AN ALGORITHM FOR MULTI SUBJECT FMRI ANALYSIS BASED ON THE SVD AND PENALIZED RANK-1 MATRIX APPROXIMATION

Asif Iqbal and Abd-Krim Seghouane

Department of Electrical and Electronic Engineering
University of Melbourne, Australia

ABSTRACT

In recent years, data driven methods have been successfully used for analyzing multi-subject functional magnetic resonance imaging (fMRI) datasets. These methods attempt to learn shared spatial activation maps (SM) or voxel time courses (TC) from temporally or spatially concatenated fMRI datasets respectively. Most of the methods proposed so far do not distinguish whether a particular SM/TC is a group level component or only present in a certain subject dataset. In this paper we present a new two stage algorithm which aims to separate the joint and sub-specific information from the temporally concatenated multi-subject datasets. The proposed method is based on the singular value decomposition (SVD) and penalized rank-one matrix approximation. Simulation experiments are used to demonstrate this ability of the proposed algorithm followed by validation on real experimental task-fMRI datasets.

Index Terms— Functional magnetic resonance imaging (fMRI), data driven methods, temporal concatenation, multi-subject analysis.

1. INTRODUCTION

In many signal and image processing applications, data driven methods have been used extensively leading to state of the art performance in face recognition [1], image denoising [2, 3], and fMRI data analysis [4–9] to name a few. For example, these methods aim to learn an overcomplete basis set (a dictionary) such that each signal can be well represented by using only a few of the basis functions (atoms). For most this learning process is performed in two steps; keeping the basis constant, a linear approximation is performed for the observed signal, followed by updating the basis to minimize some cost function while fixing parameters of the linear constraint. These two steps are alternated till a stopping criteria is satisfied.

Data driven methods with sparsity constraint (dictionary learning) have been used for single subject [5–8, 10], as well as, [11–15] multi-subject fMRI data analysis. In this context, a single subject fMRI dataset is decomposed into a factor model comprised of learned basis and matrix of coefficients such that each voxels' timecourse (TC) can be represented by the linear combination of a few vectors from the learned basis. In these formulations, each vector of the learned basis corresponds to a specific brain temporal dynamic with the corresponding coefficient vector (row of the coefficient matrix) representing the spatial activity map. Typically for multi-subject (MS) fMRI analysis, dictionary learning is performed using temporally concatenated datasets resulting in group level spatial maps and subject specific temporal dynamics [11]. However, one might ask whether a specific recovered spatial map is a group level map or is only present in a certain subject dataset. Current MS data driven methods are unable to make this distinction.

In this paper we present a novel SVD and penalized rank-1 matrix approximation method which aims to separate the overall joint (common across subjects) information from the sub-specific (local) information. The proposed method accomplishes this in two stages; starting with temporally concatenated datasets, in first stage SVD is used to decompose the datasets into two low-rank matrices containing joint and sub-specific information. Second stage further refines and decomposes these matrices into joint and sub-specific temporal dynamics and activation maps respectively.

2. PROPOSED FORMULATION FOR MULTI SUBJECT FMRI ANALYSIS

2.1. Existing formulation

Consider p fMRI datasets denoted by $\mathbf{Y}_i \in \mathbb{R}^{n\times N}$, $i \in [1, p]$, with N brain voxels and n time points per voxel. Data driven methods with sparsity constraints aim to decompose $\mathbf{Y}_i$ as:

$$\mathbf{Y}_i = \mathbf{D}_i\mathbf{X}; \forall i \in [1, p]\,, \text{ with } \mathbf{D}_i \in \mathbb{R}^{n\times k}, \mathbf{X} \in \mathbb{R}^{k\times N} \quad (1)$$

with l_2 normalized columns of $\mathbf{D}_i$ and a sparse coefficient matrix $\mathbf{X}$. Under a joint learning framework, all datasets are temporally concatenated to construct $\mathbf{Y} \in \mathbb{R}^{np\times N}$ and are decomposed into a factor model (dictionary) $\mathbf{D} \in \mathbb{R}^{np\times k}$ and

