A SPATIALLY CONSTRAINED LOW-RANK MATRIX FACTORIZATION FOR THE FUNCTIONAL PARCELLATION OF THE BRAIN

Alexis Benichoux, Thomas Blumensath

ISVR - University of Southampton
SO17 1BJ Southampton, UK
firstname.lastname@soton.ac.uk,

ABSTRACT

We propose a new matrix recovery framework to partition brain activity using time series of resting-state functional Magnetic Resonance Imaging (fMRI). Spatial clusters are obtained with a new low-rank factorization algorithm that offers the ability to add different types of constraints. As an example we add a total variation type cost function in order to exploit neighborhood constraints.

We first validate the performance of our algorithm on simulated data, which allows us to show that the neighborhood constraint improves the recovery in noisy or undersampled set-ups. Then we conduct experiments on real-world data, where we simulated an accelerated acquisition by randomly undersampling the time series. The obtained parcellation are reproducible when analysing data from different sets of individuals, and the estimation is robust to undersampling.

***Index Terms*—** Clustering, Low-rank, Sparse, Matrix recovery, Brain parcellation, Neuroimaging, fMRI

1. INTRODUCTION

A spatial brain parcellation is a partition of the human cerebral cortex into distinct cortical regions. Different approaches to brain parcellation shed light on different aspects of brain organisation. Of particular interest are functional parcellations derived from functional Magnetic Resonance Imaging (fMRI) data acquired during rest, which define brain regions that display homogeneous functional activity. These parcellations are an essential first step towards the modelling of human brain functional connectivity [1].

Functional parcellation thus complement traditional parcellation approaches based on postmortem studies of cytoarchitecture [2] and are valuable alternatives to diffusion based MRI methods [3].

fMRI based cortical parcellation is a challenging task, with generic clustering approaches often performing poorly on real data, especially in single subject studies or in situations, where data quality is poor or short resting state scans are acquired. This suggests the development and use of more constrained methods that exploit prior knowledge on brain structure. Of particular interest here are spatial neighbourhood constraints that exploit local functional homogeneity of brain networks [4]. Such constraints have for example, been used with spectral clustering [5] and region growing [6]. More complex priors have also been introduced, such as a manually labelled bayesian graph [7]. The present paper develops a clustering framework for functional brain parcellation that has sufficient flexibility to exploit a host of different constraints within an optimisation framework and thus allows us to impose several type of prior constraints.

Furthermore, it has recently been shown that it is possible to use compressed-sensing methods to accelerate fMRI acquisition [8] using randomly under-sampled kt-space data, a process that allows us to effectively trade off spatial and temporal resolution of fMRI data. The second contribution of this paper is thus to show that the proposed optimization framework can also be used to derive cortical parcellations directly from randomly under-sampled data, without the need to first reconstruct the full fMRI data-set, that is, without the initial super-resolution preprocessing.

It is well known that Principal Component Analysis is equivalent to a low rank approximation of the data matrix. Sparse and low rank priors have been combined to recover highly corrupted low rank matrices [9] or block diagonal low rank matrices [10]. In this paper we address a new challenge where a low rank factorization is combined with additional constraints on the decomposition factors themselves. Although we have not yet completed a full convergence analysis, the structure of our algorithm is similar to variants of Iterative Hard Thresholding algorithms [11] and many of the theoretical results for this method can be extended to our new approach.

The paper is organized as follows : first, we describe the sparse, low-rank factorization approach and then show how additional constraints can be added using an optimization framework. We then study the proposed algorithm for both constrained and unconstrained low rank clustering. The paper ends with results of experiments using both synthetical data and human resting-state FMRI observations.

We thank the EPSRC for funding under EP/K03102/1

2. CONSTRAINED OPTIMIZATION FRAMEWORK

In this section we show how matrix factorization can be used to perform constrained statistical clustering [1].

