Q- space trajectory imaging with positivity constraints (QTI+)

Magnus Herberthson[a], **Deneb Boito**[b,c], **Tom Dela Haije**[d], **Aasa Feragen**[e], **Carl-Fredrik Westin**[f], **Evren Özarslan**[b,c,*]

[a]Department of Mathematics, Linköping University, Linköping, Sweden

[b]Department of Biomedical Engineering Linköping University, Linköping, Sweden

[c]Center for Medical Image Science and Visualization, Linköping University, Linköping, Sweden

[d]Department of Computer Science, University of Copenhagen, Copenhagen, Denmark

[e]Department of Applied Mathematics and Computer Science, Technical University of Denmark, Lyngby, Denmark

[f]Laboratory for Mathematics in Imaging, Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

Abstract

Q-space trajectory imaging (QTI) enables the estimation of useful scalar measures indicative of the local tissue structure. This is accomplished by employing generalized gradient waveforms for diffusion sensitization alongside a diffusion tensor distribution (DTD) model. The first two moments of the underlying DTD are made available by acquisitions at low diffusion sensitivity (b-values). Here, we show that three independent conditions have to be fulfilled by the mean and covariance tensors associated with distributions of symmetric positive semidefinite tensors. We introduce an estimation framework utilizing semi-definite programming (SDP) to guarantee that these conditions are met. Applying the framework on simulated signal profiles for diffusion tensors distributed according to non-central Wishart distributions demonstrates the improved noise resilience of QTI+ over the commonly employed estimation methods. Our findings on a human brain data set also reveal pronounced improvements, especially so for acquisition protocols featuring few number of volumes. Our method's robustness to noise is expected to not only improve the accuracy of the estimates, but also enable a meaningful interpretation of contrast in the derived scalar maps. The technique's performance on shorter acquisitions could make it feasible in routine clinical practice.

*Corresponding author. evren.ozarslan@liu.se (E. Özarslan).

Keywords

Diffusion; MRI; Constrained; Positive definite; QTI; Multidimensional; mddMRI; Covariance; Microscopic anisotropy

---

1. Introduction

Determining the local structure of neural tissue using diffusion MRI has already made an impact in neuroscience and radiology. Diffusion MRI's sensitivity to tissue microstructure is exploited and interpreted through models that provide a simplified picture of the complex tissue makeup. The parameters of an adequate model reflect the key characteristics of the tissue that influence the stochastic movement of the water molecules. Measuring the diffusional process and estimating such model parameters from the acquired data are the two essential components of structure determination via diffusion MRI.

In q-space trajectory imaging (QTI) (Westin et al., 2016), diffusion sensitization is achieved via general time-dependent gradient waveforms while the tissue is envisioned to have numerous non-exchanging compartments. Diffusion is characterized by a diffusion tensor within each of these compartments. Consequently, the voxel is represented by a diffusion tensor distribution (DTD) (Jian et al., 2007). QTI exploits the sensitivity of the diffusion MRI signal to the statistical moments of the parameters characterizing the microscopic domain (Özarslan et al., 2011). By doing so, QTI provides a simple means of relating the signal obtained via general gradient waveforms to the DTD, which is key for introducing meaningful MRI 'biomarkers' in QTI. Under the assumptions of the DTD picture, the effect of all measurement parameters is captured by a positive-semidefinite tensor, referred to as the b-tensor (Mattiello et al., 1994), and denoted by $\mathbf{B}_{ij}$ in this work. The level of diffusion sensitization is usually quantified by the trace of this tensor, denoted by b.

Common clinical MRI examinations of the neural tissue probe the low-b regime of the MR signal attenuation. As shown by Westin et al. (2016), the data in this regime reveal the mean and covariance tensors of the underlying DTD. The former is a 3×3 symmetric positive semidefinite matrix, while the covariance tensor has the symmetries of the fourth order elasticity tensor in mechanics (Basser and Pajevic, 2003). Once estimated, these two tensors are employed in computing several scalar measures that characterize macroscopic and microscopic anisotropies, orientational coherence and size variance of the subdomains making up the tissue. Thus, a key step in obtaining reliable estimates of these quantities involves accurate estimation of the mean and covariance tensors from the data. In this study, we investigate possible improvements in the estimates of the QTI-derived parameters when several necessary nonnegativity conditions are enforced.

Improvements due to constrained optimization have been reported for diffusion MRI models developed for traditional pulsed field gradient measurements of Stejskal and Tanner (1965). For example, diffusion tensor imaging (DTI) (Basser et al., 1994a; 1994b) has benefited from estimation schemes (Koay, 2010; Koay et al., 2006; Lenglet et al., 2006; Pennec et al., 2006; Wang et al., 2004) that ensure that the diffusion tensor is positive semidefinite—a condition that follows from the physics of diffusion. The estimation problems for models

that go beyond DTI (Jensen et al., 2005; Özarslan and Mareci, 2003; Tournier et al., 2007) have also been studied via methods that enforce relevant constraints (Barmpoutis et al., 2012; 2009; Chen et al., 2013; Ghosh et al., 2014; Qi et al., 2010; Veraart et al., 2011).

In a recent work, Dela Haije et al. (2020) considered three such prominent models, namely, spherical deconvolution (Tournier et al., 2004), diffusion kurtosis imaging (Jensen et al., 2005; Liu et al., 2004), and mean apparent propagator MRI (Özarslan et al., 2013) and formulated several sum-of-squares (SoS) constraints arising from the nonnegativity of the relevant distribution functions yielding remarkable improvements in the model estimates over earlier methods. Enforcing them in the estimation via semidefinite programming (SDP) guaranteed the fulfillment of such constraints in contrast to earlier methods that either did not account for them or imposed them "softly," i.e., did not ensure strict adherence to the relevant constraints. "Softly" imposed constraints were also initially considered for QTI in (Jeurissen et al., 2019), highlighting the interest and need for more sophisticated fitting approaches to be used with this method.

To investigate the effects of constrained optimization for the QTI technique, we devised an estimation framework that guarantees the fulfillment of three conditions that mean and covariance tensors of DTDs have to respect. Following the naming convention in Dela Haije et al. (2020), we refer to our method as QTI+. After introducing our notation and providing an overview of the QTI model, we introduce the constraints to be imposed. Several methods for estimating the mean and covariance tensors as well as a test for checking the fulfillment of one of the constraints are introduced. Simulated signals for non-central Wishart distributed DTDs (Shakya et al., 2017) are employed to compare the performance of commonly-employed methods with ours. We also provide analyses on tensor-valued diffusion encoded brain data (Szczepankiewicz et al., 2019) and assess the performance of our framework on data sets with few number of acquisitions.

2. Background

Our notation

There is a multitude of notations for tensors. Here, we describe the notations and conventions we employ. In this study, there is no need to make a distinction between contra- and covariant tensors. Thus, all indices are written as subscripts.

Scalars are denoted with italic characters, while matrices and second order tensors are denoted with boldface characters. Blackboard bold (double struck) characters are used for fourth-order tensors. For example, $\mathbb{A}_{ijk\ell}$ is a fourth order tensor whose $ijk\ell$th component is $A_{ijk\ell}$. Fourth-order tensors considered in this work can also be represented by 6×6 matrices. To make the distinction clear, we employ the following convention:

- Latin letters i, j, k, ℓ range from 1 to 3.

- Early Greek letters α, β, and γ range from 1 to 6.

Thus, $\mathbb{A}_{ijk\ell}$ and $\mathbf{A}_{\alpha\beta}$ are the fourth order and second order representations of the same tensor. When used with double struck and boldface characters, the indices are retained just

to inform about the order of the tensor, which is the number of indices and the range of those indices; they do not refer to a particular component of the tensor.

We employ the Einstein summation convention, which is usually described as "all repeated indices are summed over." E.g., $A_{ii} = \sum_{i=1}^{3} A_{ii}$ is the trace of the matrix $\mathbf{A}_{ij}$, while $C_{ik} = A_{ij}B_{jk} = \sum_{j=1}^{3} A_{ij}B_{jk}$ is the ikth component of the product of matrices $\mathbf{A}_{ij}$ and $\mathbf{B}_{ij}$.

