

Filtrado Multicanal de Ruido en DWI Mediante Variación Total e Iteraciones de Bregman

A.G. Seco de Herrera^{1,2}, J.F. Garamendi^{1,2}, E.Schiavi^{1,2,3}, J.A. Hernández-Tamames^{1,2,4}

¹Fundación CIEN-Fundación Reina Sofía, Madrid, España,
{aseco, jfgaramendi}@fundacioncien.es

²Laboratorio de Neuroimagen CTB, Universidad Politécnica de Madrid, Madrid, España.

³Departamento de Matemática Aplicada, Universidad Rey Juan Carlos, Madrid, España,

⁴Departamento de Tecnología Electrónica, Universidad Rey Juan Carlos, Madrid, España,
{emanuele.schiavi, juan.tamames}@urjc.es

Resumen

En el presente trabajo adaptamos al ámbito de las imágenes de difusión ponderada DWI (Diffusion Weighted Imaging) una técnica de eliminación de ruido recientemente propuesta para imágenes vectoriales. En un marco multicanal estudiamos el efecto del preprocesado en la estimación del tensor de difusión (DT-MRI o DTI) y en el cálculo de las medidas escalares que de él se derivan observando una clara mejora en el cálculo de la anisotropía fraccional.

1. Introducción

A partir de las imágenes de resonancia magnética (IRM) ponderadas en difusión (DWI de las siglas en inglés Diffusion Weighted Imaging) se puede medir la difusión de las moléculas de agua en un punto del cerebro y por lo tanto inferir la estructura del medio en el que se encuentran. Si la difusividad del agua es igual en todas las direcciones del espacio esta se denomina isotrópica, si no es así, la difusividad es anisotrópica. Cuando se presenta anisotropía en la difusividad, esta puede ser debida a la estructura circundante, si la molécula de agua se encuentra por ejemplo en un tracto o en un axón, la difusión se producirá a lo largo de este. Esta información de restricción del movimiento de las moléculas puede ser representada en cada punto del espacio (imagen) $x \in \Omega$ mediante un tensor D que indica el movimiento a lo largo de cada dirección del espacio (y la correlación entre estas direcciones [1]). Este tensor es una matriz 3×3 simétrica y definida positiva:

$$D(x) = \begin{pmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{pmatrix}$$

y puede ser estimado a partir de las imágenes DWI, $S_m : \Omega \subset \mathbb{R}^3 \rightarrow \mathbb{R}$ según la ecuación de Stejskal-Tanner [2]

$$S_m(x) = S_0(x) e^{-b \bar{g}_m^T D(x) \bar{g}_m} \quad (1)$$

donde S_m es la señal medida a lo largo de una direc-

ción espacial $\bar{g}_m$ (que se encuentra atenuada respecto a la señal S_0 , señal sin difusión, debido al movimiento de las moléculas de agua) y b es un parámetro dependiente de la secuencia de pulsos, la fuerza del gradiente y otras constantes. Dado que D es simétrica, sólo es necesario adquirir imágenes en 6 direcciones del espacio para calcular una solución analítica de la ecuación (1) (estimación lineal, [3]), aunque es posible (y preferible) tener un número de direcciones mucho mayor ya que la estimación lineal es muy sensible al ruido. Sin embargo, el tiempo de adquisición aumenta con el número de direcciones y provoca un mayor tiempo de exposición del paciente.

Una vez calculado el tensor D , es posible extraer una gran cantidad de características y medidas escalares. Entre las más utilizadas se encuentran la difusividad media y la anisotropía fraccional (AF) que caracteriza la anisotropía del tensor de difusión y se define por

$$AF(x) = \sqrt{\frac{3}{2} \frac{(\lambda_1 - \bar{\lambda})^2 + (\lambda_2 - \bar{\lambda})^2 + (\lambda_3 - \bar{\lambda})^2}{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}} \quad (2)$$

donde λ_1, λ_2 y λ_3 son los autovalores del tensor D y $\bar{\lambda} = (\lambda_1 + \lambda_2 + \lambda_3)/3$ es la difusividad media.

