DICTIONARY LEARNING ALGORITHM FOR MULTI-SUBJECT FMRI ANALYSIS VIA TEMPORAL AND SPATIAL CONCATENATION

Asif Iqbal and Abd-Krim Seghouane

Department of Electrical and Electronic Engineering
University of Melbourne, Australia

ABSTRACT

In recent history, dictionary learning (DL) methods have been successfully used for analyzing multi-subject functional magnetic resonance imaging. These algorithms try to learn group-level spatial activation maps (SM) or voxel time courses (TC) from temporally or spatially concatenated fMRI datasets respectively. However, in multi-subject fMRI studies, we are interested in both group-level TCs as well as SMs. In this paper, we propose a DL algorithm which combines temporally and spatially concatenated fMRI datasets to learn not only the shared TC/SM pairs but also the subject-specific ones. We do this by separating group-level information and sub-specific information from each subject fMRI dataset. Performance of the proposed algorithm is illustrated using simulated as well as experimental task fMRI datasets.

Index Terms— functional magnetic resonance imaging (fMRI), dictionary learning, temporal concatenation, spatial concatenation, multi-subject analysis.

1. INTRODUCTION

In recent years, dictionary learning (DL) algorithms have been extensively used in signal and image processing fields. These methods have been applied to problems like face recognition [1, 2], image denoising [3, 4], and fMRI data analysis [5–10]. In DL framework, given a set of training signals, the aim is to learn a dictionary $\mathbf{D}$ that can represent each signal using the linear combination of only a few atoms from the dictionary. The training signals are modeled as $\mathbf{y}_i = \mathbf{D}\mathbf{x}_i + \boldsymbol{\epsilon}_i$, where $\mathbf{y}_i$ is the i^{th} training signal, $\mathbf{D}$ is the dictionary, $\mathbf{x}_i$ is the i^{th} sparse coefficient vector and $\boldsymbol{\epsilon}_i$ is the representation error. Starting with a set of training signals $\mathbf{Y}$, in most cases, the DL problem is solved by alternating between a sparse coding stage followed by dictionary update stage leading to the minimization of a specific cost function.
Under DL formulation, an fMRI dataset $\mathbf{Y}$ is decomposed into a dictionary matrix $\mathbf{D}$ and a sparse coefficient matrix $\mathbf{X}$ such that each voxels' time course (TC) can be approximated using only few atoms from the dictionary. Under such formulations, the learned dictionary atoms represent significant neuronal temporal dynamics and the rows of sparse coefficient matrix represent the respective spatial activation maps (SM) [11]. The DL methods have been extended to analyze multi-subject (MS) [11–14] fMRI datasets as well. These methods generate group level SMs and sub-specific temporal dynamics by using temporally concatenated datasets [13], or learn group-level dynamics and subject-level SMs by using spatial concatenation [12]. However, we might be interested in learning both group-level temporal dynamics and spatial maps which are of particular interest in task-based fMRI studies.
In this paper we present a DL algorithm which decomposes the MS fMRI datasets into a shared dictionary/sparse code pair as well as sub-specific ones leaving us with multiple dictionary/sparse code pairs containing shared as well as unique sources of information about the analyzed fMRI datasets.

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

2. DICTIONARY LEARNING FORMULATION FOR MULTI-SUBJECT FMRI DATASETS

2.1. Dictionary Learning Formulation

Starting with an fMRI dataset $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_N]$, containing N variables (brain voxels) with n observations (time points), where $\mathbf{y}_i \in \mathbb{R}^n$ contains the observations for i^{th} variable. According to the sparse representation theory, all variables in $\mathbf{Y}$ can be compactly represented as $\mathbf{Y} = \mathbf{D}\mathbf{X}$, where $\mathbf{D} \in \mathbb{R}^{n\times K}$ is the dictionary, $\mathbf{X} \in \mathbb{R}^{K\times N}$ is the sparse coefficient matrix. In DL methods, the aim is to learn such a dictionary which makes this compact representation possible. A typical DL problem is formulated as:

