# Successive Estimation Method of Locating Dipoles based on QR Decomposition using EEG Arrays

# Successive Estimation Method of Locating Dipoles based on QR Decomposition using EEG Arrays

By

## Yiming Wang

B.Sc, EE Dept, Shandong University

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Masters in Applied Science

Master of Applied Science (2007)                    McMaster University

Electrical and Computer Engineering                    Hamilton, Ontario

TITLE:     Successive Estimation Method of Locating Dipoles based on QR Decomposition using EEG Arrays

AUTHOR:   Yiming Wang, B. Sc. (Shandong University)

SUPERVISOR:   Dr. Kon Max Wong

NUMBER OF PAGES:   x, 60

**Dedications:**

*To my parents:*

*Bin Wang and Shuxin Han*

# Abstract

EEG is a noninvasive technique useful for the human brain mapping and for the estimation of neural electrical activities in human brain. A goal of processing EEG signals of a subject is the localization of neural current sources in human brain known as *dipoles*. Although this location estimation problem can be modeled as a particular kind of parameter estimation problem as in array signal processing, the nonlinear structure of an EEG electrode array, which is much more complicated than a traditional sensor array, makes the problem more difficult.

In this thesis, we formulate the inverse problem of the forward model on computing the scalp EEG at a finite set of sensors from multiple dipole sources. It is observed that the geometric structure of the EEG array plays a crucial role in ensuring a unique solution for this problem. We first present a necessary and sufficient condition in the model of a single rotating dipole, that guarantees its location to be uniquely determined, when the second-order statistic of the EEG observation is available. In addition, for a single rotating dipole, a closed-form solution to uniquely determine its position is obtained by exploiting the geometrical structure of the EEG array.

In the case of multiple dipoles, we suggest the use of the Maximum Likelihood (ML) estimator, which is often considered optimum in parameter estimation. We propose an efficient localization algorithm based on QR decomposition. Depending on whether or not the probability density functions of the dipole amplitude and the noise are available, we utilize the non-coherent ML or the LS as the criterion to develop a unified successive localization algorithm, so that solving the original multi-

dipole optimization problem can be approximated by successively solving a series of single-dipole optimization problems. Numerical simulations show that our methods have much smaller estimation errors than the existing RAP-MUSIC method under non-ideal situations such as low SNR with small number of EEG sensors.

# Acknowledgement

In the first place the author wishes to express her deep appreciation to her supervisor, Professor Kon Max Wong, for his expert supervision, guidance and encouragement throughout the course of this thesis, which have been inspiring to her growth as a student and have been illuminating the way of her independent research.

The author is very grateful to Dr. Jian-Kang Zhang for continuously providing helpful suggestions and insightful discussions, and to Dr. A. Jeremic for his constructive comments on the thesis. She would like to thank all her colleagues here in the Signal Processing Group at McMaster University, especially Qiuyuan Huang, Rong Chai, and Jing Liu, for their friendship and support during occasional stressful times.

She is also deeply indebted to her parents Bin Wang and Shuxinhan, her boyfriend Yangyang Li, and all other family members for their understanding, love and unceasing faith in her along the way.

# Contents

# List of Figures

# Notations

| | |
|---|---|
| $E\{\cdot\}$ | Expectation |
| $\text{tr}[\cdot]$ | Trace |
| $\log$ | Logarithm function with $e$ base |
| $\det$ | Determinant |
| $[\cdot]^T$ | Transpose of a matrix or a vector |
| $[\cdot]^{-1}$ | inverse of a matrix |
| $[\cdot]^{\dagger}$ | Pseudo inverse of a matrix |
| $p\,(\mathbf{f}|\mathbf{a})$ | Probability density function of $\mathbf{f}$ condition on $\mathbf{a}$ |
| $\otimes$ | Kronecker product |
| $\text{VEC}[\cdot]$ | Vectorization of a matrix |
| $\mathcal{N}(\mu, \mathbf{C})$ | Gaussian distribution with mean $\mu$ and covariance $\mathbf{C}$ |
| $[\mathbf{A}]_{kk}$ | the $k$-th diagonal entry of a matrix $\mathbf{A}$ |
| $\mathbf{I}_M$ | $M \times M$ identity matrix |
| $\arg\{\cdot\}$ | a scaler, vector or matrix which satisfies the conditions enclosed within |
| $\nabla$ | gradient |

Text Conventions:

Throughout this thesis, we use the following notation: Matrices and column vectors are denoted by uppercase boldface characters (e.g., $\mathbf{A}$) and lowercase boldface characters (e.g., $\mathbf{b}$), respectively.

# Chapter 1

# Introduction

## 1.1 Overview

EEG stands for electroencephalography which is a safe and quick way of recording the electrical activity of the brain. The simultaneous activity of thousands and thousands of neurons in the brain can be detected and recorded with electrodes placed at the scalp, which are usually small metal discs. EEG signal was first measured by Hans Berger in 1924, and has been used in the study of brain function and diseases for several decades. What we are mainly interested in, however, is the evoked po-
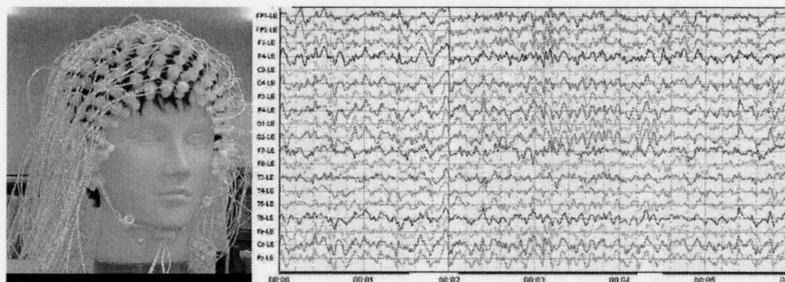


Figure 1.1: Multichannel EEG Electrodes and Recordings (Picture taken partly from [19] )

tential that can be reliably measured using EEG. Evoked potential, as distinct from

spontaneous potentials, refers to the brain response to an external or internal stimuli, such as visual, auditory or somatosensory stimulation. When a particular stimulus is applied, some corresponding brain areas are activated, inducing the evoked potential that is distinguishable from the background of ongoing EEG.

One major task of evoked potential or EEG analysis is thus to find out where these activated areas are within the human brain [33] [9]. In other words, our goal is to localize the "brain sources" that generate EEG signals. Such information is of great importance for both clinical and research applications [8][33]. For example, accurate location estimation of an epileptic focus in the brain can be used to plan surgery for its removal. Likewise, information about the brain regions that process a particular signal (such as an auditory tone) is of great value to brain functioning studies.

The so called brain sources of EEG are widely accepted as localized current sources in the cerebral cortex. In order to localize them, there are two types of source models generally used: the Equivalent Current Dipole (ECD) model [32] and the distributed source model [7]. The ECD model assumes small numbers of current dipoles to approximate the flow of electrical current in a small brain area, whereas the distributed source model assumes a large number of current dipoles scattered in a limited source space.

A dipole behaves like a vector, and generates electro-magnetic field that gives an equivalent description of the compound activity of all neuronal elements in their vicinity which are orientated in parallel to the dipole axis. Much of the literature uses ECD model to represent neural current sources, since it is simple to apply and many dipole fitting algorithms have been developed especially when the spatio-temporal EEG information is available. Therefore, we explore dipole localization methods based on this model. It is noteworthy that brain activity does not actually consist of discrete sets of physical current dipoles, but rather that the dipole is a convenient representation for coherent activation of a a large number of neural cells.

Another consideration in the localization problem is to construct a model of the
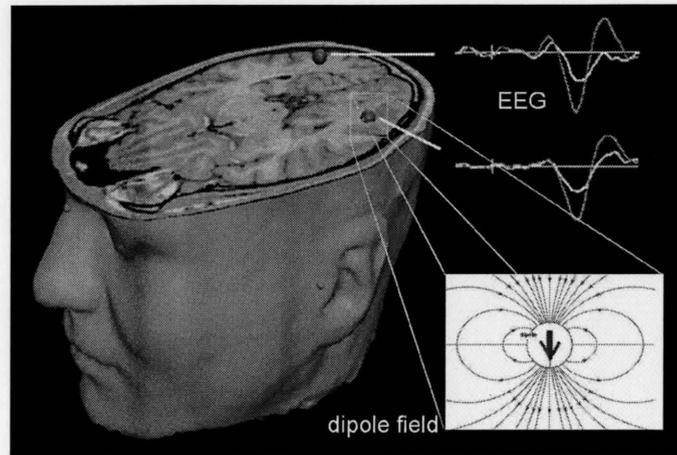
Figure 1.2: Equivalent Brain Sources (Pictures taken partly from [20] )

head. Apparently, the simplest model is a sphere, which contains concentric layers with different electrical conductivities, respectively the scalp, skull, cerebrospinal fluid and the brain. More sophisticated head models can be created using finite elements or boundary elements [6].

Once these models are selected, the location of the dipoles can be calculated by using nonlinear numerical optimization methods. Attempts to estimate ECDs based on EEG measurements dated back to Shaw in 1955 [35]. Scherg and Von Cramon [33] push forward the dipole fitting problem a big step, by introducing the spatio-temporal methods. Later, signal subspace-based methods were developed by Mosher, et al. [26] [27], which aroused great interests due to their ability to accurately locate multiple dipole sources. Meanwhile, Bayesian statistic methods were being employed [10] [18].

## 1.2   Motivation and Contribution of the Thesis

Although this location estimation problem can be modeled as a particular kind of parameter estimation problem as in array signal processing, the nonlinear structure

of an EEG electrode array, which is much more complicated than a traditional sensor array, makes the problem more difficult. For example, we cannot directly apply the uniqueness condition in determining the direction of arrival (DOA) by a linear array, into uniquely determining the location of the dipoles from multichannel EEG signals. Nevertheless, unique solutions to the localization problem do exist, at least in a mathematical sense, if we adopt a proper mapping model and adequately restrict the number of equivalent dipoles. Studies in [2] stated that two distinct dipoles could not generate the same potential map on a continuous set of scalp points. However, given an EEG array geometry, there has not been much study on the uniqueness of localizing the dipole.

In this thesis, we first consider the uniqueness of localizing a single dipole. We propose a geometric structure of an array with the minimum number of sensors to assure that a dipole can be uniquely determined based on the second order statistics. In addition, for a single dipole, a closed-form solution to uniquely determine its position and variances of its moment and noise is obtained by exploiting the geometrical structure of the EEG array.

Once a unique relationship between the scalp-recorded EEG and the dipoles is ensured, we put the emphasis of our research on the effective dipole localization algorithms. Due to the non-convexity of the dipole localization problem, in order to obtain a global solution, it has been suggested [37] [21] that the continuous location variables should be quantized, and then an exhaustive search performed over all grid points. *However, as the number of sources and the number of quantization level increase, the searching approach is computationally intractable.*

Other researchers proposed to apply signal subspace-based methods [26] [27], which under certain conditions, have demonstrated good and reliable accuracy in the localization of a restricted number of independent dipoles. However, the performances of these subspace methods are highly dependent on the sufficiency and underlying structure of observed EEG data. *If the EEG measurement interval is too*

*short, or if the EEG signals are recorded from only a small number of sensors, an accurate estimation of the signal subspace will fail, leading to unacceptable errors in localization.*

In this thesis, we suggest the use of the Maximum Likelihood (ML) estimator, which is often considered optimum in parameter estimation. We aim at developing a computational-efficient ML method for multiple dipole localization. By using the maximum likelihood estimator, either in stochastic signal model or in deterministic signal model, as a criterion, we formulate the dipole location estimation problem by the QR decomposition into an optimization problem, in which the objective function has the following two typical features:

1. The parameter matrix to be estimated can be segmented according to its columns into such a block matrix that, any two blocks are uncorrelated with respect to the parameters;

2. The original objective function can be decomposed into a sum of a series of sub-functions with the present sub-function only involving in the previous blocks.

These two properties enable us to reduce the original multi-source optimization problem into a series of single-source optimization problems so that the localization of all dipoles can be efficiently implemented. Furthermore, our method is model-independent, thus can be extended to a specific class of nonlinear optimization problems with the above two properties.

## 1.3   Organization of the Thesis

The thesis is structured as follows:

- In Chapter 2, the forward dipole-EEG mapping model is examined based on two categories of dipole descriptions (fixed-orientation and rotating orientation). Its inverse problem is then formulated; i.e., given the observations of EEG in an

array of finite sensors, we want to localize the dipoles. Several classical localization algorithms are introduced to solve the problem and their performance are evaluated.

- In Chapter 3, we investigate the uniqueness of dipole localization for a single dipole. Based on the rotating dipole model and the infinite homogeneous head model, we design a geometrical structure of an array with the minimum number of sensors to guarantee a unique localization of the dipole and present a closed-form solution for this localization problem.

- In Chapter 4, we propose a QR decomposition-based successive searching algorithm to efficiently implement the generalized likelihood ratio test. We also propose the non-coherent maximum likelihood estimator for dipole localization. The two algorithms are applicable to both dipole models.

- In Chapter 5, several numerical examples are provided. Given short snapshots of simulated EEG from a small number of EEG sensors, our methods have much smaller estimation errors than the other classical methods.

- In Chapter 6, conclusion on this work and suggestion for future work are presented.

# Chapter 2

# Background and Problem

# Statement

EEG-dipole model will be considered in detail in this chapter. We first introduce the dipole-to-EEG mapping principle, formulate the forward problem respectively according to two different dipole models, and then investigate the inverse problem, i.e., localization of dipoles from EEG observations, by specifying the assumptions throughout the thesis and formulating mathematically the problem. After that, we examine some popular dipole localization algorithms.

## 2.1 Forward Problem

In order to estimate the location of a dipole, it is necessary to assume a model of the dipole and a model of the head. The forward problem is thus the combination of both the dipole and head models, and to calculate the EEG produced by dipole sources when all the model parameters are known. We start with a single-dipole-single-snapshot case, and then extend to the case of multiple-dipole-spatiotemporal-EEG.

### 2.1.1   Physics of EEG

The scalp EEG signal is transmitted from the dipole sources and received from a finite set of electrodes on the surface of human head. Considering the dipole current flow is relatively slow in physical terms and the brain is a good conductor, the physics of EEG can be described by the quasi-static approximation of Maxwell's equations. The magnetic field $\mathcal{B}(\omega)$ at location $\omega$ can be related to the quasi-static divergence free current flow $\mathcal{J}(\ell')$ at location $\ell'$ through the well-known Biot-Savart law [4] [17]

$$\mathcal{B}(\omega) \ = \ \frac{\kappa}{4\pi} \int \mathcal{J}(\ell') \times \frac{\omega - \ell'}{\parallel \omega - \ell' \parallel^3} d\ell' \tag{2.1}$$

where $\kappa$ is the permittivity of free space. We can partition the total current density $\mathcal{J}(\ell')$ in the head volume into two current flows of distinct physiological significance:

$$\mathcal{J}(\ell') \ = \ \mathcal{J}^P(\ell') + \mathcal{J}^V(\ell') \tag{2.2}$$

with $\mathcal{J}^P(\ell')$ denoting a primary (driving) current flow related to the original neural activity and $\mathcal{J}^V(\ell')$ denoting a volume (passive) current flow that results from the effect of the electric field in the volume on extra-cellular charge carriers:

$$\mathcal{J}^V(\ell') \ = \ \zeta(\ell')\mathcal{E}(\ell') = -\zeta(\ell')\nabla\mathcal{F}(\ell') \tag{2.3}$$

where $\zeta(\ell')$ is the conductivity function of the head tissues, and $\mathcal{E}(\ell')$ is the electric filed, which equals the negative gradient of the electric potential $\mathcal{F}(\ell')$. Therefore, we can rewrite the Biot-Savart law above as a sum of contributions from the primary and volume current. After some tedious derivations [12] [5], a primary potential at location $\omega$ can be calculated as follows,

$$\mathcal{F}(\omega) \ = \ \frac{1}{4\pi\zeta} \int_{Head} \mathcal{J}^P(\ell') \cdot \frac{\omega - \ell'}{\parallel \omega - \ell' \parallel^3} d\ell' \tag{2.4}$$

Here, the primary potential $\mathcal{F}(\omega)$ is the solution for the infinite homogeneous medium of unit conductivity $\zeta$ due to the primary current $\mathcal{J}^P(\ell')$. Eq.(2.4) is the essential model for the forward problem for EEG.

