Ordinary Least Squares and its applications

Dr. Mauro Zucchelli

University Of Verona

December 5, 2016
1. Introduction to Ordinary Least Squares
2. OLS examples: fitting of a straight line
3. OLS examples: fitting of Wavelet coefficients
4. Diffusion MRI
5. Diffusion Tensor Imaging
6. Real Spherical Harmonics
Contents

1 Introduction to Ordinary Least Squares

2 OLS examples: fitting of a straight line

3 OLS examples: fitting of Wavelet coefficients

4 Diffusion MRI

5 Diffusion Tensor Imaging

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

\[ \psi_{j,n}(t) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{r - 2^j n}{2^j}\right) \]  

\[ f = \sum_{j=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} c_{j,n} \psi_{j,n} \]  

\[ c_{j,n} = \langle f, \psi_{j,n} \rangle = \int_{-\infty}^{\infty} f(t) \psi_{j,n}(t) dt \]
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.
Given a generic function \( y(x) \) and a basis \( B \) we can represent \( y \) as

\[
y(x_i) = \sum_{n=0}^{\infty} c_n B_n(x_i)
\]  

(4)

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

\[
y(x_i) = \sum_{n=0}^{N} c_n B_n(x_i)
\]  

(5)

This problem has an equivalent matrix representation
Ordinary Least Squares

- \( y \) is a vector \( s \times 1 \) where \( s \) is the number of samples
- \( c \) is a vector \( N \times 1 \)
- \( B \) is a matrix \( s \times N \)

\[
B = \begin{bmatrix}
B_0(x_0) & B_1(x_0) & \ldots & B_n(x_0) \\
B_0(x_1) & B_1(x_1) & \ldots & B_n(x_1) \\
\vdots & \vdots & \ddots & \vdots \\
B_0(x_s) & B_1(x_s) & \ldots & B_n(x_s)
\end{bmatrix}
\] (6)
Ordinary Least Squares

The vector $y$ can be calculated as

$$y = Bc$$  \hspace{1cm} (7)

The goal of OLS is finding the vector $c$ that minimize

$$\sum_{i=0}^{s} (y(x_i) - \sum_{n=0}^{N} c_n B_n(x_i))^2$$  \hspace{1cm} (8)

or, using the matrix notation, as

$$\arg \min_c \| y - Bc \|^2$$  \hspace{1cm} (9)
Ordinary Least Squares

In order to find $c$ it is necessary to perform some algebraic operations

$$
\|y - Bc\|^2 =
= (y - Bc)^T (y - Bc) =
= (y^T - c^T B^T)(y - Bc) =
= y^T y - y^T Bc - c^T B^T y + c^T B^T Bc =
= y^T y - 2c^T B^T y + c^T B^T Bc
$$

(10)

Note: This is equivalent to the second order equation

$$
y^2 - 2c(by) + c^2 b^2
$$

(11)
Ordinary Least Squares

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

\[
\|\mathbf{y} - \mathbf{Bc}\|^2 = \\
= (\mathbf{y} - \mathbf{Bc})^T (\mathbf{y} - \mathbf{Bc}) = \\
= (\mathbf{y}^T - \mathbf{c}^T \mathbf{B}^T)(\mathbf{y} - \mathbf{Bc}) = \\
= \mathbf{y}^T \mathbf{y} - \mathbf{y}^T \mathbf{B} \mathbf{c} - \mathbf{c}^T \mathbf{B}^T \mathbf{y} + \mathbf{c}^T \mathbf{B}^T \mathbf{B} \mathbf{c} = \\
= \mathbf{y}^T \mathbf{y} - 2\mathbf{c}^T \mathbf{B}^T \mathbf{y} + \mathbf{c}^T \mathbf{B}^T \mathbf{B} \mathbf{c}
\]

Note: This is equivalent to the second order equation

\[
y^2 - 2c(by) + c^2b^2
\]
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 $\nabla_c$

$$\nabla_c(y^T y - 2c^T B^T y + c^T B^T B c) =$$

$$-2B^T y + 2B^T B c$$

And find the minimum

$$-2B^T y + 2B^T B c = 0$$

$$B^T B c = B^T y$$

$$c = (B^T B)^{-1} B^T y$$
Ordinary Least Squares

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

$$y_{\text{noise}} = \mathcal{N}(y, \sigma) \quad (14)$$
Contents

1 Introduction to Ordinary Least Squares
2 OLS examples: fitting of a straight line
3 OLS examples: fitting of Wavelet coefficients
4 Diffusion MRI
5 Diffusion Tensor Imaging
6 Real Spherical Harmonics
Case one: straight line

