A Spatial Regularized Regression Method for Basal Ganglia Division Using fMRI Signals

Yiming Zhang
Department of Electrical and Computer Engineering
The University of British Columbia
ymingz@ece.ubc.ca

Abstract

The parcellation of brain Regions-of-Interest(ROI) is a challenging problem in functional magnetic resonance imaging (fMRI) studies. Labeling of ROIs based on anatomical structures is a time consuming task and the accurate definition of some region boundaries is still in debate. To relax the demands of prior knowledge and improve the computational efficiency, data driven approaches are of great interest. In addition, for the purpose of connectivity estimation, ROI definition approaches should be able to select functional and spatially homogeneous voxels within each region. In this paper, a connectivity based method using spatial regularized regression is proposed. A non-separable L-1 norm optimization problem is formulated and solved by smooth proximal gradient method (SPG) and alternating direction method of multipliers(ADMM). With the application to Basal ganglia division using fMRI signals, the proposed method is able to efficiently select the connectivity homogeneous voxels within Basal ganglia nuclei.

1 Introduction

Functional magnetic resonance imaging (fMRI) is one of the most popular non-invasive neuroimaging technologies that measures brain activities by detecting associated changes in blood flow. Most of fMRI studies are concerned with the detection of foci of the activation, called activity studies. Recently, the interest of interactions between different brain regions has been growing, i.e. connectivity studies. Generally, two ways to access the brain connectivity networks are voxel based approach and Regions-of-interest (ROI) based approach.

Voxel level approaches are well developed such as the correlation threshold method which generate the connectivity map with thousands of voxels involved. The voxel-based approaches are more sensitive to the noise, which make the results unstable. It's often difficult to interpret the results and incorporate with the neurophysiological knowledge . In addition, some sophisticated approaches including graphical models which are not able to handle a large number of variables favor more the ROI-based connectivity modeling approaches with a small number of variables involved. To tackle with these concerns, the ROI-based approaches are usually adopted. However, we have to carefully define the ROIs with accurate voxels selected. Otherwise, the signal fluctuations within a single ROI may be the result of the influence of multiple underlying brain networks or cognitive states.

The brain ROIs could be defined based on their anatomical structures or functional [1]. However, delineation of a ROI is a challenging task. The exact parcellations of the cortex and the precise boundaries of many brain areas are still in debate. Recently, several data driven methods have been developed based on functional specializations, such as clustering approach based on a distance measure [2] and seed region-growing method using brain spatial features [3]. For the purpose of connectivity estimation, ROI definition approaches should be able to select functional and spatially homogeneous voxels within each ROI. The connectivity based ROI parcellation methods have been of great interest. A connectivity based segmentation of ROIs has been proposed in [4]. In their method, canonical correlation analysis (CCA) is applied to define ROIs as sets of voxels with similar connectivity patterns to other ROIs. Barnes et.al adopt the community detection algorithm in graph theory to define the brain area using resting-state functional connectivity networks [5].

Among those brain regions whose boundaries are difficult to be identified, Basal ganglia receives a lot of attentions. Basal ganglia is widely accepted as a primary substrate in brain for the motor, cognitive, and emotional processing [6]. In basal ganglia nuclei, there are two striatal parts that play difference roles: dorsal striatum and ventral striatum. Dorsal striatum is important for performance while ventral striatum is critical for learning. Understanding the location and functional connectivity patterns of basal ganglia divisions would improve cognitive neuroscience investigations [7]. Unfortunately, a clearly anatomical boundary between these two main striatal parts has not been well defined.

