per every three elements of a dipole so that $\Gamma$ is stated as

$$\Gamma = \begin{bmatrix} \gamma_1 I_3 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \gamma_n I_3 \end{bmatrix}$$ (5.7)

where $I_3 \in \mathbb{R}^{3 \times 3}$.

This in turn will change the cost function of MUSIC, sLORETA solution, $\gamma$ update rule of both M-SBL, and our proposed method as follows.

**MUSIC cost function** :

Let SVD of each $A_i$ be $svd(A_i) = U_i \Sigma_i V_i^*$ and the minimum eigenvalue of $X$ be $\sigma_{\text{min}}(X)$. Then,

$$L_{\text{MUSIC}}(i) = \frac{1}{\sigma_{\text{min}}(U_i^* U_n U_i)}.$$ (5.8)

It is changed from comparing the scalar value of subspace correlation to comparing the minimum eigenvalues between the subspace correlation matrices. Eigenvalue can be interpreted as a magnitude of the matrix and the cost function can be also changed to comparing the maximums between the matrices. Here, however, we compared the minimum eigenvalues.

**sLORETA solution** :

The original sLORETA solution is

$$\hat{X} = A^* A + \lambda I)^{-1} Y = RX$$

where $\hat{X} \in \mathbb{R}^{3n \times N}$. Every three rows of the sLORETA solution represent the value of one dipole in $\{x, y, z\}$ directions. To render the image on the $n$ voxels, we sum up every three rows of the sLORETA $\hat{X}_i \in \mathbb{R}^{3 \times N}$ to one row, $\hat{x} \in \mathbb{R}^{1 \times N}$.

**M-SBL $\gamma$ update rule** :

The changed M-SBL $\gamma$ update rule is

$$\gamma_i^{(t+1)} = \left( \frac{1}{N} \text{Tr}(X_i^* X_i) \right)^{\frac{1}{2}} \left( \text{Tr}(A_i^*(\lambda I + A_i^* A_i)^{-1} A_i) \right)^{\frac{1}{2}}, \quad \Gamma_i = \gamma_i I_3$$ (5.9)
where $\Gamma_i \in \mathbb{R}^{3 \times 3}$ is an $i$th diagonal element matrix in $\Gamma$. Because of using a matrix $A_i$, not a vector $a_i$, the term inside the trace $\text{Tr}\{\cdot\}$ becomes a matrix not a scalar value. Therefore, the cost function and resulted update rule have changed from exploiting a scalar value to summing up every three values.

The iteration finishes if the change between the current and previous numerator of the (5.9) is under the stopping criterion of $0.003$

$$\frac{\|\text{numer}^{(t-1)} - \text{numer}^{(t)}\|_2}{\|\text{numer}^{(t)}\|_2} < 0.003$$

(5.10)

where $\text{numer}^{(t)}$ is the value of numeration $\text{Tr}(X^*X)$ in $t$-th iteration.

**SPL $\gamma$ update rule**:

As same as the change in M-SBL, the update rule of SPL has changed as follows:

$$\gamma_i^{(t)} = \left(\frac{1}{N} \frac{\text{Tr}(X_i^*X_i)}{\text{Tr}(A_i^*Q(Q^*A_i^tA_i^*Q + \epsilon I)^{-1}Q^*A_i)}\right)^{\frac{1}{2}}, \quad \Gamma_i = \gamma_i I_3.$$  

(5.11)

The stopping criterion of SPL is exactly same as M-SBL. From our experience, despite the converging iteration number changes dependent on SNR and RIP condition of A matrix, usually 100-300 iterations are required for SPL to converge with our leadfield matrices used in the experiments.
Chapter 6. Experimental Result

6.1 Simulation Results

In our simulation, a leadfield matrix $A \in \mathbb{R}^{m \times n}$ is generated by Brainstorm [41]. An input signal $X \in \mathbb{R}^{n \times N}$ with $\text{rank}(X) = r \leq k$ is randomly generated where $k$ is the dipole source number. After generating noiseless data, we add zero mean white Gaussian noise to have measurements with various signal to noise ratio from 25dB to 40dB, $\text{SNR} \in \{25\text{dB}, 30\text{dB}, 35\text{dB}, 40\text{dB}\}$. The static simulation parameters are as follows: $m = 151, n = 24220$, and $N = 375$. Throughout the simulation, we compare the four algorithms, MUSIC, sLORETA, M-SBL, and SPL (proposed method) under both correlated and uncorrelated signal conditions.