2.1. Clustering data with 1-sparse low-rank matrix factorization

Given N time series of M samples let $\mathbf{X} \in \mathbb{R}^{MN}$ be the data matrix. The K regions of interests are modelled with K cluster centres resulting in a matrix $\mathbf{D} \in \mathbb{R}^{MK}$, such that each time series can be expressed as one cluster centre plus a residual vector. Denoting by $\mathbf{E} \in \mathbb{R}^{MN}$ the matrix of residual vectors we obtain the following factorization

$$\mathbf{X} = \mathbf{DS} + \mathbf{E}, \tag{1}$$

where the assignment matrix $\mathbf{S} = (S_1, \dots S_n) \in \{0,1\}^{KN}$ contains one and only one non zero entry per column. We denote by 1-sparse the constraint on $\mathbf{S}$ that only one non zero coefficient per column is allowed. As fMRI data is only known up to an unknown scaling factor, we relax the binarry constraint on $\mathbf{S}$ and allow the non-zero entries to be real numbers. A low-rank approximation without the binary constraint on $\mathbf{S}$ is usually obtained by the Eckart-Young theorem [13]. In order to combine the low-rank approximation with a sparsity constraint, the optimization framework

$$\min_{\mathbf{S}} \|\mathbf{X} - \mathbf{DS}\|_2, \tag{2}$$

will be constraint using a 1-sparse constraint on the columns in $\mathbf{S}$, which can be enforced by a hard thresholding step after each iteration.

2.2. Total variation as a neighborhood prior

The novelty in this paper is to include additional constraints through a modification of the cost function (2). We here wish to take the spatial neighbourhood structure into account. Given a set of neighbours $f_1(n), f_2(n), f_3(n), \dots, f_6(n)$ at each spatial position of the corteical surface, we add a distance constraint defined as the mixed norm [14] of a linear transformation $\mathcal{M}$ of $\mathbf{S}$:

$$\|\mathcal{M}\mathbf{S}\|_{1,2} = \sum_{n=1}^{N} \sqrt{(S_n - S_{f_1(n)})^2 + \dots + (S_n - S_{f_6(n)})^2} \tag{3}$$

This is a simple approximation to the total variation norm on irregular sample grids[2]. In the particular case of clustering on a two dimensional flat grid, this neighborhood prior corresponds to the total variation norm, i.e. the ℓ_1 norm of the gradient of the image. Adding this total variation constraint, the cost function becomes

$$\begin{cases} \min_{\mathbf{S}} \|\mathbf{X} - \mathbf{DS}\|_2 + \lambda \|\mathcal{M}\mathbf{S}\|_{1,2} \\ \text{s.t.} \quad \forall n \leq N \quad \|S_n\|_0 = 1 \end{cases} \tag{4}$$

3. DESCRIPTION OF THE ALGORITHM

The matrix decomposition problem can be understood as a biconvex problem, that can be split into two convex problems and optimised using an alternate optimisation scheme

- estimate $\mathbf{D}$ knowing the assignment matrix $\mathbf{S}$
- estimate $\mathbf{S}$ knowing the cluster centres $\mathbf{D}$.

We choose to update the cluster centres using a straightforward least square regression. To optimise $\mathbf{S}$ we then use a gradient descent step to optimise the cost function (inclusive the spatial neighbourhood constraint) and follow this by a hard thresholding step to enforce the sparsity constraint.

3.1. Least square estimates

The matrix $\mathbf{D}$ can be updated using the least square estimate $\mathbf{D} = \mathbf{XS}^T(\mathbf{SS}^T)^{-1}$, or, as we do here, using gradient descents to avoid computational limitations.

