False positives detection in connectomics through hierarchical sparsity

...and résumé

by

Matteo Frigo

INRIA Sophia Antipolis - 19th May 2017

Curriculum Vitae

University of Verona (Italy)

2015 B.Sc. Applied Mathematics

2017 M.Sc. Mathematics

- Functional Analysis, Differential Geometry, Stochastic Analysis, ...
- Numerical Analysis, Polynomial approximation, Optimisation, ODE/PDE/SDE/BVP, ...
- Machine Learning, Digital Geometry Processing, Research and Modelling

Complementary activities

Internships:

- Image classification through time series analysis
- 3D Super Resolution of clinical MRI images

Others:

- Co-founder of the OctaveArena group
  (dev of ODE solvers for GNU Octave)
- Student representative

Bachelor thesis

Maximum Likelihood versus Gibbs Sampling for regime switching analysis of financial time series

$$
\begin{aligned}
y_t - \mu_{S_t} &= \varphi\left(y_{t-1} - \mu_{S_{t-1}}\right) + \varepsilon_t, \quad t = 1, 2, \ldots, T \\
\varepsilon_t &= \text{i.i.d. } \mathcal{N}(0, \sigma^2_{S_t}), \\
\mu_{S_t} &= \mu_1 S_{1,t} + \mu_2 S_{2,t} + \mu_3 S_{3,t}, \\
\sigma_{S_t} &= \sigma_1 S_{1,t} + \sigma_2 S_{2,t} + \sigma_3 S_{3,t}, \\
S_t &\in \{1, 2, 3\} \\
p_{ij} &:= \mathbb{P}\left(S_t = j | S_{t-1} = i\right) \quad i, j = 1, 2, 3.
\end{aligned}
$$

- Di Persio and Frigo, 2015, WSEAS Transactions on Business and Economics
- Di Persio and Frigo, 2016, Journal of Computational and Applied Mathematics

Programming skills

Languages:

- Matlab/Octave proficient, Scilab basic knowledge
- Python+Numpy good knowledge, Cython basic knowledge
- Basic knowledge of Java, C and C++

Tools:

- Diffusion MRI analysis tool
- Pattern Recognition tools
- Freefem++

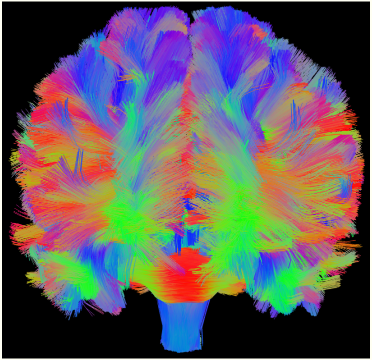
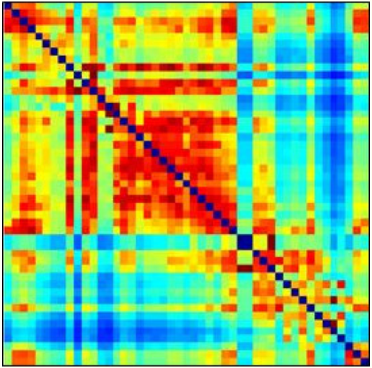
Master thesis

Joint work with...

- Prof. Giandomenico Orlandi
- Prof. Jean-Philippe Thiran
- Prof. Alessandro Daducci
- Muhamed Barakovic

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

The goal: structural connectivity

 $\Longrightarrow$ 

Quantify the connections within the brain

The problem: false positives

Tractography-based connectomes are dominated by false-positive connections

Maier-Hein et al., 2016 (bioarxiv)

- 96 distinct tractography pipelines

The problem: false positives

Tractography-based connectomes are dominated by false-positive connections

Maier-Hein et al., 2016 (bioarxiv)

- 96 distinct tractography pipelines
- *"most algorithms routinely extracted many false positive bundles"*

The problem: false positives

Tractography-based connectomes are dominated by false-positive connections

Maier-Hein et al., 2016 (bioarxiv)

- 96 distinct tractography pipelines
- *"most algorithms routinely extracted many false positive bundles"*
- *"Tractography identifies more invalid than valid bundles"*

The problem: false positives

Tractography-based connectomes are dominated by false-positive connections

Maier-Hein et al., 2016 (bioarxiv)

- 96 distinct tractography pipelines
- *"most algorithms routinely extracted many false positive bundles"*
- *"Tractography identifies more invalid than valid bundles"*
- **"Tractography is fundamentally ill-posed"**

The problem: false positives

Connectome sensitivity or specificity: which is more important?

