

Connectivity-informed Sparse Classifiers for fMRI Brain Decoding

Bernard Ng, Viviana Siless, Gaël Varoquaux, Jean-Baptiste Poline, Bertrand Thirion, Rafeef Abugharbieh

▶ **To cite this version:**

Bernard Ng, Viviana Siless, Gaël Varoquaux, Jean-Baptiste Poline, Bertrand Thirion, et al.. Connectivity-informed Sparse Classifiers for fMRI Brain Decoding. Pattern Recognition in Neuroimaging, Christos Davatzikos, Moritz Grosse-Wentrup, Janaina Mourao-Miranda, Dimitri Van De Ville, Jul 2012, London, United Kingdom. hal-00726656

HAL Id: hal-00726656

https://inria.hal.science/hal-00726656

Submitted on 31 Aug 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Connectivity-informed Sparse Classifiers for fMRI Brain Decoding

Bernard Ng, Viviana Siless, Gael Varoquaux,
Jean-Baptiste Poline, and Bertrand Thirion
Parietal team, INRIA Saclay
Gif-Sur-Yvette, France
e-mail: bernardyng@gmail.com

Rafeef Abugharbieh
Department of Electrical and Computer Engineering
The University of British Columbia
Vancouver, Canada

***Abstract*—In recent years, sparse regularization has become a dominant means for handling the curse of dimensionality in functional magnetic resonance imaging (fMRI) based brain decoding problems. Enforcing sparsity alone, however, neglects the interactions between connected brain areas. Methods that additionally impose spatial smoothness would account for local but not long-range interactions. In this paper, we propose incorporating connectivity into sparse classifier learning so that both local and long-range connections can be jointly modeled. On real data, we demonstrate that integrating connectivity information inferred from diffusion tensor imaging (DTI) data provides higher classification accuracy and more interpretable classifier weight patterns than standard classifiers. Our results thus illustrate the benefits of adding neurologically-relevant priors in fMRI brain decoding.**

Keywords-connectivity, DTI, fMRI, sparse classifiers

I. INTRODUCTION

The use of functional magnetic resonance imaging (fMRI) for brain decoding has gained substantial research interest in the past decade [1]. With the signal intensity at each voxel or brain area treated as a feature, the goal of brain decoding is to discriminate different cognitive states based on the brain signal patterns. Due to the high dimensionality of fMRI data and the typically small sample sizes, special caution must be taken to control for overfitting [1]. The traditional approach for dealing with the curse of dimensionality is to apply univariate feature selection in isolating brain areas that display significant activation or discriminability [2]. However, this approach discards the collective information encoded by the brain signal patterns [1]. To circumvent this limitation, multivariate feature selection techniques, such as support vector machines with recursive feature elimination (SVM-RFE) [3], have been proposed but tend to be computationally expensive.

Recently, sparse regularization has become a primary means for handling overfitting [4-11]. By enforcing sparsity, classifier weights associated with irrelevant features are shrunk to exactly zero. This approach thus enables feature selection and classifier learning to be simultaneously and cooperatively performed. Imposing sparsity is typically achieved by incorporating a least absolute shrinkage and selection operator (LASSO) penalty into classifier learning [5]. However, in the presence of correlated features, LASSO tends to arbitrarily select one feature among each correlated set, which is especially problematic for fMRI decoding since signals of interacting brain areas are certainly correlated [5, 6]. To encourage correlated brain areas to be jointly selected, incorporation of an elastic net penalty has been proposed [5, 6], but its effectiveness appears limited by the strong noise in fMRI data [7]. Structured sparsity [8] has also been explored to enforce joint selection of predefined groups of voxels (e.g. voxels grouped by anatomical regions or functional parcels [8]), but these penalties alone do not account for the local spatial correlations between neighboring voxels [7, 9, 10].

To explicitly model local spatial correlations, total variation regularization has been explored, which penalizes the spatial gradient of the classifier weight patterns [9]. A Bayesian approach for incorporating a spatiotemporal prior while enforcing sparsity has also been put forth [11]. Furthermore, a recent method called, GraphNet, which imposes spatial smoothness through a graph-constrained elastic net penalty [10] as well as a generalized sparse regularization (GSR) strategy [7] that additionally facilitates incorporation of structured sparse penalties have been proposed. Encouraging spatial smoothness accounts for local correlations but neglects the longer range interactions. Since it is now well established that the brain is organized into distributed areas that are interconnected by white matter fiber pathways [12], integrating connectivity into sparse classifier learning may provide additional relevant information in improving identification of task discriminant brain areas.

