Ordinary Least Squares and its applications

Dr. Mauro Zucchelli

University Of Verona

December 5, 2016

Contents

Contents

OLS and Wavelet transform

- Every mother Wavelet $\psi(t)$ generate a basis which can be used to represent any function $f(t)$

$$
\begin{aligned}
\psi_{j,n}(t) &= \frac{1}{\sqrt{2^j}}\psi(\frac{r-2^j n}{2^j}) && (1)\\
f &= \sum_{j=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} c_{j,n}\psi_{j,n} && (2)\\
c_{j,n} &= \langle f, \psi_{j,n}\rangle = \int_{-\infty}^{\infty} f(t)\psi_{j,n}(t)dt && (3)
\end{aligned}
$$

OLS and Wavelet transform

There are several ways to calculate the Wavelet coefficients $c_{j,n}$

- Calculate the scalar product $\langle f, \psi_{j,n} \rangle$
- Utilizing the DWT

The coefficients can be also found using the Ordinary Least Squares (OLS) method

Ordinary Least Squares

Given a generic function $y(x)$ and a basis B we can represent y as

$$y(x_i) = \sum_{n=0}^{\infty} c_n B_n(x_i) \tag{4}$$

For practical reason this summation is often truncated to the order N

$$y(x_i) = \sum_{n=0}^{N} c_n B_n(x_i) \tag{5}$$

This problem has an equivalent matrix representation

Ordinary Least Squares

- $\mathbf{y}$ is a vector $s \times 1$ where s is the number of samples
- $\mathbf{c}$ is a vector $N \times 1$
- $\mathbf{B}$ is a matrix $s \times N$

$$\mathbf{B} = \begin{bmatrix} B_0(x_0) & B_1(x_0) & \dots & B_n(x_0) \\ B_0(x_1) & B_1(x_1) & \dots & B_n(x_1) \\ \vdots & \vdots & \ddots & \vdots \\ B_0(x_s) & B_1(x_s) & \dots & B_n(x_s) \end{bmatrix} \tag{6}$$

Ordinary Least Squares

The vector $\mathbf{y}$ can be calculated as

$$\mathbf{y} = \mathbf{B}\mathbf{c} \tag{7}$$

The goal of OLS is finding the vector $\mathbf{c}$ that minimize

$$\sum_{i=0}^{s}(y(x_i) - \sum_{n=0}^{N} c_n B_n(x_i))^2 \tag{8}$$

or, using the matrix notation, as

$$\arg\min_{\mathbf{c}} \|\mathbf{y} - \mathbf{B}\mathbf{c}\|^2 \tag{9}$$

Ordinary Least Squares

In order to find $\mathbf{c}$ it is necessary to perform some algebraic operations

$$\begin{aligned}
&\|\mathbf{y}-\mathbf{B}\mathbf{c}\|^2 = \\
&= (\mathbf{y}-\mathbf{B}\mathbf{c})^T(\mathbf{y}-\mathbf{B}\mathbf{c}) = \\
&= (\mathbf{y}^T-\mathbf{c}^T\mathbf{B}^T)(\mathbf{y}-\mathbf{B}\mathbf{c}) = \\
&= \mathbf{y}^T\mathbf{y}-\mathbf{y}^T\mathbf{B}\mathbf{c}-\mathbf{c}^T\mathbf{B}^T\mathbf{y}+\mathbf{c}^T\mathbf{B}^T\mathbf{B}\mathbf{c} = \\
&= \mathbf{y}^T\mathbf{y}-2\mathbf{c}^T\mathbf{B}^T\mathbf{y}+\mathbf{c}^T\mathbf{B}^T\mathbf{B}\mathbf{c}
\end{aligned} \tag{10}$$

