MR Center, Dept. of Clinical Neuroscience
Karolinska Institutet, Stockholm, Sweden

OPTIMISATION STRATEGIES IN DIFFUSION TENSOR MR IMAGING

Stefan Skare

Stockholm 2002

All previously published papers were reproduced with permission from the publisher.

Figure 11 and Figure 12 courtesy of Carlo Pierpaoli and Peter Basser, NIH, Bethesda, USA.
Figure 13 and Figure 14 courtesy of Susumu Mori, Johns Hopkins, Baltimore, USA.

Linguistic editing by Janet Holmén

Published and printed by Karolinska University Press
Box 200, SE-171 77 Stockholm, Sweden

ISBN 91-7349-175-6

ABSTRACT

With diffusion MR imaging, the Brownian motion (or self-diffusion) of water molecules is measured. In the corresponding ADC (Apparent Diffusion Coefficient) image, each image element (voxel) represents the average diffusion. In MRI, the diffusion is measured with diffusion gradients along certain directions in space. The diffusion-weighting gradients are often incorporated in an echo planar imaging (EPI) pulse sequence. Depending on the type of tissue being imaged, the measured diffusion may be isotropic, i.e. equal for all directions of the diffusion gradient, as seen, e.g., in grey matter and the cerebrospinal fluid (CSF). In contrast, in white matter, the diffusion is higher along the nerve fibres than across because water molecules moving along fibres are not hindered – this is referred to as anisotropic diffusion. In diffusion tensor imaging (DTI), the diffusion, for each voxel, is represented by an ellipsoid with a certain shape and orientation. For e.g. grey matter the ellipsoid is a sphere, whereas for white matter it is elongated in one direction like a cigar. From the diffusion tensor the mean diffusion (size of the ellipsoid) and the degree of anisotropy (shape of the ellipsoid) may be calculated.

In this thesis different strategies to improve the image quality and reliability of the DTI data and diffusion anisotropy maps are presented. These include simulation studies to determine 1) which anisotropy index is most insensitive to noise and 2) which diffusion scheme (i.e. which set of diffusion gradients to use in the scanning process) minimises the variance and bias of the calculated DTI data. The simulation results were also complemented with data from phantoms and volunteers. Additionally, two major types of image artefacts in diffusion weighted single-shot echo planar imaging (DW SS-EPI) and means of correcting them have been investigated. The first artefact is signal dropout, predominantly in the mid-lower part of the brain, due to brain motion. The second is eddy currents induced by the diffusion gradients that cause the DW images to be distorted differently depending on the direction of the diffusion gradient. The distortions are translation, scaling and shear effects in the phase encoding direction of the image. These distortions, together with patient motion, result in anatomical mismatch between the different DW images used for the calculation of the diffusion tensor data. A new distortion correction method that corrects for this mismatch has been developed. DTI has been performed in a study comparing schizophrenic patients and normal controls with respect to diffusion anisotropy, mean diffusion and morphological differences.

Keywords: Diffusion tensor, magnetic resonance imaging, optimisation, artefacts

ACKNOWLEDGEMENTS

This work has been carried out at MR-Center, Dept. of Clinical Neuroscience, Karolinska Institutet, Stockholm. I wish to express my sincere gratitude and appreciation to all that in one way or another have contributed to this work, and especially to:

Bo Nordell, my former supervisor at Medical Radiation Physics, Stockholm University, my boss at the MR section of Dept. Hospital Physics, co-worker and friend. Thank you for introducing me into the MR field and to all the top scientists you know world-wide at the yearly ISMRM meetings. But also for at all times being supportive in all respects. It has also been great fun working with you on the MR compatible incubator project, diving at the Great Barrier Reef and skiing on glaciers in Whistler.

Jesper Andersson, my supervisor, for being a great, supportive mentor and friend. I'm very grateful for taking your supervisor job so seriously that you even in part changed your research field to diffusion tensor imaging. For teaching me all the mathematics, statistics and image processing to a depth I probably never would have achieved otherwise.

Tie-Qiang Li, my second supervisor, for helping me getting started in the diffusion tensor imaging field and clearing up my first two messy manuscripts to something publishable. Also, for your hospitality during my visit at your lab in Stanford and for the vineyard tours. It is really unfortunate that you don't work at our lab anymore.

Martin Ingvar and Olof Flodmark, scientific and clinical heads of MR Center, for letting me do part-time research although employed by the Karolinska Hospital, but also for always being supportive and delegating responsibilities in a way that makes one feel important as well as appreciated.

All the "white coats" who work at the MR Center: Åsa L., Åsa A., Christina, Mitra, Annetthe, Roberth, Birgitta, Susanne, Yords and Dan. Thank you all for contributing to a very nice working climate where clinical duties and physics experiments have been nicely intermingled.

All the other MR physicists: Magnus, Peter, Gunther, Andreas, Tomas J., Tobias and Mats for being great co-workers and contributing to a fun critical mass of MR physicists.

The "diffusion" doctors at the Neuroradiology Dept., Anders L. and Pernille for fruitful discussions.

Susumu Mori, Carlo Pierpaoli and Peter Basser for allowing me to use their figures for the summary. Thanks also for taking the time to find them and send them to me by e-mail.

My NF (Naturvetenskapliga Föreningen) friends in the badminton team – Björn, Sebastian, Magnus, Anders O., Patriq and Tim, for forcing me to exercise at least one hour per week.

All the other friends from NF, Grodorden, Naturvetarspexet, Klubbmästeristyrelsen. Thank you for all the friendship and great fun in the 90's. It will really be nice to look back on this era in the future.

My eternal friends from the youth, Christina and Jonas. We should really see each other more often.

Finally, I wish to thank my parents Ingvar and Lizbet for the generous support and care from the very start, Petra for being a wonderful wife and life companion, and my two-year old son Ludvig, for giving life and joy a new meaning.

LIST OF ABBREVIATIONS

γ	The gyromagnetic constant. ^{1}H: 42.576 MHz/T = 267.5 rad/sT
λ_1, λ_2, λ_3	The eigenvalues of **D**
ΔB_0	Inhomogeneity of the B_0 field
$\langle \mathbf{D} \rangle$, $\bar{\mathbf{D}}$, Trace(**D**)/3, ADC_{mean}	Mean diffusivity
b	Diffusion gradient dependent variable. Determines the amount of diffusion weighting
B_0	The static magnetic field
CSF	CerebroSpinal Fluid
D, ADC	Diffusion measured in one direction. Apparent Diffusion Coefficient
DEC	Directional Encoded Colour. Colour definition for displaying diffusion tensor images
DT, **D**	Diffusion Tensor
DTI	Diffusion Tensor Imaging
DWI	Diffusion Weighted (component) Image
EC	Eddy Currents
ECG	ElectroCardioGram containing signals from the heart
EPI, SS-EPI	(Single-shot) Echo Planar Imaging pulse sequence
fMRI	Functional Magnetic Resonance Imaging
FOV	Field Of View. The size of the MR image (unit: cm)
FSE, SS-FSE	(Single-shot) Fast Spin Echo pulse sequence
GM	Grey matter
G_x, G_y, G_z	The magnetic field gradients
MR, MRI	Magnetic Resonance Imaging
Pixel	Image element
PSF	Point spread function
RA, FA, VR, A_{major}, LI, A_{dd}	Various diffusion anisotropy indices
rBW, prBW	Receiver bandwidth, pseudo receiver bandwidth
RF	Radio Frequency pulse
RMS	Root Mean Square
ROI	Region Of Interest. An area defined in an image
R-R	Time interval between (the R-wave of) two heart beats
SE	Spin Echo pulse sequence. Also refers to the actual echo for the SE, FSE and EPI pulse sequences.
SNR	Signal-to-Noise Ratio
Spins	In MRI: The (magnetic moment of) ^{1}H protons in water and fat
Spiral, SS-Spiral	(Single-shot) Spiral pulse sequence
STE	Stimulated Echoes. Occurs in FSE sequences
T_1-w	T_1-weighted
T_2-w	T_2-weighted
TD	Trigger Delay. Time after the heart beat (or pulse) that imaging starts
TE	MR scan parameter. Controls the extent of T_2-weighting
TR	MR scan parameter. Controls the extent of T_1-weighting
TW	Trigger Window. Time when the MR scanner stops and waits for the next heart beat (or pulse)
TWB	Thesis Writing Block
WM	White Matter. The nerve fibres
Voxel	Image volume element. A pixel with a thickness

LIST OF PUBLICATIONS

I S Skare, TQ Li, B Nordell, M Ingvar
Noise considerations in the estimation of diffusion anisotropy
Magnetic Resonance Imaging, 18(6):659-669 (2000)

II S Skare, M Hedehus, M Moseley, TQ Li
Condition number as a measure of noise performance of diffusion tensor data acquisition schemes with MRI
Journal of Magnetic Resonance, 147(2):340-352 (2000)

III S Skare and J LR Andersson
On the effects of gating in diffusion imaging of the brain using single-shot EPI
Magnetic Resonance Imaging, 19(8):1125-1128 (2001)

IV I Agartz, J LR Andersson, S Skare
Abnormal brain white matter in schizophrenia: A diffusion tensor imaging study
Neuroreport, 12(10):2251-2254 (2001)

V J LR Andersson and S Skare
A model-based method for retrospective correction of geometric distortions in diffusion weighted EPI
Neuroimage, 16(1) *in press* (May 2002)

TABLE OF CONTENTS

TO MY SON,
LUDVIG

Whenever a new finding is reported to the world, people say:
"It is probably not true"

Later on, when the reliability of the new finding has been fully confirmed, people say:
"OK, it may be true but it has no real significance"

At last, when even the significance of the finding is obvious to everybody, people say:
"Well, it might have some significance, but the idea is not new"

Michel de Montaigne
1533 - 1592

1. The diffusion tensor

This summary does not cover any clinical aspects of diffusion or historic remarks on diffusion MR imaging. In Sweden alone, three theses[1-3] have recently been presented, covering these aspects of the subject.

The introduction below (1.1) gives a basic presentation of the diffusion tensor without involving the MR measurement framework at all. In section (1.2) different *scalar* or *single valued* measures derived from the diffusion tensor are discussed, still without mention of how we quantify it by diffusion MR imaging. These measures are useful because they provide ways to combine the diffusion tensor, consisting of 6 unique elements, to a single quantity that can be represented as a grey scale image. In the following section (1.3) we focus on ways to visualise the diffusion tensor data in an image without breaking it down to a scalar measure. Last, in section 1.4, a brief overview of *fibre tracking* or *diffusion tractography* is given.

1.1 INTRODUCTION

Basser et al.[4,5] proposed a tensor model for diffusion. Mathematically the diffusion tensor is a symmetric and positive definite 2nd-order tensor which can be represented as a 3×3 matrix.

$$\mathbf{D} = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix} \qquad [1]$$

where the three upper and lower off-diagonal elements are identical; $D_{xy} = D_{yx}$, $D_{xz} = D_{zx}$, and $D_{yz} = D_{zy}$. The diffusion tensor, **D**, therefore contains six unique elements.

Diffusion often occurs in non-uniform systems, where there is a macroscopic flux of a certain type of particles, e.g. water molecules. This can be described using Fick's first law:

$$\mathbf{j} = -\mathbf{D} \cdot \nabla \mathrm{C}(\mathbf{r},t) = -\begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix} \begin{bmatrix} \frac{\partial \mathrm{C}(\mathbf{r},t)}{\partial x} \\ \frac{\partial \mathrm{C}(\mathbf{r},t)}{\partial y} \\ \frac{\partial \mathrm{C}(\mathbf{r},t)}{\partial z} \end{bmatrix} \qquad [2]$$

where $\mathbf{r} = [x, y, z]^T$, $C(\mathbf{r},t)$ the particle concentration, $\nabla C(\mathbf{r},t)$ the concentration gradient and $\mathbf{j}$ the particle flux density.

The minus sign in Eq. [2] indicates that the diffusing mass flows in the direction of decreasing concentration. To explain why diffusion should be modelled by a matrix (rather than a vector or scalar) let us look at the following example in 2D. I.e. for a brief moment the diffusion tensor is a 2×2 matrix. Consider a drop of ink in a homogeneous medium that at time $t = 0$ has the concentration $C(\mathbf{r}, 0) = \delta(\mathbf{r})$, where δ denotes the *Dirac function*. Assume also that the ink has exactly the same properties as the medium. Due to random thermal motion, known as *Brownian motion*, the ink will spread into the rest of the medium, as time passes. In Figure 1, that tiny drop is illustrated by the small black circle in the left panels. The particle flux density is indicated with dashed arrows and the concentration gradient with solid arrows. If the medium is isotropic, i.e. the diffusion of water molecules is equal in all directions, the diffusion tensor $\mathbf{D}$ is given by the expression in Figure 1a. With this tensor the flux goes in the direction of the concentration gradient $\nabla\mathbf{C}$. If we could freeze the molecules after a certain time, we would see a blurry drop of ink with the shape of a circle. On the other hand, if the media is anisotropic, i.e. the water molecules can move more easily in a certain direction, the diffusion tensor will look different. If e.g. the diffusion in the *y*-direction is twice of that in the *x*-direction, the tensor is given by the expression in Figure 1b. For this tensor, the flux is no longer parallel to the gradient, and the ink will spread out and form an ellipse. Last, if the medium have the same anisotropy as in Figure 1b but is rotated 30°, the diffusion tensor and the flux changes according to Figure 1c. Note again, that the direction of the flux is not equal to the concentration gradient.

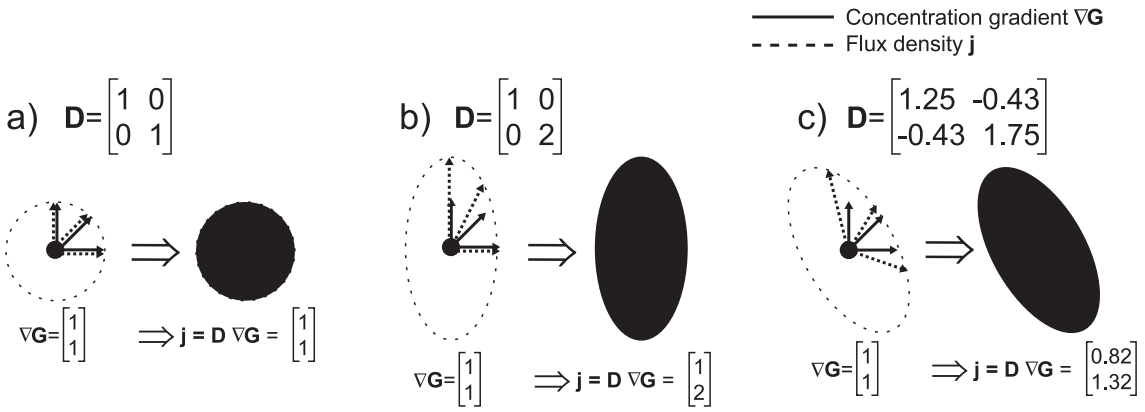

Figure 1. An illustration of Fick's first law for three different two-dimensional (!) tensors.

The 2D case was only deliberated here for motivating the matrix form of diffusion and for presenting Fick's first law in a clearer way in Figure 1. In reality, diffusion occur of course in 3D and the 3×3 tensor in Eq. [1] is the form always used in MRI.

Instead of using Fick's law it is more useful in MRI to describe the diffusion as a distribution function. Assume that we can mark a diffusing water molecule located at position **0** at time $t = 0$.

The general form of the probability of finding a water molecule at time t at position **r**, is given by[6,7]

$$\mathrm{f}(\mathbf{r},t) = \frac{1}{\sqrt{|\mathbf{D}|(4\pi t)^3}} \cdot \mathrm{e}^{-\left(\mathbf{r}^T\mathbf{D}^{-1}\mathbf{r}\right)/4t} \qquad [3]$$

where $|\mathbf{D}|$ in the denominator denotes the determinant of the diffusion tensor. It is outside the scope of this thesis to derive the proof for why the random motion of particles gives rise to a Gaussian distribution. A 1D projection of the function in Eq. [3] is indicated for a given tensor at three different times in Figure 2. For isotropic diffusion, e.g. in a glass of water, the off-diagonal elements of **D** are all zero and the diagonal elements are equal.

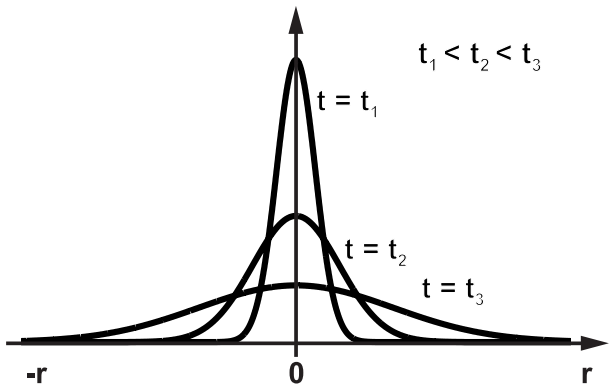

Figure 2. Probability distributions of finding a water molecule (originating at location **r** = **0** at time t = 0) at different locations at time $t = t_1$, t_2 and t_3.

$$\mathbf{D} = D\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad [4]$$

In this case the diffusion process could as well be described by a single scalar, D. That is, if we *know* that the diffusion is isotropic, then there is no need to measure diffusion in more than one direction since it will be equal to D in all other directions. To visualise the diffusion tensor in Eq. [4] we can, based on Eq. [3], create an iso-probability surface. For a given tensor, **D**, time $t = \tau$ and an arbitrarily chosen probability, p_{iso}, of finding a water molecule at position **r**, we can create the iso-probability surface by finding all corresponding **r**'s. In the following equation we have omitted the normalisation factor in Eq. [3] for simplicity.

$$\begin{aligned} & p_{iso} = \mathrm{e}^{-\left(\mathbf{r}^T\mathbf{D}^{-1}\mathbf{r}\right)/4\tau} \Leftrightarrow \mathbf{r}^T\mathbf{D}^{-1}\mathbf{r} = -\ln\left(p_{iso}\right)\cdot 4\tau \Leftrightarrow \\ & [x,y,z]\frac{1}{D}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = -\ln\left(p_{iso}\right)\cdot 4\tau \Leftrightarrow \\ & x^2 + y^2 + z^2 = -\ln\left(p_{iso}\right)\cdot 4D\tau = R^2 \end{aligned} \qquad [5]$$

Since p_{iso} can only take values between 0 and 1, the factor "$-\ln(p_{iso})$" is always positive, as is of course also τ. The last row of Eq. [5] is easily recognised as the formula for a sphere of radius R, where R depends on time τ, the scalar diffusion constant D and the chosen level of probability p_{iso}. The "diffusion sphere" is shown in Figure 3a.

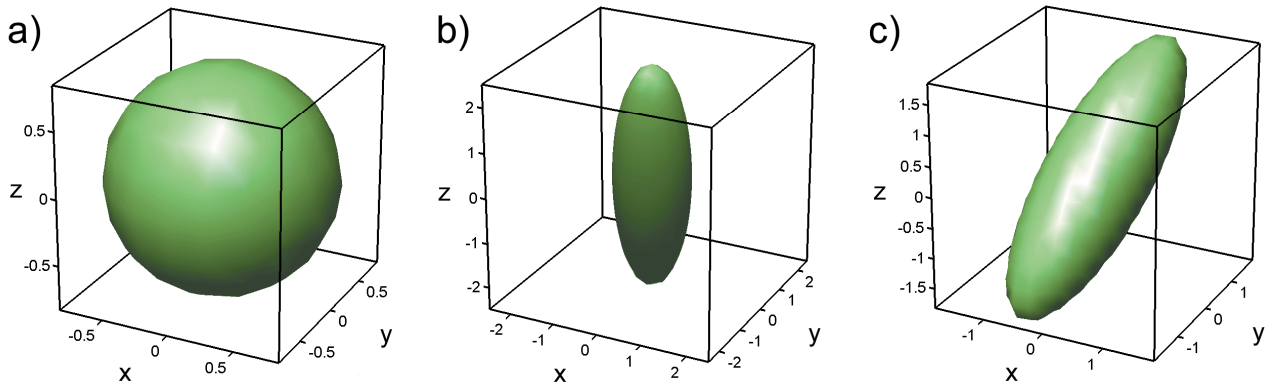

Figure 3. Three iso-probability surfaces corresponding to the diffusion tensors in Eq. [4], [6] and [9], respectively. The surfaces represent the locations in space where the probability of finding a water molecule at time τ is equal to p_{iso}. The axes are in arbitrary units but consistent between a)-c).

Let us now look at other diffusion tensors. In a homogeneous isotropic medium there is no preferred direction of molecular motion. In contrast, for an *an*isotropic medium the molecular mobility varies spatially and depends on the orientation of the medium. In biological tissues, such as brain white matter (WM), the mobility of water is higher along the WM fibre bundles than across. The fibres hinder the water movement. The exact mechanism is not known[8], but the theory of tortuosity[9,10] has gained rather wide acceptance for the experimental conditions (i.e. b values, section 2.1) normally used in diffusion MR imaging[11].

Consider e.g. the case where the root-mean-squared (RMS) displacement of water molecules is three times higher along the z direction than in the directions perpendicular to it. The diffusion tensor is then given by

$$\mathbf{D} = D\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3^2 \end{bmatrix} \qquad [6]$$

Using this tensor in Eq. [5] we now get

$$\mathbf{r}^T\mathbf{D}^{-1}\mathbf{r} = -\ln(p_{iso})\cdot 4\tau \Leftrightarrow [x,y,z]\frac{1}{D}\begin{bmatrix}1&0&0\\0&1&0\\0&0&3^2\end{bmatrix}^{-1}\begin{bmatrix}x\\y\\z\end{bmatrix} = -\ln(p_{iso})\cdot 4\tau$$
$$\Leftrightarrow [x,y,z]\begin{bmatrix}1&0&0\\0&1&0\\0&0&1/3^2\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix} = -\ln(p_{iso})\cdot 4D\tau \Leftrightarrow \frac{x^2}{1}+\frac{y^2}{1}+\frac{z^2}{3^2} = -\ln(p_{iso})\cdot 4D\tau \quad [7]$$

where the right hand side of the last row is again a constant. This is recognised as the standard equation for an ellipsoid with relative radii 1-1-3 and is depicted in Figure 3b.

In the third and last example we assume that the direction of higher diffusion is no longer collinear with our coordinate system, but rotated around the x and z axes. The rotation matrix **R** for an arbitrarily chosen rotation around the x axis of 45° followed by a rotation around the z axis of 15° is given by

$$\mathbf{R} = \begin{bmatrix}\cos(15^\circ)&\sin(15^\circ)&0\\-\sin(15^\circ)&\cos(15^\circ)&0\\0&1&1\end{bmatrix}\begin{bmatrix}1&0&0\\0&\cos(45^\circ)&\sin(45^\circ)\\0&-\sin(45^\circ)&\cos(45^\circ)\end{bmatrix} =$$
$$= \begin{bmatrix}0.966&0.183&0.183\\-0.259&0.683&0.683\\0&-0.707&0.707\end{bmatrix} \quad [8]$$

The rotated tensor will now look like this

$$\mathbf{D} = \mathbf{R}D\begin{bmatrix}1&0&0\\0&1&0\\0&0&3^2\end{bmatrix}\mathbf{R}^T =$$
$$= D\begin{bmatrix}0.966&0.183&0.183\\-0.259&0.683&0.683\\0&-0.707&0.707\end{bmatrix}\begin{bmatrix}1&0&0\\0&1&0\\0&0&3^2\end{bmatrix}\begin{bmatrix}0.966&-0.259&0\\0.183&0.683&-0.707\\0.183&0.683&0.707\end{bmatrix} = \quad [9]$$
$$= D\begin{bmatrix}1.268&1&1.035\\1&4.732&3.864\\1.035&3.864&5\end{bmatrix}$$

Performing the same calculations as in Eq. [5] and [7] for this tensor we get

$$0.970x^2 + 0.585y^2 + 0.556z^2 - 0.222xy - 0.230xz - 0.859yz = -\ln(p_{iso})\cdot 4D\tau \quad [10]$$

which is also the equation for an ellipsoid, but one whose principal axes do not coincide with the coordinate system. This rotated tensor is drawn in Figure 3c. Only when the tensor and coordinate axes coincides will the cross terms (xy, xz, yz) vanish.

These three examples show how the diffusion tensor describes both the shape of the probability distribution of the molecular diffusion, and its orientation. We can easily imagine other diffusion tensors whose ellipsoids have other shapes. For instance, if the diffusion is high in a plane and low in the remaining dimension, as in the case of water diffusion between two parallel sheets, the ellipsoid would have the shape of a discus.

