

BSMCCA: A BLOCK SPARSE MULTIPLE-SET CANONICAL CORRELATION ANALYSIS ALGORITHM FOR MULTI-SUBJECT FMRI DATA SETS

Abd-Krim Seghouane, Asif Iqbal and Nandakishor Desai

Department of EEE
Melbourne School of Engineering
University of Melbourne, Australia

ABSTRACT

Multiple-set canonical correlation analysis (mCCA) is a generalization of canonical correlation analysis (CCA) to three or more sets of variables. It aims to study the relationships between several sets of variables and it subsumes a number of interesting multivariate data analysis techniques as special cases. The quality and interpretability of the mCCA components are likely to be affected by the usefulness and relevance of each set of variables. Therefore, it is an important issue to identify each set of significant variables that are active in the relationships between sets. In this paper mCCA is extended to address the issue of variable set selection. Specifically a block sparse multiple set canonical correlation analysis (BSmCCA) algorithm is proposed to combine mCCA with ℓ_2 -norm type penalty in a unified framework. Within this framework sets of variables that are not necessarily relevant are removed. This makes BSmCCA a flexible method for analyzing for Multi-Subject functional magnetic resonance imaging (fMRI) data sets. The performances of the proposed BSmCCA algorithm are illustrated through on block design paradigm finger tapping fMRI datasets.

Index Terms— Multiple-set canonical correlation analysis, fMRI, penalized rank one approximation, variable set selection.

1. INTRODUCTION

A number of different approaches have been proposed for the study of brain function through the analysis functional magnetic resonance imaging (fMRI) data sets. These approaches can be broadly divided into two main classes: model-based or data-driven. Model-based methods through the general linear model (GLM) [1] have been widely used. They use the a priori knowledge about the properties of the data; i.e.; the hemodynamic response function (HRF) and the experimental paradigm; i.e.; the stimulus function, to investigate the goodness-of-fit of the model and make inferences about regional brain activities. The use of these methods can however

be limited for a number of reasons among them the absence of experimental paradigm; when studying resting state or naturalistic paradigms such as continuous listening or watching a movie.

Data-driven methods have also been successfully applied to fMRI data analysis. Among the justification for their suitability for fMRI data analysis is the minimization of the assumptions on the underlying structure of the problem. These methods mainly try to decompose the observed data based on a factor model and a specific constraint. Different constraints have led to different data-driven methods. The maximum variance constraint has led to principal component analysis (PCA) [2], the independence constraint has led to spatial ICA (sICA, for the format of the data described above) and temporal ICA (tICA) [3] and the sparsity constraint has led to dictionary learning [4].

The maximum correlation constraint which leads to canonical correlation analysis (CCA) [5] has also been successfully used for fMRI data analysis. It has for example been used to find latent sources in single subject fMRI data by taking advantage of the spatial or temporal autocorrelation in the data [6] or improve the specificity and sensitivity of dictionary learning methods for fMRI by accounting for the autocorrelation in the fMRI signals [7, 8]. Its extension to multiple data sets, termed multiset canonical correlation analysis (mCCA) [9] has also successfully been used in association with other methods for the analysis of multiple fMRI data sets. It has for example been successfully used in conjunction with dictionary learning for multi-subject fMRI data analysis in [10] and in conjunction with ICA to maintain the correspondence of the source estimates across different subjects in [11, 12, 13]. When working with multi-subjects data sets, the general canonical components obtained using the standard mCCA involves all the individual subjects data sets as mCCA doesn't perform group variable selection. This makes it difficult to interpret these components without using subjective judgment. To ease this drawback of mCCA, a new method called block sparse mCCA (BSmCCA) is introduced in this paper. This method can be seen as a generalization of sparse canonical correlation analysis to three or more sets of variables.

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

BsmCCA combines the power of multi-block data analysis of mCCA and the interpretability of group variables selection [14][15]. Considering that some data sets or group of variables may be irrelevant or useless, the objective of BsmCCA is to find linear combinations of block variables (block components) such that the block components that are assumed to be connected are highly correlated.