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

The problem can be also viewed as

\[ y = Bc \]  \hspace{1cm} (15)

where \( c = [r, q]^T \) and

\[ B = \begin{bmatrix} x_0 & 1 \\ \vdots & \vdots \\ x_s & 1 \end{bmatrix} \]  \hspace{1cm} (16)

resulting in

\[ \begin{bmatrix} y_0 \\ \vdots \\ y_s \end{bmatrix} = \begin{bmatrix} r \cdot x_0 + q \\ \vdots \\ r \cdot x_s + q \end{bmatrix} \]  \hspace{1cm} (17)
Case one: straight line

The problem can be solved using OLS

\[ c = (B^T B)^{-1} B^T y \]  (18)

The line coefficients \( r \) and \( q \) are respectively \( c[0] \) and \( c[1] \)

\[ y_{fit} = c[0] \times x + c[1] \]  (19)
Contents

1. Introduction to Ordinary Least Squares
2. OLS examples: fitting of a straight line
3. OLS examples: fitting of Wavelet coefficients
4. Diffusion MRI
5. Diffusion Tensor Imaging
6. Real Spherical Harmonics
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

\[ f = \Psi c \] (21)

where \( c \) are the wavelet coefficients and

\[ \Psi = \begin{bmatrix} \psi_{-J,-N}(t_0) & \ldots & \psi_{J,N}(t_0) \\ \vdots & \ddots & \vdots \\ \psi_{-J,-N}(t_s) & \ldots & \psi_{J,N}(t_s) \end{bmatrix} \] (22)
Case two: wavelet coefficients

The problem can be solved using OLS as

\[ c = (\Psi^T \Psi)^{-1} \Psi^T f \]  \hspace{1cm} (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

\[ c = (\Psi^T \Psi + \lambda \mathbf{1})^{-1} \Psi^T f \]  \hspace{1cm} (24)

where \( \mathbf{1} \) is the identity matrix of size \( n_c \times n_c \) and \( \lambda \) is a small positive number.
Case two: wavelet coefficients

Considering two families of wavelets

- Haar wavelet

\[
\psi(t) = \begin{cases} 
0, & \text{if } t < 0 \lor t \geq 1 \\
1, & \text{if } 0 \leq t < \frac{1}{2} \\
-1, & \text{if } \frac{1}{2} \leq t < 1
\end{cases}
\]  \quad (25)

- Mexican hat

\[
\psi(t, \sigma) = \frac{2}{\sqrt{2*\pi^{\frac{1}{4}}}} \left(1 - \frac{t^2}{\sigma^2}\right) \exp \left(-\frac{t^2}{2\sigma^2}\right)
\]  \quad (26)
Case two: wavelet coefficients
Contents

1 Introduction to Ordinary Least Squares

2 OLS examples: fitting of a straight line

3 OLS examples: fitting of Wavelet coefficients

4 Diffusion MRI

5 Diffusion Tensor Imaging

6 Real Spherical Harmonics
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 connect the different region of the brain and form the White Matter.
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.
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.
The diffusion signal is measured as a function of the \( b \)-value

\[
b = (\Delta - \delta/3)(\gamma\delta G)^2
\]

We can define \( q = \frac{\gamma\delta G}{2\pi} \), the \( q \)-value

And the effective diffusion time \( \tau = \Delta - \delta/3 \)

\[
b = 4\pi^2 \tau q^2
\]
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.
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 ⇒ more signal from low diffusion areas.
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 ⇒ more signal from low diffusion areas.
Diffusion MRI

\[ b = 0 \text{s/mm}^2 \]

\[ b = 1000 \text{s/mm}^2 \]

\[ b = 2000 \text{s/mm}^2 \]
Diffusion MRI

\[ b = 1000 \text{s}/\text{mm}^2 \]
A higher number of points is generally better in order to characterize the diffusion process.

There are physical limitations (time, gradient strength, noise).
The diffusion signal $E(q)$ is linked to the probability density function of the water molecules displacement $P(r)$

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

This pdf is also called Ensemble Average Propagator (EAP)

The EAP holds two important properties

$$P(r) \geq 0 \quad \forall \ r$$

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

\[ ODF(u) = \int_0^\infty P(ru) r^2 dr \]

- Where \( u \) is a unit vector representing a direction.
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**
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.
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

1 Introduction to Ordinary Least Squares
2 OLS examples: fitting of a straight line
3 OLS examples: fitting of Wavelet coefficients
4 Diffusion MRI
5 Diffusion Tensor Imaging
6 Real Spherical Harmonics
The most widely used reconstruction model is the Diffusion Tensor Imaging (DTI) model. DTI models the diffusion signal as a multivariate Gaussian function:

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

Where $D$ is a $3 \times 3$ symmetric matrix and the DTI EAP can be calculated as:

$$P(r) = \frac{1}{\sqrt{(4\pi\tau)^3|D|}} \exp \left( \frac{-r^T D^{-1} r}{4\tau} \right)$$
DTI fitting using OLS

