Spatial-temporal Source Reconstruction for Magnetoencephalography

Jing Kan

Submitted for the degree of Doctor of Philosophy

University of York

Department of Computer Science

March. 2011

Abstract

Magnetoencephalography (MEG) is a new non-invasive technique for the functional imaging of the human brain. It has been widely used in both research and clinical applications, for it has several superior properties, including a high-temporal resolution with no interference from the bone or the head-like fluid to the signal spatial transformation.

In this thesis, we aim to develop a framework for MEG spatial-temporal current course reconstruction by introducing classical methods from the pattern recognition theory into medical imaging. These applications provide a new angle for research in MEG source reconstruction with the solution for source reconstruction at a single point, and improvements of the reconstruction on spatially and temporally. The whole thesis is based on three topics, which are designed to be parts of an integrated reconstruction process, and each of them are interrelated, rather than independent from each other.

We firstly introduce the source reconstruction method at a single time point using the basis function extraction. In light of the assumption that the Laplacian eigenvectors of mesh can be the analogous to the basis functions that represent the cortex mesh; we build a new model to describe the current source that is distributed on each mesh vertex. This model consists of analogous basis functions and unknown weighted coefficients. In terms of experiment results, this algorithm shows good reconstructed property to the single stimulus, as well as the supercial stimulus on the cortical surface.

Secondly, with respect to the spatial reconstructed sources by basis function method from the last topic, we build a new solution for improving the spatial-resolution of MEG source reconstruction at a single time

point by introducing a classical method (the Bayesian super-resolution method) from the pattern recognition theory. Although the approach is designed based on the reconstruction from basis functions, it is also feasible for other spatial reconstruction methods to improve the spatial-resolution. From the numerical experiment results, it is apparent that the spatial resolution has been effectively improved.

Then, the MEG measurement system in the temporal field is assumed to be a linear dynamic system where the classical methods, Kalman filter and Kalman smoother, are applied as the solution for the estimation of source in time course. The Kalman filter is used to estimate the dynamic state while the Kalman smoother is applied for correcting the source distribution of the hidden state with the EM algorithm. This approach shows superior performance to solve the inverse problem. It extends the improvement in source reconstruction using the temporal field.

We construct the synthetic data as well as apply the real MEG data throughout all the experimental test of my work.

In summary, this thesis builds three algorithms, which aim to reconstruct the MEG source distribution on spatial and temporal field respectively aided by methods from pattern recognition. This work provides a new angle of using the pattern recognition theory for MEG source reconstruction. Meanwhile, we also explore a new direction for applying the theory of pattern recognition. This work not only provides a good integration between these two fields, but also encourage future interactions.

Contents

List of Figures

I would like to dedicate this thesis to my dearest parents, my dearest Zed, and the stubborn me.

Acknowledgements

Grateful thanks to Prof. Richard Wilson who is my ideal PHD supervisor. I have learnt how to research with precise and critical attitude from you. A special thank you for giving me the sage advice, insightful help and patient encouragement when I was stuck in two biggest difficulties nearly desperate during my study and life in 2008 and 2010. Without your unconditional help and support, I wasn't able to conquer those difficulties and continue my study with rapid speed.

Grateful thanks to Prof. Gary Green who always give me the steadfast support, profound advice and practical opportunities during my MEG study. Thanks all the colleagues and friends in YNiC, Will Woods, Rebecca Millman, Andre Gouws, Heidi Baseler, who patiently and kindly offered excellent technical support and help.

Grateful thanks to my dear parents who devote all their effort to my education, and actively supported me in my determination to find and realise my potential. A very special thank you for unconditionally providing me financial support generously with your whole-life saving during my 5 years PHD study in UK.

Grateful thanks to Zeyi He for inspiring and encouraging me to chase my PHD dream 5 years ago, and been looking after me during my PHD study in UK. I could not survive without your practical and emotional support, especially during the biggest difficulties in June. 2008 and April. 2010. I am very grateful for the time, encouragement, intelligence you dedicated to my study, and a biggest thank you for your tolerance to my bad temper.

A special thank you to Jin Yan, Bihui Xu, Claire Schofield, Jun Yue, Rebecca Millman, Ming Luo, Gul-E-Saman, Oswald Aldrian, Xinya Zhu and Yitao Guo for being so supportive in myriad ways to complete my thesis.

Declaration

I declare that all the work in this thesis is solely my own, except where attributed and cited to another authors.

Chapter 1

Introduction

Since it was first introduced by the University of Illinois physicist D. Cohen in 1968, Magnetoencephalography (MEG) has become a well-established and non-invasive technique for mapping the brain activity by recording the magnetic fields from the head. The electrical currents of the human brain generated by the neuronal activities provide measurements of magnetic eld (which are extremely weak in the range between 10^{-12} and 10^{-15} FT). MEG technique has several superior properties, such as a high temporal resolution up to several *kHZ* (up to $1ms$). Additionally, it is a non-invasive brain mapping technique which is not affected by the bone or head-like fluid in the process of signal spatial transformation. All these advantages render MEG a competitive brain imaging technique that can be used for scientific research and in clinical application throughout the whole life of patients (Preissl, 2005).

So far, there are a number of widely applied approaches to address the classical problem of reconstruction of the current sources in the brain from the limited number of MEG measurements in this field, including beamforming method and minimum-norm method. However, since the MEG source reconstruction problem is fundamentally an ill-posed inverse problem which is technically unsolvable, the existing approaches may obtain a reasonable reconstruction under some particular circumstances, but such results by no means may represent the true image at all times. Therefore, it is important to keep exploring better solutions for the MEG source reconstruction problem.

In our study, we sufficiently utilize the knowledge of pattern recognition in solving the problem of MEG source reconstruction, and provide a new perspective of solving this reconstruction problem. In our research, source reconstruction is based on the cortical surface mesh which is extracted from the corresponding MRI scan. Meanwhile, instead of assuming dipoles as the conventional sources, we innovatively assume that the current sources are distributed on each vertex of the mesh of the cortical surface at a single time point. In addition to these novel assumptions in spatial field which was explained previously, we also expanded the solution in temporal field as a linear dynamic system. During the process of the research in spatial field, multiple algorithms from computer graph theory, pattern recognition, and computer vision are applied here, such as *basis function*, *super-resolution*, *normalized cut*. In order to obtain the solution for temporal field, the classical algorithms Kalman filter as well as Kalman smoother were used.

The structure of the thesis is shown as follows: firsly, we discuss the basic knowledge of the eld of medical imaging such as where the origin of different technologies stems from. How they provided the fitting circumstance for the development of MEG was covered in the *Chapter 2 Literature review*. The MEG development history as well as the machine properties will be introduced at the same time. Also, the basic properties of the algorithms used in the later research are introduced in this part. Secondly, in *Chapter 3 Basis Functions Source Model Applied to MEG Source Reconstruction*, we focus on source reconstruction at spatial resolution at a single time point. The assumption that current sources are distributed on the cortical mesh vertices are made firstly. Following this, the geometry of the mesh is analysed with the basic functions produced as the mesh representation. And, a global basis function source reconstruction model will also be illustrated with a discussion of the relevant results. Thirdly, in *Chapter 4 Spatial Improvement of MEG source reconstruction with Bayesian Super-resolution*, we apply super-resolution algorithm in order to obtain a high-resolution image from a set of low-resolution images of the same scene into the MEG source spatial reconstruction distributed on the interpolated high-resolution cortical mesh. Fourthly, we expanded the reconstruction from spatial field into the temporal field in *Chapter 5 MEG image estimation via Kalman smoother*. Assuming that the process of MEG measurement of the brain sources is a dynamic system, the classical solution Kalman filter as well as the Kalman filter are applied to estimate a hidden high-resolution source distribution directly from the coil sensors of MEG. And then the EM algorithm is introduced to estimate the unknown parameter set of the model. Finally, we will discuss the significance of this research, as well as the advantages and disadvantages of the three designed algorithms in *Chapter 6 Conclusions*. Moreover, the associated future work will be discussed there.

In summary, the three topics from *Chpater 3* to *Chapter 5* are directly relevant to each other and should not be treated as independent work. In other words, the work in *Chapter 4* cannot proceed without the result from *Chapter 3*, and the work in *Chapter 5* cannot progress without the result from *Chapter 3* and *Chapter 4*.

Chapter 2

Literature review

2.1 Overview of medical imaging

One of the most important tools of modern medicine, brain imaging, in the clinical context referred as *clinical imaging* or *radiology* , are the technique and processes used for exploring the structure and functional status of part of, or the whole, human body; for clinical purposes(diagnosis and treatment) and medical research (Udupa and Herman, 2000), (Hajnal et al., 2001).

2.1.1 Brief introduction of medical imaging

From the view of functional use, current methods of medical imaging can be divided into two groups: one is *morphologic imaging* which focuses on imaging the internal structure of the human body anatomically, e.g. MRI, CT, X-ray, while the other one is *functional imaging* which is implemented for better understanding and observation of the functional status and change of human body, such as EEG, MEG, PET, SPECT and fMRI. In clinical and medical science, the *structural imaging* is usually combined with the *functional imaging* for advanced functional observation of human body (Papanicolaou, 2009), (Preissl, 2005), (Yokogawa, 2009), (Simon and Mattson, 1996), (Haacke et al., 1999).

For *structural imaging*, here is a brief introduction for these specific methods:

- X-ray

Also known as *X-radiation*, this is one type of medical imaging which actually uses electromagnetic (EM) radiation and generates 2D images of the human body. The wavelength of X-ray is in the range of $10 \sim 0.01$ nanometers, which corresponds to frequencies in the range of $3 \times 10^{16} Hz \sim 3 \times 10^{19} Hz$ and the energy in the range of $120eV \sim 120keV$. X-ray is a form of ionizing radiation, this imaging technique poses a health hazards during imaging acquisition. X-ray is applied in clinics for the pathology of skeletal system, soft tissue for some disease , gallstones, kidney stones (which are not always visible) etc. The most notable example of X-ray is chest X-ray, which is very effective in the diagnosis of diseases of the lung, such as pneumonia, lung cancer or pulmonary edema. However, for the imaging of soft tissue, X-ray has less advantages than CT and MRI which only produces a 2D projection of the tissue (Whaites and Roderick, 2002), (Squire and Novelline, 1997), (Charles Hodgman, 1961). Moreover, X-ray is listed as one kind of carcinogens by American government in 2005 (government, 2005).

- CT

Also referred to as *X-ray computed tomography* or *computer assisted tomography(CAT)*. CT builds three-dimensional images employing tomography through computer geometric reconstruction based on X-ray techniques (Dictionary, 2009). Tomography is a technique that employs the penetrating wave to obtain reconstructed image from the sections, so it is also called *tomographic reconstruction* (Herman, 2009),(Tomography). First, X-rays are used to obtain a large series of 2D slices with a single axis of rotation. Since the different tissues of the human body have different radiodensity when the X-ray goes through, a 3D tomographic image is reconstructed by CT from the slices. CT was first introduced in the 1970s (Herman, 2009), (Udupa and Herman, 2000),(Beatles, 2005).

X-ray generates overlapping projective slices with two dimensions, while CT provides three dimensional tomographic reconstruction of human body so that the information on the sagittal plane, coronal plane and transverse plane can be displayed individually. 3D tomographic imaging has a spatial-resolution of up to 0.5 mm, and shows advantages for detecting structure in the head, chest and heart. However, for soft tissue contrast, MRI performs better than CT. Also, as

CT is based on X-ray which is ionizing radiation, it is a health hazard. (Udupa and Herman, 2000), (Brenner et al., 2001), (Nelson, 2009).

- MRI

  Magnetic Resonance Imaging(MRI), also referred to Nuclear Magnetic Resonance Imaging(NMRI), in contrast to CT and X-ray, is one type of medical imaging with no ionizing radiation, and is used for visualizing the detailed internal structure of the human body by implementing the properties of nuclear magnetic resonance . Basically, around 2/3 of the human body consists of water in which each water molecule contains 2 hydrogen atoms(essentially protons). Within a powerful magnetic field, the nuclear magnetization of the hydrogen atoms in the human body is aligned along the direction of the magnetic field. Radio Frequency(RF) is used to produce an electromagnetic field which is able to alter the nuclear magnetization. In other words, the proton in spin-low state obtains the appropriate energy from an RF pulse, known as the resonance frequency, to flip the spin. Turning off the RF, the proton decays from spin-up state to spin-down state and the difference energy is released as a photon which produces a signal, this signal can be detected by the scanner. Compared with CT, MRI provides greater contrast of the soft tissue of the body. Moreover, MRI provides superior features for neurological (brain), musculoskeletal, cardiovascular, and oncological (cancer) imaging. However, MRI shows no better result on imaging of the lung compared with X-ray. And, CT, also no better result on imaging of the liver, prostate, pancreas and adrenal gland compared with CT. However, the MRI scanner of it is much more expensive. For clinical use, the magnetic field of the MRI ranges from 1.5 T(Tesla) to 3T, and can be up to 7T in medical research. There are some other associated imaging techniques used for medical imaging based on the imaging theory of MRI, such as Diffusion-MRI, structural MRI, etc. (Haacke et al., 1999), (Simon and Mattson, 1996), (Adams et al., March, 2009), (Squire and Novelline, 1997)

For *functional imaging*, here is a brief introduction for these specific methods:

- fMRI

Functional MRI is a new kind of functional brain imaging which measures the hemodynamic response(the change of blood flow and blood oxygenation) related to neural activity in the brain or spinal cord of the human body. Basically, the amount of blood oxygen is increased and the consumption of energy from glucose is increased as the neural cells become active. The increase of blood flow occurs approximately $1 \sim 5$ seconds after the neural cells became active. The homodynamic response peaks around $4 \sim 5$ seconds later. The BOLD (Blood Oxygen Level Dependence) response is well correlated with changes in the hemodynamic response. The change of the BOLD signal detected by fMRI is actually an indirect measure of neural activity. fMRI was first applied to the human body in the 1990s. It has advantages such as being a non-invasive record of functional brain signals, high-spatial resolution(up to 1 mm) and a superior signal record from all regions of the brain rather than only from the cortical surface. However, as the BOLD response detection is an indirect measurement of neural activity , this measurement is susceptible to the influence of non-neural events in the brain. Moreover, fMRI has poor temporal resolution approximately 5 seconds for a particular response compared with MEG/EEG . It is therefore hard to distinguish the different neural activities occurring within a short time frame. (Bandettini et al., 1993), (Logothetis et al., 2001), (Laureys et al., 2009).

- PET

  Positron Emission Tomography, is one kind of medical imaging that uses radiation detector to detect pairs of gamma ray which are generated from a radionuclide. It produces images of functional processes in the body. One kind of radionuclide, a short-lived isotope which emits positrons, is used as the radioactive tracer and is introduced into the human body as part of a biologically active molecule(e.g. Fludeoxyglucose (FDG) is commonly used (Fratt, 2003)(et al., 2005)). After waiting a period after the biologically active molecule is injected into the human body, the isotope decays and emits the positrons with opposite charge to the electrons in the body. These positrons encounter electrons and both of them are annihilated. A PET scanner includes a ring of detector units which receive gamma rays produced by annihilation events. PET is always applied along with CT or MRI so both the metabolic and anatomic information can be detected. Moreover, PET provides body functional imaging in a 3-D spatial field or even 4-D images with 3D spatial field and time fields (Ter-Pogossian et al., 1975), (Young et al., 1999), (Young et al., 1999)

PET scaning is widely applied in both clinical practice and research for functional imaging of the human body. For instance, in clinical use, PET is a great tool in oncology, especially for the imaging of tumors and searches for metastasis. Also, PET is a powerful tool for the research of cardiac and brain function. However, PET involves the exposure to ionizing radiation to a slightly extent than chest X-ray and CT. (Ter-Pogossian et al., 1975), (Young et al., 1999).

- SPECT

  The full name of SPECT is *Single Photon Emission Computed Tomography*. It is a kind of nuclear tomographic medical imaging using gamma rays. It provides a 3D image of the human body by injection of a radioisotope into the bloodstream of the subject. The radioisotope emits gamma rays which can be acquired by a gamma camera which captured a series of 2D images with multiple angles. The computed tomography is then applied for the reconstruction of a 3D image of the human body, which is similar to CT and PET. Furthermore, SPECT is similar to PET with respect to the application of a radioisotope as a tracer as well as detection of gamma rays. The difference between them is the SPECT uses a gamma-emitted radioisotope to detect the gamma rays directly while PET detects gamma ray indirectly with a positron-emitted radioactive tracer. In practice, SPECT is widely used to observe biochemical and physiological processes as well as size and volume organs, e.g. SPECT is used to tumor imaging, bone imaging, and functional cardiac and brain imaging (Frankle et al., 2005), (Amen et al., 2008).

- EEG

  Defined as Electroencephalograph, measures the electronic field surrounding head via a sensitive system of sensors and amplifiers located outside the scalp.

  Although methods like EEG and MEG are not designed for producing images primarily, the data obtained from these technologies is still suitable to be represented as maps which can be reconstructed as brain images.

- MEG

  MEG is explained carefully in the following part.

Within all the various medical imaging methods currently available , some of them are able to be used for brain imaging, such as CT, MRI, PET and SPECT. Meanwhile, some of them are specifically designed for brain imaging, such as fMRI, EEG and MEG. All these methods permit functional and anatomical studies of the human brain without opening the skull which provides a powerful tool for clinical and medical relevant research. MEG, which will be introduced next, has significant advantages over other methods of brain imaging because if produces more accurate functional information of human brain with much higher temporal domain and is non-invasive. Meanwhile, the target area of this new magnetic imaging technique has to be expanded from the brain to other areas of human body, eg, magnetocardiogram (MCG)(Baule and McFee, 1963) is the technique measures the magnetic fields produced by electrical activity in the heart. The potential of MEG outlined above may make this magnetic imaging technique one of most advantageous means of medical imaging in the future .

2.1.2 MEG vs EEG: similarities and difference

Neural current sources in the brain generate the external magnetic fields and scalp surface potentials. Modern non-invasive technologies, high-resolution electroencephalography (EEG) and magnetoencephalography (MEG) techniques allow spatio-temporal investigation respectively of these magnetic fields and potentials in the human brain. The principle characteristics of MEG and EEG are quite similar: first, the signals for MEG and EEG are both caused by the same neurophysiological event but expressed as different forms; secondly, the temporal resolution of both MEG and EEG is as high as a millisecond, and thirdly, measurements of MEG and EEG both have linear relationships with the strength of current sources distribution but non-linear relationships with the sources location (Wendel et al., 2009) (Baillet et al., 2001).

It is worthy to note that both MEG and EEG models are based on the Maxwell's equation i.e. they are based on the relationship between the current source distribution of interest and the measurement at the sensor array. This problem is described as the *forward problem* for both MEG and EEG, whose linearity can be expressed as the inner product of the *leadfield* (which is introduced in the later part: *2.2.5. Forward formula and inverse problem of MEG*) and the current source distribution of interest (Tripp, 1983). Since the majority of the inverse methods for MEG and EEG are based on linear algebraic formulations, the framework for the solution of the forward problem is a matrix formulation (Baillet et al., 2001) (Dale and Sereno, 1993) (Darvas et al., 2004) (Hämäläinen et al., 1993).

Although EEG and MEG signals originate from the same neurophysiological processes, there are important differences. The scalp surface matching for MEG and EEG are different in terms of the different means of sensors locations. For EEG, the sensor locations can be used instead of the head-shape since the sensors are attached on the scalp of the subject. For MEG, the sensors are located in the helmet in which each sensor probably does not exactly match the scalp of the subject. Thus, the head-shape is first co-registrated on a structural MRI, then the same transformation is applied onto the sensors. Moreover, distortions exist when the magnetic field and the potentials pass through the brain to the external surface of the scalp, where they can be measured. Magnetic fields are less distorted than electric fields by the skull and scalp, which results in a better spatial resolution of MEG. Further, since electric and magnetic fields are oriented perpendicular to each other, the directions of highest sensitivity are orthogonal to each other. Whereas scalp EEG is sensitive to both tangential and radial components of a current source in a spherical volume conductor, MEG detects only its tangential components. MEG therefore measures activity in the sulci selectively, whereas scalp EEG measures activity both in the sulci and at the top of the cortical gyri, but appears to be dominated by radial sources. And, according to the work of Barth D.S and colleagues, it is notable that scalp EEG is sensitive to extracellular volume currents produced by postsynaptic potentials, while MEG primarily detects intracellular currents associated with these synaptic potentials because the field components generated by volume currents tend to cancel out in a spherical head model (Barth et al., 1986). Therefore, MEG is more sensitive to superficial cortical activity, which makes it useful for the study of neocortical epilepsy, since the decay of magnetic fields as a function of distance is more distinct than for electric fields. Finally, EEG relies on a reference that makes interpretation of the data difficult to process; while MEG is reference-free. (Mosher et al., 1999) (Cohen and Cuffin, 1983) (Barth et al., 1986).

2.2 Introduction of Magnetoencephalography(MEG)

2.2.1 Brief introduction

Magnetoencephalography(MEG) is a new non-invasive tool for functional imaging of the brain. Compared with other brain imaging techniques, MEG detects the extremely faint magnetic fields generated in the human brain with no ionizing radiation. Furthermore, MEG provides functional mapping of the whole brain with outstanding temporal resolution.

The different functional states of brain are represented, in terms of the measurement of changing of magnetic field around the scalp by MEG, . Therefore, the neuronal activities that evoke the magnetic field can be measured directly. In order to understand the functional brain spatio-temporally, MEG result are co-registered with the corresponding structural MRI result so that the biological function combines with anatomical structure. The combination of MEG and structural imaging (eg. MRI) is known as magnetic source imaging (MSI). It is remarkable that MSI is able to provide information about the functional brain temporally together with the spatial functional localization.

Because of the highly sensitive qualification and precision manufacturing, the purchase of MEG apparatus and relevant services are very costy. So far, there are three manufacturers in the world that produce MEG apparatus, they are CTF-MEG (Canada), Elekta-Neuromag (Finland) and Yokogawa (Japan).

2.2.1.1 Neural basis of MEG

The human brain consists of two hemispheres which are separated by the longitudinal fissure. Furthermore, the hemispheres are divided into lobes by two deep grooves. The Rolandic fissure cuts vertically the outer part of both hemispheres, and the Sylvian fissure runs almost horizontal. Thus, the cortex is separated with four lobes in both hemispheres; frontal , parietal, temporal and occipital, respectively(showed in the Fig 2.1).

Each lobe can be mapped functionally. And, the cortex has the total surface area of approximately $2500cm^2$ which is highly folded to fit in the skull compartment. MEG

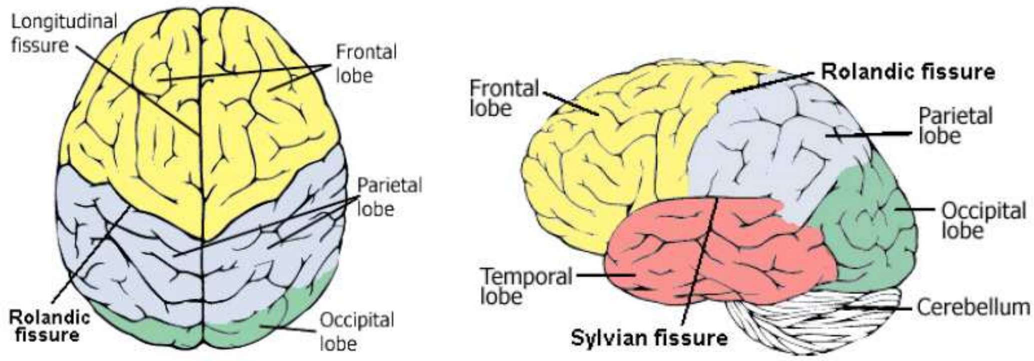

Figure 2.1: The human brain consists of two hemispheres which is separated by the longitudinal fissure. Furthermore, the hemispheres are divided into lobes by two deep grooves. The Rolandic fissure cuts vertically the outer part of both hemispheres, and the Sylvian fissure runs almost horizontal. Thus, the cortex is separated with four lobes in both hemispheres, frontal , parietal, temporal and occipital, respectively (Picturs taken from AMA Health Insight: http://www.ama-assn.org)

studies are usually covered with the uppermost layer of the brain. (Hämäläinen et al., 1993)

The neurons and glial cells are the principle components for building the brain. The glial cells provide the main physical structure of the brain as well as the transport of the nutrients between blood vessels and the brain tissue. The large number of $10^{10} - 10^{11}$ neurons mainly process the information of the brain. (Phillips, 2000)

The magnetic fields measured by the MEG sensors are contributions from both the *primary current*, produced by current flow in apical dendrites in cortical pyramidal neurons and representing the *neural and microscopic passive cellular current*, and the *volume or secondary current* , which is generated from the *macroscopic electric field* (Papanicolaou, 2009) (Tripp, 1983). Since the primary current represents the neural activity with a given cognitive process, it is considered as the sources of interested in MEG. Therefore, the general concept of MEG source estimation, reconstruction and localization are based on reconstructing the underlying primary sources. In the context of MEG source reconstruction, the aim of the forward problem and the inverse problem is both to estimate the primary current sources (Papanicolaou, 2009) (Tripp,

1983) (Baillet et al., 2001).

Primary current sources generate volume currents. If no primary current exsits, then no volume current can exist. But here can be closed loop primary currents that generate no volume currents. It is noteworthy that volume currents tend to cancel out in a spherical volume conductor, in which case the MEG measurement is only detected from the primary source(Barth et al., 1986).

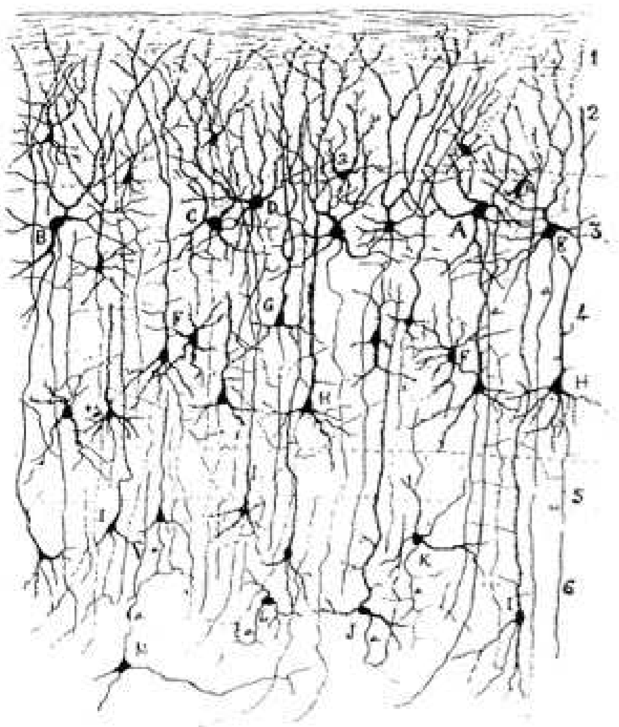

Figure 2.2: A neurone consists of three principle parts: the cell body is as the ”processor” which contains the nucleus; the dendrites which are like the thread extensions are the ”receivers” which receive stimuli from other neurones; and the axon is as the ”transmitter” which is a single long fibre carrying the impulse from the cell body to others. Primary currents are produced by current flow in apical dendrites in cortical pyramidal neurons (neu, 1993).

If the primary sources as well as the surrounding conductivity distribution are known, it is feasible to calculate the magnetic field (by MEG) /electrical field(by EEG) in terms of Maxwell's equation. This forward problem, specified for MEG, is explained in *Chapter 3: Forward problem*.

It is noteworthy that there is the special case that the volume current generates the magnetic field with equal magnitude but opposite direction. For instance, the contribution of volume currents is then canceled, and the measurement of MEG only contains the magnetic field caused by primary current(Hämäläinen et al., 1993) (Phillips, 2000).

2.2.1.2 MEG technique introduction

MEG measures the magnetic field surrounding the head via an extremely sensitive system of sensors and amplifiers located outside scalp, such as superconducting quantum interference devices (SQUIDs), or magnetometer. (Papanicolaou, 2009), (Preissl, 2005), (Baillet et al., 2001), (Cohen, 1968).

The measured magnetic field is mainly generated by the electrical activity in brain. MEG source reconstruction from the measured magnetic field localizes 3D pattern of neuronal activity of the cortex spatio-temporally (Preissl, 2005), (Srikantan et al., 2006), (Kishida, 2009). However, it is a typically ill-posed inverse problem which is theoretically insoluble. In term of this feature, the users of MEG face a choice of various possible inverse solutions which could be used for processing the measured data. ('́Ozmen et al., 2007),(Preissl, 2005).

For the functional brain imaging, although techniques, such as fMRI, PET, etc. show outstanding spatial resolution , MEG presents a superior temporal resolution which complements the weakness of brain imaging temporally (Rodriguez et al., 2003), (Baryshnikov et al., 2004).

2.2.1.3 Application of MEG

MEG is widely used for both in clinical practice and research. The relevant research using MEG includes linguistic, visual, auditory and tactile activity. Also, MEG is involved in the research of connection between visual, auditory activity and cognitive function along with linguistic study during information processing. In medicine, MEG is mostly used in the diagnosis of epilepsy and localization of the epileptic focus before surgery. MEG is involved in the diagnosis of diseases, such as epilepsy, brain tumor, stroke, brain trauma, Alzheimer's disease(AD), and Parkinson's disease.

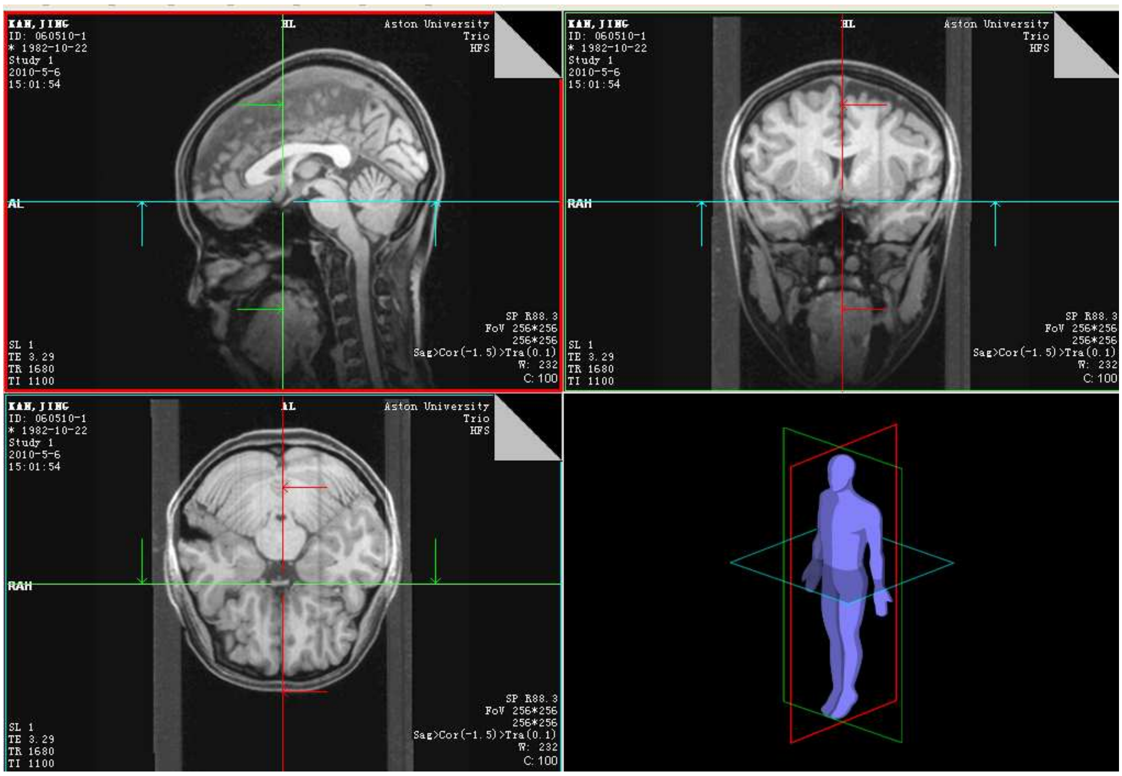

Figure 2.3: The demonstration of structural MRI

Figure 2.4: A demonstration of an MEG scanner and subject

2.2.1.4 Comparison of MEG with other medical imaging apparatus

Measurement Function	Device	Feature
Morphologic brain imaging	CT	1. measurement with ionizing radiation; 2. high spatial resolution
	MRI	1. measurement with powerful magnetic field and radio frequency; 2. high spatial resolution 3. low temporal resolution
Electric physiological functional brain imaging	MEG	1. non-invasive, natural environment of measurement 2. high temporal resolution(up to 1 millisecond); 3. high spatial resolution(up to 1 cm)
	EEG	1. non-invasive, natural environment of measurement 2. high temporal resolution(up to 1 millisecond); 3. directly marker of brain neural activity;
Metabolic functional brain imaging	fMRI	1. high spatial resolution 2. low temporal resolution 3. record changes in blood flow, indirectly marker of neural activity
	PET	1. measurement requires injection of a positron-emitting radioisotope as tracer, 2. high spatial resolution 3. low temporal resolution 4. record metabolic activity, indirectly marker of brain neural activity;
	SPECT	1. measurement requires injection of a gamma-emitting radioisotope as tracer, 2. high spatial resolution 3. low temporal resolution 4. record changes in blood flow, indirectly marker of neural activity

Figure 2.5: Comparison of brain imaging methods

Within the methods of functional brain imaging , there are three main processes for the description of brain activity, which are the neural signal, blood flow and metabolism. The neural signal is the most direct and basic and the blood flow and metabolism are both depended on the neural activity. MEG is a technique which directly measures the neural signal. fMRI measures the signal related to blood flow, and PET and SPECT is for measuring metabolism. SPECT is similar to PET in its use of radioactive tracer material and detection of gamma rays. In contrast with PET, however, the tracer used in SPECT emits gamma radiation that is measured directly, whereas PET tracer emits positrons which annihilate with electrons up to a few millimeters away.

Fig 2.8 provides a table to compare the current brain imaging methods, whilst it indicates the advantages of MEG compared with any other methods, especially if it is absolutely non-invasive and has outstanding temporal resolution(up to 1 ms). However, since methods such as fMRI have high spatial resolution but low temporal resolution, it is popular to combine the advantages of both methods. Additionally, the technique of EEG is similar to MEG. The main difference is that the skull and the tissue surrounding the brain affect the magnetic fields measured by MEG much less than they affect the electrical impulses measured by EEG. The advantage of MEG over EEG is therefore greater accuracy owing to the minimal distortion of the signal.

2.2.2 Apparatus

2.2.2.1 Recording principle

The neural signal generates a current which induces the magnetic flux surrounding the cortex.

The flux of the cortical magnetic field can be detected by extremely sensitive sensors surrounding the head surface, shown in Fig 3.1 . The type of these special sensors are a set of either magnetometers or gradiometers. Each sensor, for instance , a magnetometer, is a loop of wire, or coil, which is located parallel to the head surface. As the flux lines thread through the coil, the corresponding current is generated by induction in the coil. This current on the coil is proportional to the flux which can be thought of as the expression of the magnetic induction from the inside brain. Using a special amplifier, SQUIDs, the weak induced currents on the coil can be converted into a high-amplitude voltage. In this way, the scalp magnetic field is recorded by each sensor every millisecond. Assuming there are a sufficient number of the sensors located at regular interval places surrounding the head, the corresponding neural source distribution can be measured and determined all over the cortex. Since the magnetic field evoked in the cortex as well as the induced current on the coil of sensor are extremely feeble, MEG apparatus must be housed in a magnetically shielded room(MSR) to attenuate any noise from the external environment. Furthermore, both the sensor and the SQUIDs are placed under the condition of superconduction in order to operate correctly .

It is worth noting that any flux line threading through the coil of a sensor can be divided into two parts, a perpendicular component to the coil plane and a tangential component to it. Then, the current on the sensors is induced by the perpendicular part only. This also indicate the flux lines which is perpendicular to the coil plane induce the strongest current on the sensor does not respond to the other components.(Papanicolaou, 2009)

2.2.2.2 Industrial structure of MEG

- Special sensors

  In any location around the cortex, the magnetic field corresponding to the cortical neural signal can be thought as the combination of different types of field, eg, uniform component, first-gradient component, second-gradient component etc. In practice, there are two types of sensors used for the MEG measurement which are magnetometers or gradiometers.

  a). Magnetometer

2.2.2.3 Comparison of MEG with other medical imaging apparatus

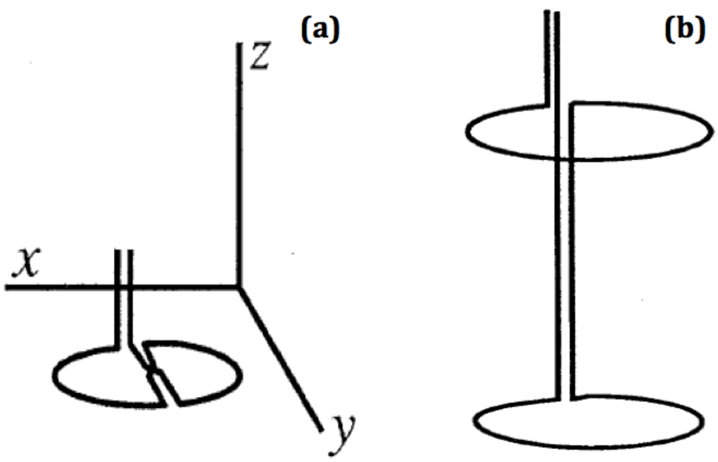

Figure 2.6: Figure *a* shows the structure of the planar gradiometer; while figure *b* shows the structure of the axial gradiometer, (Hämäläinen et al., 1993).

Each magnetometer (shown in Fig 2.6) is placed parallel to the head surface which detects the total field induced in the located area, including uniform, or zero-gradient, component, first gradient component, etc.

b). Gradiometer

A gradiometer (shown in Fig 2.6) has the outstanding advantage of reducing noise while detecting MEG measurements. The gradiometers of first-order reject the uniform component, further the gradiometers of second-gradient reject the uniform component and first component, and so on. Current MEG system uses the gradiometer of the first order, which normally has two structural types, axial gradiometer and planar gradiometer.

In summary, the magnetometer is good at detecting all types of the signal while the gradiometer is good at reducing the noise. This thesis mainly discusses the MEG system with magnetometers(comparison of the magnetometer and gradiometer comparison of the axial gradiometer and the planar gradiometer). In practice, gradiometer are more widely used as the sensors of MEG rather than magnetometers.

Due to the current on the coil of sensors induced by extremely weak magnetic flux in the brain, the environmental condition for sensors must be set as superconducting so that they have no resistance. This can be achieved when the temperature of the coil is reduced close to absolute zero(-273.15 degrees on the Celsius scale). In the MEG system, the sensors are placed within a thermal isolated dewar which is filled with liquid helium. These working conditions can keep the temperature of coil around 4K(-269.15 degrees on the Celsius scale) which is sufficient for superconducting. (Papanicolaou, 2009)

- SQUIDs

Since both the magnetic flux of the brain and the induced current on the magnetometer(which is proportional to the flux) are extremely weak, an amplifier must be implemented to detect the signal. Conventional amplifiers are not able to achieve this task because of their intrinsic thermal noise.

In the late 1960s, the superconductive quantum interference device, so-called SQUIDs, was co-invented by James.E. Zimmerman. SQUIDs were soon applied to research from airborne submarine sensors to gravitational waves. Then, the first experiment in applying SQUIDs to measure the magnetocardiogram was conducted in MIT. It is the use of SQUIDs which makes MEG measurement practical (Baillet et al., 2001), (Cohen, 1968).

- Magnetically shielded room(MSR)

  No matter how strong the magnetic signal evoked from the cortex, are no matter how accurate the induced current of the magnetometer is converted into the high-amplitude voltage, all these types of signals in these processes are too weak to compare with major types of magnetic noises in the outside world , e.g. urban noise or the earth's magnetic field. Normally, the magnetic field generated in the brain is on the order of several tens of femtoTesla(fT, or 10^{-15} Tesla), while the earth's magnetic field is around several microTesla(10^{-6} Tesla). To avoid the major interference from the external environment, the MEG apparatus is placed within a magnetically shielded room(MSR) for isolation.

  Various types of materials with different magnetic permeability, eg, mu-mental, aluminum, copper, etc, are used to construct the MSR in successive layers. As a result of this structure, the MSR is able to shield against the noise with *both low-frequency-(as low as 0.1Hz by $\geq$ 40 dB)-and high-frequency-(up to 1GHz at 60 dB-signals)* (Papanicolaou, 2009).

2.2.3 History

2.2.3.1 Development of MEG

The first trial of MEG was conducted by University of Illinois physicist David Cohen in 1968 who used a copper induction coil as a magnetometer (Cohen, 1972). A shielded room was used to reduce the measurement noise. However, this measurement result was too poor with too much noise, to be applied into in practice. Later, the invention of SQUIDs accelerated the development of the MEG technique. In MIT, David Cohen then built a better shielded room and applied SQUID detectors for MEG measurement with the cooperation of James E.Zimmerman. The result was as clear as EEG this time which marked the start of MEG research (Cohen, 1972).

At the beginning, MEG apparatus was manufactured with a single SQUID detector. To take measurements, the single SQUIDs detector was successively used for measure the magnetic field at a number of designated position on the subject's scalp. This was very much of inconvenient for making measurements as well as the influence of the accuracy of measurement. Then, the manufacturers developed the structure of MEG apparatus by increasing the number of sensors and SQUIDs in a larger thermally isolated Dewar to cover a larger area of the scalp. The MEG Dewar at present is made as a helmet-shaped which is almost able to cover the whole scalp of subjects. MEG systems are currently produced with the number of sensor arrays from 100 to 300, such as 128-channel, 248-channel and 306-channel. Of course, the more channels it has, the more accurate information can be obtained from MEG measurements.

The first MEG apparatus in UK was installed at Aston University in 1999 which started the MEG research in the UK. At present, there are more than 10 academic institutes in UK, who own MEG apparatus for research, eg, Aston Brain Centre in Aston University, York Neuroimaging Centre (YNiC), Cardiff University Brain Repair Imaging Centre (CUBRIC), University of Nottingham, University of Oxford, The Wellcome Trust Centre for Neuroimaging of UCL , Cognitive Neuroimaging (CCNi) in University of Glasgow . Further, more and more British academic institutes joined in cooperation with the institutes above for collaborative research. Because of restrictions in the NHS presently, MEG is not able to be applied clinically directly in UK, and if is only for research use. However, MEG has already been significantly applied in both medically and research legally in some other countries, such as USA, China and Japan.

2.2.4 Head model of MEG

MEG is concerned with the study of the brain and a number of different head models are used. The different assumptions of the head models in MEG directly reflects the nature of the geometry and electrical conductivity of brain. The induced internal current includes primary currents and secondary currents which both affect the brain at the same time, and so the application of different head models is important for source localization. There are some types of head models which are commonly applied in the analysis, such as the spherical model,the boundary element model(BEM), and finite element models(FEM). It is obvious that different head models are designed to have a focus on different research angles, and approximation is used to decrease computational complexity. In this case, there are some intrinstic errors which exist within each head model and affect the design of source model, accuracy of source reconstruction and computation efficiency (Papanicolaou, 2009) (Baillet et al., 2001).

Generally, MEG head models can be divided into two classes, spherical head models and realistic head models. Finite element models(FEM) and boudary element models(BEM) are classified as realistic head model. However, the boundary between these two classes are not absolute, there are plenty of models designed to combine both these features, such as the spherical BEM model . In the following part, some head models are introduced (Papanicolaou, 2009).

2.2.4.1 Spherical model

The spherical model assumes the head is a single sphere or multiple concentric spheres. By using the appropriate structural imaging scan of the subject, the best fit sphere is found for analysis. Additionally, the sphere is assumed to be homogenous and isotropy. This indicates the conductivity of each volume in the brain is assumed to be the same, moreover, the conductivity of each volume is assumed to be independent to the current direction. This kind of head model is frequently used for clinical studies as it is easy to generate and efficient to analyse with sufficient accuracy.

The spherical head model shows good accuracy for the focus on the head area where the curvature of the local brain surface approximately matches part of a sphere. However, for regions where the curvature diverges from a sphere, for instance, the temporal lobe , the results are less satisfactory. (Papanicolaou, 2009)

2.2.4.2 Boundary element model

The boundary element model(BEM) approach is a model which consists of a set of nested surfaces which are basically composed of three layers, inner skull, outer skull and scalp surface. BEM assumes homogenous and isotropy for any layers and compartment of the layers. However, the constraint of symmetry on the spherical model has been relaxed. In this case, this model is able to more accurately describe the geometry of the individual subject. In BEM, the primary current is not considered as symmetric as in the spherical model. Furthermore, the potential from different layers are computed and added to calculate the internal current. In addition, the volume current is considered to affect the neural activity and external magnetic field measured by MEG. Thus, BEM results in a better description of the brain neural activity. (Hamalainen and Sarvas, 1987)

However, since the results of BEM are not a distinct improvement in significance and accuracy, but are computationally expensive and time consuming for accurate indentification of different boundaries in complex brain tissue, it is more frequently applied in research rather than in clinical settings. (Papanicolaou, 2009)

2.2.4.3 Finite Element Model(FEM)

The Finite Element Model(FEM) is one type of numeric approach which "allows the use of anatomically realistic head models and the increased computational power that they require has became readily available"(Schimpf et al., May 2002). Since the FEM presents the realistically complicated non-homogenous head conductor, it is often used for the forward problem.

The electronic field of the behavior of the neural sources in the brain can be described from Poisson equation (Schimpf et al., May 2002) (Plonsey, 1969) :

$$\nabla \cdot \sigma \nabla V = \nabla \cdot \hat{J}_i = p \tag{2.1}$$

where $\hat{J}_i$ is the applied current density(A/m^2), σ is the conductivity of the volume (Ωm)ˆ-1, and V is the electric potential. A class of numerical methods use a set of basis functions to approximately model the potential throughout the brain volume. In terms of the Eqn 2.1, the approximation of the potential can be calculated by minimizing a weighted avaerage of the residual(Schimpf et al., May 2002):

$$\sum_{i=1}^{n} [\int_{\Omega} (\nabla \cdot \sigma \nabla N_i) W_j d\Omega] a_i - \int_{\Omega} p W_j d\Omega = 0 \tag{2.2}$$

$$j = 1, 2, ..., n \tag{2.3}$$

$$V \approx \sum_{i=1}^{n} N_i a_i \tag{2.4}$$

where W_j are the weighting functions, N_i are the basis functions, a_i are the degrees of freedom(DOFs) which are used for fitting these basis functions to the potential. The dvergence theorem is applied to the integral. The basis function are generally set as a set of polynomials in 3-space, which are defined locally over subregions as a finite number of *elements*. Also, the weighted function are set as same as the basis functions.

Although as the numeric method, FEM can represent more realistic domain, the sources in this model are an approximation of the *ideal dipole*. However, the mathematical idealization does not exist in practice (Schimpf et al., May 2002) (Wolters, 40).

2.2.5 Forward formula and inverse problem of MEG

Generally, MEG is a type of non-invasive technique reflecting the electromagnetic status of the internal brain with high temporal resolution(up to millisecond). The focus of MEG is detecting the internal current source information, including direction, strength and locations ,by the measurement of magnetic field on scalp surface.This problem in MEG research is given a number of names depending on the specific goal, including ’source estimation’, ’source localization’, ’source reconstruction’, or ’source imaging’. In order to successfully tackle this problem, it is usually necessary to use information from morphological imaging techniques such as MRI. This combination is named as Magnetic Source Imaging(MSI).

2.2.5.1 Fundamental equation of MEG

The concept of MEG sensing is to detect currents flowing in the brain from the magnetic flux recorded at a number of superconductive coils placed near the scalp. In the source space Ω', the magnetic field generated at a location r on the scalp is given by the Biot-Savart law (Baillet et al., 2001) (Preissl, 2005) (Papanicolaou, 2009):

$$\mathbf{B}(\mathbf{r}) = \int_{\Omega'} \frac{\mu_0}{4\pi} \frac{\mathbf{j}(\mathbf{r}') \times (\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} d\Omega' \tag{2.5}$$

here, $\mathbf{r}$ is the position where we measure the magnetic field; $\mathbf{r}^{'}$ is a position in the source space ; $\mathbf{j}(\mathbf{r}')$ is the internal current element which including both primary current and volume current(Barth et al., 1986); **B** is the measured magnetic field and μ_0

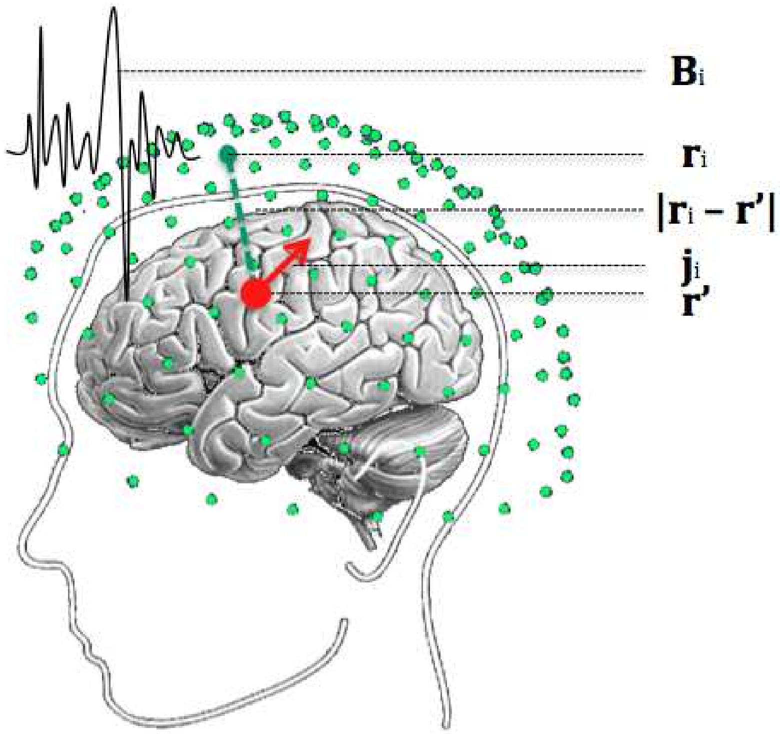

Figure 2.7: The diagram defining the source space Ω', the position of the sensor array $\mathbf{r}_i$, the position in the source space $\mathbf{r}'$ and the MEG measurement on the sensor i, $\mathbf{B}$.

is the magnetic constant; Under the spherical head model, the coils are placed radially around the origin of the coordinate system, and so the normal to coil i is given by $\mathbf{r_i}/r_i$ (shown in Fig 2.7). Under this model the problem has a simplified form where the magnetic flux detected by coil i is:

$$\begin{aligned} b_i &= \frac{\mathbf{r}_i}{r_i} \cdot \mathbf{B}(\mathbf{r}_i) \\ &= \int_{\Omega'} \frac{\mu_0}{4\pi} \frac{\mathbf{r}_i \cdot \mathbf{j}(\mathbf{r}') \times (\mathbf{r}_i - \mathbf{r}')}{r_i |\mathbf{r}_i - \mathbf{r}'|^3} d\Omega' \\ &= \int_{\Omega'} \frac{\mu_0}{4\pi} \frac{(\mathbf{r}_i - \mathbf{r}') \times \mathbf{r}_i}{r_i |\mathbf{r}_i - \mathbf{r}'|^3} \cdot \mathbf{j}(\mathbf{r}') d\Omega' \\ &= \int_{\Omega'} \mathbf{l_i}(\mathbf{r}') \cdot \mathbf{j}(\mathbf{r}') d\Omega' \end{aligned} \tag{2.6}$$

where $\mathbf{l_i}(\cdot)$ is the *leadfield* of coil i which indicates the connectivity between the measrement of magneticfield at $\mathbf{r}_i$ and the source location $\mathbf{r}$:

$$\mathbf{l_i}((r)') = \frac{\mu_0}{4\pi} \frac{(\mathbf{r}_i - \mathbf{r}') \times \mathbf{r}_i}{r_i |\mathbf{r}_i - \mathbf{r}'|^3} \tag{2.7}$$

2.2.5.2 Forward and inverse problem

Since the leadfield does not depend on currents or coil responses, MEG source reconstruction can be approached in two different ways which are the so-called *inverse problem* and *forward problem*. The forward problem involves 'computing the scalp potentials or external magnetic field at a finite set of sensor locations for putative source configuration'. This problem can be solved by a unique solution, which can also be said as 'well-posed' . On the other hand, the *inverse problem* is based on 'estimating the configuration of brain sources that account for the recorded magnetic field on the head surface'. (Mosher et al., 1999) It is practical to estimate the information of sources, eg, geometrical configuration on a 2D surface from the measured magnetic field. However, the provided information is not enough to determine the sources on a 3D cortical surface which may have multiple possibilities. In this case, the inverse problem of MEG is theoretical unsolvable or 'ill-posed'. In another words, it is underdetermined problem.

In terms of the characteristics of the ill-posed inverse problem of MEG, there are a number of methods of source estimation which models the internal sources based on different mathematical algorithms and assumptions, for example, the principle of some methods is to reduce the number of unknown to be less or lower than the number of known coil responses, making the problem well-posed. Since each of these methods are based on different the assumption, it is common that the results from different methods are different despite using the same set of data. Since the ill-posed inverse problem is insoluble itself, each method is heavily depend on the assumptions used. Addtionally, the situation is worse in practice. For instance, when the head model is assumed to be homogenous in conductivity and approximately spherical, the sources which are oriented radially, produce no magnetic field outside the head. Accordingly, the sources in the central brain where most directions are approximately radial, as well as other radial oriented sources in the brain, are not able to be reconstructed clearly by general MEG source estimation. This is one of the reasons that MEG source estimation is insensitive to the deep sources. Therefore, the design and application of various source models appropriately is the important consideration in MEG source reconstruction (Papanicolaou, 2009), (Preissl, 2005) ('Ozmen et al., 2007). We now review some source models from the literature.

2.2.6 MEG source modeling

2.2.6.1 The equivalent current dipole(ECD)

Historically, this was the first inverse solution to be developed for MEG source estimation. Mathmatically, a single ECD (equivalent current dipole) is assumed to be a pair of current sources with an infinitesimal separation (Papanicolaou, 2009). The standard method of estimating a single source is to determine the ECD (equivalent current dipole) by a non-linear least-squares research(Tuomisto et al., 1982). The dipole is assumed to be dynamic and can be adjusted to optimise the goodness-of-fit and find the unknown direction, strength and location of the dipole at each time point.

Generally, the observed measurment of magnetic field and predicted magnetic field by the estimated dipole information (eg, direction, strength and location) are incorporated into a cost function which measures the goodness-of-fit. The goal is then to find the dipole location which minimizes the cost function. Since the magnetic field has a non-linear relationship to the dipole position, it is difficult to find an analytical solution. Accordingly , numerical and iterative methods are applied to the problem. The initial setting of the parameters of the dipole sources directly influence the result of estimation, and may generate local minimum (Scherg and von Cramon, 1986) (Cuffin, 1985). The process of searching for the minimum cost function must be designed to find the global minimum and avoid the local minima. It is apparent that the global minimum provides optimal parameter estimation for cost function while the local minimums provide the sub-optimum.

In this method, there are usually a small number of dipoles, and the number of unknown quantities are generally lower than the number of measurements taken. The ECD is therefore a overdetermined problem. The simplest situation for the ECD is that only one source is assumed to be in the region of interest at one single time point. The biggest challenge for the ECD is the difficulty of determining the number of dipoles in the brain at a single time point. The most popular way is to keep adding dipoles to the possible regions and observe the change the magnetic field until the solution becomes stable or there are no more notable changes occur.

