Exploiting sparsity in diffusion MRI

Alessandro Daducci
(work in collaboration with Dr. Yves Wiaux)

30/03/2012
University of Verona

Outline

- Diffusion and why it is important
- Diffusion MRI
- Compressed Sensing (CS) framework
- Application of CS-based techniques in diffusion MRI to reduce acquisition times

What is diffusion?

- **Random movement of molecules** from regions of high concentration to regions of lower concentration due to thermal agitation
  - **EXAMPLE**: in a glass of water, molecules diffuse randomly and freely, only constrained by the boundaries of the container

- First noted b
  - *"…random r*

What is diffusion?

- **Random movement of molecules** from regions of high concentration to regions of lower concentration due to thermal agitation
  - **EXAMPLE**: in a glass of water, molecules diffuse randomly and freely, only constrained by the boundaries of the container

- First noted by **Robert Brown** in 1828
  - *"…random motion without any apparent cause…"*

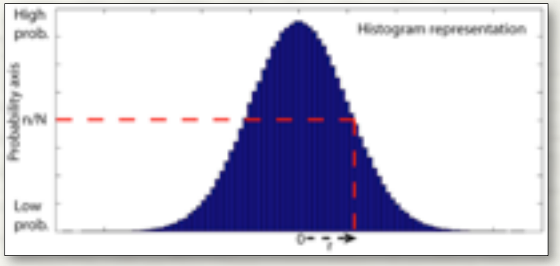
- Formally described by **Albert Einstein** in 1905

$$\langle r^2 \rangle \propto t \, \mathrm{D}$$

mean displacement of molecules ← $\langle r^2 \rangle$

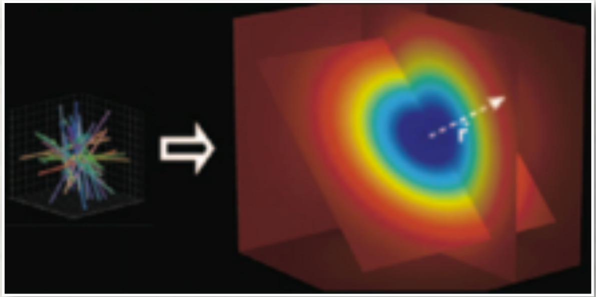
t → observation time

D → coefficient of diffusion of the medium



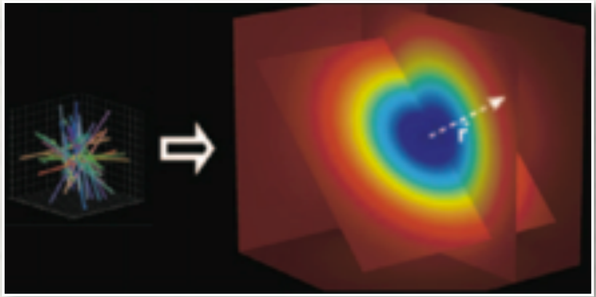
What happens in the brain?

- **Cerebrospinal Fluid (CSF)**
  - Displacements follow a **gaussian distribution** in every direction (**isotropic**)
  - Variance depends on the level of restriction of the fluid

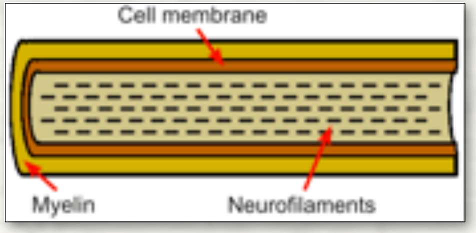


What happens in the brain?

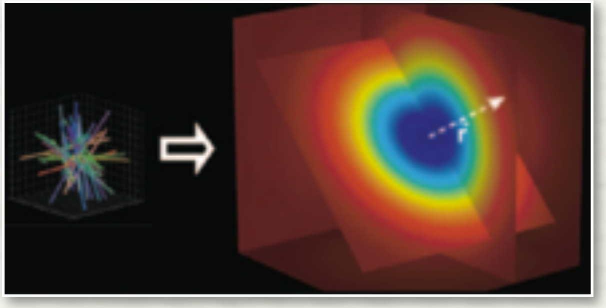
- **Cerebrospinal Fluid (CSF)**
  - Displacements follow a **gaussian distribution** in every direction (**isotropic**)
  - Variance depends on the level of restriction of the fluid



- **White-matter: neuronal tissues**

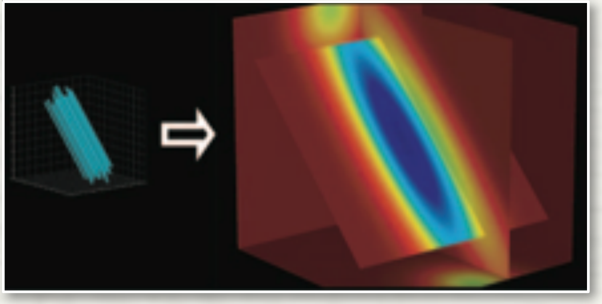


What happens in the brain?



- **Cerebrospinal Fluid (CSF)**
  - Displacements follow a **gaussian distribution** in every direction (**isotropic**)
  - Variance depends on the level of restriction of the fluid

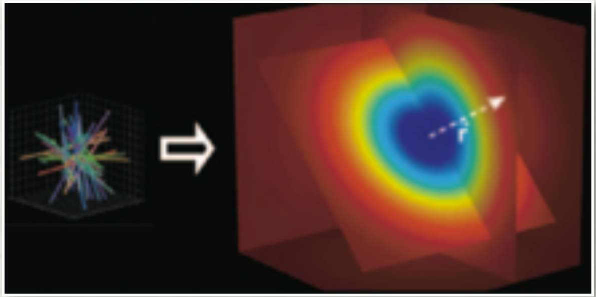
- **White-matter: neuronal tissues**



  - Diffusion is **more restricted** in directions *perpendicular* to the main axis of the fiber than *along* the fiber itself
  - Displacements still follow a gaussian distribution, but **anisotropic**

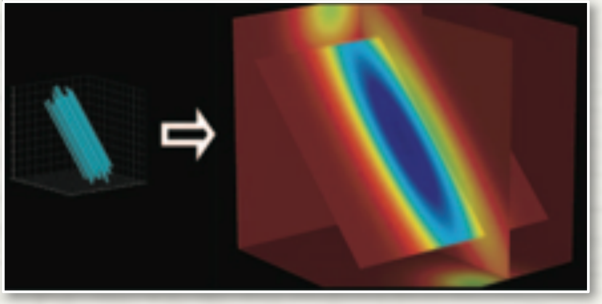
What happens in the brain?

- **Cerebrospinal Fluid (CSF)**
  - Displacements follow a **gaussian distribution** in every direction (**isotropic**)
  - Variance depends on the level of restriction of the fluid

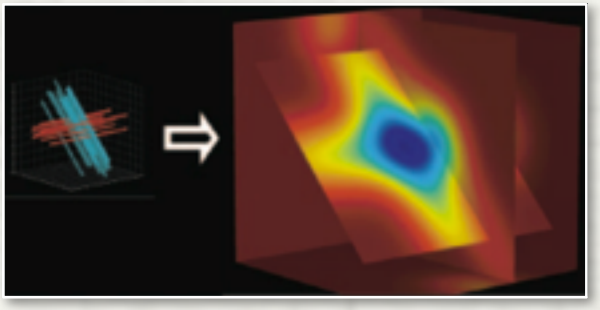


- **White-matter: neuronal tissues**

  - Diffusion is **more restricted** in directions *perpendicular* to the main axis of the fiber than *along* the fiber itself
  - Displacements still follow a gaussian distribution, but **anisotropic**

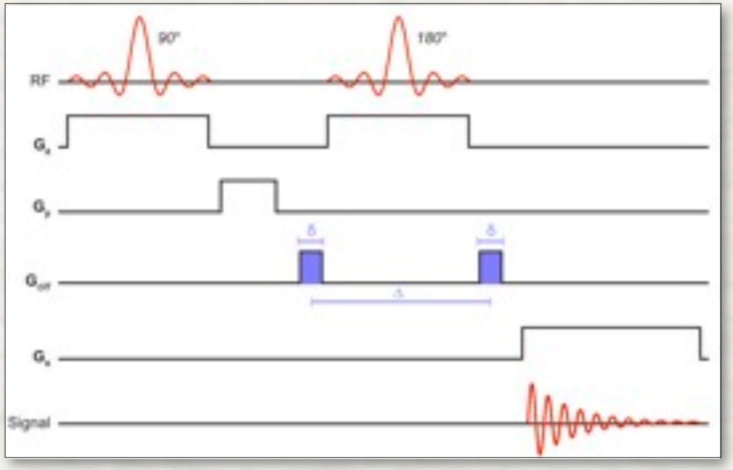


- **NOTE: regions with crossing fibers**
  - This situation **cannot** be described by a gaussian distribution!
  - Sophisticated techniques are required



Diffusion MRI in a nutshell

- MRI sequences are made sensitive to diffusion by inserting two additional magnetic field gradient pulses



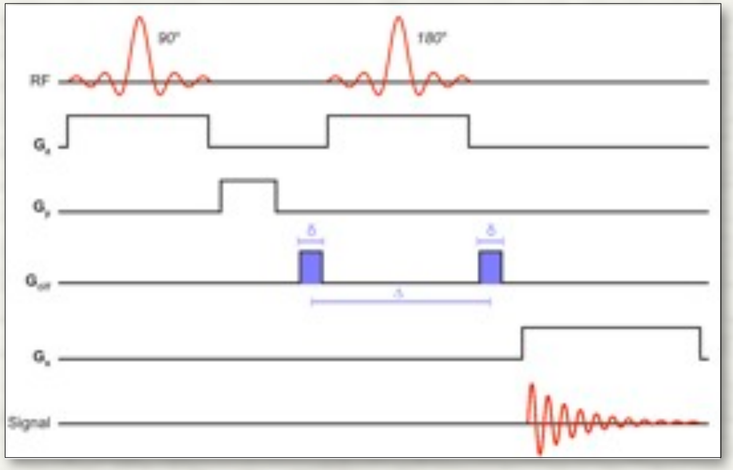
- The goal is to **dephase** spins which move in a given direction
- Signal **decays** exponentially as:

$$\text{signal} \propto e^{-b\,D\,t}$$

- *b* is called **b-value** and controls the diffusion weighting (contrast) of images

Diffusion MRI in a nutshell

- MRI sequences are made sensitive to diffusion by inserting two <u>additional</u> magnetic field gradient pulses

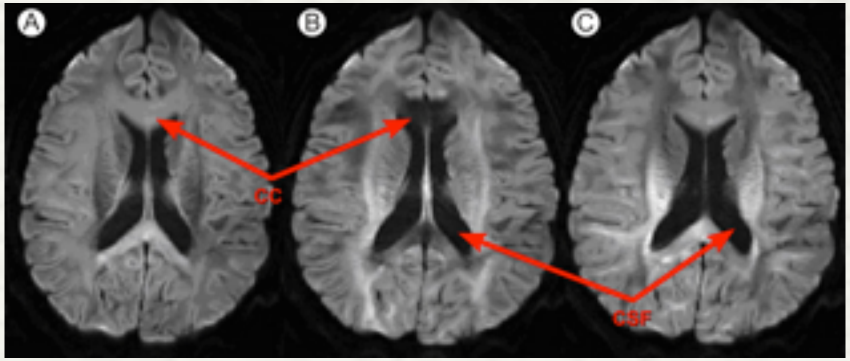


  - The goal is to **dephase** spins which move in a given direction
  - Signal **decays** exponentially as:

$$\text{signal} \propto e^{-b\,D\,t}$$

  - *b* is called **b-value** and controls the diffusion weighting (contrast) of images

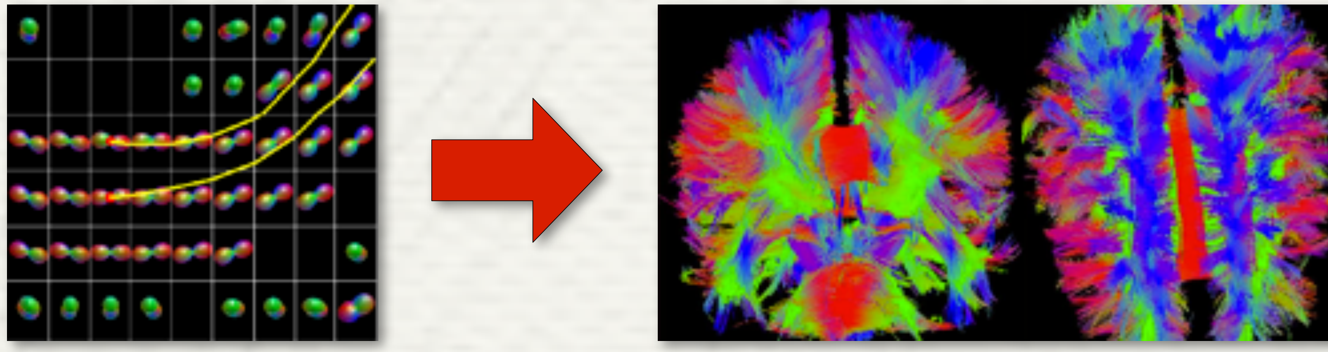
- Diffusion is <u>strongly dependent</u> on the gradient direction:



I-S L-R A-P

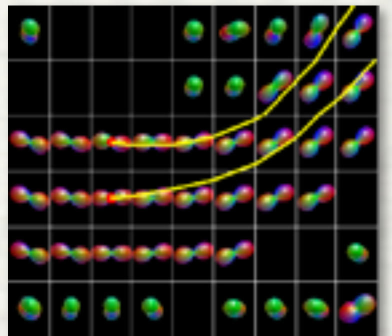
Main applications

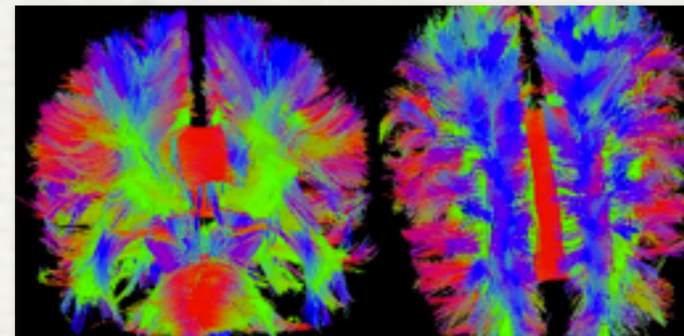
- **Fiber-tracking**: algorithms which infer axonal trajectories inside the brain by exploiting the diffusion information in each voxel



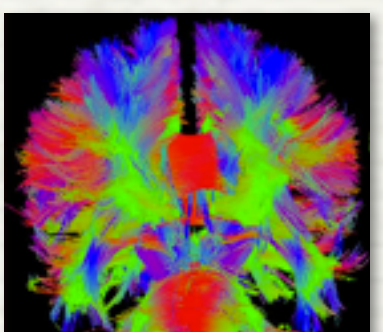
Main applications

- **Fiber-tracking**: algorithms which infer axonal trajectories inside the brain by exploiting the diffusion information in each voxel



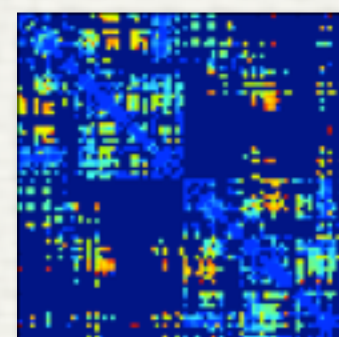


- **Connectivity analysis**: in-vivo and non-invasive assessment of structural wiring of the brain



fiber-tracking

cortical segmentation



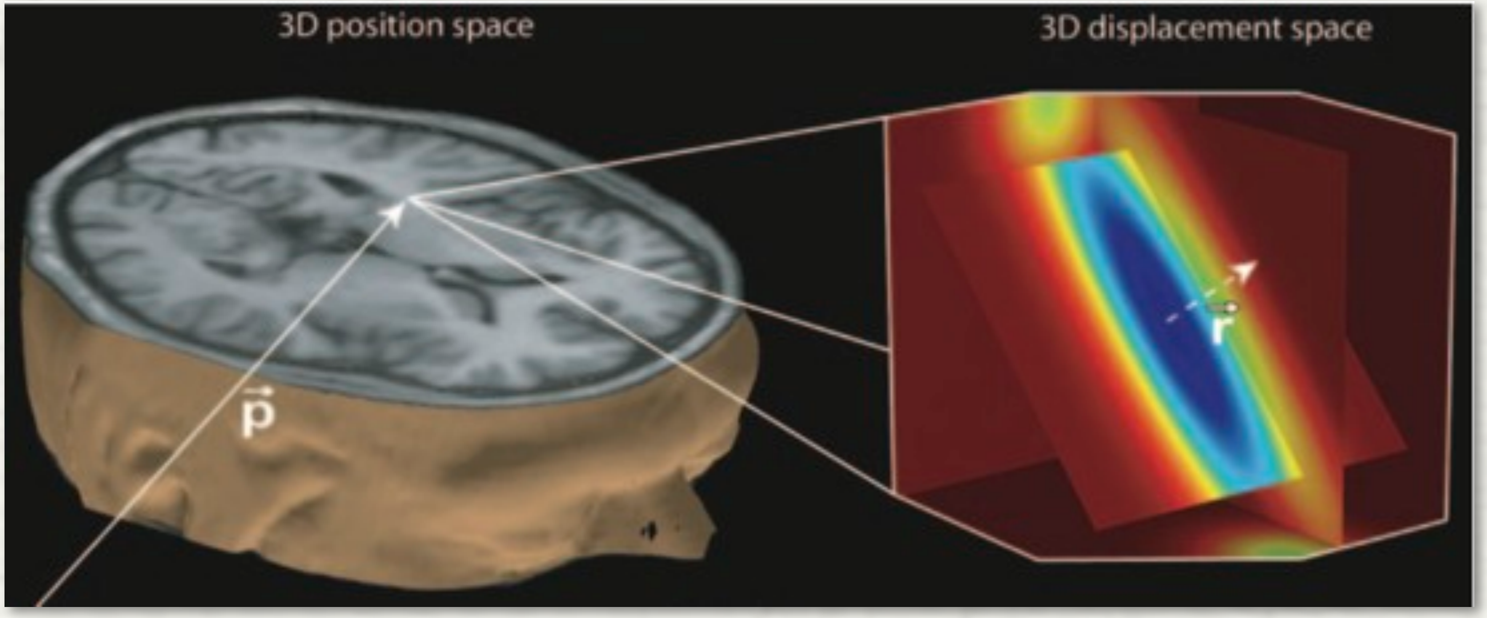
adjacency matrix

EAP, ODF and fODF

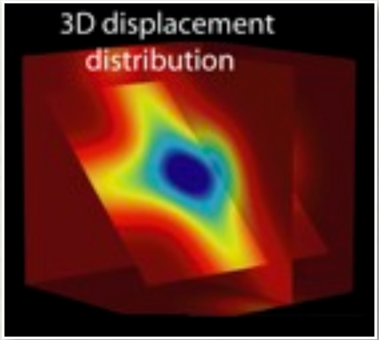
- **EAP** **(Ensemble Average Propagator)**
  - for every voxel, it's the *3D PDF* giving the probability of a given displacement ➡ diffusion MRI is a <u>6D modality</u>
  - related to the *signal attenuation* **E** by a 3D FFT:

$$\mathrm{P}(\vec{r}) = \int_{\mathbb{R}^3} E(\vec{q})\, e^{-2\pi i \vec{q}\cdot\vec{r}} d\vec{q}$$

q-space



EAP, ODF and fODF



- **EAP (Ensemble Average Propagator)**
  - for every voxel, it's the *3D PDF* giving the probability of a given displacement ➡ diffusion MRI is a 6D modality
  - related to the *signal attenuation* **E** by a 3D FFT:

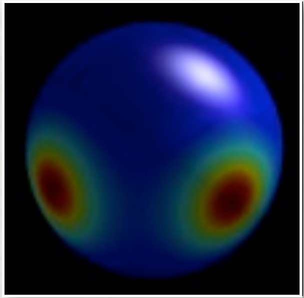
$$\mathrm{P}(\vec{r}) = \int_{\mathbb{R}^3} E(\vec{q})\, e^{-2\pi i \vec{q}\cdot\vec{r}} d\vec{q}$$

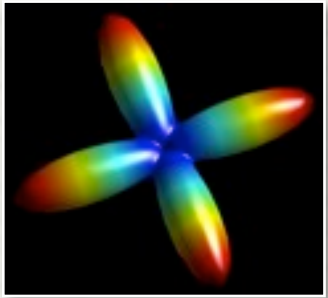
q-space

- **ODF (Orientation Distribution Function)**
  - probability of diffusion *along a given direction*:

$$\mathrm{ODF}(\hat{r}) = \int_{\mathbb{R}_+} P(r, \hat{r})\, r^2\, dr$$

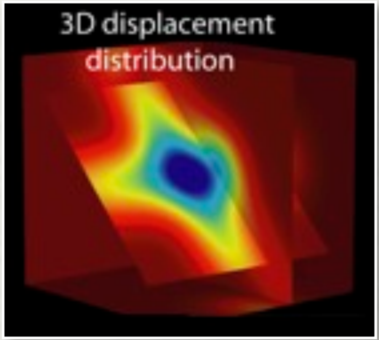
  - function on the unit sphere





EAP, ODF and fODF

- **EAP (Ensemble Average Propagator)**
  - for every voxel, it's the *3D PDF* giving the probability of a given displacement ➡ diffusion MRI is a <u>6D modality</u>
  - related to the *signal attenuation* **E** by a 3D FFT:



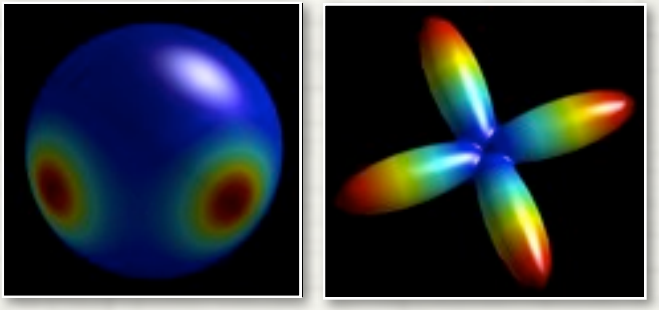
$$\mathrm{P}(\vec{r}) = \int_{\mathbb{R}^3} E(\vec{q})\, e^{-2\pi i \vec{q}\cdot\vec{r}} d\vec{q}$$

q-space

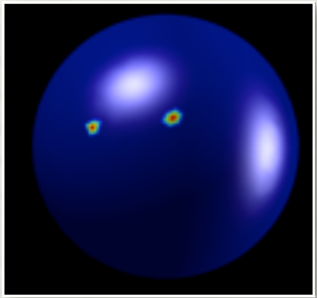
- **ODF (Orientation Distribution Function)**
  - probability of diffusion *along a given direction*:

$$\mathrm{ODF}(\hat{r}) = \int_{\mathbb{R}_+} P(r,\hat{r})\, r^2\, dr$$

  - function on the unit sphere



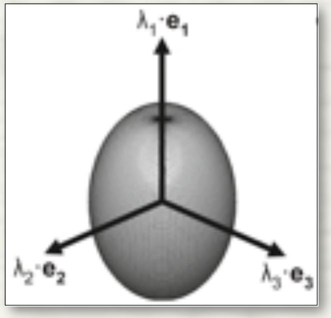
- **fODF (fiber ODF)**
  - sum of *spikes* identifying the fiber directions, with amplitudes corresponding to the *volume fractions*
  - function on the unit sphere



Diffusion TENSOR Imaging (DTI)

- **Assumption**: displacements follow a *gaussian distribution*

- Fully characterized by its **covariance matrix**
  - 3x3 symmetric positive semi-definite matrix, called Diffusion Tensor
  - plotted as an **ellipsoid** by means of its *eigenvalues* and *eigenvectors*
  - **6 degrees of freedom** (3 rotations + 3 variances)

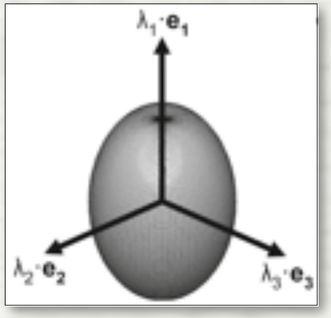


- The Diffusion Tensor can be estimated by *least-squares*
  (at least 6 DWI images along different directions are needed)

Diffusion TENSOR Imaging (DTI)

- **<u>Assumption</u>**: displacements follow a *gaussian distribution*

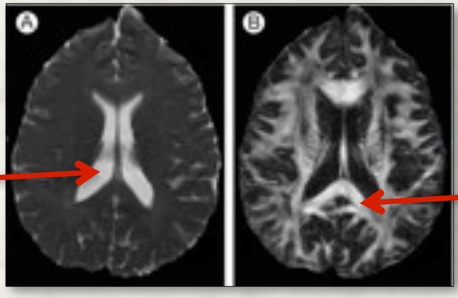
- Fully characterized by its **covariance matrix**
  - 3x3 symmetric positive semi-definite matrix, called Diffusion Tensor
  - plotted as an **ellipsoid** by means of its *eigenvalues* and *eigenvectors*
  - **6 degrees of freedom** (3 rotations + 3 variances)



- The Diffusion Tensor can be estimated by *least-squares*
  (at least 6 DWI images along different directions are needed)

- DTI scalar measures:

Mean Diffusivity:
high values = fast diffusion

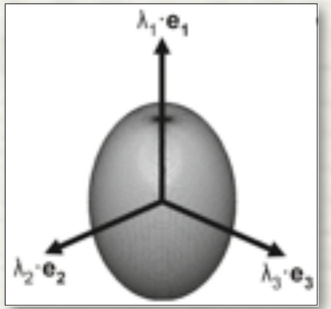


Fractional Anisotropy:
high values = fiber bundles

Diffusion TENSOR Imaging (DTI)