In Figure 4, six unique tensor component images of an axial slice of the brain are shown according to their position in the diffusion tensor matrix. The diagonal elements are similar except for intensity differences in the brain white matter (near the centre of the image) due to anisotropy. Note that the off-diagonal images have been scaled differently compared to those on the diagonal.

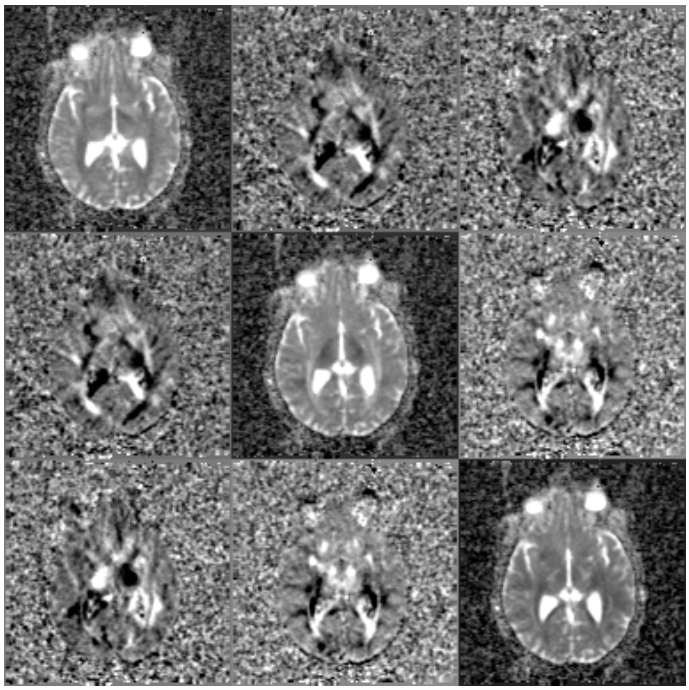

Figure 4. Actual tensor components of an axial slice in the brain arranged according to their position in the diffusion tensor. Note the symmetry of the off-diagonal elements. The grey scale level of the off-diagonal images differs from that of the diagonal ones.

1.2 SCALAR ROTATIONALLY INVARIANT MEASURES DERIVED FROM THE DIFFUSION TENSOR

In reality, tensor data are not presented as in Figure 4 since it gives little understanding to the viewer. Instead, one reduces the six-dimensional diffusion tensor information into different scalar measures that are each physically meaningful and can each be depicted as a grey scale map. Specifically, these measures, and functions of them, are independent of the orientation of both tissue structure and the image scan plane. This is referred to as *rotational invariance*. Images of rotational invariant measures, such as those demonstrated below, will have the same intensity for the same anatomical location regardless of the orientation of tissue (patient) in the scanner and of the image scan plane. In contrast, a diffusion-weighted image measured along one single direction is *not* rotationally invariant. Nor are any of the elements in the diffusion tensor. Here follows a description of three types of scalar measures calculated from the diffusion tensor. The last of them is very seldom used, but included here for completeness.

1.2.1 Mean diffusivity

We have in the previous section depicted the RMS displacement (iso-probability surfaces) for three different diffusion ellipsoids (Figure 3). The average of the radii of the ellipsoid can be considered as a "scale factor" of the diffusion ellipsoid. This is referred to as *mean diffusivity* and will hereafter be denoted by $\langle \mathbf{D} \rangle$. For example, in the CSF, $\langle \mathbf{D} \rangle$ is about 2×10^{-3} mm^2/s. For comparison, in grey and white matter it is three times lower, about 0.7×10^{-3} mm^2/s. $\langle \mathbf{D} \rangle$ is similar (it just happens to be) in grey and white matter despite the fact that the diffusion is isotropic in grey matter and highly anisotropic (cigar-shaped) in white matter. To calculate $\langle \mathbf{D} \rangle$ from the diffusion tensor one can simply average the diagonal elements

$$\langle \mathbf{D} \rangle = \frac{D_{xx} + D_{yy} + D_{zz}}{3} = \frac{\mathrm{Trace}(\mathbf{D})}{3} \qquad [11]$$

Images of $\langle \mathbf{D} \rangle$ are therefore often called Trace-maps. Another commonly used notation is $\bar{\mathbf{D}}$ (pronounced "dee-bar"). The trace of the diffusion tensor map in Figure 4 is shown in Figure 8a.

1.2.2 Diffusion anisotropy

The second scalar measure is diffusion anisotropy and comes in many flavours. Unlike $\langle \mathbf{D} \rangle$, which can only be calculated in more or less one way, several diffusion anisotropy indices have been proposed in the literature, many of them summarised in paper 1[12]. Common to all of them is that the anisotropy index should depend on how anisotropic the diffusion is, i.e. how much the diffusion ellipsoid deviates from a sphere. At the same time, anisotropy indices should be independent of the orientation of the diffusion ellipsoid – rotational *in*variance. The diffusion tensors illustrated in Figure 3b and c therefore have the same anisotropy although their orientation differs. Before continuing, we need to understand the eigenvalues and eigenvectors of the diffusion tensor, of which the former is the basis for all anisotropy indices.

λ and $\boldsymbol{\varepsilon}$ are called an eigenvalue and eigenvector, respectively, of a square n×n matrix $\mathbf{A}$ if

$$\mathbf{A}\boldsymbol{\varepsilon} = \lambda\boldsymbol{\varepsilon} \quad ; \boldsymbol{\varepsilon} \neq \mathbf{0} \qquad [12]$$

The eigenvalues of the matrix can be calculated e.g. using the QR method (see Appendix, section 5.1). For our case with a 3×3 diffusion tensor, the three eigenvalues and eigenvectors may be calculated easily by solving *the characteristic equation* (see Appendix, section 5.2). So what do they tell us? The eigenvalues of the tensor are the diffusion constants (unit: m^2/s) along the three principal axes of the diffusion ellipsoid. The

eigenvectors are vectors parallel to the axes of the ellipsoid and are defined in the image coordinate system. Hence, the eigenvalues of the two tensors in Eq. [6] and [9] are identical; namely $\lambda_1 = 9$, $\lambda_2 = \lambda_3 = 1$. A two-dimensional example with an ellipse (rather than an ellipsoid) rotated away from the frame of reference (x,y) is depicted in Figure 5. The square roots of the eigenvalues are proportional to the RMS displacements of the water molecules.

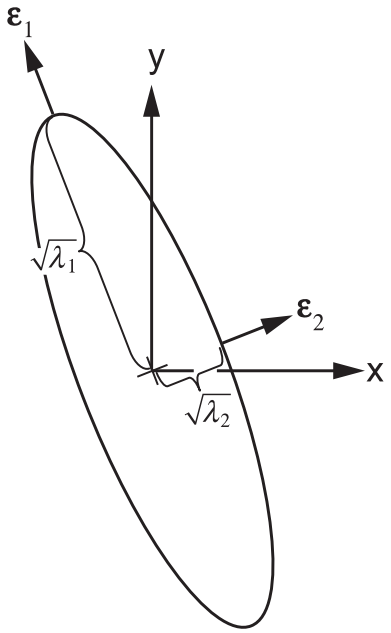

Figure 5. The relation between the shape and orientation of the diffusion ellipsoid and its eigenvalues and eigenvectors. Here shown in two dimensions for clarity. The square roots of the eigenvalues (λ_i) correspond to the RMS displacement of the water molecules, which is proportional to the radii of the ellipsoid. The eigenvectors, $\boldsymbol{\varepsilon}_i$, correspond to the principal axes of the ellipsoid expressed in the imaging coordinate system (x,y). Specifically, $\boldsymbol{\varepsilon}_1$, points in the direction of highest diffusion constant, λ_1.

After the eigenvalues of the tensor have been calculated, we can construct a rotationally invariant anisotropy index based on the eigenvalues and no longer dependent on the orientation of the tensor. A very intuitive definition of an anisotropy index is the ratio between the largest and the smallest eigenvalue as follows

$$A_{ratio} = \frac{\lambda_1}{\lambda_3} \qquad [13]$$

where λ_1 is the largest and λ_3 is the smallest eigenvalue. This index will be equal to 1 if the diffusion tensor is isotropic since $\lambda_1 = \lambda_2 = \lambda_3$. However, taking the ratio makes the definition very numerically unstable and prone to noise. A more stable index is the relative anisotropy index, RA[13]. It is defined as

$$RA = \frac{1}{\sqrt{6}} \frac{\sqrt{\sum_{i=1,2,3} \left(\lambda_i - \bar{\lambda}\right)^2}}{\bar{\lambda}} \qquad [14]$$

The numerator is, apart from a scale factor of $\sqrt{2}$, the standard deviation of the eigenvalues. The denominator is the mean diffusivity and is used to normalise with the size of the ellipsoid. Therefore RA represents the ratio of the anisotropic and the isotropic part of **D**. It is zero for isotropic diffusion and approaches 1 when $\lambda_1 \gg \lambda_2 \approx \lambda_3$. (Note:

The normalisation factor in Eq. [14] differs from the original definition where the max value of RA was $\sqrt{2}$). For the presentation of diffusion anisotropy as a grey scale map, the scale factor is of no importance. In contrast, it is important that people use the same scale (preferably between 0 and 1), and clearly state the anisotropy index used, when reporting anisotropy values in tables. Otherwise it will be harder to compare results and draw conclusions. The RA of the diffusion tensor map (Figure 4) is depicted in Figure 8b.

Another commonly used anisotropy index is the fractional anisotropy index FA[13], defined as

$$FA = \sqrt{\frac{3}{2}\frac{\sum_{i=1,2,3}\left(\lambda_i - \bar{\lambda}\right)^2}{\sum_{i=1,2,3}\lambda_i^2}} \qquad [15]$$

FA measures the fraction of the total "magnitude" of **D** that can be ascribed to anisotropic diffusion[13].

In summary, different anisotropy indices have slightly different physical interpretations. Furthermore, several groups have shown that different proposed indices differ in how strongly they are affected by image noise (see section 1.2.4).

1.2.3 Skewness

The next scalar invariant measure, skewness, has been proposed for DTI by e.g. Basser[7] and Bahn[14]. The rationale behind it is that a diffusion anisotropy index cannot distinguish between a cigar-shaped ($\lambda_1 > \lambda_2 \approx \lambda_3$) and a discus-shaped ellipsoid ($\lambda_1 \approx \lambda_2 > \lambda_3$). The skewness measure could, potentially, contribute with some information. The (unnormalised) skewness is defined as[7]

$$\text{Skewness}(\boldsymbol{\lambda}) = \frac{1}{3}\sum_{i=1,2,3}\left(\lambda_i - \bar{\lambda}\right)^3 \qquad [16]$$

For a cigar-shaped ellipsoid the skewness is positive, while it is negative for the discus-shaped ellipsoid (Note: This is the opposite of what is claimed in ref [7]). In Figure 6 the normal distribution function, which has zero skewness, is compared with two other distributions with negative and positive skewness, respectively. The skewness of the eigenvalues may be represented as a grey scale image, in which case high intensity would indicate fibres and low intensity tissue with flat membrane-like structure. The usefulness of this image is not too clear. How many true discus-shaped ellipsoid structures is one expected to find in the brain? The relationship between the eigenvalues and the scalar measures discussed is summarised in Figure 7.

In Figure 8c-d the skewness of the eigenvalues calculated from data in Figure 4 is presented. To better distinguish positive and negative skewness, only positive skewness values are shown in c) and only negative in d). Bright areas in Figure 8c indicate cigar-shaped diffusion ellipsoids. Cylinder symmetry can be appreciated especially in the occipital white matter region (*forceps major*). In Figure 8d the voxels with negative skewness are overlaid on the RA map to better show the location of the somewhat isolated discus blobs. The findings are not very conclusive.

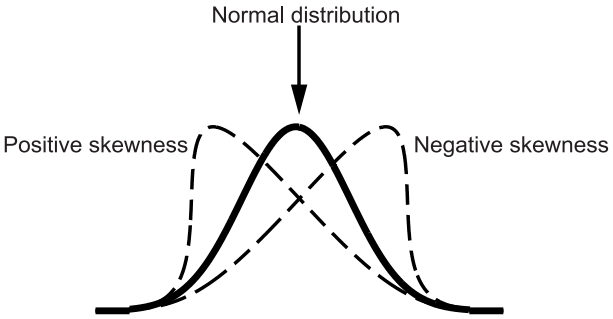

Figure 6. The Normal distribution function with zero skewness compared with two other distributions with positive and negative skewness. For the distribution between the three eigenvalues, positive and negative skewness corresponds to cigar and discus-shaped diffusion ellipsoids, respectively.

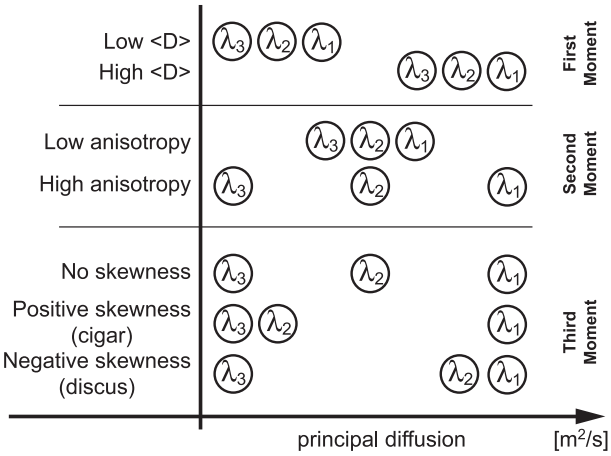

Figure 7. The relation between the eigenvalues of the diffusion tensor and the three scalar measures: mean diffusion (first moment), anisotropy (second moment) and skewness (third moment).

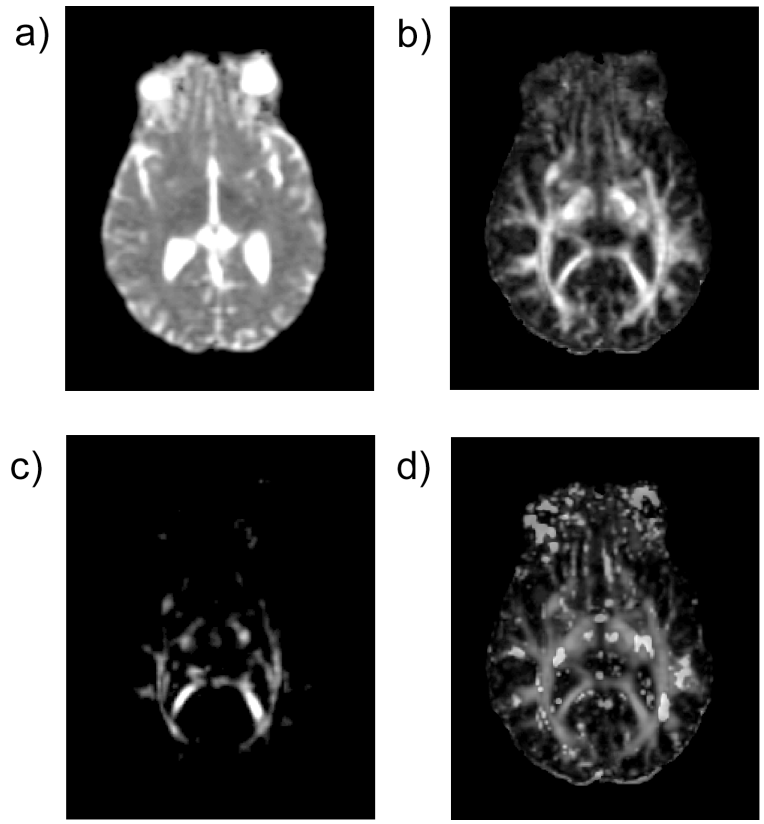

Figure 8. Images of the scalar measures of the diffusion tensor data depicted in Figure 4. a) mean diffusivity, b) RA anisotropy. The skewness data has been separated in c) positive and d) negative skewness values for better visualisation. The RA anisotropy map in b) is used as an underlay to help the viewer locate the positive skewness blobs in d).

1.2.4 Reality check I – Noise induced bias

In reality, noise is unfortunately always present in our data. This results in an overestimation of anisotropy for areas where the diffusion is isotropic. The reason is that any deviation from the identity matrix (i.e. true isotropy) will give rise to unequal eigenvalues.

In Figure 9 (Fig. 4 in paper 1) several anisotropy indices have been compared with respect to a) bias and b) sensitivity to noise.

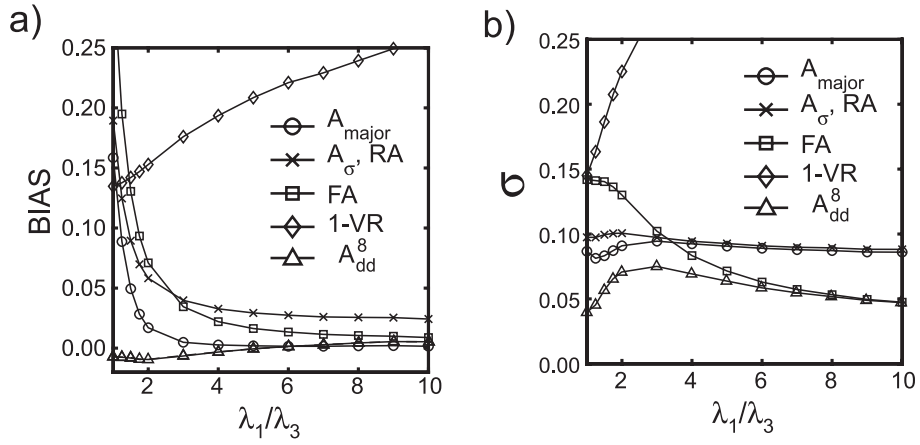

Figure 9. Monte Carlo simulations of bias and standard deviations (σ) due to noise for different anisotropy indices as a function of the ratio λ_1/λ_3 (Note: Here $\lambda_2 = \lambda_3$). Both bias and σ increase for isotropic diffusion for all indices except for the A^8_{dd}, which has nearly zero bias thanks to the comparisons with eigenvectors in neighbouring voxels.

Among the indices based on only eigenvalues, the A_{major} or RA index seems to perform better than the FA. Other groups have found that FA is the most stable index. The differences in findings may at least to some extent be explained by the handling of negative eigenvalues (see next section, 1.2.5). In Figure 9, no eigenvalue filtering has been performed. As a result of this the VR index, which is very sensitive to negative eigenvalues, differs completely from the behaviour of other indices. The index proposed in paper 1 is A^8_{dd}, which is a variant of the lattice index, LI, proposed by Pierpaoli et al.[15,16]. The advantage of this index is that its use of eigenvectors and comparisons of ellipsoid collinearity with neighbouring voxels make the A^8_{dd} index almost bias free even for isotropic structures. The drawback is that the comparison with neighbouring voxels imposes a partial volume effect that reduces the effective image resolution. The A_{dd} index is best suitable if the resolution is higher in-plane than through plane, which often is the case. In paper 1 an alternative "time-domain" A_{dd} index is also proposed that perform the same tensor collinearity calculations between different tensors measured after each other instead of between spatially adjacent tensors. The time-domain A_{dd} index is free from the partial volume effects.

1.2.5 Reality check II – Negative eigenvalues

Noise in the DWIs will perturb the elements of the diffusion tensor in such a way that it may no longer be positive definite which is equivalent to not all eigenvalues being larger than zero. A tensor is positive definite only if

$$\begin{cases} D_{xx}D_{yy} - D_{xy}^2 > 0 \\ D_{xx}D_{zz} - D_{xz}^2 > 0 \\ D_{yy}D_{zz} - D_{yz}^2 > 0 \end{cases} \quad [17]$$

which will be fulfilled if the diagonal elements are larger than the off-diagonal elements of the tensor. Negative eigenvalues have no physical meaning since negative diffusion cannot exist – the water molecules cannot be *less* than still!

It has been rather quiet about this issue in the literature even though everybody who develops a program for calculation of anisotropy needs to deal with, or at least be aware of, this problem in some way.

One could argue that it is fair to set negative eigenvalues to zero, just to barely make them physical, and then calculate the anisotropy. This will cause an undesirable bias in the anisotropy and is not really a good solution. Another approach is to exclude voxels with negative eigenvalues, i.e. we let those voxels be black. This will, however, result in quite a few removed voxels, the number of which depends on the SNR. As seen in Figure 10, the probability of getting one or more negative eigenvalues (per tensor) increases with diffusion anisotropy and with the noise level (given in % of the T_2-w image).

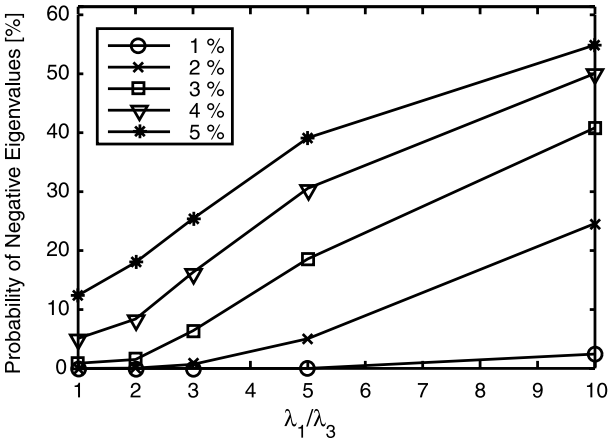

Figure 10. The probability of obtaining negative eigenvalues as function of the ratio λ_1/λ_3 (note: here $\lambda_2 = \lambda_3$) for different noise levels (defined in %). The probability is substantially higher for anisotropic diffusion and for higher noise levels (i.e. low SNR).

This implies that predominantly anisotropic regions in the image (i.e. white matter) will be jumbled with black "non-data" voxels, which cannot be distinguished from true diffusion isotropy. Instead of performing this kind of non-linear filtering, a better way is to increase the SNR so that the risk of negative eigenvalues becomes negligible. This can be done by making more measurements or by increasing the voxel size, but also by smoothing the DWI data by a moderate Gaussian smoothing filter of e.g. 2 mm.

1.3 VISUALISING THE DIFFUSION TENSOR

1.3.1 Attempts to see it all

From the tensor data (Figure 4) diffusion ellipsoids in each voxel can be calculated to transform the data into an interpretable form without losing information. For a commonly used image resolution of 128×128 pixels we would end up with 16384 diffusion tensors. It is practically impossible to draw this many ellipsoids in an image without making it cluttered and impossible to interpret with the unaided eye. Several attempts have been made to combine the information about anatomical position and the diffusion tensor information.

One way is to use an MR image as an anatomical underlay, line out a small rectangular area of interest, which is then zoomed up to show a reduced number of diffusion tensors[15,17] (see Figure 11). With this approach it is at least possible to see the ellipsoids, although they may not be easy to interpret. For example, from a certain view angle it is impossible to distinguish a spherical ellipsoid from a highly anisotropic one pointing straight towards the viewer.

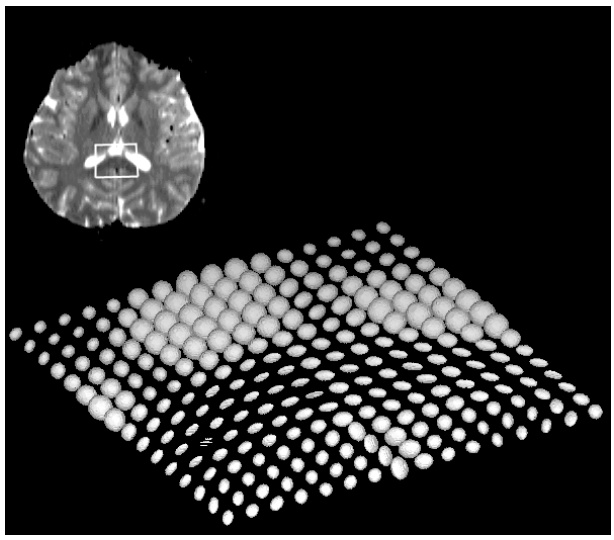

Figure 11. Visualisation of diffusion tensors in an area of the brain defined on the T_2-w image on the upper-left. The large spherical ellipsoids correspond to the high isotropic diffusion of CSF, while the cigar-shaped ellipsoids below correspond to the water diffusion in the fibres of *corpus callosum* running left-right in the image. (Courtesy of Carlo Pierpaoli. Fig 7 in ref [17])

1.3.2 Using colours

Another approach is to colour encode the diffusion tensor data. The first approach, proposed by Douek et al., was to measure diffusion along two perpendicular image axes in the coronal plane and colour code them as[18]

$$\begin{aligned} \text{Red:} \quad & \frac{D_{xx}}{D_{\tilde{x}\tilde{x}}} < 0.8 \\ \text{Green:} \quad & 0.8 < \frac{D_{xx}}{D_{\tilde{x}\tilde{x}}} < 1.2 \\ \text{Blue:} \quad & 1.2 < \frac{D_{xx}}{D_{\tilde{x}\tilde{x}}} \end{aligned} \qquad [18]$$

The intensity was proportional to the diffusion constant while the hue (i.e. colour nuance) showed the directionality of diffusion.

