On DTI models

Alexandre Savio

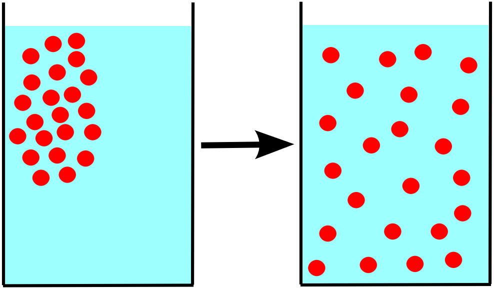
Grupo de Inteligencia Computacional, UPV/EHU

GIC, 2012

EUSKO JAURLARITZA
GOBIERNO VASCO

Outline

What is diffusion? I



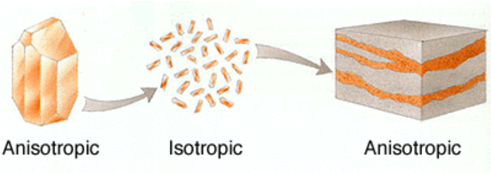
Wikimedia

Random motion of particles due to thermal energy

What is diffusion? II

- Water molecules collide and move (net displacement)
- **Displacement** described by diffusion coefficient (D)

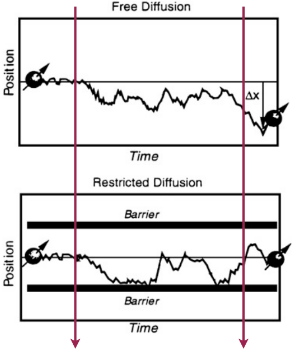
What is diffusion? III



Wikimedia

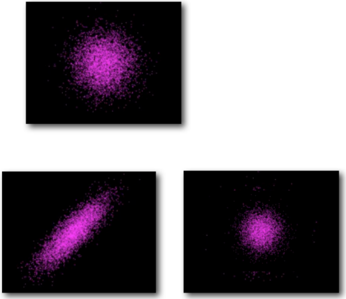
Normally, diffusion is **isotropic**

Apparent Diffusion



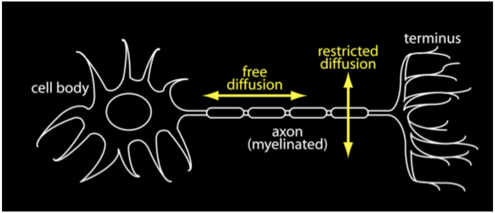
"Looks" like free diffusion

"Doesn't look" like free diffusion



Diffusion in WM is anisotropic

- Diffusion is restricted by tissue boundaries, membranes, etc.
- Directionality of diffusion tells us about fibre integrity/structure and orientation



Karla Miller

Fick's first law of diffusion

- The physical law that explains this phenomenon is called **Fick's first law** (Fick, 1855a, b), which relates the diffusive flux to any concentration difference through the relationship

$$\frac{\partial \phi}{\partial t} = D \frac{\partial^2 \phi}{\partial x^2}$$

- For more than 1 dimension:

$$\mathbf{J} = -D \nabla \phi \tag{1}$$

where $\mathbf{J}$ is the net particle flux (vector), ϕ is the particle concentration, and the constant of proportionality, D, is called the **"diffusion coefficient"**.

- Considering the **fluid density** ρ:

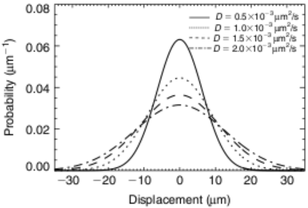
$$\mathbf{J} = -\rho D \nabla \phi \tag{2}$$

D for H_2O

- The diffusion coefficient for **water** at $37^o C$ is approximately [11]:

$$D = 3 \cdot 10^{-9} m^2/sec.$$

Diffusion displacement distribution



Diffusion MRI Book [14]

- The Gaussian **displacement distribution** plotted for various values of the diffusion coefficient when the diffusion time was taken to be 40 ms.
- Larger diffusion coefficients lead to broader displacement probabilities suggesting increased diffusional mobility.

Einstein described the motion

- Using the displacement distribution concept...
- Explicit relationship between the mean-squared displacement of the ensemble of molecules, characterizing its Brownian motion, and the classical diffusion coefficient, D, appearing in Fick's law (Einstein, 1905, 1926).
- For an ensemble of molecules in n-dimensional space:

$$\left\langle x^2 \right\rangle = 2nD\Delta \tag{3}$$

where $\left\langle x^2 \right\rangle$ is the mean-squared displacement of particles during a diffusion time, Δ, and D is the same classical diffusion coefficient appearing in Fick's first 1.

Diffusion as a Gaussian

- For a barrier-free medium, diffusion can be described by a **Normal distribution** with $N(0, 2tD)$, i.e.
    - Zero-mean displacement
    - **Variance proportional to time and the diffusion coefficient**

History of the Brownian motion and MRI I

- **Robert Brown** (1827): the first one to discover random motions of pollen grains while studying them through microscope.
- **Adolf Eugen Fick** (1855)**:** Fick managed to double-publish his law of diffusion, as it applied equally to physiology and physics. His work led to the development of the direct Fick method for measuring cardiac output.
- **Einstein** (1910): coherent description of diffusion emerged, identifying the diffusion coefficient in Fick's law and the variance of the particle displacement distribution that describes the migration of particles.Wikipedia

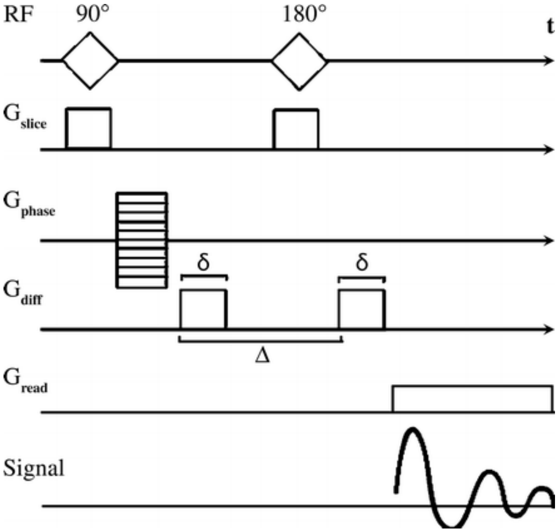
History of the Brownian motion and MRI II