Note: This is equivalent to the second order equation

$$y^2-2c(by)+c^2b^2 \tag{11}$$

Ordinary Least Squares

In order to find $\mathbf{c}$ it is necessary to perform some algebraic operations

$$\begin{aligned}
&\|\mathbf{y} - \mathbf{B}\mathbf{c}\|^2 = \\
&= (\mathbf{y} - \mathbf{B}\mathbf{c})^T(\mathbf{y} - \mathbf{B}\mathbf{c}) = \\
&= (\mathbf{y}^T - \mathbf{c}^T\mathbf{B}^T)(\mathbf{y} - \mathbf{B}\mathbf{c}) = \\
&= \mathbf{y}^T\mathbf{y} - \mathbf{y}^T\mathbf{B}\mathbf{c} - \mathbf{c}^T\mathbf{B}^T\mathbf{y} + \mathbf{c}^T\mathbf{B}^T\mathbf{B}\mathbf{c} = \\
&= \mathbf{y}^T\mathbf{y} - 2\mathbf{c}^T\mathbf{B}^T\mathbf{y} + \mathbf{c}^T\mathbf{B}^T\mathbf{B}\mathbf{c}
\end{aligned} \tag{10}$$

Note: This is equivalent to the second order equation

$$y^2 - 2c(by) + c^2b^2 \tag{11}$$

Ordinary Least Squares

- Since this equation is convex, it presents only one point in which the derivative is zero and this point is a minimum
- We can calculate the derivative by calculating the gradient $\nabla_{\mathbf{c}}$

$$
\begin{aligned}
&\nabla_{\mathbf{c}}(\mathbf{y}^T\mathbf{y} - 2\mathbf{c}^T\mathbf{B}^T\mathbf{y} + \mathbf{c}^T\mathbf{B}^T\mathbf{B}\mathbf{c}) = \\
&-2\mathbf{B}^T\mathbf{y} + 2\mathbf{B}^T\mathbf{B}\mathbf{c}
\end{aligned}
\tag{12}
$$

- And find the minimum

$$
\begin{aligned}
-2\mathbf{B}^T\mathbf{y} + 2\mathbf{B}^T\mathbf{B}\mathbf{c} &= 0 \\
\mathbf{B}^T\mathbf{B}\mathbf{c} &= \mathbf{B}^T\mathbf{y} \\
\mathbf{c} &= (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{y}
\end{aligned}
\tag{13}
$$

Ordinary Least Squares

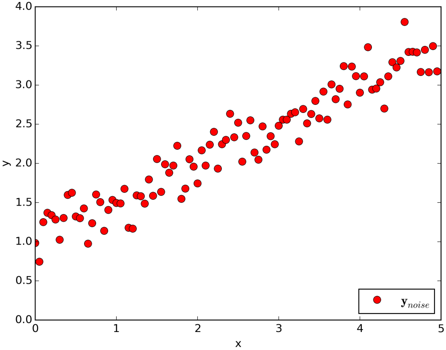
- OLS guarantees to provide the minimum error $\|\mathbf{y} - \mathbf{Bc}\|^2$
- This is true also when the observations $\mathbf{y}$ are corrupted with Gaussian noise

$$\mathbf{y}_{noise} = \mathcal{N}(\mathbf{y}, \sigma) \tag{14}$$

Contents

Case one: straight line

- Linear regression problem
- The data consist of s observations of $\mathbf{y} \in \mathcal{R}$, function of the independent variable $\mathbf{x}$
- The goal is to find the straight line $\mathbf{y} = r\mathbf{x} + q$ that better approximates the data



Case one: straight line

The problem can be also viewed as

$$\mathbf{y} = \mathbf{B}\mathbf{c} \tag{15}$$

where $\mathbf{c} = [r,\ q]^T$ and

$$\mathbf{B} = \begin{bmatrix} x_0 & 1 \\ \vdots & \vdots \\ x_s & 1 \end{bmatrix} \tag{16}$$

resulting in

$$\begin{bmatrix} y_0 \\ \vdots \\ y_s \end{bmatrix} = \begin{bmatrix} r \cdot x_0 + q \\ \vdots \\ r \cdot x_s + q \end{bmatrix} \tag{17}$$

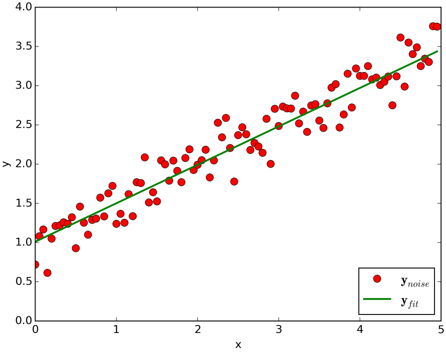
Case one: straight line

The problem can be solved using OLS

$$\mathbf{c} = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{y} \qquad (18)$$

