LOCAL SPARSE COMPONENT ANALYSIS FOR BLIND SOURCE SEPARATION: AN APPLICATION TO RESTING STATE FMRI

Gilson Vieira[1], Edson Amaro Jr[1] and Luiz A. Baccala[1]

***Abstract*— We propose a new Blind Source Separation technique for whole-brain activity estimation that best profits from FMRI's intrinsic spatial sparsity. The Local Sparse Component Analysis (LSCA) combines wavelet analysis, group-separable regularizers, contiguity-constrained clusterization and principal components analysis (PCA) into a unique spatial sparse representation of FMRI images towards efficient dimensionality reduction without sacrificing physiological characteristics by avoiding artificial stochastic model constraints. The LSCA outperforms classical PCA source reconstruction for artificial data sets over many noise levels. A real FMRI data illustration reveals resting-state activities in regions hard to observe, such as thalamus and basal ganglia, because of their small spatial scale.**

I. INTRODUCTION

Functional Magnetic Resonance Imaging (FMRI) has allowed new approaches for spatial and temporal brain activation research by addressing functional connectivity (FC). By combining voxels into predefined regions of interest (ROI) from coarse/uncustomized atlases, FC often leads to inappropriate functional spatial variability characterization [10]. Blind Source Separation (BSS), on the other hand, has received attention for consistently estimating FC while dispensing with the need of *a priori* ROI delineation by decomposing data into a set of hidden states and their corresponding spatial components [12], [1], [13], [14]. However, current BSS is unsuited for characterizing local brain image features as it imposes no constraint on spatial component extension. Another BSS weakness lies in overlooking the multiscale nature of brain images which may be of growing importance as higher resolution images become available.

Our approach combines the best of both anatomical and functional worlds thru Local Sparse Component Analysis (LSCA) to decompose the data into a small number of well localized spatial components. The related technique known as Sparse Component Analysis (SCA) only assumes sparsity in the rows of the observation matrix or in the columns of the mixing matrix [3], [13] whilst our method imposes more structure by assuming that the columns of the mixing matrix act as point-spreading functions. The key hypothesis is that any spatial component with bounded support can be perfectly described by few wavelet coefficients forming localized spatial patterns. In this case, the grouping procedure can be accomplished by using a contiguity-constrained clustering algorithm whose measure of dissimilarity spatially limits the set of clusters eligible for merging during the iteration process [16]. This even allows estimating the number of spatial components.

From the main model (1), we briefly discuss its solution via the vector soft-thresholding operator (8) which generalizes the scalar soft-thresholding operator in [9]. Simulations against traditional PCA show its superiority (Sec. IV). Further illustration is given by applying the method to real resting-state data (Sec. V).

II. PROBLEM STATEMENT

Consider the following BSS model to describe whole brain FMRI activity:

$$\mathbf{z}_t = \quad \mathbf{A}\mathbf{x}_t + \mathbf{v}_t, \tag{1}$$

where $\mathbf{z}_t = [z_{1,t}, \ldots, z_{M,t}]^{\mathrm{T}}$, $1 \leq t \leq N$, is the vector of observed signals, $\mathbf{x}_t = [x_{1,t}, \ldots, x_{K,t}]^{\mathrm{T}}$ is a vector of hidden states and $\mathbf{v}_t = [v_{1,t}, \ldots, v_{M,t}]^{\mathrm{T}}$ is a vector describing small-scale spatial variations assumed i.i.d. random variables. To simplify, we further assume $\mathbf{v}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{R} = \sigma^2 \mathbf{I}_{M\times M}\right)$. The $M \times K$ mixing matrix $\mathbf{A}$ characterizes the effect of K physiologically meaningful sources

$$\mathbf{s}_t = \mathbf{A}\mathbf{x}_t = \sum_{k=1}^{K} \mathbf{a}_k x_{k,t}, \tag{2}$$

where $\mathbf{a}_k$ is the k-th column of $\mathbf{A}$ and the goal is to estimate $\mathbf{x}_t$ without knowing $\mathbf{A}$.