- **Assumption**: displacements follow a *gaussian distribution*

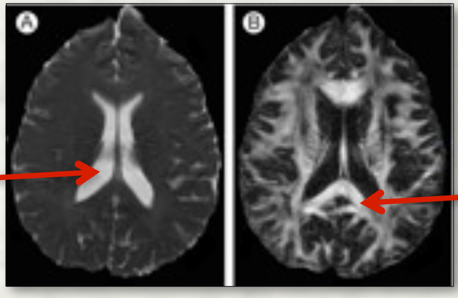
- Fully characterized by its **covariance matrix**
  - 3x3 symmetric positive semi-definite matrix, called Diffusion Tensor
  - plotted as an **ellipsoid** by means of its *eigenvalues* and *eigenvectors*
  - **6 degrees of freedom** (3 rotations + 3 variances)



- The Diffusion Tensor can be estimated by *least-squares*
  (at least 6 DWI images along different directions are needed)

- DTI scalar measures:

Mean Diffusivity:
high values = fast diffusion



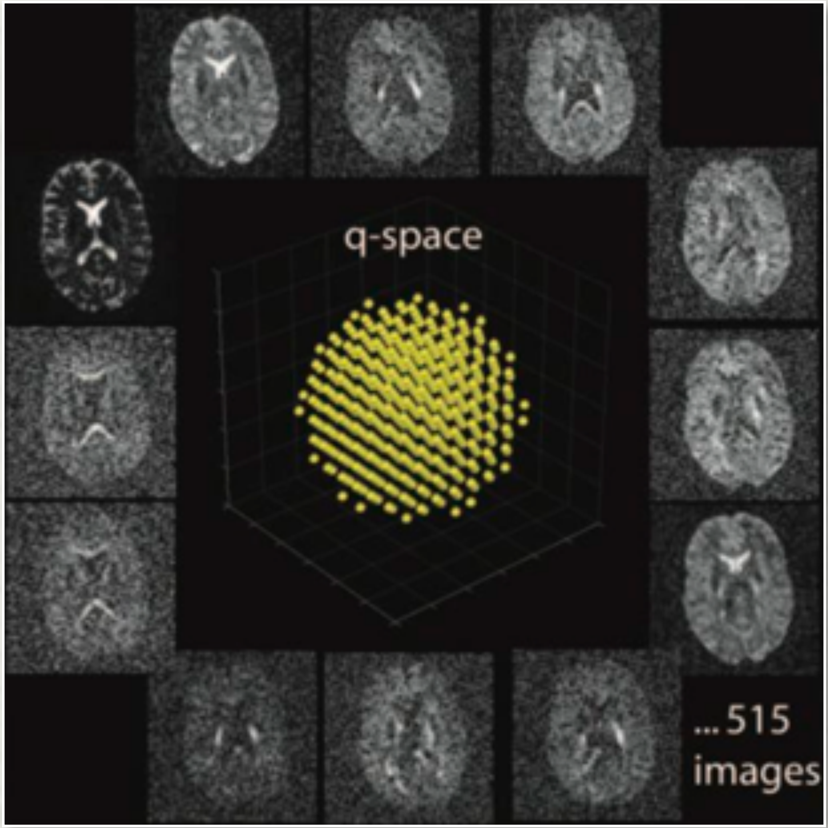
Fractional Anisotropy:
high values = fiber bundles

▲ **Fast acquisitions**: only 6 samples needed

▼ **Crossing fibers** cannot be modeled

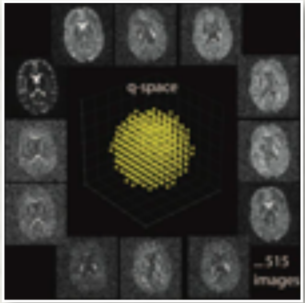
Diffusion SPECTRUM Imaging (DSI)

- Measures directly the displacement of water molecules making **no assumptions** about the underlying diffusion
  - acquire data on a *3D cartesian grid*
  - *FFT* the data in q-space to obtain the **EAP**
  - radial integration to obtain the **ODF**
  - ODF maxima extraction to have the **fODF**



Diffusion SPECTRUM Imaging (DSI)

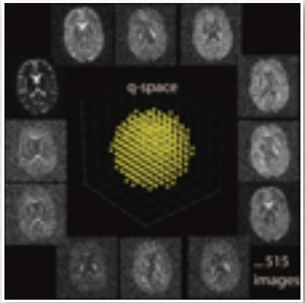
- Measures directly the displacement of water molecules making **no assumptions** about the underlying diffusion
  - acquire data on a *3D cartesian grid*
  - *FFT* the data in q-space to obtain the **EAP**
  - radial integration to obtain the **ODF**
  - ODF maxima extraction to have the **fODF**



- To recover the EAP, the q-space must be properly sampled:
  - 515 images usually acquired (acquisition time: DSI ≃ 60 min, DTI ≃ 3 min) forming a 11x11x11 cartesian grid (all points inside a sphere of radius 5)
  - high diffusion gradient strengths (*b-value* up to 8000 s/mm^2, DTI ≃ 1000 s/mm^2)

Diffusion SPECTRUM Imaging (DSI)

- Measures directly the displacement of water molecules making **no assumptions** about the underlying diffusion
  - acquire data on a *3D cartesian grid*
  - *FFT* the data in q-space to obtain the **EAP**
  - radial integration to obtain the **ODF**
  - ODF maxima extraction to have the **fODF**



- To recover the EAP, the q-space must be properly sampled:
  - 515 images usually acquired (acquisition time: DSI ≃ 60 min, DTI ≃ 3 min) forming a 11x11x11 cartesian grid (all points inside a sphere of radius 5)
  - high diffusion gradient strengths (*b-value* up to 8000 s/mm^2, DTI ≃ 1000 s/mm^2)

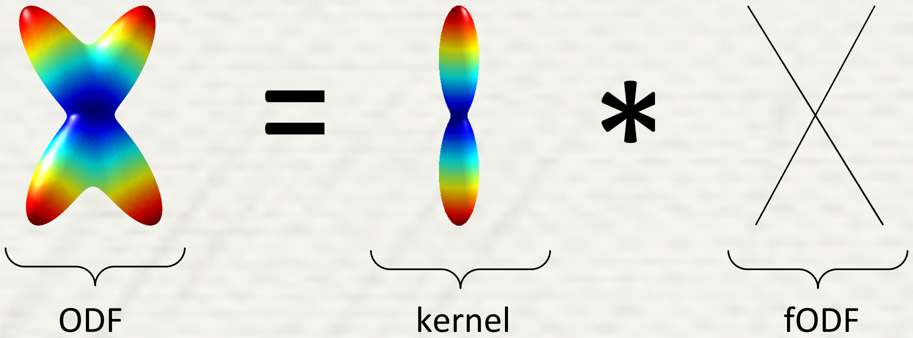
▲ Recovers the full **EAP** (full description of diffusion process)

▼ **Very long acquisitions**, no clinically feasible

▼ **Powerful gradients** are required

Spherical deconvolution (CSD)

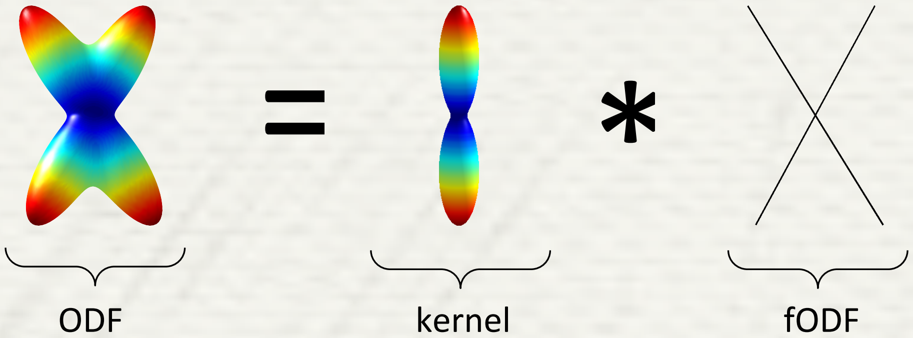
- The ODF can be seen as a **convolution on the sphere**



- The **kernel** characterizes the diffusion of a *single fiber*

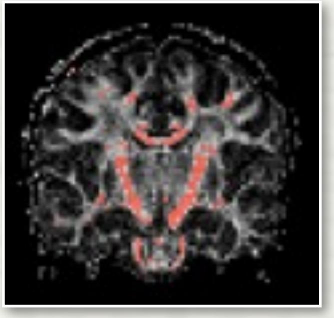
Spherical deconvolution (CSD)

- The ODF can be seen as a **convolution on the sphere**



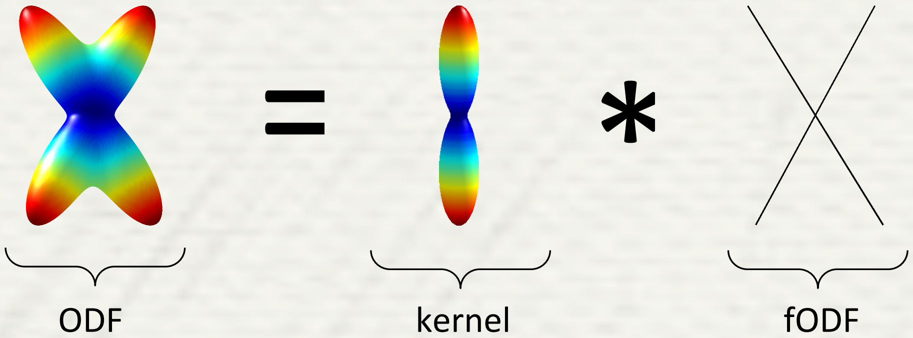
- The **kernel** characterizes the diffusion of a *single fiber*

- It can be easily estimated from the data
    1) identify known areas of the brain with just one fiber;
    2) fit a tensor in each voxel and averaging

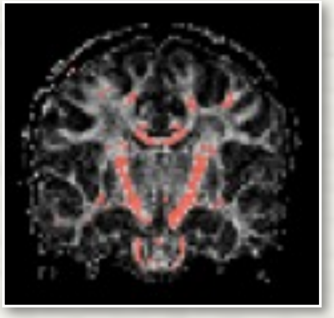


Spherical deconvolution (CSD)

- The ODF can be seen as a **convolution on the sphere**



- The **kernel** characterizes the diffusion of a *single fiber*



- It can be easily estimated from the data
    1) identify known areas of the brain with just one fiber;
    2) fit a tensor in each voxel and averaging

▲ Reduces the number of samples **down to 60**

▼ Sensitive to noise

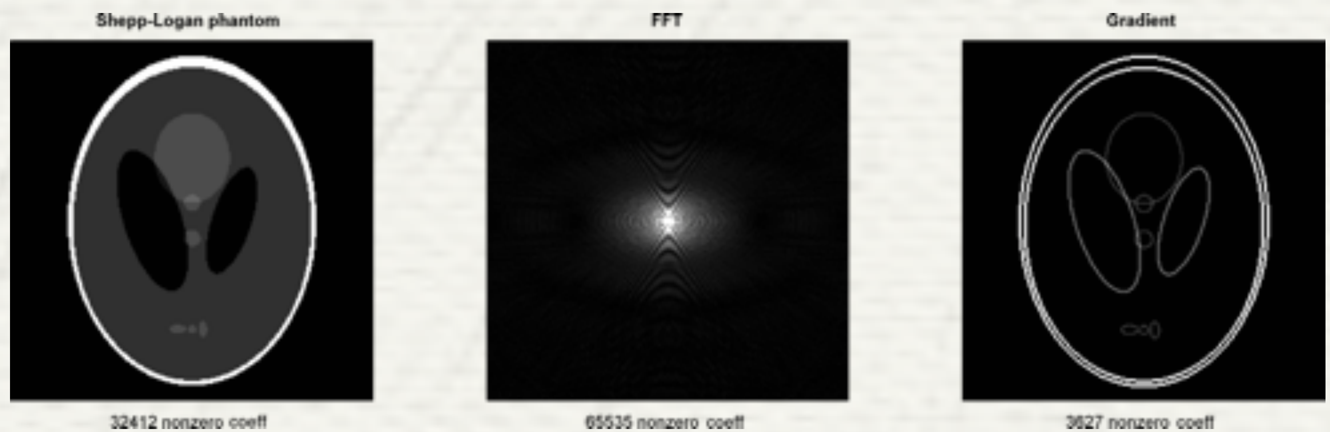
▼ Assumes the same diffusion properties across the brain

Compressed Sensing in a nutshell

- According to the **CS theory**, it is possible to recover a signal from less samples than those required by the Nyquist criterion, **provided that the signal is sparse** in some *representation*

Compressed Sensing in a nutshell

- According to the **CS theory**, it is possible to recover a signal from less samples than those required by the Nyquist criterion, **provided that the signal is sparse** in some *representation*



