CSMSDL: A COMMON SEQUENTIAL DICTIONARY LEARNING ALGORITHM FOR MULTI-SUBJECT FMRI DATA SETS ANALYSIS

Abd-Krim Seghouane and Asif Iqbal

Department of Electrical and Electronic Engineering
Melbourne School of Engineering
The University of Melbourne, Australia

ABSTRACT

Sequential dictionary learning algorithms has gained widespread acceptance in functional magnetic resonance imaging (fMRI) data analysis. However, many problems in fMRI data analysis involve the analysis of multiple-subject fMRI data sets and the existing algorithms do not extend naturally to this case. In this paper we propose an algorithm dedicated to multiple-subject fMRI data analysis. The algorithm is named SMSDL for sequential multi-subject dictionary learning and differs from existing dictionary learning algorithms in its dictionary update stage. This algorithm is derived by using a variation of the power algorithm in the dictionary update stage to extract the common information among the multiple-subject fMRI data sets. The results of the proposed dictionary learning algorithm is a set of time courses which are common to the whole group of subjects and an individual spatial response pattern for each of the subjects in the group. The performance of the proposed algorithm are illustrated through a simulation and an application on real fMRI datasets.

Index Terms— Functional magnetic resonance imaging (fMRI), multi-subjects, dictionary learning, sparsity.

1. INTRODUCTION

Sparse modeling offers an attractive framework for signal processing and computer visions where parsimonious representation is considered advantageous. The assumption underlying this modeling framework is that natural signals or images can be expressed as a sparse linear combination of vectors of an appropriately chosen basis. The basis can be predefined and based on a known transform such as the Fourier transform, the discrete cosine transform (DCT) or the discrete wavelet transform (DWT) or it can be learned by being trained on a specific class of signals or images. Dictionary learning algorithm allow to obtain a sparse modeling using this latter approach. They consists of two main tasks a sparse coding and dictionary tuning. The former strives to find a minimal decomposition given a redundant basis or dictionary while the latter aims to learn the overcomplete dictionary given a set of training samples.

Dictionary learning is an increasingly used data-driven approach to analyze fMRI data [1, 2, 3, 4, 5, 6, 7, 8]. Given an fMRI data matrix $\mathbf{Y}$ formed by vectorizing each time series observed in every voxel creating a matrix $n \times N$ where n is the number of time points and N the number of voxels ($\approx 10,000 - 100,000$) [9], dictionary learning approaches, like the widely used general linear model (GLM), assume a linear multivariable model for the fMRI data $\mathbf{Y} = \mathbf{DX}$ where the matrix $\mathbf{D}$ is the dictionary and $\mathbf{X}$ is a sparse matrix of latent variables. The difference between the two approaches is that the matrix $\mathbf{D}$ is estimated in dictionary learning approaches whereas the design matrix is specified in the GLM. As a data-driven approach to model the observed data, dictionary learning algorithms are suitable for the analysis of fMRI data as they minimize the assumptions on the underlying structure of the problem by decomposing the observed data using on a factor model and and a specific constraint.

Dictionary learning algorithms have mainly been applied so far to single subject fMRI data. Many problems in fMRI data analysis however involve the study of data collected from multiple subjects [10, 11, 12]. The application of dictionary learning algorithms to multiple subjects is not straightforward due to the estimation of a different dictionary for each subject. In this article, we consider the development of a dictionary learning algorithms for simultaneously analyzing fMRI data sets from a group of subjects. The proposed algorithm is derived using spatial concatenation as in [13] and differs from existing dictionary learning algorithms in its dictionary update stage. This stage is derived by using a dedicated variation of the power algorithm [14] instead of the SVD for rank one matrix approximation to extract the common information among the multiple-subject fMRI data sets. This algorithm is named CSMSDL for common sequential multi-subject dictionary learning. As a result the proposed algorithm yield a set of temporal patterns (time courses) which are common to the whole group and, for each time course, a separate spatial map for each subject.