Since ECD is the most simple method of MEG source estimation, and has a long history, the ECD is the principal method used in clinical work (Papanicolaou, 2009). Specificly, the ECD shows success on 'the localization of interictal spike, the localization of language-specific cortical region, presugical localization of the early cortical evoked response'. (Papanicolaou, 2009)

The algorithm is as follows:

- Noise estimation

  Since MEG measures the very weak magnetic field outside of the brain which is normally smaller than $\frac{1}{10^8}$ of the earth magnetic field, the measurement is easy to be interfered by different noises. Before building the source model, it is necessary to reduce the noise. Here, some conventional ways are introduced for reducing of the MEG measurement. Firstly, using the Magnetically Shielded room as the noise reducing way. This is a effective way for reduce the general environmental noise. It is capable to decrease the external magnetic field by $100dB$ at $1Hz$. Then, the Reference sensors are used for reducing the noise.

With this way, there are a set of sensors are particularly used for noise measurement by locating far away from the head of the subject. Some of the gradient field is calculated by detecting the distant environmental noise. Also, using gradiometer as the noise reducing way. In terms of the structure and function of the gradiometer illuminated in the previous part, the sensor takes the difference between the magnetic fields measured by two consisted coils. Since the magnetic field generates by the brain is not homogenous, the noises generated from the long distance effectively reduced by Gradiometers.

Moreover, averaging and filtering are used for noise reduction. It is assumed that the measurement from different channels are independent so that the noise from each channel is not correlated with each other. Accordingly, noise covariance $\Sigma = diag(\sigma_1, \cdots, \sigma_n)$ where σ_i contains both the source noise source and the environmental noise. For the noise averaging method, the background activities are collected and saved before the source measurment so that this can be set as the average value of noise. Then both the source measurment and average value of noise can be applied for standard deviation. It is generally useful, but shows weakness when random inherent noise appears of some independent time point for σ_i. For the filtering method, the typical cut-off frequency of MEG measurement is $0.03 - 1.0Hz$ for the high-pass filter and $40 - 400Hz$ for the low-pass filter. However, the filter is applied in practice for filtering the measurement before the noise estimation which may bring inaccuracy for noise estimation(Hämäläinen et al., 1993).

Besides these , there are still several kinds of methods for noise estimation. For instance, Masaki Kawakatsu developed the ICA approach for MEG noise reduction which produces many different components and worked effectively to reconstruct the single evoked responses based on the objective criterion(Kawakatsu, 2003).

- Model building

  To descirible the fitting between the measurement and the magnetic field generated from the predicted sources, the following equation is applied:

$$g = 1 - \sum_{i=1}^{n} (b_i - \hat{b_i})^2 / \sum_{i=1}^{n} b_i^2 \tag{2.8}$$

where n is the number of channels involved in measrement, b_i is the noisy measurement of magnetic field on channel i , and $\hat{b_i}$ is the corresponding magnetic field produced by predicted ECDs. If the value of $g = 0$ approximately, it indicates the predicted data matches the measurement data. However, if $g = 1$, it means the predicted source model of ECD doesn't describe the measurements at all and the results are similar to a the generated magnetic field of zero. This fitting is analogous to the linear regression analysis.

Moreover, the χ *-square* distribution is applied for the test of goodness-of-fit:

$$\chi^2_{obs} = \sum_{i=1}^{n} \frac{(b_i - \hat{b}_i)^2}{\sigma_i^2} \tag{2.9}$$

which assumes Gaussian errors for the measurement data. The probability P_{obs} of the observed Chi-squared value directly reflects the goodness-of-fit for the model. It indicates that if P_{obs} is close to 1, the model well describes the internal sources and here is no need for adding extra dipoles. In the contrast, if P_{obs} alternates between $0 \sim 1$, it means the model is not satisfied and the further approaches need to be applied, such as adding extra dipoles.

It is notable that the small number of sources n generally give rise to the quick and unstable alternating value between $0 \sim 1$ when σ increases. In this case, if the noise is overall underdetermined, the modell is easily to be affected with unaccurate results. Underdetermination of the noise σ passes a more complex part to modelling which should have been taken by noise, whereas, overdetermination of noise σ may lose the possible variation in the detail of the source.

2.2.6.2 Multidipole model

The ECD can be generalized to a multiple dipolar source with spatial separation. In (Hari et al., 1984), it is shown if the distances of multiple, simultaneously active sources are sufficiently large from first and second sematosensory regions, SI and SII, as well as the source directions favorable , there is less overlapping from the multiple sources and it is feasible to fit each source distinctly by the Equivalent Current Dipole (ECD). Similarly, the sources can be separated clearly with ECD if the sources change with time. However, if these conditions of sources are not met and the sources overlap in both the spatial and temporal fields, the ECD must be extended to be the *spatial-temporal multidipole model* as a solution(Hämäläinen et al., 1993).

In contrast to the single equivalent current dipole which deals with the source in the temporal domain separately, the multiple dipole model models the sources in the spatio-temporal field together. Basically, the multiple dipolar sources are assumed to be able to alter the strength but maintain the position, and optionally maintain the direction throughout the time interval of interest. Then, a predicted magnetic field is produced to match the measurement. The number of multiple dipolar sources(unknown) are generally lower than the number of measurements of the magnetic field, thus this multi-dipole model is also an overdetermined problem(Hämäläinen et al., 1993).

Compared with ECD, which depends on the initial guess of the values of the ECDs, the multidipole model solves the highly complex optimization by selecting the starting parameters (initial estimate of the solutions) randomly from either the cortical surface or the grid of brain volume.

The predicted data and measured data of the magnetic field is denoted as B_{jk} and M_{jk}. $j = 1, \cdot, n$ indicates the number of sensors, and $k = 1, \cdots, m$ indexes the time intervals t_k. And this multi-dipole model is formulated as follows:

$$S = \| \mathbf{M} - \mathbf{B}(x_1, \cdots, x_q) \|_F^2 \tag{2.10}$$

Here, the equation is the minimum for the conventional least-square error function where $x_1, \cdots, x_q$ indicates the unknown parameters in this model. There are p dipoles assumed located on $\mathbf{r}_d$, where $d = 1, \ldots, p$. Specificly, p_1 dipoles are fixed-orientation, and p_2 dipoles are with variable-orientation, where $p = p_1 + p_2$. And, $r = p_1 + 2p_2$ dipoles wave forms are retrieved.

Additionally, $\| \cdot \|_F^2$ denotes the square of the Frobenius norm:

$$\| \mathbf{A} \|_F^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} A_{ij}^2 = Tr(\mathbf{A}^T \mathbf{A}) \tag{2.11}$$

Furthermore, the magnetic field calculated by predicted multidipoles can be written as:

$$\mathbf{B}^{(n\times m)} = \mathbf{G}^{(n\times 2p)}(\mathbf{r}_1, \cdots, \mathbf{r}_p)\mathbf{R}^{(2p\times r)}\mathbf{Q}^{(r\times m)} \tag{2.12}$$

where **G** represents the *Gain matrix* composed of the unit dipoles as all these variables are calculated as matrices, the dimension of these matrices are indicated on the superscripts. These unit dipoles are indicated on the spherical coordinate system with $(\mathbf{e}_\theta, \mathbf{e}_\phi)$:

$$G_{j,2d-1} = b_i(\mathbf{r}_d, \mathbf{e}_\theta) \tag{2.13}$$

$$G_{j,2d} = b_j(\mathbf{r}_d, \mathbf{e}_\phi) \tag{2.14}$$

with $j = 1, \cdots, n$, $d = 1, \cdots, p$

where $b_j(\mathbf{r}_d, \mathbf{e}_\theta)$ and $b_j(\mathbf{r}_d, \mathbf{e}_\phi)$ are the magnetic field produced by the predicted unit dipole on $\mathbf{r}_d$ distributed on the different directions of $\mathbf{e}_\theta$ and $\mathbf{e}_\phi$.

The first p_1 rows of **Q** represents the time series of amplitude at t_k of the dipoles with fixed orientations. And, the remaining $2p_2$ rows are the time series of the two components of these variable-orientation dipoles.

R contains the differentiation between fixed- and variable-orientation dipoles:

$$\mathbf{R} = \begin{pmatrix} \cos\beta_1 & 0 & \cdots & & \cdots & \cdots & 0 \\ \sin\beta_1 & 0 & \cdots & & \cdots & \cdots & 0 \\ 0 & \cos\beta_1 & 0 & & \cdots & \cdots & 0 \\ 0 & \sin\beta_1 & 0 & & \cdots & \cdots & 0 \\ & & & \ddots & & & \\ 0 & \cdots & \cdots & & 0 & \cos\beta_{p_1} & 0 \\ 0 & \cdots & \cdots & & 0 & \sin\beta_{p_1} & 0 \\ 0 & \cdots & \cdots & & 0 & \cdots & \mathbf{I}^{(2p_2)} \end{pmatrix} \tag{2.15}$$

where the fixed dipoles form angles β_k with respect to e_θ, $k = 1, \cdots, p_1$. And, $\mathbf{I}^{(2p_2)}$ is a identity matrix with the size of $2p_2 \times 2p_2$. If all the multidipoles are with the variable orientations, $r = 2p_2 = 2p$ and $\mathbf{R} = \mathbf{I}^{(2p)}$(Hämäläinen et al., 1993).

Before the minimizing of equation 2.10,the value of p_1 and p_2 need to be chosen for fixed- and variable-orientation dipoles seperately, and the correct model is selected as well. Then, the key problem is to solve the nonlinear optimization for $\mathbf{r}_d$, $d = 1, \cdots, p$ (Hämäläinen et al., 1993).

The algorithm for determining the location and orientation of multiple sources through the highly complex optimization can easily lead to inaccurate solutions if enough different initial dipole values were not tested (Papanicolaou, 2009). The methods applied so far are mainly based on heuristic methods, and the reasonable solutions depend on both the expertise and physiological intuition(Hämäläinen et al., 1993). The so-called MUSIC(multiple signal classification) and RAP-MUSIC(recursively applied and projected MUSIC) are efficient approaches identifying each source seperately with resursive procedures rather than searching for multiple sources simultaneously (Mosher and Leahy, 1999). The MUSIC approach is based on indentifying the multiple local maxima in a single function, while RAP-MUSIC implements a search for one source as the global maxima with a resursive procedure for the cost functions of multiple sources (Papanicolaou, 2009).

The above two methods are based on the equvalent current dipole(s) assumption for source estimation which has limitations in practice: First, there are difficulties in localizing extended sources with ECDs; secondly, it is difficult to estimate the number of dipoles in advance; and thirdly, the methods shows insensitivity to dipole time-courses and errors in dipole location, especially for deep sources.

2.2.6.3 Current-distribution models

In the current distribution models, the whole brain or cortical surface are assumed to be a *source space* composed of a large number of elements. A triangular mesh is generally applied to constitute the source space on the cortical surface, while tetrahedral or hexahedral lattices are used to represent the interior volumes of the head. Additionally, a single dipole is located on each vertex of the mesh or the lattice point. Since the number of unknown sources in source space (generally several thousand) are much more than the quantity of measurements from sensors, this model is actually a *underdetermined problem*, or *ill-posed problem*. In term of this ill-posed character, the calculation of the minimum of a cost function which provides the optimal source estimation should be based upon sufficient priors as constraints. This prior is essentially a model of expected current distribution. The smoothness of the source distribution (explained as variation of sources on the spatial field), is able to be used as one type of prior. The smaller norm indicates that the sources distributed in source space, while the larger norm indicates a less smooth source distribution. (Papanicolaou, 2009) (Hämäläinen et al., 1993).

1. **Minimum-norm estimation (MNE)**

   The minimum-norm method, derives its name from the minimization of the difference or norm between predicted and actual magnetic field measurement. The concept *smallest* emphasized here depends on both the condition of measurement and the minimization of cost function.

   According to the equation 3.2 and equation 2.5 of Biot-Savart Law, there is a linear relationship between the internal source distribution and the measured magnetic field outside of the scalp, which can be explained simply as:

$$\mathbf{B} = \mathbf{L}\mathbf{J} \tag{2.16}$$

   where **B** is a $m \times 1$ matrix representing the magnetic field measurement outside of head; **J** is $n \times 1$ source current matrix with fixed locations and orientation; and **L** is the leadfield with the size of $m \times n$, which accounts for the information of the conductivity distribution of the head as well as the geometry to connect **B** and **J**. Specifically, each column of leadfield **L** provides the forward solution for a single source to the measurement, in other words, it shows the signal produced by all the sensors for a single source alone with unit strength (Hauk, 2004).

   Thus, **B** obtained from the sensors' recording , **L** is generally determined by the head geometry. According to the equation 2.16, the principal problem is to solve the unknown source current distribution based on the ill-posed (or under-determined) character with a non-unique solution. This presents the possibility that the sources distribution **J** produced current measurement may contain any primary current distribution $\mathbf{J}_2$ which the measured sensors are not sensitive to, such as radial sources. This can be explained mathematically as follows (Hauk, 2004):

$$\mathbf{J} = \mathbf{J}_1 + \mathbf{J}_2 \tag{2.17}$$

$$\mathbf{L}\mathbf{J}_1 = \mathbf{B}; \quad \mathbf{L}\mathbf{J}_2 = 0 \tag{2.18}$$

A unique solution for this inverse problem stated in equation 2.16 is to consider both the constraints on the predicted sources as well as the constriaint on the magnetic field from predicted sources which can be explained as follows:

For the predicted sources, the constraints for unique solution can be indicated as:

$$\hat{\mathbf{J}} = \min_{\mathbf{J}}[(\mathbf{J} - \hat{\mathbf{J}}_0)^T \mathbf{C}_s (\mathbf{J} - \hat{\mathbf{J}}_0)] \tag{2.19}$$

where $\hat{\mathbf{J}}$ represents the estimated solution, $\hat{\mathbf{J}}_0$ is a priori approximation of the source solution, and $\mathbf{C}_s$ is a weighting matrix which provides the prior information with the source space, such as convariance of the source or the approximate estimation of the location.

Meanwhile, the constraint for the magnetic fields predicted from the estimated sources are as follows:

$$\hat{\mathbf{J}} = \min_{\mathbf{J}}[(\mathbf{L}\mathbf{J} - \mathbf{B})^T (\mathbf{L}\mathbf{J} - \mathbf{B})] \tag{2.20}$$

where $\mathbf{LJ}$ are the magnetic field produced by the estimated sources, while $\mathbf{B}$ is the measured magnetic field.

If the matrix $\mathbf{C}_s$ is positive definite which is invertible, the solution can be inferred as:

$$\hat{\mathbf{J}} = \hat{\mathbf{J}}_0 + \mathbf{C}_s^{-1} \mathbf{L}^T (\mathbf{L}\mathbf{C}_s^{-1}\mathbf{L}^T)^{-1} (\mathbf{B} - \mathbf{L}\hat{\mathbf{J}}_0) \tag{2.21}$$

If no prior source estimation is set, the equation can be reduced as follows:

The weighting matrix represents prior information about the source, which can be incorporated to locate the source accurately. In practice, this prior information can be obtained with the assisting from other brain imaging mathod, such as fMRI (AM. et al., 2000). Nevertheless, if there is no location bias for the source, or the source can be expected at any location in the source space, $\mathbf{C}_s$ can be set as the identity matrix which means all the location in source space are provided with equal weight. So that the equation above can be simplified as (Hauk, 2004):

$$\hat{\mathbf{J}} = \mathbf{L}^T(\mathbf{L}\mathbf{L}^T)^{-1}\mathbf{B} \tag{2.22}$$

This is the standard minimum-norm least-square estimate for $\hat{\mathbf{J}}$.

2. **Regularization method**

Since the simple minimum-norm estimation generally favors surface source estimation, the sources in deeper locations requires more power for generating a measurable signal at the sensor location. In this case, the leadfield normalization is applied to the minimum-norm method to improve the estimation for deep sources.

However, since noise is usually present in the measurement together with the sources, the constraint equation 2.20 mentioned above can be written as:

$$(\mathbf{L}\hat{\mathbf{J}} - \mathbf{B})^T\mathbf{C}_b(\mathbf{L}\hat{\mathbf{J}} - \mathbf{B}) = \varepsilon > 0 \tag{2.23}$$

where ε represents a part of the data that can not be explained clearly and is due to the noise. Here, $\mathbf{C}_b$ is a positive definite weighting matrix which reflects the known basic information of sensors or the *reliability* of the sensors (e.g. by their standard deviations or covariances).

When ε reaches the optimal value λ, the required *regulaization parameter* is obtained as λ. The minimum-norm equation can then be written as:

$$\hat{\mathbf{J}} = \hat{\mathbf{J}}_0 + \mathbf{C}_s^{-1}\mathbf{L}^T(\mathbf{L}\mathbf{C}_s^{-1}\mathbf{L}^T + \lambda\mathbf{C}_d^{-1})^{-1}(\mathbf{B} - \hat{\mathbf{J}}_0) \tag{2.24}$$

The weighted minimum-norm estimation is derived as the equation above (Hauk, 2004), (Wagner et al., 1996), (Anders et al., 1993), (Phillips et al., 2002). Additonally, without the priori model $\hat{\mathbf{J}}_0$, as well as assumed equal weight to all the sensors and the source space(both $\mathbf{C}_d$ and $\mathbf{C}_s$) , the equation can be written as:

$$\hat{\mathbf{J}} = \mathbf{L}^T(\mathbf{L}\mathbf{L}^T + \lambda\mathbf{I}_d)^{-1}\mathbf{B} \tag{2.25}$$

which is the standard regularized minimum-norm least-squares estimate.

2.2.6.4 Beamformers

- **Introduction**

  A beamformer is a actually a spatial filter that combines linearly the output of the sensors' array so that the signal of interest can be enhanced and the background noise is suppressed. In other words, a beamformer allows the source of interest to pass through each volume-grid node, or cortical surface, while the non-interested sources, i.e. *noise,* are rejected. (Papanicolaou, 2009).(Preissl, 2005) (Singh et al., 2002)

  The beamformers are based on the concentration of the current sources on specific target locations. The particular parameters of these spatial filters are selected so that certain properties of the current sources, such as location, resolution , etc. are properly optimized. A weight is assigned for each sensor as the scalar of the measured contribution. Based on all these weights , as well as the information of predetermined target locations, the strength and orientations of sources of interest can be estimated. (Hillebrand et al., 2005) (Papanicolaou, 2009)

  Beamformers are actually divided into two types, adaptive and non-adaptive beamformers. Generally, the non-adaptive beamformers use a fixed set of weights to combine the signals from the sensors in the array. For example, the location of the sensors in space and the wave direction of interest are primarily applied. In contrast, the adaptive beamformers apply the unfixed weight which combine the properties information of the signal directly acquired from the array of sensor. In this case, the rejection of unwanted signals can be effectively improved. In other words, the main feature of the adoptive Beamforming method apart from the non-adaptive method is to adjust its performance to suit differences in its environment. (Papanicolaou, 2009)(Preissl, 2005) (Singh et al., 2002) (Barry et al., 1988)

  Compared with the minimum-norm method, beamforming does not need the prior knowledge of the sources of interest, such as locations, and has better spa-

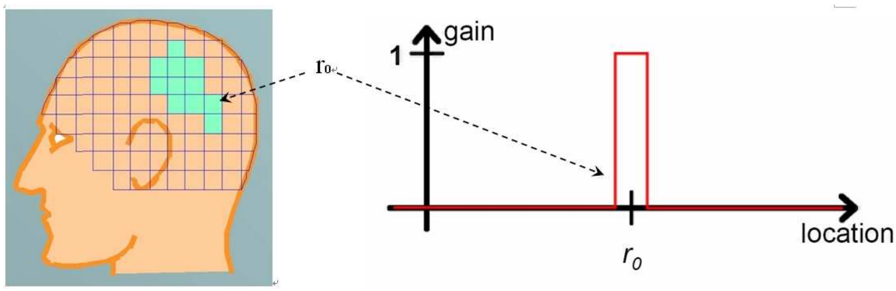

Figure 2.8: Beamforming methods as a spatial filter (Hillebrand et al., 2005)

.

tial resolution since they provide less overlap for the reconstructed sources, and are suitable with the estimation of both deep sources and superficial sources with no location bias. However, beamforming shows limitation in the case that estimated sources are correlated in the temporal field. In other words, erroneous and unstable reconstructed results are generated by applying the beamforming method for estimating the temporally correlated sources. This limitation is downplayed by many investigators who claims it is unlikely to have highly correlated brain activities in practice. Therefore, the beamforming method can be generally used(Papanicolaou, 2009). However, it indeed causes the inaccuracy of reconstruction for some estimation of temporally correlated sources, for example, 'the highly correlated brain activity involves auditory stimulation and the bilateral generators of the auditory m100 component, whereby bilateral activity is expected within milliseconds when stimuli are presented to a single ear'(Papanicolaou, 2009).

- **Filter design**

The noisy measurements of magnetic fields on the scalp stimulated by the internal sources can be represented as follows:

$$\mathbf{B}(t) = \mathbf{L}\,\mathbf{J}(t) + \mathbf{n}(t) \tag{2.26}$$

where $\mathbf{B}(t)$ indicates the measurement of magnetic field on time point t; $\mathbf{L}$ is the leadfield, $\mathbf{J}_(t)$ represents the current source at time point t. $\mathbf{n}(t)$ is the added noise in the time course assumed as zero-mean $(E(\mathbf{n}) = 0)$. Meanwhile, the current sources associated with different dipoles are assumed to be uncorrelated, giving the covariance matrix for measurement of magnetic fields in the time series:

$$\mathbf{C}(\mathbf{B}(t)) = E\{[\mathbf{B}(t) - \bar{\mathbf{B}}(t)][\mathbf{B}(t) - \bar{\mathbf{B}}(t)]^T\} \tag{2.27}$$

Assuming at a single time point t the estimated source signal on the specific voxel k is equivalent to the product of weights and measured magnetic field as follows:

$$\mathbf{J}_k(t) = \mathbf{w}_k\, \mathbf{B}(t) \tag{2.28}$$

For the weight $\mathbf{w}_k$ which governs the spatial filter of the spatial field Ω segmented as volumes, the value of power gain is set as 1 at specific voxel k_0 and zero elsewhere;

The ideal filter is

$$\mathbf{w}_k\, \mathbf{L}_k = \begin{cases} 1 & for \quad k = k_0 \\ 0 & for \quad k \neq k_0 \end{cases} \qquad k \in \Omega \tag{2.29}$$

The power at voxel is $\mathbf{S} = \mathbf{w}^T\mathbf{C}\mathbf{w}$ which is minimized to subject to $\mathbf{w}_k \cdot \mathbf{L}_k = 1$. then the beamformer weight can be calculated as:

$$\mathbf{w}_k = \frac{\mathbf{C}^{-1}\mathbf{L}_k}{\mathbf{L}_k^T\mathbf{C}^{-1}\mathbf{L}_k} \tag{2.30}$$

2.3 Kalman filter

2.3.1 Brief introduction

Since 1960s, the Kalman filter has been the subject of research and application based on the publication of R.E.Kalman (Kalman, 1960) on a recursive solution to the discrete-data linear filtering problem.

The Kalman filter is a set of mathematical equations providing an effective recursive means for estimating the state of a process by minimizing the mean of the square error. Basically, the Kalman filter is capable to estimate states in the past, present and future of the dynamic system with hidden states. Also, the Kalman filter can be applied to the estimation of the missing state, and the measurement of the estima- tion quality. The origin of the Kalman filter can be explained both in a probabilistic way and a computational way respectively.(Welch and Bishop, 2006)

Generally, the Kalman filter applied for the state estimation can be divided into two types, the discrete Kalman filter and the extended Kalman filter(EKF), which are used for describing the linear system and non-linear systems, respectively. The problem of interest is to estimate the state $x \in \mathcal{R}^n$ of a discrete-time controlled process that is governed by a *linear* stochastic difference equation. The discrete Kalman filter is applied in this case. However, the extended Kalman filter(EKF) is applied to the process where the relationship between measurement and estimation is non-linear. Since the MEG system in my research is assumed as the linear dynamic system, the discrete Kalman filter is introduced here(Welch and Bishop, 2006) (Brown and Hwang, 1992.) (Grewal and Andrews, 1993) (Sorenson, 1970).

2.3.2 The discrete Kalman filter

- The estimated process introduction

  The Kalman filter is applied to the general problem of trying to estimate the state $x \in \mathcal{R}^n$ of a discrete-time controlled process that is governed by a *linear* stochastic difference equation. In other words, two necessary linear models a dynamic model and a measurement model) to describe the process states are built as follows. The dynamic model, equation 2.31 describes the dynamic relationship between the different process states, while the measurement model,equation 2.32 describes the relationship between the measurement and the estimation. (Welch and Bishop, 2006):

$$x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1} \tag{2.31}$$

$$z_k = Hx_k + v_k \tag{2.32}$$

Here, x_k is the state at step k, u_{k-1} is the optional control input to the state x_k, z_k is the measurement on the step k. Note here that we only observe z_k and x_k is the hidden state which we would like to estimate it. While w_{k-1} and v_k are the noise for the estimation and measurement which are assumed to be independent with each other and with normal distributions separately:

$$p(w) \sim \mathcal{N}(0, Q) \tag{2.33}$$

$$p(v) \sim \mathcal{N}(0, R) \tag{2.34}$$

The noise covariance Q of dynamic process and the noise covariance R of measurement are assumed to be the constant value although they might change with the time step or the measurement in practice.

Also, A is the weight matrix which relates the state on the previous time step $k-1$ and the current time step k and models the dynamics of the system. B is the weight matrix to relate the optional control input u_{k-1} with the state x_k and H is the weight matrix which relates the state x_k and the corresponding measurement z_k. It is notable that A and H are assumed to be constant although they might change with the time step or the measurement in practice.

Following the introduction of (Welch and Bishop, 2006), the origins of Kalman filter can be explained in two ways, the computational origins and the probabilistic origins.

- The computational origins of Kalman filter

For the time step k, it is assumed that $\hat{x}_k^- \in \mathcal{R}^n$ is *a prior* state estimate at step k, and $\hat{x}_k \in \mathcal{R}^n$ is *a posteriori* state estimate at the step k with respect to the measurement z_k. Then, the estimate errors for *a prior* and *a posteriori* are set separately:

$$e_k^- \equiv x_k - \hat{x}_k^- \tag{2.35}$$

$$e_k \equiv x_k - \hat{x}_k \tag{2.36}$$

In terms of the estimate errors above, the *a prior* estimate error covariance is set as:

$$P_k^- = E[e_k^- e_k^{-T}] \tag{2.37}$$

and the *a posteriori* estimate error covariance is set as:

$$P_k = E[e_k e_k^T]. \tag{2.38}$$

The *a posteriori* state estimate $\hat{x}_k$ can be represented with a linear combination of an *a prior* estimate $\hat{x}_k^-$ and a weighted difference between an actual measurement z_k and the corresponding predicted measurement $H\hat{x}_k^-$ which shows as following(Welch and Bishop, 2006):

$$\hat{x}_k = \hat{x}_k^- + K(z_k - H\hat{x}_k^-) \tag{2.39}$$

It is notable that the difference $z_k - H\hat{x}_k^-$ is also called the *residual* which reflects the discrepancy between the predicted measurement $H\hat{x}_k^-$ and the actual measurement z_k. The larger value of the *residual* indicates the larger difference between $H\hat{x}_k^-$ and z_k, in contrast, the zeros of the *residual* means the two are in agreement completely.

In equation 2.39, weight matrix K is the *gain* which is used for minimizing the *a posteriori* error covariance 2.38, and the form of which is given by:

$$\begin{aligned} K_k &= P_k^- H^T (H P_k^- H^T + R)^{-1} \\ &= \frac{P_k^- H^T}{H P_k^- H^T + R} \end{aligned} \tag{2.40}$$

R and P^- are two components governs the changing trend of K. When the measurement error covariance R tends to be zero, the actual measurement z_k is more and more *trustable*, while the predicted measurement $H\hat{x}_k^-$ is less and less *trustable*. In other words, this can be indicated as:

$$\lim_{R_k \to 0} K_k = H^{-1} \tag{2.41}$$

And, when the *a priori* estimate error covariance P_k^- tends to be zeros, the predicted measurement $H\hat{x}_k^-$ is more and more *trustable*while the actual measurement z_k is less and less *trustable*. This can also be shown as:

$$\lim_{P_k^- \to 0} K_k = 0 \tag{2.42}$$

- The Probabilistic Origins of the Filter

  From the introduction of the probabilistic origins of the Kalman filter in (Jacobs, 1993) (Maybeck, 1979) (Brown and Hwang, 1992.), the *a posteriori* state estimate $\hat{x}_k$ and error covariance P_k can be written as:

$$E[x_k] = \hat{x}_k \tag{2.43}$$

$$E[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T] = P_k \tag{2.44}$$

  and the state distribution at time point k can be indicated as:

$$p(x_k|z_k) \sim \mathcal{N}(E[x_k], E[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T]) = \mathcal{N}(\hat{x}_k, P_k) \tag{2.45}$$

  where the *a posteriori* state estimate $\hat{x}_k$ represents the mean of the state distribution with respect of the condition 2.33 and 2.34 are satisfied. In addition, the *a posteriori* estimate error covariance P_k represents the variance of the state distribution.

- The Discrete Kalman Filter Algorithm

  In terms of the pre-knowledge of Kalman filter indicates above, the classical Kalman filter can be divided into two groups, *time update* part and *measurement* *update* part. There are equations which describes the current state and error covariance estimates to obtain the *a priori* state estimates for the next step of time series in *time update* part, shown as the following equation 2.46, 2.47:

$$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_{k-1} \tag{2.46}$$

$$P_k^- = AP_{k-1}A^T + Q \tag{2.47}$$

while the equations in the *measurement update* provide the corrected feedback which obtains an improved *a posteriori* estimate from the *a priori* estimate, showed as following equations 2.48, 2.49, 2.50:

$$K_k = P_k^- H^T (HP_k^- H^T + R)^{-1}) \tag{2.48}$$

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-) \tag{2.49}$$

$$P_k = (I - K_k H)P_k^- \tag{2.50}$$

In the other words, this two groups of Kalman filter can be described as *prediction* step and *correction* step respectively. Both these *predict-correct* algorithm which is used for solving problem numerically can be presented as Fig 2.9.

One appealing feature of the Kalman filter is its *recursive nature*. The process is repeated that estimating the new *a prior* state with respect to the previous *a posteriori* estimated state until the *a posteriori* error covariance is located on the acceptable region.

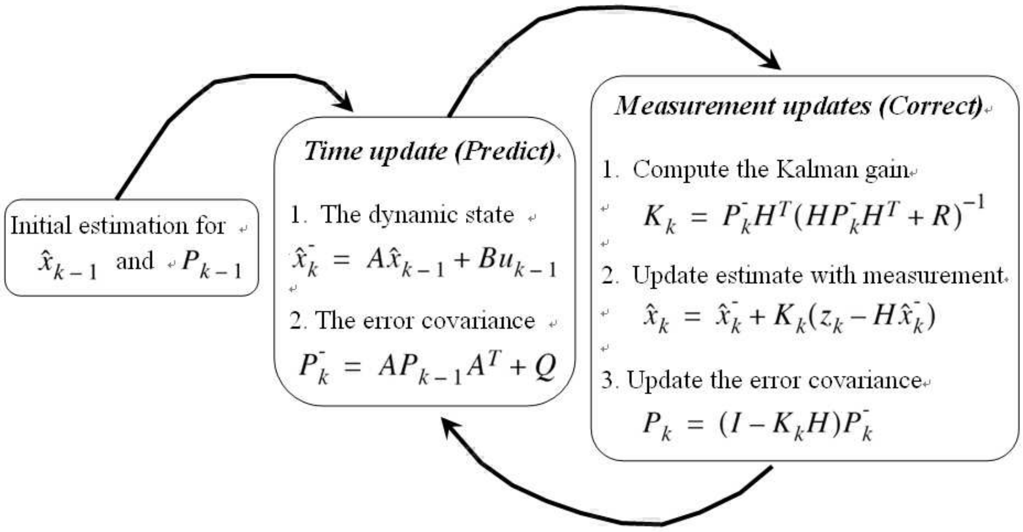

Figure 2.9: The discrete Kalman filter cycle. The *time update* indicates the current state estimate ahead in time. The *measurement update* indicates the estimation by an corresponding measurement at that specific time point

2.4 Kalman smoother

The Kalman filter above indicates the solution for the estimation of the state of the dynamic system with the Markov property that the state depends on the previous state but not any others. Based on the Kalman filter, there is no need to consider all the states at previous times , and, for the estimation of the state and the uncertainty(covariance) on specic time point t, it is feasible to obtain the solution from only the status on previous one time point $t-1$ as well as the noisy observation $x^\tau = z_1, , z_\tau$ for the specic time point t. It is notable that the difference between t and τ generally provides the process with variable uses. For instance, if τ is equal to the current time point t, the process is called *filtering*; if τ is smaller than t, the process is called *predicting*; and if τ is larger than t, the process is called *smoothing*. Here, from the explanation of Kalman lter presented above, the *Kalman* smoother equations are derived which is capable of predicting the state at the specific time point t with better accuracy(less noisy) by assuming that the state depends on the next state as well as the previous.

(Kalman, 1960), (Jazwinski, 1970,) . The aim is to calculate the probability:

$$p(x_t|z^\tau) \tag{2.51}$$

This probability in Eqn 2.51 is assumed to be the Gaussian distribution in which the main problem is focus on the calculation of its mean $\hat{x}_t^\tau$ and covariance P_t^τ:

$$\hat{x}_t^\tau = E[x_t|z^\tau] \tag{2.52}$$

$$P_t^\tau = E[\tilde{x}_t^\tau \tilde{x}_t^\tau|z^\tau] \tag{2.53}$$

where $\tilde{x}_t^\tau = x_t - \tilde{x}_t^\tau$ indicates the state prediction error.

Also, the Eqn 2.52 can be written as:

$$\hat{x}_t^\tau = E[x_t|z^\tau] = E[x_t|x_{t+1} = \hat{x}_{t+1}^\tau, x^\tau] \tag{2.54}$$

For calculating the mean $\hat{x}_t^\tau$ and covariance P_t^τ , firstly, we can write the density function as:

$$\begin{aligned} p(x_t, x_{t+1}|z^\tau) &= \frac{p(x_t, x_{t+1}, z^t, z_{t+1}, \cdots, x_\tau)}{p(z^\tau)} \\ &= \frac{p(z_{t+1}, \cdots, z_\tau|x_t, x_{t+1}, z^t)p(x_{t+1}|x_t, z^t)p(x_t|z^t)p(z^t)}{p(z^\tau)} \\ &= \frac{p(z_{t+1}, \cdots, z_\tau)p(x_{t+1}|x_t)p(x_t|z^t)}{p(z_{t+1}, \cdots, z_\tau|z^t)} \end{aligned} \tag{2.55}$$

Continuously, we can write the following function:

$$p(x_t|x_{t+1}, z^\tau) = \frac{p(x_t, x_{t+1}|z^\tau)}{p(x_{t+1}|z^\tau)} \tag{2.56}$$

With the calculation and inference from (Welling), the Kalman smoother equations are obtained as Eqn 5.21 , Eqn5.22 and Eqn 5.23 :

$$\hat{x}_t^\tau = \hat{x}_t^t + J_t(\hat{x}_{t+1}^\tau - \hat{x}_{t+1}^t) \tag{2.57}$$

$$J_t = P_t^t A^T [P_{t+1}^t]^{-1} \tag{2.58}$$

and

$$P_t^\tau = P_t^T + J_t(P_{t+1}^\tau - P_{t+1}^t) J_t^T \tag{2.59}$$

The way to applying the Kalman smoother equation is separated into two steps. Firstly, with the full set of term measurements, the Kalman filter is applied forward from the state at initial time point till the state at time τ is reached (where $t < \tau$). Then, the process is moved backward by applying the Kalman smoother equations until state at the time t is estimated. Since all the state factors, such as $\hat{x}_{x+1}^{t+1}$, $\hat{x}_{t+1}^{t}$, P_{t+1}^{t+1} and P_t^{t+1}, $t = 1 \cdots \tau$ are stored in the former step, it is easier for Kalman smoother equations to apply them directly in the later step (Welling).

Comparing with the state estimation by Kalman filter (in the former step), the estimated results indicate improved accuracy with less noise which is so-called *smoothing* since more measurement in the time sequence are applied for processing. The Kalman smoother effectively enhances the estimation of the hidden dynamic system.

2.5 EM algorithm

2.5.1 General introduction to the EM algorithm

In the Kalman filter model, there are a group of unknown parameters, such as [μ, Σ, A, B, R, Q], which may need to be estimated for further processing. EM algorithm is generally applied as the method for solving this.

Since its inception in 1977, EM algorithm has been widely applied as a general purpose method for maximum-likelihood estimation(MLE) in the variety of *incomplete data* problems. This name was been given by Dempster, Laird and Rubin in their fundamental paper in 1977 (Dempster et al., 1977). The full-name of EM algorithm is called the *Expection-Maximization* algorithm which indicates the two steps of the method, the *expectation step* or the *E-step* and the *Maximization step* or the *M-step*. The missing data is estimated in the former step by filling the unknown parameters with their expectation values. Then, in the latter step the new parameters are re-estimated in terms of the estimation of the missing data of the last step. This procedure is proceeded iteratively until reaching the convergence (McLachlan and Krishnan, 1996).

2.5.2 Maximum-likelihood estimation (MLE)

The EM algorithm is actually the extensive application of the interactive computation of maximum likelihood estimation(MLE). In this case, the Maximum-likelihood estimation(MLE) is introduced here firstly.

Governed by a set of unknown parameters Θ , there is a density function $p(x|\Theta)$ describing the distribution of state, for instance, p might be from the family of Gaussian distribution, and the set of parameters Θ is actually the mean and covariance. There is a set of the observation data sampled from the distribution above, showed as $\mathcal{X} = x_1, \cdots, x_N$ which is with the size of N , are assumed as independent and identically distributed with respect to the distribution p. Therefore, this density function of $\mathcal{X}$ can be written as following function:

$$p(\mathcal{X}|\Theta) = \prod_{i=1}^{N} p(x_i|\Theta) = \mathcal{L}(\Theta|\mathcal{X}) \tag{2.60}$$

This function, $\mathcal{L}(\Theta|\mathcal{X})$ is so-called the likelihood function of $\mathcal{X}$ in which the data $\mathcal{X}$ is xed but the set of parameters Θ are unknown. The goal of Maximum-likelihood estimation(MLE) is to estimate the appropriate value of Θ which is able to maximize the likelihood function $\mathcal{L}$. This can be presented as following function for estimating the Θ^* :

$$\Theta^* = \arg\max_{\Theta} \mathcal{L}(\Theta|\mathcal{X}) \tag{2.61}$$

Since the Eqn 2.60 generally leads to a complicated calculation, $log(L(\Theta|\mathcal{X}))$ is preferred for the maximization instead for easy analysis. Simply speaking, if the function $p(x|\Theta)$ is a single Gaussian distribution with the set of parameter $\Theta = (\mu, \sigma^2)$, the problem can be solved easily by setting the derivation to be zero and estimating the parameters μ and σ^2 directly afterward. however, if the distribution is more complicated than the single Gaussian distribution, it is usually to need more elaborate techniques rather than the analytical expressions (Bilmes, 1998).

2.5.3 EM algorithm

Finding the *incomplete data* is the key problem for the EM algorithm. In the later calculation, we assume $\mathcal{Z}$ to be the complete data which cannot be fully observed, $\mathcal{X}$ is denoted the *observation*. And $\mathcal{Y}$ denotes the *unobservable data* or *missing data*. The complete data is therefore represented as $\mathcal{Z} = (\mathcal{X}, \mathcal{Y})$ where the joint density function of complete data can be indicated as:

$$p(\mathcal{Z}|\Theta) = p(\mathcal{X}, \mathcal{Y}|\Theta) = p(\mathcal{Y}|\mathcal{X}, \Theta)p(\mathcal{X}|\Theta) \tag{2.62}$$

with respect to the Eqn 2.62, the relevant likelihood function can be defined as:

$$\mathcal{L}(\Theta|\mathcal{Z}) = \mathcal{L}(\Theta|\mathcal{X}, \mathcal{Y}) = p(\mathcal{X}, \mathcal{Y}|\Theta) \tag{2.63}$$

which is so-called the *complete-data likelihood*. Since the missing information is assumed to be unknown, but governed by an underlying distribution, the incomplete-data likelihood function can be referred as $\mathcal{L}(\Theta|\mathcal{X})$.

The two steps of EM algorithm can be explained as follows:

1. **E-step**

   Calculate the expectation of the complete-data log-likelihood $E[\log p(\mathcal{Z}|\Theta)] = E[\log p(\mathcal{X}, \mathcal{Y}|\Theta)]$ given the observation $\mathcal{X}$ and the parameter estimation in the current step with respect to the missing data $\mathcal{Y}$:

$$\begin{aligned} Q(\Theta, \Theta^{(i-1)}) &= E[\log \mathcal{L}(\Theta|\mathcal{Z})|\mathcal{X}, \Theta^{(i-1)}] \\ &= E[\log p(\mathcal{X}, \mathcal{Y}|\Theta)|\mathcal{X}, \Theta^{(i-1)}] \\ &= \int_{y \in \mathbf{Y}} f(y|\mathcal{X}, \Theta^{(i-1)}) \log p(\mathcal{X}, y|\Theta) dy \end{aligned} \tag{2.64}$$

   Here, the density function $f(y|\mathcal{X}, \Theta^{(i-1)})$ in Eqn 2.64 above is marginal distribution of the unobserved data given the observed data $\mathcal{X}$ and the parameters in the current step; and, $\mathbf{Y}$ is the possible space of y . Generally, this density function can be determined in terms of the current problem. Also, it can be applied as the form of $f(y|\mathcal{X}, \Theta^{(i-1)}) = f(y|\mathcal{X}, \Theta^{(i-1)})f(\mathcal{X}|\Theta^{(i-1)})$. Since $f(\mathcal{X}|\Theta^{(i-1)})$ is not depend on Θ, the density function is actually not affected by this extra factor.

2. **M-step**

   Choose parameter Θ^i to be any value of $\Theta \in \Omega$ (where Ω is the parameter space) that maximizes $Q(\Theta, \Theta^{(i-1)})$ in the former step(Maximum likelihood):

$$\Theta^{(i)} = \arg\max_{\Theta} Q(\Theta, \Theta^{(i-1)}) \tag{2.65}$$

   After $\Theta^{(i)}$ is obtained, the **E-step** and **M-step** are then carried out again with $\Theta^{(i)}$. These two steps are processed iteratively as each iteration is guaranteed to increase the log-likelihood until the difference (McLachlan and Krishnan, 1996):

$$L(\Theta^{(i)} - \Theta^{(i-1)}) \tag{2.66}$$

   become convergence.

EM algorithm and its extensions are the standard tools for applying the statistical methods to solve the incomplete data problems currently. It has been widely used for variable practical implements. For instance, medical imaging, regression, robust statistical modeling , survival analysis, factor analysis, nite mixture analysis, and so on (Bilmes, 1998), (McLachlan and Krishnan, 1996).

2.6 Bayesian image super-resolution

2.6.1 Introduction to super-resolution

Super-resolution(also written as superresolution in some articles) is one of the classical computer vision methods which has important applications in the field of remote sensing, satellite imagery, medical imaging, military surveillance and face recognition. The principle of the method is to reconstruct the high-resolution image from a set of low-resolution images. In other words, it is possible to estimate the high-frequency information of the scene above the Nyquist limit of the individual source images when the relevant distorted low-resolution images are provided(Tipping and Bishop, 2007).

In super-resolution, the low-resolution images are assumed as discretized versions of a high-resolution image with various distortions in the production of the low-resolution images. Further more, there are a number of different high-resolution images which may generate the set of known low-resolution images given which provides plenty of possible solutions to the reconstruction. It is important to determine appropriate forms of prior regularization, based on the practical situation, so that the optimal solution can be identified with reasonable parameters of the model . The model is applied as the constraint for the final high-resolution image. These constraints make the problem more tractable so that an appropriate solution of high-resolution images are more likely than all others.

In the problem of super-resolution, information from the low-resolution pixels is crucial in order to generate an accurate high-resolution image. Image registration is required on the subpixel level. Moreover, since the process is to reconstruct the frequency information above the Nyquist limit of the low-resolution images, the pixels of each low-resolution images should not be located on the same grid(so-called co-located) when there is no prior as the restriction. Otherwise, there will not be further information which can be extracted from these known low-resolution images which leads to poor reconstruction of the high-resolution image. If that is the case, the best method of reconstruction is to average the information of the pixels on each low-resolution image which at least obtains a denoised result(Dalton, 2004).

Before applying the method of super-resolution, the relevant low-resolution images can be obtained by either of the following methods. Firstly, they may be generated by the infinitely high-resolution real world,such as: a hand-hold camera, or a detector array which is not sufficiently dense to adequately sample the scene with the desired field of view . Besides, they can be produced from the innately high-resolution images with the known transformation, such as rotation, downsampling and blur (Tipping and Bishop, 2007), (Dalton, 2004). The evaluation of the reconstructed result of super-resolution depends on whether there is reliable relevant high-resolution image existing. If the low-resolution images are produced by a sensor as the first case explained above, it is impossible to find the original high-resolution image for comparison. If a high-resolution is available, it is reasonable to use the known high-resolution image for testing. By downsampling the low-resolution images via a set of known transformations , and using the super-resolution to reconstruct the high-resolution image we can compare with the original high-resolution image. However, there is still problem here since it is difficult to make sure the downsampling actually models the physical sensor.(Irani and Peleg, 1991).

2.6.2 Previous work on super-resolution

Improvement of image resolution depends on the physical properties of the sensors, such as the spatial response, optics and the density of the detectors. In the ideal case, downsampling is the only difference from the low-resolution image to the high-resolution image. However, the image motions may also be present, such as translation, rotation, or more complex geometric distortions. This is similar to the situation to take the images continually of the same subject using a hand-held camera so that the distortions, eg, translations, rotations, are produced in these images. It is worthy of note that the situation that the scene changes itself is not considered here(Tipping and Bishop, 2007).

This problem of image reconstruction has been addressed by a number of algorithms. The earlier research on super-resolution dates back to the work of a frequency domain approach by Tsai and Huang (Huang and Tsai, 1984). Since then, there have been a number of papers published which the problem.

D. Gross (Gross, 1986) estimated the high-resolution image with the assumption that both the imaging process and precise relative shifts of the input pictures are known. Then, the interpolation is applied for merging a set of the low-resolution pictures and a blurred image is obtained with higher spatial sample rate. A restoration filter is built by applying pseudo-inverse techniques to a matrix representing the blur operator. It is directly used for de-blurring that image to obtain the high-resolution image (Irani and Peleg, 1991).

The imaging process of the super-resolution can be represented as the following model (Keren et al., 1988):

$$g_k(m,n) = \sigma_k(h(f(x,y)) + \eta_k(x,y)) \tag{2.67}$$

where g_k is the k_{th} observed (low-resolution) image, f is the original scene which is the desired image(high-resolution), (x,y) is representing the pixel coordinate for the high-resolution image, while (m,n) is representing the pixel on the low-resolution image after reconstruction,h is a blurring operator, σ_k is the nonlinear function which describe the relation between the high-resolution function and low-resolution image in $k_t h$ frame , η_k is the added noise;

Peleg and co-workers approach (Peleg et al., 1987) (Keren et al., 1988) is based on the inversion of a transform from a assumed high-resolution image to a sequence of simulated low-resolution images. Specically, the high-resolution image above comes from an initial guess, while an error function measures the difference between simulated low-resolution images and the actual ones observed. The results of this approach shows plausibility and high-sensitivity on the noised-images in practice.

Irani and Peleg (Irani and Peleg, 1991) described a approach which is inspired from the reconstruction of computer aided tomography(CAT), which has a resemblance to superresolution. ”In tomography, images are reconstructed from their projections in many directions”. This property can be directly applied to the super-resolution since the multiple low-resolution images can be assumed as the projections of the different images of the same scene and are used for the reconstruction of high-resolution image via the approach. The low-resolution images are registered firstly with the uniform motion of translation and rotation by a proven method. The initial high-resolution image is guessed with respect of the information above . Additionally, it is used for simulating a set of synthetic low-resolution images corresponding to the actual low-resolution images with the Point Spread Function(PSF) of the sensor manually measured by a control image. Ideally, if the recovered high-resolution is correct, the simulated low-resolution images should be as same as the actual low-resolution images. The error function between these two groups of images is recursively optimized to recover the best high-resolution image. Additionally, this approach shows good results as long as the image can be divided into regions each of which is subjected to a uniform motion (Keren et al., 1988).

2.6.3 MAP method of super-resolution

Maximum a posterior(MAP) estimation is popularly used in super-resolution. Firstly, the initial registration of a set of low-resolution images is found and kept fixed in the process. A probabilistic model describing the high-resolution image is generated and maximum likelihood is applied to find the high-resolution image. However, some tricky situations may exist. There is not sufcient high-frequency information obtained from the low-resolution images if the high-resolution image contains too few pixels, or it becomes ill-conditioned if the high-resolution image contains too many pixels. In this case, the a prior distribution over the high-resolution image can be applied as a regularization term. With these regularization terms, the maximum likelihood solution is regularized and the problems above are tackled(Nguyen et al., 2001),(Smelyanskiy et al., 2000), (Capel and Sserman, 2000),(Hardie and Barnard, 1997).

2.6.4 Bayesian image super-resolution

Tipping and Bishop (Tipping and Bishop, 2007) have tried to improve the super-resolution by applying Bayesian method. Within their work, the Bayesian super-resolution shows a resemblance to the MAP approach as both of them are using a Gaussian prior, moreover, optimizing the registration parameters(including the transformation and rotation of the low-resolution images) are part of the maximization process. However, the Bayesian super-resolution method also has distinguishing properties beyond the previous approaches.

Firstly, with this Bayesian treatment of the super-resolution, the image registration parameters can be estimated in terms of the Bayesian marginalization on the unknown high-resolution image. In this case, the registration information of the low-resolution images, such as rotation, transformation and even the downsampling value, can be estimated beforehand. Additionally, the unknown point spread function (PSF) can also be estimated before the reconstruction of the high-resolution images. The point spread function (PSF) is applied as the process to obtain the low-resolution images by smoothing the high-resolution image. In previous approaches, PSF is generally assumed as known in advance. For instance, the PSF is estimated only by the low-resolution images and is kept fixed in the imaging process (Capel and Sserman, 2000) , or is approximately measured from the simulated process of scanning and imaging (Irani and Peleg, 1991). Whereas, this assumption does not work realistically in practice since the PSF is not able to be determined accurately without the information of the high-resolution image. The Bayesian marginalization provides a coherent and single framework in which the PSF can be determined along with the registration parameters as well as the high-resolution image. This gives more reasonable estimation assumptions for the image reconstruction (Tipping and Bishop, 2003).

Instead of the registration parameters, PSF and high-resolution image are estimated and optimized in a joint process as MAP super-resolution performs, Bayesian marginalization allows the registration parameters as well as PSF to be estimated in advance. With the optimizations of them, the high-resolution image can be reconstructed with accuracy. Also, Tipping and Bishop (Tipping and Bishop, 2003) presented the positive results of the Bayesian super-resolution by comparing with super-resolution via MAP.

In our study, the Bayesian super-resolution has been applied for the MEG source reconstruction distributed on the cortical mesh with high spatial resolution. With respect to the advantages of Bayesian marginalization illuminated above, the relevant setting and the estimation process of the reconstruction approach can be proceeded realistically and accurately, which provides further possibility to improve the quality of the MEG signal reconstruction.

Chapter 3

Basis Functions Source Model Applied to MEG Source Reconstruction

3.1 Brief introduction

Magnetoencephalography (MEG) is a new and non-invasive technique for the functional imaging of the human brain that has been widely used in both research and clinical application, such as intractable epilepsy, schizophrenia, depression, Parkinson's and Alzheimers diseases. The principle of the technique is to measure the magnetic field surrounding head that via the extremely sensitive sensors located outside the scalp, i.e. superconducting quantum interference devices (SQUIDs), which is shown in Fig 3.1. The measured magnetic field is mainly generated by the electronic activity in the brain (Preissl, 2005) (Kishida, 2009) (Srikantan et al., 2006). Based on the corresponding MRI scan, MEG produces a spatial-temporal pattern of the electronic activity in the cortex. Although techniques, such as fMRI, show outstanding spatial resolution, MEG provides superior temporal resolution that complements the weakness of brain imaging in the time domain (Rodriguez et al., 2003) (Baryshnikov et al., 2004).

The MEG source reconstruction from the measured magnetic field is typically an ill-posed inverse problem that is theoretically insoluble without additional information (Preissl, 2005). By now, there are many classical methods exist in this field which have been applied widely, such as the beamfroming method (Barry et al., 1988) (Rodriguez et al., 2003), and the minimum-norm method. However, there are limitations in

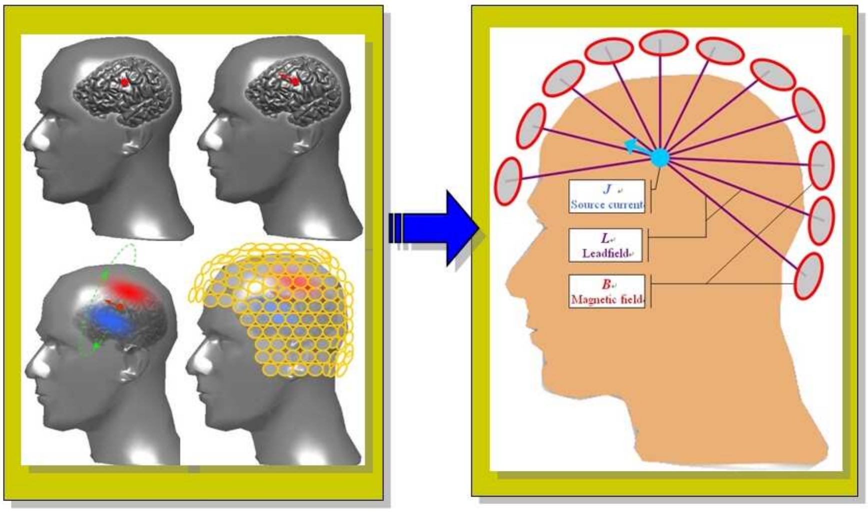

Figure 3.1: This figure shows the feature of MEG technique and the principle components of MEG data processing. The picture on the left indicates the origin of MEG: The measured magnetic field is mainly generated by the electronic activity in brain. And MEG is applied to measure the magnetic field surrounding head via the extremely sensitive sensors located outside scalp; The right picture indicates three principle components govern the MEG data processing: the measurement of magnetic field from sensors, denoted as ’B’, the current source ’J’ inside of the brain with individual direction and strength and the leadfield configuration ’L’ which connects the linear relationship between ’J’ and ’B’.

the accuracy of data using these classical methods. For instance, '*under the condition that in certain states of anesthesia, coma, and epilepsy, the beamformer formulation may prove to be in conflict with the actual state of the brain*' (Preissl, 2005). These algorithms can produce implausible results, which means that there is a gap between the actual dynamic state of the brain and the result of these methods, which affects the reliability of the MEG technique in clinical applications.

Therefore, this topic aims to explore a new solution to tackle the accuracy problem discussed above and attempts to bridge this gap. In this paper, we try to implement the MEG spatial-temporal source reconstruction through the global basis function source model. This chapter has been organized as follows. First, the forward model is introduced as well as a general description of the physics involved. All approaches of MEG signal reconstruction are based on the essential knowledge of the forward formula. Then, we demonstrate the process of cortical mesh extraction based on MRI scan and discuss the structure of each component of the model. Then, we introduce basis function source model as the solution of source reconstruction. Finally, this extended source model has been implemented to solve the inverse problem for MEG. Moreover, the robust stability of this MEG reconstruction solution is investigated in two ways. One is to compare it with the classical method of minimum-norm. The other is to apply the algorithm to signals with varying noise levels. The results show robustness to noise interference and better performance than minimum-norm. This method provides a new approach to the MEG signal reconstruction.

