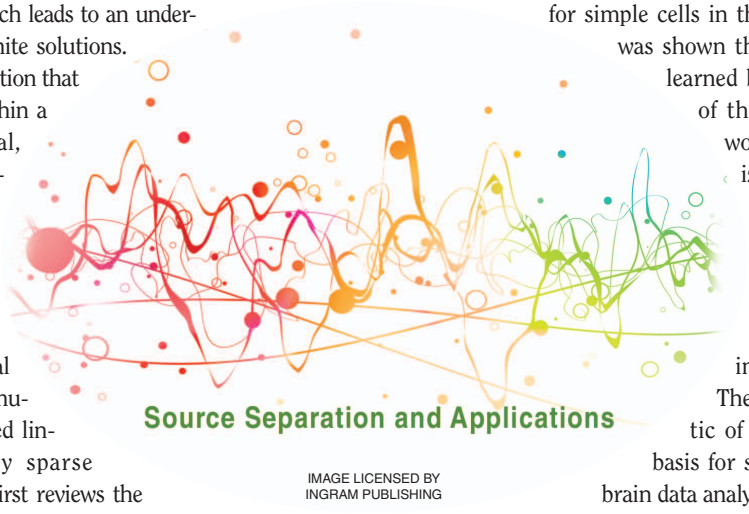
Yuanqing Li, Zhu Liang Yu, Ning Bi, Yong Xu, Zhenghui Gu, and Shun-ichi Amari

Sparse Representation for Brain Signal Processing

A tutorial on methods and applications

In many cases, observed brain signals can be assumed as the linear mixtures of unknown brain sources/components. It is the task of blind source separation (BSS) to find the sources. However, the number of brain sources is generally larger than the number of mixtures, which leads to an underdetermined model with infinite solutions. Under the reasonable assumption that brain sources are sparse within a domain, e.g., in the spatial, time, or time-frequency domain, we may obtain the sources through sparse representation. As explained in this article, several other typical problems, e.g., feature selection in brain signal processing, can also be formulated as the underdetermined linear model and solved by sparse representation. This article first reviews the probabilistic results of the equivalence between two important sparse solutions—the 0-norm and 1-norm solutions. In sparse representation-based brain component analysis including blind separation of brain sources and electroencephalogram (EEG) inverse imaging, the equivalence is related to the recoverability of the sources. This article also focuses on the applications of sparse representation in brain signal processing, including components extraction, BSS and EEG inverse imaging, feature selection, and classification. Based on functional magnetic resonance imaging (fMRI) and EEG data, the corresponding methods and experimental results are reviewed.



Digital Object Identifier 10.1109/MSP.2013.2296790
Date of publication: 7 April 2014

INTRODUCTION

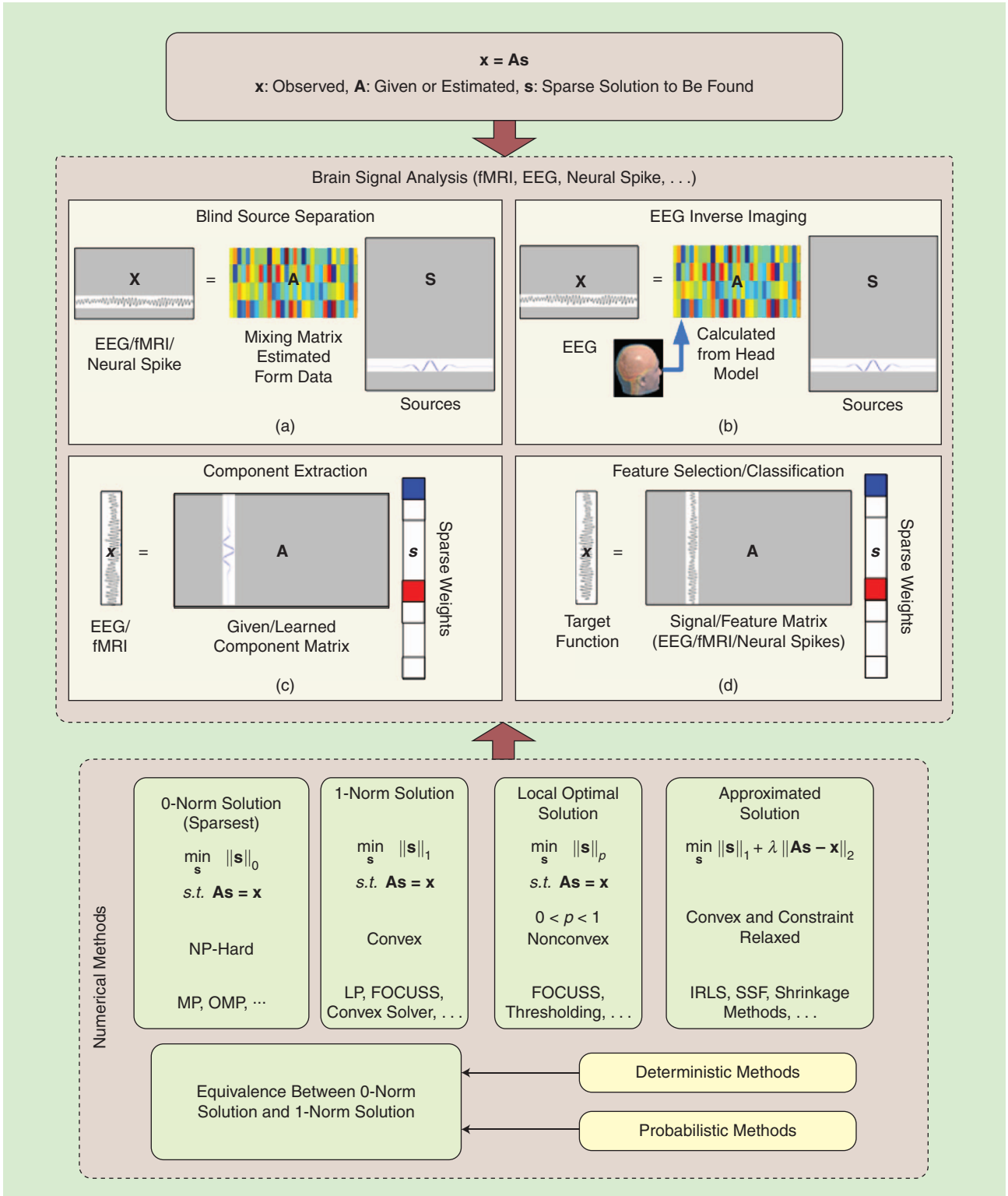
In recent years, sparse representation has received a great deal of attention in brain signal processing. Many biological findings support sparse representation/coding in the brain. For example, for simple cells in the primary visual cortex, it was shown that a set of receptive fields learned by maximizing the sparsity of the output of a neural network model is spatially localized, oriented, and selective to the spatial structure at a specific scale similar to cortical simple cells [1]. Sparsity of the neural response has been observed in neurons and in fMRI [2]. Therefore, sparsity characteristic of brain activities provides a basis for sparse representation-based brain data analysis. Many problems in brain signal processing can be formulated by the sparse representation models

$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad \text{(noiseless model)}, \tag{1}$$
$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{v} \quad \text{(noisy model)}, \tag{2}$$

where $\mathbf{x} \in R^n$ is a given signal vector, $\mathbf{A} \in R^{n \times m}$ $(n < m)$ is a basis/dictionary matrix, $\mathbf{s} \in R^m$ is the sparse coefficient vector to be found, and $\mathbf{v} \in R^n$ represents the noise. The basis matrix $\mathbf{A}$ can be randomly generated, or produced, from the union of several known bases such as Fourier and wavelet bases, and it can also be estimated from the data. Equations (1) and (2) can be in a matrix format, in which $\mathbf{x}$, $\mathbf{s}$, and $\mathbf{v}$ are replaced by a signal matrix, a coefficient matrix and a noise matrix, respectively.