Different assumptions about $\mathbf{x}_t$ and $\mathbf{A}$ lead to different solutions to (1). Neuroscientists's concerns about functional brain components centers around requiring $\mathbf{a}_k$ spatial localization. Thus we solve (1) by assuming that $\mathbf{s}_t$ lies in $\mathbf{B}^s_{1,1}$, a particular kind of Besov space chosen for containing smooth functions with localized singularities [15] so that the problem of estimating $\mathbf{s}_t \in \mathbf{B}^s_{1,1}$ from $\mathbf{z}_t$ reduces to minimizing the functional:

$$f(\mathbf{s}_t) = \|\mathbf{z}_t - \mathbf{s}_t\|_2^2 + \|\mathbf{s}_t\|_{s,1} = \|\mathbf{z}_t - \mathbf{s}_t\|_2^2 + \sum_{j,\mathbf{k}} \lambda_{j,\mathbf{k}} |\hat{s}_{j,\mathbf{k},t}|, \tag{3}$$

where $\hat{s}_{j,\mathbf{k},t} = \langle \mathbf{s}_t, \psi_{j,\mathbf{k}} \rangle$ and $\{\psi_{j,\mathbf{k}}; j, \mathbf{k} \in \mathbf{Z}^+ \times \mathbf{Z}^d\}$ constitutes an orthonormal wavelet basis [8]. The functional (3) has a unique global minimum for each t, but it must be adjusted to provide a consistently solution for all t.

III. LOCAL SPARSE COMPONENT ANALYSIS

Under matrix notation (1) satisfies $\mathbf{Z} = \mathbf{S} + \mathbf{V}$, where $\mathbf{Z} \equiv (z_{k,t})_{k,t}$ is the $M \times N$ matrix of observations and

*This work was supported in part by CNPQ grant 307163/2013-0 to L.A.B.

[1] Inter-institutional Grad Program on Bioinformatics, Universidade de São Paulo, São Paulo, Brazil. Corresponding author: gilson.vieira@usp.br

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

$\mathbf{S} = \Phi\mathbf{A}\mathbf{X}$ is the matrix of spatial components. The first step is to calculate a sparse representation of $\mathbf{Z}$ under an orthonormal wavelet transformation $\mathbf{\Phi}$

$$\hat{\mathbf{Z}} = \hat{\mathbf{S}} + \hat{\mathbf{V}} = \mathbf{\Phi}\mathbf{Z}, \tag{4}$$

where $\hat{\mathbf{S}} = \mathbf{\Phi}\mathbf{A}\mathbf{X}$. The transform $\mathbf{\Phi}$ aims to sparsify the spatial component representation, so that $\mathbf{s}^k = \mathbf{a}_k\mathbf{x}^k$, where $\mathbf{x}^k$ is the k-th row of $\mathbf{X}$, has few significant coefficients in $\hat{\mathbf{S}}$, located on a set $I_k \subset \{1, \ldots, M\}$, to be clustered. Appendix VI-A provides a best basis wavelet packet analysis criterion for reducing model complexity.

Before clustering the rows of $\hat{\mathbf{Z}}$ to obtain I_k, one should denoise $\hat{\mathbf{Z}}$ considering that $\mathbf{s}_t = \mathbf{A}\mathbf{x}_t \in \mathbf{B}^s_{1,1}$. At any fixed time t, the l_1-norm regularization in (3) (basis pursuit [4]) takes the form of a fully separable penalized regression in the wavelet domain

$$\min_{\hat{\mathbf{s}}_t} \frac{1}{2}\sum_k (\hat{z}_{k,t} - \hat{s}_{k,t})^2 + \lambda \sum_k |\hat{s}_{k,t}|. \tag{5}$$

by whose first derivative, one can show its optimum is attained via the soft-thresholding operator [9]

$$\hat{s}_{k,t} = \text{sign}(\hat{z}_{k,t}) \max(|\hat{z}_{k,t}| - \lambda, 0). \tag{6}$$

Using (6) to solve (5) for all t at once can result in setting coefficients to zero for some values of t but not for other values of t even when λ is large, unless one ties all $\hat{s}_{i,t}$'s together for all values of t. To do so, we introduce group-separable regularizers [17], where the thresholding operator applies to appropriate jointly chosen groups of variables that are simultaneously either shrunken or ignored together.