En el presente trabajo se estudia el efecto que produce el preprocesado de las imágenes de DWI en el posterior cálculo del tensor mediante el análisis de las medidas escalares que de él se derivan. Siguiendo a Kim et al.[4] y McGraw et al. [5] hemos considerado un modelo de ruido aditivo gaussiano. En Tristán-Vega y Aja-Fernández consideran un modelo de ruido de tipo Rician para realizar la eliminación de ruido en HARDI (High Angular Resolution Diffusion Imaging) [6], donde valores altos de b (aprox. 3000 s/mm²) provoca que la relación señal ruido (SNR) sea baja. En el caso de DTI, las imágenes utilizadas en este trabajo han sido obtenidas con un b menor (aprox. 1000 s/mm²), obteniendo por tanto mayor SNR, por lo que la hipótesis de ruido aditivo gaussiano puede considerarse suficientemente adecuada [5]. El modelo de eliminación de

ruido utilizado se plantea en un marco variacional donde se define un funcional de energía y el mínimo de dicho funcional representa la imagen sin ruido buscada.

En este trabajo consideramos una modificación del modelo de ROF (introducido por Rudin, Osher y Fatemi [7]) para eliminar ruido. Este modelo supera claramente los modelos clásicos lineales como el filtrado gaussiano (por su propiedad de conservación de los bordes), pero tiene también algunos inconvenientes como la pérdida de contraste y textura de la imagen filtrada. Para solventar esas desventajas Osher et al., [8] proponen un procedimiento iterativo denominado 'inverse scale space' que mejora el modelo de ROF. Más recientemente, Bresson et al. [9] generalizan el proceso de 'inverse scale space' a imágenes multicanal. En el presente trabajo se adapta dicho modelo a su aplicación a imágenes DWI considerando cada imagen S_m un canal. En la figura 1 mostramos un ejemplo de eliminación de ruido con inverse scale aplicado a una imagen DWI real.

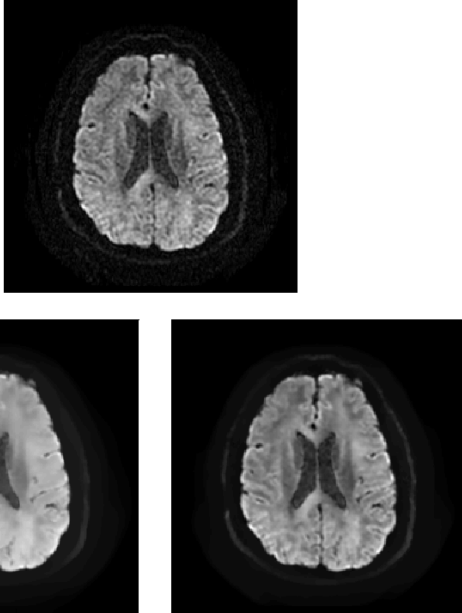


Figura 1. Arriba un corte de la imagen DWI original ruidosa en la dirección $\bar{g}_{15}=(0.335962, 0.0429951, 0.940894)$. Abajo a la izquierda, el mismo corte tras la primera iteración del proceso de 'inverse scaling' y, abajo a la derecha, tras la segunda iteración.

Este artículo está organizado en 4 apartados. El segundo apartado introduce el procedimiento de 'inverse scaling' a través del término regularizante de variación total (VT) y la notación básica necesaria para trabajar en el marco multicanal. La resolución numérica se presenta en el tercer apartado y en el último mostramos algunos resultados preliminares y trabajos futuros.

2. Inverse Scaling

Sea $\bar{f} : \Omega \subset \mathbb{R}^N \rightarrow \mathbb{R}^M$, siendo $\bar{x} \rightarrow \bar{f}(\bar{x}) := (f_1(\bar{x}), \dots, f_M(\bar{x}))$, una función vectorial donde cada función escalar $f_m : \Omega \rightarrow \mathbb{R}, 1 \leq m \leq M$, es uno de los M canales y representa cada una de las direcciones de la imagen DWI. Suponemos $\bar{f} \in L^\infty(\Omega; \mathbb{R}^M)$, i.e. $\bar{f}$ esencialmente acotada. Los detalles de los espacios funcionales de Lebesgue se pueden encontrar en el libro de Brezis [10].