- *"the impact of FPs and FNs on empirical connectomes indicate that specificity is at least twice as important as sensitivity"*

Zalesky, A. et al., 2016 (NeuroImage)

Anatomical accuracy of brain connections derived from diffusion MRI tractography is inherently limited

- *"The methods that show the highest sensitivity show the lowest specificity, and vice versa"*

Thomas, C. et al., 2014 (PNAS)

Forward model

$$y = Ax + \varepsilon$$

Given

- x: vector of weights
- A: dictionary for the fibres
- ε: systematic+random error

Obtain

- y: acquired dMRI data

Forward model: COMMIT

Convex Optimisation Modelling
for Microstructure Informed Tractography

Daducci et al., 2015 (IEEE TMI)

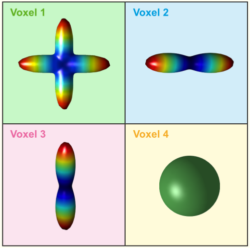
Forward model: COMMIT

Convex Optimisation Modelling
for Microstructure Informed Tractography

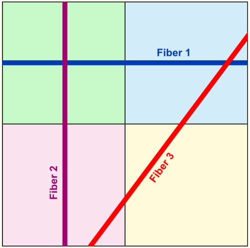
Daducci et al., 2015 (IEEE TMI)

Input:

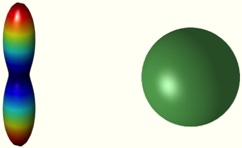
Acquired DWI



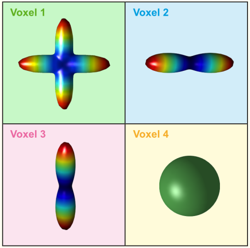
Tractography

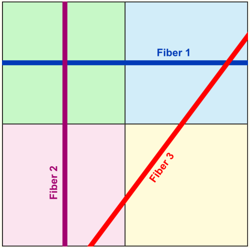


Microstructure model



Forward model: COMMIT





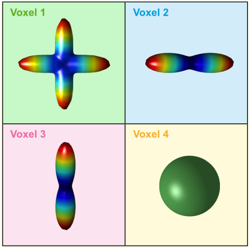
$\mathbf{y} =$

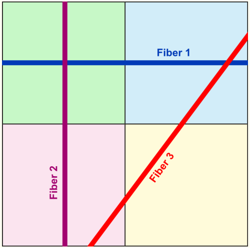
$\mathbf{A} =$

$\mathbf{x} =$

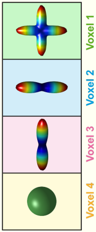
?	?	?	?	?	?	?

Forward model: COMMIT





$\mathbf{y} =$

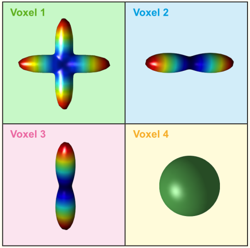


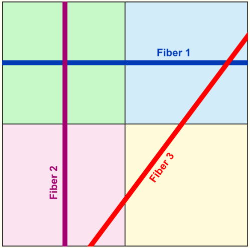
$\mathbf{A} =$

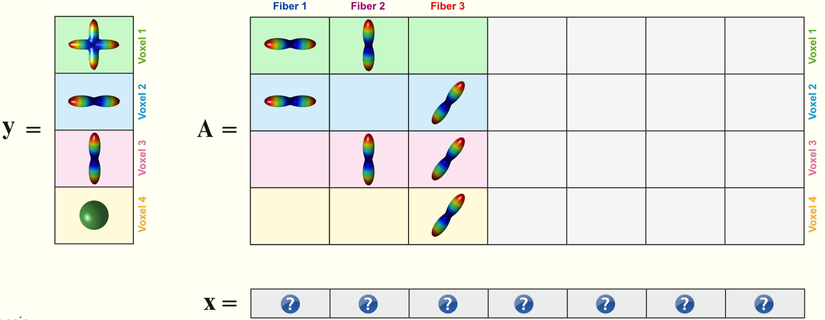
$\mathbf{x} =$

?	?	?	?	?	?	?