$$\min_{\mathbf{D},\mathbf{X}} ||\mathbf{Y} - \mathbf{D}\mathbf{X}||_F^2 \text{ s.t. } \parallel \mathbf{x}_i \parallel_0 \leq s, \ , \|\mathbf{d}_k\|_2 = 1 \qquad (1)$$

where $\mathbf{d}_k$ is the k^{th} column of $\mathbf{D}$, $\parallel . \parallel_F$ is the Frobenius norm, $\parallel . \parallel_0$ is the ℓ_0 quasi-norm, counting the number of nonzero coefficients, $\|.\|_2$ is the ℓ_2 norm, and the sparsity constraint $s \ll K$. The constraint on $\mathbf{D}$ keeps the atoms from getting arbitrarily large which could've lead to small values of $\mathbf{x}_i$ [3]. The resulting $\mathbf{D}$ and $\mathbf{X}$ matrices contain K dense

TCs and K sparse group level SMs respectively [11].

In the next section, we propose a DL algorithm which decomposes the multi-subject fMRI datasets into shared and unique dictionary/sparse code pairs containing shared and sub-specific TCs and SMs respectively.

2.2. Proposed Dictionary Learning Algorithm

In the proposed DL algorithm, our aim is to decompose p subject fMRI datasets in a structured way, i.e. we aim to represent each voxels' time course from $\mathbf{Y}_i$ (i-th subject dataset) as a linear combination of a few atoms from $\mathbf{D}_0$ (shared) and $\mathbf{D}_i$ (sub-specific) dictionaries such that

$$\mathbf{Y}_i \simeq \tilde{\mathbf{D}}_i \tilde{\mathbf{X}}_i = [\mathbf{D}_0, \mathbf{D}_i] \begin{bmatrix} \mathbf{X}_0 \\ \mathbf{X}_i \end{bmatrix} = \mathbf{D}_0\mathbf{X}_0 + \mathbf{D}_i\mathbf{X}_i \quad (2)$$

where the shared dictionary $\mathbf{D}_0$ and sub-specific dictionary $\mathbf{D}_i$ are matrices of size $n \times K_0$ and $n \times K_i$ respectively, $\mathbf{X}_0$ is of size $K_0 \times N$ representing the sparse codes of $\mathbf{Y}_i$ over $\mathbf{D}_0$ and $\mathbf{X}_i$ is of size $K_i \times N$ representing the sparse codes of $\mathbf{Y}_i$ over $\mathbf{D}_i$. As a result of such formulation, we not only require $\mathbf{D}_0/\mathbf{X}_0$ pair to capture the shared information across multi-subject datasets, but also require $\mathbf{D}_i/\mathbf{X}_i$ pair to capture the sub-specific unique information as well. To achieve this goal, we propose to solve the following minimization problem:

$$\begin{aligned} \min_{\tilde{\mathbf{D}}_i, \tilde{\mathbf{X}}_i} \sum_{i=1}^{p} \{ \frac{1}{2} \| \mathbf{Y}_i - \mathbf{D}_0\mathbf{X}_0 - \mathbf{D}_i\mathbf{X}_i \|_F^2 + \frac{\eta}{2} \| \mathbf{D}_i^\top \mathbf{A}_i \|_F^2 \} \\ \text{s.t. } \|\mathbf{x}_i^m\|_0 \leq s_i, \|\mathbf{x}_0^m\|_0 \leq s_0, \|\mathbf{d}_k\|_2 = 1 \\ \forall \quad i = 1, 2, \ldots, p \text{ and } m = 1, 2, \ldots, N \end{aligned} \quad (3)$$

where $\mathbf{x}_i^m$ is the m-th column of $\mathbf{X}_i$, s_i, s_0 are the signal sparsity parameters, η is the incoherence penalty parameter, and p is the number of subjects selected for analysis. The first term in (3) is the representation error term and second is the sub-dictionary incoherence term. Here $\mathbf{A}_i = [\mathbf{D}_0, \mathbf{D}_1, \ldots, \mathbf{D}_{i-1}, \mathbf{D}_{i+1}, \ldots, \mathbf{D}_p]$ is the concatenation of all except currently updating dictionary. The incoherence term in the objective function is included to learn incoherent dictionaries which have been shown to improve the effectiveness of sparse representation [15].