La idea es obtener una descomposición de los datos de la forma

$$\bar{f} = \bar{u} + \bar{v}$$

donde $\bar{u}$ es la señal sin ruido y $\bar{v}$ es el ruido. Una buena aproximación para resolver el problema está basada en la minimización de la variación total del funcional de ROF (VROF) [7]:

$$J(\bar{u}) = |\bar{u}|_{BV} + \frac{1}{2\lambda} \|\bar{f} - \bar{u}\|_2^2 \quad (3)$$

para un parámetro escalar real $\lambda > 0$, donde $BV(\Omega; \mathbb{R}^M)$ denota el espacio de las funciones vectoriales de variación acotada y $|\cdot|$ es la seminorma BV definida por $|\bar{u}|_{BV} = |D\bar{u}|(\Omega)$, i.e., la variación total de $\bar{u}$.

Una función $\bar{u}$ de variación acotada puede ser caracterizada como sigue: decimos que $\bar{u} \in BV(\Omega; \mathbb{R}^M)$ si y sólo si existen medidas de Radon $\bar{\mu}_1, \bar{\mu}_2$ y $\bar{\mu}_3$, $\bar{\mu}_i = (\mu_i^1, \dots, \mu_i^M), i = 1, \dots, 3$, definidas en Ω con masa total finita en Ω tal que:

$$\sum_{k=1}^M \int_{\Omega} u^k \operatorname{div} \bar{\xi}^k dx = - \sum_{k=1}^M \sum_{i=1}^3 \int_{\Omega} \bar{\xi}^k dD_i u^k,$$

para toda $\bar{\xi} \in C_c^1(\Omega; \mathbb{R}^{3 \times M})$. Entonces, el gradiente generalizado de $\bar{u}$, $D\bar{u}$ es una matriz $3 \times M$ de medidas $D_i u^k = \mu_i^k$ con variación total finita:

$$|D\bar{u}|(\Omega) = \sup \left\{ \sum_{k=1}^M \int_{\Omega} u^k \operatorname{div} \bar{\xi}^k dx : \bar{\xi} \in C_c^1(\Omega; \mathbb{R}^{3 \times M}), \|\bar{\xi}\|_{\infty} \leq 1 \right\} \quad (4)$$

(ver Ambrosio, Fusco y Pallara, [11] para más detalles).

Una vez obtenido el mínimo $\bar{u}$, es de esperar que $\bar{v} = \bar{f} - \bar{u}$ contenga sólo la información procedente del ruido. Sin embargo, $\bar{v}$ también contiene información estructural que debería estar contenida en $\bar{u}$. Para solventar este problema, Osher et al. en [8] proponen tener en cuenta no sólo el valor de la intensidad entre la imagen filtrada y la ruidosa sino también las normales de los conjuntos de nivel ('level sets') en cada imagen. Para ello, primero se resuelve el problema de minimización asociado a (3) para obtener $\bar{u}_1$ y (con abuso de notación) $\bar{n}_1 = \nabla \bar{u}_1 / |\nabla \bar{u}_1|$. La consideración propia de este término es mediante argumentos de dualidad [12]. Posteriormente se considera el siguiente funcional:

$$J(\bar{u}) = |\bar{u}|_{BV} - \int_{\Omega} \bar{n}_1 \cdot \nabla \bar{u} + \frac{1}{2\lambda} \|\bar{f} - \bar{u}\|_2^2 \quad (5)$$

Obsérvando que $\bar{f} = \bar{u}_1 + \bar{v}_1$ y, utilizando las ecuaciones de Euler-Lagrange y el problema de minimización asociado a (3), se obtiene que [8]

$$- \int_{\Omega} \bar{n}_1 \cdot \nabla \bar{u} = - \frac{1}{\lambda} \int_{\Omega} \bar{u} \bar{v}_1$$

pudiendo caracterizar el mínimo de (5) como el mínimo de

$$J(\bar{u}) = |\bar{u}|_{BV} + \frac{1}{2\lambda} \|\bar{f} + \bar{v}_1 - \bar{u}\|_2^2 \quad (6)$$

donde $\bar{v}_1 = \bar{f} - \bar{u}_1$. Este proceso puede ser repetido de manera iterativa obteniendo el procedimiento denominado 'inverse scaling':