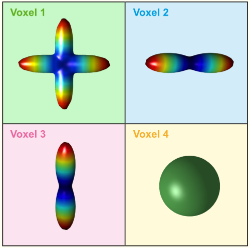
Forward model: COMMIT

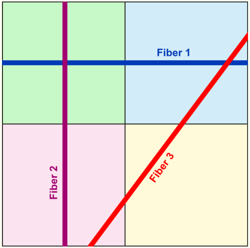




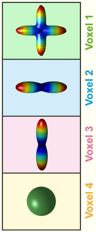


Forward model: COMMIT

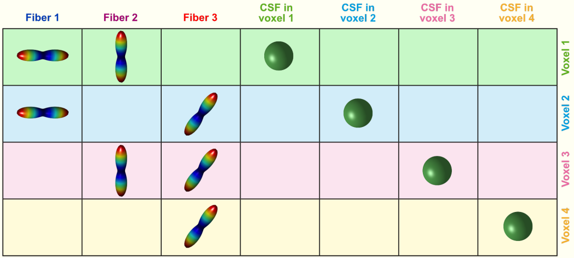




$\mathbf{y} =$

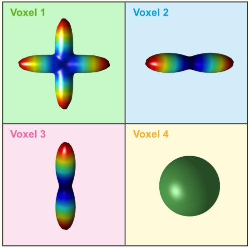


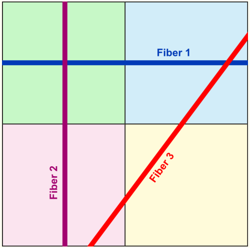
$\mathbf{A} =$



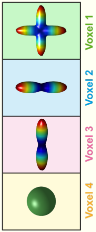
$\mathbf{x} =$

Forward model: COMMIT

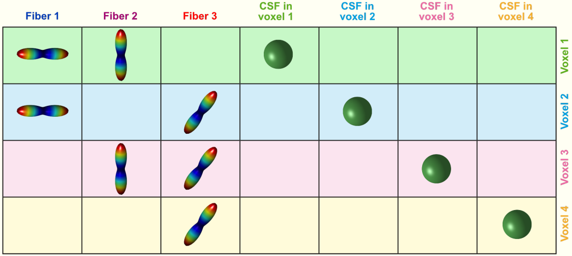




$\mathbf{y} =$



$\mathbf{A} =$



$\mathbf{x} =$

Inverse problem

$$y = Ax + \varepsilon$$

Recover x from the acquired data y

$$x^* = \underset{x \in \mathbb{R}_+^c}{\operatorname{argmin}} \frac{1}{2} \|Ax - y\|_2^2$$

Inverse problem

$$y = Ax + \varepsilon$$

Recover x from the acquired data y

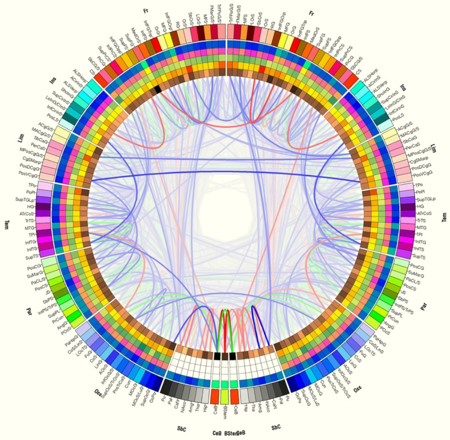
$$x^* = \operatorname*{argmin}_{x \in \mathbb{R}_+^c} \frac{1}{2} \|Ax - y\|_2^2 + \lambda \Omega(x)$$

$\Omega : \mathbb{R}^c \to \mathbb{R}$ anatomy-based penalty

$\lambda \geq 0$ regularisation parameter

Anatomical prior knowledge

Brain hierarchy

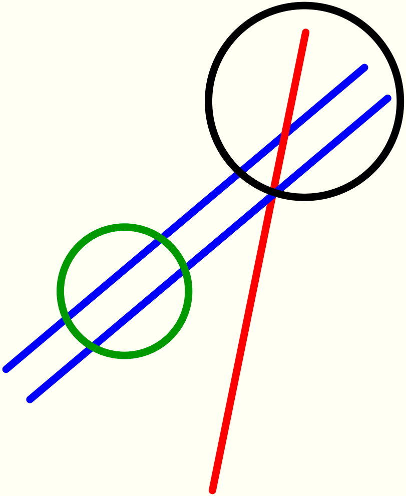


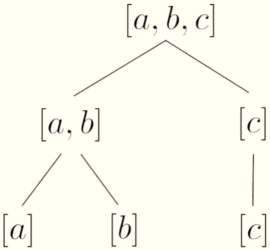
© John Darrell Van Horn

Connections within the brain are endowed with a hidden hierarchical pattern which should be exploited.

- Zhou et al., 2006
- Duarte-Carvajalino et al., 2012
- Moreno-Dominguez et al., 2012

Abstract model





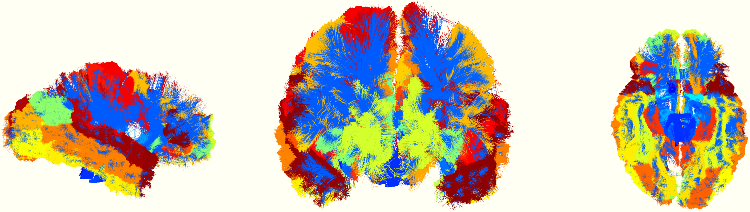
$$\mathcal{G} = \Big(\big[[a], [b], [a, b], [c], [a, b, c] \big], \preceq \Big)$$

E.g.:

$$[a] \preceq [a, b] \preceq [c]$$
$$[a, b, c] \npreceq [c]$$

Clustering

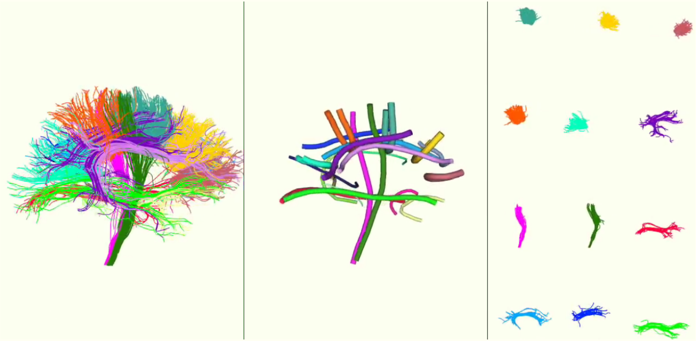
First level: Tract Querier



Wassermann et al., *The white matter query language: a novel approach for describing human white matter anatomy*, 2016