The line coefficients r and q are respectively $c[0]$ and $c[1]$

$$\mathbf{y}_{fit} = c[0] * \mathbf{x} + c[1] \qquad (19)$$



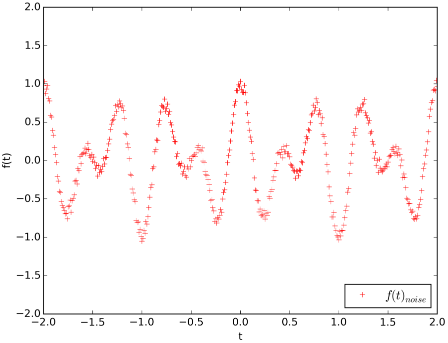
Contents

Case two: wavelet coefficients

Given a function $f(t)$ we want to represent it using a family of wavelets up to the orders J and N

$$f(t) = \sum_{j=-J}^{J} \sum_{n=-N}^{N} c_{j,n}\psi_{j,n}(t) \quad (20)$$

In this case the representations will not be perfect, because of the truncation of the infinite series



Case two: wavelet coefficients

We can view the problem as

$$\mathbf{f} = \Psi \mathbf{c} \tag{21}$$

where **c** are the wavelet coefficients and

$$\Psi = \begin{bmatrix} \psi_{-J,-N}(t_0) & \dots & \psi_{J,N}(t_0) \\ \vdots & \ddots & \vdots \\ \psi_{-J,-N}(t_s) & \dots & \psi_{J,N}(t_s) \end{bmatrix} \tag{22}$$

Case two: wavelet coefficients

The problem can be solved using OLS as

$$\mathbf{c} = (\Psi^T \Psi)^{-1} \Psi^T \mathbf{f} \tag{23}$$

Depending on the number of samples and coefficients, the square matrix $\Psi^T \Psi$ could not be invertible

In this case it is necessary to introduce some regularization, conditioning the diagonal of the matrix

$$\mathbf{c} = (\Psi^T \Psi + \lambda \mathbf{1})^{-1} \Psi^T \mathbf{f} \tag{24}$$

where $\mathbf{1}$ is the identity matrix of size $n_c \times n_c$ and λ is a small positive number

Case two: wavelet coefficients

Considering two families of wavelets

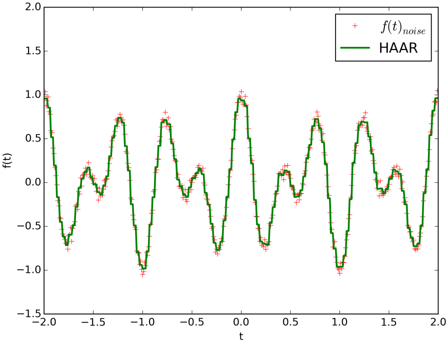
- Haar wavelet

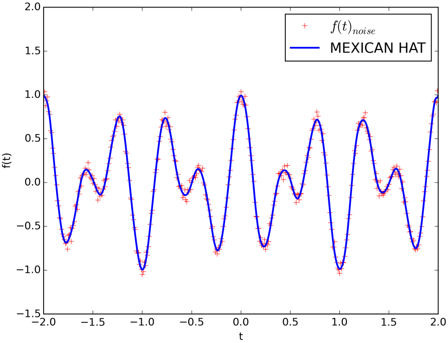
$$\psi(t) = \begin{cases} 0, & \text{if } t < 0 \vee t \geq 1 \\ 1, & \text{if } 0 \leq t < \frac{1}{2} \\ -1, & \text{if } \frac{1}{2} \leq t < 1 \end{cases} \tag{25}$$

- Mexican hat

$$\psi(t, \sigma) = \frac{2}{\sqrt{2 * \sigma}\pi^{\frac{1}{4}}} \left(1 - \frac{t^2}{\sigma^2}\right) \exp\left(-\frac{t^2}{2\sigma^2}\right) \tag{26}$$