1. Inicializar $\bar{u}_0 \equiv 0, \bar{v}_0 \equiv 0$.
2. Para $j = 0, 1, 2, \dots$
 - Calcular $\bar{u}_{j+1}$ como un mínimo de la energía de VROF modificado:

$$J(\bar{u}) = |\bar{u}|_{BV} + \frac{1}{2\lambda} \|\bar{f} + \bar{v}_j - \bar{u}\|_2^2 \quad (7)$$

- Actualizar $\bar{v}_{j+1}$:

$$\bar{v}_{j+1} = \bar{v}_j + \bar{f} - \bar{u}_{j+1}.$$

El principal punto a tener en cuenta en el proceso anterior es el criterio de parada. Se puede observar (ver Osher et al [8]) que, fijado $\lambda > 0$, existe un valor de j , digamos j^* , tal que la distancia de Bregmann (véase Osher et al., [8]) entre la imagen calculada (aproximada) $\bar{u}_{j^*}$ y la imagen verdadera (exacta) libre de ruido, es mínima (ver [8] para más detalles). La sucesión $\{\bar{u}_j\}_{j=1..j^*}$ (donde $\bar{u}_1$ es la solución de ROF) hace decrecer monótonamente la distancia de Bregmann. Desde un punto de vista práctico, basta elegir λ lo suficiente grande para que j^* corresponda a la segunda iteración de manera que la primera iteración sea una imagen sobresuavizada (ver figura 1 abajo izquierda) incluyendo la información estructural básica (bajas frecuencias) sin detalles (y ruido). En la iteración siguiente se introducen altas frecuencias, en particular bordes y detalles de pequeña escala (ver figura 2).

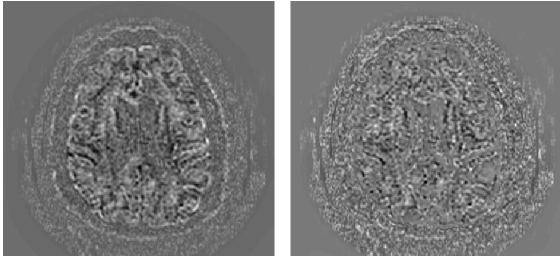


Figura 2. A la izquierda diferencia entre la imagen ruidosa original y la primera iteración del proceso de 'inverse scaling'. A la derecha diferencia entre la imagen ruidosa original y la segunda iteración. Se ve claramente la presencia de información estructural de la izquierda, mientras que en la imagen de la derecha esta es menor.

3. Resolución Numérica

Chambolle [12] observa que la minimización de la variación total (escalar) puede ser vista como un algoritmo de proyección en un conjunto convexo adecuado. Siguiendo a Bresson en su formulación vectorial [9] y usando (4), el modelo modificado de VROF (7) se escribe de la siguiente forma:

$$\inf_{\bar{u}} \sup_{|\bar{\xi}| \leq 1} \left\{ \int_{\Omega} \bar{u} \operatorname{div} \bar{\xi} dx + \frac{1}{2\lambda} \int_{\Omega} (\bar{f} + \bar{v}_j - \bar{u})^2 dx \right\}$$

y la condición de optimalidad viene dada por (la demostración se encuentra en [12]):

$$\bar{u} = \bar{f} + \bar{v}_j - \lambda \operatorname{div} \bar{\xi}$$

siendo $\bar{\xi}$ solución de las ecuaciones de Euler-Lagrange:

$$\nabla (\lambda \operatorname{div} \bar{\xi} - (\bar{f} + \bar{v}_j)) - |\nabla (\lambda \operatorname{div} \bar{\xi} - (\bar{f} + \bar{v}_j))| \bar{\xi} = 0$$

El sistema anterior se resuelve usando una generalización [9] del algoritmo de punto fijo propuesto por Chambolle [12] lo que origina un esquema de descenso de gradiente semi implícito: Sea $\tau > 0$, y $\bar{\xi}^0 = 0$; calculamos $\bar{\xi}^{n+1}$ como

$$\bar{\xi}^{n+1} = \frac{\bar{\xi}^n + \tau \nabla (\lambda \operatorname{div} \bar{\xi}^n - (\bar{f} + \bar{v}_j))}{1 + \tau |\nabla (\lambda \operatorname{div} \bar{\xi}^n - (\bar{f} + \bar{v}_j))|}$$