Since then, new colour schemes have been proposed. Defining $\mathbf{v}_{max}$ as the eigenvector in the tensor that corresponds to the largest eigenvector – i.e. the direction of highest diffusion, one could assign colour to the x-y-z components of $\mathbf{v}_{max}$. To avoid noisy eigenvectors in isotropic areas, $\mathbf{v}_{max}$ is usually weighted or filtered with some anisotropy index. This algorithm was named *directionally encoded colour* (DEC)[19,20]. The method has later been improved further by Pajevic et al., who have also made a thorough comparison of different DTI colouring approaches[21,22]. Two colour-coded coronal slices, created using the method of Pajevic et al., are shown in Figure 12.

1.4 FIBRE TRACKING

Fibre tracking, or fibre tractography as it is also named, has become a major research area within DTI during recent years. The idea is to use the directional information of $\mathbf{v}_{max}$ (previous paragraph) to find the neuronal fibre pathways. By doing so for the entire brain, ideally a three-dimensional map showing the locations of the white matter fibres could be obtained. Among many possible applications that can be envisioned, one is e.g. to see if two regions in the brain are connected with a WM fibre tract. This may be of interest in fMRI studies, to see if there is a direct connection between two activated areas.

The problem with the technique is that very high resolution and high SNR are required to obtain reliable DTI data without partial volume effects. In reality the voxel sizes of DTI data are usually not smaller than about 2×2×2 mm. The diameter of the fibre bundles is often smaller than that and partial volume effects will occur. This leads to an apparent reduction in anisotropy and an error in the estimated principal diffusion direction $\mathbf{v}_{max}$. Also, SNR is a limiting factor, causing the orientation of the principal eigenvector ($\mathbf{v}_{max}$) to deviate from its expected value. These problems must be addressed to obtain reasonably accurate tracks.

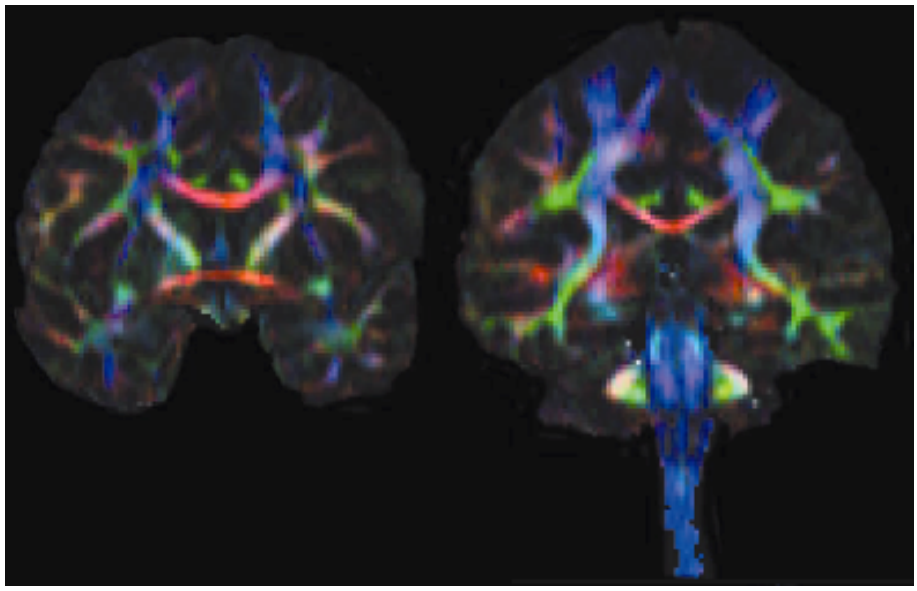

Figure 12. Coronal view of different white matter structures visualised with the directionally encoded colour scheme (DEC). The different colours denote different orientations of the eigenvector corresponding to the largest eigenvalue. (Courtesy of Carlo Pierpaoli. Modified from Fig. 7 in ref [21])

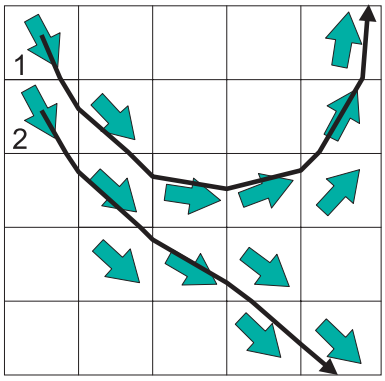

Figure 13. Illustration of the FACT (Fibre Assigning using Continuous Tracking) method. The arrows in each voxel represent the direction of the largest eigenvector. The black arrows denote the tracked path using FACT. (Courtesy of S. Mori. Fig. 1 in ref[23])

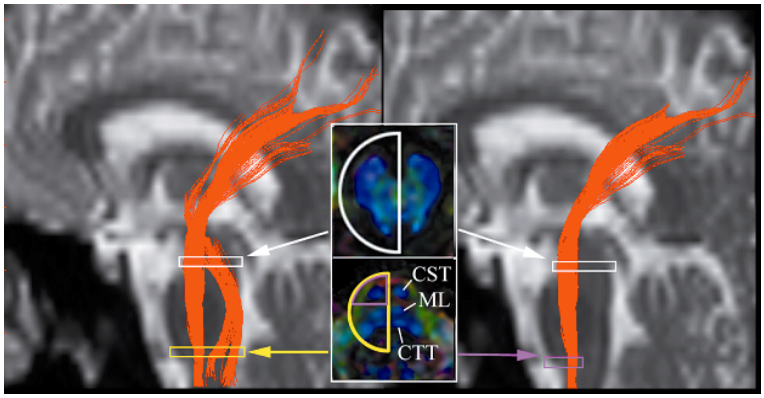

Figure 14. The fibres to be viewed are selected by defining two ROIs. Only fibres passing both ROIs are shown. The position of the lower ROI differs in a) and b), which has an impact on the result. (Courtesy of S. Mori. Fig. 1 in ref [24])

Mori et al.[25] have developed a method called FACT (Fibre Assigning using Continuous Tracking) where the track through the image volume is continuous onto the discrete DTI data. Tracking is started in a chosen voxel (or rather a ROI containing several voxels) in WM from which the direction of $\mathbf{v}_{max}$ for the ellipsoid of that voxel is followed into the next voxel. The starting point of the next voxel is now not in the centre of that voxel but rather at the intercept between the two voxels. This is illustrated in Figure 13. The tracking will go on until one of the following termination criteria is met[26]: 1) the FA anisotropy is lower than 0.25; or 2) the dot product between $\mathbf{v}_{max}$ of the current voxel and the next is less than 0.75 (remember: dot product is 0 for perpendicular vectors and ±1 for parallel vectors). The dot product termination criterion imposes limitations on the change in

direction for adjacent voxels, namely that the change of fibre direction may not exceed θ = arccos(0.75) = 41°.

Rather than defining a ROI from which fibres are tracked, one can follow fibres starting from each and every voxel within the WM. When the results are displayed, only those fibres that pass through *two* chosen ROIs are shown. An example of this is given in Figure 14, where the lower ROI location differs in a) and b), which clearly affects the results.

Other groups have also developed fibre-tracking techniques with different ways to improve tracking reliability. These include data interpolation[27,28] and regularisation techniques such as the "spaghetti" model[29,30] and the continuous tensor field approximation model[31].

For the spaghetti model, where the spaghetti is the fibre bundles (fascicles), the analogy is the following (after Poupon et al.). Before cooking, the spaghetti is stiff. After it is put in boiling water its curvature increases with time. The cooking effect on the spaghetti is assessed by the bending energy of the spaghetti by integrating the curvature. This can be extended to a whole plate of spaghetti[30], i.e. to the entire brain. The tracking method searches for the fascicle set that is the best trade-off between adherence to the DTI data and low bending energy. For each voxel, a plane spanned by the eigenvectors corresponding to the two largest eigenvalues can be defined. With this model the underlying fascicle is argued to be in the direction within this plane that leads to the lowest curvature of the fascicle.

Yet another tracking method is the continuous tensor field approach, where one creates a continuous field for each of the six unique tensor components. How this is done in detail is outside the scope of this overview but the interested reader is referred to Pajevic et al.[31] and Aldroubi et al.[32]. Basically they use B-splines as basis functions to approximate the DTI-data, achieving higher SNR at the expense of resolution. Small fibre structures with high curvature may be lost in the approximation. Additionally, the algorithm does not ensure that the tensors of the continuous field are positive definite, which means that negative eigenvalues may still occur.

The need for validation of fibre tracking techniques has resulted in studies comparing the fibre tracks with histology, e.g. in the myocardium[33-35]. Histology involves dissection, freezing and fixation etc., which will hamper the comparisons. Another group[36] has injected Mn^{2+} locally in the optic tract of rats, making it visible on T_1-weighted images. This made it possible to compare tracking results from DTI data acquired during the same experiment. They found an RMS angular error of 10-50° (quite high!), depending on the SNR, which ranged between 5 and 30. Unfortunately, from this kind of validation on rats it is hard to judge how reliable fibre tracking data is in humans.

2. Diffusion tensor imaging

2.1 DIFFUSION GRADIENTS AND THE B VALUE

2.1.1 Derivation of the b value

Diffusion in MRI is encoded by diffusion gradients, which have been incorporated in various pulse sequences. The most common type of diffusion gradient design is the Stejskal-Tanner scheme[37] where the first diffusion gradient is applied between the excitation pulse and the refocusing 180° pulse, and the second between the 180° pulse and the echo. Initially this was implemented for the spin echo pulse sequence depicted in Figure 15, but today the readout gradient (see figure) is often replaced by an EPI readout train.

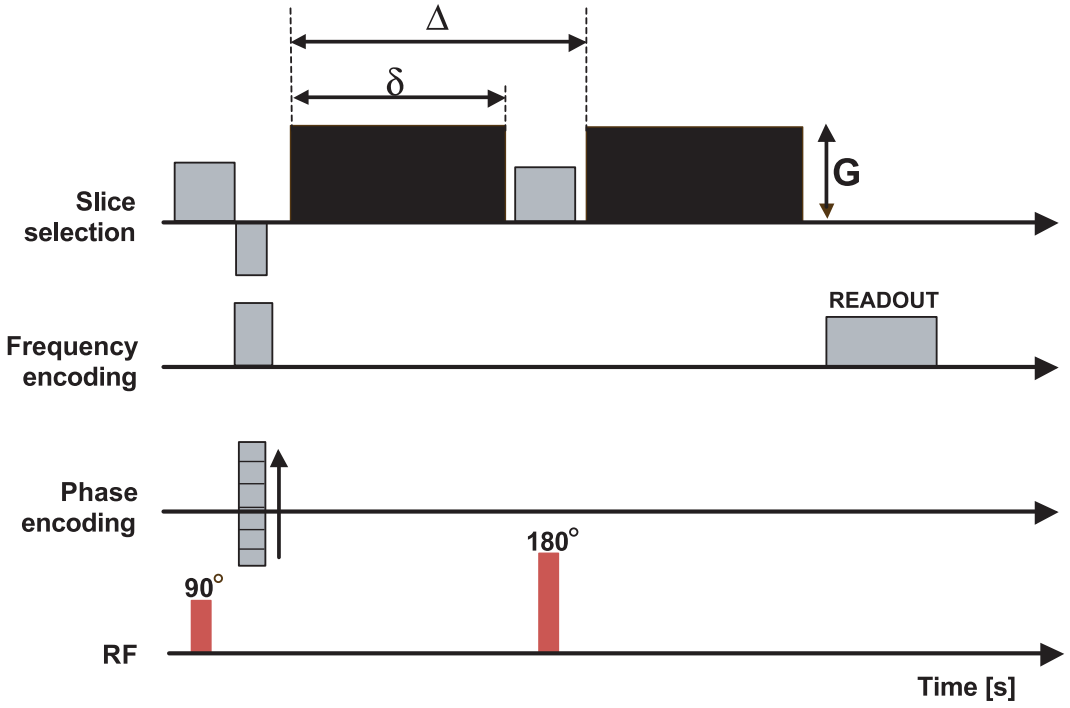

Figure 15. Stejskal-Tanner diffusion gradients (black) in a spin echo sequence. The duration (δ), amplitude (G) and time shift between the gradients (Δ) affects the amount of diffusion weighting – the "b" value. Note that the diffusion gradients may be applied on any of the three imaging axes or on a combination thereof. The readout gradient is nowadays often replaced by an EPI readout.

The diffusion gradients introduce phase shifts of the spins as a function of position. They can be applied in the x, y or z direction or a linear combination of these. If the gradients are e.g. applied in the z direction, the phase shift due to the first gradient is

$$\varphi_1 = \gamma \int_0^{\delta} G \cdot z \, \mathrm{d}t = \gamma \cdot G \cdot \delta \cdot z_1 \qquad [19]$$

where z_1 is the position of spin or molecule which is assumed to be still during the time δ when the diffusion gradient is applied. This assumption is only valid if $\delta \ll \Delta$, i.e. the diffusion process is negligible during the application of the diffusion gradients (not as illustrated in Figure 15). This requirement is usually not met for human MR scanners today where the gradient amplitude is limited to 20-50 mT/m and where a fairly large δ is needed to obtain the desired diffusion weighting. Nevertheless, this assumption makes the derivation of the *b* value reasonably simple and serves to provide a basic understanding.

After the first diffusion gradient has been switched off the 180° pulse is applied, which changes the sign of the phase ($\varphi_1 \Rightarrow -\varphi_1$). The second diffusion gradient will then produce a phase shift of

$$\varphi_2 = \gamma \int_{\Delta}^{\Delta+\delta} G \cdot z \, \mathrm{d}t = \gamma \cdot G \cdot \delta \cdot z_2 \qquad [20]$$

where z_2 is the position of the spin during the second diffusion gradient. For a "static" spin, i.e. a spin that doesn't move between excitation and the echo, $z_1 = z_2$ and the net phase of the two diffusion gradients is

$$\varphi = \varphi_1 + \varphi_2 = \gamma \cdot G \cdot \delta \left(-z_1 + z_2 \right) = 0 \qquad [21]$$

For spins that move (diffuse) during the time interval between the gradients, the phase will be non-zero. Since diffusion is a random process one has to consider the entire spin population in each voxel. Depending on the displacement path along the diffusion gradient, each individual spin *j* will get a certain net phase, φ_j. The resulting net magnetisation *M* for a voxel is the vector sum of the magnetic moments μ of each of the *N* spins within that voxel

$$M = \mu \left| \sum_{j=1}^{N} \mathrm{e}^{\mathrm{i} \cdot \varphi_j} \right| \qquad [22]$$

This is also illustrated in Figure 16 below. In the (hypothetical) absence of diffusion all $\varphi_j = k$ (constant) and the net magnetisation is maximal

$$M_0 = \mu \left| \sum_{j=1}^{N} \mathrm{e}^{\mathrm{i} \cdot k} \right| = \mu \cdot N \qquad [23]$$

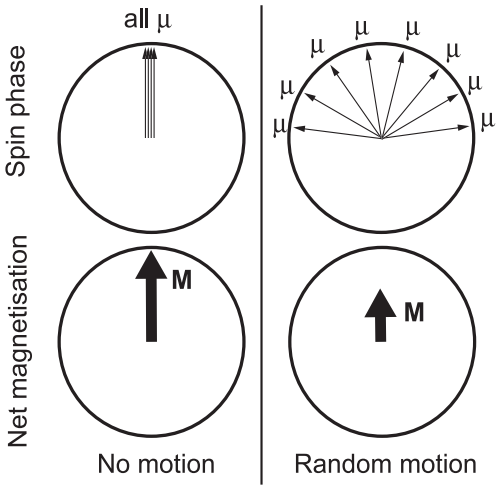

Figure 16. If no motion has occurred during the application of the diffusion gradients the phase of the magnetic moments (μ) of the spins in each voxel are still parallel and the net magnetisation **M** (i.e. the signal) is unaffected by the diffusion gradients. In contrast, if motion has occurred the spins are dephasing randomly in the transverse plane. The vector sum of all magnetic moments now results in a lower net magnetisation.

Let $P(z_2 \mid z_1, \Delta)dz_2$ be the conditional probability of finding a spin, originating at z_1 at $t = 0$, between z_2 and z_2+dz_2 at $t = \Delta$. Then, by combining Eq. [21], [22] and [23] we get

$$M = M_0 \int e^{i \cdot \gamma G \delta (z_2 - z_1)} P(z_2 | z_1, \Delta) \; dz_2 \qquad [24]$$

where $P(z_2 \mid z_1, \Delta)$ for this 1D case is given by (cf. Eq. [3], set $\mathbf{D} = D$ and $|\mathbf{D}| = D^3$)

$$P(\mathbf{r}, t) = \frac{1}{\sqrt{4\pi D \Delta}} \cdot e^{-(z_2 - z_1)^2 / 4D\Delta} \qquad [25]$$

Combining Eq. [24] and [25] gives the following

$$\frac{M}{M_0} = e^{-(\gamma \cdot \delta \cdot G)^2 \cdot \Delta \cdot D} = e^{-b \cdot D} \qquad [26]$$

Equation [26] shows the degree of signal attenuation due to diffusion as a function of gradient amplitude, duration and time interval between the two diffusion gradients. The expression for the *b* value is here $b = (\gamma \cdot \delta \cdot G)^2 \Delta$, which again is only valid if $\delta \ll \Delta$. The *b* value may be considered as the parameter that controls the amount of diffusion weighting in an image, analogous to TE's and TR's role for T_2-weighting and T_1-weighting, respectively.

A more general expression for the *b* value, which will not be derived here, is[38]

$$\frac{M}{M_0} = e^{-\left[\gamma^2 \int_0^{TE} \left(\int_0^t g(t') \, dt' \right)^2 dt \right] \cdot D} = e^{-b \cdot D} \qquad [27]$$

where TE is the echo time. This expression does not assume $\delta \ll \Delta$ and the diffusion gradients may be more than two and can be combined without invalidating the expression. For the Stejskal-Tanner diffusion gradient scheme (Figure 15), the diffusion gradient function g(t) is

$$g(t) = \begin{cases} G & , 0 \le t < \delta \\ 0 & , \delta \le t < \Delta \\ -G & , \Delta \le t < \Delta + \delta \end{cases} \qquad [28]$$

Putting Eq. [28] into [27] gives

$$b = \gamma^2 \int_0^{\Delta+\delta} \left(\int_0^t g(t')\, dt' \right)^2 dt = \ldots = (\gamma \cdot G \cdot \delta)^2 \cdot \left(\Delta - \frac{\delta}{3} \right) \qquad [29]$$

And so, we have arrived at the familiar expression for the *b* value. By letting $\delta \ll \Delta$, Eq. [29] becomes identical to the *b* factor derived in Eq. [26].

It should be noted that in the general expression for the *b* value, g(*t*) can also include the imaging gradients for an even more accurate calculation of the *b* value. However, the amplitude of the imaging gradients is much smaller than that of the diffusion gradients, and thus the diffusion effect due to the imaging gradients can be ignored (the error of omitting them is usually about 1%).

2.1.2 Optimal b value

In the absence of noise and round-off errors (images are normally stored in integer format), the choice of *b* value does not matter when calculating the ADC. However, due to noise the choice of *b* value becomes important. A too low *b* value results in a diffusion-induced signal attenuation that is comparable to the variance of the DW data. The implication is that the ADC will be estimated with low precision. On the other hand, a too high *b* value yields high signal attenuation and the signal may drop below the system noise level. This results in an underestimation for high ADCs because the system noise is higher than the true DW signal.

Since the amount of signal attenuation is determined by the product $b \cdot \mathrm{ADC}$ there is an optimal *b* value for each ADC. This optimum has been found to be $b \cdot \mathrm{ADC} = 1.1$, which corresponds to a signal drop of the DWI to $e^{-1.1} = 33\%$ of the T_2-w image[39]. Higher *b* values require longer TE, which decreases the SNR. Taking this into account, it has been found that an optimum value is about $b \cdot \mathrm{ADC} = 0.85$, or 43% of the T_2-w signal[40]. For diffusion studies in the brain the mean diffusivity is about 0.7×10^{-3} mm^2/s corresponding to an optimum *b* value of ~1200 s/mm^2. In white matter the ADC ranges from about 0.35 to 1.3×10^{-3} mm^2/s depending on the direction of measured diffusion[41]. It is not possible to optimise the *b* value for all WM voxels simultaneously, and therefore a fixed "trade-off" *b* value is normally used, common for all diffusion-encoding directions. Frequently, $b = 1000$ s/mm^2 is chosen for DTI brain studies.

2.2 ESTIMATING THE DIFFUSION TENSOR FROM DW IMAGES

2.2.1 Calculating the apparent diffusion coefficient (ADC)

It is evident from Eq. [27] that if one acquires two MR images with different b values, the unknown diffusion coefficient, D, can be calculated. If M_0 and M_1 are the signal intensities in a voxel for the two images with $b = b_0$ and b_1, respectively, D is calculated by

$$D = \mathrm{ADC} = \frac{\ln\left(M_1 / M_0\right)}{b_0 - b_1} \qquad [30]$$

The diffusion coefficient is most often denoted ADC (Apparent Diffusion Coefficient) because we really measure an overall coefficient of diffusion obtained by experiment, in which more than one diffusion mechanism is contributing to the total flux[42]. Performing the same calculations for all voxels will result in an ADC map (or image), which is a real-value map representing the ADC along the direction of the diffusion gradient.

Often the lower b value is chosen equal to zero, i.e. without diffusion gradients (T_2-weighted). In this way it is possible to re-use that image when calculating the ADC in another diffusion encoding direction.

2.2.2 Calculating the diffusion tensor from ADC data

By substituting the scalar D (or ADC measured in one direction) in Eq. [29] with the diffusion tensor $\mathbf{D}$, the expression in Eq. [27] becomes[4]

$$\frac{M}{M_0} = \mathrm{e}^{-\gamma^2\delta^2\left(\Delta-\delta/3\right)g^2\mathbf{q}^T\mathbf{D}\mathbf{q}} \qquad [31]$$

In order to calculate the six unknown elements of the diffusion tensor we need to measure the ADC in at least six non-collinear diffusion gradient directions, i.e. a minimum of 7 images (one T_2-w + 6 DW). Usually the magnitude of the diffusion gradients in a diffusion scheme is chosen to be identical for all directions. If they are not, $g^2\mathbf{q}^T\mathbf{D}\mathbf{q}$ in Eq. [31] should be replaced with $\mathbf{g}^T\mathbf{D}\mathbf{g}$. In this context, however, Eq. [31] will however hold. Assume now that we have acquired and calculated the ADC in n non-collinear directions ($n \geq 6$), each one denoted $\mathbf{q}_\mathbf{i} = [q_{ix},\ q_{iy},\ q_{iz}]^T$ $(i = 1 \ldots n)$, where $\|\mathbf{q}\| = 1$. The obtained ADC measured along the i:th direction has the following relation to the diffusion tensor $\mathbf{D}$

$$\mathrm{ADC}_i = \mathbf{q}_i^T\mathbf{D}\mathbf{q}_i \qquad [32]$$

Now we want to obtain **D**, or more specifically the six unique elements of **D**, namely $\mathbf{d}=\begin{bmatrix} D_{xx} & D_{yy} & D_{zz} & D_{xy} & D_{xz} & D_{yz} \end{bmatrix}$. We know all $\mathbf{q}_i$ from the design of our diffusion pulse sequence and we have in each voxel measured n ADC values, here defined as $\mathbf{y} = [\mathrm{ADC}_1 \;\; \mathrm{ADC}_2 \ldots \mathrm{ADC}_n]^T$. Expanding Eq. [32] to an equation system with n rows (one for each direction i) we can construct a matrix **X** that is solely dependent on the directions $\mathbf{q}_1..\mathbf{q}_n$ and gives us a simple matrix expression of the mapping between the unknown **d** and the measured **y**.

$$\mathbf{y} = \mathbf{X}\mathbf{d} + \mathbf{e} \qquad [33]$$

where **X** is defined as

$$\mathbf{X} = \begin{bmatrix} q_{1x}^2 & q_{1y}^2 & q_{1z}^2 & 2q_{1x}q_{1y} & 2q_{1x}q_{1z} & 2q_{1y}q_{1z} \\ \ldots & & & & & \\ q_{nx}^2 & q_{ny}^2 & q_{nz}^2 & 2q_{nx}q_{ny} & 2q_{nx}q_{nz} & 2q_{ny}q_{nz} \end{bmatrix} \qquad [34]$$

and where **e** is the error term. The estimate of the unknown **d**, denoted $\hat{\mathbf{d}}$, can be calculated from

$$\hat{\mathbf{d}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} \qquad [35]$$

Here $\hat{\mathbf{d}}$ is the voxel-wise least squares estimate of the tensor data from the measured data, **y**.