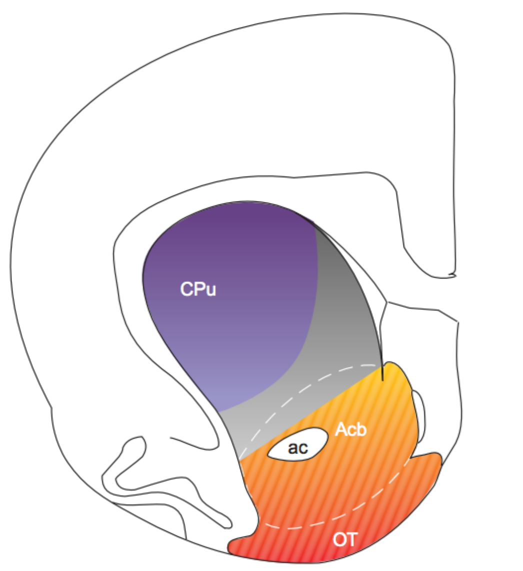


Figure 1: Various ways fo subdividing Basal ganglia.(Duplicated from [6])

As shown in Figure 1, Basal ganglia consists of caudate-putamen complex(CPu), nucleus accumbens(Acb) and striatal elements of the olfactory tubercle(OT). Some researchers consider the border between CPu and Acb (upper white dashed line), while others insist that dorsal striatum is only the purple zone and other areas, including orange and gray zones consist ventral striatum [8].

Previous studies have revealed that subregions in the basal ganglia are interconnected with different regions in cerebral cortex [9]. For instance, dorsal and ventral striatum ROIs were reported to have different patterns of functional connectivity with the hippocampus, i.e. ventral striatum receives connections from hippocampus while dorsal striatum does not [9]. As a result, we can use connectivity based method to divide basal ganglia area into subregions.

In this paper, a data-driven approach that combines brain connectivity patterns and spatial features is proposed to divide Basal ganalia. Spatial regularized regression model is performed in whole basal ganglia region to obtain the connectivity weight of each voxel. Hierarchical clustering with a distance measure that combines spatial information and connectivity weight is then applied to separate the whole region into two divisions. The results demonstrate that the proposed method could efficiently define the brain subregions and is potential for the connectivity based brain parcellationd.

2 Materials and Method

In this section, we will first introduce the spatial regularized regression model to learn the brain connectivity networks with the candidate voxels, and then we will define the ROI by grouping the voxels with the same connectivity features as well as the similar spatial properties.

2.1 fMRI data and subject

A healthy subject was recruited from Pacific Parkinson's Research Center (PPRC) at the University of British Columbia (UBC). All the experiments were approved by the Ethics Board at UBC. A 3 Tesla scanner (Philips Gyroscan Intera 3.0T; Philips Medical Systems, Netherlands) equipped with a head-coil was used to collect data in the resting state. fMRI was sampled at 0.5 Hz. Predefined Basal ganglia and Hippocampus regions were selected in our study. Each region contained a large number of voxels.

2.2 Spatial regularized regression model

Suppose there are J voxels with P time points in the target brain region we want to divide. Let $X = \{\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_J}\}$ be time courses of the target region, where each $\mathbf{x_i}$ denoting time course of one voxel. To access the different connectivity properties of subregions in the target region, one ancillary region is adopted denoted as $Y = \{y_1, y_2, \ldots, y_P\}^T$ where y_i representing mean value among voxels at time point i since we take this region as a whole. To define basal ganglia divisions, X contains the voxels in Basal ganglia and Y represent hippocampus. Then a spatial regularized regression model can be defined as:

$$Y = X\beta + \lambda \sum_{j=1}^{J-1} |\beta_{j+1} - \beta_j| + \epsilon, \tag{1}$$

where β is a column vector that represents weights of connectivity, $\Omega(\beta) = \sum_{j=1}^{J-1} |\beta_{j+1} - \beta_j|$ is the standard fused lasso penalty [9] which severs as a distance penalty of connectivity and ϵ represents the residual vector which assumed to be Gaussian distributed.

