A spatio-temporal solution for the EEG/MEG inverse problem using group penalization methods

Tian Siva Tian[*] and Zhimin Li

The inverse problem encountered in electroencephalography (EEG) and magnetoencephalography (MEG) studies refers to estimating neural activity given limited scalp-recorded data. We propose a spatio-temporal solution using group penalization approaches. This proposed method is based on the assumption that the underlying sources of EEG/MEG measurements are smooth in the temporal domain, and focal in the spatial domain. It transforms the spatio-temporal problem to a high-dimensional linear regression problem with grouped predictors using a basis expansion. Then an iterative group elastic net algorithm is utilized to localize and estimate the source time courses. The proposed approach is shown to be effective on simulations and human MEG studies.

Keywords and phrases: Inverse problem, Spatio-temporal data, Group elastic net, EEG/MEG.

1. INTRODUCTION

Over the past few decades, scalp electroencephalography (EEG) and magnetoencephalography (MEG) have been widely used in detecting spontaneous or evoked-response brain activities in normal or clinical populations due to their non-invasive nature. When a subject is receiving a stimulus, neurons in the functional region of the brain, such as the somatosensory cortex or visual cortex, are activated and then generate small currents. These small current flows produce a small magnetic and a small electric field that are mutually orthogonal. EEG and MEG measure the voltage and the magnetic fields, respectively, using a set of sensors at various locations on or near the scalp of the subject.

One of the most important uses for EEG/MEG is to determine the locations of these functional regions and how they respond to stimuli. Other important uses include detecting and localizing epileptiform activities, distinguishing different neural activity patterns between different groups of patients, determining the malfunctioning brain regions for patients with mental diseases and so on. The accomplishment of these goals requires creating a sequence of accurate brain activity maps over time. That is, one wishes to find out the *source* to the electric/magnetic fields. This involves

[*]Corresponding author.

deriving the electric current density (Sarvas, 1987), and this problem is called the "inverse problem".

The EEG/MEG inverse problem has no unique solution making it by definition "ill-posed" (Baillet et al., 2001). This is because there is an infinite number of electric current distributions which can lead to the same EEG/MEG measurements (Nunez, 1981; Sarvas, 1987). In addition, the current EEG/MEG devices have only a few hundred sensors, but the number of sources is much larger because each neuron can be treated as a potential source. The number of sensors is generally too insufficient to provide a precise reconstruction of the source current (Grave de Peralta Menendez and Gonzalez Andino, 1998).

In order to make the problem manageable and improve computational efficiency, many methods discretize the entire distribution, and generate a fine grid that covers the region of interest in the brain. Then mutually orthogonal electric dipoles with unknown amplitudes are attached at each grid point. Existing methods designed for solving this problem can be mainly grouped into three categories: the dipole fitting methods, the scanning methods and the imaging methods. Dipole fitting and scanning methods assume that the sources reside in only a few activated grid points, and then they attempt to decide among the possible dipole locations where the most appropriate location is for the sources so as to best describe the measured time courses. One challenge of these methods is that it is usually difficult to choose the number of sources (Bolstad et al., 2009). Examples of the dipole fitting methods include the equivalent current dipole methods (Scherg and von Cramon, 1986; Hämäläinen et al., 1993; Yamazaki et al., 2000) and some Bayesian methods (Jun et al., 2005; Sorrentino et al., 2009). Examples of the scanning methods include the multiple signal classification (MUSIC) method (Mosher et al., 1992), the beamforming method (VanVeen et al., 1997), maximum likelihood estimation (Dogandžić and Nehorai, 2000), etc.

The imaging methods are also referred to as the nonparametric methods. These methods attempt to estimate the dipole sources all together based on the assumption that the primary sources can be represented as linear combinations of neuron activities (Barlow, 1994). That is, the inverse problem can be expressed using the following nonparametric linear model:

$$(1) \qquad \mathbf{y}(t) = \mathbf{X}\boldsymbol{\beta}(t) + \mathbf{e}(t),$$

where $\mathbf{y}(t) = [y_1(t), y_2(t), \dots, y_n(t)]^T$ is a set of EEG/MEG time courses measured by n sensors. The forward operator $\mathbf{X}$ is an n-by-p matrix representing the propagation of the voltage/magnetic field. It contains information about positions and orientations of dipoles, and how they are represented at the sensor level. Most commonly, $\mathbf{X}$ is computed by a simple spherical head model (Mosher et al., 1999), so it is a "known" design matrix in Equation (1). The $\boldsymbol{\beta}(t) = [\beta_1(t), \beta_2(t), \dots, \beta_p(t)]^T$ is a set of time courses representing dipole activity. As mentioned earlier, the cortical area is divided into a fine grid of small cells, say $p/3$ cells. Each cell is assigned a current dipole with unknown amplitudes. The dipolar moment orientation is decomposed to x, y, and z directions in a 3-D coordinate system. Therefore, the total dimension of $\mathbf{X}$ is p, and each dimension (column) represents the lead field for one dipole at one direction. The $\mathbf{e}(t) = [e_1(t), e_2(t), \dots, e_n(t)]^T$, is a set of additive noise time courses. The goal is to estimate $\boldsymbol{\beta}(t)$ given $\mathbf{y}(t)$ and $\mathbf{X}$.

Typically p can be several thousands, whereas n is at most a few hundred, i.e., $p \gg n$. Furthermore, $\boldsymbol{\beta}(t)$ is assumed to have two characteristics: spatial focality and temporal smoothness. Spatial focality means that there should be only a small number of compact regions in the entire cortex staying active at one time point. That is, among p $\beta_j(t)$'s only a small portion related to the stimuli has significant amplitudes, while the rest associated with inactive regions should be normal static brain waves or noise perturbations. Temporal smoothness means that in the source regions the $\beta_j(t)$'s should be relatively smooth over time even though they can change rapidly.

Regularization is a commonly used technique for the problem at hand, and the solution can be expressed as:

$$(2) \qquad \arg\min_{\boldsymbol{\beta}} \left\{ \|\mathbf{y}(t) - \mathbf{X}\boldsymbol{\beta}(t)\|_2^2 + \lambda PEN(\boldsymbol{\beta}(t)) \right\},$$

where $PEN(\cdot)$ is a penalty function and $\lambda \geq 0$ is a fixed penalty parameter that controls the degree of regularization.

Earlier methods solve Equation (2) separately at each time point. That is, suppose there are s measured time points. To solve Equation (2), one has to solve s optimization problems individually. One of the most well-known methods is the minimum norm estimate (MNE) (Hämäläinen and Ilmoniemi, 1994), which imposes the L_2-penalty on $\boldsymbol{\beta}$, i.e., $PEN(\boldsymbol{\beta}) = \|\boldsymbol{\beta}\|_2^2$, where $\boldsymbol{\beta}$ is $\boldsymbol{\beta}(t)$ in the matrix form. Some modifications include the weighted minimum norm estimate (WMNE) (Dale and Sereno, 1993; Iwaki and Ueno, 1998) and the low resolution electrical tomography (LORETA) (Pascual-Marqui et al., 1994). Both methods utilizes a weighted L_2-penalty, $PEN(\boldsymbol{\beta}) = \|\mathbf{W}\boldsymbol{\beta}\|_2^2$, instead of the regular L_2-penalty, where $\mathbf{W}$ is a p-by-p weighting matrix. One advantage of L_2 methods is that they can be very efficient due to the computational efficiency of the L_2-norm minimization. However, applying L_2-norm to the entire $\boldsymbol{\beta}$ space produces a well-know blurring effect on the spatial domain of the reconstructed sources.

One alternative is the L_1-penalty, i.e., $PEN(\boldsymbol{\beta}) = \|\boldsymbol{\beta}\|_1$ [e.g., the minimum current estimate (MCE) (Matsuura and Okabe, 1995; Uutela et al., 1999; Lin et al., 2006)]. The FOCal Underdetermined System Solver (FOCUSS) (Gorodnitsky and Rao, 1997) is a recursive procedure of WMNE with an adjustable weighting matrix. Rao and Kreutz-Delgado (1999) has shown that FOCUSS is equivalent to an L_l-norm solver with $l \leq 1$. However, it is claimed that FOCUSS is very sensitive to noise (Ou et al., 2009). In general, L_1 methods can produce better solutions in terms of identifying sources due to the nature of the L_1-penalty. However, they also introduce substantial discontinuities to the temporal domain. Hence, the "spiky-looking" time courses will be observed.