In this paper, we propose incorporating connectivity into sparse classifier learning to explicitly model the integrative property of the brain [12]. Both GraphNet and GSR permit integration of connectivity, but entail converting discrete labels into continuous variables using e.g. optimal scoring (OS) [10] or graph embedding (GE) [13] for classifier learning. In this work, we show that the learning of connectivity-informed sparse classifiers can be posed as a single unified optimization and solved using existing sparse optimizers [8, 14] with only a minimal change in the gradient update. For validation, we build a connectivity-informed sparse logistic regression (SLR) classifier with connectivity inferred from diffusion tensor imaging (DTI) data. On a face viewing task, we demonstrate that incorporating connectivity improves classification accuracy and result interpretability.

II. METHODS

Restricting our attention to linear classifiers so that the relevance of the features can be directly discerned from the magnitude of linear classifier weights [8], the general

problem of classifier learning can be posed as finding the optimal $d\times1$ weight vector, $\mathbf{w}$, and an intercept, b, such that the $n\times1$ predicted labels, $f(\mathbf{X}\mathbf{w}+b)$, best match the $n\times1$ ground truth labels, $\mathbf{y}$, associated with the $n\times d$ feature matrix, $\mathbf{X}$. n is the number of samples, d is the number of features, and $f(\cdot)$ is a function that maps $\mathbf{X}\mathbf{w}+b$ to the label space. In the context of fMRI decoding, we treat the signal intensity of a voxel or brain area as a feature with different cognitive states being the labels. Since $n << d$ typically, controlling overfitting through e.g. regularization is crucial.

A. *Overview of Sparse Regularization*

Linear classifier learning with regularization can be expressed as the following optimization problem:

$$\min_{\mathbf{w},b} E\left(f\left(\mathbf{X}\mathbf{w}+b\right),\mathbf{y}\right)+\lambda R(\mathbf{w}), \qquad (1)$$

where $E(\cdot)$ is a mis-classification loss (see Section II-C), $R(\mathbf{w})$ is a regularization penalty, and λ balances the influence of $E(\cdot)$ and $R(\mathbf{w})$. To enforce sparsity, the most widely-used approach is to set $R(\mathbf{w})$ to $\|\mathbf{w}\|_1$, which is the well known LASSO penalty [5]. However, LASSO's difficulties with correlated features render it less effective for fMRI decoding [5, 6]. To encourage joint selection of correlated features, elastic net has been proposed, which is equivalent to adding a ridge penalty to LASSO, i.e. $\alpha\|\mathbf{w}\|_1+(1-\alpha)\|\mathbf{w}\|_2^2$, with the relative weightings between the two terms controlled by α. However, the performance of elastic net in jointly selecting correlated features highly depends on the level of noise in the data [7]. To explicitly model local spatial correlations, total variation regularization, $\alpha\|\nabla\mathbf{w}\|_1$, has been explored [9]. GraphNet [10], which comprises a graph-constrained elastic net penalty, $\alpha\|w\|_1+(1-\alpha)\mathbf{w}^T\mathbf{L}\mathbf{w}$, can also be used with spatial smoothness enforced through a graph Laplacian matrix, $\mathbf{L}$. Alternatively, one may employ GSR [7], which involves a penalty of the form: $\alpha J(\mathbf{w})+(1-\alpha)\|\mathbf{\Gamma}\mathbf{w}\|_2^2$, with $J(\mathbf{w})$ being a structured sparsity term and spatial contiguity imposed through $\mathbf{\Gamma}$. Overviews of structured sparsity can be found in [7] and [8]. To exploit existing optimizers that are designed for sparse regression models in learning prior-informed sparse classifiers, GraphNet employs OS [10] to transform discrete class labels, $\mathbf{y}$, into continuous variables, $\mathbf{z}$, while GSR uses GE [13] for this purpose. The effects of applying OS and GE prior to classifier learning, as opposed to directly optimizing (1), are unclear. In this work, we present a simple approach to mitigate the need for converting $\mathbf{y}$ into continuous variables so that the learning of sparse classifiers informed by prior information, such as connectivity, can be performed as a single optimization.