It's noted that the standard fused lasso penalty could be formalized as

$$\Omega(\beta) = \sum_{j=1}^{J-1} |\beta_{j+1} - \beta_j| = ||F\beta||_1, \tag{2}$$

where

$$F_{ij} = \begin{cases} 1 & j = i + 1 \\ -1 & j = i \\ 0 & \text{otherwise}, \end{cases} \tag{3}$$

As a result, the model could be formed as,

$$Y = X\beta + \lambda ||F\beta||_1 + \epsilon \tag{4}$$

Our problem can be generalized to a optimization problem,

$$\min_{\beta} \quad \frac{1}{2}||X\beta - y||_2^2 + \lambda ||F\beta||_1. \tag{5}$$

This objective function consists of one L_2 norm term and one L_1 norm term. Both of them are convex. However, it's not easy to directly use proximal method to solve this problem due to the difficulty of calculating the gradient of proximal operator. Alternatively, some other approaches could be adopted such as CVX (i.e., transforming to a symmetric cone program and solving with an interior-point method)[11], smooth proximal gradient method (SPG) which obtain the smooth approximation to the non-smooth structured sparsity-inducing penalties [12], and alternating direction method of multipliers(ADMM) [13]. In this paper, ADMM and SPG methods will be introduced and formulated to solve this problem.

2.3 ADMM algorithm

ADMM is a simple but powerful algorithm that is suitable for distributed convex optimizations, and in particular for large scale problems arising in nowadays applied statistics and machine learning[13]. ADMM takes the form of *decomposition-coordination* procedure, in which the solutions to small local subproblems are coordinated to find a solution to a large global problem. ADMM, in general, is a method that combines the benefits of dual decomposition and augmented Lagrangian methods for constrained optimization [13]. The general form of ADMM could be written as

$$\begin{array}{ll} \text{minimize} & f(x)+g(z) \\ \text{subject to} & Ax+Bz=c \end{array}$$

where we assume f and g are convex functions. Using ADMM, we could split the optimization problem into two parts which are often easier to solve. Then we iteratively execute three steps: x-minimization step, z-minimization step and a dual variable update.

Let $f(\beta)=(1/2)||X\beta-Y||_2^2$ and $g(z)=\lambda||z||_1$. It's easy to prove that f and g are convex, which satisfy the assumption of ADMM. So the problem could be formed as,

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}||X\beta-Y||_2^2+\lambda||z||_1 \\ \text{subject to} & F\beta-z=0 \end{array} \tag{6}$$

Algorithm 1 ADMM for Generalized Lasso Penalty (modified from algorithm in [13])

Input:
X, Y, λ, ρ
Initialization
$\beta^0=z^0=u^0=0$
Iteration:
For $k=0,1,2,\ldots$, until convergence
1. $\beta^{k+1}=(X^TX+\rho F^TF)^{-1}(X^TY+\rho F^T(z^k-u^k))$
2. $z^{k+1}=S_{\lambda/\rho}(F\beta^{k+1}+u^k)$
3. $u^{k+1}=u^k+Fx^{k+1}-z^{k+1}$

Output:
$\hat{\beta}=\beta^{k+1}$

where $\rho>0$ is the augmented Lagrangian parameter, the *soft thresholding operator* S is defined as

$$S_{\lambda/\rho}(a)=\begin{cases} a-\lambda/\rho & a>\lambda/\rho \\ 0 & |a|\leq\lambda/\rho \\ a+\lambda/\rho & a<-\lambda/\rho \end{cases} \tag{7}$$

2.4 SPG algorithm

Smoothing proximal gradient(SPG) method is designed for dealing with a broad family of sparsity-inducing penalties of complex structures. It introduces a smooth approximation to the structured sparsity inducing penalty instead of optimizing the original objective function directly. Additionally, it is applicable to both uni- and multi-task sparse structures. For spatial regularized penalty(Equ. 2), it could be taken as a general special form of graph-guided-fused-lasso penalty, for which SPG algorithm is suitable.

To implement SPG algorithm, we first rewrite the generalized lasso penalty as

$$||F\beta||_1=\max_{||\alpha||_\infty\leq1}\alpha^TF\beta \tag{8}$$

where $\alpha\in Q=\{\alpha|\|\alpha\|_\infty\leq1\}$ is a vector of auxiliary variables associated with $\|F\beta\|_1$, then we could get a smooth approximation to $\|F\beta\|_1$ according to [14],

$$f_\mu(\beta)=\max_{\alpha\in Q}(\alpha^TF\beta-\mu d(\alpha)) \tag{9}$$

where μ is a positive smoothness parameter and $d(\alpha)$ is a smoothing function defined as $\frac{1}{2}\|\alpha\|_2^2$. According to [14], $f_\mu(\beta)$ is smooth in β with a simple form of the gradient