2. BACKGROUND: CCA AND MCCA

2.1. Canonical correlation analysis: CCA

CCA is a standard approach for studying the relationships between two sets of random variables. Given two random vectors $\mathbf{x} = (x_1, \dots, x_p)^\top$ and $\mathbf{y} = (y_1, \dots, y_q)$ and suppose n i.i.d samples of \mathbf{x} and \mathbf{y} denoted by $\mathbf{X} \in R^{n \times p}$ and $\mathbf{Y} \in R^{n \times q}$, respectively, have been collected. Assuming that the columns of both \mathbf{X} and \mathbf{Y} have been centered and scaled, CCA aims to find two directions of projection $\mathbf{a}_1 \in R^p$ and $\mathbf{b}_1 \in R^q$ so that

$$(\mathbf{a}_1, \mathbf{b}_1) = \arg \min_{\mathbf{a}_1, \mathbf{b}_1} \mathbf{a}_1^\top \Sigma_{xy} \mathbf{b}_1$$

$$\text{subject to } \mathbf{a}_1^\top \Sigma_{xx} \mathbf{a}_1 = 1 \quad \text{and} \quad \mathbf{b}_1^\top \Sigma_{yy} \mathbf{b}_1 = 1 \quad (1)$$

where Σ_{xx} , Σ_{yy} and Σ_{xy} are covariance and cross-covariance matrices. The vectors \mathbf{a}_1 and \mathbf{b}_1 are called the first pair of canonical vectors while the variables $\mathbf{a}_1^\top \mathbf{X}$, $\mathbf{b}_1^\top \mathbf{Y}$ are called the first pair of canonical components and $\rho_1 = \mathbf{a}_1^\top \Sigma_{xy} \mathbf{b}_1$ is referred to as the first canonical correlation. Given $d \leq \min(\text{rank}(\mathbf{X}, \mathbf{Y}))$, d of such canonical vectors $\mathbf{A} \in R^{p \times d}$ and $\mathbf{B} \in R^{q \times d}$ are obtained using a deflationary approach with the following additional constraints, first

$$\mathbf{A}^\top \Sigma_{xx} \mathbf{A} = \mathbf{B}^\top \Sigma_{yy} \mathbf{B} = \mathbf{I}_d \quad (2)$$

which traduces the none correlation of the canonical components, these components are zero mean and unit variance. Second

$$\mathbf{A}^\top \Sigma_{xy} \mathbf{B} = \mathbf{R} = \text{diag}(\rho_1, \dots, \rho_d) \quad (3)$$

which traduces the correlation of the corresponding canonical components only. The closed form solutions for \mathbf{A} and \mathbf{B} obtained as the solution of a constrained optimization problem using Lagrange multipliers results in the first d eigenvectors of $\Sigma_{xx}^{-1} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx}$ and $\Sigma_{yy}^{-1} \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy}$ respectively. CCA results in vectors \mathbf{a} and \mathbf{b} that are not sparse, and these vectors are not unique if p and q exceeds n . In applications where p and q are large, one possible interest is in finding projections vectors that leads to high correlation but that are also sparse [16][17][18]. This is realized by imposing a sparsity constraint to (1).

2.2. Multiset Canonical correlation analysis: mCCA

CCA between two sets of random variables can be extended to three or more sets of variables in different ways [9]. In an extension of the CCA principle where correlation between two

canonical components is maximized, mCCA generalizes the CCA principle to optimize an objective function of the correlations between pairs of canonical components associated to the different sets of random variables with respect to the canonical vectors so that an overall correlation is maximized. Given n realizations of J sets of p_j dimensional zero mean random vectors \mathbf{x}_j so that \mathbf{X}_j is of size $n \times p_j$, the canonical vectors $\mathbf{a}_1, \dots, \mathbf{a}_J$ are determined as

$$\mathbf{a}_1, \dots, \mathbf{a}_J = \arg \max_{\mathbf{a}_1, \dots, \mathbf{a}_J} \sum_{j,k=1}^J g(\mathbf{a}_j^\top \Sigma_{jk} \mathbf{a}_k) \quad (4)$$