Analogous to fitting a straight line using only two data points, using only six directions ($n = 6$) will give us no information on how well the tensor fits to the ADC data (contained in **y**).

2.3 CALCULATING THE MEAN DIFFUSION FROM ADC DATA

Consider now that we use *less* than six directions. Then the elements of the tensor cannot be uniquely determined. However, to calculate $\langle \mathbf{D} \rangle$, the mean diffusivity, full tensor information is not needed. According to Eq. [11], it is sufficient to calculate the diagonal elements, D_{xx}, D_{yy} and D_{zz}. Two diffusion gradient schemes that allow calculation of the mean diffusion are illustrated in Figure 17, where the thick solid lines denote the actual direction of the diffusion gradient. In Figure 17a, diffusion encoding is made along the imaging axes, with gradients

$$\mathbf{q}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{q}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \mathbf{q}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \qquad [36]$$

To prove that averaging of the ADC values measured along these directions will give $\langle \mathbf{D} \rangle$, we put Eq. [36] into [32]

$$\mathrm{ADC}_1 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = D_{xx} \qquad [37]$$

Analogously, $\mathrm{ADC}_2 = D_{yy}$ and $\mathrm{ADC}_3 = D_{zz}$. Figure 17b depicts the tetrahedral gradient set, proposed by Conturo et al.[43], where the gradients vectors are

$$\begin{aligned} \mathbf{q}_1 &= \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad \mathbf{q}_2 = \frac{1}{\sqrt{3}} \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} \\ \mathbf{q}_3 &= \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix} \quad \mathbf{q}_4 = \frac{1}{\sqrt{3}} \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix} \end{aligned} \qquad [38]$$

Again, using these vectors in Eq. [32] gives

$$\begin{cases} \mathrm{ADC}_1 = \frac{1}{3}\left(D_{xx} + D_{yy} + D_{zz} + 2\left(D_{xy} + D_{xz} + D_{yz} \right) \right) \\ \mathrm{ADC}_2 = \frac{1}{3}\left(D_{xx} + D_{yy} + D_{zz} + 2\left(D_{xy} - D_{xz} - D_{yz} \right) \right) \\ \mathrm{ADC}_3 = \frac{1}{3}\left(D_{xx} + D_{yy} + D_{zz} + 2\left(-D_{xy} - D_{xz} + D_{yz} \right) \right) \\ \mathrm{ADC}_4 = \frac{1}{3}\left(D_{xx} + D_{yy} + D_{zz} + 2\left(-D_{xy} + D_{xz} - D_{yz} \right) \right) \end{cases} \qquad [39]$$

$$\Rightarrow \frac{1}{4}\sum \mathrm{ADC}_i = \frac{1}{3}\left(D_{xx} + D_{yy} + D_{zz} \right) = \langle \mathbf{D} \rangle$$

Note that, since the displacement probability of diffusion is symmetric, it does not matter if the gradient is applied along the solid line or in the opposite dashed-dotted line.

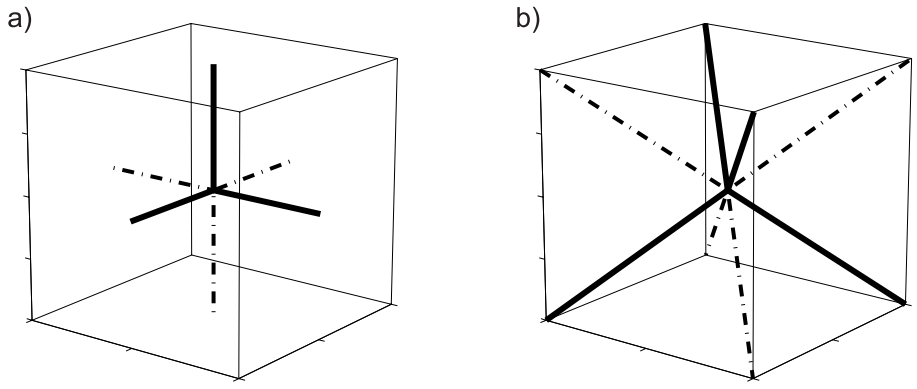

Figure 17. Diffusion schemes used to measure the mean diffusion. In a) the diffusion is measured along the three imaging axes (solid lines) whereas the tetrahedral scheme shown in b) uses diffusion encoding in four directions. For each of the four directions full amplitudes of the gradients along the imaging axes are employed simultaneously. Provided that the imaging axes and the physical gradient axes of the scanner coincide, this gives a three times higher *b* value for a given diffusion time. Since the diffusion is symmetric, it does not matter if the gradients are applied all along the solid lines or all along the dashed-dotted line.

The advantage of using the tetrahedral scheme is that, for a given diffusion gradient duration δ, the effective gradient in each direction is $\sqrt{3}$ times higher since the physical *x*, *y* and *z* gradients are applied simultaneously. This results in a 3 times higher *b* value per δ — or vice versa, for a desired *b* value, the δ can be reduced and thereby minimise the TE, which in turn increases the SNR. This is only true, however, if the image scan plane is normal to one of the physical gradients. For oblique scan planes, the improvement in *b* value efficiency is smaller.

2.4 DIFFUSION SCHEMES

A diffusion scheme is, as mentioned earlier, a set of *n* diffusion gradient directions in the directions $\mathbf{q}_1 ... \mathbf{q}_n$. To what extent does the choice of directions and the number of directions affect the accuracy of the estimation of the tensor? This issue has been dealt with in paper 2[44]. Diffusion schemes proposed in the literature[17,27,40,45-48] as well as new ones have been compared. The main conclusions from paper 2 are:

1) Given the same total number of measurements, it is better to measure the diffusion in many directions rather than do more averages of DWIs in fewer directions. Because of noise, the eigenvalues are not rotationally invariant. The accuracy of the eigenvalues and diffusion anisotropy varies with the direction of the tensor relative to the image plane (Figure 18a-b). The smaller number of directions used, the larger the variations.
2) For any *n*, the optimum direction set is to distribute them isotropically in space. This was first suggested by Jones et al.[40] but confirmed here.
3) The condition number of the matrix **X** (see Eq. [34]), which is solely determined by the chosen diffusion scheme, is to a large extent proportional to the noise properties of that scheme. In other words one can predict if a certain scheme will produce good or bad images just by looking at the condition number.

A discussion of the condition number and other possible indices of "good diffusion schemes" are deliberated in the Appendix (5.3).

In Figure 18, the uncertainty of FA, $\sigma(FA|\theta,\varphi)$, as a function of the orientation of the ellipsoid ($\lambda_1 = 10$, $\lambda_2 = \lambda_3 = 1$) is shown for three diffusion scheme. With 30 directions any orientation of the tensor ellipsoid results in the same variance of FA, where as for $n = 6$, it does not.

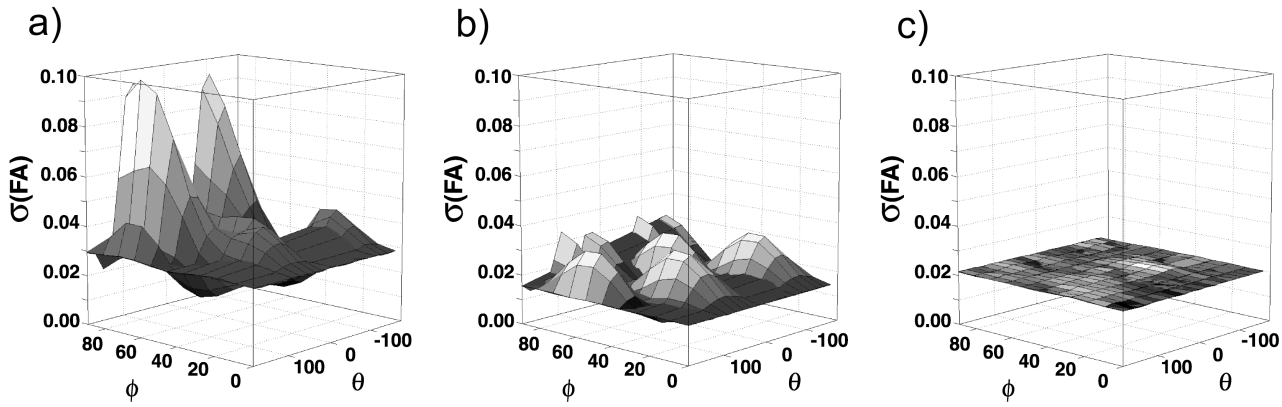

Figure 18. The standard deviation of the FA anisotropy index vs. the orientation of the diffusion tensor ellipsoid, defined by the angles θ and ϕ for three diffusion schemes. The same number of measurements has been used for each point in each of the three plots. a) Scheme #1, b) scheme #10 and c) scheme #8 in paper 2. In a) and b) the number of diffusion encoding directions is 6. In c) the number of directions is 30. The directions are isotropically distributed in b) and c), which has been found to be optimal. Scheme #1 was found to be the most noise-sensitive of the diffusion schemes investigated in paper 2.

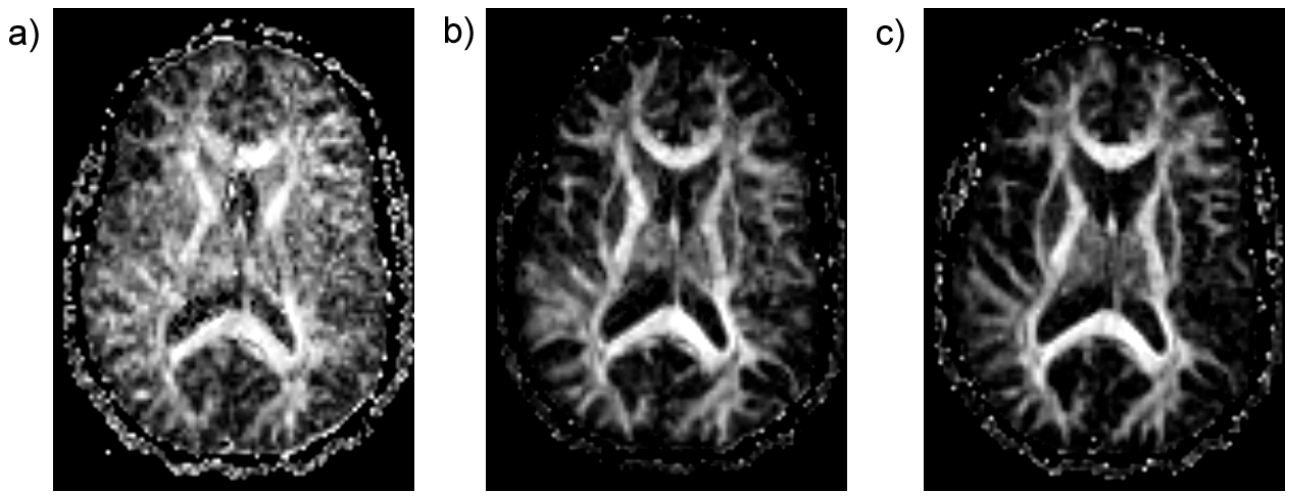

Figure 19. FA maps of an axial slice of the brain scanned with the same diffusion schemes as in Figure 18 (see caption of this figure for more details).

3. Problems with single-shot pulse sequences

One may divide pulse sequences into two groups - multi-shot and single-shot pulse sequences. Multi-shot sequences include spin echo (SE), fast spin echo (FSE), gradient echo (GRE) and multi-shot echo planar imaging (MS-EPI) – requiring several excitations to fill the k-space.

SE, FSE and GRE are practically the only pulse sequences used clinically for conventional imaging. They provide good resolution, image contrast and signal-to-noise (SNR) and are virtually free from geometric distortions. Typical scan times are a few minutes for a set of about 20 slices. The reason they are not so commonly used today for diffusion imaging is primarily the long scan times for a complete DTI-scan and the troublesome ghosting artefact caused by patient motion, described further in section 4.1. With the multi-shot EPI sequence and FSE with long echo trains the scan time can be reduced. However the ghosting problem remains.

Single-shot means that after a single excitation all lines/echoes in k-space are acquired. An image is obtained within about a tenth of a second, making these types of pulse sequences ideal for studying dynamic processes, or acquiring several image volumes with different diffusion gradients as in DTI. Single-shot (SS) sequences include SS-spin echo EPI (further on referred to as SS-EPI), SS-spiral and SS-FSE. The SS-EPI sequence is currently the predominant diffusion imaging technique. However, SS techniques are not without drawbacks — there is no free lunch.

3.1 IMAGE BLURRING, PSF

With single-shot imaging techniques the acquisition time needed to fill the k-space is on the order of 70 ms for an image with a matrix size of 128×128 using SS-EPI or SS-spiral methods. The acquisition time is about 400 ms with the SS-FSE pulse sequence. For comparison, the acquisition time for a single echo/line is about 8 ms with conventional sequences. The T_2 (or T_2^* for GRE) decay during such a long readout time will lead to a significant signal reduction from the first to the last echo. This leads to a widened point spread function (PSF), which especially for SS-FSE implies a blurry image. The long acquisition time for data in the phase encoding direction is equivalent to a very low "pseudo"-receiver bandwidth (prBW) in this direction that has further implications for the image quality (see sections 3.2, 3.4, 3.5, 4.4). The blurring of the SS-FSE image can be appreciated in Figure 20b, where the phase encoding direction is left-right. This effect may be even worse for a T_2-w SS-FSE in a diffusion experiment, because the potential need to spoil the stimulated echo signal component implies an even faster signal decay (see 4.2).

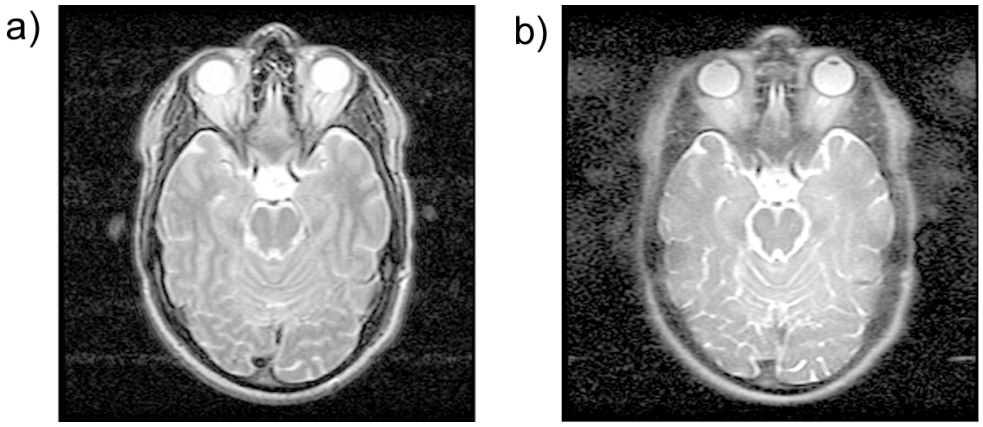

Figure 20. Image blurring due to T_2-decay during the readout train. a) FSE sequence with only 8 echoes gives substantially less blurring than b) SS-FSE with 72 echoes (256x128 resolution with half Fourier imaging). The blurring effect would be even higher if the stimulated echo (STE) component, which is inclined by a longer T_1/T_2-mix decay time, had not coherently added to the signal. For DW-SS-FSE the STE interfere with the spin echoes and causes artefacts (see section 4.2).

Off-resonance effects, such as eddy currents and magnetic susceptibility variations can lead to image blurring, which is characteristic for DW-Spiral imaging. These off-resonance effects can give rise to different image artefacts in DW-EPI. Since DW-spiral is less commonly used, in the following sections the discussion of off-resonance effects will be focused on DW-EPI.

3.2 SUSCEPTIBILITY ARTEFACTS

3.2.1 Problem

Inhomogeneities in the static magnetic field as well as local field changes due to susceptibility effects near tissue-air boundaries in the body give rise to an undesired spatially varying magnetic field. For conventional imaging the field inhomogeneity effects are low or moderate. On the other hand, for single-shot techniques with low pseudo receiver bandwidth in the phase encoding direction, this is a severe problem near tissue-air boundaries (not to mention the case where metals are present in the image volume!).

For SS-FSE imaging each echo in the echo train is refocused using an RF pulse that cancels the effect of magnetic field inhomogeneities. This is not the case for SS-EPI, where the phase and frequency encoding gradients are used to acquire the entire k-space data in a single echo. The long readout and the reduced bandwidth in the phase encoding direction for SS-EPI can lead to geometric distortions. The image artefacts become more severe with higher image resolution due to the further increased readout length and reduced prBW. In Figure 21, the effect of susceptibility is depicted for b) 4-shot EPI and c) single-shot EPI. An image acquired with conventional FSE sequence is also given in Figure 21a for comparison. The effect of the reduced prBW in c) compared to b) is evident as shown by the arrows.

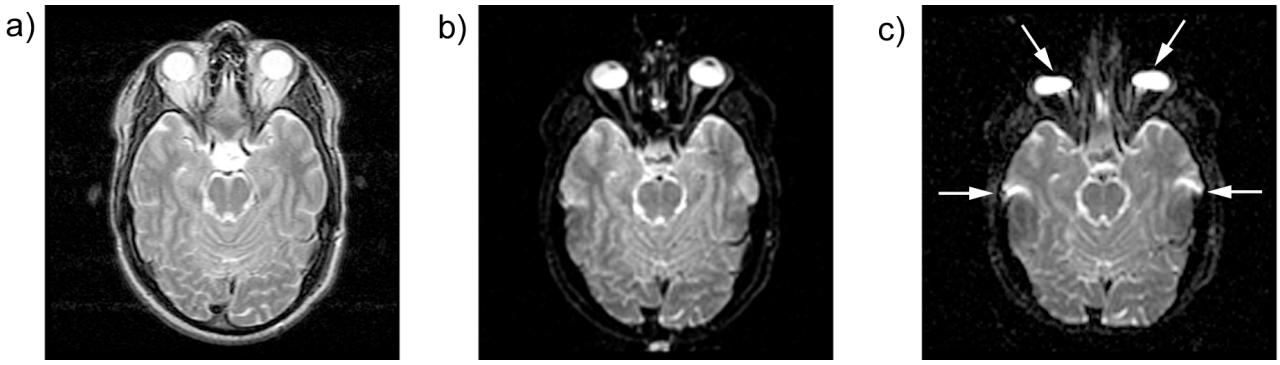

Figure 21. Susceptibility artefacts at tissue-air boundaries. a) FSE is insensitive to these small local changes in the magnetic field. In b) a 4-shot SE-EPI sequence have been used and the distortions are moderate while in c) they are substantial as indicated with the arrows.

3.2.2 Solutions

There are at least three types of solutions that can be applied to diffusion imaging, one of which has been developed "in-house".

The first approach relies on measuring the static magnetic field inhomogeneities, ΔB_0, with a field mapping technique from which the displacements for each voxel can be calculated. The ΔB_0 field map can be calculated by acquiring two (or more) GRE images with different TE. From the phase difference between these two images the ΔB_0 is given by[49]

$$\Delta B_0(x,y,z)=\frac{\Delta\varphi(x,y,z)}{\gamma\cdot\Delta TE} \qquad [40]$$

Due to the phase wrapping at 2π, the phase map, $\Delta\varphi(x,y,z)$, must first be unwrapped[50]. This is particularly difficult in regions with low SNR and high susceptibility, which are unfortunately also the areas where it is most important to obtain an accurate field map. Yet another obstacle is tissues or areas surrounded by air or bone. If there is no brain tissue pathway between the areas along which reliable unwrapping can be performed, there may potentially be an $n\cdot 2\pi$ phase difference between these areas that cannot be resolved. If the phase wrapping can be resolved properly, a displacement map can be readily calculated using ΔB_0 from Eq. [40]. The "undistorted" image can then be constructed by resampling the distorted image at the points indicated by the displacement map. Alternatively, more sophisticated methods can be used that automatically perform a *Jacobian modulation* and correct also for intensity variations and effects in the frequency encoding direction[51].

Reber et al.[52] have also shown that it is possible to calculate a field map from EPI images with the same distortions as the actual EPI data. High reliability in the field maps was obtained thanks to the use of several EPI images with a range of TEs allowing the phase

unwrapping to be performed in the time domain rather in the spatial domain. Thereby, the unwrapping problem for isolated tissues is eliminated.

Another approach that does not rely on field maps is the use of multi-reference scans. As a calibration step, N_{ph} (number of phase-encoding steps) EPI readout trains are collected in the absence of phase blips during the readout. For each of the N_{ph} scans a phase encoding gradient placed *before* the EPI train is stepped as in the conventional GRE manner. By doing so, phase maps for *each echo* in the EPI train are obtained. Based on this kind of data, Wan et al.[53] and later also Chen et al.[54,55] were able to correct both intensity and geometric distortions. In fact, this method (Chen et al.) can correct *any* off-resonance effect, including chemical shift (see 3.4) and eddy current effects (see 4.4). This approach nonetheless has its drawbacks too. The scan time of the calibration step is $N_{ph} \cdot \mathrm{TR}$. For 30 slices with 128×128 resolution the minimum TR becomes about 10 seconds yielding a calibration scan time of over 20 minutes. No doubt, patient movements during this period will also have an impact on the result. The long calibration time may still be justified for high quality and lengthy dynamic EPI based studies in e.g. fMRI and DTI.

The third approach, which uses neither field maps nor pre-scan calibration, was first suggested by Chang et al.[56] for conventional SE images, followed by Bowtell et al. for EPI[57]. Unlike the other methods, the correction is performed on the reconstructed magnitude images. The idea is that the distortions differ depending on whether the k-space is sampled bottom-up or top-down, which only is a matter of the sign of the phase encoding blips. By acquiring data twice using both k-space trajectories, one gets in the first image stretching in areas that are compressed in the second image, and vice versa.

We have recently developed Bowtell's method further. An example of an image corrected for susceptibility using our method is given in Figure 22, with a) k-space sampled bottom-up, b) top-down and c) the corrected image, based solely on the two magnitude images in a) and b).

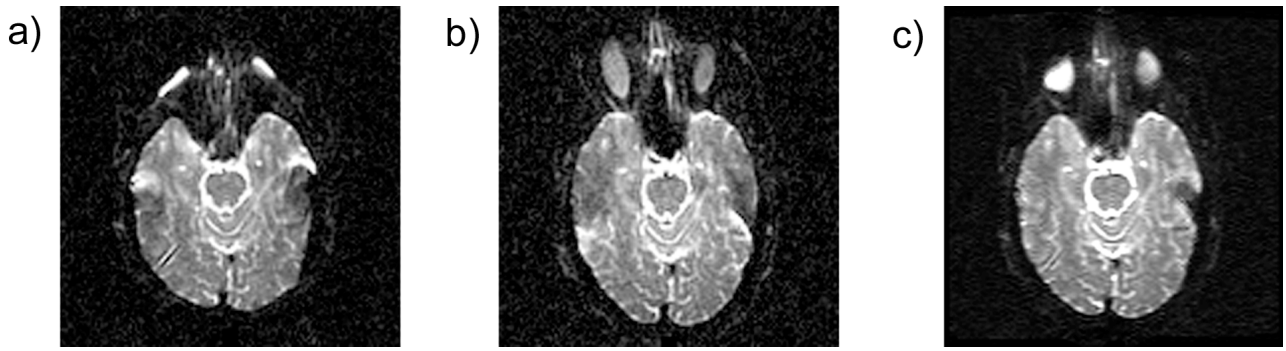

Figure 22. An axial slice of the brain scanned with SS-EPI (res. 128×128). In a) k-space has been sampled "bottom-up" while in b) it has been sampled top-down. No other changes of scan parameters were made. It is evident that areas that are squeezed in a) are stretched in b). Our susceptibility correction method, based only on these two images, gives the corrected image in c).

3.3 SNR

Each line in k-space is acquired in 1-2 ms for SS-EPI and about 4 ms for SS-FSE. This is a prerequisite to keep the above-mentioned acquisition times down to a minimum. This implies that a high receiver bandwidth (rBW) has to be used since $\text{rBW} \propto 1/t_{\text{acq}}$, which in turn leads to a decrease in SNR, which is proportional to $1/\sqrt{\text{rBW}}$. Therefore, a feature common to all single-shot sequences is a comparatively lower SNR. Because of this reduction in SNR, as well as to avoid excessive geometric distortions, the image resolution for single-shot sequences is seldom chosen larger than 128×128. This lower in-plane resolution is usually not a limitation since the through-plane resolution, i.e. the slice thickness, is generally lower anyway. Typical values for diffusion weighted SS-EPI are an image FOV of 24 cm, resolution of 128×128 and slice thickness of 3-5 mm, resulting in a voxel size of 1.8×1.8×3-5 mm.