Case two: wavelet coefficients

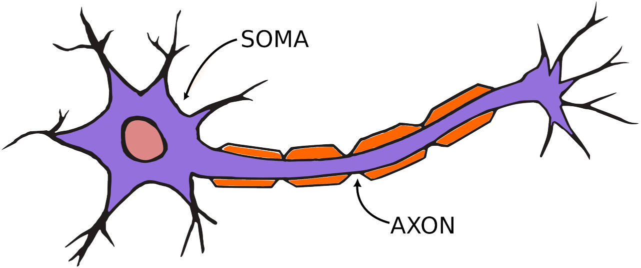




Contents

The neuron

- The most fundamental component of the nervous system is the neuron
- The neuron can be subdivided in two parts: the cellular body (soma) and the axon
- The axons connect the different neurons in the brain

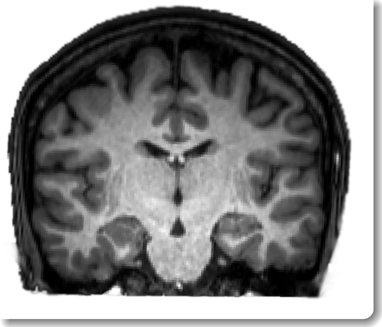


White Matter organization

- The neuron bodies are mainly clustered in the outer region of the brain, the brain cortex or Gray Matter
- The cortex is subdivided in functional regions (e.g. motor cortex)
- The axons are grouped in bundles that which connect the different region of the brain and form the White Matter

Magnetic Resonance Imaging

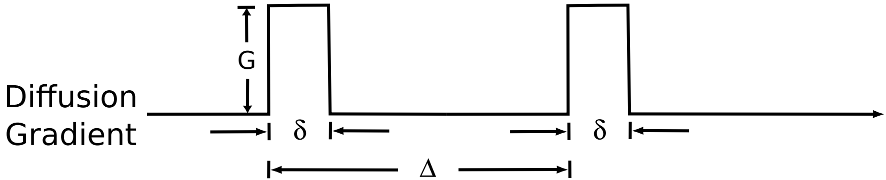
- Magnetic Resonance Imaging (MRI) is one of the only non-invasive technique which enables to characterize soft tissues *in-vivo*
- With traditional MRI techniques it is possible to obtain clear images of the principal brain tissues
- However with standard MRI resolution it is impossible to observe the white matter fibers which diameter is in the range of the micrometers



Diffusion Magnetic Resonance Imaging

- The goal of Diffusion MRI is to characterize the brain tissue microstructure by observing the water molecules diffusion profile
- By studying the principal direction of diffusion in each voxel it is possible to estimate the local orientation of the WM fibers
- Propagating the local information for all the voxels it is possible to reconstruct streamlines that generally follows the WM topography

Diffusion MRI: Pulse Gradient Spin Echo



- The diffusion signal is measured as a function of the b-value

$$\mathbf{b} = (\Delta - \delta/3)(\gamma\delta\mathbf{G})^2$$

- We can define $\mathbf{q} = \frac{\gamma\delta\mathbf{G}}{2\pi}$, the q-value
- And the effective diffusion time $\tau = \Delta - \delta/3$

$$\mathbf{b} = 4\pi^2\tau\mathbf{q}^2$$

Diffusion MRI

- By changing the b-value and the direction of the pulse it is possible to measure the diffusion dependent MR signal
- In DMRI the signal is attenuated more in the direction where the water molecules are more free to diffuse
- Higher b-values corresponds to higher signal attenuation $\Rightarrow$ more signal from low diffusion areas

Diffusion MRI

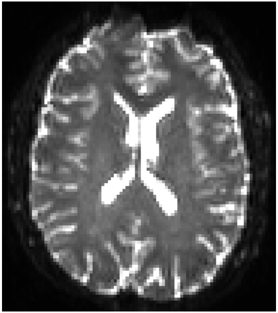
- By changing the b-value and the direction of the pulse it is possible to measure the diffusion dependent MR signal
- In DMRI the signal is attenuated more in the direction where the water molecules are more free to diffuse
- Higher b-values corresponds to higher signal attenuation $\Rightarrow$ more signal from low diffusion areas