1053-5888/14/$31.00©2014IEEE

For example, we often assume that the observed brain signals are the linear mixtures of brain sources, where both the mixing matrix, i.e., the basis matrix A in (1), and the brain sources are unknown/to be estimated [see Figure 1(a)]. This hypothesis has been demonstrated for EEG signals [3]. Generally, the number of brain sources is larger than the number of the mixtures. The brain sources can be assumed to be sparse in a domain such as the time domain or the time-frequency domain. Through sparse



[FIG1] The framework for the applications of sparse representation in brain signal analysis.

representation BSS, we can identify the brain sources based on the observed brain signals [4]. The brain sources can also be obtained and localized by sparse representation-based EEG inverse imaging, where the mixing matrix $\mathbf{A}$ is first estimated based on a head model, and the brain sources are then separated and localized [see Figure 1(b)]. For component extraction, a channel of brain signal can be treated as a linear mixture of a sparse set of dynamic components, each of which corresponds to a column of the basis matrix $\mathbf{A}$ [2] [see Figure 1(c)]. In feature selection, a target function (e.g., a stimulus function in an fMRI experiment) can be linearly regressed using a data/feature matrix $\mathbf{A}$, of which the data for each column are derived from a feature dimension (e.g., a voxel in the fMRI data). In other words, each row of the matrix $\mathbf{A}$ is a feature vector [5], as shown in Figure 1(d). The number of features (e.g., the voxels in fMRI data) is generally larger than the number of observations (e.g., the scanning time points for each voxel in the fMRI data). For the classification of brain signals, we may also employ the model shown in Figure 1(d), in which the target function is a test sample/feature vector and each column of the data matrix is a training sample/feature vector of a certain class [6]. These typical problems in brain signal processing can be solved under the framework of sparse representation.

THIS ARTICLE FOCUSES ON THE APPLICATIONS OF SPARSE REPRESENTATION IN BRAIN SIGNAL PROCESSING, INCLUDING COMPONENTS EXTRACTION, BSS AND EEG INVERSE IMAGING, FEATURE SELECTION, AND CLASSIFICATION.

The sparsity of a vector $\mathbf{s}$ can be measured by ℓ_p norm $\|\mathbf{s}\|_p$, which is defined as $\left(\sum_{i=1}^{m}|s_i|^p\right)^{1/p}$ for $0 < p \le 1$, and the number of nonzeros of $\mathbf{s}$ for $p = 0$. As shown in Figure 1, there are main four classes of algorithms for finding a sparse solution to (1):

- the algorithms for 0-norm minimization, including orthogonal matching pursuit (OMP) and matching pursuit (MP), where the corresponding 0-norm solution, denoted as $\mathbf{s}^{(0)}$, is the sparsest among all possible solutions of (1)
- the algorithms for 1-norm minimization, e.g., linear programming (LP) [7], where the corresponding 1-norm solution, denoted as $\mathbf{s}^{(1)}$, is also sparse but may not be the sparsest and is relatively easy to obtain
- the algorithms for p-norm minimization ($0 < p < 1$), e.g., focal underdetermined system solver (FOCUSS) and the thresholding algorithm [8], where the p-norm minimization is a nonconvex optimization problem with a global optimal solution that is difficult to be found [9] but is still of interest [10] (some studies have discussed $(1/2)$-norm minimization [8])
- the algorithms for finding an approximate solution through relaxing the constraints, including least absolute shrinkage and selection operator (LASSO), iterative reweighted least square (IRLS), separable surrogate functionals (SSFs), and the iterative shrinkage and thresholding algorithm (ISTA).

Previous studies have provided detailed reviews on the sparse representation algorithms [9]. These algorithms have been applied in image processing, compressive sensing, and BSS [9]. In brain signal processing, the choice of an algorithm depends on the specific task. For instance, if BSS/EEG inverse imaging is to be performed, it may be beneficial to use LP algorithms because there have been a lot of corresponding recoverability results and there is no need to set any regularization parameters for the algorithms. For feature selection/classification, all of the above-described sparse representation algorithms may be used, as the objective is generally to improve the classification accuracy, and only part of the relevant features may be required. Considering both the processing speed and accuracy, if the number of equations is small (e.g., $<1{,}000$), greedy algorithms can produce good results with high speed. For a larger number of equations, approximation methods such as ISTA, fast ISTA, and SSF are preferred.

When sparse representation is applied to brain signal processing, an important objective is to find the sparsest solution, i.e., the 0-norm solution. For instance, under the assumption that the brain sources are sparse, we may obtain the brain sources by finding the 0-norm solution through either sparse representation-based BSS or EEG inverse imaging. Although 0-norm minimization is NP-hard [9], the 0-norm solution can be obtained by 1-norm minimization in many cases. Equivalence between the 0-norm solution and the 1-norm solution thus becomes a key problem (see Figure 1). For sparse representation-based BSS and EEG inverse imaging, the equivalence is related to the recoverability of the brain sources. As shown in Figure 1, this problem can be analyzed using two different types of methods: deterministic methods (see [9] and the references therein) and probabilistic methods (see [4], [7], and [11]–[14]). Compared with the deterministic methods, the conditions under which the two sparse solutions are equivalent with a high probability in the probabilistic methods may be weaker. This is because the conditions obtained by the deterministic methods are sufficient for the equivalence but the probabilistic conditions are not.

RECOVERABILITY RESULTS BASED ON PROBABILISTIC METHODS

In BSS and EEG inverse imaging, it is necessary to consider the recoverability of the brain sources. As stated above, this problem is transformed into the equivalence between the 0-norm solution and the 1-norm solution when sparse representation is applied. Herein, we review several equivalence results obtained by probabilistic methods.

PROBABILISTIC METHODS BASED ON SUFFICIENT CONDITIONS

In references such as [10], [13], and [14], a set of sufficient conditions for the equivalence between the 0- and 1-norm solutions were proposed with respect to the basis matrix and the sparsity of the 0-norm solution. The probability that these sufficient conditions would hold for a random basis matrix was

subsequently analyzed. For instance, in [10], it was shown that the proposed sufficient conditions, in particular, the CS conditions, hold for a random basis matrix with an overwhelming probability. In [14], a probability result for the equivalence was obtained based on both an invertibility condition and a magnitude condition, which were imposed on a partial Fourier matrix and a particular polynomial, respectively. To illustrate the probabilistic method, we present several results related to a so-called restricted isometry property (RIP) condition [13], [15].

RIP can be described, as shown in [15], via the following definitions.

DEFINITION 1

For each integer $j = 1, 2, \cdots, n$, the matrix $\mathbf{A} \in R^{n\times m}$ satisfies the RIP of order j if there exists a $\delta_j \in (0,1)$ such that

$$\left(1-\delta_j\right)\| \mathbf{s} \|_2^2 \le \| \mathbf{As} \|_2^2 \le \left(1+\delta_j\right)\| \mathbf{s} \|_2^2 \tag{3}$$

holds for all j-sparse vectors $\mathbf{s}$. A vector is said to be *j-sparse* if it has at most j nonzero entries. The isometry constant $\bar{\delta}_j$ of a matrix $\mathbf{A}$ is defined as the smallest δ_j satisfying (3).