3.4 CHEMICAL SHIFT IN THE PHASE ENCODING DIRECTION

Fat and water differ in resonance frequencies by 3.5 ppm, which equivalent to about 220Hz at 1.5T. This will be a source of chemical shift artefacts in the image, i.e. fat and water signals originating from the same point will end up in different voxels in the image. The effect is inversely proportional to the rBW. Unlike conventional sequences, where the chemical shift artefact is manifest as a shift of the fat signal in the frequency encoding direction, the fat signal is shifted in the phase encoding direction in EPI. This is because in the frequency encoding direction the rBW may be e.g. ±62.5kHz or higher, which is substantially higher than the chemical shift between water and fat, whereas in the phase encoding direction the prBW is typically on the order of 1 kHz. In the case of FOV = 23 cm, the shift of the fat signal will be 23 cm × 220 Hz / 1 kHz = 5.1 cm!

There are two ways to avoid the fat signal: either a narrow bandwidth fat-saturating RF pulse can be applied before each excitation, or one can use a spectral spatial RF excitation pulse that only excites the water. Two SS-EPI images acquired using these imaging parameters are shown in Figure 23. Fat saturation has been employed in a) but not in b).

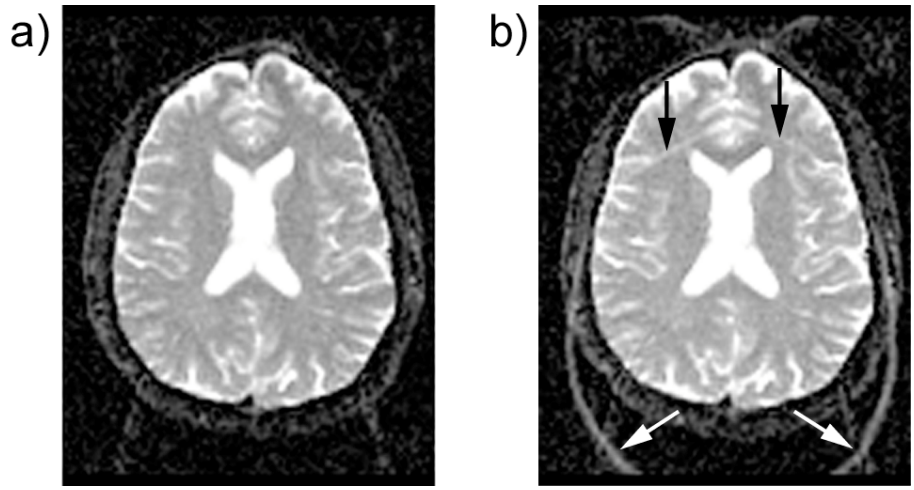

Figure 23. SS-EPI a) with and b) without fat-saturation. The arrows in b) indicate the fat signal belonging to the scalp.

3.5 MAXWELL TERMS

The static magnetic field and the imaging gradients determine the local magnetic field. This relation is normally denoted as

$$B(x,y,z) = B_0 + G_x x + G_y y + G_z z \qquad [41]$$

This is not completely true, however. The Maxwell equations require the overall magnetic field to satisfy the following conditions

$$\begin{cases} \nabla \bullet \mathbf{B} = 0 \\ \nabla \times \mathbf{B} = \mathbf{0} \end{cases} \qquad [42]$$

To fulfil the second requirement[58], Eq. [41] has to be rewritten as[59,60]

$$\begin{aligned} B(x,y,z) &= B_0 + G_x x + G_y y + G_z z \\ &+ \frac{1}{8B_0} G_z^2 \left(x^2 + y^2\right) + \frac{1}{2B_0}\left(G_x^2 + G_y^2\right) z^2 \\ &- \frac{1}{2B_0}\left(G_y G_z yz + G_x G_z xz\right) + \ldots \end{aligned} \qquad [43]$$

Here the spatial dependence of order $1/B_0^2$ and above has been neglected. For scan parameters used in multi-shot imaging these Maxwell terms are negligible, but for single-shot imaging with strong readout gradients in combination with low prBW in the phase encoding direction they may be significant. Norris and Hutchison[58] have also demonstrated Maxwell effects at low B_0 field strength.

3.5.1 Problem

When a readout gradient is applied in an *axial* EPI scan, G_z is zero and the only "extra" term in Eq. [43] is

$$B_C = \frac{1}{2B_0}\left(G_x^2 + G_y^2\right)z^2 \qquad [44]$$

This term is quadratically dependent on the slice location z and is zero for the centre slice in the image volume. This leads to three effects[59]:
1) *Nyquist ghosting* due to shift of odd and even echoes in the EPI readout train. Appreciable ghosting will be present if

$$|z| \geq \frac{1}{g_{ro}}\sqrt{\frac{\Delta\nu\ B_0}{\gamma N_{ro}}} \qquad [45]$$

where $\Delta\nu$ is the full receiver bandwidth, N_{ro} the number of data points in the readout direction, g_{ro} the readout gradient amplitude and γ is the gyromagnetic constant in Hz/T. Using common DW SS-EPI scan parameters (B_0 = 1.5T, resolution = 128×128, g_{ro} = 11.6 mT/m, $\Delta\nu$ = ±62.5kHz = 125 kHz) ghosting will occur if the slice is acquired at z = 51 cm from the centre! This is far outside the usable imaging volume.

2) *Parabolic shift*
The shift from the true location of the object in the phase encoding direction will have a quadratic dependence on the z distance from the isocentre. The shift will be

$$\delta_{pe}(z) = \frac{\gamma FOV_{pe} g_{ro}^2}{2B_0}\left(\frac{2}{3}t_1 + t_2\right)z^2 \qquad [46]$$

where t_1 is the readout ramp time and t_2 the readout plateau time for one echo (Figure 24). With the same EPI imaging parameters as before and with t_1 = 128 µs and t_2 = 1024 µs, the shift will be $\delta_{pe}(z) = 0.49 \cdot z^2$. For 35 slices of 4 mm, $|z|$ will for the first and last slice be $35 \cdot 4/2$ = 70 mm, which according to Eq. [46] will result in a shift of about 2.5 mm. This effect is shown in Figure 25, where a spherical water phantom has been scanned with SS-EPI using the scan parameters above. An image scanned with a conventional SE sequence is shown for comparison (left). At isocentre (Figure 25a), the position of the phantom in the phase encoding direction is the same, whereas at z = 70 mm (Figure 25b) the phantom is shifted by about 6 mm. This is higher than predicted from Eq. [46]. However, at 70 mm off-centre bad shimming also contributes to shift in the image (and scaling), making it difficult to isolate the Maxwell effect.

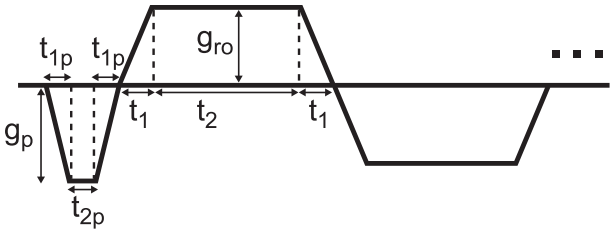

Figure 24. The first part of the EPI readout train with the parameters given in Eq. [46]-[47] outlined.

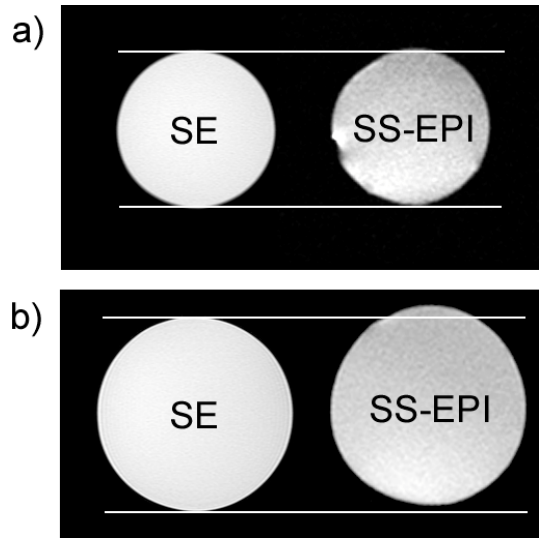

Figure 25. A spherical water phantom scanned with a conventional SE sequence and an SS-EPI sequence at two locations. a) z = 0 and b) z = 70 mm from isocentre. The white lines indicate the top and bottom edge of the phantom scanned with the SE sequence. In b) the phantom scanned with SS-EPI is shifted due to the Maxwell parabolic shift.

3) *Signal attenuation*

The phase evolution in k-space caused by the Maxwell terms is z^2 dependent. This means than for any slice thickness $\Delta z > 0$, there will be a phase dispersion causing signal attenuation (similar to diffusion!). The signal attenuation at slice position $z = z_0$ is

$$\left|\frac{S_{z=z_0}}{S_{z=0}}\right| = \frac{1}{\Delta z}\left|\int_{z_0-\Delta z/2}^{z_0+\Delta z/2} e^{\,i\varphi(z)}\, dz\right|$$

where [47]

$$\varphi(z) = \frac{\gamma}{2B_0}\left(\begin{array}{l}\left(M-\frac{1}{2}\right)\left(\frac{2}{3}t_1+t_2\right)g_{ro}^2 \\ +\left(\frac{2}{3}t_{1p}+t_{2p}\right)g_p^2\end{array}\right)z^2$$

where M is the echo number when the central k-space line is sampled. t_{1p} is the ramp time and t_{2p} the plateau time for the pre-phasing gradient before the readout. g_p is the gradient amplitude (Figure 24). The effect is very different depending on whether full or partial k-space is sampled. If full k-space is sampled with 128 phase encoding steps then M = 64. For a slice thickness of Δz = 4 mm, the signal attenuation becomes about 14% at slice position z = 70 mm. In most cases, however, about 60% of k-space is sampled to minimise blurring effects and to obtain more slices per TR. Then M may for example be equal to 8 and the Maxwell induced signal attenuation is then only 0.2%.

3.5.2 Solutions

The first Maxwell effect can be ignored because it lacks significance for common scan parameters. The parabolic shift and the signal attenuation might not.

The parabolic shift can be solved by implementing a correction method in the pulse sequence that adjusts the receiver phase during the echo train[59]. Alternatively, if an undistorted image volume is available, the shift can be corrected by rewriting the image registration software so that it in addition to rigid body adjustments also models a z^2-dependent parameter.

The signal attenuation effect is easily avoidable by using fractional k-space sampling, with early collection of the central k-space lines. This is what most DTI users do anyway, so this problem may not be frequently encountered. Nevertheless, *if* full k-space is sampled this effect must be corrected by pulse sequence modifications, since attenuation due to phase dispersion cannot be recovered after the signal has been acquired – analogous to signal attenuation for diffusion imaging. We will not go into this in detail, but the interested reader is recommended to read the paper by Zhou et al.[59].

4. Image artefacts in diffusion imaging

4.1 PATIENT MOTION: IMAGE GHOSTING USING MULTI-SHOT PULSE SEQUENCES

The rationale for using multi-shot pulse sequences for diffusion imaging is, as discussed earlier, that the geometric distortion is much less and the SNR is higher than in single-shot pulse sequences. This allows the use of higher image resolution with low geometric distortions.

4.1.1 Problem

For multi-shot pulse sequences the k-space is acquired in several segments, ultimately one line at a time as for the conventional spin echo sequence. Due to patient motion, sudden phase shifts in k-space occur between adjacent phase encoding lines acquired after different excitations. For non-diffusion weighted imaging, the motions of the patient must be quite large to introduce a phase shift sufficient to cause image ghosting. However, when diffusion gradients are applied, even minor motion will cause severe phase shifts. Fixation tools and cooperable subjects are usually not sufficient and the acquired DWI becomes practically useless. As an example a conventional DW-SE image of a healthy volunteer, instructed to remain as still as possible, is shown in Figure 26b, with the T_2-w SE for comparison (Figure 26a).

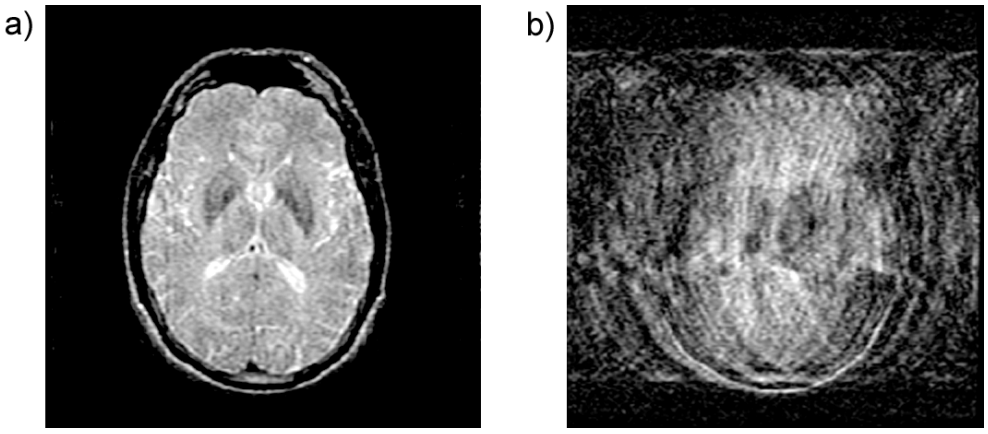

Figure 26. a) T2w SE image and b) corresponding DW-SE image with ghosting artefacts due to minor head motions during the acquisition. Images acquired on a healthy volunteer instructed to remain as still as possible.

4.1.2 Solutions

To correct for these phase shifts in k-space, different types of *navigator echo* techniques have been employed for diffusion imaging. Use of navigator echoes was first pioneered by Anderson and Gore[61]. They added an echo before the actual echo and the phase encoding gradient (Figure 27a). The navigator echo will contain phase information induced by the patient. The actual echo, which is preceded by the phase encoding gradient, will contain phase information from both motion and the phase encoding gradient. The phase of the navigator echo can then be subtracted from the phase of the actual echo. Provided that the two echoes have been acquired closely in time, the phase shift induced by motion in the frequency direction will be greatly reduced.

As a natural next step to correct for motion also in the phase encoding direction, Butts et al.[62] introduced a pair of orthogonal navigator echoes (Figure 27b) applied sequentially in the phase and frequency encoding direction. This allowed the correction of motion also in the phase encoding direction. Unlike the correction along k_x in k-space, the correction in the k_y direction requires interpolation and regridding before Fourier transformation. The pair navigator technique was not guaranteed to find the correct phase shifts, however.

A year later Butts et al.[63] improved the technique by implementing a 2D spiral navigator echo prior to the multi-shot EPI readout (Figure 27c). The spiral navigator echo sampled a circle in the centre of k-space with a radius long enough to include the location of maximum signal in the presence of motion. With this technique, the actual k-space centre could be found in cases where the orthogonal navigator echoes failed. A variable density gridding kernel was also proposed to reduce the residual ghosting due to regridding errors in the k_y direction.

Based on the same idea of sampling an area in the centre of k-space, Atkinson[64] implemented a 2D EPI based navigator echo, this time with the navigator placed after the multi-shot EPI readout (Figure 27d).

The other way to avoid ghosting artefacts is to use a single-shot sequence where there is no need for navigator echoes. This is what most groups tend to do.

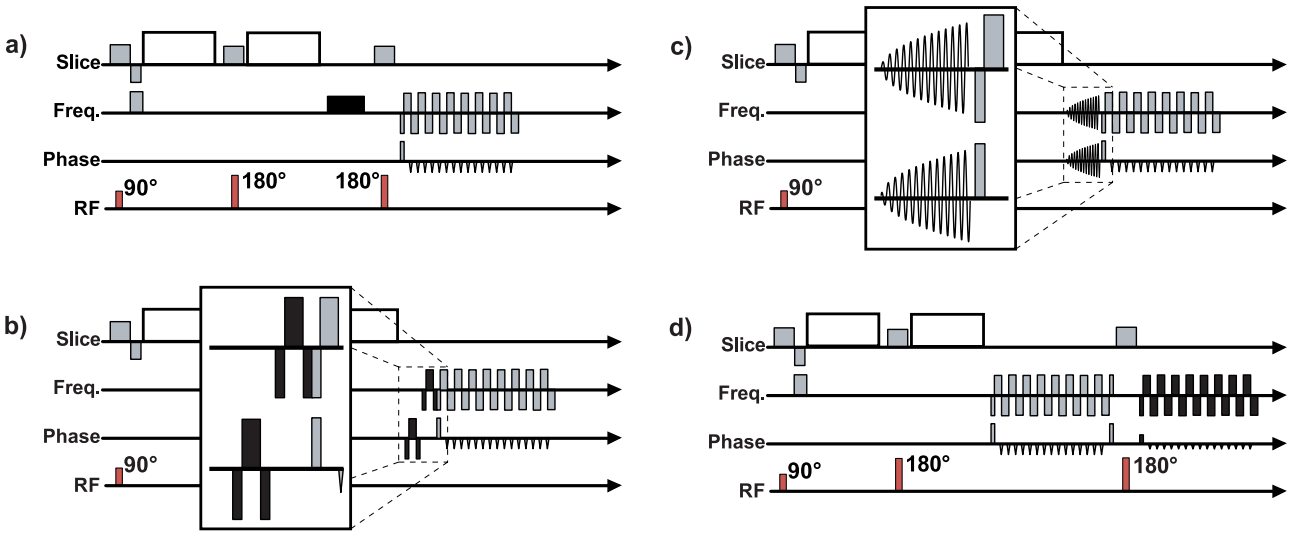

Figure 27. Different navigator techniques for correcting phase errors leading to image artefacts as in Figure 26b. The navigator echo gradients are illustrated in black. a) A single navigator correcting for motion in one direction. b) Two orthogonal navigator echoes able to correct for in-plane translational motion. c) Spiral navigator echoes scanning the centre of k-space – an improved way of correcting in-plane motion. d) A navigator EPI readout placed after the actual EPI readout. As in c), the central part of k-space is sampled.

4.2 SPIN ECHOES AND STIMULATED ECHOES IN SS-FSE

The SS-FSE sequence is robust and has none of the geometrical distortions that plague SS-EPI. Another property of SS-FSE is that, apart from the normal spin echoes, stimulated echoes (STE) will also be present. STE will occur if the refocusing pulses used in the readout train have flip angles <180°. Even if 180° pulses are employed, the existence of smaller flip angles at the edges of the slice profile creates STE. Because of this a portion of the transverse magnetisation is placed along the longitudinal axis after each refocusing pulse. After three refocusing pulses this effect contributes to the signal increasingly in subsequent echoes. If the 90°-excitation pulse is phase shifted by 90° relative to the following refocusing pulses, the phase of the transverse magnetisation and the first refocusing pulse will be parallel[65,66] (the Carr-Purcell-Meilboom-Gill (CPMG) condition). In that case, coherent addition of transverse magnetisation is obtained of the (T_2-decaying) spin echoes (SE) and the (T_1/T_2-mix-decaying) stimulated echoes (STE). This causes the signal decay more slowly allowing for long echo trains.

4.2.1 Problem

When diffusion gradients are inserted in an SS-FSE (or a FSE) sequence the CPMG condition is very hard to fulfil. This is because patient motion or minimal mismatch between the two diffusion gradients will change the phase of transverse magnetisation. When the CPMG condition is not fulfilled the phase of the SE and the STE are non-coherent, causing unpredictable signal interference that leads to image artefacts as in Figure 28b.

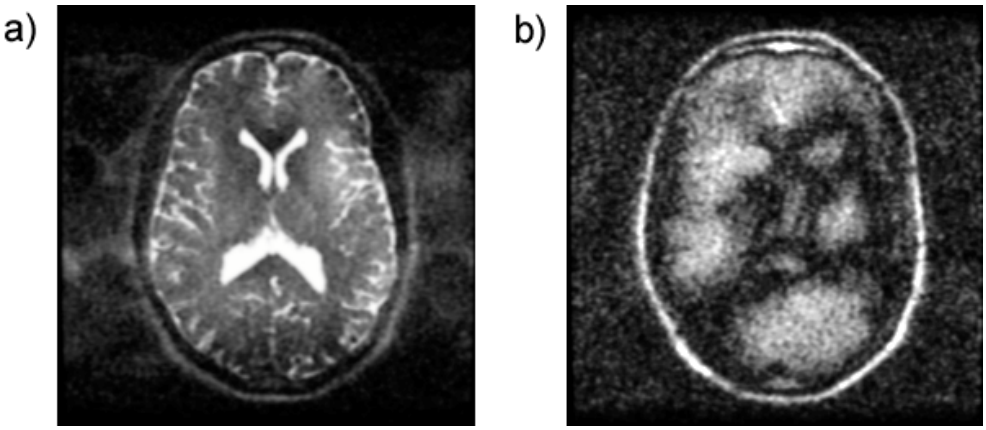

Figure 28. The DW-SS-FSE sequence with a) b = 0 s/mm^2 and b) b = 500 s/mm^2. As soon as the diffusion gradients are applied, as in b), the phase between the stimulated echoes (STE) and the spin echoes (SE) are no longer coherent. This yields artefacts in form of irregular dark stripes in the image.

4.2.2 Solutions

Different strategies have been used to avoid the effects of phase differences of the SE and STE components. One is to remove the STE component by changing the crusher gradient areas for consecutive refocusing pulses[67] or by using asymmetric crushers[65,68] (Figure 29a). This reduces the signal by about 50%[69], which together with the increased PSF of SS-FSE results in rather blurred and noisy images.

A second strategy is to split the SE and STE in time during the readout[65,66,70] (Figure 29b). The k-space then consists of two halves in the frequency encoding direction, containing signal from the SE and STE, respectively. These can be reconstructed individually and the magnitude images in the image domain finally added phase-insensitively. This requires, however, that either the receiver bandwidth or the readout time be doubled. The first alternative decreases the SNR by ~30% and the second increases the echo train length, which also yields SNR reduction as well as more pronounced blurring.

Recently, Norris et al. proposed an online correction method for FSE[69] (should be applicable to SS-FSE too) that did not cause these SNR reductions. Using orthogonal navigator echoes followed by correction gradients preceding the readout train, the phase was monitored and corrected before the start of the readout (Figure 29c). In this way, the CPMG condition was possible to fulfil.

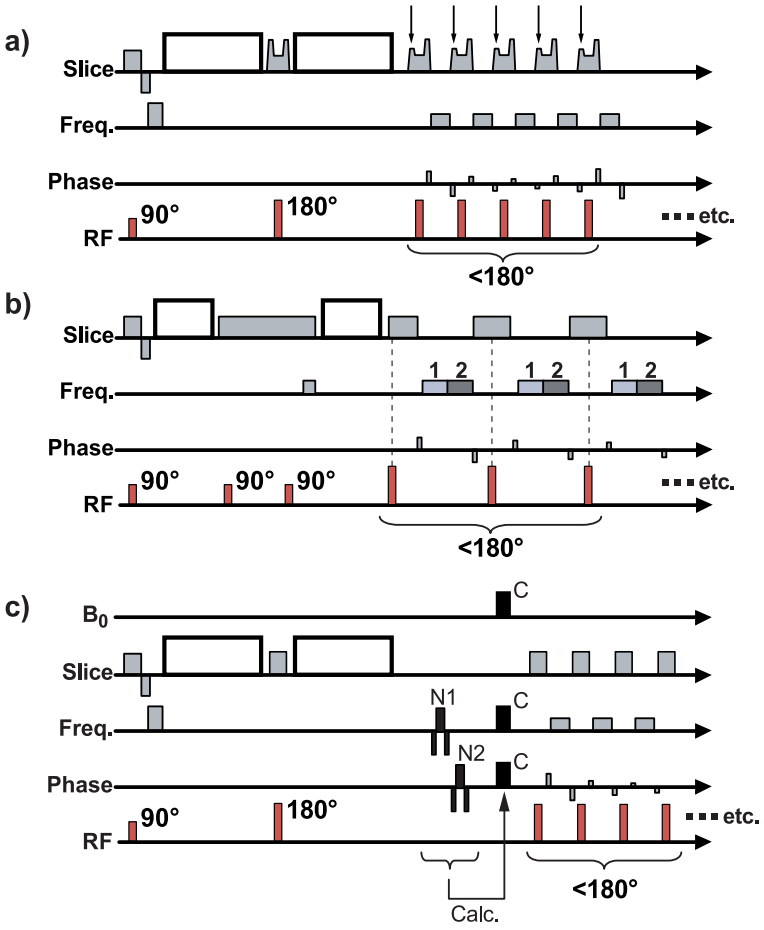

Figure 29. Methods to avoid the destructive interference between STE and SE in SS-FSE. a) The use of asymmetric crushers (see arrows) spoils the STE component. This results in a signal loss of about 50%. b) Separating the SE and the STE in time for each echo to store them in separate k-spaces. The reconstructed magnitude images are then averaged together. This requires either a doubled receiver bandwidth or an almost doubled total readout time, both leading to reduction in SNR. c) An online correction method to fulfil the CPMG condition before the readout train starts, to make the SE and STE coherent despite the diffusion gradients. Two navigator echoes (N1 & N2) are used to detect the phase error after the diffusion gradients. From this data the amplitudes of the correction gradients (C) are calculated.