3.2 Forward problem

The concept of MEG sensing is to detect currents flowing in the brain from the magnetic flux recorded at a number of superconductive coils placed near the scalp. The magnetic field generated at a location **r** on the scalp is given by the Biot-Savart law:

$$\mathbf{B}(\mathbf{r}) = \int_{\Omega'} \frac{\mu_0}{4\pi} \frac{\mathbf{j}(\mathbf{r'}) \times (\mathbf{r} - \mathbf{r'})}{|\mathbf{r} - \mathbf{r'}|^3} d\Omega' \qquad (3.1)$$

Where Ω' is the volume in which the currents reside. Under the spherical sensor model, the coils are placed radially around the origin of the coordinate system, and so the normal to coil i is given by $\mathbf{r_i}/r_i$. $\mathbf{r}_i$ is the position of the coil. It is noticeable that

'J' represents the total current (primary currents + volume currents)(Barth et al., 1986). As the measurement is assumed on only the radial component of the magnetic field at single homogeneous spheroid, the majority of contributions of the volume currents vanish and the MEG measurement are only from the primary term approximately in this case. The magnetic flux detected by coil i is then:

$$\begin{aligned} b_i &= \frac{\mathbf{r}}{r} \cdot \mathbf{B}(\mathbf{r}_i) \\ &= \int_{\Omega'} \frac{\mu_0}{4\pi} \frac{\mathbf{r} \cdot \mathbf{j}(\mathbf{r}') \times (\mathbf{r} - \mathbf{r}')}{r_i |\mathbf{r} - \mathbf{r}'|^3} d\Omega' \\ &= \int_{\Omega'} \frac{\mu_0}{4\pi} \frac{(\mathbf{r} - \mathbf{r}') \times \mathbf{r}}{r |\mathbf{r} - \mathbf{r}'|^3} \cdot \mathbf{j}(\mathbf{r}') d\Omega' \\ &= \int_{\Omega'} \mathbf{l_i}(\mathbf{r}') \cdot \mathbf{j}(\mathbf{r}') d\Omega' \end{aligned} \tag{3.2}$$

where $\mathbf{l_i}(\cdot)$ is *leadfield* of coil i (shown in Fig 3.2), with

$$\mathbf{l_i}(r)' = \frac{\mu_0}{4\pi} \frac{(\mathbf{r}_i - \mathbf{r}') \times \mathbf{r}_i}{r_i |\mathbf{r}_i - \mathbf{r}'|^3} \tag{3.3}$$

The problem is therefore essentially a linear one; the coil flux is a linear combination of the leadfield components and the currents. And, the leadfield $\mathbf{l}_i(\mathbf{r}')$, the factor indicates the connectivity between the measrement of magneticfield at $\mathbf{r}_i$ and the source location $\mathbf{r}$ can be pre-computed with the expression of the product of radial detectors' information and the constant permeability of the head.

3.3 Cortical mesh extraction

3.3.1 Graph representation of mesh

The discrete structure of cortical surface can be expressed as a triangulated mesh M that can be used to approximate the cortical surface embedded in Euclidean space $\mathbb{R}^k$. It is composed of a topological part $M = (V, E, F)$ and a geometrical realization $\mathcal{M} = (\mathcal{V}, \mathcal{E}, \mathcal{F})$ (Gabriel, 2007).

The topology M of the mesh is composed of : - Vertices: this is an abstract set of indices $V \simeq 1, \ldots, N$;

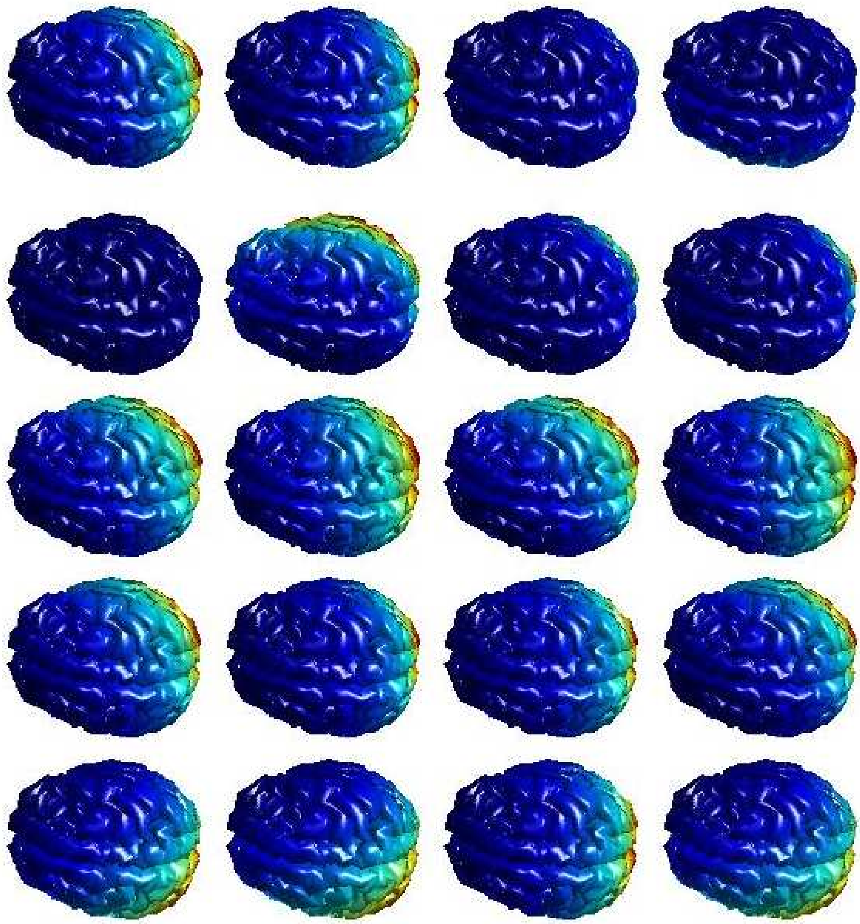

Figure 3.2: Each single plot shows the pattern of leadfield distributed on a surface reconstructed by 248-sensor points for an single mesh vertex. These 20 plots demonstrate the leadfield pattern for first 20 vertices of cortex mesh. The region responding signal from strong to weak on the color map is represented by the color from red to blue.

- Edges: this is a set of pair of vertices $E \subset V \times V$ which is assumed to be symmetric:

$$(i, j) \in E \Longleftrightarrow i \sim j \Leftrightarrow (j, i) \in E. \tag{3.4}$$

- Faces: This is a collection of 3-tuples of vertices $F \subset V \times V \times V$ with the relationship between any two of the three:

$$(i, j, k) \in F \Longrightarrow (i, j), (j, k), (k, i) \in E. \tag{3.5}$$

with the assumption that no isolated edges exist:

$$\forall (i, j) \in E, \exists k, (i, j, k) \in F. \tag{3.6}$$

The adjacency matrix A can be used to express the connection relationship(if they connect as a edge) between any two vertices of the mesh. A is a large sparse symmetric matrix where $A_{ij} = 1$, if $(i, j) \in E$, and $A_{i,j} = 0$, otherwise.

Meanwhile, with the information of the vertices and edges of mesh above, a undirected graph $\mathfrak{G} = (V, E)$ is constructed for the representation of the cortex. The geometric realization $\mathcal{M}$ is defined through the spacial localization of the set of vertices, $\mathcal{V}$, which in our study is stored as a $N \times 3$ matrix. N is the number of vertices with each row $[\mathcal{V}_{i,1}, \mathcal{V}_{i,2}, \mathcal{V}_{i,3}]$ stores the localization information of ith vertex in 3D. Additionally, the face $\mathcal{F}$ is stored as a $M \times 3$ matrix where M is the number of faces and a row $[\mathcal{F}_{j,1}, \mathcal{F}_{j,2}, \mathcal{F}_{j,3}]$ represents the indices of a face. $\mathcal{F}_{j,1}$, $\mathcal{F}_{j,2}$ and $\mathcal{F}_{j,3}$ indicate the indices of the vertices which construct the face j. $\mathcal{M}$ can be displayed as a 3D surface on the computer screen. Fig 3.3 shows the 3D display of the cortical mesh, with a zoom on the faces of the mesh.

3.3.2 Obtain the triangular mesh of grey matter from MRI

The entire 3D brain volume is a large and detailed structure and it difficult to accurately reconstruct currents within this volume using a small number of magnetic fluxes at the coils. To simplify the problem, we can assume that the currents flow only in the cortex, the outside surface of the brain (the grey matter). In other words, This essentially reduces the problem to a reconstruction problem over the cortex surface. The current

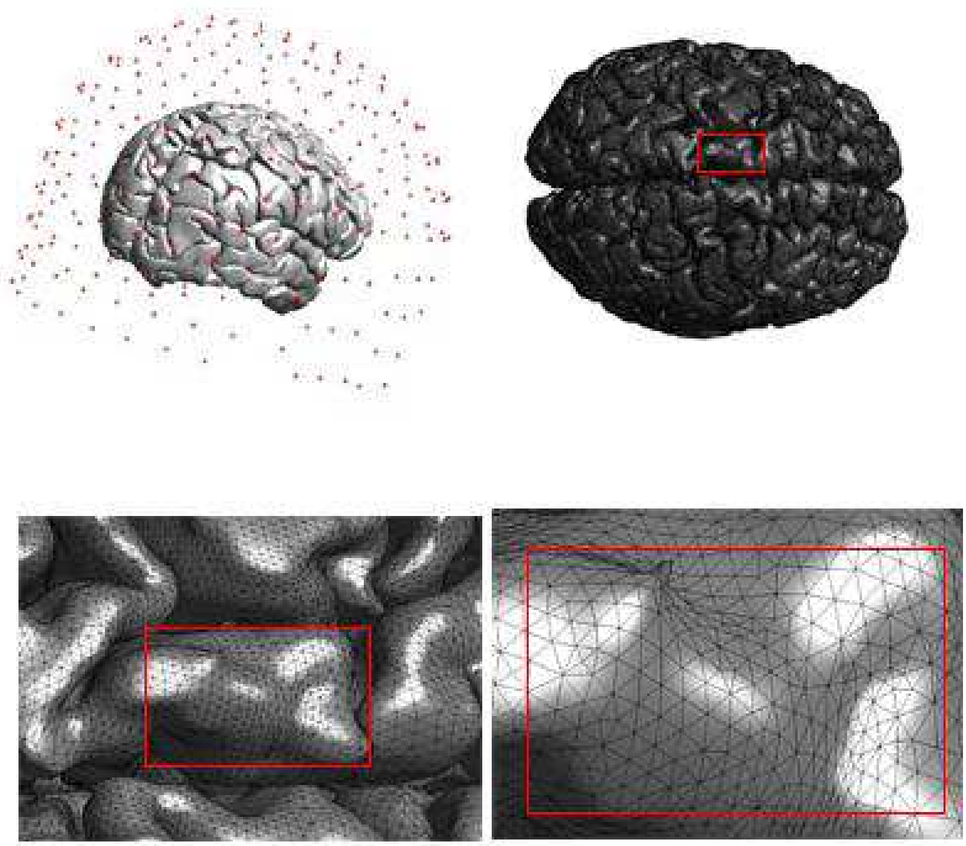

Figure 3.3: Top Left: shows the outside surface of the brain (the grey matter) and the sensor set located outside of cortex; Top Right: extracted mesh of the cortex from MRI using FreeSurfer, the resolution of the mesh is: 262,658 vertices, 525,308 faces (the part with red circle is emphasized for observation); Bottom: these two figures show the zoomed images of the emphasizing part of the mesh figure of Top Right.

sources are assumed to distributed on the whole cortical surface rather than the brain volumes, and all the current sources occur on the deep volumes are projected on the cortical surface. This approximation simplified the problem, however, it may generate the uncertain inaccuracy when the deep current sources are projected on the cortical surface. In order to construct a model of the cortex, we need a structural scan of the brain, which is achieved through a magnetic resonance imaging (MRI) scan. This scan is usually taken when an MEG scan is conducted since it is used to relate MEG responses to structural brain features.

Extraction of the cortex from an MRI scan is a well studied problem and there are a number of software tools which can perform this task. We use *FreeSurfer* (5.0.0) for this process.

Freesurfer is a set of software tools specifically for reconstruction of the brain's cortical surface from the structural MRI, as well as embedding the functional MRI data onto the reconstructed cortical surface, based on the study of the cortical and sub-cortical anatomy. The tools recognize and construct models of the boundary between the cortical gray matter, white matter as well as the pial surface. Based on these reconstructed model, an array of anatomical measures is generated, e.g. cortical thickness, surface area, curvature, and surface normal at each point on the cortex (Dale et al., 1999). For the better visualization, the surfaces can be inflated and/or flattened (Fischl et al., 1999). Moreover, *a cortical surface-based atlas has been defined based on average folding patterns mapped to a sphere*. Based on a high-dimentional non-linear registration, the surfaces can be aligned with this atlas. The spherical atlas naturally forms a coordinate system in which point-to-point correspondence between subjects can be achieved (Fischl and Dale, 2000). Since the MEG research is based on the large-sized data analysis, *Freesurfer* is very ideal as its pipeline is automated.

With the application of this software, a mesh description of the cortical surface is generated in terms of a set of 3D points and and adjacency matrix which describes the topology of the surface (5.0.0). The resulting mesh which is assumed to be an undirected graph are showed in Fig 3.3.

Finally we must perform an alignment step to bring the mesh in registration with the MEG data. This is achieved using fiducial markers in the MRI and MEG scans. The result of this process is an adjacency graph $\mathbf{A}$ describing the cortex topology as well as geometry, and aligned with the MEG data.

The resulting mesh defines a discretization over the cortex of the brain. In this case, the neural current sources are assumed located on the vertices of the cortical mesh. On each vertex, there is one current vector with independent strength and direction, showed as Fig 3.4. In the case that no current evoked on that region of cortical mesh, the current vector on the vertices there shows zero for both direction and strength. Our task is then not to find a continuous current distribution, but rather to find an estimate of the current at each discrete points, i.e. each vertex of the mesh. We can therefore formulate a modified forward problem

$$\mathbf{b}_i = \sum_{n \in V_N} \mathbf{l}_i(\mathbf{x}_n)\mathbf{j}(\mathbf{x}_n) \tag{3.7}$$

where V_N is the vertex set of the mesh, and x_n is the position of the n^{th} vertex. If we define the leadfield matrix as:

$$\mathbf{L} = \begin{pmatrix} l_{1,x}(\mathbf{r}_1) & \cdots & l_{1,x}(\mathbf{r}_N) & l_{1,y}(\mathbf{r}_1) & \cdots & l_{1,y}(\mathbf{r}_N) & l_{1,z}(\mathbf{r}_1) & \cdots & l_{1,z}(\mathbf{r}_N) \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ l_{i,x}(\mathbf{r}_1) & \cdots & l_{i,x}(\mathbf{r}_N) & l_{i,y}(\mathbf{r}_1) & \cdots & l_{i,y}(\mathbf{r}_N) & l_{i,z}(\mathbf{r}_1) & \cdots & l_{i,z}(\mathbf{r}_N) \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ l_{I,x}(\mathbf{r}_1) & \cdots & l_{I,x}(\mathbf{r}_N) & l_{I,y}(\mathbf{r}_1) & \cdots & l_{I,y}(\mathbf{r}_N) & l_{I,z}(\mathbf{r}_1) & \cdots & l_{I,z}(\mathbf{r}_N) \end{pmatrix} \tag{3.8}$$

together with the cortical distributed current distribution:

$$\mathbf{J} = (\mathbf{J}_x, \mathbf{J}_y, \mathbf{J}_z)^T \tag{3.9}$$

where the number of sensor arrays is I, and the number of the mesh vertices is N. We can write the linear forward problem as:

$$\mathbf{B} = \mathbf{LJ} \tag{3.10}$$

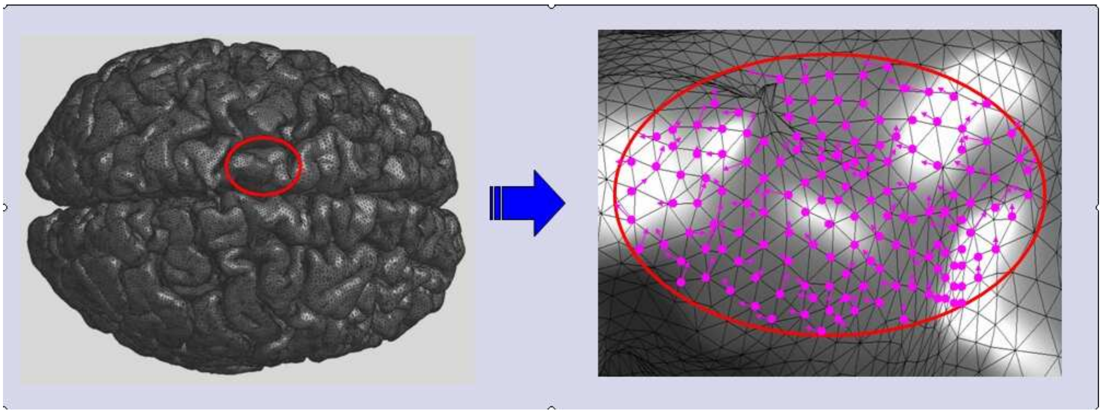

Figure 3.4: This figure indicates the assumption of current distribution. The neural current sources are assumed discretized on the vertices of the cortical mesh. On each vertex, there is one current vector with independent strength and direction. If there are no current sources evoked in that region, the current vectors there shows zero to both strength and orientation.

3.4 Geometrical expression of cortex by basis functions

3.4.1 The graph Laplacian

The cortical mesh Laplacian plays important role in our MEG current source reconstruction algorithm. As a branch of the mathematics that is concerned with characterizing the structural properties of graphs using the eigenvectors and eigenvalues of the adjacency or Laplacian matrices, the Laplacian of the graph has been widely studied by the spectral graph theory (Chung, 1997), (Cvetkovic et al., 1997).

The eigendecomposition of a graph provides us with a set of eigenvalues and eigenvectors which describe the structure of the graph. We begin by constructing the Laplacian of the graph (Gabriel, 2007):

$$\mathcal{L} = \mathcal{D} - \mathcal{A} \tag{3.11}$$

where $\mathcal{D}$ is the degree matrix, a diagonal matrix represents the connection degree of each vertex showing in the diagonal elements), which is also the combinatorial weights

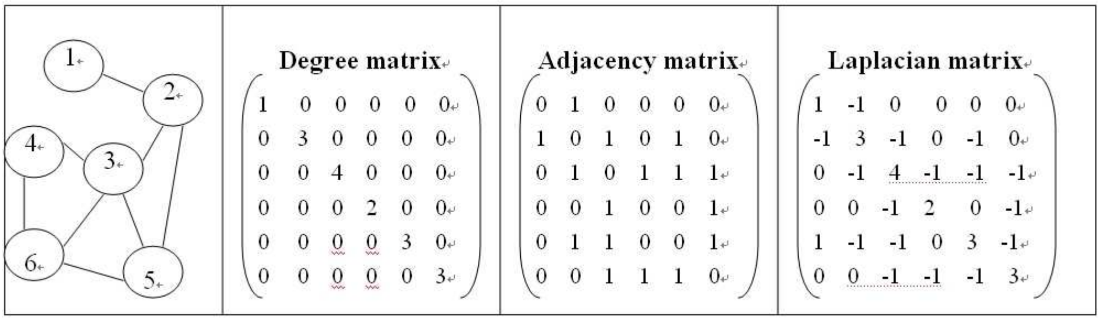

Figure 3.5: The figure shows an example to introduce the Laplacian matrix, adjacency matrix and degree matrix for a graph. A patch of graph (V=6 , E=8) is showed in the first grid. According to this graph, the associated degree matrix is extracted in the second grid. It is a diagonal matrix with connection degree of each vertex showing in the diagonal elements. The corresponding adjacency matrix is showed in the third grid. It is a symmetrical matrix and each element indicates the adjacency relation between related two vertices. The corresponding Laplacian matrix is showed in the fourth grid which is a matrix representation of the graph which is calculated by the difference between degree matrix and adjacency matrix.

of the mesh depends only on the topology (V, E) of the mesh;

$$\forall(i,j) \in E,\ \omega_{i,j} = 1 \quad ;\ \mathcal{D} = diag_i(d_i),\ \textit{with}\ d_i = \sum_j w_{ij} \tag{3.12}$$

while $\mathcal{A}$ is the adjacency matrix. The Fig 3.5 gives an small example of Laplacian matrix, adjacency matrix and degree matrix of a graph. We then compute the eigendecomposition.

3.4.2 Analogy of basis function for the cortical mesh: Laplacian eigenvectors corresponding to the smallest eigenvalues

We extend the idea of spatial basis functions to develop the current source model which represents the neural current arbitrary spatial distribution on the cortex. This model describes the current distribution using a set of global basis functions (Partha and Mitra, 2005). In fact, there are various types of basis function which can be applied for the solution above. In particular, spherical harmonic basis function, which apply to a spherical head model (Partha and Mitra, 2005). Here, we develop basis function specifically for our non-spherical cortical mesh. In the light of graph theory, we choose the eigenvectors corresponding to the smallest eigenvalues as the analogous of basis functions, showed as the Eqn 3.13.

$$\mathcal{L} = \sum_i \lambda_i \phi_i \phi_i^T \tag{3.13}$$

Here, i is the index of the mesh vertices. The eigenvectors are orthogonal $\phi_i^T \phi_i = 1$ and naturally form a set of basis functions over the graph. We can therefore use these to reconstruct any signal over the surface of the cortex. There are some of the benefits of using these basis functions: firstly, they are tailored to the cortical surface mesh; secondly, they are including the information of the topology of the cortical mesh; thirdly, the scale of the basis function set can be selected depending on the eigenvalues straightforwardly.

The current signal is constructed as three components on x, y and z orientation:

$$j_x(i) = \sum_{t=1}^{T} a_{xt} \phi_t(i); \tag{3.14}$$

$$j_y(i) = \sum_{t=1}^{T} a_{yt}\phi_t(i); \tag{3.15}$$

$$j_z(i) = \sum_{t=1}^{T} a_{zt}\phi_t(i); \tag{3.16}$$

where T is the number of the Laplacian eigenvectors corresponding to the smallest eigenvalues s we choose.

According to the Eqn 3.14, Egn 3.15 and Eqn 3.16, the currents J can be written as :

$$\mathbf{J} = \tilde{\mathbf{\Phi}}\mathbf{a} \tag{3.17}$$

where

$$\mathbf{\Phi} = (\Phi_1 \ \cdots \ \Phi_t \ \cdots \ \Phi_T) \tag{3.18}$$

$$\tilde{\mathbf{\Phi}} = \begin{pmatrix} \mathbf{\Phi} & O & O \\ O & \mathbf{\Phi} & O \\ O & O & \mathbf{\Phi} \end{pmatrix} \tag{3.19}$$

and

$$\mathbf{a} = \left(\ a_{x1} \cdots a_{xt} \cdots a_{xT} \ \ a_{y1} \cdots a_{yt} \cdots a_{yT} \ \ a_{z1} \cdots a_{zt} \cdots a_{zT} \ \right)^T \tag{3.20}$$

The currents $\mathbf{J}$ rely on two components, the basis functions $\mathbf{\Phi}$ and the corresponding coefficients $\mathbf{a}$. It is worthy of note that the basis functions $\mathbf{\Phi}$ here represents the geometrical information of the cortical mesh and the corresponding coefficient $\mathbf{a}$ represents the information of the variety of the current sources distributed on the cortical mesh. The solution of the current $\mathbf{J}$ reconstruction problem is then to find the right coefficients $\mathbf{a}$ in 3-space. This is actually a typical *inverse problem*.

The meshes describing the cortex are generally with a large number of vertices and edges. It is a difficult computational problem to decompose such large graphs using standard eigenanalysis techniques. To solve this problem, we begin by partitioning the mesh into left and right hemispheres; since these parts of the cortex are largely separate, there is limited connectivity between the two and we can decompose the parts individually, saving a large amount of computation. Also, the mesh graphs are sparse, the Lancosz method (Saad, 1992) can be used to decompose the graph efficiently. For reasons explained in the next section, we only require a limited set number of eigenvectors, making this method particularly efficient.

3.5 Basis functions source model for MEG reconstruction

The inverse problem for MEG is the problem of finding a set of currents in the cortex which give the correct magnetic fluxes at the coils. Since the cortical mesh typically has several thousands vertices and the the number of MEG sensor are limited(the MEG machine in our experiment is with 248 sensors), the problem of reconstructing the current at each vertex is severely under-constrained. We can only hope to construct a much lower resolution version of the signal from the coil responses, so it does not make sense to use rapidly varying basis functions in the reconstruction. Furthermore, the eigenvectors corresponding to the largest eigenvectors are mainly representing the variational information on the cortical surface which associated to signal noise. For this reason, we concentrate on the smoother basis functions; for the Laplacian, these are the eigenvectors corresponding to the smallest eigenvalues. By choosing the correct number of basis functions, T, we can get an under-constrained problem which we can fit with least-squares and is resistant to noise. Refer to the Eqn 3.10, Eqn 3.17, we can write the forward problem as:

$$\begin{aligned} \mathbf{B} &= \mathbf{L}\mathbf{J} \\ &= \mathbf{L}\tilde{\mathbf{\Phi}}\mathbf{a} \end{aligned} \tag{3.21}$$

With respect to the known $\mathbf{L}\tilde{\mathbf{\Phi}}$ and the measurement $\mathbf{B}$, we are trying to compute $\mathbf{a}$. This problem has the same structure as Eqn 3.10. Therefore, $\mathbf{L}\tilde{\mathbf{\Phi}}$ can be assumed as a

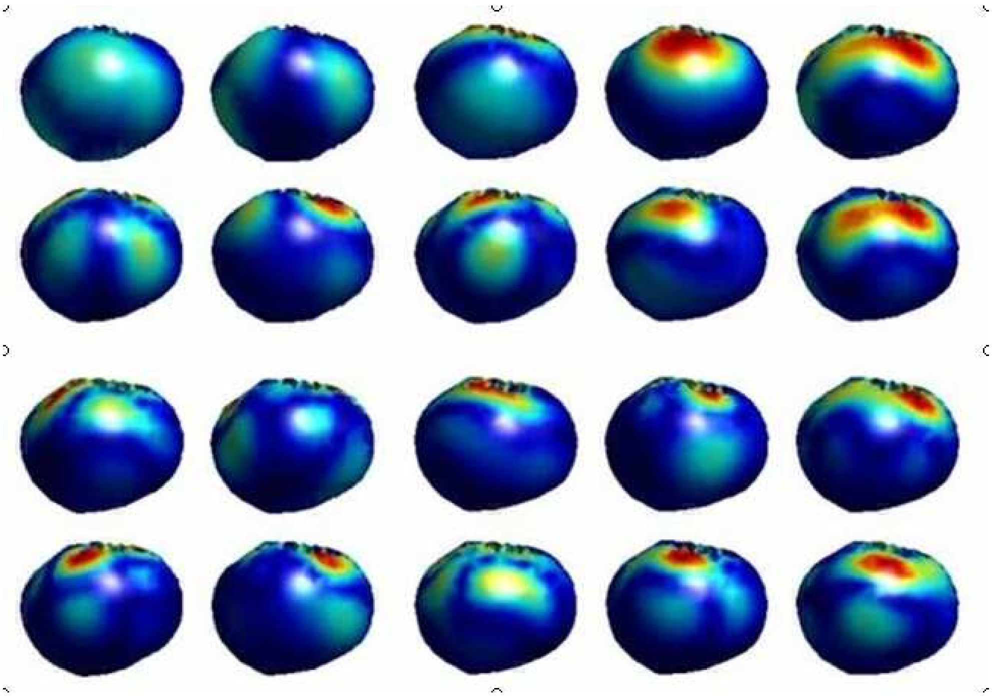

Figure 3.6: This figure shows the pattern of new leadfield W. Each cortical image indicates leadfield of the case that a single current located on one individual vertex with the set orientation and strength but the value of current distributed on other vertices show as zero. It also indicates the sensitivity of the sensor set to one geometrical point (eg, the individual vertex of the mesh) on the cortex. The colour from red into blue presents the change of the response from strong to weak. This figure only provides the leadfield pattern for first 20 vertices of all the vertices of the cortical mesh.

new leadfield, denoted as $\mathbf{W} = \mathbf{L}\tilde{\mathbf{\Phi}}$ (shows in Fig 3.6). Then the equation above can be written as:

$$\mathbf{B} = \mathbf{W}\mathbf{a} \tag{3.22}$$

which is also a typical inverse problem.

We can obtain a numerically stable estimate of $\mathbf{a}$ by solving the set of linear equations using LU-decomposition.

$$\begin{aligned} \mathbf{a} &= \mathbf{W^{-1}B} \\ &= (\mathbf{L\tilde{\Phi}})^{-\mathbf{1}}\mathbf{B} \end{aligned} \tag{3.23}$$

Choosing the basis functions is crucial to the solution of the inverse problem. If we are interested in reconstructing the global current distribution in the brain, then we need to select large-scale basis functions. These are easily found as they correspond to the eigenvectors with the smallest eigenvalues from the Laplacian. Since the eigenvectors corresponding to the top few largest eigenvalues generally reflect variational information of the graph , in contrast, the eigenvectors corresponding to the top few smallest eigenvalues representing the smooth information which is required as the general information of the geometry of the mesh in our study. On the other hand, in order to provide a well-conditioned solution to the inverse problem, there is also a limit to the number of basis functions we can select. To avoid the over determination, the total number of coefficients T for each component we can determine must be less than the number of sensor responses, 248(in the presence of noise). Since each basis function is used to reconstruct x, y and z components, we have $3T < I$ where I is the number of coils. In this case, we have 248 coils and choose $T = 82$, shows in Fig 3.7. This method of basis functions source model for MEG reconstruction is also called as ”basis function method” in the following thesis.

3.6 Results

3.6.1 Toy example

A toy example is applied to test the basis function method. Here, instead of using the spheroid cortical surface mesh, a surface mesh (vertices: 1026, faces: 2048) of a sphere is applied firstly, shown as Fig 3.8.

Two diffused sources distribution are embedded on the surface mesh of the sphere, and assumed as the simulated current sources. According to the Biot-savart law shown in Eqn 3.1, the measurements of magnetic field are generated separately with/ without the Gaussian noised added, shown in Eqn 3.9. Basis function method is then applied

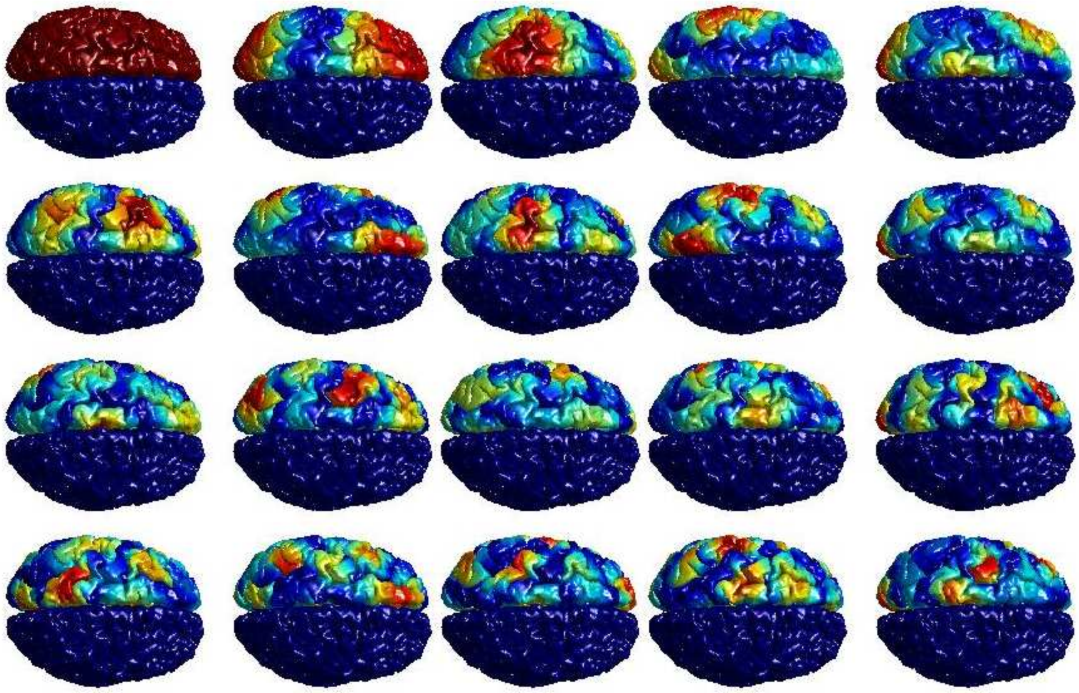

Figure 3.7: The pattern of first 20 smallest eigenvectors on left-hemisphere of cortical mesh. The basis functions is corresponding to the eigenvectors with the smallest eigenvalues from the Laplacian. The color from blue to red shows the cortex effected by the corresponding basis function from weak to strong. The portion with red color of each image shows the location that the corresponding basis function mainly represents The first color map shows the uniform information of the background, and the rest of the color maps show the detail geometrical information.

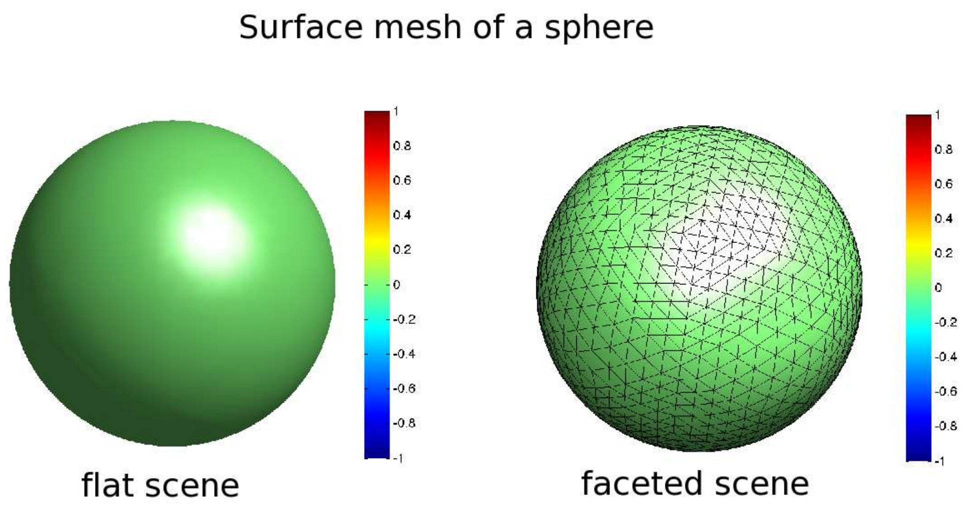

Figure 3.8: This figure shows the surface mesh of a sphere used for the toy example. The left shows the flat scene of sphere, and the right shows the faceted scene of sphere.

with respect to the geometrical information of the surface mesh as well as the measurements.

From the reconstruction results of this toy example, shown in Fig 3.9, it is explicit that the reconstructions via basis function method can provide the correct position of the simulated current sources.

3.6.2 Synthetic results

For better evaluating the Basis function method, two groups of simulated current sources are generated for synthetic experiment, i.e. articial current source distribution and realistic current source distribution in *Appendix B*. For the former type (called as *synthetic sources A*) , the fixed current source values are set on 30 particular vertices of mesh we choose but the values of current sources on other vertices are set as zero; while in the process of generating the later one(called as *synthetic source B*), the current source distribution on the cortical mesh are from the random results of previous current source reconstruction of the real MEG data with random stimuli on cortical surface at one time point. The detailed information of these two groups of simulated current sources is given in *Appendix B*.

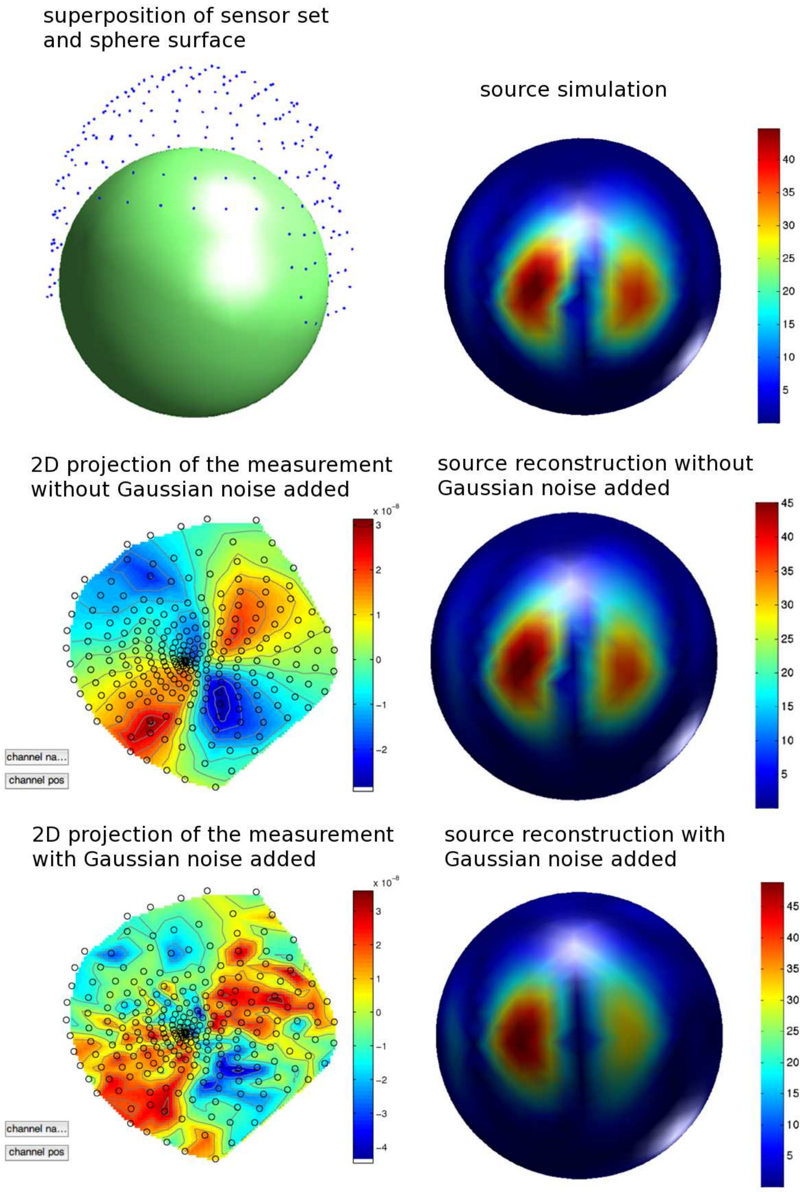

Figure 3.9: This figure demonstrates the reconstruction results of basis function method on the sphere mesh. In the first row, the left pattern shows the superposition of the sensor set and sphere surface, and the right one shows the simulated source distribution. In the second row, the left pattern demonstrates the 2D projected map of the coil measurement without the Gaussian distribution added; and the right pattern shows the reconstruction result by the basis function method. In the second row, the left pattern demonstrates the 2D projected map of the coil measurement with the Gaussian distribution added; and the right pattern shows the reconstruction result by the basis function method.

According to the manual of software *MNE* ((MNE)), the size of cortical surface mesh of the gray matter is generally set as 2600 vertices. This mesh can be obtained by simplifying the realistic head model (262658 vertices) that is generated from a segmentation of a T1 MRI image by *FreeSurfer* (5.0.0) which produces the cortical surface mesh from the MRI images with high spatial resolution ((5.0.0)). Then, the temporal and spatial correlations of the activity can be observed. The figures (Fig B.2, Fig B.3, Fig B.5 and Fig B.6) show the maps of cortical activity and the example time courses of the MEG measurement.

3.6.2.1 Reconstruction of simulated current sources

Before showing the reconstruction results, we firstly introduce the Minimum-norm method which is used for the reconstruction comparison here.

The minimum-norm estimation technique is one of the classical methods used for MEG signal processing, especially for no reliable a priori information about current source generations is available. The unique solution to the inverse problem shown in Eqn 3.10 can be found by combining constraints on the solution and constraints on the data it predicts. The following two equations described these two constraints separately.

For the solution, the general formulation in a linear framework is shown as:

$$(\mathbf{J} - \mathbf{J}^{-})^T \mathbf{C}_s (\mathbf{J} - \mathbf{J}^{-}) = min \qquad (3.24)$$

where $\mathbf{J}$ is the estimated current source, $\mathbf{J}^{-}$ is an a priori approximation of the solution and $\mathbf{C_s}$ is a weighting matrix which provides apriori knowledge about the approximate locations or covariances of current sources; meanwhile, the constraints on the data it predicts, the general formulation is shown as:

$$(\mathbf{LJ} - \mathbf{B})^T (\mathbf{LJ} - \mathbf{B}) = min \qquad (3.25)$$

where $\mathbf{L}$ is the leadfield matrix, $\mathbf{LJ}$ are the predicted data, and $\mathbf{B}$ are the measured data. When the matrix $\mathbf{C_s}$ is positive definite, the estimated current source of this problem is (Hauk, 2004):

$$\mathbf{J} = \mathbf{J}^{-} + \mathbf{C_s}^{-1} \mathbf{L}^T (\mathbf{L} \mathbf{C_s}^{-1} \mathbf{L}^T)^{-1} \mathbf{B} \qquad (3.26)$$

We assume no apriori knowledge about the reconstruction, the algorithm can be described as (Menendez et al., 1998):

$$\mathbf{J} = \mathbf{C_s}^{-1}\mathbf{L}^T(\mathbf{L}\mathbf{C_s}^{-1}\mathbf{L}^T)^{-1}\mathbf{B} \tag{3.27}$$

In this case, $\mathbf{C_s}$ is the identity matrix. The main difference between the minimum-norm method and our method is that we extract a set of smooth basis functions from the mesh with respect to the spatial organization of the signal and can be used to condition the result (Hauk, 2004).

Fig 3.11 and Fig 3.13 show the comparison between the original current source distribution, as well as the corresponding reconstructed results by basis function method and the Minimum-norm method for *synthetic data A* and *synthetic data B*, respectively (Hauk, 2004).

3.6.2.2 Noise-robustness evaluation of simulated current sources

Since one of the most important aspects of the signal reconstruction is the resistance to random noise which is always present in the MEG signal, noise-robustness is applied as an important property to measure the goodness of a algorithm of MEG source reconstruction. In our experiment, the goodness of noise-rubustness of basis function method is observed by comparing with the results of Minimum-norm method.

Firstly, we obtained the simplified triangled mesh M of the brain which is with 2600 vertices(1300 vertices for left hemisphere and 1300 vertices for right hemisphere) using an MRI scan of the subject. The coil responses are produced from an MEG scan of the same subject from a single epoch and time-slice of the scan. Here, the number of sensors are assumed to be 248 (for *4-D Neuroimaging 248-channel MEG*). We select the basis functions via the eigendecomposition of the mesh Laplacian and pick the eigenvectors corresponding to the first 41 smallest eigenvalues.

Here, we analyze the reconstruction under noisy conditions. Since the MEG measurement environment is assumed to be full of different types of noise, the noise type here is applied as the most general case, zero-mean Gaussian distribution. Based on the synthetic current sources distribution, *synthetic data A* and *synthetic data B*, the environmental noised condition can be simulated by adding Gaussian noise with fixed

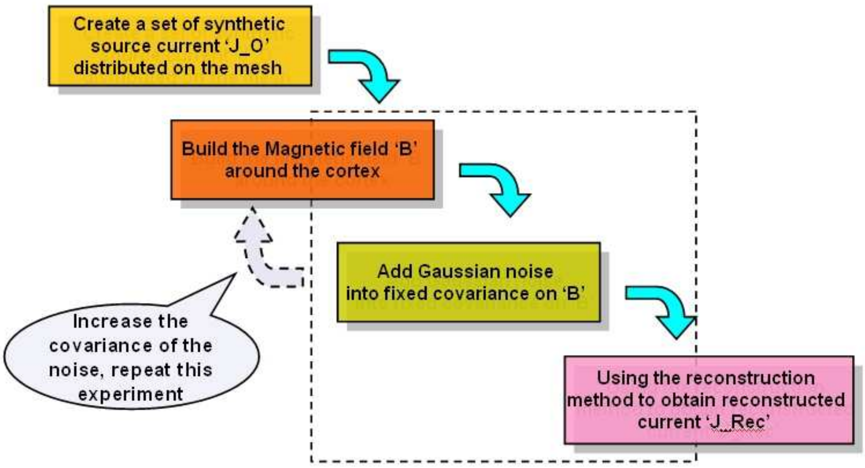

Figure 3.10: The figure shows the pipeline of the synthetic experiment of the noise-robustness comparison between the minimum-norm method and the basis function method. Based on the synthetic sources distribution, *synthetic data A* and *synthetic data B*, the environmental noised condition can be simulated by adding Gaussian noise with fixed covariance value to the coil responses before reconstruction. Increasing the covariance value of noise recursively as 100 iterations. In each iteration, undertaking the reconstruction with respect to the noised response with different covariance. The pattern of the square root of error variance between the reconstruction and the original current sources can be shown for basis function method and Minimum-norm method in figure 3.12, figure 3.14 and figure 3.17.

covariance value to the coil responses before reconstruction. The pipeline of the experiment is shown in Fig 3.10. In terms of this pipeline, keep increasing the covariance value of noise recursively as 100 iterations. With respect to the noised response with different covariance, undertaking the reconstruction in each iteration. These results can be used to obtain the patterns of the square root of error variance between the reconstruction to the original current source distribution. Here, Fig 3.12 indicates the comparison of noise-robustness between basis function method and Minimum-norm method for *synthetic data A*; and Fig 3.14 indicates the same comparison for *synthetic data B* which present the noise-robustness evaluation of both basis function method and Minimum-norm method.

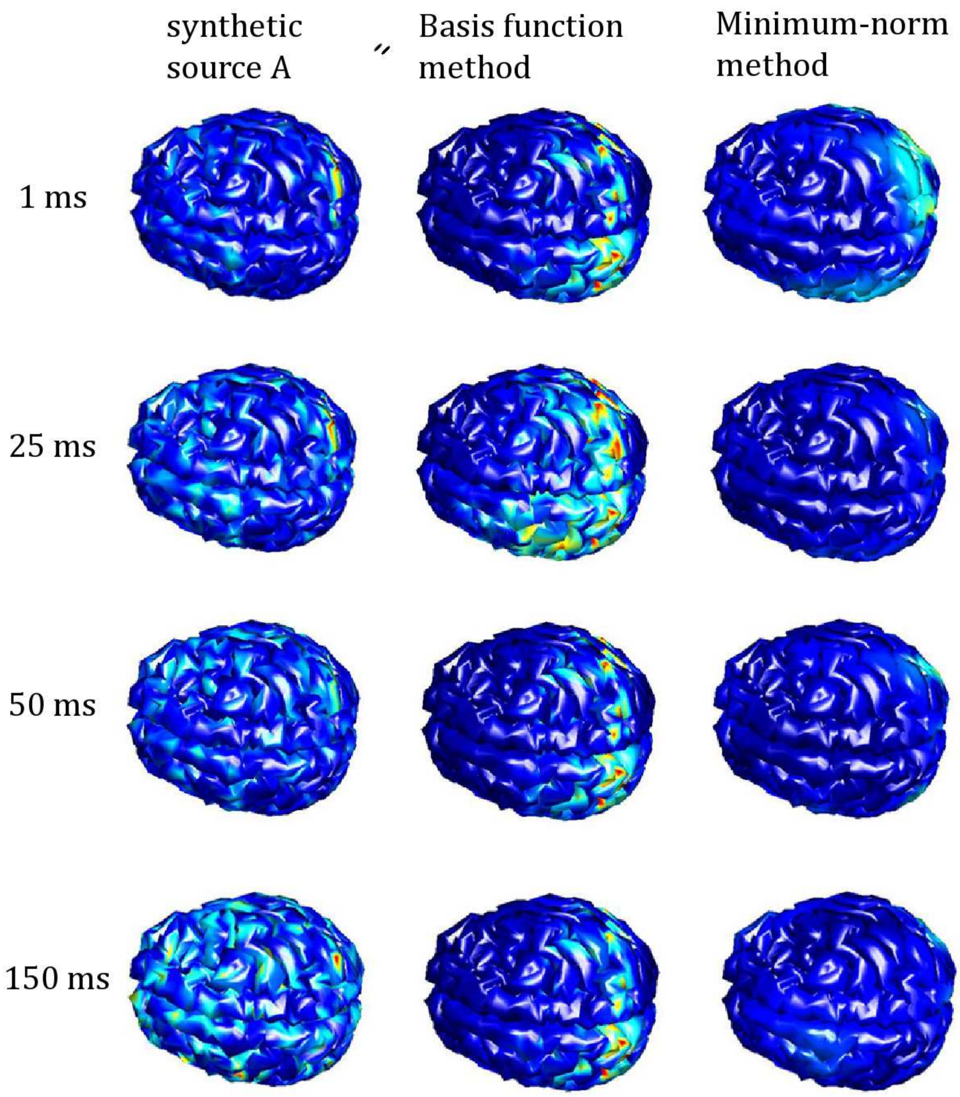

Figure 3.11: This figure shows the comparison spatial visualization of the synthetic original current sources, reconstruction by basis function method and the minimum-norm method. The color from red to blue show the intensity of current source strength from strong to weak. The first column illustrates the synthetic original current source pattern, the middle column illustrates the reconstruction by basis function method, and the right column illustrates the reconstruction by minimum-norm method. With respect to two types of synthetic sources we create in *Appendix 2*, it is notable that the *artificial source* (the *synthetic source A*) are applied in the first column, the time point(in *ms*) 1, 25, 50, 150 of *realistic source* are applied from the first row to the last row, respectively.

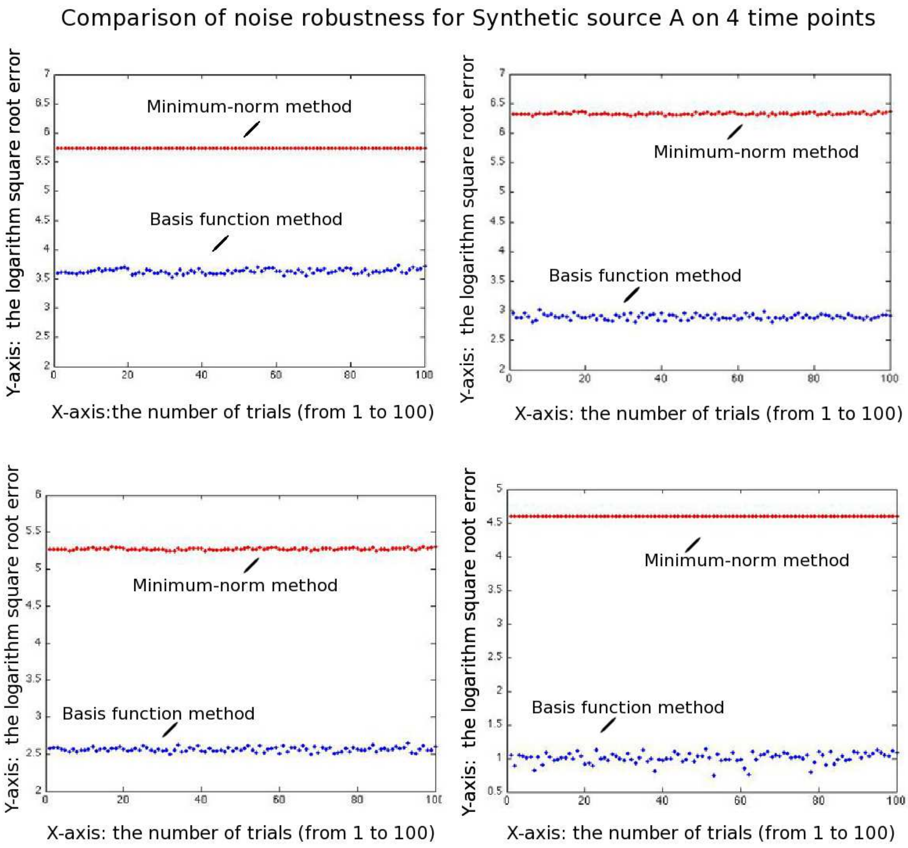

Figure 3.12: These 4 patterns indicate the comparison of noise robustness between the Minimum-norm method and basis function method for the 4 current sources used in Fig 3.11. According to the pipeline shown in Fig 3.10, we add 100 increased different covariance values of zero-mean Gaussian noise (from 0.01 to 0.1) to the measurement of the *synthetic source A* (*artificial source* in *Appendix 2*), and observe the noise robustness of these two methods. In these 4 patterns, X-axis shows the number of trials from 1 to 100, Y-axis shows the log square root error of reconstruction; the dots in blue show the log square root error of reconstruction for basis function method, and the dots in red show the log square root error of reconstruction for Minimum-norm method. Left-up,Right-up, Left-bottom and Right-bottom are for the current sources on time point (in *ms*): 1, 25, 50 and 150, respectively.

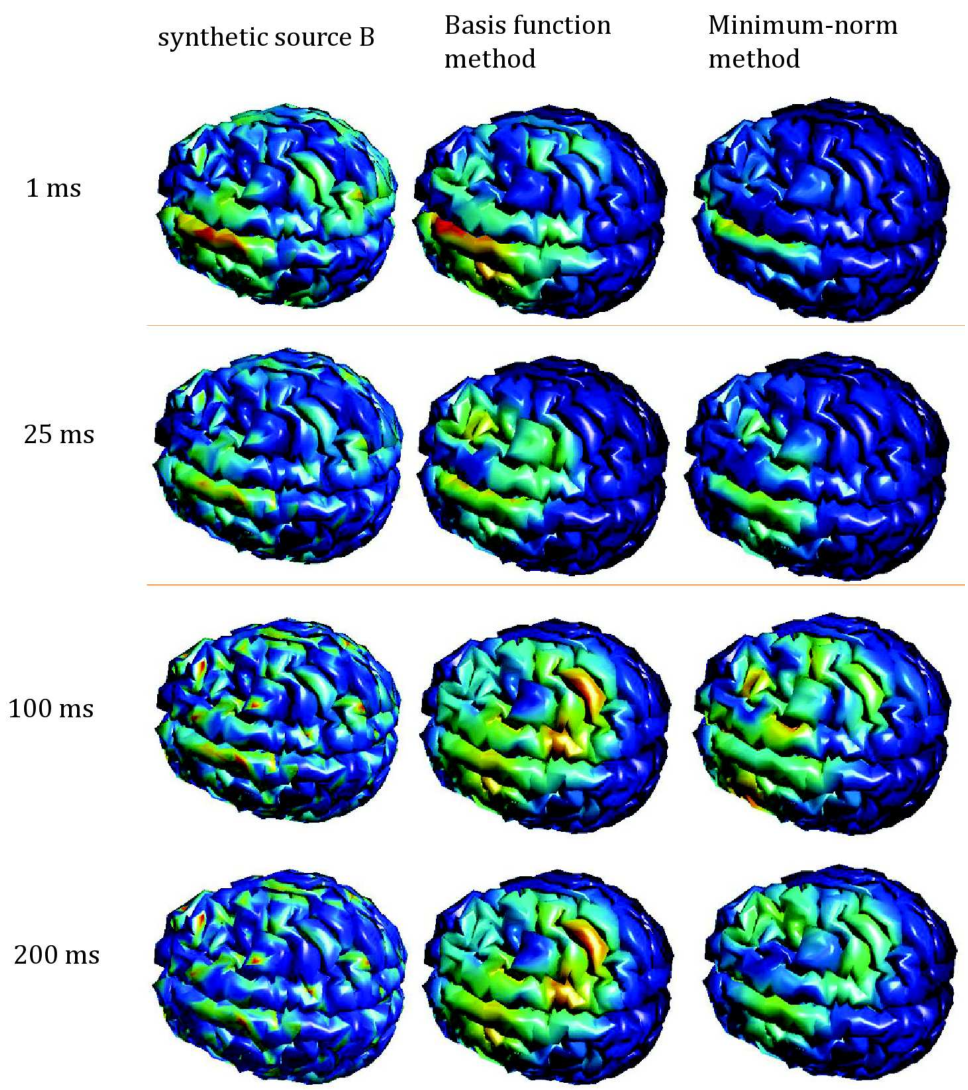

Figure 3.13: This figure shows the comparison spatial visualization of the synthetic original source, reconstruction by basis function method and the minimum-norm method. The color from red to blue show the intensity of source strength from strong to weak. The first column illustrates the synthetic original current source pattern, the middle column illustrates the reconstruction by basis function method, and the right column illustrates the reconstruction by minimum-norm method. With respect to two types of synthetic sources we create in *Appendix 2*, it is notable that the *synthetic source*(*synthetic source B*) are applied in the first column, the time point (in *ms*): 1, 25, 100 and 200 of *realistic source* are applied from the first row to the last row, respectively.