Let $\hat{\mathbf{z}}^k$ and $\hat{\mathbf{s}}^k$ be the k-th row of $\hat{\mathbf{Z}}$ and $\hat{\mathbf{S}}$ respectively. The aim of group shrinkage is to promote group sparsity in $\hat{\mathbf{S}}$ by changing the l_1 penalization to an l_2 penalization in (3) and solving M minimization separate problems

$$\min_{\hat{\mathbf{s}}^k} \frac{1}{2}\|\hat{\mathbf{z}}^k - \hat{\mathbf{s}}^k\|_2^2 + \lambda_k \|\hat{\mathbf{s}}^k\|_2, \tag{7}$$

whose solution is determined by the vector soft-thresholding operator [6]

$$\hat{\mathbf{s}}^k = \frac{\max(\|\hat{\mathbf{z}}^k\|_2 - \lambda_k, 0)}{\|\hat{\mathbf{z}}^k\|_2}\hat{\mathbf{z}}^k, \tag{8}$$

for $1 \le k \le M$. Appendix VI-C contains the appropriate λ_k values based on the estimate for σ^2 in Appendix VI-B.

In the third step group membership I_k is determined using a contiguity-constrained clustering algorithm to $\hat{\mathbf{S}}$ rows and $\mathbf{s}^k$ is estimated by applying $\mathbf{\Phi}^{-1}$ to each cluster:

$$\mathbf{s}^k = \sum_{i \in I_k} \phi_i^{-1}\hat{\mathbf{s}}^i, \tag{9}$$

where ϕ_i^{-1} is the i-th column of $\mathbf{\Phi}^{-1}$ ($\mathbf{\Phi}^{\mathrm{T}}$, for wavelet transforms). Appendix VI-D provides a measure of dissimilarity enforcing contiguity-constraint clustering by combining the instantaneous correlation between the rows of $\hat{\mathbf{S}}$ and the physical distance between the ϕ_ks' centers of mass.

Finally, the hidden states $\mathbf{x}^k$ are estimated using the scores produced by the first principal component of $\mathbf{s}^k$ and $\mathbf{A}$ is obtained by maximizing the log-likelihood of (1) given $\mathbf{x}^k$.

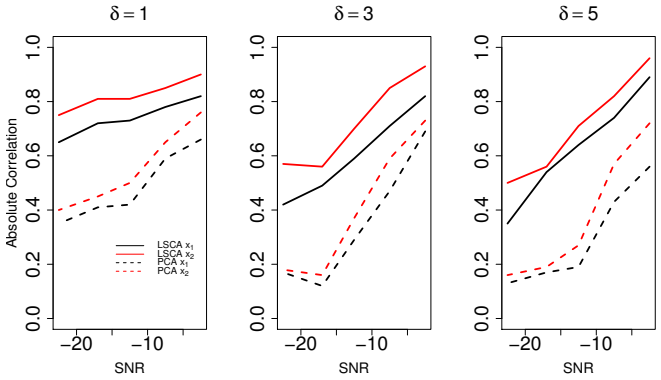


Fig. 1. Comparison of efficiency of LSCA and PCA in recovering source temporal information. Lines represent the mean correlation between the simulated hidden state $x_{k,t}$ and the estimated hidden state $\hat{x}_{k,t}$ across 30 simulations.

IV. NUMERICAL ILLUSTRATION

To gauge performance data were simulated using (1) on an $M = 64 \times 64$ grid of side length equal to $\Delta s = 0.30$ for a pair of correlated sources (equal to 0.5) whose relative distance δ is varied in addition to the amount of background additive noise. An isotropic bivariate gaussian centered at $\boldsymbol{\mu}_1 = [27 + \delta, 27 + \delta]$ and spreading over $\mathbf{\Sigma}_1 = 3\mathbf{I}_2$, where $\mathbf{I}_2$ is the 2×2 identity matrix, was used as point spread function for the first source. For the second source, we use an anisotropic bivariate gaussian centered at $\boldsymbol{\mu}_2 = [37-\delta, 37-\delta]$ and spreading over $\mathbf{\Sigma}_2 = \begin{bmatrix} 9 & 0 \\ 0 & 1 \end{bmatrix}$.

