

Imaging Task-related Neural Activity by MEG Decoding

Submitted in partial fulfillment of the requirements for

the degree of

Doctor of Philosophy

in

Electrical and Computer Engineering

Jinyin Zhang

Carnegie Mellon University
Pittsburgh, PA

August, 2013

Abstract

Recent advances in Magnetoencephalography (MEG) provide a significant new approach to study neural activity in humans. In many MEG studies different brain states are considered, which differ in the effect of an interested mental process only. These brain states are compared to determine significant neural activity associated with the interested mental process, which is referred to as task-related neural activity.

To investigate the task-related activity using MEG, we should first check whether this activity is detectable in MEG recordings. We put forward that MEG decoding can be used as an instrument to address this problem. If high decoding accuracy is achieved, strong task-related activity exists. However, low accuracy might be because of the poor performance of decoding algorithms. Thus it is important to develop high performance decoding algorithms to efficiently extract the discriminant information between brain states. In this thesis, we propose a clustering linear discriminant analysis (CLDA) algorithm for high performance MEG decoding based on small training sets.

MEG decoding tells whether the task-related activity is detectable. However, it does not uncover the neural activity in the brain space. To localize the cortical regions that produce task-related activity, we propose a discriminant pattern source localization (DPSL) algorithm. DPSL consists of two major steps. First, discriminant analysis is applied to find a filter to optimally distinguish different brain states. Next, the gain of the filter is computed at each source location to reveal the activation map of task-related sources. Since the discriminant analysis algorithms in the first step are particularly designed to be robust to noise, DPSL can efficiently reduce the impact of noise and accurately identify the task-related sources.

As human mental processes are controlled by distributed cortical networks, the task-related activity often appears in multiple cortical regions. However, a number of new MEG applications have

recently emerged and suggest a need to decode different brain states by signals arising from a specific cortical region. Towards this goal, we propose a region-of-interest-constrained discriminant analysis (RDA) algorithm, which integrates linear classification and beamspace transformation into a unified framework by formulating a constrained optimization problem.

Acknowledgement

I am grateful to my advisor Xin Li. His insightful vision, knowledgeable guidance and continuous support help me eventually find this interesting research topic and complete the thesis work. His passion and dedication to the work have a profound influence on my professional and personal development.

I would like to acknowledge my collaborators Wei Wang, Anto Bagic, Gustavo Sudre, Douglas Weber, Jennifer Collinger and Stephen Foldes from University of Pittsburgh and the Center for the Neural Basis of Cognition (CNBC). This work cannot be done without their support.

I would like to thank Vijayakumar Bhagavatula, Byron Yu and Bin He for serving on my committee and offering invaluable advices for this work. In particular, I'm grateful to Byron for referring me to join the CNBC program, where I got the opportunity for a systematic and professional training in neuroscience.

The work in this thesis is supported by National Science Foundation and the ECE department of Carnegie Mellon University. I am grateful to these sponsors for their financial support.

My life at CMU has been interesting, fun and enjoyable. It has been filled with my lovely families and friends. I am grateful to all of my friends for sharing the time with me. I'm indebted to my parents and my husband for their unlimited encouragement, support and love. Last but not the least, I'm indebted to my little Jaden, you make everything in my world meaningful!

Table of Contents

Chapter 1	Introduction	1
1.1	Problem Statement.....	3
1.1.1	Problem 1	3
1.1.2	Problem 2	4
1.1.3	Problem 3	5
1.2	Thesis Contribution and Organization.....	7
Chapter 2	Background	10
2.1	Introduction to MEG	10
2.2	MEG Decoding.....	12
2.3	MEG Source Localization	13
Chapter 3	Brain States Decoding.....	17
3.1	Motivation	17
3.2	Background.....	19
3.3	Clustering Linear Discriminant Analysis	21
3.3.1	Feature Clustering	21
3.3.2	Discriminant Analysis.....	27
3.4	Experimental Results.....	29
3.4.1	Experimental Setup	29
3.4.2	Data Preprocessing.....	30
3.4.3	Feature Clustering	34
3.4.4	Movement Decoding.....	36
3.5	Discussion	38
3.5.1	Correlation Structure	38

3.5.2	Block Diagonal Approximation	40
3.6	Summary	43
Chapter 4	Task-related Source Localization	45
4.1	Motivation	45
4.2	Discriminant Pattern Source Localization	47
4.2.1	Discriminant Analysis	47
4.2.2	Source Imaging Based on Linear Transform Features	49
4.2.3	Source Imaging Based on Non-linear Transform Features	53
4.3	Time/Frequency Extension of DPSL.....	56
4.4	Simulation Studies.....	58
4.4.1	Simulation Setup	58
4.4.2	Evaluation Method	61
4.4.3	Simulation Results.....	62
4.5	Experimental Studies.....	65
4.5.1	Experimental Setup and Data Preprocessing.....	66
4.5.2	Spatio-temporal Source Imaging Results	67
4.6	Discussion	68
4.7	Summary	70
Chapter 5	Region of Interest Constrained Decoding	71
5.1	Motivation	71
5.2	ROI-Constrained Discriminant Analysis.....	73
5.3	Numerical Solver.....	78
5.4	Algorithm Summary.....	80
5.5	Simulation Studies.....	82
5.5.1	Simulation Setup	82
5.5.2	Simulation Results.....	85
5.6	Experimental Studies.....	90

5.6.1	Experimental Setup and Data Preprocessing.....	90
5.6.2	ROI Decoding Results.....	92
5.7	Discussion	99
5.8	Summary	101
Chapter 6	Thesis Summary & Future Work	102
6.1	Summary	102
6.2	Future Work.....	104
6.2.1	Application to Stroke Rehabilitation.....	104
6.2.2	Other Research Directions.....	107
	Bibliography.....	109

List of Figures

Figure 1-1. Compare different brain imaging methods for (a) invasiveness, and (b) spatial and temporal resolution. The color indicates the level of invasiveness. Red represents high invasiveness, and blue represents non-invasiveness. The imaging methods for comparison include: single unit recordings (SU), electrocorticography (ECoG), positron emission tomography (PET), single photon emission computed tomography (SPECT), functional magnetic resonance imaging (fMRI), electroencephalography (EEG) and magnetoencephalography (MEG).....	2
Figure 1-2. The aggregate of all proposed algorithms.	7
Figure 2-1. A whole-head MEG system (Neuromag®, Elekta AB, Sweden) used in our experiments.	11
Figure 2-2. The relationship between MEG measurements and the neuron sources within the brain. .	14
Figure 3-1. A simple example of within-class scatter matrix for four MEG features $\{x_1 x_2 x_3 x_4\}$ and the corresponding similarity graph.	22
Figure 3-2. A simplified diagram of the experimental setup for our MEG-based movement decoding. A human subject first holds the wrist at the center to start a trial. After a peripheral target onset, the subject moves (or imagines moving) the wrist to the target direction and holds that position until the peripheral target disappears. Next, the human subject waits for the target to re-appear at the center and then moves (or imagines moving) the wrist back to the center.....	29
Figure 3-3. The 74 selected MEG channels located on top of the sensorimotor area (two gradiometers at each location). The symbols “L”, “R” and “F” represent left, right and front, respectively.....	31

Figure 3-4. Fisher scores (logarithmic scale) of the wavelet coefficients for SubA Overt case. The six subplots correspond to the wavelet levels D1-D6 and A6. The horizontal axes represent the index of 74 selected channels (i.e., gradiometers located on top of the sensorimotor area). The vertical axes represent the index of wavelet coefficients. Red color indicates high Fisher scores, while blue color indicates low Fisher scores. 32

Figure 3-5. (a) The spatial distribution of wavelet coefficients from the first gradiometer, and (b) the spatial distribution of wavelet coefficients from the second gradiometer. In both (a) and (b), the first four columns correspond to the four movement directions, and the last column shows the Fisher scores (FC). Each color map was calculated by averaging the wavelet coefficients over all trials of the same class. Each row corresponds to the wavelet coefficients associated with the same time-frequency window. Red color indicates large value and blue color indicates small value. All plots of wavelet coefficients share the same color scale, and all plots of Fisher scores share the same color scale. 33

Figure 3-6. The quality function $q(K)$ where K represents the number of clusters. 34

Figure 3-7. The adjacency matrices are almost block diagonal for all data sets. The MEG features are ordered based on the feature clustering results where all features in the same cluster are grouped together. 35

Figure 3-8. Movement decoding accuracy estimated by using leave-one-out cross-validation: (a) decoding results of two movement directions (left and right), and (b) decoding results of four movement directions (up, down, left and right). 37

Figure 3-9. Feature groups determined by the spectral clustering algorithm. The horizontal axes represent the indexes of wavelet coefficients. They can also be considered as the indexes of time windows, since the corresponding wavelet basis functions are associated with the same low-frequency band and have local support in different time windows. The vertical axes represent the indexes of 74 selected channels (i.e., the gradiometers on top of the sensorimotor area). Different colors indicate different feature groups. 39

Figure 3-10. The upper bound of the CLDA error e_{UP} is close to the minimal error e_{OPT} that can be

achieved by an optimal classifier, if the condition number ξ_0 of the matrix Σ_0 is sufficiently small.	43
Figure 4-1. (a) Flow of the conventional inferential statistics approaches. (b) Flow of the proposed discriminant pattern source localization algorithm.	45
Figure 4-2. Spatial locations are shown for two task-related dipole sheets (i.e., 100 current dipoles $\{\mathbf{q}_{T,n}; n = 1,2,\dots,100\}$ modeling the task-related neural activity) and one common dipole sheet (i.e., 50 current dipoles $\{\mathbf{q}_{C,n}; n = 1,2,\dots,50\}$ modeling the common neural activity for both brain states) in our simulation studies.	60
Figure 4-3. (a) Spatial map of the t-statistic is estimated by IS. (b) Spatial map of the spatial filter gain is estimated by DPSL.	63
Figure 4-4. (a) Spatial locations are shown for the top 100 dipoles found by IS. (b) Spatial locations are shown for the top 100 dipoles found by DPSL.	64
Figure 4-5. The Error Distances of IS and DPSL for simulation data sets with different SNR.	65
Figure 4-6. The spatial-temporal map of the filter gain is estimated by DPSL for different time windows.	68
Figure 5-1. Red color indicates the spatial location of the ROI (i.e. the left precentral and the left postcentral gyri).	82
Figure 5-2. The decoding accuracies of RDA are represented as a function of λ (i.e., the constraint of the power ratio p_Φ/p_Φ defined in (5.13)).	85
Figure 5-3. Shown in red are the 36 channels on top of ROI where each dot represents a pair of gradiometers.	86
Figure 5-4. The decoding accuracies of BDA are represented as a function of λ (i.e., the lower bound of the power ratio p_Φ/p_Φ defined in (5.13)).	88
Figure 5-5. Movement decoding accuracies (both mean and standard deviation) of Subject 1 are plotted for LDA-W, LDA-S, SDA, BDA and RDA. For BDA and RDA, the accuracies are plotted as a function of λ (i.e., the constraint of the power ratio p_Φ/p_Φ defined in (5.13)).	93

Figure 5-6. Movement decoding accuracies (both mean and standard deviation) of Subject 2 are plotted for LDA-W, LDA-S, SDA, BDA and RDA. For BDA and RDA, the accuracies are plotted as a function of λ (i.e., the constraint of the power ratio p_{Φ}/p_{Φ} defined in (5.13)). 94

Figure 5-7. DPSL is applied to find the spatial locations of discriminant sources for LDA-W, LDA-S and RDA of Subject 1. Red color indicates the spatial locations where the gain of the spatial filter $y = \mathbf{w}^T \cdot \mathbf{x}$ in (5.1) is large. In other words, the MEG signals generated by the sources at these locations can pass the spatial filter $y = \mathbf{w}^T \cdot \mathbf{x}$ and contribute to decoding. 96

Figure 5-8. DPSL is applied to find the spatial locations of discriminant sources for LDA-W, LDA-S and RDA of Subject 2. Red color indicates the spatial locations where the gain of the spatial filter $y = \mathbf{w}^T \cdot \mathbf{x}$ in (5.1) is large. In other words, the MEG signals generated by the sources at these locations can pass the spatial filter $y = \mathbf{w}^T \cdot \mathbf{x}$ and contribute to decoding. 97

Figure 6-1. Schematic illustration of system design and data flow..... 106

List of Tables

Table 3-1. Number of successful trials per class of each data set-----	30
Table 3-2. Frequency band of wavelet coefficients at different levels-----	31
Table 4-1. Error distance of source localization for IS and DPSL-----	65
Table 5-1. Decoding accuracies of the algorithms for comparison-----	87

Chapter 1

Introduction

As early as the end of the eighteenth century, a German physician and neuroanatomist Franz Joseph Gall has proposed that all behavior emanates from the brain [1]. Until nowadays, however, the human brain still represents a barely-explored new world. Recent advances in brain imaging provide an increasingly important approach to investigate the neural system in human brains in both research and clinical care. In Figure 1-1, we list a range of common brain imaging methods. As shown in Figure 1-1(a), some methods have little or even no impact to human health. This non-invasive nature makes these methods highly preferable to be used in research, especially for the studies involving healthy volunteers. Magnetoencephalography (MEG) is a noninvasive modality that measures magnetic field generated by electrical neural activity [3]. As shown in Figure 1-1(b), MEG records brain activity with high temporal and spatial resolution. It is a valuable technique complementary to other noninvasive recording modalities, such as EEG and fMRI [8], [9]. In this thesis, we narrow down the brain imaging method to MEG.

Brain is believed to be the most complex organ in the human body. A living brain never stops “thinking”. No matter conscious or unconscious, overwhelmed neural activities are going on within a living brain. In order to focus on the neural activity associated with an interested mental task and avoid the impacts of the other factors, many MEG studies consider different brain states in the experimental design. These brain states differ in the effect of the interested mental task only [10], [11], [22]. For instance, consider an example where we are interested in the mental process that modulates hand movement directions. In the experiments, a human subject is asked to move the hand to different directions: left or right. Brain states under the two conditions are recorded and compared to

investigate the neural activity associated with the mental process that modulates hand movement directions. Neural activity associated with the interested mental task is referred to as task-related activity. In this thesis, we seek to image and analyze the task-related neural activity using MEG. In the rest of this chapter, we first present the problems that will be investigated in the thesis, and then outline the major contribution and the overall structure of the thesis.

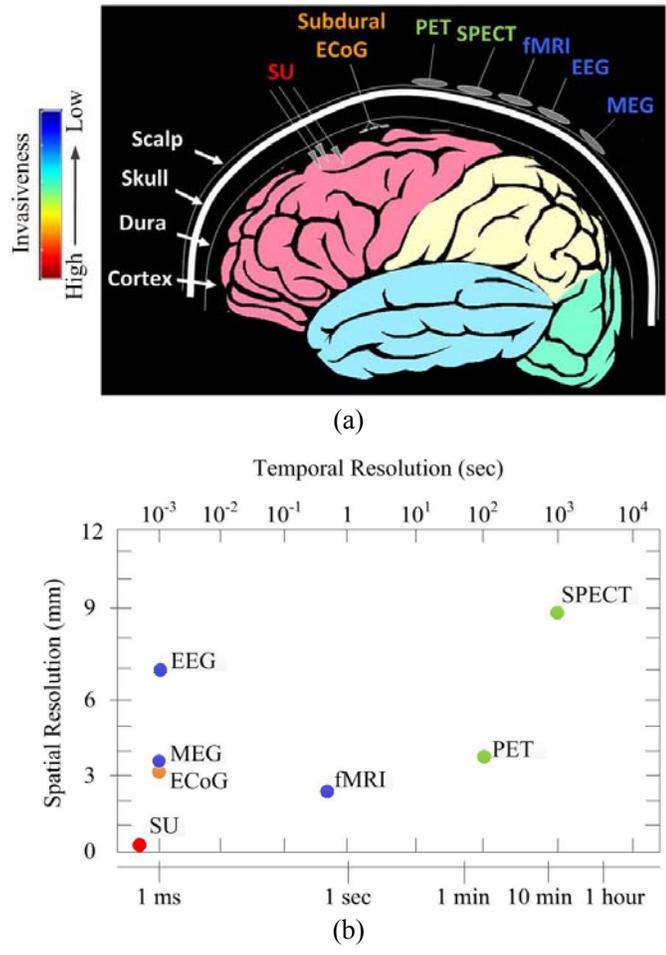


Figure 1-1. Compare different brain imaging methods for (a) invasiveness, and (b) spatial and temporal resolution. The color indicates the level of invasiveness. Red represents high invasiveness, and blue represents non-invasiveness. The imaging methods for comparison include: single unit recordings (SU), electrocorticography (ECoG), positron emission tomography (PET), single photon emission computed tomography (SPECT), functional magnetic resonance imaging (fMRI), electroencephalography (EEG) and magnetoencephalography (MEG).

1.1 Problem Statement

1.1.1 Problem 1

MEG is a powerful platform for recording neural activity with high temporal and spatial resolution. However, it is not sensitive to all types of neural activity. For example, MEG is sensitive to neural activity originating in the cortical sulci, whereas it is not good at recording the activity originating at the top of the cortical gyri [3]. Besides, MEG signals are extremely weak and, hence, a small noise or artifact can significantly distort them [3]. Therefore, it is possible that the task-related activity cannot be detected in MEG, although from neuroscience point of view the experiment is well designed and strong task-related activity is elicited. To investigate the task-related activity using MEG, we should first check whether the task-related activity is detectable in MEG recordings. In other words, we need to verify whether the MEG recordings associated with different brain states are distinguishable. This is a preliminary problem for most MEG studies. If the answer is yes, it is comfortable for the researchers to continue with further analysis. Otherwise, the researcher may need to change the experimental paradigm to induce stronger neural activity, or reconsider whether MEG is an appropriate imaging technique for the study. Thus, our first question is:

1. *Can we distinguish the MEG recordings associated with different brain states?*

This problem can be solved by MEG decoding (i.e. decoding brain states based on MEG recordings), which aims to train a discriminant function to predict different brain states. If the prediction can be made, the difference between brain states should exist. The prediction accuracy (i.e. the decoding accuracy) can be used as a quantitative measurement for the distinguishability between the MEG recordings associated with different brain states. High decoding accuracy indicates big difference. However, low accuracy might not because of little difference between MEG recordings of different brain states. It might be because of the poor performance of the decoding algorithm. Thus it is important to develop high performance decoding algorithms to efficiently extract the discriminant

information between MEG recordings associated with different brain states. Various signal processing and machine learning techniques, e.g., support vector machine [30], [38], [39], linear discriminant analysis (LDA) [31], [36], [37], logistic regression [32], etc., have been applied for MEG decoding. However, due to the high temporal and spatial resolutions, MEG recordings contain a large number of samples even within a short time period, which are generally transformed into a huge number of features. On the other hand, in practice the collected training data are often limited. Namely, many problems posed by MEG decoding are high-dimensional but with a small training set. In this case, the decoding algorithm must be carefully designed to prevent the decoder from over-fitting the training data. The challenging issue here is how to constrain our MEG features by their unique characteristics so that we can develop an efficient (i.e., requiring few training trials, and providing high decoding accuracy) decoding algorithm.

1.1.2 Problem 2

Answers to the first problem tell whether MEG successfully records the difference between brain states caused by task-related activity. However, it does not uncover the neural activity in the brain space. The idea that different cortical regions are specialized for different functions is now accepted as one of the cornerstones of modern neuroscience [1]. Thus, to understand the task-related activity in the brain space, it is important to relate the neural activity to particular cortical regions of the brain. For example, in order to investigate the neural activity induced by watching flashing lights, brain states under two experimental conditions are considered. In one condition the subject watches flashing light, while in the other condition the subject is in darkness. Comparing brain states under the two conditions, we find out that neural activity in the primary visual cortex is different, which indicates that watching flashing lights would activate primary visual cortex. Comparing different brain states and localizing task-related activity is important not only in neuroscience studies, but also in clinical research [12], [13]. Comparing the brain states of patients with healthy subjects, it is possible to identify the cortical areas that are affected by pathology. The second question we ask is:

2. *Can we localize the cortical regions that produce the task-related activity based on MEG recordings associated with different brain states?*

To solve this problem, inferential statistics (IS) based approaches have been applied in the literatures [14], [15], [22]. These approaches first apply a MEG source localization algorithm [16]-[20] to find the sources associated with each brain state. Once the sources associated with each brain state are estimated, statistical hypothesis testing is applied to determine if the sources found from the first phase carry task-related information. These approaches heavily rely on the accuracy of the source localization algorithm that is applied during the first phase. However, it is well-known that MEG source localization is profoundly underdetermined due to the limited observability of MEG measurements [9]. Hence, it is almost impossible to perfectly find all sources, especially if the signal-to-noise ratio is low. In most cases, a source localization algorithm can only capture the dominant sources, which are not necessarily associated with task-related activity. For this reason, the traditional IS-based approaches do not guarantee to identify the sources corresponding to task-related activity. It, in turn, poses an immediate need to re-think the fundamental strategy of source localization and develop a new algorithm for the task-related source localization problem.

1.1.3 Problem 3

Human mental processes are controlled by distributed cortical networks and, hence, after solving problem 2 we notice that task-related activity often appears in multiple cortical regions [26]. However, a number of new MEG applications have recently emerged and suggest a need to extract signals generated by task-related activity from specific cortical regions. For instance, applying MEG decoding to neurorehabilitation has attracted significant interest [87]-[92]. This emerging technique focuses neurorehabilitation on a target cortical region of interest, such as a region with dysfunction. The goal is to provide feedback of neural activity within this region so that the patient can learn how to produce activation patterns to facilitate the plasticity in a targeted cortical region. In this case, decoding MEG signals from the target cortical region only could be beneficial. Otherwise, even if the

decoding accuracy is high, the rehabilitation process may not effectively train the target cortical region and induce the desired function recovery. However, it is difficult, if not impossible, to force only the target cortical region to generate task-related activity in the experiments. Because human actions are controlled by distributed cortical networks, task-related activities often appear in multiple cortical regions [26]. In addition, some inevitable neural activity correlated with the experimental paradigm may produce discriminant signals. For example, in a visually guided motor study, a subject is asked to move the left or the right hand by following a visual cue. Although the intention is to activate task-related activity in the motor cortex, the visual cue for different hands induces task-related activity in the visual cortex as well. Therefore, we need to develop a new tool to extract the discriminant information from pre-specified cortical regions to distinguish different brain states while suppressing signals originating outside of the pre-specified cortical regions. Our next question is:

3. *Can we extract the MEG signals generated by the task-related activity within a specific cortical region only?*

These pre-specified cortical regions are referred to as regions of interest (ROI). One possible approach to address this problem is to first extract the MEG signals from the ROI by source localization or spatial filtering, and then construct a decoder to distinguish different brain states based on the signals from the ROI [33], [77], [94]-[96]. Such a two-step approach heavily depends on the results of the first step that may not perfectly capture all signals originating inside the ROI and simultaneously remove all signals originating outside the ROI. Taking source localization as an example, it is well-known that the source localization problem is profoundly underdetermined due to the limited observability of MEG measurements [9]. In most cases, these algorithms can only capture the dominant sources, but not necessarily the discriminant sources inside ROI. If the discriminant sources within ROI are not accurately estimated, we cannot achieve a high accuracy in the following decoding process. Another possible approach to address this problem is using only MEG channels on top of the ROI for decoding [31], [34], [35]. All MEG channels capture signals that may originate anywhere in the brain. Therefore, limiting the channels does not necessarily limit the ROI on the

cortex. For this reason, a simple channel selection approach cannot efficiently extract the MEG signals from the ROI and, hence, does not offer an optimal solution for this problem.

1.2 Thesis Contribution and Organization

This thesis provides several tools to answer the aforementioned questions related to task-related neural activity. Specifically, as shown in Figure 1-2, we propose three novel algorithms clustering linear discriminant analysis (CLDA), discriminant pattern source localization (DPSL) and region of interests constrained discriminant analysis (RDA) to address the three problems presented in Section 1.1, respectively.

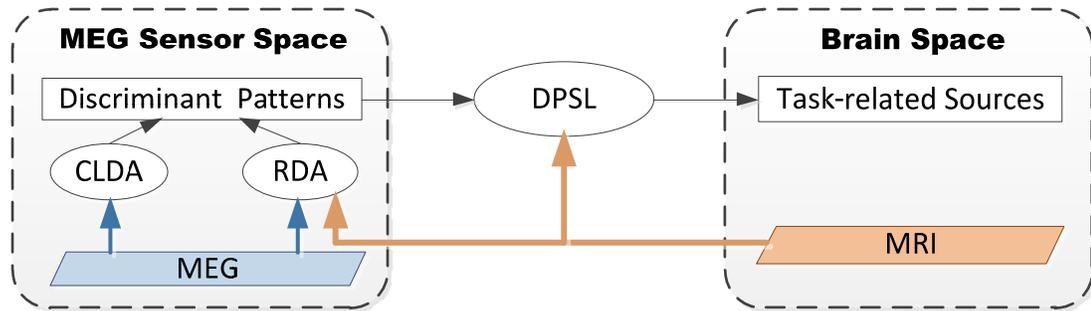


Figure 1-2. The aggregate of all proposed algorithms.

CLDA and RDA are designed for brain states decoding. Taking MEG signals as the input, CLDA aims to extract the optimal discriminant pattern existing in the MEG sensor space, which leads to the highest decoding accuracy. Different from CLDA, which makes use of only information in the MEG sensor space, RDA considers both MEG recordings in the sensor space and the underlying brain structures. RDA specifies the ROI based on MRI images, and seeks to detect the optimal discriminant pattern that generates by sources within the ROI. DPSL localizes the sources associated with task-related activity based on MEG recordings and MRI images. It bridges the discriminant pattern in MEG sensor space to the task-related neural activity in the brain space. These algorithms are briefly summarized as follows.

1. CLDA: If the feature space is high-dimensional and the training data are limited, in order to

achieve high decoding accuracy, the decoding algorithm must be carefully designed to avoid over-fitting the training data. CLDA applies a spectral clustering algorithm to automatically partition all MEG features into several groups where the within-group correlation is maximized and the between-group correlation is minimized. As such, the covariance matrix of all features can be approximated as a block diagonal matrix, thereby facilitating us to accurately extract the correlation information required by decoding from a small set of training data.

2. DPSL: Unlike most traditional algorithms that first use source localization algorithms to find sources associated with each brain state and then apply statistical hypothesis testing to determine task-related sources, DPSL starts from discriminant analysis in MEG sensor space. Once the optimal discriminant pattern existing in MEG sensor space is determined, the task-related sources are found by mapping the optimal pattern into the brain space. Since most discriminant analysis algorithms are particularly designed to reduce the impact of noise and capture the optimal discriminant pattern, by applying discriminant analysis in the first step DPSL can efficiently reduce the impact of both external noise (e.g., due to external magnetic sources) and internal interference (e.g., due to non-task-related neural activity).
3. RDA: RDA formulates a constrained optimization to find the optimal spatial filter that can accurately distinguish different brain states based on the MEG signals originating inside the ROI. In the optimization formulation, the cost function is derived by using linear classification techniques, which aim to maximize the discriminant information carried by the output of the spatial filter so that different brain states can be accurately separated. The constraint is designed by borrowing the idea of beamspace transformation, which guarantees that the spatial filter passes signals from the ROI while attenuating the signals originating outside the ROI. By simultaneously considering the decoding accuracy and the ROI constraint, RDA is able to identify the optimal discriminant pattern associated with the ROI. In addition, a numerical solver is developed to solve the non-convex optimization problem posed by RDA with guaranteed global convergence.

In Chapter 2, we briefly review the background knowledge related to this thesis work. Next In

Chapter 3, Chapter 4 and Chapter 5, we present the three algorithms CLDA, DPSL and RDA, respectively. Finally, in Chapter 6 we summarize the thesis, and discuss several potential research directions of the future work.

Chapter 2

Background

2.1 Introduction to MEG

When brain processes information, electrophysiological currents flow within and outside neural cells, thus producing electric and magnetic fields that are accessible to external measurements. MEG is a noninvasive technique that measures magnetic fields produced by neural currents in the living human brain. It is widely believed that MEG signals originate essentially from large cortical assemblies of pyramidal cells, with currents from post-synaptic potentials flowing orthogonally to the local cortical surface [3].

MEG was first measured using a copper induction coil as the detector in 1968 [4]. Four years later, a single superconducting quantum interference device (SQUID) was used successfully to record a human magnetic alpha rhythm with satisfactory signal-to-noise ratio [5]. It was in the second half of 1990s that systems able to simultaneously map magnetic fields over the whole scalp, namely whole-head system, became available [6]. The availability of whole-head systems has raised the interest of scientists and clinicians, and today MEG is used routinely in numerous hospital laboratories as a methodology to investigate the brain function and brain diseases. A 306-channel whole-head MEG system (Neuromag®, Elekta AB, Sweden) as in Figure 2-1 was in the experiments of this work.

Since MEG signal is a direct measurement of neural activity, it has extremely high temporal resolution (better than 1ms) [6], which is much better than the indirect measurements like fMRI (in the range of seconds) [11]. Besides, magnetic fields are less distorted than electric fields by the skull and the scalp and, hence, MEG has a better spatial resolution than the electric fields recordings, such as

EEG [3].



Figure 2-1. A whole-head MEG system (Neuromag®, Elekta AB, Sweden) used in our experiments.

MEG is a powerful platform for recording neural activity with high temporal and spatial resolution. However, it is not good at recording some types of neural activity. First, MEG is not sensitive to the neural activity at the top of the cortical gyri [3]. Second, MEG primarily detects intracellular currents associated with the synaptic potentials, while it is not sensitive to extracellular volume currents produced by postsynaptic potentials [7]. Third, the decay of magnetic fields as a function of distance is more pronounced than for electric fields. MEG is therefore more sensitive to superficial cortical activity [3].

Finally, MEG signal is extremely weak. Typically it is 50~500 fT. This is about 8~9 orders of magnitude smaller than the geomagnetic field generated by the earth and 1~2 orders of magnitude smaller than the magnetic field caused by eye movement [3]. For this reason, a small noise or artifact can significantly distort the MEG signals and eventually produce misleading results after data analysis. To alleviate the impact of noises and artifacts, a number of specific devices and instruments are adopted in an MEG system. For instance, magnetically shielded rooms have been designed to

reduce ambient interference, SQUID devices have been used to detect weak magnetic field and achieve high signal-to-noise ratio, and gradiometer coils have been applied to measure the MEG signals generated by human brain and simultaneously suppress the magnetic field caused by external noise sources.

2.2 MEG Decoding

Although conventional studies leave the answers to many important questions unclear (for example, how accurately and efficiently can a brain state be inferred? Is it possible to decode concealed thoughts or even unconscious mental states?), many human neuroimaging studies have provided strong evidence for a close link between the mind and the brain. It should, at least in principle, be possible to decode what an individual is thinking from their neural activity [28].

MEG noninvasively measures human neural activity with high temporal and spatial resolution. MEG decoding (i.e. decoding brain states based on MEG recordings) has been extensively studied over the past decade, especially in the community of brain computer interfaces (BCIs) [29]-[35]. Although some researchers work on decoding continuous brain states [27], such as decoding the brain states corresponding to continuous movement trajectory, in this thesis we further narrow down the scope of MEG decoding to the problems that decode a limited set of discrete brain states.

Decoding MEG in a limited set of choices is essentially a classification problem. For example, in an experiment a human subject is asked to perform two tasks: thinking left or right movements of the wrist. In this case, decoding brain signals into one of the two movement directions is equivalent to separating the signals into two classes, which leads to a binary classification. Supervised learning techniques are widely used to address this problem [28]-[35]. Such kind of approach requires a data collection phase during which human subjects repeatedly execute some training tasks. A discriminant function is then derived from the collected data sets through a learning process. This function is eventually used in the application to identify different brain states of the subjects.

Typically the learning process is composed of two interconnected problems: feature extraction

and classification. Many feature extraction methods developed for MEG decoding describe brain signals by spectral and/or spatial features such as power spectrum densities [36], [38], autoregressive model coefficients [30], [39], output power of spatial filters [32], [40], etc. Most of these methods rely on the assumption that brain signals are stationary. To further facilitate feature extraction for non-stationary brain signals, short time Fourier transform and wavelet transform have been used [43]-[47]. Once the extracted features are available, various classifiers (e.g., support vector machine [30], [38], [39], linear discriminant analysis (LDA) [31], [36], [37], logistic regression [32], etc.) have been applied to decode the brain states information.

As introduced in Section 2.1, a whole-head MEG system records brain signals using hundreds of channels. Besides, due to the high temporal resolution, MEG recordings contain a large number of time samples even within a short time period. Therefore, MEG signals are generally transformed into a huge number of features. However, in practice the collected training data are often limited. Namely, the classification problem posed by MEG decoding is high-dimensional but with a small training set. In this case, the decoding algorithm must be carefully designed to avoid over-fitting. One possible approach to address this issue is to reduce the dimension of features. For this reason, many feature extraction methods used by MEG decoding are followed by a feature selection process (e.g., score each feature individually and then select features with high scores) [30], [39], [48]. Another possible approach is to reduce the complexity of the classifier. As mentioned in [49], simple linear classifiers are preferred over complicated nonlinear classifiers for many applications. LDA is one of the popular linear methods and it has been widely applied in the literature [31], [36], [37].

2.3 MEG Source Localization

MEG measures the magnetic field on scalp surface. In order to analyze the underlying neural activity in the brain space, we need to model the relationship between the measured magnetic field and the neural current distribution within the brain.

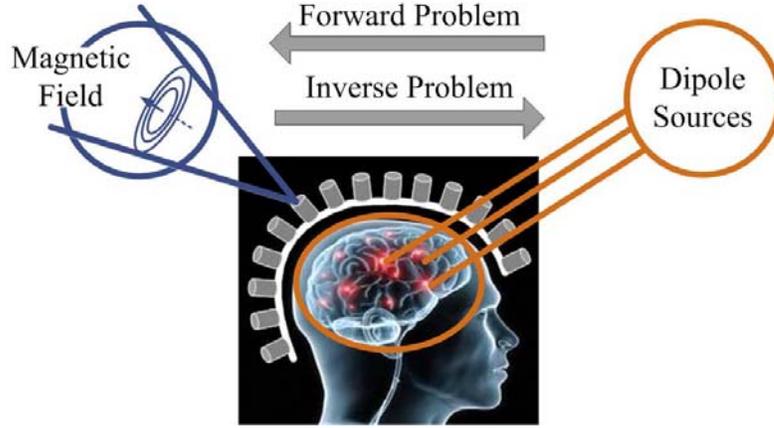


Figure 2-2. The relationship between MEG measurements and the neuron sources within the brain.