The uncorrelated input signals are generated differently if their originated dipole sources are different. This means that the signals from the different dipole sources are showing different traits in temporal direction. However, even the signal is uncorrelated, one dipole source will give the same signal for all three coordinate directions, $\{x, y, z\}$. Conversely, the correlated input signals mean that all signals are showing the exactly same traits in temporal direction.

For example, there are five dipole sources, $k = 5$, whose locations are randomly generated. If we have five uncorrelated dipole sources, then the $\text{rank}(X) = r = k = 5$. The success rates are calculated 30 times and averaged where the maximum learning iteration number per each calculation is fixed to 10000 for M-SBL, and SPL. Though it varies under the conditions, our proposed algorithm SPL usually converges in 100-300 iterations under the given stopping criterion noticed in (5.10).

We counted the estimation as success: (1) when the estimated location in the candidate pool exactly recovers the true source location and (2) when the estimated location in the candidate pool is near the true source location within its Euclidean distance of 1cm. The candidate pool is composed of $2k$ numbers of the estimated source locations which have the highest spectrum value among all locations. The location with low or zero value means that it is unlikely to have a source input in the position. For example, after we have a spectrum of estimated locations, we sort the location spectrum in a descendent order, from top to bottom. Finally, if the true source number is $k = 5$, we select $2k = 10$ locations from the ordered estimated locations to make the candidate pool. The coordinate information of the brain is gathered from the anatomical MRI coordinates dataset in Brainstorm.

Success Rate Comparison

Here, we first compare the average success rates of algorithms under the various SNR and signal traits (correlated, uncorrelated). For fair comparison, given the most favorable condition, we assume that we know the exact number of dipole sources, $k = 5$. Therefore, even in the case that the input signals are perfectly correlated ($r = 1$), we give the right rank ($r = 5$) and this is used in MUSIC and SPL algorithms.

In figure 6.1 and figure 6.2, we can compare the average success rate of our algorithm with that of the others. Under the condition of uncorrelated sources, in figure 6.1, MUSIC shows almost perfect recovery of the true source location regardless of SNR variations. However, when the condition changes
to the correlated signal, in figure 6.2, the performance of MUSIC significantly reduces so that it even
fails to locate the true sources one among five. Besides, the performance does not be recovered even with
the increasing SNR. This is already predicted in the theoretical analysis.

The other algorithm, sLORETA recovers only one among five sources in average. Here, sLORETA
seem not to depend on the signal traits but their success rates are significantly inferior compared to the
our proposed method.

Meanwhile, our proposed method shows little dependency upon the signal traits. The success rate
has slightly been reduced in the correlated signal case but still showing a steady performance. M-SBL
shows the similar tendency but the performance of our proposed method beat that of M-SBL in most of
conditions.

Figure 6.1: Success rate comparison with varying with SNR(dB)∈ {25, 30, 25, 40}: five uncorrelated
signal sources (k = 5) given true rank (r = 5)

Rank Selection Effect

Figure 6.3 and figure 6.4 illustrate the effect of wrongly selected number of sources, namely the rank
of the signal X. Here, we intentionally selected an incorrect rank for the measurement matrix. For the
case of underspecifying the rank, given rank becomes r = 3 and for the case of overspecifying the rank,
given rank becomes r = 15 in both correlated and uncorrelated signal traits. We compared them with
the results given true rank. Note that the actual rank of correlated and uncorrelated signal case is r = 5,
not r = 3 nor r = 15. Here, since we have 5 input signals, we treat the true rank is 5 in the correlated
case also.

Comparing with the case when the exact rank are resolved, as expected in 3.1.2 before, underspec-
ifying the rank seriously spoils the performance of MUSIC. On the other hand, overspecifying the rank
does not show an evident impairment to the results. In figure 6.3, MUSIC shows a bad success rate
even though the input signals are uncorrelated perfectly. This profile also holds when the signals are
correlated.

On the contrary, our proposed algorithm SPL hardly changes upon the error in rank selection. It
Figure 6.2: Success rate comparison with varying with SNR(dB)∈{25, 30, 25, 40}: five correlated signals (k = 5) given true rank (r = 5) shows a robust recovery in both correlated and uncorrelated signal. Recall that MUSIC and SPL share similar parametric estimation steps:

\[ \text{MUSIC} : \quad \mathcal{L}_{\text{MUSIC}}(\hat{\theta}) = \frac{1}{\sigma_{\text{min}}(U_i^*U_nU_n^*U_i)} \]

\[ \text{SPL} : \quad \gamma_i^{(t)} = \left( \frac{\frac{1}{N} \text{Tr}(X_i^*X_i)}{\text{Tr}(A_i^*Q(A_i^*Q + \epsilon I)^{-1}Q^*A_i)} \right)^{\frac{1}{2}}, \quad \Gamma_i = \gamma_i I_3 \]