Algorithm 1 Constrained alternate low-rank clustering

1: $\mathbf{D}^0 \in \mathbb{R}^{MK}, \mathbf{S}^0 \in \mathbb{R}^{NK}, \rho = 10^{-2}$
2: **for** $k \leq k_{\max}$ **do**
$\mathbf{C}_k = \mathbf{D}_{k-1} + \nu(\mathbf{X} - \mathbf{D}_{k-1}\mathbf{S}_{k-1})\mathbf{S}_{k-1}^T$
$\mathbf{Z}_k = \mathbf{S}_{k-1} + \mu\left(\mathbf{C}^T(\mathbf{X} - \mathbf{C}_k\mathbf{S}_{k-1}) - \rho\nabla\mathcal{L}_{\mathbf{Z}_k}\right)$
$\mathbf{S}_k = \mathcal{H}(\mathbf{Z}_k)$
$\mathbf{D}_k = \text{normalize}(\mathbf{C}_k)$
3: **end for**

Here μ and ν are step size parameters and ρ is a function of λ.

3.2. Regularized mixed-norm gradient update

To avoid issues with the discontinuous gradient of the ℓ_1 norm, we introduce a regularization parameter $\varepsilon > 0$ to obtain a differentiable constraint, as often used in image denoising algorithms [16]

$$\mathcal{L}(\mathbf{S}) = \sum_{1 \leq n \leq N} \sqrt{(S_n - S_{f_1(n)})^2 + \dots (S_n - S_{f_6(n)})^2 + \varepsilon^2)} \tag{5}$$

Several algorithms have been proposed to perform mixed norm minimization, and generalized forward-backward splitting [17] allows to combine sparsity and total variation constraint using alternate proximal descent. Our experiments on synthetic data showed that gradient descent obtains the same parcellation for $\varepsilon = 0.02$ as with generalized forward-backward proximal splitting. We chose a regularized mixed

[1] For more details, see the companion paper in this conference [12]
[2] More refined approximations can be used [15].

norm to cope with large data, since the observed convergence is a lot faster with a gradient update than with proximal update.

3.3. Sparsity constraint

Instead of the usual hard thresholding operator that keeps elements above some magnitude, we perform a pruning step, keeping only the largest coefficient of each column of $\mathbf{S}$. The resulting scaling ambiguity between the cluster centres and the assignment scalars is controlled by a normalization of the columns of $\mathbf{D}$.

The resulting process is summarized in Algorithm 1.

4. NUMERICAL EXPERIMENTS

One of the difficulties of brain parcellation is that no ground truth is available to evaluate the performances of different approaches. Therefore we first conduct a series of experiments on synthetic data to ensure the convergence of our algorithm, then present the obtained results on real world data.

4.1. Dice similarity

The matrix factorization framework (1) suffers from both scaling and permutation ambiguities. We controlled the scaling with a normalization step, on the other hand, the permutation ambiguity is inherent to the clustering problem and cannot be avoided without additional assumptions. Hence we measure the performance of our parcellation up to a permutation of the assignments within each cluster, and a global permutation of the cluster centres.

In brain imaging, assignment errors are typically measured using Dice similarity [18] between two sets X and Y defined by $d(X,Y) = 2\frac{\text{card}\, X\cap Y}{\text{card} X + \text{card} Y}$.

Between two assignment matrices $\mathbf{S}$ and $\mathbf{S}'$, we first find for each cluster of $\mathbf{S}$ the cluster of $\mathbf{S}'$ with the maximum dice similarity, then take the mean over all clusters of $\mathbf{S}$.

4.2. Experiments on synthetic data

In this section we show experimentally that the constraint allows to cluster data in noisy and undersampled set ups.

4.2.1. Description of the data

We generate a $N = 64 \times 64$ spatial grid containing 40 clusters, with non homogeneous partition of the observations as such: the $K = 40$ cluster centres are dropped at random, and we let each region spread randomly until the grid is full. We set the number of time samples to $M = 1000$. The matrix $\mathbf{D} \in \mathbb{R}^{MK}$ is generated using a gaussian centred distribution of unit variance, then each column of $\mathbf{D}$ is normalized. The residual matrix $\mathbf{E} \in \mathbb{R}^{MN}$ is generated with a similar distribution, of variance 0.1. With this construction the average signal to noise ratio between $\mathbf{X}$ and $\mathbf{DS}$ is between 0 dB and 3 dB. The experiments are averaged among 25 trials, and the confidence interval is below ± 0.02.