A great deal of research has been devoted to prevent the emergence of spiky L_1 solutions and diffuse L_2 solutions. Remedies include many hierarchical Bayesian approaches [see, for example, (Baillet and Garnero, 1997; Sato et al., 2004; Nummenmaa et al., 2007)]. Some of them considered the spatial focality and temporal smoothness by employing sparse priors and/or smoothing priors (Friston et al., 2008; Daunizeau et al., 2006; Nummenmaa et al., 2007). Various L_h-norm methods with $0 < h < 1$ and $1 < h < 2$ were also discussed (Auranen et al., 2005; Jeffs et al., 1987). Ou et al. (2009) proposed an $L_1 L_2$-regularization, which imposed the L_1-penalty to the spatial domain and L_2-penalty to the temporal domain. Bolstad et al. (2009) proposed an event sparse penalty procedure, which represented $\boldsymbol{\beta}(t)$ using both spatial and temporal bases. These methods can solve the "spiky-looking" problem in the temporal domain. However, they may produce scattering estimated sources, which may deliver implausible results physiologically (Haufe et al., 2011). That is, sources are tightly grouped together and do not smoothly transit from one cell to another.

In this paper, we propose a Spatio-Temporal Penalization (STP) approach that takes into account both smoothness and focality assumptions in the temporal and spatial domains, respectively. The proposed method can potentially produce a smooth transition in the spatial domain, and hence, reduce the scattering effect. The temporal smoothness is guaranteed by the basis expansion of the source time courses. Then the temporal bases are combined with the forward operator $\mathbf{X}$ containing the spatial information. This combined new forward operator then contains both spatial and temporal information of the sources. The spatio-temporal problem is then transformed to a high-dimensional linear regression problem with grouped predictors. An L_1 and an L_2 penalties are placed on the grouped coefficients through an iterative group elastic net (IGEN) approach. The employment of the elastic net penalties ensures a moderate degree of the spatial focality. The application of basis representation guarantees the temporal smoothness, and the elastic net penalties ensure a moderate degree of the focality of the spatial domain.

The rest of the paper is organized as follows. We present the general scheme for the STP approach and its implementation details in Sections 2 and 3. We describe two possible basis functions for STP in Section 4. We demonstrate the effectiveness of STP on two simulated examples and a human MEG example in Sections 5 and 6. Some tuning parameters are also discussed in Section 5. Section 7 discusses potential issues for future investigation.

2. METHODOLOGY

Consider Equation (1) and we assume that $\mathbf{e}(t)$ is independent white noise over time. Time-dependent noise models are considered in Huizenga et al. (2002); Bijma et al. (2005), but we will leave it for future investigation. The $\boldsymbol{\beta}(t)$ can be projected onto a subspace that is spanned by its temporal orthogonal basis functions. Assuming that most energy lies in the first q basis functions, we write Equation (1) as follows:

$$(3) \qquad y_i(t) = \sum_{j=1}^{p} x_{ij}\beta_j(t) + e_i(t)$$
$$= \sum_{j=1}^{p} x_{ij}b(t)^T\eta_j + e_i(t), \quad i = 1, \ldots, n,$$

where $b(t) = [b_1(t), b_2(t), \ldots, b_q(t)]^T$ is the q-dimensional basis of $\beta_j(t)$ and η_j is its corresponding coefficient vector. Let $x^*_{ijt} = x_{ij}b(t)$. Note that $\mathbf{X}$ contains the spatial information of sources, so x^*_{ijt} in turn contains both spatial and temporal information. Then Equation (3) becomes:

$$(4)$$
$$y_i(t) = \sum_{j=1}^{p} x^{*T}_{ijt}\eta_j + e_i(t) = X^*_{it}\boldsymbol{\eta} + e_i(t), \quad i = 1, 2, \ldots, n,$$

where $X^*_{it} = [x^{*T}_{i1t}, \ldots, x^{*T}_{ipt}] \in \mathbb{R}^{pq}$ and $\boldsymbol{\eta} = [\eta_1^T, \eta_2^T, \ldots, \eta_p^T]^T \in \mathbb{R}^{pg}$. We express Equation (4) in the matrix form:

$$(5) \qquad \mathbf{y}^* = \mathbf{X}^*\boldsymbol{\eta} + \mathbf{e}^*,$$

where $\mathbf{y}^* = vec[\mathbf{y}(t)] = [\mathbf{y}_1^T, \ldots, \mathbf{y}_n^T]^T$ is an ns-vector of the new response variable. Here $\mathbf{y}_i$, $i = 1, \ldots, n$, is the ith EEG/MEG time course measured over s time points. The $\mathbf{e}^* = vec[\mathbf{e}(t)] = [\mathbf{e}_1^T, \mathbf{e}_2^T, \ldots, \mathbf{e}_n^T]^T$ is an ns-vector of the new white noise variable, where $\mathbf{e}_i$, $i = 1, \ldots, n$, is the ith discretized noise time course. We have $\mathbf{X}^* = \mathbf{X} \otimes \mathbf{B}$, i.e.,

$$\mathbf{X}^* = \begin{bmatrix} x_{11}\mathbf{B} & \cdots & x_{1p}\mathbf{B} \\ \vdots & \vdots & \vdots \\ x_{n1}\mathbf{B} & \cdots & x_{np}\mathbf{B} \end{bmatrix} = \begin{bmatrix} x^{*T}_{11t_1} & \cdots & x^{*T}_{1pt_1} \\ \vdots & \vdots & \vdots \\ x^{*T}_{11t_s} & \cdots & x^{*T}_{1pt_s} \\ x^{*T}_{21t_1} & \cdots & x^{*T}_{2pt_1} \\ \vdots & \vdots & \vdots \\ x^{*T}_{n1t_s} & \cdots & x^{*T}_{npt_s} \end{bmatrix},$$

where $\mathbf{B} \in \mathbb{R}^{s \times q}$ is $\boldsymbol{\beta}(t)$'s basis functions in the matrix form, i.e., the (k, r)th element of $\mathbf{B}$ is $b_r(t_k)$, and more explicitly, each block element, $x^*_{ijt_k}$ (e.g., $x^*_{11t_1} = [x_{11}b_1(t_1), x_{11}b_2(t_1), \ldots, x_{11}b_q(t_1)]$), is a q-vector. This way, the original problem has been transformed to a high-dimensional linear regression problem, where the new design matrix $\mathbf{X}^*$ is ns-by-pq. This $\mathbf{X}^*$ consists of p groups, and each group contains q group members. The temporal information for $\beta_j(t)$ resides within each group, and the original spatial information still remain between groups. Hence, $\mathbf{X}^*$ contains both spatial and temporal information.

We write $\mathbf{X}^* = [\mathbf{x}_1^*, \mathbf{x}_2^*, \ldots, \mathbf{x}_p^*]$, where each column vector is a block matrix, i.e., $\mathbf{x}_j^* = [x_{1j}\mathbf{B}^T, x_{2j}\mathbf{B}^T, \ldots, x_{nj}\mathbf{B}^T]^T$. Since $\mathbf{x}_j^*$ contains temporal information of the jth potential source, it is desired to treat predictors in each $\mathbf{x}_j^*$ as a group in the estimation procedure. We are only interested in estimating the functional forms of active $\beta_j(t)$'s. Hence, those inactive $\beta_j(t)$'s can be set to zero. That is, elements in a coefficient vector, η_j, $j = 1, \ldots, p$, should be either all zero or mostly nonzero. Put another way, columns in $\mathbf{x}_j^*$ should be either included or excluded as a group in the model. Therefore, some group penalization methods may be considered.

One solution is to use a group Lasso technique, which can shrink inactive groups to be exact zero. However, sources in active areas are highly correlated to each other, the group Lasso may tend to select only one of them and does not care which one is selected (Zou and Hastie, 2005). This is equivalent to impose an L_1-penalty to the spatial domain, and hence, may result in the well-known "scattering" sources. Instead, we propose to use the following penalized regression model:

$$(6)$$
$$\hat{\boldsymbol{\eta}} = \arg\min_{\boldsymbol{\eta}} \left\{ \|\mathbf{y}^* - \mathbf{X}^*\boldsymbol{\eta}\|_2^2 + \lambda_1 \sum_{j=1}^{p} \|\eta_j\|_1 + \lambda_2 \sum_{j=1}^{p} \|\eta_j\|_2^2 \right\},$$

where λ_1 and λ_2 are penalty parameters controlling the degrees of regularization.

Equation (6) is a group version of the elastic net problem (Zou and Hastie, 2005). It can be written as solving a group Lasso problem (Yuan and Lin, 2006):

$$(7) \qquad \hat{\boldsymbol{\eta}} = \arg\min_{\boldsymbol{\eta}} \left\{ \|\mathbb{Y}^* - \mathbb{X}^*\boldsymbol{\eta}\|_2^2 + \gamma \sum_{j=1}^{p} \|\eta_j\|_1 \right\},$$

where $\mathbb{Y}^* = \begin{pmatrix} \mathbf{y}^* \\ \mathbf{0}_p \end{pmatrix}$, $\mathbb{X}^* = \frac{1}{\sqrt{1+\lambda_2}}\begin{pmatrix} \mathbf{X}^* \\ \sqrt{\lambda_2}\mathbf{I} \end{pmatrix}$, and $\gamma = \frac{\lambda_1}{\sqrt{1+\lambda_2}}$.