QTI

In this work, we are interested in the statistical properties of a distribution of diffusivity tensors $\mathbf{D}_{ij}$, represented by a family of samples $\{\mathbf{D}_{ij}^{(n)}\}$. Here, since it represents a diffusivity, which is proportional to second moment of displacements, each second order tensor $\mathbf{D}_{ij}^{(n)}$ is symmetric and positive semidefinite. With $\langle\cdot\rangle$ indicating mean (expectation) value, we would like to estimate the mean diffusivity tensor[1]

$$\widehat{\mathbf{D}}_{ij} = \langle\mathbf{D}_{ij}\rangle$$

and the covariance, which in this case becomes the fourth order (covariance) tensor $\mathbb{C}$, defined as

$$\mathbb{C}_{ijk\ell} = \langle\mathbf{D}_{ij} - \widehat{\mathbf{D}}_{ij}\rangle \otimes \langle\mathbf{D}_{k\ell} - \widehat{\mathbf{D}}_{k\ell}\rangle.$$

This tensor has the so called minor ($C_{ijk\ell} = C_{jik\ell}, C_{ijk\ell} = C_{ij\ell k}$) and major ($C_{ijk\ell} = C_{k\ell ij}$) symmetries, which result in $\mathbb{C}_{ijk\ell}$ having 21 independent components. It is (as usual) possible to express $\mathbb{C}_{ijk\ell}$ in terms of the (second) moment tensor $\mathbb{M}_{ijk\ell} = \langle\mathbf{D}_{ij} \otimes \mathbf{D}_{k\ell}\rangle$ through the relationship

$$\mathbb{C}_{ijk\ell} = \mathbb{M}_{ijk\ell} - \widehat{\mathbf{D}}_{ij} \otimes \widehat{\mathbf{D}}_{k\ell}.$$

$\mathbb{M}_{ijk\ell}$ has the same symmetries and degrees of freedom as $\mathbb{C}_{ijk\ell}$.

The QTI signal's dependence on the b-matrix $\mathbf{B}_{ij}$ is given by (Westin et al., 2016)

$$S(\mathbf{B}_{ij}) = S_0 \exp\left(-B_{ij}\widehat{D}_{ij} + \frac{1}{2}B_{ij}B_{kl}C_{ijk\ell}\right), \tag{1}$$

where S_0 is the signal with no diffusional attenuation, i.e., when $\mathbf{B}_{ij} = \mathbf{0}$. Thus, given a family of measurement tensors $\{\mathbf{B}_{ij}^{(n)}\}_{n=1}^{N}$ and the corresponding signal values $S_1, S_2, \ldots, S_N$ with $S_n = S(\mathbf{B}_{ij}^{(n)})$, the task, given the model (1), is to produce estimates of S_0, $\widehat{\mathbf{D}}_{ij}$ and $\mathbb{C}_{ijk\ell}$.

---

[1]The mean is defined in the traditional sense. For a family of N tensors, $\widehat{\mathbf{D}}_{ij} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{D}_{ij}^{(n)}$.

The Voigt notation

The diffusivity tensors $\mathbf{D}_{ij}$ in three-dimensional space are symmetric second order tensors. The set of all symmetric second order tensors forms a vector space V of dimension six, and this space is equipped with a natural scalar product: $< \mathbf{A}_{ij}, \mathbf{B}_{ij} >= A_{ij}B_{ij}$. Hence one can introduce an orthonormal basis $\{\mathbf{e}_{ij}^{(\beta)}\}_{\beta = 1}^{6}$ and express any tensor in V as

$$A_{ij} = \sum_{\beta = 1}^{6} a_\beta e_{ij}^{(\beta)}.$$

These six coordinates a_β are customarily put into a vector with six elements, and this is referred to as the Voigt notation. See Appendix A for our choice for the basis.

This approach yields various representations of the covariance tensor $\mathbb{C}_{ijk\ell}$ as well. Because of the symmetries of $\mathbb{C}_{ijk\ell}$, this tensor can be regarded as a symmetric mapping $V \rightarrow V$, and in turn (given an orthonormal basis for V) represented as a symmetric 6×6 matrix, which is consistent with $\mathbb{C}_{ijk\ell}$ having 21 degrees of freedom. This matrix will be denoted by $\mathbf{C}_{\alpha\beta}$ where $1 \leq \alpha, \beta \leq 6$ as described above.

We can proceed in a similar manner. The set of symmetric mappings $V \rightarrow V$ constitute a vector space of dimensions 21 and, again, given an orthonormal-basis for this space, any tensor $\mathbb{C}$ (with the appropriate symmetries) can be represented by a vector with 21 elements.

3. Theory

As mentioned in the previous section, given a family of measurements tensors $\{\mathbf{B}_{ij}^{(n)}\}_{n = 1}^{N}$, the set of corresponding signal values $\{S_n\}_{n = 1}^{N}$ and the model (1), the task is to produce estimates of $S_0, \widehat{\mathbf{D}}_{ij}$ and $\mathbb{C}_{ijk\ell}$. Assuming approximately Gaussian noise, this is achieved by finding the $S_0, \widehat{\mathbf{D}}_{ij}$ and $\mathbb{C}_{ijk\ell}$, which minimize the 'error,' i.e.,

$$\underset{S_0, \widehat{\mathbf{D}}_{ij}, \mathbb{C}_{ijk\ell}}{\operatorname{argmin}} \sum_{n = 1}^{N} \left| S_n - S_0 e^{- B_{ij}^{(n)} \widehat{D}_{ij} + \frac{1}{2} B_{ij}^{(n)} B_{kl}^{(n)} C_{ijk\ell}} \right|^{2}. \tag{2}$$

Here we make two remarks:

1. Even if there is a global minimum, it is not easy to specify in advance a minimizing routine, which is guaranteed to find the minimum.

2. If a minimum is found, the obtained estimates of $\widehat{\mathbf{D}}_{ij}$ and $\mathbb{C}_{ijk\ell}$ may be unacceptable.

We start by addressing the second issue in the following subsection.

3.1. Positivity conditions

There are a number of positivity conditions one can impose on the estimates, which have to be met in order for their interpretation to be physically reasonable. Here, we will give three such conditions, which are independent in the sense that any two of them do not imply the other. See Appendix B.

The first condition is on $\widehat{\mathbf{D}}_{ij}$, namely, it should represent a diffusivity and thus $\widehat{\mathbf{D}}_{ij}$ (in addition to being symmetric) is positive semi-definite. We express this condition as $\widehat{\mathbf{D}}_{ij} \succeq 0$.

The second condition is similar. From the fact that $\mathbb{C}_{ijk\ell}$ represents a covariance (so that $C_{ijkl} A_{ij} A_{kl} \geq 0$ for all symmetric matrices $\mathbf{A}_{ij}$), it is necessary that when cast as a 6×6 matrix $\mathbf{C}_{\alpha\beta}$, this matrix should also be positive semi-definite, i.e., $\mathbf{C}_{\alpha\beta} \succeq 0$.

The third condition is on $\mathbb{M}_{ijk\ell}$, i.e., concerns $\langle \mathbf{D}_{ij} \otimes \mathbf{D}_{k\ell} \rangle$. Since $\mathbb{M}_{ijk\ell}$ is the mean of tensor products of diffusion tensors, each of which is positive semi-definite, this property is carried over to $\mathbb{M}_{ijk\ell}$. The conclusion is that for any vector $\boldsymbol{u}_i \in \mathbf{R}^3$, the symmetric second order tensor, whose ijth component is $M_{ijk\ell} u_k u_\ell$ should be positive semi-definite. In other words, for any pair of vectors $\boldsymbol{v}_i$ and $\boldsymbol{u}_i$, we must have $M_{ijk\ell} v_i v_j u_k u_\ell \geq 0$.

We shall refer to these three conditions as '(d)', '(c)', and '(m)' where the letters indicate the tensor on which the conditions are imposed. To summarize, our conditions are, then,

(d) $\widehat{\mathbf{D}}_{ij} \succeq 0$,

(c) $\mathbf{C}_{\alpha\beta} \succeq 0$, and

(m) for all $\boldsymbol{v}_i$ and $\boldsymbol{u}_i$, $M_{ijk\ell} v_i v_j u_k u_\ell \geq 0$.

Let us also remark that the condition $S_0 \geq 0$ is obviously also required, but that it need not be imposed explicitly (this can be inferred from the fact that all $S_n \geq 0$).

