fMRI unmixing via properly adjusted Dictionary Learning

Yannis Kopsinis[1], Harris Georgiou[1], Sergios Theodoridis[1]

[1]University of Athens, Dept. of Informatics and Telecommunications, Athens, Greece.

Friday, Sep 5, 2014

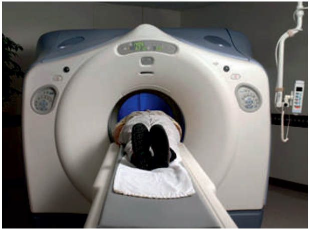
Outline

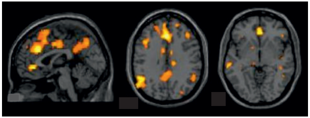
1. functional Magentic Resonance Imaging (fMRI)
2. fMRI data unmixing
3. Dictionary based fMRI data unmixing
4. Performance using synthetic data

functional Magnetic Resonance Imaging (fMRI)

Function

- A Non-invasive tool for detecting brain activity along time
- exploits the different magnetic properties of oxygen-saturated versus oxygen-desaturated hemoglobin
- It detects localized changes in the hemodynamic flow of oxygenated blood
- Indirect way to measure brain activity.
- Aim: To "understand" the brain.

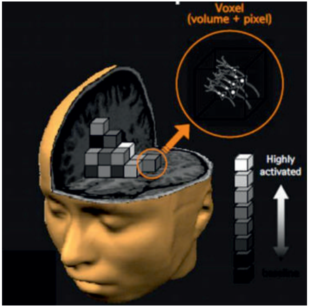


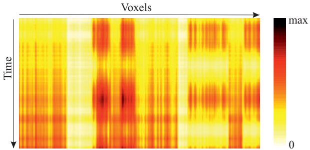


functional Magnetic Resonance Imaging (fMRI)

fMRI Data

- Typical Sampling characteristics: 3-D grid of 3-5 mm of elementary 'cubes' (*voxels*), typically 64x64x48 voxels per time instance, is acquired every 1-2 seconds.
- The full amount of data is collected in a data matrix $\boldsymbol{Y} \in \mathbb{R}^{t,n}$. Each column of $\boldsymbol{Y}$ represents the evolution in time of the values of a certain voxel.

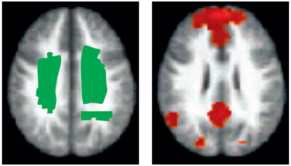




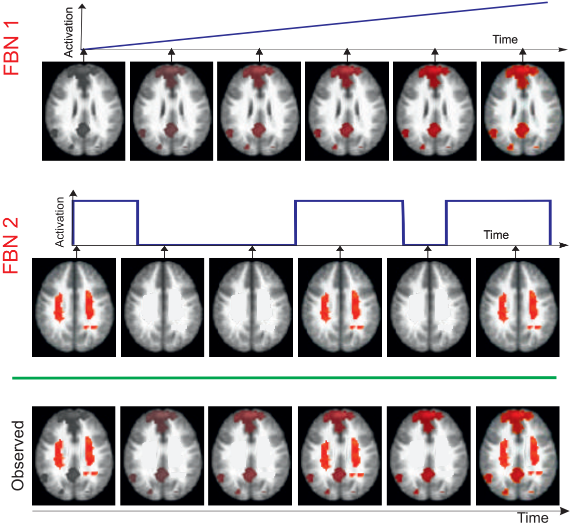
Functional Brain Networks (FBNs)

Characterization

- Comprise a number of *segregated* specialized small brain regions
- Each FBN forms a Spatial Map.
- Exhibit strong functional connectivity, which is expressed as strong coherence in the activation time-pattern Time Course.
- They are related to low-level brain functions.
- Examples of FBNs: the visual, sensorimotor, auditory, default-mode, dorsal attention



Functional Brain Networks (FBNs)