B. *Connectivity-informed Sparse Classifiers*

To integrate connectivity into sparse classifier learning, we consider the following formulation:

$$\min_{\mathbf{w},b} E\left(f\left(\mathbf{X}\mathbf{w}+b\right),\mathbf{y}\right)+\lambda\left(\alpha J(\mathbf{w})+(1-\alpha)\mathbf{w}^T\mathbf{L}\mathbf{w}\right), \qquad (2)$$

where $\mathbf{L}$ is a $d\times d$ weighted graph Laplacian matrix with $\mathbf{L}_{ij} = -\mathbf{l}_{ij}$ for $i \neq j$, $\mathbf{L}_{ij} = \Sigma_i \mathbf{l}_{ij}$ for $i = j$, and $\mathbf{l}_{ij}$ being the connectivity between features i and j (Section II-C). We note that $\mathbf{L}$ being a Laplacian matrix is not essential and other priors may be employed [10]. If $\mathbf{L} = \mathbf{\Gamma}^T\mathbf{\Gamma}$, the penalty in (2) reduces to that of GSR's. The difference is that (2) directly takes Laplacian matrices as input, whereas GSR would require computing a matrix square root. Also, if $J(\mathbf{w}) = \|\mathbf{w}\|_1$, the penalty in (2) becomes that of GraphNet's. To optimize (2), if $E(\cdot)$ is differentiable with a gradient, $\mathbf{g}(\mathbf{w})$, one may use proximal gradient [8] with only a simple modification to the gradient required: $\mathbf{g}_{\text{mod}}(\mathbf{w}) = \mathbf{g}(\mathbf{w})+2\lambda(1-\alpha)\mathbf{L}\mathbf{w}$. In this work, we focus on the case where $J(\mathbf{w}) = \|\mathbf{w}\|_1$, which we optimize using two metric projection in combination with limited memory BFGS [14]. This permits an approximation of the Hessian of $E(\cdot)$ to be exploited in learning the step direction, which helps speed up convergence [14]. We describe how connectivity-informed SLR classifiers can be built by optimizing (2) next.

C. *Connctivity-informed Sparse Logistic Regression*

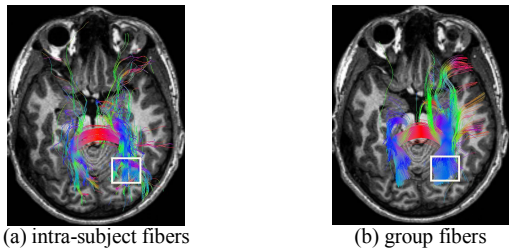
The mis-classification loss of LR is given by [5]:

$$E\left(f\left(\mathbf{X}\mathbf{w}+b\right),\mathbf{y}\right)=\frac{1}{n}\sum_{i=1}^{n}\ln\left(1+\exp\left(-\mathbf{y}_i\left(\mathbf{w}^T\mathbf{x}_i+b\right)\right)\right), \qquad (3)$$

where $\mathbf{x}_i$ is a $d\times1$ vector with $\mathbf{x}_i^T$ being the i^{th} row of $\mathbf{X}$. Since (3) is differentiable:

$$\nabla_{\mathbf{w}} E\left(f\left(\mathbf{X}\mathbf{w}+b\right),\mathbf{y}\right)=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{-\mathbf{x}_i\mathbf{y}_i}{1+\exp(\mathbf{y}_i(\mathbf{w}^T\mathbf{x}_i+b))}\right), \qquad (4)$$

incorporation of connectivity in learning a SLR classifier can be achieved by applying e.g. two metric projection [14] with $J(\mathbf{w}) = \|\mathbf{w}\|_1$ in (2) and $2\lambda(1-\alpha)\mathbf{L}\mathbf{w}$ added to (4). As a measure of connectivity, we use the fiber count between brain areas derived from DTI tractography [12]. Due to crossing fibers as well as how diffusion tensors are more isotropic near gray matter brain areas, the location of the fiber end points is subjected to errors (Fig. 1(a)). To partially alleviate these limitations, we employ a group fiber estimation strategy [15] (Section III) to improve end point localization (Fig. 1(b)).



(a) intra-subject fibers (b) group fibers

Figure 1. Fiber estimation strategies. Fibers passing through a box around the FFA displayed. (a) Fibers estimated from a single exemplar subject's tensor volume. (b) Fibers estimated from subject-average tensor volume.