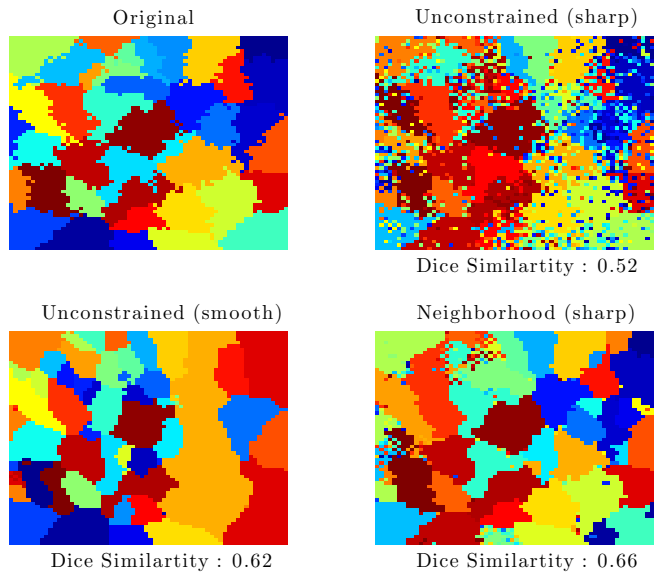

Fig. 1. Comparison of the assignment matrices on synthetic data for an undersampling ratio of 35% between unconstrained and neighborhood constrained frameworks

We simulate undersampling by artificially discarding randomly selected Fourier coefficients of each time series. Note that this process does not correspond to the reality of kt space undersampling, it is designed to study the behavior of alternate low rank factorization with additional constraint. We see the results as a proof of concept more than a practical opportunity for the moment, and keep in mind that a more realistic set-up will need to be further investigated.

4.2.2. Choice of the parameters

The convergence is observed in 50 iterations, we set the maximum iteration to 200. The gradient steps are set to 10^{-2} empirically. It is also worth noting that unlike usual descent algorithms the constraints might be uneffective if the gradient step ρ is too small, due to the 1-sparse hardthresholding step.

4.2.3. Smoothing kernel

We simulate the preprocessing of the data with spatial smoothing obtain by convolution with a two dimensional 12×12 gaussian kernel of unit variance. We will refer to *sharp* method if a framework is applied to the original data and *smooth* method when applied to the resulting smoothed data. This operation eases the spatial clustering at the centre of the clusters. Our goal is to compare this operation with a spatial total variation constraint in undersampling conditions.

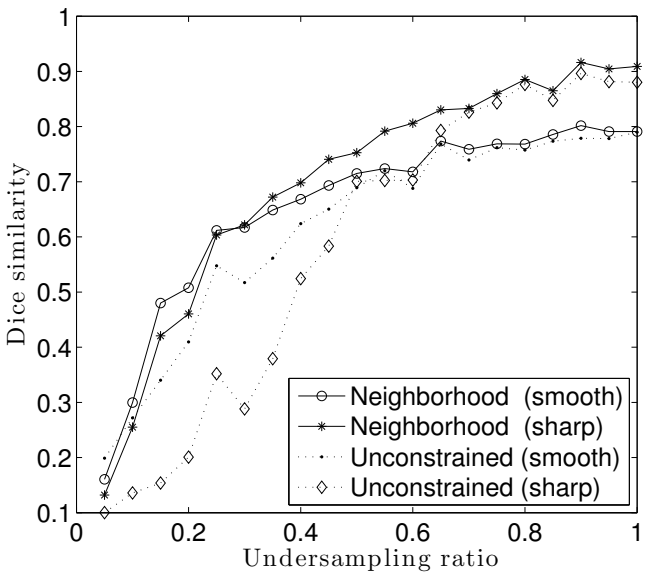

Fig. 2. Role of the neighborhood constraint : Dice similarity of the assignments for different level of undersampling

4.2.4. Performance of the unconstrained low-rank clustering

As shown Fig. 2, the smoothing kernel tends to lower the performances in sufficiently sampled data. The neighborhood constrained and the unconstrained methods on sharp data performs similarly if the undersampling ratio is above $50\ \%$.