This work was supported by the Australian Research Council through Grant FT. 130101394

sparse coefficient matrix $\mathbf{X} \in \mathbb{R}^{k \times N}$. The problem can be formulated as [16]:

$$\min_{\mathbf{D},\mathbf{X}} ||\mathbf{Y} - \mathbf{D}\mathbf{X}||_F^2 + \lambda \sum_{q=1}^{N} \|\mathbf{x}_q\|_1 \text{ s.t. } \forall\, i, l,\ \|\mathbf{d}_{il}\|_2 \leq 1 \quad (2)$$

where λ is the sparsity controlling parameter, $\mathbf{d}_{il}$ represents l^{th} column of matrix $\mathbf{D}_i$, and $\|\cdot\|_1$ and $\|\cdot\|_2$ are the column-wise l_1 and l_2 vector norms respectively. This problem can be efficiently solved by using online dictionary learning (ODL) algorithm [17]. The resulting $\mathbf{D}$ and $\mathbf{X}$ matrices contain k dense temporal dynamics and k sparse group level spatial maps (SM) respectively.

Learning SMs using this formulation lacks the ability to distinguish group-level (joint) SMs from sub-specific ones. In the next section, we propose a novel dictionary learning algorithm which decomposes the multi-subject fMRI datasets into matrices containing joint information and sub-specific information which are then further refined to extract most significant joint and sub-specific TCs and SMs respectively.

2.2. Proposed Data Driven Method

Starting with $\mathbf{Y} \in \mathbb{R}^{np \times N}$ containing normalized and temporally concatenated fMRI datasets, in the first stage, we aim to decompose it into the sum of three low-rank matrices; $\mathbf{Y} = \mathbf{J} + \mathbf{I} + \mathbf{E}$ containing joint info, sub-specific (individual) info, and representation error matrix respectively. The idea is to use singular value decomposition (SVD) to capture the most significant information from $\mathbf{Y}$ in $\mathbf{J}$. Starting with $\mathbf{J} = \mathbf{I} = \mathbf{E} = \mathbf{0}$, we construct $\mathbf{J}$ as

$$\mathbf{J} = \mathbf{U}\,(:, 1 : r_J)\, \mathbf{\Sigma}\,(1 : r_J, 1 : r_J)\, \mathbf{V}\,(:, 1 : r_J)^\top \quad (3)$$

where $\mathbf{U}$, $\mathbf{V}$ contain the left and right singular vectors and $\mathbf{\Sigma}$ contains the singular values of matrix $\mathbf{X} = \mathbf{Y} - \mathbf{I}$. Here the resulting $\mathbf{J}$ is the best r_J-rank approximation of $\mathbf{X}$ in terms of the Frobenius norm [18]. The matrix $\mathbf{I}$ is then constructed subject-wise as best r_I-rank approximation of $\mathbf{Z}_i = \mathbf{Y}_i - \mathbf{J}_i$. Experimental results have shown that only a few iteration of these steps are enough instead of repeating till convergence. At this point, most significant joint and sub-specific info would have been captured in $\mathbf{J}$ and $\mathbf{I}$ respectively.

Stage two of our algorithm constitutes refining the information present in $\mathbf{J}$ and $\mathbf{I}_i$ into k sparse spatial maps and their corresponding k smooth temporal dynamics [19]. Generally speaking, we aim to minimize the following cost function:

$$\min \frac{1}{2}\|\mathbf{G}_0 - \mathbf{A}\mathbf{B}\|_F^2 + \sum_{m=1}^{k} \left(\alpha_1 \|\mathbf{b}^m\|_1 + \alpha_2\, \mathbf{a}_m^\top\, \mathbf{\Omega}\, \mathbf{a}_m\right) \quad \text{s.t. } \|\mathbf{a}_m\|_2 = 1 \quad (4)$$

where $\mathbf{a}_m$ and $\mathbf{b}^m$ are the columns and rows from matrices $\mathbf{A} \in \mathbb{R}^{n \times k}$ and $\mathbf{B} \in \mathbb{R}^{k \times N}$ respectively, $\|\cdot\|_1$ is l_1-norm, $\mathbf{\Omega} \in \mathbb{R}^{n \times n}$ is a non-negative definite roughness penalty matrix [7] [19], α_1 is the sparsity, and α_2 is the atom smoothness controlling parameter. Here $\mathbf{G}_0$ can be either $\mathbf{J}$ or $\mathbf{I}_i$. The hemodynamic response of brain voxels is a smooth and low frequency signal [19] without going through sharp and abrupt changes. To force the learned atoms to be smooth as well, we included $\mathbf{a}_m^\top\, \mathbf{\Omega}\, \mathbf{a}_m$ into our cost function. This term becomes large if there are sharp changes in the vector $\mathbf{a}_m$, thus increasing α_2 will force the atoms to be smooth.