En nuestros experimentos numéricos elegimos $\tau = 1/8$ como sugiere Chambolle para mantener la estabilidad. El operador divergencia se define como $(\bar{\xi}_m) = ((\bar{\xi}_m)^x, (\bar{\xi}_m)^y, (\bar{\xi}_m)^z)$, con $1 \leq m \leq M$ el número de canal (en nuestro caso el número de direcciones), en el vóxel (i, j, k) :

$$\begin{aligned} (\operatorname{div} (\bar{\xi}_m))_{i,j,k} &= \\ &= \begin{cases} (\bar{\xi}_m)_{i,j,k}^x - (\bar{\xi}_m)_{i-1,j,k}^x & \text{si } 1 < i < N_x \\ (\bar{\xi}_m)_{i,j,k}^x & \text{si } i = 1 \\ -(\bar{\xi}_m)_{i-1,j,k}^x & \text{si } i = N_x \end{cases} \\ &+ \begin{cases} (\bar{\xi}_m)_{i,j,k}^y - (\bar{\xi}_m)_{i,j-1,k}^y & \text{si } 1 < j < N_y \\ (\bar{\xi}_m)_{i,j,k}^y & \text{si } j = 1 \\ -(\bar{\xi}_m)_{i,j-1,k}^y & \text{si } j = N_y \end{cases} \\ &+ \begin{cases} (\bar{\xi}_m)_{i,j,k}^z - (\bar{\xi}_m)_{i,j,k-1}^z & \text{si } 1 < k < N_z \\ (\bar{\xi}_m)_{i,j,k}^z & \text{si } k = 1 \\ -(\bar{\xi}_m)_{i,j,k-1}^z & \text{si } k = N_z \end{cases} \end{aligned} \quad (8)$$

y el operador gradiente, $(\nabla g)_{i,j,k} = ((\nabla g)_{i,j,k}^x, (\nabla g)_{i,j,k}^y, (\nabla g)_{i,j,k}^z)$, aplicado a una función g evaluado en el punto (i, j, k) como sigue:

$$\begin{aligned} (\nabla g)_{i,j,k}^x &= \begin{cases} g_{i+1,j,k} - g_{i,j,k} & \text{si } i < N_x \\ 0 & \text{si } i = N_x \end{cases} \\ (\nabla g)_{i,j,k}^y &= \begin{cases} g_{i,j+1,k} - g_{i,j,k} & \text{si } j < N_y \\ 0 & \text{si } j = N_y \end{cases} \\ (\nabla g)_{i,j,k}^z &= \begin{cases} g_{i,j,k+1} - g_{i,j,k} & \text{si } k < N_z \\ 0 & \text{si } k = N_z \end{cases} \end{aligned}$$

donde (N_x, N_y, N_z) son el número de vóxeles en la dirección (x, y, z) de cada canal.

4. Resultados y Discusión

El preprocesado descrito anteriormente se ha aplicado en imágenes adquiridas por un escáner General Electric de 3T equipado con una antena de 8 canales. Las imágenes DWI se obtuvieron con una secuencia single-shot spin-eco EPI (FOV = 24cm, TR=9100, TE=88.9, grosor del corte = 3mm, espaciado=0.3, tamaño de la matriz=128x128,

NEX=2). Cada imagen DWI consta de una imagen con $b=0/\text{mm}^2$ y 15 imágenes con $b=1000\text{s}/\text{mm}^2$ cada una de ellas correspondiente a una dirección orientada de acuerdo a Jones et al. [13].

En la figura 3 se aprecia la influencia en imágenes de anisotropía fraccional obtenidas a partir del tensor mediante la fórmula 2. Con el preprocesado previo aparece una mayor claridad en la anisotropía correspondiente a los tractos que sin el preprocesado aparecen demasiado difuminados.