- **Nuclear magnetic resonance** was first described and measured in molecular beams by Isidor Rabi in 1938, and in 1944, Rabi was awarded the Nobel Prize in physics for this work.
- In 1946, Felix Bloch and Edward Mills Purcell expanded the technique for use on liquids and solids, for which they shared the Nobel Prize in Physics in 1952.
- Purcell had worked on the development of radar during World War II at the Massachusetts Institute of Technology's Radiation Laboratory.
- His work during that project on the production and detection of radio frequency power and on the absorption of such RF power by matter laid the background for Rabi's discovery of NMR.Wikipedia

History of the Brownian motion and MRI III

- **Erwin Hahn** (1950): the spin echoes were sensitive to the effects of diffusion. He reported the reduction of signal of the spin echo and explained it in terms of the dephasing caused by translational diffusion of spins subjected to local magnetic field gradients due to inhomogenities in the magnetic field.
- **Carr and Purcell** (1952): showed that NMR spin echoes could be sensitized to diffusion in a way that permits its direct measurement.
- **Stejskal and Tanner** (1965): introduced many innovations that made modern diffusion measurements by NMR and MRI possible. Wikipedia

Pulsed field gradient spin-echo MR



Hagmann et al.

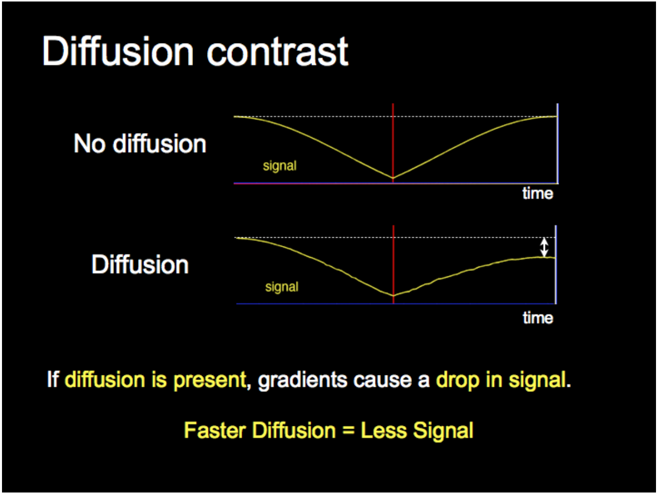
The b-value

- A diffusion gradient can be represented as a 3D vector q whose orientation is in the direction of diffusion and whose length is proportional to the gradient strength.
- The gradient strength, or more often the diffusion weighting, is sometimes expressed in terms of the b value, which is proportional to the product of the square of the gradient strength q and the diffusion time interval: [11]

$$b \sim q^2 \cdot \Delta$$

$$S(b) \propto e^{-{}^{TE}/{}_{T2}} \times e^{-bD}$$

The diffusion MR signal attenuation



Diffusion contrast

If **diffusion is present**, gradients cause a **drop in signal**.

Faster Diffusion = Less Signal

Content of DTI data

1. 4D MR image, let it be M DWI volumes.
2. **bvals**: $1xM$ vector with the value of b for each acquisition
3. **bvecs**: $3xM$ matrix where each column is a 3D unit vector that represents the direction in subject space of the diffusion gradient. $(0,0,0)$ if $b=0$

Acquisition parameters optimization

- There are many decisions to be made when choosing parameters for a DWI imaging experiment:
    - Number of **directions** of diffusion weighting.
    - Number of **b-values** for diffusion gradients.
    - Echo train (**measurement time**) duration
    - Tradeoffs with **resolution**
    - Flair DW EPI for elimination of **CSF contamination**

The Stejskal-Tanner equation

- The Stejskal-Tanner equation for a homogeneous medium

$$\mu = S_0 * exp(-b * D)$$

  - S_0 is the signal with no diffusion gradients applied
  - D is the diffusion coefficient
- The b-value units are actually $^{sec}/_{mm^2}$, this ensures that the measured signal attenuation is unitless.

The diffusion tensor model I

- The assumption made in the diffusion tensor model [4] is that local diffusion within a voxel may be characterized with a 3D Gaussian distribution, whose covariance matrix is proportional to the diffusion tensor, $\mathbf{D}$.

The diffusion tensor model II

- The resulting diffusion-weighted signal, μ_i along a gradient direction $\mathbf{r}_i$, with b-value b_i is modeled as:

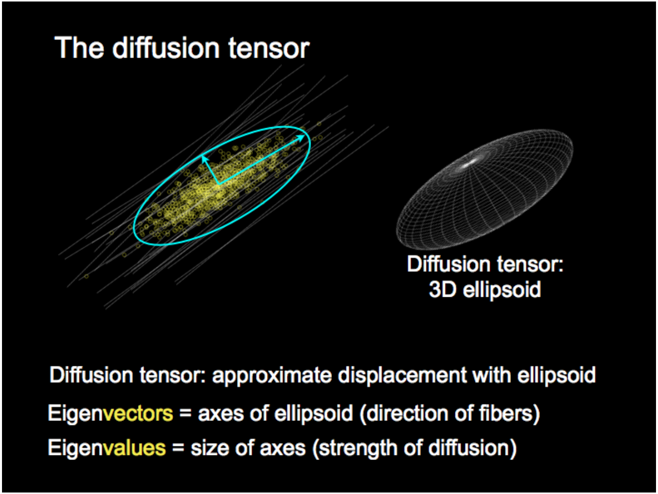
$$\mu_i = S_0 \exp\left(-b_i \mathbf{r}_i^T \mathbf{D} \mathbf{r}_i\right) \tag{4}$$

where S_0 is the signal with no diffusion gradients applied. $\mathbf{D}$, the diffusion tensor is:

$$\mathbf{D} = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{bmatrix}. \tag{5}$$

- In DTI, a diffusion tensor at a voxel is a 3×3 positive-definite symmetric matrix D.