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

2. BACKGROUND

Given a set of signals $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_N]$, a learned dictionary is a collection of vectors or atoms $\mathbf{d}_k$, $k = 1, ..K$ that can be used for optimal linear representation. Usually the objective is to find a linear representation for the set of signals $\mathbf{Y}$

$$\{\mathbf{D}, \mathbf{X}\} = \arg\min_{\mathbf{D},\mathbf{X}} \parallel \mathbf{Y} - \mathbf{D}\mathbf{X} \parallel_F^2$$

where $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, ..., \mathbf{d}_K]$, that makes the total representation error as small as possible. This optimization problem is ill-posed unless extra constraints are imposed on the dictionary $\mathbf{D}$ and coefficients $\mathbf{X}$. The common constraint on $\mathbf{X}$, is that each column of $\mathbf{X}$ is sparse where the name "sparse codes". Let the sparse coefficient vectors $\mathbf{x}_i$, $i = 1, ..., N$ constitute the columns of the matrix $\mathbf{X}$, with this constraint, the above objective can be re-stated as the minimization problem

$$\min_{\mathbf{D},\mathbf{X}} ||\mathbf{Y} - \mathbf{D}\mathbf{X}||_F^2 \text{ s.t. } \parallel \mathbf{x}_i \parallel_0 \leq s, \; \forall\, 1 \leq i \leq N, \quad (1)$$

where the $\mathbf{x}_i$'s are the column vectors of $\mathbf{X}$, $\parallel . \parallel_F$ is the Frobenius norm, $\parallel . \parallel_0$ is the l_0 quasi-norm, which counts the number of nonzero coefficients and $s \ll K$. To prevent $\mathbf{D}$ from being arbitrarily large and therefore have arbitrarily small values of $\mathbf{x}_i$, it is common to constrain $\mathbf{D}$ to belong to set $D = \{\mathbf{D} \in \mathbf{R}^{n \times K} : \; \parallel \mathbf{d}_k \parallel_2 = 1 \quad \forall k\}$, where $\parallel . \parallel_2$ is the l_2 norm and $\mathbf{d}_k$ is the k^{th} column of $\mathbf{D}$. Finding the optimal s corresponds to a problem of model order selection that can be resolved using a univariate linear model section criterion [15][16][17]. The generally used optimization strategy, not necessarily leading to a global minimum consists in splitting the problem into two stages which are alternately solved within an iterative loop. These two stages are, first, the sparse coding stage, where $\mathbf{D}$ is fixed and the sparse coefficient vectors are found by solving

$$\hat{\mathbf{x}}_i = \arg\min_{x_i} \parallel \mathbf{y}_i - \mathbf{D}\mathbf{x}_i \parallel^2; \quad (2)$$

$$\textit{subject to} \parallel \mathbf{x}_i \parallel_0 \leq s \quad i = 1, ..., N.$$

In practice, the sparse coding stage is often approximately solved by using either a greedy pursuit or convex relaxation approach [18]. The dictionary update stage where $\mathbf{X}$ is fixed and $\mathbf{D}$ is derived by solving

$$\mathbf{D} = \arg\min_{D} \parallel \mathbf{Y} - \mathbf{D}\mathbf{X} \parallel_F^2 \quad (3)$$

followed by normalizing its columns constitutes the second stage [19].

Sequential update methods break the global minimization (3) into K sequential minimization problems. In the method proposed in [20] each column $\mathbf{d}_k$ of $\mathbf{D}$ and its corresponding row of coefficients $\mathbf{x}_k^{row}$ are updated based on a rank-1 matrix approximation of the error for all the signals when $\mathbf{d}_k\mathbf{x}_k^{row}$ is removed

$$\{\mathbf{d}_k, \mathbf{x}_k\} = \arg\min_{\mathbf{d}_k, \mathbf{x}_k^{row}} \|\mathbf{E}_k - \mathbf{d}_k\mathbf{x}_k^{row}\|_F^2. \quad (4)$$

where $\mathbf{E}_k = \mathbf{Y} - \sum_{i=1, i\neq k}^{K} \mathbf{d}_i\mathbf{x}_i^{row}$. To avoid the loss of sparsity in $\mathbf{x}_k^{row}$ that will be created by the direct application of the SVD on $\mathbf{E}_k$, the SVD of $\mathbf{E}_k^R = \mathbf{E}_k\mathbf{I}_{w_k}$ is taken instead and only the the none zeros elements of $\mathbf{x}_k^{row}\mathbf{I}_{w_k}$, where $w_k = \{i | 1 \leq i \leq N; \mathbf{x}_k^{row}(i) \neq 0\}$ and $\mathbf{I}_{w_k}$ the $N \times |w_k|$ submatrix of the $N \times N$ identity matrix obtained by retaining only those columns whose index numbers are in w_k, are considered.