- Let $x \in \mathbb{R}^n$ be the signal to be recovered from the $m \ll n$ linear measurements $y = \Phi x \in \mathbb{R}^m$ s.t. $\alpha = \Psi x$ is sparse.
  If Φ and Ψ obey some randomness and incoherence conditions, then x can be recovered by solving an **inverse problem** like:

$$\underset{x}{\operatorname{argmin}} \, ||\Psi x||_1 \quad \text{s.t.} \quad ||\Phi x - y||_2 \leq \epsilon$$

sparsity basis — *sensing* basis

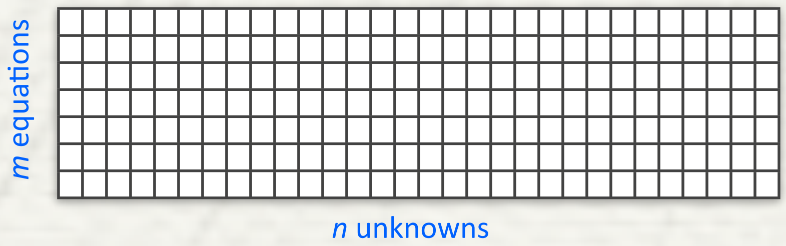
- **NOTE 1**: ℓ_1 is a convex relaxation for ℓ_0 ("true definition" of sparsity)

$$\underset{x}{\operatorname{argmin}} \; ||\Psi\, x||_0 \;\; \text{s.t.} \;\; ||\Phi\, x - y||_2 \leq \epsilon$$

- **NOTE 1**: ℓ_1 is a <u>convex relaxation</u> for ℓ_0 ("true definition" of sparsity)

$$\underset{x}{\text{argmin}} \; ||\Psi x||_0 \;\; \text{s.t.} \;\; ||\Phi x - y||_2 \leq \epsilon$$

- **NOTE 2**: since $\Phi \in \mathbb{R}^{m \times n}$ and $m \ll n$, then the system is under-determined and there are <u>infinite solutions</u>

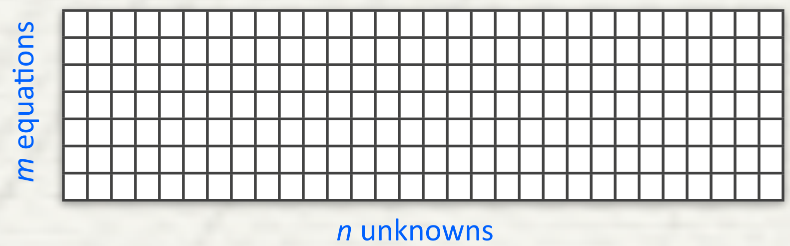


If there is a unique sparse solution, CS will find it!

- **NOTE 1**: ℓ_1 is a <u>convex relaxation</u> for ℓ_0 ("true definition" of sparsity)

$$\underset{x}{\operatorname{argmin}} \; ||\Psi\, x||_0 \;\; \text{s.t.} \;\; ||\Phi\, x - y||_2 \leq \epsilon$$

- **NOTE 2**: since $\Phi \in \mathbb{R}^{m \times n}$ and $m \ll n$, then the system is under-determined and there are <u>infinite solutions</u>



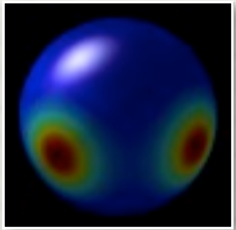
If there is a unique sparse solution, CS will find it!

- **NOTE 3**: several variants

$\underset{x}{\operatorname{argmin}} \; \|\|\Psi\, x\|\|_1 \;\; \text{s.t.} \;\; \|\|\Phi\, x - y\|\|_2 \leq \epsilon$	BPDN
$\underset{x}{\operatorname{argmin}} \; \|\|\Phi\, x - y\|\|_2 \; + \lambda \, \|\|\Psi\, x\|\|_1$	ℓ_1 regularized Least Squares
$\underset{x}{\operatorname{argmin}} \; \|\|\Phi\, x - y\|\|_2 \;\; \text{s.t.} \;\; \|\|\Psi\, x\|\|_1 \leq \tau$	LASSO
$\underset{x}{\operatorname{argmin}} \; \|\|x\|\|_{TV} \;\; \text{s.t.} \;\; \|\|\Phi\, x - y\|\|_2 \leq \epsilon$	Total Variation minimization

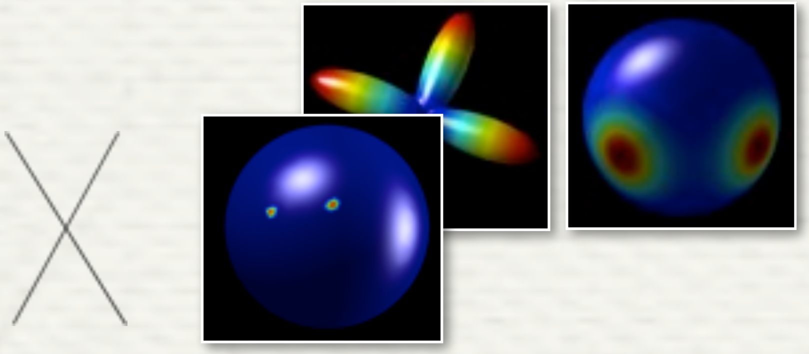
Sparsity in diffusion MRI

- The **ODF** is not sparse at all...



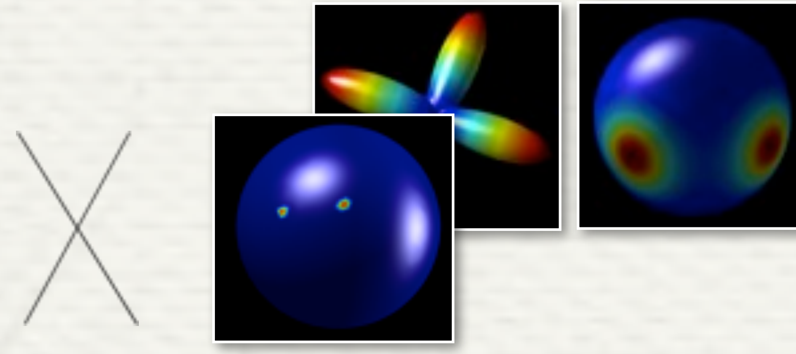
Sparsity in diffusion MRI

- The **ODF** is not sparse at all...
- ...but the **fODF** is very sparse!



Sparsity in diffusion MRI

- The **ODF** is not sparse at all...
- ...but the **fODF** is very sparse!



- Recast the reconstruction problem as:

$$\min \; \|x\|_1 \quad \text{s.t.} \quad \begin{cases} \|Ax - y\|_2 < \epsilon \\ x_i \geq 0 \end{cases}$$

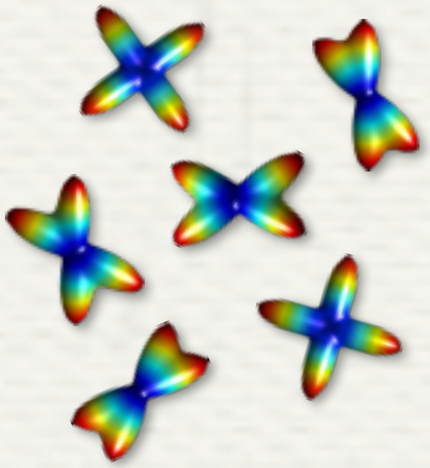
Ψ is the **identity** in this setting!

where:

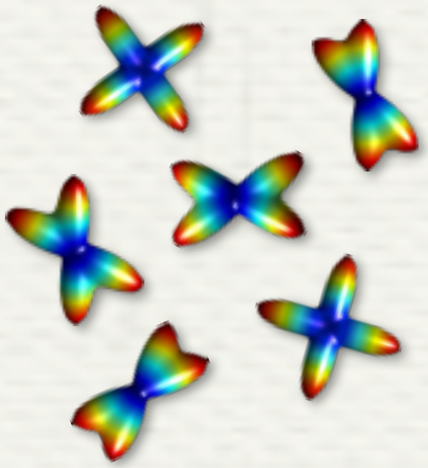
- y is the acquired **MRI data**
- A is a **dictionary of atoms** (as in Matching Pursuit) built from the single-fiber kernel estimated from the data and rotated along each possible direction

- x is the **fODF** we want to recover (contributions, i.e. *volume fractions*, of each atom to the final ODF)
- ϵ can be statistically estimated from data

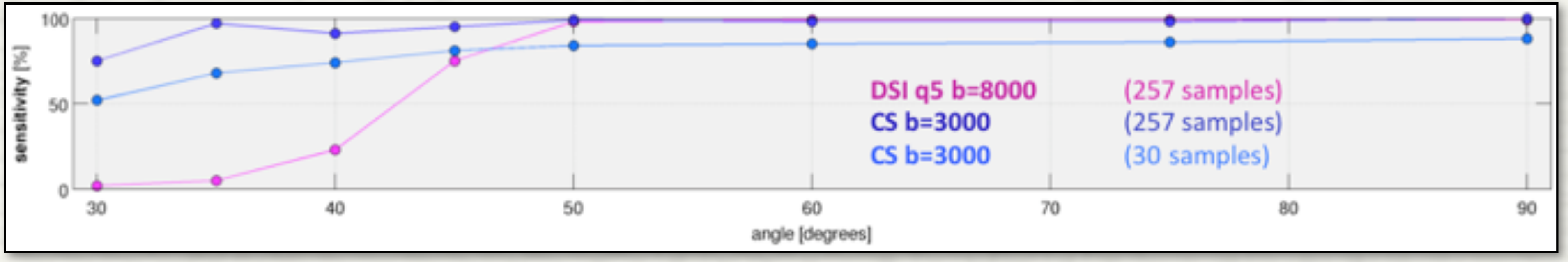
- **Simulations** of voxel configurations with:
  - 2 fibers crossing at given *angles*, ranging from 30° to 90°
  - different combinations of *noise*, *b-value, volume fractions* and *orientation* in space of the same configuration
  - 1000 *repetitions* with different realizations of the noise



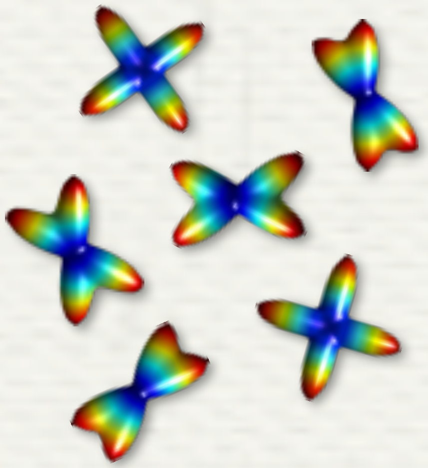
- **Simulations** of voxel configurations with:
  - 2 fibers crossing at given *angles*, ranging from 30° to 90°
  - different combinations of *noise*, *b-value, volume fractions* and *orientation* in space of the same configuration
  - 1000 *repetitions* with different realizations of the noise



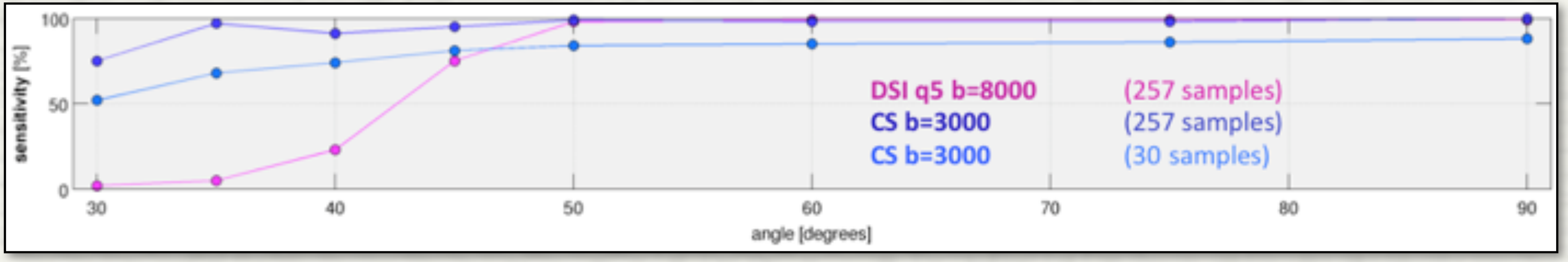
▲ CS-based reconstruction performs even better than DSI with only **10% of the samples**!



- **Simulations** of voxel configurations with:
  - 2 fibers crossing at given *angles*, ranging from 30° to 90°
  - different combinations of *noise*, *b-value, volume fractions* and *orientation* in space of the same configuration
  - 1000 *repetitions* with different realizations of the noise



▲ CS-based reconstruction performs even better than DSI with only **10% of the samples**!



▼ Unfortunately, already done...

B.A. Landman et al. "*Resolution of crossing fibers with constrained compressed sensing using diffusion tensor MRI*". **NeuroImage** (November 2011)

State-of-the-art method

- In [Landman, 2011] they formulated the problem as:

$$\begin{aligned} \min \quad & ||Ax - y||_2^2 + \lambda ||x||_1 \\ \text{s.t.} \quad & x_i \geq 0 \end{aligned}$$

where λ is *empirically* set to $0.1 \cdot ||2A^T y||_\infty$.

State-of-the-art method

- In [Landman, 2011] they formulated the problem as:

$$\begin{aligned} \min \quad & ||Ax - y||_2^2 + \lambda||x||_1 \\ \text{s.t.} \quad & x_i \geq 0 \end{aligned}$$

where λ is *empirically* set to $0.1 \cdot ||2A^T y||_\infty$.