The diffusion tensor model III



The diffusion tensor

Diffusion tensor:
3D ellipsoid

Diffusion tensor: approximate displacement with ellipsoid

Eigenvectors = axes of ellipsoid (direction of fibers)

Eigenvalues = size of axes (strength of diffusion)

Karla Miller

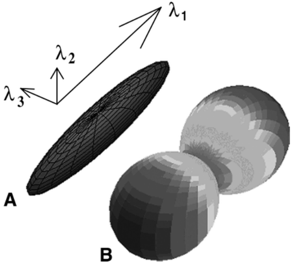
The diffusion tensor model IV

- It can be represented by its decomposition as:

$$D = \lambda_1 g_1 g_1^T + \lambda_2 g_2 g_2^T + \lambda_3 g_3 g_3^T,$$

where $\lambda_1 \geq \lambda_2 \geq \lambda_3$ and g_1, g_2, g_3 are the eigenvalues and eigenvectors of D

The diffusion tensor model V

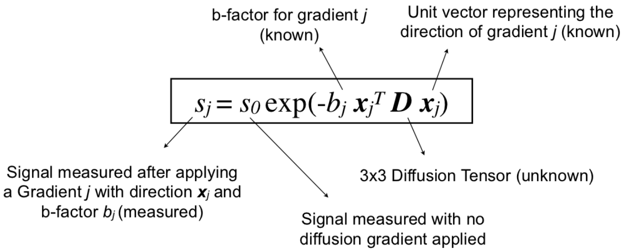


Hagmann et al.

- In A, the diffusion tensor is shown as an ellipsoid (an isosurface) with its principal axes along the eigenvectors (λ_1, λ_2, λ_3).
- In B, the diffusion tensor is shown as an orientation distribution function.

Calculating the tensor I

Diffusion Tensor Model. In each voxel:



$$\boxed{s_j = s_0 \exp(-b_j \, \boldsymbol{x}_j^T \boldsymbol{D} \, \boldsymbol{x}_j)}$$

b-factor for gradient *j* (known)

Unit vector representing the direction of gradient *j* (known)

Signal measured after applying a Gradient *j* with direction $\boldsymbol{x}_j$ and b-factor b_j (measured)

Signal measured with no diffusion gradient applied

3x3 Diffusion Tensor (unknown)

Calculating the tensor II

$$\mu_i = S_0 \exp\left(-b_i \mathbf{r}_i^T \mathbf{D} \mathbf{r}_i\right) \tag{6}$$

by taking log on both sides and expanding.
let it be $m$measures:

$$\Downarrow$$

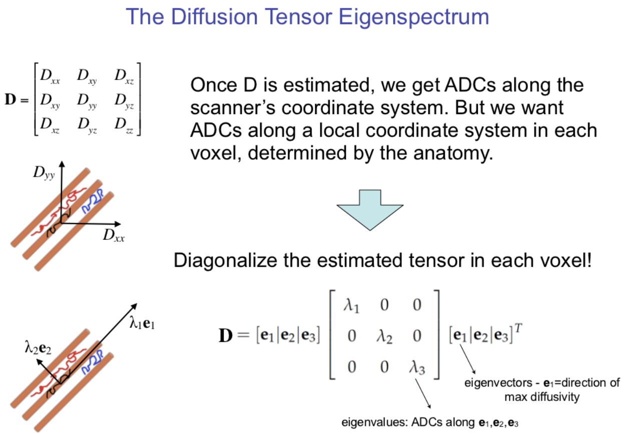
$$\begin{pmatrix} -b_1 r_{11} r_{11} & -2b_1 r_{11} r_{m12} & -b_1 r_{m1} r_{m1} & 1 \\ -b_2 r_{21} r_{21} & -2b_2 r_{21} r_{22} & & 1 \\ \cdots & \cdots & & \cdots \\ -b_m r_{m1} r_{m1} & -2b_m r_{m1} r_{m2} & & 1 \end{pmatrix} \begin{pmatrix} D_{11} \\ D_{12} \\ D_{22} \\ \log(S_0) \end{pmatrix} = \begin{pmatrix} \log(\mu_1) \\ \log(\mu_2) \\ \cdots \\ \log(\mu_m) \end{pmatrix}$$

linear equation

$$\Downarrow$$

$$\mathbf{Mx} = \mathbf{y} \Rightarrow \hat{\mathbf{x}} = \left(\mathbf{M}^T \mathbf{M}\right)^{-1} \mathbf{M}^T \mathbf{y}$$

Then its eigenvectors



The Diffusion Tensor Eigenspectrum

$$\mathbf{D} = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{bmatrix}$$

Once D is estimated, we get ADCs along the scanner's coordinate system. But we want ADCs along a local coordinate system in each voxel, determined by the anatomy.

Diagonalize the estimated tensor in each voxel!