In the optimization (4), the definition of the function g leads to specific methods defined in [9]. For example, the function $g(x) = x$ corresponds to the sum of correlation method (SUMCOR) and the function $g(x) = x^2$ corresponds to the sum of squared correlations method (SSQCOR). Furthermore, these functions are not constrained, but several natural choices for constraints under which to carry out the optimizations (4) can be taken, among them $\mathbf{a}_i^\top \mathbf{a}_i = 1$, $\sum_{i=1}^J \mathbf{a}_i^\top \mathbf{a}_i = 1$ and $\mathbf{a}_i^\top \Sigma_{ii} \mathbf{a}_i = 1$. For the subsequent stages, the canonical vectors are further constrained to be uncorrelated to the ones generated in the previous stages. In all cases defined in [9], mCCA reduces to CCA when $J = 2$. Similar to the two sets case, the linear combinations $\mathbf{u}_j = \mathbf{X}_j \mathbf{a}_j$, $j = 1, \dots, J$ represent the canonical correlation components associated to the canonical vectors $\mathbf{a}_1, \dots, \mathbf{a}_J$. In what follows the focus is put on the SUMCOR method.

3. BLOCK SPARSE MCCA

3.1. mCCA as a rank one matrix approximation

In what follows all Σ_{jj} are assumed to be full rank and the data considered are the whitened data sets $\mathbf{X}_j = \mathbf{X}_j \Sigma_{jj}^{-1/2}$ so that there are no difference between the constraints $\sum_{i=1}^J \mathbf{a}_i^\top \mathbf{a}_i = 1$ and $\sum_{i=1}^J \mathbf{a}_i^\top \Sigma_{ii} \mathbf{a}_i = 1$. This constraint is the natural extension from the two sets case and is the only one considered in this paper. Let \mathbf{X} denote the $n \times p$, where $p = \sum_{j=1}^J p_j$ row block matrix, $\mathbf{X} = [\mathbf{X}_1, \dots, \mathbf{X}_J]$.

Proposition 1: Let the $p \times d$ matrix \mathbf{A} be partitioned conformably with the partition of \mathbf{X} , that is, $\mathbf{A} = [\mathbf{A}_1^\top, \dots, \mathbf{A}_J^\top]^\top$, where \mathbf{A}_j is a $p_j \times d$ matrix. In SUMCOR, \mathbf{A} comprises the first d right singular vectors associated to the d largest singular values of \mathbf{X} .

Proof: Under the SUMCOR objective and the constraint $\sum_{i=1}^J \mathbf{a}_i^\top \mathbf{a}_i = 1$, the first set of canonical vectors is obtained as the solution of the optimization

$$\mathbf{a}_1, \dots, \mathbf{a}_J = \arg \max_{\mathbf{a}_1, \dots, \mathbf{a}_J} \sum_{j,k=1}^J \mathbf{a}_j^\top \Sigma_{jk} \mathbf{a}_k - \lambda \left(\sum_{i=1}^J \mathbf{a}_i^\top \mathbf{a}_i - 1 \right)$$

By setting the derivative with respect \mathbf{a}_j to zero we get

$$\sum_{k=1}^J \Sigma_{jk} \mathbf{a}_k = \lambda \mathbf{a}_j, \quad j = 1, \dots, J$$

or

$$\mathbf{X}^\top \mathbf{X} \mathbf{a} = \lambda \mathbf{a}, \quad \mathbf{a} = (\mathbf{a}_1, \dots, \mathbf{a}_J).$$

since $\Sigma_{jk} = \mathbf{X}_j^\top \mathbf{X}_k$. Therefore \mathbf{a} is the first eigenvector of $\mathbf{X}^\top \mathbf{X}$ and thus the first left singular vector of \mathbf{X} . Adding the orthogonality constraint between subsequent sets of canonical vectors, we have

$$\mathbf{A} = \arg \max_{\mathbf{A}} \text{tr} \left(\mathbf{A}^\top \mathbf{X}^\top \mathbf{X} \mathbf{A} \right)$$

subject to the restriction $\mathbf{A}^\top \mathbf{A} = \mathbf{I}_d$. Furthermore we can easily see that the left singular vector is given by

$$\mathbf{u} = \sum_{j=1}^J \mathbf{X}_j \mathbf{a}_j$$

and is also of unit norm as the whitened data sets are considered.