▼ Performs very well (**30 samples**)

State-of-the-art method

- In [Landman, 2011] they formulated the problem as:

$$\begin{aligned} \min \quad & ||Ax - y||_2^2 + \lambda ||x||_1 \\ \text{s.t.} \quad & x_i \geq 0 \end{aligned}$$

where λ is *empirically* set to $0.1 \cdot ||2A^T y||_\infty$.

▼ Performs very well (**30 samples**)

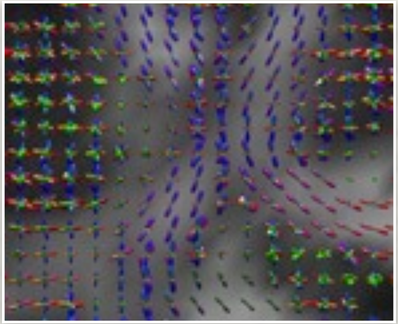
▲ However, there's still **room for improvements**:

(i) it turns out that sparsity is not properly exploited, since by definition:

volume fractions

$$\sum_i x_i = 1 \quad \Rightarrow \quad ||x||_1 = 1$$

(ii) they do not exploit any spatial correlation among voxels, which instead is the case in real world



Proposal 1: enhancing sparsity

- Re-weighted ℓ_1 minimization:

$$\begin{aligned} \min \quad & ||Ax - y||_2^2 + \lambda ||Wx||_1 \\ \text{s.t.} \quad & x_i \geq 0 \end{aligned}$$

$$\begin{aligned} \min \quad & ||Wx||_1 \\ \text{s.t.} \quad & \begin{cases} ||Ax - y||_2 < \epsilon \\ x_i \geq 0 \end{cases} \end{aligned}$$

  - initial weights $W_i^{(0)} = 1, \ \forall i$
  - at iteration $\tau + 1$, the problem is solved with weights computed from previous solution:

$$W_i^{(\tau+1)} = \frac{1}{|x_i^{(\tau)}| + \epsilon}, \ \forall i$$

Proposal 1: enhancing sparsity

- Re-weighted ℓ_1 minimization:

$$\begin{aligned} \min \quad & ||Ax - y||_2^2 + \lambda ||Wx||_1 \\ \text{s.t.} \quad & x_i \geq 0 \end{aligned}$$

$$\begin{aligned} \min \quad & ||Wx||_1 \\ \text{s.t.} \quad & \begin{cases} ||Ax - y||_2 < \epsilon \\ x_i \geq 0 \end{cases} \end{aligned}$$

  - initial weights $W_i^{(0)} = 1, \ \forall i$
  - at iteration $\tau + 1$, the problem is solved with weights computed from previous solution:

$$W_i^{(\tau+1)} = \frac{1}{|x_i^{(\tau)}| + \epsilon}, \ \forall i$$

$\tau = 0$

Proposal 1: enhancing sparsity

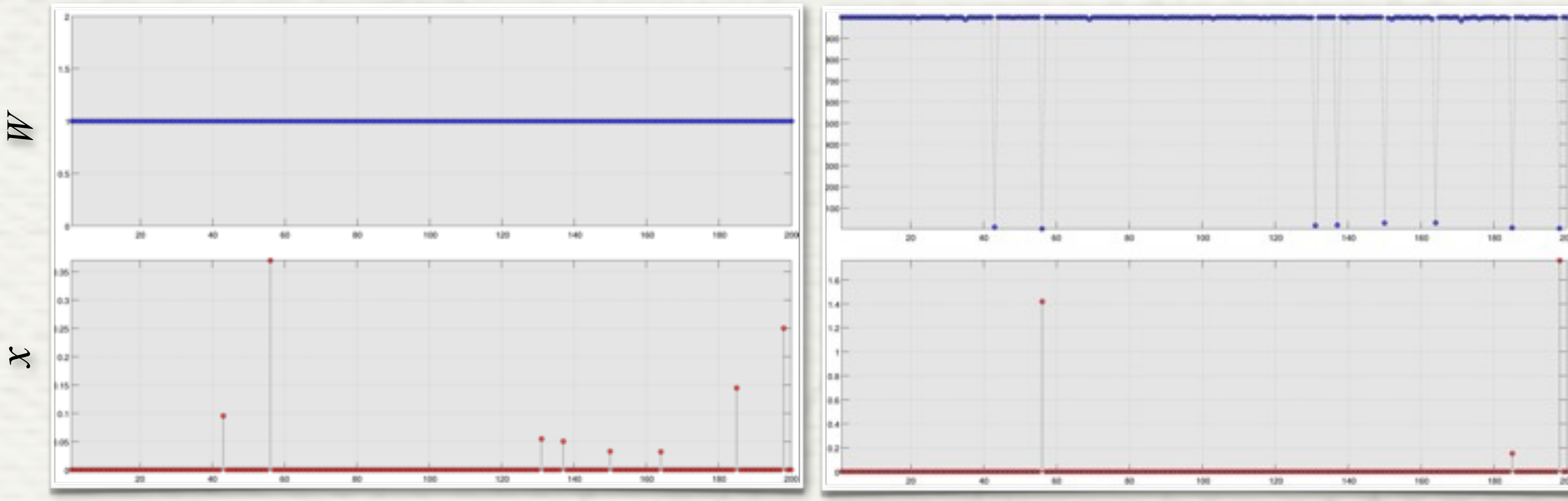
- Re-weighted ℓ_1 minimization:

$$\begin{aligned} \min \quad & \|Ax - y\|_2^2 + \lambda \|Wx\|_1 \\ \text{s.t.} \quad & x_i \geq 0 \end{aligned}$$

$$\begin{aligned} \min \quad & \|Wx\|_1 \\ \text{s.t.} \quad & \begin{cases} \|Ax - y\|_2 < \epsilon \\ x_i \geq 0 \end{cases} \end{aligned}$$

  - initial weights $W_i^{(0)} = 1, \quad \forall i$
  - at iteration $\tau + 1$, the problem is solved with weights computed from previous solution:

$$W_i^{(\tau+1)} = \frac{1}{|x_i^{(\tau)}| + \epsilon}, \quad \forall i$$



$\tau = 0$ $\tau = 1$

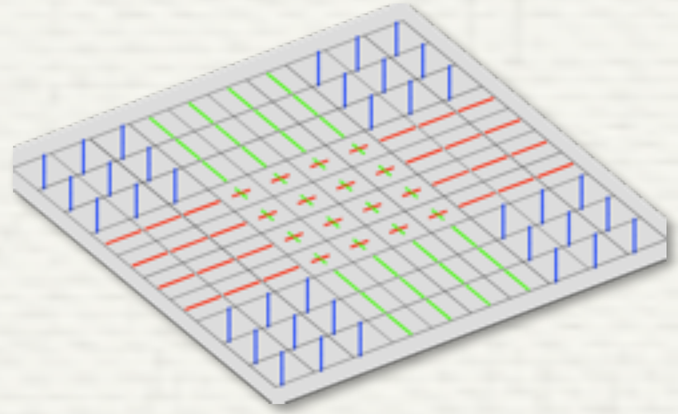
Comparison settings

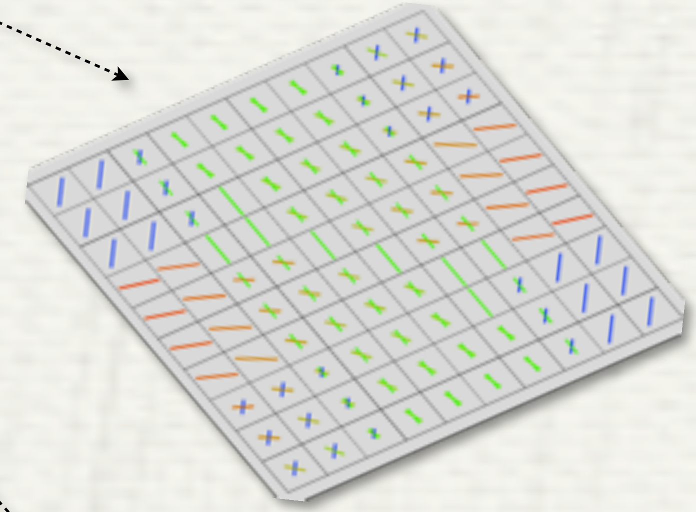
- <u>3 synthetic fields</u>

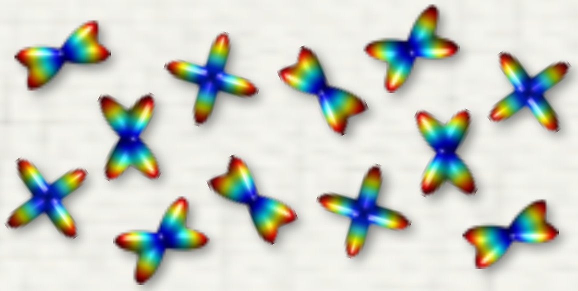
  (1) 90° fiber crossing

  (2) more realistic fiber crossing

  (3) single-voxel configurations crossing at given angles







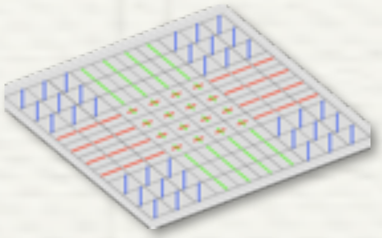
- <u>3 quality measures</u>

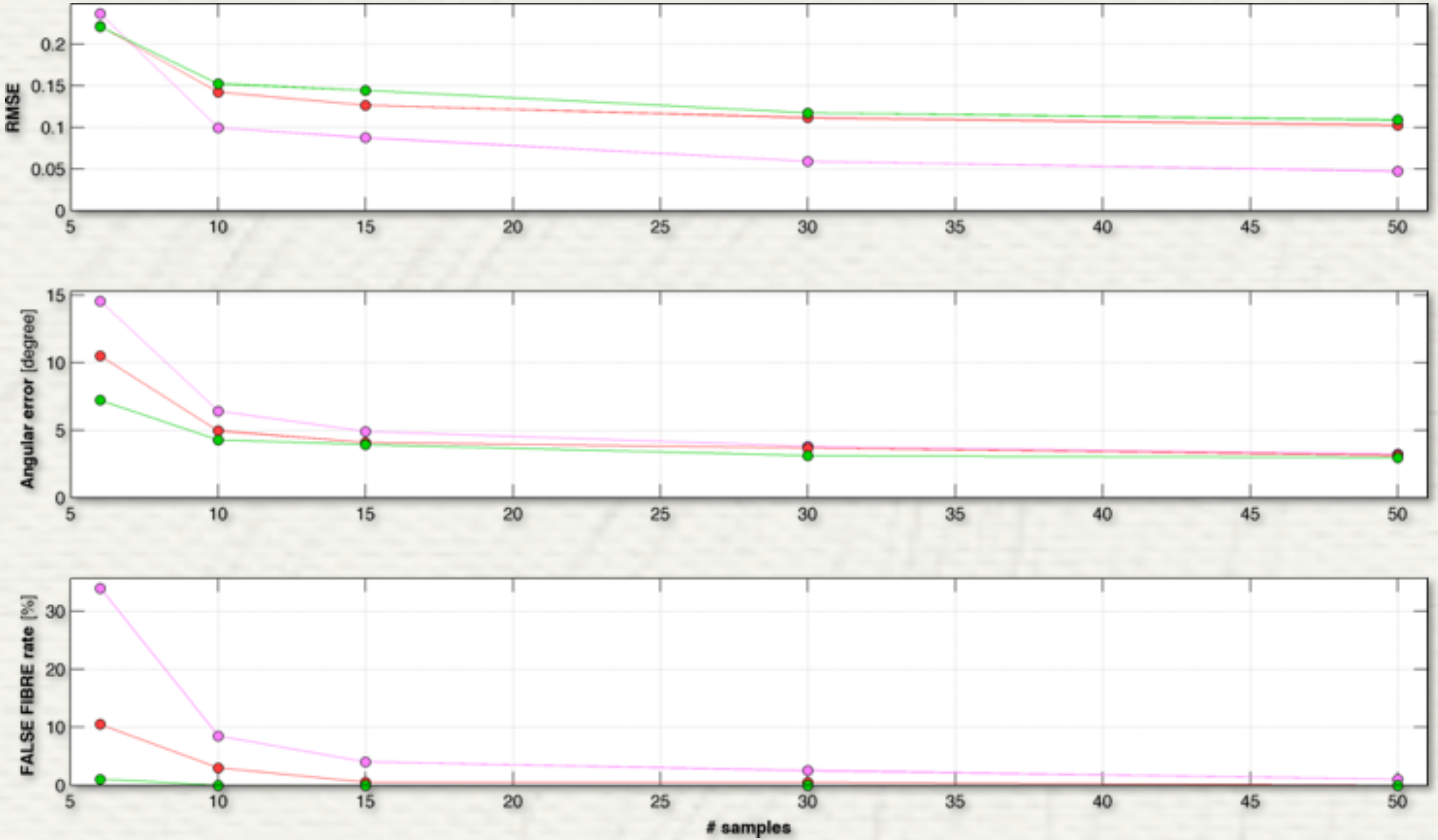
  (1) **MSE** with the ground-truth ODFs