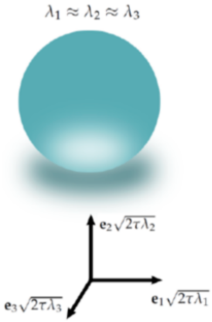
$$\mathbf{D} = [\mathbf{e}_1|\mathbf{e}_2|\mathbf{e}_3] \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} [\mathbf{e}_1|\mathbf{e}_2|\mathbf{e}_3]^T$$

eigenvectors - $\mathbf{e}_1$=direction of max diffusivity

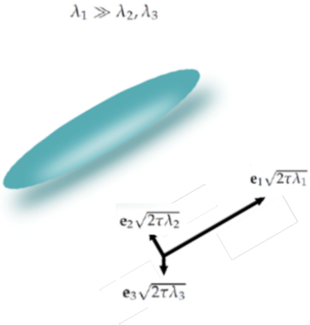
eigenvalues: ADCs along $\mathbf{e}_1$,$\mathbf{e}_2$,$\mathbf{e}_3$

Iso vs. Aniso I

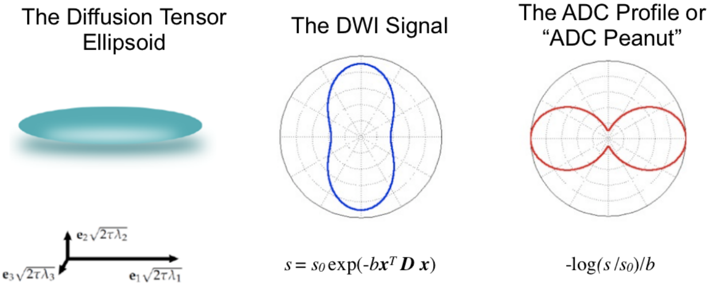
Isotropic voxel



Anisotropic voxel



Representing diffusion



Useful quantities from the diffusion tensor I

- Fractional anisotropy:

$$FA = \sqrt{\frac{1}{2}} \frac{\sqrt{(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}}.$$

The FA measures the fraction of the magnitude of D that can be related to anisotropic diffusion in a mean-squared sense (i.e. the extent of deviation from isotropic diffusivity in all direction).
Its magnitude is also rotationally invariant, and independent from sorting of the eigenvalues.

Useful quantities from the diffusion tensor II

- Mean diffusivity:

$$MD = \frac{\mathrm{Tr}\,(D)}{3} = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3}.$$

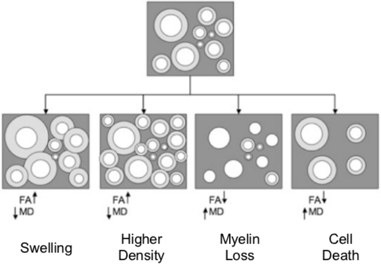
- Principal diffusion direction:
  - What direction is the greatest diffusion along? Info about principal fibre orientation.

Alright, now I can measure fibre integrity!!

www.wordle.net

Do not overinterpret!

But do not overinterpret your results! Always keep in mind that the DTI model is an oversimplification of reality! Different configurations can have same effect on FA, MD!



Stamatios Sotiropoulos

- The same can be said about any other existing brain diffusion model.

DTI data for the tensor model

- As the tensor model has 7 unknown parameters (S_0 and D_{ij}) it is needed to acquire at least:
  - 1 volume without gradient weighting. $b = 0$ (T2 weighted only)
  - 6 gradient weighted volumes
  - better if directions equally distributed in the sphere

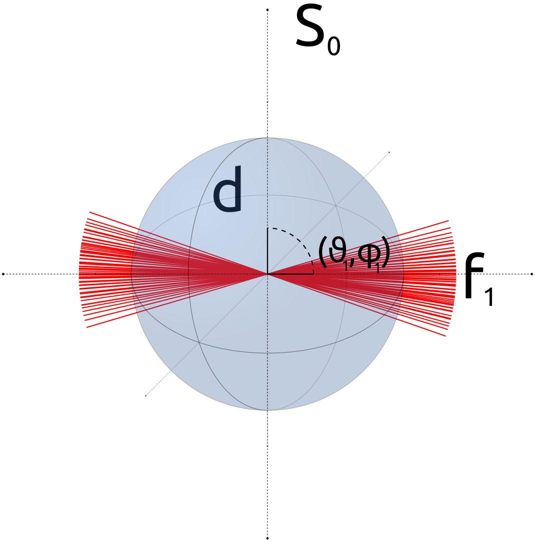
The multi-tensor model I

- Using a multi-tensor model it is possible to estimate more than 1 “fibre” direction.
- Utilize constraints to make the model identifiable, although it is an ill-posed problem.

Outline

1. Introduction
2. The diffusion tensor model
3. The simple partial volume (PVM) model
    - The model
    - Initializing and model fitting
    - The sampling method
    - The improvements
    - Noise estimation
    - Model selection
4. Summary and questions

The PVM model I



The PVM model II

- Imposes geometrical constraints on the different compartments:
    - The ball is a perfectly isotropic tensor
    - The stick (from which orientations are estimated) is infinitely anisotropic.
    - f is the volume fraction that represent how much of the the signal is explained by the corresponding stick.
- a.k.a. ball & stick model

The PVM model equation I

$$\mu_i = S_0 \left((1 - f) \cdot e^{-b_i d} + f \cdot e^{-b_i d \mathbf{r}_i^T \mathbf{RAR}^T \mathbf{r}_i} \right) \tag{7}$$

- S_0 is the non-diffusion weighted signal value,
- d is the diffusivity,
- b_i and $\mathbf{r}_i$ are the b-value and gradient direction associated with the i^{th} acquisition,
- f and $\mathbf{RAR}^T$ are the fraction of signal contributed by, and anisotropic diffusion tensor along, the fiber orientation (θ, ϕ).
- That is, $\mathbf{A}$ is fixed as:

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

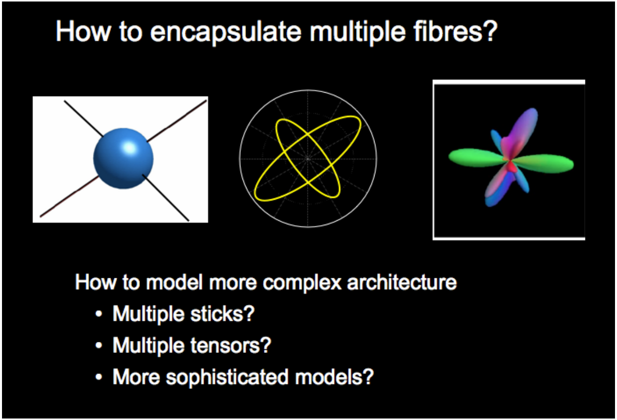
and $\mathbf{R}$ rotates $\mathbf{A}$ to (θ, ϕ).