Recently, it was proven that the equivalence holds if a j-sparse solution exists for (1) and the isometry constant $\bar{\delta}_j < (1/3)$ [16]. When the matrix $\mathbf{A}$ is given randomly, using the Johnson–Lindenstrauss lemma and width theory, R. DeVore et al. provided a probability estimation that (3) holds as below [13].

THEOREM 1

Suppose that for given n, m, and $0 < \delta < 1$, the probability distribution generating the $n \times m$ random matrices $\mathbf{A}$ satisfies the concentration inequality, i.e., $\forall \imath \in (0,1)$ and $\forall \mathbf{s} \in R^m$, there exists a $c_0(\varepsilon) > 0$, such that

$$P(|\| \mathbf{As} \|_2^2 - \| \mathbf{s} \|_2^2| \ge \varepsilon \| \mathbf{s} \|_2^2) \le 2e^{-nc_0(\varepsilon)}, \tag{4}$$

where the probability is taken over all $n \times m$ matrices $\mathbf{A}$. Then there exist constants $c_1, c_2 > 0$ depending only on δ such that the RIP condition (3) holds for $\mathbf{A}$ with the prescribed δ and any $j \le c_1 n/\log(m/j)$ with a probability $\ge 1 - 2e^{-c_2 n}$ [13].

Remark 1

There are a number of random matrices including random Gaussian and Bernoulli matrices, with distributions satisfying the concentration inequality (4) [13]. Using Theorem 1 and the bound of the isometry constant that guarantees the equivalence, we may obtain a lower bound for the equivalence probability. For example, let $\delta = \delta_j < (1/3)$ and $\mathbf{s}^{(0)}$ be j-sparse with $j \le c_1 n/\log(m/j)$. Following Theorem 1 and Theorem 3.1 in [16], $P\left(\mathbf{s}^{(1)} = \mathbf{s}^{(0)}\right) \ge 1 - 2e^{-c_2 n}$. Because the parameters c_1 and c_2 in Theorem 1 are difficult to explicitly express, the probability (or the lower bound) that the RIP condition holds is difficult to obtain, especially for fixed n and m. In fact, the probability in Theorem 1 shows an overwhelming likelihood with respect to n. The recoverability results based on the RIP condition can be extended to a noisy case (2), as shown in [16].

EQUIVALENCE PROBABILITY ESTIMATION

Different from the above approach, Li et al. proposed an alternative method to directly estimate the probability that the two sparse solutions are equivalent [4], [7], [11], [12]. Several equivalence probability estimates were presented in [11] for a fixed basis matrix and in [12] for a random basis matrix both with a random 0-norm solution. We present several equivalence probability estimations and the corresponding simulation results for a fixed basis matrix with a random 0-norm solution.

THEOREM 2 [12]

Let $\mathbf{A} \in R^{n\times m}$, $\mathbf{s}^{(0)} \in R^m$, and $\mathbf{s}^{(1)}$ be the 1-norm solutions satisfying (1) with $\mathbf{x} := \mathbf{As}^0$. Then $\mathbf{s}^{(0)} = \mathbf{s}^{(1)}$, if and only if the optimal value of the following optimization problem is less than 1/2

$$\max \sum_{1\le k\le m} \left[\operatorname{sign}(s_k^{(0)}) z_k\right]_+ \text{ s.t. } \mathbf{Az} = 0, \| \mathbf{z} \|_1 = 1, \tag{5}$$

where $[y]_+ = \max(y, 0)$ for all $y \in R$.

It is also NP-hard to check the sufficient and necessary condition in Theorem 2 and to precisely specify the set for which the maximum in (5) is computed [11], [12]. However, from Theorem 2, given the basis matrix $\mathbf{A}$ the recoverability of the 0-norm solution through 1-norm minimization depends only on the index set and the signs of its nonzero entries, i.e., the sign vector/pattern. The equivalence probability can be obtained by determining how many sign vectors can be recovered through 1-norm minimization, as illustrated in the following three cases [11].

Case 1

The number of nonzero entries for the 0-norm solution $\mathbf{s}^{(0)}$ is fixed, e.g., j. In this case, there are a total of $2^j C_m^j$ sign vectors with 0-norm j. Suppose that q_j is the number of sign vectors with 0-norm j which can be recovered by 1-norm minimization. We have

$$p_j = P\left(s^{(1)} = s^{(0)} \,|\, \| s^{(0)} \|_0 = j\right) = \frac{q_j}{2^j C_m^j}, \tag{6}$$

where q_j can be obtained by checking the equivalence of each sign vector with 0-norm j and its corresponding 1-norm solution. This can be performed either by Theorem 2 or by directly comparing the 0-norm solution with its corresponding 1-norm solution obtained by 1-norm minimization.

Case 2

The number of nonzero entries for $s^{(0)}$ is unknown, but the probability that every entry of $s^{(0)}$ is equal to zero is known and denoted as α. In this case, $P\left(\| s^{(0)} \|_0 = j\right) = C_m^j (1-\alpha)^j \alpha^{m-j}$ and

$$P\left(s^{(1)} = s^{(0)}\right) = \sum_{j=0}^{m} C_m^j (1-\alpha)^j \alpha^{m-j} p_j. \tag{7}$$

Case 3

All entries for $s^{(0)}$ are drawn from a Laplacian distribution with the probability density function $(\lambda/2)\exp(-\lambda|x|)$. In this case,

$$P\left(s^{(1)} \approx s^{(0)}\right) = \sum_{0 \le j \le m} C_m^j (1-\alpha_\lambda)^j \alpha_\lambda^{m-j} p_j, \tag{8}$$

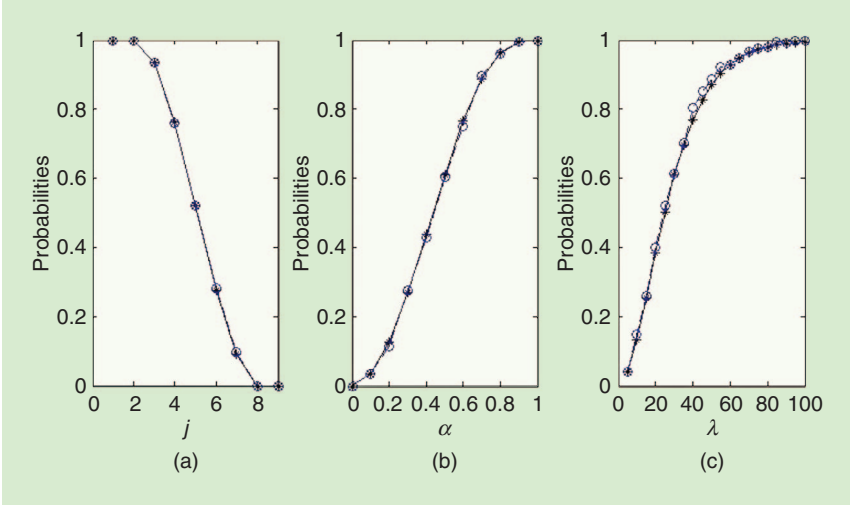
where $\alpha_\lambda = P\left(|s_k^{(0)}| < \varepsilon_0\right) = 1 - \exp(-\lambda\epsilon_0)$ and $\varepsilon_0 > 0$ is a sufficiently small constant.

We review several simulation results [11]. The basis matrix $\mathbf{A} \in R^{7\times 9}$ was arbitrarily given in advance. Figure 2 shows the estimated equivalence probability curves (solid curves with "*") and the true equivalence probability curves (dashed curves with "*o*") for the above three cases. Note that each true probability value was obtained by randomly generating 1,000 corresponding source vectors (0-norm solutions) and counting the number of source vectors recovered by 1-norm minimization. It follows from Figure 2 that the probability estimates from (6)–(8) accurately reflected the true probability that a 0-norm solution was recovered by 1-norm minimization in Cases 1–3, respectively.