In this regime, unlike most clustering algorithms, our framework without any neighboring constraint nor spatial smoothing is able to recover the parcellation with a Dice similarity higher than 0.75. Although the problem formulation of (1) differs from the known theoretical guarantees of sparse and low rank recovery, this results suggest an identifiability of the problem suitable for a compressed-sensing framework.

4.2.5. Performance depending on the undersampling ratio

We conducted the experiments for different sampling ratios between 2.5% and 95%. We present Fig. 2 the resulting Dice similarity averaged over 30 different datasets with the same spatial grid. If the undersampling ratio is below $50\ \%$, the neighborhood constrained method on both sharp and smoothed data performs significativly better than the unconstrained method. Moreover, the similarity of the performances when applied to smooth and sharp data shows that the neighborhood contraint behave like a smoothing spatial constraint. In contrast, when using the unconstrained method, there is a gap between the performances on the smooth and sharp data.

4.2.6. Comment on the role of the neighborhood constraint

A detailed illustration of the resulting assignments is displayed Fig. 1. We chose a configuration where the smooth and sharp neighborhood constrained methods perform with a comparable Dice similarity around 0.6, for an undersampling ratio of 35%. The neighborhood constraint preserves the edges of each cluster, whereas heavy spatial smoothing loses more precision far from the cluster centred.

5. EXPERIMENTS ON RESTING STATE FMRI DATA

In the section we illustrate the possibilities of our clustering algorithm using datasets provided by the Human Connectome Project.

5.1. Description of the data

We use resting state FMRI data acquired with a 2 mm isotropic spatial resolution and a 1.4 s temporal resolution. The data was processed using the project's preprocessing standards, involving cortical surface modelling, bias field correction, structural registration, noise reduction, surface smoothing.

The resulting data contain $N = 64984$ time series of length $M = 4500$, for 66 subjects. In order to validate our method by cross-validation we splitted the data into two data matrices averaged over 33 subjects.

5.2. Comparison between two datasets

As there is no ground truth available, we use the stability between the estimation on two different datasets as a performance measure. The unconstrained method shows a Dice similarity of 0.49. We illustrated the stability of the clustering Fig. 3. In order to visualize the results we split the clustering into connected components, and discard all regions with less than 200 elements. As a result the Dice similarity is improved to 0.51.

For an undersampling ratio of 40%, the unconstrained method fails completely with a Dice similarity of 0.04 compared to the full sampled data, whereas the total variation constrained method reachs a Dice similarity of 0.24.

6. CONCLUSION AND PERSPECTIVES

We presented a new matrix factorization interpretation of the widely studied statistical clustering problem. We proposed an algorithm which tackles the corresponding optimization problem, with an experimental study of an example brain parcellation on real data.

The quality of the results obtained on simulated data suggests a theoretical study of the guarantees of identification of low-rank 1-sparse unconstrained problem. In addition, the neighbouring constraint is a successful example of the use of

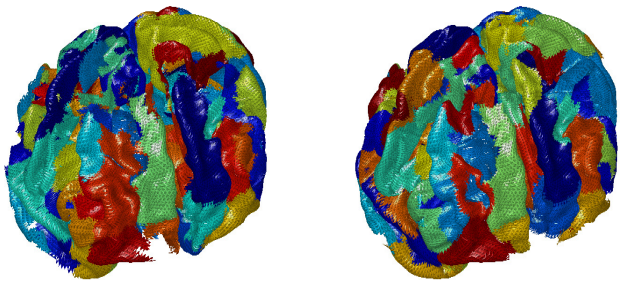

Fig. 3. Comparison of the brain parcellation results obtained between the two datasets

cost functions for the regularization of brain parcellation using more complex priors.