Clustering

Deeper levels: QuickBundlesX



Garyfallidis et al., "*QuickBundlesX: Sequential clustering of millions of streamlines in multiple levels of detail at record execution time.*, 2016"

Nomenclature

- Valid/Invalid Bundle

  VB : fascicle belonging to the ground truth

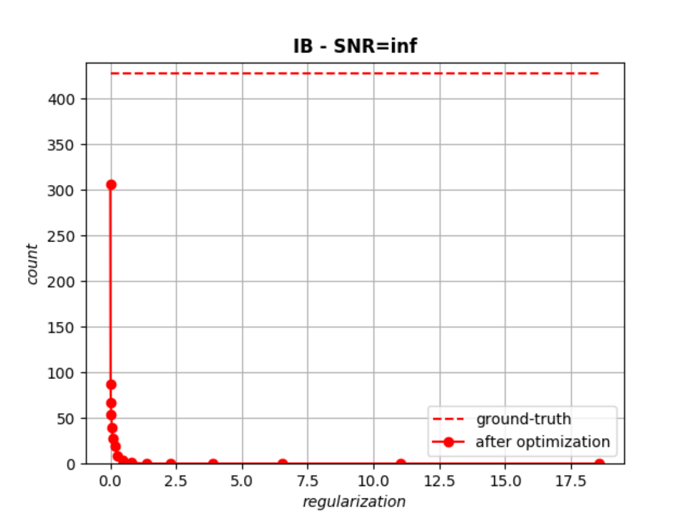
  IB : fascicle made of false positives

- Valid/Invalid Connection

  VC : fibre belonging to a VB

  IC : fibre belonging to an IB

Toy example



IB - SNR=inf
400
350
300
250
200
150
100
50
0
count
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
regularization
ground-truth
after optimization

Formalisation

The problem

Find $x^* \in \mathbb{R}^c$ such that

$$x^* = \operatorname*{argmin}_{x \in \mathbb{R}^c_+} \frac{1}{2} \|Ax - y\|_2^2 + \lambda \Omega(x)$$

The problem

Find $x^* \in \mathbb{R}^c$ such that

$$x^* = \underset{x \in \mathbb{R}^c_+}{\operatorname{argmin}} \frac{1}{2} \|Ax - y\|_2^2 + \lambda \Omega(x)$$

RLS: Regularised Least Squares

Penalty term Ω

Objective:

Cancel invalid bundles

$$\Downarrow$$

Sparsity in the space of fibres

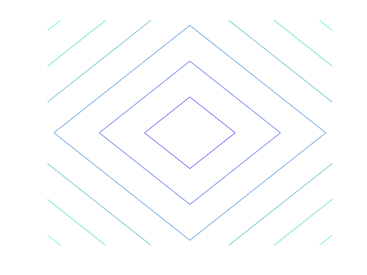
Penalty term Ω

Objective:

Cancel invalid bundles

$\Downarrow$

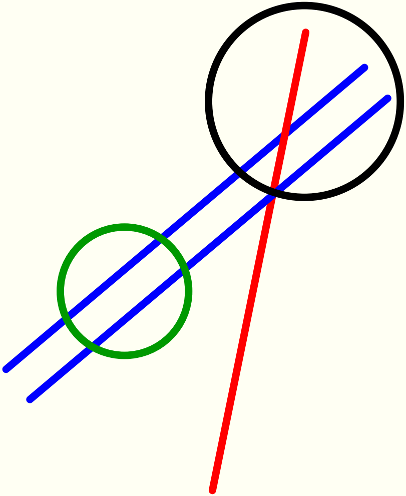
Sparsity in the space of fibres



Classical sparsity:

$\Omega(x) = \|x\|_1$

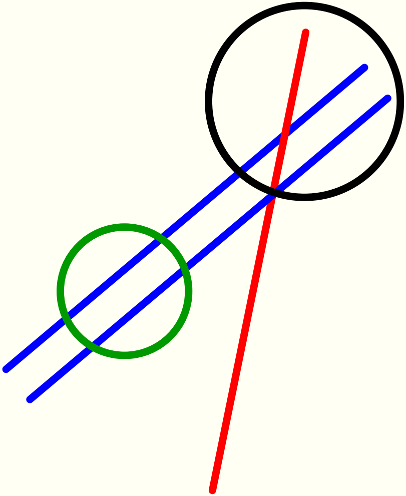
Hierarchical sparsity



$$\Omega(x) = \|X_{\mathcal{G}}\|_1 = \sum_{g \in \mathcal{G}} \left\|x_{|g}\right\|_2$$

$$g = [a, b], \quad x = [x_a, x_b, x_c]$$
$$\Rightarrow x_{|g} = [x_a, x_b, 0]$$

Hierarchical sparsity



$$\Omega(x) = \|X_{\mathcal{G}}\|_1 = \sum_{g \in \mathcal{G}} \left\|x_{|g}\right\|_2$$

$$g = [a, b]\,, \quad x = [x_a, x_b, x_c]$$
$$\Rightarrow x_{|g} = [x_a, x_b, 0]$$

$$\Omega(x) = \|[x_a, 0, 0]\|_2 + \|[0, x_b, 0]\|_2 + \|[x_a, x_b, 0]\|_2 + \|[0, 0, x_c]\|_2 + \|[x_a, x_b, x_c]\|_2$$