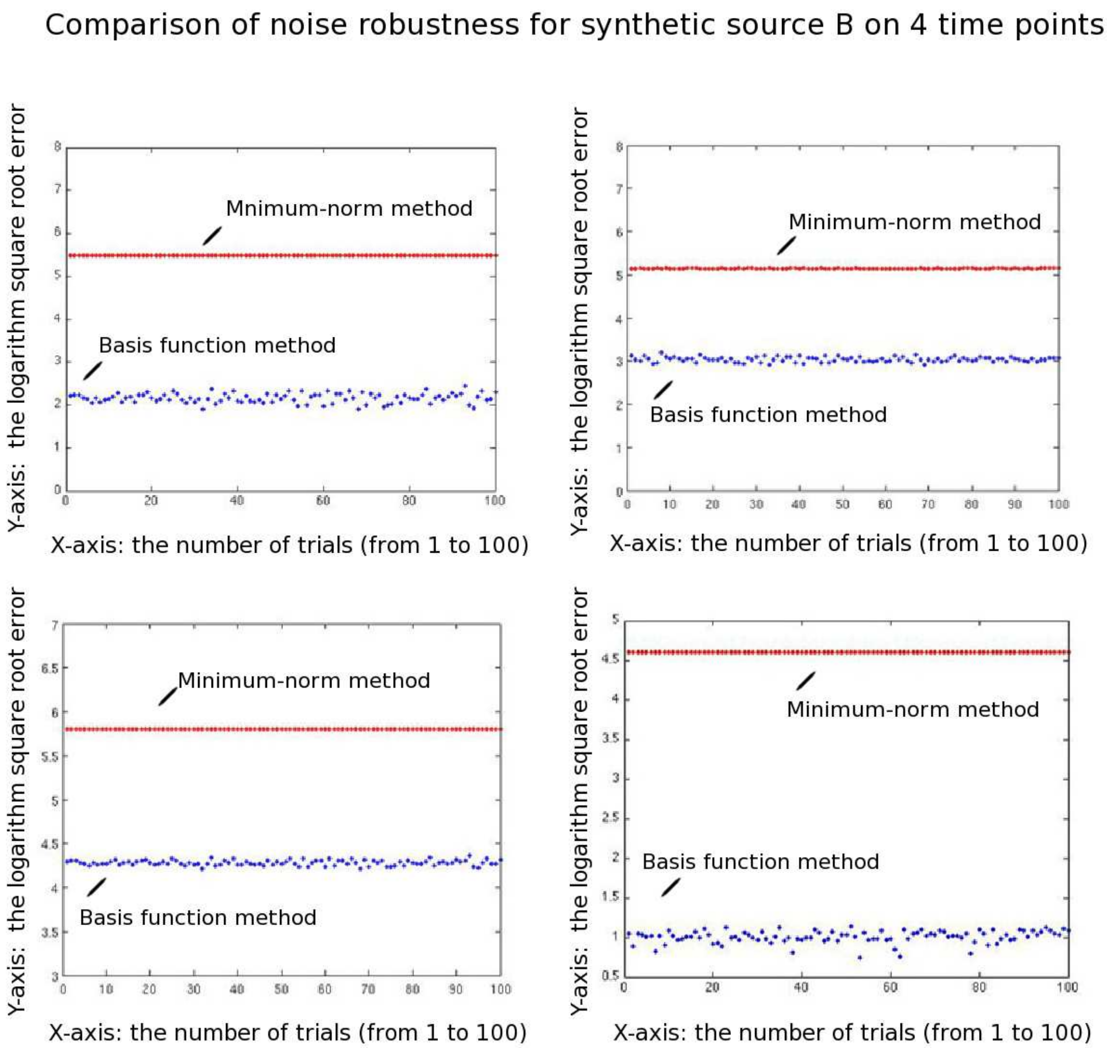

Figure 3.14: These 4 patterns indicate the comparison of noise robustness between the Minimum-norm method and basis function method for the 4 current sources used in Fig 3.13. According to the pipeline shown in Fig 3.10, we add 100 increased different covariance values of zero-mean Gaussian noise (from 0.01 to 0.1) to the measurement of the *synthetic source B* (*synthetic source* in *Appendix 2*), and observe the noise robustness of these two methods. In these 4 patterns, X-axis shows the number of trials from 1 to 100, Y-axis shows the log square root error of reconstruction; the dots in blue show the log square root error of reconstruction for basis function method, and the dots in red show the log square root error of reconstruction for Minimum-norm method. Left-up, Right-up, Left-bottom and Right-bottom are for the current sources on time point (in *ms*): 1, 25, 100 and 200, respectively.

From the reconstruction results shown in Fig 3.11 and Fig 3.13, it indicates that the basis function method can produce the source reconstruction as good as Minimum-norm method although the reconstruction results still exist distortions and instability comparing with the original source patterns. Furthermore, basis function method shows superior performance on noise robustness rather than Mninimum-norm method in terms of the results shown in Fig 3.12, as well as Fig 3.14. However, the basis function method is not entirely suitable for the current source localization and the reconstruction of the whole brain, as shown by the results in Fig 3.11 and Fig 3.13. These unsatisfactory reconstructed results are mainly because of the basis function method is basically a ill-posed inverse problem (showed in Eqn 3.22). The larger region of cortical surface used for reconstruction leads to more plausible and less accurate results.

3.7 Localizing the source reconstruction into the region of interest(ROI)

Based on the reason above, it is worth to try to reduce the reconstructed region to be close to the region of interest(ROI) so that the accuracy is assumed to be increased. In order to do this, we need to find localized basis function based on the geometrical information of ROI. Here, the normalized cut method is used for segmentation of cortical surface mesh to obtain the mesh of ROI.

The normalized cut method has been widely applied for graph segmentation. The use of the Fiedler vector(eigenvector associated with the 2nd smallest eigenvalue of the graph Laplacian) for the purpose of data clustering is one of the most important applications of spectral graph theory in image analysis and pattern recognition (Shi and Malik, 2000), (Belkin and Niyogi, 2003), (Sarkar and Boyer, 1996). In our study, the normalized cut method is applied to mesh segmentation for obtaining a region of interest (ROI) with respect to the features of basis function method illustrated above.

By removing the edges of the connection, a graph $G = (V, E)$ is easily partitioned into two disjoint sets, A and B, with $A \bigcup B = V$ and $A \bigcap B = \emptyset$. Also, the degree of dissimilarity between these two segments can be calculated as a *cut*. Since we do not have the weighted edges in the cortical surface mesh, the total weight of the removed edges can be assumed as Adjacency matrix of the mesh in our problem:

$$cut(A, B) = A_{uv}; \quad u \in A, v \in B. \tag{3.28}$$

The *minimum cut* of this graph G is the optimal solution of bipartition which is a well-studied problem with plenty of existing efficient algorithm in graphic research.

However, the normalized cut method is a unbiased measure of disassociation between subgroups of a graph and provides a nice property which avoids the unnatural bias for partitioning out small sets of points which indicates in Wu and Leahy's method (Wu and Leahy, 1993).

The idea of normalized cut method is to calculate the cut cost as a fraction of the total edge connections to all the nodes in the graph (this disassociation is so-called the *normalized cut* or *Ncut*) instead of the computation of total edge weight connecting two partitions. The *Ncut* can be showed as:

$$Ncut(A, B) = \frac{cut(A, B)}{assoc(A, V)} + \frac{cut(A, B)}{assoc(B, V)} \tag{3.29}$$

where $assoc(A, V) = \sum_{u \in A, t \in V} \omega(u, t)$ is the total connection of the nodes in A to all the nodes in the graph G and $assoc(B, V)$ is defined similarly.

Meantime, a measure for total normalized association for a give partition can be defined as:

$$Nassoc(A, B) = \frac{assoc(A, A)}{assoc(A, V)} + \frac{(B, B)}{assoc(B, V)} \tag{3.30}$$

where $assoc(A, A)$ and $assoc(B, B)$ are total weights of edges connecting nodes within A and B separately. This reflects how tightly on average nodes within the group are connected to each other which is an unbiased measure.

These disassociation and association of a partition can be then related as follows:

$$\begin{aligned} Ncut(A,B) &= \frac{cut(A,B)}{assoc(A,V)} + \frac{cut(A,B)}{assoc(B,V)} \\ &= \frac{assoc(A,V) - assoc(A,A)}{assoc(A,V)} + \frac{assoc(B,V) - assoc(B,B)}{assoc(B,V)} \\ &= 2 - (\frac{assoc(A,A)}{assoc(A,V)} + \frac{assoc(B,B)}{assoc(B,V)}) \\ &= 2 - Nassoc(A,B). \end{aligned} \tag{3.31}$$

Thus, the optimal solution for the partition is to minimize normalized cut which directly leads to maximize the normalized association. According to the Eqn 3.31, these two can be satisfied simultaneously. Shi and Malik then indicated the detail process how normalized cut is computed efficiently by solving a generalized eigenvalue problem (Shi and Malik, 2000).

In our study, the normalized cut has been applied by the following steps with respect to the work of Shi and Malik (Shi and Malik, 2000):

1. The mesh of cortical surface, also assumed as a weighted graph $G = (V, E)$, is constructed with the matrices of vertices V and edges E. The weights, referred as the elements of the Adjacency matrix, reflect the connecting state between two vertices of the mesh.

2. Solve for the eigenvectors with the smallest eigenvalues of the system which can be transformed into a standard eigenvalue problem of :

$$D^{-\frac{1}{2}}(D - A_{uv})D^{-\frac{1}{2}}x = \lambda x \tag{3.32}$$

It is worth to note that from the previous work the Laplacian of the cortical surface $\mathcal{L}$ can be directly introduced here for calculation, $\mathcal{L} = D - A_{uv}$ (where D is degree matrix of the mesh, and A_{uv} is the Adjacency matrix). Additionally, there are some properties which simplifies the computation of the segmentation. Firstly, the eigensystems must be sparse since the mesh is only locally connected; secondly, our segmentation only needs the eigenvectors corresponding to the first few smallest eigenvectors; moreover, the precision requirement for the eigenvectors are low. The Lanczos method is used for solve the eigendecomposition with these properties;

3. Use the eigenvector with the second smallest eigenvalue to segment the graph;

4. Decide if the current segmentation should be partitioned and do the segmentation recursively if necessary.

It is notable that the continuous mesh segmentation should not be more than 5 times for a single hemisphere for avoiding the over determination , since 2^5 is smaller than the number of basis function 41 but 2^6 is larger than it, this hierarchical relationship is shown as Fig 3.15. Also. Fig 3.16 shows the segments of the cortical mesh by 5-level normalized cut.

The figures, Fig 3.18 and Fig 3.19, illustrate the comparison of source reconstructions between the basis function method for the whole cortical surface and the partition of ROI by the Normalized cut method for the synthetic sources on one time point selected from *synthetic source A* (at time point 1, refer to Fig 3.11) and *synthetic source B* (at time point 1, refer to Fig 3.13), respectively. In Fig 3.18, the current sources are distributed on both the left and right hemisphere with respect to the pre-knowledge. Therefore, the normalized cut is applied on the whole cortical surface mesh with *level-1*. And, in Fig 3.18, the current sources are distributed on only right hemisphere with respect to the pre-knowledge. The normalized cut is then applied on right hemisphere with *level-1*. These results indicate that when the reconstruction model is segmented as close as the region that the current sources actual locate on, the reconstruction is capable to obtain the more accurate result. In other words, the change of the reconstruction on the model from the global cortical surface into the local region based on the ROI and additional selection of proper local region effect the accuracy and goodness of the result directly. This is a crucial feature when the basis function method is applied into the MEG source reconstruction.

3.7.1 Application to the real data

We use the real MEG data of visual expression based on *Appendix 3*. Firstly, we obtained the cortical surface mesh with 262658 vertices and 565782 faces from the

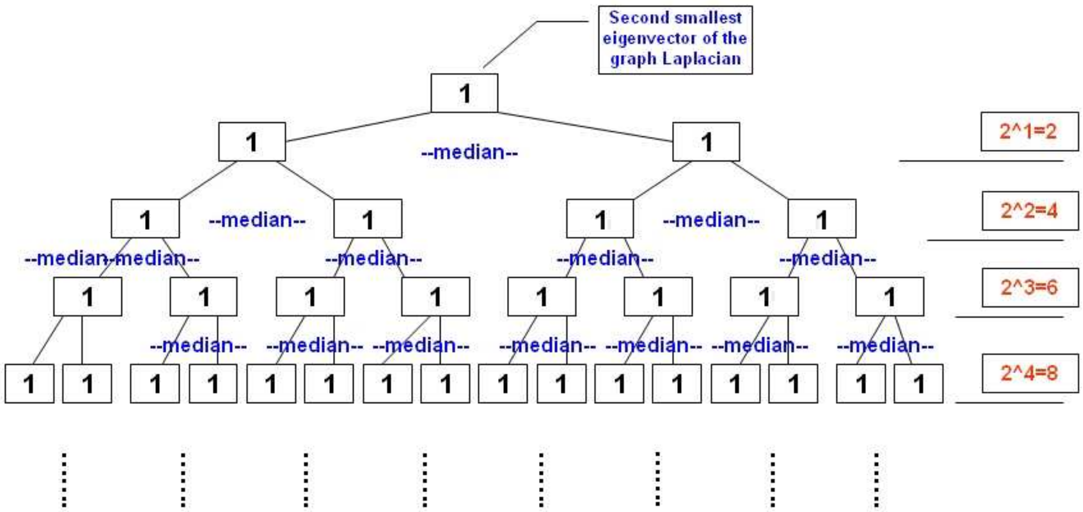

Figure 3.15: This figure shows the hierarchy of mesh segmentation with normalized cut method. From the top to the bottom, the original cortical surface mesh can be segmented into 2 partitions in terms of the Fiedler vector (this is called *level-1*); and each partition can be segmented as the same way (called as *level-2*), and so on. To avoid the over determination, the total number of partition should not be more than the number of basis functions. In our problem, the number of basis function is 41, therefore, the partition of normalized cut should not be more than *level-5*.

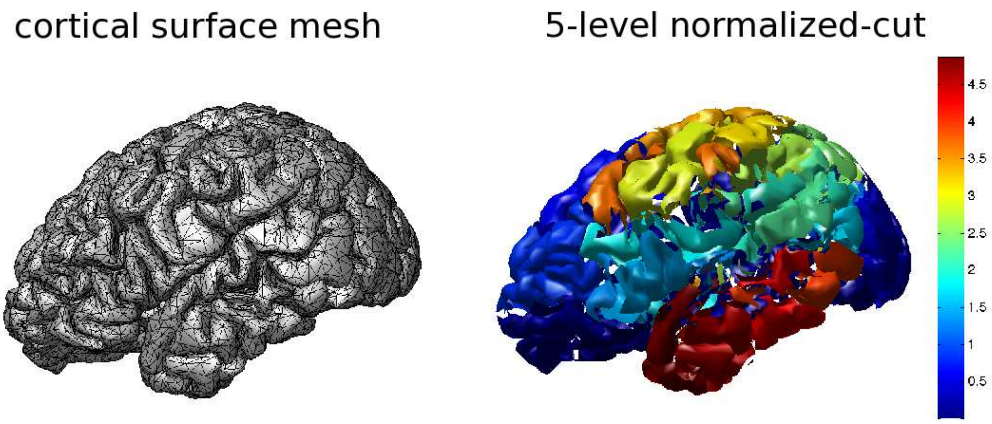

Figure 3.16: This figure demonstrates the mesh partitions of the cortical surface mesh by *level-5* normalized cut method. With respect to the segmentation theory, the number of the partitions of *level-5* normalized cut is $2^5 = 32$ The left pattern shows the cortical surface mesh, and the right pattern shows the mesh partitions where the segments present in different color indicated the different segments by normalized cut method.

structural MRI scan of the same subject by *Freesurfer*(5.0.0). Since the coordinate system of MRI cortical surface is different from the MEG coordinates, the coordinate registration is processed as the first step(with the special solution provided by YNiC).

Specifically, this *coregistration* between the coordinate of structural MRI and MEG system should be based on a set of at least 3 points whose coordinates are known in both systems. These points are called *fiducials*. In terms of these *fiducials*, a position of any point on one of these two spaces can be convert to the other by the rigid transformation matrices (Rotation and Translation). These fiducial points are located in both scans using special markers introduced on the head during the scanning process. Meanwhile, the same transformation is applied to the sensors position of MEG as well.

However, the spatial resolution of the mesh obtained from MRI is too large for the realistic or reasonable solution. The simplified mesh is therefore generated by the software *Remesh* ((Remesh, 2008)). In terms of the mesh resolution selected for MEG analysis in software MNE (MNE), we apply a spatial resolution for the mesh M with 2600 vertices and 5192 faces . Secondly, the measurement of MEG signals are represented as a $96 \times 248 \times 813$ matrix, where 96 indicates the number of different

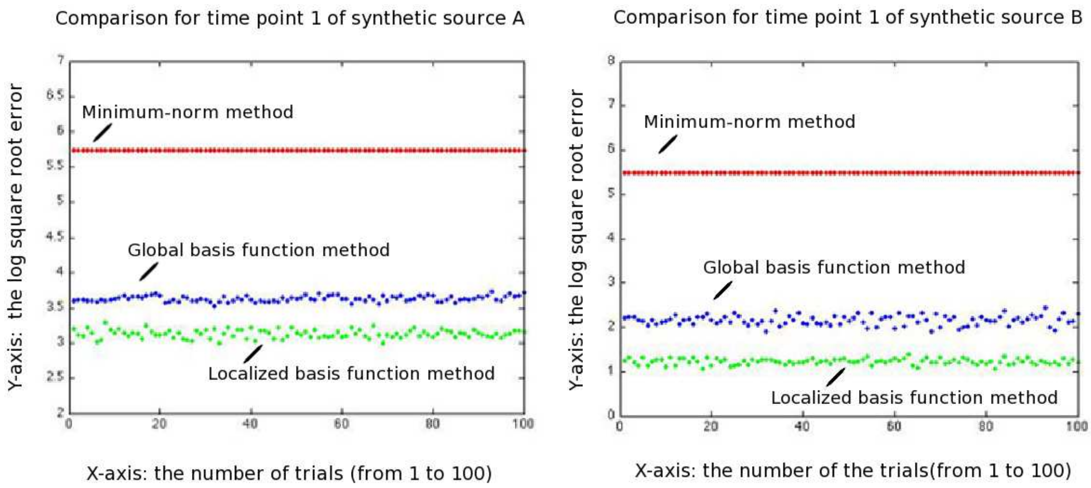

Figure 3.17: These 2 patterns indicate the comparison of noise robustness between the Minimum-norm method, basis function method for the whole cortical surface and for the partition of ROI. The left: the comparison for time point 1 of *synthetic source A*; the right: the comparison for time 1 of *synthetic source B*. According to the pipeline shown in Fig 3.10, we add 100 increased different covariance values of zero-mean Gaussian noise (from 0.01 to 0.1) to the measurement of the synthetic source, and observe the noise robustness for these three methods. In these 2 patterns, X-axis shows the number of trials from 1 to 100, Y-axis shows the log square root error of reconstruction; the dots in blue show the log square root error of reconstruction for basis function method, the dots in red show the log square root error of reconstruction for Minimum-norm method, and dots in green show the log square root error of reconstruction for localized basis function method.

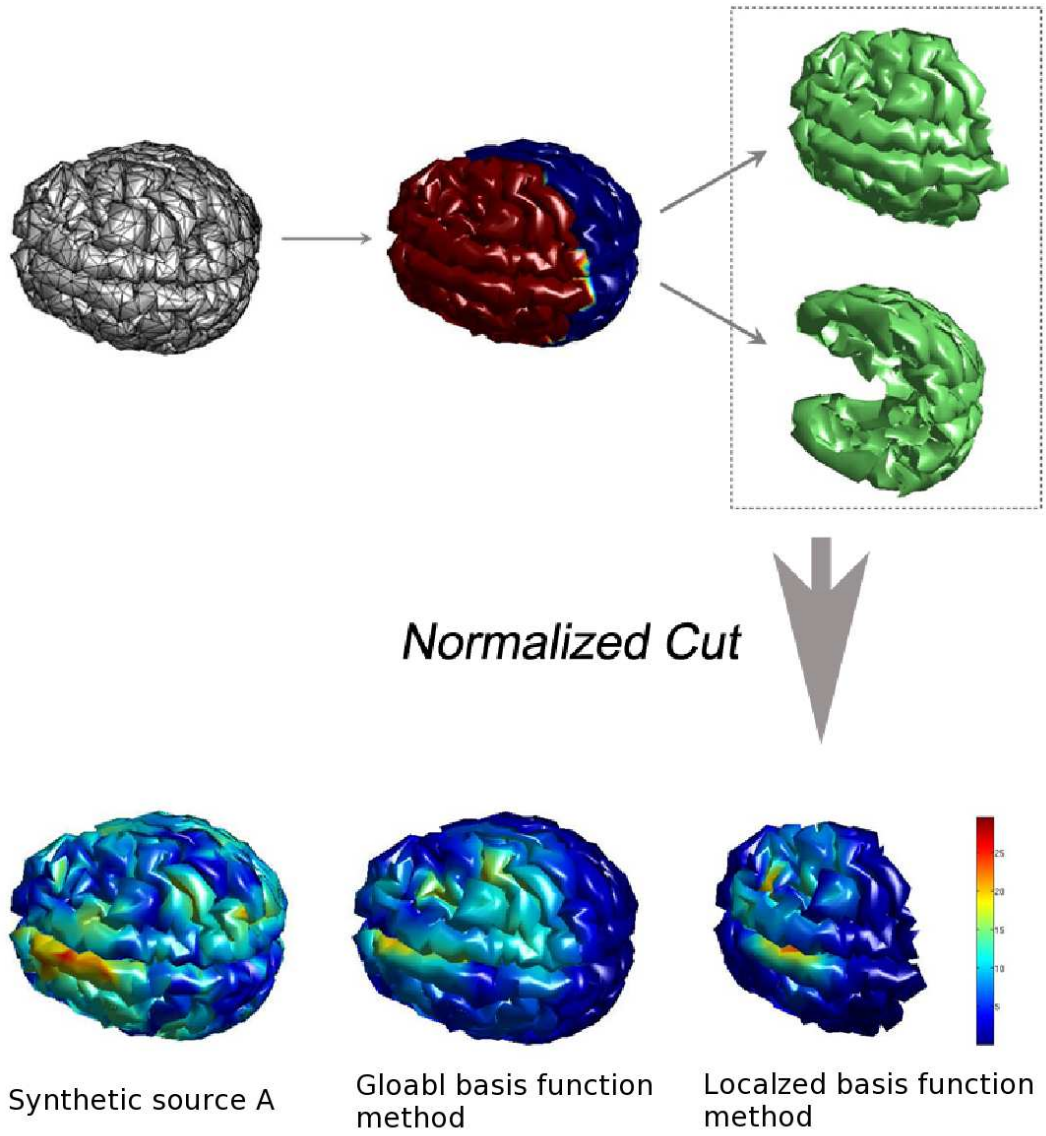

Figure 3.18: The top part in this figure demonstrates the process of the normalized cut method applied on the whole cortical surface mesh with *level-1*. The bottom part of the figure shows the comparison of source reconstructions between the basis function method for the whole cortical surface and the partition of ROI by the Normalized cut method. The left: the original source pattern; the middle: the basis function reconstruction based on the whole cortical surface; the right: the basis function reconstruction on the partition of ROI obtained by normalized cut method. The source distribution at one time point used here is selected from *synthetic source A* (at time point 1, refer to Fig 3.11).

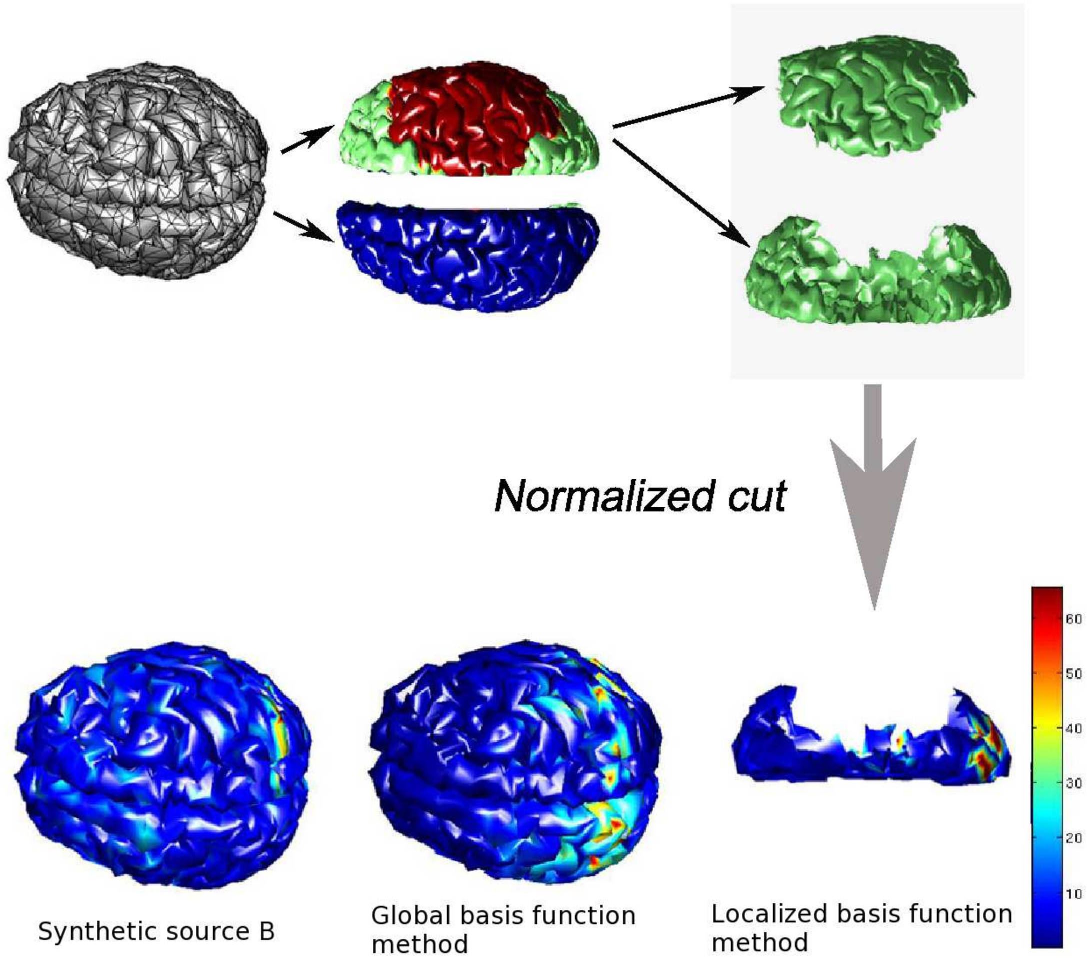

Figure 3.19: The top part in this figure demonstrates the process of the normalized cut method applied on the whole cortical surface mesh with *level-1*. The bottom part of the figure shows the comparison of source reconstructions between the basis function method for the whole cortical surface and the partition of ROI by the Normalized cut method. The left: the original source pattern; the middle: the basis function reconstruction based on the whole cortical surface; the right: the basis function reconstruction on the partition of ROI obtained by normalized cut method. The source distribution at one time point used here is selected from *synthetic source B* (at time point 1, refer to Fig 3.13).

stimuli, 248 indicates the number of sensors and 813 indicates the continuous time instants. The visualization of this measurement matrix is showed in Fig C.1.

For applying the basis function method for spatial reconstruction, we choose the measurement of the particular stimulus on particular time point and process the reconstruction. In this experiment, we choose the stimulus 3 of 96, and select the time point 20 ,45 , 70, 95 , 120 , 145 ,170 , 195 , 220 , 245 , 270, 295, 320 , 345,370 ,395 , 420 , 445 , 470 , 495 , 520, 545 , 570 , 595, 620 , 645, 670 , 695, 720 , 745 separately for the reconstruction. The Fig 3.20 shows the reconstruction results of this trail.

Since it is impossible to have an absolute correct sources location for the goodness evaluation of our method in the real MEG experiment, we introduce the source reconstruction of the same trial by fMRI and cognition estimation based on the stimulus knowledge we have. Here, Cindy C. Hagan's fMRI result for the same experiment (Hagan et al., 2009) are applied here for the reconstruction result comparison and checking. It is clear transient visual changes are occurs in the posterior superior temporal sulcus (STS) from(Hagan et al., 2009). The reconstruction results of our method, showed in Fig 3.20, illustrate the correct source location with respect to these fMRI reconstruction.

3.8 Discussion

Firstly, the basis function method has the weakness on the assumption that all sensors are perfectly set tangentially to the conducting sphere. Since in real data, more realistic head model (like BEM) is more and more used and there are no a priori on the sensor array orientation, meanwhile, the cortical mesh typically has several thousands vertices and the the number of MEG sensor are limited(the MEG machine in our experiment is with 248 sensors(in 4D-Imaging 248-sensor MEG machine), the problem of reconstructing the current at each vertex is severely under-constrained. In this case, the source reconstruction from the basis function method can be assumed as the low-resolution version of source reconstruction. Then, it is crucial to apply some further method for the correction and resolution improvement, such as super-resolution method and Kalman smoother method illuminated in the *Chapter 4* and *Chapter 5* , as a complement for estimating the source distribution of higher spatial resolution

Figure 3.20: In terms of the MEG data of facial expression (showed as the $96 \times 248 \times 813$ matrix) ,this figure shows the reconstruction results on 3 of 96, and select the time point (in *ms*): 20 ,45 , 70, 95 , 120 , 145 ,170 , 195 , 220 , 245 , 270, 295, 320 , 345, 370, 395, 420, 445, 470, 495, 520, 545, 570, 595, 620, 645, 670, 695, 720, 745 for the reconstruction. First row shows the results on 20, 45, 70, 95, 120; second row shows the results on 145, 170, 195, 220, 245; the third row shows the results on 270, 295, 320, 345, 370; the fourth row shows the results on 395, 420, 445, 470, 495; the fifth row shows the results on 520, 545, 570, 595, 620; the sixth row shows the results on 645 , 670, 695, 720, 745.

with better accuracy. In the assumption of super-resolution method, the distortions of reconstruction caused by the assumption that sensors are perfectly tangential to the conducting sphere in basis function method are assumed as distorted low-resolution current sources contain insufficient information about the original current sources.

Secondly, the current experimental configuration is : InterCore2(1,8GHz), Linux system(2.6.34.1)-32bit, matlab $7.9.0(R2009b) \times 32$ edition , RAM: 4GB. With is configuration, the computational costs of basis function method is low since we have reduced the number of the unknown in this ill-posed inverse problem. Specifically, the *backslash* function of Matlab has been applied the matrices division which effectively leads to fast computation (Matlab, 2009).

3.9 Conclusion

The aim of this topic is to explore a new method of the MEG source spatio-temporal reconstruction based on modelling the neural current sources with extended basis functions. In light of the assumption that the Laplacian eigenvectors of mesh can be analogous to its basis functions that represent the cortex mesh, we build a new model to describe the current sources distributed on each mesh vertex. This model consists of analogous basis functions and unknown weighted coefficients. Using the leadfield, the weighted coefcients can be calculated according to the forward formulae of MEG. The distributed neural current sources on mesh are then reconstructed according to the basis functions model. Expanding this process from a single time point to continuous time stretches, we are able to obtain the spatio-temporal the reconstructed current source that is distributed on cortical mesh vertices. This provides a smooth and well-conditioned reconstruction problem that can be solved directly by an inverse method. The results are more physically plausible than the minimum-norm method while being resistant to noise. Moreover, in terms of the experimental results, this algorithm shows good reconstructed property in response to the single stimulus, as well as the supercial stimulus on the cortical surface. However, it generates ambiguous and inaccurate results when the cortical current sources are distributed over multiple sites on the brain or deep in the head. In conclusion, the algorithm is mainly effective for the distributed and superficial current sources rather than the single or/and deep current sources of the cortex.

The further improvements should be focussed on :

- The basis function algorithm is based on the basis functions extracted from cortical surface mesh and their associated coefficients. The basis functions rely on the geometry of the cortical surface mesh. It will be interesting to test the robustness of the algorithm to the cortical surface mesh with difference spatial resolution.
- In our work, we apply a part of eigenvectors corresponding to the first smallest eigenvalues as the analogue of basis functions which are used for the representation of the geometry of the cortical mesh. However, there are plenty of other types of basis functions, e.g. radial basis function(RBF), spherical harmonic basis function(SHBF) which can be directly applied in our algorithm instead of the eigenvectors set. It is possible that some type of basis function can provide a superior result for this basis function reconstruction.
- In terms of the weakness of normalized-cut showed in (Belkin and Niyogi, 2003) and (Sharma et al., 2009), it has been noticed that this does not guarantee good clusters as the normalized cut is computed recursively irrespective of the global structure of the data in the practical mesh segmentation. It might be interesting to research more on mesh segmentation to find more specialized algorithm so that the source-distributed mesh can be cut more accurately with respect to the range of stimuli smoothed.

Chapter 4

Spatial Improvement of MEG reconstruction with Bayesian Super-resolution

4.1 Introduction

Super-resolution is one of the classic pattern recognition methods that are used for high-quality image recovery from a set of low-resolution images. The principle of the method is to improve the image by the inversion of a transformation from some unknown high-resolution image into the observed low-resolution images (Tipping and Bishop, 2003). This approach applies a regularization process for the ill-posed inverse problem.

MEG source reconstruction can be achieved by the basis function method presented in the previous chapter. The reconstructed current source at a single time point is distributed on a triangular mesh of the cortical surface obtained from the structural brain imaging, e.g., MRI, for the same subject. The advantage of MEG over fMRI scans is that MEG has a high temporal resolution; in other words, we can obtain a rapid sequence of images. The goal here is to use these image sequences in conjunction with the previous source reconstruction method, to improve the spatial resolution of MEG. This problem can be described as obtaining a source distribution on the higher-resolution cortical surface mesh from several continuous current source frames distributed on the original cortical mesh in the temporal field, which resembles the concept of super-resolution for image reconstruction. Bayesian super-resolution is able to provide a reasonable estimation for the spatial resolution improvement (Tipping and Bishop, 2003) , (Nara et al., 2006).

4.2 High-resolution mesh extraction

Since we aim to reconstruct the current source distributed on a mesh with higher spatial resolution, this new cortical mesh is produced firstly. The old cortical mesh used for basis function reconstruction is called *low-resolution mesh* here (noted as **M**), in contrast, the new mesh with high spatial resolution is called the *high-resolution mesh* (noted as $\mathbf{M}^+$).

The *high-resolution mesh*, $\mathbf{M}^+$ can be interpolated geometrically from the *low-resolution mesh* **M**. The *low-resolution mesh*, **M**, used here is actually the triangular mesh representing the surface of grey matter extracted from the MRI scan of the same subject with V vertices and F faces. The new mesh, $\mathbf{M}^+$, is associated with **M** but with higher spatial resolution, can be generated via the approach of interpolation on the basis of **M** by adding one vertex in the center of each mesh triangle and linking it with the surrounding triangular vertices, shown as Fig 4.1 , specifically. In this case, each new vertex and associated three edges are constructed in this way. Moreover, the coordinate of each new-added vertex is the average of the coordinate of the three vertices surrounded. Thus, the high-resolution mesh $\mathbf{M}^+$ associated with **M** is constructed with $(V + F)$ vertices and $3F$ faces are shown as Fig 4.2, where the group of interpolated new vertices is showed as the $F \times 3$ matrix $\mathbf{V}_{ip}$. In other words, there is the $V \times 3$ matrix $\mathbf{V}_{LR}$ of the vertices on the low-resolution mesh **M** in 3 dimensions, and the $(V + F) \times 3$ matrix $\mathbf{V}_{HR}$ of the vertices on the high-resolution mesh $\mathbf{M}^+$. All the structure of the matrix **M** and $\mathbf{M}^+$ are shown as follows:

$$\mathbf{V}_{LR} = \begin{pmatrix} \mathbf{V}_{LRx} \\ \mathbf{V}_{LRy} \\ \mathbf{V}_{LRz} \end{pmatrix} \tag{4.1}$$

$$\mathbf{V}_{HR} = \begin{pmatrix} \mathbf{V}_{LR} \\ \mathbf{V}_{ip} \end{pmatrix} = \begin{pmatrix} \mathbf{V}_{HRx} \\ \mathbf{V}_{HRy} \\ \mathbf{V}_{HRz} \end{pmatrix} \tag{4.2}$$

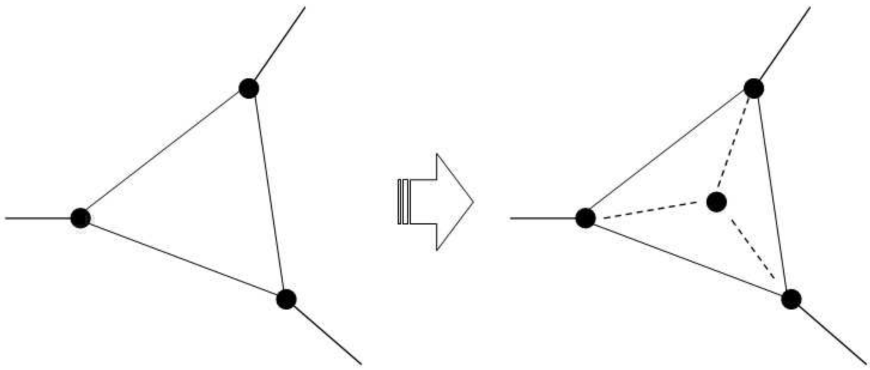

Figure 4.1: This figure shows how to interpolate a new vertex on *high-resolution mesh* $\mathbf{M}^+$ from the vertices on *low-resolution mesh* **M**. Each new vertex is added in the centre of the triangular face. The location of the new vertex is obtained by averaging the locations of surrounded 3 triangular vertices. This figure can also be used to indicate the averaging process of $\mathbf{T}_t^+$. The new element corresponding to the interpolated new vertex on M^+ is obtained by averaging the elements of $\mathbf{T}_t$ on surrounded 3 triangular vertices(which from the low-resolution mesh M).

In other words, the vertices of high-resolution mesh **M** belong to two groups, the vertices from the low-resolution mesh and the interpolated new vertices.

4.3 Revising reconstructed current source on low-resolution mesh using the Kalman filter

4.3.1 Revision of reconstructed current source

In this Chapter, we aim at improving the spatial resolution of source distribution at one particular time point (also called as *target time point*) from the low-resolution mesh **M** to high-resolution mesh $\mathbf{M}^+$ by the Bayesian super-resolution method. We firstly need to work out a set of low-resolution source distribution on a continuous time series (including the *target time point*). These reconstructions on **M** are implemented by the basis function method.

The reconstructions on the low-resolution mesh **M** by the basis function method are

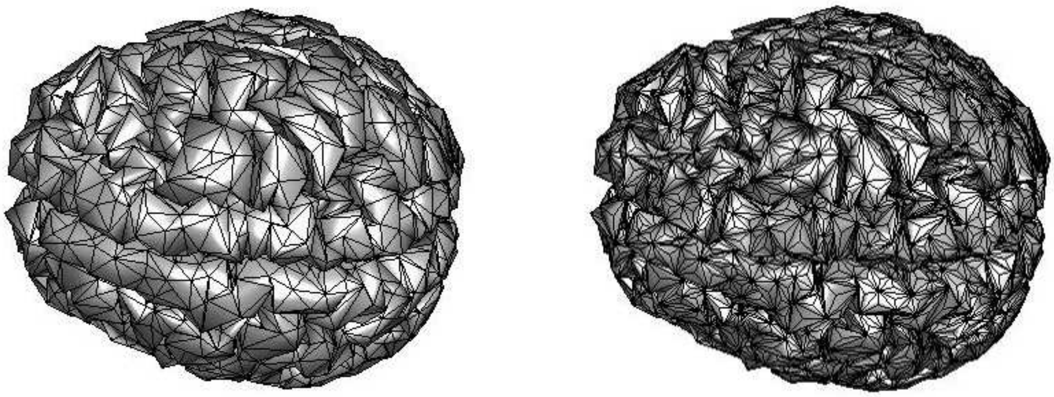

Figure 4.2: This figure shows the *low-resolution mesh*, **M** on the left (with V vertices and F faces) and the *high-resolution mesh*, $\mathbf{M}^+$ on the right with $V^+ = V + F$ vertices and $F^+ = 3F$ faces. **M** has 2600 vertices and 5192 faces; $\mathbf{M}^+$ is with 7792 vertices and 15576 faces.

calculated for each single time point firstly with the discrete state, and then combined together through the time series as the dynamic system. The idea is to provide a series of predictions for the currents at the target time-point (using the time series), and then combine these predictions using super-resolution. Therefore, it is essential to smooth these current sources over the time sequence to obtain a more accurate prediction of the signals. This process of smoothness can provide a more rational prior for the later reconstruction on M^+. For estimating the states in the past, present and future, the Kalman filter can be applied in a straightforward way for the reconstruction and smoothness of the current sources in time series with respect to the Markov property.

4.3.2 Kalman filter

From the study of Kalman filter in *Chapter 2*, the real-time MEG state and measurement can be described as the following (Welch and Bishop, 2006):

$$\mathbf{x}_t = \mathrm{A}\mathbf{x}_{t-1} + \omega \tag{4.3}$$

$$\mathbf{z}_t = \mathrm{B}\mathbf{x}_t + \upsilon \tag{4.4}$$

Here $\mathbf{x}_t$ is the the hidden state and $\mathbf{z}_t$ is the observation. $\mathbf{A}$, and $\mathbf{B}$ are the relevant coefficient matrices for $\mathbf{x}_{t-1}$ and $\mathbf{x}_t$, respectively. ω is the state noise , and υ is the observation noise.

4.3.2.1 Smoothing the successive source distribution J on mesh M

The state vector $\mathbf{x}_t$ of the 3D reconstructed current source at the single time point t is given by Eqn 4.5 as follows:

$$\mathbf{x}_t = \begin{pmatrix} J_x^{(t)} \\ J_y^{(t)} \\ J_z^{(t)} \\ \upsilon_x^{(t)} \\ \upsilon_y^{(t)} \\ \upsilon_z^{(t)} \end{pmatrix} \tag{4.5}$$

where $\mathbf{J}_t = (J_x^{(t)}, J_y^{(t)}, J_z^{(t)})$, is the estimated current of the single point t , and $\mathbf{v}_t = (\upsilon_x^{(t)}, \upsilon_y^{(t)}, \upsilon_z^{(t)})$ is the associated rate of change of the current.

The state matrix is:

$$\mathbf{A} = \begin{pmatrix} \mathbf{I} & \mathbf{I} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \tag{4.6}$$

and the observation matrix is

$$\mathbf{H} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \end{pmatrix} \tag{4.7}$$

$\mathbf{Q}$ is the covariance matrix of the state process and $\mathbf{R}$ is the covariance matrix of the observation noise (which is determined from data). We then get the Kalman update equations as follows:

$$\mathbf{x_t^-} = \mathbf{A}\mathbf{x_{t-1}} \tag{4.8}$$

$$\mathbf{P_t^-} = \mathbf{A P_{t-1} A^T} + \mathbf{Q} \tag{4.9}$$

$$\mathbf{K_t} = \mathbf{P_t^- H^T (H P_t^- H^T + R)^{-1}} \tag{4.10}$$

$$\mathbf{x_t} = \mathbf{x_t^-} + \mathbf{K_t (J_t - H_{x_t^-})} \tag{4.11}$$

$$\mathbf{P_t} = \mathbf{(I - K_t H) K_t^-} \tag{4.12}$$

where $\mathbf{J_t}$ represents the observed current source in specific time point (which includes 3 dimensional data). By applying this process, we obtain a noise-reduced set of estimates $\mathbf{x}_t$ for the current source.

4.3.2.2 Smoothing associated basis function coefficients $\mathbf{a}$ of source distribution J

The Kalman filter method works by smoothing the source distribution **J** (size: $M \times 1$, M is the number of mesh vertices) on **M** at the successive time points as a recursive process. Since the basis function method is used for the source reconstruction on each single time point, the coefficients of basis function $\mathbf{a}$ can be used for simplifying this process. In terms of the basis function method explained in last Chapter, the source distribution **J** at one single time point is consist of two components shown as following equation, basis function set $\tilde{\mathbf{\Phi}}$ and the corresponding coefficients $\mathbf{a}$ (also shown as Eqn 3.17 in Chapter 3):

$$\mathbf{J} = \tilde{\mathbf{\Phi}} \mathbf{a} \tag{4.13}$$

Since the basis function set $\tilde{\mathbf{\Phi}}$ are fixed as we choose, the only variable in the equation above is the associated coefficient $\mathbf{a}$ for each basis functions. The structure of $\mathbf{a}$ (size: $3T \times 1$, T is the number of basis function set) for each time point is shown as follows (also shown as Eqn 3.20 in Chapter 3):

$$\mathbf{a} = \begin{pmatrix} a_{x1} \cdots a_{xt} \cdots a_{xT} & a_{y1} \cdots a_{yt} \cdots a_{yT} & a_{z1} \cdots a_{zt} \cdots a_{zT} \end{pmatrix}^T \tag{4.14}$$

With respect to the process of Kalman filter explained above, $\mathbf{a}$ for the successive time points can be smoothed to obtain a group of new coefficients, $\mathbf{a}^{new}$. Therefore, $\mathbf{J}^{new}$, the smoothed source distributions on the successive time points, are generated according to the Eqn 4.13.

As the size of $\mathbf{a}$ is much smaller than the size of $\mathbf{J}$ ($3T \ll M$), the approach that smoothing $\mathbf{a}$ via Kalman filter method instead of directly using the source distribution $\mathbf{J}$ reduces the computation complexity effectively with the same smoothing results.

4.4 Applying Bayesian super-resolution on improving spatial resolution of MEG source reconstruction

4.4.1 Selecting a prior

In terms of our problem, the prior an important constraint for the solution. As we discussed in Chapter 3 , the cortical mesh Laplacian can be constructed as follows (Chung, 1997), (Cvetkovic et al., 1997):

$$\mathcal{L} = \mathcal{D} - \mathcal{A} \tag{4.15}$$

where $\mathcal{A}$ is the adjacency matrix and $\mathcal{D}$ is the degree matrix, a diagonal matrix represents the degree of each vertex shown by in the diagonal elements.

In the following part, $i = 1, \cdots, M$ denotes the vertex index of $\mathbf{M}$ and $j = 1, \cdots, N$ denotes the vertex index of $\mathbf{M}^{+}$.

Since both the high resolution mesh $\mathbf{M}^{+}$ and the low resolution mesh $\mathbf{M}$ are obtained previously, the corresponding mesh Laplacian $\mathbf{L}$ and $\mathbf{L}^{+}$ can be generated. The prior on the high resolution mesh $\mathbf{M}^{+}$ is given as a Normal distribution, showed as Eqn 4.16;

$$p(x) = \mathcal{N}(x|0, \mathbf{Z}_x) \tag{4.16}$$

On the high resolution mesh $\mathbf{M}^{+}$, the covariance $\mathbf{Z}_x$ of this distribution is assumed as a heat kernel. The reason for using the heat kernel is that the anisotropic diffusion across the cortical mesh with time for the synthetic source is captured by the heat kernel and this enforce smoothness between adjacent mesh vertices.

$$\mathbf{Z}_x = \hat{A} \begin{pmatrix} \exp(-L^+_{1,1}\alpha) & \cdots & \cdots & \exp(-L^+_{1,N}\alpha) \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \exp(-L^+_{N,1}\alpha) & \cdots & \cdots & \exp(-L^+_{M,M}\alpha) \end{pmatrix} \tag{4.17}$$

Note this prior is fixed with a constant prior α (which we set as 0.5 in our experiment).

In light of the approach of super resolution used by (Tipping and Bishop, 2003) , we build a standard mathematical model to describe the relationship between current sources on the low-resolution mesh **M** and the high-resolution mesh $\mathbf{M}^+$ (indicated as Eqn 4.18) where the process of generating low resolution frames from the high resolution frame can be assumed by applying a time shift (namely the current sources varying between high-resolution frame at time t_0 and high-resolution frames at time point t) for convolving with point spread function (PSF) and downsampling to the lower resolution mesh.

$$\mathbf{J}^{(t)} = \mathbf{W}^{(t)}\mathbf{J}^+ + \epsilon^{(t)} \tag{4.18}$$

where $\epsilon^{(t)}$ is a vector of independent Gaussian random variables $\epsilon_i^{(t)} \sim \mathcal{N}(0, \beta^{-1})$ with zero mean and precision(inverse variance) β . This is used to represent the noise terms between the generative model and observed data; t represents the time of the frames, $\mathbf{J}^{(t)}$ represents the current source distributed on the low-resolution mesh **M** at the time point t; **W** is the transformation matrix; Our goal is to estimate the high-resolution frame at a particular time point t_0, $\mathbf{J}^+$.

The transformation matrix **W** in Eqn 4.18 can be defined in the following steps:

$$\tilde{\mathbf{W}}^{(t)} = \mathbf{S}\mathbf{T}_t \tag{4.19}$$

The downsampling **S** in Eqn 4.19 is defined in a straightforward way by the heat kernel:

$$\mathbf{S} = \mathbf{P}\mathbf{E}_\gamma \tag{4.20}$$

where

$$\mathbf{E}_\gamma = \begin{pmatrix} \exp(-L^+_{1,1}\gamma) & \cdots & \cdots & \exp(-L^+_{1,N}\gamma) \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \exp(-L^+_{N,1}\gamma) & \cdots & \cdots & \exp(-L^+_{M,M}\gamma) \end{pmatrix} \tag{4.21}$$

So, here $\mathbf{P}$ is a downsampling operator which only picks out the vertices from the low-resolution mesh $\mathbf{M}$; $\mathbf{E}_\gamma$ is the point spread function(so-called PSF), with the unknown parameter γ in 3-space. In our case, the PSF describes the limited resolution of the low-resolution current distribution with respect to the high resolution distribution with respect to the geometrical information of the high-resolution mesh (refer to the mesh Laplacian $\mathbf{L}^+$), i.e. the parameter γ in turn is a measure of the smoothness degree.

The transfer matrix $\mathbf{T}_t$ is used to describe a shift of the high-resolution frame from the *target time point* t_0 to another time t. The process for calculate $\mathbf{T}_t$ is shown as follows. We firstly get the current sources varying $\mathcal{T}_t$ on low-resolution frames between the *target time* t_0 and any other time point t in the time sequence we choose, shown as Eqn 4.22. Therefore, $\mathbf{J}_t$, the current source in each time point t, has a corresponding $\mathbf{T}_t$ which represents the proportional shift with the current source on *target time point*, $\mathbf{J}_{t0}$. Similar to the averaging process for getting new vertices of $\mathbf{M}^+$(referr to Fig 4.1), each 3 elements of $\mathbf{T}_t$ that corresponding to the current source located on the triangular vertices of each face of mesh $\mathbf{M}$ are averaged to obtain the new T_{tn} that is corresponding to the current source located on the new vertex generated in M^+. Then, for each $\mathbf{J}^+_t$, we produce a $\mathbf{T}^+_t$ with the same size. Diagonalizing the elements of $\mathcal{T}^+_t$ into the square matrix with other elements set as zeros, current source varying $\mathcal{T}_t$ is obtained for the following calculation (shown in Eqn 4.25).

$$\begin{aligned} \mathcal{T}_t &= \mathbf{J}_t./\mathbf{J}_{t_0} \\ &= (\mathcal{T}_{t1} \;\; \mathcal{T}_{t2} \;\; \cdots \;\; \mathcal{T}_{tM})^T \end{aligned} \tag{4.22}$$

$$\mathbf{T}^{v_{fn}}_t = \frac{(T^{(v_{f1})}_t + T^{(v_{f2})}_t + T^{(v_{f3})}_t)}{3} \tag{4.23}$$

$$\mathbf{T}^+_t = \begin{pmatrix} \mathcal{T}_t \\ \mathbf{T}^{v_{fn}}_t \end{pmatrix} \tag{4.24}$$

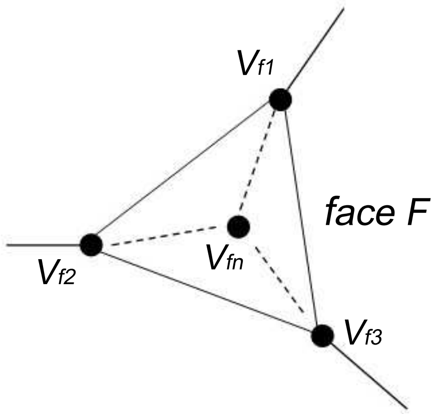

Figure 4.3: This figure demonstrates the example face F of $\mathbf{M}$ for averaging $\mathcal{T}$ to generate $\mathbf{T}_t^{v_{fn}}$ (also refer to Eqn 4.24). The triangular vertices v_{f1}, v_{f2} and v_{f3} are the vertices from $\mathbf{M}$. V_{fn} is the interpolated new vertex of $\mathbf{M}^+$ which is generated from v_{f1}, v_{f2} and v_{f3}.

$$\mathbf{T_t} = \begin{pmatrix} T_{t1} & \cdots & \cdots & 0 \\ \vdots & T_{t2} & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & T_{tN} \end{pmatrix} \tag{4.25}$$

The transformation matrix $\mathbf{W}^{(t)}$ in Eqn 4.19 govern by the matrix $\mathbf{S}$ (including the downsampling matrix $\mathbf{P}$ and point spread function (PSF)) as well as transformation matrix $\mathbf{T}_t$ can be finally presented as the normalized form in Eqn 4.26. With the normalization, the $\mathbf{J}^{(t)}$ can be ensured to be transformed from $\mathbf{J}^+$ with the same scale.

$$W_{ij}^{(t)} = \tilde{W}_{ij}^{(t)} / \sum_{j'} \tilde{W}_{ij'}^{(t)} \tag{4.26}$$

4.5 Estimation

4.5.1 Posterior estimation

The likelihood of an individual scan (on the low-resolution mesh) is as follows(Tipping and Bishop, 2003) :

$$p(J^{(t)}|J^{+},\gamma) = (\beta/2\pi)^{M/2}\exp\left\{-\frac{\beta}{2}\|\mathbf{J}^{(t)} - \mathbf{W}^{(t)}\mathbf{J}^{+}\|^{2}\right\} \tag{4.27}$$

Therefore assuming conditional independence of the scans(on the low-resolution mesh) given the high resolution scan, we get(Tipping and Bishop, 2003) :

$$p(J^{+}|J^{(t)},\gamma) = \mathcal{N}(\mu,\mathbf{\Sigma}) \tag{4.28}$$

Where,

$$\mathbf{\Sigma} = \left[\frac{1}{A}\exp(\alpha\mathbf{L}^{+}) + \beta\sum_{t}\mathbf{W}^{(t)T}\mathbf{W}^{(t)}\right]^{-1} \tag{4.29}$$

$$\mu = \beta\mathbf{\Sigma}\sum_{t}\mathbf{T}_{t}^{T}\mathbf{W}^{(t)T}\mathbf{J}^{(t)} \tag{4.30}$$

μ is the optimized current source distributed on the high resolution mesh $\mathbf{M}^{+}$ which is equivalent to $\mathbf{J}^{+}$ that we are trying to calculate by the Bayesian super-resolution method. For computing this, it is essential to estimate the unknown parameter γ .