The SUMCOR problem can then be reformulated as the following rank one matrix approximation optimization problem

$$\min_{\mathbf{u}, \tilde{\mathbf{a}}} \|\mathbf{X} - \mathbf{u} \tilde{\mathbf{a}}^\top\|_F^2 \quad (5)$$

where $\tilde{\mathbf{a}} = \lambda \mathbf{a}$. The subsequent pairs $(\mathbf{u}_k, \lambda_k \mathbf{a}_k)$, $1 < k \leq J$, provide best rank one approximations of the corresponding residual matrices or rank one matrix deflation. For example, $\lambda_2 \mathbf{u}_2 \mathbf{a}_2^\top$ is the best rank one approximation of $\mathbf{X} - \lambda_1 \mathbf{u}_1 \mathbf{a}_1^\top$.

3.2. Block sparse mCCA via penalization rank one matrix approximation

The key idea of the proposed block sparse mCCA algorithm is the observation that the optimization (5) is a least squares problem. For fixed \mathbf{u} , the optimal $\tilde{\mathbf{a}}$ is the least squares coefficient vector of regressing the columns of \mathbf{X} on \mathbf{u} . Introducing block sparsity on $\tilde{\mathbf{a}}$ in such a context is the familiar group variable selection problem in least squares regression. A group sparsity on $\tilde{\mathbf{a}}$ has two main advantages. First, the canonical correlation components are more interpretable as the irrelevant or negligible $\tilde{\mathbf{a}}_i$'s will not appear in $\tilde{\mathbf{a}}$. Second the canonical correlation component \mathbf{u} will not lose much in terms of the sum of correlation it explains.

To achieve group sparsity on $\tilde{\mathbf{a}}$, we propose to use the ℓ_2 -norm type penalty [14][15] in (5) to promote shrinkage and sparsity of the groups $\tilde{\mathbf{a}}_i$'s

$$\min_{\mathbf{u}, \tilde{\mathbf{a}}} \frac{1}{2} \|\mathbf{X} - \mathbf{u} \tilde{\mathbf{a}}^\top\|_F^2 + \alpha \sum_{i=1}^J \sqrt{p_i} \|\tilde{\mathbf{a}}_i\|_2 \quad (6)$$

where α is the tuning parameter that can be adjusted using a model selection criterion [19, 20, 21, 22]. The minimization of (6) with respect to \mathbf{u} and $\tilde{\mathbf{a}}$ under the constraint $\|\mathbf{u}\|_2 = 1$ can be obtained via an iterative algorithm. Consider first the problem of optimizing over \mathbf{u} for a fixed $\tilde{\mathbf{a}}$.

Proposition 2: For a fixed $\tilde{\mathbf{a}}$, the \mathbf{u} that minimizes (6) and verifies $\|\mathbf{u}\|_2 = 1$ is $\mathbf{u} = \mathbf{X} \tilde{\mathbf{a}} / \|\mathbf{X} \tilde{\mathbf{a}}\|$.

Proof: From (6) we have

$$\begin{aligned} \mathbf{u} &= \arg \min_{\mathbf{u}} \frac{1}{2} \|\mathbf{X} - \mathbf{u} \tilde{\mathbf{a}}^\top\|_F^2 + \alpha \sum_{i=1}^J \sqrt{p_i} \|\tilde{\mathbf{a}}_i\|_2 \\ &\propto \arg \min_{\mathbf{u}} -\mathbf{u}^\top \mathbf{X} \tilde{\mathbf{a}} + \frac{1}{2} \mathbf{u}^\top \mathbf{u} \tilde{\mathbf{a}}^\top \tilde{\mathbf{a}} \end{aligned} \quad (7)$$

Proposition 2 is obtained by deriving (7) with respect to \mathbf{u} and imposing the norm one constraint.