The choice of λ_1 and λ_2 can be based on a cross-validation (CV) on a 2-dimensional space (Zou and Hastie, 2005). For example, we first specify a set of candidate values to λ_2, e.g., $\lambda_2 \in (0, 0.01, 0.1, 1, 10)$. For each value of λ_2, run Equation (7) and obtain an entire path of λ_1. Then one can select λ_1 by a K-fold CV and choose the λ_2 that gives the smallest CV error.

3. ITERATIVE GROUP ELASTIC NET

There is a potential practical problem with the implementation of the group elastic net. After basis expansion, $\mathbf{X}^*$ may become ultra-large. For example, we have $n = 247$, $p = 15,372$, $s = 228$, and $q = 7$ in our MEG example in Section 6. Then $\mathbf{X}^*$ is a 56,316-by-107,604 matrix and $\mathbb{X}^*$ is 163,920-by-107,604. Equation (7) is hard to compute and problems such as memory allocation will arise.

Therefore, we propose an ad-hoc iterative group elastic net (IGEN) algorithm that is more satisfactory in our settings. The idea is to resample subareas from the entire cortex and apply a group elastic net to each sample to identify active sources in that sample. Once many samples are examined, a candidate pool of possible active sources can be formed by simply putting together all historically selected sources, applying a majority vote to this candidate pool, and selecting a small portion of the sources in the pool. This algorithm is similar to an ensemble method, which uses multiple models to obtain better predictive performance than from any of the constituent models (Opitz and Maclin, 1999), and can be more robust to noise.

The following steps describe the STP approach. First, IGEN randomly selects a subset of variables from $\mathbf{X}$ (i.e., columns of $\mathbf{X}$) with equal probabilities. Intuitively, this step is to divide the cortical area into subareas, which can have overlap with each other. Other sampling techniques may also be considered, such as even-spaced sampling, block sampling (i.e., dividing the cortex into small blocks and sampling one or a few blocks each time), and weighted sampling (i.e., assigning higher probabilities to regions of interest), but will be left to our future investigation. Second, we expand the selected subset using the method described in Section 2. Third, treating temporal coefficients for one $\beta_j(t)$ as a group, a group elastic net (GEN) is applied to identify a small number of potential active groups. Repeat these three steps a sufficiently large number of times. Ideally, one wishes that the number of iterations could be large enough to guarantee all columns in $\mathbf{X}$ are examined multiple times. In practice, some moderately large number can be chosen to obtain a good result. We explain the choice of this number in Section 5.3. Fourth, a candidate pool is formed by putting together selected groups from all previous iterations, and a majority vote is applied to this candidate pool to further narrow down the "important groups". That is, count the number of times a group is selected, put these counted numbers for all groups in descending order, and choose the first several groups with the largest number of time being selected in previous iterations. These groups are considered as more "important" or "active". The idea of using a majority voting rule is based on the assumption that some real active dipoles can always be identified even in different sets of dipole samples, but some less active ones may be selected only when the more active ones are not in the sample. A majority vote can prevent selecting inactive dipoles. Then we apply GEN again to the groups resulting from the majority vote and obtain the estimated coefficient vectors. Finally, we transform everything back to the original spatio-temporal space to obtain the estimated source matrix, $\hat{\boldsymbol{\beta}}$.

Denote the maximum sub-sample size for one subset of $\mathbf{X}$ as p_{max}. That is, each subset can only contain less than or equal to p_{max} variables of $\mathbf{X}$. The ad-hoc IGEN algorithm can be described as follows:

1. Initialize the number of iterations L and the maximum sub-sample size p_{max}.
2. For $l = 1$ to L
   (a) Draw a random sample of size p' from $\mathbf{X}$, where $p' \leq p_{max}$.
   (b) Expand the subset of data by Equation (4) or (5).
   (c) Fit GEN (6) to the expanded data from Step 2(b) and select $m^{(l)}$ groups.
3. Apply a majority vote to the L sets of selected groups and obtain the m most voted groups.
4. Apply GEN to these m most voted groups from Step 3 to select m_0 groups and obtain their estimated coefficient vector, $\hat{\eta}_j$'s, $j = 1, \ldots, m_0$.
5. Calculate the estimated source time courses, $\hat{\boldsymbol{\beta}} = \mathbf{B}\hat{\boldsymbol{\eta}}^T$, where $\hat{\boldsymbol{\eta}}^T = [\hat{\eta}_1^T, \ldots, \hat{\eta}_{m_0}^T]^T$.

There are several tuning parameters in this algorithm. First, the ideal number of iterations L should be sufficiently large so that all columns of $\mathbf{X}$ can be examined several times with different sample combinations, especially when p_{max} is relatively small. This requires $Lp_{max} \gg p$. However, we demonstrate that in practice the predictive accuracy is not affected much as long as L is roughly more than twice as large as p/p_{max} in Section 5.3. Second, the maximum size of a sub-sample, p_{max}, can depend on the computational resource. In Section 5.3 we empirically show that the prediction accuracy is not very sensitive to the choice of p_{max} either, but if p_{max} is too large the computational cost increases. We also illustrate the relationship between L and p_{max} relative to the predictive accuracy. Third, the number of groups selected by the majority voting rule can be as large as the size of the union of all selected groups, i.e., $\bigcup_{l=1}^L m^{(l)}$. But in our study, we use $m = 10\%p$ based on the assumption that only a small number of locations in the cortex are responsible for the measured signals. Fourth, $m^{(l)}$ and m_0 need to be justified. In Step 2(c) $m^{(l)}$ varies from iteration to iteration and it is automatically chosen by GEN based on minimizing the 10-fold cross-validated error rate on the training data. In Step 4 m_0 is also decided in a similar manner. Presumably, m_0 can be as large as m.

The next problem is that whether or not IGEN performs as well as the standard GEN, if not better. We compare the performance of IGEN to GEN using a simulated small data set in Section 5.

4. BASIS SELECTION

STP requires an approximation of the signal space by its temporal basis. One problem is to choose an appropriate basis. Here we examine one data-dependent basis and one data-independent basis.

The data-dependent basis is the right orthogonal matrix of the singular-value decomposition (SVD) of the EEG/MEG data. Let $\mathbf{Y}$ be the measured EEG/MEG data in the matrix form. That is, each row of $\mathbf{Y}$ represents a time course for all time instants. Then the SVD of $\mathbf{Y}$ can be expressed as

$$\mathbf{Y} = \mathbf{U}\mathbf{D}\mathbf{V}^T,$$

where the orthogonal matrix $\mathbf{U}$ can be viewed as the spatial pattern of the EEG/MEG measurements, while the orthogonal matrix $\mathbf{V}$ can be viewed as the temporal pattern. The diagonal matrix $\mathbf{D}$ contains the singular values in a descending order. Since $\mathbf{y}(t)$ represents the measured time courses evoked by the signal sources, it is natural to assume that $\boldsymbol{\beta}(t)$ and $\mathbf{y}(t)$ share the same temporal bases. In particular, we approximate $\beta_j(t)$'s using the first q bases from $\mathbf{V}$, which corresponds to the first q largest singular values. These q bases are considered to have most of the information in active signals.

There are many options for the data-independent basis. Some commonly used ones include wavelet and Fourier bases. We demonstrate the natural cubic splines (NCS) in this paper.

One issue is to determine the value of q. We use a fixed number of $q = 9$ for both bases in our simulations and MEG study. Our variance analysis shows that $q = 9$ is adequate in both simulation and MEG studies in this paper, because the first nine eigenvectors contain about 80% variance of $\mathbf{Y}$. Note that the specification of q can be different for different subareas in practice, since q controls the flexibility of the functions. Presumably a larger q may be assigned to subareas with higher source activity. Therefore, a two-step approach may be considered. First, assign a fixed q to the entire cortex, i.e., all p $\beta_j(t)$'s, and then apply STP to obtain a rough estimate of the active subareas. Second, reassign a larger q to the active subareas and smaller q to inactive subareas, and conduct STP again to obtain a more accurate estimation. In particular, when prior knowledge is available, one can be more confident to specify a larger q to the potential active regions. Intuitively, q can be larger in studies with high frequency stimuli than that in studies with low frequency stimuli. To reduce computational cost, we suggest that the reassigned q for the active subareas only needs to be slightly larger than the initial q and the reassigned q for the inactive subareas can be as low as 3.