The objective in (3) is non-convex, however, an approx. solution is possible when all but one of the variables are fixed. Thus its minimization can be carried out in an alternating optimization fashion, i.e. fixing $(\mathbf{D}_0, \mathbf{D}_i)$, first optimize for $(\mathbf{X}_0, \mathbf{X}_i)$ followed by updating $(\mathbf{D}_0, \mathbf{D}_i)$ with fix $(\mathbf{X}_0, \mathbf{X}_i)$ repeating till convergence. The details of these minimizations are given in coming sections.

2.2.1. *Sparse Coding Stage*

With dictionaries $(\mathbf{D}_0, \mathbf{D}_i)$ and sub-specific sparse codes $\mathbf{X}_i$ fixed, we first update $\mathbf{X}_0$, by minimizing

$$\hat{\mathbf{X}}_0 = \min_{\mathbf{X}_0} \frac{1}{2} \|\mathbf{E}_{te} - \mathbf{D}_{te}\mathbf{X}_0\|_F^2; \text{ s.t. } \|\mathbf{x}_0^m\|_0 \leq s_0 \quad (4)$$

where $\mathbf{E}_{te} = \frac{1}{\sqrt{p}} \left[\mathbf{E}_1^\top, \mathbf{E}_2^\top, \ldots, \mathbf{E}_p^\top\right]^\top$ is the *temporally concatenated* residual matrix containing subject level residuals $\mathbf{E}_i = \mathbf{Y}_i - \mathbf{D}_i \mathbf{X}_i$ and $\mathbf{D}_{te} \in \mathbb{R}^{np \times K_0}$ contains *temporally concatenated* p copies of the shared dictionary $\mathbf{D}_0$. Here we have included a factor of $1/\sqrt{p}$ to indirectly control the entries of $\mathbf{X}_0$. After $\mathbf{X}_0$ update, we fix it and update $\mathbf{X}_i$ for all p subject by solving:

$$\hat{\mathbf{X}}_i = \min_{\mathbf{X}_i} \frac{1}{2} \|\mathbf{B}_i - \mathbf{D}_i\mathbf{X}_i\|_F^2; \text{ s.t. } \|\mathbf{x}_i^m\|_0 \leq s_i \quad (5)$$

where $\mathbf{B}_i = \mathbf{Y}_i - \mathbf{D}_0\mathbf{X}_0$. Here we opt to use the Orthogonal Matching Pursuit (OMP) algorithm [16] to efficiently solve (4) and (5). In our implementation, we iterated the complete sparse coding stage two times as further iterations did not lead to significant improvement.

2.2.2. *Dictionary Update Stage*

After sparse coding stage, with fix sparse matrices $(\mathbf{X}_0, \mathbf{X}_i)$ and sub-specific dictionaries $\mathbf{D}_i$s, we solve for shared info dictionary $\mathbf{D}_0$ by solving:

$$\hat{\mathbf{D}}_0 = \min_{\mathbf{D}_0} \frac{1}{2} \|\mathbf{E}_{sp} - \mathbf{D}_0\mathbf{X}_{sp}\|_F^2 + \frac{\eta}{2} \|\mathbf{D}_0^\top \mathbf{A}_0\|_F^2 \text{ s.t. } \|\mathbf{d}_k\|_2 = 1 \quad (6)$$

where $\mathbf{E}_{sp} = [\mathbf{E}_1, \mathbf{E}_2, \ldots, \mathbf{E}_p]$ is the *spatially concatenated* residual matrix containing subject level residuals $\mathbf{E}_i = \mathbf{Y}_i - \mathbf{D}_i \mathbf{X}_i$ and $\mathbf{X}_{sp}$ contains *spatially concatenated* p copies of the shared sparse code matrix $\mathbf{X}_0 \in \mathbb{R}^{K_0 \times pN}$. After $\mathbf{D}_0$ update, keeping it fixed, the p sub-specific $\mathbf{D}_i$s are found by solving:

$$\hat{\mathbf{D}}_i = \min_{\mathbf{D}_i} \frac{1}{2} \|\mathbf{B}_i - \mathbf{D}_i\mathbf{X}_i\|_F^2 + \frac{\eta}{2} \|\mathbf{D}_i^\top \mathbf{A}_i\|_F^2 \text{ s.t. } \|\mathbf{d}_k\|_2 = 1 \quad (7)$$

where $\mathbf{B}_i = \mathbf{Y}_i - \mathbf{D}_0\mathbf{X}_0$. Both (6) and (7) are constrained convex quadratic optimization problems and we adopt the projected gradient method [17] to sequentially update all atoms of the dictionaries. Thus, the update rule for (6) is given by:

$$\begin{aligned} \mathbf{d}_{k_0}^{t+0.5} = \mathbf{d}_{k_0}^t + \mu_{k_0} \left(\mathbf{E}_{sp} \mathbf{x}_{sp}^{k_0 \top} - \mathbf{D}_0 \mathbf{X}_{sp} \mathbf{x}_{sp}^{k_0 \top} - \eta \mathbf{A}_0 \mathbf{A}_0^\top \mathbf{d}_{k_0}^t \right) \\ \mathbf{d}_{k_0}^{t+1} = \mathbf{d}_{k_0}^{t+0.5} / \|\mathbf{d}_{k_0}^{t+0.5}\|_2 \end{aligned} \quad (8)$$

where $\mathbf{d}_{k_0}^t$ is k_0-th atom of $\mathbf{D}_0$ at t-th iteration, $\mathbf{x}_{sp}^{k_0}$ is the k_0-th row of $\mathbf{X}_{sp}$, $\mathbf{A}_0 = [\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_p]$, and $\mu_{k_0} = 1/\|\mathbf{x}_{sp}^{k_0}\mathbf{x}_{sp}^{k_0 \top}\|_2$ is the step-size parameter. Similarly the update rule for (7) is given by:

$$\begin{aligned} \mathbf{d}_{k_i}^{t+0.5} = \mathbf{d}_{k_i}^t + \mu_{k_i} \left(\mathbf{B}_i \mathbf{x}_i^{k_i \top} - \mathbf{D}_i \mathbf{X}_i \mathbf{x}_i^{k_i \top} - \eta \mathbf{A}_i \mathbf{A}_i^\top \mathbf{d}_{k_i}^t \right) \\ \mathbf{d}_{k_i}^{t+1} = \mathbf{d}_{k_i}^{t+0.5} / \|\mathbf{d}_{k_i}^{t+0.5}\|_2 \end{aligned} \quad (9)$$

where $\mathbf{A}_i$ is the concatenation of all but currently updating dictionary and $\mu_{k_i} = 1/\|\mathbf{x}_i^{k_i}\mathbf{x}_i^{k_i \top}\|_2$. The complete learning method is summarized in algorithm 1.

Algorithm 1: The proposed algorithm

Input: fMRI datasets $\mathbf{Y}_i$, K_0, K_i, s_0, s_i, η, and *noIt*

Initialization: Initialize $\mathbf{D}_0$, and $\mathbf{D}_i$ with ℓ_2-normalized random vectors.

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

- Fix $\mathbf{D}_0$, $\mathbf{D}_i$ and use OMP to solve (4) for $\mathbf{X}_0$ and (5) for $\mathbf{X}_i\ \forall\, i = 1, \ldots, p$.
- Fix $\mathbf{X}_0$, $\mathbf{X}_i$ and sequentially update dictionaries $\mathbf{D}_0$ using (8) and $\mathbf{D}_i$ using (9) $\forall\, i = 1, \ldots, p$.

Output: $\mathbf{D}_0, \mathbf{X}_0, \mathbf{D}_i, \mathbf{X}_i$

Table 1. Mean, median, and standard deviation of most correlated TCs and SMs w.r.t. GrTr over 100 trials.