Second the problem of optimizing over $\tilde{\mathbf{a}}$ for fixed \mathbf{u} is considered.

Proposition 3: For a fixed \mathbf{u} , the $\tilde{\mathbf{a}}$ that minimizes (6) is obtained by iteratively applying

$$\tilde{\mathbf{a}}_j = \left(1 - \frac{\alpha \sqrt{p_j}}{\|\mathbf{u}^\top \mathbf{X}_j\|_2} \right)_+ \mathbf{u}^\top \mathbf{X}_j \quad (8)$$

where $(x)_+$ is $\max(x, 0)$ to $\tilde{\mathbf{a}}_j$, $j = 1, \dots, J$.

Proof: From (6) we have

$$\begin{aligned} \tilde{\mathbf{a}}_j &= \arg \min_{\tilde{\mathbf{a}}_j} \frac{1}{2} \|\mathbf{X} - \mathbf{u} \tilde{\mathbf{a}}^\top\|_F^2 + \alpha \sum_{i=1}^J \sqrt{p_i} \|\tilde{\mathbf{a}}_i\|_2 \\ &\propto \arg \min_{\tilde{\mathbf{a}}_j} -\mathbf{u}^\top \sum_{i=1, i \neq j}^J \mathbf{X}_i \tilde{\mathbf{a}}_i - \mathbf{u}^\top \mathbf{X}_j \tilde{\mathbf{a}}_j + \alpha \sqrt{p_j} \|\tilde{\mathbf{a}}_j\|_2 \\ &+ \frac{1}{2} \|\mathbf{u}\|_2^2 \sum_{i=1, i \neq j}^J \tilde{\mathbf{a}}_i^\top \tilde{\mathbf{a}}_i + \frac{1}{2} \|\mathbf{u}\|_2^2 \tilde{\mathbf{a}}_j^\top \tilde{\mathbf{a}}_j \\ &+ \alpha \sum_{i=1, j \neq i}^J \sqrt{p_i} \|\tilde{\mathbf{a}}_i\|_2. \end{aligned} \quad (9)$$

By deriving (9) with respect to $\tilde{\mathbf{a}}_j$, and accounting for the norm one of \mathbf{u} we easily obtain the necessary and sufficient condition for $\tilde{\mathbf{a}}$ to be a solution (6)

$$\tilde{\mathbf{a}}_j + \alpha \sqrt{p_j} \frac{\tilde{\mathbf{a}}_j}{\|\tilde{\mathbf{a}}_j\|_2} = \mathbf{u}^\top \mathbf{X}_j \quad \forall \tilde{\mathbf{a}}_j \neq 0 \quad (10)$$

$$\|\mathbf{u}^\top \mathbf{X}_j\|_2 \leq \alpha \sqrt{p_j} \quad \forall \tilde{\mathbf{a}}_j = 0 \quad (11)$$

It can easily be verified that the solution to expressions (10) and (11) is (8).

The above derivations leads to the following iterative procedure for minimizing (6).

BSmCCA Algorithm:

1. Initialize: apply the SVD to \mathbf{X} to obtain an initial \mathbf{u} and $\tilde{\mathbf{a}} = \lambda \mathbf{a}$ as indicated in (5).
2. Update $\tilde{\mathbf{a}}$: While $\|\tilde{\mathbf{a}}_{iter} - \tilde{\mathbf{a}}_{(iter-1)}\| > \epsilon$ iterate (8) for $j = 1 : J$
3. Update \mathbf{u} using $\mathbf{u} = \mathbf{X}\tilde{\mathbf{a}} / \|\mathbf{X}\tilde{\mathbf{a}}\|$.
4. Repeat 2 and 3 until convergence.
5. Standardize $\tilde{\mathbf{a}} = \tilde{\mathbf{a}} / \|\tilde{\mathbf{a}}\|$ to obtain λ .