4.5.2 Energy function of Bayesian super-resolution

According to the theory of Tipping and Bishops paper(Tipping and Bishop, 2003), the critical step is to marginalize out the unknown high-resolution image from the known equations, so that the probability of the registration parameters and point spread function(PSF) are assumed to be correct, and the marginal likelihood function for low-resolution images is shown in the form:

$$p(J|\{s_k,\theta_k\},\gamma) = \mathcal{N}(0,\mathbf{Z}_y) \tag{4.31}$$

where

$$\mathbf{Z}_y = \beta^{-1}\mathbf{I} + W\mathbf{Z}_x W^{T} \tag{4.32}$$

Here, $\{s_k, \theta_k\}$ is the registration parameters; γ is the parameter of PSF; J and W are the vector and matrix of the stacked $\mathbf{J}^t$ and $\mathbf{W}^t$, respectively. With the marginalization shown in Eqn 4.31 and Eqn 4.32, the marginal likelihood can be rewrote as Eqn 4.33 after some *standard matrix manipulations* are performed on it:

$$\log p(y|\{s_k, \theta_k\}, \gamma) = -\frac{1}{2}[\beta \sum_t (\vec{\mathbf{J}}^{(t)} - \mathbf{W}^{t)} \vec{\mu})^2 + \vec{\mu}^T \mathbf{Z}_x^{-1} \vec{\mu} + \log|\mathbf{Z}_x| - \log|\mathbf{\Sigma}| - KM \log \beta] \tag{4.33}$$

Eqn 4.34 must be optimized to obtain the most likely combination of registration parameters and PSF, therefore, we define an energy function from the marginal log-likelihood as:

$$E = -\frac{1}{2}[\beta \sum_t (\vec{\mathbf{J}}^{(t)} - \mathbf{W}^{t)} \vec{\mu})^2 + \vec{\mu}^T \mathbf{Z}_x^{-1} \vec{\mu} + \log|\mathbf{Z}_x| - \log|\mathbf{\Sigma}| - KM \log \beta] \tag{4.34}$$

However, it is notable that a few differences exist between our problem and image super-resoluion explained previously. In our problem, we are working with the mesh and not the image grids. Also, there is no alignment problem between the successive frames here. Instead, we are dealing with the problem that current sources may vary among the frames in the time series (without rotation and blur in the imaging problem, which is related to $\{s_k, \theta_k\}$) because the current sources are dynamic. In this case, we only need to estimate the parameter of PSF, γ ,in the optimization. This will be explained in the later part of this Chapter.

4.5.3 Parameter optimization

As the Bayesian super-resolution method explained by Tipping and Bishops paper (Tipping and Bishop, 2007) , Bayesian marginalization allows the registration parameters as well as PSF to be estimated in advance so that the high-resolution frame generated afterward can be estimated with superior accuracy. In this case, the parameter γ of point spread function (PSF) are estimated firstly by optimizing the energy function Eqn 4.34.

$$\frac{\partial E}{\partial \gamma} = \left(\sum \frac{\partial E}{\partial W_{ij}^{(t)}} \frac{\partial W_{ij}^{(t)}}{\partial \gamma} \right) \doteq 0 \tag{4.35}$$

The calculation of $\frac{\partial \tilde{W}_{ij}^{(t)}}{\partial \gamma}$ are presented as follows:

According to Eqn 4.26, the denominator $\sum_{j'} \tilde{W}_{ij'}$ is a normalization factor which should be constant regarded less of the choice of parameters, so that we have:

$$\frac{\partial W_{ij}^{(t)}}{\partial \gamma} = \frac{\partial W_{ij}^{(t)}}{\partial \tilde{W}_{ij}^{(t)}} \frac{\partial \tilde{W}_{ij}^{(t)}}{\partial \gamma} \tag{4.36}$$

with

$$\frac{\partial W_{ij}^{(t)}}{\partial \tilde{W}_{ij}^{(t)}} = \frac{1}{\sum_{j'} \tilde{W}_{ij'}^{(t)}} \tag{4.37}$$

and

$$\frac{\partial \tilde{W}_{ij}^{(t)}}{\partial \gamma} = \mathbf{P}\mathbf{L}^{+} \exp(-\mathbf{L}^{+}\gamma)\mathbf{T}_t \tag{4.38}$$

Meanwhile, $\frac{\partial E}{\partial W_{ij}^{(t)}}$ in Eqn 4.35 can be expanded as the following expression with multiple factors:

$$\begin{aligned} \frac{\partial E}{\partial W_{ij}^{(t)}} = -\frac{1}{2}\Bigg[& -2\beta\left(\sum_t (\mathbf{J}^{(t)} - \mathbf{W}^{(t)}\mu)^T (\mathbf{W}^{(t)} \frac{\partial \mu}{\partial W_{ij}^{(t)}})\right) - 2\beta(\mathbf{J}^{(t)} - \mathbf{W}^{(t)}\mu)^T \mu_{ij}^{*} \\ & + 2\mu^T \exp \alpha \mathbf{L}^{+} \frac{\partial \mu}{\partial W_{ij}^{(t)}} - \mathrm{Tr}\left(\mathbf{\Sigma}^{-1} \frac{\partial \Sigma}{\partial W_{ij}^{(t)}}\right)\Bigg] \end{aligned} \tag{4.39}$$

The factors in Eqn 4.39 can be calculated as:

- μ_{ij}^{*} can be written as:

$$(\mu_{ij}^{*})_t = \begin{cases} \mu_j & \text{if } t = i \\ 0 & \text{otherwise} \end{cases}$$

- In terms of Eqn 4.30, $\frac{\partial \mu}{\partial W_{ij}^{(t)}}$ can be written as follows:

$$\frac{\partial \mu}{\partial W_{ij}^{(t)}} = -\beta \mathbf{\Sigma} \mathbf{W}_{ij}^{*(t)} \mu + \beta \mathbf{\Sigma} \mathbf{J}_{ij}^{*(t)} \tag{4.40}$$

where

$$(\mathbf{J}^*_{ij})_t = \begin{cases} J_j & \text{if } t = j \\ 0 & \text{otherwise} \end{cases}$$

and

$$(\mathbf{W}^{*(t)}_{ij})_{ab} = \begin{cases} W^{(t)}_{ij} & \text{if } a = j \text{ and } b = j \\ W^{(t)}_{ib} & \text{if } a = j \text{ and } b \neq j \\ W^{(t)}_{ia} & \text{if } b = j \text{ and } a \neq j \\ 0 & \text{otherwise} \end{cases}$$

The equations above indicate that μ (size: $N \times 1$) reduces the size to $(\mu^*_{ij})_t$ (size: $M \times 1$) with the rest of the elements are cancelled; **J** (size: $M \times 1$) increase the size to $(\mathbf{J}^*_{ij})_t$ (size: $N \times 1$) with the new elements set as zeros for the calculation need; and **W** (size: $M \times N$) increase the size to $\mathbf{W}^{*(t)}_{ij}$ (size: $N \times N$) with the new elements set as zeros, for the calculation need.

- $\frac{\partial \Sigma}{\partial W^{(t)}_{ij}}$ can be expanded as :

$$\frac{\partial \mathbf{\Sigma}}{\partial W^{(t)}_{ij}} = -\beta \mathbf{\Sigma} \mathbf{W}^{*(t)}_{ij} \mathbf{\Sigma} \tag{4.41}$$

With optimized parameter 'γ' for the point spead function(PSF), the high-resolution frame μ can be estimated with improved accuracy.

It is worth to emphasize that all the calculations above are implemented in 3 dimensions (components x,y and z).

4.6 Results

In the following part, the simulated data as well as the real MEG data are both applied for the evaluation of the Bayesian super-resolution method on improving MEG spatial resolution. Since we are working on the high-resolution source distribution on mesh $\mathbf{M}^+$, the corresponding low-resolution source frames on mesh **M** at the same time course must be obtained firstly as a prior. The method used for reconstructing the low-resolution source frames on successive time points are the basis function method explained in the last Chapter.

The experimental configuration (the configuration currently is : InterCore2(1,8GHz), Linux system(2.6.34.1)-32bit, matlab $7.9.0(R2009b) \times 32$ edition, RAM: 4GB).

4.6.1 Synthetic results

In the part of simulated experiment, two groups of simulated current sources are generated for synthetic experiment, i.e. artificial source distribution and realistic source distribution in *Appendix B*. For the former type (called as *synthetic sources A*), the fixed current source values are set on 30 particular vertices of mesh we choose but the values of current sources on other vertices are set as zero; while in the process of generating the later one(called as *synthetic source B*), the source distribution on the cortical mesh are from the selected results of previous source reconstruction of the real MEG data with random stimuli on cortical surface at one time point. The detailed information of these two groups of simulated current sources is given in *Appendix B*.

Since we are working on a successive time period, the number of time points are firstly set as $\mathbb{T} = 31$, and the time point 3 of 31 are set as the *target time point*. The basis function method are then applied to get 31 low-resolution source frames in terms of $\mathbb{T}$. The details of these reconstructions can refer to the result part of *Chapter 3*. The figures, Fig 4.4 and Fig 4.11, show these reconstructions of low-resolution source frames at 31 successive time points for *synthetic source A* and *synthetic source B*, respectively. With respect to the results of the basis function method, the coefficient matrices of the basis functions $\mathbf{a}$ are selected, and smoothed by the Kalman filter along $\mathbb{T}$ time points. The figures, Fig 4.5 and Fig 4.6, demonstrate the smoothing process of 3rd element and 12th element of coefficient matrix $\mathbf{a}$ in successive 31 time points by Kalman filter (in 3-space, with components x, y and z) for *synthetic source A* ; while the figures, Fig 4.12 and Fig 4.13, show are the same process for *synthetic source B*, respectively. The size of low-resolution current source matrix $\mathbf{J}$ on a single time point is 7800×1 which is much bigger than the size of coefficient matrix, 246×1. Therefore, the approach that smoothing $\mathbf{a}$ instead of directly smoothing the source distribution $\mathbf{J}$ effectively reduces the computation complexity.

After that, we obtain the high-resolution mesh $\mathbf{M}^+$ by geometrically interpolating from the low-resolution mesh $\mathbf{M}$. Each new vertex is added in the centre of the triangular face(referred as Fig 4.1). The new mesh $\mathbf{M}^+$ is obtained with 7792 vertices(

represented as the 7792×3 matrix, shown in Eqn 4.2) and 15576 faces (represented as the 15576×3 matrix).

After then, using the Laplacian of the mesh $\mathbf{M}$ and $\mathbf{M}^+$ we extracted, the Bayesian super-redsolution method is applied for improving the spatial resolution of MEG source reconstruction. Since all the factors applied in this calculation process are large size matrices which leads to the much expensive computation, there are some approaches applied here:

- Selecting the matrices of the factors contains a number of zeros, such as transformation matrix $\mathbf{T}_t$, and transform them to be the format *sparse matrix*. Thus, all the relevant calculation of them are transformed to be the *sparse matrix calculation* in *Matlab* with effectively reduce the computation expense rather than the full-matrix computation;

- The calculation is involved into the unaffordable expensive computation when calculating the matrix of Σ which is the inverse of large size matrix(size 7792×7792) with less sparse. Therefore, the software *SPAI* (Grote and Hagemann) is applied for calculating the approximation of the inverse of this large size matrix.

With the optimization process illuminated above, the optimized parameter γ_x, γ_y and γ_z for 3-space can be obtained by the analysis on each coordinate separately. The figures, Fig 4.7, Fig 4.8 and Fig 4.9, demonstrate this optimization process for *Synthetic source A* on x, y and z component, respectively. And, the figures, Fig 4.14, Fig 4.15 and Fig 4.16, show the same experiment results for *Synthetic source B*.

With respect to the optimized parameter γ, the high-resolution source frame on the *target time point* are produced via the Bayesian super-resolution method. The figures, Fig 4.10 and Fig 4.17, show the comparison between the high-resolution current source simulation(referred to *Appendix B*) and the reconstruction by the Bayesian super-resolution method for *synthetic source A* and *synthetic source B*, separately.

From the reconstruction results of simulated experiment above, shown in Fig 4.17 and Fig 4.17, the source reconstruction by the estimated parameters we obtained lead to the distortion of the location and strength of the original simulated current source.

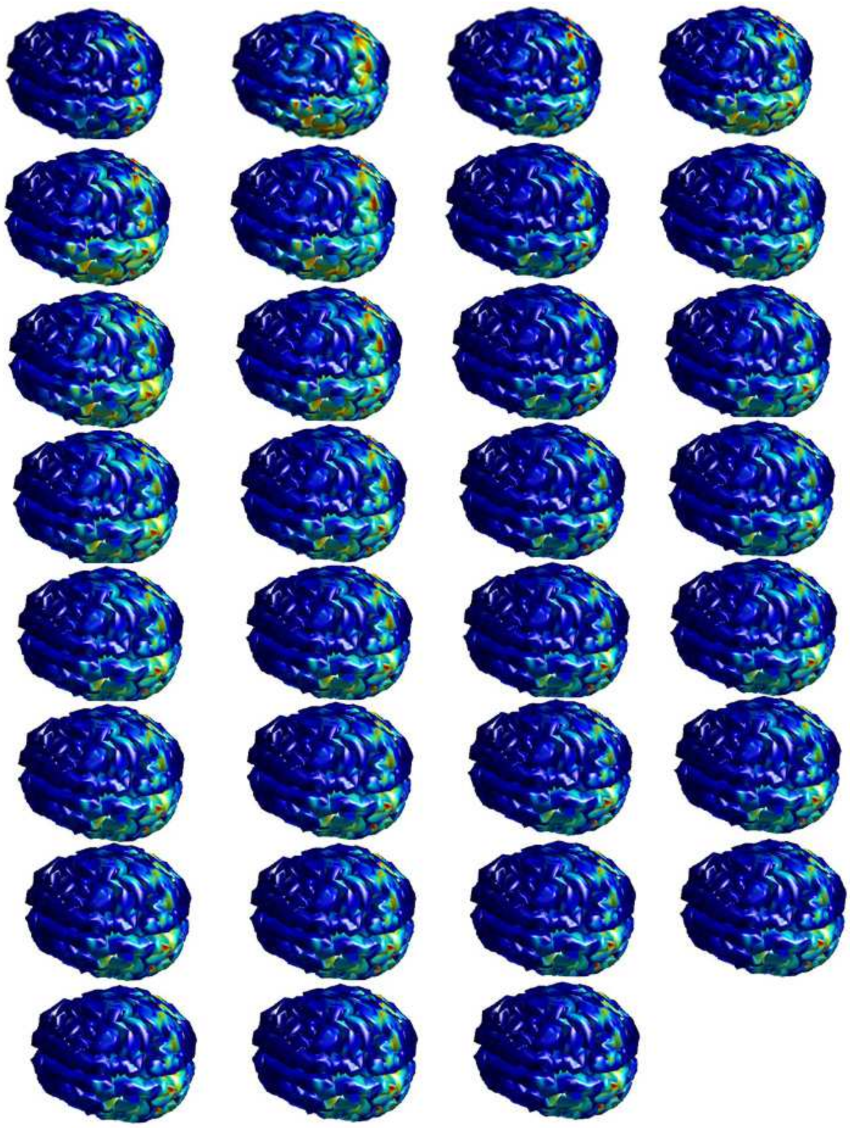

Figure 4.4: This figure demonstrates the reconstructions of low-resolution current source frames at 31 successive time points for *synthetic source A*. The algorithm applied for these reconstructions are the Basis function method. The 1st row: the reconstructions from 1 ms to 4 ms; the 2nd row: the reconstructions from 5 ms to 8 ms; the 3rd row: the reconstructions from 9 ms to 12 ms; the 4th row: the reconstructions from 13 ms to 16 ms; the 5th row: the reconstructions from 17 ms to 20 ms; the 6th row: the reconstructions from 21 ms to 24 ms; the 7th row: the reconstructions from 25 ms to 28 ms; the 8th row: the reconstructions from 29 ms to 31 ms.

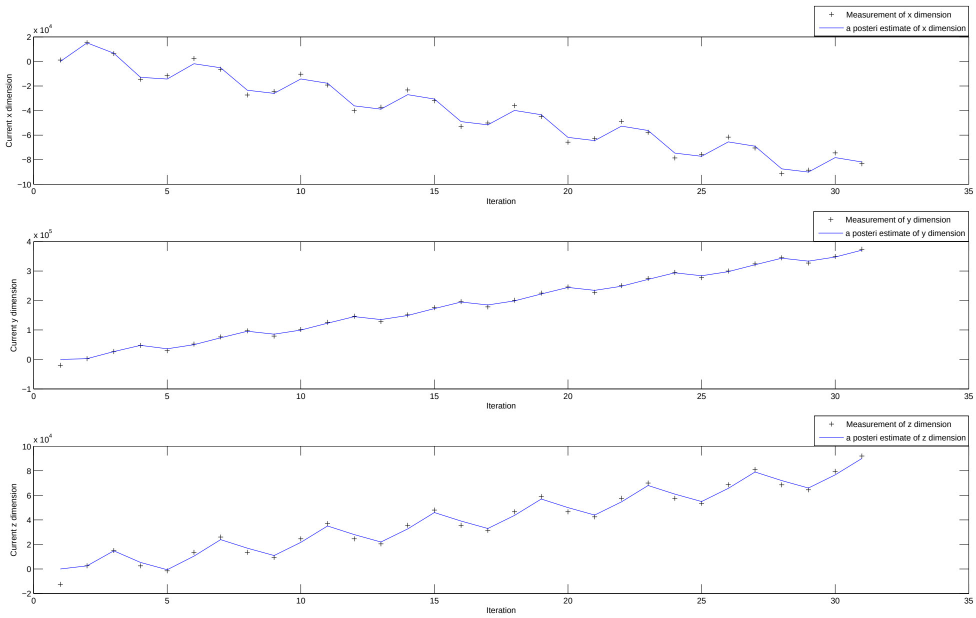

Figure 4.5: This figure shows the smoothing process of 3rd element of coefficient matrix $\mathbf{a}$ in successive 31 time points by Kalman filter(in 3-space, with components x, y and z) for *synthetic source A*. This $\mathbf{a}$ is generated for mesh $\mathbf{M}$ by the basis function method over the time sequence to obtain a more accurate prediction of the current sources. The '*' shows the original reconstructions of $\mathbf{a}$ by the basis function method; the blue line shows *the posterior estimate by Kalman filter estimation on each component of 3 dimensions*.

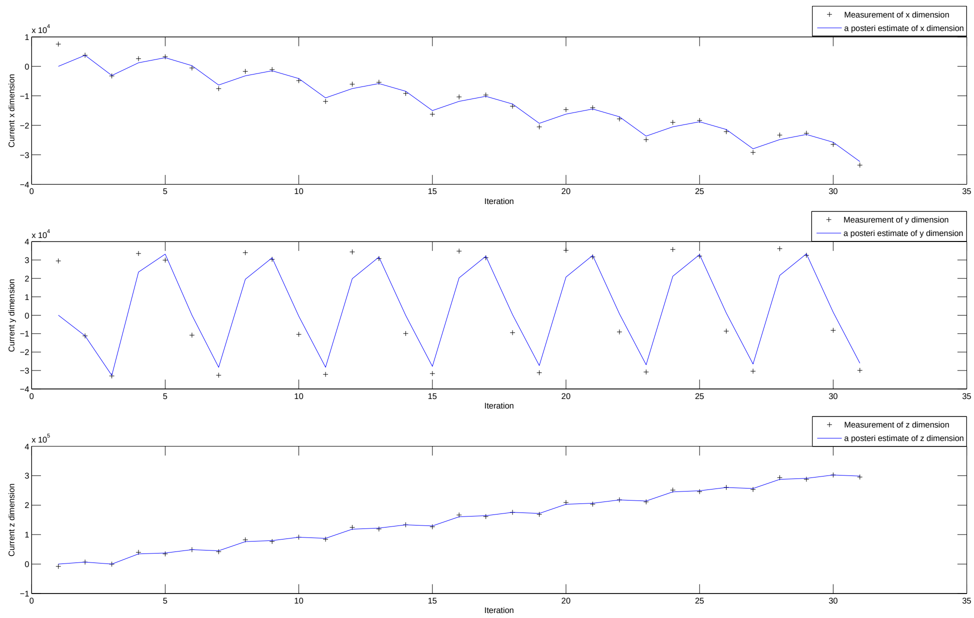

Figure 4.6: This figure shows the smoothing process of 12th element of coefficient matrix $\mathbf{a}$ in successive 31 time points by Kalman filter (in 3-space, with components x, y and z) for *synthetic source A*. This $\mathbf{a}$ is generated for mesh $\mathbf{M}$ by the basis function method over the time sequence to obtain a more accurate prediction of the current sources. The '*' shows the original reconstructions of $\mathbf{a}$ by the basis function method; the blue line shows *the posterior estimate by Kalman filter estimation on each component of 3 dimensions*.

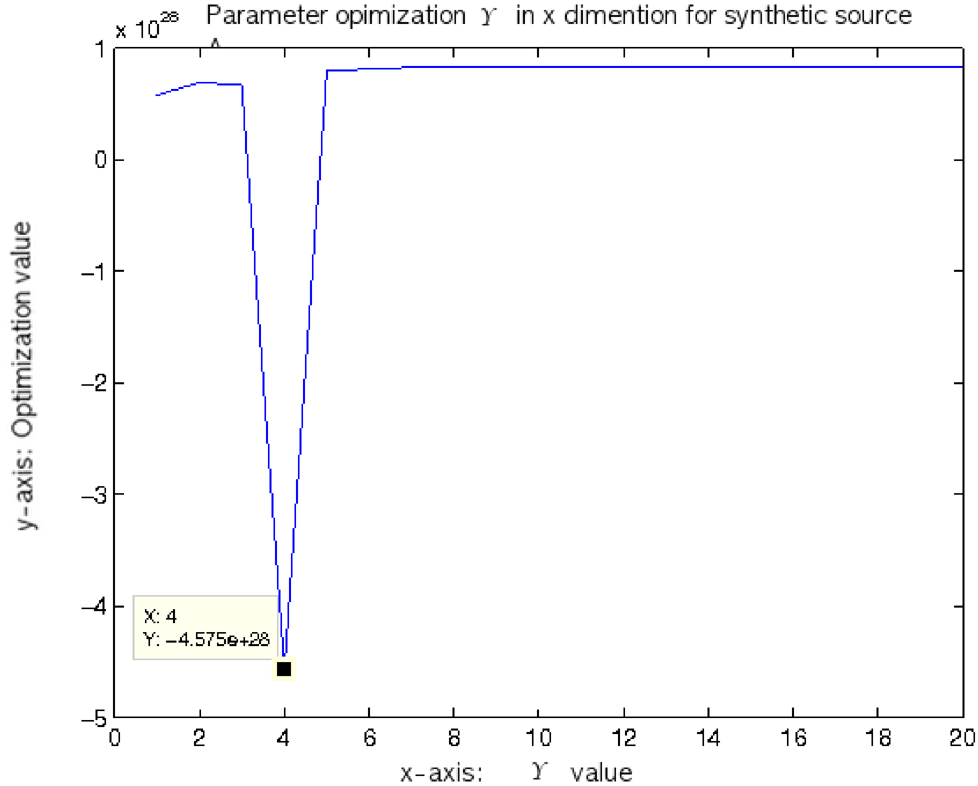

Figure 4.7: This figure shows the optimization results for γ in x dimensions for *synthetic source A*. With the optimization calculation explained in *Parameter optimization*, the minimum values are obtained numerically. The x-axis shows the numerical value of x, and y-axis shows the approximation of E. With y-axis reaches to minimum, the corresponding value in x-axis are the optimization of x.

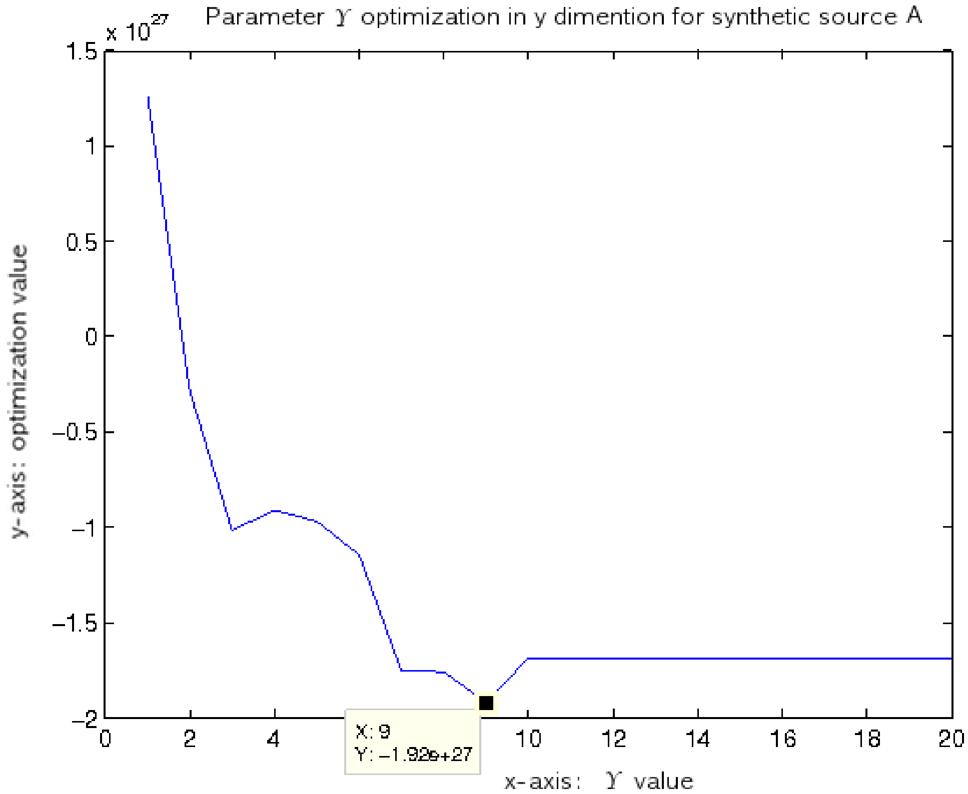

Figure 4.8: This figure shows the optimization results for γ in y dimensions for *synthetic source A*. With the optimization calculation explained in *Parameter optimization*, the minimum values are obtained numerically. The x-axis shows the numerical value of y, and y-axis shows the approximation of E. With y-axis reaches to minimum, the corresponding value in x-axis are the optimization of y.

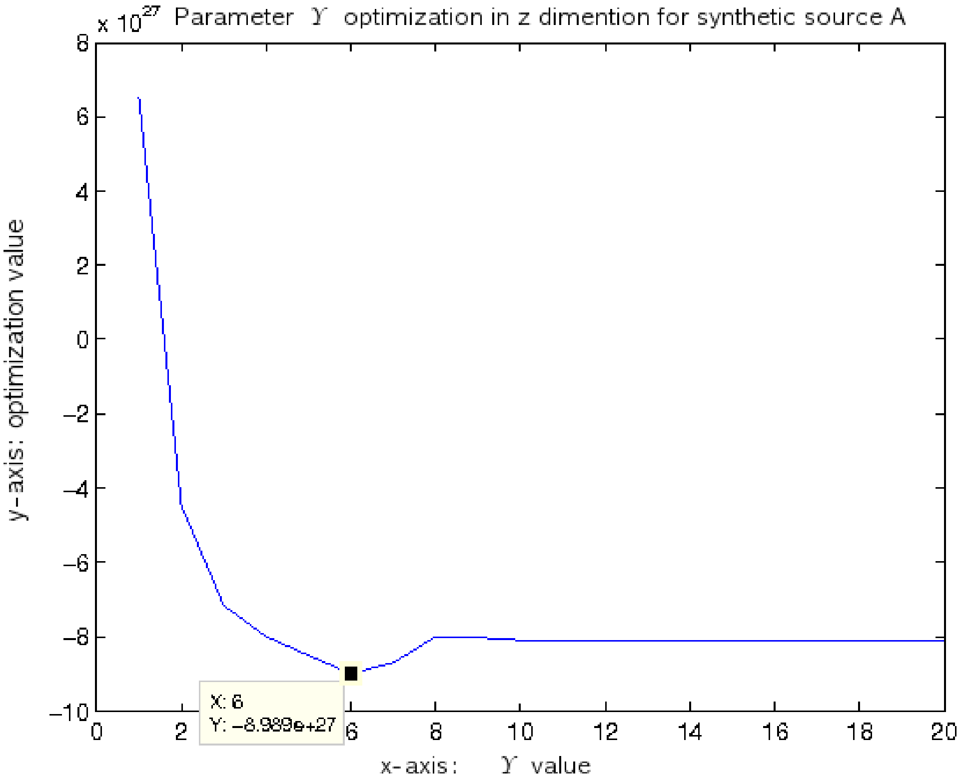

Figure 4.9: This figure shows the optimization results for γ in z dimensions for *synthetic source A*. With the optimization calculation explained in *Parameter optimization*, the minimum values are obtained numerically. The x-axis shows the numerical value of z, and y-axis shows the approximation of E. With y-axis reaches to minimum, the corresponding value in x-axis are the optimization of z.

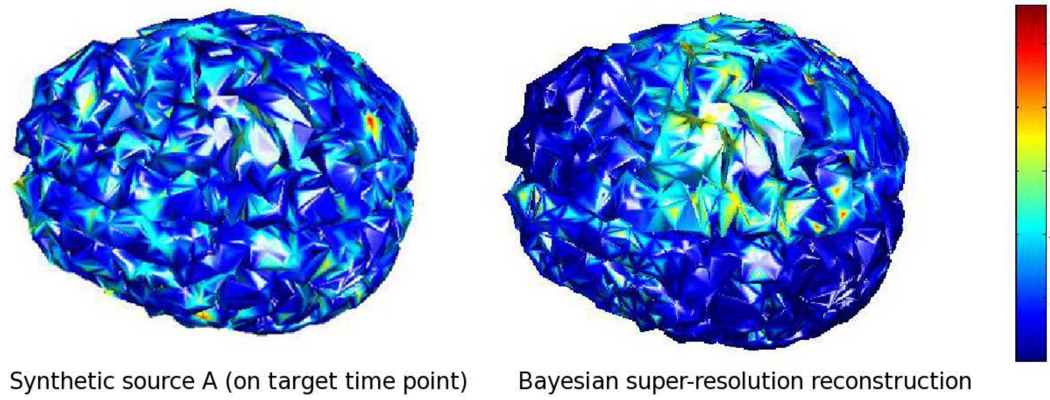

Figure 4.10: This figure shows the comparison of the simulated current source pattern of *synthetic source A* on high-resolution mesh $\mathbf{M}^+$ (on the left) and the reconstruction result by the Bayesian super-resolution method (on the right) at the *target time point*.

4.6.2 Application to the real MEG data

We get the real MEG data of visual expression based on *Appendix 3*. Firstly, we obtained the cortical surface mesh with 262658 vertices and 565782 faces from the structural MRI scan of the same subject by *Freesurfer* (http://surfer.nmr.mgh.harvard.edu/) (5.0.0). Since the coordinate of MRI cortical surface are different with the MEG coordinate, the coordinate registration is processed as the first step(with the special solution provided by YNiC). However, the spatial resolution of mesh obtained from MRI is too large for a realistic or reasonable solution. The simplified mesh is therefore generated by the software *Remesh* (*http://remesh.sourceforge.net/*). In terms of the mesh resolution selected for MEG analysis in (MNE), we apply the reasonable spatial resolution for mesh **M** with 2600 vertices and 5192 faces . Secondly, the measurement of MEG signals is represented as a $96 \times 248 \times 813$ matrix, where 96 indicates the number of different stimulus, 248 indicates the number of sensors and 813 indicates the continuous time instants. The visualization of this measurement matrix is shown in Fig C.1.

We choose 30 continuous time points from 813 continuous time instants, and 70 th stimulus from 96 different stimulus as the *target time point*, then we obtain a 248×30 matrix B^{set} which represents the measurement of magnetic field of 70th stimulus during the continuous time points the time point 20 ,45 , 70, 95 , 120 , 145 ,170 , 195 ,

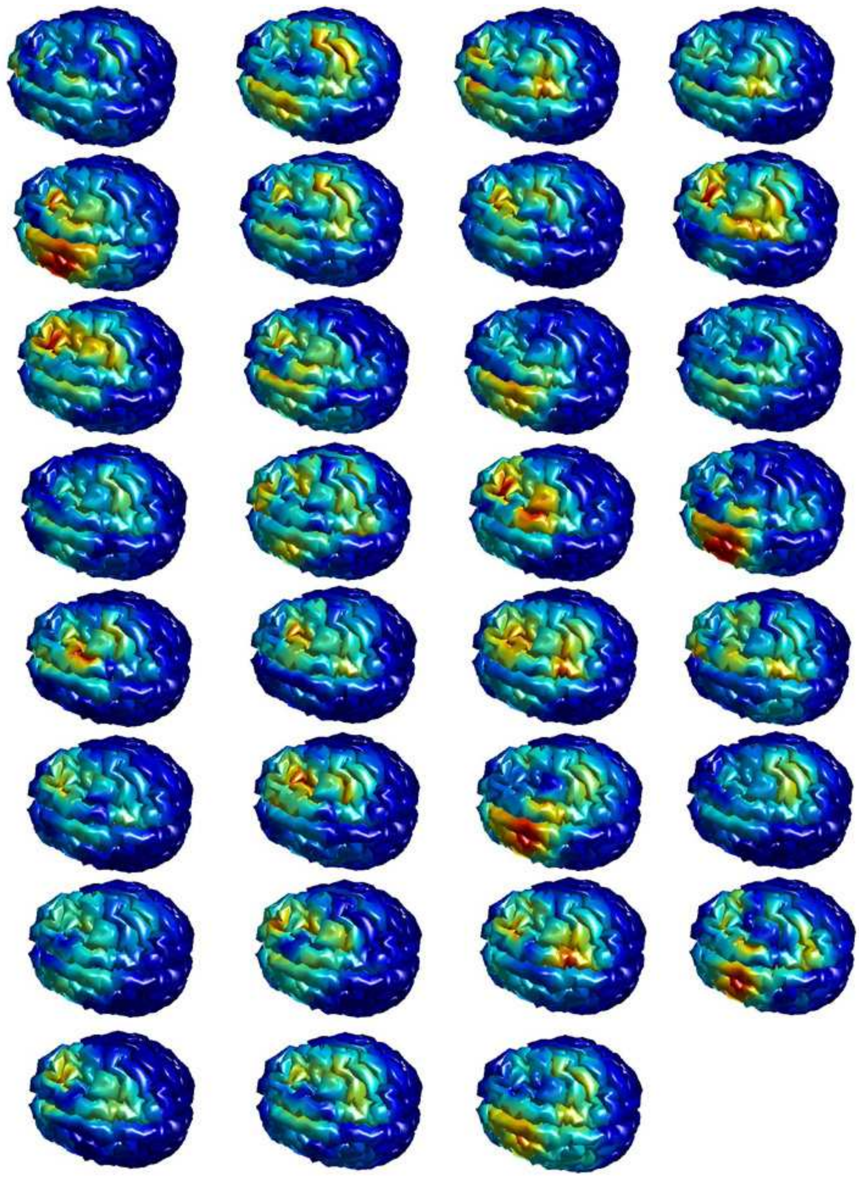

Figure 4.11: This figure demonstrates the reconstructions of low-resolution current source frames at 31 successive time points for *synthetic source B*. The algorithm applied for these reconstructions are the Basis function method. The 1st row: the reconstructions from 1 ms to 4 ms; the 2nd row: the reconstructions from 5 ms to 8 ms; the 3rd row: the reconstructions from 9 ms to 12 ms; the 4th row: the reconstructions from 13 ms to 16 ms; the 5th row: the reconstructions from 17 ms to 20 ms; the 6th row: the reconstructions from 21 ms to 24 ms; the 7th row: the reconstructions from 25 ms to 28 ms; the 8th row: the reconstructions from 29 ms to 31 ms.

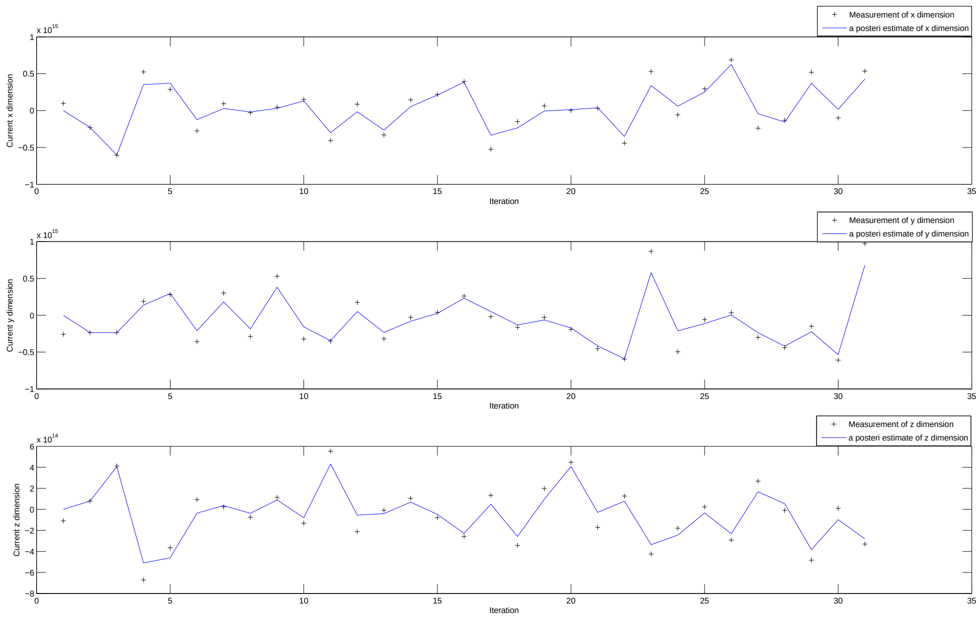

Figure 4.12: This figure shows the smoothing process of 3rd element of coefficient matrix $\mathbf{a}$ in successive 31 time points by Kalman filter (in 3-space, with components x, y and z) for *synthetic source B*. This $\mathbf{a}$ is generated for mesh $\mathbf{M}$ by the basis function method over the time sequence to obtain a more accurate prediction of the current source. The '*' shows the original reconstructions of $\mathbf{a}$ by the basis function method; the blue line shows *the posterior estimate by Kalman filter estimation on each component of 3 dimensions*.

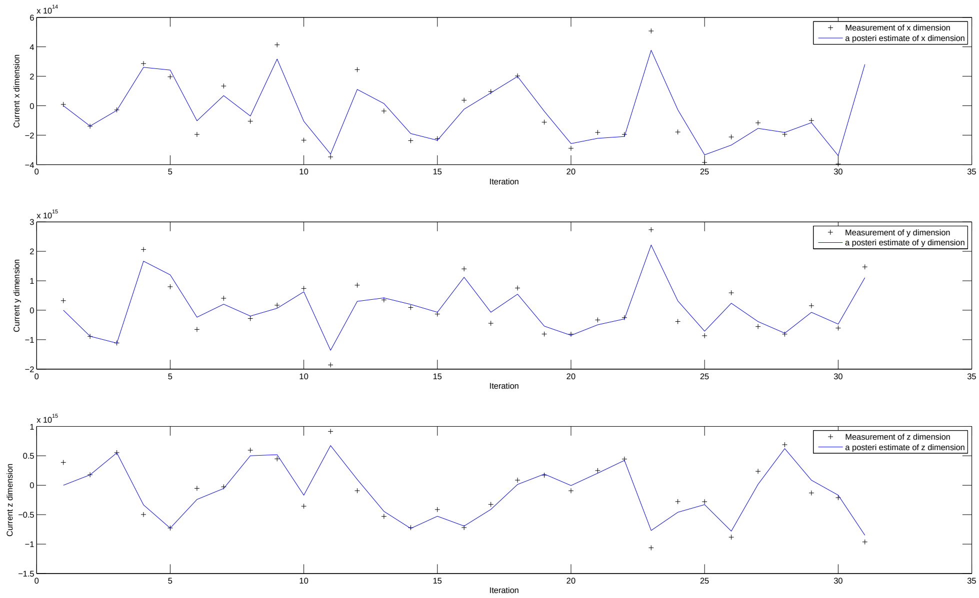

Figure 4.13: This figure shows the smoothing process of 12th element of coefficient matrix $\mathbf{a}$ in successive 31 time points by Kalman filter(in 3-space, with components x, y and z) for *synthetic source B*. This $\mathbf{a}$ is generated for mesh $\mathbf{M}$ by the basis function method over the time sequence to obtain a more accurate prediction of the current source. The '*' shows the original reconstructions of $\mathbf{a}$ by the basis function method; the blue line shows *the posterior estimate by Kalman filter estimation on each component of 3 dimensions*.

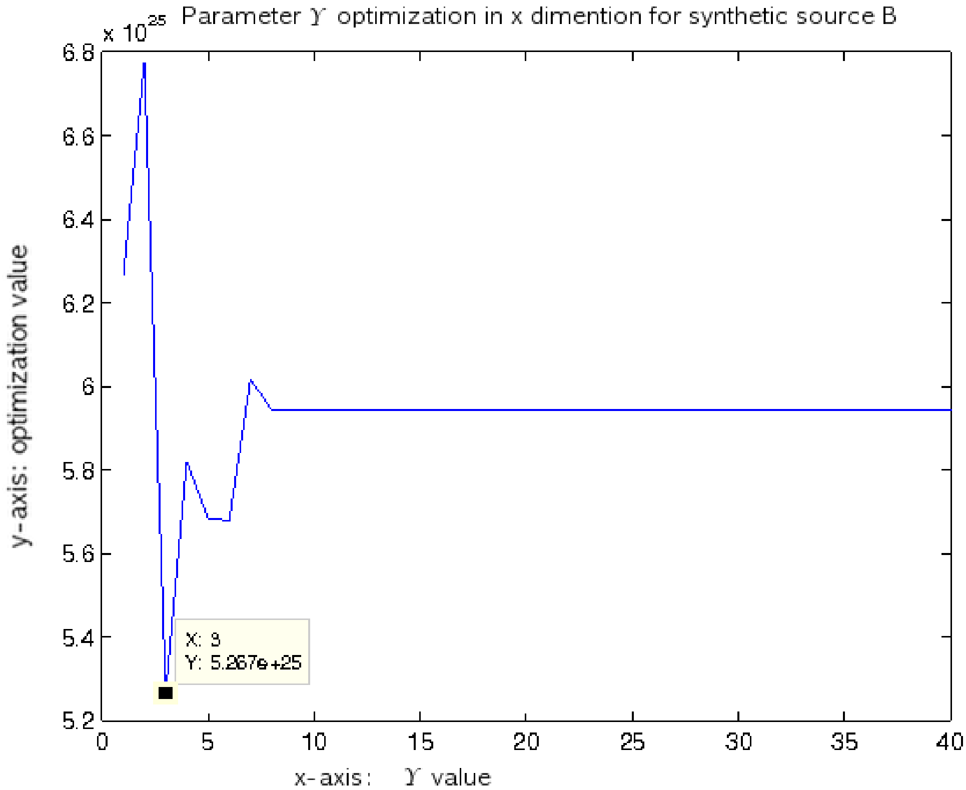

Figure 4.14: This figure shows the optimization results for γ in x dimensions for *synthetic source B*. With the optimization calculation explained in *Parameter optimization*, the minimum values are obtained numerically. The x-axis shows the numerical value of x, and y-axis shows the approximation of E. With y-axis reaches to minimum, the corresponding value in x-axis are the optimization of x.

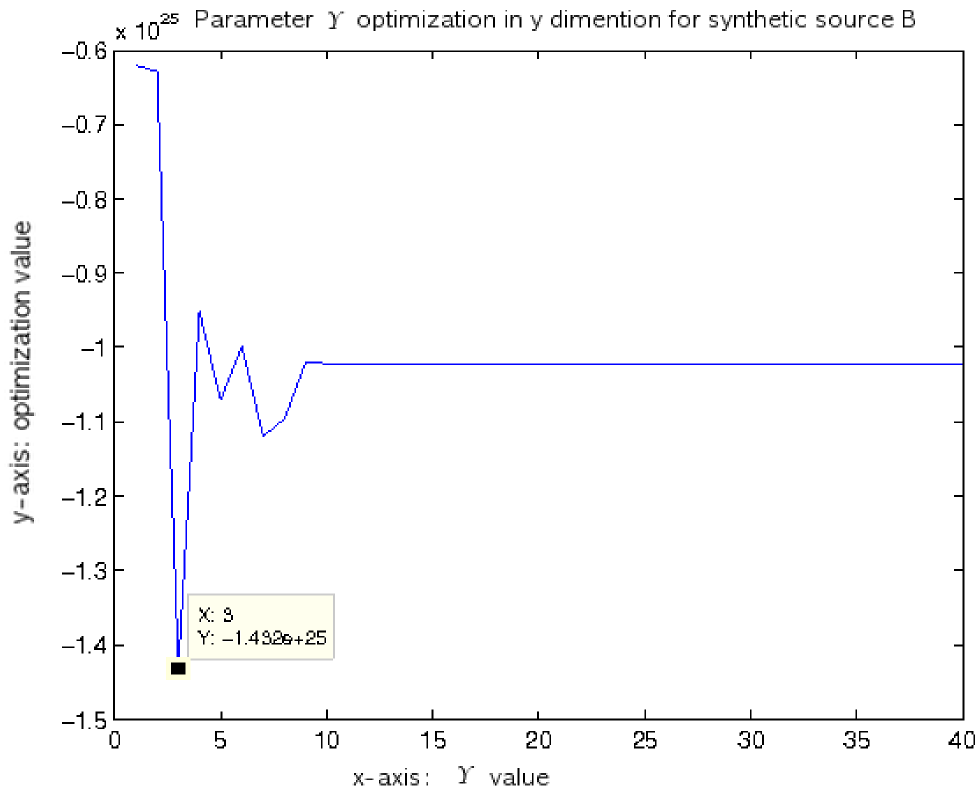

Figure 4.15: This figure shows the optimization results for γ in y dimensions for *synthetic source B*. With the optimization calculation explained in *Parameter optimization*, the minimum values are obtained numerically. The x-axis shows the numerical value of y, and y-axis shows the approximation of E. With y-axis reaches to minimum, the corresponding value in x-axis are the optimization of y.

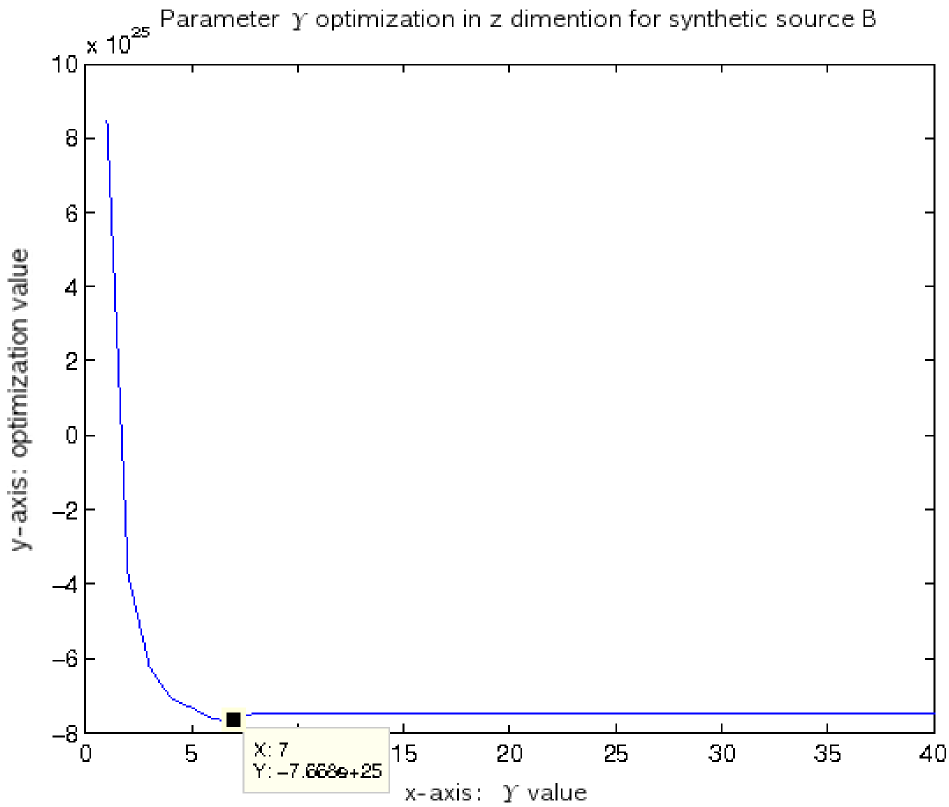

Figure 4.16: This figure shows the optimization results for γ in z dimensions for *synthetic source B*. With the optimization calculation explained in *Parameter optimization*, the minimum values are obtained numerically. The x-axis shows the numerical value of z, and y-axis shows the approximation of E. With y-axis reaches to minimum, the corresponding value in x-axis are the optimization of z.

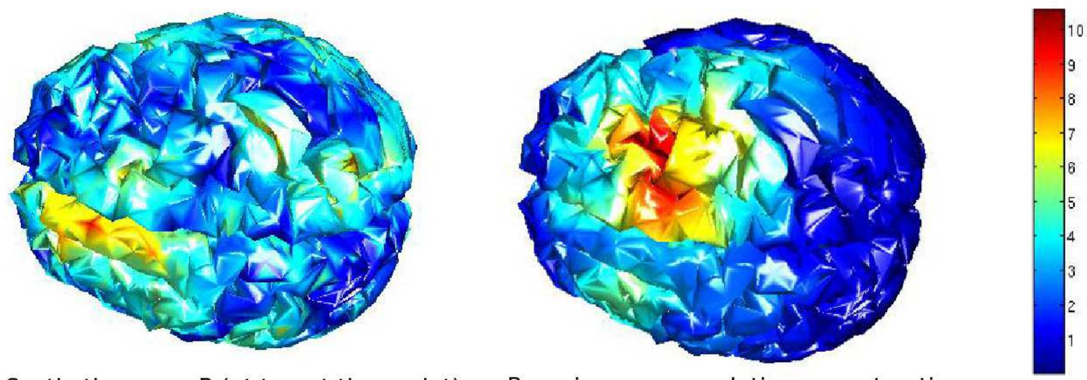

Figure 4.17: This figure shows the comparison of the simulated source pattern of *synthetic source B* on high-resolution mesh $\mathbf{M}^+$ (on the left) and the reconstruction result by the Bayesian super-resolution method (on the right) at the *target time point*.

220 , 245 , 270, 295, 320 , 345,370 ,395 , 420 , 445 , 470 , 495 , 520, 545 , 570 , 595, 620 , 645, 670 , 695, 720 , 745 for the reconstruction from the MEG data ($96 \times 248 \times 813$ matrix). Then, the basis function method is applied here for the source reconstruction on the low-resoluion mesh for B^{set} so that we get the 7800×1 matrix $\mathbf{J}^{real}$.

And then, based on the Laplacian of the mesh $\mathbf{M}$ and $\mathbf{M}^+$ extracted from *Matlab*, the Bayesian super-resolution method is applied for improving the spatial resolution of MEG source reconstruction and $\mathbf{j}^{real}$, the current source distributed on the high-resolution mesh $\mathbf{M}^+$, is generated. The real data process has the same calculation difficulty as the synthetic experiment: the calculation of large size matrices of algorithm factors leads to the much expensive computation. To solve this problem, a sparse matrix calculation is applied for the faster computation and the software *SPAI* (Grote and Hagemann) is used for calculating the approximation of the inverse of the large-size covariance matrix Σ, which are the same solutions applied in the synthetic experiment.

Fig 4.18 shows the reconstructions by the basis function method of low-resolution current source frames at 31 successive time points; Fig 4.19 and Fig 4.20 demonstrates the smoothing process of 3rd and 12th element of coefficient matrix $\mathbf{a}$ in successive 30 time points by Kalman filter (in 3-space, with components x, y and z); Fig 4.24 shows the comparison of reconstruction results by the basis function method and by the Bayesian super-resolution method . Both the color pattern and 2D signal pattern of the reconstruction result on the *target time point*. Also, Fig 4.25 demonstrates the wave patterns of the reconstructed current source $\mathbf{J}_{70}$ (at the time point 70 of 813) on $\mathbf{M}^+$ in 3-space.

Since it is impossible to have an absolute correct source location for the goodness evaluation of our method in the real MEG experiment, we refer the reconstruction results of the same trial by fMRI and cognition estimation based on the stimulus knowledge we have. According to Cindy C. Hagan's fMRI result for the same experiment (Hagan et al., 2009), the transient visual changes are occurs in the posterior superior temporal sulcus (STS) from(Hagan et al., 2009), which approximately match the result of the basis function method on the low-resolution mesh **M**, however, the reconstruction by the Bayesian super-resolution method shows the distortion of the current source location, refer to Fig 4.24.

4.7 Discussion

In terms of the reconstruction results of simulated experiment as well as applying to the real data , shown in Fig 4.10, Fig 4.17 and Fig 4.24, the estimated parameters we obtained do not lead to perfect reconstruction results. Here, the possible reason is provided for the incorrect reconstruction. The approximation of inverse of the large size matrix leads to the inaccurate reconstruction. In terms of the Eqn 4.29 and Eqn 4.30, covariance matrix Σ need to be calculated for the computation of estimated high-resolution source frame μ. This computation is to search for the inverse of large size matrix, which is too expensive to calculate precisely in practice. In this case, an approximation of the inverse of large-size matrix is applied in this step for the convenience to computation, i.e. software *SPAI* (Grote and Hagemann) is used to generate the inverse of the large size sparse matrix (*given a sparse matrix* A *the* SPAI *Algorithm computes a sparse approximate inverse* M *by minimizing* $\|AM - I\|$ *in the Frobenius norm. The approximate inverse is computed explicitly and can then be applied as a preconditioner to an iterative method.*). The difference produced by this approximation with the real inverse matrix may lead to the inaccurate result of the reconstruction.

Also, the main experimental configuration we use is : InterCore2(1,8GHz), Linux system(2.6.34.1)-32bit, matlab $7.9.0(R2009b) \times 32$ edition , RAM: 4GB. Since there

Figure 4.18: In terms of the MEG data of facial expression (showed as the $96 \times 248 \times 813$ matrix) ,this figure shows the reconstruction results on low-resolution mesh **M** by the basis function method at 3 of 96, and select the time point (in *ms*): 20, 45, 70, 95, 120, 145, 170, 195, 220, 245, 270, 295, 320 , 345, 370, 395, 420, 445, 470, 495, 520, 545, 570, 595, 620 , 645, 670, 695, 720, 745 for the reconstruction. First row shows the results on the time point (in *ms*): 20, 45, 70, 95, 120; second row shows the results on the time point (in *ms*): 145, 170, 195, 220, 245; the third row shows the results on the time point (in *ms*): 270, 295, 320, 345, 370; the fourth row shows the results on the time point (in *ms*): 395, 420, 445, 470, 495; the fifth row shows the results on the time point (in *ms*): 520, 545, 570, 595, 620; the sixth row shows the results on the time point (in *ms*): 645 , 670 , 695, 720 , 745.