The cost function in (4) is a penalized low rank-*k* matrix approximation problem and is biconvex. It can be approximately minimized via *k* penalized rank-1 matrix approximations via matrix deflation, i.e, by replacing $\mathbf{G}_0$ in (4) by the residual matrix $\mathbf{G}_m = \mathbf{G}_{m-1} - \mathbf{a}_m\mathbf{b}^m$ with $m = [1, \cdots, k]$. Thus, minimizing (4) is equivalent to minimizing

$$\min \frac{1}{2}\|\mathbf{G}_{m-1} - \mathbf{a}_m\mathbf{b}^m\|_F^2 + \alpha_1\|\mathbf{b}^m\|_1 + \alpha_2\, \mathbf{a}_m^\top\, \mathbf{\Omega}\, \mathbf{a}_m \quad \text{s.t. } \|\mathbf{a}_m\|_2 = 1 \quad (5)$$

We can approx. solve (5) by using alternating minimization, i.e. keeping $\mathbf{b}^m$ fixed, $\mathbf{a}_m$ that minimizes (5) is given by

$$\hat{\mathbf{a}}_m = \arg\min_{\mathbf{a}_m} \frac{1}{2}\ \text{tr}\ (\|\mathbf{G}_{m-1}\|_F^2 - 2\, \mathbf{a}_m^\top\, \mathbf{G}_{m-1}\, \mathbf{b}^{m\top} + \|\mathbf{a}_m\|_2^2\, \|\mathbf{b}^m\|_2^2) + \alpha_2\, \mathbf{a}_m^\top\, \mathbf{\Omega}\, \mathbf{a}_m \quad (6)$$

Taking derivative w.r.t. $\mathbf{a}_m$ and equating to zero leads to the solution given by

$$\hat{\mathbf{a}}_m = (\mathbf{I}\, \|\mathbf{b}^m\|_2^2 + \alpha_2\, \mathbf{\Omega})^{-1}\, \mathbf{G}_{m-1}\mathbf{b}^{m\top} \quad \hat{\mathbf{a}}_m = \hat{\mathbf{a}}_m / \|\hat{\mathbf{a}}_m\|_2 \quad (7)$$

where $\mathbf{I}$ is identity matrix of appropriate size. Similarly, keeping $\mathbf{a}_m$ fixed, the solution that minimizes (5) is derived from

$$\hat{\mathbf{b}}^m = \arg\min_{\mathbf{b}^m} \frac{1}{2}\ \text{tr}\ (\|\mathbf{G}_{m-1}\|_F^2 - 2\, \mathbf{a}_m^\top\, \mathbf{G}_{m-1}\, \mathbf{b}^{m\top} + \|\mathbf{a}_m\|_2^2\, \|\mathbf{b}^m\|_2^2) + \alpha_1\|\mathbf{b}^m\|_1 \quad (8)$$

whose solution is

$$\hat{\mathbf{b}}^m = \text{sgn}(\mathbf{a}_k^\top \mathbf{G}_{m-1}) \odot \max(0, |\mathbf{a}_k^\top \mathbf{G}_{m-1}| - \alpha_1 \mathbf{1}_N) \quad (9)$$

where $\odot$ is the Hadamard product and $\mathbf{1}_N$ is a row vector of dimension N. Equations (7) and (9) are very similar to the power algorithm [18] equations, which if initialized randomly, converge almost surely to a least square rank-1 fit. Our experiments have shown that using only $2 - 3$ iterations of (7) and (9) are enough to get a good solution. The entire procedure is summarized in algorithm 2. Thus in stage two, using algorithm 2, we decompose the matrices $\mathbf{J}$ and $\mathbf{I}_i$ as $\mathbf{J} = \mathbf{A}^J\mathbf{B}^J$ and $\mathbf{I}_i = \mathbf{A}_i^I\mathbf{B}_i^I\ \forall i = [1, \ldots, p]$, where $\mathbf{A}^J \in \mathbb{R}^{np \times k_J}$ (TCs), $\mathbf{B}^J \in \mathbb{R}^{k_J \times N}$ (SMs), $\mathbf{A}_i^I \in \mathbb{R}^{n \times k_I}$ (TCs), and $\mathbf{B}_i^I \in \mathbb{R}^{k_I \times N}$ (SMs). The complete procedure is outlined in algorithm 1.