The PVM model equation II

- Noise is modeled as Gaussian, where the parameter set ω now has five free parameters:

$$\omega = (S_o, d, f, \theta, \phi).$$

More than one fibre? I

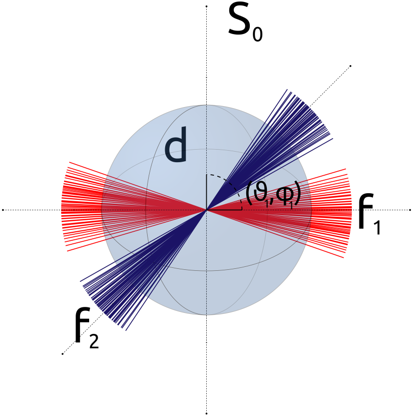


How to encapsulate multiple fibres?

How to model more complex architecture

- Multiple sticks?
- Multiple tensors?
- More sophisticated models?

More than one fibre? II



$$\mu_i = S_0 \left(\left(1 - \sum_{j=1}^{N} f_j \right) \cdot e^{-b_i d} + \sum_{j=1}^{N} f_j \cdot e^{-b_i d \mathbf{r}_i^T \mathbf{R_j} \mathbf{A} \mathbf{R_j}^T \mathbf{r}_i} \right), \qquad (8)$$

More than one fibre? III

- f_j and $\mathbf{R_j A R_j}^T$ are the fraction of signal contributed by, and anisotropic diffusion tensor along, the jth fiber orientation(θ_j, ϕ_j).
- That is, **A** is fixed as:

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and $\mathbf{R_j}$ rotates **A** to (θ_j, ϕ_j).

How many fibres should the model have?

- The number of fibres is an initial input parameter of the model.
- The number of fibres should be supported by the data.
- Grossly: *1, 2 or 3 depending on how many directions your data has*.
- The model has at least one fibre.
- The following fibres can be determined using Automatic Relevance Determination (ARD).

Automatic Relevance Determination (ARD)

- ARD is a known technique used in Bayesian modelling from the neural network literature.
- The benefit of ARD is that any unnecessary parameters are automatically forced to zero.
- ARD requires the use of a certain type of prior on a parameter whose relevance needs to be determined.
- In FSL Xfibres, a $\log(f_j)$ prior distribution is used when $j > 1$.

The assumptions I

- Each of these parameters is subject to a prior distribution, which are chosen to be noninformative except for where we ensure positivity:

$$\begin{aligned}
\mathscr{P}(\theta, \phi) &\propto \sin(\theta) \\
\mathscr{P}(S_o) &\sim \mathscr{U}(0, \infty) \\
\mathscr{P}(f) &\sim \mathscr{U}(0, 1) \qquad (9) \\
\mathscr{P}(d) &\sim \Gamma(a_d, b_d) \qquad (10)
\end{aligned}$$

- There will be a *distribution*, $H(\theta_j, \phi_j)$, of fiber orientations in the voxel.
  - In order to estimate this distribution we must build a model which, given this distribution, could predict the diffusion-weighted MR measurements.

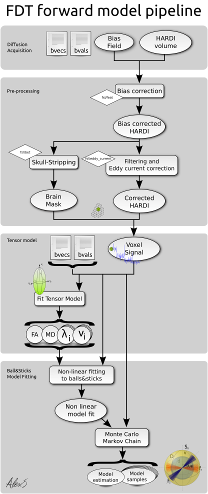
The assumptions II

- Such a model clearly requires some assumptions.
  - That the MR signal from the voxel is the sum of the signal from arbitrarily small subvoxels
  - The signal from each subvoxel behaves as described by Eq. 7.

Outline

1. Introduction
2. The diffusion tensor model
3. The simple partial volume (PVM) model
    - The model
    - Initializing and model fitting
    - The sampling method
    - The improvements
    - Noise estimation
    - Model selection
4. Summary and questions

The BedpostX pipeline



FDT forward model pipeline

Diffusion Acquisition

bvecs

bvals

Bias Field

HARDI volume

Pre-processing

fsl/feat

Bias correction

Bias corrected HARDI

fsl/bet

Skull-Stripping

fsl/eddy_current

Filtering and Eddy current correction

Brain Mask

Corrected HARDI

Tensor model

bvecs

bvals

Voxel Signal

Fit Tensor Model

FA

MD

λ_i

v_i

Ball&Sticks Model Fitting

Non-linear fitting to balls&sticks

Non linear model fit

Monte Carlo Markov Chain

Model estimation

Model samples

AlexS

The initialization

- Fit the tensor model to initialize a few parameters of the model:
    - S_0
    - $d = MD * 2$
    - (θ_j, ϕ_j) from $\mathbf{g_1}$, $\mathbf{g_2}$, $\mathbf{g_3}$
    - f_j could be ${}^{FA}/_{num\_fibres}$

The nonlinear optimization approach

- This will give a fit for the rest of parameters of the model: sensitive to initial values.
- Options:
    - Gradient descent
    - Conjugate gradient
    - Levenberg-Marquardt
        - requires first-order partial derivatives of the model.
        - the Hessian can be calculated with the Jacobian matrix

Outline

Why taking samples?

- When fitting a parametrized model to data [5], there are two general approaches which may be taken.
    1. Look for the set of parameters ω which best fit the data: a point estimate of the parameters.
    2. The second approach is to associate a *pdf* with the parameters.
        1. In the Bayesian framework, this distribution is called the posterior distribution on the parameters given the data

Point estimate

- A special case of this is Maximum Likelihood estimation, where we look for the set of parameters which maximize the probability of seeing this realization of the data given the model and its parameters:

$$\omega_{ML} = \arg\max_{\Omega} \mathscr{P}(Y|\omega, M) \tag{11}$$

where Y is the data and M is the model.