3.2. Linearizing the equation and the least squares solution

As mentioned above, it is not trivial to ensure that a global minimum to (2) is found. However, there is a related problem for which a global minimum is guaranteed to be found. Namely, by taking the logarithm of (1), the model is linearized as

$$\ln S(\mathbf{B}_{ij}) = \ln S_0 - B_{ij}\widehat{D}_{ij} + \frac{1}{2} B_{ij} B_{k\ell} C_{ijk\ell}. \tag{3}$$

Due to the heteroscedasticity caused by taking the logarithm of the signal, the minimization problem arising from (3) is the weighted problem (Basser et al., 1994a; Bevington and Robinson, 2003)

$$\underset{S_0, \widehat{\mathbf{D}}_{ij}, \mathbb{C}_{ijk\ell}}{\text{argmin}} \sum_{n=1}^{N} S_n^2 \left| \ln(S_n) - \ln(S_0) + B_{ij}^{(n)} \widehat{D}_{ij} - \frac{1}{2} B_{ij}^{(n)} B_{kl}^{(n)} C_{ijk\ell} \right|^2, \tag{4}$$

which is an approximation to (2); see Appendix C. Using the Voigt notation, the unknowns determining $\ln(S_0)$, $\widehat{\mathbf{D}}_{ij}$ and $\mathbb{C}_{ijk\ell}$ can be stacked into a vector $\boldsymbol{x}$ with $1 + 6 + 21 = 28$ components, i.e., $\boldsymbol{x} = (x_1, \ldots, x_{28})^\top$. The components of $\boldsymbol{x}$ could be determined through[2]

$$\ln(S_0) = x_1$$

$$\widehat{\mathbf{D}}_{ij} \doteq \frac{1}{\sqrt{2}} \begin{pmatrix} \sqrt{2}x_2 & x_5 & x_6 \\ x_5 & \sqrt{2}x_3 & x_7 \\ x_6 & x_7 & \sqrt{2}x_4 \end{pmatrix}$$

$$\mathbf{C}_{\alpha\beta} \doteq \frac{1}{\sqrt{2}} \begin{pmatrix} \sqrt{2}x_8 & x_{11} & x_{12} & x_{17} & x_{18} & x_{14} \\ x_{11} & \sqrt{2}x_9 & x_{13} & x_{19} & x_{15} & x_{20} \\ x_{12} & x_{13} & \sqrt{2}x_{10} & x_{16} & x_{21} & x_{22} \\ x_{17} & x_{19} & x_{16} & \sqrt{2}x_{23} & x_{26} & x_{27} \\ x_{18} & x_{15} & x_{21} & x_{26} & \sqrt{2}x_{24} & x_{28} \\ x_{14} & x_{20} & x_{22} & x_{27} & x_{28} & \sqrt{2}x_{25} \end{pmatrix},$$

and Eq. (4) can be formulated as the weighted linear least squares (WLLS) problem

$$\mathbf{A}\boldsymbol{x} = \boldsymbol{y} \tag{5}$$

where the vector $\boldsymbol{y} = (S_1 \ln(S_1), \ldots, S_N \ln(S_N))^\top$ contains the weighted signals and the $N \times 28$ matrix $\mathbf{A}$ is formed by the signal values S_n and the measurement tensors $\mathbf{B}_{ij}^{(n)}$.

Without further restrictions, the minimizing vector $\boldsymbol{x}$ can easily be found by standard linear regression routines. However, we also give two other formulations, which are equivalent to (5) in the unconstrained case, but differ when it comes to imposing the positivity constraints (d), (c), and (m).

3.3. Quadratic programming (QP) for the linearized problem

First, we note that since the least squares solution minimizes $\|\mathbf{A}\boldsymbol{x} - \boldsymbol{y}\|^2$, i.e.,

$$\boldsymbol{x}^\top \mathbf{A}^\top \mathbf{A}\boldsymbol{x} - 2\boldsymbol{y}^\top \mathbf{A}\boldsymbol{x} + \boldsymbol{y}^\top \boldsymbol{y} \tag{6}$$

this can also be solved using quadratic programming (Nocedal and Wright, 2006). Note that if $\mathbf{A}$ has full rank, $\mathbf{Q} = \mathbf{A}^\top \mathbf{A}$ is positive definite. Through the substitution $\boldsymbol{c} = -2A^\top\boldsymbol{y}$, the (least squares) solution to (5) can also be found as the solution to

$$\underset{\boldsymbol{x}}{\text{argmin}} \, \boldsymbol{x}^\top \mathbf{Q}\boldsymbol{x} + \boldsymbol{c}^\top \boldsymbol{x}. \tag{7}$$

---

[2]Here, "$\doteq$" is used to indicate that the following matrix is just one representation of the tensor in a particular basis.

Here, we ignored the constant factor $y^\top y$ as we are interested in the minimizing argument.

3.4. Semidefinite programming (SDP) for the linearized problem

Eq. (7) can be further reformulated. First, we note that the minimizing argument x can be found by minimizing an auxiliary variable t under the condition $t \quad x^\top Q x + c^\top x$, i.e., we are interested in

$$\underset{t,\,x}{\operatorname{argmin}}t, \text{ subject to } t - x^\top Q x - c^\top x \geq 0\,. \tag{8}$$

With $\mathbf{P}$ being a square matrix such that $\mathbf{P}^\top \mathbf{P} = \mathbf{Q}$, and with $\mathbf{I}$ being the identity matrix of the same size as $\mathbf{Q}$, this can be formulated as

$$\underset{t,\,x}{\operatorname{argmin}}t, \text{ subject to } \begin{pmatrix} \mathbf{I} & \mathbf{P}x \\ x^\top \mathbf{P}^\top & t - c^\top x \end{pmatrix} \succeq 0 \tag{9}$$

which shows that we can employ SDP as well to solve this problem (see Appendix D).

3.5. Imposing positivity conditions: Nonlinear least squares with (d) and (c) constraints (NLLS(dc))

Comparing the original problem (2), and the various linearized versions (4),(7), and (8), they all differ when it comes to imposing the positivity constrains (d), (c), and (m).

For Eq. (2), it is possible to impose conditions (d) and (c), by utilizing the Cholesky decomposition, i.e., the fact that any symmetric positive semi-definite matrix $\mathbf{A}$ can be written $\mathbf{A} = \mathbf{L}\mathbf{L}^\top$, where $\mathbf{L}$ is a lower triangular matrix with positive diagonal entries.

To this end, we cast $\mathbb{C}_{ijk\ell}$ in its 6×6 matrix form $\mathbf{C}_{\alpha\beta}$. We also introduce a fourth order tensor $\mathbb{B}'_{ijk\ell}$ whose $ijk\ell$th component is $B_{ij}B_{k\ell}$ and its 6×6 matrix form is $\mathbf{B}'_{\alpha\beta}$. We can use the ansatz

$$\begin{aligned} \widehat{D}_{ij} &= L_{ik}L_{jk} \text{ and} \\ C_{\alpha\beta} &= \Lambda_{\alpha\gamma}\Lambda_{\beta\gamma} \end{aligned} \tag{10}$$

where both $\mathbf{L}_{ij}$ and $\mathbf{\Lambda}_{\alpha\beta}$ are lower triangular matrices with positive diagonal entries. The problem (2), then, becomes

$$\underset{S_0,\,\mathbf{L}_{ij},\,\mathbf{\Lambda}_{\alpha\beta}}{\operatorname{argmin}} \sum_{n=1}^{N} \left| S_n - S_0 e^{-B_{ij}^{(n)}L_{ik}L_{jk} + \frac{1}{2}B_{\alpha\beta}^{'(n)}\Lambda_{\alpha\beta}\Lambda_{\beta\gamma}} \right|^2, \tag{11}$$

which guarantees that (d) and (c) (but not necessarily (m)) are satisfied.

3.6. Imposing positivity conditions: SDP with (d) and (c) constraints (SDP(dc))

The estimation schemes based on the linearized version of the model (weighted linear, quadratic programming, and semidefinite programming) described above, which all produce global minima, differ when it comes to imposing the positivity constraints. In particular,

the semi-definite programming (SDP) framework is particularly well-suited for imposing (d) and (c). With $\boldsymbol{x} = (x_1, \dots, x_{28})^\top$ as before, the conditions (d) and (c), namely that $\widehat{\mathbf{D}}_{ij}$ and $\mathbf{C}_{\alpha\beta}$ are nonnegative, then fit directly into the SDP framework, i.e., we can impose (d) and (c) by solving

$$\underset{t,\,\boldsymbol{x}}{\mathrm{argmin}}\, t, \text{ subject to} \begin{pmatrix} \boldsymbol{I} & \boldsymbol{Px} & 0 & 0 \\ \boldsymbol{x}^\top \boldsymbol{P}^\top & t - \boldsymbol{c}^\top \boldsymbol{x} & 0 & 0 \\ 0 & 0 & \widehat{\mathbf{D}}_{ij} & 0 \\ 0 & 0 & 0 & \mathbf{C}_{\alpha\beta} \end{pmatrix} \succeq 0. \tag{12}$$

Again, condition (m) is not imposed, which is mostly because $\mathbb{M}_{ijk\ell}$ depends quadratically on $\langle \mathbf{D}_{ij} \rangle$.