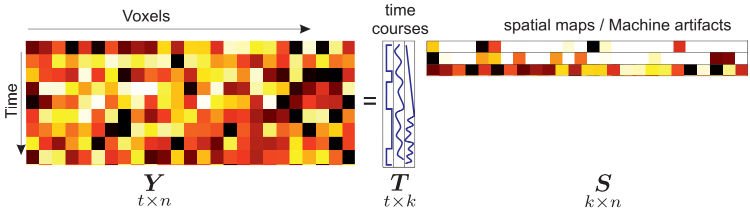
Functional Brain Networks (FBNs)

Mathematical modeling

$$\boldsymbol{Y} = \boldsymbol{T}_{\cdot,1}\boldsymbol{S}_{1,\cdot} + \boldsymbol{T}_{\cdot,2}\boldsymbol{S}_{2,\cdot} + \ldots + \boldsymbol{T}_{\cdot,k}\boldsymbol{S}_{k,\cdot} = \sum_{i=1}^{k}\boldsymbol{T}_{\cdot,i}\boldsymbol{S}_{i,\cdot} = \boldsymbol{T}\boldsymbol{S},$$

where $\boldsymbol{T}$, $\boldsymbol{S}$ are $(t \times k)$ and $(k \times n)$ matrices respectively whereas the columns $\boldsymbol{T}_{\cdot,i}$ representing time courses and the rows $\boldsymbol{S}_{i,\cdot}$ represents spatial maps.

- Spatial maps are Sparse
- Artifacts related to the acquisition process and machine-induced artifacts can be modeled in the same way, however they are Dense.



fMRI Unmixing

General Task

- Factor the data matrix as $\boldsymbol{Y} \approx \boldsymbol{A}\boldsymbol{B}$ aiming at getting $\boldsymbol{A} \approx \boldsymbol{T}$ and $\boldsymbol{B} \approx \boldsymbol{S}$
- There is an infinite number of such factorizations
- A priori information information about the characteristics of $\boldsymbol{B}$ and/or $\boldsymbol{A}$ need to get imposed

fMRI Unmixing

General Task

- Factor the data matrix as $\boldsymbol{Y} \approx \boldsymbol{A}\boldsymbol{B}$ aiming at getting $\boldsymbol{A} \approx \boldsymbol{T}$ and $\boldsymbol{B} \approx \boldsymbol{S}$
- There is an infinite number of such factorizations
- A priori information information about the characteristics of $\boldsymbol{B}$ and/or $\boldsymbol{A}$ need to get imposed

fMRI unmixing approaches

- **Conventional fMRI unmixing via ICA:** Factor $\boldsymbol{Y} \approx \boldsymbol{A}\boldsymbol{B}$ under the constraint that the columns of $\boldsymbol{B}$ are realizations of a random variable having independent components, e.g. [Calhoun 2006]
- **Recent advances:** Factor $\boldsymbol{Y} \approx \boldsymbol{A}\boldsymbol{B}$ under the constraint that $\boldsymbol{B}$ exhibits certain structured sparsity characteristics, e.g. [Varoquaux 2011], [Lee 2011], [Abraham 2013], [Abolghasemi 2013].

fMRI Unmixing

General Task

- Factor the data matrix as $\boldsymbol{Y} \approx \boldsymbol{A}\boldsymbol{B}$ aiming at getting $\boldsymbol{A} \approx \boldsymbol{T}$ and $\boldsymbol{B} \approx \boldsymbol{S}$
- There is an infinite number of such factorizations
- A priori information information about the characteristics of $\boldsymbol{B}$ and/or $\boldsymbol{A}$ need to get imposed

fMRI unmixing approaches

- **fMRI unmixing via Dictionary Learning:** Factor $\boldsymbol{Y} \approx \boldsymbol{A}\boldsymbol{B}$ under the constraint that the columns of $\boldsymbol{B}$ are **sparse**.

$$\min_{\boldsymbol{A},\boldsymbol{B}} \|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{B}\|_{\mathsf{F}}^2, \text{ s.t. } \|\boldsymbol{B}_{\cdot,j}\|_0 \leq K, \; j = 1 \cdots n, \quad (1)$$

- K-SVD: One of the best performing DL method.
- **Fast and Incoherent Dictionary Learning** (FIDL), [Abolghasemi 2013]:
  1. Incoherence is imposed in $\boldsymbol{A}$,
  2. relatively low complexity per iteration,
  3. improved performance to both synthetic and real fMRI data.