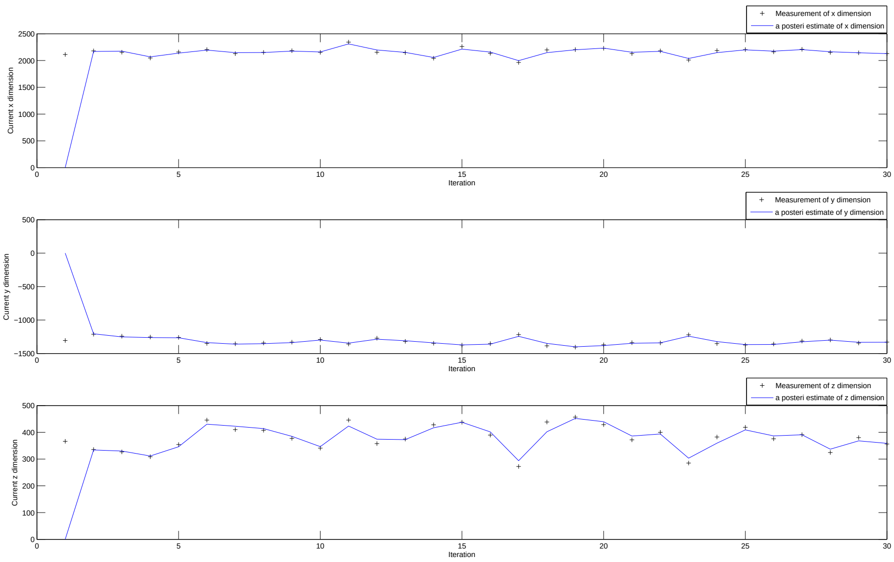

Figure 4.19: This figure shows the smoothing process of 3rd element of coefficient matrix $\mathbf{a}$ in successive 31 time points by Kalman filter (in 3-space, with components x, y and z) for the real MEG data of facial expression (showed as the $96 \times 248 \times 813$ matrix). This $\mathbf{a}$ is generated for mesh $\mathbf{M}$ by the basis function method over the time sequence to obtain a more accurate prediction of the current source. The '*' shows the original reconstructions of $\mathbf{a}$ by the basis function method; the blue line shows *the posterior estimate by Kalman filter estimation on each component of 3 dimensions.*

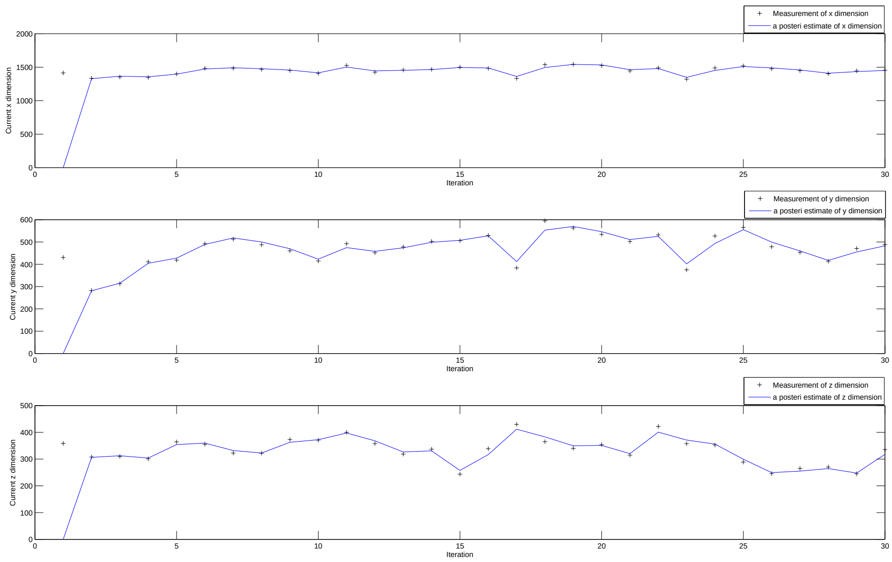

Figure 4.20: This figure shows the smoothing process of 12th element of coefficient matrix $\mathbf{a}$ in successive 31 time points by Kalman filter (in 3-space, with components x, y and z) for the real MEG data of facial expression (showed as the $96 \times 248 \times 813$ matrix). This $\mathbf{a}$ is generated for mesh $\mathbf{M}$ by the basis function method over the time sequence to obtain a more accurate prediction of the current source. The '*' shows the original reconstructions of $\mathbf{a}$ by the basis function method; the blue line shows *the posterior estimate by Kalman filter estimation on each component of 3 dimensions.*

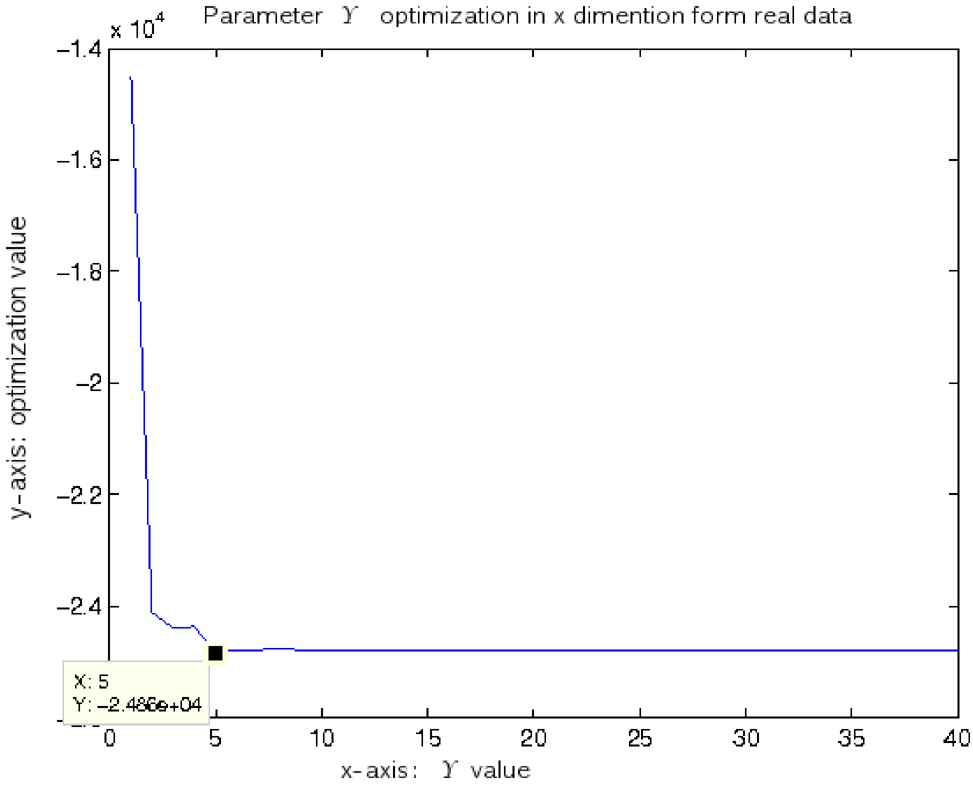

Figure 4.21: This figure shows the optimization results for γ in x dimensions for the real MEG data of facial expression (showed as the $96 \times 248 \times 813$ matrix). With the optimization calculation explained in *Parameter optimization*, the minimum values are obtained numerically. The x-axis shows the numerical value of x, and y-axis shows the approximation of E. With y-axis reaches to minimum, the corresponding value in x-axis are the optimization of x.

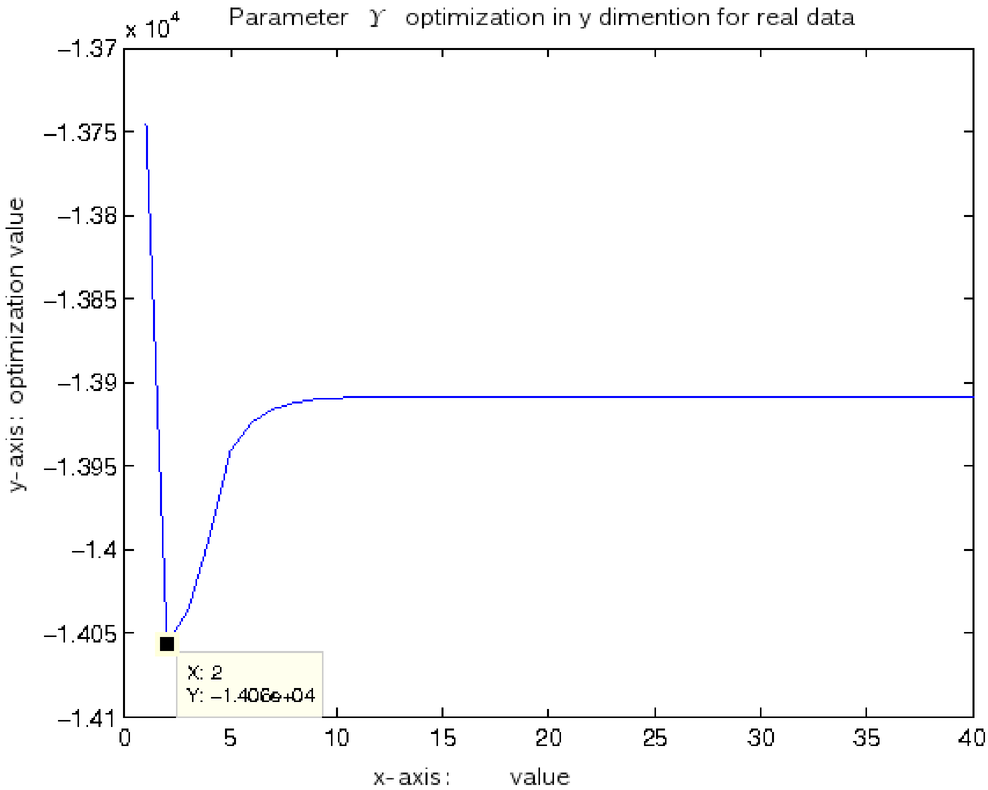

Figure 4.22: This figure shows the optimization results for γ in y dimensions for the real MEG data of facial expression (showed as the $96 \times 248 \times 813$ matrix). With the optimization calculation explained in *Parameter optimization*, the minimum values are obtained numerically. The x-axis shows the numerical value of y, and y-axis shows the approximation of E. With y-axis reaches to minimum, the corresponding value in x-axis are the optimization of y.

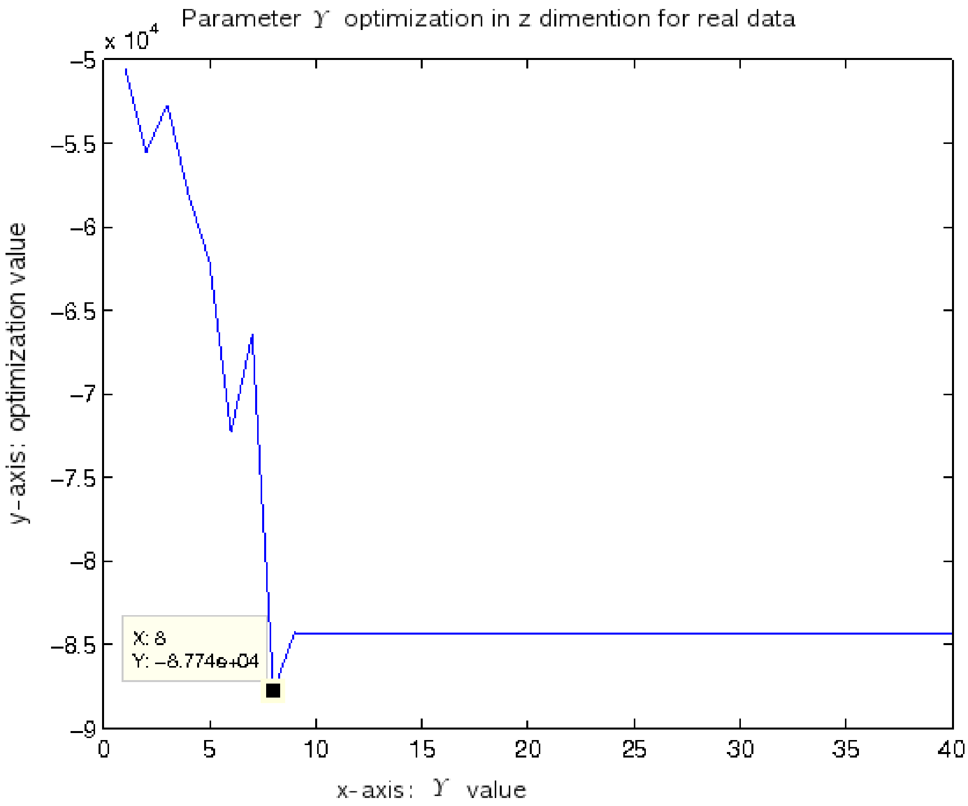

Figure 4.23: This figure shows the optimization results for γ in z dimensions for the real MEG data of facial expression (showed as the $96 \times 248 \times 813$ matrix). With the optimization calculation explained in *Parameter optimization*, the minimum values are obtained numerically. The x-axis shows the numerical value of z, and y-axis shows the approximation of E. With y-axis reaches to minimum, the corresponding value in x-axis are the optimization of z.

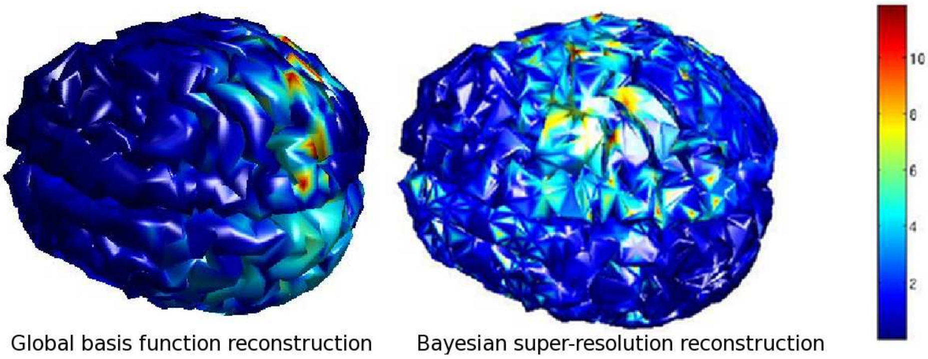

Figure 4.24: This figure demonstrate the comparison of reconstruction results at *target time point:* 70 *of* 813 by the basis function method on $\mathbf{M}$ (on the left) and by the Bayesian super-resolution method on $\mathbf{M}^+$ (on the right).

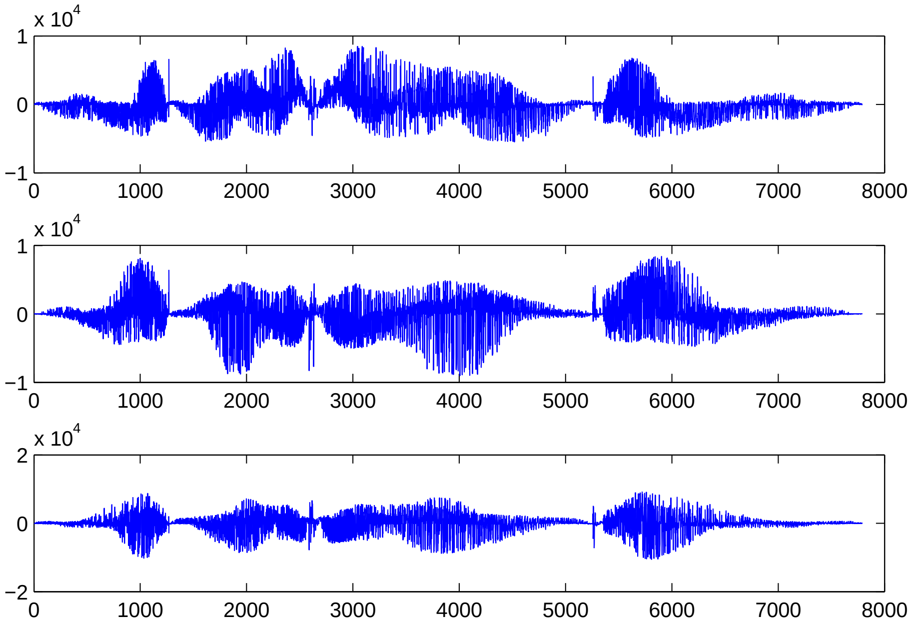

Figure 4.25: This figure demonstrate the wave pattern of the reconstructed current source $\mathbf{J}_{70}$ (at the time point 70 of 813) on $\mathbf{M}^+$ in 3-space. From the top to the bottom, the patterns show the wave pattern of $\mathbf{J}_{70}$ in x, y and z dimension, respectively. The x-axis indicates the index of vertex of $\mathbf{M}^+$, the y-axis indicates the amplitude of the current source.

are plenty of loop calculation associated with the large size matrix calculation (including the full matrix and sparse matrix calculation) , the implement of the algorithms are with huge time and space computation complexity which generates a critical restriction of the application of the method. In the real application, it is crucial to decrease the computation cost so that the algorithm can be applied more efficiently and realistically on the real MEG source reconstruction, e.g. we can upgrade the advanced configuration, as well as explore more reasonable format to store the variables(such as the state noise covariance matrix $\mathbf{Q}$, instead of storing the full matrices).

4.8 Conclusion

In summary, the main contribution of the algorithm designed in this Chapter is to build a new solution for improving the spatial-resolution of MEG source reconstruction at a single time point by introducing a classical method (Bayesian super-resolution method) from the pattern recognition theory. This approach is applied based on the MEG spatial reconstruction with basis function method which is elaborated in *Chapter 3* of the thesis. However, it could also be applied to other spatial reconstruction methods to improve the spatial-resolution.

As a competitive brain imaging technique, MEG shows superior temporal resolution (up to 1 ms). However, one of the weaknesses is that the spatial resolution is reduced. This method can be applied complemented by the spatial resolution of MEG source reconstruction using the time series of signals.

The mathematical framework of the method provides sound logic and an adequate description of the inverse problem of MEG. From the numerical experiment results of parameter estimation, it is explicit that the spatial resolution has effectively been improved.

However, instead of analysing the data of image, the method here is used for processing the problem of source distribution on the 3D cortical surface mesh that increase the computation complexity and inaccuracy of the reconstruction results immensely. This problem is reflected by the results generated from synthetic data as well as real MEG data.

Moreover, the parameter estimation and the optimization of high-resolution source distribution contain a number of large matrix calculations. Although some of them can be simplied by sparse matrix methods, there are many other factors that require large and full size matrices, which lead to expensive calculations and an increase in time and space complexity. Some effective software for matrix calculation can be used, such as *SPAI* (Grote and Hagemann) for the inverse problems of large matrix. This limitation also affects the widely application of the super-resolution algorithm.

Therefore, there are some possible extensions that can be achieved in further work. Firstly, with respect to the expensive computation mentioned above, it is still feasible to either upgrade the experimental configuration (the configuration currently is : InterCore2(1,8GHz), Linux system(2.6.34.1)-32bit, matlab $7.9.0(R2009b) \times 32$ edition , RAM: 4GB), or to develop the structure of matrix calculation mathematically to improve the efficiency of the application of this Bayesian super-resolution method.

Furthermore, the high-resolution mesh $\mathbf{M}^+$ we applied for the Bayesian super-resolution method is interpolated from the original mesh $\mathbf{M}$ and directly used for the reconstruction process. It will be beneficial if we can create $\mathbf{M}^+$ more accurately, which can better represent the cortical surface realistically as a part if the future work. Moreover, this smoothing method needs to be carefully designed to decrease the distortion of the information of cortical surface to the minimum.

Chapter 5

MEG image estimation via Kalman smoother

5.1 Brief introduction

In the last two Chapters, we have tackled the MEG source reconstruction problem and improved the spatial resolution of the reconstruction based on the MEG measurement using the basis function method and the Bayesian super-resolution method. In this Chapter we will use the Kalman smoother to provide a direct reconstruction.

Assuming the MEG system as a dynamic system, the Kalman filter is applied to correct the original frames with low-resolutions in a particular time sequence in *Chapter 4: Spatial Improvement of MEG source reconstruction with Bayesian Super-resolution*. As a classical tool for smoothing the state of a dynamic system, the Kalman smoother can be applied to the MEG study for improving individual state estimation in the temporal field, using the data from other time frames.

Since the Kalman filter and the Kalman smoother both are both able to produce the estimation of the state of a dynamic system, it is feasible to apply the Kalman smoothing theory into the estimation of event-related dynamics in brain imaging. M.P. Tarvainen and his colleagues tried to solve the estimation of Nonstationary EEG on event-Related Synchronization (ERS) with a Kalman smoother approach (Tarvainen et al., 2004). Also, the Kalman filter and smoother have been successfully used to perform the estimation with high dimensionality as well as to solve the inverse problem on EEG-fMRI fusion of paradigm-free activity(Deneux and Faugeras, 2010) .

In this Chapter, we present a Kalman smoother approach, utilizing a fixed-interval smoother, to estimate a high resolution MEG current source in the temporal field. We use the basis function source model (Chapter 3) which is integrated with the Kalman filter. However, this estimation still needs smoothing to improve the accuracy in the temporal field. The mathematical framework of Kalman smoother for measuring magnetic field and conditions of dynamic system is developed from the last two Chapters of (Welling) (Welch and Bishop, 2006). Then, the EM algorithm can be used to estimate the parameter set. It is worthy to note that this approach makes it possible to estimate the hidden *high-resolution* image directly from the coil sensors of MEG.

In the later part of this Chapter, the dynamic system is built based on the integration of the basis function source model and the Kalman filter. The MEG system is described as a dynamic system, which provides the prior conditions for Kalman smoothing. Then, the Kalman smoother will be introduced for estimating the current source frames with high-resolution in the temporal field. Next, the parameter set is estimated by applying the EM algorithm. Finally, the current source reconstruction experiments based on the Kalman smoother method were conducted again using synthetic data and real MEG data .

5.2 Noisy linear dynamic system

5.2.1 Noisy linear dynamic model

We have a strong assumption of the cortical distributed current source model for MEG inverse problem that the current source are embedding on the cortical surface and oriented tangentially to it. The magnetic field generated by the current sources which are tangential to the cortical surface are decayed. Thus, the MEG measurement from the sensor set at single time point can be described as a noisy linear dynamic system with the current source, showed as Eqn 5.1 .

$$\mathbf{B}(t) = \mathbf{L}\,\mathbf{J}(t) + \mathbf{n}(t) \tag{5.1}$$

where $\mathbf{B}(t)$ indicates the measurement of magnetic field on time point t; $\mathbf{L}$ is the leadfield, $\mathbf{J}_(t)$ represents the current source at time point t. $\mathbf{n}(t)$ is the added noise in the time course assumed as zero-mean $(E(\mathbf{n}) = 0)$(Olivier et al., 2001).

Also, as MEG has a high temporal-resolution(up to 1 ms), we assume that the current sources are linked between frames and can be modelled as a dynamic system. Therefore, the Kalman filter and Kalman smoother are considered as the conventional solution for estimating as well as smoothing the state of this dynamic system in the temporal field.

Before we start the design of the algorithm, the following priors must be set firstly as the conditional assumptions of dynamic system for later processing of Kalman smoothing. The measured signals are modeled as an output of a parametric model with time-varying parameters (Tarvainen et al., 2004).

5.2.2 Prior setting of dynamic process

From the study of Kalman filter in *Chapter 2*, the real-time MEG state and measurement can be described as the following:

$$\mathbf{x}_t = \mathrm{A}\mathbf{x}_{t-1} + \omega \tag{5.2}$$

$$\mathbf{z}_t = \mathrm{B}\mathbf{x}_t + \upsilon \tag{5.3}$$

Here $\mathbf{x}_t$ is the the hidden state and $\mathbf{z}_t$ is the observation. $\mathbf{A}$, and $\mathbf{B}$ are the relevant coefficient matrices for $\mathbf{x}_{t-1}$ and $\mathbf{x}_t$, respectively. ω is the state noise , and υ is the observation noise. Both of them are assumed as the zero-mean Gaussian distribution:

$$\omega \sim \mathcal{N}(0, \mathbf{Q}) \tag{5.4}$$

$$\upsilon \sim \mathcal{N}(0, \mathbf{R}) \tag{5.5}$$

where $\mathbf{Q}$ is the $2n \times 2n$ covariance matrix of the state noise. n is the number of vertices of cortical mesh. The structure of $\mathbf{Q}$ is shown in Eqn 5.10; and $\mathbf{R}$ is the $m \times m$ covariance matrix of the observation noise. m is the number of sensors.

In the context of our problem, we would like to find the information of current $\mathbf{x}_t$ which includes not only current $\mathbf{j}_x$ but also the *rate-of-change* $\mathbf{v}_t$ of the current in the time course:

$$\mathbf{x_t} = \begin{pmatrix} \mathbf{j}_t \\ \mathbf{v}_t \end{pmatrix} = \begin{pmatrix} \mathbf{j}_{xt} \\ \mathbf{j}_{yt} \\ \mathbf{j}_{zt} \\ \mathbf{v}_{xt} \\ \mathbf{v}_{yt} \\ \mathbf{v}_{zt} \end{pmatrix} \tag{5.6}$$

here, $\mathbf{j}_t = [\mathbf{j}_{xt} \;\; \mathbf{j}_{yt} \;\; \mathbf{j}_{zt}]^T$ represents *currents embedding on the cortical surface at time point t on 3 dimensions, respectively* ; $\mathbf{v}_t = [\mathbf{v}_{xt}, \mathbf{v}_{yt}, \mathbf{v}_{zt}]^T$ represents *currents rate of change at time point t on 3 dimensions, respectively*. Of course, we observe the coil response which represents as following equation:

$$\mathbf{z}_t = \mathbf{b}_t \tag{5.7}$$

The state transition matrix $\mathbf{A}$ just gives us $\mathbf{j}_t = \mathbf{j}_{t-1} + \mathbf{v}_{t-1}$ and $\mathbf{v}_t = \mathbf{v}_{t-1}$, so

$$\mathbf{A} = \begin{pmatrix} \mathbf{I} & \mathbf{I} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \tag{5.8}$$

And the observation matrix $\mathbf{b}_t$ gives the coil responses from a particular current distribution at time point t. In the simplest form this is the leadfield matrix operation on $\mathbf{j}_t$,so

$$\mathbf{b}_t = \mathbf{L}\mathbf{j}_t + \upsilon \tag{5.9}$$

There are two types of noise present which are the covariance matrix Q of the state noise and the covariance matrix **R** of the observation noise. We assume that Q is smoothed over our mesh and can be modelled as with respect to the element matrix $\mathbf{E}_\gamma$ in Eqn 4.21:

$$\mathbf{Q} = \beta \begin{pmatrix} \mathbf{E}_\alpha & & & & & 0 \\ & \mathbf{E}_\alpha & & & & \\ & & \mathbf{E}_\alpha & & & \\ & & & \mathbf{E}_\alpha & & \\ & & & & \mathbf{E}_\alpha & \\ 0 & & & & & \mathbf{E}_\alpha \end{pmatrix} \tag{5.10}$$

where

$$\mathbf{E}_\alpha = \begin{pmatrix} \exp(-L^+_{1,1}\alpha) & \cdots & \cdots & \exp(-L^+_{1,N}\alpha) \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \exp(-L^+_{N,1}\alpha) & \cdots & \cdots & \exp(-L^+_{M,M}\alpha) \end{pmatrix} \tag{5.11}$$

In terms of Eqn 5.10, there are two parameters α and β which directly govern the trend of $\mathbf{Q}$. $\mathbf{Q}$ is the same noise covariance matrix as we used in the super-resolution method and based on the mesh heat kernel (refer to Eqn 4.17). Therefore, β is the Gaussian constant for the covariance matrix $\mathbf{Q}$ which has the linear relationship with $\mathbf{Q}$. The larger value of β leads to the larger values of the elements of $\mathbf{Q}$; and vice versa. The α is the constant prior of the heat kernel; the heat kernel models the local interactions between neighbouring elements of the mesh. The larger the value of α, the larger the scale of correlations on the mesh. Therefore, the problem of searching for unknown parameters $\mathbf{R}$ and $\mathbf{Q}$ is transformed into searching for $\mathbf{R}$, α and β in the later part of this Chapter.

The basis set algorithm of MEG source reconstruction(explained in Chapter 3) is used for re-writing the dynamic state in following way:

$$\mathbf{j}_x = \mathbf{\Phi a}_x,\ \mathbf{j}_y = \mathbf{\Phi a}_y,\ \mathbf{j}_z = \mathbf{\Phi a}_z \tag{5.12}$$

We can do the same for our rates-of-change:

$$\mathbf{v}_x = \mathbf{\Phi c}_x,\ \mathbf{v}_y = \mathbf{\Phi c}_y,\ \mathbf{v}_z = \mathbf{\Phi c}_z \tag{5.13}$$

As this process is supposed to be a linear transform, we can estimate the parameters directly:

$$\mathbf{x}_t^{new} = \begin{pmatrix} \mathbf{a}_{xt} \\ \mathbf{a}_{yt} \\ \mathbf{a}_{zt} \\ \mathbf{c}_{xt} \\ \mathbf{c}_{yt} \\ \mathbf{c}_{zt} \end{pmatrix} \tag{5.14}$$

with A unchanged. However, the noise will change in this new representation of the problem where each component $\exp(-\mathbf{L}^{+}\alpha)$ will be transformed into:

$$\mathbf{\Phi}^{T}\exp(-\mathbf{L}^{+}\alpha)\mathbf{\Phi} = \exp(-\mathbf{\Lambda}\alpha) \tag{5.15}$$

where $\mathbf{\Lambda}$ is a diagonal matrix of the eigenvalues of the high-resolution cortical mesh $\mathbf{M}^{+}$:

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_{1,1} & \cdots & \cdots & 0 \\ \vdots & \lambda_{2,2} & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & \lambda_{M,M} \end{pmatrix} \tag{5.16}$$

With respect to the equations above, we can write the noise $\mathbf{Q}$ as follows:

$$\begin{aligned}\mathbf{Q} &= \beta \begin{pmatrix} \exp(-\mathbf{\Lambda}\alpha) & & & & & 0 \\ & \exp(-\mathbf{\Lambda}\alpha) & & & & \\ & & \exp(-\mathbf{\Lambda}\alpha) & & & \\ & & & \exp(-\mathbf{\Lambda}\alpha) & & \\ & & & & \exp(-\mathbf{\Lambda}\alpha) & \\ 0 & & & & & \exp(-\mathbf{\Lambda}\alpha) \end{pmatrix} \\ &= \beta\exp(-\mathbf{\Lambda}_f\alpha) \end{aligned} \tag{5.17}$$

where

$$\mathbf{\Lambda}_f = \begin{pmatrix} \mathbf{\Lambda} & & & & & 0 \\ & \mathbf{\Lambda} & & & & \\ & & \mathbf{\Lambda} & & & \\ & & & \mathbf{\Lambda} & & \\ & & & & \mathbf{\Lambda} & \\ 0 & & & & & \mathbf{\Lambda} \end{pmatrix} \tag{5.18}$$

and

$$\exp(-\mathbf{\Lambda}\alpha) = \begin{pmatrix} \exp(-\lambda_{1,1}\alpha) & \cdots & \cdots & 0 \\ \vdots & \exp(-\lambda_{2,2}\alpha) & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & \exp(-\lambda_{M,M}\alpha) \end{pmatrix} \tag{5.19}$$

Finally, according to the basis set algorithm in Chapter 3, the observation model can be represented as follows:

$$\mathbf{b}_t = \mathbf{L} \begin{pmatrix} \mathbf{\Phi} & & \mathbf{0} \\ & \mathbf{\Phi} & \\ \mathbf{0} & & \mathbf{\Phi} \end{pmatrix} \begin{pmatrix} \mathbf{a}_x \\ \mathbf{a}_y \\ \mathbf{a}_z \end{pmatrix} + \upsilon \tag{5.20}$$

It is worthy noting that the number of basis functions here can be considerably larger than before. The inaccuracy generated by the over-determined can be corrected by applying Kalman smoother for the estimation later.

5.3 Application of Kalman smoother

5.3.1 Brief introduction of Kalman smoother

As we discussed in *Chapter 1 : Application of Kalman smoother*, the Kalman smoother can be used to estimate the hidden state of a Gaussian process. Based on the Markov property of Kalman filter, the state depends on the previous state but not any others. However, for the estimation of the state and the uncertainty(covariance) at a specific time point t, it is feasible to obtain the solution from only the status on previous one time point $t-1$ as well as the noisy observation $\mathbf{x}^{\tau} = \mathbf{z}_1, \cdots, \mathbf{z}_{\tau}$ for the specic time point t. It is notable that the difference between t and τ generally provides the process with variable uses. For instance, if τ is equal to the current time point t, the process is called *filtering*; if τ is smaller than t, the process is called *predicting*; and if τ is larger than t, the process is called *smoothing*. In other words, if the measured data is not processed in real time or if a small lag in the processing is allowed, the future observations can also be used in the state estimation. Since more measurement in the time sequence are applied for processing in this case, it is reasonable to expect the estimates to be more accurate. This is called a smoother (Kalman, 1960), (Jazwinski, 1970,), (Deneux and Faugeras, 2010).

With the calculation and inference from (Welling), the Kalman smoother equations are obtained as Eqn 5.21 , Eqn5.22 and Eqn 5.23 :

$$\hat{\mathbf{x}}_t^{\tau} = \hat{\mathbf{x}}_t^{t} + \mathbf{J}_t(\hat{\mathbf{x}}_{t+1}^{\tau} - \hat{\mathbf{x}}_{t+1}^{t}) \tag{5.21}$$

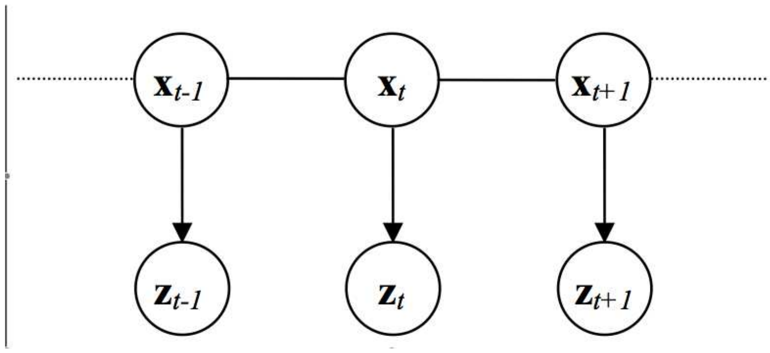

Figure 5.1: This figure shows the relationship between the observed states $\mathbf{z}$ and source state $\mathbf{x}$ in the time course. The current source $\mathbf{x}$ is assumed to be the hidden state which depends both on the state at time $t-1$ and $t+1$ and the observed state $\mathbf{z}$ depends on the hidden state $\mathbf{x}$ only. Since there is on state after it, $\mathbf{x}_\tau$ is assumed to be the final state which only depends on the state $\mathbf{x}_{\tau-1}$ before it.

$$\mathbf{J}_t = \mathbf{P}_t^t \mathbf{A}^T [\mathbf{P}_{t+1}^t]^{-1} \tag{5.22}$$

$$\mathbf{P}_t^\tau = \mathbf{P}_t^t + \mathbf{J}_t(\mathbf{P}_{t+1}^\tau - \mathbf{P}_{t+1}^t)\mathbf{J}_t^T \tag{5.23}$$

The way to apply the Kalman smoother equation is separated into two steps. Firstly, with the full set of term measurements, the Kalman filter is applied forward from the state at initial time point till the state at time t is reached (where $t < \tau$). Then, the process is moved backward by applying the Kalman smoother equations until state at the time t is estimated. Since all the state factors, such as $\hat{\mathbf{x}}_{\mathbf{t}+1}^{t+1}$, $\hat{\mathbf{x}}_{\mathbf{t}+1}^{t}$, $\mathbf{P}_{t+1}^{t+1}$ and $\mathbf{P}_{t+1}^{t}$, $t = 1 \cdots \tau$ are stored in the former step, it is easier for Kalman smoother equations to apply them directly in the later step (Welling).

5.3.2 Application of Kalman smoother

The process can be represented as graphical model showed in Fig 5.1. The unobserved state $\mathbf{x}_t$ depends both on the state at time $t-1$ and $t+1$ and the observed state depends on the hidden state only.

Here, we consider the expression of the states with final state $\mathbf{x}_\tau$ and earlier state $\mathbf{x}_t$, respectively.

Firstly, it is noticeable that the final state $\mathbf{x}_\tau$ only depends on the state before it, $\tau - 1$, since there is no state after it, so

$$p(\mathbf{x}_\tau|\{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_\tau\}) = p(\mathbf{z}_\tau|\mathbf{x}_\tau)p(\mathbf{x}_\tau|\{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{\tau-1}\})/p(\{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{\tau-1}\}) \tag{5.24}$$

This is identical to the Kalman filter, so we can find $\mathbf{x}_\tau$ using the normal Kalman filter equations:

$$\mathbf{x}_t^- = \mathbf{A}\mathbf{x}_{t-1} \tag{5.25}$$

$$\mathbf{P}_t^- = \mathbf{A}\mathbf{P}_{t-1}\mathbf{A}^T + \mathbf{Q} \tag{5.26}$$

$$\mathbf{K}_t = \mathbf{P}_t^-\mathbf{B}^T(\mathbf{B}\mathbf{P}_t^-\mathbf{B}^T + \mathbf{R})^{-1} \tag{5.27}$$

$$\mathbf{x}_t = \mathbf{x}_t^- + \mathbf{K}_t(\mathbf{z}_t - \mathbf{B}\mathbf{x}_t^-) \tag{5.28}$$

$$\mathbf{P}_t = (\mathbf{I} - \mathbf{K}_t\mathbf{B})\mathbf{P}_t^- \tag{5.29}$$

Then, for a time t earlier in the sequence, we can write

$$\begin{aligned} p(x_t, x_{t+1}, \{z_1, \ldots, z_\tau\}) &= p(\{z_{t+1,\ldots,z_\tau}\}|x_t, x_{t+1}, \{z_1, \ldots, z_t\}) \\ &\quad \times p(\{z_1, \ldots, z_t\}, x_{t+1}|x_t)p(x_t|\{z_1, \ldots, z_t\}) \\ &= p(\{z_1, \ldots, z_\tau\}, x_{t+1}|x_t, \{z_1, \ldots, z_t\})p(x_t|\{z_1, \ldots, z_t\}) \\ &= p(\{z_{t+1,\ldots,z_\tau}\}, x_{t+1}|x_t)p(x_t|\{z_1, \ldots, z_t\}) \\ &= p(\{z_{t+1}, \ldots, z_\tau\}|x_{t+1})p(x_{t+1}|x_t)p(x_t|\{z_1, \ldots, z_t\}) \end{aligned} \tag{5.30}$$

According to Eqn 5.30, it is apparent that the distribution of $\mathbf{x}_t$ depends on three components : $p(\{z_{t+1}, \ldots, z_\tau\}|x_{t+1})$; $p(x_{t+1}|x_t)$ and $p(x_t|\{z_1, \ldots, z_t\})$. The first component $p(\{z_{t+1}, \ldots, z_\tau\}|x_{t+1})$ comes from the measurement; $p(x_t|\{z_1, \ldots, z_t\})$ can be found from the Kalman filter; and $p(x_{t+1}|x_t)$ which we can find from the state transition $\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \omega$. We now use $*$ to denote the best estimate. We have $\mathbf{x}^*_\tau = \mathbf{x}_\tau$.

$$\mathbf{J}_t = \mathbf{P}_t\mathbf{A}^T[\mathbf{P}^-_t]^{-1} \tag{5.31}$$

$$\mathbf{x}^*_t = \mathbf{x}_t + \mathbf{J}_t(\mathbf{x}^*_{t+1} - \mathbf{x}^-_{t+1}) \tag{5.32}$$

$$\mathbf{P}^*_t = \mathbf{P}_t + \mathbf{J}_t(\mathbf{P}^*_{t+1} - \mathbf{P}^-_{t+1})\mathbf{J}^T_{t-1} \tag{5.33}$$

5.4 Parameter estimation

Following the work presented previously (Shumway and Stoffer, 1982) , (Shumway and Stoffer, 1992), (Welling), (Ghahramani and Hinton, 1996), the EM algorithm is applied to find the parameters of the method. For the EM algorithm, we consider the states $\mathbf{x}_t$ as hidden variables, while $\{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_\tau\}$ are the observation. The joint probability of the complete data is showed as:

$$p(\{\mathbf{z}\}^\tau_1\{\mathbf{x}\}^\tau_1) \equiv p(\{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_\tau\}, \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_\tau\}) = p(\mathbf{x}_1)\prod_{t=2}^{\tau} p(\mathbf{x}_t|\mathbf{x}_{t-1})\prod_{t=1}^{\tau} p(\mathbf{z}_t|\mathbf{x}_t) \tag{5.34}$$

and we know that

$$p(\mathbf{z}_t|\mathbf{x}_t) = \mathcal{N}(\mathbf{B}\mathbf{x}_t, \mathbf{R}) \tag{5.35}$$

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{A}\mathbf{x}_{t-1}, \mathbf{Q}) \tag{5.36}$$

$$p(\mathbf{x}_1) = \mathcal{N}(\mu, \mathbf{\Sigma}) \tag{5.37}$$

We proceed to estimate the parameters $\{\mathbf{R}, \mathbf{Q}\}$ by determining the log likelihood of the expectation of the joint probability density function(pdf) over the posterior density '$\mathcal{L}$'

$$\begin{aligned}\mathcal{L} &= E[\log(p(\{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_\tau\}, \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_\tau\}))] \\ &= \log\det\mathbf{\Sigma} + (\tau-1)\log\det\mathbf{Q} + \tau\log\det\mathbf{R} + [(\mathbf{x}_1-\mu)^T\Sigma^{-1}(\mathbf{x}_1-\mu) \\ &\quad + \sum_{t=2}^{\tau}(\mathbf{x}_t - \mathbf{A}\mathbf{x}_{t-1})^T\mathbf{Q}^{-1}(\mathbf{x}_t - \mathbf{A}\mathbf{x}_{t-1}) + \sum_{t=1}^{\tau}(\mathbf{z}_t - \mathbf{B}\mathbf{x}_t)^T\mathbf{R}^{-1}(\mathbf{z}_t - \mathbf{B}\mathbf{x}_t)]\end{aligned} \tag{5.38}$$

We now describe the process of parameters estimation by an *E-step* and *M-step*, respectively. First we find the log-likelihood by computing the expectation in Eqn 5.38, and then maximize the log-likelihood to find the best parameters.

- **The E-step**

  Since the probability $p(\mathbf{x}_t|\{\mathbf{z}\}_1^\tau)$ is assumed to be Gaussian, we are using $\mathbf{x}_t^\tau$ to denote the state estimate $E[\mathbf{x}_t|\{\mathbf{z}\}_1^\tau]$ that depends on the *past and future observations* (for the Kalman smoother) , and $\mathbf{P}_t^\tau$ to denote covariance estimate $E(\tilde{\mathbf{x}}_t^\tau\tilde{\mathbf{x}}_t^\tau|\{\mathbf{z}\}_1^\tau)$ where $\tilde{\mathbf{x}}_t$ represents the state prediction error between the state and its estimate. Then, the objective function Eqn 5.38 contains a number of terms need to be calculated in *E-step* (Shumway and Stoffer, 1982) , (Shumway and Stoffer, 1992), (Ghahramani and Hinton, 1996):

$$E[\mathbf{x}_t|\{\mathbf{z}\}_1^\tau] = \mathbf{x}_t^\tau \equiv \mathbf{x}_t^* \qquad t = 1, \cdots, \tau \tag{5.39}$$

$$E(\tilde{\mathbf{x}}_t^\tau\tilde{\mathbf{x}}_t^\tau|\{\mathbf{z}\}_1^\tau) = \mathbf{P}_t^\tau \equiv \mathbf{P}_t^* \qquad t = 1, \cdots, \tau \tag{5.40}$$

$$E[\mathbf{x}_t\mathbf{x}_t|\{\mathbf{z}\}_1^\tau] = \mathbf{P}_t^\tau + \mathbf{x}_t^\tau\mathbf{x}_t^\tau \equiv \mathbf{M}_{t,t} \qquad t = 1, \cdots, \tau \tag{5.41}$$

$$E[\mathbf{x}_t\mathbf{x}_{t-1}|\{\mathbf{z}\}_1^\tau] = \mathbf{P}_{t,t-1}^\tau + \mathbf{x}_t^\tau\mathbf{x}_{t-1}^\tau \equiv \mathbf{M}_{t,t-1} \qquad t = 2, \cdots, \tau \tag{5.42}$$

  We can obtain the following Kalman filter forward recursions (Ghahramani and Hinton, 1996):

$$\mathbf{x}_t^{t-1} = \mathbf{A}\mathbf{x}_{t-1}^{t-1} \tag{5.43}$$

$$\mathbf{P}_t^{t-1} = \mathbf{A}\mathbf{P}_{t-1}^{t-1}\mathbf{A}^T + \mathbf{Q} \tag{5.44}$$

$$\mathbf{K}_t = \mathbf{P}_t^{t-1}\mathbf{B}^T(\mathbf{B}\mathbf{P}_t^{t-1}\mathbf{B} + \mathbf{R})^{-1} \tag{5.45}$$

$$\mathbf{x}_t^t = \mathbf{x}_t^{t-1} + \mathbf{K}_t(\mathbf{z}_t - \mathbf{B}\mathbf{x}_t^{t-1}) \tag{5.46}$$

$$\begin{aligned}\mathbf{P}_t^t &= \mathbf{P}_t^{t-1} - \mathbf{K}_t\mathbf{B}\mathbf{P}_t^{t-1} \\ &= (\mathbf{I} - \mathbf{K}_t\mathbf{B})\mathbf{P}_t^{t-1}\end{aligned} \tag{5.47}$$

Following (Shumway and Stoffer, 1982), the items in Eqn 5.39 Eqn 5.40 and Eqn 5.41 can be calculated by a set of backward recursions of Kalman smoother:

$$\mathbf{J}_{t-1} = \mathbf{P}_{t-1}^{t-1}\mathbf{A}(\mathbf{P}_t^{t-1})^{-1} \tag{5.48}$$

$$\mathbf{x}_{t-1}^{\tau} = \mathbf{x}_{t-1}^{t-1} + \mathbf{J}_{t-1}(\mathbf{x}_t^{\tau} - \mathbf{A}\mathbf{x}_{t-1}^{t-1}) \tag{5.49}$$

$$\mathbf{P}_{t-1}^{\tau} = \mathbf{P}_{t-1}^{t-1} + \mathbf{J}_{t-1}(\mathbf{P}_t^{\tau} - \mathbf{P}_t^{t-1})\mathbf{J}_{t-1}^T \tag{5.50}$$

The quantity in Eqn 5.42 is so-called the *lag-one covariance smoother*, which is given by the following recursion (Shumway and Stoffer, 1982), (Welling):

$$\mathbf{P}_{t-1,t-2}^{\tau} = \mathbf{P}_{t-1}^{t-1}\mathbf{J}_{t-2}^T + \mathbf{J}_{t-1}(\mathbf{P}_{t,t-1}^{\tau} - \mathbf{A}\mathbf{P}_{t-1}^{t-1})\mathbf{J}_{t-2}^T \tag{5.51}$$

which is initialized by:

$$\mathbf{P}_{\tau,\tau-1}^{\tau} = (\mathbf{I} - \mathbf{K}_{\tau}\mathbf{B})\mathbf{A}\mathbf{P}_{\tau-1}^{\tau-1} \tag{5.52}$$

Also, there is the relation between $\mathbf{M}_{t-1,t}$ and $\mathbf{M}_{t,t-1}$ showed as follows (Shumway and Stoffer, 1982), (Welling):

$$\mathbf{M}_{t-1,t} = \mathbf{M}_{t,t-1}^T \tag{5.53}$$

Moreover, when we observe the log-likelihood function Eqn 5.38, it is notable that $\mathcal{LL}$ contains the typical term $\mathbf{x}_t\mathbf{Q}^{-1}\mathbf{x}_t$ which can be replaced as follows:

$$\begin{aligned} E[\mathbf{x}_t\mathbf{Q}^{-1}\mathbf{x}_t] &= E[\sum_{ij} Q_{ij}^{-1} x_{ti}x_{tj}] \\ &= \sum_{ij} Q_{ij}^{-1} E[x_{ti}x_{tj}] \\ &= \mathrm{Tr}(\mathbf{Q}^{-1}\mathbf{M}_{t,t}) \end{aligned} \tag{5.54}$$

So, taking the expectation value (and omitting terms without $\mathbf{Q}$ and $\mathbf{R}$ which do not interest us), we get:

$$\begin{aligned} \mathcal{L} \;=\; & (\tau-1)\log\det\mathbf{Q} + \tau\log\det\mathbf{R} \\ & + \sum_{t=2}^{\tau}[\mathrm{Tr}(\mathbf{Q}^{-1}\mathbf{M}_{t,t}) + \mathrm{Tr}(\mathbf{A}^T\mathbf{Q}^{-1}\mathbf{A}\mathbf{M}_{t-1,t-1}) - \mathrm{Tr}(\mathbf{A}^T\mathbf{Q}^{-1}\mathbf{M}_{t-1,t}) \\ & - \mathrm{Tr}(\mathbf{Q}^{-1}\mathbf{A}\mathbf{M}_{t,t-1})] \\ & + \sum_{t=1}^{\tau}[\mathrm{Tr}(\mathbf{B}^T\mathbf{R}^{-1}\mathbf{B}\mathbf{M}_{t,t}) + \mathrm{Tr}(\mathbf{R}^{-1}\mathbf{z}_t\mathbf{z}_t^T) - \mathrm{Tr}(\mathbf{B}^T\mathbf{R}^{-1}\mathbf{x}_t^*\mathbf{z}_t^T) \\ & - \mathrm{Tr}(\mathbf{R}^{-1}\mathbf{B}\mathbf{z}_t\mathbf{x}_t^{*T})] \end{aligned} \tag{5.55}$$

- **The M-step**

  The parameters $\{\mathbf{R},\mathbf{Q}\}$ are estimated in the M-step by taking the corresponding partial derivative of the log-likelikhood function $\mathcal{L}$ which is equivalent to zero for the optimal value. All the matrix calculation in the following are referred to the *Appendix A*.

  Firstly, for finding the covariance matrix of the observation noise υ, we have:

$$\frac{d\mathcal{L}}{d\mathbf{R}} = -\tau\mathbf{R} + \sum_{t=1}^{\tau}\mathbf{B}\mathbf{M}_{t,t}^T\mathbf{B}^T + \mathbf{z}_t\mathbf{z}_t^T - \mathbf{B}\mathbf{z}_t\mathbf{x}_t^{*T} - \mathbf{x}_t^*\mathbf{z}_t^T\mathbf{B}^T = 0 \tag{5.56}$$

$$\mathbf{R}^{new} = \frac{1}{\tau}\sum_{t=1}^{\tau}\mathbf{B}\mathbf{M}_{t,t}\mathbf{B}^T + \mathbf{z}_t\mathbf{z}_t^T - \mathbf{B}\mathbf{z}_t\mathbf{x}_t^{*T} - \mathbf{x}_t^*\mathbf{z}_t^T\mathbf{B}^T \tag{5.57}$$

Then, for finding the covariance matrix of the state noise ω, we have:

$$\frac{d\mathcal{L}}{d\mathbf{Q}} = -(\tau-1)\mathbf{Q}^T + \sum_{t=2}^{\tau}[\mathbf{M}_{t,t} + \mathbf{A}\mathbf{M}_{t-1,t-1}\mathbf{A}^T - \mathbf{M}_{t-1,t}\mathbf{A}^T - \mathbf{A}\mathbf{M}_{t,t-1}] = 0 \tag{5.58}$$

$$\mathbf{Q}^{new} = 1/(\tau-1)\sum_{t=2}^{\tau}[\mathbf{M}_{t,t} + \mathbf{A}\mathbf{M}_{t-1,t-1}\mathbf{A}^T - \mathbf{M}_{t-1,t}\mathbf{A}^T - \mathbf{A}\mathbf{M}_{t,t-1}] \tag{5.59}$$

Moreover, in terms of the Eqn 5.17, the covariance matrix $\mathbf{Q}$ of state noise contains two unknown parameters (β, α). Applying the following process of calculation, it is feasible to estimate the optimized values , respectively.

If we let $\mathbf{Q}_\alpha = \mathbf{Q}/\beta = \exp[-\alpha\mathbf{\Lambda}_f]$ and then we have

$$\log\det\mathbf{Q} = \log\beta\exp[-\alpha\sum_i\lambda_i] = \log\beta - \alpha\,\mathrm{Tr}(\mathbf{\Lambda}_f) \tag{5.60}$$

$$\mathrm{Tr}(\mathbf{Q}^{-1}\mathbf{M}_{t,t}) = \beta\,\mathrm{Tr}(\mathbf{Q}_\alpha^{-1}\mathbf{M}_{t,t}) \tag{5.61}$$

$$\frac{d\mathbf{Q}_\alpha^{-1}}{d\alpha} = \mathbf{\Lambda}_f\mathbf{Q}_\alpha^{-1} \tag{5.62}$$

So the differentials of the log-likelihood are:

$$\begin{aligned}\frac{d\mathcal{L}}{d\beta} &= (\tau-1)/\beta - \frac{1}{\beta^2}\sum_{t=2}^{\tau}[\mathrm{Tr}(\mathbf{Q}_\alpha^{-1}\mathbf{M}_{t,t}) + \mathrm{Tr}(\mathbf{A}^T\mathbf{Q}_\alpha^{-1}\mathbf{A}\mathbf{M}_{t-1,t-1}) \\ &\quad - \mathrm{Tr}(\mathbf{A}^T\mathbf{Q}_\alpha^{-1}\mathbf{M}_{t-1,t}) - \mathrm{Tr}(\mathbf{Q}_\alpha^{-1}\mathbf{A}\mathbf{M}_{t,t-1})] \\ &= (\tau-1)/\beta - \frac{1}{\beta^2}\mathrm{Tr}(\mathbf{Q}_\alpha^{-1}\tilde{\mathbf{M}})\end{aligned} \tag{5.63}$$

Where

$$\tilde{\mathbf{M}} = \sum_{t=2}^{\tau}[\mathbf{M}_{t,t} + \mathbf{A}\mathbf{M}_{t-1,t-1}\mathbf{A}^T - \mathbf{M}_{t-1,t}\mathbf{A}^T - \mathbf{A}\mathbf{M}_{t,t-1}] \tag{5.64}$$

and

$$\begin{aligned}\frac{d\mathcal{L}}{\alpha} &= -(\tau-1)\operatorname{Tr}(\mathbf{\Lambda}_f) + \frac{1}{\beta}\sum_{t=2}^{\tau}[\operatorname{Tr}(\mathbf{Q}_\alpha^{-1}\mathbf{\Lambda}_f\mathbf{M}_{t,t}) + \operatorname{Tr}(\mathbf{A}^T\mathbf{Q}_\alpha^{-1}\mathbf{\Lambda}_f\mathbf{A}\mathbf{M}_{t-1,t-1}) \\ &\qquad - \operatorname{Tr}(\mathbf{A}^T\mathbf{Q}_\alpha^{-1}\mathbf{\Lambda}_f\mathbf{M}_{t-1,t}) - \operatorname{Tr}(\mathbf{Q}_\alpha^{-1}\mathbf{\Lambda}_f\mathbf{A}\mathbf{M}_{t,t+1})] \\ &= -(\tau-1)\operatorname{Tr}(\mathbf{\Lambda}_f) + \frac{1}{\beta}\operatorname{Tr}(\mathbf{Q}_\alpha^{-1}\mathbf{\Lambda}_f\tilde{\mathbf{M}})\end{aligned} \tag{5.65}$$

We have to solve the two equations:

$$\frac{d\mathcal{L}}{d\alpha} = 0 \tag{5.66}$$

and

$$\frac{d\mathcal{L}}{d\beta} = 0 \tag{5.67}$$

According to Eqn 5.66 and Eqn 5.67, Eqn 5.63 and Eqn 5.64 can be written as following two equations:

$$\frac{1}{\beta} = (\tau-1)\operatorname{Tr}(\mathbf{\Lambda}_f)/\operatorname{Tr}(\mathbf{Q}_\alpha^{-1}\,\Lambda_f\,\tilde{\mathbf{M}}) \tag{5.68}$$

$$\operatorname{Tr}(\mathbf{\Lambda}_f)\operatorname{Tr}(\mathbf{Q}_\alpha^{-1}\tilde{\mathbf{M}}) = \operatorname{Tr}(\mathbf{Q}_\alpha^{-1}\mathbf{\Lambda}_f\tilde{\mathbf{M}}) \tag{5.69}$$

Since the factor $\operatorname{Tr}(\mathbf{Q}_\alpha^{-1}\tilde{\mathbf{M}})$ can be written as according to previously setting, then we have the new representation of $\operatorname{Tr}(\mathbf{Q}_\alpha^{-1}\tilde{\mathbf{M}})$:

$$\begin{aligned}\operatorname{Tr}(\mathbf{Q}_\alpha^{-1}\tilde{\mathbf{M}}) &= \operatorname{Tr}(\frac{1}{\beta}\exp(\alpha\mathbf{\Lambda_f}\tilde{\mathbf{M}}) \\ &= \frac{1}{\beta}\operatorname{Tr}(\exp\alpha\mathbf{\Lambda}_f\tilde{\mathbf{M}})\end{aligned} \tag{5.70}$$

Then , with respect to the Eqn 5.70, the Eqn 5.69 can be written as follows:

$$\begin{aligned}\operatorname{Tr}(\mathbf{\Lambda}_f)\frac{1}{\beta}\operatorname{Tr}(\exp(\alpha\mathbf{\Lambda}_f)\tilde{\mathbf{M}}) &= \operatorname{Tr}(\mathbf{Q}_\alpha^{-1}\mathbf{\Lambda}_f\tilde{\mathbf{M}}) \\ &= \operatorname{Tr}(\frac{1}{\beta}\exp(\alpha\mathbf{\Lambda}_f)\mathbf{\Lambda}_f\tilde{\mathbf{M}})\end{aligned} \tag{5.71}$$

The left-side and right-side of the Eqn 5.71 can be written as :

$$\begin{aligned}\text{Left-side} &= \mathrm{Tr}(\boldsymbol{\Lambda}_f)\,\mathrm{Tr}(\ \exp(\alpha\boldsymbol{\Lambda}_f)\ \tilde{\mathbf{M}}\ ) \\ &= \mathrm{Tr}(\boldsymbol{\Lambda}_f)\Big[\sum_i^m\sum_j^m \tilde{M}_{ij} + \sum_i^m(\exp(\alpha\lambda_{i,i})-1)\tilde{M}_{i,i}\Big]\end{aligned} \tag{5.72}$$

and

$$\begin{aligned}\text{Right-side} &= \mathrm{Tr}(\exp(\alpha\boldsymbol{\Lambda}_f)\ \mathbf{K}_m) \\ &= \sum_i^m \lambda_{i,i}\tilde{M}_{i,i}\exp(\alpha\lambda_{i,i})\end{aligned} \tag{5.73}$$

where

$$\mathbf{K}_m = \begin{pmatrix} \lambda_{1,1}\tilde{M}_{1,1} & & & 0 \\ & \lambda_{2,2}\tilde{M}_{2,2} & & \\ & & \ddots & \\ 0 & & & \lambda_{m,m}\tilde{M}_{m,m} \end{pmatrix} \tag{5.74}$$

From the above inference of Eqn 5.71 and Eqn 5.68, the unknown parameters set (α, β) can be estimated numerically.