Diffusion MRI

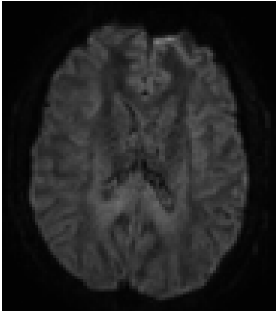
- By changing the b-value and the direction of the pulse it is possible to measure the diffusion dependent MR signal
- In DMRI the signal is attenuated more in the direction where the water molecules are more free to diffuse
- Higher b-values corresponds to higher signal attenuation $\Rightarrow$ more signal from low diffusion areas

Diffusion MRI

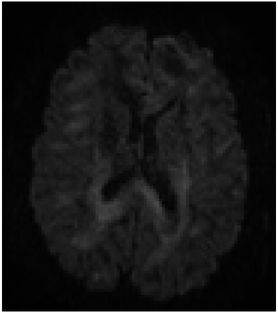
$b = 0s/mm^2$



$b = 1000s/mm^2$

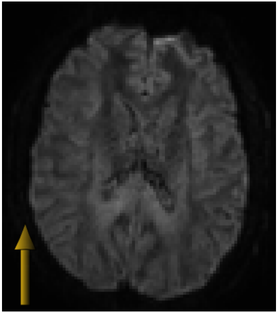


$b = 2000s/mm^2$

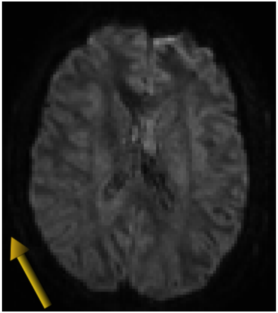


Diffusion MRI

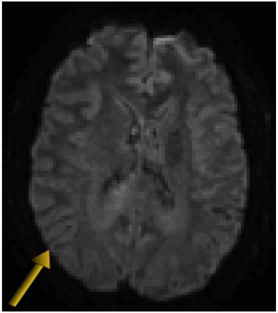
$b = 1000s/mm^2$



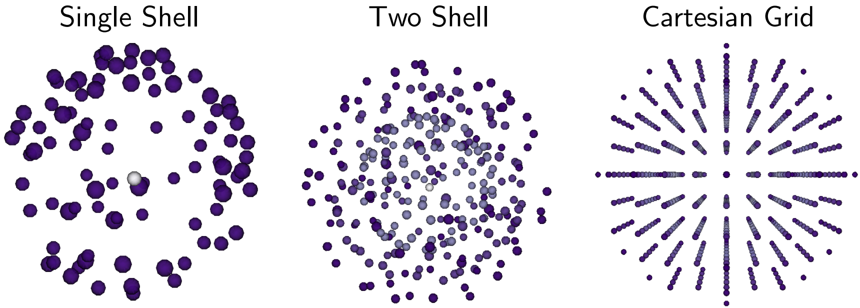
$b = 1000s/mm^2$



$b = 1000s/mm^2$



Diffusion MRI: sampling schemes



- A higher number of points is generally better in order to characterize the diffusion process
- There are physical limitations (time, gradient strength, noise)

Diffusion Ensemble Average Propagator

- The diffusion signal $E(\mathbf{q})$ is linked to the probability density function of the water molecules displacement $P(\mathbf{r})$

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

- This pdf is also called Ensemble Average Propagator (EAP)
- The EAP holds two important properties

$$P(\mathbf{r}) \geq 0 \ \forall \ \mathbf{r}$$

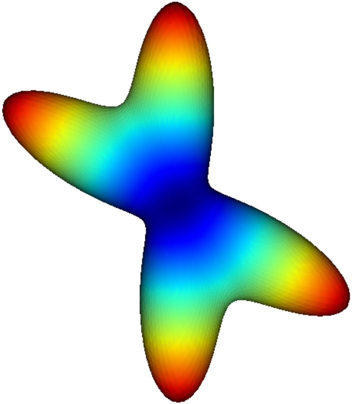
$$\int_{\mathbf{r}\in\mathbb{R}^3} P(\mathbf{r})d\mathbf{r} = 1$$