5. SIMULATION STUDIES

In this section we demonstrate the relative performance of STP with the aforementioned two bases to two commonly used EEG/MEG inverse solvers using two simulated examples. In both simulations, we create two scenarios with different SNR levels. One is a lower noise level scenario (Scenario 1) and the other is a higher noise level scenario (Scenario 2). The two counterparts are the L_2-norm based MNE method (Dale and Sereno, 1993) and the L_1-norm based MCE method (Lin et al., 2006). In order to demonstrate the effectiveness of IGEN, we also implement the standard GEN algorithm to STP (STP-GEN) in Simulation I.
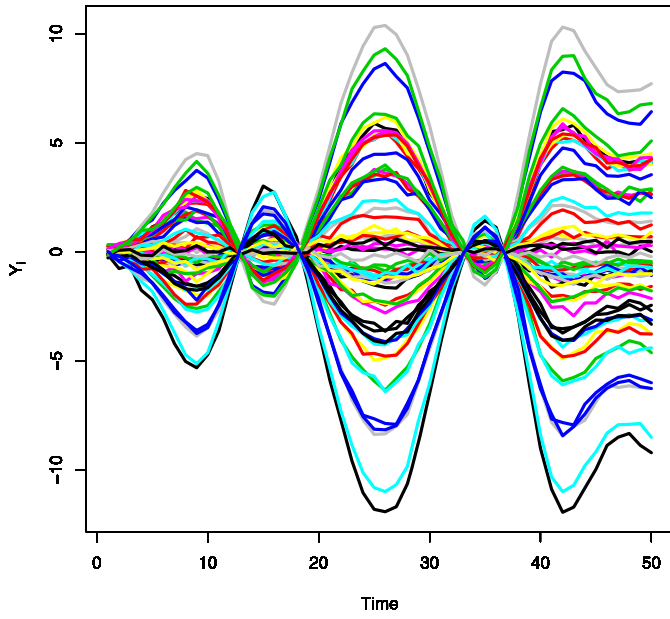
Note that in both the synthetic and the MEG examples (i.e., examples in this section and Section 6), we use $\hat{\beta}_j(t)$ and a subscript to represent the estimated $\beta_j(t)$ by different methods. For example, $\hat{\beta}_{j,NCS}(t)$ represents the estimated $\beta_j(t)$ using STP with a NCS basis and $\hat{\beta}_{j,SVD}(t)$ represents the estimated $\beta_j(t)$ using STP with the basis from SVD of $\mathbf{Y}$. Similarly, $\hat{\beta}_{j,MNE}(t)$ and $\hat{\beta}_{j,MCE}(t)$ represent the estimated $\beta_j(t)$'s using MNE and MCE, respectively.
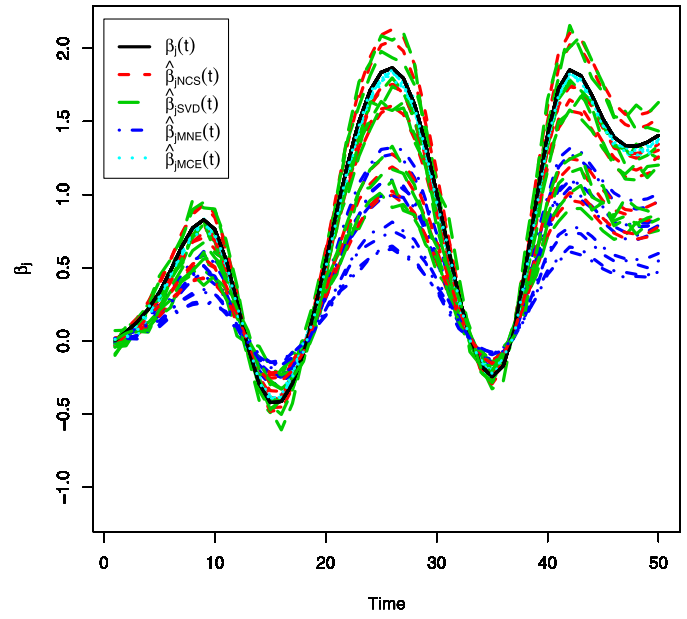
5.1 Simulation I

In this section we simulate a relatively small data set. We let $n = 60$, $s = 50$, $p = 100$. That is the forward operator $\mathbf{X}$ is a 50-by-100 matrix, and each element is generated from a standard normal distribution. Among the 100 functions, the first 10 are signals and have exactly the same form. They are generated from an 11-dimensional NCS basis. Elements in the coefficient vector are generated from $\mathcal{N}(0,1)$. The remaining 90 functions are noise and are generated randomly from $\mathcal{N}(0,0.01)$. We create two scenarios: a low noise level scenario (Scenario 1) and a high noise level scenario (Scenario 2). In Scenarios 1 and 2, elements from the white noise, $e_i(t)$, are generated independently from $\mathcal{N}(0,0.1)$ and $\mathcal{N}(0,1)$, respectively. The responses, $y_i(t)$'s, are obtained by Equation (5), and they are plotted in Figure 1 (a) and (c), respectively.

Using an NCS basis to approximate $\beta_j(t)$'s matches the data generation mechanism. Therefore, STP with NCS basis should outperform STP with SVD basis. Figure 1 (b) and (d) show the true and estimated signal time courses by STP, MNE and MCE in the two scenarios. $\hat{\beta}_{j,NCS}(t)$ and $\hat{\beta}_{j,SVD}(t)$ estimated $\beta_j(t)$ considerably well in Scenario 1. The shapes of both $\hat{\beta}_{j,NCS}(t)$'s and $\hat{\beta}_{j,SVD}(t)$'s followed the shape of the signal $\beta_j(t)$. In particular, STP with NCS basis produced smoother curves. MCE also produced good estimations of $\beta_j(t)$'s. But MNE tended to underestimate the signals. The shape of $\hat{\beta}_{j,MNE}(t)$ followed the signal $\beta_j(t)$ but the magnitude was much smaller. STP performed considerably well in Scenario 2 too, while MNE and MCE were unsatisfactory due to the underestimation of the signal $\beta_j(t)$.
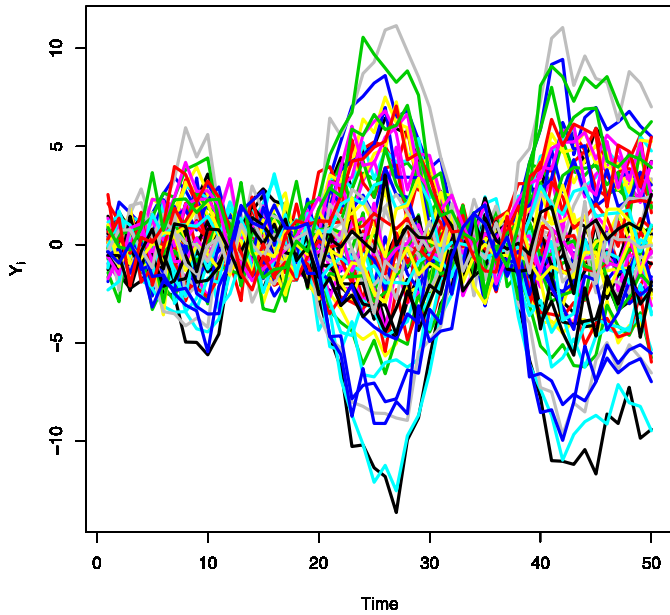
For evaluation purposes, we use two measures to measure the performance of the inverse solvers:
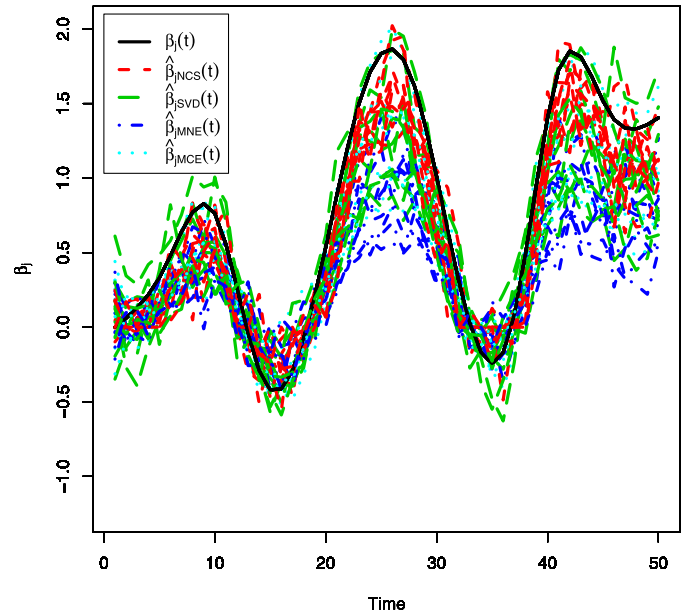
(a) Low noise: $y_i(t)$'s

(b) Low noise: $\hat{\beta}_j(t)$'s at the source location

(c) High noise: $y_i(t)$'s

(d) High noise: $\hat{\beta}_j(t)$'s at the source location

Figure 1. The measured signal, $y_i(t)$'s, from hypothetical 60 sensors, and the true, $\beta(t)$, and estimated source time courses, $\hat{\beta}_j(t)$'s, by different methods in Simulation I.