In [21, 22, 23] an alternative dictionary update stage that leads to a substantial improved performance dictionary learning algorithm was proposed. Within this dictionary update stage it is proposed to re-update all the entries of $\mathbf{x}_k^{row}$ and the sparsity pattern information. The resulting algorithm is a variant of the power method where the estimates $\mathbf{d}_k$ and $\mathbf{x}_k^{row}$ are given by

$$\mathbf{d}_k = \frac{\mathbf{E}_k\mathbf{x}_k^{row^\top}}{||\mathbf{E}_k\mathbf{x}_k^{row^\top}||_2}. \quad (5)$$

$$\mathbf{x}_k^{row} = \text{sgn}(\mathbf{d}_k^\top\mathbf{E}_k) \circ \left(|\mathbf{d}_k^\top\mathbf{E}_k| - \frac{\alpha}{2}\mathbf{1}_{(N)}^\top\right)_+ \quad (6)$$

where $\circ$, $| \; . \; |$, sgn(.), $(.)_+$ define the Hadamard product, the component-wise absolute value, the component-wise sign and the component-wise $\max(0, x)$ respectively. The $\mathbf{1}_N$ is a vector of ones of size N. Below we propose an extension of the algorithm proposed in [21] to learn a single dictionary dictionary for multi-subjects fMRI data analysis.

3. SEQUENTIAL DICTIONARY LEARNING FOR MULTI-SUBJECT FMRI DATA ANALYSIS

Direct application of the dictionary learning algorithm described above on each fMRI data set separately will generate multiple dictionaries and avoid taking into consideration the common information across the multiple data sets while performing the analysis. For the development of the proposed dictionary learning algorithm to simultaneously analyzing fMRI data from a group of p subjects we consider the spatial concatenation of the fMRI data sets $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_p]$ of size $n \times pN$ where n is the number of time points, N the number of voxels and p the number of subjects [13] and the model

$$[\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_p] \simeq \mathbf{D}[\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_p] \quad (7)$$

where the dictionary $\mathbf{D}$ is matrix of size $n \times K$ that contains the corresponding set of common time courses across all the subjects and the sparse codes $\mathbf{X}$ is a matrix of size $K \times pN$ that contains spatial maps associated to each of the temporal time course. Instead of iterating (7) and (6) in the dictionary update to generate $\mathbf{D}$ we propose the following dictionary update stage to account for the common information among the

p fMRI data sets.
The proposed extension of [21] for the analysis of multisubjects data sets is obtained by the application of this algorithm to spatially concatenated fMRI data sets $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_p]$. The sparse coding stage is obtained in this case by solving (2) for the $\mathbf{x}_i'$s, $i = 1, ..., N$ whereas the dictionary update stage is obtained from

$$\begin{aligned} \{\mathbf{d}_k, \mathbf{x}_k^{row}\} &= \arg\min_{\mathbf{d}_k, \mathbf{x}_k^{row}} \|\mathbf{E}_k - \mathbf{d}_k \mathbf{x}_k^{row}\|_F^2 \\ &+ \alpha \|\mathbf{x}_k^{row}\|_1 \end{aligned} \quad (8)$$

where $\mathbf{E}_k = \left[\mathbf{E}_{k_1} \mathbf{E}_{k_2} ... \mathbf{E}_{k_p}\right]$ with $\mathbf{E}_{k_j} = \mathbf{Y}_j - \sum_{i=1, i\neq k}^{K} \mathbf{d}_i \mathbf{x}_{i_j}^{row}$ and $\mathbf{x}_k^{row} = \left[\mathbf{x}_{k_1}^{row}, \mathbf{x}_{k_2}^{row}, ..., \mathbf{x}_{k_p}^{row}\right]$ corresponds to the spatial concatenation of the p subjects spatial maps associated with the atoms $\mathbf{d}_k$ and α is a non-negative penalty parameter controlling the amount of sparsity in $\mathbf{x}_k^{row}$ (increasing α increases the amount of sparsity in $\mathbf{x}_k^{row}$).
For fixed $\mathbf{d}_k$ and $\|\mathbf{d}_k\|_2 = 1$, the $\mathbf{x}_{k_j}^{row}$ that minimizes (8) is derived from