3.7. An SDP scheme for checking if condition (m) is fulfilled ((m)-check)

Addressing (11) or (12), we get (initial) estimates of S_0, $\langle \mathbf{D}_{ij} \rangle$, $\mathbb{C}_{ijk\ell}$ and $\mathbb{M}_{ijk\ell}$, where the constraints (d) and (c) are imposed. To determine whether condition (m) needs to be imposed, we first check whether it is violated or not. Hence, we pose the question: for a given estimate of $\mathbb{M}_{ijk\ell}$, is it true that

$$\text{for all } \boldsymbol{v}_i, \boldsymbol{u}_i, \qquad \boldsymbol{M}_{ijk\ell}\, v_i\, v_j\, u_k\, u_l \geq 0 \, ? \tag{13}$$

Again, this can be investigated with SDP by addressing a feasibility problem. As explained in Appendix E, it is possible from $\mathbb{M}_{ijk\ell}$ to construct a 9×9 matrix $\mathbf{M}$ where each entry is a first order polynomial in the parameters $\ell_1, \ell_2, \dots, \ell_9$, and check (using SDP) whether there are feasible solutions to the problem

$$\underset{\ell}{\min}\, 0, \text{ subject to } \mathbf{M}(\boldsymbol{\ell}) \succeq 0. \tag{14}$$

Here we have put the parameters ℓ_i into a vector: $\boldsymbol{\ell} = (\ell_1, \ell_2, \dots, \ell_9)^\top$. This expression differs from the earlier adaptations of the SDP method in that we are only interested in finding out whether a solution fulfilling all the constraints exists. Thus, the function to be minimized is unimportant, and is taken to be 0 by choice. If[3] the SDP routine finds a vector $\boldsymbol{\ell}$ condition (m), i.e., (13) is satisfied.

3.8. Imposing positivity conditions: SDP with (c) and (m) constraints (SDP(dcm))

In the case when condition (m) is violated, it is imposed in the following way. From the estimate at hand, we fix S_0 and $\widehat{\mathbf{D}}_{ij}$, i.e., $x_1, x_2, \dots, x_7$,[4] so that $\mathbb{M}_{ijk\ell}$ is linear in the remaining variables $x_8, \dots, x_{28}$. These are then re-estimated to ensure that both (c) and (m)

---

[3]We do not have strict equivalence, see Appendix E

[4]This may seem like a restriction. In our experience, however, the estimates of $x_1, x_2, \dots, x_7$ are relatively 'stable' as compared to $x_8, \dots, x_{28}$.

are satisfied. Again, this can be accomplished with SDP, and we refer the reader to Appendix F for the formulation. In short, with $x_1, \ldots, x_7$ fixed, we set $\tilde{x} = (x_8, \ldots, x_{28})^\top$ and $\ell = (\ell_1, \ell_2, \ldots, \ell_6)^\top$ and can assert (c) and (m) by solving

$$\min_{t, \tilde{x}, l} t, \text{ subject to} \begin{pmatrix} \mathbf{I} & \widetilde{\mathbf{P}}\tilde{\mathbf{x}} & 0 & 0 \\ \tilde{\mathbf{x}}^T\widetilde{\mathbf{P}}^T & t - \tilde{\mathbf{c}}^T\tilde{\mathbf{x}} & 0 & 0 \\ 0 & 0 & \mathbf{C}_{\alpha\beta} & 0 \\ 0 & 0 & 0 & \mathbf{M}(l) \end{pmatrix} \geq 0. \tag{15}$$

Note that we refer to this scheme as "SDP(dcm)" as it relies on a previous estimate of $\widehat{\mathbf{D}}_{ij}$, which is positive semidefinite. Thus, the end result is guaranteed to fulfill all three conditions (d), (c), and (m).

3.9. Rank deficient estimation

Normally, it is assumed that the matrix $\mathbf{A}$ in the (weighted) least squares problem (5) is of full rank, which in our case is $1 + 6 + 21 = 28$. This is achieved by having a sufficiently rich family of measurement tensors $\{\mathbf{B}_{ij}^{(n)}\}_{n=1}^{N}$, which 'spans the parameter space.' However, for practical reasons there is a trade-off since there is also a desire to keep the measurement protocol short.

Acquisition protocols could feature measurements having $\mathbf{B}_{ij}$ tensors with quite general features (e.g., non-axisymmetric or anisotropic rank-3 tensors), which could offer some benefits (Herberthson et al., 2019). However, in the current practice, it is quite common to use measurement tensors which fall into one of the following three classes: (i) LTE (linear tensor encoding) where each measurement tensor is the outer product of some vector with itself, implying the eigenvalues of each such $\mathbf{B}_{ij}^{(n)}$ being $\{\lambda^{(n)}, 0, 0\}$ for some $\lambda^{(n)} > 0$. (ii) PTE (planar tensor encoding), where each such measurement tensor $\mathbf{B}_{ij}^{(n)}$ has eigenvalues $\{\lambda^{(n)}, \lambda^{(n)}, 0\}$, $\lambda^{(n)} > 0$. (iii) STE (spherical tensor encoding) where each such measurement tensor is proportional to the identity matrix.

For protocols that use measurements of only type (i) and (iii), i.e., LTE and STE, this will lead to a rank deficient matrix $\mathbf{A}$, with (maximum) rank $1 + 6 + 16 = 23$. The reason for this is that with measurement tensors of type (i), i.e., LTE, the measurements are only sensitive to the completely symmetric part of $\mathbb{C}_{ijk\ell}$, and the space of such tensors has dimension 15. Furthermore, since all isotropic (STE) measurement tensors, i.e., tensors of type (iii), are proportional to each other, they will only be capable of measuring one more dimension in the parameter space.

This raises two questions. First, how does this affect the estimates and the routines to find these? The observation is that the solution is non-unique and also that derived matrices like $\mathbf{Q} = \mathbf{A}^\top\mathbf{A}$ become singular (positive semidefinite but not positive definite). There are various ways to handle this challenge; the most common with degenerate least squares problem is perhaps to pick the solution vector with minimal norm. This can be achieved by employing

a pseudoinverse or using a so called subspace reduction. When integrated into the WLLS method, we refer to the technique as "WLLS(ss)."

The second question is: how are the presented results affected by the rank deficiency of the design matrix? Indeed, because of this degeneracy, many $\mathbb{C}_{ijk\ell}$ tensors are equivalent in terms of their ability to represent the data. However, as shown in Appendix G, all scalar measures to be used in this work are insensitive to this degeneracy, the exception being the Frobenius norm involving the covariance tensor $\mathbb{C}_{ijk\ell}$ in the case of rank deficient estimation.

4.   Methods

4.1.   Implementation

In this section, we describe our strategy incorporating the techniques described above into a unified framework. The framework contains the following steps:

1.   SDP(dc): See Section 3.6 and Eq. (12). The result of this step could be taken as the final result. However, it can also be treated as an initial estimate and fed into Step 2, for which heteroscedasticity is not an issue. It can also be fed into Step 3 (and 4 if necessary) for imposing condition (m).

2.   NLLS(dc): See Section 3.5 and Eq. (11). This step employs the original (nonlinear) model, which should in general reduce the residues (in the nonlinear form of the model) obtained via SDP(dc). In rare cases when NLLS(dc) fails to produce an improvement over SDP(dc), which can occur when the modified Cholesky decomposition leads to poor initial estimates, the SDP(dc) outcome is retained. The result of this step could be used as the final result. However, if condition (m) is to be imposed, further analysis is necessary.

3.   (m)-check: See Section 3.7 and Eq. (14). If the voxel satisfies condition (m), no further step is necessary. If not, the next step is employed.

4.   SDP(dcm): See Section 3.8 and Eq. (15).

All the fitting routines were implemented in Matlab (The Mathworks Inc, Natick, Massachussets). For SDP we used CVX, a package for specifying and solving convex programs in Matlab (Grant and Boyd, 2008; 2014). In steps 1, 3, and 4 CVX calls the solver MOSEK version 9.1.9 (MOSEK ApS, Denmark). For the non-linear fit in step 2 we used the Matlab routine *lsqcurvefit*. For the standard QTI analysis, the multidimensional dMRI toolbox, provided at https://github.com/markus-nilsson/mddmri was employed. The estimation methods were also independently implemented in Mathematica (Wolfram Research Inc., Champaign, IL, USA) to check for consistency.