4.3 BRAIN MOTION: SIGNAL LOSS DUE TO BRAIN MOTION

4.3.1 Problem

By using single-shot sequences such as EPI, ghost artefacts due to patient motion discussed in section 4.1 can be avoided without the need to use navigator echoes. This, and the high acquisition speed, are the main reasons why this sequence has been so popular for diffusion MR imaging. But even if the effects of patient motion can be avoided, there are other sources of motion that may still lead to artefacts in the DWI.

The brain might not normally be thought of as a moving tissue like other organs in the body, e.g. the heart or the abdomen. However, due to the arterial pulse wave, the brain experiences a vertical force during systole. This force causes primarily the part of the brain below the corpus callosum to move. Peak velocities of about 1-1.5 mm/s and displacements on the order of 0.1-0.13 mm have been reported[71]. Brain motions that are present during the operation of the diffusion gradients will cause a pseudo-diffusion effect. Considering that signal attenuation in diffusion imaging occurs for molecular displacements on the order of a few microns, these motions are an order of magnitude larger and will result in substantial signal loss. The brain motions are present only during a part of the heart cycle. In Figure 30 (Fig. 1 in paper 3), one slice for an image volume

acquired 18 times consecutively is presented. In this figure no gating (see 4.3.2) has been carried out and a fixed TR of 8 s has been used. One notices that there is a fair chance of obtaining a single DWI without this artefact. However, for a whole brain DTI scan, with 6 or even more DWI volumes and 20-40 slices in each volume, it is very likely that these artefacts will be present in the acquired data. This will lead to bias and a bad estimate of the diffusion tensor.

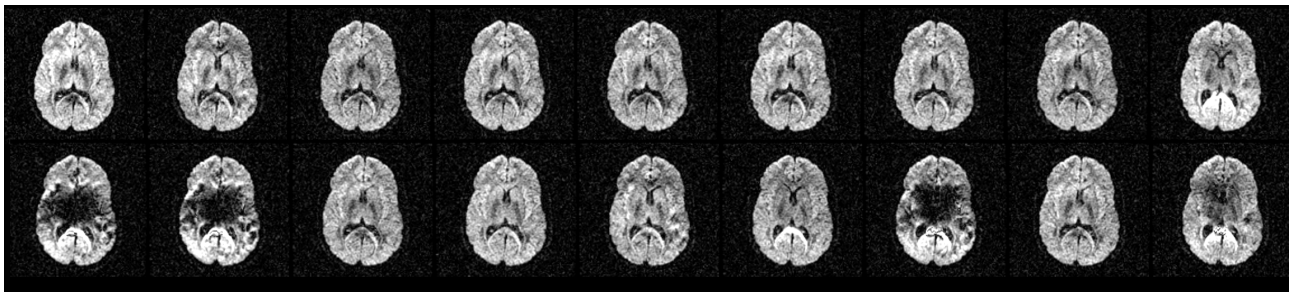

Figure 30. Brain motions in DW-SS-EPI. One slice (of an image volume) acquired 18 times consecutively, separated in time with a TR of 8 sec. Since the TR and the heart rate are not synchronised, slices will be acquired randomly in the heart cycle, which will jeopardise the image quality. Of these 18 images about 3-4, or ≈ 20%, suffer from severe brain motion artefacts. The diffusion gradient has been applied in the *z* (through-plane) direction for all images.

4.3.2 Solutions

All MR systems can be run synchronously to the heartbeats so that each echo in the pulse sequence is acquired in the same phase of the heart cycle. This is called *gating* or *triggering*. By this means, it is possible to avoid imaging during systole when the brain is moving.

One can measure the so-called R-wave in the ECG, either directly using electrodes on the chest (cardiac gating), or indirectly by measuring the pulse wave in the finger (peripheral or pulse gating). The latter is often used because it is easier to handle. From the onset of the trigger point (either the R-wave or the pulse wave) one can choose a trigger delay (TD) after which imaging starts. Additionally, because the heart rate may vary, one also needs to define a trigger window (TW), usually expressed in percent of the R-R interval time, during which the scanner waits for the next pulse. This is schematically illustrated in Figure 31.

Greitz et al. have shown that brain motions occur within the 200-250 ms after the R-peak[71], corresponding to approximately 20-25% of the R-R interval time. This is in agreement with Figure 30 where 20-25% of the images suffer from brain motion artefacts.

As delineated in Figure 31, one needs to use different TDs depending on whether cardiac or pulse gating is employed. This has implications for the available imaging time between two heartbeats. If pulse gating is used, a minimum TD may be used and the major part of brain motions are in the TW interval. For DW SS-EPI the acquisition is about 150 ms for one slice. Based on practical experience, it is possible to acquire three slices per heartbeat with pulse gating provided that the pulse rate is below about 70 bpm. For almost any

patient, it is possible to acquire two slices per R-R interval, unless the pulse rate exceeds about 100 bpm. As an example, the effective TR for a DTI scan with 30 slices should be no more than 15 heartbeats (usually written as TR = 15×R-R ≈ 12-15 sec).

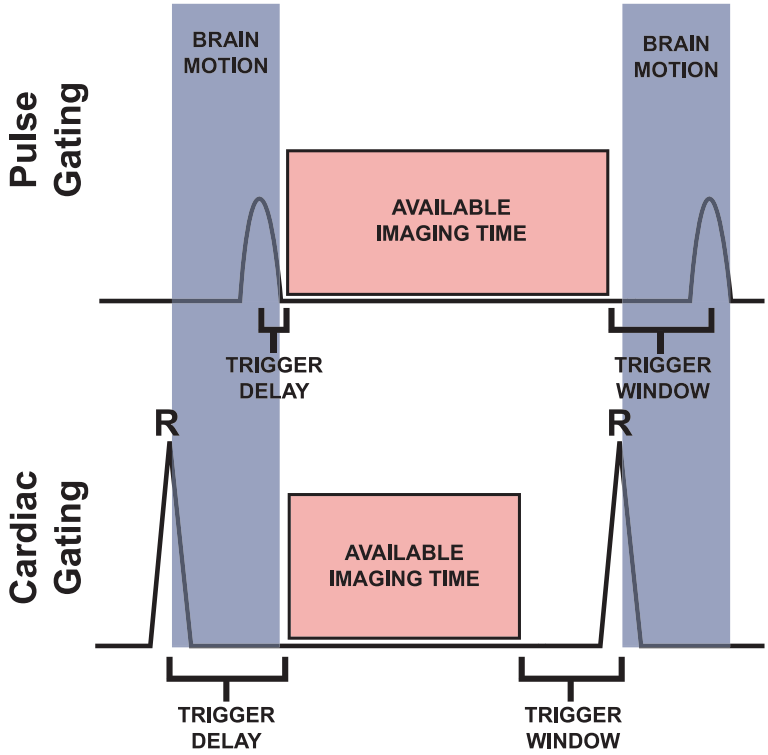

Figure 31. Gating techniques and the impact on the available imaging time (AIT). Brain motions occur very soon after the R-wave of the heart and last for about 200-250 ms. Depending on if cardiac or pulse gating is employed the delay (TD) after the trigger point and when imaging starts differ. Imaging may then be performed until the trigger window (TW) starts, during which period the scanner stops acquiring data and waits for the next pulse. The AIT is longer for pulse gating, which in most cases leads to a shorter total scan time.

A common argument against gating is that the scan time is increased. This is because fewer slices per second may be acquired because of the TD and TW during which no scanning can be performed. Nevertheless, bearing in mind the artefacts in Figure 30, the results from paper 3 have shown that the variance and bias are much smaller when triggering is performed[72] for areas in the brain inferior to *corpus callosum* where brain motions predominantly occur. Consequently, for a given variance in the DTI data, the number of measurements needed may be about 2-3 times higher for the ungated case. Thus, the total required scan time becomes in fact *longer* if gating is *not* used, despite its higher acquisition rate.

Furthermore, even when the number of measurements (and the SNR) approaches infinity, the brain motion will still induce significant bias in the DTI data yielding an overestimation of mean diffusion and anisotropy.

4.4 EDDY CURRENTS

When a gradient is applied, the change in magnetic field, d**B**/dt, induces an eddy current (EC) which in turn creates an extra magnetic field. This will for a "normal" trapezoidal gradient cause a smoother gradient waveform than the expected one. To get the desired waveform, *pre-emphasis* is used, where the gradients are ramped up in a way that compensates for this (Figure 32).

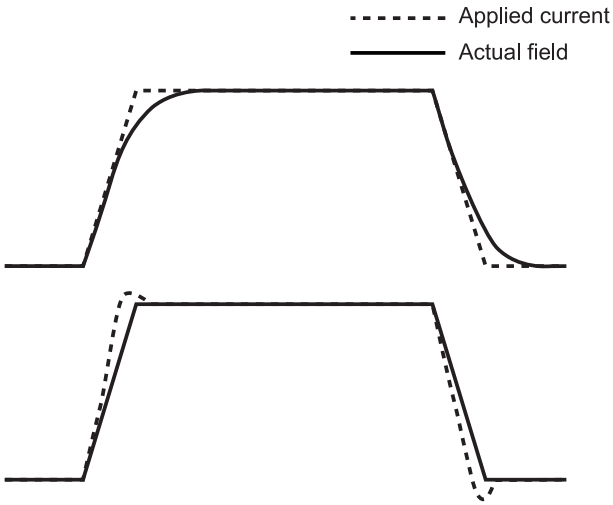

Figure 32. Eddy current pre-emphasis. a) The dashed line denotes the current applied in the gradient coil. Due to eddy currents during the ramp times, the actual field gradient obtained differs from the nominal (solid line). b) Employing gradient *pre-emphasis* the applied current during the ramps is adjusted so that the actual magnetic field gradient becomes close to the nominal shape indicated by the dashed line in a).

The eddy currents decay multi-exponentially with different time constants. There are fast decaying components such as the ones shown in Figure 32 and longer ones whose half-life may be on the order of a second. Many manufacturers have a pre-emphasis system with three time constants whose amplitudes are adjusted by the routine service program.

4.4.1 Problem

The pre-emphasis system provided by the manufacturer is typically not sufficient to correct for the ECs induced by the very strong diffusion gradients applied in the EPI sequence. Especially the long-term eddy currents are difficult to correct, partly due to inter-slice effects. The actual effects of eddy current components depend on the proportions of amplitudes and time constants (how fast they decay). Eddy currents present during the acquisition (50-100 ms for SS-EPI), will be superimposed on the imaging gradients and cause the k-space trajectory to deviate from its desired path. Strong and fast-decaying eddy currents *during* the acquisition will cause non-equidistant sampling of k-space, resulting in a blurry image.

For eddy current components that are approximately constant during the acquisition, the effects in k-space and in the image domain are linear and can be corrected using post-processing methods on the reconstructed magnitude images (see 4.4.4). In the rest of this section we assume constant or long-term EC and see how this affects the k-space and the image.

Residual eddy currents in the slice selection direction will induce a slice-wise change in the static magnetic field (Note: Non-zero EC gradients in the phase and frequency encoding directions *at the iso-centre* will also contribute). This will change the precessing frequency of the spins, which in k-space is equivalent to a phase accumulation along the k-space trajectory. For EPI, this phase change between consecutive data points along the k_x direction is low thanks to the high bandwidth. On the other hand, the time between consecutive points along k_y is N_x times longer, why the phase accumulation in the k_y direction may be significant. Such a linear phase trend in k-space is equivalent to a shift of the object in the phase encoding direction in the image domain. This is illustrated in Figure 33a. The shift measured in pixels can be calculated using

$$\Delta y = \frac{\gamma \cdot \Delta B_0 \cdot T}{2\pi} \qquad [48]$$

where ΔB_0 = the slice-wise field change of B_0 and T = the total readout time of k-space. This eddy current will be denoted EC_{tr} (tr = translation).

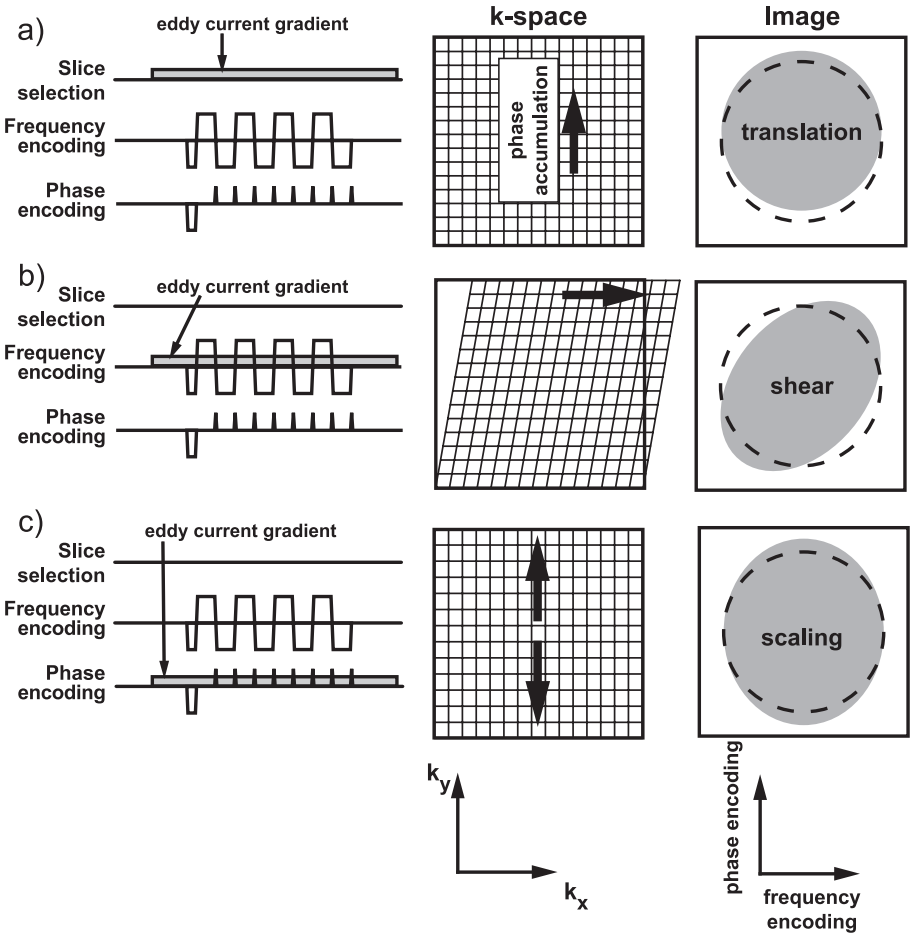

Figure 33. Eddy currents (EC) during the EPI readout. a) An EC in the slice selection direction will cause a linear phase shift in the k_y direction in k-space. This corresponds to a shift of the object in the phase encoding direction (here *y*) in the reconstructed image. b) An EC in the frequency encoding directions will cause the k-space to be sheared resulting in a shear of the object in the image in the phase encoding direction. c) Finally, an EC in the phase encoding direction makes sampling density of k-space to change in the k_y direction. This causes the effective FOV to change in the MR image, which is equivalent to a scaling effect of the object.

The slice selection eddy current can be divided in two parts, one termed B_0 eddy (described above) and the other z eddy current. The distinction is shown in Figure 34. The average EC_{slice} for a given slice will induce, as mentioned earlier, a shift of the image – this is the B_0 eddy current (here EC_{tr}). *Within* the slice the EC_{slice} will induce a dephasing due to the fact that spins at the left edge of the slice will precess slower than spins on the right edge – this is z eddy current (here EC_{int}, int = intensity) and will result in a signal intensity reduction. Clearly, the slice thickness will have an impact on this effect. Thinner slices will have less EC_{int}.

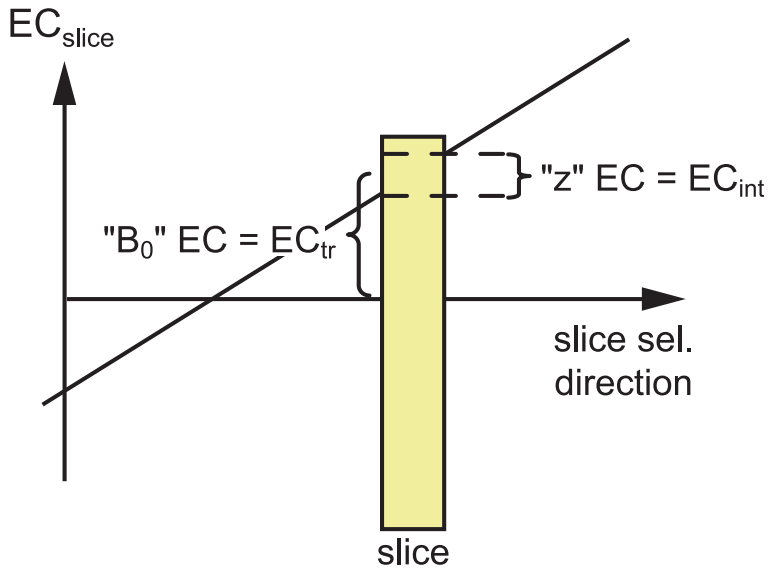

Figure 34. The two effects of eddy currents in the *z* (slice) eddy current. For a given slice, the major effect (pointed out in Figure 33a) is due to the local change of the static magnetic field that causes an shift of the object in the image. The other effect is that the magnetic field will vary *through* the slice. This causes the spin within the slice to dephase, leading to an irrecoverable loss of signal. The magnitude of signal attenuation varies with slice thickness but is usually rather low.

An eddy current in the frequency encoding direction (Figure 33b) causes the sampling step size to increase for odd echoes and to decrease for even echoes. This leads to a shift to the right (left if the eddy current gradient is negative) for consecutive k_x lines, which results in a sheared sampling of k-space. In the image domain the object becomes sheared in the phase encoding direction. This eddy current will here be denoted EC_{sh}.

Finally, an eddy current gradient in the phase encoding direction, as depicted in Figure 33c, will cause the effective area of the phase encoding blips to increase. Note that each line in k-space is no longer sampled at the same k_y, but rather in a zig-zag manner through k-space, which induces some blurring in the image. The dominant effect is however that k-space is more sparsely sampled in the k_y direction. This is equivalent to a decreased FOV in the image domain, which will make the object magnified or scaled in the phase encoding direction. The situation is vice versa if the eddy current gradient is negative or if k-space is sampled top-down (negative blips) instead of bottom-up. This eddy current will be denoted EC_{sc}.

Because these EC effects depend on the direction of the diffusion gradient, there will be a mismatch not only between the DWIs and the T_2-w EPI but also between the different DWIs. As a result, the diffusion tensor calculated from these images as well as the mean diffusion and anisotropy maps will be blurred. For the latter, there will also be artefactually high anisotropy values along tissue edges.

4.4.2 Pulse sequence solutions

Several groups have attempted to minimise the EC by applying "extra" diffusion gradients or using other diffusion gradient configurations in the sequence. For instance Alexander et al.[73] employed bipolar gradients, i.e. a diffusion gradient pair before *and* after the 180° pulse, and have shown that this decreases the EC (Figure 35a). Unfortunately, this gradient configuration gives less diffusion weighting and hence requires that the duration of the DW gradients is longer, which increases TE and decreases the SNR as well as the number of slices/TR. To avoid increasing TE the insertion of one extra diffusion gradient with opposite sign *before* the excitation has also been proposed[73] (Figure 35b). Since this is placed before the excitation, obviously, it will not contribute to the diffusion encoding. Calamante et al.[74] improved the technique by combining the use of extra gradients with a non-phase encoded reference scan for k-space phase correction to remove the EC_{tr} effects.

Yet the seemingly most promising solution was proposed by Reese et al.[75] (based on Feinberg and Jakab[76] and later improved by Heid[77]). They have suggested a gradient combination applied in a double spin echo EPI sequence (i.e. with two 180° pulses) that has higher *b* value efficiency and therefore shorter TE compared to the above-mentioned methods. The use of two 180° pulses allows the diffusion gradient train to be designed in a way that makes the diffusion encoding close to the Stejskal-Tanner gradient pair. At the same time, the gradients actually played out are similar to the bipolar gradient configuration in terms of eddy currents (Figure 35c). By adjusting the length of the first gradient, different eddy current time constants may be nulled.

Another pulse sequence modification has been proposed by Jezzard et al.[78]. Instead of using a low-eddy current diffusion pulse sequence designed as above, the eddy currents are measured during the diffusion scan using two reference scans. After filling k-space (single or multi-shot-wise) the first reference scan is acquired by setting the phase encoding amplitudes to zero, followed by the second reference scan where the phase/frequency encoding axes are swapped (see Figure 35d). Consider e.g. the data acquired using the readout-train shown in Figure 35d-2, where the phase encode blips have been set to zero. Under ideal circumstances each echo should be identical (T_2^* dependent decay between successive echoes excluded), and represent the Fourier-transform of a projection of the object onto the frequency-encoding axis. The amplitude of the 1D inverse Fourier-transform applied to any of these echoes should then represent that projection. The corresponding phase will be due to a multitude of causes, but is generally dominated by "machine factors" such as slight timing errors.

However, the phase evolution between consecutive acquisitions is directly related to the deviation of the local field from its "tentative" value such that $\Delta B_0 = \Delta\varphi/\gamma t$ where $\Delta\varphi$ denotes the phase evolution during the time interval t at a specific location. So, if we assume there is an x-gradient (over and above the readout-gradient) caused by eddy-currents we would expect that to cause a phase evolution between the "complex projections" corresponding to consecutive echoes. Furthermore, let us assume this gradient causes a higher magnetic field at the right side of the object: we would then expect the phase-evolution between the two projections to be greater there than at the left side. Hence, the assumption behind the method of Jezzard et al. is that the subtraction of the phase of two consecutively acquired "complex projections" should yield a straight line, the slope of which is proportional to the eddy-current induced gradient in that direction. The intercept of this line is expected to coincide with the centre of the projection (assuming the centre of the FOV coincides with the iso-centre of the scanner in this direction) and any deviation from that is considered to constitute an eddy-current induced B_0 component.

In addition, Jezzard et al. can assess the development over time of the eddy-current induced gradients. Consider the straight line resulting from the subtraction of the phase of the "complex projections" corresponding to echoes #1 and #3. Furthermore, consider the corresponding line for echoes #3 and #5. If all had remained constant between the times corresponding (roughly) to the midpoints of the acquisitions of these echo-pairs we would have expected these lines to be identical (except for noise). A difference in slope between these lines would indicate a change, over this time interval, in eddy-current induced gradient along this direction, whereas a disagreement in intercept would suggest a change in the B_0 component. Everything that has been written above applies equally well when the readout-gradient is applied in the y-direction (Figure 35d-3) (the phase-encoding direction of the "true" acquisition scheme, Figure 35d-1). Hence, the eddy-current induced gradient in that direction is calculated in an equivalent manner. Both schemes yield estimates of the B_0 component and Jezzard et al. simply use the average. Once the eddy-current induced gradients have been characterised for each slice and diffusion gradient, the data acquired with the "proper" sequence are corrected accordingly. The method is fairly time efficient for multi-shot EPI, but for single-shot EPI the scan time is increased by a factor of three.