$$\nabla f_\mu(\beta) = F^T \alpha^* \tag{10}$$

where α^* is the optimal solution to the smooth approximation penalty, here we assume that for any μ, $f_\mu(\beta)$ is convex and continuously differentiable. For the optimal solution α^*, we have

$$\alpha^* = S(\frac{F\beta}{\mu}) \tag{11}$$

where S is the projection operator defined as follows

$$S(x) = \begin{cases} x & -1 \leq x \leq 1 \\ 1 & x > 1 \\ -1 & x < -1 \end{cases} \tag{12}$$

Then with smooth approximation of the L-1 norm penalty, we could apply proximal gradient to solve this problem.

Let

$$h(\beta) = g(\beta) + f_\mu(\beta) = \frac{1}{2}\|Y - X\beta\|_2^2 + \lambda f_\mu(\beta) \tag{13}$$

then the gradient is given as

$$\nabla h(\beta) = X^T(X\beta - Y) + \lambda F^T \alpha^* \tag{14}$$

$\nabla h(\beta)$ is Lipschitz-continuous with Lipschitz constant

$$L = \lambda(X^T X) + L_\mu = \lambda_{max}(X^T X) + \frac{\|F\|^2}{\mu} \tag{15}$$

where $\lambda_{max}(X^T X)$ is the largest eigenvalue of $(X^T X)$.

Then the algorithm is formulated

Algorithm 2 SPG for Generalized Lasso Penalty (modified from algorithm in [12])

Input:

X, Y, F, β^0, Lipschitz constant L, desired accuracy ϵ, λ

Initialization

set $\mu = \frac{\epsilon}{2D}$ where $D = \max_{\alpha \in Q} \frac{1}{2}\|\alpha\|_2^2$, $\theta_0 = 1$, $w^0 = \beta$

Iteration:

For $t = 0, 1, 2, \ldots$, until convergence of β^t:

1. Compute $\nabla h(w^t) = X^T(Xw^t - Y) + \lambda F^T \alpha^*$
2. Solve the proximal operator:

$$\beta^{t+1} = h(w^t) + \langle \beta - w^t, \nabla h(w^t) \rangle + \frac{L}{2}\|\beta - w^t\|_2^2$$

3. Set $\theta_{t+1} = \frac{2}{t+3}$.
4. Set $w^{t+1} = \beta^{t+1} + \frac{1-\theta_t}{\theta_t}\theta_{t+1}(\beta^{t-1} - \beta^t)$

Output:

$\hat{\beta} = \beta^{t+1}$

2.5 Clustering voxels according to connectivity weight and spatial position

In this clustering analysis, a new distance measure that combines normalized Euclidean distance and normalized connectivity weight differences between voxels is defined. The distance defined between voxel x and y is

$$d_{x,y} = \frac{\gamma(\beta_x - \beta_y)}{\|\beta\|} \cdot \frac{1}{d}\sqrt{\sum_{j=1}^{3}(x_j - y_j)^2} + (1 - \gamma)\frac{1}{d}\sqrt{\sum_{j=1}^{3}(x_j - y_j)^2} \tag{16}$$

where x_j and y_j are vectors containing 3-D coordinates of each corresponding voxel, β is the connectivity weight, $\|\beta\| = \sum_{i=1}^{J} |\beta_i|$ is L_1 norm normalization factor of β, $d = \sum_{i=1}^{J} \sum_{j=i+1}^{J} d_{ij}$ is normalization factor of Euclidean distance where d_{ij} represents the Euclidean distance between voxel i and j, $\gamma \in [0, 1]$ is a parameter of the weight of connectivity in this distance measure. With this distance defining, we use agglomerative hierarchical clustering method and stop when the number of clusters decreases to 2.