The problem

Find $x^* \in \mathbb{R}^c$ such that

$$x^* = \underset{x \in \mathbb{R}^c}{\operatorname{argmin}} \underbrace{\frac{1}{2} \|Ax - y\|_2^2}_{\text{smooth}} + \underbrace{\lambda \Omega(x) + \iota_{\geq 0}(x)}_{\text{non smooth}}$$

HNNLS: Hierarchical Non Negative Least Squares

Numerical optimisation

Proximal operator

Proximal operator of f:

$$\mathbf{prox}_f(x) = \underset{y \in \mathbb{R}^c}{\operatorname{argmin}} \frac{1}{2} \|x - y\|_2^2 + f(y)$$

Proximal operator

Proximal operator of f:

$$\mathbf{prox}_f(x) = \underset{y \in \mathbb{R}^c}{\operatorname{argmin}} \frac{1}{2} \|x - y\|_2^2 + f(y)$$

If $f(x) = \iota_S(x)$ with S convex set:

$$\mathbf{prox}_f(x) = \Pi_S(x)$$

Proximal operator

Proximal operator of f:

$$\mathbf{prox}_f(x) = \underset{y \in \mathbb{R}^c}{\operatorname{argmin}} \frac{1}{2} \|x - y\|_2^2 + f(y)$$

If $f(x) = \iota_S(x)$ with S convex set:

$$\mathbf{prox}_f(x) = \Pi_S(x)$$

$$\mathbf{prox}_{\iota_{\geq 0}}(x) = \Pi_{\geq 0}(x)$$

Proximal of Ω

Let $\mathcal{G}$ tree structure with order $\preceq$

$$\Omega(x) = \sum_{g \in \mathcal{G}} \omega_g \left\| x_{|g} \right\|$$

Compute $\mathbf{prox}_{\Omega}(x)$

1. Set $v = x$.
2. For $g \in \mathcal{G}$ following $\preceq$ do

$$v_{|g} \longleftarrow v_{|g} - \Pi_{\|\cdot\|_* \leq \omega_g} \left(v_{|g} \right).$$

Jenatton et al., "Proximal methods for sparse hierarchical dictionary learning", 2010, ICML-10

Davis-Yin splitting scheme(2015)

Objective:

$$x^* = \underset{y \in \mathbb{R}^c}{\operatorname{argmin}} \overbrace{f(y)}^{\text{s}} + \overbrace{g(y) + h(y)}^{\text{ns}}$$

Algorithm:

$$\begin{aligned} y &= \mathbf{prox}_{\gamma g}(x) \\ z &= \mathbf{prox}_{\gamma h}(2x - y - \gamma \nabla f(y)) \\ x^+ &= x - y + z \end{aligned}$$

Davis-Yin splitting scheme(2015)

Objective:

$$x^* = \underset{y \in \mathbb{R}^c}{\operatorname{argmin}} \overbrace{f(y)}^{\text{s}} + \overbrace{g(y) + h(y)}^{\text{ns}}$$

Algorithm:

$$\begin{aligned} y &= \mathbf{prox}_{\gamma g}(x) \\ z &= \mathbf{prox}_{\gamma h}(2x - y - \gamma \nabla f(y)) \\ x^+ &= x - y + z \end{aligned}$$

Rate of convergence:

$$\mathcal{O}(1/k)$$

Non negativity

Definition (Absolute norm)

A norm $\Omega : X \to \mathbb{R}$ is called *absolute* if $\forall u, v \in \mathbb{R}^N$ such that $|u_j| \leq |v_j|$ for all j implies $\Omega(u) \leq \Omega(v)$.

Theorem (Proximal operator of absolute norms)

Let $w \in \mathbb{R}^n$ and $\lambda > 0$. Consider an absolute norm Ω. We have

$$\operatorname*{argmin}_{z \in \mathbb{R}^n_+} \left[\frac{1}{2} \|w - z\|_2^2 + \lambda \Omega(z) \right] = \operatorname*{argmin}_{z \in \mathbb{R}^n} \left[\frac{1}{2} \|[w]_+ - z\|_2^2 + \lambda \Omega(z) \right].$$

Non negativity

Definition (Absolute norm)

A norm $\Omega : X \to \mathbb{R}$ is called *absolute* if $\forall u, v \in \mathbb{R}^N$ such that $|u_j| \leq |v_j|$ for all j implies $\Omega(u) \leq \Omega(v)$.