SNR dB	Algorithm	TCs			SMs		
		Mean	Median	STD	Mean	Median	STD
-10	Proposed	**0.98**	**0.98**	**0.02**	**0.87**	**0.88**	**0.05**
	CODL	0.95	0.95	0.03	0.79	0.82	0.14
-15	Proposed	**0.92**	**0.96**	**0.08**	**0.69**	**0.66**	**0.18**
	CODL	0.68	0.68	0.23	0.44	0.27	0.34

3. EXPERIMENTAL RESULTS

In this section we use simulated and real multi-subject task-fMRI datasets to illustrate our proposed algorithms' ability to learn shared (group-level) info and unique (sub-specific) info separately while providing performance comparison with CODL [13] algorithm as well.

3.1. Simulation Study

In simulation study, we generate $p = 6$ fMRI datasets using the publicly available SimTB toolbox [18]. We simulated spatial maps (SM) of size (100×100) voxels and their respective time courses (TC) having 150 time points with a repetition time TR $= 2$ secs. Each subject dataset $\mathbf{Y}_i \in \mathbb{R}^{150 \times 10^4}$ consisted of 4 TC/SM pairs with 1 sub-specific pair (unique info) and 3 group-level pairs (shared info). To simulate spatial variability in the group-level SMs, we introduced random translations ($\mu = 0, \sigma = 2$ voxels) in x and y directions, scaling ($\mu = 1, \sigma = 0.03$), and rotations ($\mu = 0, \sigma = 2.5$ degrees), where μ and σ are the mean and standard deviation of a Gaussian distribution. Similarly temporal variability across subjects was also introduced. A sample of simulated TC/SM pairs is shown in Fig 1 a). Here, the pairs $(1, 2, 3)$ are present in all datasets while $(4 - 9)$ are the unique sub-specific pairs. AWGN noise corresponding to SNR $= \{-10, -15\}$ dB was introduced into the datasets and resulting noisy datasets were passed to the proposed and CODL algorithm for decomposition.

The datasets were decomposed by the proposed algorithm into shared info pair $(\mathbf{D}_0/\mathbf{X}_0)$ and sub-specific pairs $(\mathbf{D}_i/\mathbf{X}_i)$. The learning parameters $(K_0, K_i, s_0, s_i, \eta)$ were set to $(10, 5, 2, 1, 10)$ and the algorithm was iterated $noIt = 20$ times. For a fair comparison, we opted not to temporally reduce the datasets to compare with CODL [13], which essentially reduces to ODL [19] when applied to full datasets. Thus we used the temporally concatenated datasets $\mathbf{Y}$ as input to the ODL to learn a dictionary of size 900×20 with $\lambda = 0.15$, batch size of 200 and 50 iterations. We tried multiple parameters for both algorithms and selected the ones resulting in best overall performance in terms of correlation between recovered TC/SMs and their ground truth (GrTr) counterparts.

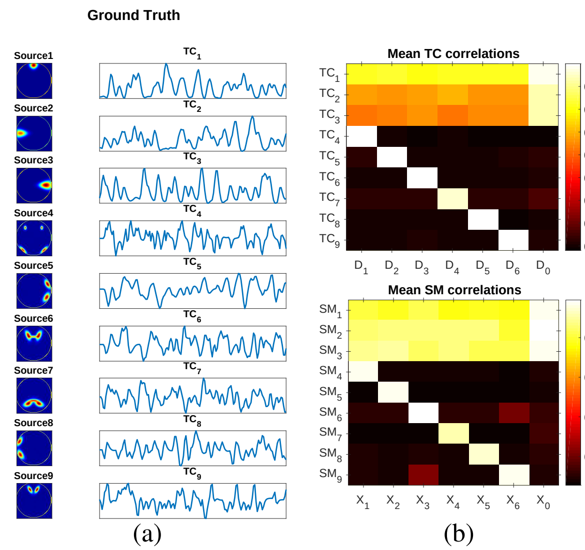
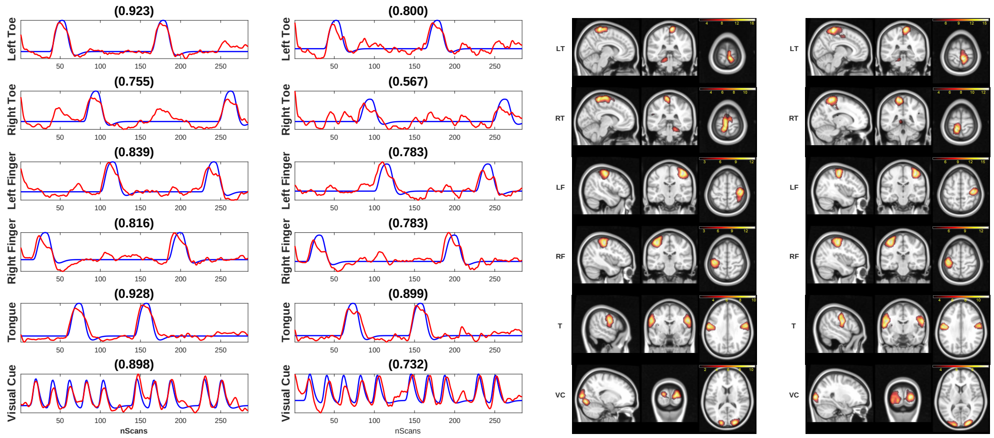