The forward model in MEG computes the magnetic field at given MEG sensor locations, when the neural current distribution is known. To simplify the model, a set of current dipoles are commonly used to approximate the neural current sources in the brain. Derived from the Maxwell Equations, the relationship between the external magnetic field and the moment of a set of current dipoles can be described by the following linear equations:

$$\begin{bmatrix} s_1 \\ \vdots \\ s_M \end{bmatrix} = \begin{bmatrix} a_{1,1}^1 & a_{1,2}^1 & a_{1,3}^1 & \cdots & a_{N,1}^1 & a_{N,2}^1 & a_{N,3}^1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ a_{1,1}^M & a_{1,2}^M & a_{1,3}^M & \cdots & a_{N,1}^M & a_{N,2}^M & a_{N,3}^M \end{bmatrix} \cdot \begin{bmatrix} q_{1,1} \\ q_{1,2} \\ q_{1,3} \\ \vdots \\ q_{N,1} \\ q_{N,2} \\ q_{N,3} \end{bmatrix}, \quad (2.1)$$

where M is the total number of MEG channels, N is the total number of current dipoles, $\mathbf{s} = [s_1 \dots s_M]^T \in R^M$ is the magnetic field at given MEG sensor locations, $\mathbf{q}_n = [q_{n,1} \ q_{n,2} \ q_{n,3}]^T \in R^3$ denotes the moment of the n th dipole and $\mathbf{A} = [a_{1,1}^1 \ a_{1,2}^1 \ a_{1,3}^1 \ \dots \ a_{N,1}^1 \ a_{N,2}^1 \ a_{N,3}^1; \dots; a_{1,1}^M \ a_{1,2}^M \ a_{1,3}^M \ \dots \ a_{N,1}^M \ a_{N,2}^M \ a_{N,3}^M]^T \in R^{M \cdot 3N}$ is the transformation matrix from the dipole moments to the external magnetic field. This matrix is called leadfield matrix. It can be calculated according to the geometrical structure and the conducting medium of the head.

To compute the leadfield matrix \mathbf{A} , we need to model the head geometry. Spherical head models have been widely used in the MEG community, where a sphere is fitted to the head geometry

[9]. This model is popular due to the simplicity for computation and the reasonably good performance in practice. However, in reality the head is not spherical and has inhomogeneous tissue properties. By using the anatomical structure data of the head such as MRI from individuals, it is possible to construct a more detailed head model by isolating different components of the head using automatic segmentation techniques [84].

Finally, in order to set current dipoles \mathbf{q}_n to valid locations, we need another model for the source space. Typically, the current sources are considered to distribute within the brain volume or be constrained at the cortical surface [9]. In the volume case, the brain is gridded using a 3D lattice of voxels. All sources are assumed to lie in these voxels. In the cortical surface case, a mesh is created from cortical surface, which comprise a continuous triangular tessellation. Sources are constrained to the vertices of the mesh. The cortical surface model derives from the assumption that MEG data originate essentially from large cortical assemblies of pyramidal cells, with currents from post-synaptic potentials flowing orthogonally to the local cortical surface [3].

The head model, the source space model, and the forward model in (2.1) allow us to compute the magnetic field given a set of current sources. On the other hand, to estimate the current sources given MEG measurements, we must solve an inverse problem, namely the source localization problem. If M equals $(3N)$ in (2.1) and the MEG system is noise free, it is easy to estimate the sources by solving the linear equations posed by the forward model. In reality, however, a MEG system has only hundreds of channels, whereas there are billions of neurons in an average adult brain. Although a sampled source space is considered in most source space model, the number of sources is still significantly larger than the number of MEG channels. Therefore, the source localization problem is profoundly underdetermined. The inherent ill-posedness, together with the noises and artifacts in MEG recording, makes the source localization problem very challenging.

There are indeed an infinite number of solutions to the source localization problem that fit given MEG recordings well. Therefore, priori assumptions about the sources are formulated implicitly or explicitly to find solutions with specific properties. Based on different assumptions, a large number of source localization algorithms have been developed during the past several decades, including

dipole fitting [9], multiple signal classification (MUSIC) [16], beamforming [17], minimum norm estimation (MNE) [18], sLORETA [19], minimum current estimation (MCE) [20], etc. These techniques have been successfully applied to many practical MEG problems.

Chapter 3

Brain States Decoding

3.1 Motivation

As was discussed in Section 1.1, to investigate the task-related activity using MEG, we should first check whether the task-related activity is detectable in MEG recordings. This problem can be solved by MEG decoding. If high decoding accuracy is achieved, the task-related activity should be well recorded by MEG. However, low accuracy might be because of the poor performance of the decoding algorithms. Thus developing high performance decoding algorithms to efficiently extract the discriminant information between MEG recordings associated with different brain states is one of our major works, which will be discussed in this chapter.

MEG decoding has been extensively studied over the past decade, especially in the community of brain computer interfaces (BCIs) [29]-[35]. Various signal processing and machine learning algorithms, e.g., support vector machine [30], [38], [39], linear discriminant analysis (LDA) [31], [36], [37], logistic regression [32], etc., have been applied to decode the brain states information.

Due to the high temporal resolution (i.e. 1ms typically) and multiple spatial measurements (i.e. hundreds of channels for a whole-head MEG system), MEG recordings contain a large number of samples even within a short time period. Therefore, MEG signals are generally transformed into a huge number of features. However, in practice it is impossible to collect too many training data. First, the human experiment is time consuming and expensive. Second, human subjects are easy to feel anxious or tired when repeating the same task for a long time, which may introduce numerous undesired artifacts in the brain signal. Thus the collected training data are often limited. Namely, the problem posed by MEG decoding is high-dimensional but with a small training set. In this case, the

decoding algorithm must be carefully designed to prevent the decoder from over-fitting the training data. To address this over-fitting problem, feature selection [30], [39] and regularization [31], [38] are often applied. For instance, diagonal LDA (DLDA) [37] and regularized LDA (RLDA) [31], [38] have been used in a few MEG decoding systems. Both of them pose extra constraints on the features to address the aforementioned dimensionality issue. In particular, DLDA assumes mutual independence among all features so that their covariance matrix can be approximated as a diagonal matrix. On the other hand, RLDA applies a Bayesian inference where a simple prior with diagonal covariance matrix is assumed for all features. In other words, both DLDA and RLDA rely on the prior knowledge that all features are mutually independent. While these two methods have been successfully applied to many practical MEG decoding problems, they may not guarantee high decoding accuracy if the underlying prior knowledge does not represent the actual correlation structure of features.

In this chapter, we propose a clustering linear discriminant analysis (CLDA) algorithm for MEG based brain states decoding. Unlike DLDA or RLDA, CLDA utilizes a unique group structure to model the correlation information of MEG features. It partitions all features into several groups where the within-group correlation is maximized and the between-group correlation is minimized. As such, the covariance matrix of all features can be approximated as a block diagonal matrix, thereby facilitating us to accurately extract the correlation information required by movement decoding from a small set of training data. Note that the traditional DLDA method can be conceptually viewed as a special case of CLDA where each group only contains a single feature and, hence, no within-group correlation is modeled. From this point of view, the proposed CLDA algorithm is a generalized version of DLDA. It aims to achieve improved decoding accuracy by accurately capturing the correlation information among all features. As will be demonstrated by the experimental results in Section 3.4, CLDA achieves superior decoding accuracy over other traditional approaches. The average accuracy of CLDA is 87% for single-trial movement decoding of four directions (i.e., up, down, left and right).

An important contribution of this algorithm is to apply a spectral clustering algorithm [61]-[64] to automatically identify the underlying group structure of MEG features and assign each feature to

the appropriate group. The spectral clustering method first represents the correlation information of MEG features in form of a similarity graph. Next, an optimal partition is constructed to split the graph into several sub-graphs (i.e., groups) based on its Laplacian matrix. The optimal number of groups is automatically determined by measuring the “quality” of the clustering results [63]. In this work, the spectral clustering algorithm is used, since it is not sensitive to the error of the correlation model estimated from a small set of training data, as is demonstrated by both theoretical studies and application examples in the machine learning community [61]-[64].

In addition, several theoretical aspects of the proposed CLDA algorithm are further examined in order to explain the reason why CLDA outperforms other traditional decoding techniques. An error bound is derived to quantitatively assess the approximation accuracy of the block diagonal covariance matrix and its impact on the final decoding accuracy. It can be shown that the decoding error of CLDA is directly related to the condition number of a normalized covariance matrix Σ_0 . If the condition number of Σ_0 is sufficiently small, the accuracy of CLDA is close to that of an optimal classifier. These results provide theoretical evidence to support the practical utility of the proposed CLDA method.

The remainder of this chapter is organized as follows. In Section 3.2, we briefly review the background of LDA, and then propose our CLDA algorithm in Section 3.3. The efficiency of CLDA is demonstrated by a number of experimental examples in Section 3.4. Several theoretical and practical aspects of CLDA are further discussed in Section 3.5. Finally, we conclude in Section 3.6.

3.2 Background

In this section, we briefly review the background of LDA and two of its modified versions: DLDA and RLDA. Consider two sets of training data $\{\mathbf{x}_{n,1}; n = 1, 2, \dots, N_1\}$ and $\{\mathbf{x}_{n,2}; n = 1, 2, \dots, N_2\}$ corresponding to two classes, where $\mathbf{x}_{n,k} = [x_{1,n,k} \ x_{2,n,k} \ \dots \ x_{M,n,k}]^T$ is the feature vector of the n th trial from the k th class, M stands for the number of features, and N_1 and N_2 represent the numbers of training samples for these two classes respectively. The key idea of LDA is to find the optimal

projection direction $\mathbf{p}_{OPT} \in R^M$ so that the between-class scatter is maximized and the within-class scatter is minimized [52]. Define the within-class scatter matrix $\mathbf{S}_W \in R^{M \times M}$ as

$$\mathbf{S}_W = \sum_{n=1}^{N_1} (\mathbf{x}_{n,1} - \boldsymbol{\mu}_1) \cdot (\mathbf{x}_{n,1} - \boldsymbol{\mu}_1)^T + \sum_{n=1}^{N_2} (\mathbf{x}_{n,2} - \boldsymbol{\mu}_2) \cdot (\mathbf{x}_{n,2} - \boldsymbol{\mu}_2)^T, \quad (3.1)$$

where $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ stand for the mean of $\{\mathbf{x}_{n,1}; n = 1, 2, \dots, N_1\}$ and $\{\mathbf{x}_{n,2}; n = 1, 2, \dots, N_2\}$, respectively. If \mathbf{S}_W in (3.1) is non-singular, \mathbf{p}_{OPT} can be determined as [52]

$$\mathbf{p}_{OPT} \propto \mathbf{S}_W^{-1} \cdot (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2). \quad (3.2)$$

Once \mathbf{p}_{OPT} is found, the following discriminant function can be constructed for two-class classification

$$\mathbf{p}_{OPT}^T \cdot \left(\mathbf{x} - \frac{\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2}{2} \right) = \begin{cases} \geq 0 & (1st\ Class) \\ < 0 & (2nd\ Class) \end{cases}. \quad (3.3)$$

The aforementioned two-class LDA can be extended to handle multiple classes. More details of LDA can be found in [52].

If there are a sufficient number of training samples, \mathbf{S}_W in (3.1) is an accurate estimator of the covariance matrix and LDA yields the optimal projection direction \mathbf{p}_{OPT} that maximizes classification accuracy. However, if only a small number of training samples are available for a high-dimensional feature space, it is extremely difficult to accurately estimate the covariance matrix required by LDA. To address this dimensionality issue, DLDA [37] and RLDA [31], [38] have been proposed. DLDA assumes mutual independence among all features, thereby forcing \mathbf{S}_W to be diagonal. Alternatively, RLDA adds an additional regularization term to the estimator: $\mathbf{S}_W + \lambda \cdot \mathbf{I}$, where \mathbf{I} is an identity matrix and $\lambda \geq 0$ is a regularization parameter that is typically determined by cross-validation. Both DLDA (with a diagonal within-class scatter matrix) and RLDA (using a diagonal covariance matrix to model the prior distribution for Bayesian inference) rely on the prior knowledge that all MEG features are mutually independent. While these two methods have been successfully applied to a broad range of practical applications, they cannot guarantee high decoding accuracy if the underlying prior knowledge does not represent the actual correlation structure of MEG features. In addition, as will be demonstrated by the experimental results in Section 3.4, simple feature selection (e.g., feature selection using Fisher criterion [52]) does not lead to high decoding accuracy either. These

observations, therefore, motivate us to develop a new CLDA algorithm to achieve improved classification accuracy by carefully modeling the mutual correlation among all features.

3.3 Clustering Linear Discriminant Analysis

The proposed CLDA algorithm relies on a unique group structure to extract the correlation information required for movement decoding. Namely, we assign each feature to the appropriate group so that the within-group correlation is maximized and the between-group correlation is minimized. Note that such a feature clustering task is not trivial, since the correlation information extracted from training data is likely to be inaccurate, especially if only a limited number of training samples are available for a high-dimensional feature space. In other words, the challenging issue here is how to develop a robust clustering scheme that is not sensitive to the error of the correlation model estimated from a small set of training data.

In this section, we propose to borrow the spectral clustering algorithm [61]-[64] from graph theory [60] to address the aforementioned challenge on feature clustering. Spectral clustering is one of the most important clustering techniques developed by the machine learning community. It first forms a similarity graph based on the mutual correlation of different features. Next, the features are partitioned into several groups based on the Laplacian matrix of the similarity graph. Here, the spectral clustering algorithm is selected, because it can provide robust performance, even if the input data are noisy, as is demonstrated by both theoretical studies and application examples in the machine learning community [61]-[64]. Hence, the spectral clustering algorithm perfectly fits the needs of our proposed feature clustering problem. Based upon spectral clustering, a modified LDA algorithm (i.e., CLDA) is further proposed for MEG decoding using the grouped features. In what follows, we will describe the technical details of these algorithms and highlight their novelties.

3.3.1 Feature Clustering

Given a set of MEG features $\{x_m; m = 1, 2, \dots, M\}$, the goal of feature clustering is to partition all

features into several groups such that the features in the same group are similar and different features in different groups are dissimilar to each other. In our application, correlation is the criterion to quantitatively measure the “similarity” between different features. Namely, we want to maximize the within-group correlation and simultaneously minimize the between-group correlation.

To mathematically define the aforementioned feature clustering problem, we represent all features $\{x_m; m = 1, 2, \dots, M\}$ in form of a weighted undirected graph $G = (X, E)$ that is referred to as *similarity graph* in [64]. In this graph G , each vertex represents a feature x_m where $m \in \{1, 2, \dots, M\}$. Two vertices x_m and x_n are connected by an edge e_{mn} , if and only if the correlation between these two features is non-zero. The weight w_{mn} of e_{mn} is equal to the correlation coefficient

$$w_{mn} = \frac{|\mathbf{S}_{W,m,n}|}{\sqrt{\mathbf{S}_{W,m,m}} \cdot \sqrt{\mathbf{S}_{W,n,n}}} \quad (m, n = 1, 2, \dots, M), \quad (3.4)$$

where $\mathbf{S}_{W,m,n}$ stands for the (m, n) -th element of the within-class scatter matrix \mathbf{S}_W in (3.1). For each vertex x_m , there is a self-edge e_{mm} and the weight w_{mm} is equal to 1. In (3.4), the correlation coefficient is defined by the within-class scatter matrix and is always non-negative. Figure 3-1 shows a simple example of within-class scatter matrix for four features $\{x_1, x_2, x_3, x_4\}$ and the corresponding similarity graph.

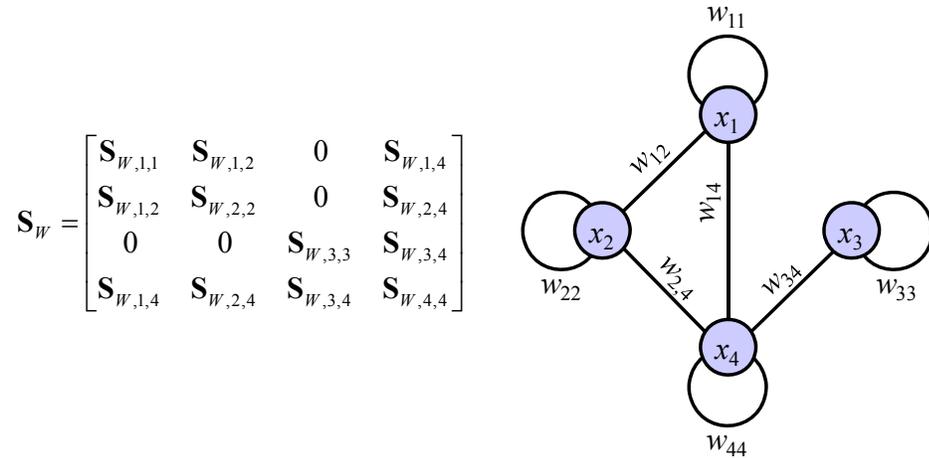


Figure 3-1. A simple example of within-class scatter matrix for four MEG features $\{x_1, x_2, x_3, x_4\}$ and the corresponding similarity graph.

Based on the similarity graph $G = (X, E)$, we want to partition G into several sub-graphs such that the edges between different sub-graphs have small weights (i.e., the corresponding features are weakly correlated) and the edges within the same sub-graph have large weights (i.e., the corresponding features are strongly correlated). Such a partition can be constructed by using the Laplacian matrix of G . In what follows, we will first define several important terminologies in graph theory [60], [64].

The *adjacency matrix* $\mathbf{W} \in R^{M \times M}$ of the similarity graph G is defined as

$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1M} \\ w_{12} & w_{22} & \cdots & w_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ w_{1M} & w_{2M} & \cdots & w_{MM} \end{bmatrix}. \quad (3.5)$$

Namely, the (m, n) -th element of \mathbf{W} is the weight w_{mn} of the edge e_{mn} . If the two vertices x_m and x_n are not connected, w_{mn} is simply set to zero. Since the similarity graph G is undirected, the adjacency matrix \mathbf{W} is symmetric. Based on the adjacency matrix \mathbf{W} , the *degree* of a vertex x_m is defined as

$$d_m = \sum_{n=1}^M w_{mn} \quad (m = 1, 2, \dots, M). \quad (3.6)$$

Remember that the weight w_{mn} is non-zero, if and only if the vertices x_m and x_n are connected. Hence, the degree d_m in (3.6) is determined by all edges that are connected to x_m . The degree matrix $\mathbf{D} \in R^{M \times M}$ is defined as a diagonal matrix with $\{d_m; m = 1, 2, \dots, M\}$ on its diagonal

$$\mathbf{D} = \text{diag}(d_1, d_2, \dots, d_M). \quad (3.7)$$

Now, we are ready to define the Laplacian matrix $\mathbf{L} \in R^{M \times M}$ of the similarity graph G

$$\mathbf{L} = \mathbf{I} - \mathbf{D}^{-1/2} \cdot \mathbf{W} \cdot \mathbf{D}^{-1/2}. \quad (3.8)$$

It can be shown that the Laplacian matrix \mathbf{L} in (3.8) is positive semi-definite. All of its eigenvalues are within the interval $[0, 1]$. In particular, $\lambda = 0$ is one of its eigenvalues (i.e., the smallest eigenvalue) and the corresponding eigenvector is

$$\mathbf{v} = [\sqrt{d_1} \quad \sqrt{d_2} \quad \cdots \quad \sqrt{d_M}]^T. \quad (3.9)$$

More details on Laplacian matrix can be found in [60], [64].

If the similarity graph G can be exactly partitioned into K sub-graphs $\{G_k; k = 1, 2, \dots, K\}$ (i.e., there is no edge connecting any two sub-graphs) and all vertices are appropriately ordered (i.e., the vertices in the same sub-graph are grouped together), it is straightforward to verify that the Laplacian matrix \mathbf{L} of the graph G is block diagonal

$$\mathbf{L} = \text{diag}(\mathbf{L}_1, \mathbf{L}_2, \dots, \mathbf{L}_K), \quad (3.10)$$

where \mathbf{L}_k is the Laplacian matrix of the sub-graph G_k . Since \mathbf{L}_k is a Laplacian matrix, $\lambda = 0$ is one of its eigenvalues and we represent the corresponding eigenvector as \mathbf{v}_k . Similar to (3.9), \mathbf{v}_k can be determined by the degrees of the vertices in the sub-graph G_k . Based on these observations, we can conclude that the Laplacian matrix \mathbf{L} in (3.10) has K different eigenvectors that is associated with the same eigenvalue $\lambda = 0$

$$\mathbf{V} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{v}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{v}_K \end{bmatrix}, \quad (3.11)$$

where each column of the matrix $\mathbf{V} \in R^{M \times K}$ in (3.11) is an eigenvector of the Laplacian matrix \mathbf{L} , and the symbol $\mathbf{0}$ denotes a zero vector (i.e., all elements in $\mathbf{0}$ are zero).

Next, we normalize each row of the matrix \mathbf{V} to unit length, resulting in

$$\tilde{\mathbf{V}} = \begin{bmatrix} \mathbf{1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{1} \end{bmatrix}, \quad (3.12)$$

where the symbol $\mathbf{1}$ denotes a vector in which all elements are one. Studying the matrix $\tilde{\mathbf{V}} \in R^{M \times K}$ in (3.12), one would notice that each row of $\tilde{\mathbf{V}}$ corresponds to a vertex x_m of the similarity graph G (i.e., the MEG feature x_m). We can conceptually consider the m th row of $\tilde{\mathbf{V}}$ as the coordinate of the m th feature x_m in a “transformed” feature space R^K . For all features in the same sub-graph G_k , their coordinates are identical. Hence, we can use the “normalized” eigenvectors in (3.12) to partition all MEG features into K groups (e.g., by applying K-means clustering [52]). Such a feature clustering scheme is based on spectral graph theory [60] and, hence, is referred to as spectral clustering in the

literature [61]-[64].

While the above discussion covers the key idea of the proposed feature clustering based on spectral graph theory, there are three important implementation issues that should be further considered. First, our previous discussion assumes that all vertices of the similarity graph G are appropriately ordered so that the Laplacian matrix \mathbf{L} in (3.10) is block diagonal. In practice, this is never the case and the correct ordering of all vertices should be the output of the clustering algorithm. Note that if the ordering of the vertices is changed, the Laplacian matrix of the similarity graph G can be written as a linear transformation of the block diagonal matrix \mathbf{L} in (3.10): $\mathbf{P}\cdot\mathbf{L}\cdot\mathbf{P}^T$ where \mathbf{P} is a permutation matrix (i.e., an identity matrix with its rows re-ordered [60]). Compared to the matrix \mathbf{L} , the matrix $\mathbf{P}\cdot\mathbf{L}\cdot\mathbf{P}^T$ has identical eigenvalues. The eigenvectors of \mathbf{L} and $\mathbf{P}\cdot\mathbf{L}\cdot\mathbf{P}^T$ only differ by a simple permutation. In other words, the aforementioned properties for eigenvalues and eigenvectors still hold and, hence, the spectral clustering scheme can be applied to appropriately identify the sub-graphs, even if all vertices in the similarity graph G are arbitrarily ordered.

Second, the sub-graphs $\{G_k; k = 1, 2, \dots, K\}$ are not necessarily disconnected in most practical applications, since the MEG features x_m and x_n in different sub-graphs can be weakly correlated and, hence, the weight w_{mn} of the edge e_{mn} is not exactly zero. In addition, since the weight w_{mn} in (3.4) is calculated from the within-class scatter matrix \mathbf{S}_W in (3.1) based on a set of training data, the estimation of w_{mn} can be inaccurate, especially if the feature space is high-dimensional and only a small set of training data are available. In these cases, the Laplacian matrix \mathbf{L} in (3.8) is not exactly block diagonal. However, the spectral clustering result is not sensitive to the small perturbation presented in the correlation model. In other words, even if the Laplacian matrix \mathbf{L} deviates from the ideal (i.e., block diagonal) case, spectral clustering can still yield the correct clustering results, as is demonstrated by both theoretical studies and application examples in the machine learning community [61]-[64]. For instance, a detailed perturbation analysis of the invariant subspace (i.e., the subspace spanned by eigenvectors) has been studied in [64] for spectral clustering. It provides theoretical evidence that the spectral clustering method offers robust performance and, hence, perfectly fits the needs of our proposed feature clustering problem.

Third, the number of clusters (i.e., K) is not known in advance when the proposed feature clustering scheme is applied to movement decoding. In other words, the optimal value of K must be automatically determined as part of the clustering procedure. To this end, we borrow the concept of quality function $q(K)$ from [63]. Namely, $q(K)$ is defined to measure the quality of different clustering results with different K values

$$q(K) = \sum_{k=1}^K \left\{ \frac{f(G_k, G_k)}{f(G, G)} - \left[\frac{f(G_k, G)}{f(G, G)} \right]^2 \right\}, \quad (3.13)$$

where the function $f(A, B)$ measures the similarity between two sub-graphs A and B based on their weights

$$f(A, B) = \sum_{\substack{x_m \in A \\ x_n \in B}} w_{mn}. \quad (3.14)$$

It has been empirically demonstrated by a broad range of simulated and real-world examples in [63] that a large $q(K)$ implies an improved clustering result. Hence, we can repeatedly perform feature clustering with different K values and calculate the “quality” $q(K)$ in (3.13) for the clustering results. The optimal value of K is then determined by finding the largest quality function $q(K)$. In Section 3.4.3, we will show an experimental example of the quality function $q(K)$ over K .

Algorithm 3-1: Feature Clustering

1. Start from the training data $\{\mathbf{x}_{n,1}; n = 1, 2, \dots, N_1\}$ and $\{\mathbf{x}_{n,2}; n = 1, 2, \dots, N_2\}$ corresponding to the MEG features $\{x_m; m = 1, 2, \dots, M\}$ of two classes.
2. Calculate the within-class scatter matrix $\mathbf{S}_W \in R^{M \times M}$ in (3.1). Construct the similarity graph G . Calculate the adjacency matrix $\mathbf{W} \in R^{M \times M}$ in (3.4)-(3.5), the degree matrix $\mathbf{D} \in R^{M \times M}$ in (3.6)-(3.7), and the Laplacian matrix $\mathbf{L} \in R^{M \times M}$ in (3.8).

For each $K \in \{1, 2, \dots, M\}$

3. Find the K smallest eigenvalues and the corresponding K eigenvectors. Form the matrix $\mathbf{V} \in R^{M \times K}$ where each column is one of the K eigenvectors.
4. Normalize each row of the matrix \mathbf{V} to unit length, resulting in the matrix $\tilde{\mathbf{V}} \in R^{M \times K}$.

5. Consider each row of $\tilde{\mathbf{V}}$ as the coordinate of a MEG feature in the space R^K , and apply K-means clustering [52] to partition the features $\{x_m; m = 1, 2, \dots, M\}$ into K groups.
6. Calculate the quality function $q(K)$ in (3.13).

End For

7. Find the optimal value K_{OPT} at which the quality function $q(K)$ reaches its maximum. Use the clustering result at K_{OPT} to partition all features into K_{OPT} groups.

Algorithm 3-1 summarizes the major steps of the proposed feature clustering method. During the clustering procedure, the optimal number of clusters is automatically determined by evaluating the quality function in (3.13). Once the features are appropriately partitioned into K groups, a modified LDA algorithm (i.e., CLDA) can be applied for movement decoding using the grouped features. The details of CLDA will be discussed in the next section.

3.3.2 Discriminant Analysis

Once the MEG features $\{x_m; m = 1, 2, \dots, M\}$ are partitioned into K groups, all features are ordered according to their group assignment. Since the mutual correlation between different groups is almost zero, the within-class scatter matrix $\mathbf{S}_W \in R^{M \times M}$ can be approximated by a block diagonal form

$$\mathbf{S}_{WB} = \text{diag}(\mathbf{S}_{W,1}, \mathbf{S}_{W,2}, \dots, \mathbf{S}_{W,K}), \quad (3.15)$$

where $\mathbf{S}_{W,k}$ stands for the within-class scatter matrix for the features in the k th group. The key idea of CLDA is to estimate the block diagonal matrix \mathbf{S}_{WB} in (3.15), and use it to replace \mathbf{S}_W in (3.2) to calculate the optimal projection direction \mathbf{p}_{OPT} . Since the block diagonal matrix \mathbf{S}_{WB} is more constrained than the original within-class scatter matrix \mathbf{S}_W , \mathbf{S}_{WB} is less sensitive to the dimensionality issue posed by high-dimensional feature space and small training data set. In other words, the proposed CLDA can efficiently approximate the within-class scatter matrix using a block diagonal form and then accurately estimate the optimal projection direction \mathbf{p}_{OPT} , even if the feature space is high-dimensional and the training data are limited. This is the primary advantage of CLDA over LDA.

On the other hand, CLDA can be viewed as a direct extension of DLDA [37]. Unlike DLDA

that approximates the within-class scatter matrix \mathbf{S}_W by a diagonal matrix and completely ignores the correlation between different features, the proposed CLDA is able to automatically identify the critical correlation information (i.e., by partitioning all features into different groups) and then accurately extract the correlation (i.e., by estimating a block diagonal within-class scatter matrix) from a small set of training data. For this reason, CLDA can capture the mutual correlation of MEG features more accurately than DLDA. Hence, it is expected to achieve superior decoding accuracy over DLDA.

Algorithm 3-2: Clustering Linear Discriminant Analysis (CLDA)

1. Start from the training data $\{\mathbf{x}_{n,1}; n = 1, 2, \dots, N_1\}$ and $\{\mathbf{x}_{n,2}; n = 1, 2, \dots, N_2\}$ corresponding to the MEG features $\{x_m; m = 1, 2, \dots, M\}$ of two classes.
2. Apply Algorithm 3-1 to partition all features into K_{OPT} groups.
3. Order all features according to their group assignment. Construct the block diagonal within-class scatter matrix \mathbf{S}_{WB} in (3.15).
4. Replace \mathbf{S}_W in (3.2) by \mathbf{S}_{WB} to calculate the optimal projection direction \mathbf{p}_{OPT} .
5. Create the two-class classifier for movement decoding based on the discriminant function in (3.3).

Algorithm 3-2 summarizes the simplified flow of the proposed CLDA method for two classes. It should be noted that Algorithm 3-2 can be easily extended to multiple classes by following the standard flow of multi-class LDA [52]. Since the extension to multi-class CLDA is straightforward, we will not present its details in this thesis.

The efficiency of the proposed CLDA algorithm will be demonstrated by several experimental examples in the next section. As will be shown in Section 3.4, CLDA results in substantially higher decoding accuracy than other traditional approaches, including LDA with feature selection, DLDA and RLDA. In addition, a theoretical study will be presented in Section 3.5.2 to derive the error bound of the proposed block diagonal approximation. It, in turn, further explains why CLDA achieves superior accuracy for decoding.

3.4 Experimental Results

In this section, CLDA is applied to MEG-based movement decoding and its performance is compared to other traditional decoding techniques on five human subjects. All experimental procedures are approved by the Institutional Review Boards of the University of Pittsburgh and Carnegie Mellon University. All experiments are performed in accordance with the approved protocol. In what follows, we will describe the experimental setup and the movement decoding results in detail.

3.4.1 Experimental Setup

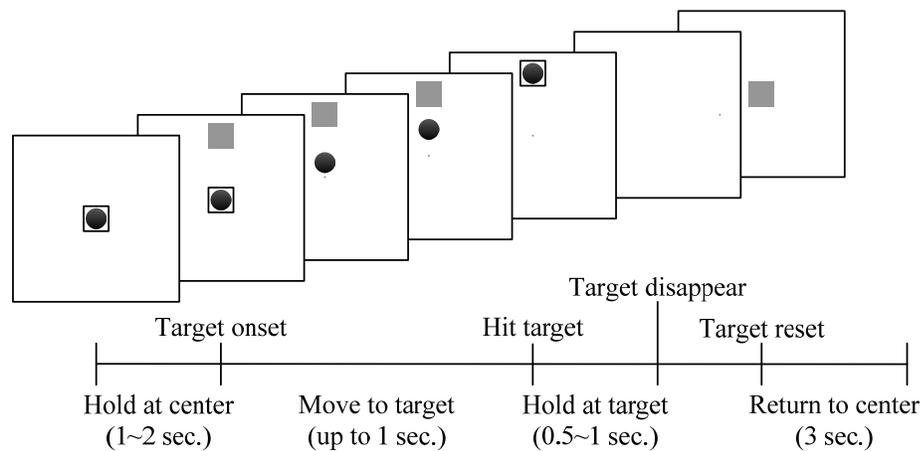


Figure 3-2. A simplified diagram of the experimental setup for our MEG-based movement decoding. A human subject first holds the wrist at the center to start a trial. After a peripheral target onset, the subject moves (or imagines moving) the wrist to the target direction and holds that position until the peripheral target disappears. Next, the human subject waits for the target to re-appear at the center and then moves (or imagines moving) the wrist back to the center.

In our experiment, five human subjects perform a four-target center-out task with their wrist holding an MEG-compatible joystick. During overt movements, subjects are instructed to move the cursor from the center target to one of the four locations (i.e., up, down, left or right) by making wrist movements (i.e., radial deviation, ulnar deviation, flexion and extension) while keeping the rest of the body in a relaxed position. A successful repetition is characterized by reaching one of the four

peripheral targets within a pre-specified time window after the onset of the target and holding the cursor position there without overshooting, as shown in Figure 3-2. Only successful repetitions are used for our off-line data analysis. During imagined movements, subjects are instructed to imagine making the wrist movements to one of the four targets displayed on the screen, while the cursor moves from the center position to the target automatically. For both overt and imagined conditions, subjects are instructed to keep their gaze at the center of the screen, and only attend to the targets using their peripheral vision.

During the experiment, MEG data are acquired by using a 306-channel whole-head MEG system (Elekta Neuromag[®]) with 1 kHz sampling frequency. In addition, electrooculography (EOG) is used to monitor eye blinks and eye movements. Electromyography (EMG) of wrist flexor and extensor muscles is recorded to make sure that no movement happened during the imagined sessions. All trials with EOG or EMG contamination are rejected.

All five subjects perform both overt and imagined movements in the experiment, resulting in a total of ten data sets. The number of successful trials in each data set is not the same after rejecting contaminated repetitions. This number is adjusted for each data set among the four classes (i.e., up, down, left and right) such that each class in the same data set has the same number of successful trials. In other words, we discard the last few trials for the classes with more trials than the others. Table 3-1 summarizes the data set size for the aforementioned experimental setup.

Table 3-1. Number of successful trials per class of each data set

Subject ID	SubA	SubB	SubC	SubD	SubE
Overt	81	84	155	123	93
Imagined	75	179	129	126	98

3.4.2 Data Preprocessing

The recorded MEG signals are processed by the signal space separation (SSS) method [65] to

remove the interference signals due to magnetic impurities (e.g., sensor electronics, electrical activities from arm muscles, etc.). SSS also compensates the signal distortions caused by head movement. Next, a notch filter is applied to remove the 60 Hz power line interference. A linear approximation is then determined by least-squares fitting for each channel and each trial, and the linear trends are subtracted from the recorded MEG signals. In this study, although the MEG signals contain 306 channels, only 74 channels are used for movement decoding. These 74 channels correspond to the gradiometers located on top of the sensorimotor area, as shown in Figure 3-3. They are expected to carry useful information about the motor activity in which we are interested.

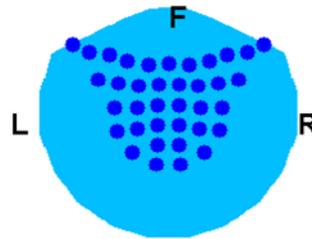


Figure 3-3. The 74 selected MEG channels located on top of the sensorimotor area (two gradiometers at each location). The symbols “L”, “R” and “F” represent left, right and front, respectively.