Algorithm 1: Proposed Data Driven Method

Input: $\mathbf{Y}$, p, r_J, r_I, k_J, k_I, α_1, α_2, $noIt$

Stage 1: $\mathbf{J} \leftarrow \mathbf{0}$, $\mathbf{I} \leftarrow \mathbf{0}$

for $it = 1 : noIt$ **do**

  Compute $\mathbf{X} = \mathbf{Y} - \mathbf{I}$,

  Find $\mathbf{J}$ as best r_J-rank approx. of $\mathbf{X}$ using SVD.

  for $i = 1 : p$ **do**

    Compute $\mathbf{Z}_i = \mathbf{Y}_i - \mathbf{J}_i$,

    Find $\mathbf{I}_i$ as best r_I-rank approx. of $\mathbf{Z}_i$ using SVD.

Stage 2:

Use Algorithm 2 to refine $\mathbf{J}$ into k_J-rank matrix pairs as $\mathbf{J} = \mathbf{A}^J \mathbf{B}^J$,

for $i = 1 : p$ **do**

  Use Algorithm 2 to refine $\mathbf{I}_i$ into k_I-rank matrix pairs as $\mathbf{I}_i = \mathbf{A}_i^I \mathbf{B}_i^I$,

Output: $\mathbf{J}$, $\mathbf{I}$, $\mathbf{A}^J$, $\mathbf{B}^J$, $\mathbf{A}_i^I$, $\mathbf{B}_i^I$

Algorithm 2: Refinement Algorithm

Input: $\mathbf{G}_0$, k, $noIt$, α_1, α_2

Initialize $\mathbf{A}$ and $\mathbf{B}$ from $\mathbf{G}_0$,

for $m = 1 : k$ **do**

  for $it = 1 : noIt$ **do**

    use (7) to get $\mathbf{a}_m$,

    use (9) to get $\mathbf{b}^m$,

  Compute $\mathbf{G}_m = \mathbf{G}_{m-1} - \mathbf{a}_m \mathbf{b}^m$,

  Store the pair as $\mathbf{A}(:, m) = \mathbf{a}_m$ and $\mathbf{B}(m, :) = \mathbf{b}^m$,

Output: $\mathbf{A}$, $\mathbf{B}$

The parameters α_1 and α_2 can be obtained using cross validation or model selection criteria and the parameters $k_J \geq r_J$ and $k_I \geq r_I$. Here r_J and r_I are directly related to how much joint and sub-specific information we want to retain and k_J and k_I control the total number of TC/SM components learnt from $\mathbf{J}$ and $\mathbf{I}$ respectively.

3. EXPERIMENTAL RESULTS

In the following section we have used simulated fMRI datasets to establish the working of the proposed algorithm along with a comparison with CODL [11] algorithm. The proposed algorithm is then validated using motor task fMRI datasets acquired from HCP Q1 release [20]. The details of these experiments are given in their respective sections.

3.1. Simulation Study

In the simulation study, our objective is to show that our proposed algorithm is capable of separating the joint information and sub-specific information from a temporally concatenated multi-subject fMRI datasets efficiently. To do so, we started by generating $p = 6$ fMRI datasets using the publicly available SimTB toolbox [21]. The simulated spatial maps (SM) were (100×100) voxels in size and the time courses (TC) had 150 time points with repetition time TR $= 2$ sec/sample. The single-subject datasets $\mathbf{Y}_i \in \mathbb{R}^{150 \times 10^4}$ were created where each dataset contained the linear combinations of 4 SM/TC pairs, with 3 common (joint info) and 1 unique pair (sub-specific info). We introduced spatial variability in the common SMs by creating random translations ($\mu = 0, \sigma = 2$ voxels) in x and y directions, rotations ($\mu = 0, \sigma = 2.5$ degrees), and spreads ($\mu = 1, \sigma = 0.03$), where μ and σ represent mean and standard deviation respectively. We also introduced temporal variability across subjects in a similar fashion. Each resulting dataset was corrupted by AWGN to make the overall SNR $= \{-10, -15\}$ dB. For comparisons, the ground truth (GT) for common SMs/TCs were generated by taking the mean of each SM/TC common pair for all subjects. All noisy datasets were normalized to make $||\mathbf{Y}_i||_F = 1$, temporally concatenated into a big matrix $\mathbf{Y} \in \mathbb{R}^{900 \times 10^4}$ and passed to the proposed and CODL algorithms for decomposition.