The wavelet package (see VI-A) employed a library of 2D Haar functions with wavelet coefficients up to level 3. The regularization parameter λ (see VI-C) was chosen to represent the upper limit of an interval with 95% of confidence for σ^2. We set $r = 9$ (see VI-D) yet results are similar for all neighbourhood sizes $r = 3, \ldots, 9$. The hidden states were assumed zero mean and unit variance independent random variables. The figure of merit for source reconstruction was the average value of the correlation between the simulated hidden state $x_{k,t}$ and the estimated hidden state $\hat{x}_{k,t}$ across 30 repetitions of each parameter combination. For PCA the results were compared using the largest correlation between $x_{k,t}$ and the identified hidden states.

A. Simulation results

Fig. 1 shows the LSCA superiorirty over PCA for $N = 250$ over the noise levels ($\sigma^2 = 1, 3, 5, 7, 9$) corresponding to SNR $= -2.5, -7.5, -12.5, -17.5$ and -22.5 db (SNR $= 10\log_{10}(\text{VAR}(\mathbf{s})/\sigma^2)$ with $\mathbf{s} = \text{vec}([\mathbf{A}\mathbf{x}_1 \cdots \mathbf{A}\mathbf{x}_N])$. This suggests that LSCA outperforms traditional PCA and that much information can be recovered even under very low SNR.

V. EXPERIMENTAL RESULTS

For real data illustration, we used fMRI images from one volunteer under a resting state protocol with results not only in accord with current resting state theories but most

importantly revealing subcortical regions hard to capture using standard methods because of their small spatial scale.

A. Data Acquisition Details

Whole brain volumes of FMRI data (TR $=$ 600sec, TE $=$ 33ms, 32 slices, FOV $=$ 247 $\times$ 247 mm, matrix size 128 $\times$ 128, in plane resolution 1.975 $\times$ 1.975 mm, slice thickness 3.5mm with 1.8mm of gap) was acquired on a 3T system under resting condition using a Multiplexed Echo Planar Imaging sequence (multi-band accelerator factor of 4) [11]. A high-resolution T1 image was also acquired with an MPRAGE sequence (TR $=$ 2500 ms, TE $=$ 3.45 ms, inversion time = 1000 ms, 256 $\times$ 256 mm FOV, 256 $\times$ 256 in-plane matrix, 1 $\times$ 1 $\times$ 1 mm voxel size, 7° flip angle).

B. Preprocessing

After motion and slice time correction and temporal high pass filtering (allowing fluctuations above 0.005Hz) using FEAT v5.98, the FMRI data was aligned to the grey-matter mask via FreeSurfer's automatic registration tools (v. 5.0.0) resulting in extracted BOLD signals at regions with preponderantly neuronal cell bodies.

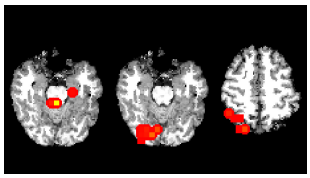
C. Real Data Results

Using 3D Haar functions to generate the sparsifying transform up to level 3 (other paramenters as in IV), nine well localized spatial components emerged from 28 image slices of FMRI images covering cortical and subcortical structures ($M = 458752$ voxels) and comprising visual (inferior occipital sulcus (IOS)) and sensory (superior parietal gyrus (SPG)) cortices, the default mode network (precuneus (PC), anterior cingulate (AC), inferior frontal gyrus (IFG)), the attention network (frontal pole (FP)), the basal gaglia (BG) and thalamus (THA), see Fig. 2. As expected, these components reflect most of data variability and coincide with traditional resting state regions observed across individuals, different data acquisition and analysis techniques [2]. Furthermore, small regions such as the thalamus and the basal ganglia, often overlooked by ICA and seed based techiniques [7], [2], clearly appeared. This suggests that the present wavelet approach can identify structures often obscured by standard methods.

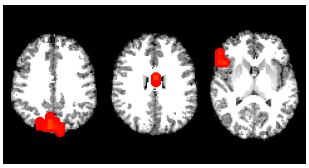
Fig. 3 illustrates the signals and their wavelet cross spectrum (WCS) for (a) the right and left estimated SPG and (b) right estimated IOS and AC. As expected, these regions are very correlated and their correlation varies over time. The time interval with the highest level of correlation is disjoint among these pairs of regions, indicating that the method successfully distinguishes sources that are not independent and which may even be nonstationary.