$$\mathbf{x}_{k_j}^{row} = \arg\min_{\mathbf{x}_k^{row}} \|\mathbf{x}_{k_j}^{row}\|^2 + \alpha \|\mathbf{x}_{k_j}^{row}\|_1 - 2\mathbf{d}_k^\top \mathbf{E}_{k_j} \mathbf{x}_{k_j}^{row\top} \quad (9)$$

which gives

$$\mathbf{x}_{k_j}^{row} = \mathrm{sgn}(\mathbf{d}_k^\top \mathbf{E}_{k_j}) \circ \left(|\mathbf{d}_k^\top \mathbf{E}_{k_j}| - \frac{\alpha}{2} \mathbf{1}_{(N)}^\top\right)_+. \quad (10)$$

For fixed $\mathbf{x}_k^{row}$, the $\mathbf{d}_k$ that minimizes (8) is derived from

$$\begin{aligned} \mathbf{d}_k &= \arg\min_{\mathbf{d}_k} -2\mathbf{d}_k^\top \mathbf{E}_k \mathbf{x}_k^{row\top} + \|\mathbf{d}_k\|^2 . \|\mathbf{x}_k^{row}\|^2 \\ &= \arg\min_{\mathbf{d}_k} -2\mathbf{d}_k^\top \sum_{j=1}^{p} \mathbf{E}_{k_j} \mathbf{x}_{k_j}^{row\top} \\ &+ \|\mathbf{d}_k\|^2 . \|\mathbf{x}_k^{row}\|^2 \end{aligned} \quad (11)$$

which with the constraint $\|\mathbf{d}_k\|_2 = 1$ gives

$$\mathbf{d}_k = \frac{\sum_{j=1}^{p} \mathbf{E}_{k_j} \mathbf{x}_{k_j}^{row\top}}{|| \sum_{j=1}^{p} \mathbf{E}_{k_j} \mathbf{x}_{k_j}^{row\top} ||_2}. \quad (12)$$

The dictionary learning algorithm obtained using this approach is presented in table 1. Its dictionary update stage is a variation of the power algorithm for computing the SVD where the updates of $\mathbf{d}_k$ and $\mathbf{x}_k^{row}$ are found by iterating (10) and (12) until convergence. A model selection criterion or cross-validation can be used for the selection of the penalty parameter α. From the simulation we found that a single iteration of (10) and (12) was enough instead of iterating until convergence and it is the approach adopted in the paper. The computation cost of this iteration is $O(pnN)$. From (11) we can further observe that the common atom $\mathbf{d}_k$ is proportional to the average to the individual atoms $\mathbf{d}_{k_j}, j = 1, ..., p$ associated to the data sets $\mathbf{Y}_1, ..., \mathbf{Y}_p$.

Table 1. Stepwise description of the proposed sequential multi-subject dictionary learning algorithm by average