If we assume that the primary current exists only at a specific point, i.e., the primary current source is an equivalent current dipole located at $\ell$ with moment $\mathbf{v}$, its primary current distribution can be represented as

$$\mathcal{J}^P(\ell') \;=\; \mathbf{v}\delta(\ell' - \ell) \tag{2.5}$$

where $\delta$ is the Dirac delta function. Then the primary potential for the dipole is reduced to a remarkably simple form

$$\mathcal{F}(\omega) \;=\; \frac{1}{4\pi\zeta}\frac{(\omega - \ell)}{\parallel \omega - \ell \parallel^3}\mathbf{v} \tag{2.6}$$

## 2.1.2   Single Dipole Model

Consider the EEG measurement taken at $M$ discrete locations on the head surface, from an array of $M$ sensors. Assume the EEG data is generated by a single dipole and corrupted by noise. Then the general forward model at the $t$th temporal snapshot can be described as

$$\mathbf{f}_t \;=\; \mathcal{G}(\ell)\mathbf{v}_t + \mathbf{e}_t \tag{2.7}$$

in which,

- $\mathbf{f}_t$ denotes an $M \times 1$ measurement collection vector from the $M$ sensors at the $t$th time slot;

- each dipole, as a vector, has two kinds of 3-dimensional parameters to describe its activity: location $\ell = [\ell_x, \ell_y, \ell_z]^T$ and moment $\mathbf{v}_t = [v_x(t), v_y(t), v_z(t)]^T$, representing the strength of neural signals in the three spatial directions in the reference coordinates. $\ell$ is supposed to be fixed during the observation interval, whereas $\mathbf{v}_t$ may vary with time.

- $\mathcal{G}(\ell)$ denotes an $M \times 3$ gain matrix, relating the dipole source to the EEG signals, which depends on the dipole location $\ell$ but not on time $t$;

- $\mathbf{e}_t$ is an $M \times 1$ measurement noise vector.

In this thesis, we adopt the simple infinite homogeneous head model. In fact, more complicated models can be chosen [27]. However, it does not affect the localization algorithm. From Eq.(2.6), the gain matrix $\mathcal{G}$ has the following structure:

$$\mathcal{G}(\boldsymbol{\ell}) = \frac{1}{4\pi\zeta} \begin{bmatrix} \frac{\omega_{1x}-\ell_x}{d_1{}^3} & \frac{\omega_{1y}-\ell_y}{d_1{}^3} & \frac{\omega_{1z}-\ell_z}{d_1{}^3} \\ \frac{\omega_{2x}-\ell_x}{d_2{}^3} & \frac{\omega_{2y}-\ell_y}{d_2{}^3} & \frac{\omega_{2z}-\ell_z}{d_2{}^3} \\ \vdots & \vdots & \vdots \\ \frac{\omega_{mx}-\ell_x}{d_m{}^3} & \frac{\omega_{my}-\ell_y}{d_m{}^3} & \frac{\omega_{mz}-\ell_z}{d_m{}^3} \end{bmatrix} \tag{2.8}$$

where

- $\zeta$ denotes the constant isotropic conductivity of the head;

- $\boldsymbol{\omega}_i = [\omega_{ix}, \omega_{iy}, \omega_{iz}]^T$ denotes the $i$th sensor's location vector, known on the surface of the head sphere;

- $d_i$ denotes the distance between the dipole and $i$th sensor; i.e.,
  $d_i = \sqrt{(\omega_{ix} - \ell_x)^2 + (\omega_{iy} - \ell_y)^2 + (\omega_{iz} - \ell_z)^2}$.

## 2.1.3 Multiple dipoles model

Now, we consider a more complicated model with multiple dipoles. Suppose we still have $M$ EEG sensors, each having $N$ snapshot measurements to localize $K$ dipole sources. The spatial-temporal EEG measurement matrix $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \cdots, \mathbf{f}_N]$ can be expressed by

$$\mathbf{F} = \mathbf{G}\mathbf{V} + \mathbf{E} \tag{2.9}$$

Here,

- $\mathbf{G}$ is an $M \times 3K$ gain matrix; i.e., $\mathbf{G} = [\mathcal{G}(\boldsymbol{\ell}_1), \mathcal{G}(\boldsymbol{\ell}_2), \cdots, \mathcal{G}(\boldsymbol{\ell}_K)]$ with each matrix $\mathcal{G}(\boldsymbol{\ell}_k)$ for $k = 1, 2, \cdots, K$ having the same structure as Eq.(2.8);

$$\bullet \; \mathbf{V} = \begin{bmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \\ \vdots \\ \mathbf{V}_K \end{bmatrix}$$, where the $\mathbf{V}_k$ denotes a $3 \times N$ collection matrix of all the $N$

moment vectors for the $k$th dipole; i.e.,

$$
\begin{aligned}
\mathbf{V}_k &= [\mathbf{v}_k(t_1), \mathbf{v}_k(t_2), \cdots, \mathbf{v}_k(t_N)] \\
&= \begin{pmatrix} v_{kx}(t_1) & v_{kx}(t_2) & \cdots & v_{kx}(t_N) \\ v_{ky}(t_1) & v_{ky}(t_2) & \cdots & v_{ky}(t_N) \\ v_{kz}(t_1) & v_{kz}(t_2) & \cdots & v_{kz}(t_N) \end{pmatrix}
\end{aligned}
$$

$\bullet \; \mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_N]$ denotes an $M \times N$ noise matrix.

Since the moment vector $\mathbf{v}$ can be furthermore decomposed into two parts: a unit-norm orientation vector $\boldsymbol{\mu}$ and a scalar amplitude $s$, such that $\mathbf{v}_k = \boldsymbol{\mu}_k s_k = [\mu_{kx}, \mu_{ky}, \mu_{kz}]^T s$, there are two typical categories of dipole moment models.

## (1) Fixed-Orientation Moment

Some researchers [33] argue that, physiologically a dipole's orientation should not rotate because the dipole model represents a fixed neuroanatomical structure. Therefore in this model, $\boldsymbol{\mu}$ is supposed to be fixed during the observation interval, and only its amplitude varies with time. Therefore, if we let $\boldsymbol{\theta}_k$ denote all the fixed parameters for the $k$th dipole, $\boldsymbol{\theta}_k = [\boldsymbol{\ell}_k^T, \boldsymbol{\mu}_k^T]^T$, and put all the items bearing these fixed parameters into one matrix, the forward model of fixed-orientation moment can be rewritten

from Eq.(2.9) as

$$
\begin{aligned}
\mathbf{F} &= \left[ \underbrace{\mathcal{G}(\ell_1)\boldsymbol{\mu}_1}_{\mathbf{a}(\boldsymbol{\theta}_1)}, \ \underbrace{\mathcal{G}(\ell_2)\boldsymbol{\mu}_2}_{\mathbf{a}(\boldsymbol{\theta}_2)}, \ \cdots, \ \underbrace{\mathcal{G}(\ell_K)\boldsymbol{\mu}_K}_{\mathbf{a}(\boldsymbol{\theta}_K)} \right] \begin{bmatrix} \mathbf{s}_1^T \\ \mathbf{s}_2^T \\ \vdots \\ \mathbf{s}_K^T \end{bmatrix} + \mathbf{E} \\
&= \underbrace{[\mathbf{a}(\boldsymbol{\theta}_1), \mathbf{a}(\boldsymbol{\theta}_2), \cdots, \mathbf{a}(\boldsymbol{\theta}_K)]}_{\mathbf{A}(\boldsymbol{\Theta}_f)} \mathbf{S} + \mathbf{E} \\
&= \mathbf{A}(\boldsymbol{\Theta}_f)\mathbf{S} + \mathbf{E} \tag{2.10}
\end{aligned}
$$

where $\boldsymbol{\Theta}_f = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_K]$.

**(2) Rotating Moment**

Physically, since two nearly collocated dipoles with independent time series are not distinguishable, they may be modelled as a single dipole with a rotating orientation [27]. In this model, both arguments $\boldsymbol{\mu}$ and $s$ vary with time, and the forward model Eq. (2.9) becomes,

$$
\begin{aligned}
\mathbf{F} &= \underbrace{\left[ \mathcal{G}(\ell_1), \ \mathcal{G}(\ell_2), \ \cdots, \ \mathcal{G}(\ell_K), \right]}_{\mathbf{G}(\boldsymbol{\Theta}_r)} \begin{bmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \\ \vdots \\ \mathbf{V}_K \end{bmatrix} + \mathbf{E} \\
&= \mathbf{G}(\boldsymbol{\Theta}_r)\mathbf{V} + \mathbf{E} \tag{2.11}
\end{aligned}
$$

where $\boldsymbol{\Theta}_r = [\ell_1, \ell_2, \cdots, \ell_K]$.

## 2.2 Solution to the Inverse Problem

Once the dipole and head model are selected, we can solve the inverse problem, that is, we can determine the dipole parameters from EEG measurements.

## 2.2.1 Assumption and Problem Statement

Throughout this thesis, we make the following assumptions:

1. The number of dipoles $K$ is known at the receiver side;

2. The gain matrix $G$ has full column rank.

3. Noise matrix $E$ is zero-mean white Gaussian, distributed as $\mathcal{N}(0, \sigma^2 I)$, and it is uncorrelated with the dipole activities.

For Assumption 1, the number of dipoles $K$ can be estimated from the effective rank of the data matrix or using information-theoretic criteria [3]. In practice, expert data analysts often run several model orders and select results based on physiological plausibility. Since the determination of the number of dipoles is well researched, this issue is excluded in our research interests, and Assumption 1 is taken to be valid. Assumption 2 implies that the dipoles' activities are not correlated, which enable us to uniquely identify the dipole locations if the number of sensors is large enough. In addition, noise comes mainly from the background activity in neurons. Therefore, Assumption 3 can be justified by the large number of neurons normally active throughout the brain and has been validated in [13].

Our goal in this thesis is to solve the inverse problem of the forward model described by Eq. (2.9), i.e., we intend to solve the following problem:

**Problem 2.1.** *Given the observations of EEG $F$ from sensors corrupted with noises, we want to develop an efficient algorithm to uniquely estimate the locations $\ell_k$ of all the dipoles in Eq.(2.9).*

## 2.2.2 Overview of Localization Methods

In this overview, we only focus on two classes of localization methods currently used among researches. This is because these two classes of methods have greater stability and superior performance to all other methods.

## MAXIMUM LIKELIHOOD

The principle of Maximum Likelihood (ML) parameter estimation is to find the parameter values that make the observed data most likely. It is intuitively appealing and has remained one of the most powerful methods in estimation theory. In fact, the method of maximum likelihood may be applied to any estimation problems, provided that we formulate the joint probability density function of the available set of observed data as a function of parameters of interest. It was first used by Fisher. For detailed discussion of this technique, see [30] [36]. Recall the general dipole model

$$\mathbf{f} = \mathbf{G}(\ell)\mathbf{v} + \mathbf{e}$$

The method of maximum likelihood is based on the principle that we should estimate the parameter $\ell$ by its most plausible values, given the observed sample vector $\mathbf{f}$. In other words, the maximum-likelihood estimator of $\ell$ is the value for which the conditional joint probability density function $p(\mathbf{f}|\ell)$ is maximized, i.e., the ML estimate of $\ell$ is given by

$$\hat{\ell}_{ml} = \arg\ \max_{\ell}\ p(\mathbf{f}|\ell) \tag{2.12}$$

For dipole localization, we can adopt the ML method in [?] in which some of the irrelevant parameters in the likelihood function are removed, arriving at a concentrated maximum likelihood problem, i.e., $p(\mathbf{f}|\ell)$ is replaced by a concentrated function $\text{Tr}[(\mathbf{I} - \mathbf{G}(\mathbf{G}^{\mathrm{T}}\mathbf{G})^{-1}\mathbf{G}^{\mathrm{T}})\hat{\mathcal{R}}_f]$, where $\hat{\mathcal{R}}_f$ is the estimated covariance matrix of the observation vector, obtained by

$$\hat{\mathcal{R}}_f\ =\ \frac{1}{N}\sum_{n=1}^{N}\mathbf{f}_n\mathbf{f}_n^T \tag{2.13}$$

The ML estimate of the dipole moment can then be derived by a simple least-square fit, $\hat{\mathbf{v}} = (\hat{\mathbf{G}}^T\hat{\mathbf{G}})^{-1}\hat{\mathbf{G}}^T\mathbf{f}$, where $\hat{\mathbf{G}}$ is used instead of $\mathbf{G}(\hat{\ell})$ for notational convenience.

Since the mathematical models in dipole localization are highly nonlinear, the multi-variant nonlinear optimization process makes the ML estimation computationally intractable. As the number of sources and the number of quantization levels

increase, the optimization objective function results in increasing chances of trapping in the local minima. The author in [10] proposed maximum likelihood methods for localization of fixed dipoles by modeling the dipole moments as a linear combination of parametric or non-parametric basis functions, and further developed ML-based methods under spatially correlated noise with unknown covariance [11]. The disadvantage of these methods is that the basis functions should be known a priori, and that they are not universal to all dipoles.

## SUBSPACE-BASED METHODS

In principle, the subspace-based methods, which operate on the second order statistics, find optimum peaks of their objective function by employing certain projections on the estimated signal subspace, or alternatively on the estimated noise subspace. Mosher and Leahy [26] [27] applied Multiple Signal Classification (MUSIC) [34] into estimating the location of the dipoles.

**A. MUSIC**    Consider the EEG observation model in Eq.(2.10),

$$\mathbf{F} = \mathbf{A}(\Theta)\mathbf{S} + \mathbf{E}$$

Since we assume the dipole activities are linearly independent among each other, both matrix $\mathbf{A}$ and $\mathbf{S}$ are of full rank $K$. We also observe that each column of the data matrix $\mathbf{F}$ is formed as a linear combination of the manifold vectors from each of the dipole sources. Therefore, an eigen-decomposition performing to the correlation matrix of the noiseless data yields orthogonal basis that span the same subspace, referred to as the signal subspace, as the manifold vectors $\mathbf{a}(\ell_k), k = 1, 2, \cdots, K$ with true dipole parameters. This is the fundamental of the MUSIC algorithm. The best we can do is to search for the gain vectors which are closest to being orthogonal to the noise subspace. Accordingly, we form an orthogonal projector [29] such that

$$\mathbf{P}_n = \Omega_n(\Omega_n^T \Omega_n)^{-1}\Omega_n^T \tag{2.14}$$

where $\Omega_n$ denotes the noise subspace given by the eigen-decomposition of the signal

covariance matrix,

$$\mathcal{R}_f = \Omega\Lambda\Omega^T = [\Omega_s\ \Omega_n] \begin{bmatrix} \Lambda_s & \\ & \Lambda_n \end{bmatrix} \begin{bmatrix} \Omega_s^T \\ \Omega_n^T \end{bmatrix} \tag{2.15}$$

with $\Omega_s$ denoting the signal subspace and $\Lambda$ denoting the diagonal matrix of corresponding eigenvalues. In practice, the true value of $\mathcal{R}_f$ is seldom known. Therefore, an estimate $\hat{\mathcal{R}}_f$ is obtained by averaging the outer products of the data over the total number of snapshots $N$ such that

$$\hat{\mathcal{R}}_f = \frac{1}{N}\mathbf{FF}^T \tag{2.16}$$

We can now use $\hat{\mathbf{P}}_n$ to project the gain vectors $\mathbf{a}(\theta_k)$ onto the estimated noise subspace spanned by the column vectors of $\hat{\Omega}_n$. The values of $\hat{\theta}_k$, $k = 1, \cdots, K$, which give the $K$ highest peaks of the quantity $1/ \parallel \hat{\mathbf{P}}_n\mathbf{a}(\theta_k) \parallel^2$ are the estimated parameters of the $K$ dipole sources, i.e.,

$$\hat{\boldsymbol{\theta}}_{k\ music} = arg\ \min_{\boldsymbol{\theta}_k}\ \parallel \hat{\mathbf{P}}_n\mathbf{a}(\theta_k) \parallel^2 \tag{2.17}$$

Alternatively, a metric to describe the closeness between the column vectors $\mathbf{a}(\boldsymbol{\theta}_k), k = 1, 2, \cdots, K$ and the estimated signal subspace $\hat{\Omega}_s$, defined by

$$\mathcal{C}_s\left\{\mathbf{a}(\boldsymbol{\theta}_k), \hat{\Omega}_s\right\} = \frac{\mathbf{a}(\boldsymbol{\theta}_k)^T\hat{\Omega}_s\hat{\Omega}_s^T\mathbf{a}(\boldsymbol{\theta}_k)}{\mathbf{a}(\boldsymbol{\theta}_k)^T\mathbf{a}(\boldsymbol{\theta}_k)} \tag{2.18}$$

can be employed resultingly in a modified MUSIC [26], i.e.,

$$\hat{\boldsymbol{\theta}}_{k\ music} = arg\ \max_{\boldsymbol{\theta}_k}\ \mathcal{C}_s\{\mathbf{a}(\boldsymbol{\theta}_k),\ \hat{\Omega}_s\} \tag{2.19}$$

In applying MUSIC for dipole localization, errors in estimating the signal subspace can lead to errors in the location of the peaks in the MUSIC metric. In addition, as the dimension of the signal subspace increases, automatically finding several local maxima in the MUSIC metric become more complex.