2.6 Overall work flow

To summarize, the overall work flow is as follows,

Step 1. Load the whole fMRI signals with the target voxels and ancillary pre-defined ROI. Check the spatial continuous of the whole region data. If the candidate voxels are not spatially continuous, run Gaussian spatial smoother on the data to make it spatially continuous.
Step 2. Apply the spatial regularized regression on X and get the connectivity weight β between X and Y.
Step 3. Apply hierarchical clustering approach to cluster the voxels with similar connectivity weights and close spatial positions, stop when with at last two clusters
Step 4. Use pointer to get the original spatial position of each cluster and then check the spatial continuity within each cluster, if not continuous, run Gaussian spatial smoother on the data to make it spatial continuous, generate two spatial continuous clusters.
Step 5. Define each cluster as two ROIs.

The overall work is basically a clustering analysis. In clustering process, we cluster the object through connectivity weights obtained from generalized lasso penalty and also spatial position.

3 Results

In this section, we redefine and separate Basal ganglia divisions based on the approach above. Along with ADMM and SPG methods, we also tried CVX on the optimization problem and compare the results with three methods.

3.1 Data Intake

In the fMRI data we used for analysis, there are two pre-defined ROIs that represent dorsal striastum and ventral striastum. In whole Basal ganglia area, there are 1221 voxels.

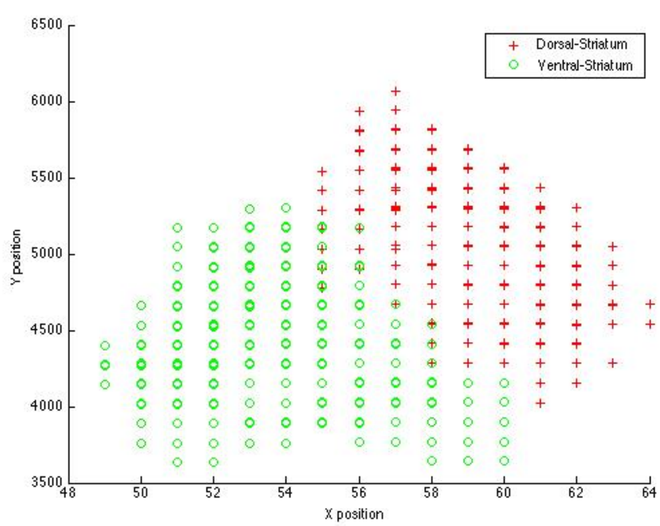


Figure 2: Pre-difined division in Basal ganlia

Figure 2 shows the pre-defined divisions in Basal ganglia of one healthy subject from our fMRI data base, the Z-position is missing due to the fact that all voxels have same Z-position in our data. From this result we can see the boundary between the two pre-defined ROIs is not very clear, and also, this boundary does not consider the different connectivity pattern between two striastum which could result in inaccurate parcellation. In the following part, spatial regression will be performed to get the connectivity weight of each voxel. Then connectivity weights will be incorporated with spatial positions to get the separation of basal ganglia.

3.2 Results and Discussions

In this part, the pacellation results solved by ADMM, SPG and also CVX will be compared and discussed.

To evaluate parcellation performance , we introduce parcellation result on same data using modularity approach in [15]. Modularity is one measure of the structure of networks and graphs which is often designed measure the strength of strength of division of a net work into clusters. In [15], they proposed a method that based on functional brain network modularity to separate brain region. This method also incorporate functional connectivity with spatial information of voxels and biomedical meaning of their results has been confirmed. So we could use modularity parcellation result as a reference to interpret our results.

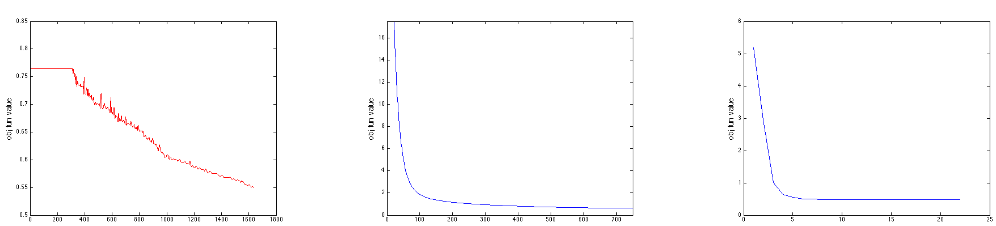
After several experiments, we found the choice of parameter λ in spatial regularized model(Equation 2) barely has effects on our final results, so λ is set as 1 in the following experiment results. However, the choice of parameter α in our self-defined distance(Equation 16) is crucial in final parcellation results, the differences of pacellation performance between different choices of α will be shown later.