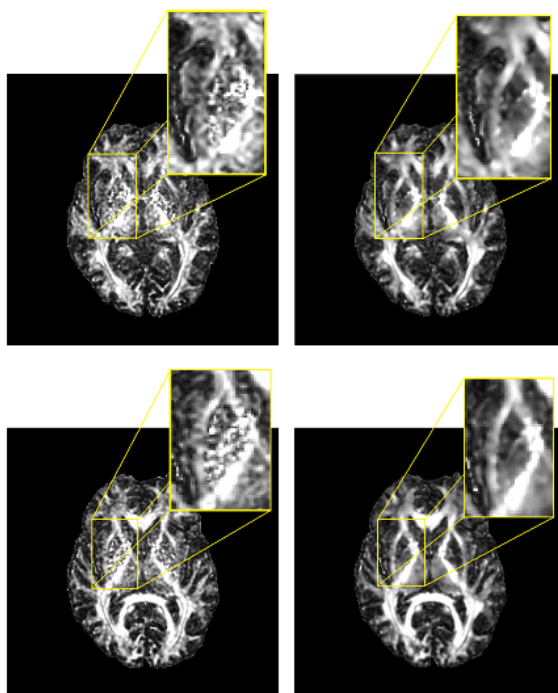


Figura 3. Dos cortes de una imagen de anisotropía fraccional. A la izquierda las imágenes obtenidas antes del preprocesado. A la derecha tras el preprocesado.

5. Conclusiones y Líneas Futuras

Se ha adaptado el modelo de 'inverse scaling' multicanal para ser aplicado a las imágenes de DWI antes de realizar el cálculo del tensor, obteniéndose resultados preliminares muy satisfactorios a la hora de estimar posteriormente medidas escalares como la anisotropía fraccional. Futuras mejoras del modelo propuesto radican en la consideración de ruido tipo Rician (véase [6]), propio de la MRI, y en la introducción la norma L^1 en el término de fidelidad de los datos del funcional de energía (3). También se esperan buenos resultados en la orientación de las fibras para lo que se utilizarán adquisiciones con un mayor número de direcciones.

Agradecimientos

Este trabajo ha sido financiado parcialmente por los proyectos: TEC2009-14587-C03-03 y TEC2009-14587-C03-02.

A Dña. Eva Alfayate, técnico de rayos de la Fundación Reina Sofía, y Juan Álvarez Linera, jefe del área de Neuroimagen de la Fundación CIEN, sin cuya colaboración este

trabajo no hubiera sido posible.

Referencias

- [1] Le Bihan D, Mangin JF, Poupon C, Clark CA, Pappata S, Molko, Chabriet H. Diffusion Tensor Imaging: Concepts and Applications. Journal of Magnetic Resonance Imaging, vol 13, 2001, pp 534-546.
- [2] Stejskal EO, Tanner JE. Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient. Journal of Chemical Physics vol 42, 1965. pp 288-292.
- [3] Westin CF, Maier S, Mamata H, Nabavi A, Jolesz F, Kikinis R. Processing and visualization for diffusion tensor mri. Journal of Medical Image Analysis, vol. 6, 2002, pp. 93-108.
- [4] Kim Y, Thompson P, Vese L. HARDI data denoising using vectorial total variation and logarithmic barrier. Journal of Inverse Problems and Imaging, vol 4, n 2, 2010, pp 273-310.
- [5] McGraw T, Vemuri B, Ozarslan E. Variational denoising of diffusion weighted MRI. Journal of Inverse Problems and Imaging, vol 3 n 4, 2009, pp 625-648.
- [6] Tristán-Vega A, Aja-Fernández S. DWI filtering using joint information for DTI and HARDI. Journal of Medical Image Analysis, vol14, 2010, pp 205-218
- [7] Rudin LI, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms, Journal of Physica D, vol 60, 1992, pp 259-268.
- [8] Osher S, Burger M, Goldfarb D, Xu J, Yin W. An iterative regularization method for total variation based image restoration. Journal of Multiscale Modeling and Simulation, vol 4, 2005. pp 460-484.
- [9] Bresson X, Chan TF. Fast dual minimization of the vectorial total variation norm and applications to color image processing. Journal of Inverse Problems and Imaging vol 2, n 4, 2008, pp 455-484.
- [10] Brezis H. Análisis Funcional. Alianza Editorial Sa, 2007 (ISBN-10 8420680885).
- [11] Ambrosio L, Fusco N, Pallara D. Functions of bounded variation and free discontinuity problems. Oxford Mathematical Monographs. T, 2000 (ISBN-10 0198502451).
- [12] Chambolle A. An algorithm for Total variation Regularization and Denoising. Journal of Math. Imaging Vision, vol 20, 2004, pp 89-97.
- [13] Jones DK, Horsfield MA, Simmons A. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. Journal of Magnetic Resonance in Medicine vol 42 (3), 1999, pp 515-525.