**B. RAP-MUSIC**      Mosher and Leahy [28] therefore proposed the Recursively Applied and Projected (RAP) MUSIC by using a recursive procedure in which each

dipole is found as the global maximizer of a recursively modified cost function. By doing so, a framework for the RAP-MUSIC was established based on the theory of principal angles [14], which then aroused great interests due to their ability to accurately locate multiple dipoles. In RAP-MUSIC, the successively localized dipole sources form an intermediate projection matrix first, and then by searching over a reduced subspace obtained from the projection's orthogonal subspace complement, the next dipole is found. For the estimation of each dipole, the RAP-MUSIC method utilizes a modified criteria to find each global maxima,

$$\hat{\boldsymbol{\theta}}_k = \arg \max_{\boldsymbol{\theta}_k} \mathcal{C}_s\{\boldsymbol{\Pi}_{k-1}^\perp \mathbf{a}(\boldsymbol{\theta}_k), \ \boldsymbol{\Pi}_{k-1}^\perp \hat{\boldsymbol{\Omega}}_s\} \tag{2.20}$$

where $\boldsymbol{\Pi}_{k-1}^\perp$ is the projector onto the previous $(k-1)$ signal subspaces

$$\boldsymbol{\Pi}_{k-1}^\perp = \mathbf{I} - \hat{\mathbf{A}}_{k-1}(\hat{\mathbf{A}}_{k-1}^T \hat{\mathbf{A}}_{k-1})^{-1}\hat{\mathbf{A}}_{k-1}^T$$

with $\hat{\mathbf{A}}_{k-1}$ formed from the manifold estimates of the previous $k-1$ recursions, $\hat{\mathbf{A}}_{k-1} = [\mathbf{a}(\hat{\boldsymbol{\theta}}_1), \cdots, \mathbf{a}(\hat{\boldsymbol{\theta}}_{k-1})]$.

However, the subspace methods have their own drawbacks. They are highly dependent on the sufficiency and underlying structure of observed data. When either the number of sensors or the number of snapshots is not very large, or the signal to noise ratio is low, the error resulting from estimating signal subspace basis leads to a significant estimation error of the dipole locations.

# Chapter 3

# Uniqueness of the Determination

# of A Single Dipole Location

Prior to developing any new localization algorithm, we must ensure that there is a unique relationship between the scalp-recorded EEG and the dipoles. Studies in [2] stated that two distinct dipoles could not generate the same potential map on a continuous set of scalp points. However, since EEG data are recorded only on a few discrete sensor locations, we do not know if a unique inverse mapping exists. Given an EEG array geometry, the uniqueness of the localization has to be addressed.

In this chapter, given the observations of EEG in an array of finite sensors from a single rotating dipole, we want to uniquely localize the dipole. We design a geometrical structure of an array with the minimum number of sensors to guarantee a unique localization of the dipole. In addition, we present a closed-form solution for this localization problem.

## 3.1   Array Geometrical Structure

Unlike a uniform linear array, the EEG array has no fixed manifold structure such as shift invariance or Vandermonde [34], but the gain matrix has the following geomet-

rical property.

**Proposition 1.** *The gain matrix $\mathcal{G}$ defined by Eq.(2.8) satisfies*

$$[\mathcal{G}\mathcal{G}^T]_{ii} = (\frac{1}{4\pi\zeta})^2 \frac{1}{d_i^4} \quad i = 1, 2 \cdots, M$$

*where $d_i$ denotes the distance between the dipole and ith sensor; i.e.,*

$d_i = \sqrt{(\omega_{ix} - \ell_x)^2 + (\omega_{iy} - \ell_y)^2 + (\omega_{iz} - \ell_z)^2}$. *This property holds universally for any gain matrix in an infinite homogenous model and it is independent of the array structure.*

Proposition 1, which can be verified by a direct calculation, plays a key role in uniquely determining the location for a single rotating dipole when the second-order statistic of observations is available. Consider the rotating dipole model with only one dipole. Since the moment is independent in each dimension $\{x, y, z\}$ and in time, we have

$$E\{\mathbf{v}\mathbf{v}^T\} = E\{s^2\boldsymbol{\mu}\boldsymbol{\mu}^T\} = \sigma_v^2\mathbf{I}_3$$

where $\sigma_v$ denotes the variance of the intensity. Therefore, the covariance matrix of the measurement vector $\mathbf{f}$ is given by

$$\mathcal{R}_f = E\{\mathbf{f}\mathbf{f}^T\} = \sigma_v^2\mathcal{G}\mathcal{G}^T + \sigma^2\mathbf{I} \tag{3.1}$$

Now, a natural question is

**Problem 3.1.** *what is the minimum number of sensors and how these senors are positioned on the head surface to uniquely determine the above five unknown parameters, $\sigma_v, \sigma, \ell_x, \ell_y$ and $\ell_z$ if $\mathcal{R}_f$ is given?*

On one hand, there are totally five parameters in Eq.(3.1) to be determined. On the other hand, $\mathcal{R}_f$ is a symmetric matrix and has $(1 + M)M/2$ non-redundant elements. Hence, it seems that an immediate answer to the above problem is to choose the number of sensors $M$ such that

$$\frac{(1 + M)M}{2} \geq 5$$

However, this is far from sufficient. Take 3 sensors for instance. It is the minimum number of sensors to satisfy the condition. Since any 3 arbitrary sensors can be viewed as being placed in a circle around the surface of spherical head, this two dimensional planar structure is not able to identify the parameter information in a three dimensional environment. For example, we cannot discriminate two dipoles if they are symmetrically distributed to the array plane. Therefore, three coplanar sensors lead to an ambiguity of location estimation. In fact, such argument on three coplanar sensors can be extended to a general situation. Therefore, we have

**Proposition 2.** *An array of sensors positioned on the same plane cannot provide a unique localization of the dipole.*

Proposition 2 tells us that we need at least four sensors not belonging to a plane to uniquely localize a dipole. To keep the array geometry structure as simple as possible, we employ four sensors placed at the following positions as showed in Fig. 3.1: $\omega_1 = [-\beta, 0, 0]^T$, $\omega_2 = [\beta, 0, 0]^T$, $\omega_3 = [0, \beta, 0]^T$, and $\omega_4 = [0, 0, \beta]^T$, where $\beta$ is the head radius.

## 3.2   Closed-form Solution

Our purpose in this section is to obtain a closed-form solution which uniquely determine the variances of the intensity and the noise, and the position of the dipole. Utilizing Eq.(3.1) with 4 sensors, we can have a system of 10 equations, among which we carefully select 5 equations to sufficiently determine the 5 unknown variables.
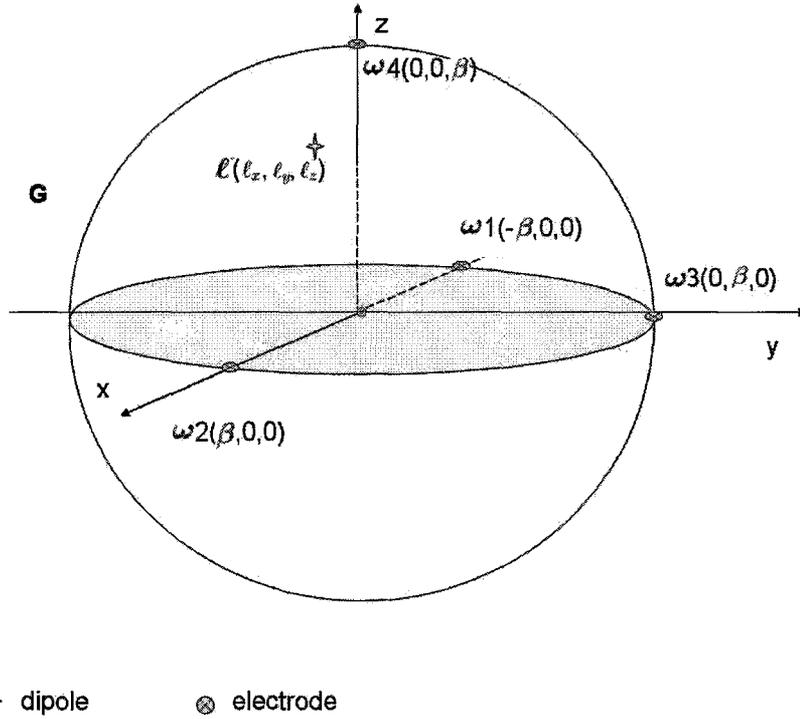
Figure 3.1: Illustration of 4 electrodes in a particular structure

Employing Proposition 1 and equating the diagonal elements of Eq.(3.1) yield

$$\frac{\sigma_v}{4\pi\zeta\sqrt{[\mathcal{R}_f]_{11} - \sigma^2}} = \rho^2 + \beta^2 + 2\beta\ell_x \qquad (3.2a)$$

$$\frac{\sigma_v}{4\pi\zeta\sqrt{[\mathcal{R}_f]_{22} - \sigma^2}} = \rho^2 + \beta^2 - 2\beta\ell_x \qquad (3.2b)$$

$$\frac{\sigma_v}{4\pi\zeta\sqrt{[\mathcal{R}_f]_{33} - \sigma^2}} = \rho^2 + \beta^2 - 2\beta\ell_y \qquad (3.2c)$$

$$\frac{\sigma_v}{4\pi\zeta\sqrt{[\mathcal{R}_f]_{44} - \sigma^2}} = \rho^2 + \beta^2 - 2\beta\ell_z \qquad (3.2d)$$

where we denote $\rho^2 = \ell_x^2 + \ell_y^2 + \ell_z^2$. In addition, comparing some off-diagonal elements of Eq.(3.1), we can also obtain

$$[\mathcal{R}_f]_{12} = (\frac{\sigma_v}{4\pi\zeta})^2 \frac{\rho^2 - \beta^2}{(d_1 d_2)^3} \tag{3.3a}$$

$$[\mathcal{R}_f]_{13} = (\frac{\sigma_v}{4\pi\zeta})^2 \frac{\rho^2 + \beta(\ell_x - \ell_y)}{(d_1 d_3)^3} \tag{3.3b}$$

$$[\mathcal{R}_f]_{14} = (\frac{\sigma_v}{4\pi\zeta})^2 \frac{\rho^2 + \beta(\ell_x - \ell_z)}{(d_1 d_4)^3} \tag{3.3c}$$

$$[\mathcal{R}_f]_{23} = (\frac{\sigma_v}{4\pi\zeta})^2 \frac{\rho^2 - \beta(\ell_x + \ell_y)}{(d_2 d_3)^3} \tag{3.3d}$$

$$[\mathcal{R}_f]_{24} = (\frac{\sigma_v}{4\pi\zeta})^2 \frac{\rho^2 - \beta(\ell_x + \ell_z)}{(d_2 d_4)^3} \tag{3.3e}$$

$$[\mathcal{R}_f]_{34} = (\frac{\sigma_v}{4\pi\zeta})^2 \frac{\rho^2 - \beta(\ell_y + \ell_z)}{(d_3 d_4)^3} \tag{3.3f}$$

Hence, solving the original Problem 3.1 is equivalent to uniquely solving Eqs.(3.2) and (3.3). Note that in all these equations, the covariance matrix $\mathcal{R}_f$ is of true value. In other words, we assume $\mathcal{R}_f$ is exactly known. If an estimate $\hat{\mathcal{R}}_f$ is used, then none of these equations hold. Then by carefully selecting 5 equations out of Eqs.(3.2)- (3.3), we can derive unique solution for the 5 unknowns, which is necessarily consistent with the solution of the other equations. In fact, we choose Eqs.(3.2) and Eq.(3.3a) for the sake of the simplest solving procedure, which is given in detail as follows.

First of all, it is straightforward to solve for $\sigma$. Let the eigen-decomposition of covariance $\mathcal{R}_f$ be

$$\mathcal{R}_f = \Omega \Lambda \Omega^T = \Omega \begin{bmatrix} \Lambda_v & \\ & \lambda_e \end{bmatrix} \Omega^T$$

where $\Lambda_v$ denotes the diagonal matrix containing the 3 largest eigenvalues of the covariance matrix $\mathcal{R}_f$ and $\lambda_e$ is the smallest eigenvalue. Since $\mathcal{G}\mathcal{G}^T$ is a $4 \times 4$ matrix with rank 3, combining this with Eq.(3.1), we see that $\lambda_e = \sigma^2$. Thereafter, the other four parameters can be obtained by solving Eqs(3.2) and (3.3a). On one hand, substituting Eq.(3.2a) and Eq.(3.2b) into Eq.(3.3a) results in

$$\rho^2 = \frac{[\mathcal{R}_f]_{12}}{4\pi\zeta\{([\mathcal{R}_f]_{11} - \sigma^2)([\mathcal{R}_f]_{22} - \sigma^2)\}^{3/4}} \sigma_v + \beta^2 \tag{3.4}$$

On the other hand, adding Eq.(3.2a) and Eq.(3.2b) leads to

$$\rho^2 \;=\; \frac{1}{8\pi\zeta}\Big[\frac{1}{\sqrt{[\mathcal{R}_f]_{11}-\sigma^2}}+\frac{1}{\sqrt{[\mathcal{R}_f]_{22}-\sigma^2}}\Big]\sigma_v - \beta^2 \tag{3.5}$$

Combining Eq.(3.8) and Eq.(3.9) yields,

$$\sigma_v = \frac{2\beta^2}{Q_2 - Q_1}$$

where $Q_1$ and $Q_2$ are defined as

$$Q_1 \;=\; \frac{1}{4\pi\zeta}[\mathcal{R}_f]_{12}\{([\mathcal{R}_f]_{11}-\sigma^2)([\mathcal{R}_f]_{22}-\sigma^2)\}^{-3/4} \tag{3.6a}$$

$$Q_2 \;=\; \frac{1}{8\pi\zeta}\{([\mathcal{R}_f]_{11}-\sigma^2)^{-1/2}+([\mathcal{R}_f]_{22}-\sigma^2)^{-1/2}\} \tag{3.6b}$$

So far, $\sigma$, $\sigma_v$ and $\rho^2$ have been obtained. Substituting these terms back into Eq.(3.2), we come up with the immediately solutions for the dipole location as follows:

$$\ell_x \;=\; \frac{1}{2\beta}(\rho^2+\beta^2-\frac{\sigma_v}{4\pi\zeta\sqrt{[\mathcal{R}_f]_{22}-\sigma^2}}) \tag{3.7a}$$

$$\ell_y \;=\; \frac{1}{2\beta}(\rho^2+\beta^2-\frac{\sigma_v}{4\pi\zeta\sqrt{[\mathcal{R}_f]_{33}-\sigma^2}}) \tag{3.7b}$$

$$\ell_z \;=\; \frac{1}{2\beta}(\rho^2+\beta^2-\frac{\sigma_v}{4\pi\zeta\sqrt{[\mathcal{R}_f]_{44}-\sigma^2}}) \tag{3.7c}$$

In fact, without using eigen value decomposition, we can also uniquely determine $\sigma^2$. This is shown in Section 3.3. As a conclusion, we state the following property.