Dictionary Learning

K-SVD steps

Matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are estimated via iteratively repeating two processing stages:

First Stage Keep $\boldsymbol{A}$ fixed and update $\boldsymbol{B}$, column by column, via sparse codings of $\boldsymbol{Y}_{\cdot,j}$, $\forall j$, i.e.,

$$\min_{\boldsymbol{B}_{\cdot,j}} \|\boldsymbol{Y}_{\cdot,j} - \boldsymbol{A}\boldsymbol{B}_{\cdot,j}\|_F^2 \text{ subject to } \|\boldsymbol{B}_{\cdot,j}\|_0 \leq K, \ \forall j$$

Approximate solution: OMP

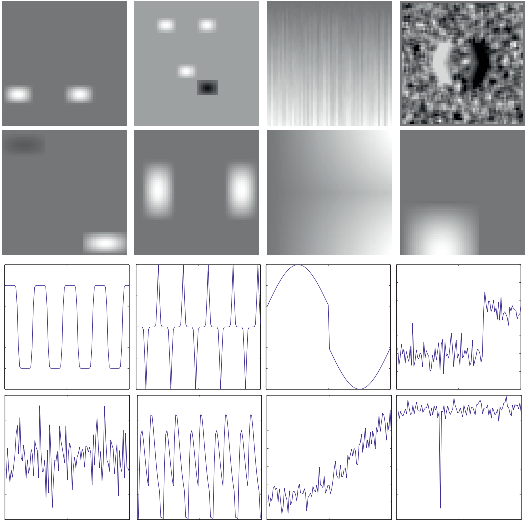
Second Stage Keep the support of $\boldsymbol{B}$ fixed and update sequentially the column-row pairs $\boldsymbol{A}_{\cdot,j}$ and $\boldsymbol{B}_{j,\cdot}$ to reduce $\|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{B}\|_F^2$

$$\min_{\boldsymbol{a},\boldsymbol{b}} \|\boldsymbol{E} + \boldsymbol{A}_{\cdot,j}\boldsymbol{B}_{j,\cdot} - \boldsymbol{a}\boldsymbol{b}^T\|_F^2 \text{ s.t. } \operatorname{supp}(\boldsymbol{B}_{j,\cdot})$$

Closed form solution: truncated SVD.

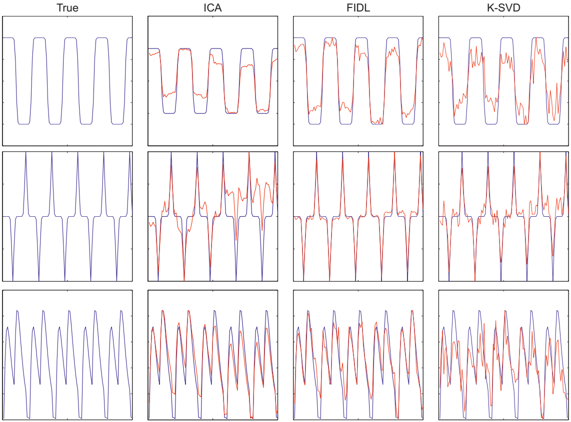
Synthetic fMRI data

[N. Correa, T. Adali, Yi-Ou Li, and V.D. Calhoun - 2005]



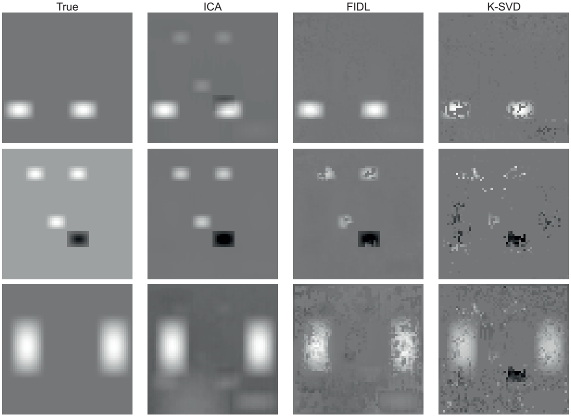
Some results

	Ca	Cm	Cam
fastICA	0.776	0.849	0.812
FIDL	0.836	0.792	0.814
K-SVD	0.798	0.697	0.747