5.5 Results

In the following part, the simulated data as well as the real MEG data are both applied for the evaluation of the Kalman smoother method to MEG source reconstruction on high-resolution mesh $\mathbf{M}^+$. Since this method is based on a successive time sequence, we choose the measurement on the time period of interest: $(1, \cdots, t, \cdots, \tau)$ with the length of τ. In terms of the experimental configuration we have(the experimental configuration is : Inter Core 2(1,8GHz), Linux system(2.6.34.1)-32bit, matlab $7.9.0(R2009b) \times 32$ edition, RAM: 4GB), we defined $\tau = 16$. The source distribution on time point t is what we try to reconstruct by the Kalman smoother method here. The initial current source distribution on 1st time point (namely J_{t1}) is come from the reconstruction result by the Bayesian super-resolution method on corresponding time point (refer to the last Chapter). Moreover, the interpolated high-resolution mesh $\mathbf{M}^+$ generated in last Chapter is directly used here.

5.5.1 Synthetic results

In this simulated experiment, we applied two groups of simulated current sources as well as the corresponding measurement as same as last two Chapters, i.e. artificial source distribution and realistic source distribution in *Appendix B* which are also called as *synthetic source A* and *synthetic source B*. It is worth to note that the current sources in successive 16 time points we selected are as same as the ones in last Chapter. The only difference is only the first 16 of 31 time points are selected because of the restriction of experimental configuration. And, the *target time point* is 3 of 16 for both *synthetic source A* and *synthetic source B*.

The whole process of simulation experiment is based on the frame of EM algorithm. Starting with the *E-step*, the high-resolution current source frames in successive 15 time points are estimated *forward* with the initial guess of the state noise covariance $\mathbf{Q}$ and observation noise covariance $\mathbf{R}$ in terms of the theory of Kalman filter. The corresponding factors of the Kalman filter, are calculated, such as $\mathbf{K}_t$, $\mathbf{P}_t^t$ and $\mathbf{x}_t^t$. Then, the Kalman smoother is applied for the *backward* estimation based on all of these factors calculated above. Then, according to the previous section, the observation noise covariance $\mathbf{R}$ and the state noise covariance $\mathbf{Q}$ are estimated in the *M-step* . The entire process is recursive and continues untill the estimation of $\mathbf{R}$ and $\mathbf{Q}$ tend to be convergent. Furthermore, the unknown parameter set α and β which are associated with $\mathbf{Q}$ are optimized numerically in terms of the best estimation of α, shown in Fig 5.3 and Fig 5.6 for *synthetic source A* and *synthetic source B*, respectively. Then, β can be calculated with respect to the optimized α, refer to Eqn 5.68. In terms of the above results as well as the known parameter set: R, α and β, we applied for the synthetic current source generation, Table 5.1 shows the comparison of the setting parameters and the reconstructed parameters for *synthetic source A* and *synthetic source B*. Table 5.1 indicates that the instability still exists in parameter reconstruction of Kalman smoother method, where the parameter $\mathbf{R}$ can be provided with satisfied reconstruction while reconstruction of α and β exist errors to the original parameters.

The figures, Fig 5.2 and Fig 5.5, demonstrate the overlapping pattern of MEG measurement on the selected 16 successive time points for *synthetic source A* and *synthetic source B*, respectively. The line in different color indicates the measurement on individual time point. The figures, Fig 5.4 and Fig 5.7, indicates the comparison

Table 5.1: Reconstructed parameters results

Parameter type	synthetic original	for *synthetic source A*	for *synthetic source B*
Frobenius Norm of **R**	1	1	1
α	0.5	8.2	1.6
β	1	0.7314	1.013

of the simulated current source pattern on high-resolution mesh M^{+} (on the left) and the reconstruction result by the Kalman filter method (on the right) at the *target time point* for the *synthetic source A* and the *synthetic source B*, separately. Fig 5.4 does not show the exactly correct location of the *synthetic source A*. Whereas, Fig 5.4 indeed reconstruct the main location of the *synthetic source B* although the strength of the current source is lower than the original current source. The Table 5.2 shows the comparison of logarithm of RMS (root mean square) error for the Kalman smoother method and the Super-resolution method to the *synthetic source A* and *synthetic source B* at the *target time point*, respectively. From the table, it is clear that the reconstruction results by the Kalman smoother method is superior than the super-resolution method.

Table 5.2: Logarithm of RMS error results

data type	for the Kalman smoother method	for the super-resolution method
for *synthetic source A*	3.2576	7.9077
for *synthetic source B*	7.7807	32.3870

5.5.2 Application to the real MEG data

We get the real MEG data of visual expression based on *Appendix 3*. Firstly, we obtained the cortical surface mesh with 262658 vertices and 565782 faces from the structural MRI scan of the same subject by *Freesurfer* (http://surfer.nmr.mgh.harvard.edu/) (5.0.0). Since the coordinate of MRI cortical surface are different with the MEG coordinate, the coordinate registration is processed as the first step(with the special solution provided by YNiC). However, the spatial resolution of mesh obtained from MRI is too

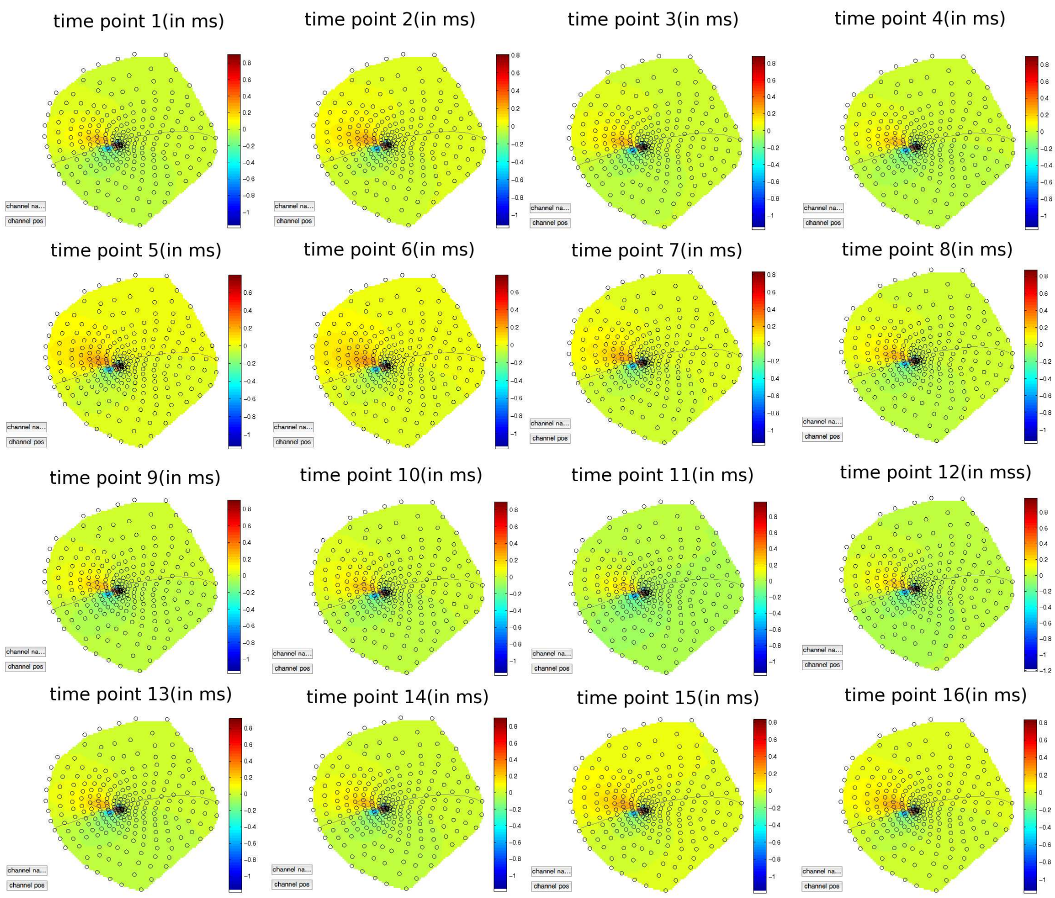

Figure 5.2: This figure demonstrates 2D projected pattern of MEG sensor measurement on the selected 16 successive time points(in *ms*) for *synthetic source A*.

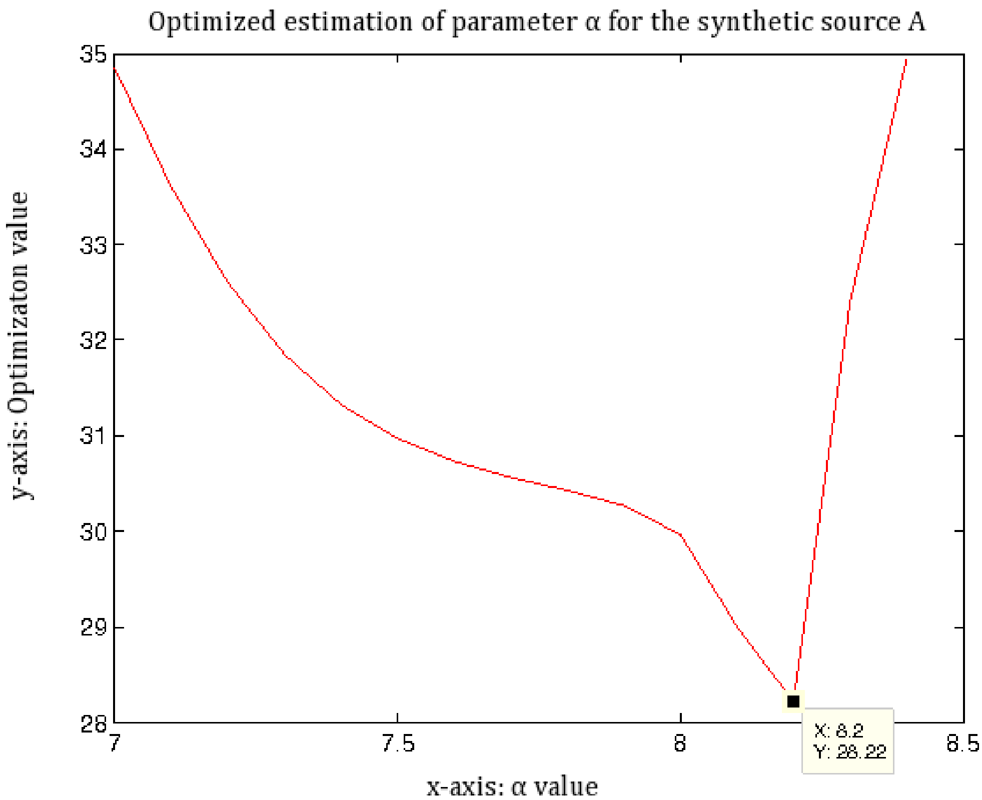

Figure 5.3: This figure shows the optimized estimation of unknown parameter α for *synthetic source A*. The x-axis represents the variable range of α, and the y-axis represents the logarithm of $\frac{d\mathcal{L}}{d\alpha}$ (refer to Eqn 5.69). With the minimum on the y-axis, we can obtain the optimized α on the corresponding x-axis.

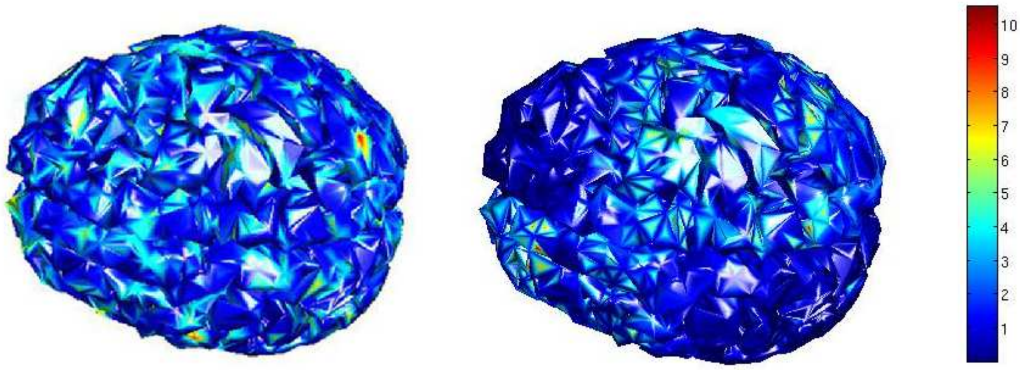

Figure 5.4: This figure shows the comparison of the simulated current source pattern of *synthetic source A* on high-resolution mesh M^{+} (on the left) and the reconstruction result by the Kalman smoother method (on the right) at the *target time point*.

large for the realistic or reasonable solution. The simplified mesh is therefore generated by the software *Remesh* (*http://remesh.sourceforge.net/*). In terms of the mesh resolution selected for MEG analysis in (MNE), we apply the reasonable spatial resolution for mesh **M** is with 2600 vertices and 5192 faces . Secondly, the measurement of MEG signals are represented as a $96 \times 248 \times 813$ matrix, where 96 indicates the number of different stimulus, 248 indicates the number of sensors and 813 indicates the sequence of time instants. The visualization of this measurement matrix is showed in Fig C.1.

The experiment process of the MEG real data is similar with the synthetic experiment we illuminated above but just implementing with the real MEG data instead of the synthetic ones. We also choose 16 successive time points (which are as same as the first 16 time points of 31 in the *Application to the real MEG data* in *Chapter 3*) from 813 time instants, and 70 th stimulus from 96 different stimulus, then we obtain a 248×813 matrix $\mathbf{B}^{real}$. The framework of Kalman smoother is then constructed and EM algorithm is used for the parameter optimization. With the proper parameter estimation with EM algorithm, the optimized estimation of current source distributed on the high-resolution mesh $\mathbf{M}^{+}$, 7792×3 matrix $\mathbf{j}_{new}^{real}$, is produced finally.

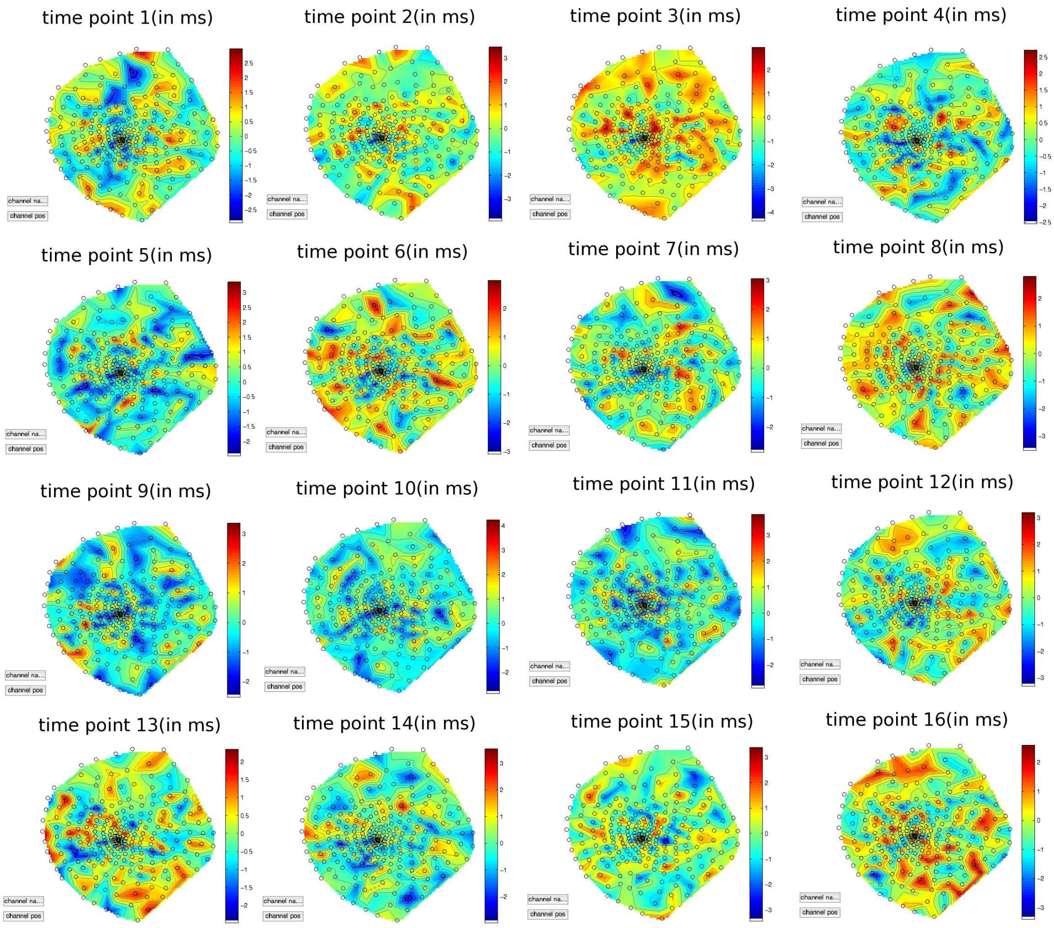

Figure 5.5: This figure demonstrates 2D projected pattern of MEG sensor measurement on the selected 16 successive time points(in *ms*) for *synthetic source B*.

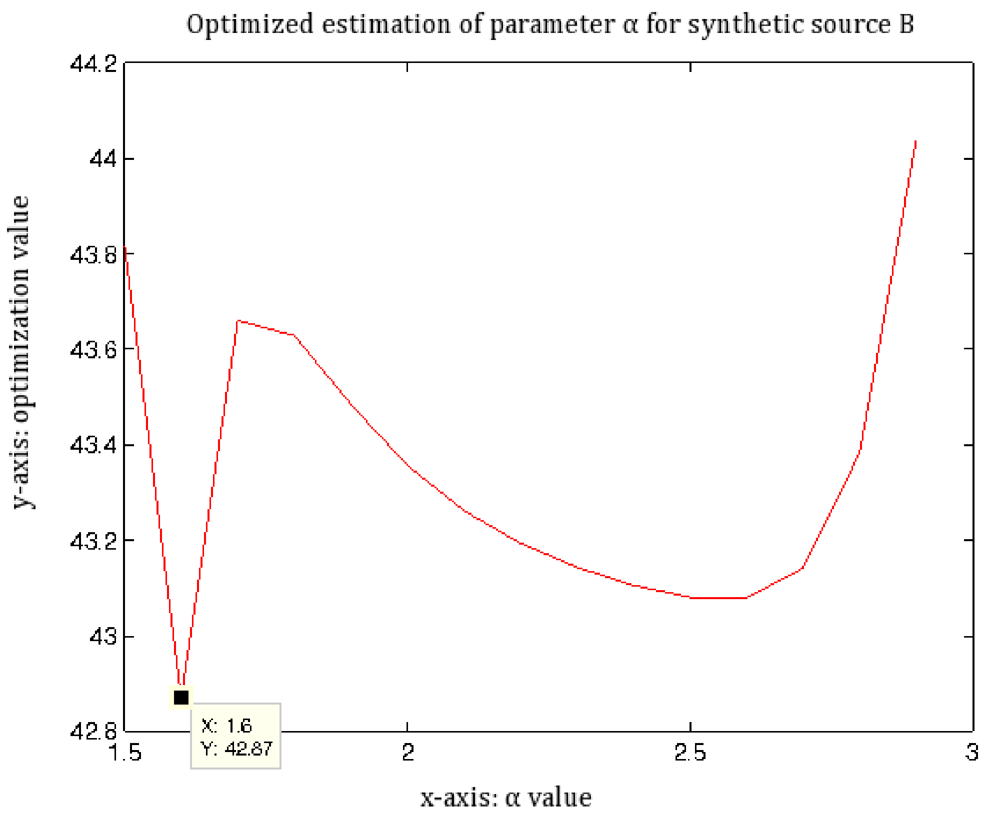

Figure 5.6: This figure shows the optimized estimation of unknown parameter α for *synthetic source B*. The x-axis represents the variable range of α, and the y-axis represents the logarithm of $\frac{d\mathcal{L}}{d\alpha}$ (refer to Eqn 5.69). With the minimum on the y-axis, we can obtain the optimized α on the corresponding x-axis.

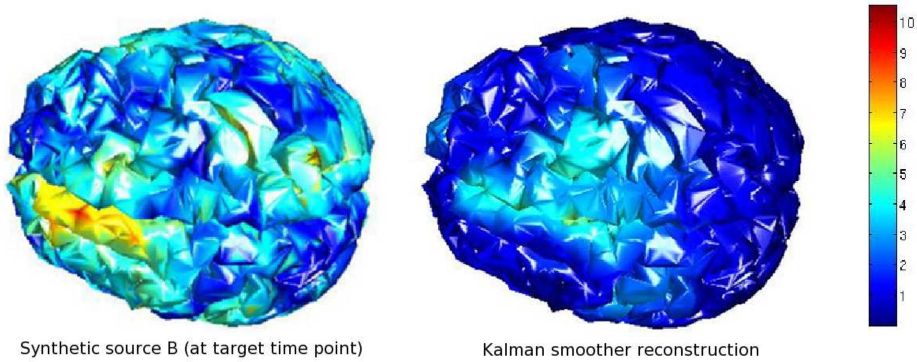

Figure 5.7: This figure shows the comparison of the simulated current source pattern of *synthetic source B* on high-resolution mesh M^{+} (on the left) and the reconstruction result by the Kalman smoother method (on the right) at the *target time point*.

5.6 Discussion

The reconstructions shown in Fig 5.4, Fig 5.7 and Fig 5.10 do not show the satisfied results. This is possibly caused by the reason that the method is a integration of the Kalman smoother and the Basis function source model. The application of this method may affected by any inaccuracy caused by the basis function source model to the MEG source reconstruction at the initial time point. In the context of this reason, it is feasible that to apply the reconstructed results via other solutions as the initial estimation for the Kalman smoother method in the future work. Additionally, with respect to Eqn 5.6, the velocity of the source variation in the temporal field $\mathbf{v_t}$ is assumed as the same value. It might not adaptable in practice. This assumption of current source variation with uniform velocity may also produce the inaccuracy of the reconstruction. However, based on the current simulation results, it is notable that the reconstruction of *synthetic source B* shows better performance rather than the result of *synthetic source A*, where the source distribution is mainly reconstructed at the correct location. From this performance, we can also conclude that the algorithm may more effective and sensitive to the distributed and supercial current source rather than the single or/and deep current source of the cortex.

Since the Kalman smoother method is designed for the high-resolution current source frames, there are a number of larger size matrix calculation in the method. This

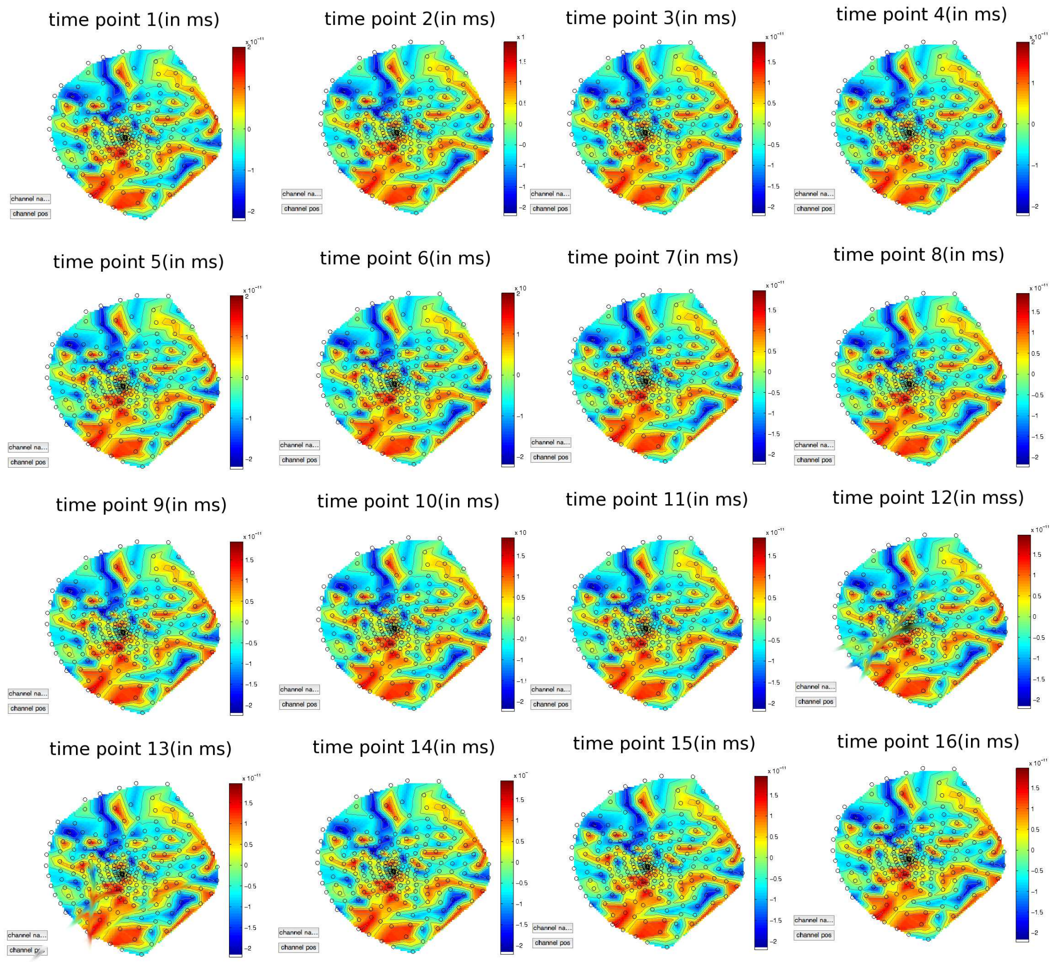

Figure 5.8: This figure demonstrates 2D projected pattern of MEG sensor measurement on the selected 16 successive time points(in *ms*) for the real MEG data of facial expression (showed as the $96 \times 248 \times 813$ matrix).

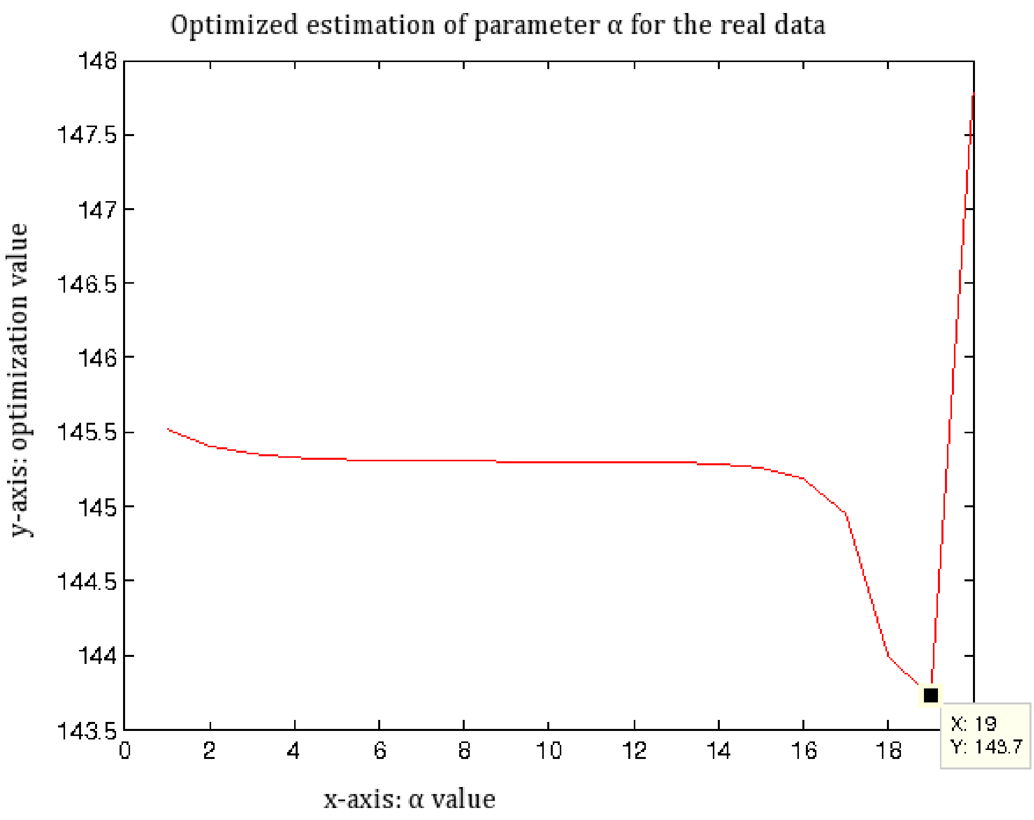

Figure 5.9: This figure shows the optimized estimation of unknown parameter α for the real MEG data of facial expression (showed as the $96 \times 248 \times 813$ matrix). The x-axis represents the variable range of α, and the y-axis represents the logarithm of $\frac{d\mathcal{L}}{d\alpha}$ (refer to Eqn 5.69). With the minimum on the y-axis, we can obtain the optimized α on the corresponding x-axis.

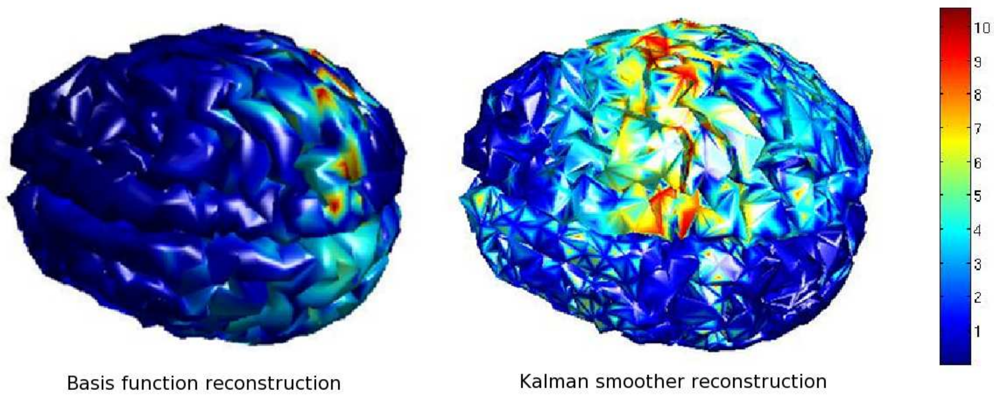

Figure 5.10: This figure demonstrate the comparison of reconstruction results at *target time point:* 70 *of* 813 by the basis function method on $\mathbf{M}$ (on the left) and by the Kalman smoother method on $\mathbf{M}^+$ (on the right).

leads to expensive computation complexity. These problems may affect the accuracy of the result as well as the application of the method in the real world. Also, during the session of unknown parameters estimation, the unknown parameters β and α are applied for determining covariance matrix $\mathbf{Q}$ of the state noise ω (refer to Eqn 5.2 and Eqn 5.10). By applying EM algorithm to the expectation of the joint probability density function (pdf) $\mathcal{L}$ (refer to Eqn 5.38), the optimized estimated α and β are obtained in turn. Since we are using numerical analysis to estimate the optimized α as well as β so that to obtain the minimization of $\mathcal{L}$ (refer to Eqn 5.66 and Eqn 5.67), it is worth to be careful about the selection of possible range of α which is the difficult part in this step. If the possible range of α we select is not large enough, it is possible that the obtained estimation is corresponding to the local minimization of $\mathcal{L}$, so that generates the incorrect estimation of α, and then affects the accuracy of estimated β. In practical, the solution is to select the possible range of α as large as possible with respect to the experimental circumstance so that to avoid this inaccurate estimation.

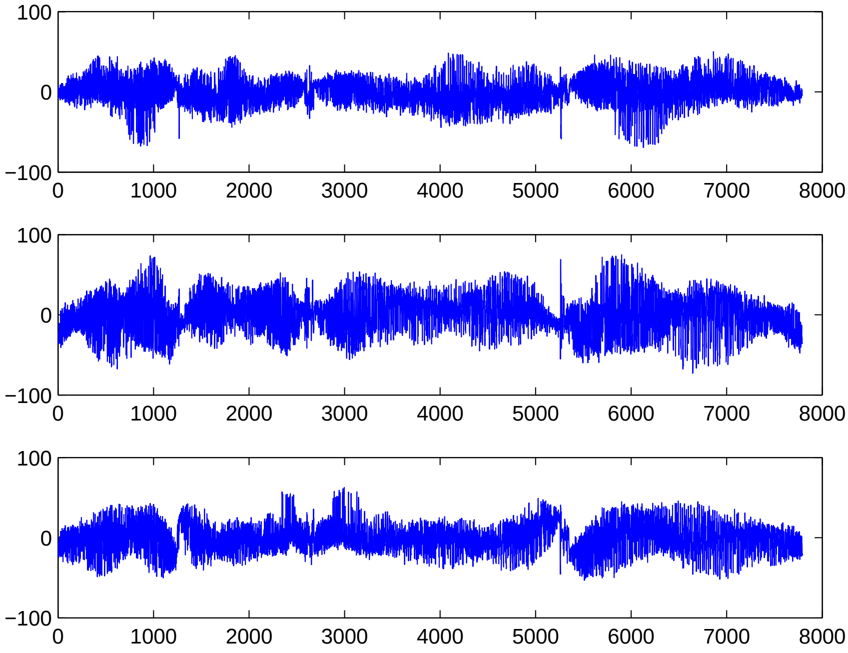

Figure 5.11: This figure demonstrate the wave pattern of the reconstructed current source $\mathbf{J}_{70}$ (at the time point 70 of 813) on $\mathbf{M}^{+}$ in 3D by the Kalman smoother method. From the top to the bottom, the patterns show the wave pattern of $\mathbf{J}_{70}$ in x, y and z dimension, respectively. The x-axis indicates the index of vertex of $\mathbf{M}^{+}$, the y-axis indicates the amplitude of the current source.

5.7 Conclusion

In this Chapter, the Kalman lter and Kalman smoother are applied for correcting the reconstruction in the temporal eld with assuming that MEG system is a linear dynamic system. Here, we summarize the novelties of the algorithm and the contribution to the MEG source reconstruction.

We sufficiently utilize the forward model of MEG system and assume it as a linear dynamic system to design an approach based on previous work. With respect to the reconstruction from the last chapter, the transformation matrix *Leadfield* on the high-resolution mesh M^+ , as well as the measurement of the magnetic field from MEG, are directly used for the estimation of the source distribution as a continuous time series.

We assume there is an unknown hidden cortical activity in this dynamic process. The Kalman filter is used to estimate the dynamic state while the Kalman smoother is applied for correcting the source distribution of the hidden state with EM algorithm. From the intrinsic property of the Kalman filter as well as the framework of this method, it is apparent that this approach is advantageous to solve the inverse problem. Based on the source reconstruction results, the Kalman smoother method shows superior performance for MEG source reconstruction, as shown in Fig 5.7. However, it still shows instability and inaccuracy of reconstruction on different types of current source, shown in Fig 5.4 and Fig 5.10. These may be related to the following limitations when the method is applied in practice. As our algorithm is based on the assumption of a linear dynamic system, it does not cope well with the strong nonlinearity in the model. In other words, the non-linear signals may not be reconstructed properly using this method. Additionally, the computational complexity in time and space remains high. A number of large full-matrix and sparse matrix calculations decrease the efficiency of the method. It is possible to either upgrade the experimental conguration or to develop the structure of matrix calculation mathematically to optimize the calculation in future work. Moreover, with the advanced conguration, the estimations can be extended to more than 10 time points which can effectively improve the estimation accuracy.

Chapter 6

Conclusions and future work

In this chapter, we summarize the main contributions of the thesis, as well as discuss the possible directions of the work that can be improved in the future. Especially, the novel ideas of our work on MEG spatial-temporal source reconstruction are emphasized. Then, we discuss the advantages and limitations of the theoretical models we design and their related applications in real MEG source reconstruction experiment, i.e. the basis function reconstruction algorithm, the Bayesian super-resolution algorithm and the Kalman smoother estimation algorithm. Moreover, applying our approaches into the real-world MEG application, we further extend our findings.

6.1 Contributions

6.1.1 Novel idea combined both pattern recognition and MEG source reconstruction

The novel idea of the thesis is to introduce classical pattern recognition methods, e.g. basis function extraction, super-resolution method and Kalman filter, to solve the problem of MEG source reconstruction, in other words, using a new angle of pattern-recognition as the solution to reconstruct the MEG current sources. The whole design of this thesis work is based on the MEG spatial reconstruction at a single time point. Rather than applying the classical source distribution, e.g. dipoles or current source volumes of the brain, we assume that the 3D source distributed on each vertex of the original cortical surface mesh(generated from the MRI scan of the subject). Then, the basis function algorithm is applied to spatially reconstruct the source distribution at the specific time point. Subsequently, another method from pattern-recognition, a super-resolution algorithm, is introduced to expand the reconstructed source distribution from the original mesh into the interpolated high-resolution mesh, through the process of which the spatial resolution of the reconstruction is developed. Furthermore, as the MEG measuring system is assumed to be a linear dynamic system, one of the classical solutions, Kalman smoother, is finally used to improve the temporal resolution of this source reconstruction based on the high- resolution mesh. In summary, the thesis combines the use of both some classical methods of pattern recognition and the MEG spatial-temproal source reconstruction in order to achieve a highly sensitive spatial and temporal reconstruction. The design of the thesis aims to bring together the three related topics (*Chapter* 3, 4 and 5) together and present them as an integrated process rather than independent topics.

6.1.2 Spatial source reconstruction by Basis function

Specifically, for the basis function algorithm elaborated in *Chapter 3*, we explore a new method of MEG source reconstruction based on modeling the current source with extended basis functions. This algorithm shows a good possibility to reconstruct the source using the basis functions set and the corresponding coefcients rather than the classical Beamforming or minimum-norm methods. This algorithm provides a smooth and well-conditioned reconstruction problem which can be solved directly by an inverse method. The results are more physically plausible than the minimum-norm method and are resistant to noise. Fig 6.1 shows the comparison of source reconstructions between the basis function method for the whole cortical surface and the partition of ROI by the Normalized cut method at one particular time-point.

Moreover, corresponding the smallest eigenvectors is not the only basis function set for this basis function algorithm. It would be interesting to try other types of basis functions and evaluate the efficiency of each type of basis functions as part of the future work.

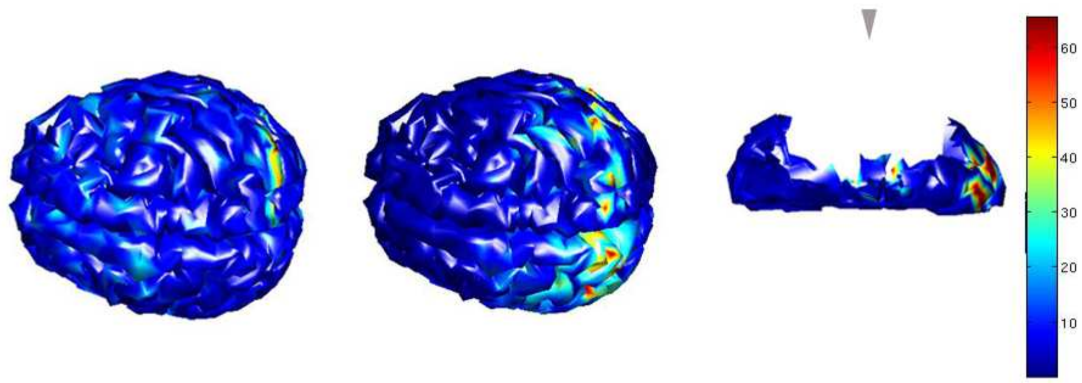

Figure 6.1: The figure shows the comparison of source reconstructions between the basis function method for the whole cortical surface and the partition of ROI by the Normalized cut method at one particular time-point for synthetic source A. The color from red to blue show the intensity of source strength from strong to weak. The left: the original source pattern; the middle: the basis function reconstruction based on the whole cortical surface; the right: the basis function reconstruction on the partition of ROI obtained by normalized cut method. The source distribution at one time point used here is selected from *synthetic source A* (at time point 1, refer to Fig 3.11).

6.1.3 Spatial resolution improvement with Bayesian super-resolution

For the Bayesian super-resolution algorithm elaborated in *Chapter 4*, a new solution is constructed for improving the spatial-resolution of MEG source reconstruction at a single time point by introducing a classical method (Bayesian super-resolution method) from pattern-recognition theory. This approach is applied based on the MEG spatial reconstruction with basis function method elaborated in *Chapter 3* of thesis.

The results from synthetic data as well as the real MEG data show reasonable estimation of the parameters which restrict the assumption on the source distribution on high-resolution cortical mesh. From the quantitative experimental results, it is apparent that the spatial resolution has been effectively improved. However, since the parameter estimation as well as the optimization of high-resolution source distribution contain a number of large matrix calculations. Although some of them can be simplied by sparse matrix calculation, there remain many other full matrices that lead to expensive calculations in terms of temporal and spatial complexity. Fig 6.2 indicates the high-resolution mesh $\mathbf{M}^+$ interpolated from the low-resolution mesh $\mathbf{M}$, as well as the comparison between the original synthetic source distribution on low resolution mesh $\mathbf{M}$ and the source reconstruction on the high-resolution mesh $\mathbf{M}^+$ at one particular time-point.

6.1.4 Temporal source reconstruction by Kalman smoother

For the Kalman smoother algorithm elaborated in *Chapter 5*, we applied the Kalman lter and Kalman smoother to correct the reconstruction temporally, keeping in mind the assumption that MEG system is a linear dynamic system.

Based on previous work, the forward model of MEG system and the assumption of a linear dynamic system are sufficiently utilized to design the approach. We assume that there is an unknown and hidden cortical activity in this dynamic process. The Kalman filter and Kalman smoother are applied respectively for the state estimating forward and the state correcting backward. The MEG measurement of magnetic field are directly used for the source reconstruction in the temporal field here. Fig 6.3 shows the comparison between the original synthetic source distribution and the source reconstruction on the high-resolution mesh $\mathbf{M}^+$ at one particular time-point.

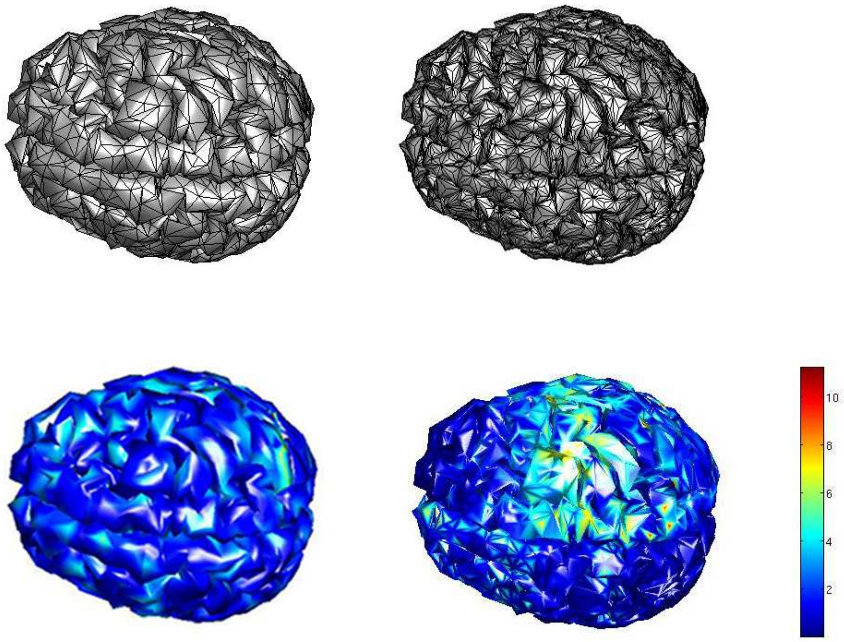

Figure 6.2: The top part of the figure shows the *low-resolution mesh*, **M** on the left (with V vertices and F faces) and the *high-resolution mesh*, $\mathbf{M}^{+}$ on the right with $V^{+} = V + F$ vertices and $F^{+} = 3F$ faces. **M** has 2600 vertices and 5192 faces; $\mathbf{M}^{+}$ is with 7792 vertices and 15576 faces; the bottom part shows the comparison of the simulated source pattern of *synthetic source B* on high-resolution mesh $\mathbf{M}^{+}$ (on the left) and the reconstruction result by the Bayesian super-resolution method (on the right) at the *target time point* for synthetic source A. The color from red to blue show the intensity of source strength from strong to weak.

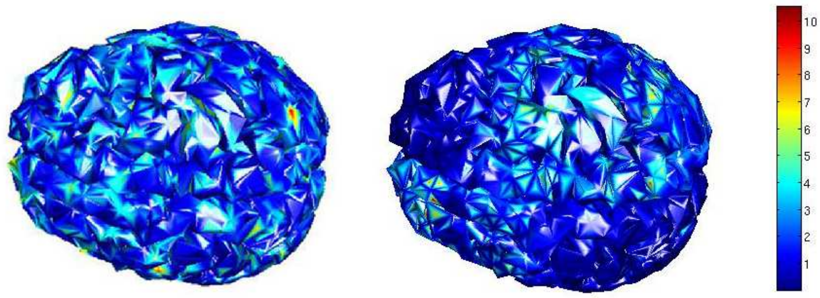

Figure 6.3: The figure shows the comparison of the simulated source pattern of *synthetic source A* on high-resolution mesh M^{+} (on the left) and the reconstruction result by the Kalman filter method (on the right) at the *target time point*.

However, the computational complexity in time and space remains high for this algorithm to be practically applicable. Also, since we assume that the MEG system as a linear dynamic system, this leads to the insensitivity of the reconstruction of the current sources with nonlinear relationship, in other words, the non-linear signals may not be reconstructed properly using this method.

6.1.5 Summary

The main contribution of this thesis can be summarized into the following three points:

Firstly, the thesis makes the connection between the field of pattern recognition, graph theory and medical imaging (specifically on MEG source reconstruction). The research process of pattern recognition and MEG source reconstruction share great similarity between each other, and specifically it is to mathematically build and optimize the research target by applying specific algorithms to the observed information. On one hand, pattern recognition is a well-developed research area that fully contains a variety of algorithms. On the other hand, MEG source reconstruction is a research field with full potential for the further research is finding the solution of the methodology. The combination of these two fields in the thesis opens a new window for the MEG source reconstruction problem from a novel angle.

Secondly, our work provides new possibilities for the further MEG research on solving the reconstruction problem. MEG source reconstruction is an ill-posed inverse problem, which is theoretically unsolvable. The current methodologies of MEG source reconstruction, such as Minimum-norm method, Beamforming method and the equivalent current dipole (ECD) method, have their intrinsic weaknesses. For instance, Minimum-norm method requires obtaining prior information of the current source distribution, which is difficult in practice. Beamforming method is not sensitive to the current sources, which have high temporal correlations. And for the equivalent current dipole (ECD) method, it is quite difficult to estimate the number of dipoles in advance, meanwhile, ECD method shows insensitivity to the localization to deep source (Preissl, 2005). All these intrinsic weaknesses provide room for improvement in such a research field and the possibility of exploring new solutions by applying the knowledge of the new research field, such as pattern recognition.

Thirdly, we have made the contribution specifically on applying the basis function method, super-resolution method and Kalman smoother method into MEG source reconstruction. Instead of concentrating on the current source variation in a conventional way, the novel idea of the basis function method is to focus on the geometrical information of the cortical surface which are described by a set of basis functions (mesh Laplacian eigenvectors corresponding to the smallest eigenvalues), while the variable information of the current source is defined as the corresponding coefficients. This design combines the MEG source reconstruction with the application of basis functions. Meanwhile, the graph theory is fully applied in this combination. Then, the idea of Bayesian super-resolution is borrowed from the image processing into the MEG source estimation on the high-resolution cortical mesh. Instead of working on the grids of multiple same-scene images(with low/high resolution), our research is based on the MEG source frames(with low/high resolution cortical mesh) at different time points. This application to MEG source reconstruction provides a great similarity to the process of image estimation by Bayesian super-resolution. This is another good example of combination of MEG source reconstructions and the classical pattern recognition.

Furthermore, in the application of Kalman smoother into the MEG source reconstruction, the signal processing of MEG system can be interpreted as a dynamic system over the course of time. The property of dynamic system, i.e. Markov property, can be directly applied into the detailed analysis. With this application, MEG source reconstruction can be implemented in a straightforward manner by utilizing the MEG measurement of the magnetic field.

6.2 Discussion

In summary, three methods of MEG source reconstruction have been designed and evaluated in the thesis, which are the Basis function method, the Bayesian super-resolution method and the Kalman smoother method. Since all of this research is aimed at application to real MEG analysis, it is worth comparing these three methods from the viewpoint of intrinsic features efficiency. Although the basis function method is limited for the reconstruction of the whole cortical surface, it presents the superior advantage of the simplicity of the process as well as the easy computation complexity. With the proper localization of the region of interest(ROI), it is feasible that the basis function method can be more competitive than any other two methods in the real MEG application. The Bayesian super-resolution method is designed for MEG source reconstruction in light of the super-resolution method applied to image processing. The mathematical framework of the method provides a framework for the inverse problem of MEG. Instead of analyzing image data, the method here is used for processing the problem of source distribution on the 3D cortical surface mesh which highly increases the computation complexity of the reconstruction results. This feature of expensive computation with unstable result make the Bayesian super-resolution method less effective than other two methods. Additionally, the Kalman smoother method is based on the assumption of a linear dynamic system and directly estimates the source frame in a successive time points with respect to the measurement. The Kalman filter is used for estimating the dynamic state while the Kalman smoother is applied for correcting the source distribution of the hidden state with the EM algorithm. The properties of the Kalman filter makes this approach advantageous for solving the inverse problem. However, non-linear signals may not be reconstructed properly by it. Moreover, the method still contains a number of large matrix calculation which affects the accuracy and efficiency of its application in the real world. Since the mathematical framework, as well as its calculation are not as complex as the Bayesian super-resolution method, the Kalman smoother method presents advantages in practice rather than the Bayesian super-resolution method.

6.3 Future work

In the theoretical design of the thesis, 'J' represents the total current (primary currents + volume currents)(Barth et al., 1986). In our assumption, as the measurement is assumed on only the radial component of the magnetic field at single homogeneous spheroid, the majority of contributions of the volume currents vanish and the MEG measurement are only from the primary term approximately in this case. However, for the more accurate reconstruction with more realistic situation, the volume current cannot be disregarded. The further source modeling with both primary currents and volume currents is one of the future work I would like to concentrate on.

Additionally, during the whole experiment process of the thesis work , the expensive computation in both space and time complexity stands as a signicant problem which affects the application efficiency.

One possible solution is to decrease the computation cost. Since there are plenty of loop calculation associated with the large size matrix calculation (including the full matrix and sparse matrix calculation) , the implement of the algorithms leads to huge time and space computation complexity. It is crucial to decrease the computation cost so that the algorithm can applied more efciently and realistically in real MEG source reconstructions; The main experimental conguration we use is: InterCore2(1,8GHz), Linux system(2.6.34.1)-32bit, matlab 7.9.0(R2009b) 32 edition , RAM: 4GB. It is possible that we can upgrade to an advanced conguration for higher computational power. Alternatively, it is feasible to explore a more reasonable format to store the variables, e.g. the state noise covariance matrix Q, instead of storing the full matrices.

Another solution is to combine the experimental model with real physiological models. The thesis work is based on synthetic experiment, instead of applying real MEG data on facial emotion for the evaluation due to limited research time and resources. It is signicant to apply the algorithm with more different types of real cortical stimulus, such as random spatiotemporally smooth activity spread over the cortex , the deep source inside of the cortex, single source distributed on the cortex surface. The efciency of algorithm in practice can therefore be evaluated more close to reality.

Appendix A

Matrix calculus reference

The following part indicates the reference of some matrix differentials and calculus for the matrix calculation in thesis:

$$\frac{d \log \det \mathbf{X}}{d\mathbf{X}} = \mathbf{X}^{-1} \tag{A.1}$$

$$\frac{d \operatorname{Tr}(\mathbf{AB})}{d\mathbf{A}} = \mathbf{B}^{T} \tag{A.2}$$

$$\frac{d \operatorname{Tr}(\mathbf{AXB})}{d\mathbf{X}} = \mathbf{A}^{T}\mathbf{B}^{T} \tag{A.3}$$

$$\frac{d \operatorname{In}(\det(\mathbf{X^k}))}{d\mathbf{X}} = k\mathbf{X}^{-T} \tag{A.4}$$

$$\frac{d \log \det(\mathbf{X}^{-1})}{d\mathbf{X}} = \mathbf{X}^{-T} \tag{A.5}$$

$$C \operatorname{tr}(\mathbf{X}) = \operatorname{tr}(C\mathbf{X}) \tag{A.6}$$

$$\operatorname{tr}(\mathbf{A}) = \sum_{i=1} \lambda_i \tag{A.7}$$

$$\det(\exp(\mathbf{A})) = \exp(\operatorname{tr}(\mathbf{A})) \tag{A.8}$$

$$\frac{d \exp t\mathbf{X}}{dt} = \mathbf{X} \exp t\mathbf{X} \tag{A.9}$$

$$(\beta\mathbf{A})^{-1} = \beta^{-1}\mathbf{A}^{-1} \tag{A.10}$$

$$\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA}) \tag{A.11}$$

$$\mathrm{tr}(\mathbf{ABC}) = \mathrm{tr}(\mathbf{BCA}) = \mathrm{tr}(\mathbf{CAB}) \tag{A.12}$$

$$\frac{d\,\mathrm{tr}(\mathbf{A}(\gamma)\mathbf{B})}{d\gamma} = \mathrm{tr}\,(\frac{\mathbf{A}(\gamma)}{d\gamma}\mathbf{B}) \tag{A.13}$$

Appendix B

Synthetic source generation

The synthetic source generation actually combines the methods and results from *Chapter 3 Basis Functions Source Model Applied to MEG Source Reconstruction, Chapter 4 Spatial Improvement of MEG source reconstruction with Bayesian Super- resolution* and *Chapter 5 MEG image estimation via Kalman smoother*. It has been used within simulated experiments for all these three chapters. Therefore, when the synthetic source is firstly been applied in *Chapter 3*, there are several pieces of work in the source generation process which need to be referred to.