Calculating the posterior distribution I

- The posterior distribution on the parameters given the data:

$$\mathscr{P}(\omega|Y,M) = \frac{\mathscr{P}(Y|\omega,M)\,\mathscr{P}(\omega|M)}{\mathscr{P}(Y|M)} \tag{12}$$

- This posterior density allows us to ask the question of any hypervolume $\mathscr{V}$ in parameter space Ω:
  - "What is our belief given the measured data that the true value of ω is in $\mathscr{V}$?".
- In the one-dimensional case, this question becomes, for any (ω_0, ω_1):
  - "What is our belief that ω lies between ω_0 and ω_1?".

Calculating the posterior distribution II

- These questions, and their answers, represent the **uncertainty** we have in the values of the parameters ω.

Calculating the posterior distribution III

- Unfortunately, calculating this *pdf* is seldom straightforward.

Random sampling the posterior pdf I

- One solution to this problem is to draw **samples** in parameter space from the joint posterior distribution, implicitly performing the integrals numerically.
- For example:
    - We may repetitively **choose random sets of parameter values** and choose to accept or reject these samples according to a criterion based on the value of the numerator in Eq. 12.

Random sampling the posterior pdf II

- It can be shown (e.g., [9]) that a correct choice of this criterion will result in the accepted samples being distributed according to the *joint posterior pdf* (Eq. 12).
- Such schemes generate independent samples from the posterior pdf, e.g.:
  - rejection sampling
  - importance sampling
- Any marginal distributions may then be generated by examining the samples from only the parameters of interest.
- However, these kinds of sampling schemes tend to be painfully slow
  - particularly in high-dimensional parameter spaces, as samples are proposed at random, and thus each has a very small chance of being accepted.

Markov Chain MonteCarlo (MCMC) I

- Markov Chain MonteCarlo (MCMC) (e.g., [8, 9]) methods are a class of algorithms for sampling from probability distributions based on constructing a Markov chain that has the desired distribution as its equilibrium distribution.
- Samples drawn from the posterior are no longer independent of one another, but the high probability of accepting samples allows for many samples to be drawn and, in many cases, for the posterior *pdf* to be built in a relatively short period of time.
- Usually it is not hard to construct a Markov Chain with the desired properties [wikipedia].
- The more difficult problem is to determine how many steps are needed to converge to the stationary distribution within an acceptable error.

Markov Chain MonteCarlo (MCMC) II

  - A good chain will have rapid mixing, i.e. the stationary distribution is reached quickly starting from an arbitrary position.
- Typical use of MCMC sampling can only approximate the target distribution, as there is always some residual effect of the starting position.
- More sophisticated MCMC-based algorithms such as coupling from the past can produce exact samples, at the cost of additional computation and an unbounded (though finite in expectation) running time.

Examples of MCMC algorithms I

- Random walk algorithms
  - **Metropolis–Hastings algorithm**: Generates a random walk using a proposal density and a method for rejecting proposed moves.
  - **Gibbs sampling**: Requires that all the conditional distributions of the target distribution can be sampled exactly. Popular partly because when this is so, the method does not require any 'tuning'.
  - **Slice sampling**: Depends on the principle that one can sample from a distribution by sampling uniformly from the region under the plot of its density function. This method alternates uniform sampling in the vertical direction with uniform sampling from the horizontal 'slice' defined by the current vertical position.

Examples of MCMC algorithms II

- **Multiple-try Metropolis**: A variation of the Metropolis–Hastings algorithm that allows multiple trials at each point. This allows the algorithm to generally take larger steps at each iteration, which helps combat problems intrinsic to large dimensional problems.

Examples of MCMC algorithms III

- More sophisticated algorithms use some method of preventing the walker from doubling back.
  - **Successive over-relaxation**: A Monte Carlo version of this technique can be seen as a variation on Gibbs sampling; it sometimes avoids random walks.
  - **Hybrid Monte Carlo (HMC)**: Tries to avoid random walk behaviour by introducing an auxiliary momentum vector and implementing Hamiltonian dynamics where the potential function is the target density. The momentum samples are discarded after sampling. The end result of Hybrid MCMC is that proposals move across the sample space in larger steps and are therefore less correlated and converge to the target distribution more rapidly. Some variations on slice sampling also avoid random walks.

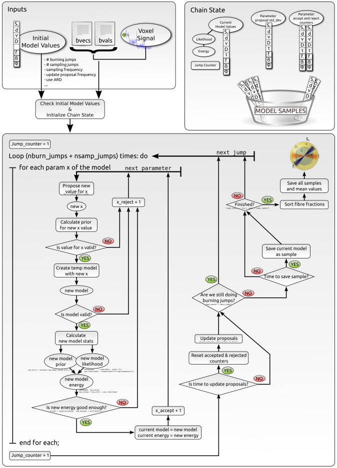
Examples of MCMC algorithms IV

- **Langevin MCMC and other methods** that rely on the gradient (and possibly second derivative) of the log posterior avoid random walks by making proposals that are more likely to be in the direction of higher probability density.

Examples of MCMC algorithms V

- Changing dimension
    - The reversible-jump method is a variant of Metropolis–Hastings that allows proposals that change the dimensionality of the space. This method was proposed in 1995 by Peter Green of Bristol University.
    - Markov chain Monte Carlo methods that change dimensionality have also long been used in statistical physics applications, where for some problems a distribution that is a grand canonical ensemble is used (e.g., when the number of molecules in a box is variable).