Note that the sparsity penalty term of SPL, \( \frac{1}{N} \text{Tr}(X_i^*X_i) \), compensates the drawbacks even though our proposed method has the MUSIC criterion, \( A_i^*Q \). Here, both \( U_n^* \) term in MUSIC cost function and \( Q \) term in SPL cost function are actually working as a same role. They are both comprise of the unit vectors spanning an estimated noise subspace:

\[ \text{rank}(X) = r, \quad \text{svd}(Y) = U\Sigma V^*, \quad U_n = U(r + 1 : \text{end}, :) \] (6.1)

where \( Y = AX + E \).

For the fair comparison, we let \( U_n \) and \( Q \) be generated by same process as (6.1) where the rank \( r \) is not given in practice so it is carefully selected by examining the eigenvalue, \( \Sigma \). However, our purpose is to compare the effect of incorrect rank selection on both algorithms so we decided the appropriate rank in advance as stated before.

From figure 6.5 to figure 6.8, the results are shown by images. In the uncorrelated input signal case, it is clearly seen that the MUSIC result becomes impaired due to the rank deficiency. Meanwhile, our proposed algorithm shows only a small change in intensity. In the correlated input signal case, MUSIC fails to separate the sources but giving the position in-between. However, our proposed algorithm resolves five positions with some artifacts under which is not shown here. We added an arrow in the results shown in the figure 6.8 to guide another hidden position which is hard to see. Though it shows 3-5 artifacts under, many of their intensities are relatively smaller than the ones in the front which represent the true sources.
Figure 6.3: Success rate comparison with varying SNR(dB) $\in \{25, 30, 25, 40\}$: five uncorrelated signal sources ($k = 5$) given wrong rank ($r = 3, r = 15$).

Figure 6.4: Success rate comparison with varying SNR(dB) $\in \{25, 30, 25, 40\}$: five correlated signal sources ($k = 5$) given wrong rank ($r = 3, r = 15$).
Figure 6.5: MUSIC image of five uncorrelated signal sources \((k = 5)\) with SNR 40dB: (a) Ground truth, (b) True rank \((r = 5)\) (c) Wrong rank \((r = 3)\).

Figure 6.6: SPL image of five uncorrelated signal sources \((k = 5)\) with SNR 40dB, (a) Ground truth, (b) True rank \((r = 5)\) (c) Wrong rank \((r = 3)\).

Figure 6.7: MUSIC image of five correlated signal sources \((k = 5)\) with SNR 40dB: (a) Ground truth, (b) True rank \((r = 5)\) (c) Wrong rank \((r = 3)\).

Figure 6.8: SPL image of five correlated signal sources \((k = 5)\) with SNR 40dB, (a) Ground truth, (b) True rank \((r = 5)\) (c) Wrong rank \((r = 3)\).
Anti-correlation

To examine if there is a difference in performance between the correlated and anti-correlated signal, we calculated the success rates and compared them in figure 6.9. There are total 4 signals and each two of them is anti-correlated. There seems to be no performance difference between the correlated and anti-correlated signal.

![Graph showing success rate comparison with varying SNR](image)

Figure 6.9: Success rate comparison with varying with SNR(dB) ∈ {25, 30, 25, 40}: four anti correlated signals \((k = 4)\) given true rank \((r = 4)\)

Preconditioning Effect

In figure 6.10 and figure 6.11, the effect of preconditioning is discussed. Note that all previous results from our proposed algorithm are already exploiting the precondition of the leadfield matrix \(A\). Just for convenience, we dropped the term ‘preconditioned’ in the legends of graphs and notated simply ‘SPL’.

However, here we exclusively notate the ‘preconditioned’ SPL, abbreviated ‘SPL precond.’ and ‘non-preconditioned SPL’, abbreviated ‘SPL’ in the graphs. As it is clearly seen in the figure, obviously the preconditioned SPL shows a better results all over the conditions.
Figure 6.10: SPL success rate comparison with and without preconditioning the leadfield matrix $A$, varying SNR(dB)$\in\{25, 30, 25, 40\}$: five uncorrelated signal sources ($k = 5$) given true rank ($r = 5$).