Remark 2

Based on the equivalence probability estimates described above, it is still NP-hard to obtain the exact values for a large m. However, approximations can be obtained by applying a sampling method to the set of sign vectors [17]. In addition to the above-described equivalence problem, another important problem in sparse representation is the uniqueness of the 0-norm solution of (1). In [7], it was shown that for model (1), a 0-norm solution with less than n nonzeros is unique with a probability of one if the basis matrix $\mathbf{A}$ is randomly given.

In brain signal processing, noise generally can not be neglected. According to the discussion in [7], the 1-norm solution of (1) is robust to noise to some degree. In particular, for a given $\mathbf{A}$, there exists an $M > 0$ such that $\|\mathbf{s}_v^{(1)} - \mathbf{s}^{(1)}\|_1 < M\|\mathbf{v}\|_1$, where $\mathbf{s}_v^{(1)}$ is the 1-norm solution of the noisy model (2). Therefore, for a case with low noise, we can determine the 0-norm solution through 1-norm minimization, provided that the equivalence holds. Furthermore, the above probability estimations on the equivalence have been extended to the noise case in [17]. Simulation results showed that these probability estimations of the equivalence hold at noise levels of approximately 18 dB. However, for brain signals, it is difficult to estimate the noise level, which may be very high. Further studies are needed to demonstrate the effectiveness of the above-described probability estimations and to evaluate the brain sources obtained by sparse representation-based BSS/EEG inverse imaging. Additionally, the above-described probability estimates are based on the sparsity of the 0-norm solution, which is generally difficult to directly determine. It has been shown that the sparsity of sources can be estimated based on the sparsity of the mixtures [17].



[FIG2] Curves for estimated equivalence probabilities (solid curves with "*") and true equivalence probabilities (dash-dotted curves with "*o*"). (a) Case 1, (b), Case 2, and (c) Case 3.

FUNCTIONAL MRI DATA ANALYSIS

In this section, we discuss two applications of sparse representation in fMRI data analysis: the modification of a general linear regression model (GLM) with statistical parameter mapping (SPM), and brain decoding.

MODIFIED GLM-SPM APPROACH BASED ON SPARSE REPRESENTATION

An important objective of fMRI data analysis is to detect the weak blood-oxygen-level dependent (BOLD signal from the noisy data and localize the activated regions in the brain. GLM-SPM is a common method for fMRI data analysis [18] that is based on

$$\mathbf{x}_i = \mathbf{G}\beta_i + \mathbf{e}_i, \tag{9}$$

where $\mathbf{x}_i \in R^N$ is a time series of the ith voxel $(i = 1, \ldots, M)$, $\mathbf{G} \in R^{N\times K}$ is a so-called design matrix of which each column corresponds to an explanatory variable related to the specific experimental conditions under which the data were collected, $\beta_i \in R^K$ is an unknown weight vector to be estimated for each voxel, and $\mathbf{e}_i \in R^N$ is a noise vector. At the end of the GLM learning process, the statistical parameter map is obtained based on t- or F-statistics calculated using the regression coefficients, and it can be used to display the activated brain areas and the importance of each voxel.

Several studies have been conducted to improve the GLM-SPM method based on sparse representation. One method for improving GLM is to modify the design matrix to overcome its drawbacks. The design matrix is generally constructed using the canonical hemodynamic response function (HRF). However, this function does not fully reflect the individual and experimental variances that occur during the task period [2]. To overcome this issue, Hu et al. proposed an SPM-ICA framework with a design matrix composed of the components learned by an independent component analysis (ICA) algorithm [19]. In [2], a data-driven sparse GLM method was proposed

for maximum likelihood (ML) estimation of spatially adaptive design matrices and sparse response signals. In this method, the BOLD signal at a specific voxel can be regarded as a linear combination of a sparse set of dynamic components [see Figure 1(c)]. A K-SVD-based iterative algorithm was used to identify these components and to perform sparse coding at the same time. Based on the learned design matrix and sparse response signals, a statistical test was then used to detect the activated voxels. Using simulated and real fMRI data, this method was demonstrated to adapt to individual variations better than canonical HRF and spatial ICA.

SPARSE REPRESENTATION-BASED MVPA METHODS FOR BRAIN DECODING

Multivariate pattern analysis (MVPA) has recently become a popular approach for analyzing fMRI data. MVPA approaches open the possibility of separating and localizing spatially distributed patterns, which are generally too weak to be detected by univariate methods, such as GLM [20]. By effectively pooling the information available across many fMRI voxels, MVPA methods allow the perceptual, cognitive, and behavioral parameters or features to be decoded. The results of brain decoding can be used to further assess how precisely cognitive information is encoded by the activity of neural populations within the whole brain [21]. Because there are far more voxels (e.g., 30,000) than fMRI volumes/scanning time points for each of the voxels (e.g. 1,000), voxel selection plays an important role in the MVPA-based brain decoding of fMRI data [see Figure 1(d)]. It has been shown that sparse representation is an effective method for voxel selection. The sparseness leads to a simple prediction function useful for avoiding overfitting. For instance, a LASSO regression was used to reconstruct muscle activity from human cortical fMRI data, where the correlated voxels were selected through the learned sparse weights [22]. In the following, we briefly present a sparse representation-based voxel selection algorithm and its two variants/extensions. Specific details for these algorithms can be obtained from [5] and [23].

MVPA APPROACHES OPEN THE POSSIBILITY OF SEPARATING AND LOCALIZING SPATIALLY DISTRIBUTED PATTERNS, WHICH ARE GENERALLY TOO WEAK TO BE DETECTED BY UNIVARIATE METHODS, SUCH AS GLM.

LINEAR PROGRAMMING-BASED FEATURE SELECTION ALGORITHM

Let $\mathbf{A} \in R^{n \times m}$ denote an fMRI data matrix, where each column is the time series of a voxel. Let $\mathbf{x} \in R^n$ denote a stimulus/task function convolved with an HRF. A stimulus function can be constructed by setting its value to one when the stimulus is available; otherwise the value is set to zero. The following algorithm is designed to identify the columns of $\mathbf{A}$ (i.e., voxels in the fMRI data) that are relevant to $\mathbf{x}$.

Algorithm 1 (LP-Based Feature Selection Algorithm):

Step 1: For $k = 1, \cdots, K_0$ (a predefined integer, e.g., 100), perform Steps 1.1 and 1.2.

Step 1.1: Randomly choose L (e.g., $0.3n$) rows from the matrix $\mathbf{A}$ to construct an L by m submatrix denoted as $\mathbf{A}_k$, with the corresponding L entries of $\mathbf{x}$ forming a column vector denoted by $\mathbf{x}_k \in R^L$.

Step 1.2: Solve the optimization problem (denote the optimal solution as $\overline{s}^{(k)}$),

$$\min ||\mathbf{s}||_1, s.t.\ \mathbf{A}_k\mathbf{s} = \mathbf{x}_k. \tag{10}$$

Step 2: Let

$$\mathbf{s} = \frac{1}{K_0}\sum_{k=1}^{K_0} \overline{\mathbf{s}}^{(k)}. \tag{11}$$

Step 3: Using the weight vector $\mathbf{s}$, we either select a fixed number (e.g., 100) of voxels with large absolute weights or select voxels with absolute weights higher than a given positive constant θ_0. The threshold parameter θ_0 can be chosen in various ways, e.g., through the cross-validation method [5].