III. MATERIALS

Task fMRI data were collected from 10 healthy subjects. Each subject performed a passive face viewing task over a period of ~5 min, where ten 18 s blocks of face stimulus interleaved with nine 18 s control blocks of triangles were displayed. Resting state (RS) fMRI data of ~7 min were also collected. Data were acquired at multiple centers using 3T scanners with TR = 2200 ms, TE = 30 ms, and flip angle = 75^o. Slice timing correction, motion correction, temporal detrending, and spatial normalization, were performed on the task fMRI data using the SPM8 software. Similar preprocessing was performed on the RS-fMRI data except a band-pass filter with cutoff frequencies at 0.01 to 0.1 Hz was applied. White matter and cerebrospinal fluid confounds were regressed out from the gray matter voxel time courses.

To account for the delay in hemodynamic response in labeling the fMRI volumes as associated with viewing faces or triangles, we generated a regressor for each stimulus by convolving a boxcar function (having an amplitude of one and time-locked to stimulus) with a hemodynamic response function, and extracted volumes in which the regressor value exceeded 0.5. Only volumes of the last nine blocks for each stimulus were used to ensure equal sample sizes for both stimulus classes. This results in 72 samples per class.

DTI data were collected from the same 10 subjects. Acquisition sequence similar to that in [16] was used with TR = 15000 ms, TE = 104 ms, flip angle = 90^o, 36 gradient directions, and b-value = 1300 s/mm^2. Tensor estimation and fiber tractography were performed at both intra-subject and group level using MedINRIA [17]. For generating group fibers, we first warped all subjects' tensor volumes onto that of an arbitrary subject. We then averaged the tensor volumes across subjects, performed fiber tractography on the average tensor volume, and warped the group fibers back onto each subject's native space. To enable fiber count computation, we divided the brain into P parcels with the parcel template warped onto each subject's B_0 volume. Parcellation was performed by concatenating RS-fMRI time courses across subjects and applying Ward clustering [18]. P was set to 500 to permit finer brain partitioning than facilitated by standard brain atlases (typically $P < 150$). This choice of P provides a balance between functional localization and robustness to errors in the location of fiber end points. The parcel template was further used for the generation of task fMRI parcel time courses by averaging voxel time courses within each parcel.

IV. RESULTS AND DISCUSSION

For validation, we considered the task of discriminating fMRI volumes associated with viewing faces against viewing triangles. The average intensity within each parcel was taken as a feature with each volume being a sample. 9-fold cross-validation was performed to compute classification accuracy. Using 9 folds ensures that volumes within the same stimulus block would not be erroneously assigned to both the training and test set, which inflates classification accuracy due to correlations between temporally adjacent volumes. We compared connectivity-informed SLR with anatomical connectivity prior learned using the group fiber estimation strategy (GASLR) against SVM [2], LR, LR with ridge penalty $\lambda\|\mathbf{w}\|_2^2$ (RLR), SLR $\lambda\|\mathbf{w}\|_1$, LR with elastic net penalty $\lambda\left(\alpha\|\mathbf{w}\|_1+(1-\alpha)\|\mathbf{w}\|_2^2\right)$ (ENLR), and connectivity-informed SLR with intra-subject anatomical connectivity prior (ASLR). Since local spatial correlations between voxels are implicitly modeled in generating functional parcels, comparison against total variation regularization was not performed. To set the parameters α and λ for each training fold, 8-fold internal cross validation [2] was performed with $\alpha = \{0.1, 0.3, 0.5, 0.7, 0.9\}$ and $\lambda = \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}\}$. For qualitative assessment, we performed 9-fold cross validation on the whole sample set to determine the optimal α and λ. We then generated 100 subsamples of size 72 (i.e. half of the total number of samples) and learned a classifier for each subsample with α and λ fixed to the optimal value. The classifier weights across subsamples were averaged to identify parcels consistently assigned higher weights.

Results of the contrasted classifiers are shown in Fig. 2. Only classifier weight patterns of an exemplar subject are displayed due to space limitation. Among the contrasted methods, SVM obtained the lowest average classification accuracy over subjects, and the classifier weight patterns seemed spatially spurious. Using LR resulted in 2% higher accuracy compared to SVM, but the classifier weight patterns appeared similar. Controlling overfitting with a ridge penalty led to a further 2% increase in accuracy, but only a slight improvement in localization of relevant parcels. Enforcing sparsity using SLR provided an additional 1% increase in accuracy over using a ridge penalty with more localized classifier weight patterns obtained. However, rather small weights were assigned to parcels in the fusiform face area (FFA), which is known to be specialized for face recognition. Using LR with elastic net resulted in a 0.5% accuracy gain over SLR and more weights assigned to parcels within the FFA, but localization power was reduced compared to SLR. Incorporating intra-subject anatomical connectivity into SLR actually led to lower accuracy than the contrasted sparse classifiers. We suspect the decrease in accuracy was due to errors in fiber tractography, which could have a major impact on the connectivity estimates. By using the group fiber extraction strategy to improve connectivity estimation and integrating the resulting connectivity prior into SLR, the highest classification accuracy was attained with weights consistently assigned to parcels within the FFA.