Figure 6.11: SPL success rate comparison with and without preconditioning the leadfield matrix $A$, varying SNR(dB)$\in\{25, 30, 25, 40\}$: five correlated signal sources ($k = 5$) given true rank ($r = 5$).
6.2 Real Experiments

The real epilepsy data is gathered from five to eight years old Ronaldic epilepsy patient while the right hand seizure occurred. Its sampling rate is $600\,Hz$ and it is filtered through the $0.05 \sim 200\,Hz$ bandpass filter. Sensors are located under the international 10-20 system. Leadfield matrix, $A \in \mathbb{R}^{m \times 3n}$, is obtained by CURRY 7.0 (CURRY, Compumedics/Neuroscan El Paso, TX, USA) and the measurement data matrix is $Y \in \mathbb{R}^{m \times N}$ where the experimental parameters were $m = 23$, $n = 24220$, and $N = 200$. The temporal measurements are shown in figure 6.12. You can see the firing signals which seem to be caused by the seizure.

![Figure 6.12: Epileptic seizure wave.](image)

We performed ICA (independent component analysis) on the measurement data and removed the components which seem not to affect the real signals. This is done by using EEGLAB [42]. There was 22 components and we removed eight components manually; i.e. (6, 7, 9, 13, 16, 19, 21, 22). This components are shown in figure 6.13.

![Figure 6.13: ICA results. There are total 22 components and we removed eight components manually; i.e. (6, 7, 9, 13, 16, 19, 21, 22).](image)
In our real data, the epileptic seizure is known to be induced in bilateral locations of the brain simultaneously. In figure 6.14, we showed the reconstructed results using sLORETA, MUSIC, M-SBL, and our proposed algorithm. It is clearly seen that our proposed method provides accurate and reasonable locations whereas all other methods fail. For example, sLORETA provided very blurry reconstruction. MUSIC shows a similar result with the simulation result under correlated signal which is shown in figure 6.7. MUSIC seems to suffer from the correlated signal trait of epileptic seizure. M-SBL identified only one source locations and moreover it also shows some artifacts in the other areas of the brain (results not shown). However, our proposed method accurately identified the correlated two source locations without any artifacts.

Figure 6.14: Real experimental results using (a) sLORETA, (b) MUSIC, (c) M-SBL (d) SPL (our proposed algorithm).

To see the temporal changes of the signal sources, we made a window of size 10 and showed the results of the windowed data while moving the window through the timeline. Figure 6.15 shows the results from 90 to 105. For example, the first figure is the results of 90 \( \sim \) 99 and the second figure is the results of 91 \( \sim \) 100. The results shows that the seizure inducing sources appear on the bilateral area of the brain simultaneously and moves around the central sulcus of the brain. These seem to be reasonable because it is known that Ronaldic epilepsy seizures start around the central sulcus of the brain.
Figure 6.15: Real experimental results using SPL (our proposed algorithm) with moving window of size 10. These are the results from 90 to 105 timepoints.
Chapter 7. Discussion

In Wipf et al [13], they demonstrated the fundamental x and γ space duality between both cost functions of M-SBL and standard sparse recovery framework. This duality offered an avenue for further improvements suggesting alternative means of building cost functions and algorithms. The key aspect of their paper is to figure out that the only difference between M-SBL and subspace-based hybrid greedy algorithms is an additional non-separable log| · | (log determinant) penalty which seems to induce a high performance. Based on the results, they suggested that if one searches the regularization terms beyond the separable ones, there might be a more suitable prior choice which will bring much higher performances though it does not neatly fit into current frameworks.

In present paper, we extended and improved their works to analyze the origin of M-SBL’s high performance furthermore as well as its limitations. We revealed that the role of the non-separable penalty term in M-SBL, i.e. log|AI + AΓA∗| indeed acts as a good rank proxy for rank(AΓ1/2). This naturally imposes another constraint on the solution space compared to the other methods. Finally, we demonstrated that there exists better rank penalty based on the geometric argument so that the proposed algorithm is able to exploit a deterministic penalty improving the performance. We then provided the theoretical reasons that the conditions for removing the local minimizers are more favorable for our proposed method than for M-SBL.

Success rate and the effect of rank selection were used in its performance evaluation and this evaluation led to the following conclusions. First, our proposed method tends to be powerful over both correlated and uncorrelated signal traits. Second, it is robust over the rank selection issues. It is worth to note that even though it shares the same criterion with MUSIC, another existing sparsity penalty seems to compensate the drawbacks.

Still, there are several remaining issues. One is its performance change which highly depends upon decreasing SNR. Our proposed algorithm shows performance impairment under the bad SNR while MUSIC shows good performance under some restricted conditions. Again, this is quite mysterious because they share the same criterion, A∗Q. The only two differences between their cost functions are the sparsity inducing component and Γ including component which implicitly affects the update rule. The relationship between MUSIC and our proposed algorithm can be investigated by studying trade-offs of changing each component while tracking down its EM steps. Changing the way of eigenvalue comparison process in (5.8) and (5.11) between σ_{min}{·} and Tr{·} also seems to change the performance.