Setting $\alpha = 0$ in the *BSmCCA* above algorithm reduces step 2 to $\tilde{\mathbf{a}}$ to the minimizer of (5) and therefore the normal SUMCOR algorithm. The computation cost of each step of this algorithm is $O(np)$. The BSmCCA algorithm is particularly suited when $p > n$.

4. EXPERIMENTAL RESULTS

For experimental verification of our method, we have used three subject fMRI datasets from a block-paradigm right finger tapping (RFT) task from [4]. During the image acquisition, subjects were asked to perform right finger tapping task for 15 sec followed by a 72 sec resting period. This task was repeated 4 times. For details of image acquisition process, reader is referred to [4].

Preprocessing steps were done using SPM-12¹ in Matlab which included head motion correction, coregistration, normalization to standard MNI template, resampled to $2 \times 2 \times 2 \text{ mm}^3$ voxels and spatial smoothing was performed using a $8 \times 8 \times 8 \text{ mm}^3$ full-width at half-maximum (FWHM) Gaussian kernel. To improve SNR and the low frequency scanner drifts, we used a discrete cosine transform (DCT) with a cutoff frequency of (1/128) Hz. High frequency noise was removed by smoothing the time courses using 1.5 s FWHM Gaussian Kernel. The dataset for all subjects was spatially down sampled 8 times to reduce the computation time for the algorithm.

After the pre-processing, the fMRI volumes are reshaped into matrices, where each row corresponds to one voxel and the columns correspond to different time instants. The data are then centered and whitened using PCA. To demonstrate the novelty of BSmCCA, we simulate a noisy fMRI data-set and use it in the analysis. Thus, we get 4 data-sets X_1, \dots, X_4 of size $n \times p_j (j = 1, \dots, 4)$. We then carry out BSmCCA to obtain the canonical vectors a_1, \dots, a_4 , corresponding to the 4 datasets. By selecting an appropriate value of λ , the transform corresponding to X_4 reduces to zero. This implies that the fourth data-set is an out-lier and hence, is removed from the analysis by reducing its canonical vector to zero. For the above data-set, we use $\alpha = 2$.

After the BSmCCA, we combine the canonical vectors of all the data-sets and compute PCA, and select its first compo-

¹<http://www.fil.ion.ucl.ac.uk/spm/>

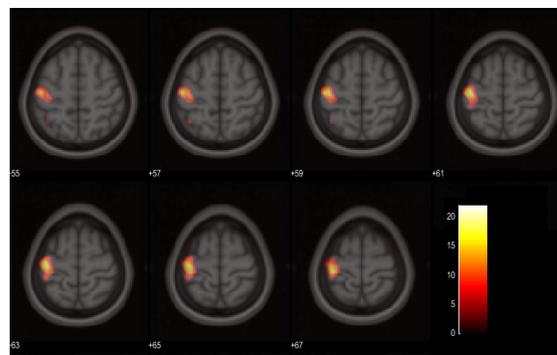


Fig. 1. The group activation map for block-paradigm RFT data using BSmCCA

nent to generate the canonical vector that represents all the data-sets. This is then used to project all the data and construct a mean group activation map that represents the group of patients. From the figure 1, it can be observed that the activation maps contain strong activation in the motor cortex area as also seen in [23, 24, 25]. Thus, the BSmCCA,

- Successfully removes the noisy patient from the analysis, by reducing its canonical vector to zero.
- Constructs a mean activation map that represents the group of patients, excluding noise.

5. CONCLUSION

A new multiple-set canonical correlation analysis (mCCA) method is proposed in this paper. The proposed method is based on the SUMCOR method for mCCA and allows the identification of significant sets of variable that are active in the relationships between sets of variables. The algorithm is derived by establishing the link between the set of canonical vectors and rank one matrix approximation. The selection of the most significant sets of variables is obtained via penalized rank one matrix approximation where the sum of ℓ_2 -norm of variables sets penalty is used to shrink the irrelevant sets of variables to zero. While only the SUMCOR methods is addressed in this paper, the proposed method can easily be extended to the SSQCOR method. Note that the proposed algorithm can also be seen as a block sparse principal component analysis (PCA) algorithm.