Several previous studies have shown that brain signals are non-stationary [47]. Discrete wavelet transform (DWT) is a powerful tool for non-stationary signal analysis [43]-[47]. It decomposes a signal into its time-frequency representation with different resolutions [43]. Every wavelet coefficient corresponds to the signal energy in a particular time-frequency window. In this study, we apply DWT with second-order Symlet wavelet function [70] to decompose the MEG signals from each channel and each trial to multiple resolution levels (i.e. D1~D6 and A6). Table 3-2 lists the frequency bands corresponding to different wavelet levels for our data analysis.

Table 3-2. Frequency band of wavelet coefficients at different levels

Wavelet level	D1	D2	D3	D4	D5	D6	A6
Frequency (Hz)	250-500	125-250	62.5-125	31.2-62.5	15.6-31.2	7.8-15.6	0-7.8

Previous neuroscience research on MEG-based BCI demonstrates that significant power modulation of MEG activity is observed in three different frequency bands [31]: (1) ≤ 7 Hz (low-frequency band), (2) 62-87 Hz, and (3) 10-30 Hz. In [31], the authors further mention that movement directions can be inferred from the low-frequency band only, but not from the other two frequency bands. In our study, we employ Fisher criterion to score the wavelet coefficients at different levels (i.e., different frequency bands), and find that most highly-scored coefficients are from the level A6 corresponding to the low-frequency band (≤ 7.8 Hz). An example of the Fisher scores for SubA Overt case is shown in Figure 3-4. This observation from our data sets is exactly consistent with the previous research in [31]. Therefore, in the following data analysis, we only consider the wavelet coefficients at the level A6.

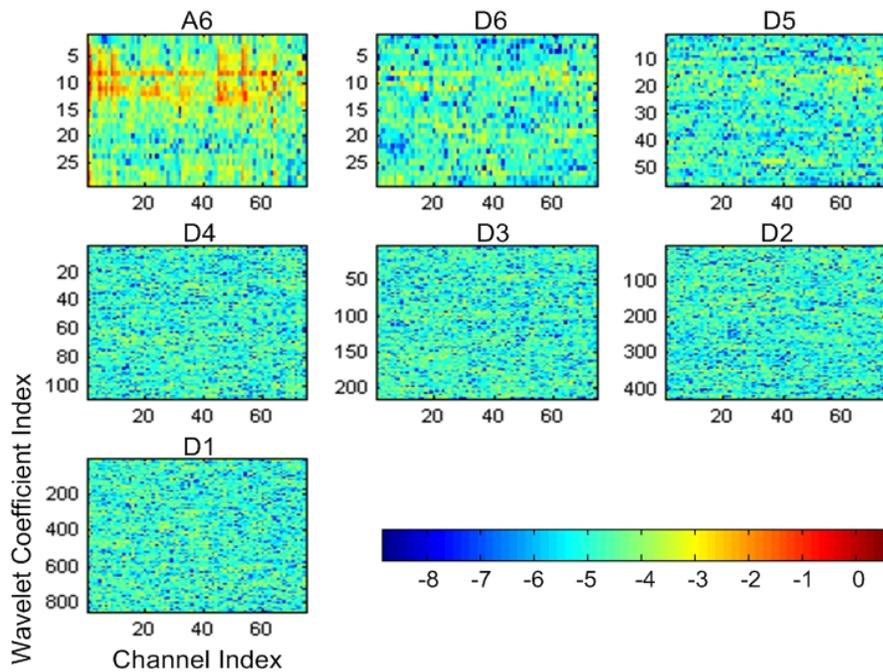


Figure 3-4. Fisher scores (logarithmic scale) of the wavelet coefficients for SubA Overt case. The six subplots correspond to the wavelet levels D1-D6 and A6. The horizontal axes represent the index of 74 selected channels (i.e., gradiometers located on top of the sensorimotor area). The vertical axes represent the index of wavelet coefficients. Red color indicates high Fisher scores, while blue color indicates low Fisher scores.

In addition, the important neural activity that carries movement information is observed during a short time window [31]. Thus we only consider the time window $t \in [0.2 \text{ s}, 0.6 \text{ s}]$, where $t = 0 \text{ s}$ represents target onset, in our movement decoding. Six wavelet coefficients at the level A6 corresponding to the selected time-frequency window are used to represent the features for each channel. Here, each time-frequency window is around 60 ms in length and covers the low frequency band ($\leq 7 \text{ Hz}$). Since 74 channels are considered in total, the dimensionality of the feature space is: $M = 6 \times 74 = 444$. Each wavelet coefficient (i.e., the feature) is correlated to the signal energy in a specific time-frequency window of a given channel. Taking the overt case of SubC as an example, Figure 3-5 shows the spatial distribution of the selected wavelet coefficients for four different classes.

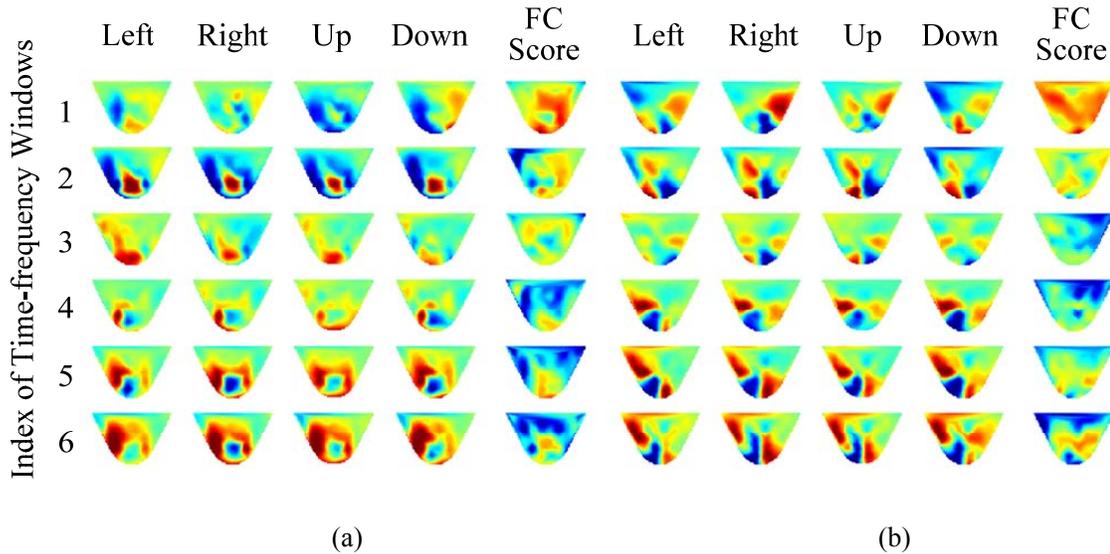


Figure 3-5. (a) The spatial distribution of wavelet coefficients from the first gradiometer, and (b) the spatial distribution of wavelet coefficients from the second gradiometer. In both (a) and (b), the first four columns correspond to the four movement directions, and the last column shows the Fisher scores (FC). Each color map was calculated by averaging the wavelet coefficients over all trials of the same class. Each row corresponds to the wavelet coefficients associated with the same time-frequency window. Red color indicates large value and blue color indicates small value. All plots of wavelet coefficients share the same color scale, and all plots of Fisher scores share the same color scale.

3.4.3 Feature Clustering

Given the MEG features extracted in the previous section, we apply Algorithm 3-1 to partition these features into several groups. The optimal number of groups (i.e., K_{OPT}) is automatically determined by evaluating the clustering quality in Algorithm 3-1. Note that K_{OPT} can be different for different data sets. In Figure 3-6, we show an example of the quality function $q(K)$ for the overt case of SubA. For this example $K_{OPT}=5$, where the quality function $q(K)$ achieves the maximum.

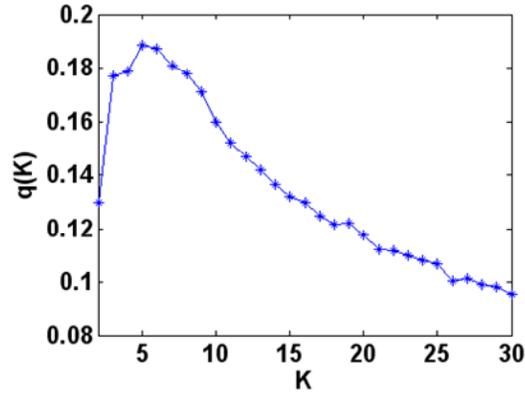


Figure 3-6. The quality function $q(K)$ where K represents the number of clusters.

Figure 3-7 shows the adjacency matrices (i.e., \mathbf{W} in (3.5)) for all ten data sets. In Figure 3-7, the MEG features are ordered based on the feature clustering results where all features in the same cluster are grouped together.

Studying Figure 3-7, we have two important observations. First, once the features are appropriately ordered, the adjacency matrices are almost block diagonal. Remember that the adjacency matrix \mathbf{W} in (3.5) contains the correlation coefficients for all features. Hence, a block diagonal \mathbf{W} implies that different features in different groups are uncorrelated. In other words, a unique group structure of feature correlation exists for all data sets collected by our experiment. Second, the proposed feature clustering algorithm (i.e., Algorithm 3-1) successfully identifies the appropriate feature groups for our data sets. Both the optimal number of groups and the appropriate group assignment for each feature are successfully found by Algorithm 3-1. Such a feature clustering scheme will eventually lead to superior decoding accuracy of the proposed CLDA algorithm, as will be discussed in detail in the next section.

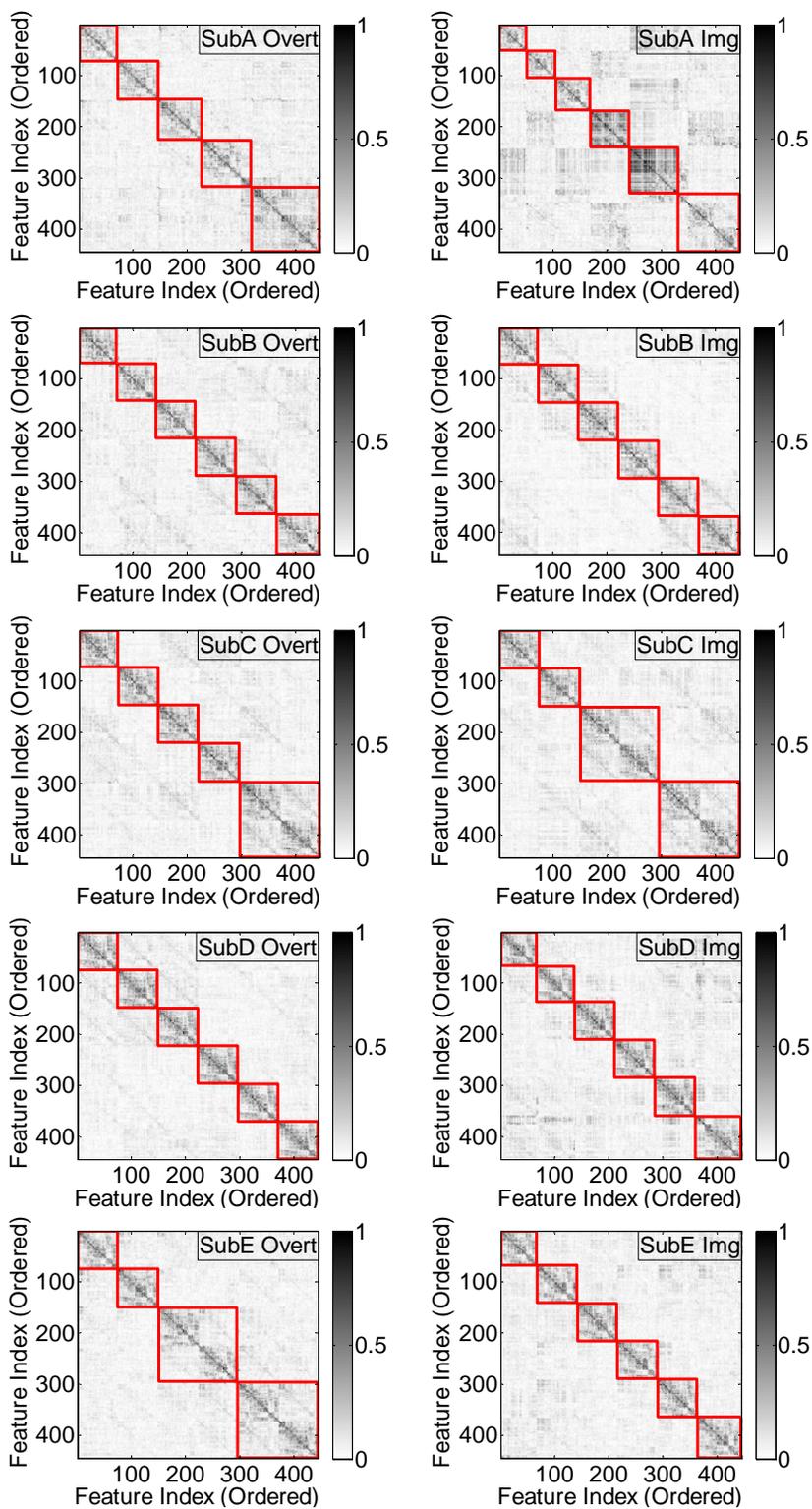


Figure 3-7. The adjacency matrices are almost block diagonal for all data sets. The MEG features are ordered based on the feature clustering results where all features in the same cluster are grouped together.

3.4.4 Movement Decoding

We implement four different decoding algorithms for the purpose of comparison. Since CLDA is a variation of LDA, in this study we only compare it with traditional LDA-based methods, although a number of other classification algorithms (e.g., support vector machine [30], logistic regression [32]) have been applied for MEG decoding as well. For each decoding method, its accuracy is estimated by using leave-one-out cross-validation [52], where feature selection and/or feature clustering are repeatedly applied for each run within the cross-validation loop.

1. **FC-LDA:** Apply Fisher criterion [52] to select a set of important features for dimension reduction. Next, LDA is used for movement decoding. During the feature selection phase, the optimal number of required features is determined via an extra cross-validation step using the training data only.
2. **DLDA:** Assume mutual independence among all features and force the within-class scatter matrix \mathbf{S}_W in (3.1) to be diagonal. Next, LDA is used for movement decoding based on the diagonal approximation of \mathbf{S}_W .
3. **RLDA:** An additional regularization term is added to the within-class scatter matrix: $\mathbf{S}_W + \lambda \cdot \mathbf{I}$. The regularization parameter $\lambda \geq 0$ is determined via an extra cross-validation step using the training data only. Next, LDA is used for movement decoding based on the regularized scatter matrix $\mathbf{S}_W + \lambda \cdot \mathbf{I}$.
4. **CLDA:** Algorithm 3-2 is applied for movement decoding.

First, we consider a simple two-class decoding problem where the movement direction is either left or right. Figure 3-8(a) shows the accuracy of the aforementioned four decoding algorithms. Note that CLDA outperforms the other three methods for all data sets. The decoding accuracy of CLDA is above 90% for all overt cases and it is above 80% for all imagined cases. The average decoding accuracy of CLDA is 97.3% and 94.5% for overt and imagined cases, respectively.

Next, we consider the four-class decoding problem where the movement direction can be up, down, left or right. The accuracy of four-class movement decoding is shown in Figure 3-8(b). Similar

to the two-class case, CLDA offers the best decoding accuracy in all test cases. The average decoding accuracy of CLDA is 90.2% and 83.7% for overt and imagined cases, respectively.

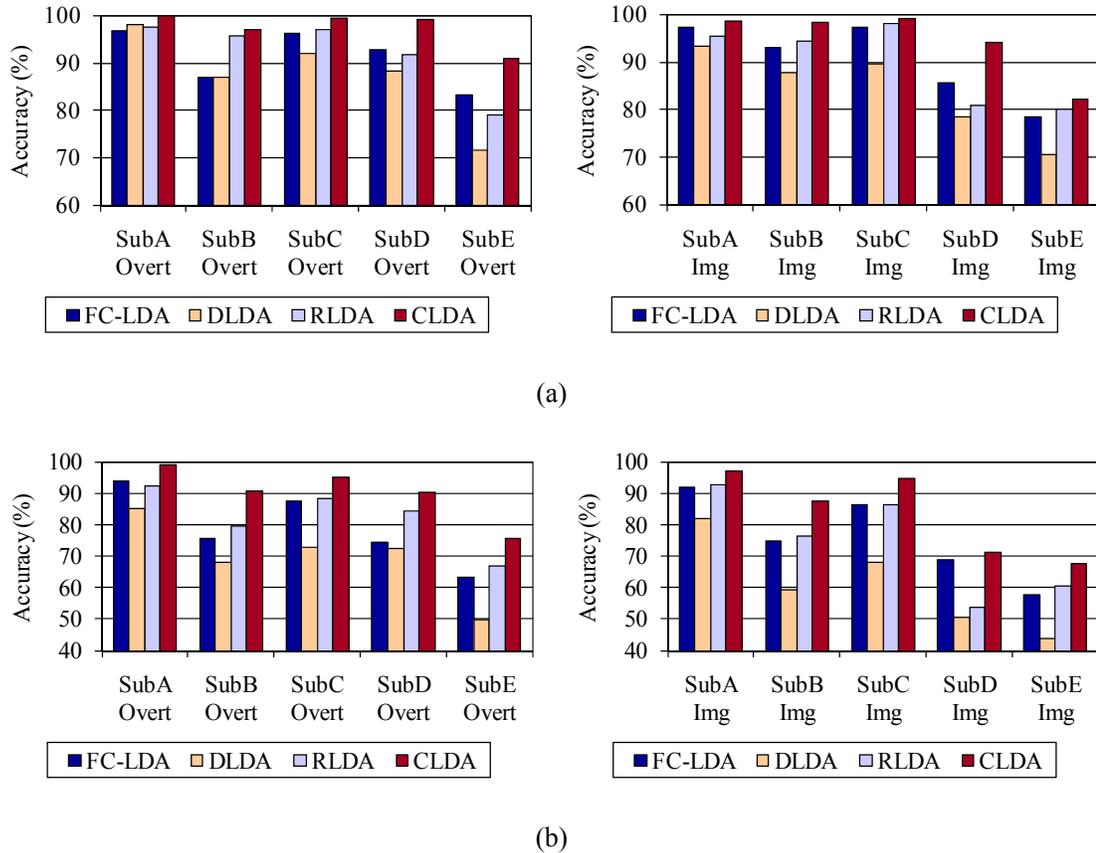


Figure 3-8. Movement decoding accuracy estimated by using leave-one-out cross-validation: (a) decoding results of two movement directions (left and right), and (b) decoding results of four movement directions (up, down, left and right).

As previously mentioned, both DLDA and RLDA rely on the prior knowledge that all features are mutually independent. Unlike DLDA or RLDA, CLDA is able to automatically identify the critical correlation structure (i.e., by partitioning all features into different groups) and then accurately extract the correlation (i.e., by estimating a block diagonal within-class scatter matrix) of all features. It, in turn, results in improved decoding accuracy over DLDA and RLDA.

On the other hand, FC-LDA applies Fisher criterion to select a small subset of important features. Such a feature selection method can be pessimistic, when there are a large number of

important features and the training data are limited. In other words, given a small set of training data, FC-LDA has to filter out many useful features in order to reduce the dimensionality of the feature space and avoid over-fitting. These useful features carry the information that is needed for decoding. Since they are simply ignored by FC-LDA, the resulting decoding error becomes large. This is the primary reason why FC-LDA is less accurate than CLDA in our experiments.

3.5 Discussion

In this section, we further discuss the advantages and limitations of the proposed CLDA algorithm from both theoretical and practical points of view. In particular, we aim to explain the reason why the correlation of our MEG features shows a unique group structure and the reason why CLDA offers superior decoding accuracy by approximating a block diagonal within-class scatter matrix.

3.5.1 Correlation Structure

To understand the unique correlation structure of our MEG features, Figure 3-9 plots the group assignment of each feature that is determined by the proposed feature clustering algorithm (i.e., Algorithm 3-1). In our experiment, each feature (i.e., each wavelet coefficient) is associated with a particular time window of a particular channel, because the corresponding wavelet basis functions are in the same low-frequency band and have local support in different time windows.

Studying Figure 3-9, we notice that most features of the same group are in the same time window or adjacent time windows, but from different channels. Remember that our proposed feature clustering algorithm attempts to maximize the within-group correlation and minimize the between-group correlation. It, in turn, implies that the MEG features from the same time window but different channels are strongly correlated, while the features from different time windows are weakly correlated. In other words, our MEG data sets present a strong spatial correlation but a weak temporal correlation.

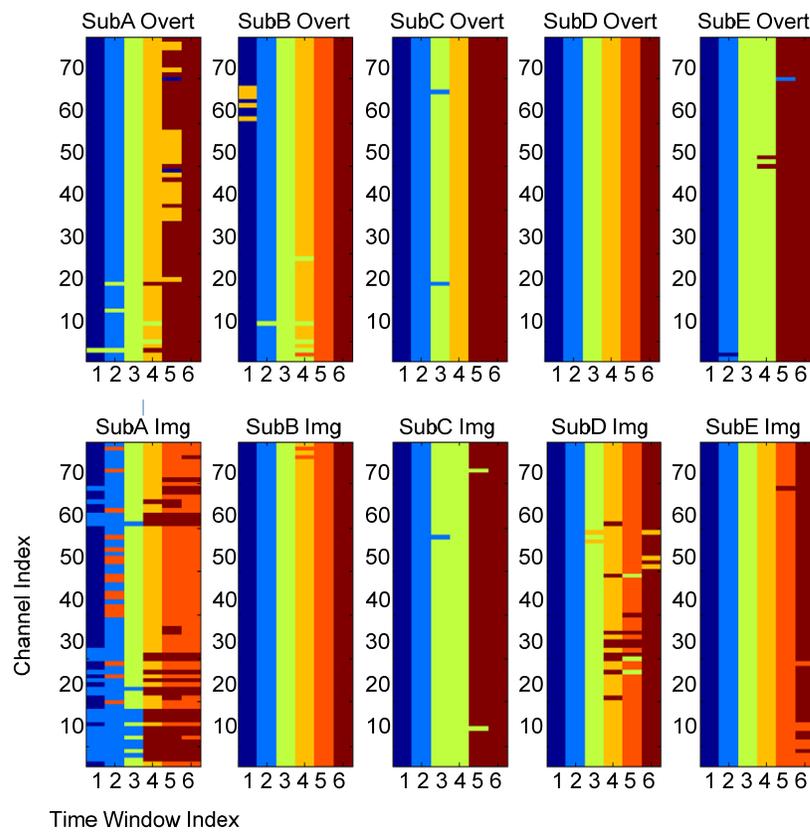


Figure 3-9. Feature groups determined by the spectral clustering algorithm. The horizontal axes represent the indexes of wavelet coefficients. They can also be considered as the indexes of time windows, since the corresponding wavelet basis functions are associated with the same low-frequency band and have local support in different time windows. The vertical axes represent the indexes of 74 selected channels (i.e., the gradiometers on top of the sensorimotor area). Different colors indicate different feature groups.

Similar observations of strong spatial correlation for MEG data have been reported in many other applications [3]. From the physics point of view, MEG signals are generated by the primary current sources inside the brain. Based on Maxwell's equations, the magnetic field created by the same current source can propagate to multiple MEG sensors (i.e., the superconducting quantum interference devices) corresponding to different channels at different locations [3]. This is an important reason why a strong spatial correlation has been observed for MEG measurement data.

On the other hand, to explain the weak temporal correlation, we consider the following model for the time-domain signal $s_i(t)$ of the i th channel

$$s_i(t) = v_i(t) + n_i(t), \quad (3.16)$$

where $v_i(t)$ stands for the phase-locked evoked response associated with the stimulus and $n_i(t)$ denotes the ongoing activity. The linear model in (3.16) has been used in several previous studies to analyze the measured neural signals in time domain [66]-[67]. The evoked response $v_i(t)$ in (3.16) can be estimated by averaging $s_i(t)$ over all trials corresponding to the same stimulus [67]. Once $v_i(t)$ is estimated and subtracted from $s_i(t)$, the resulting residue $n_i(t)$ does not contain event-related information and is often modeled as white noise [66]-[67]. In other words, while $s_i(t)$ is a colored signal due to the evoked response $v_i(t)$, $n_i(t)$ is white once $v_i(t)$ is removed from $s_i(t)$. It, hence, explains the weak temporal correlation that we observe, since the within-class scatter matrix \mathbf{S}_W in (3.1) is mainly determined by $n_i(t)$.

Finally, it is important to note that while the between-group correlation is weak after feature clustering, it is not exactly zero, as shown in Figure 3-7. Namely, the within-group scatter matrix \mathbf{S}_W is not exactly block diagonal. Since CLDA uses a block diagonal approximation to model \mathbf{S}_W for movement decoding, it is important for us to study the accuracy of such an approximation and, most importantly, its impact on the final decoding accuracy. For this reason, a comprehensive discussion will be provided in the next section to study the accuracy of the proposed block diagonal approximation.

3.5.2 Block Diagonal Approximation

To theoretically analyze the quality of the block diagonal approximation, we consider a simple two-class movement decoding problem with the following assumptions:

1. The prior probability for each class (i.e., the probability for each class to occur) is identical.
2. The probability distributions of MEG features for both classes are multivariate Gaussian. These distributions have different mean values (denoted as \mathbf{m}_1 and \mathbf{m}_2 respectively), but share the

same covariance matrix Σ .

These two assumptions are not necessarily valid for all movement decoding problems; however, they define a simple classification problem for which we can show many insights on the proposed block diagonal approximation.

Given the aforementioned two-class decoding problem, it is straightforward to verify that the minimal decoding error (i.e., the probability of misclassification) of an optimal classifier is [56]

$$e_{OPT} = 1 - \varphi(\rho_{OPT}), \quad (3.17)$$

where $\varphi(\bullet)$ denotes the cumulative distribution function of standard Gaussian distribution (i.e., zero mean and unit variance), and ρ_{OPT} is defined as

$$\rho_{OPT} = \frac{1}{2} \cdot \sqrt{(\mathbf{m}_2 - \mathbf{m}_1)^T \cdot \Sigma^{-1} \cdot (\mathbf{m}_2 - \mathbf{m}_1)}. \quad (3.18)$$

Studying (3.17)-(3.18), we would have two important observations. First, if the difference of \mathbf{m}_1 and \mathbf{m}_2 increases, ρ_{OPT} in (3.18) increases and, hence the decoding error in (3.17) decreases. Second, the decoding error also decreases, if the variance of the MEG features (measured by the covariance matrix Σ) decreases. These two observations are consistent with our intuition. Namely, the decoding error is small, if the two classes are substantially different (measured by the difference of \mathbf{m}_1 and \mathbf{m}_2) or the trial-to-trial variation is small (measured by the covariance matrix Σ).

On the other hand, if the covariance matrix Σ is approximated by its block diagonal form Σ_B , the decoding error of the proposed CLDA algorithm becomes

$$e_B = 1 - \varphi(\rho_B), \quad (3.19)$$

where

$$\rho_B = \frac{1}{2} \cdot \frac{(\mathbf{m}_2 - \mathbf{m}_1)^T \cdot \Sigma_B^{-1} \cdot (\mathbf{m}_2 - \mathbf{m}_1)}{\sqrt{(\mathbf{m}_2 - \mathbf{m}_1)^T \cdot \Sigma_B^{-1} \cdot \Sigma \cdot \Sigma_B^{-1} \cdot (\mathbf{m}_2 - \mathbf{m}_1)}}. \quad (3.20)$$

Eq. (3.19)-(3.20) can be derived by directly following the results in [56]. To study the difference between e_{OPT} in (3.17) and e_B in (3.19), we further define

$$\begin{aligned}
r &= \frac{\rho_B}{\rho_{OPT}} \\
&= \frac{(\mathbf{m}_2 - \mathbf{m}_1)^T \cdot \Sigma_B^{-1} \cdot (\mathbf{m}_2 - \mathbf{m}_1)}{\sqrt{(\mathbf{m}_2 - \mathbf{m}_1)^T \cdot \Sigma_B^{-1} \Sigma \Sigma_B^{-1} \cdot (\mathbf{m}_2 - \mathbf{m}_1) \cdot (\mathbf{m}_2 - \mathbf{m}_1)^T \cdot \Sigma^{-1} \cdot (\mathbf{m}_2 - \mathbf{m}_1)}}, \quad (3.21) \\
&= \frac{\boldsymbol{\delta}_0^T \boldsymbol{\delta}_0}{\sqrt{(\boldsymbol{\delta}_0^T \Sigma_0 \boldsymbol{\delta}_0) \cdot (\boldsymbol{\delta}_0^T \Sigma_0^{-1} \boldsymbol{\delta}_0)}}
\end{aligned}$$

where

$$\boldsymbol{\delta}_0 = \Sigma_B^{-1/2} \cdot (\mathbf{m}_2 - \mathbf{m}_1) \quad (3.22)$$

$$\Sigma_0 = \Sigma_B^{-1/2} \cdot \Sigma \cdot \Sigma_B^{-1/2}. \quad (3.23)$$

The value of r in (3.21) indicates the decoding accuracy of CLDA, as compared to an optimal classifier. If r is close to 1, the accuracy of CLDA is close to the maximal accuracy that can be achieved by the optimal classifier.

Based on the Kantorovich inequality [56], [59], we can obtain a lower bound of r

$$r \geq \frac{2 \cdot \sqrt{\xi_0}}{1 + \xi_0}, \quad (3.24)$$

where ξ_0 represents the condition number of the matrix Σ_0 in (3.23) based on L_2 norm. Combining (3.19), (3.21) and (3.24), we can derive the upper bound of the decoding error for CLDA

$$e_{OPT} \leq e_B \leq e_{UP} = 1 - \varphi \left(\frac{2 \cdot \sqrt{\xi_0} \cdot \rho_{OPT}}{1 + \xi_0} \right). \quad (3.25)$$

Figure 3-10 shows the relation between e_{OPT} and e_{UP} when the condition number ξ_0 takes different values. Note that the accuracy of CLDA strongly depends on ξ_0 . If ξ_0 is sufficiently small, e_{UP} is close to e_{OPT} . Namely, the accuracy of CLDA is close to that of an optimal classifier. In the extreme case, if the covariance matrix Σ is exactly block diagonal, we have $\Sigma_B = \Sigma$, $\Sigma_0 = \mathbf{I}$ and $\xi_0 = 1$. Therefore, CLDA is equivalent to the optimal classifier and both of them yield the same decoding accuracy. The aforementioned results provide theoretical guidelines to quantitatively assess the quality of the proposed block diagonal approximation for CLDA.

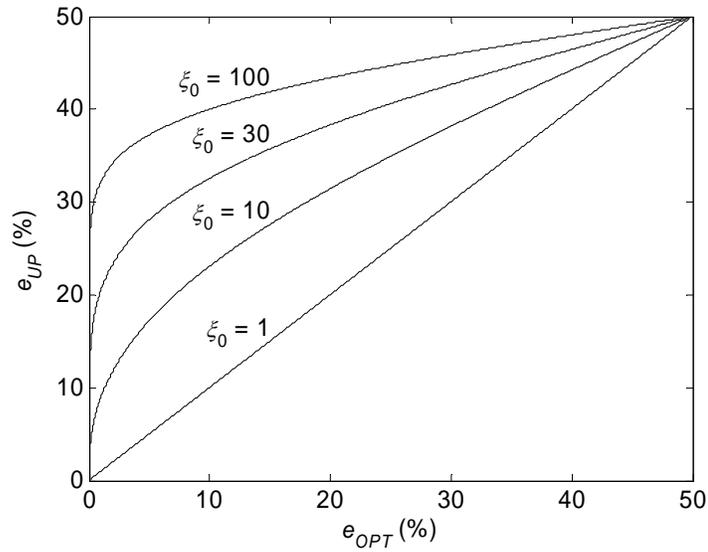


Figure 3-10. The upper bound of the CLDA error e_{UP} is close to the minimal error e_{OPT} that can be achieved by an optimal classifier, if the condition number ξ_0 of the matrix Σ_0 is sufficiently small.

3.6 Summary

To investigate task-related activity using MEG, a preliminary problem is whether the task-related activity is detectable in the MEG recordings. In this chapter, we put forward that MEG decoding can be used as an instrument to address this problem. High decoding accuracy indicates that strong task-related activity exists. However, low accuracy might be because of the poor performance of the decoding algorithm. Thus it is important to develop high performance decoding algorithms to efficiently extract the discriminant information between MEG recordings associated with different brain states.

Due to the high temporal and spatial resolution, MEG signals are generally transformed into a huge number of features. However, in practice the collected training data are often limited. Namely, the problem posed by MEG decoding is high-dimensional but with a small training set. In this case, the decoding algorithm must be carefully designed to prevent the decoder from over-fitting the training data. In this chapter, we propose a clustering linear discriminant analysis (CLDA) algorithm

for MEG decoding. CLDA applies a spectral clustering algorithm to partition all MEG features into several groups where the within-group correlation is maximized and the between-group correlation is minimized. As such, the covariance matrix can be approximated as a block diagonal form that can be accurately estimated from a small set of training data. The efficiency of CLDA is studied by both theoretical analyses and practical examples. Our MEG-based movement decoding demonstrates that the average accuracy of CLDA is 87% for single-trial movement decoding of four directions (i.e., up, down, left and right).

Chapter 4

Task-related Source Localization

4.1 Motivation

In the previous chapter, we present an algorithm for MEG decoding with a small training set. MEG decoding can be used to check whether task-related activity is detectable in the MEG recordings. However, it does not uncover the underlying task-related neural activity in the brain space. In this chapter, we focus on the problem of localizing the sources associated with task-related activity. In other words, we want to estimate the activation map of the neural sources that differentiate two different brain states. This is referred to as task-related source localization in the thesis. It is an important tool that helps us to compare different brain states and study the task-related neural activity.

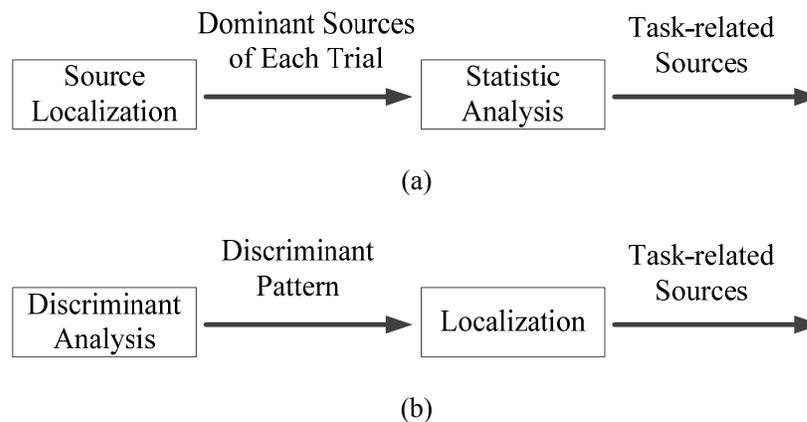


Figure 4-1. (a) Flow of the conventional inferential statistics approaches. (b) Flow of the proposed discriminant pattern source localization algorithm.