4.2.   Simulations

We performed simulations to assess the impact of adding different constraints to the estimation of S_0, $\widehat{\mathbf{D}}_{ij}$ and $\mathbb{C}_{ijk\ell}$. We considered an independent method for the generation of the diffusion signals to be fitted with both the available and proposed methods. For this task, we chose the non-central Wishart distribution whose mean and covariance tensors can

be derived analytically. As discussed elsewhere (Herberthson et al., 2019; Jian et al., 2009; 2007), the diffusion MR signal is the Laplace transform of the underlying DTD. For the case of non-central Wishart distributions, the result is provided by Mayerhofer (2013). Given some measurement tensor $\mathbf{B}_{ij}$, the signal for a non-central Wishart distribution with the non-centrality matrix[5] $\mathbf{\Omega}_{ij}$, the scale parameter (degrees of freedom) p, and scale matrix $\mathbf{\Sigma}_{ij}$ is given by[6] (Shakya et al., 2017)

$$S(\mathbf{B}) = S_0 \mid \mathbf{I} + \mathbf{\Sigma B} \mid^{-p} \exp(-\mathrm{Tr}[\mathbf{B}(\mathbf{I} + \mathbf{\Sigma B})^{-1}\mathbf{\Omega}]) \tag{16}$$

where "$\mid \cdot \mid$" indicates matrix determinant and we dropped the subscripts for brevity. We employed this expression for the signal to find $\widehat{D}_{ij}$ and $C_{ijk\ell}$ in the model (1), the result being

$$\widehat{D}_{ij} = p\Sigma_{ij} + \Omega_{ij} \tag{17}$$

$$
\begin{aligned}
C_{ijk\ell} = &\frac{p}{2}(\Sigma_{i\ell}\Sigma_{jk} + \Sigma_{ik}\Sigma_{j\ell}) \\
&+ \frac{1}{2}(\Sigma_{jk}\Omega_{i\ell} + \Sigma_{i\ell}\Omega_{jk} + \Sigma_{j\ell}\Omega_{ik} + \Sigma_{ik}\Omega_{j\ell}).
\end{aligned}
\tag{18}
$$

The derivation of the above expressions is provided in Appendix H. The non-central Wishart distribution simulated here has higher order cumulants, which is what we expect to have in neural tissue as well. However, we provide the expressions for the first two cumulants, which are to be estimated using the model.

In our simulations, we took 0.7 $\mu m^2/ms$ for $\frac{1}{3}\mathrm{Tr}(\widehat{\mathbf{D}}_{ij})$, and set $\mathbf{\Sigma}_{ij} = \widehat{\mathbf{D}}_{ij}/(5p)$. We performed two simulations, first having isotropic $\widehat{\mathbf{D}}_{ij}$ with $p = 2$, while in the other we took $p = 4$ and the eigenvalues of $\widehat{\mathbf{D}}_{ij}$ to be 0.6, 0.2, and 1.3 $\mu m^2/ms$. Note that p determines the asymptotic behavior of the signal decay curve; see (16). For $p = 2$. one obtains a signal decay consistent with Debye-Porod law, which is the expected decay for diffusion in porous media measured via Stejskal-Tanner pulse sequence featuring narrow pulses (Sen et al., 1995). For wide pulses, such slow decay is replaced by a steeper one (Özarslan et al., 2018). Fig. 1 shows the joint distributions of Mean Diffusivity (MD) and Fractional Anisotropy (FA) for the tensor distributions whose averages are the anisotropic and isotropic $\widehat{\mathbf{D}}_{ij}$ considered in the simulations.

Two sets of measurement tensors $\mathbf{B}_{ij}$ were used to generate the signal for the simulations. The shorter protocol having 56 measurements is referred to as p56s. This protocol combines $\mathbf{B}_{ij}$ tensors of rank 1 and 3, i.e., linear (LTE) and spherical (STE) encodings, and one measurement without diffusion encoding. A longer protocol, referred to as p217, consisting of 217 measurements was also considered. This longer protocol combines encoding tensors of rank 1, 2, and 3 as well as 13 measurements without diffusion weighting. The two

---

[5]Here we are following the notation of Letac and Massam (1998). In this work, which emphasizes the relation to gamma distributions, the Wishart distribution is written $\gamma_{p,\Sigma}$. It is related to the more common notation $W_d(p, \Sigma)$ by $\gamma_{p,\Sigma} = W_d(2p, \Sigma/2)$.
[6]We note and correct an error in the order of $\mathbf{\Sigma}$ and $\mathbf{B}$ matrices in (Shakya et al., 2017).

protocols are summarized in Table 1. The protocol p56s can be found at http://github.com/filip-szczepankiewicz/fwf_seq_resources/tree/master/GE. The longer protocol p217 is a subset of the one described in Szczepankiewicz et al. (2019) and available at https://github.com/filip-szczepankiewicz/Szczepankiewicz_DIB_2019. In particular, the repeated STE measurements were removed from the full protocol in Szczepankiewicz et al. (2019). Note that the shorter protocol p56s leads to rank-deficient design matrices while the matrices associated with the long protocol p217 are not rank-deficient.

Noisy Gaussian and Rician distributed signals were synthesized by adding Gaussian noise to the real and to the real and imaginary parts of the analytical signals obtained from (16), respectively. The standard deviations of the Gaussian noise were taken to be $\sigma = [0, 0.020\ 0.056, 0.092, 0.128, 0.164, 0.200]$, which correspond to SNR[7] values of SNR $= 1/\sigma = [\infty, 50, 18, 11, 8, 6, 5]$ for the non-diffusion weighted signal, where this signal was taken to be $S_0 = 1$. For each standard deviation, 1000 noisy signals were generated and then fitted using the various QTI and QTI+ estimation schemes. We compared the results produced by the different methods using metrics derived from $\widehat{\mathbf{D}}_{ij}$ and $\mathbb{C}_{ijkl}$. These metrics involved both a direct measure of the distance between the analytical and the estimated tensors, given by the Frobenius norm of the difference between the reference and estimated tensors normalized with the Frobenius norm of the reference tensors, $||\Delta\widehat{\mathbf{D}}_{ij}||\ /\ ||\widehat{\mathbf{D}}_{ij}||$ and $||\Delta\mathbb{C}_{ijkl}||\ /\ ||\mathbb{C}_{ijkl}||$, and invariants obtained from the estimated $\widehat{\mathbf{D}}_{ij}$ (fractional anisotropy (FA), mean diffusivity (MD), and macroscopic anisotropy (C_M)) as well as those that utilize additional information from the covariance tensor (microscopic anisotropy (C_μ), size variance (C_{MD}), and microscopic orientation coherence (C_c)).

4.3. Experimental data

Four subsets of the data set publicly available at https://github.com/filip-szczepankiewicz/Szczepankiewicz_DIB_2019 and described in Szczepankiewicz et al. (2019) were used to test the proposed framework. One subset was formed with the 217 samples previously described in Table 1, i.e., protocol p217. Further subsets containing 39, 56, and 81 measurements produced the protocols p39, p56, and p81, respectively. These are also summarized in Table 1. The samples in p56 and p81 were chosen with the purpose of mimicking the protocols found at http://github.com/filip-szczepankiewicz/fwf_seq_resources/tree/master/GE. Having to pick samples out of an existing dataset, we randomly selected the measurements from the ones available with the goal of keeping reasonably spread measurement directions while making sure that the design matrix will have rank 23. Fig. 2 shows the sample distributions for p217, p81, p56, and p39.

On these four datasets, we fitted the QTI model using Eqs. (5) (with and without subspace implementation), (12), (11), and (15). For each fit we then checked where the conditions (d), (c), and (m) were violated. Conditions (d) and (c) were considered satisfied if the eigenvalues of the estimated $\widehat{\mathbf{D}}_{ij}$ and $\mathbf{C}_{\alpha\beta}$ were non-negative. However, we consider that a simple check done on the raw eigenvalues of the two tensors might mistake a violation of

---

[7]Our definition of the SNR is the same as that in other studies on noise in MRI (Gudbjartsson and Patz, 1995; Koay et al., 2009).

the two conditions with numerical errors arising from limited tolerances in the employed fitting routines. For example, a tensor having eigenvalues 2, 1, and -10^{-8} can still be considered nonnegative if the proximity of the negative eigenvalue to 0 is smaller than numerical tolerance. To overcome this ambiguity, we introduced a metric we refer to as "negativity index," which in essence is a normalized and dimensionless indicator of the positivity violations. For any symmetric matrix, we calculate the eigenvalues $\lambda_1, \ldots, \lambda_N$ and form the quotient

$$\text{NI} = \sum_{i, \, \lambda_i < 0}^{N} \lambda_i^2 \Bigg/ \sum_{i}^{N} \lambda_i^2 \tag{19}$$

i.e., the sum in the numerator is only taken over the negative eigenvalues. Note that this measure is insensitive to scalings of the underlying matrix. When NI is below 5×10^{-4}, the nonnegativity condition (d) or (c) is deemed to be fulfilled. To check whether condition (m) is fulfilled, we employed the scheme described in Appendix E.