V. CONCLUSIONS

We proposed informing sparse classifier learning with connectivity for fMRI decoding. We showed that with only a minor change to the gradient, integration of connectivity is facilitated. On real data, we demonstrated that incorporating connectivity inferred from a group fiber estimation strategy improves classification accuracy over standard classifiers, such as SVM and LR, as well as a number of sparse linear classifiers. Classifier weights were also more consistently assigned to neurologically-relevant brain areas. Our results thus provide further evidence for the presence of intrinsic relationships between brain structure and function.

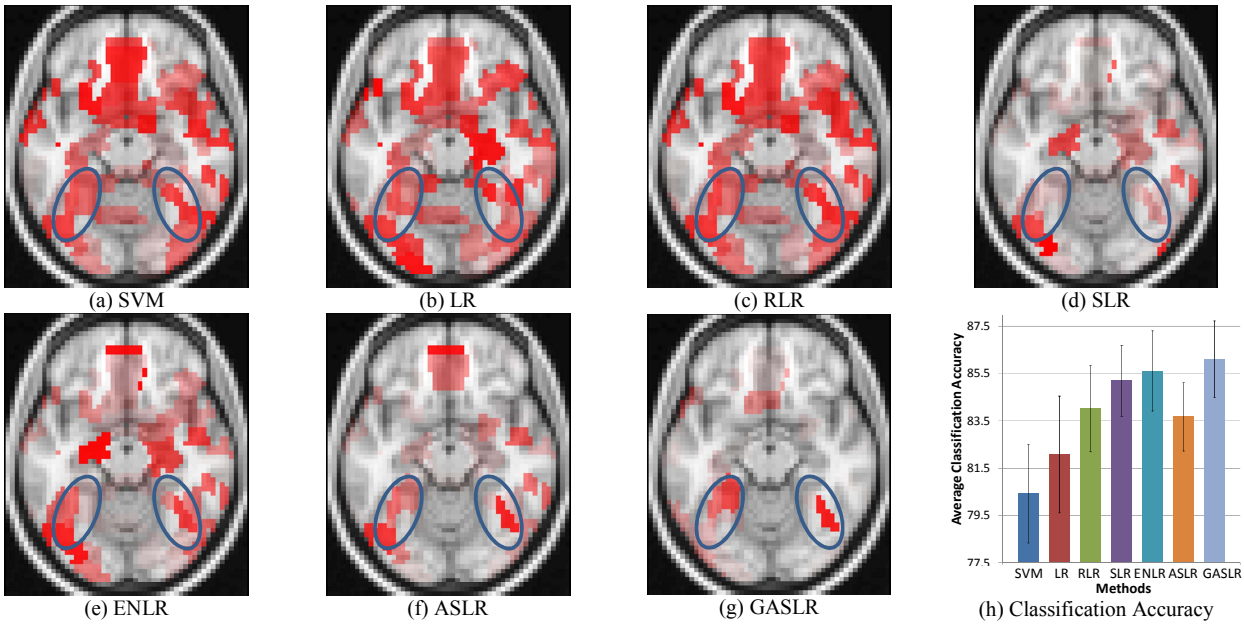


Figure 2. Classifier comparison. (a)-(g) Classifier weights averaged over 100 subsamples for an exemplar subject. Only positive weights displayed for clarity. The fusiform face area (FFA), which is involved with face recognition, is roughly encircled in blue. (h) Average classification accuracy over subjects. Adding a group connectivity prior resulted in the highest accuracy among all contrasted methods with classifier weights localized in the FFA.

ACKNOWLEDGMENT

This work was supported by the ANR grant, BrainPedia ANR-10-JCJC 1408-01, and the Berkeley INRIA Stanford grant. The data were acquired within the IMAGEN project. J.B. Poline was partly funded by the IMAGEN project, which receives funding from the E.U. Community's FP6, LSHM-CT-2007-037286. This manuscript reflects only the authors' views and the Community is not liable for any use that may be made of the information contained therein.

REFERENCES

[1] J.D. Haynes and G. Rees, "Decoding Mental States from Brain Activity in Humans," Nat Rev. Neurosci., vol. 7, pp. 523–534, 2006.