**Proposition 3.** *For a rotating single dipole model, the variances of intensity and the noise and the position of the dipole can be uniquely determined using the second order statistics if and only if an EEG array has at least four sensors not belonging to one plane. Moreover, under this sufficient condition with four sensors, closed-form solutions uniquely determining the variances of intensity and the noise and the position of the dipole are given by Eq.(3.7).*

## 3.3    Uniqueness Validation

In this section, a complementary proof is presented, without using eigen-decomposition, that $\sigma^2$ exists necessarily and uniquely. As a result, all the subsequent solutions are unique. For convenience, $\sigma_v = 1$ is assumed. On one hand, substituting Eq.(3.2a) and Eq.(3.2b) into Eq.(3.3a) results in

$$\rho^2 \quad = \quad \frac{[\mathcal{R}_f]_{12}}{4\pi\zeta\{([\mathcal{R}_f]_{11} - \sigma^2)([\mathcal{R}_f]_{22} - \sigma^2)\}^{3/4}}\sigma_v + \beta^2 \tag{3.8}$$

On the other hand, adding Eq.(3.2a) and Eq.(3.2b) leads to

$$\rho^2 \quad = \quad \frac{1}{8\pi\zeta}[\frac{1}{\sqrt{[\mathcal{R}_f]_{11} - \sigma^2}} + \frac{1}{\sqrt{[\mathcal{R}_f]_{22} - \sigma^2}}]\sigma_v - \beta^2 \tag{3.9}$$

Subtracting Eq.(3.9) from Eq.(3.8), we can write

$$\frac{1}{\sqrt{[\mathcal{R}_f]_{11} - \sigma_2}} + \frac{1}{\sqrt{[\mathcal{R}_f]_{22} - \sigma_2}} - \frac{2[\mathcal{R}_f]_{12}}{[([\mathcal{R}_f]_{11} - \sigma_2)([\mathcal{R}_f]_{22} - \sigma_2)]^{3/4}} - 4\beta^2 = 0 \tag{3.10}$$

If we define the left hand side to be $f(\sigma^2)$, obviously, the domain of $\sigma^2$ is $0 < \sigma^2 <$
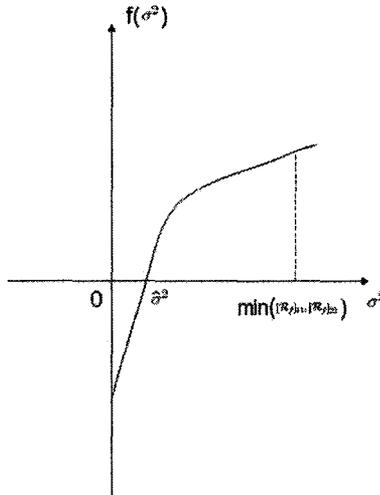


Figure 3.2: Illustration of monotone $f(\sigma^2)$

$\min\{[\mathcal{R}_f]_{11}, [\mathcal{R}_f]_{22}\}$. Moreover, the following properties about function $f(\sigma^2)$,

$$f(0) < 0 \tag{3.11}$$

$$f(\min\{[\mathcal{R}_f]_{11}, [\mathcal{R}_f]_{22}\}) = +\infty \tag{3.12}$$

can be easily verified by

$$\frac{1}{\sqrt{[\mathcal{R}_f]_{11}}} + \frac{1}{\sqrt{[\mathcal{R}_f]_{22}}} < \frac{1}{\sqrt{[\mathcal{R}_f]_{11} - \sigma_2}} + \frac{1}{\sqrt{[\mathcal{R}_f]_{22} - \sigma^2}} = 2(\rho^2 + \beta^2)$$

and

$$\frac{2[\mathcal{R}_f]_{12}}{[([\mathcal{R}_f]_{11})([\mathcal{R}_f]_{22})]^{3/4}} > \frac{2[\mathcal{R}_f]_{12}}{[([\mathcal{R}_f]_{11} - \sigma^2)([\mathcal{R}_f]_{22} - \sigma^2)]^{3/4}} = 2(\rho^2 - \beta^2)$$

In addition, $f(\sigma^2)$ is a monotonically increasing function within the valid region of $\sigma^2$. This is because

$$f'(\sigma^2) = \frac{1}{2}([\mathcal{R}_f]_{11} - \sigma^2)^{-3/2} + \frac{1}{2}([\mathcal{R}_f]_{22} - \sigma^2)^{-3/2}$$

$$-\frac{3[\mathcal{R}_f]_{12}}{2}[([\mathcal{R}_f]_{11} - \sigma^2)([\mathcal{R}_f]_{22} - \sigma^2)]^{-7/4}([\mathcal{R}_f]_{11} + [\mathcal{R}_f]_{22} - 2\sigma^2) > 0$$

Consequently, a rough illustration of function $f(\sigma^2)$ can be shown in Fig. 3.2, from which we deduce that there exists one and only solution for Eq.(3.10) such that $f(\sigma^2) = 0$. Then a unique solution of $\sigma^2 > 0$ is ensured.

# Chapter 4

# Successive Estimation Methods based on QR Decomposition

In this chapter, using the non-coherent Maximum Likelihood (ML) estimator or the Least Square (LS) estimator as a criterion, the multiple dipole localization is formulated as a nonconvex optimization problem. Then, exploiting the QR decomposition, we propose a unified recursive algorithm to approximately calculate the optimal solution by successively and efficiently solving a series of optimization sub-problems. To facilitate the discussion, we will first introduce the concept of QR decomposition, with a focus on the matrix version of the Gram-Schmidt orthogonal process.

## 4.1 QR Decomposition

In linear algebra, the QR decomposition (also called the QR factorization) of a matrix is a decomposition of the matrix into an orthogonal and a triangular matrix. Given

a full rank $M \times K$ matrix $\mathbf{A}$, its QR-decomposition is of the form

$$\mathbf{A} = \mathbf{QR}$$

$$= \begin{bmatrix} \mathbf{q}_1, & \mathbf{q}_2, & \cdots, & \mathbf{q}_K \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1K} \\ 0 & r_{22} & \cdots & r_{2K} \\ \vdots & & & \\ 0 & 0 & \cdots & r_{KK} \end{bmatrix}$$

where $\mathbf{R}$ is a $K \times K$ upper triangular matrix, and $\mathbf{Q}$ is an $M \times K$ matrix with orthonormal columns, i.e.,

$$\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$$

There are several methods for computing the QR-decomposition. Here we introduce the **Gram-Schmidt** process. In this procedure, we successively form orthonormal columns of $\mathbf{Q}$ from the columns of $\mathbf{A}$, beginning at the first column. Since each $\mathbf{q}_i$ is a vector of unit norm, the first column $\mathbf{q}_1$ is defined as

$$\mathbf{q}_1 = \frac{\mathbf{a}_1}{\| \mathbf{a}_1 \|}$$

Then since each column vector $\mathbf{a}_k$, for $k = 1, 2, \cdots, K$, is in the space spanned by $\{\mathbf{q}_1, \mathbf{q}_1, \cdots, \mathbf{q}_K\}$, we can write $\mathbf{a}_k$ as a linear combination of the orthonormal basis $\mathbf{q}_1, \mathbf{q}_1, \cdots, \mathbf{q}_K$ in the following manner:

$$\mathbf{a}_1 = \langle \mathbf{a}_1, \mathbf{q}_1 \rangle \mathbf{q}_1 \tag{4.1a}$$

$$\mathbf{a}_2 = \langle \mathbf{a}_2, \mathbf{q}_1 \rangle \mathbf{q}_1 + \langle \mathbf{a}_2, \mathbf{q}_2 \rangle \mathbf{q}_2 \tag{4.1b}$$

$$\vdots$$

$$\mathbf{a}_k = \langle \mathbf{a}_k, \mathbf{q}_1 \rangle \mathbf{q}_1 + \langle \mathbf{a}_k, \mathbf{q}_2 \rangle \mathbf{q}_2 + \cdots + \langle \mathbf{a}_k, \mathbf{q}_k \rangle \mathbf{q}_k \tag{4.1c}$$

where notation $\langle \mathbf{a}_k, \mathbf{q}_k \rangle = \mathbf{a}_k^T \mathbf{q}_k$ denotes the inner product of these two vectors. As a result, $\mathbf{q}_2$ can be represented in terms of $\mathbf{a}_1$ and $\mathbf{a}_2$ by

$$\mathbf{q}_2 = \frac{(\mathbf{I} - \mathbf{q}_1\mathbf{q}_1^T)\mathbf{a}_2}{\| (\mathbf{I} - \mathbf{q}_1\mathbf{q}_1^T)\mathbf{a}_2 \|} = \frac{\mathbf{\Pi}^{(1)}\mathbf{a}_2}{\| \mathbf{\Pi}^{(1)}\mathbf{a}_2 \|} \tag{4.2}$$

where $\mathbf{\Pi}^{(1)}$ is the projection matrix formed by the first column vector $\mathbf{q}_1$. To generalize, for $k = 1, 2, \cdots, K$, the $k$th column vector $\mathbf{q}_k$ can be written in terms of $\mathbf{a}_1, \cdots, \mathbf{a}_k$ as follows:

$$\mathbf{q}_k = \frac{(\mathbf{I} - \mathbf{Q}_{k-1}\mathbf{Q}_{k-1}^T)\mathbf{a}_k}{\parallel (\mathbf{I} - \mathbf{Q}_{k-1}\mathbf{Q}_{k-1}^T)\mathbf{a}_k \parallel} = \frac{\mathbf{\Pi}^{(k-1)}\mathbf{a}_k}{\parallel \mathbf{\Pi}^{(k-1)}\mathbf{a}_k \parallel} \tag{4.3}$$

where

$$\begin{cases} \mathbf{Q}_0 = \mathbf{0} \\ \mathbf{Q}_k = [\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_k] \end{cases} \qquad \begin{cases} \mathbf{\Pi}^{(0)} = \mathbf{I} \\ \mathbf{\Pi}^{(k)} = \mathbf{I} - \mathbf{Q}_k\mathbf{Q}_k^T \end{cases}$$

Here, $\mathbf{\Pi}^{(k)}$ is a projection matrix with the following properties,

$$\mathbf{\Pi}^{(k)^T} = \mathbf{\Pi}^{(k)}$$

$$\mathbf{\Pi}^{(k)}\mathbf{\Pi}^{(k)} = \mathbf{\Pi}^{(k)}$$

In addition, the matrix $\mathbf{R}$ can be obtained directly by Eq.(4.1),

$$\mathbf{R} = \begin{bmatrix} \langle \mathbf{a}_1, \mathbf{q}_1 \rangle & \langle \mathbf{a}_2, \mathbf{q}_1 \rangle & \cdots & \langle \mathbf{a}_M, \mathbf{q}_1 \rangle \\ 0 & \langle \mathbf{a}_2, \mathbf{q}_2 \rangle & \cdots & \langle \mathbf{a}_M, \mathbf{q}_2 \rangle \\ \vdots & & & \\ 0 & 0 & \cdots & \langle \mathbf{a}_M, \mathbf{q}_K \rangle \end{bmatrix}$$

so that the $k$th diagonal element of $\mathbf{R}$ can be expressed by

$$[\mathbf{R}]_{kk} = \parallel (\mathbf{I} - \mathbf{Q}_{k-1}\mathbf{Q}_{k-1}^T)\mathbf{a}_k \parallel = \parallel \mathbf{\Pi}^{(k-1)}\mathbf{a}_k \parallel \qquad k = 1, 2, \cdots, K \tag{4.4}$$

We would like to emphasize the following two points on the QR decomposition, which play an important role in developing our recursive algorithms.

- $\mathbf{\Pi}^{(k)}$ depends only on the first $k$ column vectors of $\mathbf{A}$, $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_k$.

- The $k$th diagonal entry $[\mathbf{R}]_{kk}$ also depends only on the first $k$ column vectors of $\mathbf{A}$, $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_k$.

# 4.2 Successive-ML Localization Methods based on QR Decomposition

In this section, assuming some stochastic properties of the dipoles are known a priori, we develop a computationally efficient localization method on the basis of the non-coherent ML criterion. We will focus on the fixed-orientation dipole model. Then a modified algorithm is developed for the model of rotating dipoles.

## 4.2.1 Fixed-orientation Dipole Model

Under the observation model described by Eq.(2.10), we assume that the dipole amplitude $\mathbf{S}$ and the noise $\mathbf{E}$ are independent and are zero-mean Gaussian distributed with covariance matrices $\mathbf{I}_N \otimes \Sigma_S$ and $\mathbf{I}_N \otimes \sigma^2 \mathbf{I}_M$ respectively, where $\Sigma_S$ is a $K \times K$ matrix. With this assumption, the $n$th column $\mathbf{f}_n$ of the observation matrix $\mathbf{F}$ in Eq.(2.10) is zero-mean Gaussian random vector with covariance matrix $\Sigma_F = \mathbf{A}\Sigma_S\mathbf{A}^T + \sigma^2\mathbf{I}$. Therefore, the likelihood function, i.e., the conditional probability density function of $\mathbf{f}_n$ can be expressed as

$$p\left(\mathbf{f}_n | \mathbf{A}\right) = \frac{\exp\{-\frac{1}{2}\mathbf{f}_n^T \Sigma_F^{-1} \mathbf{f}_n\}}{(2\pi)^{\frac{M}{2}} \det^{\frac{1}{2}} \Sigma_F} \tag{4.5}$$

Since we have assumed the dipole activities as well as noise are temporally independent, the columns of $\mathbf{F}$ are statistically independent. Then, the covariance matrix of $\mathbf{F}$ is $\mathbf{I}_N \otimes \Sigma_F$. As a result, the probability density function of the whole observation matrix $\mathbf{F}$ conditioned on $\mathbf{A}$ can be derived by:

$$
\begin{aligned}
p\left(\mathbf{F} \mid \mathbf{A}\right) &= p\left(\text{VEC}[\mathbf{F}] \mid \mathbf{A}\right) \\
&= \frac{\exp\{-\frac{1}{2}\text{VEC}[\mathbf{F}]^T (\mathbf{I}_N \otimes \Sigma_F)^{-1}\text{VEC}[\mathbf{F}]\}}{(2\pi)^{\frac{MN}{2}} \det^{\frac{1}{2}}(\mathbf{I}_N \otimes \Sigma_F)} \\
&= \frac{\exp\{-\frac{1}{2}\sum_{n=1}^{N}\mathbf{f}_n^T \Sigma_F^{-1}\mathbf{f}_n\}}{(2\pi)^{\frac{MN}{2}} \det^{\frac{N}{2}}(\Sigma_F)}
\end{aligned}
\tag{4.6}
$$

where $\text{VEC}[\mathbf{F}]$ denotes the vectorization of $\mathbf{F}$, i.e., $\text{VEC}[\mathbf{F}] = [\mathbf{f}_1^T, \cdots, \mathbf{f}_N^T]^T$. Because Eq.(4.6) can be written in terms of matrix trace, we come up with the likelihood

function

$$p\left(\mathbf{F} \mid \mathbf{A}\right) = \frac{\exp\{-\frac{1}{2}\mathrm{tr}[\mathbf{F}^T(\mathbf{A}\Sigma_S\mathbf{A}^T + \sigma^2\mathbf{I})^{-1}\mathbf{F}]\}}{(2\pi)^{\frac{MN}{2}}\det^{\frac{N}{2}}(\mathbf{A}\Sigma_S\mathbf{A}^T + \sigma^2\mathbf{I})} \qquad (4.7)$$

According to the ML principle, we aim to maximize the likelihood function. To further simplify the objective function in Eq.(4.7), we need the following two lemmas.