Traditionally, inferential statistics have been applied to estimate task-related sources [14], [15], [22]. As shown in Figure 4-1(a), these techniques are composed of two phases. First, a source

localization algorithm (e.g., MNE [18]) is used to estimate the source distribution associated with each brain state. Next, statistical hypothesis testing is applied to determine if the sources estimated from the first phase carry the information to distinguish different brain states.

The aforementioned techniques based on inferential statistics heavily rely on the accuracy of the source localization algorithm that is applied during the first phase. In other words, the hypothesis testing in the second phase is meaningful, if and only if MEG sources are accurately estimated in the first phase. In practice, as discussed in Section 2.3, it is well-known that the source localization problem is profoundly underdetermined due to the limited observability of MEG measurements (i.e., the limited number of MEG sensors relative to the number of possible neural sources) [9]. Hence, it is difficult to estimate all sources in the first phase, especially if the signal-to-noise ratio (SNR) is low. In most cases, a source localization algorithm such as MNE [18] can capture the *dominant* sources which are *not* necessarily associated with the discriminant pattern for different brain tasks. For this reason, the traditional techniques based on inferential statistics may not correctly estimate the sources corresponding to task-related activity. It, in turn, poses an immediate need to re-think the fundamental strategy of source localization and develop a new algorithm for the problem of task-related source localization.

Towards this goal, we propose a novel discriminant pattern source localization (DPSL) algorithm to estimate the activation map of task-related sources in this chapter. As shown in Figure 4-1(b), the proposed DPSL algorithm consists of two major steps. First, discriminant analysis is applied to create a filter that can optimally differentiate different brain states in the sensor space. Second, based on the forward model of the magnetic field, the distributed activation map of the task-related sources is computed by studying the response of the filter. Unlike the traditional inferential statistics methods that estimate the dominant sources of each brain states followed by a statistical hypothesis testing to determine the task-related sources, our proposed DPSL algorithm identifies the task-related sources by applying an optimal discriminant analysis in the first step. Since the discriminant analysis algorithms are particularly designed to be robust to noise [30], [31], [34], [54], DPSL can efficiently reduce the impact of noise and accurately identify the task-related sources, as will be demonstrated by

the examples in Section 4.4.

In addition, although most traditional source localization methods [16]-[20] are designed to estimate the spatial source distribution within a short time window and a specific frequency band, it is equally important to capture the time and/or frequency dynamics of these sources in practice [21], [55]. In other words, the task-related sources may vary over time and/or across different frequency bands and, hence, new methods must be developed to estimate the source distribution over time and/or across frequency bands. The proposed framework of DPSL can be easily extended for this purpose. Particularly, discriminant analysis can be first applied to create a filter in time/frequency domain and then the task-related sources are estimated according to the response of the filter. More details of the extended DPSL algorithm will be discussed in Section 4.3.

The remainder of this chapter is organized as follows. In Section 4.2, we develop the basic framework of DPSL and then extend DPSL to time/frequency domain in Section 4.3. The accuracy of DPSL is compared with other traditional methods by several simulation examples in Section 4.4. In Section 4.5, we further demonstrate the efficiency of DPSL by studying the MEG data collected from a human subject. Finally, we summarize this chapter in Section 4.6.

4.2 Discriminant Pattern Source Localization

As mentioned in the previous section, DPSL consists of two major steps: (i) discriminant analysis, and (ii) task-related source imaging. In this section, we first describe the technical details of discriminant analysis. Next, we further discuss the source imaging methods corresponding to different feature extraction approaches used by the discriminant analysis.

4.2.1 Discriminant Analysis

Without loss of generality, we consider two brain states that are labeled as “State-A” and “State-B”, respectively. For example, these two states may correspond to the neural activity when a human subject performs two different movement tasks. We further assume that MEG signals are

recorded from M channels. We use a vector $\mathbf{x} \in R^M$ to represent the MEG features corresponding to these M channels, where x_m is the m th element of \mathbf{x} and it denotes the feature associated with the m th MEG channel. In this section, we assume that the task-related sources within a short time-frequency window is fixed and consider \mathbf{x} as a transformation of the MEG measurements over this time-frequency window. The transformation can be either linear (e.g. wavelet transform [45]) or nonlinear (e.g. power spectral density estimation [50]). In Section 4.3, we will further extend DPSL to accommodate the cases where task-related sources vary over time and/or frequency. Several examples of representing the measured MEG signals as a feature vector \mathbf{x} can be found in Section 4.4-4.5.

With the feature vector \mathbf{x} , the objective of discriminant analysis is to find a discriminant function $F(\mathbf{x})$ that can appropriately distinguish the two brain states:

$$F(\mathbf{x}) = \begin{cases} \geq 0 & (\text{if "State - A"}) \\ < 0 & (\text{if "State - B"}) \end{cases} \quad (4.1)$$

The aforementioned discriminant analysis is essentially a binary classification problem. A large number of machine learning techniques, such as linear discriminant analysis (LDA) [31], support vector machine (SVM) [30], logistic regression [32], Bayesian classifier [54], etc., can be used to construct the discriminant function $F(\mathbf{x})$ in (4.1).

Taking linear L1-norm SVM as an example, it uses the MEG features collected from repeated experimental trials as the input. Given the feature vectors $\{\mathbf{x}_{l,1}; l = 1,2,\dots,L\}$ and $\{\mathbf{x}_{l,2}; l = 1,2,\dots,L\}$ from L different trials of “State-A” and “State-B” respectively, a linear L1-norm SVM aims to find the following discriminant function to distinguish the two brain states [30]:

$$F(\mathbf{x}) = \mathbf{w}^T \cdot \mathbf{x} + c \quad (4.2)$$

where $\mathbf{w} \in R^M$ and $c \in R$ stand for the SVM coefficients and they can be found by solving a convex optimization problem:

$$\begin{aligned}
& \underset{\mathbf{w}, c, \xi}{\text{minimize}} && \sum_{l=1}^{2 \cdot L} \xi_l \\
& \text{subject to} && \|\mathbf{w}\|_1 \leq \lambda \\
& && \mathbf{w}^T \cdot \mathbf{x}_{l,1} + c \geq 1 - \xi_l \quad (l = 1, \dots, L) \\
& && -(\mathbf{w}^T \cdot \mathbf{x}_{l,2} + c) \geq 1 - \xi_{l+L} \quad (l = 1, \dots, L) \\
& && \xi_l \geq 0 \quad (l = 1, \dots, 2 \cdot L)
\end{aligned} \tag{4.3}$$

In (4.3), $\|\mathbf{w}\|_1$ denotes the L1-norm of the vector \mathbf{w} , $\{\xi_l; l = 1, 2, \dots, 2 \cdot L\}$ are slack variables and λ is a regularization parameter that can be determined by cross-validation [30]. More details on L1-norm SVM can be found in [30].

While we use linear L1-norm SVM as an example to illustrate the basic idea of discriminant analysis in this chapter, it is important to note that many other linear classification techniques used in [30], [31], [34], [54] can also be applied to find the discriminant function $F(\mathbf{x})$ in (4.1)-(4.2). Extending DPSL to other linear classification methods is straightforward and, hence, the details are not included in this thesis.

4.2.2 Source Imaging Based on Linear Transform Features

As discussed in Section 4.2.1, the feature vector \mathbf{x} is a transformation of the MEG measurements over a short time-frequency window. In this section, we consider the cases where the transformation is linear (e.g. wavelet transform [45]).

The discriminant function $F(\mathbf{x})$ in (4.2) is a linear combination of multiple MEG features $\{x_m; m = 1, 2, \dots, M\}$. Remember that the m th feature x_m is associated with the m th MEG channel. It, in turn, implies that the discriminant function $F(\mathbf{x})$ can be conceptually viewed as a spatial filter $\mathbf{w}^T \cdot \mathbf{x}$ that extracts the discriminant information from all MEG channels to distinguish different brain states. For this reason, the spatial filter $\mathbf{w}^T \cdot \mathbf{x}$ should amplify the MEG signals generated by task-related sources and simultaneously attenuate the signals coming from non-task-related sources. Based upon this assumption, we can estimate the activation map of the task-related sources by evaluating the response of the spatial filter $\mathbf{w}^T \cdot \mathbf{x}$ for every possible source. Next we will describe the proposed source localization algorithm in detail.

We represent a human brain surface as a number of vertices. Each vertex contains a current dipole that models the MEG source at the vertex. Such a vertex-based model has been widely adopted by many source localization algorithms [19]-[22]. In this chapter, we apply this model with N vertices and, hence, N current dipoles. These current dipoles generate the magnetic fields that are measured by M MEG channels. The MEG measurement $\mathbf{s}(t) \in R^M$ collected from M channels at the time t can be described as a linear combination of the dipole moments [9]:

$$\mathbf{s}(t) = \sum_{n=1}^N \mathbf{A}_n \cdot \mathbf{q}_n(t), \quad (4.4)$$

where $\mathbf{q}_n(t) \in R^3$ denotes the moment of the n th dipole at the time t , and $\mathbf{A}_n \in R^{M \times 3}$ is the leadfield matrix associated with the location of the n th dipole. The matrix \mathbf{A}_n models the influence of the n th dipole on the measured MEG signals. It can be calculated according to the geometrical structure and the conducting medium of the human head [9].

On the other hand, the feature vector \mathbf{x} represents a linear transformation of the MEG measurements over a time window. Without loss of generality, we denote this linear transformation as $\mathcal{L}(\bullet)$:

$$\mathbf{x} = \mathcal{L}(\mathbf{s}(t)). \quad (4.5)$$

In a special case where the feature vector \mathbf{x} represents the MEG measurements at a particular time, the linear transformation is simply an identify function: $\mathcal{L}(\mathbf{s}(t)) = \mathbf{s}(t)$. Since the transformation $\mathcal{L}(\bullet)$ is linear, substituting (4.4) into (4.5) yields:

$$\mathbf{x} = \mathcal{L}\left(\sum_{n=1}^N \mathbf{A}_n \cdot \mathbf{q}_n(t)\right) = \sum_{n=1}^N \mathbf{A}_n \cdot \mathcal{L}(\mathbf{q}_n(t)). \quad (4.6)$$

Define:

$$\mathbf{q}_n(t) = r_n(t) \cdot \mathbf{v}_n, \quad (4.7)$$

where the scalar $r_n(t) \in R$ and the unit vector $\mathbf{v}_n \in R^3$ represent the magnitude and the orientation of $\mathbf{q}_n(t)$ respectively. Here, we assume that the dipole orientation \mathbf{v}_n is determined by the anatomic structure of the brain [3], [23] and, hence, it does not vary over time. Such a dipole model is referred to as the fixed-orientation model [9] and has been widely used for MEG source localization in the

literature. Combining (4.6) and (4.7), we have:

$$\mathbf{x} = \sum_{n=1}^N \mathbf{A}_n \cdot \mathbf{v}_n \cdot \mathcal{L}(r_n(t)) = \sum_{n=1}^N \mathbf{A}_n \cdot \mathbf{v}_n \cdot \phi_n, \quad (4.8)$$

where $\Phi_n = \mathcal{L}(r_n(t))$ is the linear transformation of the magnitude of the n th dipole over a time window.

Next, we consider the discriminant function $F(\mathbf{x})$ in (4.2) which is essentially a spatial filter applied to the MEG feature vector \mathbf{x} . Substituting (4.8) into (4.2) yields:

$$F(\mathbf{x}) = \sum_{n=1}^N \mathbf{w}^T \cdot \mathbf{A}_n \cdot \mathbf{v}_n \cdot \phi_n + c. \quad (4.9)$$

As previously discussed, the optimal spatial filter $\mathbf{w}^T \cdot \mathbf{x}$ should amplify the MEG signals generated by task-related current dipoles and simultaneously attenuate the signals coming from non-task-related dipoles. Based upon this assumption, we propose to calculate the gain of the spatial filter $\mathbf{w}^T \cdot \mathbf{x}$ for each dipole and use it as a quantitative metric to assess the contribution of each dipole to the discriminant function $F(\mathbf{x})$:

$$g_n = \left| \mathbf{w}^T \cdot \mathbf{A}_n \cdot \mathbf{v}_n \right| \quad (n = 1, 2, \dots, N). \quad (4.10)$$

If the value of g_n in (4.10) is large, it implies that the n th dipole is “selected” by the spatial filter $\mathbf{w}^T \cdot \mathbf{x}$ and, hence, it carries the discriminant information for different brain states. On the other hand, if the value of g_n is small, it means that the n th dipole is “neglected” by the spatial filter $\mathbf{w}^T \cdot \mathbf{x}$ and, hence, it is weakly correlated with the task-related modulation. By calculating the gain values $\{g_n; n = 1, 2, \dots, N\}$ for all dipoles, we can indirectly obtain the activation map of the task-related sources. Hence, calculating the gain values in (4.10) is one of the major steps of the proposed DPSL algorithm.

Studying (4.10), one would notice that we must know the orientation \mathbf{v}_n for each dipole in order to calculate the gain g_n . In practice, there are several possible approaches that we can take to determine the orientation \mathbf{v}_n .

1. Several previous works [16], [24], [25] mention that the dipole orientation at a specific spatial location should be physiologically determined, since a current dipole represents a given neuroanatomical structure. In particular, several existing source localization algorithms assume that all current dipoles are normal to the cortical surface [25]. In this case, we can determine the

orientation \mathbf{v}_n of each dipole according to the geometrical structure of the cortical surface.

2. A close look at (4.10) reveals an important fact that the gain g_n is zero if the vector \mathbf{v}_n is orthogonal to the vector $\mathbf{A}_n^T \cdot \mathbf{w}$. Namely, the spatial filter $\mathbf{w}^T \cdot \mathbf{x}$ is orientation-selective. In general, we can decompose the dipole moment $\mathbf{q}_n(t)$ into three orthogonal components:

$$\mathbf{q}_n(t) = r_{\parallel,n}(t) \cdot \mathbf{v}_{\parallel,n} + r_{\perp,1,n}(t) \cdot \mathbf{v}_{\perp,1,n} + r_{\perp,2,n}(t) \cdot \mathbf{v}_{\perp,2,n}, \quad (4.11)$$

where $\mathbf{v}_{\parallel,n}$ represents the unit vector parallel to $\mathbf{A}_n^T \cdot \mathbf{w}$, $\mathbf{v}_{\perp,1,n}$ and $\mathbf{v}_{\perp,2,n}$ denote two other unit vectors orthogonal to $\mathbf{A}_n^T \cdot \mathbf{w}$, and these three vectors $\mathbf{v}_{\parallel,n}$, $\mathbf{v}_{\perp,1,n}$ and $\mathbf{v}_{\perp,2,n}$ form an orthonormal basis of the three-dimensional space. The spatial filter $\mathbf{w}^T \cdot \mathbf{x}$ only “picks up” one of these three components (i.e., $r_{\parallel,n}(t) \cdot \mathbf{v}_{\parallel,n}$) that is parallel to $\mathbf{A}_n^T \cdot \mathbf{w}$. Since the other two components $r_{\perp,1,n}(t) \cdot \mathbf{v}_{\perp,1,n}$ and $r_{\perp,2,n}(t) \cdot \mathbf{v}_{\perp,2,n}$ do not contribute to the gain g_n in (4.10), we simply calculate g_n by considering the orientation of $\mathbf{v}_{\parallel,n}$ only:

$$g_n = \left| \mathbf{w}^T \cdot \mathbf{A}_n \cdot \mathbf{v}_{\parallel,n} \right| \quad (n = 1, 2, \dots, N). \quad (4.12)$$

Since the unit vector $\mathbf{v}_{\parallel,n}$ is parallel to $\mathbf{A}_n^T \cdot \mathbf{w}$, it can be re-written as:

$$\mathbf{v}_{\parallel,n} = \frac{\mathbf{A}_n^T \cdot \mathbf{w}}{\left\| \mathbf{A}_n^T \cdot \mathbf{w} \right\|_2}, \quad (4.13)$$

where $\|\cdot\|_2$ denotes the L2-norm of a vector. Substitute (4.13) into (4.12) and we get:

$$g_n = \left\| \mathbf{w}^T \cdot \mathbf{A}_n \right\|_2 \quad (n = 1, 2, \dots, N). \quad (4.14)$$

Mathematically, Eq. (4.14) can also be derived from the following optimization problem:

$$\begin{aligned} & \underset{\mathbf{v}_n}{\text{maximize}} && \mathbf{w}^T \cdot \mathbf{A}_n \cdot \mathbf{v}_n \\ & \text{subject to} && \left\| \mathbf{v}_n \right\|_2 = 1 \end{aligned} \quad (4.15)$$

It results in the maximum possible gain g_n where the dipole orientation \mathbf{v}_n is parallel to the vector $\mathbf{A}_n^T \cdot \mathbf{w}$ in (4.10).

For our simulation and experimental results shown in this study, we determine the orientation \mathbf{v}_n using the second method. However, it should be noted that different choices of dipole orientation lead to different results. Hence, it is extremely important to appropriately interpret the results when

different models are applied. Algorithm 4-1 lists the major steps of the proposed DPSL algorithm.

Algorithm 4-1: DPSL for Task-related Source Imaging Based on Linear Transform Features

1. Start from the training data corresponding to two different brain states.
2. For each trial of the training data, extract the feature vector $\mathbf{x} \in R^M$ for discriminant analysis, where M represents the total number of MEG channels. The feature vector \mathbf{x} can be a collection of MEG signals at a particular time, or a linear transformation (e.g. wavelet transform [45]) of the MEG measurements over a time window. Here, we assume that a single feature is extracted from each MEG channel.
3. Build the discriminant function $F(\mathbf{x})$ in (4.2) to distinguish two different brain states using classification algorithms.
4. Compute the leadfield matrices $\{\mathbf{A}_n; n = 1, 2, \dots, N\}$ for all of the N current dipoles associated with the vertex-based model.
5. If the dipole orientation \mathbf{v}_n at the n th vertex is known, compute the corresponding gain value g_n using (4.10). Otherwise, compute the gain value g_n using (4.14).
6. The gain values $\{g_n; n = 1, 2, \dots, N\}$ of all dipoles represent the activation map of the task-related sources.

4.2.3 Source Imaging Based on Non-linear Transform Features

In Section 4.2.2, the features used by DPSL are linear transformations of MEG measurements over a time-frequency window. In addition to linear transformation, nonlinear transformation can also be applied to efficiently extract important features for discriminant analysis [50]. For instance, power spectral density (PSD), which estimates signal variance as a function of frequency, has been widely used for discriminant analysis of MEG [50]. In this section, we will discuss how to use PSD features to image task-related sources within a given time-frequency window. Our proposed approach, however, is not directly applicable to other nonlinear transformations. In our future research, new algorithms and methodologies should be further developed to address this limitation of DPSL.

Again, we use a vector $\mathbf{x} \in R^M$ to represent the MEG features associated with M channels, where x_m is the m th element of \mathbf{x} and it denotes the PSD value associated with the m th channel. However, given the discriminant function $F(\mathbf{x})$ in (4.2), we cannot directly apply (4.10) or (4.14) to compute the gain values, because the feature vector \mathbf{x} is no longer a linear combination of the dipole moments and, hence, Eq. (4.10) and (4.14) are not applicable. We must derive a different set of equations to image task-related sources based on PSD features.

Similar to Section 4.2.2, we assume a vertex-based dipole model and the MEG measurement $s_m(t)$ collected from the m th channel can be written as a linear combination of all dipole moments [9]:

$$s_m(t) = \sum_{n=1}^N \mathbf{A}_n^{(m)} \cdot \mathbf{q}_n(t), \quad (4.16)$$

where $\mathbf{q}_n(t) \in R^3$ denotes the moment of the n th dipole at the time t , $\mathbf{A}_n^{(m)} \in R^3$ is the leadfield matrix corresponding to the n th dipole and the m th channel, and N is the total number of dipoles. Based on (4.16), the auto-correlation function of $s_m(t)$ is expressed as [57]:

$$C_{s_m}(k) = \mathbb{E}[s_m(t) \cdot s_m(t+k)] = \mathbb{E}\left[\left(\sum_{n=1}^N \mathbf{A}_n^{(m)} \cdot \mathbf{q}_n(t)\right) \cdot \left(\sum_{n=1}^N \mathbf{A}_n^{(m)} \cdot \mathbf{q}_n(t+k)\right)\right]. \quad (4.17)$$

Since the PSD of $s_m(t)$ is the Fourier transform of the auto-correlation function [57], the PSD feature x_m can be represented as:

$$x_m = \mathcal{F}(C_{s_m}(k)) = \mathcal{F}\left(\mathbb{E}\left[\left(\sum_{n=1}^N \mathbf{A}_n^{(m)} \cdot \mathbf{q}_n(t)\right) \cdot \left(\sum_{n=1}^N \mathbf{A}_n^{(m)} \cdot \mathbf{q}_n(t+k)\right)\right]\right), \quad (4.18)$$

where $\mathcal{F}(\bullet)$ denotes the Fourier transform. To simplify the expression in (4.18), we assume that (i) all sources are mutually independent, and (ii) the mean value of each source is zero (i.e. $\mathbb{E}[\mathbf{q}_n(t)] = \mathbf{0}$). The first assumption has been applied for source localization in the literature [9]. The second assumption is valid if we remove the DC components of the MEG signals for all channels. This is a standard procedure that is often performed in MEG data analysis by applying a band-pass filter [30] or removing the linear trend [39]. Based on these assumptions, Eq. (4.18) is simplified as:

$$x_m = \mathcal{F}\left(\mathbb{E}\left[\sum_{n=1}^N \left(\mathbf{A}_n^{(m)} \cdot \mathbf{q}_n(t)\right) \cdot \left(\mathbf{A}_n^{(m)} \cdot \mathbf{q}_n(t+k)\right)\right]\right). \quad (4.19)$$

Similar to Eq. (4.7), we use a scalar $r_n(t)$ and a unit vector \mathbf{v}_n to represent the magnitude and the orientation of $\mathbf{q}_n(t)$ respectively. Thus, Eq. (4.19) can be re-written as:

$$x_m = \sum_{n=1}^N (\mathbf{A}_n^{(m)} \cdot \mathbf{v}_n)^2 \cdot \mathcal{F}(\mathbb{E}[r_n(t) \cdot r_n(t+k)]) = \sum_{n=1}^N (\mathbf{A}_n^{(m)} \cdot \mathbf{v}_n)^2 \cdot \varphi_n, \quad (4.20)$$

where $\varphi_n = \mathcal{F}(\mathbb{E}[r_n(t) \cdot r_n(t+k)])$ is the PSD of the n th dipole. Eq. (4.20) describes the relationship between the PSDs of dipole sources and the PSDs of MEG measurements.

Substituting (4.20) into the discriminant function (4.2) yields:

$$F(\mathbf{x}) = \sum_{m=1}^M w_m \cdot x_m + c = \sum_{n=1}^N \left(\sum_{m=1}^M w_m \cdot (\mathbf{A}_n^{(m)} \cdot \mathbf{v}_n)^2 \cdot \varphi_n \right) + c, \quad (4.21)$$

where w_m is the m th component of \mathbf{w} (i.e. $\mathbf{w} = [w_1 \ w_2 \ \dots \ w_M]^T$). As discussed in Section 4.2.2, the optimal filter $\mathbf{w}^T \cdot \mathbf{x}$ should amplify the signals generated by task-related dipoles and attenuate the signals coming from non-task-related dipoles. Thus we compute the gain of the filter $\mathbf{w}^T \cdot \mathbf{x}$ for each dipole and use it as a quantitative metric to assess the contribution of each dipole to the discriminant function $F(\mathbf{x})$:

$$g_n = \left| \sum_{m=1}^M w_m \cdot (\mathbf{A}_n^{(m)} \cdot \mathbf{v}_n)^2 \right| \quad (n = 1, 2, \dots, N). \quad (4.22)$$

As summarized in Section 4.2.2, there are several practical approaches that we can take to determine the orientation \mathbf{v}_n for each dipole in order to calculate the gain g_n . First, we can determine \mathbf{v}_n according to the geometrical structure of the cortical surface. Second, similar to (4.15), we can find \mathbf{v}_n associated with the maximum possible gain by solving the following optimization problem:

$$\begin{aligned} & \underset{\mathbf{v}_n}{\text{maximize}} && g_n \\ & \text{subject to} && g_n = \left| \sum_{m=1}^M w_m \cdot (\mathbf{A}_n^{(m)} \cdot \mathbf{v}_n)^2 \right| \\ & && \|\mathbf{v}_n\|_2 = 1 \end{aligned} \quad (4.23)$$

Eq. (4.23) can be re-written as:

$$\begin{aligned}
& \underset{\mathbf{v}_n}{\text{maximize}} && g_n \\
& \text{subject to} && g_n = \left| \mathbf{v}_n^T \cdot \mathbf{K}_n \cdot \mathbf{v}_n \right| \\
& && \mathbf{K}_n = \sum_{m=1}^M w_m \cdot (\mathbf{A}_n^{(m)})^T \cdot \mathbf{A}_n^{(m)} \\
& && \|\mathbf{v}_n\|_2 = 1
\end{aligned} \tag{4.24}$$

In (4.24), it is straightforward to prove that the matrix \mathbf{K}_n is symmetric. Therefore, according to Theorem 4.2.2 in [58], the maximal value of g_n in (4.24) is:

$$g_n = \max(|\rho(\mathbf{K}_n)|, |\eta(\mathbf{K}_n)|), \tag{4.25}$$

where $|\cdot|$ stands for the absolute value, $\eta(\cdot)$ and $\rho(\cdot)$ denote the minimal and the maximal eigenvalues of a matrix, respectively. Based upon the aforementioned discussions, Algorithm 4-2 summarizes the major steps of DPSL for source imaging with PSD features.

Algorithm 4-2: DPSL for Task-related Source Imaging Based on PSD Features

1. Start from the training data corresponding to two different brain states.
2. For each trial of the training data, extract the feature vector $\mathbf{x} \in R^M$ for discriminant analysis, where $x_m \in \mathbf{x}$ is the PSD over a given frequency band associated with the m th MEG channel and M represents the total number of MEG channels.
3. Build the discriminant function $F(\mathbf{x})$ in (4.2) to distinguish two different brain states using classification algorithms.
4. Compute the leadfield matrices $\{\mathbf{A}_n^{(m)}; m = 1, 2, \dots, M, n = 1, 2, \dots, N\}$ for all of the N current dipoles associated with the vertex-based model.
5. If the dipole orientation at a specific spatial location is known, compute the gain map using (4.22). Otherwise, compute the gain map using (4.25).
6. The gain values $\{g_n; n = 1, 2, \dots, N\}$ of all dipoles represent the activation map of the task-related sources.

4.3 Time/Frequency Extension of DPSL

In Section 4.2, DPSL is developed to compute the spatial distribution of task-related sources

within a short time-frequency window, where the task-related sources are assumed to be invariant within the window. For the case of a long time window or a wide frequency band, however, the task-related sources may vary over time/frequency. In this section, we further extend DPSL to obtain the spatio-temporal/frequency distribution of task-related sources.

In the case of a long time window, we partition the whole time window into a few short time windows, and then extract one feature from each time window and each channel. We use a vector $\mathbf{x} \in R^{M \cdot T}$ to represent all MEG features, where $x_{m,t} \in \mathbf{x}$ ($m = 1, 2, \dots, M$, $t = 1, 2, \dots, T$) denotes the feature associated with the m th MEG channel and the t th time window, and M and T represent the total number of MEG channels and the total number of time windows, respectively.

As described in Section 4.2, we first consider $x_{m,t}$ as the linear transformation (e.g. wavelet transform [45]) of time samples from the m th MEG channel and the t th time window. With the feature vector \mathbf{x} , we construct the discriminant function $F(\mathbf{x})$ in (4.2), where $\mathbf{w} = [w_{1,1} \ w_{2,1} \ \dots \ w_{M,1} \ \dots \ w_{1,T} \ w_{2,T} \ \dots \ w_{M,T}]^T \in R^{M \cdot T}$ and $c \in R$ are the SVM coefficients. Note that the dimension of the vector \mathbf{w} is different from the value used in Section 4.2. The discriminant function $F(\mathbf{x})$ is the linear combination of multiple MEG features $\{x_{m,t}; m = 1, 2, \dots, M, t = 1, 2, \dots, T\}$. Hence, $F(\mathbf{x})$ can be conceptually viewed as a spatio-temporal filter $\mathbf{w}^T \cdot \mathbf{x}$.

Once the spatio-temporal filter is known, we create a spatio-temporal gain map using (4.26) if the dipole orientation \mathbf{v}_n at a specific spatial location is known:

$$g_{n,t} = \left| \mathbf{w}_t^T \cdot \mathbf{A}_n \cdot \mathbf{v}_n \right| \quad (n = 1, 2, \dots, N; t = 1, 2, \dots, T), \quad (4.26)$$

or using (4.27) if the dipole orientation is unknown:

$$g_{n,t} = \left\| \mathbf{w}_t^T \cdot \mathbf{A}_n \right\|_2 \quad (n = 1, 2, \dots, N; t = 1, 2, \dots, T), \quad (4.27)$$

where $\mathbf{w}_t = [w_{1,t} \ w_{2,t} \ \dots \ w_{M,t}]^T \in R^M$ represents the coefficients associated with all features from the t th time window. The gain map in (4.26) or (4.27) reflects the spatio-temporal distribution of task-related sources. Based on the aforementioned discussions, Algorithm 4-3 summarizes the major steps of DPSL for spatio-temporal task-related source imaging.

Algorithm 4-3: DPSL for Spatio-temporal Task-related Source Imaging

1. Start from the training data corresponding to two different brain states.
2. For each trial of the training data, extract the feature vector $\mathbf{x} \in R^{M \cdot T}$ for discriminant analysis, where a feature $x_{m,t} \in \mathbf{x}$ is calculated by the linear transformation (e.g. wavelet transform [45]) of time samples from the m th MEG channel and the t th time window.
3. Build the discriminant function $F(\mathbf{x})$ in (4.2) to distinguish two different brain states using classification algorithms.
4. Compute the leadfield matrices $\{\mathbf{A}_n; n = 1, 2, \dots, N\}$ for all of the N current dipoles associated with the vertex-based model.
5. If the dipole orientation at a specific spatial location is known, compute the gain map using (4.26). Otherwise, compute the gain map using (4.27).
6. The gain values $\{g_{n,t}; n = 1, 2, \dots, N, t = 1, 2, \dots, T\}$ of all dipoles and all time windows represent the spatio-temporal activation map of the task-related sources.

Similarly, if the feature $x_{m,t}$ is extracted as the PSD from the m th MEG channel and the t th time window, we can obtain the corresponding gain map based on the method described in Section 4.2.3.

Moreover, instead of showing the spatio-temporal distribution of task-related sources, we can further extend DPSL to obtain the gain map showing the spatio-frequency distribution of task-related sources by extracting features from different frequency bands.

4.4 Simulation Studies

In this section, we present several simulation studies to demonstrate the superior accuracy of DPSL. Since the actual sources are exactly known in the simulation, the studies shown in this section facilitate us to make a full comparison between DPSL and other traditional methods.

4.4.1 Simulation Setup

To generate the required simulation data, we consider two different brain states: “State-A” and “State-B”. The MRI image of a human subject is used to define the source space. A fixed grid with 2-

mm spacing is generated from the gray/white matter boundary of the MRI image by using FreeSurfer [84]. We use one current dipole at each vertex of the grid to model the source within the brain. Several different current dipoles and noise components are considered in our simulation setup.

1. N_T current dipoles $\{\mathbf{q}_{T,n}; n = 1, 2, \dots, N_T\}$ are used to model the task-related neural activity. These dipoles are arranged as two dipole sheets shown in Figure 4-2, each of which contains 50 dipoles (i.e., $N_T = 100$). Their magnitude is randomly generated from two Gaussian distributions: $N(0, \sigma^2)$ for “State-A” and $N(1, \sigma^2)$ for “State-B”. The different mean values of these two Gaussian distributions model the discriminant information of the two brain states. The orientation of these dipoles is uniformly distributed over all possible directions.
2. N_C current dipoles $\{\mathbf{q}_{C,n}; n = 1, 2, \dots, N_C\}$ are used to model the common neural activity for “State-A” and “State-B”. These dipoles are arranged as one dipole sheet (i.e., $N_C = 50$), as shown in Figure 4-2. Their magnitude is randomly generated from a Gaussian distribution $N(1, \sigma^2)$ and the orientation is uniformly distributed over all possible directions.
3. N_B current dipoles $\{\mathbf{q}_{B,n}; n = 1, 2, \dots, N_B\}$ are deployed at all other vertices that are not covered by the aforementioned dipole sheets. These N_B current dipoles are used to model the random ongoing neural activity that is independent of the brain states. The magnitude of these dipoles is randomly generated from a Gaussian distribution $N(0, \sigma^2)$ and their orientation is uniformly distributed over all possible directions.
4. In addition to the current dipoles, a random vector \mathbf{n} is used to model the MEG measurement noise of all channels. In our simulation setup, we assume that the noise is statistically independent between different channels and it follows the Gaussian distribution $N(0, \sigma^2)$.

There are two important clarifications that should be made regarding the aforementioned dipole model. First, the magnitude of $\{\mathbf{q}_{T,n}; n = 1, 2, \dots, N_T\}$ follows the Gaussian distributions $N(0, \sigma^2)$ and $N(1, \sigma^2)$ where the mean values are normalized to 0 and 1 for “State-A” and “State-B” respectively. Such a normalized representation allows us to change the variance σ^2 to vary the SNR. In this work, since we are interested in the discriminant pattern corresponding to different brain states, the mean

difference between these two Gaussian distributions $N(0, \sigma^2)$ and $N(1, \sigma^2)$ for $\{\mathbf{q}_{T,n}; n = 1, 2, \dots, N_T\}$ is the “signal” of interest. On the other hand, the variations of the dipole moments $\{\mathbf{q}_{T,n}; n = 1, 2, \dots, N_T\}$, $\{\mathbf{q}_{C,n}; n = 1, 2, \dots, N_C\}$ and $\{\mathbf{q}_{B,n}; n = 1, 2, \dots, N_B\}$ and the noise vector \mathbf{n} are all considered as the “noise”. Note that the noise power is controlled by a single parameter σ for all dipoles $\mathbf{q}_{T,n}$, $\mathbf{q}_{C,n}$, and $\mathbf{q}_{B,n}$, as well as the measurement noise \mathbf{n} . In our simulation studies, we set the parameter σ to different values and perform source localization for all these different cases. It, in turn, allows us to test the robustness of the source localization algorithms with different SNR conditions.

Second, in our simulation setup, each dipole can fall in one of the following three categories: $\{\mathbf{q}_{T,n}; n = 1, 2, \dots, N_T\}$, $\{\mathbf{q}_{C,n}; n = 1, 2, \dots, N_C\}$ and $\{\mathbf{q}_{B,n}; n = 1, 2, \dots, N_B\}$. Remember that N_T equals 100 and N_C equals 50. Hence, there are 150 dipoles associated with $\{\mathbf{q}_{T,n}; n = 1, 2, \dots, N_T\}$ and $\{\mathbf{q}_{C,n}; n = 1, 2, \dots, N_C\}$, and all other dipoles are associated with $\{\mathbf{q}_{B,n}; n = 1, 2, \dots, N_B\}$.