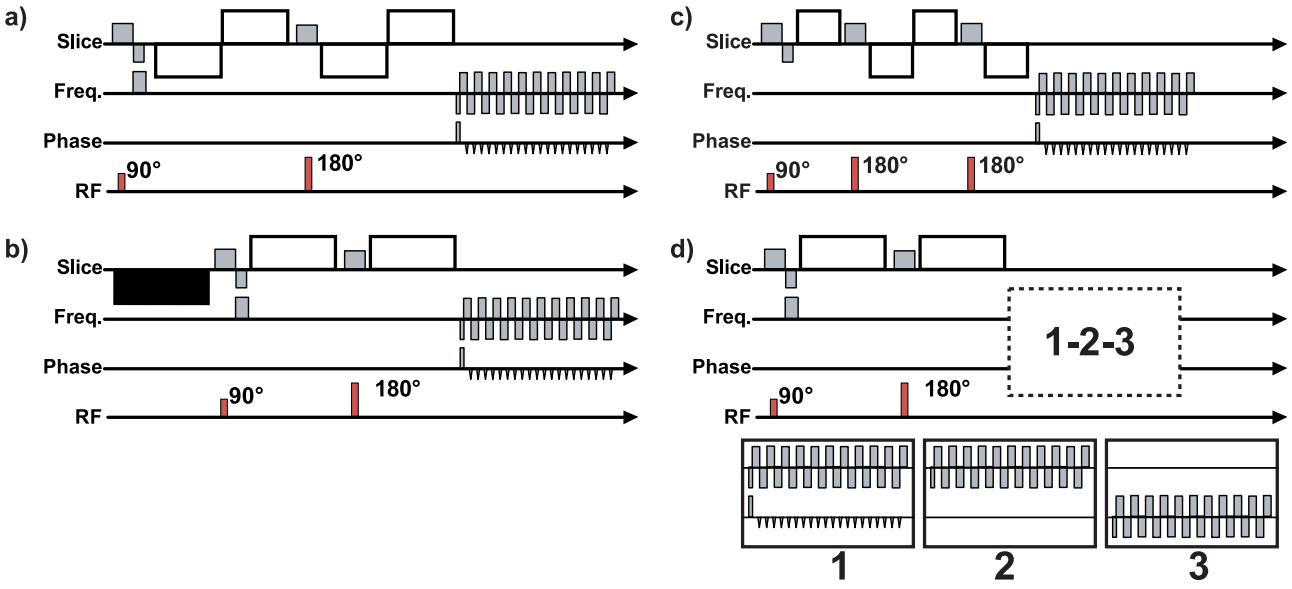

Figure 35. Diffusion gradient configurations to minimise eddy currents. a) The use of bipolar diffusion gradients reduces the EC but unfortunately also the b value. The consequence is a longer TE and reduced SNR. b) An extra gradient added before the excitation. This allows the TE to be kept constant, but has not proven to reduce the EC sufficiently. c) The double spin echo sequence. The diffusion gradients are similar to the bipolar configuration, giving substantial reduction of the EC. By employing two refocusing pulses the gradient configuration becomes similar to the standard Stejskal-Tanner scheme. d) The EC's are measured by alternating the actual EPI readout (#1) with two readouts (#2 and #3) that measure the eddy currents. From these measurements the distorted image data can be corrected. For single-shot EPI, this increases the scan time by a factor of three.

4.4.3 Gradient pre-emphasis solutions

The major drawbacks of making changes in the pulse sequence are increased TE, fewer slices/TR or longer scan times. Instead, one could try to improve the eddy current pre-emphasis calibration.

Zhou et al.[79,80], Terpsta et al.[81], Papadakis et al.[82] and Schmithorst et al.[83] have all proposed methods to determine the amplitudes of the eddy current time constants in the pre-emphasis system. The first two groups were able to use an arbitrary number of time constants, while the latter two were "limited" to three. The Papadakis group, unlike the others, estimated the EC parameters slice-wise. After the calibration step, these parameters were used in the DW-EPI sequence to modify, in real-time, the phase and frequency offset of the receiver before and during the EPI acquisition window[79].

4.4.4 Post-processing solutions

All EC effects except the intensity reduction can be corrected using post-processing methods on the magnitude MR images, provided that the EC are fairly constant during the acquisition.

The problem is to remove the EC induced distortions in the DW component images, which each has a unique geometric distortion, so that an anatomical match between the

DWIs and the T_2-w image is obtained – the images are *registered* to each other. The first obstacle is the fact that the image contrast differs between the T_2-w and the DW images. Most obvious is the presence of CSF signal in the T_2-w image. Hence it is not possible to directly use an image registration method based on similarity (as commonly done in fMRI), unless the CSF signal in the T_2-w is nulled using a FLAIR-EPI sequence.

To register two images based on similarity one can minimise the sum-of-squares difference between the images, i.e.

$$O = \sum_{i=1}^{N} \left(I_1(i) - I_2(i) \right)^2 \qquad [49]$$

where I_1 and I_2 are the images to be registered and N the number of voxels. Haselgrove and Moore[84] have made an approach based on similarity. First, the T_2-w image and a low-b value DWI ($b < 300$ s/mm^2), in which the CSF signal is still present, are registered. Then, under the assumption that the distortion is proportional to the b value, they could extrapolate to higher b values normally used (≈ 1000 s/mm^2). Rather than finding the slice-wise translation, shear and scale parameter Haselgrove use a column-wise registration. This makes it unnecessary to find the shear of the image as it is, for a single column, included in the translation term.

There are several issues with this method. First, extrapolation of distortion parameters could easily generate more distortion instead of less. Second, one is forced to acquire also a low-b value DWI, which will increase the scan time.

Bastin[85] and de Crespigny[86] have continued this approach. de Crespigny, however, used the average of the DWIs as a template instead of the T_2-w image to correct each component DWI, avoiding the need for a low-b value DWI because the image contrast is similar. Since the averaging of the distorted component DWIs for the generation of the template results in a blurred mean DWI with little remaining distortion, the corrected component DWIs are expected to match the T_2-w image fairly well.

Horsfield has proposed the use of mutual information[87,88] as the objective function in the distortion correction algorithm of the component DWI and the T_2-w image[89,90]. The mutual information (MI) measure makes very few assumptions about the information content in the images. It is therefore potentially to prefer over Haselgrove's or de Crespigny's approach. In fact, any two images (e.g. DWI vs. T_2-w or T_1-w MRI vs. PET image) can be registered to each other using this method. When two images are properly aligned to each other the mutual information is maximised. The drawback of the MI method is that it is rather slow. An execution time of about one day for 20 slices and a 6-directional diffusion scheme has been reported[90].

Bammer et al.[91] have used the MI measure to, estimate not only the affine transformations (i.e. shear, scaling and translation), but also non-linear or "elastic" transformations. The

aim was to make the DW images geometrically "look more alike the T_2-w". However, there is no physical model or reasoning behind these non-linear distortions.

4.5 PATIENT MOTION: MISMATCH BETWEEN THE DW COMPONENT IMAGES

In the previous section (4.4) a variety of techniques to correct for eddy current induced distortions were presented, all intended to minimise the mismatch between the DW images and the T_2-w image(s). However, apart from Bastin[85] who performed an in-plane correction for motion, little has been done about the mismatch due to motion. For many DTI applications high resolution and high image quality are needed and scan times of half an hour are not uncommon. During an acquisition time as long as that, patient motion *will* occur unless a stereotactic frame is used, adding another source of mismatch between the DW component images – in some cases larger than the eddy current distortions. For fMRI, where the scan times are comparable to a DTI scan, image realignment is virtually always employed. It is therefore a bit peculiar that this issue has not been debated in the DTI field, where the same problem exists.

For 3D image realignment in fMRI, all image volumes are similar and Eq. [49] may be used to find the six rigid body parameters (translation in x, y and z and rotations around the three axes) for each image volume. However, for 3D image realignment in DTI the same measure could not really be used even in the absence of EC distortions for the same reasons as for EC correction – the images are *not* similar due to diffusion anisotropy in WM. Mutual information could be used to find the motion parameters for each image volume even though this takes a lot of computation time.

A principle question of interest is, however, in which order should image realignment and EC correction be performed? The eddy current distortions act along the direction that was parallel to the phase encoding direction when that particular image was acquired. If motion has occurred and image realignment is performed before the EC correction, the image volume will be rotated so that the EC distortions no longer are in the phase encoding direction. Performing the EC correction on motion corrected images is consequently not expected to yield very good results. On the other hand, if EC correction is performed before image realignment, an in-plane rotation of the object will alter the estimation of the shear parameter, as a rotation is mathematically identical to three consecutive shears. The biased shear parameter cannot be undone by the image realignment performed afterwards.

Based on these arguments it would be desirable to correct for these two sources of mismatch simultaneously.

4.6 3D SIMULTANEOUS CORRECTION OF EDDY CURRENTS AND MOTION

In this section, an overview of the correction model developed in paper 5[92] is presented. This work has been given more space in this summary because of its relative novelty.

4.6.1 Estimated parameters and constraints

For each slice of each of the component DWIs, the three distortion types (shear, scaling, translation) are estimated. For each DW volume the six rigid motion parameters are estimated. For instance, with a diffusion scheme of 20 diffusion gradient directions and 25 slices, the total number of parameters that need to be estimated is $n_{param} = 3 \cdot n_{slices} \cdot n_{dir} + 6 \cdot n_{dir} = 1620$. Although these parameters can be estimated directly from the DTI data itself using our method (described below), we have also proposed and implemented two types of constraints that significantly reduce the number of independent parameters.

Spatial constraints
Even though it has indeed been shown that the distortions may vary for different slices in the image volume[86] it is reasonable to assume that they will vary smoothly rather than abruptly from slice to slice. Variation of the shear distortion through the slices (and likewise that of scaling and translation) could be modelled very well by e.g. a polynomial or a cosine basis set of, let us say, four parameters. This is what we call a spatial constraint. For a 3rd degree polynomial set the shear variation through the slices of a particular DWI volume can be determined by finding the coefficients, c_i

$$shear(x) = c_1 x^3 + c_2 x^2 + c_3 x + c_4 \qquad [50]$$

For each DWI volume now only 4 parameters are estimated for each distortion type instead of as many parameters as there are slices. This reduces the dimensionality of the problem in this example to $n_{param} = 3 \cdot 4 \cdot 20 + 6 \cdot 20 = 360$.

Figure 36 shows the rationale behind spatial constraints, exemplified with the shear distortion (not actual data). The estimates of the shear parameters for each slice for a specific diffusion direction are given in Figure 36a. In this example it is clear that the shear is not constant across slices. Figure 36b shows that much of the variation may be captured by a simple linear relationship. Finally in Figure 36c, where a slightly higher order model is used, described by four parameters in a cosine basis set, virtually all of the variability is accounted for.

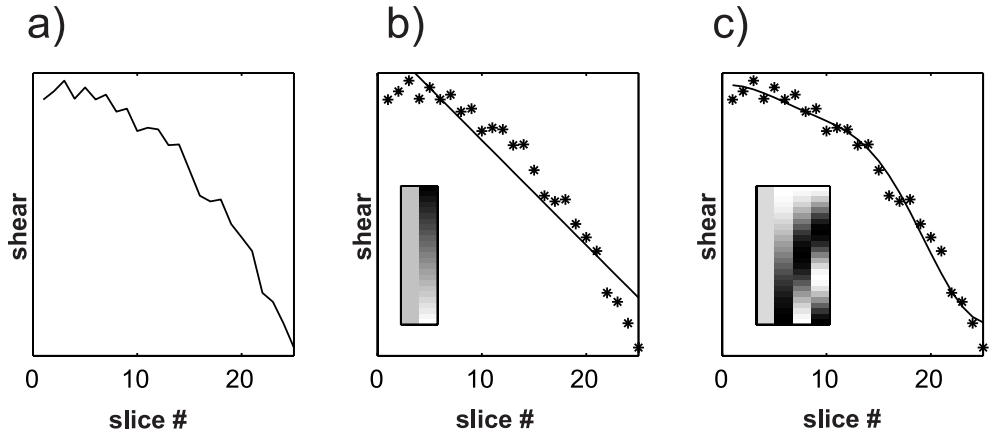

Figure 36. Spatial constraints. An example of the shear distortion as a function of slice number. a) In principle the shear parameter for each slice can be estimated. b) A major part of the variance across slices can be modelled by a linear relationship that reduces the estimated shear parameters per image volume to two. c) Using a slightly higher order model (four parameters in a cosine basis set) virtually all variance is accounted for.

Gradient constraints

The EC distortions of the different DW component image volumes are not independent of each other. There is a strong correlation between the direction of the diffusion gradients and the distortions. For instance, a diffusion gradient along the frequency encoding direction induces mainly an EC in the same direction although some cross-term effects will be present. The simplest gradient constraint - and most rigid - is to say that the EC distortions are a simple function of the diffusion gradient amplitudes. In order words, the expected distortions are

$$\begin{cases} shear(i) = c_1 G_{x,i} \\ scaling(i) = c_2 G_{y,i} \\ translation(i) = c_3 G_{z,i} \end{cases} \quad [51]$$

where i = 1…n_{dir} and $[G_{x,i}, G_{y,i}, G_{z,i}]^T$ is the i:th diffusion gradient vector in the diffusion scheme. By doing so we estimate for each distortion type only one constant c_i per slice (or spatial hyper parameter) rather than one for each DW image volume and slice (or spatial hyper parameter). Assuming spatial constraints with four hyper parameters per distortion type and 20 diffusion directions, this reduces the number of parameters to be estimated to: $n_{param} = 3 \cdot 4 + 6 \cdot 20 = 132$, of which only 12 are used to model the ECs. To also account for first order cross-terms (i.e. a diffusion gradient along x will give some EC in the y and z direction) we could set up another gradient constraint, namely

$$\begin{cases} shear(i) = c_1 G_{x,i} + c_2 G_{y,i} + c_3 G_{z,i} \\ scaling(i) = c_4 G_{x,i} + c_5 G_{y,i} + c_6 G_{z,i} \\ translation(i) = c_7 G_{x,i} + c_8 G_{y,i} + c_9 G_{z,i} \end{cases} \quad [52]$$

In this case the number of parameters to be estimated becomes $n_{param} = 9 \cdot 4 + 6 \cdot 20 = 156$. This is only slightly higher than for the gradient constraint with no EC cross-terms modelled.

Figure 37a shows for a given slice, the shear estimates vs. the strength of the *x*-component of the diffusion gradient (not actual data). Applying a gradient constraint without cross-terms (as in Eq. [51]) is equivalent to assigning a linear relationship between shear and *x*-component, which only partly explains the shear effect. If cross-terms are included (as in Eq. [52]) the shear should be plotted against *x, y* and *z*, which is hard to present. Nonetheless, in Figure 37b the shear is however plotted against *x*- and *y*-component of the diffusion gradients. (In reality the shear does not significantly depend on *z* anyway. See Fig 11, paper 5.) This explains most of the structure of the shear distortions.

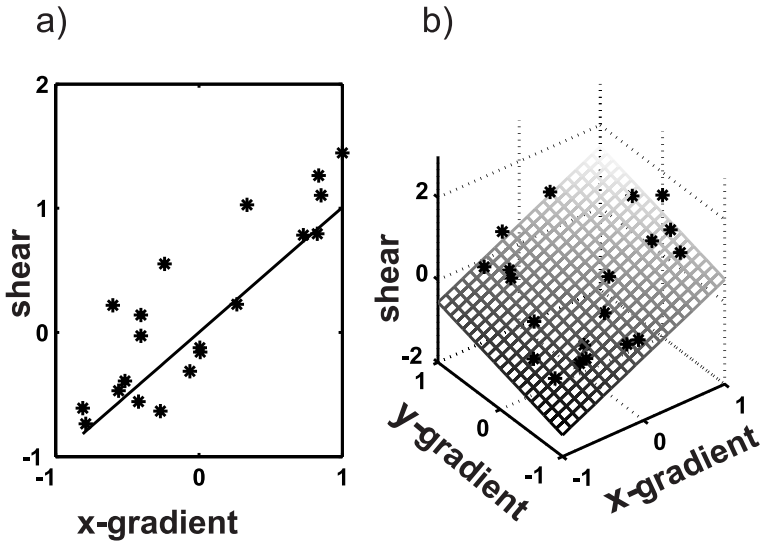

Figure 37. Gradient constraints. An example of the shear distortion vs. the x-gradient component for a 20 directional diffusion scheme. a) Instead of estimating a shear parameter for each image volume (diffusion direction) indicated by the asterisks, one can reduce the number of parameters to be estimated by assigning a linear relationship (i.e. one parameter) with the x-gradient. b) A slightly more flexible, and in practice more accurate, constraint is to account for cross-terms too. I.e. to say that the shear distortion is proportional to a linear combination of the x- and y-components of the diffusion gradients (two parameters).

4.6.2 Using the residual error of the tensor to correct for motion and eddy current distortions

The idea behind our algorithm developed in paper 5, is that the tensor model explains our DW data better in the absence than in the presence of motion and distortions. Distortions and motion will increase the residual error term given in Eq. [33] for all voxels in the image volume. This can be used as an objective function to be minimised in the simultaneous search for motion and distortion parameters (see section 4.6.1). Residual error information is obtained if a diffusion scheme with more than six unique directions is used, since we then have a voxel-wise over-determined system with more measurements than unknowns. The residual error can be derived from Eq. [33] according to the following. In each voxel we have estimated the diffusion tensor elements $\hat{\mathbf{d}}$ from the

measured ADC values contained in $\mathbf{y}$. The DW data that *exactly* corresponds to $\hat{\mathbf{d}}$ is here denoted $\hat{\mathbf{y}}$ and is given by

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\mathbf{d}} = \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} \qquad [53]$$

The difference between the actual data $\mathbf{y}$ and $\hat{\mathbf{y}}$ is the residual error, $\mathbf{e}$. The sum-of-squares errors (see Eq. [49]) can then be written as

$$\begin{aligned} O &= \mathbf{e}^T\mathbf{e} = \\ &\left(\mathbf{y} - \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}\right)^T\left(\mathbf{y} - \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}\right) = \\ &\mathbf{y}^T\left(\mathbf{I} - \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\right)^T\left(\mathbf{I} - \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\right)\mathbf{y} = \\ &= \mathbf{y}^T\mathbf{R}^T\mathbf{R}\mathbf{y} \end{aligned} \qquad [54]$$

$\mathbf{R}$ is the *residual forming matrix* and has the properties of being symmetric and *idempotent*. The symmetry implies that $\mathbf{R}^T = \mathbf{R}$ while being idempotent implies that $\mathbf{RR} = \mathbf{R}$. Therefore the expression can be simplified to

$$O = \mathbf{y}^T\mathbf{R}\mathbf{y} \qquad [55]$$

which is a *quadratic form*. This expression is valid for only one voxel. However it is possible to formulate an expression that is valid for the entire image volume.

When estimating a large number of parameters one needs an efficient search algorithm. These can be divided in two groups – those that use the derivatives of the objective function and those that do not. Whenever possible a search algorithm employing derivatives should be used.

Because it is possible to calculate the derivatives of our objective function, we may use a fast search algorithm (Leveberg-Marquardt). Our objective function is a function of the images, which in turn are multidimensional functions of the parameters to be estimated. Then by the chain rule (with inner derivatives) there will be a huge second order derivative matrix that would make the calculations very tedious and memory consuming. However, since our objective function is a quadratic form, it can be shown that the term containing the second derivative becomes nearly zero, obviating the need to calculate this matrix. This speeds up the algorithm significantly.

Besides, preliminary results indicate that the accuracy of the method is better than mutual information[93]. This finding is illustrated in Figure 38. In Figure 38a the dashed line delineates the common anatomical location (a column along the phase encoding direction of a single slice) of all columns in Figure 38b-d. Each of the 20 columns in Figure 38b-d

depicts a stacking of the intensity profiles for a 20-direction isotropic diffusion scheme. Figure 38b shows an intensity profile without any correction while Figure 38c-d shows the corresponding data after correction with MI and the residual error model, respectively. Since the dashed line is placed off-centre in Figure 38a, the shear effect is included in the comparison. It can be seen that both methods reduce geometrical distortions, but that the residual error algorithm appears to perform slightly better.

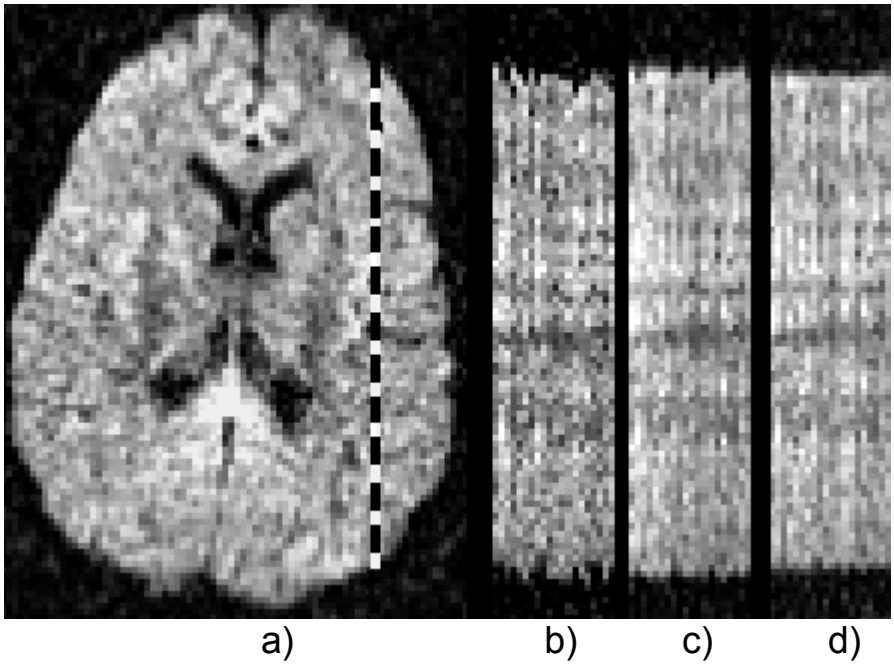

Figure 38. Eddy current correction using MI and residual error model. In b)-d) each column corresponds to one of the 20 diffusion directions at the anatomical location shown by the dashed line in a). b) without correction, c) with MI d) with residual error.

5. Appendix

5.1 GENERAL EIGENVALUE CALCULATION

One method to calculate the eigenvalues of a matrix **A**, is to use the QR method which is an iterative method based on QR-decomposition of the matrix **A**.

1. QR-decomposition

Any m×n matrix **A** ($m \geq n$) with linearly independent columns can be factored into

$$\mathbf{A}=\mathbf{QR}$$

where **Q** is an m×m orthonormal matrix and **R** is an m×n upper triangular matrix with non-zero diagonal elements.

There are several methods to orthogonalise **A**. One of the simplest is called the Gram-Schmidt orthogonalisation method and goes as follows

Let the column vectors of **A** be $\mathbf{a}_1..\mathbf{a}_n$

$$\mathbf{A}=\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}=\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}$$

Find the column vectors of $\mathbf{Q}=\begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{bmatrix}$ using the following approach (Gram-Schmidt)

Vector 1: $\mathbf{q}_1=\dfrac{\mathbf{a}_1}{|\mathbf{a}_1|}$

Vector 2: $\mathbf{q}_2=\dfrac{\mathbf{f}_2}{|\mathbf{f}_2|}$, where $\mathbf{f}_2=\mathbf{a}_2-\left(\mathbf{a}_2^T\mathbf{q}_1\right)\mathbf{q}_1$

and so on. For $\mathbf{q}_k$ the expression becomes

Vector k: $\mathbf{q}_k=\dfrac{\mathbf{f}_k}{|\mathbf{f}_k|}$, where $\mathbf{f}_k=\mathbf{a}_k-\left(\mathbf{a}_k^T\mathbf{q}_1\right)\mathbf{q}_1-\ldots-\left(\mathbf{a}_k^T\mathbf{q}_{k-1}\right)\mathbf{q}_{k-1}$

Now that **Q** is determined, **R** is calculated using

$$\mathbf{A}=\mathbf{QR}\Leftrightarrow\mathbf{R}=\mathbf{Q}^{-1}\mathbf{A}\Rightarrow\left|\begin{array}{c}\mathbf{Q}\text{ is orthonormal}\\ \Rightarrow\ \mathbf{Q}^{-1}=\mathbf{Q}^T\end{array}\right|\Rightarrow\mathbf{R}=\mathbf{Q}^T\mathbf{A}$$

continuing on next page

2. The QR method

Start by QR-decomposing **A** into $\mathbf{Q}_0$ and $\mathbf{R}_0$.

Set $\mathbf{A}_1 = \mathbf{R}_0\mathbf{Q}_0 = \mathbf{Q}_0^T\mathbf{A}_0\mathbf{Q}_0$ (note the reversed order of R and Q!!) and perform QR-decomposition of $\mathbf{A}_1$ into $\mathbf{Q}_1$ and $\mathbf{R}_1$. Continuing iteratively in n steps the general expression becomes

$$\mathbf{A}_n = \left(\prod_{i=0}^{n}\mathbf{Q}_i^T\right)\mathbf{A}\left(\prod_{i=0}^{n}\mathbf{Q}_i\right)$$

The eigenvalues of **A** are contained in the diagonal elements of $\mathbf{A}_n$ provided that n is large enough. One notices that there is no need to calculate **R**. The power series of **Q** form a rotation matrix that rotates **A** into the principal coordinate system spanned by its eigenvectors.