The Metropolis-Hastings algorithm

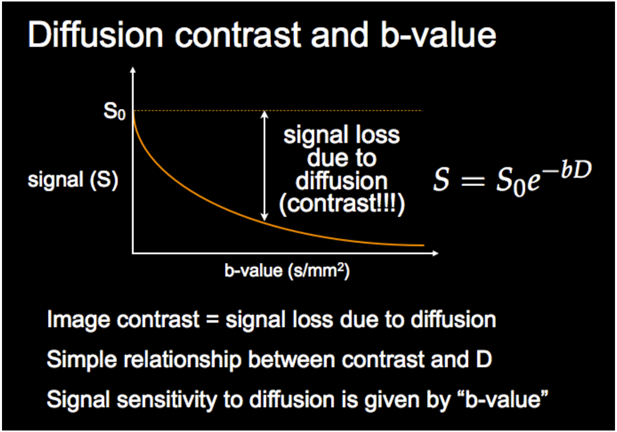


MCMC Algorithm

higher_res

Outline

The nonlinear signal decay



Diffusion contrast and b-value

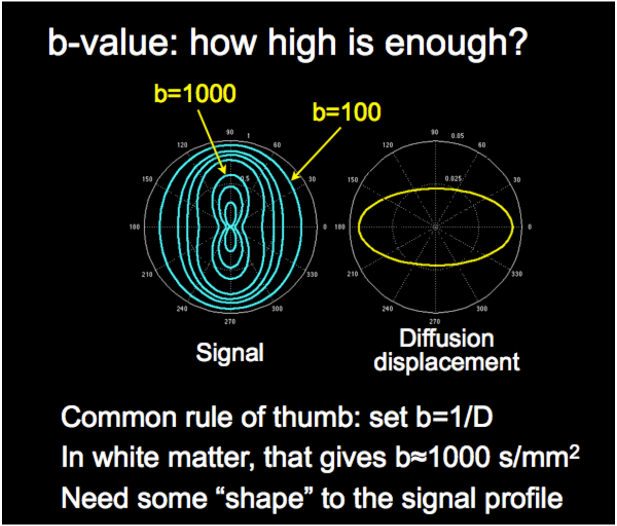
$$S = S_0 e^{-bD}$$

Image contrast = signal loss due to diffusion

Simple relationship between contrast and D

Signal sensitivity to diffusion is given by “b-value”

How many b-values?



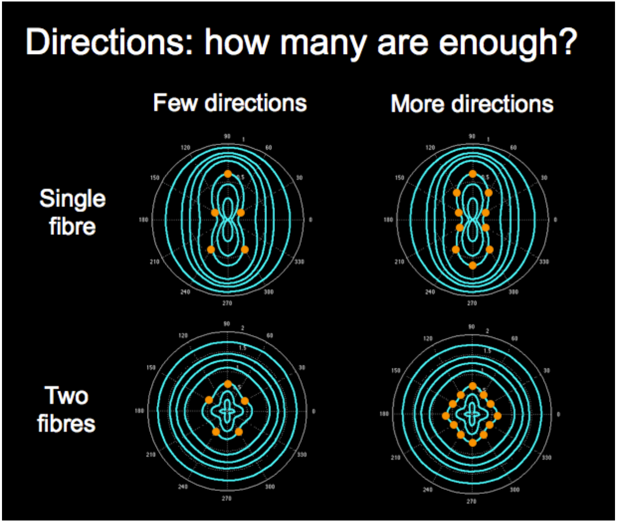
b-value: how high is enough?

Signal

Diffusion displacement

- Common rule of thumb: set b=1/D
- In white matter, that gives b≈1000 s/mm^2
- Need some "shape" to the signal profile

How many directions?



Directions: how many are enough?

Few directions

More directions

Single fibre

Two fibres

FDT Model 2 I

- An estimation of the standard deviation on d, the diffusivity parameter, which will be called d_{std}.
- The signal generated with this model is calculated as following

$$\mu_i = S_0 \left(\left(1 - \sum_{j=1}^{N} f_j \right) \cdot \exp \left(\alpha_d \cdot \log \left(\frac{\beta_d}{\beta_d + b_i} \right) \right) + \right.$$
$$\left. + \sum_{j=1}^{N} f_j \cdot \exp \left(\alpha_d \cdot \log \left(\frac{\beta_d}{\beta_d + b_i d \mathbf{r}_i^T \mathbf{RAR}^T \mathbf{r}_i} \right) \right) \right), \quad (13)$$

where $\alpha_d = \frac{d^2}{d_{std}^2}$, and $\beta_d = \frac{d}{d_{std}^2}$.

Would we need two balls? I

- To tackle the partial volume fraction effect (PVF) where CSF and GM are mixed in the same voxel, an improvement is being tested.
- Model 1 plus two parameters (ν and D) to model the free water found in cerebrospinal fluid (CSF):

$$\mu_i = S_0 \left(\nu \left(\exp\left(-b_i D\right)\right) + (1-\nu) \left(\left(1 - \sum_{j=1}^{N} f_j \right) \cdot \exp\left(-b_i d\right) + \right.\right.$$
$$\left.\left. + \sum_{j=1}^{N} f_j \cdot \exp\left(-b_i d \mathbf{r}_i^T \mathbf{R}\mathbf{A}\mathbf{R}^T \mathbf{r}_i \right) \right) \right), \quad (14)$$

where ν represents the fraction of signal contributed by free water and D the diffusivity in the free water compartment.

Would we need two balls? II

- The prior distributions on this model parameters are as follows:

$$P(S_0) \sim U(0, \infty)$$

$$P(d) \sim U(0, \infty)$$

$$P(d_{std}) \sim U(0, \infty)$$

$$P(\nu) \sim U(0, 1)$$

$$P(D) \sim N\left(3^{-3}, 5^{-6}\right)$$

$$P(f_1) \sim U(0, 1)$$

Would we need two balls? III

$$P\left(f_2^N\right) \sim ARD$$

$$P\left(\theta_1^N\right) \sim \left|\sin\left(\theta\right)\right|$$

$$P\left(\phi_1^N\right) \sim U\left(0,\infty\right)$$

- Values for D were estimated measuring an approximation of D from CSF voxels in the ventricles of a multishell dataset, where a mean of 3^{-3} and standard deviation of 5^{-6}.
- First, we will try to estimate ν from the CSF segmented volume from a T2 acquisition (using FAST).
- We might also need a hyperparameter in the prior of D to chose what is the best mean and standard deviation for the prior Gaussian curve in each voxel.