- The percentage of correctly identified $\beta_j(t)$'s. That is, the percentage of the correctly identified active sources plus the percentage of the correctly identified inactive

sources. We call this measure the *percentage of correctly identified signals*, denoted by PCIS.

- The overall mean squared error (MSE) between the true

Table 1. Comparing methods based on MSE and PCIS in Simulation I

Scenario	Criterion	STP (NCS)	STP (SVD)	MNE	MCE	STP-GEN (NCS)
Low noise	MSE (10^{-3})	51.6	56.1	178.8	45.6	54.9
	PCIS (%)	94.5	93.7	33.2	94.3	92.6
High noise	MSE (10^{-3})	58.2	60.9	196.6	82.9	67.7
	PCIS (%)	91.2	88.9	19.8	67.5	84.1

(a) β_j's at two source locations   (b) Time point 15, left hemisphere active   (c) Time point 30, right hemisphere active

Figure 2. The true source time courses at two different locations and the corresponding brain maps in Simulation II.

and the estimated solution, which is calculated by

$$MSE = E\left(\frac{1}{p}||\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}||_F^2\right),$$

where $\| \cdot \|_F^2$ is the Frobenius norm.

Since the data set in this simulation is fairly small, computation may not be a serious issue for the GEN method. Therefore, we implement the GEN algorithm within our STP framework (STP-GEN). Here we only examine the NCS basis because it has been considered superior to the SVD basis in this particular example. Table 1 lists MSE and PCIS for the three methods. In Scenario 1, MCE performed the best due to the smallest MSE and its high PCIS. STP with NCS is the second best method. Its MSE was small and its PCIS was as high as MCE if not higher. MNE was the least satisfactory method because its MSE was large and it only correctly identified 33% of the active/inactive sources. STP dominated over other methods in Scenario 2, where the noise level was high. The accuracy of MCE dropped in this scenario. It only correctly identified about 67.5% active/inactive sources, while its MSE increased approximately 90% compared to that in Scenario 1. STP still produced accurate, stable estimations. The MSE and PCIS did not change much. STP with SVD performed slightly inferior to STP with NCS basis. This followed our expectations because SVD basis tends to capture too many noise characteristics when the noise level is high.

Interestingly, IGEN outperformed the standard GEN method in both scenarios, especially when the noise level is high. This indicates that the proposed ensemble method is more robust to noise than the standard GEN method.

5.2 Simulation II

In this section we simulate the data based on a real-world MEG study. The forward operator $\mathbf{X}$ is obtained from the head model of a human subject and is normalized. It is computed using a boundary element model based on the subject's head geometric measurements. It is a 248-by-15,360 matrix. We simulate 100 active signals in two compact subareas in left and right hemispheres, respectively. Among these 100 sources, 50 are in a subarea in the left hemisphere and have the same function form. The other 50 are in a subarea in the right hemisphere and have the same function form with a slightly longer offset in time than that with subarea 1 [see Figure 2 (a)]. These 100 $\beta_j(t)$'s are simulated from a sine-exponential function:

$$(8) \qquad \beta_j(t) = 0.6\sin(5\pi t)\exp(-5t).$$

As we can see, the two peaks for the two source locations are at time points 15 and 30, respectively. Figure 2 (b) and (c) show the overlook of the true brain maps at these two peaks.

The source functions are recorded by 50 discrete time points from (0, 1). As with Simulation I, elements from the white noise, $e_i(t)$, are generated from $\mathcal{N}(0, 0.1)$ and $\mathcal{N}(0, 1)$
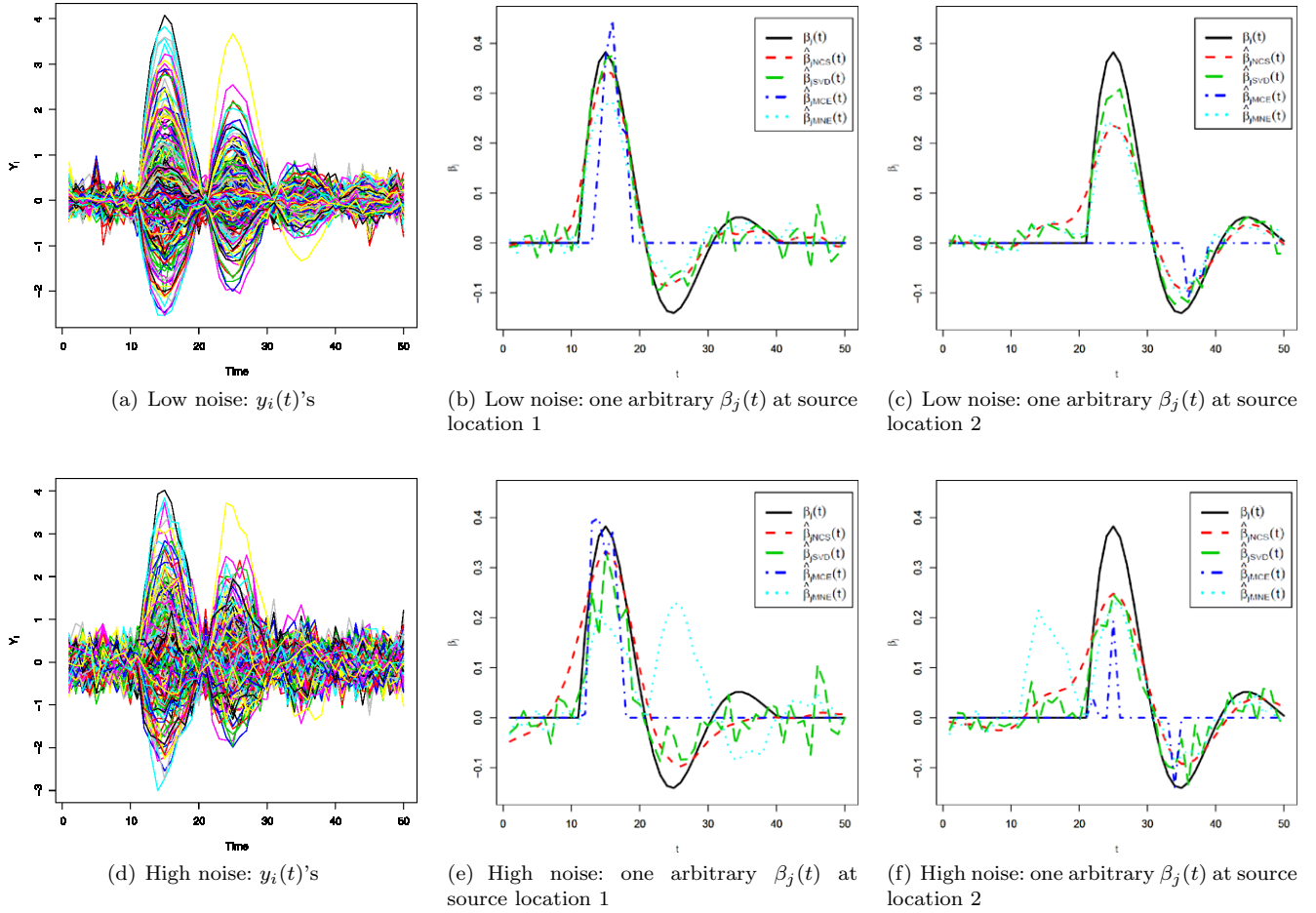
(a) Low noise: $y_i(t)$'s

(b) Low noise: one arbitrary $\beta_j(t)$ at source location 1

(c) Low noise: one arbitrary $\beta_j(t)$ at source location 2

(d) High noise: $y_i(t)$'s

(e) High noise: one arbitrary $\beta_j(t)$ at source location 1

(f) High noise: one arbitrary $\beta_j(t)$ at source location 2

Figure 3. The measured signal, $y_i(t)$'s, from hypothetical 248 sensors, and the true, $\beta(t)$'s, from two arbitrary dipoles and their corresponding estimates, $\hat{\beta}_j(t)$'s, by different methods at two locations, respectively, in Simulation II.

in Scenarios 1 and 2, respectively. The response, $\mathbf{y}(t)$, is generated by Equation (1).

Figure 3 (a) and (d) show $y_i(t)$'s from 248 hypothetical sensors in Scenarios 1 and 2, respectively. Since $\mathbf{X}$ is computed from a real head model, the measured $y_i(t)$'s are not as smooth as those in Simulation I even though they are at the same noise level. Scenario 2 is a more realistic situation because EEG/MEG data usually contain a high level of noise. This is because the brain's magnetic field (10–10^3 femtotesla) is considerably smaller than the ambient magnetic noise in an urban environment (10^8 femtotesla).