**Lemma 4.1.** *(Generalized Matrix Inverse Lemma [23]) Let an $M \times M$ matrix $\mathbf{A}$ be $\mathbf{A} = \mathbf{D} + \mathbf{B}\mathbf{C}\mathbf{B}^H$, where $\mathbf{D}$ is an $M \times M$ positive-definite matrix, $\mathbf{B}$ is an $M \times N$ matrix and $\mathbf{C}$ is an $N \times N$ positive-definite matrix. Then the inverse of $\mathbf{A}$ is*

$$\mathbf{A}^{-1} = \mathbf{D}^{-1} - \mathbf{D}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{B}^H\mathbf{D}^{-1}\mathbf{B})^{-1}\mathbf{B}^H\mathbf{D}^{-1}$$

**Lemma 4.2.** *(Matrix Determinant Lemma [1]) For an $M \times N$ matrix $\mathbf{A}$, and an $N \times M$ matrix $\mathbf{B}$, we have*

$$\det(\mathbf{I}_M + \mathbf{A}\mathbf{B}) = \det(\mathbf{I}_N + \mathbf{B}\mathbf{A})$$

By Lemmas 4.1 and 4.2, the log likelihood function of Eq.(4.7) can be simplified as

$$J_{ml}(\Theta) = \frac{1}{\sigma^2}\mathrm{tr}[(\sigma^2\Sigma_S^{-1} + \mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{F}\mathbf{F}^T\mathbf{A}] - N\log\det(\mathbf{A}^T\mathbf{A} + \sigma^2\Sigma_S^{-1}) \quad (4.8)$$

Therefore, the optimal ML estimate $\hat{\Theta}_{ml}$ can be obtained by solving the following optimization problem:

$$\hat{\Theta}_{ml} = \arg\max_{\Theta} \ J_{ml}(\Theta) \qquad (4.9)$$

Notice that

$$\mathbf{A}^T\mathbf{A} + \sigma^2\Sigma_S^{-1} = \begin{bmatrix} \sigma\Sigma_S^{-\frac{1}{2}} & \mathbf{A}^T \end{bmatrix} \begin{bmatrix} \sigma\Sigma_S^{-\frac{1}{2}} \\ \mathbf{A} \end{bmatrix} \qquad (4.10)$$

For notational simplicity, let $\mathbf{C} = [\mathbf{H} \ \mathbf{A}]^T$ with $\mathbf{H} = \sigma\Sigma_S^{-\frac{1}{2}}$. Performing the QR-decomposition to $\mathbf{C}$ yields

$$\mathbf{C} = \mathbf{Q}_C\mathbf{R}_C = \begin{bmatrix} \mathbf{Q}_H \\ \mathbf{Q}_A^{(ml)} \end{bmatrix} \mathbf{R}_C \qquad (4.11)$$

where $\mathbf{Q}_C$ is an $(M + K) \times K$ column-orthonormal matrix and $\mathbf{R}_C$ is an $(M + K) \times (M + K)$ upper triangular matrix with positive diagonal entries. The matrices $\mathbf{Q}_H$ and $\mathbf{Q}_A^{(ml)}$ are the upper $K \times K$ and lower $M \times K$ block matrices of $\mathbf{Q}_C$ respectively. Substituting Eq.(4.11) into Eq.(4.8), we obtain

$$
\begin{aligned}
J_{ml}(\boldsymbol{\Theta}) &= \frac{1}{\sigma^2} \mathrm{tr}[\mathbf{Q}_A^{(ml)T} \mathbf{F} \mathbf{F}^T \mathbf{Q}_A^{(ml)}] - N \log \det(\mathbf{R}_C^T \mathbf{R}_C) \\
&= \sum_{k=1}^{K} (\frac{1}{\sigma^2} \, \mathbf{q}_{Ak}^{(ml)T} \mathbf{F} \mathbf{F}^T \mathbf{q}_{Ak}^{(ml)} - N \log [\mathbf{R}_C]_{kk}^2) \\
&= \sum_{k=1}^{K} J_k^{(ml)}(\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_k)
\end{aligned}
\tag{4.12}
$$

where each $J_k^{(ml)}(\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_k)$ is defined by

$$
J_k^{(ml)}(\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_k) = \frac{1}{\sigma^2} \, \mathbf{q}_{Ak}^{(ml)T} \mathbf{F} \mathbf{F}^T \mathbf{q}_{Ak}^{(ml)} - N \log [\mathbf{R}_C]_{kk}^2
\tag{4.13}
$$

with $\mathbf{q}_{Ak}^{(ml)}$ denoting the $k$th column vector of $\mathbf{Q}_A^{(ml)}$.

This optimization problem $\max_{\boldsymbol{\Theta}} J_{ml}(\boldsymbol{\Theta})$ is non-convex, so that the methods such as gradient-based optimization often become trapped in local minima, resulting in significant localization errors. Moreover, since we totally have $5K$ unknown continuous parameters, when $K$ is greater than one, performing quantization to search over all grids in the feasible domain is prohibitive. In the following, we will develop a recursive algorithm to efficiently obtain a suboptimal solution of Eq.(4.9).

Now, by the Gram-Schmidt procedure, we can clearly see how $J_k^{(ml)}(\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_k)$ is associated only with the first $k$ columns of the matrix $\mathbf{A}$: For $k = 1$, we have

$$
\mathbf{q}_{C1} = \frac{\mathbf{c}_1}{\| \mathbf{c}_1 \|}
\tag{4.14a}
$$

$$
\mathbf{q}_{A1}^{(ml)}(\boldsymbol{\theta}_1) = \frac{\mathbf{a}_1}{\| \mathbf{c}_1 \|} = \frac{\mathcal{G}(\ell_1)\boldsymbol{\mu}_1}{\sqrt{\mathbf{h}_1^T \mathbf{h}_1 + \boldsymbol{\mu}_1^T \mathcal{G}^T(\ell_1)\mathcal{G}(\ell_1)\boldsymbol{\mu}_1}}
\tag{4.14b}
$$

$$
[\mathbf{R}_C]_{11} = \| \mathbf{c}_1 \| = \sqrt{\mathbf{h}_1^T \mathbf{h}_1 + \boldsymbol{\mu}_1^T \mathcal{G}^T(\ell_1)\mathcal{G}(\ell_1)\boldsymbol{\mu}_1}
\tag{4.14c}
$$

and for $k = 2, \cdots, K$,

$$\mathbf{q}_{Ck} = \frac{\hat{\mathbf{\Pi}}^{(k-1)} \mathbf{c}_k}{\parallel \hat{\mathbf{\Pi}}^{(k-1)} \mathbf{c}_k \parallel} \tag{4.14d}$$

$$\mathbf{q}_{Ak}^{(ml)}(\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_k) = \mathbf{L} \mathbf{q}_{Ck} = \mathbf{L} \frac{\hat{\mathbf{\Pi}}^{(k-1)} \mathbf{c}_k}{\parallel \hat{\mathbf{\Pi}}^{(k-1)} \mathbf{c}_k \parallel} \tag{4.14e}$$

$$[\mathbf{R}_C]_{kk} = \parallel \hat{\mathbf{\Pi}}^{(k-1)} \mathbf{c}_k \parallel = \sqrt{\mathbf{c}_k^T \hat{\mathbf{\Pi}}^{(k-1)} \mathbf{c}_k} \tag{4.14f}$$

where $\mathbf{q}_{Ck}$ denotes the $k$th column vector of $\mathbf{Q}_C$, $\hat{\mathbf{\Pi}}^{(k-1)} = \mathbf{I} - \hat{\mathbf{Q}}_{C(k-1)} \hat{\mathbf{Q}}_{C(k-1)}^T$ is the estimated projection matrix generated by the $(k-1)$ column vectors of $\hat{\mathbf{Q}}_{C(k-1)}$, with $\hat{\mathbf{Q}}_{C(k-1)} = [\hat{\mathbf{q}}_{C1}, \hat{\mathbf{q}}_{C2}, \cdots, \hat{\mathbf{q}}_{C(k-1)}]$, and $\mathbf{L}$ is a "scissor" matrix $\mathbf{L} = [\mathbf{0}_{M \times K}, \mathbf{I}_{M \times M}]$. Considering that both the column vector $\mathbf{q}_{Ak}^{(ml)}$ and $[\mathbf{R}_C]_{kk}$ are only dependent on the first $k$ dipole parameters $(\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_k)$, and every parameter $\boldsymbol{\theta}_k$ is independent of each other, then the optimization sub-function $J_k^{(ml)}$ is only dependent on the first $k$ dipole parameters $(\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_k)$. Therefore, we propose the following recursive schemes to approach Eq.(4.12), which can be expressed in terms of $J_k^{(ml)}$, i.e.,

$$\{\hat{\boldsymbol{\theta}}_1, \cdots, \hat{\boldsymbol{\theta}}_K\} = \arg \max_{\{\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_K\}} J_1^{(ml)}(\boldsymbol{\theta}_1) + J_2^{(ml)}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) + \cdots + J_K^{(ml)}(\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_K)$$

by successively optimizing a series of sub-problems:

$$\hat{\boldsymbol{\theta}}_1 = \arg \max_{\boldsymbol{\theta}_1} J_1^{(ml)}(\boldsymbol{\theta}_1)$$

$$\hat{\boldsymbol{\theta}}_2 = \arg \max_{\boldsymbol{\theta}_2} J_2^{(ml)}(\boldsymbol{\theta}_2 | \hat{\boldsymbol{\theta}}_1)$$

$$\vdots$$

$$\hat{\boldsymbol{\theta}}_K = \arg \max_{\boldsymbol{\theta}_K} J_K^{(ml)}(\boldsymbol{\theta}_K | \hat{\boldsymbol{\theta}}_1 ... \hat{\boldsymbol{\theta}}_{K-1})$$

where the notation $J_K^{(ml)}$ is equivalently defined as $J_K^{(ml)}(\hat{\boldsymbol{\theta}}_1, ..., \hat{\boldsymbol{\theta}}_{K-1}, \boldsymbol{\theta}_K)$. The above procedure can be summarized as follows:

---

1. Initialization: The estimate of $\boldsymbol{\theta}_1 = [\boldsymbol{\ell}_1^T, \boldsymbol{\mu}_1^T]^T$ is obtained by maximizing $J_1^{(ml)}(\boldsymbol{\theta}_1)$

2. Recursion: Suppose that the previous $(k-1)$ estimates $\hat{\boldsymbol{\theta}}_1, \cdots, \hat{\boldsymbol{\theta}}_{k-1}$ are all correct. Then, the estimate of $\boldsymbol{\theta}_k$ is obtained by maximizing $J_k^{(ml)}(\boldsymbol{\theta}_k | \hat{\boldsymbol{\theta}}_1 \cdots \hat{\boldsymbol{\theta}}_{k-1})$

---

## 4.2.2   Rotating-orientation Dipole Model

The same optimization strategy can be applied into the rotating dipole model as well. From Eq.(2.11), we intend to estimate $\mathbf{\Theta} = \{\ell_k, \quad k = 1, 2, \cdots, K\}$ which involves in $\mathbf{G}$, given observed data $\mathbf{F}$. With the alternative assumptions that the dipole moment $\mathbf{V}$ is zero-mean Gaussian distributed with covariance matrices $\mathbf{\Sigma}_V$, the spatio-temporal observation matrix $\mathbf{F}$ can be also described by the Gaussian distribution $\mathcal{N}(0, \mathbf{G}\mathbf{\Sigma}_V\mathbf{G}^T + \sigma^2\mathbf{I})$.

According to the ML principle, we aim to maximize the conditional probability density function $p\,(\mathbf{F}|\mathbf{V})$. After some straightforward manipulations, we come up with the log likelihood function, which is similar to Eq.(4.8), as

$$Y_{ml}(\mathbf{\Theta}) = \frac{1}{\sigma^2}\mathrm{tr}[(\sigma^2\mathbf{\Sigma}_V^{-1} + \mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{F}\mathbf{F}^T\mathbf{G}] - N\log\det(\mathbf{G}^T\mathbf{G} + \sigma^2\mathbf{\Sigma}_V^{-1}) \qquad (4.15)$$

In this case, we define the matrix $\mathbf{C}$ by $[\mathbf{H}\ \mathbf{G}]^T$ instead, with $\mathbf{H} = \sigma\mathbf{\Sigma}_V^{-\frac{1}{2}}$. Performing QR-decomposition to $\mathbf{C}$ yields,

$$\mathbf{C} = \begin{bmatrix} \mathbf{Q}_H \\ \mathbf{Q}_G^{(ml)} \end{bmatrix} \mathbf{R}_C \qquad (4.16)$$

where $\mathbf{Q}_G^{(ml)}$ is an $M \times 3K$ column-orthonormal matrix and $\mathbf{R}_C$ is an $(M + 3K) \times (M + 3K)$ upper triangular matrix with positive diagonal entries. Then by substituting Eq.(4.16) into the ML criterion and ignoring the irrelevant terms, we come up with the QR decomposition-based ML objective function,

$$Y_{ml}(\mathbf{\Theta}) = \frac{1}{\sigma^2}\mathrm{tr}[\mathbf{Q}_G^{(ml)T}\mathbf{F}\mathbf{F}^T\mathbf{Q}_G^{(ml)}] - N\log\det(\mathbf{R}_C^T\mathbf{R}_C) \qquad (4.17)$$

which is similar to Eq.(4.12). Then, the optimal ML estimate $\hat{\mathbf{\Theta}}_{ml}$ in the case of rotating dipoles can be obtained by solving the following optimization problem:

$$\hat{\mathbf{\Theta}}_{ml} = \arg\ \max_{\mathbf{\Theta}}\ Y_{ml}(\mathbf{\Theta}) \qquad (4.18)$$

We again intend to solve this problem by a recursive procedure as in the case of fixed-orientation dipoles. However, the trace here cannot be expressed by separable

columns. Recall that

$$\mathbf{G} = \left[ \begin{array}{cccc} \mathcal{G}(\boldsymbol{\ell}_1), & \mathcal{G}(\boldsymbol{\ell}_2), & \cdots, & \mathcal{G}(\boldsymbol{\ell}_K), \end{array} \right] \tag{4.19}$$

then the structure of $\mathbf{Q}_G^{(ml)}$ can be written as,

$$\mathbf{Q}_G^{(ml)} = \left[ \underbrace{\mathbf{q}_{G1}(\boldsymbol{\ell}_1), \mathbf{q}_{G2}(\boldsymbol{\ell}_1), \mathbf{q}_{G3}(\boldsymbol{\ell}_1),}_{\mathbf{Q}_{G1}(\boldsymbol{\ell}_1)} \quad \underbrace{\mathbf{q}_{G4}(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2), \mathbf{q}_{G5}(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2), \mathbf{q}_{G6}(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2),}_{\mathbf{Q}_{G2}(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2)} \quad \cdots \right]$$

$$= \left[ \begin{array}{cccc} \mathbf{Q}_{G1}(\boldsymbol{\ell}_1), & \mathbf{Q}_{G2}(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2), & \cdots, & \mathbf{Q}_{GK}(\boldsymbol{\ell}_1, \cdots, \boldsymbol{\ell}_K) \end{array} \right] \tag{4.20}$$

where $\mathbf{Q}_{Gk}^{(ml)}$ denotes the $k$th $M \times 3$ block matrix of $\mathbf{Q}_G^{(ml)}$, which is only dependent on the first $k$ dipole parameters $(\boldsymbol{\ell}_1, \cdots, \boldsymbol{\ell}_k)$, i.e., every 3 column vectors of $\mathbf{Q}_G^{(ml)}$ bear a set of common parameter. Therefore, utilizing a "Grouping" policy instead Eq.(4.18) can be expressed in the following format,

$$Y_{ml}(\boldsymbol{\Theta}) = \sum_{k=1}^{K} (\frac{1}{\sigma^2} \text{tr}[\mathbf{Q}_{Gk}^{(ml)T} \mathbf{F} \mathbf{F}^T \mathbf{Q}_{Gk}^{(ml)}] - N \sum_{i=k}^{3k} \log [\mathbf{R}_C]_{ii}^2)$$

$$= \sum_{k=1}^{K} Y_k^{(ml)}(\boldsymbol{\ell}_1, \cdots, \boldsymbol{\ell}_k) \tag{4.21}$$

where each $Y_k^{(ml)}(\boldsymbol{\ell}_1, \cdots, \boldsymbol{\ell}_k)$ is defined by

$$Y_k^{(ml)}(\boldsymbol{\ell}_1, \cdots, \boldsymbol{\ell}_k) = \frac{1}{\sigma^2} \text{tr}[\mathbf{Q}_{Gk}^{(ml)T} \mathbf{F} \mathbf{F}^T \mathbf{Q}_{Gk}^{(ml)}] - N \sum_{i=k}^{3k} \log [\mathbf{R}_C]_{ii}^2 \tag{4.22}$$