Algorithm CSMSDL
Given: $\mathbf{Y}_1, \mathbf{Y}_2$,..., and $\mathbf{Y}_p$, $\mathbf{D}_{ini}$, $\mathbf{s}$, α, $iter$ and *J*.
Set $\mathbf{D} = \mathbf{D}_{ini}$
For *i=1 to J*
1: *Sparse Coding Stage:*
Find sparse coefficients X, by approx. solving
$\hat{\mathbf{x}}_i = \arg\min_{x_i} \parallel \mathbf{y}_i - \mathbf{D}\mathbf{x}_i \parallel^2$;
subject to $\parallel \mathbf{x}_i \parallel_0 \leq s \quad i = 1, ..., pN$
2: Dictionary Update Stage:
For each column $k = 1, 2, ..., K$ in $\mathbf{D}$,
2.a: Compute the error matrices using
$\mathbf{E}_{k_j} = \mathbf{Y}_j - \sum_{i=1, i\neq k}^{K} \mathbf{d}_i \mathbf{x}_{i_j}^{row}, j = 1, ..., p$
2.b: Construct the error matrix $\mathbf{E}_k$ as
$\mathbf{E}_k = \left[\mathbf{E}_{k_1} \mathbf{E}_{k_2} ... \mathbf{E}_{k_p}\right]$
While $\|\mathbf{d}_k^{iter} - \mathbf{d}_k^{iter+1}\|_2^2 \geq \varepsilon$ iterate
2.c: Update the p block rows $\mathbf{x}_{k_j}^{row}$'s and its sparsity
$\mathbf{x}_{k_j}^{row} = \mathrm{sgn}(\mathbf{d}_k^\top \mathbf{E}_{k_j}) . \left(\|\mathbf{d}_k^\top \mathbf{E}_{k_j}\| - \frac{\alpha}{2} \mathbf{1}_{(N)}^\top\right)_+$
2.d: Update the dictionary atom $\mathbf{d}_k$ using
$\mathbf{d}_k^{iter} = \frac{\sum_{j=1}^{p} \mathbf{E}_{k_j} \mathbf{x}_{k_j}^{row\top}}{\|\| \sum_{j=1}^{p} \mathbf{E}_{k_j} \mathbf{x}_{k_j}^{row\top} \|\|_2}$
iter = iter+1
end.
Output: **D**,**X**

4. EXPERIMENTAL EVALUATION

In this section, we perform two experiments to evaluate the performance of proposed algorithm in the recovery of underlying common signal dynamics. These experiments include a simulation study and an application on task fMRI dataset acquired from [1]. Details of the experiments are given in the following sections.

4.1. Simulation Study

To demonstrate the ability of proposed algorithm in recovering the underlying common signal from a mixture of signals, we generated fMRI datasets for three different subjects. Three sinusoidal signals along with a block-design paradigm signal of length 220 were used as temporal dynamics and four box signals were used as their activation patterns as shown in Fig. 1. The visual patterns of size 10×10 were created with amplitudes of 1 at $\{2, 3, 4\} \times \{2, 3, 4\}$ for pattern *A*, $\{2, 3, 4\} \times \{7, 8, 9\}$ for pattern *B*, $\{7, 8, 9\} \times \{2, 4, 6\}$ for pattern *C*, $\{7, 8, 9\} \times \{7, 8, 9\}$ for pattern *D* and zero otherwise. The corresponding time-courses T_1, T_2, T_4 were sinusoids of frequencies $\{5, 1.5, 8\}$Hz respectively. The signal T_3 was generated by convolving the canonical HRF and a box car

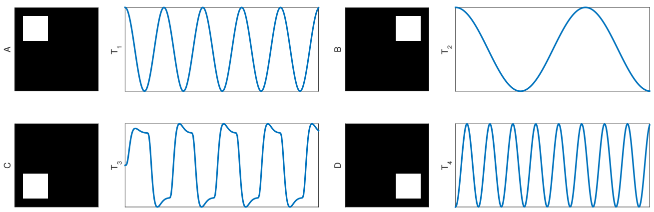

Fig. 1. Activations patterns and the respective time-courses.

signal. Three subject datasets of size 220×100 were created as $Sub_1\{A+C\}$, $Sub_2\{B+C\}$, and $Sub_3\{D+C\}$ where the combination $\{C, T_3\}$ was present in all three datasets. These three datasets were concatenated along columns to make a signal matrix $\mathbf{Y} \in \mathbb{R}^{220\times300}$, additive white Gaussian noise was used to corrupt the signals with corresponding signal to noise ratios of $\in \{0, -5, -10\}dB$. This training matrix $\mathbf{Y}$ was decomposed by our dictionary learning algorithm which was iterated 15 times to learn a dictionary of size $\mathbf{D} \in \mathbb{R}^{220\times1}$ with sparsity $s = 1$[1], and $\alpha = 0.9$. The sparse coding was performed using OMP [18]. Starting with a random dictionary, we repeated the experiment 100 times and have presented the mean correlation of learned atom and the respective sparse code vectors with the ground truth (T_3, C) in table 2. As seen in table 2, our proposed algorithm was able to recover the common signals from the mixture with high accuracy even in the presence of high noise.