The big data matrix $\mathbf{Y}$ is decomposed by the proposed method into small matrices $(\mathbf{A}^J, \mathbf{B}^J)$ containing joint info and $(\mathbf{A}_i^I, \mathbf{B}_i^I)$ containing sub-specific info. Both stages of the algorithm were iterated 3 times, learning 3 joint and 1 unique component in stage 1 and refining them into 6 joint and 2 unique components in stage 2. The tuning parameters were selected as $\alpha_1 = 0.003$ and $\alpha_2 = 0.1$. For a fair comparison, we used the full datasets instead of temporally reduced ones to compare with CODL [11], which reduces to ODL [17] under this condition. Thus, starting with $\mathbf{Y}$, we used ODL to learn a dictionary of size 900×20 with $\lambda = 0.15$, batch size of 200 and 50 iterations. We experimented with different tuning parameters for both algorithms and selected the ones giving best results in terms of the correlation between recovered sources and respective GT counterparts.

The experiment was repeated 100 times, each time with different datasets, and the most correlated recovered SM/TC w.r.t. the GT were saved. The mean and std dev of Pearson correlation coefficients are given in table 1 where it can

Table 1. Mean and std dev (Pearson correlation) of most correlated TCs and SMs w.r.t. ground truth over 100 trials.

SNR dB	Algorithm	TCs		SMs	
		Mean	STD	Mean	STD
-10	Proposed	**0.99**	**0.01**	**0.89**	**0.05**
	CODL	0.95	0.03	0.79	0.05
-15	Proposed	**0.98**	**0.01**	**0.85**	**0.06**
	CODL	0.87	0.04	0.58	0.21

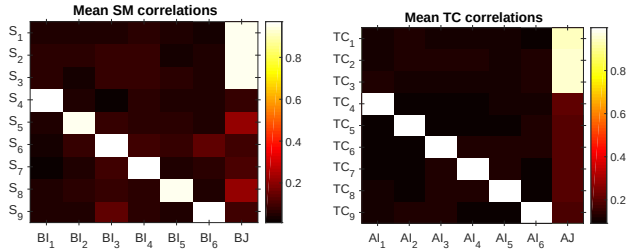


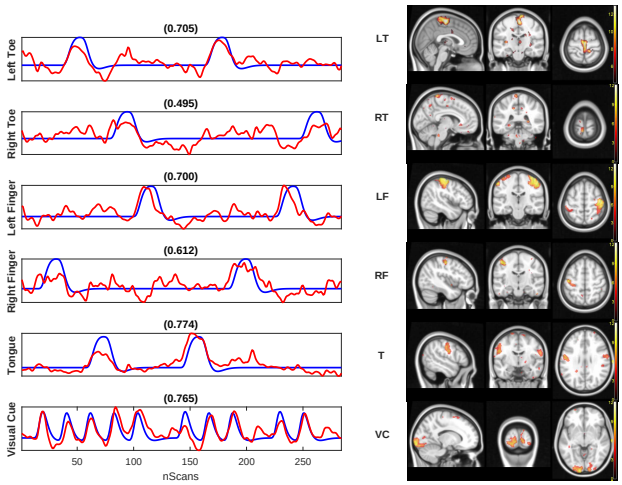
Fig. 1. The mean Ground Truth (GT) SM and TC correlation coefficients over 100 trials with respect to all recovered $\mathbf{A}^J, \mathbf{B}^J, \mathbf{A}_i^I, \mathbf{B}_i^I$ matrices. SNR $= 0$ dB.

be seen that for both noise levels, the proposed algorithm was able to outperform CODL. Furthermore, to highlight the proposed algorithm's ability to separate joint from the sub-specific information, we correlated the GT TCs with the recovered $\mathbf{A}^J$, and $\mathbf{A}_i^I$ and SMs with $\mathbf{B}^J$, and $\mathbf{B}_i^I$ and the results are shown in Fig. 1. Here it can be seen that the first 3 SM/TC pairs have been successfully extracted into the joint info matrices $\mathbf{B}^J$ and $\mathbf{A}^J$ respectively, with the sub-specific information recovered in their respective $\mathbf{B}_i^I$ and $\mathbf{A}_i^I$ matrices.