B.1 Initial simulated source set

The first step the synthetic source generation is to create the initial synthetic source distributed on the interpolated high-resolution mesh(refer to Fig 4.1 and Fig 4.2) before expanding it in the temporal field. In our research, we generate two types of initial synthetic sources distribution, i.e. artificial source distribution and realistic source distribution. In the process of generating the former one, the fixed source values are set on 30 particular vertices of mesh we choose but the values of sources on other vertices are set as zero; while in the process of generating the later one, the source distribution on the cortical mesh are from the results of previous source reconstruction of the real MEG data with from stimuli on cortical surface at one time point. The reason of doing these different types of simulation is that we can clearly see the goodness of source reconstruction of our algorithms theoretically from the synthetic experimental result of the former source, since this type of source is based on very simple case. However, it does not describe the MEG source distribution realistically. Since for the real MEG data analysis, it is impossible that we can do the comparison of the reconstruction results by our algorithms and the known absolute correct source. Therefore, it is desirable if we could simulate the synthetic source realistic so that the difference between the real-world-like sources and the reconstruction by our algorithms can be observed.

B.1.1 Initial artificial source generation

Firstly, we define the number of the interpolated high-resolution mesh(so-called M^+) vertices as N_{HR}, while the original mesh used MEG analysis from the corresponding T1 MRI scan has N_{LR} vertices (so-called M). The number of the sensors is M_s.

Then, we try to generate a $3N_{HR} \times 1$ matrix $\tilde{\mathbf{j}}_{a0}$, the initial artificial source on the interpolated high-resolution mesh:

$$\tilde{\mathbf{j}}_0^a = \begin{pmatrix} \tilde{\mathbf{j}}_{0x}^a \\ \tilde{\mathbf{j}}_{0y}^a \\ \tilde{\mathbf{j}}_{0z}^a \end{pmatrix} \tag{B.1}$$

We determine ten particular vertices of k vertices on the cortical mesh. The single 3D source value we create on this vertex is set as $[j_{kx} \quad j_{ky} \quad j_{kz}]$. Thus, for the source value in the matrix $\tilde{\mathbf{j}}_{a0}$, the elements are set as follows and other elements are set as zeros (showed in Fig B.1, Fig B.2 and Fig B.3).

$$\tilde{\mathbf{j}}_{a0}(1:10,1) = j_{kx} \qquad \tilde{\mathbf{j}}_{a0}(k+1:k+10,1) = j_{ky} \qquad \tilde{\mathbf{j}}_{a0}(2k+1:2k+10,1) = j_{kz} \tag{B.2}$$

B.1.2 Initial realistic source generation

Additionally, as we explained above, we try to create the initial source distribution on $\mathbf{M}^+$ which is more realistic so that the difference between our reconstruction and the real-world-like source are able to be observed clearly. To achieve this aim, we apply the previous experimental result of reconstruction which based on the real MEG measurement. Specifically, within the previously research experiments on MEG source reconstruction with basis function method, super-resolution method and Kalman-smoother method which are illuminated on *Chapter 3*, *Chapter 4* and *Chapter 5*, a set of real MEG measurement with several spontaneous stimulus on the cortical surface has been applied to produce a set of reconstructed source. We randomly choose one source distribution on $\mathbf{M}^{+}$ at a single time point which can be assumed as the initial realistic synthetic source $\tilde{\mathbf{j}}_{a0}$ (as a $3N_{HR} \times 1$ matrix) (The sources showed in Fig B.4, Fig B.5 and Fig B.6).

B.2 Dynamic system generation

In our simulation, the activity consists of the output of the modeling dynamic system, which has the relation as follows:

$$\tilde{\mathbf{x}}_t = \tilde{\mathrm{A}}\tilde{\mathbf{x}}_{t-1} + \tilde{\omega} \tag{B.3}$$

$$\tilde{\mathbf{z}}_t = \tilde{\mathrm{B}}\tilde{\mathbf{x}}_t + \tilde{\upsilon} \tag{B.4}$$

As we have obtain the initial sources distributed on the high-resolution cortical mesh as above, we are using the *Kalman filter* to generate the sources in the temporal field. According to the knowledge of *Kalman filter* and *Kalman smoother* showed in *Chapter 4* and *Chapter 5*, we simulate the sources in the following way. Firstly, as showed in Eqn 5.6, we set the velocity of the dynamic system as a fixed vector $\tilde{\mathbf{v}}$. We can obtain the dynamic state $\tilde{\mathbf{x}}$ by combining $\tilde{\mathbf{v}}$ with instant source $\tilde{\mathbf{v}}$:

$$\tilde{\mathbf{x_t}} = \begin{pmatrix} \tilde{\mathbf{j}}_t \\ \tilde{\mathbf{v}}_t \end{pmatrix} \tag{B.5}$$

- Following Eqn 5.2 and Eqn 5.3, we set the state noise $\tilde{\omega}$ (matrix size is as same as $\tilde{\mathbf{x}}$, $2N_{HR} \times 1$) and the measurement noise $\tilde{\upsilon}$ (matrix size is as same as the measurement at single time point, $M_s \times 1$) are both set as zero mean multivariate

Gaussian distribution, where the covariance $\tilde{\mathbf{Q}}$ of state noise and $\tilde{\mathbf{R}}$ of measurement noise are set as fixed matrices, respectively:

$$\tilde{\omega} \sim \mathcal{N}(0, \tilde{\mathbf{R}}) \tag{B.6}$$

$$\tilde{\upsilon} \sim \mathcal{N}(0, \tilde{\mathbf{Q}}) \tag{B.7}$$

Specifically, the parameters β and α are both set as fixed value, where $\beta = 1$ and $\alpha = 0.5$ (refer to Eqn 5.17), and $\tilde{\mathbf{R}}$ are set as the identity matrix.

- Define the *leadfield* from the forward model as $M_s \times 6N_{HR}$ measurement transform matrix $\tilde{\mathbf{B}}$.
- Since the velocity of the dynamic system $\tilde{\mathbf{v}}$ is set as fixed, the $6N_{HR} \times 6N_{HR}$ state transition matrix $\tilde{\mathbf{A}}$ just gives us $\tilde{\mathbf{j}}_t = \tilde{\mathbf{j}}_{t-1} + \tilde{\mathbf{v}}_{t-1}$ and $\tilde{\mathbf{v}}_t = \tilde{\mathbf{v}}_{t-1}$, so

$$\tilde{\mathbf{A}} = \begin{pmatrix} \mathbf{I} & \mathbf{I} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \tag{B.8}$$

  where $\mathbf{I}$ are the identity matrix with the size $3N_{HR} \times 3N_{HR}$.

- Combining with the initial source set introduced previously, as well as all other items set above, Kalman filter is used to obtain the generate the matrix $\tilde{\mathbf{b}}$ (with the size $M_s \times T$) as follows which contains a set of measurement of magnetic field of the synthetic source in the sampled time point T (Welch and Bishop, 2006):

$$\tilde{\mathbf{x}}_t^- = \tilde{\mathbf{A}}\tilde{\mathbf{x}}_{t-1} \tag{B.9}$$

$$\tilde{\mathbf{P}}_t^- = \tilde{\mathbf{A}}\tilde{\mathbf{P}}_{t-1}\tilde{\mathbf{A}}^T + \tilde{\mathbf{Q}} \tag{B.10}$$

$$\tilde{\mathbf{K}}_t = \tilde{\mathbf{P}}_t^- \tilde{\mathbf{B}}^T (\tilde{\mathbf{B}}\tilde{\mathbf{P}}_t^- \tilde{\mathbf{B}}^T + \tilde{\mathbf{R}})^{-1} \tag{B.11}$$

$$\tilde{\mathbf{x}}_t = \tilde{\mathbf{x}}_t^- + \tilde{\mathbf{K}}_t(\tilde{\mathbf{z}}_t - \tilde{\mathbf{B}}\tilde{\mathbf{x}}_t^-) \tag{B.12}$$

$$\tilde{\mathbf{P}}_t = (\tilde{\mathbf{I}} - \tilde{\mathbf{K}}_t\tilde{\mathbf{B}})\mathbf{P}_t^- \tag{B.13}$$

B.3 Mesh downsampling

With respect to the synthetic sources generation on the interpolated high-resolution mesh, the source simulation on the corresponding $\mathbf{M}$ can be also produced by the product of source on the $\mathbf{M}^+$ and a particular downsampling matrix $\tilde{\mathbf{P}}$. In terms of the structure of interpolated high-resolution mesh indicated in Eqn 4.2, the $3N_{HR}times1$ matrix, the source $\tilde{\mathbf{j}}_k$ distributed on $\mathbf{M}^+$ at time point k, can be rewrite as a $N_{HR} * 3$ matrix $\tilde{\mathbf{j}}_k'$ with the different dimensional components on each column. Therefore, the corresponding source distributed on mesh $\mathbf{M}$ can be down-sampled by the following equation:

$$\tilde{\mathbf{j}}_k^{LR'} = \tilde{\mathbf{P}}\ \tilde{\mathbf{j}}_k' \tag{B.14}$$

where

$$\tilde{\mathbf{P}} = [\mathbf{I}_M \quad \mathbf{0}] \tag{B.15}$$

then $\tilde{\mathbf{j}}_k^{LR'}$ is transformed back to a $N_{LR} \times 1$ matrix $\tilde{\mathbf{j}}_k^{LR}$. Thus, the $\tilde{\mathbf{j}}_k^{LR}$ can be distributed on the mesh $\mathbf{M}$ and get the color-map as Fig B.2, Fig B.5.

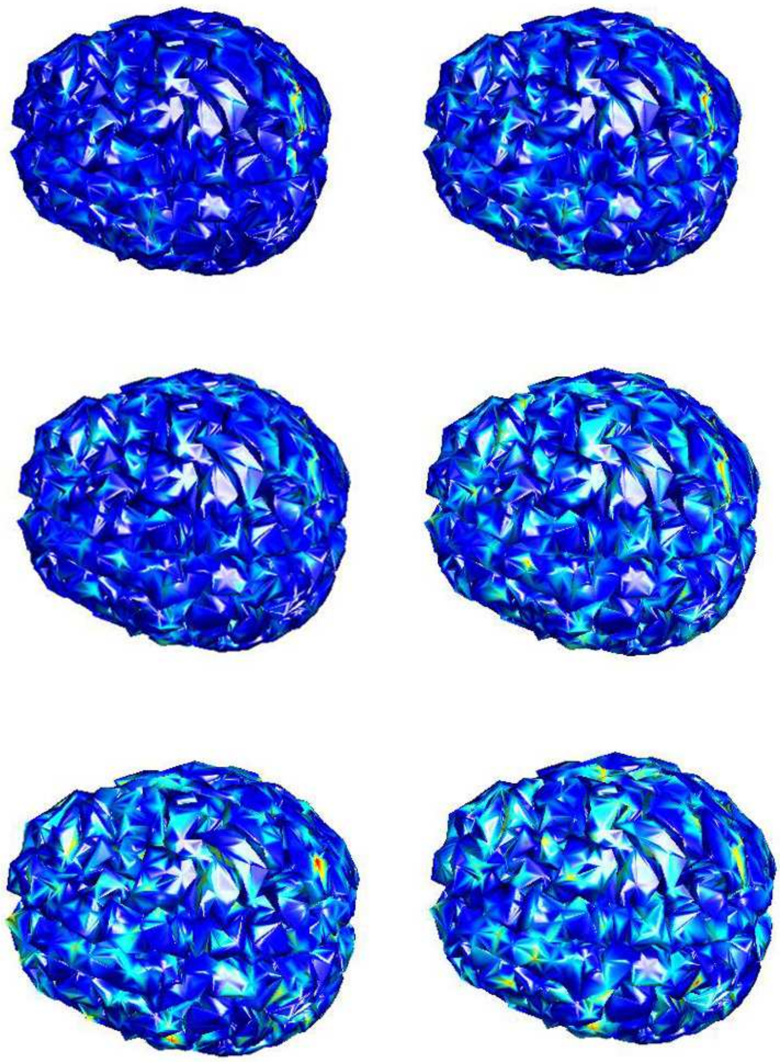

Figure B.1: Simulated data of *artificial source generation*, spatial visualization on high-resolution mesh $\mathbf{M}^{+}$. The cortical surfaces display the data at different instants: time point 1 (up˙left), 25 (up-right), 50 (middle-left), 100 (middle-right), 150 (bottom-left) and 200 (bottom-right), respectively. In this trail, only 30 currents are put artificially on 30 continuous vertices on right hemisphere. Then, the noise with multivariable Gaussian distribution are added in time course.

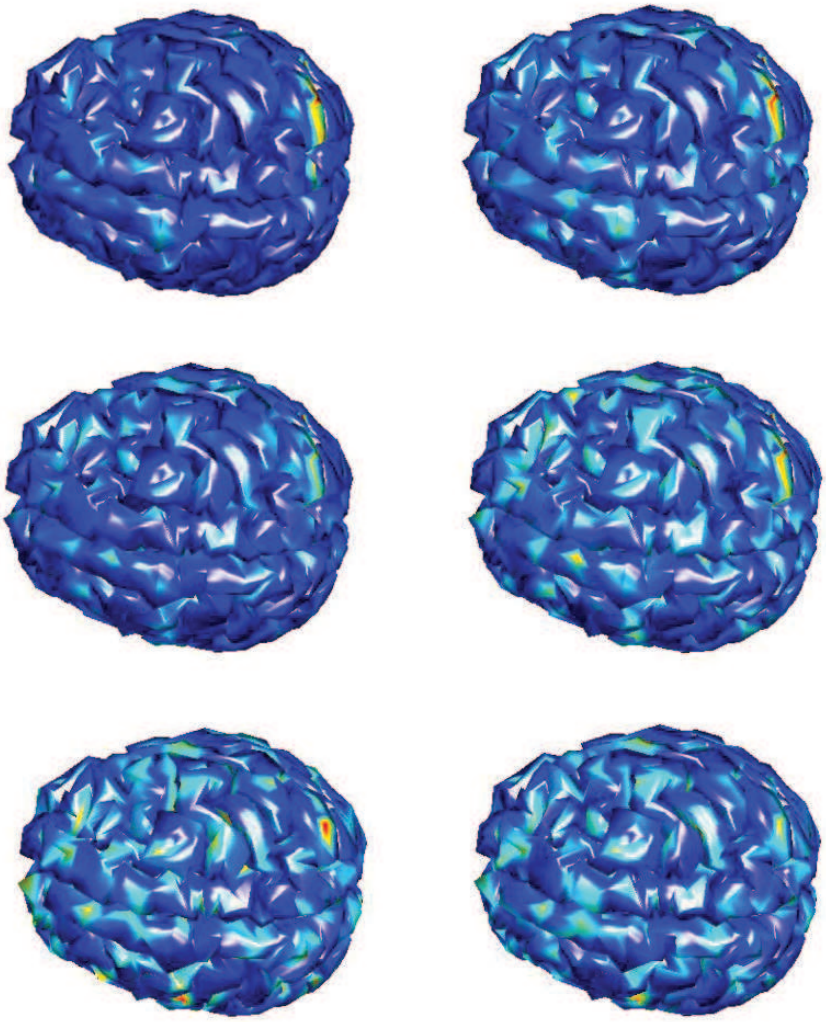

Figure B.2: Simulated data of *artificial source generation*, spatial visualization on low-resolution mesh **M**. The cortical surfaces display the data at different instants: time point 1 (up-left), 25 (up-right), 50 (middle-left), 100 (middle-right), 150 (bottom-left) and 200 (bottom-right), respectively. In this trail, only 30 currents are put artificially on 30 continuous vertices on right hemisphere. Then, the noise with multivariable Gaussian distribution are added in time course.

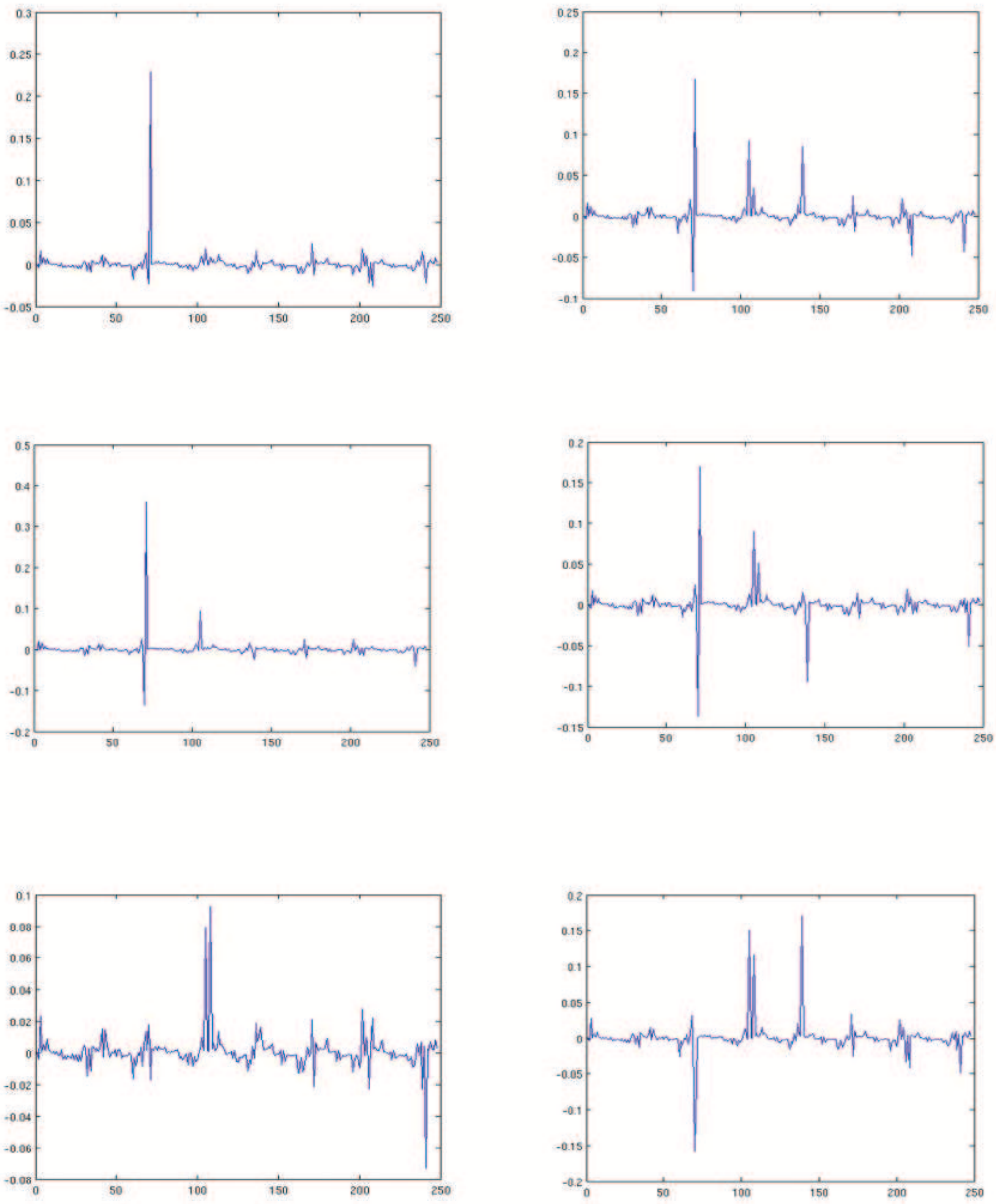

Figure B.3: Simulated data of *artificial source generation* , temporal visualization on 248 sensors for time point 1(up-left), 25(up-right), 50(middle-left), 100(middle-right), 150(bottom-left) and 200(bottom-right), respectively. X-axis shows the number of sensors, Y-axis shows the amplitude of the magnetic field.

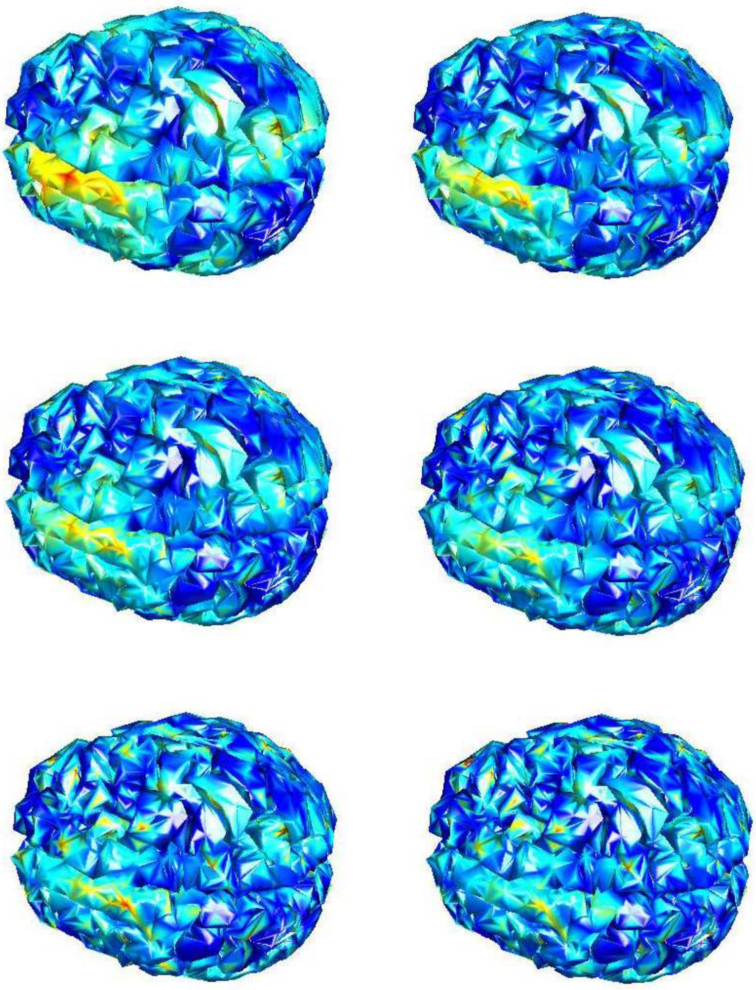

Figure B.4: Simulated data of *realistic source generation*, spatial visualization on high-resolution mesh $\mathbf{M}^{+}$. The cortical surfaces display the data at different instants: time point 25 (up-left), 50 (up-right), 75 (middle-left), 100 (middle-right), 150 (bottom-left) and 200 (bottom-right), respectively. In this trial, we use the *realistic source generation* explained beforehand. Then, the noise with multivariable Gaussian distribution are added in time course.

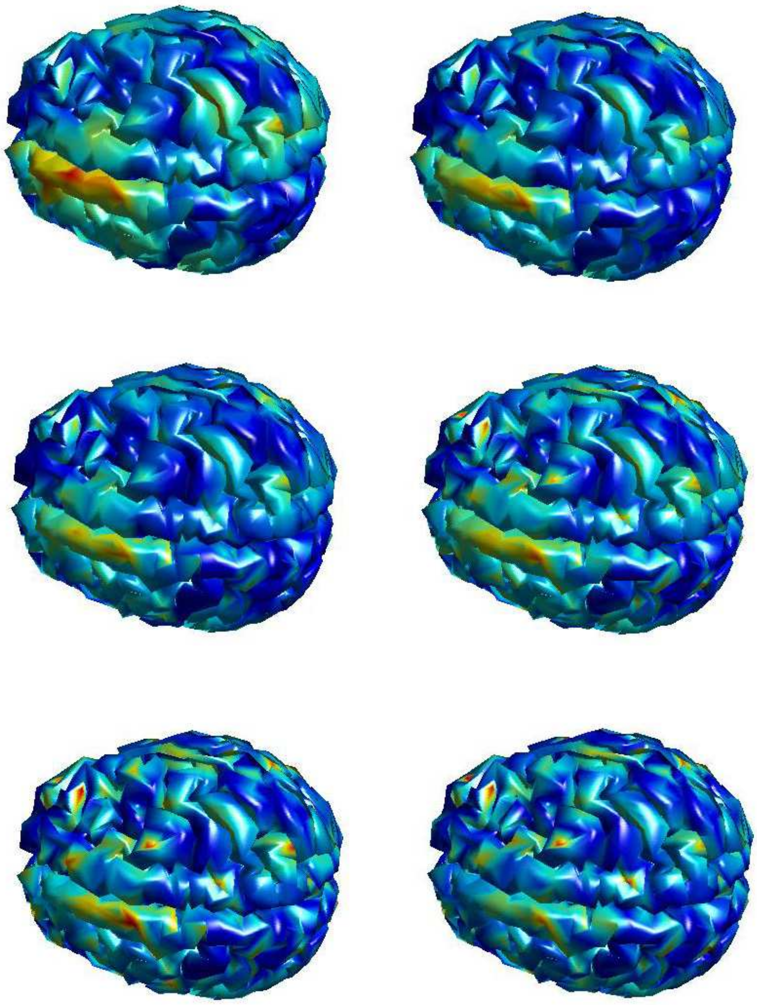

Figure B.5: Simulated data of *realistic source generation*, spatial visualization on low-resolution mesh **M**. The cortical surfaces display the data at different instants: time point 25 (up˙left), 50 (up-right), 75 (middle-left), 100 (middle-right), 150 (bottom-left) and 200 (bottom-right), respectively. In this trail, we use the *realistic source generation* explained beforehand. Then, the noise with multivariable Gaussian distribution are added in time course.

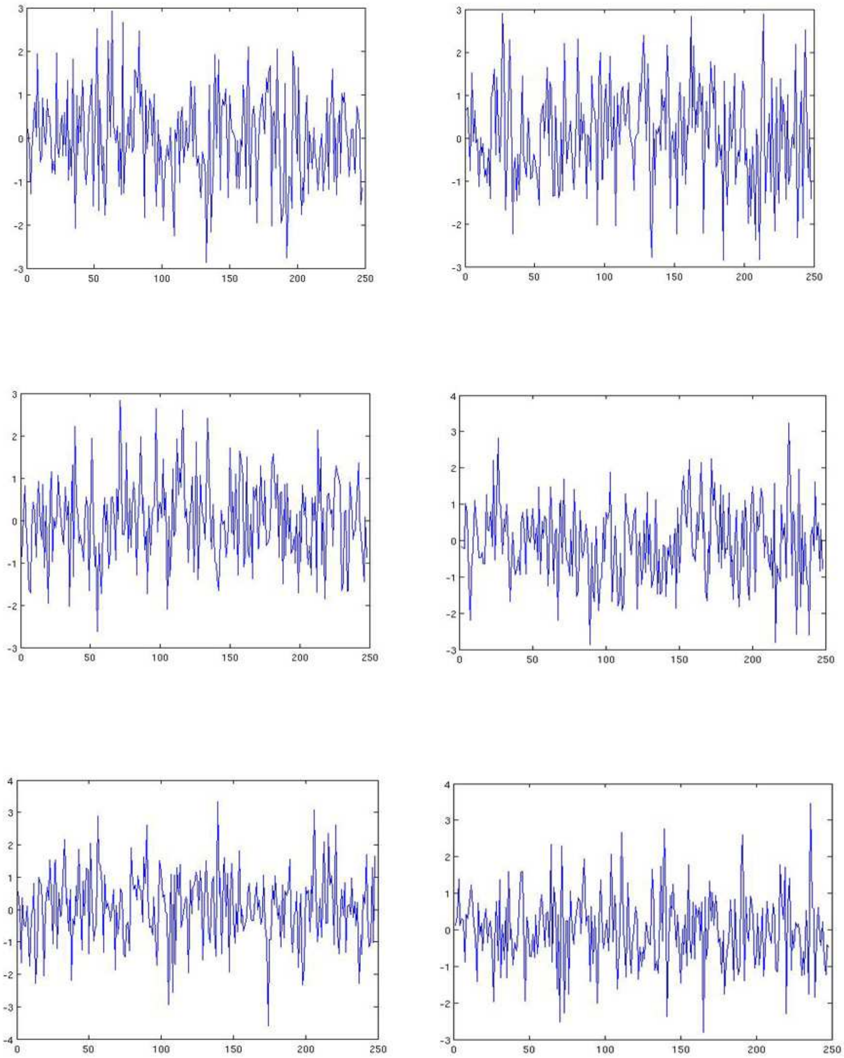

Figure B.6: Simulated data of *realistic source generation* , temporal visualization on 248 sensors for time point 25(up-left), 50(up-right), 75(middle-left), 100(middle-right), 150(bottom-left) and 200(bottom-right), respectively. X-axis shows the number of sensors, Y-axis shows the amplitude of the magnetic field.

Appendix C

MEG real data acquisition

The set of MEG measurement was kindly provided by the York Neuroimaging Centre (YNiC). Participants were shown emotionally congruent fear and minimally congruent neutral face stimuli, and responses were recorded as described by (Hagan et al., 2009).

C.1 The participants of the experiment

Twenty-eight healthy participants were recruited from the University of York and they were offered a stipend for participation in the study. Ethical approval was granted jointly by the Department of Psychology at the University of York and the York Neuroimaging Centre. All participants are right-handed, and with normal hearing and normal or corrected-to-normal vision, and were without a history of neurological injuries. Nine sets of data were excluded due to scanner problems, excessive head movements, or electrical noise in the background, which leaves nineteen data sets in the final data analysis (Males: 10, Females: 9, mean age: 24.44 (SD 4.23) years, range $19.22 \sim 33.41$ years).

C.2 Introduction of the MEG experiment

Visual stimuli of fearful and neutral facial expressions used in this study were selected from Ekman and Friesen Facial effect series (Young et al., 2002). The expressions came from two actors and two actresses whom were represented as JJ, WF, MF, and SW. All faces were presented in grey scale and hair was removed from each face so that the contrast differences between stimuli were minimized. In order to produce more discernible fearful stimuli, facial expressions were processed in some cases, for example, the female fear faces (MF) were half caricatured

To attempt to maintain the central fixation during each trial, the participants were asked to attend to the face experiments of all four actors and actresses. We included 96 additional trials in which a response (such as raising the finger) was requested from the participant. On such occasions, either the letter B or the letter R appeared in the centre of the screen directly after stimulus offset. The letter remained on the screen for 250 ms and was followed by a solid gray screen for $1,050$ ms directly after the letter offset. Response trials were followed by a dummy trial (16 for both conditions, pseudorandomly chosen and counterbalanced between conditions). Dummy trials were discarded in the overall analysis of data because of potential motor response contamination Cindy (Hagan et al., 2009).

All trials began with a black fixation cross (3×3 cm) in the centre of the screen against a solid grey background, which was presented for 500 ms. Next a visual stimulus appeared for 700 ms, immediately followed by the stimulus offset, a solid grey screen, which lasted $1,300$ ms.

C.3 MEG Data Acquisition

MEG data were acquired at the York Neuroimaging Centre using a 248-channel Magnes 3600 whole-scalp recording system (4-D Neuroimaging) with superconducting quantum interference device-based first-order magnetometer sensors. A Polhemus stylus digitizer (Polhemus Isotrak) was used to digitize the head, nose and eye orbit shapes of each participant before data acquisition to facilitate accurate co-registration with MRI data. Coils were placed in front of the left and right ears and at three equally spaced locations across the forehead to monitor head position prior to and following data acquisition. Data from four participants with head movement values of 0.75 cm or greater at two or more coils were excluded from both sensor and source-space analyses. Participants were seated during the experiment. Magnetic brain activity was digitized continuously in all three runs.

Images were projected onto a screen at a viewing distance of ≈ 70 cm and subtending a viewing angle of 8 degrees for face stimulus and 0.3 degrees for letters. Faces were presented in small size (5×9 cm) to help minimize participant eye saccades. During response trials, the letters displayed were $5mm$ cm in size to ensure that the central fixation was maintained throughout all stimulus presentations. Auditory stimuli were presented at a comfortably audible level via Etymotic Research ER30 earphones. Participants were monitored throughout the scan using a video camera situated in a magnetic shielded room.

All data were filtered using an online direct current (DC) filter and were sampled at a rate of 678.17 Hz (bandwidth 200 Hz). Standard structural MRI scans were obtained for co-registration with MEG. Images were acquired using a 3-T scanner (HD Excite; General Electric) with a whole-head coil (8-channel high T-resolution brain array). The scanner uses a 3-T 60-cm magnet. To maximize magnetic field homogeneity, an automatic shim was applied before scanning. In order to cover the whole brain, 176 parallel 1-mm 3-D sagittal planes were imaged, using an IR-prepared fast spoiled gradient recalled pulse sequence (repetition time 6.6 ms, echo time 2.8 ms, flip angle 20, and an inversion time of 450 ms). The field of view was 290×290 mm, and the matrix size was 256×256, which results in an in-plane spatial resolution of 1.13 mm. Localizer and calibration scans were performed before performing a high resolution T1 volume with voxel dimensions of $1 \times 1.13 \times 1.13$ mm. For better elimination of distortion and improved co-registration of MRI and MEG data, 3-D gradient warping corrections and edge-enhancement filters were applied. The data pattern is showed in Fig C.1.

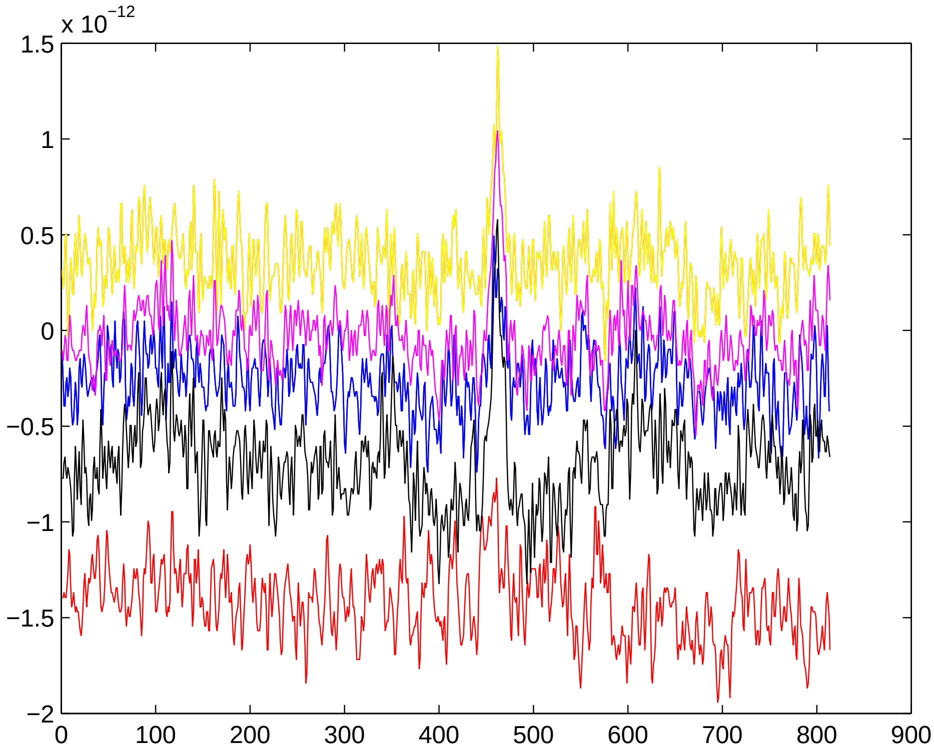

Figure C.1: This figure shows the example of the measurement of magnetic field from MEG for the real visual expression data. The original MEG data we obtained is a $96 \times 248 \times 813$ matrix which indicates *the MEG measurement on 248 sensors for 96 different stimulus within 813 continuous time instants* . This is a overlapping pattern for the time courses of measurement on particular single sensor(here we choose sensor 30, 60, 90, 120 and 150) for the stimulus 3.

References

Ramon y Cajal 1888 from Hamalainen et al. Reviews of Modern Physics, 1993. x, 13

M. Irani and S. Peleg. Improving resolution by image registration. *Graphical Models and Image Provessing*, 53:231–239, 1991. 53, 54, 55

S. Peleg, D. Keren, and L. Schweitzer. Improving image resolution using subpixel motion. *Pattern Recognition Letters*, 5:223–226, 1987. 54

Freesurfer 5.0.0. (http://surfer.nmr.mgh.harvard.edu/). Technical report, Martinos Center for Biomedical Imaging. 64, 76, 88, 118, 153

Ralph W. Adams, Juan A. Aguilar, Kevin D. Atkinson, Michael J. Cowley, Paul I. P. Elliott, Simon B. Duckett, Gary G. R. Green, Iman G. Khazal, Joaqułn Lpez-Serrano, and David C. Williamson. Reversible interactions with para-hydrogen enhance nmr sensitivity by polarization transfer. *Science*, 323:1708 – 1711, March, 2009. 6

Dale AM., AK. Liu, BR. Fischl, RL. Buckner, JW. Belliveau, JD. Lewine, and E. Halgren. Dynamic statistical parametric mapping: combining fmri and meg for high-resolution imaging of cortical activity. *Neuron*, 26:55–67, 2000. 36

D.G. Amen, C. Hanks, and J . Prunella. Predicting positive and negative treatment responses to stimulants with brain spect imaging. *Journal of psychoactive drugs*, 40 (2):131, 2008. 8

M. Dale. Anders, Martin I. Sereno, and J. Cogn. Improved localization of cortical activity by combining eeg and meg with mri cortical surface reconstruction: A linear approach. *Neurosci*, 5:162 – 175, 1993. 37

Sylvain. Baillet, John C. Mosher, and Richard M. Leahy. Electromagnetic brain mapping. *Signal Processing Magazine, IEEE,*, 18:14–30, 2001. 9, 10, 13, 14, 21, 23, 25

P.A. Bandettini, A. Jesmanowicz, E.C. Wong, and J.S. Hyde. Processing strategies for time-course data sets in functional mri of the human brain. *Magnetic Resonance in Medicine*, 30 (2):161–173, 1993. 7

D. Barry, B. Van Veen, and Kevin. M.. Buckley. Beamforming: A versatile approach to spatial filtering. *IEEE Signal Processing Magazine*, 5(2):4–24, 1988. 38, 57

D.S. Barth, W.W. Sutherling, and J. Beatty. Intracellular currents of interictal penicillin spikes: evidence from neuromagnetic mapping. *Brain Res,*, 368:36–48., 1986. 10, 13, 25, 60, 173

B. Baryshnikov, B.D. Van Veen, and R.T. Wakai. Maximum likelihood estimation of low rank signals for multiepoch meg/eeg analysis. *IEEE Transactions on Biomedical Engineering*, 51:1981–1993, 2004. 14, 57

G.M. Baule and R. McFee. Detection of the magnetic field of the heart. *American Heart Journal*, 66:95–96., 1963. 9

The Beatles. *"The Beatles greatest gift... is to science". Whittington Hospital NHS Trust. Retrieved 2007-05-07.*, 2005. 5

M. Belkin and P. Niyogi. Laplacian eigenmaps for dimentionality reduction and data representation. *Neural Comput*, 15(6):1373–1396, 2003. 83, 95

Jeff. A. Bilmes. A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. Technical report, International Computer Science Institute and Computer Science Division, Department of Electrical Engineering and Computer Science, U.C. Berkeley, 1998. 49, 51

D. Brenner, C. Elliston, E. Hall, and W. Berdon. Estimated risks of radiation-induced fatal cancer from pediatric ct . *American Journal of Roentgenology .*, 176:289–296, 2001. 6

R. G. Brown and P. Y. C. Hwang. *Introduction to Random Signals and Applied Kalman Filtering, Second Edition*. John Wiley & Sons, Inc., 1992. 41, 44

A. D. Capel and ZI. Sserman. Super-resolution enhancement of text image sequences. *15th International Conference on Pattern Recognition, 2000. Proceedings.*,, 1:600–605, 2000. 55

Ed. Charles Hodgman. *CRC Handbook of Chemistry and Physics, 44th Ed.*. USA: Chemical Rubber Co., 1961. 5

F.R.K. Chung. *Spectral Graph Theory*. American Mathematical Society, 1997. 66, 102

D. Cohen. Magnetoencephalography: Evidence of magnetic field produced by alpha rhythm current. *Science*, 161:784–786, 1968. 14, 21

D. Cohen. Magnetoencephalography: Detection of the brain's electrical activity with a superconducting magnetometer. *Science*, 175:664–666, 1972. 21

D. Cohen and B.N. Cuffin. Demonstration of useful differences between the magnetoencephalogram and electroencephalogram. *Electroencephalogr Clin Neurophysiol*, 56:38–51, 1983. 10

BN. Cuffin. A comparison of moving dipole inverse solutions using eeg's and meg's. *IEEE Transactions on Biomedical Engineering*, 32:905–910, 1985. 29

D. Cvetkovic, M. Doob, and H. Sachs. *Spectra of Graphs: Theory and Applications*. Academic Press, 1997. 66, 102

A. M. Dale and M. I. Sereno. Improved localization of cortical activity by combining eeg and meg with mri cortical surface reconstruction: a linear approach. *J. Cogn. Neurosci.*, 5:162 – 176., 1993. 10

A. M. Dale, 1. B. Fischl, and M. I. Sereno. Cortical surface-based analysis: I. segmentation and surface reconstruction. *NeuroImage*, 9:179–194, 1999. 64

Tom Dalton. Bayesian image superresolution - meng project. Master's thesis, University of York, UK, 2004. 52

F. Darvas, D. Pantazis, E. Kucukaltun-Yildirim, and R.M. Leahy. Mapping human brain function with meg and eeg: methods and validation. *NeuroImage*, 23:S289–S299., 2004. 10

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, 39(1):1–38, 1977. 48

T. Deneux and O. Faugeras. Eeg-fmri fusion of paradigm-free activity using kalman filtering. *Neural Comput.*, 22(4):906–948., 2010. 137, 142

Webster Online Dictionary. ”computed tomographyłdefinition from the merriam-webster online dictionary”. retrieved 2009-08-18., 08 2009. 5

Drake et al. *Gray's Anatomy for Students*. 2005. 7

B. Fischl and A. M. Dale. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc Natl Acad Sci U S A*, 20:11050–5, 2000. 64

B. Fischl, M. I. Sereno, and A. M. Dale. Cortical surface-based analysis: Ii: Inflation, flattening, and a surface-based coordinate systems. *NeuroImage*, 9:195–207, 1999. 64

Slifstein Mark. Frankle, W. Gordon, Peter S. Talbot, and Marc. Laruelle. Neuroreceptor imaging in psychiatry: Theory and applications. *International Review of Neurobiology,*, 67:385 440, 2005. 8

L. Fratt. Radiation testing and pet minding the radiopharmaceutical store. *Medical Imaging*, 2003. 7

P. Gabriel. Numerical mesh processing. Technical report, 2007. 60, 66

Zoubin. Ghahramani and Geoffrey E. Hinton. Parameter estimation for linear dynamical systems. Technical report, University of Toronto, 1996. 145, 146

American government. 11th report on carcinogens. 2005. 5

M. S. Grewal and A. P. Andrews. *Kalman Filtering Theory and Practice.* Upper Saddle River, NJ USA, Prentice Hall., 1993. 41

D. Gross. Super-resolution from subpixel sifted pictures. Master's thesis, Tel-Aviv University, 1986. 53

M. Grote and M. Hagemann. Spai-3.2. Technical report, 1. Department of Mathematics, University of Basel; 2. Department of Computer Science, University of Basel. 111, 125, 126, 135

E. M. Haacke, R. F. Brown, M. Thompson, and R Venkatesan. *Magnetic resonance imaging: Physical principles and sequence design.* New York: J. Wiley & Sons. ISBN 0-471-35128-8., 1999. 4, 6

C. C. Hagan, W. Woods, S. Johnson, A. J. Calder, G. G. R. Green, and A. W. Young. Meg demonstrates a supra-additive response to facial and vocal emotion in the right superior temporal sulcus. *Psychological and Cognitive Sciences, Neuroscience,* 106(47):20010–20015, 2009. 92, 126, 188, 189

J. V. Hajnal, D. J. Hawkes, and D. L. Hill. *Medical Image Registration*. CRC Press, 2001. 4

M. Hämäläinen, R. Hari, R. J. Ilmoniemi, J. Knuutila, and O. V. Lounasmaa. Magnetoencephalography-theory, instrumentation, and application to noninvasive studies of the working human brain. *Reviews of Modern Pysics*, 65:2, 1993. x, 10, 12, 14, 19, 30, 32, 33, 34, 35

M.S. Hamalainen and J. Sarvas. Feasibility of the homogeneous head model in the interpretation of neuromagnetic fields. *Physics in Medicine and Biology*, 32:91–98, 1987. 24

K. A. E. Hardie and R. C. Barnard. Joint map registration and high-resolution image estimation using a sequence of undersampled images. *IEEE Transactions on Image Processing,, 6(12)*, 6(12):1621–1633, 1997. 55

Reinikainen K Hari, R, E Kaukoranta, M. Hämäläinen, R. Ilmoniemi, A. Penttinen, J. Salminen, and D. Teszner. Somatosensory evoked cerebral magnetic fields from si and sii in man. *Electroencephalogr Clin Neurophysio*, 57:254–63, 1984. 31

Olaf Hauk. Keep it simple: a case for using classical minimum norm estimation in the analysis of eeg and meg data. *NeuroImage*, 21:1612–1621, 2004. 35, 37, 76, 77

G. T. Herman. *Fundamentals of computerized tomography: Image reconstruction from projection, 2nd edition*. Springer, 2009. 5

A. Hillebrand, KD. Singh, IE. Holliday, PL. Furlong, and GR. Barnes. A new approach to neuroimaging with magnetoencephalography. *Human Brain Mapping*, 25(2):199–211, 2005. x, 38, 39

T.S. Huang and R.Y. Tsai. Multi-frame image restoration and registration. *in Advances in Computer Vision and Image Processing*, 1:317–399, 1984. 53

O.L.R. Jacobs. *Introduction to Control Theory, 2nd Edition.* Oxford University Press, 1993. 44

A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. (New York: Academic Press), 1970,. 47, 142

R. E. Kalman. A new approach to linearfilteringand prediction problems. *Transactions of the ASME-Journal of Basic Engineering*, 82 (Series D):35–45, 1960. 40, 47, 142

Masaki Kawakatsu. Application of ica to meg noise reduction. *4th International Symposium on independent component analysis and blind signal separation (ICA2003)*, 2003. 30

D. Keren, S. Peleg, and R. Brada. Image sequence enhancement using sub-pixel displacements. *In IEEE Conference on Computer Vision and Pattern Recognition*, pages 742–746, 1988. 53, 54

Kuniharu. Kishida. Evoked magnetic fields of magnetoencephalography and their statistical property. *PHYSICAL REVIEW*, 79:011922, 2009. 14, 57

S. Laureys, M. Boly, and G. Tononi. The neurology of consciousness: Cognitive neuroscience and neuropathology. *Academic Press-Elsevier*, pages 31–42, 2009. 7

NK. Logothetis, J. Pauls, M. Augath, T. Trinath, and A. Oeltermann. Neurophysiological investigation of the basis of the fmri signal. *Nature*, 412 (6843):150–157, 2001. 7

Matlab. Matlab7.9.0(r2009b). Technical report, http://www.mathworks.com/, 2009. 94

Peter S Maybeck. *Stochastic Models, Estimation, and Control, Volume 1*. Academic Press, Inc, 1979. 44

G. J. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. Wiley-Interscience; 1 edition, 1996. 49, 51

P Menendez, G.d, R Gonzalez, and S. Andino. Distributed source models: standard solutions and new developments. *Analysis of Neurophysiological Brain Functioning.*, pages 176 – 201, 1998. 77

MNE. Mne software user guide version 2.7. Technical report. 76, 88, 118, 156

JC. Mosher and RM. Leahy. Source localization using recursively applied and projected (rap)music. *IEEE Transactions on Signal Processing Magazine, IEEE,*, 47: 332–340, 1999. 34

John C. Mosher, Richard. M. Leahy, and Paul S. Lewis. Eeg and meg: forward solutions for inverse methods. *IEEE transactions on biomadical engineering*, 46:NO:3, 1999. 10, 27

T. Nara, J. Oohama, and S. Ando. Super-resolution for meg inversion -reconstruction from the partial boundary measurement. *SICE-ICASE, 2006. International Joint Conference*, pages 3543–3547, 2006. 97

Roxanne Nelson. Thousands of new cancers predicted due to increased use of ct. *Medscape*, 2009. 6

N. Nguyen, P. Milanfar, and G. Golub. A computationally efficient superresolution image reconstruction algorithm. *IEEE Transactions on Image Processing*, 10(4): 573–583, 2001. 55

David. Olivier, Garnero. Line, and J. Varela. Francisco. A new approach to the meg/eeg inverse problem for the recovery of cortical phase-synchrony. *IPMI 2001*, pages 272–285, 2001. 138

N. ´Ozmen, F. Yerlikaya, and D Gärsoy. The inverse problem of magnetoencephalography: Source localization and the shape of a ball. *SIAM News*, Volume 40, Number 2, 2007. 14, 28

Andrew C. Papanicolaou. *Clinical magnetoencephalography and magnetic source imaging*. Cambridge university press, 2009. 4, 12, 14, 19, 20, 21, 23, 24, 25, 28, 29, 34, 35, 38, 39

P. Partha and H.Maniar Mitra. Local basis expansions for meg source localization. *IJBEM*, 7(2):30–33, 2005. 68

C. Phillips. *Source estimation in EEG*. PhD thesis, University de Liege, Belgium,, 2000. 12, 14

C. Phillips, M.D. Rugg, and K.J. Friston. Anatomically informed basis functions for eeg source localisation: Combining functional and anatomical constraints. *NeuroImage*, 16:678– 695, 2002. 37

R. Plonsey. *Bioeletric Phenomena*. New York: McGraw Hill, 1969. 24

Hubert Preissl, editor. *Magnetoencephalography*. Academic Press, 2005. 1, 4, 14, 25, 28, 38, 57, 59, 171

Remesh. http://remesh.sourceforge.net/features.html. Technical report, Remsh2.0, 2008. 88

A. Rodriguez, B. Van Veen, and R. Wakai. Statistical performance analysis of signal variance based dipole models for meg/eeg source localization and detection. *IEEE Transactions on Biomedical Engineering*, 50:137–149, 2003. 14, 57

Yousef Saad. *Numerical Methods for Large Eigenvalue Problems*. Manchester University Press, 1992. 70

S. Sarkar and K. Boyer. Quantitative measures of change based on feature orgnization: eigenvalues and eigenvectors. *Proceeding of the IEEE conference on computer vision and pattern recognition*, page 478, 1996. 83

M. Scherg and D. von Cramon. Evoked dipole source potentials of the human auditory cortex. *Electroencephalogr Clin Neurophisiol*, 65:344–360, 1986. 29

H.Paul . Schimpf, Ceon . Ramon, and Jens. Haueisen. Dipole models for the eeg and meg. *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, 49:5, May 2002. 24, 25

Avinash. Sharma, Radu. Horaud, David. Knossow, and Etienne von. Lavante. Mesh segmentation using laplacian eigenvectors and gaussian mixtures. *Fall Symposium on Manifold Learning and its Applications*, pages 50–56., 2009. 95

Jianbo. Shi and Jitendra. Malik. Normalized cuts and image segmentation. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 22: 888–905, 2000. 83, 85

R.H. Shumway and D.S. Stoffer. An approach to time series smoothing and forecasting using the em algorithm. *Journal of Time Series Analysis*, 3:253–264, 1982. 145, 146, 147

R.H. Shumway and D.S. Stoffer. Dynamic linear models with switching. *Journal of the American Statistical Association*, 86:763–769., 1992. 145, 146

Merrill. Simon and James. Mattson. *The pioneers of NMR and magnetic resonance in medicine: The story of MRI.* Ramat Gan, Israel: Bar-Ilan University Press. ISBN 0-9619243-1-4., 1996. 4, 6

K.D. Singh, G.R. Barnes, A. Hillebrand, E.M. Forde, and A.L. Williams. Task-related changes in cortical synchronization are spatially coincident with the hemodynamic response. *Neuroimage*, 16(1):103–114, 2002. 38

V. N. Smelyanskiy, D. A. Maluf P. Cheeseman, and R. D. Morris. Bayesian super-resolved surface reconstruction from images. *IEEE Conference on Computer Vision and Pattern Recognition, 2000. Proceedings.*, 1:375–382, 2000. 55

H. W. Sorenson. Least-squares estimation: from gauss to kalman. *IEEE Spectrum,*, vol. 7,:63–68, 1970. 41

LF. Squire and RA. Novelline. *Squire's fundamentals of radiology (5th ed.)*. Harvard University Press. ISBN 0-674-83339-2., 1997. 5, 6

S. Nagarajan. Srikantan, T. Hagai, and K. E. H. K. S. Attias. Graphical model for estimating stimulus-evoked brain responses from magnetoencephalography data with large background brain activity neuroimage. *NeuroImage*, 30(no.2):400–16, 2006. 14, 57

MP. Tarvainen, JK. Hiltunen, PO. Ranta-aho, and PA. Karjalainen. Estimation of nonstationary eeg with kalman smoother approach: an application to event-related synchronization (ers). *IEEE Trans Biomed Eng*, 51(3):516–524., 2004. 136, 138

M.M. Ter-Pogossian, M.E. Phelps, E.J. Hoffman, and N.A. Mullani . A positron-emission transaxial tomograph for nuclear imaging (pet). *Radiology*, 114 (1):89–98., 1975. 8

E. Tipping, Michael and Christopher M. Bishop. Bayesian image super-resolution. *Machine Learning for Signal Processing, 2007 IEEE Workshop*, 27-29:181–186, 2007. 51, 52, 53, 55, 107

M. E. Tipping and C. M. Bishop. Bayesian image super-resolution. *Advances in Neural Information Processing Systems*, pages 1303–1310, 2003. 55, 56, 96, 97, 103, 106

Tomography. *"tomography - Definition from the Merriam-Webster Online Dictionary". http://www.merriam-webster.com/dictionary/tomography. Retrieved 2009-08-12*. 5

J Tripp. *Physical concepts and mathematical models, in Biomagnetism: An Interdisciplinary Approach*. New York: Plenum,, 1983. 10, 12

T. Tuomisto, R. Hari, T. Katila, T. Poutanen, and T. Varpula. Studies of auditory evoked magnetic and electric responses: modality specificity and modelling. *Nuovo Cimento D 2*, 2:471–483, 1982. 28

J.K. Udupa and G. T. Herman. *3D Imaging in Medicine, 2nd Edition*. CRC Press, 2000. 4, 5, 6

M. Wagner, M. Fush, HA. Wischanman, K. ottenberg, and O. Doessel. Smooth reconstruction of cortical source from eeg or meg recording. *NeuroImaging*, 3:1996, 1996. 37

Greg Welch and Gary. Bishop. An introduction to the kalman filter. 2006. 41, 42, 43, 99, 137, 180

Max Welling. The kalman filter. Technical report, California Institute of Technology. 47, 48, 137, 142, 143, 145, 147

Katrina. Wendel, Outi. Väisänen, Jaakko. Malmivuo, Nevzat G. Gencer, Bart. Vanrumste, Piotr. Durka, Ratko. Magjarevic, Selma. Supek, Mihail. Lucian. Pascu, Hugues. Fontenelle, de. Peralta. Grave, and Rolando. Menendez. Eeg/meg source imaging: Methods, challenges, and open issues. *Computational intelligence and neuroscience*, pages 1687–5265, 2009. 9

Eric. Whaites and Cawson. Roderick. Essentials of dental radiography and radiology. *Elsevier Health Sciences.*, pages 15–20., 2002. 5

C.H. Wolters. The finite element method is eeg/meg source analysis. *SIAM News*, 2: 2007, 40. 25

Z. Wu and R. Leahy. An optimal graph theoretic approach to data clustering: theory and its application to image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 15(11):1101–1113, 1993. 84

Yokogawa. http://www.yokogawa.com/me/index.htm. Technical report, Yokogawa Electric Corporation, 2009. 4

A. Young, D. Perrett, A. Calder, R. Sprengelmeyer, and P Ekman. *Facial Expressions of Emotion: Stimuli and Tests (FEEST)*, 2002. 189

H. Young, R. Baum, and U. Cremerius. Measurement of clinical and subclinical tumour response using [18f]-fluorodeoxyglucose and positron emission tomography:

review and 1999 eortc recommendations. *European Journal of Cancer*, 35(13): 1773–1782., 1999. 8