VI. Conclusion

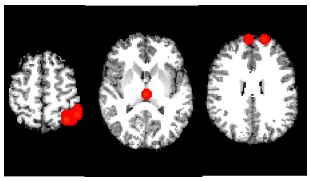
By taking advantage of BOLD signal local nature, we developed a new BSS method capable of improved FMRI dataset characterization. Due to space constraints, we have left systematic comparison to other methods and the consideration of other mother wavelets for future work.



(a) BG, right IOS and right SPG.

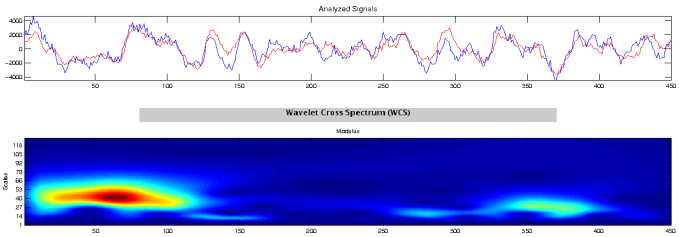


(b) PC, AC and right IFG.

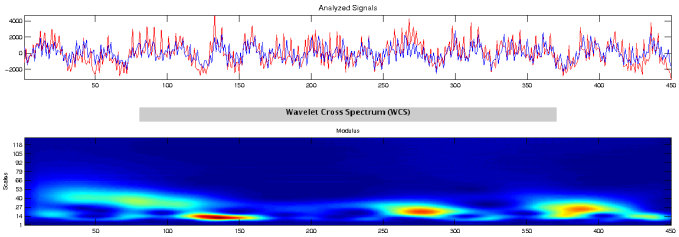


(c) left SPG, THA and FP.

Fig. 2. Location of the nine spatial components found by LSCA.



(a) Right (blue) and left SPG (red).



(b) Rigth IOS (blue) and AC (red)

Fig. 3. Figure illustrating two pairs of selected recovered signals and their respective WCS (see text).

APPENDIX

A. Proposed Extensions For Dynamic Wavelet Packets

For ease of notation, let $M = 2^J$, for $J > 0$ for the wavelet packet library $\mathcal{L} = \{W_n\}_n$ [5]. For fixed t, consider the following spatial multiresolution analysis of $\mathbf{z}_t$

$$\mathbf{z}_t = \sum_{(j,n)\in\mathbb{P}} \sum_{k=0}^{2^j-1} w_{j,n,k,t}\mathbf{w}_{j,n,k}, \qquad (10)$$

where $\mathbb{P} = \{(j,n)\} \subset \mathbb{N}\times\mathbb{N}$ is a collection of indexes and $\mathbf{w}_{j,n,k} = [W_n(2^j s_1 - k), \ldots, W_n(2^j s_M - k)]^{\mathrm{T}}$, $(j,n) \in \mathbb{P}$ and $1 \le k \le 2^J$. If $\mathbb{P}$ is such that the intervals $[2^j n, 2^j(n+1))$ form a disjoint covering of $[0, 2^J)$, the vectors $\mathbf{w}_{j,n,k}$ form a complete and orthonormal basis of $\mathbb{R}^M$ [5].

Among all bases generated by $\mathcal{L}$, it is possible to obtain a complete and orthonormal best basis prioritizing the components of $\mathbf{z}_t$ of largest variance. For this it is necessary to

define a criterion upon which the importance of each wavelet packet $\mathbf{W}_{j,n} = \{\mathbf{w}_{j,n,1}, \ldots, \mathbf{w}_{j,n,2^j-1}\}$ in (10) is judged.