Theorem (Proximal operator of absolute norms)

Let $w \in \mathbb{R}^n$ and $\lambda > 0$. Consider an absolute norm Ω. We have

$$\operatorname*{argmin}_{z \in \mathbb{R}^n_+} \left[\frac{1}{2} \|w - z\|_2^2 + \lambda \Omega(z) \right] = \operatorname*{argmin}_{z \in \mathbb{R}^n} \left[\frac{1}{2} \|[w]_+ - z\|_2^2 + \lambda \Omega(z) \right].$$

$$\begin{aligned} \mathbf{prox}_{\lambda\Omega + \iota_{\geq 0}}(w) &= \mathbf{prox}_{\lambda\Omega}([w]_+) \\ &= \mathbf{prox}_{\lambda\Omega}\left(\mathbf{prox}_{\iota_{\geq 0}}(w)\right) \end{aligned}$$

Jenatton et al., "Proximal methods for hierarchical sparse coding", 2011, JMLR

FISTA

Fast Iterative Shrinkage Thresholding Algorithm (2009)

Objective:

$$x^* = \operatorname*{argmin}_{x \in \mathbb{R}^c} \overbrace{f(x)}^{\text{s}} + \overbrace{g(x)}^{\text{ns}}$$

Algorithm ($t_0 = 1$):

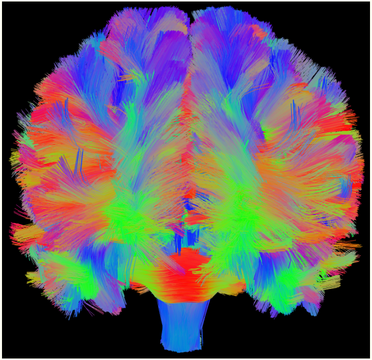
$$\begin{aligned}
x_k &= \mathbf{prox}_{t_k g}\left(x_{k-1} - t_k \nabla f(x_{k-1})\right) \\
t_{k+1} &= \frac{1 + \sqrt{1 + 4t_k^2}}{2} \\
y_{k+1} &= x_k + \left(\frac{t_k - 1}{t_{k+1}}\right)(x_k - x_{k-1})
\end{aligned}$$

Rate of convergence:

$$\mathcal{O}\left(1/k^2\right)$$

Results

Dataset



$$\|Ax - y\|_2^2$$

- A Tractography:
  Particle Filtering[1]
- y TDI map given by the fibres belonging to VBs

[1] Girard and Descoteaux. *Online filtering tractography: tracking with anatomical priors*, ISMRM 2013

Competitors

SIFT2

Smith, Robert E., et al. "*SIFT2: Enabling dense quantitative assessment of brain white matter connectivity using streamlines tractography.*" Neuroimage 119 (2015): 338-351.

LiFE

Pestilli, Franco, et al. "*Evaluation and statistical inference for human connectomes.*" Nature methods 11.10 (2014): 1058-1063.

Same goal of HNNLS

Competitors

SIFT2

Smith, Robert E., et al. "*SIFT2: Enabling dense quantitative assessment of brain white matter connectivity using streamlines tractography.*" Neuroimage 119 (2015): 338-351.

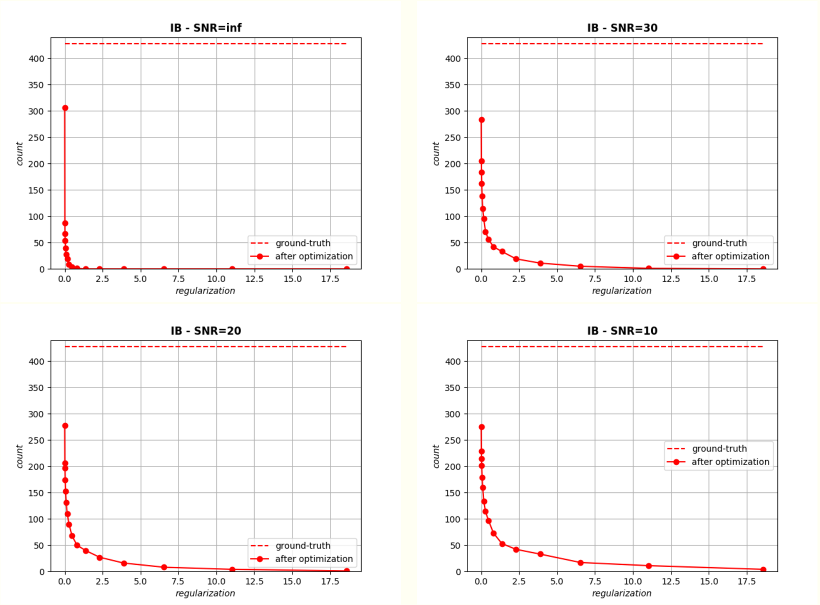
LiFE

Pestilli, Franco, et al. "*Evaluation and statistical inference for human connectomes.*" Nature methods 11.10 (2014): 1058-1063.