Some results

	Ca	Cm	Cam
fastICA	0.776	0.849	0.812
FIDL	0.836	0.792	0.814
K-SVD	0.798	0.697	0.747



fMRI suited Dictionary Learning

Tackle several issues

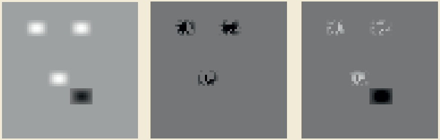
- Spatial maps are possibly split
  - identify split maps
  - merge them back
- time courses can get interrupted.
- The dense characteristics of the artifacts need to be incorporated
- Care about computational complexity

fMRI suited Dictionary Learning

Tackle several issues

- Spatial maps are possibly split
  - identify split maps
  - merge them back
- time courses can get interrupted.
- The dense characteristics of the artifacts need to be incorporated
- Care about computational complexity

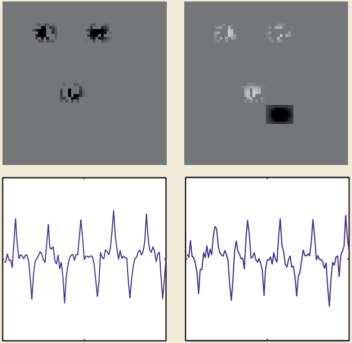
Example of split spatial maps



fMRI suited Dictionary Learning

Detect split atoms and merge them back

- Split maps corresponding to the same FBN are functionally associated, therefor should exhibit similar activation time-courses



fMRI suited Dictionary Learning

Detect split atoms and merge them back

- Split maps corresponding to the same FBN are functionally associated, therefor should exhibit similar activation time-courses

fMRI suited Dictionary Learning

Detect split atoms and merge them back

- Split maps corresponding to the same FBN are functionally associated, therefor should exhibit similar activation time-courses

- **Spatial Maps Merging:**

1. Construct, $\boldsymbol{m}$, the vector with indices, say ρ of a set of highly correlated time courses.
2. Set, $\boldsymbol{\beta}$ the union of the supports of the spatial maps indexed in $\boldsymbol{m}$.
3. Solve for $\boldsymbol{a}$ and $\boldsymbol{b}$ the optimization problem

$$\min_{\boldsymbol{a},\boldsymbol{b}} \|(\boldsymbol{A}_{\cdot,m_1}\boldsymbol{B}_{m_1,\boldsymbol{\beta}} + \cdots + \boldsymbol{A}_{\cdot,m_\rho}\boldsymbol{B}_{m_\rho,\boldsymbol{\beta}}) - \boldsymbol{a}\boldsymbol{b}^T\|_{\mathsf{F}}^2.$$

   Rank-1 best approximation (via truncated SVD).

4. Set $\boldsymbol{A}_{\cdot,m_1} = \boldsymbol{a}$ and $\boldsymbol{B}_{m_1,\boldsymbol{\beta}} = \boldsymbol{b}$
5. Re-initialize the rest of the atoms indexed in $\boldsymbol{m}$.

fMRI suited Dictionary Learning

Detect split atoms and merge them back

- Split maps corresponding to the same FBN are functionally associated, therefor should exhibit similar activation time-courses

- **Spatial Maps Merging:**

  1. Construct, $\boldsymbol{m}$, the vector with indices, say ρ of a set of highly correlated time courses.
  2. Set, $\boldsymbol{\beta}$ the union of the supports of the spatial maps indexed in $\boldsymbol{m}$.
  3. Solve for $\boldsymbol{a}$ and $\boldsymbol{b}$ the optimization problem

$$\min_{\boldsymbol{a},\boldsymbol{b}} \|(\boldsymbol{A}_{\cdot,m_1}\boldsymbol{B}_{m_1,\boldsymbol{\beta}} + \cdots + \boldsymbol{A}_{\cdot,m_\rho}\boldsymbol{B}_{m_\rho,\boldsymbol{\beta}}) - \boldsymbol{a}\boldsymbol{b}^T\|_{\mathsf{F}}^2.$$

     Rank-1 best approximation (via truncated SVD).