Due to orthonormality, the power of $\mathbf{z}_t$ is the sum of the wavelet coefficient powers

$$\begin{aligned} \mathrm{E}(\mathbf{z}_t^{\mathrm{T}}\mathbf{z}_t) = & \quad \mathrm{E}\Big(\textstyle\sum_{(j,n)\in\mathbb{P}} \sum_{k=0}^{2^j-1} (w_{j,n,k,t})^2 \mathbf{w}_{j,n,k}^{\mathrm{T}} \mathbf{w}_{j,n,k} \Big) \\ = & \quad \textstyle\sum_{(j,n)\in\mathbb{P}} \sum_{k=0}^{2^j-1} \mathrm{E}\Big(w_{j,n,k,t}^2 \Big) \\ = & \quad \textstyle\sum_{(j,n)\in\mathbb{P}} \sum_{k=0}^{2^j-1} \mathrm{VAR}(w_{j,n,k,t}), \end{aligned} \tag{11}$$

where $\mathrm{VAR}(x)$ denotes the time variance of x. The last equality in (11) follows from $\mathrm{E}(\mathbf{z}_t) = \mathbf{0}$ implies $\mathrm{E}(w_{j,n,k,t}) = 0$ for all $(j,n) \in \mathbb{P}$ e $1 \leq j \leq 2^J$.

As $\mathrm{VAR}(w_{t,j,n,k}) = \sigma^2$ if and only if $w_{t,j,n,k}$ is noise, we set the following cost function

$$C_{\sigma^2}\left(\{\mathbf{W}_{j,n}\}\right) = \sum_{k=1}^{2^j-1} \mathbf{1}\left(\mathrm{VAR}(w_{j,n,k,t}) \geq \sigma^2\right), \tag{12}$$

called dynamic thresholding cost function. Note that $C_{\sigma^2}\left(\{\mathbf{W}_{j,n}\}\right)$ is additive, and so can be used in a best basis selection algorithm of linear complexity.

B. Additive Noise Estimation

If $\mathbf{R} = \sigma^2 \mathbf{I}_{M\times M}$, then $\hat{\mathbf{v}}^k \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \mathbf{I}_{N\times N}\right)$ due to the orthogonality imposed by $\mathbf{\Phi}$, where $\hat{\mathbf{v}}^k$ is the k-th row of $\hat{\mathbf{V}}$. Assuming $\mathbf{z}_t$ is sparse under $\mathbf{\Phi}$ implies that most of $\{\hat{s}_{k,t}\}_{\forall k}$ are zero or equivalently that most of $\{\hat{z}_{k,t}\}_{\forall k}$ are just noise. This observation motivates the following estimator for σ^2

$$\hat{\sigma}^2 = \mathrm{median}_{\forall k} \hat{\mathrm{VAR}}\{\hat{z}_{k,t}\}, \tag{13}$$

where $\hat{\mathrm{VAR}}$ denotes temporal sample covariance. It is easy to see that (13) is unbiased provided that fifty percent of $\{\hat{s}_{k,t}\}_{\forall k}$ are zero.

C. Setting the regularization parameter

As a rule of the thumb we clip $\hat{z}_{k,t}$ so that most noise-related coefficients lie below (13) with high probability. If $\mathrm{VAR}\{\hat{s}_{k,t}\} = 0$, then $\hat{z}_{k,t}$ are independent normally distributed random variables, then one may readily show that this is attained if

$$\lambda_k = \frac{(N-1)^2 \hat{\sigma}^2}{\chi^2_{\alpha/2,N-1}}, \tag{14}$$

where $\chi^2_{\alpha/2,N-1}$ is the $\alpha/2$-th percentile of the chi-square distribution with $N-1$ degrees of freedom.

D. Contiguity-constrained Hierarchical Clustering Algorithm

The clustering begins with each time series $\hat{\mathbf{s}}^k$, $1 \leq k \leq M$, defining a singleton cluster. At each step, clusters pairs (A, B) from the previous step are joined if they minimize the following criterion

$$\max\left\{\mathrm{dist}(\hat{\mathbf{s}}^i, \hat{\mathbf{s}}^j) : i \in A, j \in B\right\}, \tag{15}$$

where

$$\mathrm{dist}(\hat{\mathbf{s}}^i, \hat{\mathbf{s}}^j) = \begin{cases} \infty, & \|\bar{\phi}_i - \bar{\phi}_j\|_2 > r \\ 1 - |\mathrm{cor}(\hat{\mathbf{s}}^i, \hat{\mathbf{s}}^j)|, & \text{otherwise,} \end{cases} \tag{16}$$

where $\bar{\phi}_i = \int_{\mathbb{R}^d} s|\phi_i|^2 ds / \int_{\mathbb{R}^d} |\phi_i|^2 ds$ defines de center of mass of ϕ_i. In short, the above criterium combines a complete-linkage strategy and dissimilarity measure accounting for the absolute correlation restricted to nearby sources only.