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

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

The signal equation can be rewritten as

$$E(b) = \exp(-bu^TDu)$$

$$- \frac{\ln(E(b))}{b} = u^TDu$$
DTI fitting using OLS

\[- \frac{\ln(E(b))}{b} = \begin{bmatrix} u_x & u_y & u_z \end{bmatrix} \begin{bmatrix} D_{x,x} & D_{x,y} & D_{x,z} \\ D_{x,y} & D_{y,y} & D_{y,z} \\ D_{x,z} & D_{y,z} & D_{z,z} \end{bmatrix} \begin{bmatrix} u_x \\ u_y \\ u_z \end{bmatrix} = u_x^2 D_{x,x} + 2u_x u_y D_{x,y} + 2u_x u_z D_{x,z} + 2u_y u_z D_{y,z} + u_y^2 D_{y,y} + u_z^2 D_{z,z} \]
DTI fitting using OLS

\[ -\frac{\ln(E(b))}{b} = \begin{bmatrix} u_x & u_y & u_z \end{bmatrix} \begin{bmatrix} D_{x,x} & D_{x,y} & D_{x,z} \\ D_{x,y} & D_{y,y} & D_{y,z} \\ D_{x,z} & D_{y,z} & D_{z,z} \end{bmatrix} \begin{bmatrix} u_x \\ u_y \\ u_z \end{bmatrix} \]

\[ = u_x^2 D_{x,x} + 2u_x u_y D_{x,y} + 2u_x u_z D_{x,z} + \\
+ 2u_y u_z D_{y,z} + u_y^2 D_{y,y} + u_z^2 D_{z,z} \]
DTI fitting using OLS

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

\[
\mathbf{c} = \begin{bmatrix}
D_{x,x} \\
D_{x,y} \\
D_{x,z} \\
D_{y,x} \\
D_{y,y} \\
D_{y,z} \\
D_{z,x} \\
D_{z,y} \\
D_{z,z}
\end{bmatrix}
\]

(30)

\[
\mathbf{M} = \begin{bmatrix}
u_0^2 & 2u_0xu_0y & 2u_0xu_0z & 2u_0yu_0z & u_0^2 & u_0^2 \\
u_1^2 & 2u_1xu_1y & 2u_1xu_1z & 2u_1yu_1z & u_1^2 & u_1^2 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
u_s^2 & 2u_su sy & 2u_su sz & 2u syu sz & u_s^2 & u_s^2
\end{bmatrix}
\]

(31)
DTI fitting using OLS

The coefficients can be retrieved using OLS as

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

(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 clinical 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

1 Introduction to Ordinary Least Squares

2 OLS examples: fitting of a straight line

3 OLS examples: fitting of Wavelet coefficients

4 Diffusion MRI

5 Diffusion Tensor Imaging

6 Real Spherical Harmonics
Spherical Harmonics

- The diffusion signal at a given \textit{b-value} and the ODF are symmetric function on the sphere
- \textbf{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

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

(33)

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

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

(34)

with \( \theta, \phi \) the polar representation of \( \mathbf{u} \), and \( P_l^m \) the associated Legendre Polynomial.
Real Spherical Harmonics

$l=0$

$l=1$

$l=2$

$l=3$

$l=4$
Real Spherical Harmonics

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

\[
E(b, \mathbf{u}) = \sum_{l=0}^{N} \sum_{\text{even } m=-l}^{l} c_{l,m} Y_{l,m}(\mathbf{u})
\]  

(35)

where the coefficients \( c_{l,m} \) are obtained as

\[
c = (\mathbf{Y}^T \mathbf{Y})^{-1} \mathbf{Y}^T E(b, \mathbf{u})
\]  

(36)

with \( \mathbf{Y} \) defined as

\[
\mathbf{Y} = 
\begin{bmatrix}
Y_{0,0}(\mathbf{u}_0) & Y_{2,-2}(\mathbf{u}_0) & \cdots & Y_{N,N}(\mathbf{u}_0) \\
Y_{0,0}(\mathbf{u}_1) & Y_{2,-2}(\mathbf{u}_1) & \cdots & Y_{N,N}(\mathbf{u}_1) \\
\vdots & \vdots & \ddots & \vdots \\
Y_{0,0}(\mathbf{u}_s) & Y_{2,-2}(\mathbf{u}_s) & \cdots & Y_{N,N}(\mathbf{u}_s) \\
\end{bmatrix}
\]  

(37)
Spherical Harmonics as interpolation basis

We can recover the diffusion signal in \( u \) as

\[
E(b, u) = Yc
\]  \( (38) \)

The coefficients \( c \) described the signal in the full \( S^2 \) space and not only in the set of points \( u \) used for the fitting

\[
E(b, \hat{u}) = \hat{Y}c
\]  \( (39) \)
The End