EAP-derived indices: ODF

- By averaging the radial part of the EAP it is possible to obtain the diffusion orientation profile of the water molecules displacement
- This is also known as Orientation Distribution Function (ODF)

$$ODF(\mathbf{u}) = \int_0^\infty P(r\mathbf{u}) r^2 dr$$

- Where $\mathbf{u}$ is a unit vector representing a direction



EAP reconstruction models

- In order to calculate the EAP from the diffusion signal it is necessary to calculate its Fourier transform
- If the signal is sampled in a Cartesian grid it is possible to use the Fast Fourier Transform (FFT)
- More generally, since the signal is generally sampled in a non-uniform manner, it is necessary to fit a mathematical model to the signal
- The mathematical models used in DMRI are called reconstruction models

EAP reconstruction models

- In order to calculate the EAP from the diffusion signal it is necessary to calculate its Fourier transform
- If the signal is sampled in a Cartesian grid it is possible to use the Fast Fourier Transform (FFT)
- More generally, since the signal is generally sampled in a non-uniform manner, it is necessary to fit a mathematical model to the signal
- The mathematical models used in DMRI are called reconstruction models

EAP reconstruction models

- In order to calculate the EAP from the diffusion signal it is necessary to calculate its Fourier transform
- If the signal is sampled in a Cartesian grid it is possible to use the Fast Fourier Transform (FFT)
- More generally, since the signal is generally sampled in a non-uniform manner, it is necessary to fit a mathematical model to the signal
- The mathematical models used in DMRI are called reconstruction models

EAP reconstruction models

- In order to calculate the EAP from the diffusion signal it is necessary to calculate its Fourier transform
- If the signal is sampled in a Cartesian grid it is possible to use the Fast Fourier Transform (FFT)
- More generally, since the signal is generally sampled in a non-uniform manner, it is necessary to fit a mathematical model to the signal
- The mathematical models used in DMRI are called reconstruction models

Contents

Diffusion Tensor Imaging

- The most widely used reconstruction model is the Diffusion Tensor Imaging (DTI)
- DTI models the diffusion signal as a multivariate Gaussian function

$$E(\mathbf{q}) = \exp(-4\pi^2\tau\mathbf{q}^T\mathbf{D}\mathbf{q})$$

- Where $\mathbf{D}$ is a 3×3 symmetric matrix
- The DTI EAP can be calculated as

$$P(\mathbf{r}) = \frac{1}{\sqrt{(4\pi\tau)^3|\mathbf{D}|}} \exp \frac{-\mathbf{r}^T\mathbf{D}^{-1}\mathbf{r}}{4\tau}$$

DTI fitting using OLS

In order to model the diffusion signal using DTI it is necessary to fit the diffusion tensor $\mathbf{D}$

$$\mathbf{D} = \begin{bmatrix} D_{x,x} & D_{x,y} & D_{x,z} \\ D_{x,y} & D_{y,y} & D_{y,z} \\ D_{x,z} & D_{y,z} & D_{z,z} \end{bmatrix} \tag{27}$$

The signal equation can be rewritten as

$$\begin{aligned} E(\mathbf{b}) &= \exp(-b\mathbf{u}^T\mathbf{D}\mathbf{u}) \\ -\frac{\ln(E(\mathbf{b}))}{b} &= \mathbf{u}^T\mathbf{D}\mathbf{u} \end{aligned} \tag{28}$$

DTI fitting using OLS

$$-\frac{\ln(E(\mathbf{b}))}{b} = \begin{bmatrix} u_x & u_y & u_z \end{bmatrix} \begin{bmatrix} D_{x,x} & D_{x,y} & D_{x,z} \\ D_{x,y} & D_{y,y} & D_{y,z} \\ D_{x,z} & D_{y,z} & D_{z,z} \end{bmatrix} \begin{bmatrix} u_x \\ u_y \\ u_z \end{bmatrix} \tag{29}$$

$$= u_x^2 D_{x,x} + 2u_x u_y D_{x,y} + 2u_x u_z D_{x,z} + \\ + 2u_y u_z D_{y,z} + u_y^2 D_{y,y} + u_z^2 D_{z,z}$$