[2] T. Mitchell, R. Hutchinson, R. Niculescu, F. Pereira, X. Wang, M. Just, and S. Newman, "Learning to Decode Cognitive States from Brain Images," Mach. Learn., vol. 57, pp. 145–175, 2004.

[3] S. Hanson and Y. Halchenko, "Brain Reading Using Full Brain Support Vector Machines for Object Recognition: There is no Face Identification Area," Neural Comput., vol. 20, pp. 486–503, 2008.

[4] O. Yamashita, M. Sato, T. Yoshioka, F. Tong, and Y. Kamitani, "Sparse Estimation Automatically Selects Voxels Relevant for the Decoding of fMRI Activity Patterns," NeuroImage, vol. 42, pp. 1414–1429, 2008.

[5] S. Ryali, K. Supekar, D.A. Abrams, and V. Menon, "Sparse Logistic Regression for Whole-brain Classification of fMRI Data," NeuroImage, vol. 51, pp. 752–764, 2010.

[6] M.K. Carroll, G.A. Cecchi, I. Rish, R. Garg, and A.R. Rao, "Prediction and Interpretation of Distributed Neural Activity with Sparse Models," NeuroImage, vol. 44, pp. 112–122, 2009.

[7] B. Ng and R. Abugharbieh, "Generalized Sparse Regularization with Application to fMRI Brain Decoding," Proc. International Conference on Information Processing in Medical Imaging (IPMI 2011), vol. 6801, pp. 612−623, 2011.

[8] R. Jenatton, A. Gramfort, V. Michel, G. Obozinski, F. Bach, and B. Thirion, "Multi-scale Mining of fMRI Data with Hierarchical Structured Sparsity," Proc. Int. Workshop Pattern Recognition in NeuroImaging (PRNI 2011), May 2011, pp. 69–72.

[9] V. Michel, A. Gramfort, G. Varoquaux, E. Eger, and B. Thirion, "Total Variation Regularization for fMRI-based Prediction of Behavior," IEEE Trans. Med. Imaging, vol. 30, pp. 1328–1340, 2011.

[10] L. Grosenick, B. Klingenberg, B. Knutson, and J.E. Taylor, "A Family of Interpretable Multivariate Models for Regression and Classification of Whole-brain fMRI Data," arXiv:1110.4139v1, 2011.

[11] M. van Gerven, B. Cseke, F.P. de Lange, and T. Heskes, "Efficient Bayesian Multivariate fMRI Analysis Using a Sparsifying Spatio-temporal Prior," NeuroImage, vol. 50, pp. 150–161, 2010.

[12] J.S. Damoiseaux and M.D. Greicius, "Greater than the Sum of its Parts: A Review of Studies Combining Structural Connectivity and Resting-state Functional Connectivity," Brain Struct. Funct., vol. 213, 525–533, 2009.

[13] S. Yan, D. Xu, B. Zhang, H.J. Zhang, Q. Yang, and S. Lin, "Graph Embedding and Extension: A General Framework for Dimensionality Reduction," IEEE Trans. Pat. Ana. Mach. Intell., vol. 29, pp. 40–51, 2007.

[14] M. Schmidt, D. Kim, and S. Sra, "Projected Netwon-type Methods in Machine Learning," in Optimization for Machine Learning, S. Sra, S. Nowozin, S. Wright, Eds. MIT Press, 2011, pp. 312–316.

[15] C.B. Goodlett, P.T. Fletcher, J.H. Gilmore, and G. Gerig, "Group Analysis of DTI Fiber Tract Statistics with Application to Neurodevelopment," NeuroImage, vol. 45, pp. S133–S142, 2009.

[16] D.K. Jones, S.C.R. Williams, D. Gasston, M.A. Horsfield, A. Simmons, and R. Howard, "Isotropic Resolution Diffusion Tensor Imaging with Whole Brain Acquisition in a Clinically Acceptable Time," Hum. Brain Mapp., vol. 15, pp. 216–230, 2002.

[17] N. Toussaint, J.C. Souplet, and P. Fillard, "MedINRIA: Medical Image Navigation and Research Tool by INRIA," Proc. MICCAI Workshop on Interaction in Medical Image Analysis and Visualization, pp. 1–8, 2007.

[18] V. Michel, A. Gramfort, G. Varoquaux, E. Eger, C. Keribin, and B. Thirion, "A Supervised Clustering Approach for fMRI-based Inference of Brain States," Pat. Recog., vol. 45, pp. 2041–2049, 2012.