Because of noise, the weight vector $\mathbf{s}$ obtained by a single optimization may not accurately reflect the importance of the features. Thus, we calculate an average weight vector by boosting the 1-norm minimization in Algorithm 1. With the selected voxels, decoding of the stimulus/task parameters can be performed using a classification/regression algorithm such as SVM.

OMP-BASED FEATURE SELECTION ALGORITHM

In Step 1.2 of Algorithm 1, we can replace the LP/basis pursuit (BP) algorithm with the OMP algorithm [9] to conduct sparse representation and obtain the OMP-based feature selection algorithm. For fMRI data, we find that the OMP-based algorithm is faster than Algorithm 1 because the number of time samples is generally much smaller than the number of voxels and the OMP algorithm avoids a large-scale LP problem.

SPARSE REPRESENTATION-BASED PATTERN LOCALIZATION ALGORITHM

There are three unsolved problems in Algorithm 1: the abilities to 1) select all informative features, 2) differentiate those selected features according to two stimuli classes/brain states, and 3) remove incorrectly selected features (irrelevant/noisy features). A sparse representation-based pattern localization (SPL) algorithm was used to address these problems [23]. The SPL algorithm contains a K-fold cross-validation procedure. In each fold, a recursive iterative feature elimination method relying on the weights obtained by sparse representation (e.g., Algorithm 1 or the OMP-based feature selection algorithm) is used to identify as many informative features as possible. Each iteration selects those informative features for removal. The next iteration is based on the remaining features. After the selected

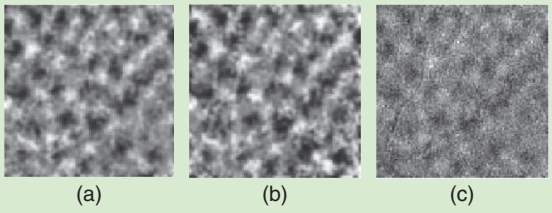
[TABLE 1] PREDICTION ACCURACIES (PEARSON CORRELATION) FOR FOUR TASK FUNCTIONS (HITS, INSTRUCTIONS, FACES, AND VELOCITY), OBTAINED BY THREE METHODS: GLM-SPM, LP, AND OMP-BASED ALGORITHMS.

TASK	SUBJECT 1			SUBJECT 2			SUBJECT 3		
	GLM-SPM	LP	OMP	GLM-SPM	LP	OMP	GLM-SPM	LP	OMP
HITS	0.18	0.31	0.33	0.35	0.39	0.47	0.25	0.33	0.33
INSTR.	0.51	0.75	0.60	0.63	0.66	0.57	0.32	0.83	0.60
FACES	0.06	0.15	0.12	–0.01	0.31	0.28	0.61	0.65	0.42
VELOCITY	0.25	0.38	0.63	0.38	0.38	0.69	0.16	0.35	0.61

features from all of the iterations are removed, the decoding accuracy based on the remaining features is close to the chance level. According to the signs of the weights, the selected features are divided into two sets corresponding to two stimuli classes/brain states. Next, two probability maps/density functions are constructed using the two classes of features selected across the K folds. Inside each probability map, the probability value of a feature is obtained by counting the number of times that the feature is selected across all folds. Finally, to remove the irrelevant features, the two probability maps are tested with a permutation test either at the individual level or at the group level if group data are available. In this way, two patterns corresponding to the two stimuli classes/brain states are obtained.

EXPERIMENTAL RESULTS

First, we individually applied Algorithm 1 and the OMP-based feature selection algorithm to the fMRI data of PBAIC 2007 for voxel selection. The details for the data can be found at http://pbc.lrdc.pitt.edu/?q=taxonomy/term/45. The subject performed several tasks in a virtual reality world during scan acquisition (e.g., Hits: whenever the subject picked up fruits or weapons; Instruction: whenever task instructions were presented; Faces: whenever the subject looked at faces; Velocity: whenever the subject was moving). For each task, a stimulus function was provided by PBAIC 2007. The corresponding task function was computed by convolving this stimulus function with an HRF, which reflected the delay of the hemodynamic responses with respect to the onsets of the stimuli. The data analysis predicted the task functions from the fMRI data.



[FIG3] The dark and bright blobs in (a)–(c) are shown for two classes of stimuli, i.e., the horizontal axis-of-motion stimuli and the vertical axis-of-motion stimuli. (a) and (b) The difference maps reconstructed using our BP-based and OMP-based SPL algorithms, respectively. (c) A difference map between the two stimulus conditions.

For each of the three subjects, two runs of data, collected in different time intervals, were analyzed. After data processing, each run of fMRI data corresponded to a matrix with 500 rows (time points) and approximately 32,000 columns (voxels). Based on the BP (Algorithm 1) or OMP-based feature selection algorithm, we performed twofold cross-validation. In each fold, voxel selection was performed based on the training data (data from Run 1/Run 2) in two steps, i.e., an initial selection based on Pearson correlation coefficients between the time series of voxels and the task function and a second selection based on our algorithm. Using the selected voxels (ranging from two to 100), we predicted the task function of the test data (data from Run 2/Run 1) through ridge regression. The prediction performance was measured as the Pearson correlation between the actual task and the predicted task functions. For the purpose of comparison, we used the GLM-SPM method to replace our method to select the voxels. The average prediction accuracies across all the numbers of the selected voxels are shown in Table 1. The high accuracy values demonstrate the effectiveness of our algorithms. Based on Table 1, our algorithms overperformed GLM-SPM method in the majority of cases, with only two exceptions: the accuracies obtained by the OMP-based method for subject 2 in the instruction task and for subject 3 in the face task. One important reason for this finding is that GLM-SPM is a univariate method, whereas our algorithms are multivariate. Multivariate methods simultaneously consider a set of variables, and their advantages have been shown in numerous studies [20]. For the face task, the results for the three subjects were significantly different, possibly because of the different levels of attention that the subjects paid to this task.

Second, we present results obtained by applying the SPL algorithm to an optical imaging data set that was collected from a macaque monkey. The detailed experimental procedure was described elsewhere [23]. Based on data from 40 horizontal and 40 vertical axis-of-motion trials, we used a leave-one-out method to search the informative pixels and obtained two probability maps. The difference between the two probability maps, which reflects the class information, is shown in Figure 3(a), (based on Algorithm 1) and Figure 3(b) (based on the OMP-based feature selection algorithm). As shown in Figure 3(c), we also obtained a differential map between the two conditions using an established method for optical imaging data analysis (the so-called differential mapping method). The dark and bright blobs in Figure 3(c) represent the two classes of informative features. In Figure 3, a comparison of (a) and (b) to (c) reveals that our algorithms can find all of the informative features and further separate them into two classes corresponding to the two experimental conditions. These results show that the columnar structures in the V1 area of the visual cortex of the monkey can be detected by the SPL algorithm.

EEG DATA ANALYSIS

In this section, we discuss the application of sparse representation in EEG data analysis including component analysis and pattern classification.

EEG COMPONENT ANALYSIS

BLIND SOURCE SEPARATION

EEG signals can be considered as the linear mixtures of unknown sources with an unknown mixing matrix. In this case, the true sources can be obtained through sparse representation-based BSS [see Figure 1(a)]. Under the condition that the sources are sparse within a domain such as the time-frequency domain, the sparse representation-based BSS can be conducted using a two-step method ([7], [11], and the references therein). The mixing matrix is first estimated using, e.g., a clustering algorithm. The sources are then obtained using a sparse representation algorithm. In the two-step method, it is difficult to estimate the mixing matrix precisely. For sparse representation-based BSS, the number of sources can be larger than the number of the mixtures and the sources can be correlated, provided that the sources are sufficiently sparse. In [4], a wavelet packet transformation was first applied to an EEG data set collected in an experiment with a modified Sternberg memory task for producing sparsity. Specifically, in the experiment, the subjects were instructed to memorize three numbers successively presented at random positions on a computer monitor. The effectiveness of the subjects' memory was evaluated using a "test number" presented 2.5 s later. Next, the ratio matrix was constructed as the mixing matrix using wavelet packet transformation coefficients. Third, the sources were estimated by 1-norm minimization. Furthermore, several pairs of almost uncorrelated sources were obtained, which showed memory-related synchronization and desynchronization.