Fig. 1. a) The simulated ground truth TC/SMs and their b) mean correlation coefficients w.r.t. $\mathbf{D}_0, \mathbf{D}_i$ and $\mathbf{X}_0, \mathbf{X}_i$ (b) over 100 trials for SNR $= 0$ dB.

The experiment was repeated 100 times with different datasets and the highest correlation coefficients of GrTr w.r.t. recovered TC/SM were saved. The overall results are given in table 1 where it is evident that the proposed algorithm was able to recover the underlying sources very effectively as compared to CODL. Now to check whether the TC/SM pairs have been separated into their respective $\mathbf{D}/\mathbf{X}$ pairs, we correlated the GrTr TCs with the recovered $\mathbf{D}_0$ and $\mathbf{D}_i$s and SMs with $\mathbf{X}_0$ and $\mathbf{X}_i$s for every trials and have presented the mean results as correlation matrices in Fig. 1 b). Here it can be seen that the sub-specific TC/SM pairs $(4 - 9)$ have been successfully recovered only in their respective $\mathbf{D}_i/\mathbf{X}_i$ pairs. Whereas, the most highly correlated shared pairs are found in $\mathbf{D}_0/\mathbf{X}_0$ pairs and the subject variability has been captured in $\mathbf{D}_i/\mathbf{X}_i$ pairs as well.

3.2. Multi-subject task fMRI Analysis

In this section we have used $p = 7$ subject motor task fMRI datasets from human connectome project (HCP) Q1 release [20] for the analysis. During image acquisition, following a visual cue, each subject was asked to move their left toe,



(a) **(b)** **(c)** **(d)**

Fig. 2. Most correlated $\mathbf{D}_0$ atoms (*red*) with their respective PTCs (*blue*) recovered by a) proposed algorithm, b) CODL. The corresponding correlation coefficients are also given inside parenthesis above each TC plot. Most informative population level z-scored ($p < 0.001$) activation maps recovered by c) proposed algorithm and d) CODL.

right toe, left finger, right finger, and tongue to map the motor areas of the brain. Repetition time of TR $= 0.72$s was used for image acquisition. Reader is referred to [8] section V-B for experimental setup and preprocessing details. After preprocessing, a brain mask was used to remove data outside the brain. Each brain volume was vectorized and placed as rows of a data matrix $\mathbf{Y}_i \in \mathbb{R}^{n\times N}, i \in \{1,\dots,p\}$ where $n = 284$ are the time points and $N = 283494$ are the total brain voxels.

The proposed algorithm was used to decompose the multi-subject datasets into a shared info $\mathbf{D}_0/\mathbf{X}_0$ pair and sub-specific ones. The learning parameters $(K_0, K_i, s_0, s_i, \eta)$ were set to $(30, 20, 1, 2, 500)$. The algorithm was iterated 10 times as the relative change in the dictionary atoms was very small after 6 iterations. Similar to the simulation section, CODL algorithm was used to learn a dictionary $\mathbf{D}_{ODL} \in \mathbb{R}^{np\times 70}$ with $\lambda = 6$ [13], batch size of 2000 and 100 iterations.