  (2) **average angular error** in estimating the orientations of the fibers

  (3) **false fiber rate**

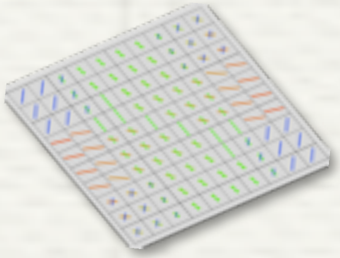
Results 1/3

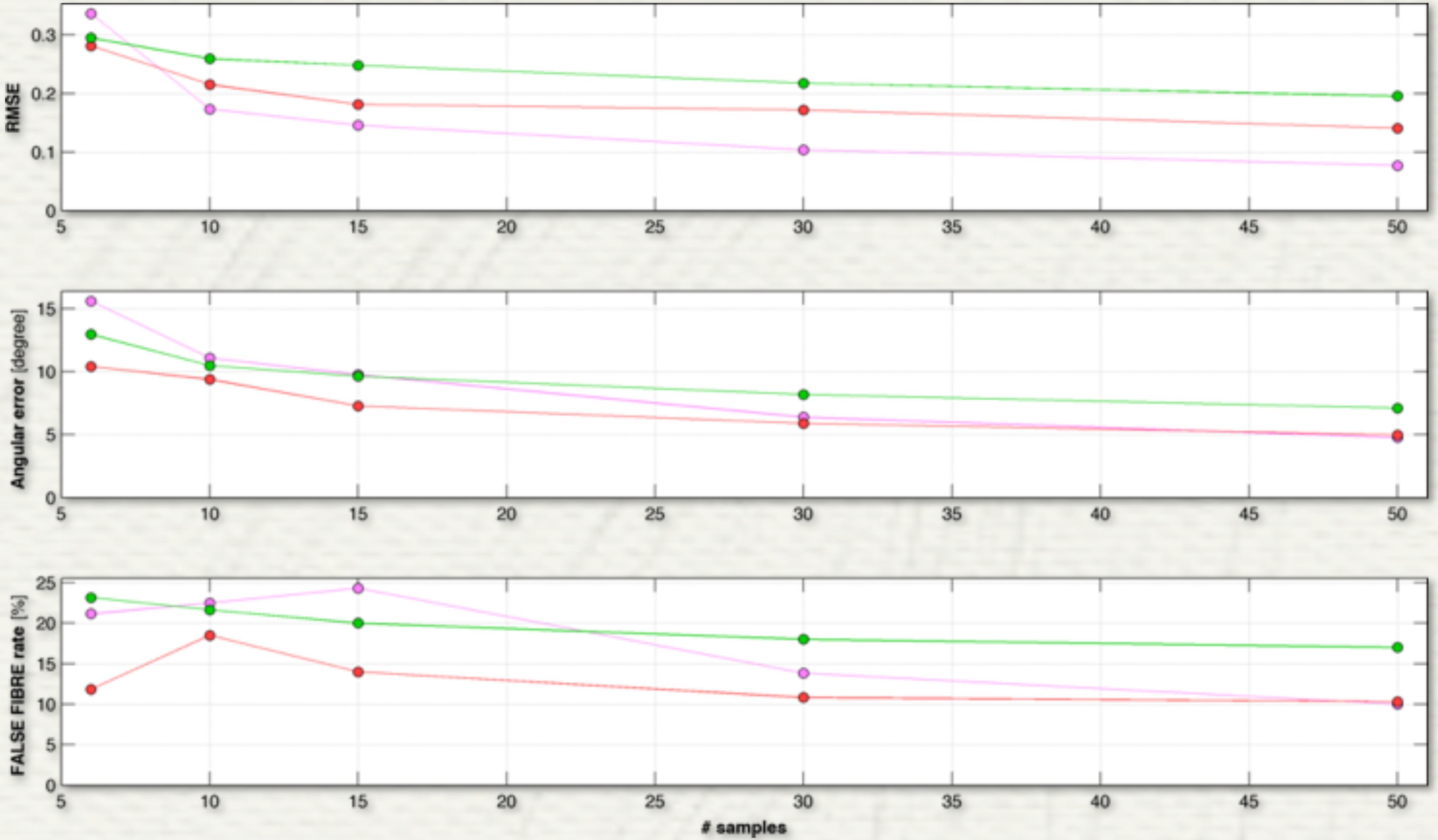




RMSE
0.2
0.15
0.1
0.05
0
5
10
15
20
25
30
35
40
45
50
Angular error [degree]
15
10
5
0
FALSE FIBRE rate [%]
30
20
10
0
samples

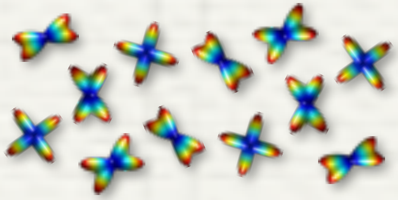
Results 2/3



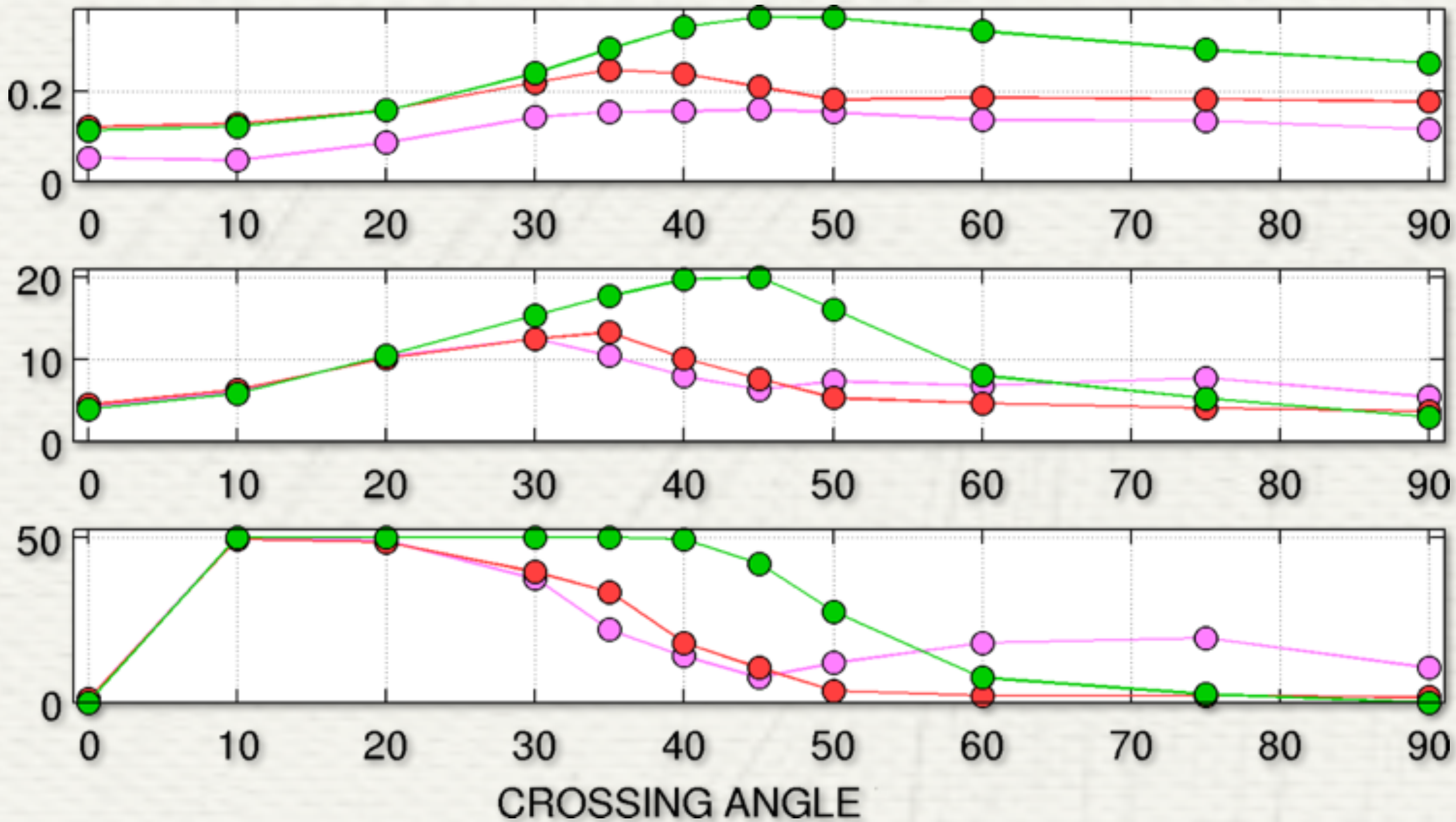


RMSE
0.3
0.2
0.1
0
5
10
15
20
25
30
35
40
45
50
Angular error [degree]
15
10
5
0
FALSE FIBRE rate [%]
25
20
15
10
5
0
samples