4.4. Synthetic data

Additional simulations, inspired by those performed in Dela Haije et al. (2020), were performed to further assess the relevance of enforcing positivity constraints for the estimation of the parameters. The S_0, $\widehat{D}_{ij}$, and $\mathbb{C}_{ijkl}$ estimated by applying SDP(dcm) on the dataset with 217 measurements were used to create a synthetic dataset according to equation (1). As explained in Dela Haije et al. (2020), such dataset can be seen as the output of an ideal preprocessing pipeline which removes any bias and artifacts in the data. Moreover, assuming that the signal reconstruction provided by the investigated model is representative of the acquired data, this dataset can effectively be seen as a collection of signals produced by many different plausible tissue specimens. Therefore, it can act as ground truth for validation purposes.

Gaussian and Rician noise with standard deviation $\sigma = 0.04$, corresponding to SNR = $1/\sigma = 25$ on an S_0 value estimated from a region of interest containing white and gray matter voxels, was added to the dataset. The noisy datasets were then subsampled to 56 measurements. The parameters were estimated through both WLLS(ss) and SDP(dcm) applied on the noisy synthetic datasets with 217 and 56 measurements.

5.   Results

5.1.   Simulations

The results of our simulations are illustrated in Fig. 3 for the isotropic $\widehat{\mathbf{D}}_{ij}$ and Fig. 4 for the anisotropic $\widehat{\mathbf{D}}_{ij}$. The analytical results are depicted via dotted lines, which can be regarded as the ground truth in cases when QTI offers an accurate representation of the analytical signal. For a comparison of the QTI model in general compared to other methods for estimating statistical descriptors, we refer to (Reymbaut et al., 2020)[8]. Here, it

---

[8]Note, however, that the work of Reymbaut et al. (2020) focuses on DTDs of axisymmetric diffusion tensors.

is observed that the QTI model allows for, and indeed may produce, negative estimates of manifestly non-negative quantities. If the model in (1) is insufficient in describing the signal, i.e., when the higher order cumulants influence the signal within the range of employed diffusion weightings, we expect a deviation of the noiseless (SNR=∞) estimates from the dotted lines. Each solid circle shows the mean value of the estimates, different colors representing different estimation methods. The standard deviations are depicted via error bars. No appreciable difference was observed between the fits obtained on the noisy Rician and Gaussian distributed signals, therefore only the results on the Rician noise are shown.

From Figs. 3 and 4, it is clear that the QTI+ estimates obtained via SDP(dc), NLLS(dc), and SDP(dcm) methods are more robust to noise. In general, this refers to smaller deviations of the mean of the presented metrics, and substantially reduced standard deviations (error bars). This is particularly evident for the derived scalar measures. The Frobenius norms exhibit more noise sensitivity, which is likely because the Frobenius norm captures the tensors in their entirety while the other parameters are sensitive only to certain features of the tensors. Therefore, larger deviations appear, as expected, in the Frobenius norms of the difference between the reference and estimated tensors. This suffices to explain as well why most of the metrics derived from the WLLS(ss) fit are acceptable for SNRs down to ≈ 20 despite complications due to the rank deficiency of the design matrix for the shorter protocol.

Concerning the QTI+ estimation methods, we note that the results produced with SDP(dc), NLLS(dc), and SDP(dcm) are not drastically different. Especially when comparing NLLS(dc) and SDP(dcm), the difference is very subtle. This is partly because the violations of condition (m) are not frequent[9] and perhaps also because satisfying condition (m) does not have a very strong influence on the estimated metrics.

Looking at specific metrics, we note that FA increases with noise when $\widehat{\mathbf{D}}_{ij}$ is isotropic. Interestingly, constrained estimation tends to reduce FA in simulations featuring anisotropic $\widehat{\mathbf{D}}_{ij}$ tensors. This could be explained considering that in absence of constraints, smaller eigenvalues would spread in the negative direction, thus incorrectly increasing the spread of the eigenvalues of $\widehat{\mathbf{D}}_{ij}$ hence the FA value, while when constraints are applied, the small eigenvalues can only grow in the positive direction, leading to a reduction in anisotropy. The same trends are evident in the C_C results as expected.

Microscopic anisotropy is perhaps the most interesting scalar measure that has prompted much interest in the development of alternative diffusion encoding methods (Cheng and Cory, 1999; Cory et al., 1990; Ianus et al., 2017; Lawrenz et al., 2010; Özarslan, 2009) that eventually led to the introduction of QTI. Note that having isotropic $\widehat{\mathbf{D}}_{ij}$ does not imply 0 microscopic anisotropy because a non-central Wishart distributed set of tensors represent an ensemble of anisotropic subdomains even if their mean is isotropic. Our simulations suggest that the microscopic anisotropy index ($C_\mu = \mu FA^2$) is also quite susceptible to noise

---

[9]The numbers of signal profiles that violated condition (m) after the NLLS(dc) fitting and were subsequently fed into the SDP(dcm) routine were highest for the simulations of the p56s protocol with anisotropic $\widehat{\mathbf{D}}_{ij}$. These numbers were, respectively, 0, 12, 92, 200, 307, 358, out of the 1000 noisy samples for each (non-zero) noise level.

when traditional QTI methods are employed. The estimates benefit greatly from constrained estimation methods.

The noise sensitivity issue is more serious for indices of size variance (C_{MD}) and coherence (C_C). In fact, the estimates are simply unreliable under noisy conditions when no constraint is employed. QTI+ estimates of these indices could make them suitable for comparative analyses.

Effects of employing shorter acquisition protocols can be assessed by comparing the two panels of Figs. 3 and 4. Remarkably, employing the shorter protocol leads to a very considerable loss of quality for unconstrained QTI estimates of C_μ for example. As far as the constrained QTI+ estimators are concerned, the influence of the protocol has a relatively minor effect. This observation is important as it suggests that QTI+ could facilitate the employment of the method in clinical investigations where the acquisition time is a critically important limitation.

5.2. Experimental data

In Fig. 5, we illustrate the extent of the violations of the three postivity conditions. For both the long protocol p217 (left) and shorter protocol p56 (right), condition (c) is violated almost everywhere within the brain parenchyma when WLLS methods are employed. Condition (d) seems to be violated mostly in the very anisotropic and coherently organized regions like in the corpus callosum. As expected, these violations do not prevail when the QTI+ methods are employed. The rank deficiency of the design matrix associated with the shorter protocol seems to have the biggest impact on condition (m). Without the formulation in the subspace, this issue manifests as violation of (m) in almost all voxels. WLLS(ss) reduces the extent of such violations considerably. Interestingly, the SDP(dc) method provides further improvement although the condition (m) is not enforced. SDP(dcm) fulfills all three conditions as expected.

Fig. 6 illustrates the maps of the scalars obtained through various estimation methods for the dataset comprising 56 volumes. Despite the apparent similarity of the maps, some differences are visible, particularly in anisotropy measures (FA, C_M, C_μ, μFA). Namely, the maps derived through constrained estimation methods shown in the last three rows appear to be smoother than those obtained via unconstrained estimation. As none of the analyses employs information from neighboring voxels, we think this is an important finding, which corroborates the noise resilience associated with the constrained estimation methods evident in the simulations. Appreciable changes are also evident in the C_C maps by way of a reduction in the apparent coherence values in CSF.

Fig. 7 shows the scalar maps obtained by fitting the considered protocols, respectively, with WLLS(ss) and SDP(dcm). Looking at both panels, one observes that the non-diffusion weighted (S_0) and mean diffusivity (MD) maps are not severely affected by the downsampling. The anisotropy maps (fractional anisotropy FA, macroscopic anisotropy C_M, and microscopic anisotropy $C_\mu = \mu\text{FA}^2$) obtained via both methods are acceptable for the 81-measurement protocol. At sparser samplings, the improvement obtained by enforcing constraints becomes clear. Such improvement is evident also when looking at the bar plots

indicating the mean absolute deviations of the scalar maps from their ground truth values, which are taken to be the maps computed on the p217 protocol. The results obtained by employing SDP(dcm) show consistent lower deviations from the respective ground truths compared to those obtained via WLLS(ss). The bar plots also reveal that the coherence (C_c) and size variance (C_{MD}) estimates have very large absolute deviations for unconstrained estimation. A further examination of the pixel values revealed that this can be attributed in part to a small number of voxels that yield values way outside their expected range ([0, 1]). This issue is remedied by the QTI+ framework.