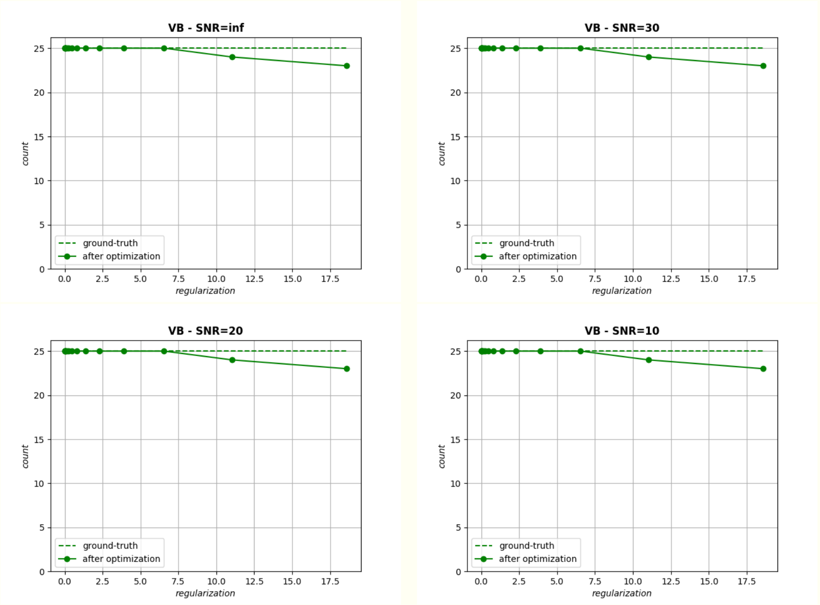
Same goal of HNNLS

No convexity, huge computational effort,
no microstructure information, ...

Results: Invalid Bundles (IB)



Results: Valid Bundles (VB)



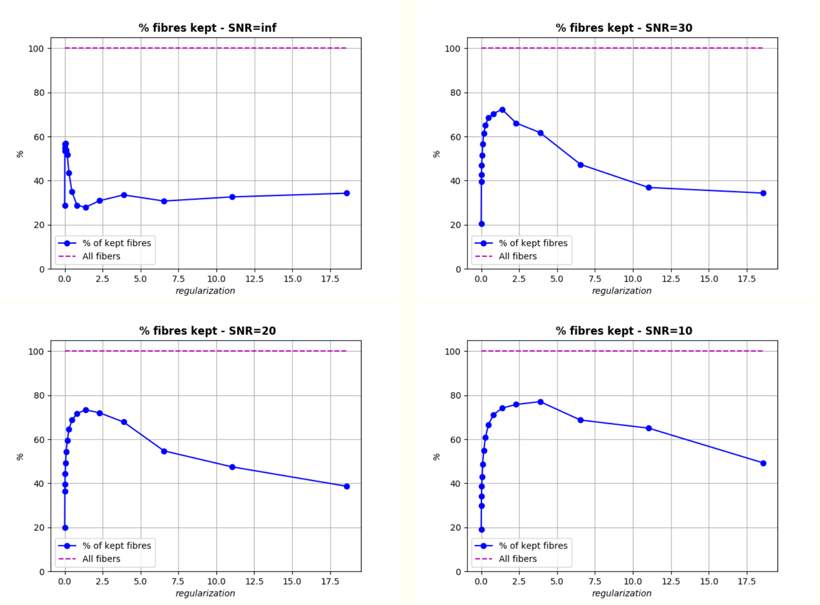
Results

$$\lambda \in \{\ 0 \quad 0.012 \quad 0.020 \quad 0.035 \quad 0.059 \quad 0.100 \quad 0.168 \quad 0.284 \quad 0.480 \quad 0.809 \quad 1.364 \quad 2.300 \quad 3.879 \quad 6.540 \quad 11.02 \quad 18.59\ \}$$

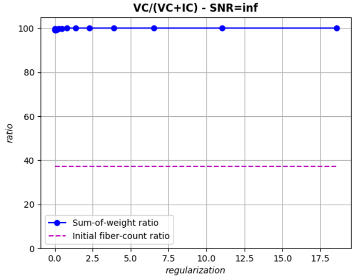
	SIFT2	LiFE	$\lambda = 0$	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
VB	25	25	25	25	25	25	25	25	25
IB	427	423	284	205	184	162	138	115	95

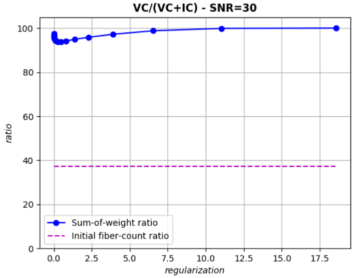
	λ_7	λ_8	λ_9	λ_{10}	λ_{11}	λ_{12}	λ_{13}	λ_{14}	λ_{15}
VB	25	25	25	25	25	25	25	24	23
IB	71	56	42	33	19	11	5	1	0