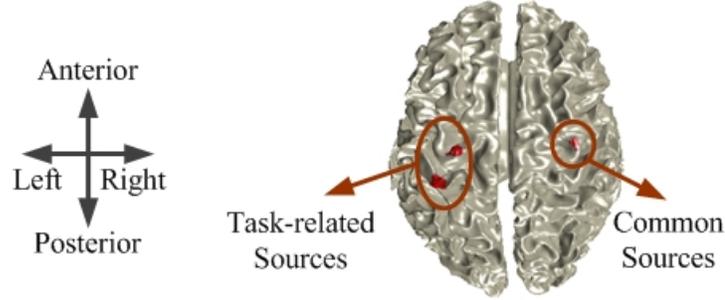


Figure 4-2. Spatial locations are shown for two task-related dipole sheets (i.e., 100 current dipoles $\{\mathbf{q}_{T,n}; n = 1, 2, \dots, 100\}$ modeling the task-related neural activity) and one common dipole sheet (i.e., 50 current dipoles $\{\mathbf{q}_{C,n}; n = 1, 2, \dots, 50\}$ modeling the common neural activity for both brain states) in our simulation studies.

Once the dipole model is set up, the leadfield of each dipole is calculated for a 306-channel whole-head MEG system (Elekta Neuromag[®]) where the anatomic structure of the brain is derived from the structural MRI data of a human subject. The MEG feature vector \mathbf{x} can be re-written as:

$$\mathbf{x} = \sum_{n=1}^{N_T} \mathbf{A}_{T,n} \cdot \mathbf{q}_{T,n} + \sum_{n=1}^{N_C} \mathbf{A}_{C,n} \cdot \mathbf{q}_{C,n} + \sum_{n=1}^{N_B} \mathbf{A}_{B,n} \cdot \mathbf{q}_{B,n} + \mathbf{n}, \quad (4.28)$$

where $\mathbf{A}_{T,n}$, $\mathbf{A}_{C,n}$, and $\mathbf{A}_{B,n}$ represent the leadfield matrices.

For testing and comparison purposes, we vary the parameter σ to implement four different SNR settings: 0 dB, -10 dB, -20 dB and -30 dB. At each SNR setting, 150 feature vectors \mathbf{x} 's are generated for each brain state to represent different trials (i.e., 150 trials for each brain state at each SNR setting). Next, we perform source localization on these simulation data, as will be discussed in detail in the following sections.

4.4.2 Evaluation Method

The accuracy of source localization using the simulated MEG data is evaluated by two methods. First, we plot the activation map generated by source localization on the cortical surface, and compare it with the actual locations of the task-related dipole sheets in Figure 4-2. Although the 2-D figures cannot precisely indicate the 3-D dipole locations, visualizing these activation maps on cortical surface provides a straightforward way to compare the results of different source localization algorithms.

Secondly, to further quantitatively measure the accuracy of a source localization algorithm, we borrow the error distance metric from [68]. It evaluates the difference between the actual and estimated dipole locations. Since our proposed DPSL results in a spatial map, we further apply a post-processing step to threshold the map and select a few locations where the activation map takes the largest values [22]. They are considered as the locations for task-related sources. Let N_E be the number of selected dipoles and these estimated dipoles are labeled as $\{\mathbf{q}_{E,n}; n = 1, 2, \dots, N_E\}$.

In the simulation setup, we have 100 current dipoles $\{\mathbf{q}_{T,n}; n = 1, 2, \dots, N_T\}$ ($N_T = 100$) modeling the task-related neural activity. For simplicity, the number of estimated dipoles is manually selected to match the actual number of dipoles that model the task-related neural activity (i.e., $N_E = N_T = 100$). In practice, it is also possible to automatically determine the appropriate value for N_E by using a statistical analysis algorithm [22].

To evaluate the accuracy of source localization, the N_T task-related dipoles $\{\mathbf{q}_{T,n}; n = 1, 2, \dots, N_T\}$ are partitioned into two groups: (i) K_D detected dipoles, and (ii) K_U (i.e., $N_T - K_D$) undetected dipoles.

For each estimated dipole $\mathbf{q}_{E,n}$, the closest dipole $\mathbf{q}_{T,n}$ from the set $\{\mathbf{q}_{T,n}; n = 1, 2, \dots, N_T\}$ is found and labeled as “detected”. After all detected dipoles are appropriately labeled, the other dipoles are labeled as “undetected”. The error distance is then calculated as [68]:

$$\frac{1}{N_E} \cdot \sum_{n=1}^{N_E} \min_i \{\|\mathbf{d}_n - \mathbf{s}_i\|_2\} + \frac{1}{K_U} \cdot \sum_{k \in J} \min_i \{\|\mathbf{s}_k - \mathbf{d}_i\|_2\}, \quad (4.29)$$

where \mathbf{s}_i denotes the coordinate of the i th actual dipole, \mathbf{d}_i stands for the coordinate of the i th estimated dipole, J represents the set of undetected dipoles, N_E is the total number of estimated dipoles, and K_U is total number of undetected dipoles. In (4.29), the first term calculates the average distance from each estimated dipole to its closest actual dipole, and the second term calculates the average distance from each undetected dipole to its closest estimated dipole. By using this definition, we can measure the average error between the estimated source locations and the actual source locations in terms of distance. More details about the formulation and meaning of the aforementioned error distance metric can be found in [68].

4.4.3 Simulation Results

To demonstrate the accuracy of the proposed DPSL algorithm, two different source localization methods are implemented for comparison purpose:

1. The traditional inferential statistics method [22]. For this approach, we apply MNE [18] to find the dominant sources for each brain state at each trial and then calculate the t-statistic [22] to identify the discriminant sources that are associated with the task-related activity. The inferential statistics method is labeled as IS in this thesis.
2. The proposed DPSL method as in Algorithm 4-1. For DPSL, we apply a linear L1-norm SVM [30] to find the discriminant function $F(\mathbf{x})$ in (4.2).

Note that IS provides a spatial map of the t-statistic and DPSL results in a spatial map of the spatial filter gain. We project both maps onto the cortical surface. Figure 4-3(a)-(b) show the t-statistic map of IS and the filter gain map of DPSL for four data sets with different SNR settings. In Figure 4-3, all maps are normalized to the interval $[0, 1]$, where large values indicate the locations with strong

task-related sources. From Figure 4-3(a), we observe that in the case of high SNR (i.e. SNR = 0 dB), IS accurately captures the locations of the task-related dipole sheets defined in Figure 4-2. However, as SNR decreases, the spatial map estimated by IS gradually deviates from the actual source distribution. When SNR is -10 dB, the spatial map of IS only shows large values around one of the two task-related dipole sheets. When SNR is -20 dB, the spatial map of IS shows large values around one task-related dipole sheet and many other locations that are far away from the task-related dipole sheets. Finally, when SNR is further reduced to -30 dB, IS cannot detect any task-related source locations. In contrast to IS, DPSL consistently identifies the task-related source locations, even if SNR is extremely low, as shown in Figure 4-3(b).

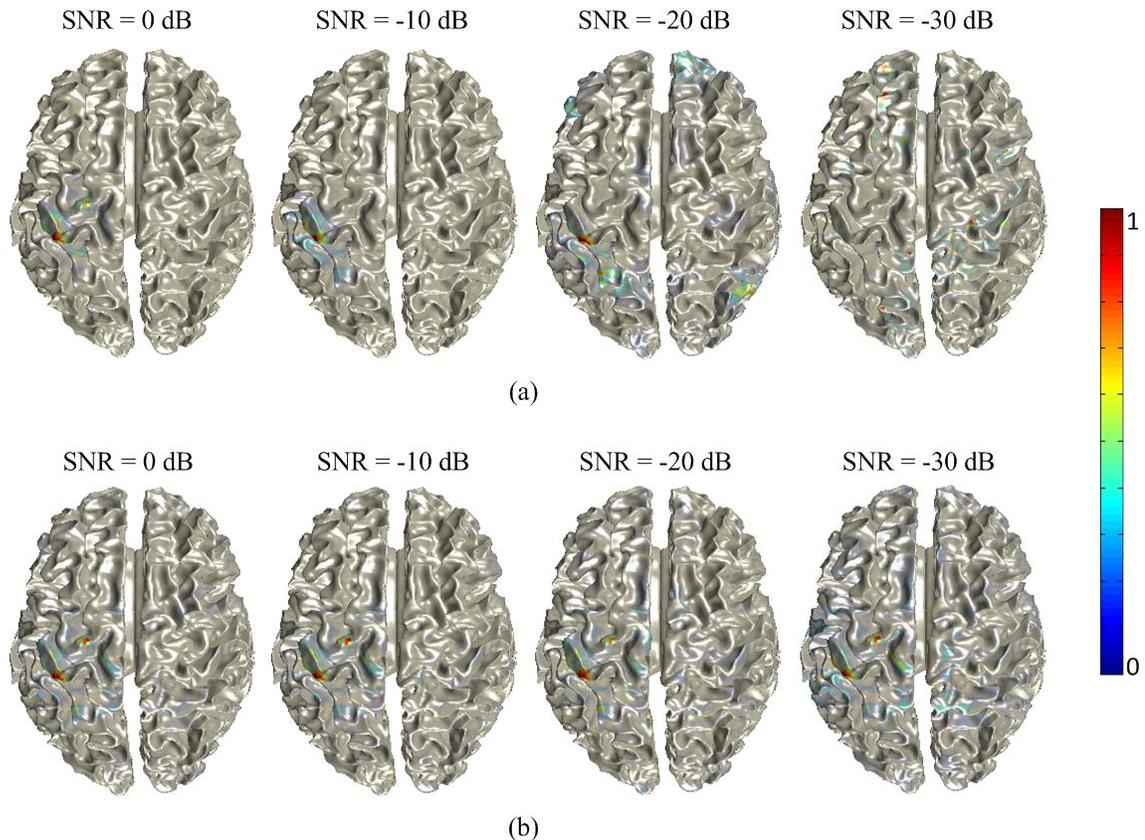


Figure 4-3. (a) Spatial map of the t-statistic is estimated by IS. (b) Spatial map of the spatial filter gain is estimated by DPSL.

To further quantitatively measure the accuracy of the source localization algorithms, we

calculate the error distance metric described in (4.29). Note that both IS and DPSL result in a spatial map. To determine the task-related source locations, we apply a post-processing step to select the top 100 dipole locations where the t-statistic (for IS) or the spatial filter gain (for DPSL) takes the largest values. Figure 4-4 shows the locations of the dipoles found by IS and DPSL.

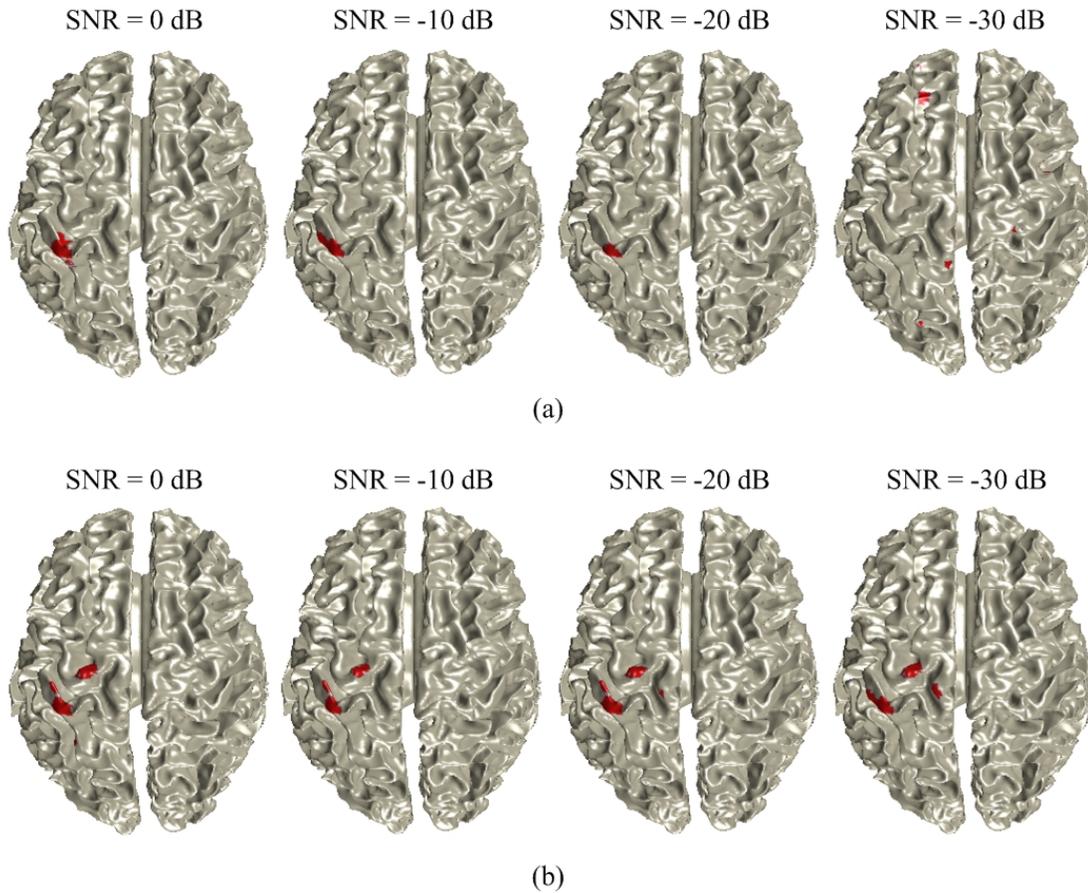


Figure 4-4. (a) Spatial locations are shown for the top 100 dipoles found by IS. (b) Spatial locations are shown for the top 100 dipoles found by DPSL.

Figure 4-5 shows the error distance for both IS and DPSL at four different SNR conditions: 0 dB, -10 dB, -20 dB and -30 dB. The detailed values are listed in Table 4-1. Studying the results in Figure 4-5, we notice that the error distance increases for both IS and DPSL as SNR decreases. However, at the same SNR, DPSL results in smaller error than IS. In particular, if SNR is high (e.g., SNR = 0 dB), both IS and DPSL yield acceptable results. However, as SNR decreases (e.g., SNR =

-30 dB), the error distance of IS substantially increases (from 1.41 cm to 10.98 cm), while the error distance of DPSL only increases slightly (from 0.45 cm to 1.05 cm). It, in turn, implies that the proposed DPSL algorithm is much less sensitive to noise than the traditional IS method. In these test cases, DPSL achieves superior accuracy, because it applies discriminant analysis to efficiently capture the discriminant pattern. As discussed in the literature [30], [31], [34], [54], most discriminant analysis algorithms (e.g., SVM) are particularly designed to reduce the impact of noise.

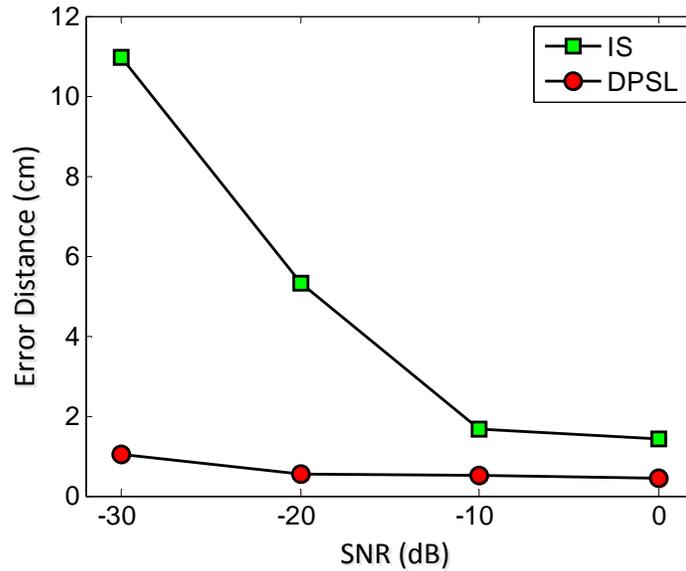


Figure 4-5. The Error Distances of IS and DPSL for simulation data sets with different SNR.

Table 4-1. Error distance of source localization for IS and DPSL

SNR (dB)	0	-10	-20	-30
Error Distance for IS (cm)	1.41	1.68	5.30	10.98
Error Distance for DPSL (cm)	0.45	0.50	0.53	1.05

4.5 Experimental Studies

In this section, we show the DPSL results for spatio-temporal task-related source imaging by

using the MEG data recorded from a human subject. All experimental procedures are approved by the Institutional Review Boards of the University of Pittsburgh and Carnegie Mellon University. All experiments are performed in accordance with the approved protocol.

4.5.1 Experimental Setup and Data Preprocessing

In our experiment, a healthy human subject performs a four-target center-out task with his right wrist holding an MEG-compatible joystick. During each trial, visual signals are presented on a screen in front of the subject. The subject is instructed to move the cursor from the center target to one of the four locations (i.e., up, down, left or right) by making wrist movements (i.e., radial deviation, ulnar deviation, flexion and extension) while keeping the rest of the body in a relaxed position. In addition, the subject is instructed to keep his gaze at the center of the screen, and only attend to the targets using his peripheral vision. A successful repetition is characterized by reaching one of the four peripheral targets within a pre-specific time window after the onset of the target and holding the cursor position there without overshooting. Only successful repetitions are used for our data analysis.

During the experiment, MEG data are acquired by using a 306-channel whole-head MEG system (Elekta Neuromag[®]) with 1 kHz sampling frequency. Electrooculography (EOG) is used to monitor eye blinks and eye movements. All trials with EOG contamination are rejected. During a separate visit, the subject takes a standard head structural MRI scan. The MRI data are co-registered with the MEG data for source localization.

In our data analysis, we consider the MEG data for two movement directions: left and right (i.e., two different brain states). There are 123 trials collected for each movement direction. The recorded MEG signals are first processed by the signal space separation (SSS) method [65] to remove the interference signals due to magnetic impurities. SSS also compensates the signal distortions caused by head movement. Next, a notch filter is applied to remove the 60 Hz power line interference. A linear approximation is then determined by least-squares fitting for each channel and each trial, and the linear trends are subtracted from the recorded MEG signals. In this study, although the MEG signals are measured by 306 channels, only the data of 204 gradiometer channels are used for the following

analysis. The other 102 magnetometer channels are not used due to their low signal-to-noise ratio.

4.5.2 Spatio-temporal Source Imaging Results

Previous neuroscience research on MEG movement decoding demonstrates that significant power modulation related to movement directions can be observed in low-frequency band (≤ 7 Hz) [31]. In addition, the important neural activity that carries movement information can be found during a short time window [31]. For these reasons, we only consider the low-frequency band for the time window $t \in [150 \text{ ms}, 330 \text{ ms}]$, where $t = 0 \text{ ms}$ represents target onset. We apply discrete wavelet transform (DWT) with second-order Symlet wavelet function [70] to decompose the MEG signals from each channel and each trial to multiple resolution levels. The aforementioned DWT results in three wavelet coefficients, which cover the low-frequency band (≤ 7 Hz) and correspond to three time windows for each channel. Here, each time window is around 60 ms in length.

Next, we apply DPSL in Algorithm 4-3 to find the spatio-temporal distribution of the task-related sources. Such an analysis allows us to study and compare the task-related source locations over different time windows. In this example, we take the wavelet coefficients as features (i.e. the feature vector contains 204×3 wavelet coefficients corresponding to 204 gradiometer channels and 3 time windows) and use linear L1-norm SVM [30] to find the discriminant function $F(\mathbf{x})$ in (4.2). Figure 4-6 shows the task-related source imaging results, i.e., the spatio-temporal map of the filter gain estimated by DPSL.

Studying Figure 4-6, we notice that the task-related sources are activated in the following order: (i) the primary visual area in Figure 4-6(a), (ii) the parietal area in Figure 4-6(b), and (iii) the contralateral sensorimotor area in Figure 4-6(c). It is important to note that strong modulation is observed in the parietal cortex and the contralateral sensorimotor area from 210 ms to 330 ms, as shown in Figure 4-6(b)-(c). These observations are consistent with the results reported by other neuroscience studies [71]-[73]. It should be further noted that although this experiment is designed to distinguish wrist movements to the left and right directions, we do not observe strong discriminant

sources in the wrist area of the sensorimotor cortex in Figure 4-6. We believe that the major activity of the wrist area is probably common across different movement directions and, hence, no strong discriminant sources are observed [74]-[75].

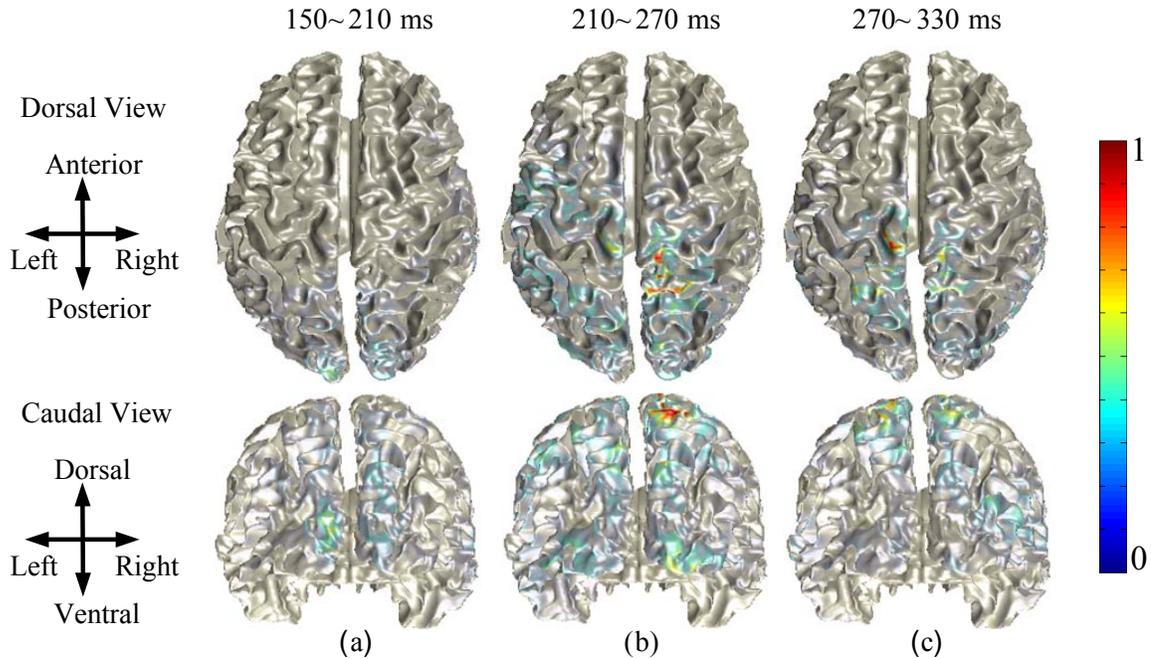


Figure 4-6. The spatial-temporal map of the filter gain is estimated by DPSL for different time windows.

4.6 Discussion

In this chapter, we propose a DPSL method to localize task-related sources. DPSL identifies the task-related sources by first applying discriminant analysis to design a filter that can optimally differentiate different brain states in MEG sensor space. Since the filter extracts the optimal discriminant pattern to distinguish different brain states, it should amplify the signals generated by task-related sources and attenuate the signals coming from non-task-related sources. Based on this idea, DPSL compute the activation map of the task-related sources by studying the response of the filter. This procedure is quite different from the traditional linear approaches for the inverse solution (e.g. MNE), which intends to find a direct linear transform from MEG measurement to the source

estimation [83]. Since the discriminant analysis algorithms are particularly designed to be robust to noise [30], [31], [34], [54], DPSL can efficiently reduce the impact of noise and the non-task-related interferences, and, hence, DPSL should be a good choice to identify the weak task-related sources from noisy MEG measurements.

However, it is worth mentioning that there remain a number of open questions related to the DPSL method. First, similar to other source localization methods, the spatial resolution of DPSL is limited by the number of MEG channels. Given the limited resolution, DPSL cannot accurately distinguish task-related and non-task-related sources that are close to each other.

Second, DPSL may not accurately detect the sources that are deep in the brain, since the filter gain can be extremely small for these deep sources. This is a well-known issue for many other source localization algorithms in the literature.

Finally, in this chapter, we demonstrate the efficiency of DPSL with both simulated and experimental MEG data. Since the actual source locations in the simulation are exactly known, simulation studies facilitate us to evaluate the performance of DPSL. However, the simulation assumes simplified head, source and noise models, which do not take into account the non-ideal nature of the sensors, realistic brain structures, the correlations in the sources and the noise, and the non-Gaussian signatures of all signals. To further demonstrate the efficiency of DPSL on practical MEG data, we should use real experimental MEG data. However, the actual source locations in the experimental data are typically unknown, which makes the evaluation based on real MEG difficult. As shown in our experimental studies, one way for evaluation of source localization results on real data is to check whether the results can be well explained by the existing neuroscience knowledge. Besides, the other functional modalities, such as fMRI and PET, may offer the potential for providing ground truth for the source locations. In addition to use real experimental MEG data, a multiple dipole phantom was used in a few studies for the evaluation of source localization algorithms [81], [82]. All of these methods can be applied in our future work to further validate the efficiency of DPSL.

4.7 Summary

To understand the task-related neural activity within the brain, it is important to localize the cortical regions that produce the task-related activity based on MEG recordings associated with different brain states. In this chapter, we propose a discriminant pattern source localization (DPSL) method to address this problem. DPSL consists of two major steps. First, discriminant analysis is applied to find a filter to optimally distinguish different brain states. Next, the gain of the filter is computed at each source location to reveal the activation map of task-related sources. DPSL is an important tool to identify the locations of task-related sources that differentiate different brain states. Even though several traditional techniques based on inferential statistics were previously developed to address the same problem, DPSL outperforms the traditional approach, as is demonstrated by our simulation studies in Section 4.4.

DPSL is closely related to MEG decoding [30], [31], [34], [54]. Many robust decoding algorithms developed in the literature can be employed by DPSL for discriminant analysis in the first step. On the other hand, DPSL can be conceptually viewed as a post-processing step for MEG decoding. It can be used to explain where the modulation signal used for decoding comes from, because the gain map obtained by DPSL directly reflects the contribution of each possible source location to the decoder.

Chapter 5

Region of Interest Constrained Decoding

5.1 Motivation

In Chapter 3, we propose an MEG decoding algorithm, which aims to extract the optimal discriminant pattern to distinguish different brain states. High decoding accuracies indicate that strong task-related activity exists in the brain. In Chapter 4, we propose an algorithm to further localize the neural sources that produce the task-related activity. After the source localization, we notice that task-related activity often appears in multiple cortical regions. Such results are observed, because human mental processes are controlled by distributed cortical networks [26].

However, a number of new MEG applications have recently emerged and suggest a need to decode different brain states by signals arising from a specific cortical region. For instance, applying MEG decoding to neurorehabilitation has attracted significant interest [87]-[90]. This emerging technique focuses neurorehabilitation on a target cortical region of interest, such as a region with dysfunction. The goal is to provide the feedback of neural activity within this region so that the patient can learn how to produce activation patterns to facilitate the plasticity in the targeted cortical region. In this case, decoding MEG signals from only the target cortical region could be beneficial. Otherwise, even if the decoding accuracy is high, the rehabilitation process may not effectively train the target cortical region and induce the desired function recovery. However, it is difficult, if not impossible, to force only the target cortical region to generate discriminant signals in the experiments. Because human actions are controlled by distributed cortical networks, the discriminant brain signals often come from multiple cortical regions [79]. In addition, some inevitable neural activity correlated

with the experimental paradigm may produce discriminant signals. For example, in a visually-guided motor study, a subject is asked to move the left or right hand by following a visual cue. The visual cue for different hands will modulate visual cortex activity, although the intention is to activate discriminant signals in motor cortex. Therefore, we need to develop a new MEG decoding tool to extract the discriminant information from pre-specified cortical regions to distinguish different brain states while suppressing signals originating outside of the pre-specified cortical regions. These pre-specified cortical regions are referred to as regions of interest (ROI) and the associated decoding problem is referred to as region-of-interest-constrained decoding (ROI decoding).

One possible approach for ROI decoding is to first extract the MEG signals from the ROI by source localization or spatial filtering, and then construct a decoder to distinguish different brain states based on these signals from the ROI [33], [77], [94]-[96]. Such a two-step approach heavily depends on the results of the first step that may not perfectly capture all signals originating from the ROI and simultaneously remove all signals originating outside of the ROI. Taking source localization as an example, it is well-known that the source localization problem is profoundly underdetermined due to the limited number of MEG sensors compared to the number of possible electrical sources [9]. Source localization algorithms choose a solution out of an infinite set of candidates based on various prior assumptions. In most cases, these algorithms can only capture the “dominant” sources, but not necessarily the discriminant sources inside the ROI. If the discriminant sources within the ROI are not accurately estimated, we cannot achieve a high accuracy in the following decoding process.

On the other hand, it has been reported in the literature that ROI decoding can be implemented by using the channels on top of the ROI only [31], [34], [35]. All MEG channels capture the signals that may originate anywhere in the brain. Therefore, limiting the channels does not necessarily limit the ROI on the cortex. For this reason, a simple channel selection approach cannot efficiently extract the MEG signals from the ROI and, hence, does not offer an optimal solution for ROI decoding.

In this chapter, we propose a ROI-constrained discriminant analysis (RDA) algorithm to solve the aforementioned ROI decoding problem. RDA formulates a constrained optimization to find the optimal spatial filter that can accurately distinguish different brain states based on the MEG signals

generated from the ROI. In our proposed optimization formulation, the cost function is derived by using linear classification techniques. It aims to maximize the discriminant information carried by the output signal of the spatial filter so that different brain states can be accurately separated. The constraint is designed by borrowing the idea of beamspace transformation [33], [77], [96]. It guarantees that the spatial filter is able to pass the signals generated by the ROI while attenuating the signals generated by other cortical regions. By simultaneously considering decoding accuracy and ROI constraint, RDA identifies the optimal discriminant pattern associated with the ROI and, hence, efficiently solves the ROI decoding problem. In addition, a numerical solver is developed in this chapter to solve the non-convex optimization problem posed by RDA with guaranteed global convergence.

The remainder of this chapter is organized as follows. We first derive the optimization formulation for the proposed RDA algorithm in Section 5.2, and then develop an efficient numerical solver to solve the RDA problem in Section 5.3. The efficiency of the RDA method is demonstrated by a number of simulation and experimental examples in Section 5.5 and Section 5.6, respectively. Finally, we discuss several practical issues related to RDA in Section 5.7, followed by a summary in Section 5.8.

5.2 ROI-Constrained Discriminant Analysis

Without loss of generality, we consider two brain states that are labeled as “State-A” and “State-B”, respectively. For example, these two states may correspond to the brain activity when a subject performs two different movement tasks. We further assume that MEG data are recorded from M channels. We use a vector $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_M]^T \in \mathbb{R}^M$ to represent the MEG features corresponding to these M channels, where x_m denotes the m th feature which is associated with the m th MEG channel. The feature vector \mathbf{x} can be a collection of MEG signals from M different channels at a particular time t . Alternatively, if a linear transformation (e.g., short-time Fourier transform [46], wavelet transform [45], etc.) is applied, \mathbf{x} can be a set of the transformed signals in frequency or wavelet domain. To

simplify our notation, we assume that a single feature is extracted from each MEG channel. However, the proposed RDA algorithm can be extended to accommodate multiple features per channel, as will be discussed in Section 5.7.

RDA seeks for a linear combination of all features as the discriminant function for decoding:

$$y = \mathbf{w}^T \cdot \mathbf{x}, \quad (5.1)$$

where $\mathbf{w} \in R^M$ defines the weight values of all MEG features. Since \mathbf{x} is the collection of features associated with the MEG channels at different spatial locations, the linear combination in (5.1) can be conceptually viewed as a spatial filter. For ROI decoding, the spatial filter must be optimally designed to meet the following two criteria:

1. ROI selectivity: The spatial filter should pass the MEG signals generated by the neural sources within ROI and attenuate the MEG signals originating from other cortical regions.
2. Decoding accuracy: The output of the spatial filter must carry sufficient discriminant information to accurately distinguish different brain states.

In what follows, we formulate a constrained optimization to find the vector \mathbf{w} in (5.1), while simultaneously satisfying the aforementioned two criteria.

Consider two sets of training data $\{\mathbf{x}_{n,1}; n = 1, 2, \dots, N_1\}$ and $\{\mathbf{x}_{n,2}; n = 1, 2, \dots, N_2\}$ where $\mathbf{x}_{n,1}$ and $\mathbf{x}_{n,2}$ are the feature vectors of the n th trial for the brain states ‘‘State-A’’ and ‘‘State-B’’ respectively. The symbols N_1 and N_2 represent the numbers of training trials for these two brain states. In order to distinguish the two brain states, we follow the idea of LDA [31], [35], [52] to maximize the ratio of the between-class scatter over the within-class scatter for the output y of the spatial filter in (5.1):

$$\underset{\mathbf{w}}{\text{maximize}} \quad \frac{\mathbf{w}^T \cdot \mathbf{S}_B \cdot \mathbf{w}}{\mathbf{w}^T \cdot \mathbf{S}_W \cdot \mathbf{w}}. \quad (5.2)$$

The between-class scatter matrix $\mathbf{S}_B \in R^{M \times M}$ and the within-class scatter matrix $\mathbf{S}_W \in R^{M \times M}$ are defined by:

$$\mathbf{S}_B = N_1 \cdot (\boldsymbol{\mu}_1 - \boldsymbol{\mu}) \cdot (\boldsymbol{\mu}_1 - \boldsymbol{\mu})^T + N_2 \cdot (\boldsymbol{\mu}_2 - \boldsymbol{\mu}) \cdot (\boldsymbol{\mu}_2 - \boldsymbol{\mu})^T \quad (5.3)$$

$$\mathbf{S}_W = \sum_{n=1}^{N_1} (\mathbf{x}_{n,1} - \boldsymbol{\mu}_1) \cdot (\mathbf{x}_{n,1} - \boldsymbol{\mu}_1)^T + \sum_{n=1}^{N_2} (\mathbf{x}_{n,2} - \boldsymbol{\mu}_2) \cdot (\mathbf{x}_{n,2} - \boldsymbol{\mu}_2)^T, \quad (5.4)$$

where $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ stand for the mean values of $\{\mathbf{x}_{n,1}; n = 1, 2, \dots, N_1\}$ and $\{\mathbf{x}_{n,2}; n = 1, 2, \dots, N_2\}$ respectively, and $\boldsymbol{\mu}$ is the mean of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$. Note that the matrix \mathbf{S}_B is positive semidefinite and the matrix \mathbf{S}_W is positive definite.

In addition to the cost function in (5.2), the spatial filter $y = \mathbf{w}^T \cdot \mathbf{x}$ must appropriately select the MEG signals generated by the sources within ROI. Towards this goal, we adopt a vertex-based model that has been widely used by MEG source localization algorithms [9], [84]. Specifically, we assume that the cortex is partitioned into a number of vertices. Each vertex contains a current dipole that models the electrical source within the vertex. Based on this model, the MEG feature vector \mathbf{x} can be represented as:

$$\mathbf{x} = \sum_{n=1}^N \mathbf{A}_n \cdot \mathbf{q}_n, \quad (5.5)$$

where $\mathbf{A}_n \in R^{M \times 3}$ is the leadfield matrix of the n th dipole, $\mathbf{q}_n \in R^3$ denotes the moment of the n th dipole, and N is the total number of dipoles. The leadfield matrix \mathbf{A}_n models the influence of the n th dipole on the measured MEG signals. It can be calculated according to the geometrical structure and the conducting medium of the human head [9].