Figure 3 (b) and (c) show the estimated time courses in the two active subareas in Scenario 1. Different from Simulation I, the recovered sources listed in these figures are from one arbitrarily chosen source in the active subareas in order to make the graphs clearer. As we can see, STP and MNE recovered the entire time courses well. MCE was able to select one peak/valley in each source location, but it missed the valley for the first source location and the peak for the second source location. In addition, its solution was

too "spiky". Figures 4 (a) and (b) show the recovered brain maps at time points 15 and 30 using STP with SVD basis in Scenario 1. As can be seen, both subareas are identified correctly.

Figure 3 (d) and (e) show the estimated time courses in the two active subareas in Scenario 2. STP performed reasonably well in this Scenario too. Sources from the two active subareas were clearly recovered. MNE was able to identify the peaks for the two source locations with strong artifacts due to the time offset. MCE missed the valley of the first source location, and the spiky effect was obvious. Figures 4 (c) and (d) show the recovered brain maps at time points 15 and 30 using STP with SVD basis in Scenario 2. The two subareas can be identified even though they are not as clean as those in Scenario 1.

Table 2 lists the MSE and PCIS from the three methods. As can be seen, STP outperformed the other two methods due to small MSE and high PCIS in both scenarios. MNE and MCR produced poor MSE and PCIS. MCE is the least satisfactory among all the methods.
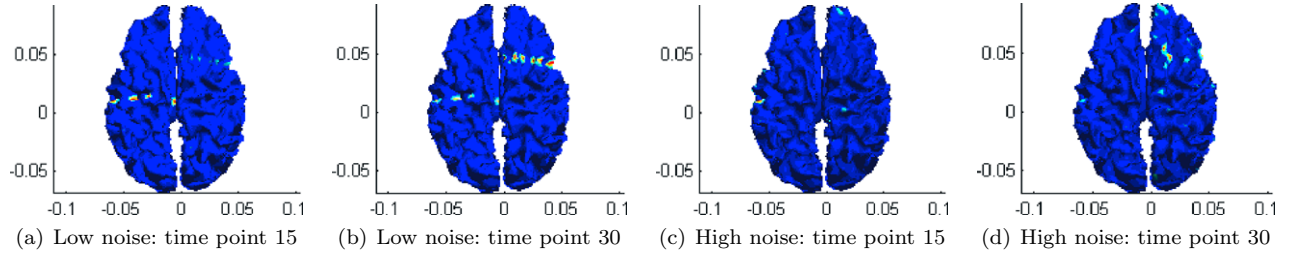
(a) Low noise: time point 15   (b) Low noise: time point 30   (c) High noise: time point 15   (d) High noise: time point 30

Figure 4. Reconstructed brain maps by STP (SVD) at two peak time points in Simulation II.

Table 2. Comparing methods based on MSE and PCIS in Simulation II

Scenario	Criterion	STP (NCS)	STP (SVD)	MNE	MCE
Low noise	MSE (10^{-3})	63.7	77.1	801.2	992.4
	PCIS (%)	95.4	94.1	60.2	86.7
High noise	MSE (10^{-3})	105.1	117.5	1011.2	1501.0
	PCIS (%)	85.6	80.3	31.0	67.2

5.3 Tuning parameters

In this section we perform sensitivity analyses to examine two of the most important tuning parameters. They are the number of iterations, L, and the maximum sub-sample size, p_{max}. Presumably, L should be as large as possible. However, it is impossible to assign an ultra-large L in practice due to the cost-effectiveness consideration. Therefore, we examine the relationship between L and p_{max} in terms of the MSE. The experiments in this section are all based on Simulation II Scenario 2.
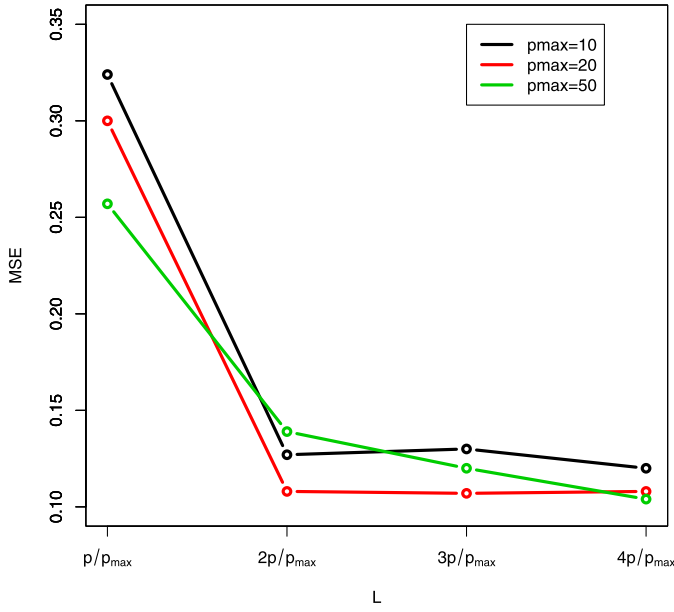
We examine four combinations of L and p_{max}: $L = \frac{p}{p_{max}}$, $\frac{2p}{p_{max}}$, $\frac{3p}{p_{max}}$, $\frac{4p}{p_{max}}$. In each combination, three values of p_{max} are examined: $p_{max} = 10, 20, 50$. In this simulation $p = 15,372$. We plot the values of L as functions of the MSE [see Figure 5 (a)]. As we can see, when $L = \frac{p}{p_{max}}$, the MSE for all three p_{max}'s are relatively large. But they all decrease as L increases. When $L \geq \frac{2p}{p_{max}}$, the MSE starts to level off. This indicates that it is reasonable to set values of L to be approximately equal to $\frac{2p}{p_{max}}$, because larger L will not significantly increase the accuracy but will dramatically increase the computational cost.

We also investigate the choice of p_{max} related to the computational cost. Letting $L = \frac{2p}{p_{max}}$, we examine four values of p_{max}: $p_{max} = 10, 20, 50, 100$. For each case, we record the CPU time in seconds on a 64-bit Dell Precision Workstation (24GB RAM, 4G quad CPU). We plot the p_{max} as a function of CPU time [see Figure 5 (b)]. As we can see, the CPU time roughly reaches the minimum when $p_{max} = 20$. When p_{max} is too small (e.g., 10), the computational cost is high because L is large. When p_{max} is too large (e.g., 50 or 100), L is small but GEN in each iteration takes a long time due to the lar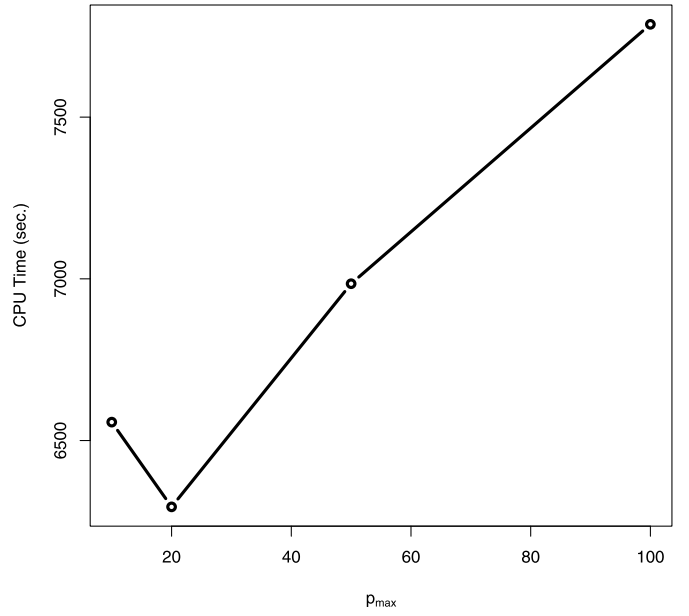ge number of variables involved. There-fore, p_{max} should be chosen to balance the computational cost and accuracy. Based on our empirical study, we suggest to use $L = \frac{2p}{p_{max}}$ and $p_{max} = 20$ for a data set with about 250 observed functions.