Figs. 8 and 9 show details for the FA and μFA maps computed on all considered protocols by employing WLLS(ss) and SDP(dcm), respectively. Looking at the (a) panels, the benefits of applying constraints are already evident. Again, we stress how the maps obtained with SDP(dcm) appear overall smoother even though the method is performed on a voxel-by-voxel basis without incorporating any intervoxel information. Panels (b) in these figures show the difference between the maps estimated with the p217 protocol, taken here as reference, and its subsets p81, p56, and p39. It is quite interesting to note that reducing the number of measurements from 56 to 39 does not drastically change the results. The histograms in panels (c) illustrate how reducing the number of available samples introduces a bias towards higher values in the anisotropy measures. We also note that constraining the fit strongly reduces the number of voxels presenting values outside the expected range ([0, 1] for FA and μFA). With respect to this, applying constraints (d), (c), and (m) seems to be insufficient to guarantee the condition μFA $\leq$ 1. We found that when μFA is greater than 1 in the results produced by SDP(dcm), the values are still very close to 1. Although one can be tempted to attribute this error to numerics, a more reasonable explanation is that μFA is formed from the estimates of $\widehat{\mathbf{D}}_{ij}$ and $\mathbb{C}_{ijkl}$, which are in a sense independent, and there is no guarantee that μFA should in fact not be greater than 1. Moreover, QTI+ only ensures some necessary constraints, but not all. Having μFA values strictly lower or equal to 1 could be added as a constraint, but from our findings this would have a very marginal effect[10].

5.3. Synthetic data

Fig. 10 shows the results obtained by fitting the synthetic brain datasets with both the WLLS(ss) and SDP(dcm) routines. The performance of the two methods was quantified through the Fobenius norms of the difference between the estimated and ground truth $\mathbb{C}_{ijkl}$ tensors, $||\Delta\mathbb{C}||$, and differences between the estimated and ground truth metrics, FA, μFA, C_c, counted for all voxels ($\approx$ 84000) in the dataset.

Looking globally at the results in panels (a) and (b), there seems to be no relevant difference between the fitting results obtained in the data corrupted with either Gaussian or Rician noise. There is moreover not a marked difference between the performance of QTI and QTI+ on the 217-measurement protocols, with QTI+ providing slightly better results. The difference in performance between WLLS(ss) and SDP(dcm) is instead highlighted in the plots showing the results on the 56-measurement protocol. There, the distance between

---

[10]Out of the $\approx$ 84000 considered voxels, only 42, 30, 14, and 21 had μFA values > 1 for the SDP(dcm) fits performed on the 217, 81, 56, and 39 measurements datasets, respectively.

the estimated and reference metrics is almost centered about zero for QTI+, while the parameter values produced with QTI exhibit a more pronounced tendency towards being over-estimated.

5.4. Run times

One of the appealing features of QTI is the computational speed at which the estimation can be performed via standard linear regression routines; it takes only a few seconds to fit the model to an entire dataset. This is definitely not the case for non-linear fitting routines, but also for the softwares currently available for semidefinite programming. As mentioned in the implementation section, we rely on an external package to solve the SDP problems. In our experience, calling this package on a voxel-by-voxel basis is inefficient, leading to prolonged computational times. A better approach involves performing the operations on a batch of voxels each time the function is called. In this case, steps 1, 3, and 4 can be performed by, for example, solving the problems on 50 voxels at a time. This provided relevant speed up when compared to performing the analysis one voxel at a time, as shown in Table 2. The table shows the run times for the different routines using different strategies on the experimental dataset ($\approx$ 84000 voxels) with, respectively, 217 and 56 measurements. The clock times were recorded on a workstation featuring a 12-core Intel Core i9-7920X CPU. The "multi voxel" implementation concerned sending 50 voxels at a time to the SDP solver.

Performing the fit on a multivoxel basis is perhaps intuitive for SDP(dc), but maybe not so much for the (m)-check and eventual SDP(dcm) steps, given the conditional step involved in the process. In the worst case scenario, one would in fact have to run the (m)-check on all batches of voxels, and then SDP(dcm) on all those batches. Even though this would still be faster than doing this process voxel-by-voxel, a faster option could be to skip the check on the (m) condition and directly perform SDP(dcm) on a multivoxel basis. However, since already after SDP(dc) (and NLLS(dc)) most of the (m) violations are resolved, and since the (m)-check appears to be faster than SDP(dcm), we find that the fastest option is to perform both the (m)-check and SDP(dcm) on a multivoxel basis.

It is well-known that non-linear fitting is typically more time-consuming than linear regression. One aspect to be considered is that having a good starting point, provided here by SDP(dc), helps in speeding up the non-linear fitting of the NLLS(dc) routine. However, we would like to remind the reader that NLLS(dc) is not a necessary step to perform in QTI+ as satisfactory results can be obtained using SDP(dc) and SDP(dcm). If truly pressed with time, one could also rely on the results produced with SDP(dc) only, as violations of the (m) condition are both infrequent and not extremely influential on the estimates.

6. Discussion

Since the inception of the DTD model (De Swiet and Mitra, 1996; Jian et al., 2007), the challenge of obtaining the underlying DTD from the MR signal has been addressed in different ways. One approach is to assume a parametric distribution, which can naturally ensure that all tensors in the DTD are positive-nonnegative. Indeed, Jian et al. (2007) have assumed a mixture of Wishart distributions for the DTD and even provided the analytical

signal for diffusion encoding via arbitrary b-tensors; that relationship can be obtained by setting $\mathbf{\Omega} = \mathbf{0}$ in (16). The Wishart distribution is the multidimensional generalization of the gamma distribution, which has been employed to represent the distribution of diffusivities for diffusion in polymer solutions (Röding et al., 2012). This approach has been adopted by Lasi̇ et al. (2014) for representing the distribution of diffusivities along all directions combined. When employed on tensor-encoded data, this approach yields visually appealing results for some of the parameters considered in this work (Szczepankiewicz et al., 2015; 2016). However, the validity of the gamma distribution is not guaranteed in all voxels, and the consequences of not fulfilling this assumption have not been understood. Moreover, this method relies on an accurate estimation of the orientationally-averaged signal, which can be challenging especially when the number of samples is limited (Afzali et al., 2020). Most recently, Magdoom et al. (2021) have introduced an alternative method, which employs a new pulsed field gradient sequence. The DTD is taken to be a tensor-valued normal distribution with non-positive-definite tensors suppressed. Due to the lack of an analytical form of the signal for this distribution, the signal is approximated using a large number of samples drawn from a given normal distribution. The mean and covariance are subsequently estimated using a least squares optimization.

Rather than employing an assumption regarding the underlying distribution like in the works mentioned above, we have employed the framework in Westin et al. (2016), which allows the estimation of the mean and covariance tensors only. It should be noted that this is not meant to represent the signal in the entire "b-space," but only its behavior at low b-values, which are probed in typical clinical acquisitions. Here, we considered imposing three constraints on the estimated mean and covariance tensors. Strictly speaking, each and every diffusion tensor in the underlying DTD has to be positive-semidefinite. However, imposing such a strong condition without attempting to solve an extremely ill-posed problem that involves the reconstruction of the actual DTD (Jian et al., 2007; Topgaard, 2019) is likely to be infeasible under noisy conditions or with limited number of diffusion encodings. Interestingly, non-negativity of each microscopic diffusion tensor implied the (m) condition, which seemed to have a minor effect in our analyses. Much of the improvement is already obtained through imposing condition (d) together with the (c) condition, which follow from the nonnegativity of the covariance tensors[11]; these conditions are valid even when the distribution is over a more general space—not necessarily the space of positive semidefinite tensors.

Satisfactory performance obtained by imposing only the conditions (d) and (c) have implications also when one decides on which estimation method to use. Our findings suggest that the (m) condition is relevant for a small portion of the voxels. Moreover, imposing the (d) and (c) conditions in the linearized version of the problem already provides a substantial portion of the overall gain. Thus, the SDP(dc) routine can be employed with relative confidence, which makes the overall estimation computationally inexpensive compared to the full framework that includes the subsequent NLLS(dc), (m) check and SDP(dcm) methods.

---

[11]Here, we remark that the diffusion tensor is the covariance matrix of net displacements.

Our work could be extended to include higher order terms in the expansion in Eq. 1. However, the next order term introduces 56 new unknowns (Westin et al., 2016), which demands longer acquisition protocols, which make them challenging for clinical studies.

Another matter that we did not address here concerns the limitations of the DTD model. The latter assumes that diffusion is described fully by a diffusion tensor and the effect of the waveforms on the signal are captured by the b-matrix. As discussed elsewhere (Özarslan et al., 2015; Yolcu et al., 2016), this would be valid for large compartments or at very short measurement times. To address this problem, one can consider the distribution of confinement tensors (Afzali et al., 2015; Yolcu et al., 2016; Zucchelli et al., 2016), which are shown to have the correct time dependence in small compartments making it the effective model for many scenarios of interest (Boito et al., 2020; Liu and Özarslan, 2019; Özarslan et al., 2017). One manifestation of this problem concerns acquisitions with isotropic b-matrices (Avram et al., 2018; Mori and van Zijl, 1995; Wong et al., 1995). In this case, the DTD model predicts the same signal for all measurements at each b-value, thus suppressing the potentially relevant information due to the non-Gaussianity of the diffusion process (Jespersen et al., 2019; de Swiet and Mitra, 1996). Similarly, the DTD model ignores the higher order cumulants of the diffusion process within each compartment; this challenge has been studied in recent works (Henriques et al., 2020; Paulsen et al., 2015). We note that the estimation methods introduced here may be instrumental in developing constrained fitting techniques for models aiming to overcome the limitations of the DTD picture.