DTI fitting using OLS

$$
\begin{aligned}
-\frac{\ln(E(\mathbf{b}))}{b} &= \begin{bmatrix} u_x & u_y & u_z \end{bmatrix} \begin{bmatrix} D_{x,x} & D_{x,y} & D_{x,z} \\ D_{x,y} & D_{y,y} & D_{y,z} \\ D_{x,z} & D_{y,z} & D_{z,z} \end{bmatrix} \begin{bmatrix} u_x \\ u_y \\ u_z \end{bmatrix} \\
&= u_x^2 D_{x,x} + 2u_x u_y D_{x,y} + 2u_x u_z D_{x,z} + \\
&+ 2u_y u_z D_{y,z} + u_y^2 D_{y,y} + u_z^2 D_{z,z}
\end{aligned} \tag{29}
$$

DTI fitting using OLS

We can recast the problem using OLS where the observations are $-\frac{\ln(E(\mathbf{b}_i))}{b_i}$ and the coefficients vector and the matrix basis are

$$\mathbf{c} = \begin{bmatrix} D_{x,x} \\ D_{x,y} \\ D_{x,z} \\ D_{y,z} \\ D_{y,y} \\ D_{z,z} \end{bmatrix} \tag{30}$$

$$\mathbf{M} = \begin{bmatrix} {u_0}_x^2 & 2{u_0}_x{u_0}_y & 2{u_0}_x{u_0}_z & 2{u_0}_y{u_0}_z & {u_0}_y^2 & {u_0}_z^2 \\ {u_1}_x^2 & 2{u_1}_x{u_1}_y & 2{u_1}_x{u_1}_z & 2{u_1}_y{u_1}_z & {u_1}_y^2 & {u_1}_z^2 \\ \vdots & & \ddots & & & \vdots \\ {u_s}_x^2 & 2{u_s}_x{u_s}_y & 2{u_s}_x{u_s}_z & 2{u_s}_y{u_s}_z & {u_s}_y^2 & {u_s}_z^2 \end{bmatrix} \tag{31}$$

DTI fitting using OLS

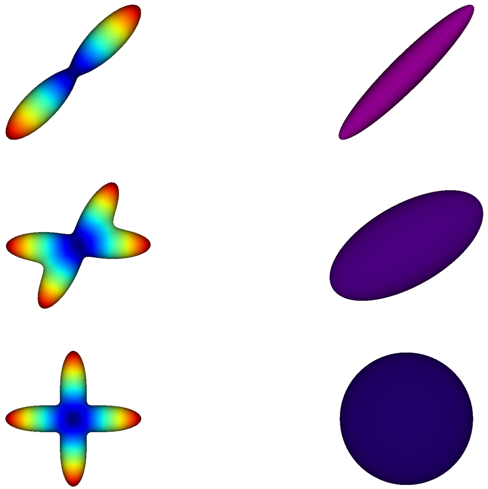
The coefficients can be retrieved using OLS as

$$\mathbf{c} = (\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T\left(-\frac{\ln(E(\mathbf{b}))}{b}\right) \tag{32}$$

- In theory only 6 samples are necessary to fit DTI coefficients
- In general, because of the noise, at least 30 samples are used in clincal practice
- OLS can be used for fitting the diffusion tensor only when all the sample are acquired using a single b-value

DTI limitations

Fiber configuration **DTI ellipsoid**



Contents

Spherical Harmonics

- The diffusion signal at a given *b-value* and the ODF are symmetric function on the sphere
- Spherical Harmonics (SH) are a useful basis for representing spherical functions
- SH can be viewed as the equivalent of the complex exponential $e^{-i\phi}$ on the sphere