REFERENCES

[1] C.F. Beckmann and S.M. Smith. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, 23(2):137–152, 2004.

[2] Bharat B Biswal and et al. Toward discovery science of human brain function. *Proc. Natl. Acad. Sci. U.S.A.*, 107(10):4734–4739, March 2010. PMID: 20176931.

[3] Pau Bofill and Michael Zibulevsky. Underdetermined blind source separation using sparse representations. *Signal Processing*, 81(11):2353–2362, November 2001.

[4] Scott Shaobing Chen, David L. Donoho, and Michael A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, January 1998.

[5] Ronald R. Coifman, Yves Meyer, Steven Quake, and M. Victor Wickerhauser. Signal processing and compression with wavelet packets. In J. S. Byrnes, Jennifer L. Byrnes, Kathryn A. Hargreaves, and Karl Berry, editors, *Wavelets and Their Applications*, number 442 in NATO ASI Series, pages 363–379. Springer Netherlands, January 1994.

[6] Patrick L. Combettes and Valrie R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Modeling & Simulation*, 4(4):1168–1200, January 2005.

[7] J. S. Damoiseaux, S. a. R. B. Rombouts, F. Barkhof, P. Scheltens, C. J. Stam, S. M. Smith, and C. F. Beckmann. Consistent resting-state networks across healthy subjects. *PNAS*, 103(37):13848–13853, September 2006. PMID: 16945915.

[8] Ingrid Daubechies, Michel Defrise, and Christine De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. arXiv e-print math/0307152, July 2003.

[9] David L. Donoho, Iain M. Johnstone, Gerard Kerkyacharian, and Dominique Picard. Wavelet shrinkage: Asymptopia? *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(2):301–369, January 1995. ArticleType: research-article / Full publication date: 1995 / Copyright 1995 Royal Statistical Society.

[10] Alan C Evans, Andrew L Janke, D Louis Collins, and Sylvain Baillet. Brain templates and atlases. *Neuroimage*, 62(2):911–922, August 2012. PMID: 22248580.

[11] David A. Feinberg, Steen Moeller, Stephen M. Smith, Edward Auerbach, Sudhir Ramanna, Matt F. Glasser, Karla L. Miller, Kamil Ugurbil, and Essa Yacoub. Multiplexed echo planar imaging for sub-second whole brain FMRI and fast diffusion imaging. *PLoS ONE*, 5(12):e15710, December 2010.

[12] K. J. Friston, C. D. Frith, P. F. Liddle, and R. S. J. Frackowiak. Functional connectivity: The principal-component analysis of large (PET) data sets. *J Cereb Blood Flow Metab*, 13(1):5–14, January 1993.

[13] Pando Georgiev, Fabian Theis, Andrzej Cichocki, and Hovagim Bakardjian. Sparse component analysis: a new tool for data mining. In Panos M. Pardalos, Vladimir L. Boginski, and Alkis Vazacopoulos, editors, *Data Mining in Biomedicine*, number 7 in Springer Optimization and Its Applications, pages 91–116. Springer US, January 2007.

[14] Gabriele Lohmann, Kirsten G. Volz, and Markus Ullsperger. Using non-negative matrix factorization for single-trial analysis of fMRI data. *NeuroImage*, 37(4):1148–1160, October 2007.

[15] S. G Mallat. *A wavelet tour of signal processing the Sparse way*. Elsevier /Academic Press, Amsterdam; Boston, 2009.

[16] F. Murtagh. A survey of algorithms for contiguity-constrained clustering and related problems. *The Computer Journal*, 28(1):82–88, January 1985.

[17] S.J. Wright, R.D. Nowak, and M. A T Figueiredo. Sparse reconstruction by separable approximation. *IEEE Transactions on Signal Processing*, 57(7):2479–2493, 2009.