6. REFERENCES

- [1] K. Friston, J. Ashburner, S. Keibel, T. Nichols, and W. E. Penny, *Statistical Parametric Mapping: The Analysis of functional Brain Images*, New York Academic, 2006.

- [2] A. Andersen, D. Gash, and M. Avion, "Principal component analysis of the dynamic response measured by fMRI: a generalized linear system framework," *Magnetic Resonance Imaging*, vol. 17, pp. 795–815, 1999.
- [3] M. McKeown and T. Sejnowski, "Independent component analysis of fMRI data: examining the assumptions," *Human Brain Mapping*, vol. 6, pp. 368–372, 1998.
- [4] 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.
- [5] H. Hotelling, "Relations between two sets of variables," *Biometrika*, pp. 321–377, 1936.
- [6] O. Friman, M. Borga, P. Lundberg, and H. Knutsson, "Exploratory fMRI analysis by autocorrelation maximization," *Neuroimage*, pp. 454–464, 2002.
- [7] 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.
- [8] M. U. Khalid and A. K. Seghouane, "Sparse dictionary learning for fmri analysis using autocorrelation maximization," *In Proceedings of IEEE International Conference on Engineering in Medicine and Biology EMBC*, pp. 4286–4289, 2015.
- [9] J. R. Kettenring, "Canonical analysis of several sets of variables," *Biometrika*, pp. 433–451, 1971.
- [10] M. U. Khalid and A. K. Seghouane, "Multi-subject fmri connectivity analysis using sparse dictionary learning and multiset canonical correlation analysis," *In Proceedings of IEEE International Symposium on Biomedical Imaging*, pp. 683–686, 2015.
- [11] N. M. Correa, T. Adali, and V. Calhoun, "Canonical correlation analysis for data fusion and group inferences," *IEEE Signal Processing Magazine*, pp. 39–50, 2010.
- [12] Y. O. Li, T. Adali, W. Wang, and V. D. Calhoun, "Joint blind source separation by multiset canonical correlation analysis," *IEEE Transactions on Signal Processing*, vol. 57, pp. 3918–3929, 2009.
- [13] Y. O. Li, T. Eichele, V. D. Calhoun, and T. Adali, "Group study of simulated driving fMRI by multiset canonical correlation analysis," *Journal of Signal Processing Systems*, vol. 68, pp. 31–48, 2012.
- [14] S. Bakin, *Adaptive regression and model selection in data mining problems*, PhD Thesis, The Australian National University, 1999.
- [15] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society, Serie B*, vol. 68, pp. 49–67, 2006.
- [16] H. Shen and J. Z. Huang, "Sparse principal component analysis via regularized low rank matrix approximation," *Journal of Multivariate Analysis*, vol. 99, pp. 1015–1034, 2008.
- [17] D. M. Witten, R. Tibshirani, and T. Hastie, "A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis," *Biostatistics*, vol. 10, pp. 515–534, 2009.
- [18] A. Aissa-El-Bey and A. K. Seghouane, "Sparse canonical correlation analysis based on rank-1 matrix approximation and its application for fmri signals," *In Proceedings of IEEE International Conference on Acoustic, Speech and Signal Processing, ICASSP*, pp. 4678–4682, 2016.
- [19] 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.
- [20] 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.
- [21] A. K. Seghouane, "A note on image restoration using Cp and MSE," *IEEE Signal Processing Letters*, vol. 15, pp. 61–64, 2008.
- [22] A. K. Seghouane, "Asymptotic bootstrap corrections of AIC for linear regression models," *Signal Processing*, vol. 90, pp. 217–224, 2010.
- [23] 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, SSP*, pp. 65–68, 2014.
- [24] 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, ISBI*, pp. 45–48, 2014.
- [25] A. K. Seghouane and M. U. Khalid, "Learning dictionaries from correlated data: Application to fMRI data analysis," *In Proceedings of IEEE International Conference on Image Processing, ICIP*, pp. 2340–2344, 2016.