Table 2. Correlation of recovered time-series and spatial maps with ground truth averaged over 100 iterations

SNR dB		Sub_1	Sub_2	Sub_3
0	T_3	0.996	0.996	0.996
	C	0.978	0.995	0.996
-5	T_3	0.963	0.963	0.963
	C	0.971	0.984	0.985
-10	T_3	0.741	0.741	0.741
	C	0.866	0.869	0.872

4.2. Block-Paradigm RFT fMRI dataset

In this section, we used three block-design right finger tapping datasets acquired from [1] to validate our proposed dictionary learning algorithms' ability in the recovery of common underlying time-series from the datasets. For detailed dataset and fMRI preprocessing information, the reader is referred to [1]. The BOLD time-series data was extracted from every voxel and arranged into a matrix $\mathbf{Y}_p \in \mathbb{R}^{n\times N}$, where $n = 156$ is the number of time points, N is the number of voxels, and p is the subject number. We concatenated all three datasets

[1]In our experimental setup, there is only one common time-series, thus learning only one dictionary atom.

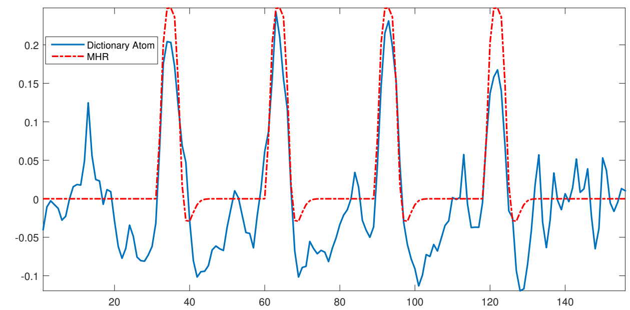

Fig. 2. Most correlated dictionary atom w.r.t. MHR

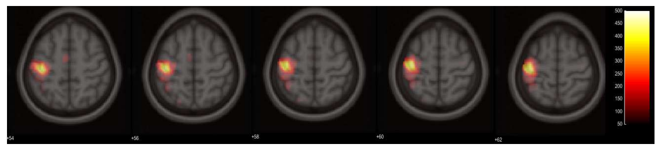

Fig. 3. Average F-statistics activation maps *(5 contiguous slices)* for right finger tapping task stimulus random field correction $p < 0.0001$.

along columns to make a matrix $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3]$. The data matrix $\mathbf{Y}$ was down-sampled by a factor of 8 in the spatial direction in order to reduce computation time during the dictionary learning stage [1]. Starting with a random dictionary $\mathbf{D} \in \mathbb{R}^{156\times60}$, we iterated the dictionary learning algorithm at max 20 times or when the difference between dictionaries became smaller than ϵ[2]. The sparse coding was performed using correlation based thresholding [1] with sparsity $s = 3$ found using MDL criterion [1]. The most correlated learned dictionary atom w.r.t the modeled hemodynamic response function (MHR), shown in Fig. 2, was used to generate the F-statistics maps for all three subject datasets. The average map at a random field correction *$p < 0.0001$* was calculated and five consecutive axial slices of the motor area are shown in Fig. 3 showing that our proposed algorithm was able to correctly localize the common activation maps in the brain motor area.

5. CONCLUSION

A new dictionary learning algorithm for simultaneously analyzing fMRI data sets from a group of subjects was proposed. Instead of the sets of time courses which would have been obtained by directly applying a dictionary learning algorithm on each fMRI data set, the proposed algorithm generates a single common set of time courses and the associated subjects spatial responses pattern. It is obtained by spatially concatenating the fMRI data sets and using a variation of the power algorithm in the dictionary update stage to extract the common information. The effectiveness of the proposed algorithm was tested both on simulated and real fMRI data sets.

[2]$||D^{i+1} - D^i||_F / ||D^i||_F < \epsilon$, where $\epsilon = 0.01$

6. REFERENCES

[1] K. Lee, S. K. Tak, and J. C. Yee, "A data driven sparse GLM for fMRI analysis using sparse dictionary learning and MDL criterion," *IEEE Transactions on Medical Imaging*, vol. 30, pp. 1176–1089, 2011.

[2] S. Zhao et al., "Supervised dictionary learning for inferring concurrent brain networks," *IEEE Transactions on Medical Imaging*, vol. 34, pp. 2036–2045, 2015.