  4. Set $\boldsymbol{A}_{\cdot,m_1} = \boldsymbol{a}$ and $\boldsymbol{B}_{m_1,\boldsymbol{\beta}} = \boldsymbol{b}$
  5. Re-initialize the rest of the atoms indexed in $\boldsymbol{m}$.

- **Time Courses Merging:** It can be done in exactly the same way.

fMRI suited Dictionary Learning

Detect split atoms and merge them back

- Split maps corresponding to the same FBN are functionally associated, therefor should exhibit similar activation time-courses
- **Spatial Maps Merging:**
  1. Construct, $\boldsymbol{m}$, the vector with indices, say ρ of a set of highly correlated time courses.
  2. Set, $\boldsymbol{\beta}$ the union of the supports of the spatial maps indexed in $\boldsymbol{m}$.
  3. Solve for $\boldsymbol{a}$ and $\boldsymbol{b}$ the optimization problem

$$\min_{\boldsymbol{a},\boldsymbol{b}} \|(\boldsymbol{A}_{\cdot,m_1}\boldsymbol{B}_{m_1,\boldsymbol{\beta}} + \cdots + \boldsymbol{A}_{\cdot,m_\rho}\boldsymbol{B}_{m_\rho,\boldsymbol{\beta}}) - \boldsymbol{a}\boldsymbol{b}^T\|_{\mathsf{F}}^2.$$

     Rank-1 best approximation (via truncated SVD).
  4. Set $\boldsymbol{A}_{\cdot,m_1} = \boldsymbol{a}$ and $\boldsymbol{B}_{m_1,\boldsymbol{\beta}} = \boldsymbol{b}$
  5. Re-initialize the rest of the atoms indexed in $\boldsymbol{m}$.
- **Time Courses Merging:** It can be done in exactly the same way.
- Merging tasks need not to be performed in each K-SVD iteration.

fMRI suited Dictionary Learning

Coping with machine-induced artifacts

- These artifacts exhibit low localization rendered, in this way, dense.
- Force a number of rows of $\boldsymbol{B}$ to be dense: $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{B} + \bar{\boldsymbol{A}}\bar{\boldsymbol{B}}$
- Sparse coding with Partly known support.

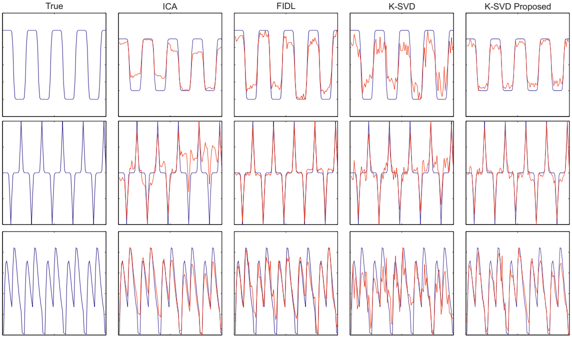
Computational Complexity issues

- Replace OMP with Batch OMP [Rubinstein 2008]:
  - Batch OMP Pre-computes $\boldsymbol{A}^T\boldsymbol{A}$ avoiding the explicit computation of the error residual in each iteration.
  - A large number of columns of $\boldsymbol{Y}$ (voxels) need to be sparse coded over the same dictionary.
  - t is larger than the number of dictionary atoms.
- Replace exact rank-1 solutions of $\min_{\boldsymbol{a},\boldsymbol{b}} \|\boldsymbol{\Gamma} - \boldsymbol{a}\boldsymbol{b}^T\|_{\mathsf{F}}^2$, with approximates:
  - e.g., with alternating optimization over $\boldsymbol{a}$ and $\boldsymbol{b}$ according to:

$$\boldsymbol{a} = \frac{\boldsymbol{\Gamma}\boldsymbol{b}}{\|\boldsymbol{\Gamma}\boldsymbol{b}\|_2}, \qquad \boldsymbol{b} = \boldsymbol{\Gamma}^T\boldsymbol{a}$$