3.2. Multi-subject task fMRI Analysis

In this section we use 3 subject motor task fMRI datasets acquired from human connectome project (HCP) Q1 release [20] to validate the proposed algorithm. Following a visual cue, each subject was asked to perform 5 tasks, i.e. flexing left toe, right toe, left finger, right finger and tongue movements to map the motor areas of the brain. For experimental setup, spatial and temporal preprocessing details, the reader is referred to [6] section V-B. After preprocessing, voxels outside the brain were removed by using a brain mask, each brain volume was vectorized and placed as rows of $\mathbf{Y}_i$ resulting in a data matrix with size $n \times N$ for each subject. Here $n = 284$ are time points and $N = 283494$ are brain voxels. The datasets were acquired using TR $= 0.72$s. Each column of $\mathbf{Y}_i$ was normalized to zero mean and unit variance followed by normalizing it to get $\|\mathbf{Y}_i\|_F = 10^3$. These matrices were then temporally concatenated ($\mathbf{Y}$) and used by the proposed algorithm to decompose the joint info and sub-specific info into separate matrices.

We iterated both stages of the algorithm 3 times, 50 joint and 10 sub-specific components were kept in stage 1 which were then further decomposed into 80 joint and 20 sub-specific components respectively. The tuning parameter used in stage 2 were set to $\alpha_1 = 0.6$ and $\alpha_2 = 0.2$. Using the timing information of the 5 tasks along with visual cue timings, we generated 6 paradigm time courses (PTC) which were compared with the average TCs recovered in the matrix $\mathbf{A}^J$. The most correlated TCs found are shown in Fig. 2 a). To check for activations, the corresponding rows of $\mathbf{B}^J$ were extracted,



(a) **(b)**

Fig. 2. a) Most correlated average TCs from $\mathbf{A}^J$ (*red*) with their respective PTCs (*blue*) recovered by the proposed algorithm. The corresponding correlation coefficients are given above each TC plot. b) Respective activation maps from $\mathbf{B}^J$.

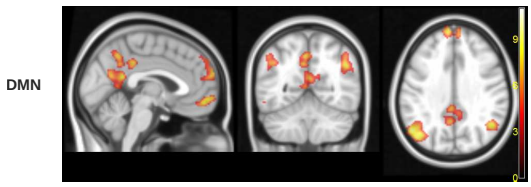


Fig. 3. Default mode network from joint info matrix $\mathbf{B}^J$.

z-scored and thresholded at $p < 0.001$ and are shown in Fig. 2 b). With careful examination, it can be seen that the first 5 activations are tightly localized in the motor cortex area, whereas, the last one (VC) shows activations in the primary visual cortex area of the brain. Upon detail examinations of all recovered activation maps in $\mathbf{B}^J$, we found the default mode network along with a few other well reported resting state networks [22] as well. The recovered DMN is shown in Fig. 3.

4. CONCLUSION

In this paper we proposed a new DL algorithm which can separate the joint and sub-specific information from multi-subject fMRI datasets. The performance of the algorithm was highlighted using a simulation experiment where the proposed algorithm was able to decompose joint SM/TC pairs and sub-specific ones into different matrices with high precision. The results generated are on par with the results recovered by the CODL algorithm [11]. The algorithm was then validated on the experimental task fMRI dataset.

5. REFERENCES

[1] T. H. Vu and V. Monga, "Fast low-rank shared dictionary learning for image classification," *IEEE Transactions on Image Processing*, vol. 26, no. 11, pp. 5160–5175, 2017.

[2] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *Image Processing, IEEE Transactions on*, vol. 15, no. 12, pp. 3736–3745, 2006.

[3] A. K. Seghouane and M. Hanif, "A sequential dictionary learning algorithm with enforced sparsity," in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, 2015, pp. 3876–3880.