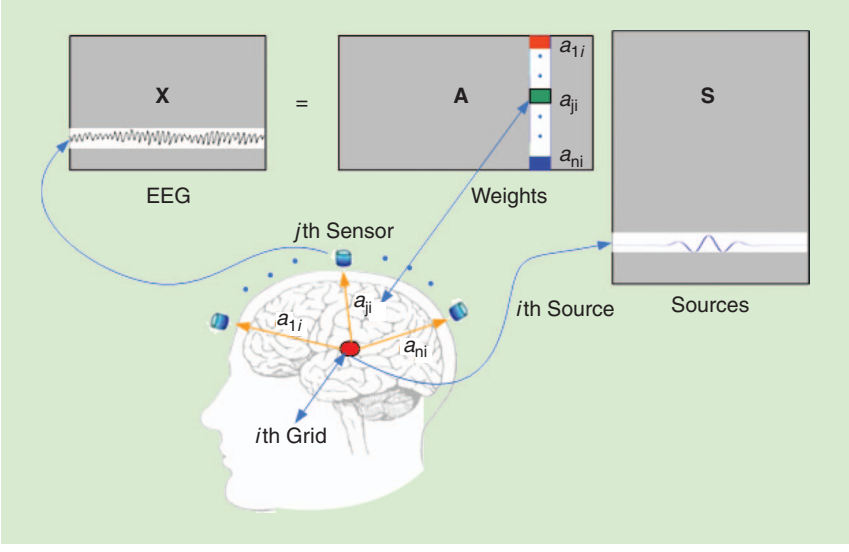
EEG SIGNALS CAN BE CONSIDERED AS THE LINEAR MIXTURES OF UNKNOWN SOURCES WITH AN UNKNOWN MIXING MATRIX. IN THIS CASE, THE TRUE SOURCES CAN BE OBTAINED THROUGH SPARSE REPRESENTATION-BASED BSS.

EEG INVERSE IMAGING

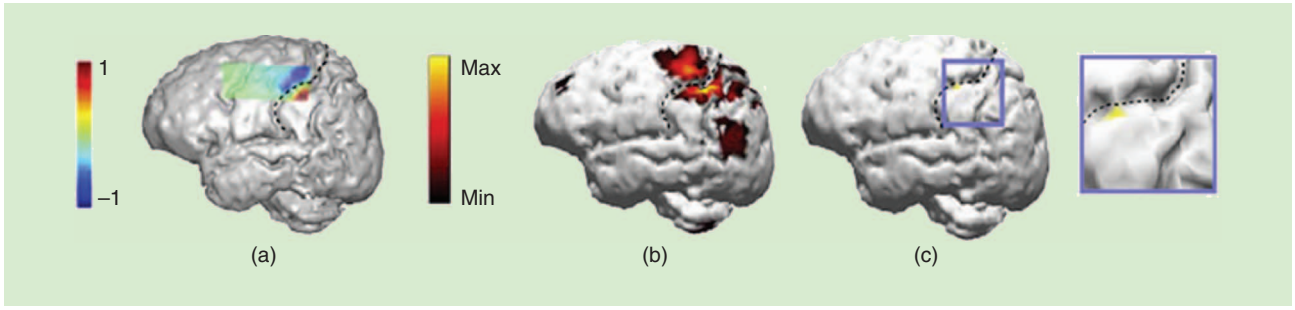
EEG signals are generally considered to be generated from the synchronized activation of cortical pyramidal neurons. Through forward modeling of brain sources and head volume conduction, EEG inverse imaging can identify these sources and their localizations [see Figures 1(b) and 4] [24]. EEG inverse imaging is useful for the study of brain mechanisms and diseases detection. For example, high-resolution EEG inverse imaging can be used to identify the origin and propagation of dynamic epileptic activity and to provide information for presurgical planning for the patients [24].

The distributed source model for inverse imaging assumes that a large number of unit dipoles are evenly positioned in the brain volume or over the cortical sheet of gray matter, and each dipole represents a candidate source [24]. Under this assumption, the first step of inverse imaging is the forward modeling of the brain sources and head volume conduction to establish a linear source-to-measurement relationship. Specifically, for the model shown in Figure 4, the jth row of $\mathbf{X}$ is an observed EEG signal from the jth sensor. The ith column of the matrix $\mathbf{A}$ [a lead field matrix (LFM)], corresponding to the ith grid, describes how a unit dipole with a certain location and orientation is related to the EEG measurements [24]. The ith row of $\mathbf{S}$ is a brain source associated with the ith grid through the ith column of $\mathbf{A}$. In practice, the number of grids can range from 3,000 to 9,000 depending on several factors, e.g., the data analysis task and whether the grids are evenly distributed in the brain volume or over the cortical sheet of gray matter. Given the configuration of the grids, the placement of the sensors and the head model, the transfer property a_{ji} between the ith grid and the jth sensor can be calculated using either a boundary element model (BEM) or a finite element model (FEM) [24]. Thus, the matrix $\mathbf{A}$ is determined. The head model can be constructed from single spherical shell, multiple spherical shells or the structural MRI. The second step is to identify the brain sources from the observed EEG data based on the linear model.

Because the number of electrodes n is much lower than the number of grids m, there are infinite solutions for the linear inverse problem. Various methods have been developed to



[FIG4] **A diagram for EEG inverse imaging.**



[FIG5] The results for EEG inverse imaging (extracted from Figure 4 in [25] with permission). (a) Direct subdural ECoG recordings, (b) imaging results from LORETA, and (c) imaging results from the sparse method. Dotted lines indicate the central sulcus.

obtain an "optimal" source estimate by introducing biophysical and/or physiological constraints. Under the assumption that the brain sources are sparse, the inverse imaging problem can be solved through sparse representation (1-norm minimization) [25]. Once the sparse brain sources are obtained, they can be localized by the corresponding grids.

Using ECoG mappings as a gold standard for evaluation, it has been shown that the results of the sparse method are comparable to those of other methods, e.g., 2-norm-based low-resolution brain electromagnetic tomography (LORETA) for EEG inverse imaging [25]. We present the partial results from this study in Figure 5. Figure 5(a) shows the direct subdural ECoG recordings, whereas (b) and (c) show the imaging results from the LORETA and sparse methods, respectively. The results are displayed with the threshold set at 70% of the maximum current density (A/mm^2). The hot spots (yellow areas) indicate that the source activities, estimated by the sparse and the LORETA methods, were located on the sensory cortex. However, the hot spot for the sparse method corresponded more closely with the area identified by the subdural ECoG map. This experiment demonstrates that the sparse method provides better specificity than the LORETA method for focal sources. However, for spatial extended sources, multiple distributed source imaging methods have been tested in well-controlled simulations [26]. For sources with large spatial extents ($10\,\text{mm}^2 - 40\,\text{cm}^2$), the LORETA method could provided effective results. In conclusion, 2-norm-based methods including LORETA are suited to imaging spatially distributed sources, whereas sparse methods are suited to imaging sparse and focal sources. Considering the different advantages of these two classes of methods, several inverse imaging algorithms combining 1-norm and 2-norm have been developed [24].