Then in order to optimize the original nonconvex objective function in Eq.(4.21), we propose the following recursive schemes of successively optimizing a series of sub-problems:

$$\hat{\boldsymbol{\ell}}_1 = \arg \max_{\boldsymbol{\ell}_1} Y_1^{(ml)}(\boldsymbol{\ell}_1)$$

$$\hat{\boldsymbol{\ell}}_2 = \arg \max_{\boldsymbol{\ell}_2} Y_2^{(ml)}(\boldsymbol{\ell}_2 | \hat{\boldsymbol{\ell}}_1)$$

$$\vdots$$

$$\hat{\boldsymbol{\ell}}_K = \arg \max_{\boldsymbol{\ell}_K} Y_K^{(ml)}(\boldsymbol{\ell}_K | \hat{\boldsymbol{\ell}}_1 ... \hat{\boldsymbol{\ell}}_{K-1})$$

To sum up, the successive ML localization algorithm for rotating dipoles based on QR decomposition is described as follows

---

1. Initialization: The estimate of $\ell_1$ is obtained by maximizing $Y_1^{(ml)}(\ell_1)$

2. Recursion: Suppose that the previous $(k-1)$ estimates $\hat{\ell}_1, \cdots, \hat{\ell}_{k-1}$ are all correct. Then, the estimate of $\ell_k$ is obtained by maximizing $Y_k^{(ml)}(\ell_k | \hat{\ell}_1 \cdots \hat{\ell}_{k-1})$

---

# 4.3 Successive LS Localization Method based on QR Decomposition

So far we have assumed the dipole activities are stochastically known. But what if the statistics of the dipole amplitude $S$ (refer to the fixed-orientation model Eq.(2.10)) is not known? In this section, with the assumption that $S$ is an unknown but deterministic variable, we apply the ML estimator in stochastic signal model, which is finalized by a Least Square (LS) successive localization algorithm based on QR decomposition.

## 4.3.1 Fixed-orientation Dipole Model

Assuming the noise matrix $E$ is zero-mean Gaussian distributed with the covariance matrix $\sigma^2 I$, the conditional probability density function of $F$ can be written as

$$p\left(F|A,S\right) = \frac{\exp\{\frac{1}{2} - \text{tr}[(F - AS)^T(\sigma^2 I)^{-1}(F - AS)]\}}{2\pi^{\frac{MN}{2}}\det^{\frac{N}{2}}(\sigma^2 I)} \qquad (4.23)$$

where $p\left(F|A,S\right)$ denotes the probability density function of $F$ conditioned on $A$ and $S$. From Eq.(4.23) it is quite straightforward to see that the ML estimator is equivalent to the following Least Square (LS) estimator for jointly estimating the location and the amplitude of the dipole:

$$\{\hat{\Theta}, \hat{S}\}_{ls} = \arg \min_{A,S} \text{tr}[(F - AS)^T(F - AS)]$$

$$= \arg \min_{A,S} \parallel F - AS \parallel_F^2 \qquad (4.24)$$

Assuming $\mathbf{A}(\Theta)$ is known and has a full column rank, a solution for the matrix $\mathbf{S}$ that minimizes Eq.(4.24) has the form of

$$\hat{\mathbf{S}}_{ls} = \mathbf{A}^\dagger \mathbf{F} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{F}$$

which can be substituted back into Eq.(4.24). As a result, the LS optimization problem (4.24) is equivalent to the following optimization problem,

$$
\begin{aligned}
\hat{\Theta}_{ls} &= \arg \min_{\mathbf{A}} \| \mathbf{F} - \mathbf{A}(\mathbf{A}^\dagger \mathbf{F}) \|_F^2 \\
&= \arg \max_{\Theta} \| \mathbf{A}\mathbf{A}^\dagger \mathbf{F} \|\|_F^2 \\
&= \arg \max_{\Theta} \operatorname{tr}[\mathbf{F}^T \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{F}]
\end{aligned}
\tag{4.25}
$$

This optimization problem is highly non-convex. Fortunately, the similar strategy that has successfully dealt with the ML estimator in Section 4.2, can be applied to the LS estimator. Applying the QR-decomposition to matrix $\mathbf{A}$ yields

$$\mathbf{A} = \mathbf{Q}_A^{(ls)} \mathbf{R}_A \tag{4.26}$$

where $\mathbf{Q}_A^{(ls)}$ is an $M \times K$ column orthonormal matrix and $\mathbf{R}_A$ is a $K \times K$ upper triangular matrix with positive diagonal entries, so that the objective function in Eq.(4.25) can be rewritten as follows. For simplicity, let

$$J_{ls}(\Theta) = \operatorname{tr}(\mathbf{Q}_A^{(ls)T} \mathbf{F}\mathbf{F}^T \mathbf{Q}_A^{(ls)}) \tag{4.27}$$

Then, we have

$$
\begin{aligned}
J_{ls}(\Theta) &= \sum_{k=1}^{K} \mathbf{q}_{Ak}^{(ls)T} \mathbf{F}\mathbf{F}^T \mathbf{q}_{Ak}^{(ls)} \\
&= \sum_{k=1}^{K} J_k^{(ls)}(\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_k)
\end{aligned}
\tag{4.28}
$$

where $J_k^{(ls)}(\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_k)$ is defined by

$$J_k^{(ls)}(\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_k) = \mathbf{q}_{Ak}^{(ls)T} \mathbf{F}\mathbf{F}^T \mathbf{q}_{Ak}^{(ls)} \tag{4.29}$$

Since the $k$th column of the matrix $\mathbf{A}$ involves only $\boldsymbol{\theta}_k$, for $k = 1, 2, \cdots, K$, according to the Gram-Schmidt process, we can clearly see that the $k$th column vector of $\mathbf{Q}_A^{(ls)}$ depends only on the first $k$ location vectors $\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_k$. The detailed structure of $\mathbf{Q}_A^{(ls)}$ and $\mathbf{R}_A$ can be expressed as, for $k = 1$,

$$\mathbf{q}_{A1}^{(ls)} = \frac{\mathbf{a}_1}{\| \mathbf{a}_1 \|} \tag{4.30a}$$

$$[\mathbf{R}_A]_{11} = \| \mathbf{a}_1 \| \tag{4.30b}$$

and for $k = 2, \cdots, K$, we have

$$\mathbf{q}_{Ak}^{(ls)} = \frac{\hat{\boldsymbol{\Pi}}^{(k-1)}\mathbf{a}_k}{\| \hat{\boldsymbol{\Pi}}^{(k-1)}\mathbf{a}_k \|} = \frac{\hat{\boldsymbol{\Pi}}^{(k-1)}\mathcal{G}(\ell_k)\boldsymbol{\mu}_k}{\| \hat{\boldsymbol{\Pi}}^{(k-1)}\mathcal{G}(\ell_k)\boldsymbol{\mu}_k \|} \tag{4.30c}$$

$$[\mathbf{R}_A]_{kk} = \| \hat{\boldsymbol{\Pi}}^{(k-1)}\mathbf{a}_k \| \tag{4.30d}$$

where $\mathbf{q}_{Ak}^{(ls)}$ denotes the $k$th column vector of $\mathbf{Q}_A$ and $\hat{\boldsymbol{\Pi}}^{(k-1)} = \mathbf{I} - \mathbf{Q}_{A(k-1)}\mathbf{Q}_{A(k-1)}^T$ with $\mathbf{Q}_{A(k-1)} = [\hat{\mathbf{q}}_{A1}, \hat{\mathbf{q}}_{A2}, \cdots, \hat{\mathbf{q}}_{A(k-1)}]$. In other words, the original objective function $J_{ls}(\boldsymbol{\Theta})$ can be decomposed into a sum of $K$ sub-functions, in which the $k$th sub-function $J_k^{(ls)}$ relies only on the parameters of the first $k$ dipoles $\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_k$. It is this property that allows us to successfully utilize a similar recursive scheme as in Section 4.2 again to approximate the LS estimator. Therefore, we propose the following recursive schemes to approach Eq.(4.24), which can be expressed in terms of $J_k^{(ls)}$, i.e.,

$$\{\hat{\boldsymbol{\theta}}_1, \cdots, \hat{\boldsymbol{\theta}}_K\} = \arg \max_{\{\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_K\}} J_1^{(ls)}(\boldsymbol{\theta}_1) + J_2^{(ls)}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) + \cdots + J_K^{(ls)}(\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_K)$$

by successively optimizing a series of sub-problems:

$$\hat{\boldsymbol{\theta}}_1 = \arg \max_{\boldsymbol{\theta}_1} J_1^{(ls)}(\boldsymbol{\theta}_1)$$

$$\hat{\boldsymbol{\theta}}_2 = \arg \max_{\boldsymbol{\theta}_2} J_2^{(ls)}(\boldsymbol{\theta}_2|\hat{\boldsymbol{\theta}}_1)$$

$$\vdots$$

$$\hat{\boldsymbol{\theta}}_K = \arg \max_{\boldsymbol{\theta}_K} J_K^{(ls)}(\boldsymbol{\theta}_K|\hat{\boldsymbol{\theta}}_1...\hat{\boldsymbol{\theta}}_{K-1}) \tag{4.31}$$

where notation $J_k^{(ls)}(\boldsymbol{\theta}_k|\hat{\boldsymbol{\theta}}_1...\hat{\boldsymbol{\theta}}_{k-1})$ is equivalent to notation $J_k^{(ls)}(\hat{\boldsymbol{\theta}}_1,...,\hat{\boldsymbol{\theta}}_{k-1},\boldsymbol{\theta}_k)$.

Actually, we are able to further simply the above problem. Substituting $\mathbf{q}_{Ak}$ Eq.(4.30c) into Eq.(4.29) results in

$$
\begin{aligned}
J_k^{(ls)}(\boldsymbol{\theta}_k|\hat{\boldsymbol{\theta}}_1...\hat{\boldsymbol{\theta}}_{k-1}) &= \frac{\boldsymbol{\mu}_k^T \mathcal{G}^T(\boldsymbol{\ell}_k)\hat{\boldsymbol{\Pi}}^{(k-1)}\mathbf{F}\mathbf{F}^T\hat{\boldsymbol{\Pi}}^{(k-1)}\mathcal{G}(\boldsymbol{\ell}_k)\boldsymbol{\mu}_k}{\boldsymbol{\mu}_k^T \mathcal{G}^T(\boldsymbol{\ell}_k)\hat{\boldsymbol{\Pi}}^{(k-1)}\mathcal{G}(\boldsymbol{\ell}_k)\boldsymbol{\mu}_k} \\
&= \frac{\boldsymbol{\mu}_k^T \boldsymbol{\Psi}_k \boldsymbol{\mu}_k}{\boldsymbol{\mu}_k^T \boldsymbol{\Phi}_k \boldsymbol{\mu}_k}
\end{aligned}
\tag{4.32}
$$

by defining

$$
\boldsymbol{\Psi}_k = \mathcal{G}^T(\boldsymbol{\ell}_k)\hat{\boldsymbol{\Pi}}^{(k-1)}\mathbf{F}\mathbf{F}^T\hat{\boldsymbol{\Pi}}^{(k-1)}\mathcal{G}(\boldsymbol{\ell}_k) \tag{4.33}
$$

$$
\boldsymbol{\Phi}_k = \mathcal{G}^T(\boldsymbol{\ell}_k)\hat{\boldsymbol{\Pi}}^{(k-1)}\mathcal{G}(\boldsymbol{\ell}_k) \tag{4.34}
$$

**Lemma 4.3.** *Given an $N \times N$ symmetric matrix $\mathbf{A}$ and an $N \times N$ positive definite matrix $\mathbf{B}$, $\frac{\boldsymbol{\mu}^T \mathbf{A}\boldsymbol{\mu}}{\boldsymbol{\mu}^T \mathbf{B}\boldsymbol{\mu}}$ is called a generalized Rayleigh quotient, which has the following property,*

$$
\max_{\boldsymbol{\mu}} \frac{\boldsymbol{\mu}^T \mathbf{A}\boldsymbol{\mu}}{\boldsymbol{\mu}^T \mathbf{B}\boldsymbol{\mu}} = \lambda_{\max}(\mathbf{A},\mathbf{B})
$$

*where $\lambda(\mathbf{A},\mathbf{B})$ is the generalized eigenvalue, which is defined by $\lambda(\mathbf{A},\mathbf{B}) = \{\lambda| \det(\mathbf{A} - \lambda\mathbf{B}) = 0\}$, $\boldsymbol{\mu}$ is the corresponding generalized eigenvector and $\lambda_{\max}(\mathbf{A},\mathbf{B})$ denotes the largest generalized eigenvalue.*

To further simplify Eq.(4.32), we need the above Lemma 4.3, whose proof is given in Appendix [14]. As a result, the optimum value of Eq.(4.32) is equal to the maximized largest generalized eigenvalue, i.e.,

$$
\max_{\theta_k} J_k^{(ls)} = \max_{\theta_k} \lambda_{k\max}(\boldsymbol{\Psi}_k, \boldsymbol{\Phi}_k) \tag{4.35}
$$

Note that $\lambda_{k\max}$ is only dependent on the first $k$ dipole parameters $(\boldsymbol{\theta}_1,\cdots,\boldsymbol{\theta}_k)$ as $J_k^{(ls)}$ is. The optimization algorithm described by Eq.(4.31) is thus can be reduced to the following procedure:

1. To localize the first dipole, we search for the optimum location that yields the maximum generalized eigenvalue,

$$\hat{\ell}_1 = \arg \ \max \lambda_1(\ell_1)$$

Meanwhile, the corresponding eigenvector $\hat{\mu}_1$ can be obtained, which is equal to the estimated dipole orientation.