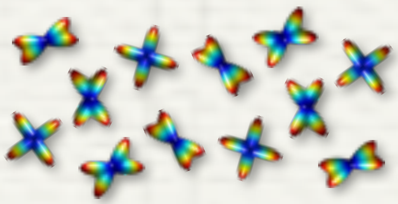
Results 3/3



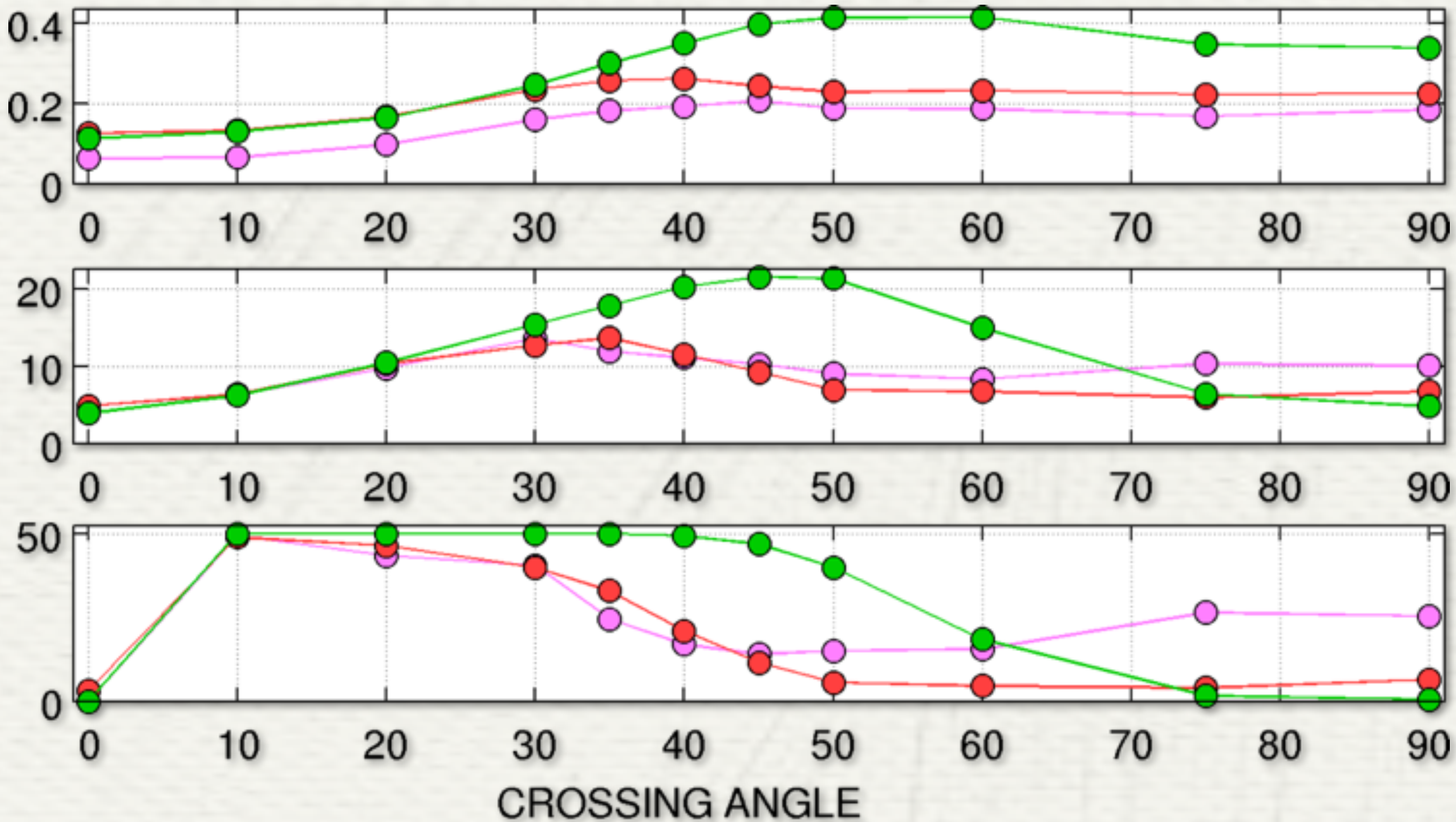
Reconstruction with: **30 samples**



Results 3/3

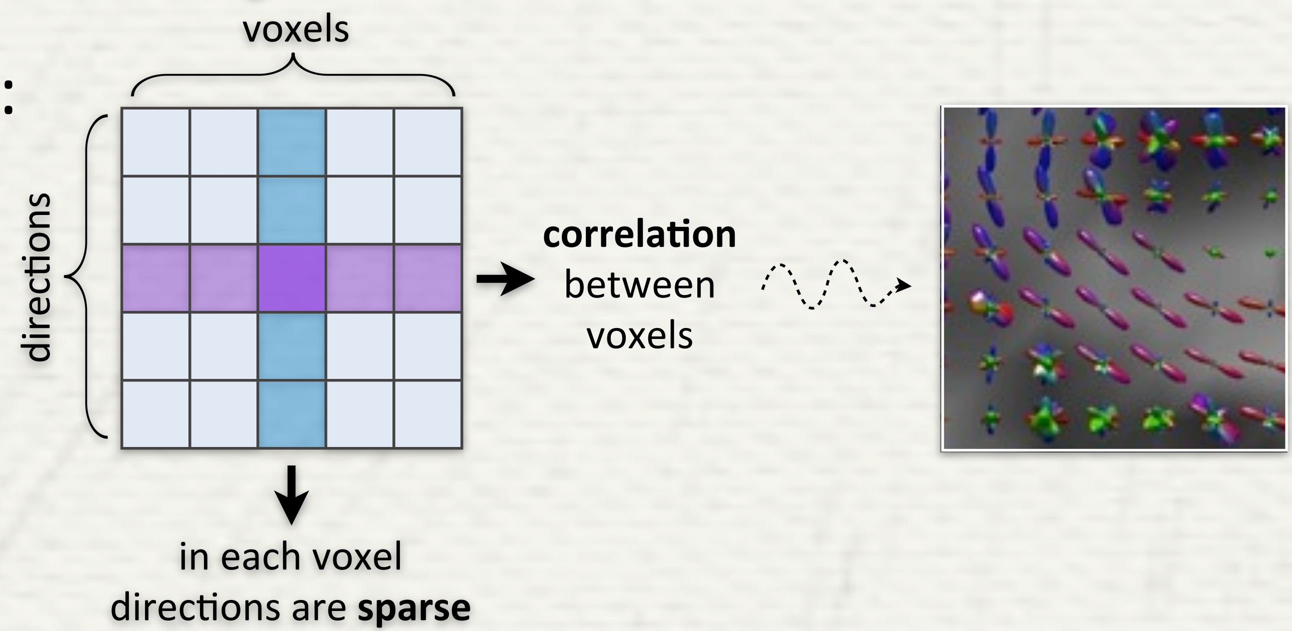


Reconstruction with: **15 samples**



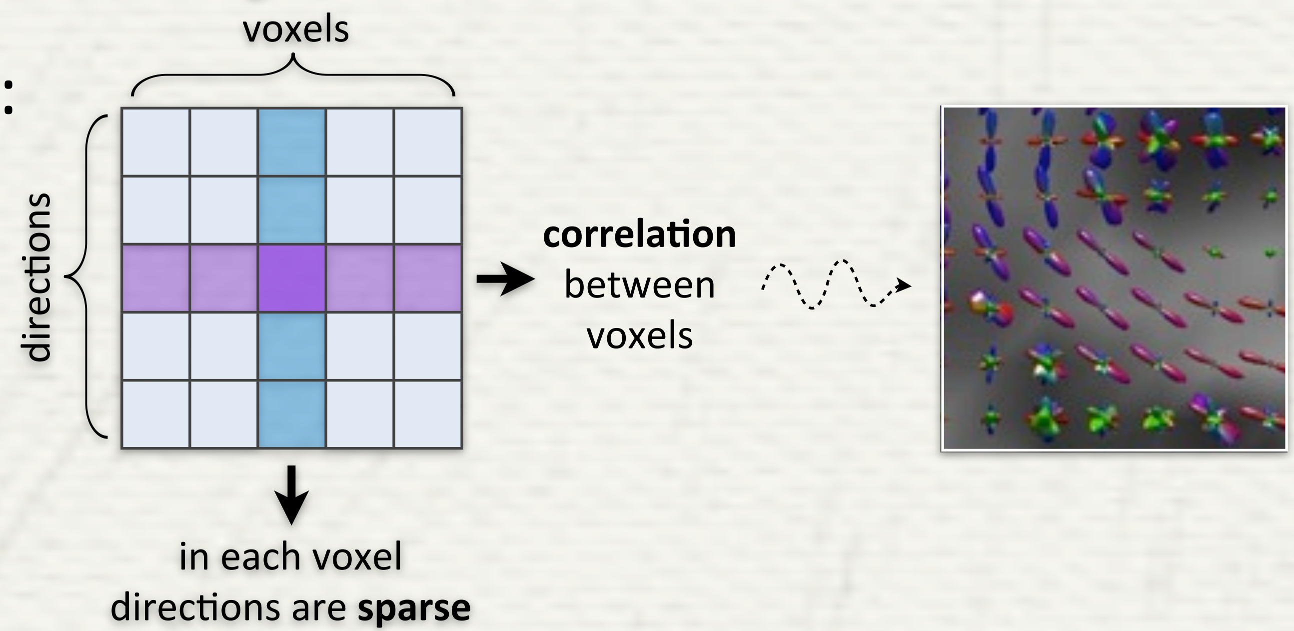
Proposal 2: spatial correlation

- Re-arranging data:



Proposal 2: spatial correlation

- Re-arranging data:



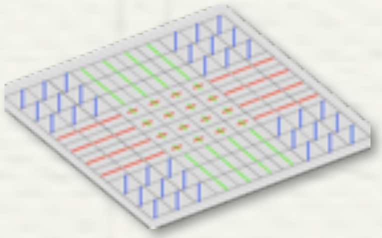
- Reconstruction problem re-formulated as:

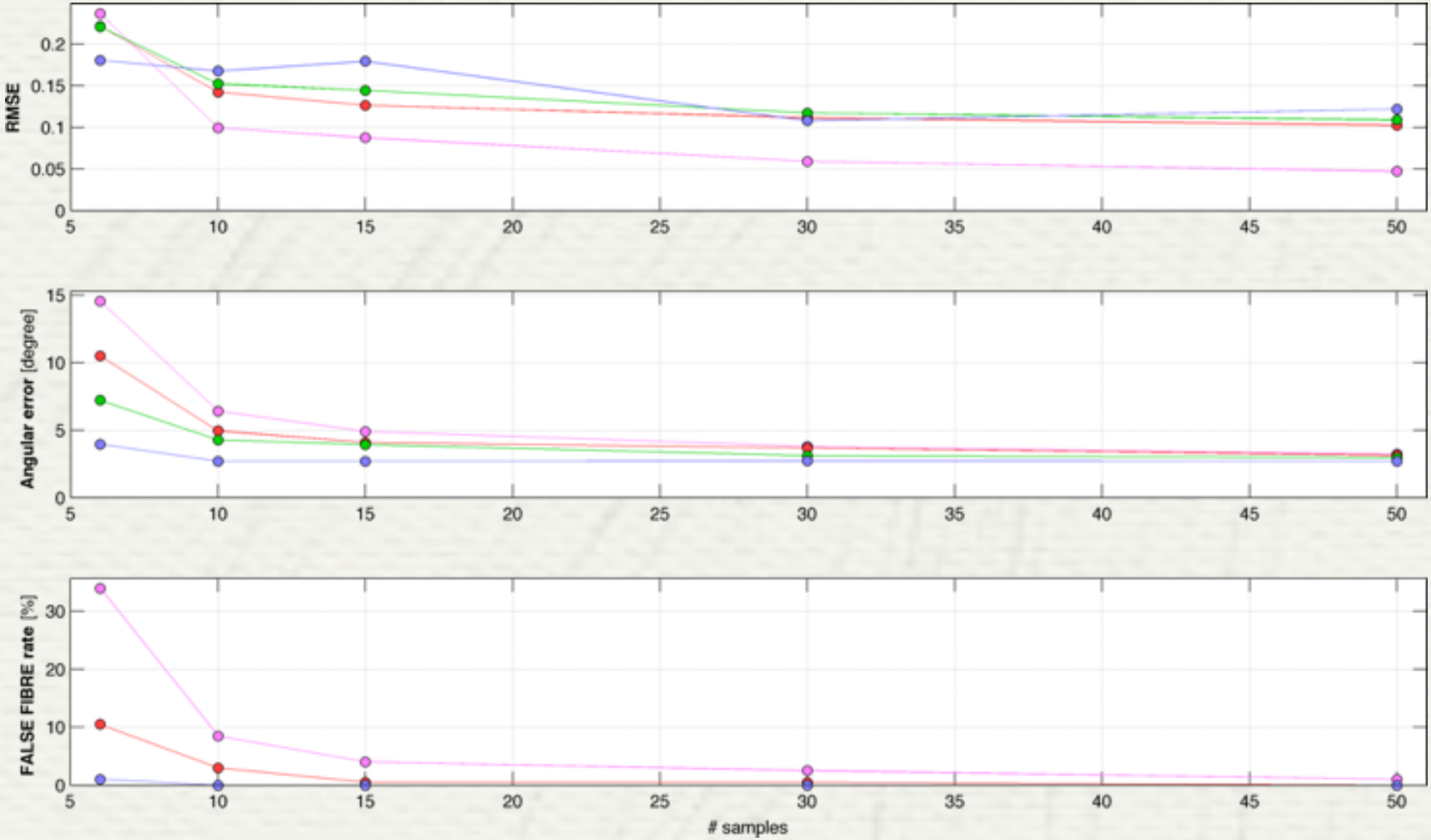
$$\min \quad \|X\|_* + \alpha \|X\|_{1,1}$$
$$\text{s.t.} \quad \begin{cases} \|\mathcal{A}(X) - Y\|_F < \epsilon \\ X_{ij} \geq 0 \end{cases}$$

  - $\|X\|_{1,1}$ tries to enforce **sparsity** in each voxel
  - $\|X\|_*$ tries to exploit the **correlation** of directions among voxels

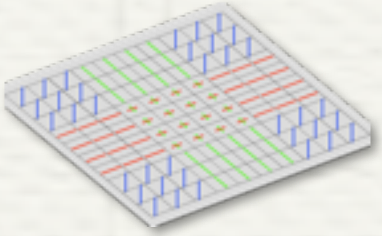
    (as before, it is a convex relaxation for **rk(A)**)

Results 1/2

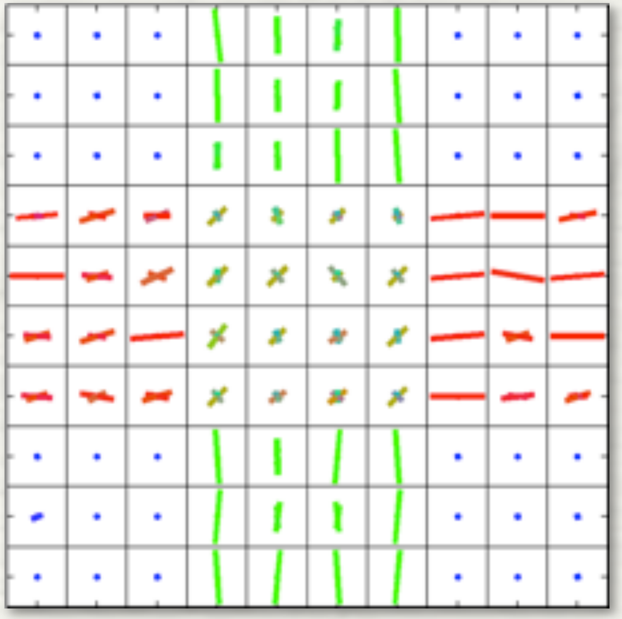



RMSE
0.2
0.15
0.1
0.05
0
5
10
15
20
25
30
35
40
45
50
Angular error [degree]
15
10
5
0
FALSE FIBRE rate [%]
30
20
10
0
samples