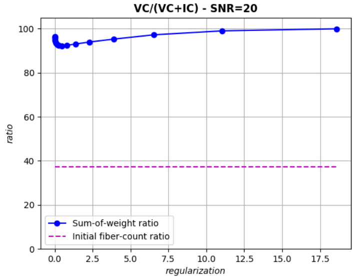
Results: Kept fibres

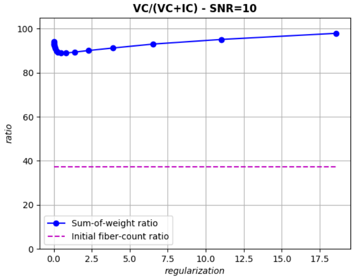


Results: Valid Connections







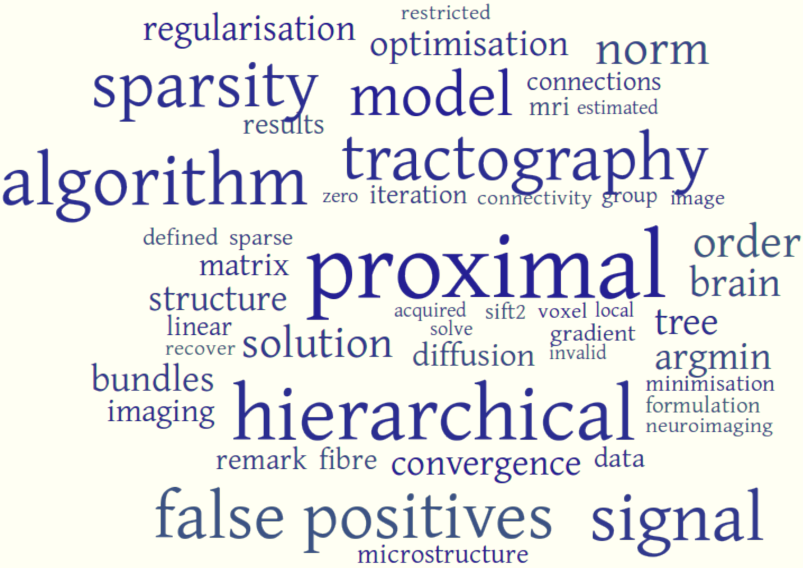


Future work

Still to be done...

- Sensitivity analysis w.r.t. tractogram
- Sensitivity analysis w.r.t. clustering thresholds
- Set up a strategy for selecting λ
- Exploit more sophisticated solvers (Canales, 2015)

Thank you


restricted
regularisation
optimisation
norm
sparsity
model
connections
mri
estimated
results
algorithm
tractography
zero
iteration
connectivity
group
image
defined
sparse
proximal
order
matrix
brain
structure
acquired
sift2
voxel
local
tree
linear
solve
gradient
recover
solution
diffusion
invalid
argmin
bundles
minimisation
hierarchical
formulation
imaging
neuroimaging
remark
fibre
convergence
data
false positives
signal
microstructure

The end

Reweighted ℓ_1

Objective:

Smart choice of w_g

Iteratively define

$$w_g = \frac{1}{\|x_g\|_1 + \varepsilon}$$

and restart the optimisation procedure.

Candes, Emmanuel J., Michael B. Wakin, and Stephen P. Boyd. "Enhancing sparsity by reweighted l1 minimization." Journal of Fourier analysis and applications, 2008

RUMBA

Robust and Unbiased Model-Based Spherical Deconvolution

$$\mathbb{P}\left(S_i|\bar{S}_i,\sigma^2,n\right)=\frac{\bar{S}_i}{\sigma^2}\left(\frac{S_i}{\bar{S}_i}\right)^n\exp\left\{-\frac{1}{2\sigma^2}\left[S_i^2+\bar{S}_i^2\right]\right\}I_{n-1}\left(\frac{S_i\bar{S}_i}{\sigma^2}\right)u(S_i),$$

$$f=f\circ\frac{H^t\left[S\circ\frac{I_n(S\circ Hf/\sigma^2)}{I_{n-1}(S\circ Hf/\sigma^2)}\right]}{H^tHf}$$

where $S_i=(Hf)_i$.

Canales-Rodríguez, Erick J., et al. *Spherical deconvolution of multichannel diffusion MRI data with non-Gaussian noise models and spatial regularization.* PloS one 10.10 (2015).

Image classification

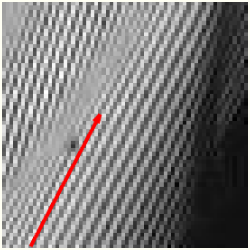
Supervisor: Prof. Manuele Bicego and Prof. Luca di Persio
Task: **classification of production of industrial butchery**

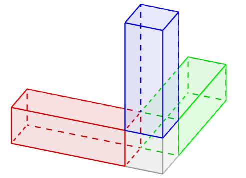
1. Edge detection
2. Parametrisation of the edge
3. Model identification (AR, ARMA, GARCH) on the parametrisation
4. Classification of the identified coefficients

Super Resolution

Supervisors: Prof. Roberto Foroni and Prof. Giandomenico Orlandi
Task: **In-plane and through-plane SR of MRI scans**

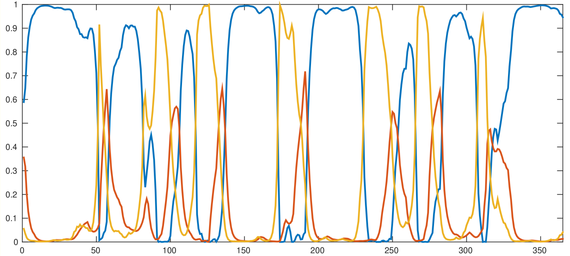
- In-plane
  1. Find the local scale (Wavelet)
  2. Detect the local orientation (TV)
  3. Convolution with tilted anisotropic Gaussian kernel
- Through-plane





Bachelor thesis

State probabilities $\mathbb{P}\left(S_t = j|\psi_t\right)$



Comparison with the benchmark