2. To localize the second dipole, we first form the projection matrix $\Pi^{(0)}$ based on $\hat{\theta}_1 = (\hat{\ell}_1, \hat{\mu}_1)$, and then search for its optimum location that yields the maximum generalized eigenvalue,

$$\hat{\ell}_2 = \arg \ \max \lambda_2(\ell_2|\hat{\ell}_1)$$

3. Iteratively, the next dipole locations can be obtained by

$$\hat{\ell}_k = \arg \ \max \lambda_k(\ell_k|\hat{\ell}_1 \cdots \hat{\ell}_{k-1})$$

where the notation $\lambda_k(\ell_k|\hat{\ell}_1, \cdots, \hat{\ell}_{k-1})$ is equivalent to $\lambda_{\max}(\Psi_k, \Phi_k)$. Note that in this thesis, we are only interested in the dipole location $\ell_k$ such that the term $\mu_k$ has been omitted. The above procedure can be summarized as follows,

---

1. Initialization: The estimate of $\theta_1 = [\ell_1^T, \mu_1^T]^T$ is obtained by maximizing $\lambda_1(\theta_1)$

2. Recursion: Suppose that the previous $(k-1)$ estimates $\hat{\theta}_1, \cdots, \hat{\theta}_{k-1}$ are all correct. Then, the estimate of $\theta_k$ is obtained by maximizing $\lambda_k(\theta_k|\hat{\theta}_1 \cdots \hat{\theta}_{k-1})$

---

## 4.3.2  Rotating-orientation Model

Under the rotating dipole model, we assume that the dipole moment $\mathbf{V}$ is unknown but deterministic, and the noise matrix $\mathbf{V}$ is zero-mean Gaussian distributed. Using the same strategy to obtain the LS estimate as in Section 4.3.1, we come up with the following optimization problem:

$$\hat{\Theta}_{ls} = \arg \ \max_{\Theta} \mathrm{tr}[\mathbf{F}^T \mathbf{G} (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{G}] \tag{4.36}$$

Applying the QR-decomposition to matrix $\mathbf{G}$ yields

$$\mathbf{G} = \mathbf{Q}_G^{(ls)} \mathbf{R}_G \tag{4.37}$$

where $\mathbf{Q}_G^{(ls)}$ is an $M \times 3K$ column orthonormal matrix and $\mathbf{R}_G$ is a $3K \times 3K$ upper triangular matrix with positive diagonal entries. Then LS estimate of $\boldsymbol{\Theta}$ in the case of rotating dipoles can be written as

$$\hat{\boldsymbol{\Theta}}_{ls} = \arg \max_{\boldsymbol{\Theta}} Y_{ls}(\boldsymbol{\Theta}) \tag{4.38}$$

with

$$Y_{ls}(\boldsymbol{\Theta}) = \text{tr}[\mathbf{Q}_G^{(ls)T} \mathbf{F} \mathbf{F}^T \mathbf{Q}_G^{(ls)}] \tag{4.39}$$

We again intend to approximately solve this non-linear optimization problem by a recursive procedure. As in Section 4.2.2, the linear independence among columns of $\mathbf{Q}_G^{(ls)}$ is not available. Therefore, we need to utilize the "Grouping" strategy to rewrite $\mathbf{Q}_G^{(ls)}$ in the following formation

$$\mathbf{Q}_G^{(ls)} = \left[ \ \mathbf{Q}_{G1}^{(ls)}(\boldsymbol{\ell}_1), \ \ \mathbf{Q}_{G2}^{(ls)}(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2), \ \cdots, \ \mathbf{Q}_{GK}^{(ls)}(\boldsymbol{\ell}_1, \cdots, \boldsymbol{\ell}_K) \ \right] \tag{4.40}$$

where $\mathbf{Q}_{Gk}^{(ls)}$ denotes the $k$th $M \times 3$ block matrix of $\mathbf{Q}_G^{(ls)}$, which is only dependent on the first $k$ dipole parameters $(\boldsymbol{\ell}_1, \cdots, \boldsymbol{\ell}_k)$, i.e., every 3 column vectors of $\mathbf{Q}_G^{(ls)}$ bear a common set of parameters. As a result, Eq.(4.39) can be expressed as

$$\begin{aligned} Y_{ls}(\boldsymbol{\Theta}) &= \sum_{k=1}^{K} \text{tr}[\mathbf{Q}_{Gk}^{(ls)T} \mathbf{F} \mathbf{F}^T \mathbf{Q}_{Gk}^{(ls)}] \\ &= \sum_{k=1}^{K} Y_k^{(ls)}(\boldsymbol{\ell}_1, \cdots, \boldsymbol{\ell}_k) \end{aligned} \tag{4.41}$$

where each addend function is defined by

$$Y_k^{(ls)}(\boldsymbol{\ell}_1, \cdots, \boldsymbol{\ell}_k) = \text{tr}[\mathbf{Q}_{Gk}^{(ls)T} \mathbf{F} \mathbf{F}^T \mathbf{Q}_{Gk}^{(ls)}] \tag{4.42}$$

In order to maximize Eq.(4.41), we need to simultaneously search over the whole parameters range, which is computationally prohibitive. As an approximation, we

propose the following recursive schemes by successively optimizing a series of sub-problems:

$$\hat{\ell}_1 = \arg \max_{\ell_1} Y_1^{(ls)}(\ell_1)$$

$$\hat{\ell}_2 = \arg \max_{\ell_2} Y_2^{(ls)}(\ell_2|\hat{\ell}_1)$$

$$\vdots$$

$$\hat{\ell}_K = \arg \max_{\ell_K} Y_K^{(ls)}(\ell_K|\hat{\ell}_1...\hat{\ell}_{K-1})$$

We summarize the successive LS algorithm based on QR decomposition for rotating dipole as follows:

---

1. Initialization: The estimate of $\ell_1$ is obtained by maximizing $Y_1^{(ls)}(\ell_1)$

2. Recursion: Suppose that the previous $(k-1)$ estimates $\hat{\ell}_1, \cdots, \hat{\ell}_{k-1}$ are all correct. Then, the estimate of $\ell_k$ is obtained by maximizing $Y_k^{(ls)}(\ell_k|\hat{\ell}_1 \cdots \hat{\ell}_{k-1})$

---

# Chapter 5

# Experimental Results and Discussions

In this chapter, we provide several numerical examples to show the performance of our closed-form solution and the QR decomposition-based methods for dipole localization. We also compare our methods with other algorithms, such as RAP-MUSIC, which is known to have good performance.

## 5.1 Experiment Set-up

Our experiments are carried out on simulated EEG data. In all simulations we use the infinite homogeneous head model (see page 10). Similar to the experiments in [26], we choose the head sphere radius $\beta = 10cm$ centered at the origin with conductivity $\zeta = 0.33s/m$. The simulated EEG is measured by $M$ number of sensors, positioned upon guideline as shown in Fig. 3.1. In addition, every channel of the EEG observation is composed of $N$ samples. According to each experiment, $M$ and $N$ will be assigned different values.

Assume the added noise is uncorrelated white Gaussian with the variance $\sigma^2$.

Then the Signal to Noise Ratio (SNR) is defined by

$$\text{SNR} = 10\log_{10}\frac{SignalPower}{NoisePower} = 10\log_{10}\frac{\text{tr}(\mathbf{G}\mathbf{V}\mathbf{V}^T\mathbf{G}^T)}{M\sigma^2}$$

where $\mathbf{G}$ is the gain matrix and $\mathbf{V}$ is the matrix describing the dipole moments as defined by Eq.(2.9). Given an SNR, the experiment is repeated $P = 500$ trials with independent noise realizations. At each time, we set the dipole parameter randomly so that an average result is calculated over all $P$ trials. To examine the performance of each localization method, we define the dipole localization error as the average root mean squared error between its ideal and the estimated locations, i.e.,

$$\delta = \frac{1}{PK}\sum_{p=1}^{P}\sum_{k=1}^{K}\parallel \ell_k - \hat{\ell}_k \parallel \tag{5.1}$$

Then, the localization error (decimeter) will be evaluated at different SNRs(dB).

## 5.2   Closed-form Solution for A Single Rotating Dipole

*Example* 1. In the first simulation, for a single rotating dipole, we compare the performance of our Unique Closed-form (UC) localization method with those of the ML, LS and MUSIC algorithms. As suggested in Proposition 3 (see page 23), at least $M = 4$ sensors are needed. Then we set the sample length to be $N = 50$. The simulation result is shown in Fig. 5.1, from which we can see that our UC algorithm achieves a much smaller estimation error than MUSIC or LS for all SNRs. We also observe that although our UC method is inferior to ML at low SNRs, it gradually gains better performance that is very close to the ML method when SNR is increasing. Moreover, the ML algorithm, as well as the MUSIC and the LS method, need an exhaustive 3-dimensional search over all possible locations in the head sphere (the number of all possible grid points that need to be searched is proportional to
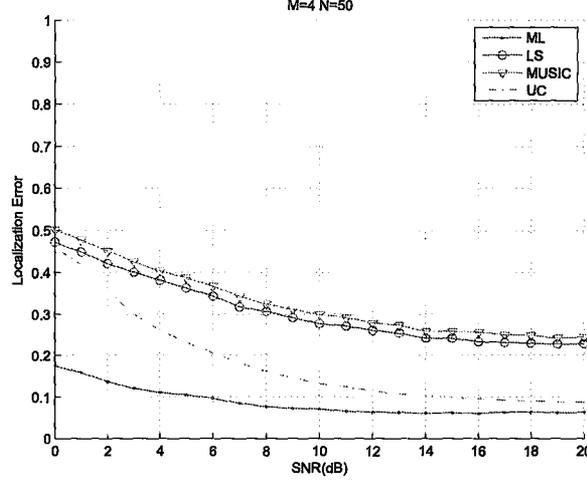
Figure 5.1: Performance of the unique closed-form solution for 1 rotating dipole

$D^3$ with $D$ denoting the grid density in each dimension), thus it is quite a time-consuming scanning process for localizing a dipole. On the other hand, for our UC method, which has instantaneous closed form solutions, the computational complexity is much smaller.

# 5.3　Successive-ML and Successive-LS Algorithm

## 5.3.1　Fixed-orientation Model

In this section, we compare the performance of our successive-ML and successive-LS algorithms based on QR decomposition with that of the well-established RAP-MUSIC method (refer to Section 2.3) for the fixed-orientation dipole model. In this case, since each dipole orientation $\boldsymbol{\mu}_k = [\mu_{kx}, \mu_{ky}, \mu_{kz}]^T$ for $k = 1, \cdots, K$ is a unit norm vector, it can be clearly represented in spherical coordinates, shown in Fig.(5.2) as $[\sin \phi_k \cos \varphi_k, \; \sin \phi_k \sin \varphi_k, \; \cos \phi_k]^T$ by only two parameters, the polar angle $\phi_k$ and the azimuthal angle $\varphi_k$.

44
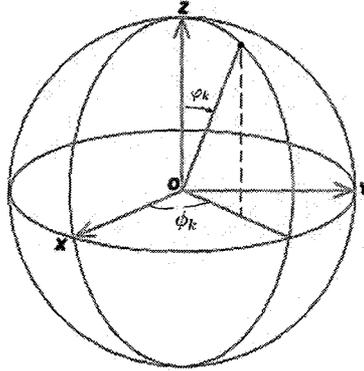
Figure 5.2: Spherical Coordinate

*Example* 2. Consider the fixed-orientation model with two dipoles. There are totally 10 parameters to be estimated; i.e., $\{\boldsymbol{\theta}_k = [\boldsymbol{\ell}_k^T, \phi_k, \varphi_k]^T, \quad k = 1, 2\}$. According to [26], the necessary number of sensors that provides sufficient observations to guarantee a valid localization is greater than 10. More generally, the number of sensors $M$ should be greater than the number of unknown parameters $5K$ in the fixed-orientation case. Based on this statement, we first choose $M = 11$ and set the sample numbers to be $N = 50$. As can be seen in Fig. 5.3, RAP-MUSIC almost fails due to insufficient number of sensors to accurately estimate the signal subspace, while our methods can still work. In addition, since we take advantage of extra information on the dipole moment and on the noise distributions, we observe that the performance of our successive-ML algorithm achieves better performance than that of the successive-LS algorithm.

*Example* 3. We then examine how their performances vary as we change the number of sensors, by increasing the number of sensors to be $M = 20$, which is adequately large, but keeping the sample length to be still the same. From Fig. 5.4 we can
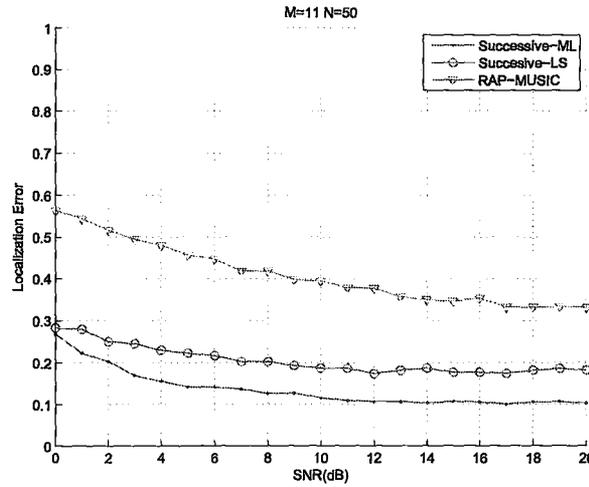
Figure 5.3: Performance comparison of Successive-ML, Successive-LS and RAP-MUSIC for 2 fixed-orientation dipoles using M=11 sensors

observe that the performance of RAP-MUSIC is much better than that with $M = 11$ sensors. But, it is still worse than our two algorithms especially for low SNRs. This is what we have expected: although the RAP-MUSIC has a similar structure to that of our successive algorithms, when the number of snapshots is not sufficiently large, the error resulting from estimating the signal-subspace basis would lead to a significant estimation error of dipole locations.

*Example* 4. We now examine how the localization performance of our methods and the RAP-MUSIC method change with the different number of samples. We fix the sensor number to be $M = 20$ and reduce the sample number to $N = 20$. The simulation result is shown in Fig. 5.5, from which we can conclude that when the number of sensors $M$ is large enough, the sample number $N$ does not cause much performance difference except under low SNR for the three recursive algorithms. This suggests that when designing an EEG array, we should utilize as many sensors as we
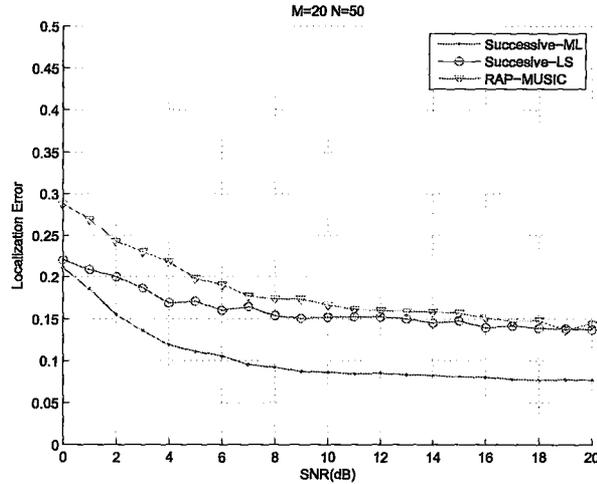
Figure 5.4: Performance comparison of Successive-ML, Successive-LS and RAP-MUSIC for 2 fixed-orientation dipoles using M=20 sensors (N=50)

can, particularly for the localization of multiple dipoles. However, our successive methods require the minimum number of sensors for an effective dipole localization, compared to the subspace methods.

## 5.3.2 Rotating-orientation Model

*Example* 5. For further illustration, we present the performance in the case of two rotating dipoles. Theoretically [26], in order to sufficiently estimate the 6 location parameters $\{\ell_k, \quad k = 1, 2\}$, the MUSIC requires at least $M = 7$ sensors. Let there be still $N = 50$ snapshots. Fig. 5.6 shows the performance comparison between the successive-ML, successive-LS and the MUSIC method. It is apparently that MUSIC is inferior to our successive methods due to insufficient sensors. Both of our methods can achieve efficient and acceptable estimation results, and the performance of the successive-ML is superior to that of the successive-LS method, as a result of using the a priori knowledge on the dipole and noise statistics.
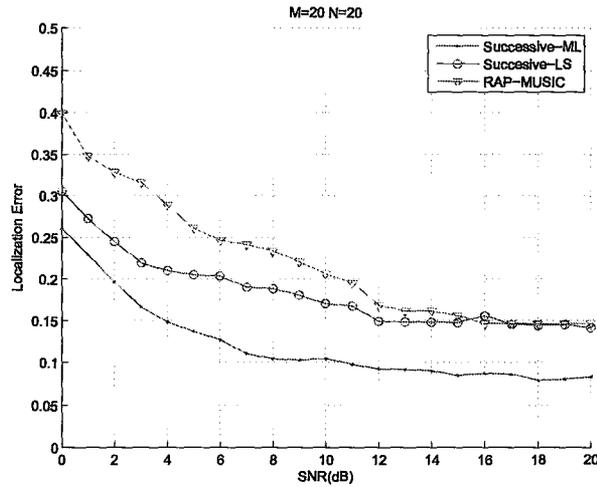
Figure 5.5: Performance comparison of Successive-ML, Successive-LS and RAP-MUSIC for 2 fixed-orientation dipoles using M=20 sensors (N=20)
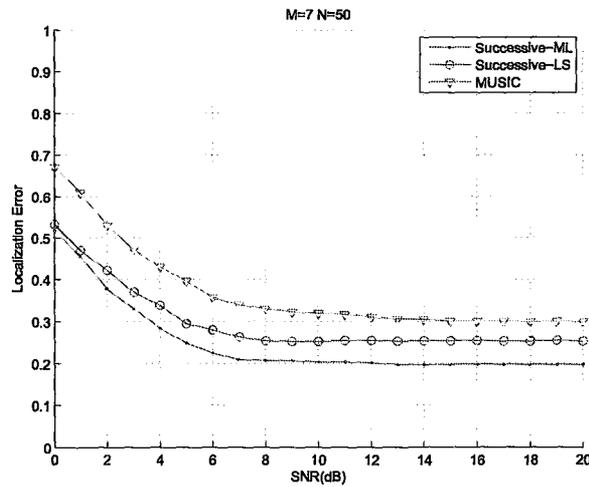


Figure 5.6: Performance comparison of Successive-ML,Successive-LS and MUSIC for 2 rotating dipoles

# Chapter 6

# Conclusion and Future Work

## 6.1 Conclusion

In this thesis, we have considered the inverse problem of the forward model on computing the scalp EEG at a finite set of sensors from multiple dipole sources, which is known as the dipole localization problem. It is observed that the geometric structure of the EEG array plays a crucial role in ensuring a unique solution for this problem.