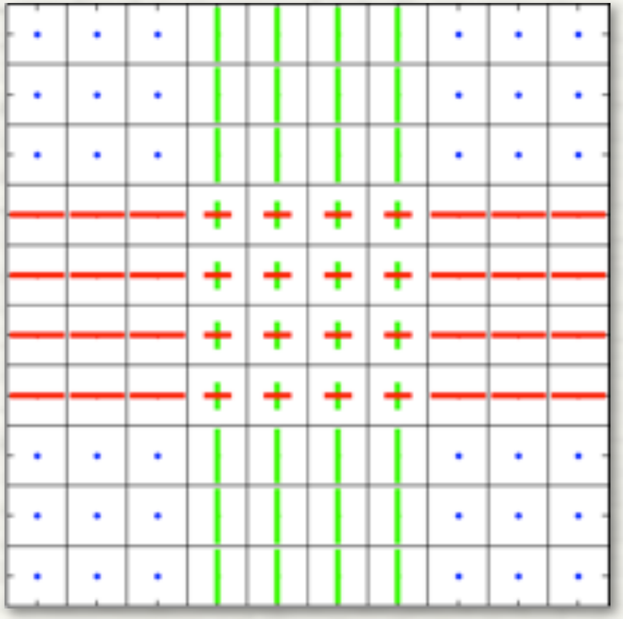
Results 1/2



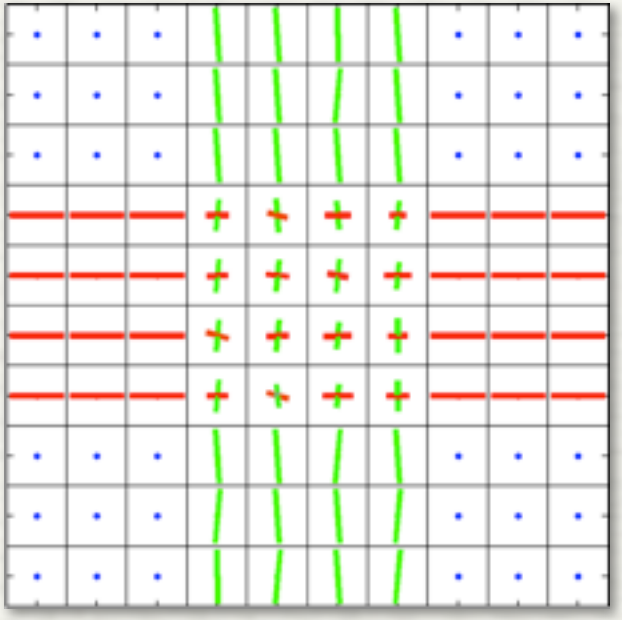
Re-weighted ℓ_1



GROUND-TRUTH

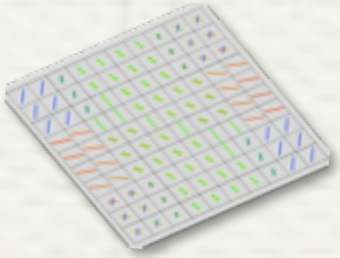


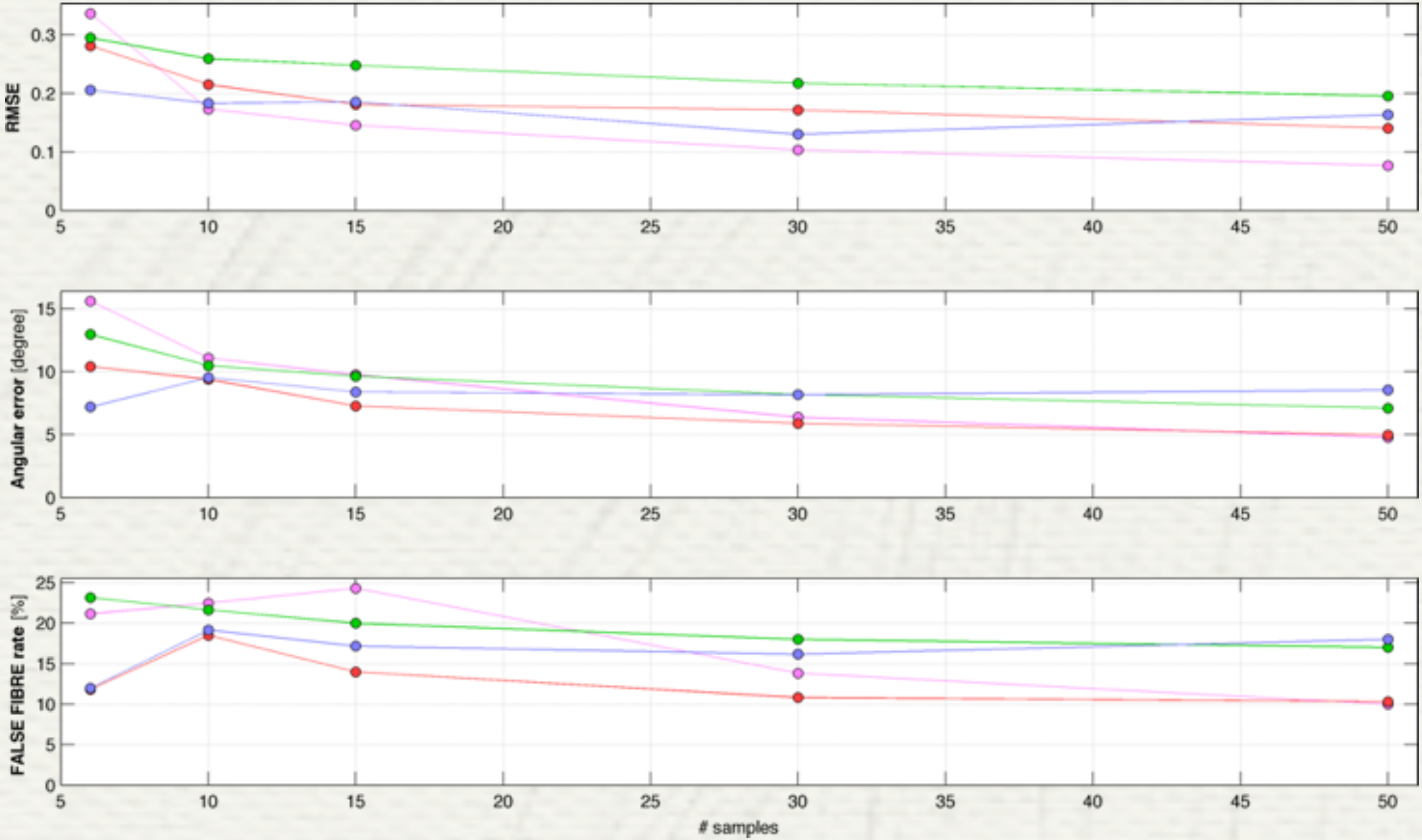
LR



(simulations with 6 samples)

Results 2/2

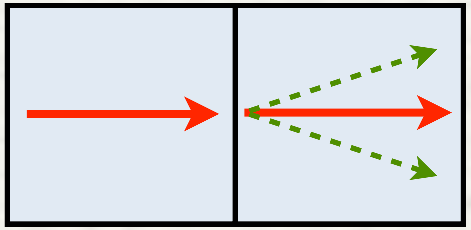


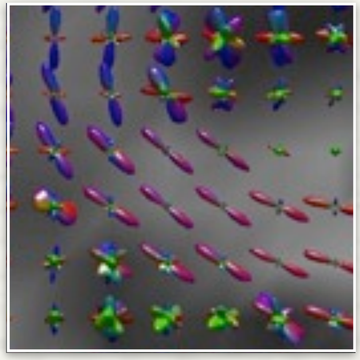


RMSE
0.3
0.2
0.1
0
5
10
15
20
25
30
35
40
45
50
Angular error [degree]
15
10
5
0
FALSE FIBRE rate [%]
25
20
15
10
5
0
samples

LR minimization: issues

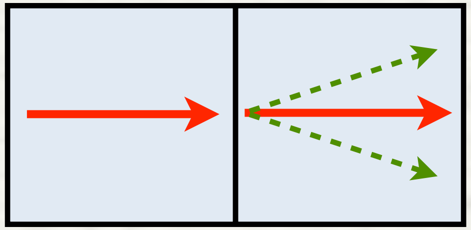
- With this formulation we do not account for the concept of “**similar directions**”

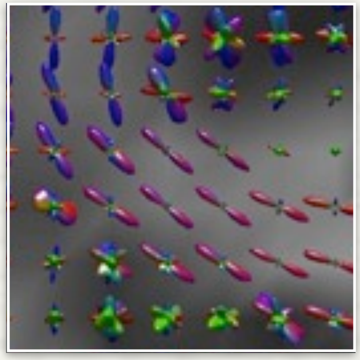




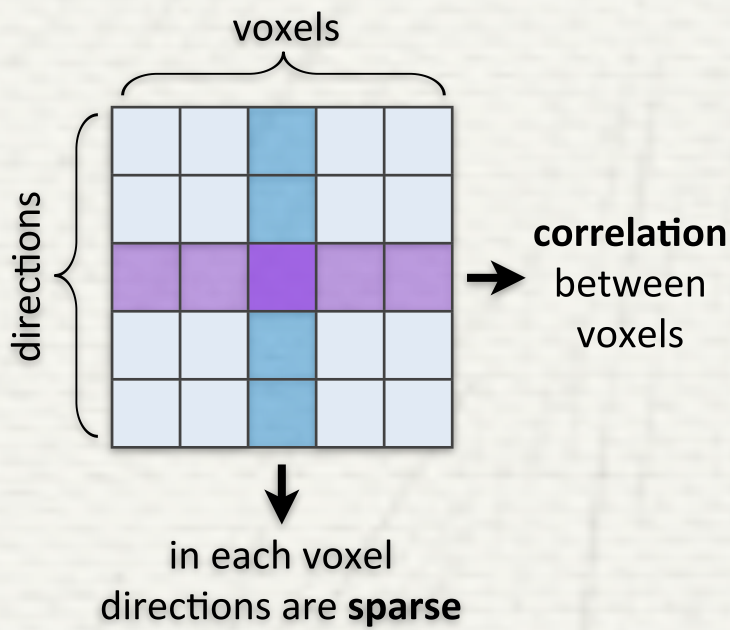
LR minimization: issues

- With this formulation we do not account for the concept of "**similar directions**"



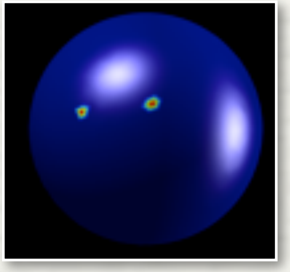


- "*Sparsity term*" and "*low-rank term*" **are in competition**!



Summary

(1) Diffusion MRI is sparse!



(2) **Re-weighted ℓ_1** improves the quality of reconstructions

- State-of-the-art: DSI = 257 samples
  L2L1 = 30 samples
  We achieve the same quality with **15 samples**

(3) Is **low-rank minimization** the key to go beyond?

- It is **very effective**, but only in very simplistic cases.
  In more realistic settings, it tends to over-simplify the structure
- Sparsity is "**competing"** with correlation
- Needs to be investigated further

Summary

(1) Diffusion MRI is sparse!

(2) **Re-weighted ℓ_1** improves the quality of reconstructions

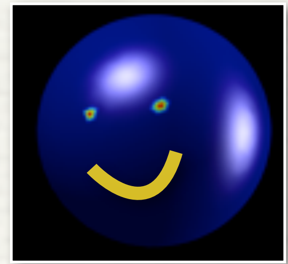
- State-of-the-art: DSI = 257 samples
  L2L1 = 30 samples
  We achieve the same quality with **15 samples**

(3) Is **low-rank minimization** the key to go beyond?

- It is **very effective**, but only in very simplistic cases.
  In more realistic settings, it tends to over-simplify the structure
- Sparsity is "**competing"** with correlation
- Needs to be investigated further

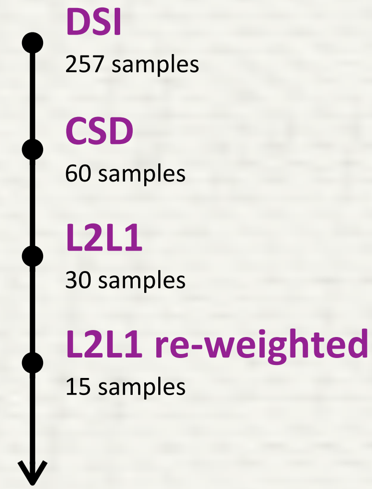
Summary

(1) Diffusion MRI is sparse!



(2) **Re-weighted ℓ_1** improves the quality of reconstructions

- State-of-the-art: DSI = 257 samples
  L2L1 = 30 samples
  We achieve the same quality with **15 samples**

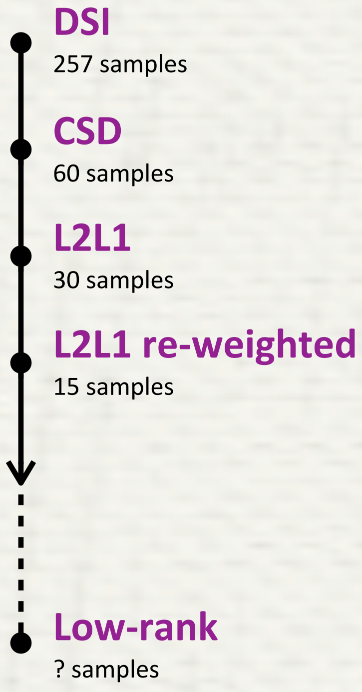


Summary

(1) Diffusion MRI is sparse!

(2) **Re-weighted ℓ_1** improves the quality of reconstructions

- State-of-the-art: DSI = 257 samples
  L2L1 = 30 samples
  We achieve the same quality with **15 samples**



(3) Is **low-rank minimization** the key to go beyond?

- It is **very effective**, but only in very simplistic cases.
  In more realistic settings, it tends to over-simplify the structure
- Sparsity is "**competing**" with correlation
- Needs to be investigated further

Lausanne "diffusion group"

Data reconstruction
Yves Wiaux and me

Fiber-tracking and structural connectivity
Alia Lemkaddem

Relationship between structural-functional networks
Alessandra Griffa

Statistical analysis of connectivity graphs
Djalel Meskaldji

Clinical studies
Elda Fischi-Gomez

Questions?

Comments?

Suggestions?