Using the task timing information, we generated 6 paradigm time courses (PTCs) by convolving the canonical HRF with the boxcar signals. As all subjects were performing the same task, we expect to find the corresponding TCs and respective SMs in the shared info pair $\mathbf{D}_0/\mathbf{X}_0$. Thus we correlated the 6 PTCs with all recovered dictionaries and found that this was indeed the case. These most correlated TCs w.r.t. PTCs from $\mathbf{D}_0$ and $\mathbf{D}_{ODL}$ are shown in Fig. 2 a) and b), followed by thr respective activation maps from $\mathbf{X}_0$ in Fig. 2 c) and the maps recovered by CODL are shown in Fig. 2 d). The activation maps recovered by both algorithm are very similar and are correctly localized in the Sensorimotor cortex and Visual cortex areas of the brain. However the recovered TCs extracted by proposed algorithm are much better matched to the PTCs as compared to CODL's recovered TCs.

To analyze the sub-specific dictionaries, we took a slightly different approach. In [21], the authors have shown the existence of ten well-established resting state networks (RSNs) [22] in the experimental task datasets. Although these RSNs might be present in all subject datasets, but their respective TCs are not bound to be similar. Based on this intuition, we expect to find these RSNs in the sub-specific $\mathbf{X}_i$s. Using the RSN templates from [22], we correlated them with every $\mathbf{X}_i$ and stored the most correlated ones for each subject. Using the most correlated RSNs from each subject, we generated 10 average RSN maps. For comparison, the correlation of these averaged RSN maps and those recovered by CODL w.r.t. the RSN templates are given in table 2. Here it can be seen that the RSN maps recovered by the proposed algorithm show better correlation w.r.t. the templates as compared to CODL.

Table 2. Correlation coefficients of most correlated spatial maps w.r.t. the RSN templates as recovered by proposed algorithm and CODL.

RSN	1	2	3	4	5	6	7	8	9	10	Mean
Proposed	0.55	0.48	**0.57**	**0.60**	**0.41**	**0.44**	**0.47**	**0.41**	**0.55**	**0.57**	**0.51**
CODL	**0.72**	**0.71**	0.43	0.47	0.31	0.34	0.36	0.31	0.49	0.37	0.45

4. CONCLUSION

In this paper we proposed a new dictionary learning algorithm which can separate the shared and sub-specific information from multi-subject fMRI datasets. A simulation example was used to highlight performance of the proposed algorithm followed by comparison with CODL. Multi-subject task fMRI datasets were also used to show that the algorithm was able to separate the shared from the sub-specific information with high precision.

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] A. Iqbal and A. K. Seghouane, "An approach for sequential dictionary learning in nonuniform noise," in *2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, 2017, pp. 1–5.

[3] 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.

[4] 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.

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

[6] 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.

[7] 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)*, pp. 683–686.

[8] 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.

[9] 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.

[10] 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.

[11] J. Lv, X. Jiang, X. Li, et al., "Sparse representation of whole-brain fMRI signals for identification of functional networks," *Medical Image Analysis*, vol. 20, no. 1, pp. 112–134, 2015.

[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] 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)*, pp. 1282–1285.

[14] 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.

[15] S. Ubaru, A. K. Seghouane, and Y. Saad, "Improving the incoherence of a learned dictionary via rank shrinkage," *Neural computation*, vol. 29, no. 1, pp. 263–285, 2017.

[16] J. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *Information Theory, IEEE Transactions on*, vol. 53, no. 12, pp. 4655–4666, 2007.

[17] P. H. Calamai and J. J. More, "Projected gradient methods for linearly constrained problems," *Mathematical programming*, vol. 39, no. 1, pp. 93–116, 1987.

[18] 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.

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

[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] S. Zhao, J. Han, J. Lv, et al., "Supervised dictionary learning for inferring concurrent brain networks," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 2036–2045, 2015.

[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.