[3] V. Abolghasemi, S. Ferdowsi, and S. Sanei, "Fast and incoherent dictionary learning algorithms with application to fMRI," *Signal, Image and Video Processing*, vol. 9, pp. 147–158, 2013.

[4] M. U. Khalid and A. K. Seghouane, "A single SVD sparse dictionary learning algorithm for fMRI data analysis," *In Proceedings of IEEE International Workshop on Statistical signal Processing*, pp. 65–68, 2014.

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

[6] M. U. Khalid and A. K. Seghouane, "Constrained maximum likelihood based efficient dictionary learning for fMRI analysis," *In Proceedings of IEEE International Symposium on Biomedical Imaging*, pp. 45–48, 2014.

[7] A. K. Seghouane and A. Iqbal, "Sequential dictionary learning from correlated data: Application to fMRI data analysis," *IEEE Transactions on Image Processing*, pp. 3002–3015, 2017.

[8] M. Ramezani, K. Marble, H. Tang, I. S. Johnsrude, and P. Abolmaesumi, "Joint sparse representation of brain activity patterns in multi-task fMRI data," *IEEE Transactions on Medical Imaging*, vol. 34, pp. 2–12, 2015.

[9] A. K. Seghouane and Y. Saad, "Prewhitening high dimensional fmri data sets without eigendecomposition," *Neural Computation*, vol. 26, pp. 907–919, 2014.

[10] V. D. Calhoun and T. Adali, "Multisubject independent component analysis of fMRI: A decade of intrinsic networks, default mode, and neurodiagnostic discovery," *IEEE Reviews in Biomedical Engineering*, vol. 5, pp. 60–73, 2012.

[11] A. K. Seghouane, A. Iqbal, and N. Desai, "BSmCCA: A block sparse multiple-set canonical correlation analysis algorithm for multi-subject fmri data sets," *In Proceedings of IEEE International Conference on Acoustic, Speech and Signal Processing, ICASSP*, pp. 6324–6328, 2017.

[12] V. D. Calhoun, T. Adali, G.D. Pearlson, and J. J. Pekar, "A method for making group inferences from functional MRI data using independent component analysis," *Human Brain Mapping*, vol. 14, pp. 140–151, 2001.

[13] M. Svensen, F. Kruggel, and H. Benali, "ICA of fMRI group study data," *Neuroimage*, vol. 16, pp. 551–563, 2002.

[14] G. H. Golub and C. f. Van Loan, *Matrix Computations*, Johns Hopkins, 1996.

[15] A. K. Seghouane and M. Bekara, "A small sample model selection criterion based on the kullback symmetric divergence," *IEEE Transactions on Signal Processing*, vol. 52, pp. 3314–3323, 2004.

[16] A. K. Seghouane and S. I. Amari, "The AIC criterion and symmetrizing the kullback-leibler divergence," *IEEE Transactions on Neural Networks*, vol. 18, pp. 97–106, 2007.

[17] A. K. Seghouane, "Asymptotic bootstrap corrections of AIC for linear regression models," *Signal Processing*, vol. 90, pp. 217–224, 2010.

[18] J. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proceedings of the IEEE*, vol. 98, pp. 948–958, 2010.

[19] M. Hanif and A. K. Seghouane, "Maximum likelihood orthogonal dictionary learning," *IEEE Workshop on Statistical Signal Processing (SSP)*, pp. 141–144, 2014.

[20] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An Anlgorithm for desiging overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, pp. 4311–4322, 2006.

[21] A. K. Seghouane and M. Hanif, "A sequential dictionary learning algorithm with enforced sparsity," *IEEE International Conference on Acoustic Speech and signal Processing, ICASSP*, pp. 3876–3880, 2015.

[22] A. K. Seghouane and A. Iqbal, "Basis expansion approaches for regularized sequential dictionary learning algorithms with enforced sparsity for fMRI data analysis," *IEEE Transactions on Medical Imaging*, pp. 1–12, 2017.

[23] M. Sadeghi, M. Babaie-Zadeh, and C. Jutten, "Dictionary learning for sparse representation: A novel approach," *IEEE Signal Processing Letters*, pp. 1195–1198, 2013.