The other issue is its increasing computational loads upon increasing number of head voxel locations. There might be two possible remedies for the problem. First is to prune the redundant hyperparameters during the process as M-SBL does. M-SBL already prunes away the hyperparameters which are seemingly unimportant and this lowers the computation loads as the iteration proceeds. However, when it comes to our proposed algorithm, this might cause giving up the strength over M-SBL.

In our proposed algorithm, it takes advantage of its cost function structure. It is easily seen in the case when a∗iQΨ(t)Q∗ai = 0. Under this situation, based on (4.13), we have the following two observations: 1) γ_i(t) → ∞ when \sum_j |x_{ij}(t)|^2 \neq 0; and 2) γ_i(t) can be arbitrary positive number C when \sum_j |x_{ij}(t)|^2 = 0 since the equality in (4.13) is satisfied regardless of the choice of C. Therefore, we define
the following $\gamma_i$ update:

$$\gamma_i^{(t)} = \begin{cases} 
\infty, & \text{if } a_i^*Q\Psi^{(t)}Q^*a_i = 0 \text{ and } \sum_j |x_{ij}^{(t)}|^2 \neq 0 \\
C > 0, & \text{if } a_i^*Q\Psi^{(t)}Q^*a_i = 0 \text{ and } \sum_j |x_{ij}^{(t)}|^2 = 0 \\
\left( \frac{\sum_j |x_{ij}^{(t)}|^2}{a_i^*Q\Psi^{(t)}Q^*a_i} \right)^\frac{1}{2}, & \text{if } a_i^*Q\Psi^{(t)}Q^*a_i \neq 0
\end{cases}$$

(7.1)

Thanks to (7.1), even though $\sum_j |x_{ij}^{(t)}|^2$ becomes erroneously zero during iterations, there is a chance to escape from such error if $a_i^*Q\Psi^{(t)}Q^*a_i = 0$. However, if we prune out the suspicious $\gamma$ or its resulting $X$ depending on the inappropriate criterion during the iteration, there is a possibility to disregard the chance to escape from the situation which is stated above.

Second, since our algorithm can be easily implemented using alternating minimization methods that have very similar structures as M-SBL, there is the other way is to settle the problem. By taking the concept of Friston et al [32], the optimization using ReML would get rid of such problem due to its characteristic using sample covariance of the data. PEB modeling would also open the possibility to improve the empirical prior updates. Having said these, it would be interesting to compare the results of each trial and resulting model for more complete analysis and improvements on our proposed algorithm.
Chapter 8. Conclusion

In this paper, we proposed a novel subspace penalized sparse learning algorithm by exploiting the rank reduction property discovered in subspace based joint sparse recovery algorithms. Compared to M-SBL, theoretical analysis showed that the proposed method avoids more local minimizers and less prone to correlated signal sources. Furthermore, the algorithm can be easily implemented using alternating minimization methods that have very similar structures as M-SBL. Experimental results demonstrated that our proposed algorithm improved the source recovery over the conventional methods such as MUSIC, minimum norm, and M-SBL algorithms under both a highly correlated and uncorrelated source conditions.
References


Summary

NEUROELECTROMAGNETIC IMAGING OF CORRELATED SOURCES USING A NOVEL SUBSPACE PENALIZED SPARSE LEARNING

EEG와 MEG로부터 얻은 신호를 바탕으로 뇌신호원을 국지화하는 문제는 활발히 연구되고 있는 분야이다. 현재 이와 같이 신호원을 찾는 문제를 풀기 위하여 MUSIC, M-SBL 등 다양한 방식의 접근이 존재한다. 이 알고리듬들은 다양한 임상사례에 적용되어왔으며 훌륭한 성능을 보여주고 있다. 그러나 미지의 신호원이 크게 동기화되어 있는 경우, 현존하는 알고리듬들은 다소 신뢰도가 떨어지는 복원 결과를 보여주며 이것이 기존 알고리듬들이 다양한 임상 환경에서 이용되지 못하는 큰 이유 중 하나이다. 이 논문에서는 다중벡터문제에서의 기저공간 기하구조를 이용하여 M-SBL을 일반화한 새로운 알고리듬을 통해 이러한 문제를 해결하고자 한다. 시뮬레이션과 실제 간질발작 신호에 대한 실험 결과들은 새로운 알고리듬이 동기화된 신호에 대해서 기존의 방법들보다 뛰어난 복원성능을 가지는 것을 보여준다.