Substituting (5.5) into (5.1), the output of the spatial filter $y = \mathbf{w}^T \cdot \mathbf{x}$ can be represented as

$$y = \mathbf{w}^T \cdot \sum_{n=1}^N \mathbf{A}_n \cdot \mathbf{q}_n. \quad (5.6)$$

We further partition y into two non-overlap components:

$$y_\Phi = \mathbf{w}^T \cdot \sum_{n \in \Phi} \mathbf{A}_n \cdot \mathbf{q}_n \quad (5.7)$$

$$y_{\bar{\Phi}} = \mathbf{w}^T \cdot \sum_{n \notin \Phi} \mathbf{A}_n \cdot \mathbf{q}_n \quad (5.8)$$

$$y = y_\Phi + y_{\bar{\Phi}}, \quad (5.9)$$

where Φ represents the set containing all current dipoles inside the ROI. In other words, y_Φ in (5.7) represents the filter output generated by the sources within the ROI and $y_{\bar{\Phi}}$ in (5.8) denotes the filter

output generated by the sources outside the ROI. The power of y_Φ and $y_{\bar{\Phi}}$ can be calculated as

$$p_\Phi = y_\Phi y_\Phi^T = \mathbf{w}^T \cdot \sum_{m \in \Phi} \sum_{n \in \Phi} \mathbf{A}_m \cdot \mathbf{q}_m \cdot \mathbf{q}_n^T \cdot \mathbf{A}_n^T \cdot \mathbf{w} \quad (5.10)$$

$$p_{\bar{\Phi}} = y_{\bar{\Phi}} y_{\bar{\Phi}}^T = \mathbf{w}^T \cdot \sum_{m \in \Phi} \sum_{n \notin \Phi} \mathbf{A}_m \cdot \mathbf{q}_m \cdot \mathbf{q}_n^T \cdot \mathbf{A}_n^T \cdot \mathbf{w}. \quad (5.11)$$

We use the ratio between p_Φ and $p_{\bar{\Phi}}$ to quantitatively measure the ROI selectivity of the spatial filter:

$$\frac{p_\Phi}{p_{\bar{\Phi}}} = \frac{\mathbf{w}^T \cdot \sum_{m \in \Phi} \sum_{n \in \Phi} \mathbf{A}_m \cdot \mathbf{q}_m \cdot \mathbf{q}_n^T \cdot \mathbf{A}_n^T \cdot \mathbf{w}}{\mathbf{w}^T \cdot \sum_{m \in \Phi} \sum_{n \notin \Phi} \mathbf{A}_m \cdot \mathbf{q}_m \cdot \mathbf{q}_n^T \cdot \mathbf{A}_n^T \cdot \mathbf{w}}. \quad (5.12)$$

In practice, the dipole moments $\{\mathbf{q}_n; n = 1, 2, \dots, N\}$ are not known. Hence, the ratio in (5.12) must be estimated with a number of assumptions. For instance, if there is no prior information about the dipole moments, the traditional beamspace transformation method calculates the ratio $p_\Phi/p_{\bar{\Phi}}$ by assuming that all dipoles are uniformly distributed and mutually independent [77]. With these assumptions, the ratio in (5.12) can be simplified as:

$$\frac{p_\Phi}{p_{\bar{\Phi}}} = \frac{\mathbf{w}^T \cdot \sum_{n \in \Phi} \mathbf{A}_n \cdot \mathbf{A}_n^T \cdot \mathbf{w}}{\mathbf{w}^T \cdot \sum_{n \notin \Phi} \mathbf{A}_n \cdot \mathbf{A}_n^T \cdot \mathbf{w}} = \frac{\mathbf{w}^T \cdot \mathbf{G}_\Phi \cdot \mathbf{w}}{\mathbf{w}^T \cdot \mathbf{G}_{\bar{\Phi}} \cdot \mathbf{w}}, \quad (5.13)$$

where

$$\mathbf{G}_\Phi = \sum_{n \in \Phi} \mathbf{A}_n \cdot \mathbf{A}_n^T \quad (5.14)$$

$$\mathbf{G}_{\bar{\Phi}} = \sum_{n \notin \Phi} \mathbf{A}_n \cdot \mathbf{A}_n^T. \quad (5.15)$$

The matrices \mathbf{G}_Φ and $\mathbf{G}_{\bar{\Phi}}$ in (5.14)-(5.15) are both positive definite [77]. Once the leadfield matrices $\{\mathbf{A}_n; n = 1, 2, \dots, N\}$ are known, $p_\Phi/p_{\bar{\Phi}}$ in (5.13) is represented as a function of the vector \mathbf{w} .

Combining (5.2) and (5.13) yields the following constrained optimization problem:

$$\begin{aligned} & \underset{\mathbf{w}}{\text{maximize}} && \frac{\mathbf{w}^T \cdot \mathbf{S}_B \cdot \mathbf{w}}{\mathbf{w}^T \cdot \mathbf{S}_W \cdot \mathbf{w}} \\ & \text{subject to} && \frac{\mathbf{w}^T \cdot \mathbf{G}_\Phi \cdot \mathbf{w}}{\mathbf{w}^T \cdot \mathbf{G}_{\bar{\Phi}} \cdot \mathbf{w}} \geq \lambda \end{aligned}, \quad (5.16)$$

where λ is a user-defined parameter. The optimization in (5.16) aims to find the optimal spatial filter $y = \mathbf{w}^T \cdot \mathbf{x}$ to distinguish two brain states, while simultaneously constraining the power ratio $p_\Phi/p_{\bar{\Phi}}$ to be

no less than λ .

Studying the optimization formulation in (5.16) reveals several important properties. First, since the left side of the constraint function in (5.16) is in the form of a generalized Rayleigh quotient, its lower bound and upper bound are equal to

$$\inf_{\mathbf{w}} \frac{\mathbf{w}^T \cdot \mathbf{G}_\Phi \cdot \mathbf{w}}{\mathbf{w}^T \cdot \mathbf{G}_{\bar{\Phi}} \cdot \mathbf{w}} = \eta(\mathbf{G}_{\bar{\Phi}}^{-1} \cdot \mathbf{G}_\Phi), \quad (5.17)$$

and

$$\sup_{\mathbf{w}} \frac{\mathbf{w}^T \cdot \mathbf{G}_\Phi \cdot \mathbf{w}}{\mathbf{w}^T \cdot \mathbf{G}_{\bar{\Phi}} \cdot \mathbf{w}} = \rho(\mathbf{G}_{\bar{\Phi}}^{-1} \cdot \mathbf{G}_\Phi), \quad (5.18)$$

respectively [76]. In (5.17)-(5.18), $\inf(\bullet)$ and $\sup(\bullet)$ denote the infimum (i.e., the greatest lower bound) and the supremum (i.e., the least upper bound) of a function, and $\eta(\bullet)$ and $\rho(\bullet)$ stand for the minimum and maximum eigenvalues of a matrix.

Second, adjusting the value of λ allows us to explore the trade-off between the discriminant power and the ROI selectivity of the spatial filter. In one extreme case, if λ is set to the minimum possible value of the constraint (i.e. $\eta(\mathbf{G}_{\bar{\Phi}}^{-1} \cdot \mathbf{G}_\Phi)$), the constraint in (5.16) is satisfied for any given \mathbf{w} and, hence, the ROI decoding in (5.16) is equivalent to the traditional LDA in (5.2) where the discriminant information is maximized and the ROI constrain is completely ignored. In the other extreme case, if λ is set to the maximum possible value of the constraint (i.e. $\rho(\mathbf{G}_{\bar{\Phi}}^{-1} \cdot \mathbf{G}_\Phi)$), there is only one possible direction of \mathbf{w} to satisfy the constraint in (5.16) and the optimal solution (denoted as \mathbf{w}^*) is independent of the cost function. In this case, the filter is designed to maximize the total signal power originating from the ROI, instead of the discriminant information generated by the ROI. In other words, it maximizes the ROI selectivity only and, hence, is unlikely to achieve a high decoding accuracy. When λ varies between these two extreme cases, the optimization in (5.16) explores the trade-off between its cost function (related to decoding accuracy) and the constraint function (related to ROI selectivity).

Third, neither the cost function nor the constraint set in (5.16) is convex. In this case, a general-purpose nonconvex optimizer is likely to settle at a local optimum and, hence, cannot efficiently find

the global optimum. To the best of our knowledge, there is no existing optimization algorithm that guarantees to find the global optimum of (5.16). Motivated by this observation, we propose an efficient numerical algorithm to solve (5.16) with guaranteed global convergence. The details of our proposed numerical solver are described in the next section.

5.3 Numerical Solver

As discussed in Section 5.2, the constraint function of (5.16) is upper bounded by $\rho(\mathbf{G}_{\bar{\phi}}^{-1} \cdot \mathbf{G}_{\phi})$ defined in (5.18). Therefore, the optimization in (5.16) is feasible, if and only if λ is not greater than $\rho(\mathbf{G}_{\bar{\phi}}^{-1} \cdot \mathbf{G}_{\phi})$. If λ equals $\rho(\mathbf{G}_{\bar{\phi}}^{-1} \cdot \mathbf{G}_{\phi})$, the globally optimal solution \mathbf{w}^* should be the dominant eigenvector of $\mathbf{G}_{\bar{\phi}}^{-1} \cdot \mathbf{G}_{\phi}$ corresponding to the eigenvalue λ . Otherwise, if λ is less than $\rho(\mathbf{G}_{\bar{\phi}}^{-1} \cdot \mathbf{G}_{\phi})$, the optimization in (5.16) is strictly feasible. Next, we will develop a numerical algorithm to find the globally optimal solution \mathbf{w}^* for this strictly feasible case.

Given the fact that $\mathbf{G}_{\bar{\phi}}$ is positive-definite, we re-write (5.16) as:

$$\begin{aligned} & \underset{\mathbf{w}}{\text{minimize}} && -\frac{\mathbf{w}^T \cdot \mathbf{S}_B \cdot \mathbf{w}}{\mathbf{w}^T \cdot \mathbf{S}_W \cdot \mathbf{w}} \\ & \text{subject to} && \mathbf{w}^T \cdot (\lambda \cdot \mathbf{G}_{\bar{\phi}} - \mathbf{G}_{\phi}) \cdot \mathbf{w} \leq 0 \end{aligned} \quad (5.19)$$

Since the cost function in (5.19) is invariant with respect to the rescaling of \mathbf{w} , we can fix its denominator to an arbitrary constant and optimize the numerator only. This allows us to transform the optimization in (5.19) to an equivalent problem:

$$\begin{aligned} & \underset{\mathbf{w}}{\text{minimize}} && -\mathbf{w}^T \cdot \mathbf{S}_B \cdot \mathbf{w} \\ & \text{subject to} && \mathbf{w}^T \cdot \mathbf{S}_W \cdot \mathbf{w} = 1 \\ & && \mathbf{w}^T \cdot (\lambda \cdot \mathbf{G}_{\bar{\phi}} - \mathbf{G}_{\phi}) \cdot \mathbf{w} \leq 0 \end{aligned} \quad (5.20)$$

As shown by the Proposition 2 of [97], strong duality holds for (5.20) and its dual problem is expressed as:

$$\begin{aligned} & \underset{\alpha, \beta}{\text{maximize}} && \alpha \\ & \text{subject to} && -\mathbf{S}_B - \alpha \cdot \mathbf{S}_W + \beta \cdot (\lambda \cdot \mathbf{G}_{\bar{\phi}} - \mathbf{G}_{\phi}) \geq 0, \\ & && \beta \geq 0 \end{aligned} \quad (5.21)$$

where the constraint $-\mathbf{S}_B - \alpha \cdot \mathbf{S}_W + \beta (\lambda \cdot \mathbf{G}_\Phi - \mathbf{G}_\Phi) \geq 0$ means that the matrix $-\mathbf{S}_B - \alpha \cdot \mathbf{S}_W + \beta (\lambda \cdot \mathbf{G}_\Phi - \mathbf{G}_\Phi)$ must be positive semidefinite.

Due to strong duality, the optimal cost function value of (5.20) is equal to that of (5.21). In addition, since the two optimization problems (5.19) and (5.20) yield the same optimal cost function values, we can conclude that the optimal cost function values of (5.19), (5.20) and (5.21) are all identical. We denote this optimal value as f^* . On the other hand, since the optimization in (5.21) is a convex semi-definite programming (SDP) problem [79], it can be solved both robustly (i.e., with guaranteed global optimum) and efficiently (i.e., with low computational cost) [79]. Once the optimization in (5.21) is solved, the optimal cost function value f^* is known and we need to further find the optimal solution \mathbf{w}^* to fully determine the spatial filter in (5.1).

Given the optimal cost function value f^* of (5.19), the following inequality holds for any vector \mathbf{w} within the feasible set of (5.19):

$$\frac{-\mathbf{w}^T \cdot \mathbf{S}_B \cdot \mathbf{w}}{\mathbf{w}^T \cdot \mathbf{S}_W \cdot \mathbf{w}} \geq f^* . \quad (5.22)$$

Remember that the matrix \mathbf{S}_W is positive definite and, hence, the scalar $\mathbf{w}^T \cdot \mathbf{S}_W \cdot \mathbf{w}$ is positive. Eq. (5.22) is equivalent to:

$$-\mathbf{w}^T \cdot \mathbf{S}_B \cdot \mathbf{w} - f^* \cdot \mathbf{w}^T \cdot \mathbf{S}_W \cdot \mathbf{w} \geq 0 . \quad (5.23)$$

The left-hand side of (5.23) reaches the minimum value 0 or, equivalently, the equality sign is reached for the inequality in (5.22), when \mathbf{w} is equal to the optimal solution \mathbf{w}^* . Therefore, the solution \mathbf{w}^* of (5.19) can be determined by solving the following optimization problem:

$$\begin{aligned} & \underset{\mathbf{w}}{\text{minimize}} && -\mathbf{w}^T \cdot (\mathbf{S}_B + f^* \cdot \mathbf{S}_W) \cdot \mathbf{w} \\ & \text{subject to} && \mathbf{w}^T \cdot (\lambda \cdot \mathbf{G}_\Phi - \mathbf{G}_\Phi) \cdot \mathbf{w} \leq 0 \end{aligned} \quad (5.24)$$

The optimal cost function value of (5.24) is zero and, hence, is invariant with respect to the rescaling of the vector \mathbf{w} . Based on this observation, we define:

$$\mathbf{w} = [w_1 \quad w_2 \quad \cdots \quad w_{M-1} \quad 1]^T = [\mathbf{v}^T \quad 1]^T , \quad (5.25)$$

where $\mathbf{v} = [w_1 \quad w_2 \quad \dots \quad w_{M-1}]^T$. In other words, since the optimization in (5.24) aims to find the direction

of the vector \mathbf{w} , we can fix the last element of \mathbf{w} to be any constant (e.g., 1). Next, we represent the M -by- M symmetric matrix $-(\mathbf{S}_B + f^* \cdot \mathbf{S}_W)$ in the form of four sub-matrices:

$$-(\mathbf{S}_B + f^* \cdot \mathbf{S}_W) = \begin{bmatrix} \mathbf{A}_0 & \mathbf{b}_0 \\ \mathbf{b}_0^T & c_0 \end{bmatrix}, \quad (5.26)$$

where $\mathbf{A}_0 \in R^{(M-1) \times (M-1)}$, $\mathbf{b}_0 \in R^{M-1}$, and $c_0 \in R$. Similarly, we represent the matrix $\lambda \cdot \mathbf{G}_{\bar{\phi}} - \mathbf{G}_{\phi}$ as:

$$\lambda \cdot \mathbf{G}_{\bar{\phi}} - \mathbf{G}_{\phi} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{b}_1 \\ \mathbf{b}_1^T & c_1 \end{bmatrix}, \quad (5.27)$$

where $\mathbf{A}_1 \in R^{(M-1) \times (M-1)}$, $\mathbf{b}_1 \in R^{M-1}$, and $c_1 \in R$.

Based on (5.25)-(5.27), we re-write the optimization problem in (5.24) as:

$$\begin{aligned} & \underset{\mathbf{v}}{\text{minimize}} && \mathbf{v}^T \cdot \mathbf{A}_0 \cdot \mathbf{v} + 2 \cdot \mathbf{b}_0^T \cdot \mathbf{v} + c_0 \\ & \text{subject to} && \mathbf{v}^T \cdot \mathbf{A}_1 \cdot \mathbf{v} + 2 \cdot \mathbf{b}_1^T \cdot \mathbf{v} + c_1 \leq 0 \end{aligned} \quad (5.28)$$

To solve (5.28), we write the corresponding dual problem [79]:

$$\begin{aligned} & \underset{r, \zeta}{\text{maximize}} && r \\ & \text{subject to} && \zeta \geq 0 \\ & && \begin{bmatrix} \mathbf{A}_0 + \zeta \cdot \mathbf{A}_1 & \mathbf{b}_0 + \zeta \cdot \mathbf{b}_1 \\ (\mathbf{b}_0 + \zeta \cdot \mathbf{b}_1)^T & c_0 + \zeta \cdot c_1 - r \end{bmatrix} \geq 0 \end{aligned}, \quad (5.29)$$

where $r \in R$ and $\zeta \in R$ are the optimization variables. The optimization in (5.29) is a convex SDP problem [79]. Solving it yields the optimal solution $\zeta = \zeta^*$. Once ζ^* is known, the optimal solution \mathbf{v}^* of (5.28) is given by [79]:

$$\mathbf{v}^* = -(\mathbf{A}_0 + \zeta^* \cdot \mathbf{A}_1)^{-1} \cdot (\mathbf{b}_0 + \zeta^* \cdot \mathbf{b}_1). \quad (5.30)$$

Substituting $\mathbf{v} = \mathbf{v}^*$ into (5.25) yields the optimal solution \mathbf{w}^* of (5.19), which is also the optimal solution of (5.16).

5.4 Algorithm Summary

Algorithm 5-1 summarizes the major steps of our proposed algorithm in Section 5.3 that finds the optimal solution \mathbf{w}^* of (5.16) with guaranteed global convergence.

Algorithm 5-1: RDA Solver

1. Start from the optimization problem in (5.16) where \mathbf{S}_B , \mathbf{S}_W , \mathbf{G}_Φ , \mathbf{G}_T and λ are known.
2. Calculate $\rho(\mathbf{G}_\Phi^{-1} \cdot \mathbf{G}_\Phi)$ in (5.18).
3. If $\lambda > \rho(\mathbf{G}_\Phi^{-1} \cdot \mathbf{G}_\Phi)$, the optimization in (5.16) is infeasible. If $\lambda = \rho(\mathbf{G}_\Phi^{-1} \cdot \mathbf{G}_\Phi)$, the optimal solution \mathbf{w}^* equals the dominant eigenvector of $\mathbf{G}_\Phi^{-1} \cdot \mathbf{G}_\Phi$. Otherwise, if $\lambda < \rho(\mathbf{G}_\Phi^{-1} \cdot \mathbf{G}_\Phi)$, follow Step 4-8 to find the optimal solution \mathbf{w}^* .
4. Solve the SDP problem in (5.21) and get the optimal cost function value f^* .
5. Compute \mathbf{A}_0 , \mathbf{b}_0 and c_0 using (5.26), and \mathbf{A}_1 , \mathbf{b}_1 and c_1 using (5.27).
6. Solve the SDP problem in (5.29) and get the optimal solution ζ^* .
7. Calculate \mathbf{v}^* using (5.30).
8. Substitute $\mathbf{v} = \mathbf{v}^*$ into (5.25) to calculate the optimal solution \mathbf{w}^* .

Algorithm 5-2 lists the major steps of the proposed RDA algorithm. As shown in the previous sections, even though the optimization problem in (5.16) is not convex, Algorithm 5-2 solves a number of convex semidefinite programming problems and guarantees to find the globally optimal solution. This is the key advantage of our proposed numerical solver over a general-purpose nonconvex optimizer, as will be further demonstrated by the experimental results in Section 5.5.

Algorithm 5-2: Region-of-interest-constrained Discriminant Analysis (RDA)

1. Start from the training data $\{\mathbf{x}_{n,1}; n = 1, 2, \dots, N_1\}$ and $\{\mathbf{x}_{n,2}; n = 1, 2, \dots, N_2\}$ corresponding to two different brain states, the user-specified ROI, and a pre-specified threshold λ .
2. Partition the structural brain image into N vertices and define the set Φ containing all current dipoles inside the ROI.
3. Calculate the leadfield matrices $\{\mathbf{A}_n; n = 1, 2, \dots, N\}$ for the current dipoles associated with the vertex-based model.
4. Calculate the matrices \mathbf{S}_B and \mathbf{S}_W based on (5.3)-(5.4).
5. Calculate the matrices \mathbf{G}_Φ and \mathbf{G}_Φ^{-1} based on (5.14)-(5.15).

6. Formulate the constrained nonlinear optimization problem in (5.16).
7. Apply Algorithm 5-1 to find the optimal solution \mathbf{w}^* of (5.16) to form the spatial filter in (5.1).

5.5 Simulation Studies

In this simulation, we will demonstrate how RDA controls the ROI selectivity by varying the value of λ . Also we will make a full comparison between RDA and other traditional methods for ROI decoding.

5.5.1 Simulation Setup

In the simulation, the MRI image of a human subject is used to define the source space. A fixed grid with 2-mm spacing is generated from the gray/white matter boundary of the MRI image by using FreeSurfer [84]. We use one current dipole at each vertex of the grid to model the source within the brain[85]. In addition, we automatically parcellate and label the cortical surface as different cortical areas by using FreeSurfer [85].

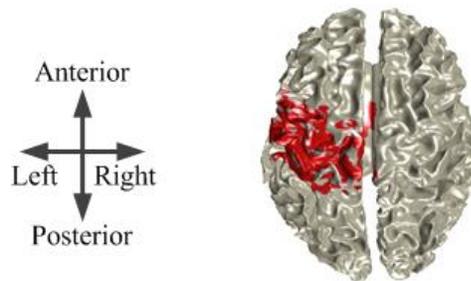


Figure 5-1. Red color indicates the spatial location of the ROI (i.e. the left precentral and the left postcentral gyri).

Two different brain states: “State-A” and “State-B” are considered in this study. We assume that the discriminant signal between “State-A” and “State-B” comes from both inside and outside the ROI. Specifically, we set the left sensorimotor area (i.e. the left precentral and the left postcentral gyri) as the ROI, which is shown in Figure 5-1. Several different current dipoles and noise components are considered in our simulation setup:

1. 20 current dipoles are used to model the discriminant sources between “State-A” and “State-B”. A half of these dipoles $\{\mathbf{q}_{ROI,n}; n = 1,2,\dots,10\}$ locate inside the ROI, and the other half $\{\mathbf{q}_{RON,n}; n = 1,2,\dots,10\}$ locate outside the ROI. Their magnitude is randomly generated from two Gaussian distributions: $N(0,1)$ for “State-A” and $N(6,1)$ for “State-B”. The different mean values of these two Gaussian distributions model the discriminant information of the two brain states. The orientation of these dipoles is uniformly distributed over all possible directions.
2. 200 current dipoles $\{\mathbf{q}_{Noi,n}; n = 1,2,\dots,200\}$ are used to model the noisy sources that is independent of the brain states. These dipole locations are randomly distributed all over the cortical surface. The magnitude of these dipoles is randomly generated from a Gaussian distribution $N(0,2^2)$ and their orientation is uniformly distributed over all possible directions.
3. In addition to the current dipoles, a vector \mathbf{n} is used to model the MEG measurement noise of all channels. In our simulation setup, we assume that the noise is statistically independent between different channels and it follows a Gaussian distribution.

Once the dipole model is set up, the leadfield of each dipole is calculated for a whole-head MEG system (Elekta Neuromag[®]). The MEG feature vector \mathbf{x} is computed by:

$$\mathbf{x} = \sum_{n=1}^{10} \mathbf{A}_{ROI,n} \cdot \mathbf{q}_{ROI,n} + \sum_{n=1}^{10} \mathbf{A}_{RON,n} \cdot \mathbf{q}_{RON,n} + \sum_{n=1}^{200} \mathbf{A}_{Noi,n} \cdot \mathbf{q}_{Noi,n} + \mathbf{n}, \quad (5.31)$$

where $\mathbf{A}_{ROI,n}$, $\mathbf{A}_{RON,n}$, and $\mathbf{A}_{Noi,n}$ represent the corresponding leadfield matrices. According to (5.31), we generate 500 feature vectors \mathbf{x} 's for each brain state to represent the trials of the training set.

In addition, to test the accuracy of a decoder derived from the training set, we generate 5000 feature vectors \mathbf{x} 's for each brain state based on (5.31) to represent the trials of the test set. This test set is referred to as TestSet-Both, as the discriminant signal is generated by the dipoles both inside and outside the ROI.

Higher decoding accuracy on TestSet-Both indicates that the associated decoding algorithm is better at extracting the discriminant information in MEG sensor space. This discriminant information may originate both inside and outside the ROI. However, ROI decoding aims to extract the optimal discriminant pattern originating inside the ROI only to distinguish different brain states. In other

words, the performance of ROI decoding should be evaluated by two aspects: (i) ROI selectivity (i.e. discriminant information originating outside the ROI should be suppressed) and (ii) discriminant signal selectivity (i.e more discriminant information originating inside the ROI should be selected). Based on these criteria, higher decoding accuracy on TestSet-Both cannot demonstrate better performance of ROI decoding. Thus, in order to evaluate the performance of a ROI decoding algorithm, we consider the other two test sets:

1. A test set with discriminant signals generated by the dipoles outside the ROI only (TestSet-RON): The MEG feature vector is computed by:

$$\mathbf{x} = \sum_{n=1}^{10} \mathbf{A}_{RON,n} \cdot \mathbf{q}_{RON,n} + \sum_{n=1}^{200} \mathbf{A}_{Noi,n} \cdot \mathbf{q}_{Noi,n} + \mathbf{n} . \quad (5.32)$$

2. A test set with discriminant signals generated by the dipoles inside the ROI only (TestSet-ROI): The MEG feature vector is computed by:

$$\mathbf{x} = \sum_{n=1}^{10} \mathbf{A}_{ROI,n} \cdot \mathbf{q}_{ROI,n} + \sum_{n=1}^{200} \mathbf{A}_{Noi,n} \cdot \mathbf{q}_{Noi,n} + \mathbf{n} . \quad (5.33)$$

For each of these test sets, we simulate 5000 trials for each of the two brain states.

Ideally, if a decoding algorithm does not select the discriminant information originating outside the ROI, its decoding accuracy on TestSet-RON should be at the chance level. That's because all discriminant information of this test set originates outside the ROI. Therefore, the ROI selectivity of an algorithm can be evaluated by the decoding accuracy on TestSet-RON. Lower accuracy on TestSet-RON indicates better ROI selectivity.

On the other hand, if a decoding algorithm is efficient at extracting the discriminant information originating inside the ROI, its decoding accuracy on TestSet-ROI should be high. That's because all discriminant information of the test set originates inside the ROI, and, hence selecting more discriminant information from ROI leads to higher decoding accuracy. Therefore, the discriminant signal selectivity of an algorithm can be evaluated by the decoding accuracy on TestSet-ROI.

In summary, a good ROI decoding algorithm should yield high decoding accuracy on TestSet-ROI, and low decoding accuracy on TestSet-RON. However, it should be noted that although TestSet-

ROI and TestSet-ROI facilitate our evaluation of a ROI decoding algorithm in the simulation, they cannot be collected in the practical MEG experiments. In practical MEG experiments, the test data and the training data follow the same distribution, and consequently, only TestSet-Both is observable. The decoding accuracy on TestSet-Both reflects the actual decoding accuracy when the decoder derived from the training data is applied.

5.5.2 Simulation Results

In Figure 5-2, the decoding accuracy of the proposed RDA on TestSet-ROI and TestSet-ROI is plotted as a function of λ (i.e., the constraint of the power ratio $p_{\Phi}/p_{\bar{\Phi}}$ defined in (5.13)). As discussed in Section 5.2, the parameter λ varies between the lower bound and the upper bound of the constraint function in (5.16). From this figure, we have several observations.

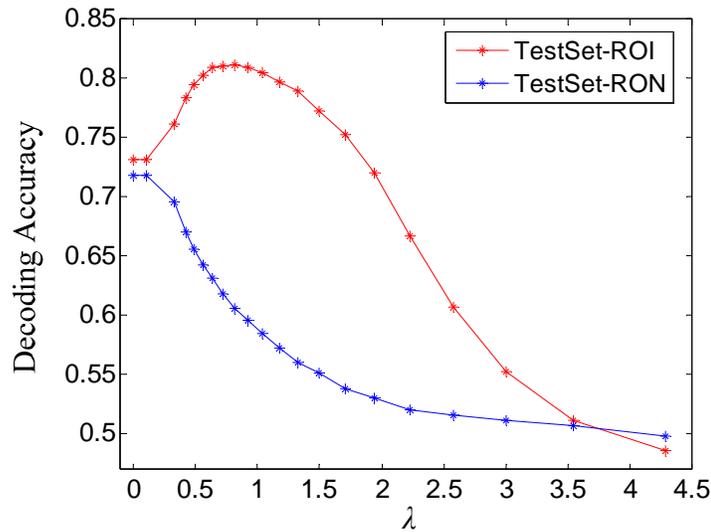


Figure 5-2. The decoding accuracies of RDA are represented as a function of λ (i.e., the constraint of the power ratio $p_{\Phi}/p_{\bar{\Phi}}$ defined in (5.13)).

First, when λ reaches the lower bound, the decoding accuracies of RDA on TestSet-ROI and TestSet-ROI are 73.7% and 72.1%, respectively. This result indicates that when the ROI constraint is completely ignored, the discriminant pattern extracted by RDA contains discriminant information originating both inside and outside the ROI. The decoding accuracies of TestSet-ROI and

TestSetRON are nearly the same, which indicates that the discriminant information used for decoding almost equally originates inside and outside the ROI in this simulation.

Second, as λ increases, the decoding accuracy of TestSet-RON decreases. This result indicates that as λ increases, less and less discriminant information originating outside the ROI is used for decoding. Based on this observation, RDA can smoothly control the ROI selectivity of the decoder by varying the value of λ .

Third, although as λ increases RDA gradually uses only the discriminant information from the ROI for decoding, the extracted discriminant information from the ROI does not keep increasing. It can be observed by the decoding accuracy of TestSet-ROI in Figure 5-2, which first increases and then decreases. This is because some signals originating outside the ROI cannot be removed without removing the discriminant signals from the ROI. Although this problem is inevitable, it can be alleviated by appropriately setting the ROI and the regions of non-interests, which will be further discussed in Section 5.7.

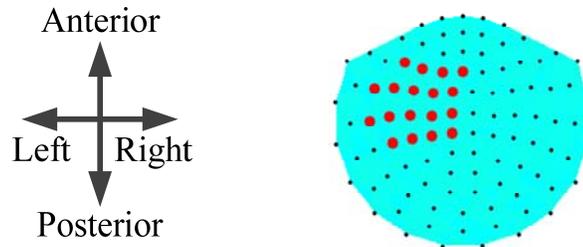


Figure 5-3. Shown in red are the 36 channels on top of ROI where each dot represents a pair of gradiometers.

Next, we implement four traditional decoding algorithms and compare RDA with these methods.

1. Linear discriminant analysis based on whole-head MEG (LDA-W): The MEG data recorded from all gradiometers are used for decoding.
2. Linear discriminant analysis based on selected channels (LDA-S): The MEG data recorded from 36 gradiometers located on top of the ROI are used for decoding. The locations of the selected channels are shown in Figure 5-3.

3. Beamspace discriminant analysis (BDA): A linear subspace is first determined by the beamspace transformation [33], [77], [96] to extract the signals originating from the given ROI. Next, LDA is applied to this linear subspace for decoding.
4. Source-space discriminant analysis (SDA): MNE [18] is applied to estimate the sources for each trial. Thereafter, only the sources within the ROI are used for decoding. Since thousands of sources within the ROI are found by MNE, Fisher criterion is applied to select the important sources as features [53], where the number of selected features is determined by leave-one-out cross validation [52] based on the training set. With these selected features, LDA is applied for classification.

Table 5-1. Decoding accuracies of the algorithms for comparison

	TestSet-Both	TestSet-ROI	TestSet-RON
LDA-W	94.8%	73.7%	72.1%
LDA-S	92.6%	77.6%	68.7%
BDA	92.6%	75.8%	67.2%
RDA	92.6%	81.6%	61.5%
SDA	50.1%	49.4%	50.2%

The decoding accuracies of LDA-W, LDA-S and SDA are shown in Table 5-1. Similar to RDA, BDA uses the power ratio p_{ϕ}/p_{ϕ} defined in (5.13) to control ROI selectivity. In Figure 5-4, the decoding accuracy of BDA is plotted as a function of this power ratio. Since RDA and BDA control the ROI selectivity of a decoder by varying a parameter, to compare their performance with LDA-S, we can consider the setting where the decoding accuracies on TestSet-Both of all these three algorithms are the same (i.e. all are 92.6%). The decoding accuracies of RDA and BDA for this setting are listed in Table 5-1. Here, we choose the same accuracy on TestSet-Both instead of TestSet-ROI and TestSet-RON, because the decoding accuracy on TestSet-Both reflects the actual decoding accuracy when the decoder derived from the training data is applied. For the same decoding accuracy

on TestSet-Both, an efficient ROI decoding algorithm should achieve low accuracy on TestSet-RON and high accuracy on TestSet-ROI.

From Table 5-1, LDA-W achieves similar high decoding accuracy on TestSet-ROI and TestSet-RON, which indicates that the discriminant information used for decoding nearly equally originates inside and outside the ROI. Such results are observed, because without considering any ROI constraint LDA-W aims to extract the optimal discriminant pattern existing in the whole feature space, and hence, the discriminant information used for decoding may originate inside and outside the ROI. Essentially, LDA-W is equivalent to an extreme case of RDA, where λ reaches the lower bound.

By using only MEG channels located on top of the ROI, LDA-S helps to select discriminant signals originating inside the ROI. As shown in Table 5-1, in comparison with LDA-W the decoding accuracy of LDA-S increases for TestSet-ROI, and decreases for TestSet-RON. However, comparing RDA and LDA-S, we find that when the decoding accuracies on TestSet-Both are the same, the decoding accuracy of RDA on TestSet-ROI is higher and on TestSet-RON is lower. This result indicates that in comparison with LDA-S, RDA is more efficient at extracting the discriminant information from the ROI while suppressing the discriminant information originating outside the ROI.