REFERENCES

[1] B. Biswal, Y. F. Zerrin, V. M. Haughton, and J. S. Hyde, "Functional connectivity in the motor cortex of resting human brain using echo-planar MRI," *Magnetic resonance in medicine*, vol. 34, no. 4, pp. 537–541, 1995.

[2] K. Zilles and K. Amunts, "Centenary of brodmann's map - conception and fate," *Nat. Rev. Neurosci.*, vol. 11, pp. 139–145, 2010.

[3] T. Behrens, H. Johansen-Berg, M. Woolrich, S. Smith, C. Wheeler-Kingshott, P. Boulby, G. Barker, E. Sillery, K. Sheehan, O. Ciccarelli, A. Thompson, J. Brady, and P. Matthews, "Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging," *Nat. Neurosci.*, vol. 6, pp. 750–757, 2003.

[4] B. Thirion, G. Flandin, P. Pinel, A. Roche, P. Ciuciu, and J.-B. Poline, "Dealing with the shortcomings of spatial normalization: Multi-subject parcellation of FMRI datasets," *Human brain mapping*, vol. 27, no. 8, pp. 678–693, 2006.

[5] R. C. Craddock, G. James, P. E. Holtzheimer, X. P. Hu, and H. S. Mayberg, "A whole brain FMRI atlas generated via spatially constrained spectral clustering," *Human brain mapping*, vol. 33, no. 8, pp. 1914–1928, 2012.

[6] T. Blumensath, T. E. Behrens, and S. M. Smith, "Resting-state FMRI single subject cortical parcellation based on region growing," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*, pp. 188–195, Springer, 2012.

[7] B. Fischl, C. van der Kouwe, A.and Destrieux, E. Halgren, F. Ségonne, D. H. Salat, E. Busa, L. J. Seidman, J. Goldstein, D. Kennedy, *et al.*, "Automatically parcellating the human cerebral cortex," *Cerebral cortex*, vol. 14, no. 1, pp. 11–22, 2004.

[8] M. Chiew, S. Smith, P. Koopmans, T. Blumensath, and K. Miller, "k-t faster: a new method for the acceleration of resting state fmri data acquisition," in *The 21st annual meeting of the ISMRM (International Society of Magnetic Resonance in Medicine) - Discovery, Innovation, and Application Advancing MR for Improved Health*, pp. 20 – 26, 2013.

[9] E. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *Journal of the ACM (JACM)*, vol. 58, no. 3, p. 11, 2011.

[10] E. Richard, P.-A. Savalle, and N. Vayatis, "Estimation of simultaneously sparse and low rank matrices," *arXiv preprint arXiv:1206.6474*, 2012.

[11] T. Blumensath, "Accelerated iterative hard thresholding," *Signal Processing*, vol. 92, no. 3, pp. 752–756, 2012.

[12] T. Blumensath, "Sparse matrix decompositions for clustering," in *The 22nd European Signal Processing Conference (EUSIPCO 2014)*, 2014.

[13] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, 1936.

[14] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259–268, 1992.

[15] M. Meyer, M. Desbrun, P. Schrder, and A. Barr, "Discrete differential-geometry operators for triangulated 2-manifolds," in *Visualization and Mathematics III* (H.-C. Hege and K. Polthier, eds.), Mathematics and Visualization, pp. 35–57, Springer Berlin Heidelberg, 2003.

[16] G. Peyré, "The numerical tours of signal processing-advanced computational signal and image processing," *IEEE Computing in Science and Engineering*, vol. 13, no. 4, pp. 90–98, 2011.

[17] H. Raguet, J. Fadili, and G. Peyré, "A generalized forward-backward splitting," *SIAM Journal on Imaging Sciences*, vol. 6, no. 3, pp. 1199–1226, 2013.

[18] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.