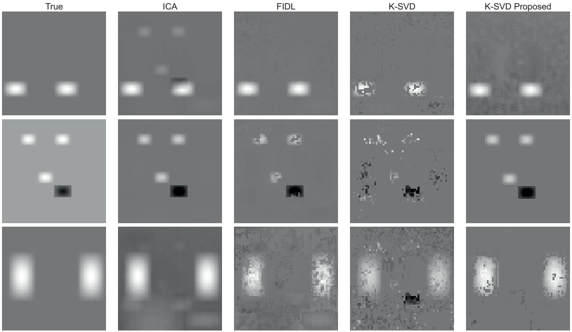
Some results

	Ca	Cm	Cam
fastICA	0.776	0.849	0.812
FIDL	0.836	0.792	0.814
K-SVD	0.798	0.697	0.747
Proposed	**0.862**	**0.861**	**0.861**



Some results

	Ca	Cm	Cam
fastICA	0.776	0.849	0.812
FIDL	0.836	0.792	0.814
K-SVD	0.798	0.697	0.747
Proposed	**0.862**	**0.861**	**0.861**



Thank you

Bibliography

- Martin A. Lindquist, “The statistical analysis of fMRI data,” *Statistical Science*, Nov. 2008.
- V. Perlbarg and G. Marrelec, “Contribution of exploratory methods to the investigation of extended large-scale brain networks in functional MRI: methodologies, results, and challenges,” *International Journal of Biomedical Imaging*, vol. 2008.
- Vince D. Calhoun and Tülay Adali, “Unmixing fMRI with independent component analysis,” *Engineering in Medicine and Biology Magazine, IEEE*, vol. 25, no. 2, pp. 79–90, 2006.
- Kangjoo Lee, Sungho Tak, and Jong Chul Ye, “A data-driven sparse GLM for fMRI analysis using sparse dictionary learning with MDL criterion,” *IEEE Transactions on Medical Imaging*, vol. 30, no. 5, pp. 1076–1089, May 2011.
- Vahid Abolghasemi, Saideh Ferdowsi, and Saeid Sanei, “Fast and incoherent dictionary learning algorithms with application to fMRI,” *Signal, Image and Video Processing*, Feb. 2013.
- Gaël Varoquaux, Alexandre Gramfort, Fabian Pedregosa, Vincent Michel, and Bertrand Thirion, “Multi-subject dictionary learning to segment an atlas of brain spontaneous activity,” in *Information Processing in Medical Imaging*, vol. 6801. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Alexandre Abraham, Elvis Dohmatob, Bertrand Thirion, Dimitris Samaras, and Gael Varoquaux, “Extracting brain regions from rest fMRI with total-variation constrained dictionary learning,” in *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2013*, pp. 607–615. Springer, 2013.
- Jeonghyeon Lee, Yong Jeong, and Jong Chul Ye, “Group sparse dictionary learning and inference for resting-state fMRI analysis of Alzheimer'S disease,” in *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*, 2013, pp. 540–543.
- M. Aharon, M. Elad, and A. Bruckstein, “K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation,” *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
- Rafael E. Carrillo, Luisa F. Polania, and Kenneth E. Barner, “Iterative algorithms for compressed sensing with partially known support,” in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. 2010, p. 3654–3657, IEEE.
- Ron Rubinstein, Michael Zibulevsky, and Michael Elad, “Efficient implementation of the k-SVD algorithm using batch orthogonal matching pursuit,” *CS Technion*, 2008.
- N. Correa, T. Adali, Yi-Ou Li, and V.D. Calhoun, “Comparison of blind source separation algorithms for fmri using a new matlab toolbox: Gift,” in *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*, March 2005, vol. 5, pp. v/401–v/404 Vol. 5.