First look at the optimization operation performance between three methods, ADMM, SPG and CVX.

	Optimization Performance		
Method	Operation Time(sec)	Num.Iteration	Optimal Obj-Fun Value
ADMM	66.8899	1637	0.5503
SPG	0.4610	2794	0.4906
CVX	0.0033	27	0.4912

Table 1: Optimization performance of three methods

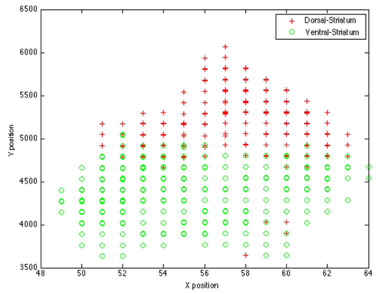
This table shows the basic information of optimization performance solving by three methods. ADMM, SPG and CVX all finished optimization task and reached same optimal objection function value level, but among them, CVX is the fastest and least iteration times while ADMM is slowest.



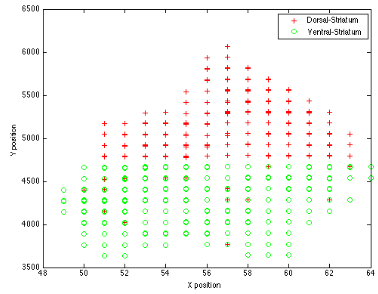
(a) ADMM (b) SPG (c) CVX

Figure 3: Objective function value change in iteration process of ADMM, SPG and CVX

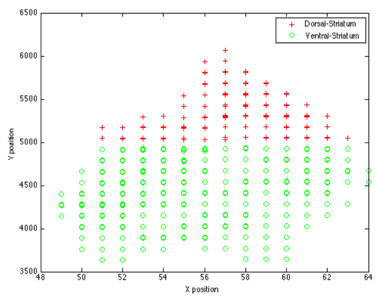
Figure 3 shows the different iteration process of these three methods. It should be clarified that the y-axis of these three plots are not at same scale : ADMM starts from a low level (10^{-1}), but the iteration process is slow and stochastic; SPG starts from high level(10^3), but the descent process is fast and nonstochastic; CVX also starts from high level(10^3), but it only takes it few steps to decent to optimal value.



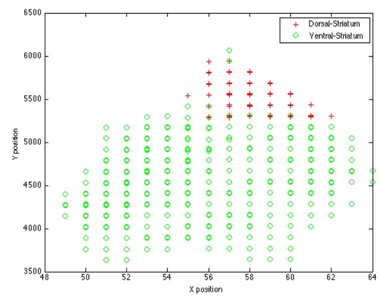
(a) Modularity



(b) $\alpha = 0.3$



(c) $\alpha = 0.5$



(d) $\alpha = 1$

Figure 4: Pacellation results with different α, with reference result by modularity method

Figure 4 shows Basal ganglia parcellation results by CVX solver with different choices of α, reference parcellation result by modularity is given. Here we just take CVX results because all the three methods reached similar optimal results and among the three approaches, CVX is the fastest. In terms of parcellation results compared with reference result, different choices of α actually decide the final parcellation result, so choosing α should be very carefully. Due to the limitation of prior knowledge of neurophysiological knowledge, quantitative evaluation cannot be given, so here we just choose α manually, automatic choosing of α is expected in future work.

4 Conclusion

In this paper, we proposed a new voxel based approach to define ROI in brain regions, using spatial regularized linear regression to set up connectivity weight for each voxel and incorporated it with spatial information to get voxels into two clusters. The approach was solved by three optimization methods, ADMM, SPG and CVX, and performed on real fMRI data of one particular biomedical problem of parcellation of divisions in Basal ganglia. Results showed the three optimization approaches solve the problem successfully while among them CVX is the fastest while ADMM is the most time-consuming. Choice of α in self-defined distance measure (Equation 16) is crucial for parcellation result. The result still need further validation and also biological interpretation.