When the second-order statistic of the EEG observation is available, we first present a necessary and sufficient condition in the model of a single rotating dipole, that guarantees its location to be uniquely determined. In addition, a geometrical structure of an array with minimum number of sensors is designed to efficiently derive a closed-form solution for the dipole's location. Although the analysis on a single dipole has not been generalized into multiple dipoles, the result obtained provides us a guideline on how to set up an EEG sensor array for the localization of multiple dipoles.

In the case of multiple dipoles, we propose an efficient localization algorithm based on QR decomposition. Depending on whether or not the probability density functions of the dipole amplitude and the noise are available, we utilize the ML or the LS as the criterion to develop a unified successive localization algorithm, so that solving

the original multi-dipole optimization problem can be approximated by successively solving a series of single-dipole optimization problems. Numerical simulations show that our methods have much smaller estimation errors than the existing RAP-MUSIC method under non-ideal situations such as low SNR with small number of EEG sensors. Although the whole thesis focused on the dipole-localization problems, the strategy we have developed can be applied into a more general family of nonlinear optimization problems, in which the objective function has the following two key properties:

1. It can be formulated into a function with respect to some variable matrix $\mathbf{A}$; i.e., $\mathcal{F}(\mathbf{A})$, where $\mathbf{A}$ can be written as a block matrix, $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_P]$ with $\mathbf{A}_i$ and $\mathbf{A}_j$ (for $i, j = 1, 2, \cdots, P$, and $i \neq j$) being independent. In other words, the parameters included in $\mathbf{A}_i$ do not overlap with those contained in $\mathbf{A}_j$.

2. The function $\mathcal{F}(\mathbf{A})$ can be decomposed into a sum of a series of sub-functions; i.e.,

$$\begin{aligned}
\mathcal{F}(\mathbf{A}) &= \mathcal{F}_1(\mathbf{A}_1) + \mathcal{F}_2(\mathbf{A}_1, \mathbf{A}_2) + \cdots + \mathcal{F}_P(\mathbf{A}_1, \cdots, \mathbf{A}_P) \\
&= \sum_{p=1}^{P} \mathcal{F}_p(\mathbf{A}_1, \cdots, \mathbf{A}_p)
\end{aligned} \tag{6.1}$$

where the QR decomposition is a very useful tool for fulfilling this procedure.

Then, solving the original optimization problem

$$\hat{\mathbf{A}} = \arg \min_{\mathbf{A}} \text{ (or max) } \mathcal{F}(\mathbf{A}) \tag{6.2}$$

can be approximated by successively solving a series of the following sub-optimization

problems:

$$\hat{A}_1 = \arg\min_{A_1} \text{ (or max) } \mathcal{F}_1(A_1)$$

$$\hat{A}_2 = \arg\min_{A_2} \text{ (or max) } \mathcal{F}_2(A_2|\hat{A}_1)$$

$$\vdots$$

$$\hat{A}_P = \arg\min_{A_P} \text{ (or max) } \mathcal{F}_P(A_P|\hat{A}_1, \cdots, \hat{A}_{P-1})$$

where the notation $\mathcal{F}_p(A_p|\hat{A}_1, \cdots, \hat{A}_{p-1})$ is equivalent to $\mathcal{F}_p(\hat{A}_1, \cdots, \hat{A}_{p-1}, A_p)$.

## 6.2   Future Work

The usefulness of our inverse method depends on how accurately the estimated dipole locations approximate to those of actual sources. The main factors that affect localization accuracy are: head-modeling errors, measurement-location errors and noise-modeling errors. Accordingly, some future work can be addressed in the following areas:

1. Solutions for multiple dipole locations are very sensitive to the above factors. Therefore a sensitivity analysis is needed on how the variation of the localization error can be apportioned, quantitatively to different modeling variations.

2. More complicated head model need to be adopted to accurately portray various structures inside the brain. It is even possible to develop an individual realistic head model for each patient or subject. Therefore, corresponding localization algorithms can be developed for the modified forward model.

3. The model established in this thesis assume the background noise to be white. However, the more realistic case is for noise to be spatially and temporally correlated. Therefore, development of an effective localization algorithm in complicated noise environment is another future target to be fulfilled.

# Appendix A

## A.1 Proof of Lemma 4.3

Given an $N \times N$ symmetric matrix $\mathbf{A}$ and an $N \times N$ positive definite matrix $\mathbf{B}$, the generalized Rayleigh quotient is defined as

$$R(\boldsymbol{\mu}) = \frac{\boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu}}{\boldsymbol{\mu}^T \mathbf{B} \boldsymbol{\mu}} \tag{A.1}$$

Note that if the direction of vector $\boldsymbol{\mu}$ is fixed, its amplitude does not affect the value of $R(\boldsymbol{\mu})$. Without loss of generality, let's consider the property of $R(\boldsymbol{\mu})$ on the ellipsoid $\boldsymbol{\mu}^T \mathbf{B} \boldsymbol{\mu} = 1$. Firstly, rewrite Eq.(A.1) as

$$(\boldsymbol{\mu}^T \mathbf{B} \boldsymbol{\mu}) R(\boldsymbol{\mu}) = \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu} \tag{A.2}$$

Then the gradient of both sides can be obtained,

$$\nabla(\boldsymbol{\mu}^T \mathbf{B} \boldsymbol{\mu}) R(\boldsymbol{\mu}) + (\boldsymbol{\mu}^T \mathbf{B} \boldsymbol{\mu}) \nabla R(\boldsymbol{\mu}) = \nabla(\boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu})$$

$$(2\mathbf{B}\boldsymbol{\mu}) R(\boldsymbol{\mu}) + (\boldsymbol{\mu}^T \mathbf{B} \boldsymbol{\mu}) \nabla R(\boldsymbol{\mu}) = 2\mathbf{A}\boldsymbol{\mu} \tag{A.3}$$

Substituting $\boldsymbol{\mu}^T \mathbf{B} \boldsymbol{\mu} = 1$ yields

$$\nabla R(\boldsymbol{\mu}) = 2[\mathbf{A}\boldsymbol{\mu} - R(\boldsymbol{\mu})\mathbf{B}\boldsymbol{\mu}] \tag{A.4}$$

In the following, we will show that $\boldsymbol{\mu}$ being the stationary point of $R(\boldsymbol{\mu})$ is the necessary and sufficient condition of $\boldsymbol{\mu}$ being the eigenvector of $\mathbf{A}\boldsymbol{\mu} = \lambda \mathbf{B}\boldsymbol{\mu}$ corresponding to $\lambda$.

(1) sufficiency: Because $\mu$ is the stationary point of $R(\mu)$, $\nabla R(\mu) = 0$. According to Eq.(A.4),

$$\mathbf{A}\mu - R(\mu)\mathbf{B}\mu \tag{A.5}$$

which shows that $\mu$ is the eigenvector of $\mathbf{A}\mu = \lambda\mathbf{B}\mu$, and its corresponding eigenvalue is equal to $R(\mu)$.

(2) necessity: Left-handed multiplying $\mathbf{A}\mu = \lambda\mathbf{B}\mu$ by $\mu^T$ results in $\mu^T\mathbf{A}\mu = \lambda\mu^T\mathbf{B}\mu$,

$$\lambda = \frac{\mu^T\mathbf{A}\mu}{\mu^T\mathbf{B}\mu} = R(\mu) \tag{A.6}$$

which is to say $\mathbf{A}\mu = R(\mu)\mathbf{B}\mu$. Combining Eq.(A.4), we see that $\nabla R(\mu) = 0$.

Therefore, Lemma 4.3 holds

$$\max_{\mu} \frac{\mu^T\mathbf{A}\mu}{\mu^T\mathbf{B}\mu} = \lambda_{\max}(\mathbf{A}, \mathbf{B})$$

Here, we introduce the computation of the "Generalized Eigenvalue" $\lambda$. $\lambda$ is defined by

$$\lambda(\mathbf{A}, \mathbf{B}) = \{\lambda|\det(\mathbf{A} - \lambda\mathbf{B}) = 0\} \tag{A.7}$$

and can be computed as

$$\det(\mathbf{A} - \lambda\mathbf{B}) = 0 \tag{A.8}$$

$$\det(\mathbf{B}^{-\frac{1}{2}}\mathbf{A}\mathbf{B}^{-\frac{1}{2}} - \lambda\mathbf{I}) = 0 \tag{A.9}$$

$$\lambda = eig(\mathbf{B}^{-\frac{1}{2}}\mathbf{A}\mathbf{B}^{-\frac{1}{2}}) \tag{A.10}$$

# Bibliography

[1] A.G. Akritas, E.K. Akritas, and G.I. Malaschonok. Various proofs of sylvester's (determinant) identity. *Mathematics and Computers in Simulation, Elsevier*, 42, No. 4:585–593, 1996.

[2] Avner Amir. Uniqueness of the generators of brain evoked potential maps. *IEEE Trans. Biomed. Eng.*, 41:2461 – 2462, 1994.

[3] X. Bai and B. He. Estimation of number of independent brain electric sources from the scalp EEGs. *IEEE Trans. Biomed. Eng.*, 53:1883 – 1892, 2006.

[4] S. Baillet, J. C. Mosher, and R. M. Leahy. Electromagnetic brain mapping. *IEEE Signal Processing Magazine*, pages 14–27, November 2001.

[5] D.A. Brody, F. H. Terry, and R. E. Ideker. Eccentric dipole in a sherical medium: Generalized expression for surface potentials. *IEEE Trans. Biomed. Eng.*, 20:141–143, 1973.

[6] H. Buchner, T. D. Waberski, M. Fuchs, H. A. Wischmann, M. Wagner, and R. Drenckhahn. Comparison of realistically shaped boundary-element and spherical head models in source localization of early somatosensory evoked potentials. *Brain Topography*, 8:137–143, Dec. 1995.

[7] H. I. Chang and O. A. Kwang. Enhancing accuracy in MEG focal source localization. *IEEE Trans. Magnetics*, 42:1387–1391, Apr. 2006.

[8] R. Cooper, J. W. Osselton, and J. C. Shaw. *EEG Technology*. London, England: Butterworths.

[9] B. N. Cuffin. EEG dipole source localization. *IEEE Engineering in Medicine and Biology*, September/October 1998.

[10] A. Dogandzic and A. Nehorai. Localization of evoked electric sources and design of EEG/MEG sensor arrays. In *Proc. Ninth IEEE SP Workshop Statistical Signal and Array Processing*, pages 228–231, Sept 1998.

[11] A. Dogandzic and A. Nehorai. Estimating evoked dipole responses in unknown spatially correlated noise with EEG/MEG arrays. *IEEE Trans. Signal Processing*, 48:13–25, 2000.

[12] M. Gamalainen, R. Hari, R. Ilmoniemi, J. Knuutila, and O. Lounasmaa. MEG theory instrumentation and applications to the noninvasive study of human brain function. *Rev. Mod. Phys.*, 65:413–497, 1993.

[13] T. Gasser, J. Mochs, and W. Kohler. Amplitude probability distribution of noise for flash-evoked potentials and robust response estimates. *IEEE Trans. Biomed. Eng.*, 33:579–584, 1986.

[14] G. H. Golub and C. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, 1983.

[15] G. H. Golub and V. Pereyra. The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. *SIAM J. Num. Anal.*, 10:413–432, 1973.

[16] A. Graham. *Kronecker Products and Matrix Calculus: with applications*. Ellis Horwood Limited, England, 1981.

[17] R. M. Gulrajani. *TBioelectricity and Biomagnetism*. Wiley, New York, 1998.

[18] David Gutierrez and A. Nehorai. Estimating brain conductivities and dipole source signals with EEG arrays. *IEEE Trans. Biomed. Eng.*, 51, Issue 12:2113 – 2122, 2004.

[19] http://www.estsp.pt.

[20] http://www.ym.edu.tw.

[21] M. Huang. Multi-start downhill simplex method for spatio-temporal source localization in MEG. *Electroenceph Clin Neurophysiol*, 108:32–44, 1998.

[22] Y. D. Huang and M. Barkat. A dynamic programming algorithm for the maximum likelihood localization of multiple sources. *IEEE Trans. Antennas and Propagation*, 40, No. 9, 1992.

[23] A. B. Israel and T. Greville. *Generalized Inverses: theory and applications.* New York: Wiley, 1974.

[24] H. Krim and M. Viberg. Two decades of array signal processing research. *IEEE Signal Processing Magazine*, pages 67–94, July 1996.

[25] A. Lapidoth and P. Narayan. Reliable communication under channel uncertainty. *IEEE Trans. Information Theory*, 44:2148–2177, Oct. 1998.

[26] J. C. Mosher, P. S. Lewis, and R. M. Leahy. Multiple dipole modeling and localization from spatio-temporal MEG data. *IEEE Trans. Biomed. Eng.*, 39:541–57, 1992.

[27] J. C. Mosher, P. S. Lewis, and R. M. Leahy. EEG and MEG: forward solutions for inverse methods. *IEEE Trans. Biomed. Eng.*, 46, No. 3:245–259, 1999.

[28] J. C. Mosher, P. S. Lewis, and R. M. Leahy. Source localization using recursively applied and projected (RAP) MUSIC. *IEEE Trans. Signal Processing*, 47:332–340, Feb. 1999.

[29] P.Lancaster and M. Tismenesky. *The Theory of Matrices*. New York: Academic Press, 1985.

[30] H. V. Poor. *An introduction to signal detection and estimation*. Springer-Verlag, 1988.

[31] R. Roy and T. Kailath. ESPRIT - estimation of signal parameters via rotational invariance techniques. *IEEE Trans. on ASSP*, 37:984–995, 1989.

[32] M. Scherg. *Fundamentals of dipole source potential analysis. In: Auditory Evoked Magnetic Field and Potentials*. Karger, Advances in audiology, vol 6, 1989.

[33] M. Scherg and D. Von Cramon. Two bilateral sources of the late AEP as identified by a spatio-temporal dipole model. *Electroenceph Clin Neurophysiol*, 62:32–44, Jan. 1985.

[34] R.O. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Trans. Antennas Propagation*, AP-34:276–280, 1986.

[35] J. C. Shaw and M. Roth. Potential distribution analysis, i. a new technique for the analysis of electrophysilogical phenomena. *Electroenceph Clin Neurophysiol*, 7:273–284, October 1955.

[36] H. L. Van Trees. *Detection, Estimation and Modulation Theory, Part I*. Wiley, New York, 1968.

[37] K. Uutela, M. Hamalaainen, and R. Salmelin. Global optimization in the localization of neuromagnetic sources. *IEEE Trans. Biomed. Eng.*, 45:139–155, Jun. 1998.