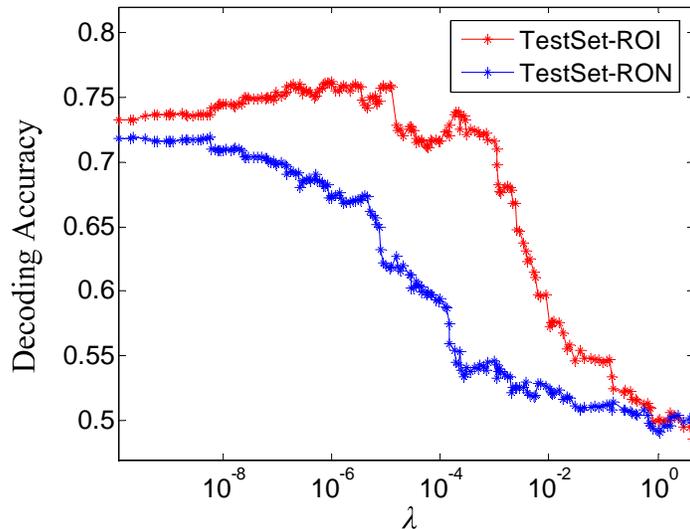


Figure 5-4. The decoding accuracies of BDA are represented as a function of λ (i.e., the lower bound of the power ratio p_ψ/p_ϕ defined in (5.13)).

Next, we compare RDA with BDA. BDA uses beam-space transformation [33], [77], [96] to extract signals generated by sources within ROI. Similar to RDA, beam-space transformation uses the power ratio p_Φ/p_Φ defined in (5.13) to control ROI selectivity. However, unlike RDA that formulates a constrained optimization problem in (5.16), beam-space transformation cannot directly set the ratio p_Φ/p_Φ to a specific value. Let $\varphi_i, i=1,2,\dots,M$ are the eigenvalues of the matrix $\mathbf{G}_\Phi^{-1}\cdot\mathbf{G}_\Phi$ arranged in descending order. Beam-space transformation selects K eigenvectors associated with the top K eigenvalues $\varphi_1,\varphi_2,\dots,\varphi_K$. Next, the ROI constraint p_Φ/p_Φ is approximated as a linear subspace spanned by the eigenvectors corresponding to these K eigenvalues. In this case, the exact value of p_Φ/p_Φ is unknown but it is lower bounded by φ_K (i.e. the minimum of the top K eigenvalues). Therefore, as K increases the lower bound of p_Φ/p_Φ decreases. In other words, BDA increases the ROI selectivity by reducing the number of selected eigenvectors K . Considering $K=M,\dots,2,1$, we plot the decoding accuracies of BDA as a function of λ in Figure 5-4. It should be noted that different from RDA here λ is the lower bound of the power ratio p_Φ/p_Φ , instead of the exact value of p_Φ/p_Φ . From Figure 5-4, as λ increases, the decoding accuracy on TestSet-ROI significantly drops, which is similar to RDA. It indicates that as λ increases, less and less discriminant information originating outside the ROI is used for decoding. However, comparing RDA with BDA, we find that RDA is more efficient than BDA at ROI decoding. Taking the location where the decoding accuracy of both methods on TestSet-Both is 92.6% as an example, the decoding accuracy of RDA on TestSet-ROI is higher and on TestSet-ROI is lower, which is shown in Table 5-1. Such results are observed, because the constraint set of (5.16) is a non-convex set, instead of a simple linear subspace. When BDA is applied, it approximates the non-convex constraint set of (5.16) by a linear subspace. Such an approximation does not fully cover the actual constraint set of (5.16). In other words, BDA attempts to decode the brain states by searching a smaller constraint set than RDA. It, in turn, results in a sub-optimal performance.

Finally, consider the result of SDA. SDA maps surface MEG signals into source space and then uses the sources within ROI for decoding. Hence, SDA should result in perfect ROI selectivity. However, as shown in Table 5-1, the decoding accuracy of SDA in the simulation is extremely low. The low decoding accuracy is observed for SDA due to several possible reasons. First, it is difficult to

accurately estimate the sources inside the ROI for each single trial. It is well-known that source localization is profoundly underdetermined. Most source localization algorithms (e.g., MNE [18]) can only capture the “dominant” sources, but not necessarily the discriminant sources inside the ROI. If the discriminant sources within the ROI are not accurately estimated, we cannot achieve a high decoding accuracy by applying SDA. Second, source localization often results in thousands of sources that are considered as the input features of the decoding, while our training data are limited. In this case, it is extremely difficult to prevent the decoder from over-fitting the training data, even though we have carefully applied feature selection in the experiments. For these reasons, although SDA is supposed to accomplish perfect ROI selectivity, it cannot efficiently extract the discriminant pattern inside the ROI to achieve a high decoding accuracy.

5.6 Experimental Studies

MEG data collected from two human subjects performing both overt and imagined wrist movements are used to compare the proposed RDA algorithm with the other traditional methods and demonstrate its efficiency. All procedures are approved by the Institutional Review Board of both Carnegie Mellon University and the University of Pittsburgh. All experiments are performed in accordance with the approved protocol.

5.6.1 Experimental Setup and Data Preprocessing

In our experiment, two healthy subjects perform a center-out movement task with their right wrist while holding an MEG-compatible joystick. Target images are generated by a computer and projected to a non-magnetic screen in front of the subjects to prompt the movement direction. Two different experimental paradigms are designed: (i) overt movement and (ii) imagined movement. During overt movement trials, the subjects are instructed to move a cursor from the center of the screen to one of the four locations (i.e., up, down, left or right) by making wrist movements (i.e., radial deviation, ulnar deviation, flexion and extension), while keeping the rest of the body relaxed. During imagined movement trials, the subjects are instructed to imagine making the wrist movements

to the target displayed on the screen, while the cursor moves from the center to the target automatically. For both overt and imagined movements, the subjects are instructed to keep their gaze at the center of the screen and only attend to the targets using their peripheral vision.

In this study, we are interested in decoding the movement direction of the right wrist from the left sensorimotor area. Hence, the left sensorimotor area (i.e. the left precentral gyrus and the left postcentral gyrus) is set as the ROI, as shown in Figure 5-1. MEG data are acquired by using a 306-channel whole-head MEG system (Elekta Neuromag[®]) with 1 kHz sampling frequency. The signals recorded by 204 gradiometers are used for decoding. Other signals recorded by 102 magnetometers are not used due to their low signal-to-noise ratio.

In addition, electrooculography (EOG) is used to monitor eye blinks and eye movements. Electromyography (EMG) of wrist flexor and extensor muscles is recorded to make sure that no movement happened during the imagined sessions. All trials with EOG or EMG contamination are rejected.

We consider the ROI decoding problem for the left and the right directions only, since the RDA algorithm is formulated for two-class classifications in this chapter. It has been demonstrated in the literature that significant power modulation related to movement directions can be observed in the low-frequency band (≤ 7 Hz) during a short time window [31]. For this reason, we only considered the low-frequency band (≤ 7 Hz) for the time window $t \in [270 \text{ ms}, 510 \text{ ms}]$, where $t = 0$ ms represents target onset. We apply discrete wavelet transform (DWT) with second-order Symlet wavelet function [70] to decompose the MEG signals from each channel and each trial to multiple levels. The DWT results in four wavelet coefficients within the selected frequency band and time window for each channel. These four wavelet coefficients correspond to four time windows $t \in [270 \text{ ms}, 330 \text{ ms}]$, $[330 \text{ ms}, 390 \text{ ms}]$, $[390 \text{ ms}, 450 \text{ ms}]$ and $[450 \text{ ms}, 510 \text{ ms}]$ and cover the low-frequency band (≤ 7 Hz).

We investigate the ROI decoding problem for each time-frequency window separately. Since there are two subjects, two experimental paradigms (i.e., overt and imagined movements) for each subject and four time-frequency windows for each paradigm, we have 16 different data sets in total. To simplify the notation, the data sets associated with the i th ($i = 1, 2$) subject and the j th ($j = 1, 2, 3, 4$)

time-frequency window are labeled as “Subi O_j ” and “Subi I_j ” for overt and imagined movements respectively. For each data set, the feature vector is 204-dimensional, corresponding to the wavelet coefficients calculated for 204 gradiometers.

During a separate visit, the same subjects take a standard structural MRI scan. The MRI data are used to define the vertices and the associated current dipoles to formulate ROI constraint. Towards this goal, the MRI data are co-registered with the MEG data. A fixed source grid with 2-mm spacing is generated from the gray/white matter boundary of the MRI by using FreeSurfer [84]. In addition, the cortical surface is parcellated and labeled as different cortical regions by using FreeSurfer.

5.6.2 ROI Decoding Results

As discussed in Section 5.5.1, the performance of ROI decoding should be evaluated by two aspects: (i) ROI selectivity and (ii) discriminant signal selectivity. For the simulation, we evaluate these two aspects by test the discriminant function derived from a decoding algorithm on different test sets. Here, for the real data, we evaluate the discriminant signal selectivity by the decoding accuracy using six-fold cross-validation [52], and assess the ROI selectivity using DPSL. Considering the spatial filter $y = \mathbf{w}^T \cdot \mathbf{x}$ in (5.1), DPSL calculates the gain of the spatial filter for the neural source (i.e., a current dipole) at each spatial location to determine whether the MEG signal generated by this particular source can pass the spatial filter. The resulting gain is normalized between 0 and 1. A non-zero gain implies that the MEG signal generated by the corresponding source can pass the spatial filter and, hence, contribute to decoding. In the ideal case of ROI decoding, the neural sources with non-zero gain values should sit inside ROI only.

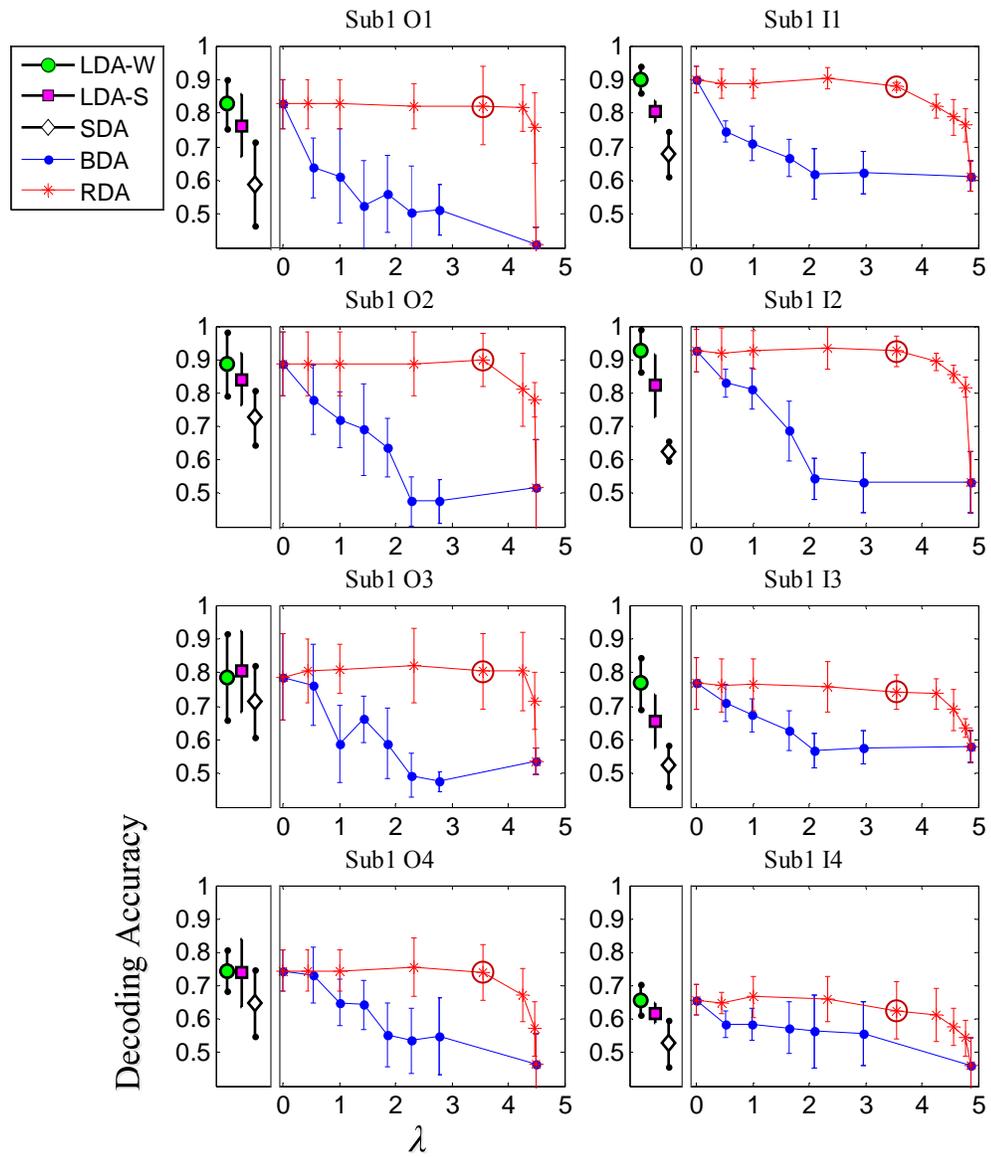


Figure 5-5. Movement decoding accuracies (both mean and standard deviation) of Subject 1 are plotted for LDA-W, LDA-S, SDA, BDA and RDA. For BDA and RDA, the accuracies are plotted as a function of λ (i.e., the constraint of the power ratio $p_{\phi}/p_{\bar{\phi}}$ defined in (5.13)).

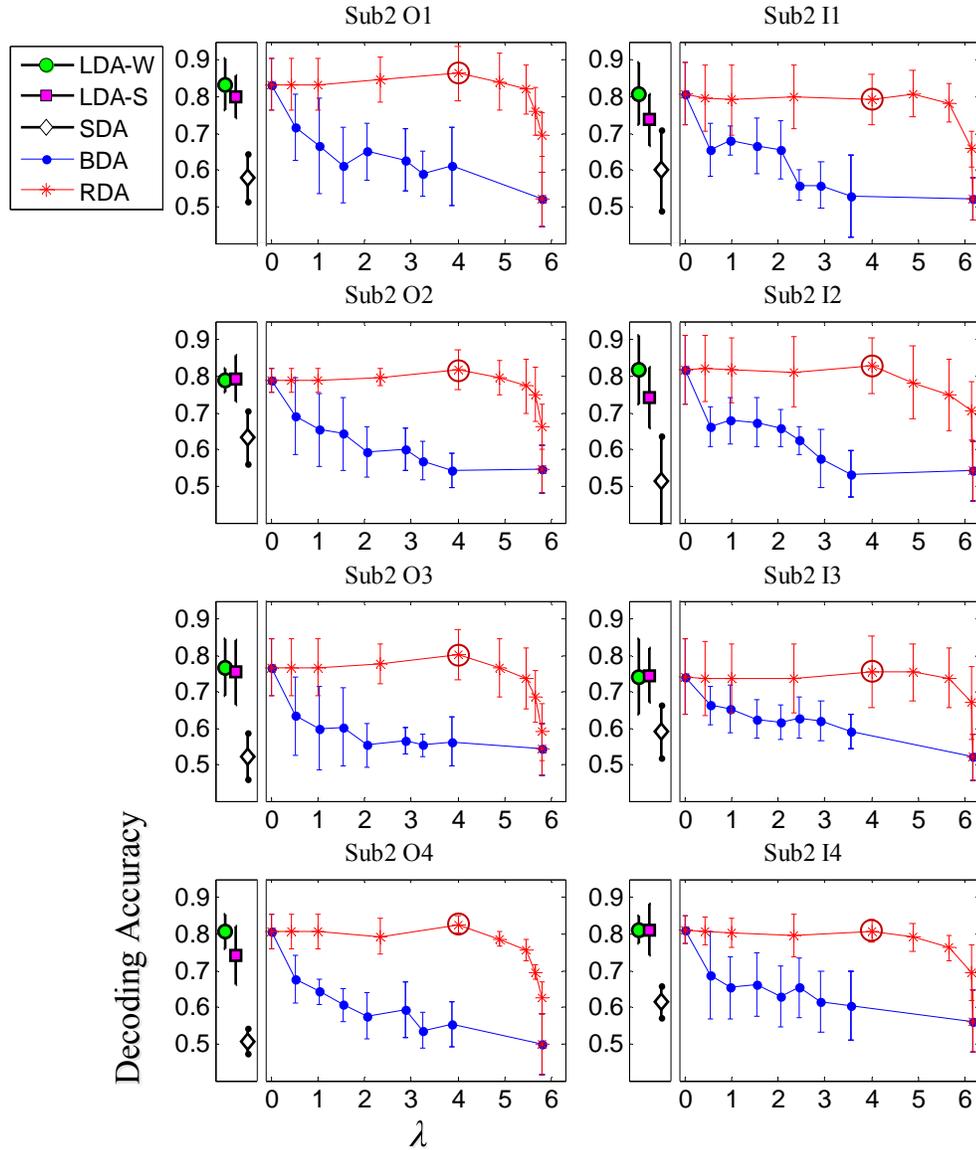


Figure 5-6. Movement decoding accuracies (both mean and standard deviation) of Subject 2 are plotted for LDA-W, LDA-S, SDA, BDA and RDA. For BDA and RDA, the accuracies are plotted as a function of λ (i.e., the constraint of the power ratio p_{Φ}/p_{Φ} defined in (5.13)).

Figure 5-5 and Figure 5-6 show the decoding accuracies of five different algorithms that we implement. The decoding accuracies for both BDA and RDA are plotted as a function of λ (i.e., the constraint of the power ratio p_{Φ}/p_{Φ} defined in (5.13)). Furthermore, the spatial locations of the selected discriminant sources corresponding to LDA-W, LDA-S and RDA are estimated by DPSL and are

shown in Figure 5-7 and Figure 5-8. Since SDA and BDA are expected to have good ROI selectivity, we do not show the DPSL results for SDA and BDA in the figures.

Studying Figure 5-5 - Figure 5-8 reveals a number of important observations. First, while LDA-W achieves a high decoding accuracy as shown in Figure 5-5 and Figure 5-6, it completely ignores the ROI constraint. Therefore, as shown in Figure 5-7 and Figure 5-8, the discriminant sources corresponding to LDA-W are widely distributed over several cortical regions. Such source localization results are observed, because our experimental paradigm involves both visual stimulus and overt/imagined movements. Without considering ROI, LDA-W picks up the discriminant information from multiple cortical regions for decoding.

By using the MEG channels located on top of ROI, LDA-S helps to constrain the discriminant sources close to the ROI. However, as shown in Figure 5-7 and Figure 5-8, LDA-S fails to constrain the discriminant sources within the ROI in most cases. In addition, because LDA-S only uses a limited number of channels for decoding, its accuracy is not as high as that of LDA-W, as shown in Figure 5-5 and Figure 5-6. These results demonstrate an important fact that directly selecting a subset of MEG channels on top of ROI is not an efficient approach to extract the discriminant information for ROI decoding.

As shown in Figure 5-5 and Figure 5-6, the decoding accuracy of SDA is extremely low. The same as in the simulation study, this low decoding accuracy is observed for SDA due to several possible reasons. First, it is difficult to accurately estimate the sources inside the ROI for each single trial. Second, source localization often results in thousands of sources that are considered as the input features of decoding, while our training data are limited. In this case, it is extremely difficult to prevent the decoder from over-fitting the training data, even though we have carefully applied feature selection in the experiments. For these reasons, although SDA is able to accomplish perfect ROI selectivity, it cannot efficiently extract the discriminant pattern inside the ROI to achieve a high decoding accuracy.

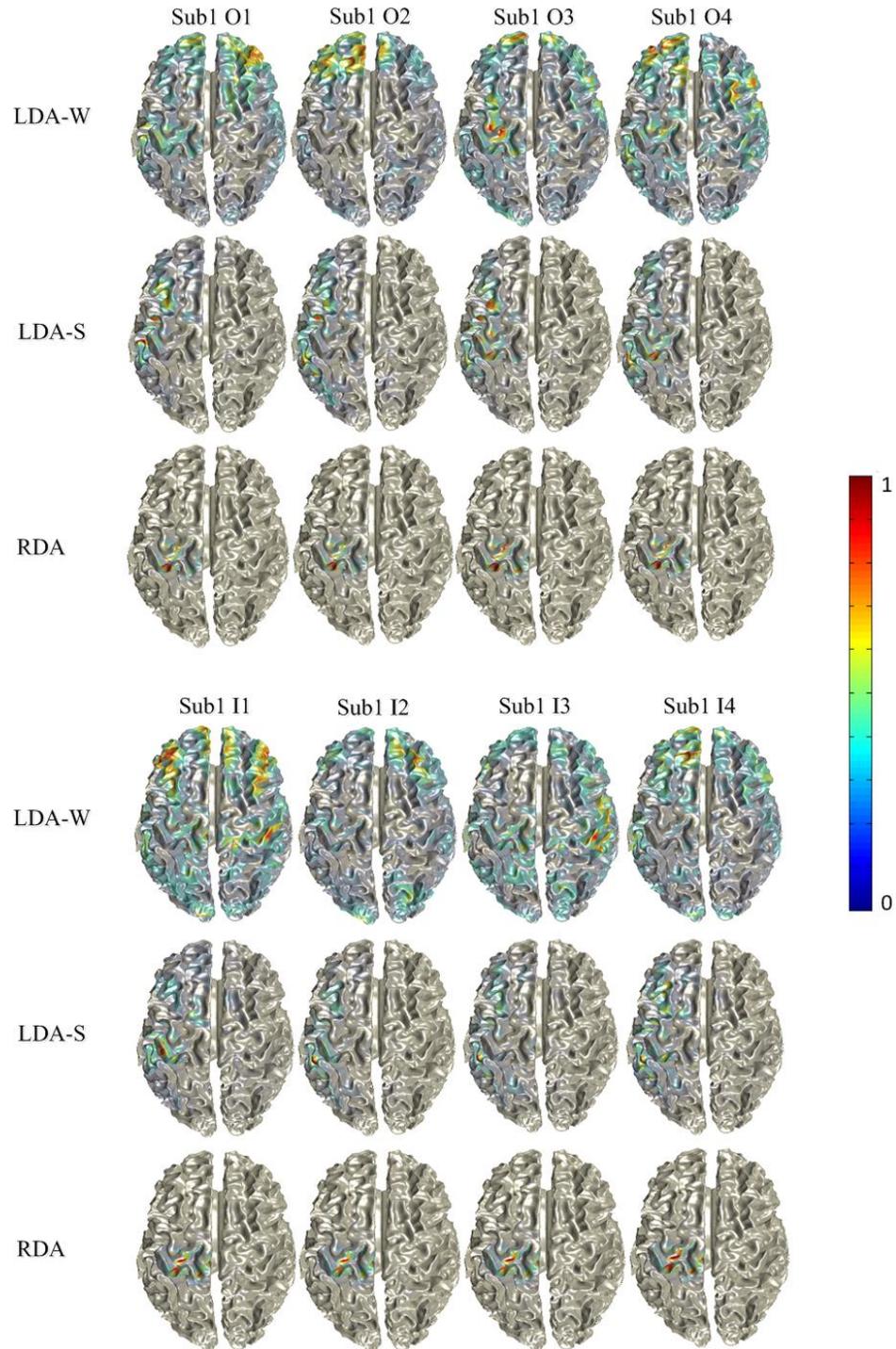


Figure 5-7. DPSL is applied to find the spatial locations of discriminant sources for LDA-W, LDA-S and RDA of Subject 1. Red color indicates the spatial locations where the gain of the spatial filter $y = \mathbf{w}^T \cdot \mathbf{x}$ in (5.1) is large. In other words, the MEG signals generated by the sources at these locations can pass the spatial filter $y = \mathbf{w}^T \cdot \mathbf{x}$ and contribute to decoding.

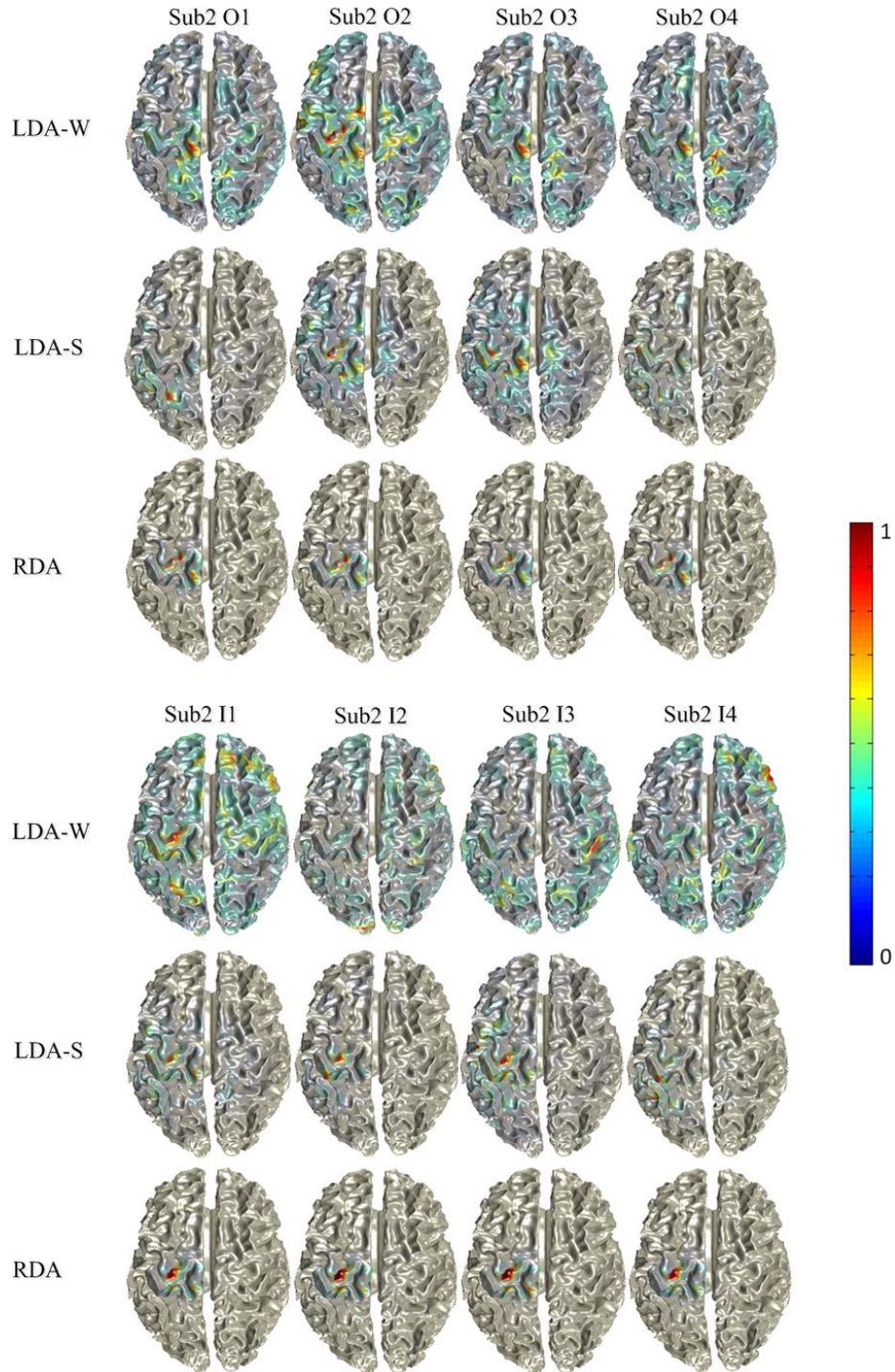


Figure 5-8. DPSL is applied to find the spatial locations of discriminant sources for LDA-W, LDA-S and RDA of Subject 2. Red color indicates the spatial locations where the gain of the spatial filter $y = \mathbf{w}^T \cdot \mathbf{x}$ in (5.1) is large. In other words, the MEG signals generated by the sources at these locations can pass the spatial filter $y = \mathbf{w}^T \cdot \mathbf{x}$ and contribute to decoding.

Next, we consider our proposed RDA algorithm and compare it with other traditional methods. As shown in Figure 5-5 and Figure 5-6, the decoding accuracy of RDA is plotted as a function of λ . If λ reaches the lower bound, RDA and LDA-W are equivalent. Otherwise, if λ reaches the upper bound, the decoding accuracy drops significantly. In this extreme case, the spatial filter $y = \mathbf{w}^T \cdot \mathbf{x}$ formed by RDA is determined by the dominant eigenvector of $\mathbf{G}_\phi^{-1} \cdot \mathbf{G}_\phi$, which maximizes the total signal power originating from the ROI, instead of the discriminant information generated by the ROI. In other words, if λ is set to its upper bound, RDA maximizes the ROI selectivity only and, hence, is unlikely to achieve a high decoding accuracy. When λ varies between these two extreme cases, RDA explores the trade-off between ROI selectivity and decoding accuracy. As λ increases from the lower bound to the upper bound, the ROI selectivity is improved while the decoding accuracy is expected to decrease.

It is also important to note that the decoding accuracy of RDA is almost unchanged, as λ varies between the lower bound and a reasonably large value (e.g. the values denoted by the red circles in Figure 5-5 and Figure 5-6). For several data sets such as Sub2 O3 and Sub2 O4, the decoding accuracy even slightly increases, as λ increases. This observation can be explained by the following two reasons. First, the discriminant information generated by multiple cortical regions may be redundant. Hence, a high decoding accuracy can be achieved by extracting the discriminant information from the ROI, instead of other cortical regions. Second, as only a small number of training trials are available in the training sets, the ROI constraint posed by RDA can act as an efficient regularization strategy to avoid over-fitting.

The spatial locations of the discriminant sources captured by RDA are shown in Figure 5-7 and Figure 5-8, when λ is set to the values denoted by the red circles in Figure 5-5 and Figure 5-6. From Figure 5-7 and Figure 5-8, we observe that when λ is large enough the proposed RDA algorithm successfully forms a spatial filter to pass the MEG signals generated by the ROI only. In other words, unlike LDA-W and LDA-S where the discriminant sources are widely distributed over multiple cortical regions, all sources corresponding to RDA appropriately sit inside the ROI, as shown in Figure 5-7 and Figure 5-8.

Finally, comparing the decoding accuracies between BDA and RDA as a function of λ (i.e., the

constraint of the power ratio p_ϕ/p_Φ) in Figure 5-5 and Figure 5-6, we notice that these two methods result in the same decoding accuracy, when the value of λ reaches its lower bound or the upper bound. Remember that the lower and upper bounds are determined by (5.17) and (5.18), respectively. When the ratio reaches the lower bound, the constraint is always satisfied for any $\mathbf{w} \in R^M$. On the other hand, when the ratio reaches the upper bound, the constraint set is a one-dimensional subspace defined by the dominant eigenvector of $\mathbf{G}_\Phi^{-1} \cdot \mathbf{G}_\phi$. In these two extreme cases, BDA and RDA are equivalent and, hence, their decoding results are identical. However, when the value of λ varies between the two bounds, RDA results in significantly improved decoding accuracies over BDA. In these cases, the constraint set of (5.16) is non-convex, instead of a simple linear subspace. When BDA is applied, it approximates the non-convex constraint set of (5.16) by a linear subspace. Such an approximation does not fully cover the actual constraint set of (5.16). In other words, BDA attempts to decode the movement information by searching a smaller constraint set than RDA. It, in turn, results in a sub-optimal decoding accuracy.

5.7 Discussion

In this chapter, we propose an RDA algorithm for ROI decoding. RDA can smoothly control the ROI selectivity of the decoder by varying a parameter λ in (5.16). Compared with the other traditional methods, when the same decoding accuracy is achieved, RDA is more efficient at extracting the discriminant signals from the ROI and suppressing the discriminant signal originating outside the ROI, as is demonstrated in Section 5.5.

It is worth mentioning that there remain a number of open questions related to the proposed RDA. First, RDA provides a tool that is able to extract the optimal discriminant pattern for decoding based on MEG signals originating from pre-specified cortical regions, where the ROI selectivity is controlled by a user-defined parameter λ . However, how to appropriately choose the parameter λ to satisfy the requirements of practical applications remains an open question. One possible solution is to plot the decoding accuracy as a function of λ based on the training data, as shown in Figure 5-5 and

Figure 5-6. Next, set a lowest acceptable decoding accuracy according to the application. For example, if the user cannot tolerate the high failure rate when the decoding accuracy is less than 80%, we set the lowest acceptable decoding accuracy as 80%. Once this value is determined, take the largest λ associated with this value as the selected λ .

Second, in this chapter, the ROI constraint aims to extract the signals from the ROI and suppress all signals originating outside the ROI. Since the ROI is typically a small area surrounded by the other cortical regions, it is difficult to eliminate all signals originating outside the ROI without removing any discriminant signals from the ROI. However, if the large cortical area outside the ROI is replaced with a small region of non-interests, the ROI constraint in (5.16) should be more efficient to control ROI selectivity. Therefore, in practice, if a small region of non-interests is known in advanced, we can replace the cortical area outside the ROI with the small region of non-interests in (5.11).

Third, RDA is limited to the cases where one feature is extracted from each MEG channel for decoding in this chapter. However, it is possible to extend RDA to handle the general cases with multiple features per MEG channel. For example, if we use the feature extraction method described in Section 5.6.1, multiple features of the same channel correspond to different time windows. In this case, the linear discriminant function $y = \mathbf{w}^T \cdot \mathbf{x}$ in (5.1) can be conceptually viewed as a spatio-temporal filter. We can use the same objective function as (5.2), but re-define p_ϕ and $p_{\bar{\phi}}$ in (5.13) as the power of the output of this spatio-temporal filter generated by the sources inside and outside the ROI, respectively. Next the optimization formulated in (5.16) is used to solve the ROI decoding problem.

Fourth, RDA is formulated for two-class classification in this chapter. It is possible to extend RDA to multiple classes by integrating it with a multi-class classification algorithm developed by the machine learning community. For example, we can use the one-versus-rest approach, which attempts to build a K -class classifier by combining K two-class classifiers [52].

Finally, the possible applications of the proposed RDA algorithm should be further explored. For example, a brain computer interface based on RDA may help to promote neuroplasticity within a target cortical region, and it may be applied to upper limb recovery for the patients with stroke [88]. Furthermore, by setting ROI to different cortical regions, RDA may be used to study and compare the

functional roles of different cortical regions when performing specific brain tasks. It, in turn, may serve as a powerful tool for pre-surgical localization of the cortical areas that are significantly modulated by the intended brain task [93]. These possible applications will be further studied in our future research.

5.8 Summary

Human mental processes are controlled by distributed cortical networks and, hence, task-related activity often appears in multiple cortical regions. However, recent MEG applications pose an emerging need of decoding brain states by signals originating from a specific cortical region. Therefore, we need to develop a new tool to extract the discriminant information from the ROI to distinguish different brain states, while suppressing signals originating outside the ROI.

In this chapter, we propose an ROI-constrained discriminant analysis (RDA) algorithm to address the aforementioned problem. RDA integrates linear classification and beamspace transformation into a unified framework by formulating a constrained optimization problem. A numerical solver is developed to solve the non-convex optimization problem posed by RDA with guaranteed global convergence. Our simulation studies and the experimental results based on human subjects demonstrate that RDA can efficiently extract the discriminant pattern from pre-specified cortical regions to accurately distinguish different brain states.