FEATURE SELECTION AND CLASSIFICATION OF EEG SIGNAL

FEATURE SELECTION

Although the EEG signals are contaminated by noise and artifacts caused by volume conduction effects, EEG patterns still have typical spatial, temporal, and spectral distribution characteristics. For example, the motor imagery of the left/right hand gives rise to contralateral attenuation (or accentuation) in mu and beta rhythm activities in EEG, known as event-related desynchronization (or synchronization) (ERD/ERS) [27], which can be located in sensorimotor areas. P300 potentials occur approximately 300 ms after the attended stimulation and exist primarily in the parietal area, which means that signals collected from certain electrodes at certain time intervals or frequency bands are easier to discriminate than other signals. Therefore, it is desirable to build a spatial/temporal/spectral filter for feature extraction/selection. Furthermore, dimension reduction based on feature extraction/selection may lead to better generalization performance for the corresponding classifier.

Feature selection from EEG data can be performed by sparse representation, as shown in Figure 1(d). Take the feature extraction for motor imagery-based ERD/ERS as an example. We often use labeled EEG data to train a common spatial pattern (CSP) filter. To distinguish the features of the two classes of data corresponding to the left- and right-hand motor imageries, respectively, the CSP algorithm finds the spatial filters that maximize the variance for one class and at the same time minimize the variance for the other class [27]. Because the CSP method is based on the optimization of signal variance, which is 2-norm, the resultant filter weights are nonsparse, which implies that all of the channels are used in the following classification. However, because the ERD/ERS is located in specific areas (e.g., the sensorimotor area), only nearby channels have good discrimination for the two classes. The other channels need to be removed before classification. In this case, sparse representation is well tailored for channel selection. Selection can be achieved by simply modifying the optimization problem in the CSP method by introducing the 1-norm of the filter weights in the objective function [28]. The lower-weighted channels can be viewed as irrelevant and can thus be removed. This method is called sparse CSP (SCSP). Experiments have shown that the classification accuracy based on SCSP is greater than the accuracy of regular CSP [28]. This outstanding performance can be explained by the fact that sparse representation reduces the number of participating channels, with a concomitant denoising effect.

CLASSIFICATION

Sparse representation-based classification (SRC) can be conducted as shown below [see Figure 1(d)]. Suppose that the basis matrix $\mathbf{A}$ is composed of two component submatrices

corresponding to the two classes, i.e., $\mathbf{A} = [\mathbf{A}_1 \,|\, \mathbf{A}_2]$. The mutual coherence of the two submatrices is defined by

$$MC(\mathbf{A}_1, \mathbf{A}_2) = \max\{|\langle \mathbf{a}_{1,i}, \mathbf{a}_{2,j} \rangle| : i=1, \ldots, N_1;\ j=1, \ldots, N_2\} \quad (12)$$

where $\mathbf{a}_{1,i}$ is the ith column of $\mathbf{A}_1$, $\mathbf{a}_{2,j}$ is the jth column of $\mathbf{A}_2$, and N_1 and N_2 are the numbers of columns in $\mathbf{A}_1$ and $\mathbf{A}_2$, respectively. The inner product of the two vectors is denoted by $\langle \cdot, \cdot \rangle$. When MC is low, i.e., the basis matrix is incoherent, a test data vector from one class can be predominantly represented by the columns of the same class in the basis matrix [6]. Thus, the classification based on the sparse coefficients, which can be obtained by sparse representation, is prone to producing the correct label. In ERD/ERS, for example, the CSP features can be used to construct the basis matrix because CSP filtering maximizes the incoherence between the two classes. Using the basis matrix for sparse regression of a test feature vector, a sparse solution is obtained for further classification. Specifically, the class label is determined by computing the energy of the coefficients for each class and assigning the class label of the larger one to the test data. SRC was applied in the data analysis of several motor imagery-based brain–computer interface (BCI) data sets and showed better classification performance than the well-known linear discriminant analysis method [6].

SPARSE REPRESENTATION CAN ALSO BE APPLIED TO OTHER BRAIN SIGNALS SUCH AS NEURAL SPIKE DATA, MAGNETOENCEPHALOGRAPHY, AND MRI.

CONCLUDING REMARKS

This article discussed the applications of sparse representation in brain signal processing, including BSS, EEG inverse imaging, feature selection, and classification. Although we mainly focused on fMRI and EEG data, sparse representation can also be applied to other brain signals such as neural spike data [29], magnetoencephalography [30], and MRI [31]. When sparse representation is used to separate sources from brain signals in BSS and EEG inverse imaging, recoverability of the sources is a basic problem. This problem can be transformed to the equivalence between the 0-norm solution and the 1-norm solution. We first reviewed the recoverability/equivalence results obtained by probabilistic methods. Next, we reviewed several fMRI studies to illustrate how to improve the GLM-SPM, a common method in fMRI data analysis, through sparse component analysis, and how to perform feature selection based on LP or OMP methods. For EEG signal processing, we reviewed several sparse representation methods and experimental results for BSS, inverse imaging, feature selection, and classification.

We identified several challenging problems for further study:

- Brain signals are highly dynamic. Several sparse representation methods have been proposed to capture the dynamic properties of the brain signals, e.g., a mixed-norm estimate method based on the structured sparsity of the sources [30]. Because of the high complexity of brain signals, this problem still needs further investigation. It may also be possible to establish time-varying sparse representation methods in which the basis matrix is time-varying.
- For high-dimensional brain signals, the existing sparse representation algorithms are generally time consuming. For some applications including BCIs, fast/real-time sparse representation algorithms are expected. One option is to reduce the dimensionality by considering the neurophysiological mechanisms of brain activities, e.g., brain functional areas, and develop or choose fast sparse representation algorithms such as fast ISTA.
- Because brain signals are highly noisy, it is challenging to evaluate the obtained brain sources. The existing recoverability results (deterministic/probabilistic) need to be extended to highly noisy cases, and indirect but more effective methods need to be developed to explore the neurophysiological reasonability of the brain sources associated with the corresponding experimental conditions.

ACKNOWEDGMENTS

This work was supported by 863 Program of China (2012AA011601), the National Foundation of Natural Science of China (91120305, 61105121, and 61175114), and the High-Level Talent Project of Guangdong Province, China.

AUTHORS

Yuanqing Li (auyqli@scut.edu.cn) received the B.S. degree in applied mathematics from Wuhan University, China, in 1988, the M.S. degree in applied mathematics from South China Normal University in 1994, and the Ph.D. degree in control theory and applications from South China University of Technology, China, in 1997. Currently, he is with the South China University of Technology as a full professor. He is the author or coauthor of more than 80 scientific papers in journals and conference proceedings. His research interests include blind signal processing, sparse representation, machine learning, BCIs, EEG, and fMRI data analysis.

Zhu Liang Yu (zlyu@scut.edu.cn) received the B.S.E.E. degree in 1995 and the M.S.E.E. degree in 1998, both in electronic engineering from the Nanjing University of Aeronautics and Astronautics, China. He received his Ph.D. degree in 2006 from Nanyang Technological University, Singapore. He joined the Center for Signal Processing, Nanyang Technological University in 2000, as a research engineer and then as a research fellow. In 2008, he joined the College of Automation Science and Engineering, South China University of Technology, where he was promoted to a full professor in 2009. His research interests include signal processing, pattern recognition, and biomedical applications.

Ning Bi (mcsbn@mail.sysu.edu.cn) received the B.S. and Ph.D. degrees in mathematics from Zhejiang University, China, in 1989 and 2002, respectively. He is currently with the School of Mathematics and Computational Science and Guangdong Province Key

Laboratory of Computational Science, Sun Yat-Sen University, Guangzhou, China. He has published more than 30 papers and coauthored the book *An Introduction to Multiband Wavelets* with Q. Sun and D. Huang. His research interests include multiband wavelets, watermarking, and compressed sensing.