Outline

1. Introduction
2. The diffusion tensor model
3. The simple partial volume (PVM) model
    - The model
    - Initializing and model fitting
    - The sampling method
    - The improvements
    - Noise estimation
    - Model selection
4. Summary and questions

Noise estimation I

- In addition to the models a likelihood with Rician noise is calculated, for that,
- A τ parameter can be included in the MCMC to estimate the Rician noise in the signal.
- The likelihood is estimated as:

$$\text{likelihood} = -M \log(\tau) - \sum_{i=0}^{N} pred_i,$$

where

$$pred_i = \log(s_i) - \tau/2 \left(s_i^2 + \mu_i^2\right) + logIo\left(\tau\left(\mu_i s_i\right)\right),$$

where M is the number of elements in the signal vector s and $logIo$ is the natural logarithm of the $0th$ order modified Bessel function of first kind

Noise estimation II

[http://mathworld.wolfram.com/BesselFunctionoftheFirstKind.html] (follows the exponential implementation of the Bessel function in [17], Ch. 6).

Outline

What model to use?

- Model selection using Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) could be used to choose a good model to fit the data.
    - $AIC = N \cdot ln(SS/N) + 2M$
    - $BIC = N \cdot ln(SS/N) + M \cdot ln(N)$
    - where M is the number of model parameters, N the number of measurements and SS the sum of squared residuals.
    - The best model minimizes BIC or AIC, BIC giving a higher penalty for more model parameters, so it is more conservative.
- Isabelle Guyon has a recent publication on model selection. [10]

Summary

- Introduction to diffusion
- The tensor model
- The FDT partial volume model
    - MCMC sampling
    - Model selection

Questions?

?

References I

Yaniv Assaf and Peter J. Basser.
Composite hindered and restricted model of diffusion (CHARMED) MR imaging of the human brain.
NeuroImage, 27(1):48–58, August 2005.

Yaniv Assaf, Yoram Cohen, Heidi Johansen-Berg, and Timothy E.J. Behrens.
Inferring microstructural information of white matter from diffusion MRI.
In *Diffusion MRI*, pages 127–146. Academic Press, San Diego, 2009.

Yaniv Assaf, Raisa Z. Freidlin, Gustavo K. Rohde, and Peter J. Basser.
New modeling and experimental framework to characterize hindered and restricted water diffusion in brain white matter.
Magnetic Resonance in Medicine, 52(5):965–978, 2004.

P J Basser, J Mattiello, and D LeBihan.

Estimation of the effective self-diffusion tensor from the NMR spin echo.
Journal of Magnetic Resonance. Series B, 103(3):247–254, March 1994.
PMID: 8019776.

T E J Behrens, M W Woolrich, M Jenkinson, H Johansen-Berg, R G Nunes, S Clare, P M Matthews, J M Brady, and S M Smith.
Characterization and propagation of uncertainty in diffusion-weighted MR imaging.
Magnetic Resonance in Medicine: Official Journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine, 50(5):1077–1088, November 2003.
PMID: 14587019.

T.E.J. Behrens, H. Johansen Berg, S. Jbabdi, M.F.S. Rushworth, and M.W. Woolrich.

Probabilistic diffusion tractography with multiple fibre orientations: What can we gain?
NeuroImage, 34(1):144–155, January 2007.

Timothy E J Behrens, Mark W Woolrich, Mark E Walton, and Matthew F S Rushworth.
Learning the value of information in an uncertain world.
Nat Neurosci, 10(9):1214–1221, 2007.

Dani Gamerman and Hedibert F. Lopes.
Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, Second Edition.
Chapman and Hall/CRC, 2 edition, May 2006.

W.R. Gilks, S. Richardson, and David Spiegelhalter.
Markov Chain Monte Carlo in Practice: Interdisciplinary Statistics.
Chapman and Hall/CRC, 1 edition, December 1995.

Isabelle Guyon, Amir Saffari, Gideon Dror, and Gavin Cawley.
Model selection: Beyond the Bayesian/Frequentist divide.

- P. Hagmann, L. Jonasson, P. Maeder, J.-P. Thiran, V. J. Wedeen, and R. Meuli.
  Understanding diffusion MR imaging techniques: From scalar diffusion-weighted imaging to diffusion tensor imaging and beyond1.
  Radiographics, 26(Supplement 1):S205–S223, October 2006.

- Tim Hosey, Guy Williams, and Richard Ansorge.
  Inference of multiple fiber orientations in high angular resolution diffusion imaging.
  Magnetic Resonance in Medicine, 54(6):1480–1489, 2005.

- S. Jbabdi, M.W. Woolrich, J.L.R. Andersson, and T.E.J. Behrens.
  A bayesian framework for global tractography.
  NeuroImage, 37(1):116–129, August 2007.

- Heidi Johansen-Berg and Timothy E. J. Behrens.
  Diffusion MRI: from quantitative measurement to in-vivo neuroanatomy.

Academic Press, 2009.

Shlomi Lifshits, Arie Tamir, and Yaniv Assaf.
Combinatorial fiber-tracking of the human brain.
NeuroImage, 48(3):532–540, November 2009.

Karla L. Miller.
Lectures and teaching material - FMRIB centre, university of oxford.
http://www.fmrib.ox.ac.uk/Members/karla/lectures-and-teaching-material.

William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery.
Numerical Recipes 3rd Edition: The Art of Scientific Computing.
Cambridge University Press, 3 edition, August 2007.

SO Rice and Nelson Wax.
Mathematical analysis of random noise.

In *Selected Papers on Noise and Stochastic Processes*. Dover Pubns, June 1954.