Acknowledge

I would like to acknowledge Aiping Liu, a Phd student in my Lab for providing me the basic idea of this project and helping me in the initial implementation. I would also like to thank Dr. Martin J. McKeown, UBC chair in parkinson's research, for providing me the precious fMRI data and also biological instructions.

Appendix

Matlab code for the algorithm in this paper could be found at `https://github.com/zym268/CPSC_546_project`, The fMRI data in this paper will not be provided, the public sample fMRI data could be found at `http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-81/www/` , but the ROI defined in this public data set could be different. If anyone is interested in this experiment and original data set or has further questions, please do not hesitate to contact me.

References

[1] Rademacher, J., Caviness, V. S., Steinmetz, H., & Galaburda, A. M. (1993). Topographical variation of the human primary cortices: implications for neuroimaging, brain mapping, and neurobiology. *Cerebral Cortex*, 3(4), 313-329.

[2]Baumgartner, R., Scarth, G., Teichtmeister, C., Somorjai, R., & Moser, E. (1997). Fuzzy clustering of gradientecho functional MRI in the human visual cortex. Part I: Reproducibility. *Journal of Magnetic Resonance Imaging*, 7(6), 1094-1101.

[3]Lu, Y., Jiang, T., & Zang, Y. (2004). A split-merge-based region-growing method for fMRI activation detection. *Human brain mapping*, 22(4), 271-279.

[4]Deleus, F., & Van Hulle, M. M. (2009). A connectivity-based method for defining regions-of-interest in fMRI data. *Image Processing, IEEE Transactions on*, 18(8), 1760-1771.

[5]Barnes, K. A., Cohen, A. L., Power, J. D., Nelson, S. M., Dosenbach, Y. B., Miezin, F. M., ... & Schlaggar, B. L. (2010). Identifying basal ganglia divisions in individuals using resting-state functional connectivity MRI. *Frontiers in systems neuroscience*, 4.

[6] Mink, J. W. (1996). The basal ganglia: focused selection and inhibition of competing motor programs. *Progress in neurobiology*, 50(4), 381-425.

[7] Barnes, K. A., Cohen, A. L., Power, J. D., Nelson, S. M., Dosenbach, Y. B., Miezin, F. M., ... & Schlaggar, B. L. (2010). Identifying basal ganglia divisions in individuals using resting-state functional connectivity MRI. *Frontiers in systems neuroscience*, 4.

[8]Di Martino, A., Scheres, A., Margulies, D. S., Kelly, A. M. C., Uddin, L. Q., Shehzad, Z., ... & Milham, M. P. (2008). Functional connectivity of human striatum: a resting state FMRI study. *Cerebral cortex*, 18(12), 2735-2747.

[9]Voorn, P., Vanderschuren, L. J., Groenewegen, H. J., Robbins, T. W., & Pennartz, C. (2004). Putting a spin on the dorsalventral divide of the striatum. *Trends in neurosciences*, 27(8), 468-474.

[10]Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., & Knight, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(1), 91-108.

[11]Grant, M., Boyd, S., & Ye, Y. (2008). CVX: Matlab software for disciplined convex programming.

[12]Chen, X., Lin, Q., Kim, S., Carbonell, J. G., & Xing, E. P. (2012). Smoothing proximal gradient method for general structured sparse regression. *The Annals of Applied Statistics*, 6(2), 719-752.

[13]Boyd, S., Parikh, N., Chu, E., Peleato, B., & Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1), 1-122.

[14]Nesterov, Y. (2005). Smooth minimization of non-smooth functions. *Mathematical programming*, 103(1), 127-152.

[15]Meunier, D., Lambiotte, R., Fornito, A., Ersche, K. D., & Bullmore, E. T. (2009). Hierarchical modularity in human brain functional networks. *Hierarchy and dynamics in neural networks*, 1, 2.