Yong Xu (yxu@scut.edu.cn) received the B.S., M.S., and Ph.D. degrees in mathematics from Nanjing University, Nanjing, China, in 1993, 1996, and 1999, respectively. He was a postdoctoral research fellow in computer science with South China University of Technology, Guangzhou, China, from 1999 to 2001, where he became a faculty member and where he is currently a professor with the School of Computer Science and Engineering. His current research interests include image analysis, image and video recognition, image quality assessment, and object recognition. He is a member of the IEEE Computer Society and the Association of Computing Machinery and a Senior Member of the IEEE.

Zhenghui Gu (zhgu@scut.edu.cn) received the Ph.D. degree from Nanyang Technological University, Singapore, in 2003. From 2002 to 2008, she was with the Institute for Infocomm Research, Singapore. In 2008, she joined the College of Automation Science and Engineering, South China University of Technology, Guangzhou, as an associate professor. Her current research interests include the fields of signal processing and pattern recognition.

Shun-ichi Amari (amari@brain.riken.jp) received the Ph.D. degree in 1963 from the University of Tokyo. He worked as a professor at Kyushu University and then at the University of Tokyo, where he is now professor emeritus. He is an ex director and senior advisor at the RIKEN Brain Science Institute. He has worked on a wide range of mathematical sciences, including information geometry and mathematical neuroscience. He served as president of the International Neural Network Society (INNS). He received the IEEE Emanuel R. Piore Award, the INNS Society Gabor Award, and the Japanese Order of Cultural Merit among many others.

REFERENCES

[1] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

[2] K. Lee, S. Tak, and J. C. Ye, "A data-driven sparse GLM for fMRI analysis using sparse dictionary learning with MDL criterion," *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1076–1089, 2011.

[3] J. Sarvas, "Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem," *Phys. Med. Biol.*, vol. 32, no. 1, pp. 11–22, 1987.

[4] Y. Li, A. Cichocki, and S.-I. Amari, "Blind estimation of channel parameters and source components for EEG signals: A sparse factorization approach," *IEEE Trans. Neural Netw.*, vol. 17, no. 2, pp. 419–431, 2006.

[5] Y. Li, P. Namburi, Z. Yu, C. Guan, J. Feng, and Z. Gu, "Voxel selection in fMRI data analysis based on sparse representation," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 10, pp. 2439–2451, 2009.

[6] Y. Shin, S. Lee, J. Lee, and H.-N. Lee, "Sparse representation-based classification scheme for motor imagery-based brain–computer interface systems," *J. Neural Eng.*, vol. 9, no. 5, p. 056002, 2012.

[7] Y. Li, A. Cichocki, and S.-I. Amari, "Analysis of sparse representation and blind source separation," *Neural Comput.*, vol. 16, no. 6, pp. 1193–1234, 2004.

[8] Z. Xu, X. Chang, F. Xu, and H. Zhang, "$l_{1/2}$ regularization: A thresholding representation theory and a fast solver," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 7, pp. 1013–1027, 2012.

[9] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, no. 1, pp. 34–81, 2009.

[10] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[11] Y. Li, S.-I. Amari, A. Cichocki, D. W. C. Ho, and S. Xie, "Underdetermined blind source separation based on sparse representation," *IEEE Trans. Signal Processing*, vol. 54, no. 2, pp. 423–437, 2006.

[12] Y. Li, S.-I. Amari, A. Cichocki, and C. Guan, "Probability estimation for recoverability analysis of blind source separation based on sparse representation," *IEEE Trans. Inform. Theory*, vol. 52, no. 7, pp. 3139–3152, 2006.

[13] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constr. Approx.*, vol. 28, no. 3, pp. 253–263, 2008.

[14] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[15] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *C. R. Math.*, vol. 346, no. 9, pp. 589–592, 2008.

[16] T. Cai and A. Zhang, "Sharp RIP bound for sparse signal and low-rank matrix recovery," *Appl. Comput. Harmon. Anal.*, vol. 35, no. 1, pp. 74–93, 2013.

[17] Y. Li, A. Cichocki, S.-I. Amari, S. Xie, and C. Guan, "Equivalence probability and sparsity of two sparse solutions in sparse representation," *IEEE Trans. Neural Netw.*, vol. 19, no. 12, pp. 2009–2021, 2008.

[18] K. J. Friston, A. P. Holmes, K. J. Worsley, J.-P. Poline, C. D. Frith, and R. S. J. Frackowiak, "Statistical parametric maps in functional imaging: A general linear approach," *Hum. Brain Mapp.*, vol. 2, no. 4, pp. 189–210, 1994.

[19] D. Hu, L. Yan, Y. Liu, Z. Zhou, K. J. Friston, C. Tan, and D. Wu, "Unified SPM-ICA for fMRI analysis," *Neuroimage*, vol. 25, no. 3, pp. 746–55, 2005.

[20] S. A. Harrison and F. Tong, "Decoding reveals the contents of visual working memory in early visual areas," *Nature*, vol. 458, no. 7238, pp. 632–635, 2009.

[21] Y. Li, J. Long, B. Huang, T. Yu, W. Wu, Y. Liu, C. Liang, and P. Sun, "Crossmodal integration enhances neural representation of task-relevant features in audiovisual face perception," *Cerebral Cortex,* 2013. [Online]. Available: http://cercor.oxfordjournals.org/content/early/2013/08/26/cercor.bht228.abstract

[22] G. Ganesh, E. Burdet, M. Haruno, and M. Kawato, "Sparse linear regression for reconstructing muscle activity from human cortical fMRI," *Neuroimage*, vol. 42, no. 4, pp. 1463–1472, 2008.

[23] Y. Li, J. Long, L. He, H. Lu, Z. Gu, and P. Sun, "A sparse representation-based algorithm for pattern localization in brain imaging data analysis," *PLoS ONE*, vol. 7, no. 12, p. e50332, 2012.

[24] B. He, L. Yang, C. Wilke, and H. Yuan, "Electrophysiological imaging of brain activity and connectivity - challenges and opportunities," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 7, pp. 1918–1931, 2011.

[25] X. Bai, V. L. Towle, E. J. He, and B. He, "Evaluation of cortical current density imaging methods using intracranial electrocorticograms and functional MRI," *Neuroimage*, vol. 35, no. 2, pp. 598–608, 2007.

[26] C. Grova, J. Daunizeau, J.-M. Lina, C. G. B E nar, H. Benali, and J. Gotman, "Evaluation of EEG localization methods using realistic simulations of interictal spikes," *Neuroimage*, vol. 29, no. 3, pp. 734–753, 2006.

[27] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, 2008.

[28] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek, "Optimizing the channel selection and classification accuracy in EEG-based BCI," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 6, pp. 1865–1873, 2011.

[29] K. Oweiss and M. Aghagolzadeh, "Detection and classification of extracellular action potential recordings," in *Statistical Signal Processing for Neuroscience and Neurotechnology*, (Statistical Signal Processing for Neuroscience and Neurotechnology), K. G. Oweiss, Ed. Amsterdam, The Netherlands: Elsevier, 2010.

[30] A. Gramfort, M. Kowalski, and M. Hämäläinen, "Mixed-norm estimates for the M/EEG inverse problem using accelerated gradient methods," *Phys. Med. Biol.*, vol. 57, no. 7, pp. 1937–1961, 2012.

[31] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 58, no. 6, pp. 1182–1195, 2007.

[SP]