6. AN MEG EXAMPLE

In this section we provide a real-world MEG example of a 44-year-old female patient with grade three left frontal astrocytoma who underwent the MEG test as part of the presurgical evaluation. The MEG test discussed here is a so-matosensory task that is designed to non-invasively identify the somatosensory areas of the patient under evaluation, with the expectation that the primary somatosensory area contralateral to the site of activation will normally be acti-vated. Data collection was done with a whole-head neuro-magnetometer containing 248 first-order axial gradiometers. During the MEG somatosensory session, 558 repeated stim-ulations were delivered to her right lower lip through a pneu-matically driven soft plastic diaphragm (diameter: 1 cm) at the pressure setting of 25 psi. Each stimulation lasted 40 ms with 450 ms epoch duration (including prestimulus baseline of 100 ms) and an interstimulus interval randomized be-tween 0.5 s and 0.6 s. The data were collected with an analog high-pass filter of 0.1 Hz, a sampling rate of 508.63 Hz and a bandwidth of 200 Hz. During data processing a bandpass filter of 2 to 40 Hz was utilized and epochs were inspected to identify artifacts, then those 558 epochs without artifacts were averaged to obtain the final ERF (event-related mag-netic field) response. Finally, a bad channel (sensor A67) was removed. The MEG device recorded three trials with 228 time points in each epoch. We choose this study because it is an important task to clinical research in presurgical mapping.
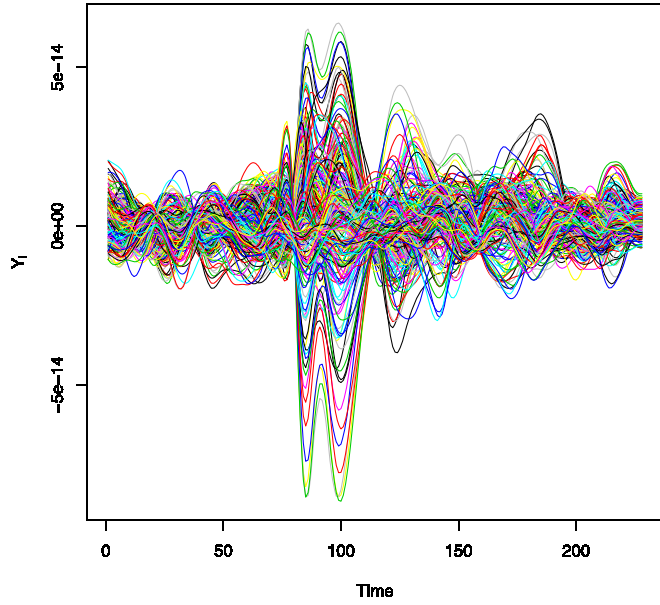
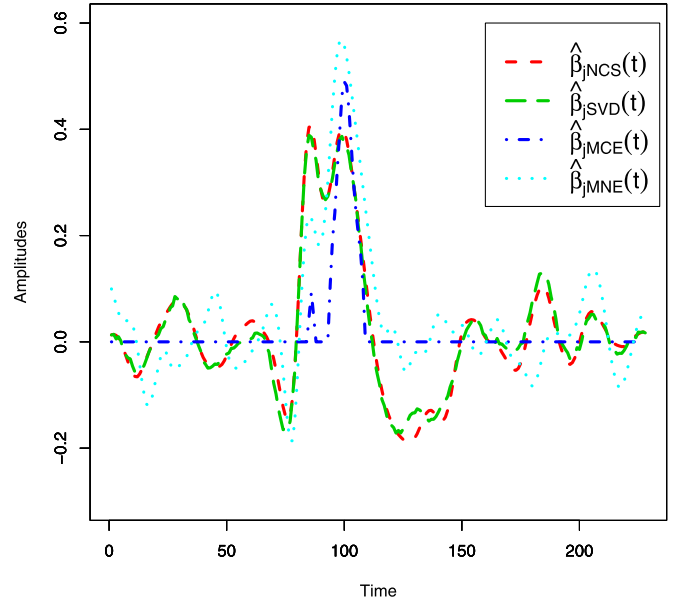(a) MSE as functions of L for three p_{max} values

(b) CPU time as a function of p_{max} (fix $L = \frac{2p}{p_{max}}$)

Figure 5. Investigating tuning parameters using Simulation II Scenario 1.

(a) MEG recordings from 247 valid channels

(b) Reconstructed source time courses for one arbitrary source

Figure 6. Human MEG recordings, $y_i(t)$'s, from 247 sensors and reconstructed time courses, $\hat{\beta}_j(t)$'s, by different methods for one arbitrary source in the somatosensory area.

After preprocessing, the measured data matrix becomes 247-by-228, where 247 is the number of good MEG channels and 228 is the number of recorded data points per epoch. That is, each row of this data matrix represents the time se-

ries of an MEG channel. From the first row to the last row, there are time series from sensor A1 to sensor A248 (except A67). The forward operator is arranged accordingly. Figure 6 (a) shows the time series from 247 valid channels.
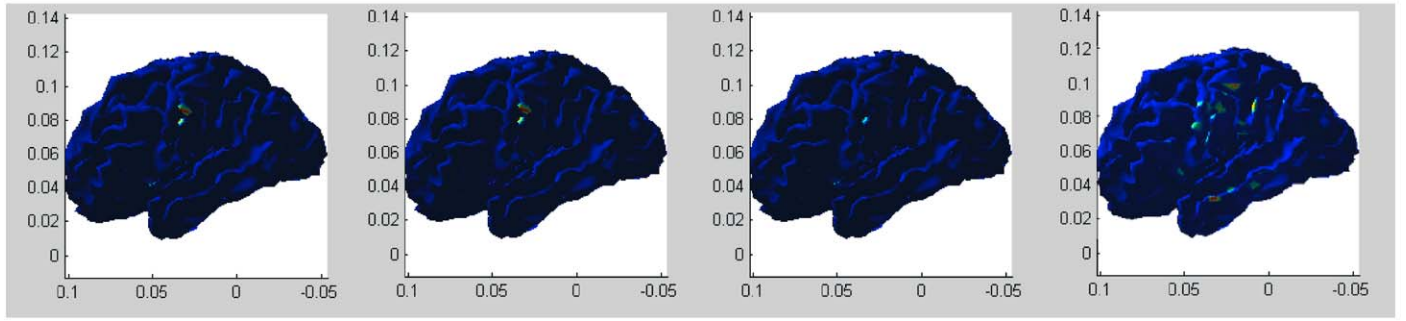
Figure 7. Reconstructed source maps at time point 99 (side views). From left to right: STP (NCS), STP (SVD), MCE, MNE.

As we can see, significant activation peaks corresponding to activations of somatosensory cortices are visible in this averaged event-related field response, and the peaks are achieved at around time points 85 and 99. We expect that the reconstructed signals should have significant activations in the somatosensory area at the corresponding two time points.

We apply STP with NCS and SVD bases ($q = 9$), MNE and MCE to the data. Figure 6 (b) shows the reconstructed time courses for one arbitrary source at the somatosensory cortex. Based on the nature of the experiment, we expect two peaks at around time points 85 and 99, respectively. As we can see, the peak at time point 99 was identified by all methods. However, MCE missed the earlier peak at time point 85 (or the magnitude of the activation was not large enough), and it also introduced discontinuity between the two peaks. MNE and the two versions of STP produced smooth time courses. Note that the MEG recordings we used in this study were preprocessed, i.e., filtered and averaged. Even with STP (SVD) [see Figure 6 (a)] the reconstructed time courses were fairly smooth.

Figure 7 shows side views of the 3-D brain maps at time point 99 by three methods because all methods seem to work well at time point 99 based on the time course study. STP with both bases performed the best because they correctly identified the somatosensory area (located at the left postcentral gyrus) and the active area was focal too. MCE also correctly identified the somatosensory area at time point 99, but its entire time courses are unsatisfactory [see Figure 6 (b)]. MNE was inferior to other methods, because the active areas from MNE were too broad. It picked out the somatosensory area but also incorrectly picked out some inactive areas.

7. CONCLUSION

We have presented a spatio-temporal penalization procedure for tackling the ill-posed EEG/MEG inverse problem. Given the measured EEG/MEG time courses at the sensor space, the goal of localizing and recovering the potential signal source time courses in the cortex is achieved by solving a nonparametric linear model.

The proposed approach utilizes the temporal bases to expand the sources and projects the temporal and spatial domains onto the same hyperplane. Then the nonparametric linear model is transformed to a high-dimensional linear regression problem. As a result, one source is related to multiple predictors. An iterative group elastic net algorithm is implemented to achieve temporal smoothness and spatial focality of the sources. The implementation of group shrinkage method allows us to select temporal information associated with an active source all together, while shrinking temporal information associated with an inactive source all towards zero. In particular, GEN results in a solution between L_1 and L_2. Hence, the reconstructed sources can smoothly transit from active areas to inactive areas. We demonstrate that STP performed considerably well over an L_2 based MNE method and an L_1 based MCE method in our simulation and MEG studies.

While STP showed advantages in terms of accuracy, a drawback is that it is more computationally expensive than MNE and MCE due to its iterative nature. However, it is claimed that this might not be a severe problem for the inverse problem, which does not require real-time solutions in most of the situations. Hence, a slightly time-consuming method is still desired as long as it can produce good accuracy. Another drawback of STP or probably many other methods based on basis representation is that using bases to represent the data will potentially lose some information. In addition, how to choose an appropriate basis may be another issue. Further investigation on fast solutions without doing basis expansion has been planned.

REFERENCES

AURANEN, T., NUMMENMAA, A., HÄMÄLÄINEN, M. S., JÄÄSKELÄINEN, I. P., LAMPINEN, J., VEHTARI, A. and SAMS, M. (2005). Bayesian analysis of the neuromagnetic inverse problem with lp-norm priors. *NeuroImage* **26**(3) 870–884.