Chapter 6

Thesis Summary & Future Work

6.1 Summary

Recent advances in MEG provide a significant new approach to study task-related neural activity in humans. However, since MEG measures the magnetic field on scalp surface, it does not directly uncover the neural activity in the brain space. To study and understand the neural activity based on MEG measurements, source localization has been identified as the most important tool. In practice, as discussed in Section 2.3, it is well-known that the source localization problem is profoundly underdetermined due to the limited observability of MEG measurements. Hence, it is difficult to estimate all sources accurately, especially if the signal-to-noise ratio is low. In most cases, a source localization algorithm can only capture the dominant sources which are not necessarily to be the sources corresponding to the task-related activity. If the sources corresponding to task-related activity are not accurately estimated, we cannot expect correct results in the following analysis. It, in turn, poses an immediate need to re-think the fundamental strategy that connects the MEG recording with the underlying task-related neural activity.

In this thesis, we put forward that MEG decoding can be used as an instrument to investigate task-related activity. MEG decoding has been extensively studied over the past decade. The objective is to predict different brain states from MEG recordings. Most MEG decoding algorithms are particularly designed to detect the optimal discriminant pattern existing in MEG sensor space. Since the discriminant pattern in MEG sensor space is generated by task-related activity in the brain space, the task-related activity should be significant if high decoding accuracy is achieved. Based on this idea,

to investigate whether task-related activity is detectable in MEG recordings, we can start from MEG decoding in the sensor space. Developing decoding algorithms with high accuracy is one of our major works. In Chapter 3, advanced signal processing and machine learning techniques are applied to extract the efficient time-frequency features and discover the optimal discriminant pattern between brain states existing in MEG sensor space. In particular, we propose a CLDA algorithm to address the small training set problem posed by MEG decoding. CLDA applies a spectral clustering algorithm to automatically partition all MEG features into several groups where the within-group correlation is maximized and the between-group correlation is minimized. As such, the covariance matrix of all features can be approximated as a block diagonal matrix, thereby facilitating us to accurately extract the correlation information required by decoding from a small set of training data.

The high accuracy of MEG decoding indicates that task-related activity is detectable in the MEG recordings. Then our next question is: can we localize the cortical regions that produce the task-related activity? In Chapter 4, we propose a DPSL algorithm to localize the sources associated with task-related activity. DPSL bridges the discriminant pattern in MEG sensor space to the task-related neural activity in the brain space. Specifically, it consists of two major steps. First, decoding technique is applied to find the discriminant pattern in MEG sensor space that can optimally differentiate different brain states. Second, based on the forward model of the magnetic field, the task-related sources are found by studying the contribution of all sources to the discriminant pattern. Since most decoding algorithms are particularly designed to be robust to noise, by applying decoding techniques in the first step DPSL can efficiently reduce the impact of noise and accurately identify the task-related sources, as demonstrated by the examples in Section 4.4.

Conventional MEG decoding uses signals originating from the whole brain. The optimal discriminant pattern is generated by the strong task-related sources distributed all over the brain. However, recent MEG applications pose an emerging need of decoding brain states by signals originating from a specific cortical region instead of inside the whole brain. To address this problem, we propose a RDA algorithm in Chapter 5. RDA integrates linear classification and beamspace transformation into a unified framework by formulating a constrained optimization problem. A

numerical solver is developed to solve the non-convex optimization problem posed by RDA with guaranteed global convergence. Our simulation studies and the experimental results based on human subjects demonstrate that RDA can efficiently extract the discriminant pattern from pre-specified cortical regions to accurately distinguish different brain states.

6.2 Future Work

In this thesis, we develop several tools to investigate task-related neural activity based on MEG. One of the most important research directions of our future work is to explore the clinical applications of these tools. In Section 6.2.1, we propose a possible application of these approaches to support motor imagery practice in stroke rehabilitation. Next in Section 6.2.2, we further present a number of promising directions that can be explored to extend this work.

6.2.1 Application to Stroke Rehabilitation

Motor impairment due to stroke is one of the major causes of permanent disability. Until recently, however, the available methods do not restore normal or near normal motor function in most patients. There is now abundant evidence that motor imagery (MI) practice may lead to enhanced functional recovery of paralyzed limbs among stroke sufferers [88], [91], [109]-[113]. The expectation of this approach is to directly manipulate brain activity to induce plasticity within the affected brain structures to improve motor function.

Although MI is promising for stroke rehabilitation, it is still limited in many respects [111]. One problem of using MI is how to feedback the patients' brain activity to confirm their engagement in the MI task on-line [87], [88], [90]. Analyzing neural signals and providing feedback in real-time is one of the core characteristics of a brain-computer interface (BCI). BCI can provide an on-line measure of MI as a neurofeedback for the patients to help them undertake MI practice with sufficient focus, which has gained particular attention for stroke rehabilitation [87], [88], [90]. In this section, we will discuss how to apply the algorithms proposed in this thesis to advance the application and

evaluation of a MEG-based BCI system for stroke rehabilitation.

It should be noted that the choice of an appropriate MI task is important for the rehabilitation process. Although some preliminary studies have been reported, it is at present still unclear which MI task may be optimal [87], [90]. In this section, we take the MI task that the patients imagine moving the wrists to the left (i.e. flexion) or to the right (i.e. extension) directions as an example [87]. However, the approach we present here is independent of the choice of MI tasks.

A crucial problem for MI practice is to encourage neural activity in a target brain area to induce neural plasticity and restore function in this area. But conventional BCIs tend to make use of signals containing rich modulation information no matter where they come from. While this could result in high decoding accuracy, it may not induce the desired behavioral changes of the stroke rehabilitation [90]. For example, the MI task that patients imagine left or right wrist movement is designed to enhance neural activity in the motor cortex. Ideally, the patient is expected to activate motor cortex to generate modulation signals that discriminate left and right movements. The BCI system will decode their brain signals into one of the two directions based on the modulation signals and provide feedback to the patients. Through the feedback decoding accuracy, the patients can efficiently monitor and adjust their brain activity of the motor cortex. However, in practice, if the patients pay attention to left or right spatial location when executing the motor imagery, their parietal cortex will also be activated during the practice. If the modulation signal from the parietal cortex is strong, it will be taken by the BCI system for decoding. As a result, the patient is trained to activate the parietal cortex rather than the motor cortex during the practice. This issue cannot be solved simply by changing the rehabilitation protocol or by training the patient. Therefore, a method is desirable that could efficiently extract the discriminant information from a target cortical area, while allowing the exclusion of discriminant signals from other cortical areas. In Chapter 5, we have shown that RDA can be used successfully for this problem.

Figure 6-1 illustrates the overall system design of a BCI for the stroke rehabilitation, which is composed of three modules. The first module is used to record the real-time MEG signals and perform several preprocessing steps, such as removing noise, compensating disturbances due to head

movements. The second module takes in the processed MEG, uses the MRI of the same patient to accurately target the brain area of interest, and then applies RDA to decode different brain states based on discriminant signals originating inside the target cortical area. Finally, the last module feeds back the decoding results to the patients. For example, for the task of imagining left or right wrist movements, we can use a ball moving to the left or to the right on a screen in front of the patient to display the decoding result.

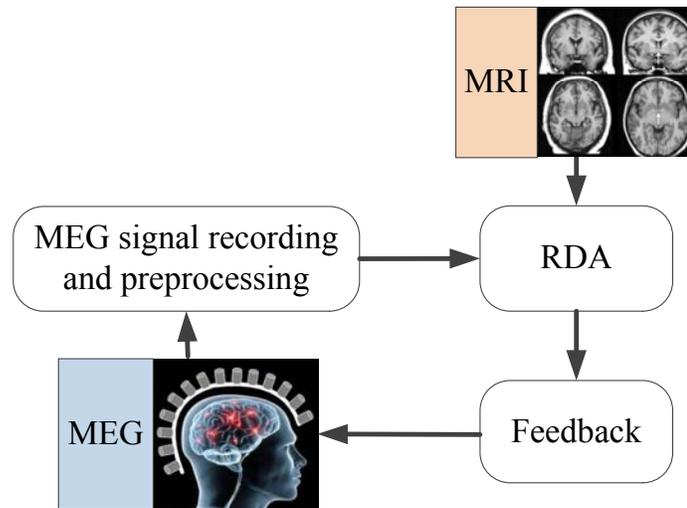


Figure 6-1. Schematic illustration of system design and data flow.

The objective of this rehabilitation is to induce neural plasticity within the target brain area to improve motor function. Therefore, during the rehabilitation process it is important to evaluate the performance of this therapy by comparing the cortical activation before and after a period of MI practice. This difference might be minor, which is difficult to be detected by the traditional methods based on MEG source localization. However, MEG decoding and DPSL proposed in this thesis should be good at the problem. In particular, we consider two brain states of the same patient corresponding to performing the same mental task before and after a session of rehabilitation therapy. The mental task can be the same as the MI task. Also it can be other tasks that are appropriate for evaluating motor activity. MEG associated with the two brain states are recorded, respectively. First, advanced decoding techniques are applied to judge the distinguishability between the MEG recordings of the two brain states. High decoding accuracy indicates significant changes of the cortical activity after this

period of rehabilitation therapy. Next, DPSL is applied to check whether the changes are within the target cortical area. If significant changes are found within the target cortical area, the rehabilitation therapy should be efficient in inducing desired neural plasticity.

6.2.2 Other Research Directions

There are a number of other promising directions that can be explored to further extend this work.

First, MEG decoding requires multiple trials associated with every brain states to derive the discriminant function. In the data collection phase, when a subject is presented with a stimulus, the stimulus onset defines a common time-reference between trials. Although the subject is asked to repeatedly execute the same training task after the stimulus onset, the time delay between the brain response and the stimulus onset (i.e. the response latency) might be different from trial to trial. Due to the variability of the response latency, MEG recordings time-locked to the stimulus onset are not properly aligned with respect to the task-related neural activity. This leads to undesired variances among training data and, hence, results in a biased discriminant function for decoding. To reduce the impact of variability caused by the response latency, more efforts can be made for synchronizing the trials. The synchronization techniques have been widely studied in speech and video processing [104], [105]. For MEG/EEG community, there are a few studies in recent years [106]. If these techniques help to align the MEG training trials, the accuracy of current MEG decoding algorithms will be further improved.

Second, due to the neural plasticity, the task-related activity might vary over time. This is especially significant for rehabilitation applications, since during the recovery process some brain areas are reactivated or reorganized. Detecting this kind of shift in data stream and dynamically adapting the decoder to the changes of the task-related activity are important for us to further improve the decoding performance. However, conventional MEG decoding does not consider the changes, which assumes the neural activity is stable over the long training session. One possible research direction is to investigate adaptive learning techniques to track and quickly adapt the brain state

changes.

Third, as mentioned in Section 4.6, DPSL has a bias towards superficial sources. That's because the sensitivity of MEG sensors is not uniform to sources across different cortical locations [107]. In fact, it follows from Maxwell's equations that the lead fields of MEG sensors have a maximum at the source space closest to the sensors, and, hence, the source localization solutions are often biased to the superficial locations to which the sensors are most sensitive [108]. This character severely impacts the applications of MEG source localization to deep sources. Improving the accuracy of deep source localization is an interesting topic for the future research.

Finally, throughout this thesis we restrict ourselves to MEG recordings. However, as the signals of EEG and MEG are quite similar, our algorithms developed for MEG could be transferred smoothly to EEG data. For example, the recording channels of an EEG system are much less than the whole-head MEG system typically, and hence the ill-posedness problem of EEG source localization is the same as, if not more challenging than, the MEG source localization. In this case, DPSL should also be a good option for the task-related source localization of EEG.

Bibliography

- [1] E. R. Kandel, J. H. Schwartz and T. M. Jessell, *Principles of Neural Science*, 4th Edition, McGraw-Hill, 2000.
- [2] A. Zani and A. M. Proverbio, *The Cognitive Electrophysiology of Mind and Brain*, Academic press, 2002.
- [3] M. Hamalainen, R. Hari, R. Ilmoniemi, J. Knuutila and O. Lounasmaa, "Magnetoencephalography – theory, instrumentation, and applications to noninvasive studies of the working human brain," *Rev. Modern Phys.*, vol. 65, no. 2, pp. 413-497, Oct. 1993.
- [4] D. Cohen, "Magnetoencephalography: evidence of magnetic fields produced by alpha rhythm currents," *Science*, vol. 161, pp. 784-786, 1968.
- [5] D. Cohen, "Magnetoencephalography: detection of the brain's electrical activity with a superconducting magnetometer," *Science*, vol. 175, pp. 664-666, 1972.
- [6] C. Gratta, V. Pizzella, F. Tecchio and G. Romani, "Magnetoencephalography – a noninvasive brain imaging method with 1 ms time resolution," *Rep. Prog. Phys.*, vol. 64, pp. 1759-1814, 2001.
- [7] D.S. Barth, W.W. Sutherling and J. Beatty, "Intracellular currents of interictal penicillin spikes: evidence from neuromagnetic mapping," *Brain Res*, vol. 368, pp. 36-48, 1986.
- [8] B. He and Z. Liu, "Multimodal functional neuroimaging: integrating functional MRI and EEG/MEG," *IEEE Rev. Biomed. Eng.*, vol. 1, no. 1, pp. 23-40, Nov. 2008.
- [9] S. Baillet, J. C. Mosher and R. M. Leahy, "Electromagnetic brain mapping," *IEEE Signal Processing Magazine*, vol. 18, no. 6, pp. 14-30, Nov. 2001.
- [10] F. Rosler, C. Ranganath, B. Roder and R. H. Kluwe, *Neuroimaging of Human Memory*, Oxford university press, 2009.

- [11] S. A. Huettel, A. W. Song and G. McCarthy, *Functional Magnetic Resonance Imaging (2nd Edition)*, Sinauer Associates, Inc. Publishers, 2009.
- [12] J. Kissler, M. M. Muller, T. Fehr, B. Rockstroh and T. Elbert, "MEG gamma band activity in schizophrenia patients and healthy subjects in a mental arithmetic task and at rest," *Clinical Neurophysiology*, vol. 111, no. 11, pp. 2079-2087, Nov. 2000.
- [13] I. Kreitschmann-Andermahr, T. Rosburg, T. Meier, H. P. Volz, H. Nowak and H. Sauer, "Impaired sensory processing in male patients with schizophrenia – a magnetoencephalographic study of auditory mismatch detection," *Schizophrenia Research*, vol 35, no. 2, pp. 121-129, Jan. 1999.
- [14] G. R. Barnes and A. Hillbrand, "Statistical flattening of MEG beamformer images," *Human Brain Mapping*, vol. 18, no. 1, pp. 1-12, Jan. 2003.
- [15] Y. Xu, G. P. Sudre, W. Wang, D. J. Weber and R. E. Kass, "Characterizing global statistical significance of spatio-temporal hot spots in magnetoencephalography/electroencephalography source space via excursion algorithms," *Statistics in Medicine*, vol. 30, no. 23, pp. 2854-2866, Jul. 2011.
- [16] J. C. Mosher, P. S. Lewis, and R. M. Leahy, "Multiple dipole modeling and localization of spatio-temporal MEG data," *IEEE Trans. Biomed. Eng.*, vol. 39, pp. 541–557, 1992.
- [17] B. Veen, W. Drongelen, M. Yuchtman and A. Suzuki, "Localization of brain electrical activity via linearly constrained minimum variance spatial filtering," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 9, pp. 867-880, Sep. 1997.
- [18] M. Hamalainen and R. Ilmoniemi, "Interpreting magnetic fields of the brain: minimum norm estimates," *Med. Biol. Eng. Comput.*, vol. 32, pp. 35-42, Jan. 1994.
- [19] R. Pascual-Marqui, "Standardized low resolution brain electromagnetic tomography (sLORETA): technical details," *Methods Findings Exp. Clin. Pharmacol.*, 24D, pp. 5-12, 2002.
- [20] K. Uutela, M. Hamalainen and E. Somersalo, "Visualization of magnetoencephalographic data using minimum current estimates," *NeuroImage*, vol. 10, pp. 173-180, 1999.
- [21] W. Ou, M. Hamalainen and P. Golland, "A distributed spatio-temporal EEG/MEG inverse

- solver,” *NeuroImage*, vol. 44, pp. 932-946, 2009.
- [22] P. Hansen, M. Kringelbach and R. Salmelin, *MEG: An Introduction to Methods*, Oxford university press, 2010.
- [23] S. Ahlfors and G. Simpson, “Geometrical interpretation of fMRI-guided MEG/EEG inverse estimates,” *NeuroImage*, vol. 22, no. 1, pp. 323-332, May 2004.
- [24] M. Scherg, “Fundamentals of dipole source potential analysis,” in *Auditory Evoked Magnetic Fields and Potentials*, vol. 6, 1990.
- [25] D. Pantazis, T. Nichols, S. Baillet, R. Leahy, “A comparison of random field theory and permutation methods for the statistical analysis of MEG data,” *NeuroImage*, vol. 25, pp. 383-394, 2005.
- [26] D. Purves, G. Augustine, D. Fitzpatrick, W. Hall, A. Lamantia, J. Mcnamara and S. Williams (Ed.), *Neuroscience*, Sinauer Associates, Inc., 2007.
- [27] T. Bradberry, F. Rong and J. Contreras-Vidal, “Decoding center-out hand velocity from MEG signals during visuomotor adaptation,” *NeuroImage*, vol. 47, pp. 1691-1700, 2009.
- [28] J.D. Haynes and G. Rees, “Decoding mental states from brain activity in humans,” *Neuroimaging*, vol. 7, pp. 523-534, 2006.
- [29] J. Wolpaw, N. Birbaumer, W. Heetderks, D. McFarland, P. Peckham, G. Schalk, E. Donchin, L. Quatrano, C. Robinson and T. Vaughan, “Brain-computer interface technology: a review of the first international meeting,” *IEEE Trans. Rehab. Eng.*, vol. 8, no. 2, pp. 164-173, Jun. 2000.
- [30] T. Lal, M. Schroder, T. Hinterberger, J. Weston, M. Bogdan, N. Birbaumer and B. Scholkopf, “Support vector channel selection in BCI,” *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1003-1010, Jun. 2004.
- [31] S. Waldert, H. Preissl, E. Demandt, B. Christoph, B. Niels, A. Aertsen and C. Mehring, “Hand movement direction decoded from MEG and EEG,” *J. Neurosci.*, vol. 28, no. 4, pp.1000-1008, Jan. 2008.
- [32] M. Grosse-Wentrup and M. Buss, “Multiclass common spatial patterns and information theoretic feature extraction,” *IEEE Trans. Biomed. Eng.*, vol. 55, no. 8, pp. 1991-2000, Aug.

2008.

- [33] M. Grosse-Wentrup, C. Liefhold, K. Gramann and M. Buss, "Beamforming in noninvasive brain-computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 4, pp. 1209-1219, Apr. 2009.
- [34] M. Gerven, A. Bahramisharif, T. Heskes and O. Jensen, "Selecting features for BCI control based on a covert spatial attention paradigm," *Neural Networks*, vol. 22, no. 9, pp. 1271-1277, Jun. 2009.
- [35] J. Zhang, G. Sudre, X. Li, W. Wang, D. Weber and A. Bagic, "Clustering linear discriminant analysis for MEG-based brain computer interfaces," *IEEE Trans. Rehab. Eng.*, vol. 19, no. 3, pp. 221-231, Jun. 2011.
- [36] G. Pfurtscheller, C. Neuper, C. Guger, W. Harkam, H. Ramoser, A. Schlogl, B. Obermaier, and M. Pregenzer, "Current trends in Graz brain-computer interface (BCI) research," *IEEE Trans. Rehab. Eng.*, vol. 8, no. 2, pp. 216-219, Jun. 2000.
- [37] J. Kronegg, G. Chanel, and S. Voloshynovskiy, "EEG-based synchronized brain-computer interfaces: a model for optimizing the number of mental tasks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 15, no. 1, pp. 50-58, Mar. 2007.
- [38] P. Shenoy, K. Miller, J. Ojemann and R. Rao, "Generalized features for electrocorticographic BCIs," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 1, pp. 273-280, Jan. 2008.
- [39] T. Lal, M. Schroder, N. Hill, H. Preissl, T. Hinterberger, J. Mellinger, M. Bogdan, W. Rosenstiel, T. Hofmann, N. Birbaumer and B. Scholkopf, "A brain computer interface with online feedback based on magnetoencephalography," in *Proc. Int. Conf. Machine Learning*, pp. 465-472, 2005.
- [40] H. Ramoser, J. Mueller-Gerking and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 441-446, Dec. 2000.
- [41] G. Blanchard and B. Blankertz, "BCI competition 2003 – data set IIa: spatial patterns of self-controlled brain rhythm modulations," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1062-

1066, Jun. 2004.

- [42] J. Ye and S. Ji, “Discriminant analysis for dimensionality reduction: an overview of recent developments,” in *Biometrics: Theory, Methods and applications*, Wiley/IEEE Press, 2009.
- [43] Q. Xu, H. Zhou, Y. Wang and J. Huang, “Fuzzy support vector machine for classification of EEG signals using wavelet-based features,” *Medical Engineering & Physics*, vol 31, pp. 858–865, 2009.
- [44] Y.P.A. Yong, N. J. Hurley, and G.C.M. Silvestre, “Single-trial EEG classification for brain-computer interface using wavelet decomposition,” in *European Signal Processing Conference (EUSIPCO)*, Sept. 2005.
- [45] B. Graimann, J. Huggins, S. Levine and G. Pfurtscheller, “Toward a direct brain interface based on human subdural recordings and waveletpacket analysis,” *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 954-962, Jun. 2004.
- [46] D. Coyle, G. Prasad and T. McGinnity, “A time-frequency approach to feature extraction for a brain-computer interface with a comparative analysis of performance measures,” *EURASIP Journal on Applied Signal Processing*, vol. 2005, no. 19, pp. 3141-3151, 2005.
- [47] I. Clark, R. Biscay, M. Echeverria and T. Virues, “Multiresolution decomposition of non-stationary EEG signals: a preliminary study,” *Comput. Biol. Med.*, vol. 25, no. 4, pp. 373–382, Jul. 1995.
- [48] M. Besserve, K. Jerbi, F. Laurent, S. Baillet, J. Martinerie and L. Garnero, “Classification methods for ongoing EEG and MEG signals,” *Biol. Res.*, vol. 40, no. 4, pp. 415–437, 2008.
- [49] K. Muller, C. W. Anderson, and G. E. Birch, “Linear and non-linear methods for brain-computer interfaces,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 165–169, Jun. 2003.
- [50] S. Sardouie and M. Shamsollahi, “Selection of efficient features for discrimination of hand movements from MEG using a BCI competition IV Data Set,” *Front Neurosci.*, vol. 6, 2012.
- [51] A. Chan, E. Halgren, K. Marinkovic and S. Cash, “Decoding word and category-specific spatiotemporal representations from MEG and EEG,” *NeuroImage*, vol. 54, pp. 3028-3039,

2011.

- [52] C. Bishop, *Pattern recognition and machine learning*, Springer, 2006.
- [53] K. Fukunaga, *Introduction to statistical pattern recognition (2nd edition)*, Academic Press, 1990.
- [54] W. Wang, G. Sudre, Y. Xu, R. Kass, J. Collinger, A. Degenhart, A. Bagic and D. Weber, “Decoding and cortical source localization for intended movement direction with MEG,” *J. Neurophysiology*, Aug. 2010.
- [55] J. Gross, J. Kujala, M. Hamalainen, L. Timmermann, A. Schnitzler and R. Salmelin, “Dynamic imaging of coherent sources: Studying neural interactions in the human brain,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 2, pp. 694-699, 2001.
- [56] P. Bickel and E. Levina, “Some theory for Fisher’s linear discriminant function, ‘naive Bayes’, and some alternatives when there are many more variables than observations,” *Bernoulli*, vol. 10, no. 6, pp. 989-1010, 2004.
- [57] J. S. Lim and A. V. Oppenheim, *Advanced Topics in Signal Processing*, Prentice Hall, 1988.
- [58] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.
- [59] D. Luenberger, *Linear and Nonlinear Programming*, Addison-Wesley, 1984.
- [60] F. Chung, *Spectral Graph Theory*, American Mathematical Society, 1997.
- [61] A. Ng, M. Jordan, and Y. Weiss, “On spectral clustering: analysis and an algorithm,” In *Proc. Advances in Neural Information Processing Systems*, pp. 849-856, 2001.
- [62] F. Bach, M. Jordan, “Learning spectral clustering,” In *Proc. Advances in Neural Information Processing Systems*, pp. 305-312, 2004.
- [63] M. Newman and M. Girvan, “Finding and evaluating community structure in networks,” *Physical Review E*, vol. 69, no. 026113, 2004.
- [64] U. Luxburg, “A tutorial on spectral clustering,” *Statistics and Computing*, vol. 17, no. 4, pp. 395-416, 2007.
- [65] S. Taulu, J. Simola and M. Kajola, “Applications of the signal space separation method,” *IEEE*

Trans. Signal Process., vol. 53, no. 9, pp. 3359-3372, Sep. 2005.

- [66] P. Karjalainen, J. Kaipio, A. Koistinen and M. Vauhkonen, "Subspace regularization method for the single-trial estimation of evoked potentials," *IEEE Trans. Biomed. Eng.*, vol. 46, no. 7, pp. 849-860, Jul. 1999.
- [67] W. Truccolo, M. Ding, K. Knuth, R. Nakamura and S. Bressler, "Trial-to-trial variability of cortical evoked responses: implications for the analysis of functional connectivity," *Clinical Neurophysiology*, vol. 113, pp. 206-226, 2002.
- [68] J. Yao and J. Dewald, "Evaluation of different cortical source localization methods using simulated and experimental EEG data," *NeuroImage*, vol. 25, pp. 369-382, 2005.
- [69] J. Zhang, G. Sudre, X. Li, W. Wang, D. Weber and A. Bagic, "Task-related MEG source localization via discriminant analysis," in *Proc. 33rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pp. 2351-2354, 2011.
- [70] D. Percival and A. Walden, *Wavelet Methods for Time Series Analysis*, Cambridge University Press, 2006.
- [71] C. Galletti, D. Kutz, M. Gamberini, R. Breveglieri and P. Fattori, "Role of the medial parieto-occipital cortex in the control of reaching and grasping movements," *Exp. Brain Res.*, vol. 153, pp. 158-170, 2003.
- [72] S. Wise, D. Boussaoud, P. Johnson and R. Caminiti, "Premotor and parietal cortex: corticocortical connectivity and combinatorial computations," *Annu. Rev. Neurosci.*, vol. 20, pp. 25-42, 1997.
- [73] B. Okuda, H. Tanaka, Y. Tomino, K. Kawabata, H. Tachibana and M. Sugita, "The role of the left somatosensory cortex in human hand movement," *Exp. Brain Res.*, vol. 106, pp. 493-498, 1995.
- [74] J. Meier, T. Aflalo, S. Kastner and M. Graziano, "Complex organization of human primary motor cortex: a high-resolution fMRI study," *Journal of neurophysiology*, vol. 100, pp. 1800-1812, 2008.
- [75] C. Toxopeus, B. De Jong, G. Valsan, B. Conway, K. Leenders and N. Maurits, "Direction of

- movement is encoded in the human primary motor cortex,” *PLoS One*, vol. 6, no. 11, 2011.
- [76] G. Golub and C. Loan, *Matrix Computations*, Johns Hopkins Univ. Press, 1996.
- [77] A. Rodriguez-Rivera, B. Baryshnikov, B. Veen and R. Wakai, “MEG and EEG source localization in beamspace,” *IEEE Trans. Biomed. Eng.*, vol. 53, no. 3, pp. 430-441, Mar. 2006.
- [78] D. Bertsekas, *Nonlinear Programming*, Athena Scientific, 1999.
- [79] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [80] M. Grant and S. Boyd, *CVX: Matlab Software for Disciplined Convex Programming*, <http://cvxr.com/cvx>, April 2011.
- [81] J. Phillips, R. Leahy and J. Mosher, “MEG-based imaging of focal neuronal current sources,” *IEEE Transactions on Medical Imaging*, vol. 16, no. 3, Jun. 1997.
- [82] R. Leahy, J. Mosher, M. Spencer, M. Huang and J. Lewine, “A study of dipole localization accuracy for MEG and EEG using a human skull phantom,” *Electroencephalography and Clinical Neurophysiology*, vol. 107, no. 2, pp. 159-173, Apr. 1998.
- [83] J. Mosher, S. Baillet and R. Leahy, “Equivalence of linear approaches in bioelectromagnetic inverse solutions,” In *Proceedings of 2003 IEEE Workshop on Statistical Signal Processing*, pp. 294-297, 2003.
- [84] A. Dale, B. Fischl and M. Sereno, “Cortical surface-based analysis: I. Segmentation and surface reconstruction,” *Neuroimage*, vol. 9, no. 2, pp. 179-194, Feb. 1999.
- [85] R. Desikan, F. Segonne, B. Fischl, B. Quinn, B. Dickerson, D. Blacker, R. Buckner, A. Dale, R. Maguire, B. Hyman, M. Albert and R. Killiany, “An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest,” *Neuroimage*, vol. 31, no. 3, pp. 968-980, Jul. 2006.
- [86] S. Ahlfors, J. Han, J. Belliveau and M. Hamalainen, “Sensitivity of MEG and EEG to source orientation,” *Brain Topogr*, vol. 23, pp. 227-232, 2010.
- [87] W. Wang, J. Collinger, M. Perez, E. Tyler-Kabara, L. Cohen, N. Birbaumer, S. Brose, A. Schwartz, M. Boninger and D. Weber, “Neural interface technology for rehabilitation: exploiting and promoting neuroplasticity,” *Phys Med Rehabil Clin N Am.*, vol. 21, no. 1, pp.

157-178, Feb. 2010.

- [88] G. Prasad, P. Herman, D. Coyle, S. McDonough and J. Crosbie, "Applying a brain-computer interface to support motor imagery practice in people with stroke for upper limb recovery: a feasibility study," *Journal of Neuroengineering and Rehabilitation*, vol. 7, pp. 60-78, 2010.
- [89] S. Foldes, W. Wang, J. Collinger, X. Li, J. Zhang, G. Sudre, A. Bagic & D. Weber, "Accessing and Processing MEG Signals in Real-Time: Emerging Applications and Enabling Technologies," In *E. W. Pang (Ed.), Magnetoencephalography*, InTech, 2011.
- [90] M. Grosse-Wentrup, D. Mattia and K. Oweiss, "Using brain-computer interfaces to induce neural plasticity and restore function," *J. Neural Eng.*, vol. 8, no. 2, pp.1-5, 2011.
- [91] N. Sharma, V.M. Pomeroy and J.C. Baron, "Motor imagery a backdoor to the motor system after stroke?" *Stroke*, vol. 37, pp. 1941-1952, 2006.
- [92] R. Dickstein, J.E. Deutsch, "Motor imagery in physical therapist practice," *Phys. Ther.*, vol. 87, pp. 942-953, 2007.
- [93] E. Patariaia, C. Baumgartner, G. Lindinger and L. Deecke, "Magnetoencephalography in presurgical epilepsy evaluation," *Neurosurg Rev*, vol. 25, pp. 141-159, 2002.
- [94] H. Yuan, A. Doud, A. Gururajan and B. He, "Cortical imaging of event-related (de)synchronization during online control of brain-computer interface using minimum-norm estimates in frequency domain," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 16, no. 5, pp.425-431, 2008.
- [95] M. Ahn, J. Hong and S. Jun, "Source space based brain computer interface," In *Proceedings of the 17th International Conference on Biomagnetism Advances in Biomagnetism (Biomag 2010)*, vol. 28, pp.366-369, 2010.
- [96] J. Gross and A. Ioannides, "Linear transformations of data in space in MEG," *Phys. Med. Biol.*, vol. 44, no. 8, pp. 2081-2097, Aug. 1999.
- [97] A. Beck and M. Teboulle, "A convex optimization approach for minimizing the ratio of indefinite quadratic functions over an ellipsoid," *Math. Program*, vol. 118, no.1, pp. 13-35, 2009.

- [98] E. Capitani, M. Laiacona, B. Mahon and A. Caramazza, "What are the facts of semantic category-specific deficits? A critical review of the clinical evidence," *Cognitive Neuropsychology*, vol. 20, pp. 213-261, 2003.
- [99] T.M. Mitchell, S.V. Shinkareva, A. Carlson, K.M. Chang, V.L. Malave, R.A. Mason and M.A. Just, "Predicting human brain activity associated with the meanings of nouns," *Science*, vol. 320, pp. 1191-1195, 2008.
- [100] S.V. Shinkareva, R.A. Mason, V.L. Malave, W. Wang, T.M. Mitchell, M.A. Just, "Using fMRI brain activation to identify cognitive states associated with perception of tools and dwellings," *PLoS One*, e1394, 2008.
- [101] A. Martin, "The representation of object concepts in the brain," *Annu. Rev. Psychol.*, vol. 58, pp. 25-45, 2007.
- [102] G. Sudre, D. Pomerleau, M. Palatucci, L. Wehbe, A. Fyshe, R. Salmelin and T. Mitchell, "Tracking Neural Coding of Perceptual and Semantic Features of Concrete Nouns," *NeuroImage* (submitted).
- [103] A.M. Chan, E. Halgren, K. Marinkovic and S.S. Cash, "Decoding word and category-specific spatiotemporal representations from MEG and EEG," *NeuroImage*, vol. 54, pp. 3028-3039, 2011.
- [104] L. Rabiner and B. Juang, *Fundamentals of speech recognition*, Prentice-Hall, Inc., 1993.
- [105] F. Zhou and F. Torre, "Canonical time warping for alignment of human behavior," *Proc. Neural Information Processing Systems*, 2009.
- [106] A. Gramfort, R. Keriven and M. Clerc, "Graph-based variability estimation in single-trial event-related neural responses," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 5, pp. 1051-1061, 2010.
- [107] A. Hillebrand and G. Barnes, "A quantitative assessment of the sensitivity of whole-head MEG to activity in the adult human cortex," *NeuroImage*, vol. 16, pp. 638-650, 2002.
- [108] F. Lin, T. Witzel, S. Ahlfors, S. Stufflebeam, J. Belliveau and M. Hamalainen, "Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm

- estimates,” *NeuroImage*, vol. 31, pp. 160-171, 2006.
- [109] J. Crosbie, S. McDonough, D. Gilmore, M. Wiggam, “The adjunctive role of mental practice in the rehabilitation of the upper limb after hemiplegic stroke: a pilot study,” *Clinical Rehabilitation*, vol. 18, pp. 60-68, 2004.
- [110] S. Braun, A. Beurskens, P. Borm, T. Schack, D. Wade, “The effects of mental practice in stroke rehabilitation: a systematic review,” *Arch Phys Med Rehabil.*, vol. 87, pp. 842-852, 2006.
- [111] S. Braun, M. Kleynen, J. Schols, T. Schack, A. Beurskens, D. Wade, “Using mental practice in stroke rehabilitation: a framework,” *Clin Rehabil.*, vol. 22, pp. 579-591, 2008.
- [112] S. Page, P. Levine, A. Leonard, “Mental Practice in Chronic Stroke: Results of a Randomised, Placebo-Controlled Trial,” *Stroke*, vol. 38, no.4, pp. 1293-1297, 2007.
- [113] R. Dickstein, J. Deutsch, “Motor imagery in physical therapist practice,” *Phys Ther*, vol. 87, no. 7, pp. 942-953, 2007.