[4] M. U. Khalid and A. K. Seghouane, "Improving functional connectivity detection in fMRI by combining sparse dictionary learning and canonical correlation analysis," in *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*, 2013, pp. 286–289.

[5] V. Abolghasemi, S. Ferdowsi, and S. Sanei, "Fast and incoherent dictionary learning algorithms with application to fMRI," *Signal, Image and Video Processing*, vol. 9, no. 1, pp. 147–158, 2015.

[6] A. K. Seghouane and A. Iqbal, "Sequential dictionary learning from correlated data: Application to fMRI data analysis," *IEEE Transactions on Image Processing*, vol. 26, no. 6, pp. 3002–3015, 2017.

[7] A. K. Seghouane and A. Iqbal, "Basis expansions approaches for regularized sequential dictionary learning algorithms with enforced sparsity for fMRI data analysis," *IEEE Transactions on Medical Imaging*, vol. 36, no. 9, pp. 1796–1807, 2017.

[8] A. K. Seghouane and A. Iqbal, "A regularized sequential dictionary learning algorithm for fMRI data analysis," in *2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP)*, 2017, pp. 1–6.

[9] A. K. Seghouane, A. Iqbal, and N. K. Desai, "BSmCCA: A block sparse multiple-set canonical correlation analysis algorithm for multi-subject fMRI data sets," in *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*. 2017, pp. 6324–6328, IEEE.

[10] M. U. Khalid and A. K. Seghouane, "A single svd sparse dictionary learning algorithm for fmri data analysis," in *2014 IEEE Workshop on Statistical Signal Processing (SSP)*, 2014, pp. 65–68.

[11] A. Mensch, G. Varoquaux, and B. Thirion, "Compressed online dictionary learning for fast resting-state fMRI decomposition," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, 2016, pp. 1282–1285.

[12] A. K. Seghouane and A. Iqbal, "CSMSDL: A common sequential dictionary learning algorithm for multi-subject fMRI data sets analysis," in *2017 IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 1–5.

[13] M. U. Khalid and A. K. Seghouane, "Multi-subject fMRI connectivity analysis using sparse dictionary learning and multiset canonical correlation analysis," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, 2015, pp. 683–686.

[14] N. K. Desai, A. K. Seghouane, and M. Palaniswami, "Multisubject fmri data analysis via two dimensional multi-set canonical correlation analysis," in *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*. 2017, pp. 468–471, IEEE.

[15] A. Iqbal, A. K. Seghouane, and T. Adali, "Shared and subject-specific dictionary learning algorithm for multi-subject fMRI data analysis (ShSSDL)," *IEEE Transactions on Biomedical Engineering*, pp. 1–10, 2018.

[16] K. K. Delgado, J. Murray, B. Rao, K. Engan, Te-Won Lee, and T. Sejnowski, "Dictionary learning algorithms for sparse representation," *Neural computation*, vol. 15, no. 2, pp. 349–396, 2003.

[17] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, *Online Learning for Matrix Factorization and Sparse Coding*, pp. 19–60, JMLR.org, 2010.

[18] Gene H Golub and Charles F Van Loan, *Matrix computations*, vol. 3, JHU Press, 2012.

[19] P. Ciuciu, J. B. Poline, G. Marrelec, J. Idier, C. Pallier, and H. Benali, "Unsupervised robust nonparametric estimation of the hemodynamic response function for any fMRI experiment," *IEEE Transactions on Medical Imaging*, vol. 22, no. 10, pp. 1235–1251, 2003.

[20] D. M. Barch, G. C. Burgess, M. P. Harms, et al., "Function in the human connectome: Task-fMRI and individual differences in behavior," *NeuroImage*, vol. 80, pp. 169–189, 2013.

[21] E. B. Erhardt, E. A. Allen, Y. Wei, et al., "SimTB, a simulation toolbox for fMRI data under a model of spatiotemporal separability," *NeuroImage*, vol. 59, no. 4, pp. 4160–4167, 2012.

[22] S. M. Smith, P. T. Fox, K. L. Miller, et al., "Correspondence of the brain's functional architecture during activation and rest," *Proceedings of the National Academy of Sciences*, vol. 106, no. 31, pp. 13040–13045, 2009.