Real Spherical Harmonics

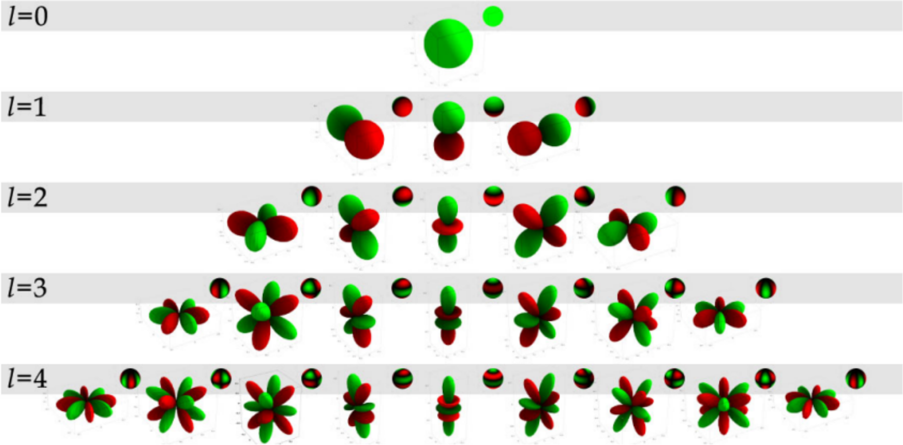
$$Y_l^m = \begin{cases} \sqrt{2} \cdot Re(\hat{Y}_l^m), & \text{if -l} \leq m < 0 \\ \hat{Y}_l^0, & \text{if } m\text{=}0 \\ \sqrt{2} \cdot Img(\hat{Y}_l^m), & \text{if } 0 < m \leq l \end{cases} \tag{33}$$

where $\hat{Y}_l^m$ is the normalized SH basis, written as

$$\hat{Y}_l^m(\theta, \phi) = \sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}} P_l^m(\cos\theta) e^{im\phi} \tag{34}$$

with θ, ϕ the polar representation of $\mathbf{u}$, and P_l^m the associated Legendre Polynomial

Real Spherical Harmonics



Real Spherical Harmonics

We can fit the diffusion signal at a given b-value and in a given direction $\mathbf{u}$ as

$$E(b,\mathbf{u}) = \sum_{l=0,\ even}^{N} \sum_{m=-l}^{l} c_{l,m} Y_{l,m}(\mathbf{u}) \tag{35}$$

where the coefficients $c_{l,m}$ are obtained as

$$\mathbf{c} = (\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{Y}^T E(b,\mathbf{u}) \tag{36}$$

with $\mathbf{Y}$ defined as

$$\mathbf{Y} = \begin{bmatrix} Y_{0,0}(\mathbf{u}_0) & Y_{2,-2}(\mathbf{u}_0) & \dots & Y_{N,N}(\mathbf{u}_0) \\ Y_{0,0}(\mathbf{u}_1) & Y_{2,-2}(\mathbf{u}_1) & \dots & Y_{N,N}(\mathbf{u}_1) \\ \vdots & \vdots & \ddots & \vdots \\ Y_{0,0}(\mathbf{u}_s) & Y_{2,-2}(\mathbf{u}_s) & \dots & Y_{N,N}(\mathbf{u}_s) \end{bmatrix} \tag{37}$$

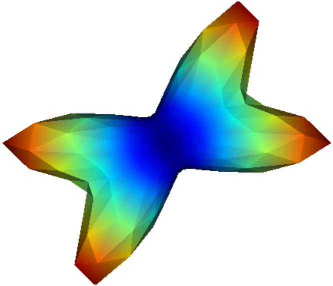
Spherical Harmonics as interpolation basis

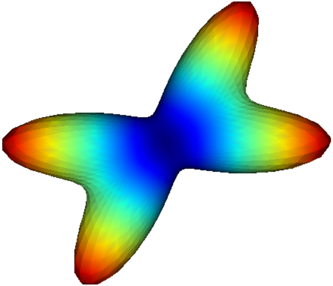
We can recover the diffusion signal in **u** as

$$E(b, \mathbf{u}) = \mathbf{Y}\mathbf{c} \tag{38}$$

The coefficients **c** described the signal in the full $\mathcal{S}^2$ space and not only in the set of points **u** used for the fitting

$$E(b, \hat{\mathbf{u}}) = \hat{\mathbf{Y}}\mathbf{c} \tag{39}$$





Fin

The End