5.2 SOLVING THE CHARACTERISTIC EQUATION

The characteristic equation can be used to calculate the eigenvalues of **D**

$$\det(\mathbf{D}-\lambda\mathbf{I}) = \begin{vmatrix} D_{xx}-\lambda & D_{xy} & D_{xz} \\ D_{yx} & D_{yy}-\lambda & D_{yz} \\ D_{zx} & D_{zy} & D_{zz}-\lambda \end{vmatrix} = 0$$

where **I** here is the 3×3 identity matrix.

For example, if we have the tensor defined by Eq. [9], the characteristic equation becomes

$$-\lambda^3 + 11\lambda^2 - 19\lambda + 9 = 0$$

which has roots $\lambda_1 = 9$, $\lambda_2 = \lambda_3 = 1$. Voila!

5.3 MEASURES FOR STABILITY OF DIFFUSION SCHEMES

In this section we attempt to clarify the significance of two different indices of "diffusion scheme goodness", the κ index suggested by Papadakis[48] and the condition number suggested in paper 2. The model for estimation of the diffusion tensor elements is given by

$$\mathbf{y} = -\frac{1}{\gamma^2 G^2 \delta^2 (\Delta - \delta/3)} \begin{bmatrix} \ln\left(\frac{s_1}{s_0}\right) \\ \ln\left(\frac{s_2}{s_0}\right) \\ \vdots \\ \ln\left(\frac{s_n}{s_0}\right) \end{bmatrix} = \begin{bmatrix} q_{1x}^2 & q_{1y}^2 & q_{1z}^2 & q_{1x}q_{1y} & q_{1x}q_{1z} & q_{1y}q_{1z} \\ q_{2x}^2 & q_{2y}^2 & q_{2z}^2 & q_{2x}q_{2y} & q_{2x}q_{2z} & q_{2y}q_{2z} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ q_{nx}^2 & q_{ny}^2 & q_{nz}^2 & q_{nx}q_{ny} & q_{nx}q_{nz} & q_{ny}q_{nz} \end{bmatrix} \begin{bmatrix} D_{xx} \\ D_{yy} \\ D_{zz} \\ D_{xy} \\ D_{xz} \\ D_{yz} \end{bmatrix} + \mathbf{e} = \mathbf{X}\mathbf{d} + \mathbf{e}$$

where $\mathbf{e}$ is an (unknown) $n\times1$ stochastic error vector.
The tensor elements are estimated in a least-squares sense by

$$\hat{\mathbf{d}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

Our concern is the precision by which we are able to estimate $\mathbf{d}$. In other words, we are concerned with $\mathrm{Cov}(\hat{\mathbf{d}})$, which is a 6×6 matrix with the variance of the estimates along the diagonal and their covariances off the diagonal. So, how can we estimate $\mathrm{Cov}(\hat{\mathbf{d}})$ given the model above? First, we combine these two equations yielding

$$\hat{\mathbf{d}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{X}\mathbf{d} + \mathbf{e}) = \mathbf{d} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{e}$$

From this, it is easily seen that the covariance matrix of $\hat{\mathbf{d}}$ can be expressed as

$$\mathrm{Cov}(\hat{\mathbf{d}}) = \mathrm{Cov}(\mathbf{d}) + \mathrm{Cov}((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{e}) = \mathrm{Cov}((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{e})$$

where the latter equality comes from $\mathbf{d}$ being a constant in our model. Using the basic equality $\mathrm{Cov}(\mathbf{A}\mathbf{x})=\mathbf{A}\mathrm{Cov}(\mathbf{x})\mathbf{A}^T$, where $\mathbf{A}$ is a constant matrix and $\mathbf{x}$ a stochastic vector, this can be expressed as

$$\mathrm{Cov}(\hat{\mathbf{d}}) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathrm{Cov}(\mathbf{e})\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}$$

In order to progress we shall now need to make certain assumptions regarding $\mathbf{e}$. Specifically we will assume that $\mathrm{Cov}(\mathbf{e})=\sigma^2\mathrm{I}$, i.e. that the variance is equal for all measurements, and that the measurements are independent (iid). Using this assumption we get

$$\mathrm{Cov}(\hat{\mathbf{d}}) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} = (\mathbf{X}^T\mathbf{X})^{-1}\sigma^2$$

Hence, using only minimal assumptions we have obtained an expression for the covariance-matrix of our estimate of the diffusion tensor elements. Furthermore, from our perspective σ^2 can be seen as a constant (the variance of the individual measurements is not affected by our choice of diffusion gradient directions). This means that we should chose a set of diffusion gradients that minimise the elements of $(\mathbf{X}^T\mathbf{X})^{-1}$.
However, there is still an ambiguity. What specifically is meant by "the elements of"?
Ideally, we would like a scalar measure that is proportional to the efficiency of the gradient set as embodied by **X**. An intuitive candidate would be the trace of $(\mathbf{X}^T\mathbf{X})^{-1}$ i.e. the sum of the variances of the elements of the tensor. An alternative candidate would the determinant of $(\mathbf{X}^T\mathbf{X})^{-1}$, denoted $|(\mathbf{X}^T\mathbf{X})^{-1}|$, which can be thought of as proportional to the volume (in 6D space) spanned by the multivariate probability density function of the estimated parameters.
Both these candidates, $\mathrm{tr}((\mathbf{X}^T\mathbf{X})^{-1})$ and $|(\mathbf{X}^T\mathbf{X})^{-1}|$ are reasonable indices that depend only on ones choice of diffusion gradient directions and that have a direct relationship to the precision by which we can estimate the diffusion tensor.

The condition number (with respect to inversion) of the resulting design matrix has a somewhat different meaning. In order to understand that we need the concept of a matrix norm. First consider the vector norm, denoted $\|\mathbf{a}\|$ for some vector **a**. The vector norm $\|\mathbf{a}-\mathbf{b}\|$ is a useful scalar measure for determining how "similar" two vectors **a** and **b** are. It is e.g. the measure used for all least-squares estimation.
Imagine now that we have two matrices **A** and **B**, and that we want to determine how "similar" they are. More specifically one often wants to know if "**A** is more similar to **B** or **C**". Hence, what we need is a matrix norm, denoted $|||\mathbf{A}|||$, by which we can measure "distance" in a matrix space. A commonly used choice of matrix-norm is the "spectral norm", often denoted $|||\mathbf{A}|||_2$, which is simply the largest singular value of the matrix **A**. A discussion of different matrix norms would lead much to far, so suffice to say $|||\mathbf{A}-\mathbf{B}|||_2 < |||\mathbf{A}-\mathbf{C}|||_2$ would indicate that **B** is more similar to **A** than is **C**.
Using the matrix-norm as a tool one can then start to gauge the error sensitivity of matrix-valued functions of matrices, and specifically the matrix inverse. Assume we have a matrix **A**, and that we want to calculate its inverse $\mathbf{A}^{-1}$ but that, due to round-off or stochastic errors, we really obtain $(\mathbf{A}+\mathbf{E})^{-1}$ where **E** is a stochastic matrix. If $|||\mathbf{A}^{-1}-(\mathbf{A}+\mathbf{E})^{-1}|||_2$ is large for a given **E** the matrix **A** is said to be poorly conditioned, and its "condition number", defined as $\mathrm{cond}(\mathbf{A}) = |||\mathbf{A}^{-1}|||\,|||\mathbf{A}|||$, will be large. It can be shown that one can obtain an upper bound on the error of ones solution to a linear equation system based on the condition number according to[94]

$$\frac{\left\|\mathbf{d}-\hat{\mathbf{d}}\right\|}{\left\|\mathbf{d}\right\|} \leq \mathrm{cond}(\mathbf{X}) \frac{\left\|(\mathbf{I}-\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)\mathbf{y}\right\|}{\left\|\mathbf{y}\right\|}$$

where symbols have been used as in the previous section. If we attempt to dress this equation in words it would be; the relative squared error of the tensor-elements is less than, or equal to, the condition number of **X** times the relative residual error of the data.

Hence, even though a formal derivation, which we have only sketched above, of why the condition number is relevant is much more complex than is the case for $\mathrm{Cov}(\hat{\mathbf{d}})$ it is still clear that they have some relation. In particular $\mathrm{tr}((\mathbf{X}^T\mathbf{X})^{-1})$, which is almost the same as Papadakis κ, should have a close relation to $\mathrm{cond}(\mathbf{X})$.

6. References

1. S Brockstedt, Nuclear magnetic resonance for diffusion imaging: Development and evaluation of spin-echo based pulse sequences for in vivo studies of water self-diffusion, *Thesis* Lund University (1998), Lund
2. B Geijer, Diffusion MRI of small ischemic brain lesions: Technical aspects, clinical experiences and diagnostic criteria, *Thesis* Lund University (2001), Lund
3. Z Chen, Experimental and clinical perspective on stroke: Evaluation by behavior, magnetic resonance imaging and morphology, *Thesis* Karolinska Institute (2001), Stockholm
4. PJ Basser, et al., Estimation of the effective self-diffusion tensor from the NMR spin echo. *J Magn Reson B* (1994) **103**(3): p. 247-254.
5. PJ Basser, et al., MR diffusion tensor spectroscopy and imaging. *Biophys J* (1994) **66**(1): p. 259-267.
6. A Einstein, Investigations on the theory of the Brownian motion (1956 (orig. 1926)).
7. PJ Basser, New histological and physiological stains derived from diffusion-tensor MR images. *Ann N Y Acad Sci* (1997) **820**: p. 123-138.
8. D Le Bihan, et al., Diffusion tensor imaging: concepts and applications. *J Magn Reson Imaging* (2001) **13**(4): p. 534-546.
9. E Sykova, The extracellular space in the CNS: Its regulation, volume and geometry in normal and pathological neuronal function. *Neuroscientist* (1997) **3**: p. 28-41.
10. C Nicholson and L Tao, Hindered diffusion of high molecular weight compounds in brain extracellular microenvironment measured with integrative optical imaging. *Biophys J* (1993) **65**(6): p. 2277-2290.
11. DG Norris, The effects of microscopic tissue parameters on the diffusion weighted magnetic resonance imaging experiment. *NMR Biomed* (2001) **14**(2): p. 77-93.
12. S Skare, et al., Noise considerations in the determination of diffusion tensor anisotropy. *Magn Reson Imaging* (2000) **18**(6): p. 659-669.
13. PJ Basser, Inferring microstructural features and the physiological state of tissues from diffusion-weighted images. *NMR Biomed* (1995) **8**(7-8): p. 333-344.
14. MM Bahn, Invariant and orthonormal scalar measures derived from magnetic resonance diffusion tensor imaging. *J Magn Reson* (1999) **141**(1): p. 68-77.
15. C Pierpaoli and PJ Basser, Toward a quantitative assessment of diffusion anisotropy. *Magn Reson Med* (1996) **36**(6): p. 893-906.
16. C Pierpaoli, Erratum: Toward a quantitative assessment of diffusion anisotropy. *Magn Reson Med* (1997) **37**(6): p. 972.
17. C Pierpaoli, et al., Diffusion tensor MR imaging of the human brain. *Radiology* (1996) **201**(3): p. 637-648.
18. P Douek, et al., MR color mapping of myelin fiber orientation. *J Comput Assist Tomogr* (1991) **15**(6): p. 923-929.
19. C Pierpaoli, Oh no! One more method for color mapping of fiber tract direction using diffusion MR imaging data. *Proceedings in* ISMRM (1997), Vancouver, Canada.
20. D Jones, et al., Full representation of white-matter fibre direction on one map via diffusion tensor analysis. *Proceedings in* ISMRM (1997), Vancouver, Canada.
21. S Pajevic and C Pierpaoli, Color schemes to represent the orientation of anisotropic tissues from diffusion tensor data: application to white matter fiber tract mapping in the human brain. *Magn Reson Med* (1999) **42**(3): p. 526-540.
22. S Pajevic and C Pierpaoli, Color schemes to represent the orientation of anisotropic tissues from diffusion tensor data: application to white matter fiber tract mapping in the human brain. *Magn Reson Med* (2000) **43**(6): p. 921.

23. R Xue, et al., In vivo three-dimensional reconstruction of rat brain axonal projections by diffusion tensor imaging. *Magn Reson Med* (1999) **42**(6): p. 1123-1127.
24. B Stieltjes, et al., Diffusion tensor imaging and axonal tracking in the human brainstem. *Neuroimage* (2001) **14**(3): p. 723-735.
25. S Mori, et al., Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Ann Neurol* (1999) **45**(2): p. 265-269.
26. S Mori, et al., Imaging cortical association using diffusion-tensor-based tracts in the human brain axonal tracking. *Magn Reson Med* (2002) **47**(2): p. 215-223.
27. TE Conturo, et al., Tracking neuronal fiber pathways in the living human brain. *Proc Natl Acad Sci U S A* (1999) **96**(18): p. 10422-10427.
28. PJ Basser, et al., In vivo fiber tractography using DT-MRI data. *Magn Reson Med* (2000) **44**(4): p. 625-632.
29. C Poupon, et al., Regularization of diffusion-based direction maps for the tracking of brain white matter fascicles. *Neuroimage* (2000) **12**(2): p. 184-195.
30. C Poupon, et al., Towards inference of human brain connectivity from MR diffusion tensor data. *Med Image Anal* (2001) **5**(1): p. 1-15.
31. S Pajevic, et al., A Continuous Tensor Field Approximation of Discrete DT-MRI Data for Extracting Microstructural and Architectural Features of Tissue. *J Magn Reson* (2002) **154**(1): p. 85-100.
32. A Aldroubi and PJ Basser, Reconstruction of vector and tensor fields from sampled discrete data, in *Contemporary mathematics*, L.D. Baggett LW, Editor (1999), American Math Society: Providence, RI. p. 1-15.
33. EW Hsu, et al., Magnetic resonance myocardial fiber-orientation mapping with direct histological correlation. *Am J Physiol* (1998) **274**(5 Pt 2): p. H1627-1634.
34. DF Scollan, et al., Histological validation of myocardial microstructure obtained from diffusion tensor magnetic resonance imaging. *Am J Physiol* (1998) **275**(6 Pt 2): p. H2308-2318.
35. AA Holmes, et al., Direct histological validation of diffusion tensor MRI in formaldehyde- fixed myocardium. *Magn Reson Med* (2000) **44**(1): p. 157-161.
36. CP Lin, et al., Validation of diffusion tensor magnetic resonance axonal fiber imaging with registered manganese-enhanced optic tracts. *Neuroimage* (2001) **14**(5): p. 1035-1047.
37. EO Stejskal and JE Tanner, Spin diffusion measurements: spin-echoes in the presence of a time-dependent field gradient. *J Chem Phys* (1965) **42**: p. 288-292.
38. D Le Bihan, Diffusion and perfusion magnetic resonance imaging, (1995), Raven Press. p. 14-15.
39. D Xing, et al., Optimised diffusion-weighting for measurement of apparent diffusion coefficient (ADC) in human brain. *Magn Reson Imaging* (1997) **15**(7): p. 771-784.
40. DK Jones, et al., Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. *Magn Reson Med* (1999) **42**(3): p. 515-525.
41. ME Moseley, et al., Diffusion-weighted MR imaging of anisotropic water diffusion in cat central nervous system. *Radiology* (1990) **176**(2): p. 439-445.
42. A Press, Dictionary of Science and Technology (1992): Academic Press.
43. TE Conturo, et al., Encoding of anisotropic diffusion with tetrahedral gradients: a general mathematical diffusion formalism and experimental results. *Magn Reson Med* (1996) **35**(3): p. 399-412.
44. S Skare, et al., Condition number as a measure of noise performance of diffusion tensor data acquisition schemes with MRI. *J Magn Reson* (2000) **147**(2): p. 340-352.
45. DK Jones, et al., Characterization of white matter damage in ischemic leukoaraiosis with diffusion tensor MRI. *Stroke* (1999) **30**(2): p. 393-397.
46. R Muthupallai, et al., Navigator Aided, Multishot EPI Diffusion Images of Brain with Complete Orientation and Anisotropy Information. *Proceedings in* ISMRM (1999), Philadelphia, USA.
47. S Skare and B Nordell, "Decahedral" gradient encoding for increased accuracy in the estimation of diffusion anisotropy. *Proceedings in* ISMRM (1999), Philadelphia, USA.

48. NG Papadakis, et al., A comparative study of acquisition schemes for diffusion tensor imaging using MRI. *J Magn Reson* (1999) **137**(1): p. 67-82.
49. P Jezzard and RS Balaban, Correction for geometric distortion in echo planar images from B0 field variations. *Magn Reson Med* (1995) **34**(1): p. 65-73.
50. Z-P Liang, A model based method for phase unwrapping. *IEEE Trans Med Imaging* (1996) **15**: p. 893-897.
51. P Munger, et al., An inverse problem approach to the correction of distortion in EPI images. *IEEE Trans Med Imaging* (2000) **19**(7): p. 681-689.
52. PJ Reber, et al., Correction of off resonance-related distortion in echo-planar imaging using EPI-based field maps. *Magn Reson Med* (1998) **39**(2): p. 328-330.
53. X Wan, et al., Reduction of geometric and intensity distortions in echo-planar imaging using a multireference scan. *Magn Reson Med* (1997) **37**(6): p. 932-942.
54. NK Chen and AM Wyrwicz, Correction for EPI distortions using multi-echo gradient-echo imaging. *Magn Reson Med* (1999) **41**(6): p. 1206-1213.
55. NK Chen and AM Wyrwicz, Optimized distortion correction technique for echo planar imaging. *Magn Reson Med* (2001) **45**(3): p. 525-528.
56. H Chang and JM Fitzpatrick, A technique for accurate magnetic resonance imaging in the presence of field inhomogeneities. *IEEE Trans Med Imaging* (1992) **11**: p. 319-329.
57. R Bowtell, et al., Correction of geometric distortion in echo planar images. *Proceedings in* SMR (ISMRM) (1994).
58. DG Norris and JM Hutchison, Concomitant magnetic field gradients and their effects on imaging at low magnetic field strengths. *Magn Reson Imaging* (1990) **8**(1): p. 33-37.
59. XJ Zhou, et al., Concomitant magnetic-field-induced artifacts in axial echo planar imaging. *Magn Reson Med* (1998) **39**(4): p. 596-605.
60. MA Bernstein, et al., Concomitant gradient terms in phase contrast MR: analysis and correction. *Magn Reson Med* (1998) **39**(2): p. 300-308.
61. AW Anderson and JC Gore, Analysis and correction of motion artifacts in diffusion weighted imaging. *Magn Reson Med* (1994) **32**(3): p. 379-387.
62. K Butts, et al., Diffusion-weighted interleaved echo-planar imaging with a pair of orthogonal navigator echoes. *Magn Reson Med* (1996) **35**(5): p. 763-770.
63. K Butts, et al., Isotropic diffusion-weighted and spiral-navigated interleaved EPI for routine imaging of acute stroke. *Magn Reson Med* (1997) **38**(5): p. 741-749.
64. D Atkinson, et al., Sampling and reconstruction effects due to motion in diffusion-weighted interleaved echo planar imaging. *Magn Reson Med* (2000) **44**(1): p. 101-109.
65. DG Norris, et al., On the application of ultra-fast RARE experiments. *Magn Reson Med* (1992) **27**(1): p. 142-164.
66. CF Williams, et al., A novel fast split-echo multi-shot diffusion-weighted MRI method using navigator echoes. *Magn Reson Med* (1999) **41**(4): p. 734-742.
67. G Liu, et al., Single-shot diffusion MRI of human brain on a conventional clinical instrument. *Magn Reson Med* (1996) **35**(5): p. 671-677.
68. DC Alsop, Phase insensitive preparation of single-shot RARE: application to diffusion imaging in humans. *Magn Reson Med* (1997) **38**(4): p. 527-533.
69. DG Norris and W Driesel, Online motion correction for diffusion-weighted imaging using navigator echoes: Application to RARE imaging without sensitivity loss. *Magn Reson Med* (2001) **45**(5): p. 729-733.
70. F Schick, SPLICE: sub-second diffusion-sensitive MR imaging using a modified fast spin-echo acquisition mode. *Magn Reson Med* (1997) **38**(4): p. 638-644.
71. D Greitz, et al., Pulsatile brain movement and associated hydrodynamics studied by magnetic resonance phase imaging. The Monro-Kellie doctrine revisited. *Neuroradiology* (1992) **34**(5): p. 370-380.
72. S Skare and JLR Andersson, On the effects of gating in diffusion imaging of the brain using single shot EPI. *Magn Reson Imaging* (2001) **19**(8): p. 1125-1128.

73. AL Alexander, et al., Elimination of eddy current artifacts in diffusion-weighted echo-planar images: the use of bipolar gradients. *Magn Reson Med* (1997) **38**(6): p. 1016-1021.
74. F Calamante, et al., Correction for eddy current induced Bo shifts in diffusion-weighted echo-planar imaging. *Magn Reson Med* (1999) **41**(1): p. 95-102.
75. TG Reese, et al., Diffusion NMR facilitated by a refocused eddy-current EPI pulse sequence. *Proceedings in* ISMRM (1998), Sydney, Australia.
76. DA Feinberg and PD Jakab, Tissue perfusion in humans studied by Fourier velocity distribution, line scan, and echo-planar imaging. *Magn Reson Med* (1990) **16**(2): p. 280-293.
77. O Heid, Eddy current-nulled diffusion weighting. *Proceedings in* ISMRM (2000), Denver, USA.
78. P Jezzard, et al., Characterization of and correction for eddy current artifacts in echo planar diffusion imaging. *Magn Reson Med* (1998) **39**(5): p. 801-812.
79. XJ Zhou, Method to reduce eddy current effects in diffusion-weighted echo planar imaging, *Patent registered* 26 Jan 1999, General Electric Company, USA
80. XJ Zhou, Method for measuring and compensating for spatially and temporally varying magnetic fields induced by eddy currents, *Patent registered* Jun 23 1998, General Electric Company, USA
81. M Terpstra, et al., Localized eddy current compensation using quantitative field mapping. *J Magn Reson* (1998) **131**(1): p. 139-143.
82. NG Papadakis, et al., Gradient preemphasis calibration in diffusion-weighted echo-planar imaging. *Magn. Reson. Med.* (2000) **44**(4): p. 616-624.
83. VJ Schmithorst and BJ Dardzinski, Automatic gradient preemphasis adjustment: A 15-minute journey to improved diffusion-weighted echo-planar imaging. *Magn Reson Med* (2002) **47**(1): p. 208-212.
84. JC Haselgrove and JR Moore, Correction for distortion of echo-planar images used to calculate the apparent diffusion coefficient. *Magn Reson Med* (1996) **36**(6): p. 960-964.
85. ME Bastin, Correction of eddy current-induced artefacts in diffusion tensor imaging using iterative cross-correlation. *Magn Reson Imaging* (1999) **17**(7): p. 1011-1024.
86. AJ de Crespigny and ME Moseley, Eddy Current Induced Image Warping in Diffusion Weighted EPI. *Proceedings in* ISMRM (1998), Sydney, Australia.
87. F Maes, et al., Multimodality image registration by maximation of mutual information. *IEEE Trans Med Imaging* (1997) **16**: p. 187-198.
88. C Studholme, et al., Automated three-dimensional registration of magnetic resonance and positron emission tomography brain images by multiresolution optimization of voxel similarity measures. *J Med Phys* (1997) **24**(1): p. 25-35.
89. MA Horsfield, Warp correcting diffusion-weighted echo planar images by mapping eddy curreent induced fields. *Proceedings in* ISMRM (1999), Philadelphia, USA.
90. MA Horsfield, Mapping eddy current induced fields for the correction of diffusion- weighted echo planar images. *Magn Reson Imaging* (1999) **17**(9): p. 1335-1345.
91. R Bammer and M Auer, Correction of eddy-current induced image warping in diffusion-weighted single-shot EPI using constrained non-rigid mutual information image registration. *Proceedings in* ISMRM (2001), Glasgow.
92. JLR Andersson and S Skare, A model-based method for retrospective correction of geometric distortions in diffusion weighted EPI. *Neuroimage* (2002) **16**(1): *in press.*
93. S Skare and JLR Andersson, Simultaneous correction of eddy currents and motion in DTI using the residual error of the diffusion tensor: Comparisons with mutual information. *Proceedings in* ISMRM (2002), Hawai'i, USA.
94. RA Horn and CR Johnson, Matrix analysis (1985): Cambridge University Press, UK.