BAILLET, S. and GARNERO, L. (1997). A bayesian approach to introducing anatomo-functional prior in the eeg/meg inverse problem. *IEEE Transactions on Biomedical Engineering* **44**(5) 374–385.

BAILLET, S., MOSHER, J. and LEAHY, R. (2001). Electromagnetic brain mapping. *IEEE Signal Processing Magazine* **18**(6) 14–30.

BARLOW, H. B. (1994). What is the computational goal of the neocortex? In Koch, C. and Davis, J. L. (eds.), *Large-Scale Neuronal Theories of the Brain*, pp. 1–22. Cambridge, MA: MIT press.

BIJMA, F., DE MUNCK, J. and HEETHAAR, R. (2005). The spatiotemporal meg covariance matrix modeled as a sum of kronecker products. *NeuroImage* **27** 402–415.

BOLSTAD, A., VEEN, B. V. and NOWAK, R. (2009). Space-time event sparse penalization for magneto-/electroencephalography. *NeuroImage* **46**(4) 1066–1081.

DALE, A. M. and SERENO, M. I. (1993). Improved localization of cortical activity by combining eeg and meg with mri cortical surface reconstruction: A linear approach. *Journal of Cognitive Neuroscience* **5** 162–176.

DAUNIZEAU, J., MATTOUT, J., CLONDA, D., GOULARD, B., BENALI, H. and LINA, J.-M. (2006). Bayesian spatio-temporal approach for eeg source reconstruction: Conciliating ecd and distributed models. *IEEE Transactions on Biomedical Engineering* **53**(3) 503–516.

DOGANDŽIĆ, A. and NEHORAI, A. (2000). Estimating evoked dipole responses in unknown spatially correlated noise with eeg/meg arrays. *IEEE Transactions on Signal Processing* **48**(1) 13–25.

FRISTON, K., HARRISON, L., DAUNIZEAU, J., KIEBEL, S., PHILLIPS, C., TRUJILLO-BARRETO, N., HENSON, R., FLANDIN, G. and MATTOUT, J. (2008). Multiple sparse priors for the m/eeg inverse problem. *NeuroImage* **39**(3) 1104–1120.

GORODNITSKY, I. F. and Rao, B. D. (1997). Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm. *Signal Processing, IEEE Transactions on* **45**(3) 600–616.

GRAVE DE PERALTA MENENDEZ, R. and GONZALEZ ANDINO, S. (1998). A critical analysis of linear inverse solutions to the neuroelectromagnetic inverse problem. *IEEE Transactions on Biomedical Engineering* **45** 440–448.

HÄMÄLÄINEN, M., HARI, R., ILMONIEMI, R. J., KNUUTILA, J. and LOUNASMAA, O. V. (1993, Apr). Magnetoencephalography-theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics* **65**(2) 413–497.

HÄMÄLÄINEN, M. and ILMONIEMI, R. J. (1994). Interpreting measured magnetic fields of the brain: estimates of current distribution. Technical report, Helsinki University of Technology. TKK-F-A599.

HAUFE, S., TOMIOKA, R., DICKHAUS, T., SANNELLI, C., BLANKERTZ, B., NOLTE, G., and MÜLLER, K. R. (2011, Apr). Large-scale eeg/meg source localization with spatial flexibility. *NeuroImage* **54**(2) 851–859.

HUIZENGA, H., DE MUNCK, J., WALDORP, L. and GRASMAN, R. (2002). Spatiotemporal eeg/meg source analysis based on a parametric noise covariance model. *IEEE Transactions on Biomedical Engineering* **49** 533–539.

IWAKI, S. and Ueno, S. (1998). Weighted minimum-norm source estimation of magnetoencephalography utilizing the temporal information of the measured data. *Journal of Applied Physics* **83**(11) 6441–6443.

JEFFS, B., LEAHY, R. and SINGH, M. (1987). An evaluation of methods for neuromagnetic image reconstruction. *IEEE Transactions on Biomedical Engineering* **34** 713–723.

JUN, S. C., GEORGE, J. S., PAŔE-BLAGOEV, J., PLIS, S. M., RANKEN, D. M., SCHMIDT, D. M. and WOOD, C. C. (2005). Spatiotemporal bayesian inference dipole analysis for meg neuroimaging data. *NeuroImage* **29**(1) 84–98.

LIN, F.-H., BELLIVEAU, J. W., DALE, A. M. and HÄMÄLÄINEN, M. S. (2006). Distributed current estimates using cortical orientation constraints. *Human Brain Mapping* **27** 1–13.

MATSUURA, K. and OKABE, Y. (1995). Selective minimum-norm solution of the biomagnetic inverse problem. *IEEE Transactions on Biomedical Engineering* **42** 608–615.

MOSHER, J., LEWIS, P. and LEAHY, R. (1992). Multiple dipole modeling and localization from spatio-temporal meg data. *IEEE Transactions on Biomedical Engineering* **39**(6) 541–557.

MOSHER, J. C., LEAHY, R. M. and LEWIS, P. S. (1999). Eeg and meg: Forward solutions for inverse methods. *IEEE Transactions on Biomedical Engineering* **46**(3) 245–259.

NUMMENMAA, A., AURANEN, T., HÄMÄLÄINEN, M. S., JÄÄSKELÄINEN, I. P., LAMPINEN, J., SAMS, M. and VEHTARI, A. (2007). Hierarchical bayesian estimates of distributed meg sources: Theoretical aspects and comparison of variational and mcmc methods. *NeuroImage* **35**(2) 669–685.

NUMMENMAA, A., AURANEN, T., VANNI, S., HÄMÄLÄINEN, M. S., JÄÄSKELÄINEN, I. P., LAMPINEN, J., VEHTARI, A. and SAMS, M. (2007). Sparse meg inverse solutions via hierarchical bayesian modeling: Evaluation with a parallel fmri study. Technical report, Helsinki, Finland. Report B65.

NUNEZ, P. (1981). *Electric Fields of the Brain: The Neurophysics of EEG*. Oxford University Press, New York, NY.

OPITZ, D. and MACLIN, R. (1999). Popular ensemble methods: An empirical study. *Journal of Artificial Intelligence Research* **11** 169–198.

OU, W., HÄMÄLÄINEN, M. and GOLLAND, P. (2009). A distributed spatio-temporal eeg/meg inverse solver. *NeuroImage* **44**(3) 932–946.

PASCUAL-MARQUI, R. D., MICHEL, C. M, and LEHMANN, D (1994). Low resolution electromagnetic tomography: A new method for localizing electrical activity in the brain. *International Journal of Psychophysiology* **18** 49–65.

RAO, B. D. and KREUTZ-DELGADO, K. (1999). An affine scaling methodology for best basis selection. *Signal Processing, IEEE Transactions on* **47**(1) 187–200. MR1729739

SARVAS, J. (1987). Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. *Physics in Medicine and Biology* **32**(1) 11–22.

SATO, M., YOSHIOKA, T., KAJIHARA, S., TOYAMA, K., GODA, N., DOYA, K. and KAWATO, M. (2004). Hierarchical bayesian estimation for meg inverse problem. *NeuroImage* **23**(3) 806–826.

SCHERG, M. and VON CRAMON, D. (1986). Evoked dipole source potentials of the human auditory cortex. *Electroencephalography and Clinical Neurophysiology* **65** 344–360.

SORRENTINO, A., PARKKONEN, L., PASCARELLA, A., CAMPI, C. and PIANA, M. (2009). Dynamical meg source modeling with multi-target bayesian filtering. *Human Brain Mapping* **30**(6) 1911–1921.

UUTELA, K., HÄMÄLÄINEN, M. and SOMERSALO, E. (1999). Visualization of magnetoencephalographic data using minimum current estimates. *NeuroImage* **10** 173–180.

VANVEEN, B., VAN DRONGELEN, W., YUCHTMAN, M. and SUZUKI, A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on Biomedical Engineering* **44** 867–880.

YAMAZAKI, T., KAMIJO, K., KENMOCHI, A., FUKUZUMI, S., KIYUNA, T., TAKAKI, Y. and KUROIWA, Y. (2000). Multiple equivalent current dipole source localization of visual event-related potentials during oddball paradigm with motor response. *Brain Topography* **12** 159–175.

YUAN, M. and LIN, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B* **68** 49–67. MR2212574

ZOU, H. and HASTIE, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B* **67** 301–320. MR2137327

Tian Siva Tian
Department of Psychology
University of Houston
Houston, TX 77004
USA
E-mail address: ttian@times.uh.edu

Zhimin Li
Center for Clinical Neurosciences
The University of Texas Health Science Center at Houston
Houston, TX 77225
USA
E-mail address: zhimin.li@uth.tmc.edu