The protocols obtained by downsampling a long protocol were not optimized as the waveforms were selected from a preexisting set. This poses an additional limitation for the downsampled protocols. However, our constrained estimation framework yielded acceptable image quality even for the shortest protocol comprising only 39 acquisitions. Thus, QTI+ could be more robust to imperfections in the experimental design as well; such imperfections are encountered, for example, due to gradient nonlinearities. More generally, a typical data set is likely to exhibit various artifacts such as Gibbs-ringing, subject motion, and frequency drift. In this case, unconstrained fitting will likely yield violations of the mathematically-necessary conditions. Employing a constrained estimation framework like QTI+ is thus expected to help alleviate the effects of artifacts. Similarly, studies have shown that identifying and discarding outliers is an effective approach for dealing with some of the confounding factors (Chang et al., 2012; Maximov et al., 2011; Tax et al., 2015; Zwiers, 2010). We would like to stress that we do not envision QTI+ to be a replacement for techniques developed to address such effects. Rather it could be part of a series of algorithms (Maximov et al., 2019) that collectively provide accurate maps of the desired parameters.

In recent years, the sensitivity of QTI-accessible quantities like μFA on various cerebral diseases including schizophrenia (Westin et al., 2016), brain tumors (Szczepankiewicz et al., 2016), epilepsy (Lampinen et al., 2020), multiple sclerosis (Andersen et al., 2020), and Parkinson's disease (Kamiya et al., 2020) have been investigated. Having reliable estimates of those quantities is of paramount importance for such studies, which could benefit particularly from the robustness of QTI+ to SNR. In fact, in the brain the signal without diffusion weighting does have some (typically T_2-weighted) contrast, which may

be amplified in the presence of pathologies. This contrast will lead to a spatially-dependent SNR, which ultimately affects the estimated parameters. Hardware-related effects such as spatially-varying sensitivity of the receiver coil are expected to contribute to this challenge. Our findings indicate that computing the sought after parameters via QTI+ could substantially reduce the SNR dependence of the findings, improving the accuracy and specificity of the estimated parameters.

Another important challenge in the translation of advanced imaging techniques into the clinical realm involves the limitations concerning the acquisition length. Recent works have attempted to address this issue by employing parsimonious data acquisition schemes (Nilsson et al., 2020). Our study demonstrates that sophisticated post-processing methods could be employed to achieve the same goal.

From a signal processing point of view, it is quite intriguing that the constrained estimation framework produces smoother maps and improved resilience to noise although the estimation is performed for each voxel independently, i.e., without employing any information from the adjacent voxels. Thus, the constrained methods do not yield a loss of image resolution, which is typically the case for routine smoothing methods. Moreover, the constraints have a very solid foundation pertaining to the mathematical properties of the estimated quantities. Consequently, the constrained estimation schemes like the ones we introduced here, do not involve any parameters that are to be decided in an *ad hoc* manner.

7.   Conclusion

In conclusion, we introduced QTI+, a new estimation framework for q-space trajectory imaging that respects three positivity conditions arising from the mathematical properties of the quantities estimated. We demonstrated that QTI+ leads to notable improvements in the accuracy and precision of the measured parameters. Although data smoothing is not employed, our framework is exceptionally robust to SNR, which has important ramifications for the interpretability of the derived parameters. The benefits of QTI+ are more conspicuous when shorter acquisition protocols with fewer number of diffusion-weighted volumes are available. Thus, our technique is expected to improve the feasibility as well as reliability, hence the diagnostic utility, of diffusion MRI measurements with generalized diffusion encoding.

Acknowledgements

Data and Code Availability Statement

In this work we use software and data that are available for academic purposes.

Appendix A.: Convention for the Voigt notation

It is customary to use the Voigt notation so that the matrix

$$\mathbf{A}_{ij} = \begin{pmatrix} x_1 & x_4 & x_5 \\ x_4 & x_2 & x_6 \\ x_5 & x_6 & x_3 \end{pmatrix}$$

is represented by the vector

$$\mathbf{v}_i = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \sqrt{2}x_4 \\ \sqrt{2}x_5 \\ \sqrt{2}x_6 \end{pmatrix}.$$

Here, $\mathbf{A}_{ij} = \sum_{\beta=1}^{6} a_\beta \mathbf{e}_{ij}^{(\beta)}$, i.e., the coordinates $a_1, a_2, \ldots, a_6$ express the matrix $\mathbf{A}_{ij}$ in the (orthonormal) basis $\{\mathbf{e}_{ij}^{(1)}, \mathbf{e}_{ij}^{(2)}, \ldots, \mathbf{e}_{ij}^{(6)}\}$. With the representation above, these basis matrices are

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\begin{pmatrix} 0 & 0 & \frac{1}{\sqrt{2}} \\ 0 & 0 & 0 \\ \frac{1}{\sqrt{2}} & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & 0 \end{pmatrix}.$$

Appendix B.: Independence of (d), (c), and (m)

Here, we take independence to mean that there are examples of tensors $\mathbf{D}_{ij}$, $\mathbb{C}_{ijk\ell}$, $\mathbb{M}_{ijk\ell}$, where two given constraints are satisfied but not the third.[12] To see that the conditions (d), (c), and (m) are independent in the above sense, we give the following examples.

(d) & (c) $\not\Rightarrow$ (m): Let $\langle \mathbf{D}_{ij} \rangle = \mathbf{0}_{ij}$ and define $\mathbf{E}_{ij}$ through $\mathbf{E}_{ij} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$. Take $\mathbb{C}_{ijk\ell} = \mathbf{E}_{ij} \otimes \mathbf{E}_{k\ell}$, which implies that also $\mathbb{M}_{ijk\ell} = \mathbf{E}_{ij} \otimes \mathbf{E}_{k\ell}$. By construction, $\mathbb{C}_{ijk\ell}$ is positive semi-definite as a symmetric mapping $V \to V$, but the choice $\mathbf{v}_i = (1, 1, 0)^\top$, $\mathbf{u}_i = (1, -1, 0)^\top$ gives an example where $M_{ijk\ell} u_i u_j v_k v_\ell = -4 < 0$.

---

[12]Not all possible constraints are independent in this sense. For instance, a possible constraint is that $\mathbb{M}$ viewed as a symmetric mapping $V \to V$ is positive semi-definite. If (d) and (c) are satisfied, then this is automatically true.

(d) & (m) $\not\Rightarrow$ (c): Again, take $\langle \mathbf{D}_{ij} \rangle = \mathbf{0}_{ij}$ and define $\mathbb{C}$ by

$$\mathbb{C}_{ijk\ell} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Then (c) is violated but on the other hand, for any vector $\boldsymbol{v}_i = (x, y, z)^\top$

$$\mathbf{M}'_{ij} = M_{ijk\ell} v_k v_\ell = \begin{pmatrix} x^2 + y^2 + z^2 & -2xy & 0 \\ -2xy & x^2 + y^2 + z^2 & 0 \\ 0 & 0 & x^2 + y^2 + z^2 \end{pmatrix}.$$

This matrix has eigenvalues $x^2 + y^2 + z^2$, $(x - y)^2 + z^2$ and $(x + y)^2 + z^2$, which means that $\mathbf{M}'_{ij}$ is positive semi-definite for any vector v_i. Hence (m) is satisfied.

(c) & (m) $\not\Rightarrow$ (d): This is immediate since if $\mathbb{M}_{ijk\ell}$ is constructed as $\mathbb{M}_{ijk\ell} = \mathbb{C}_{ijk\ell} + \langle \mathbf{D}_{ij} \rangle \otimes \langle \mathbf{D}_{k\ell} \rangle$, then $\mathbb{M}_{ijk\ell}$ is unaffected by the replacement $\langle \mathbf{D}_{ij} \rangle \to -\langle \mathbf{D}_{ij} \rangle$, and by choosing an example with $\langle \mathbf{D}_{ij} \rangle \neq \mathbf{0}$, both $\langle \mathbf{D}_{ij} \rangle$ and $-\langle \mathbf{D}_{ij} \rangle$ cannot be positive semi-definite.

Appendix C.: Heteroscedasticity compensation of the log-linearized problem.

In this appendix we motivate, very briefly the weighting of the linearized problem (4), i.e., the insertion of the weights S_n^2. With

$$\Delta_n = S_0 e^{-B_{ij}^{(n)} \widehat{D}_{ij} + \frac{1}{2} B_{ij}^{(n)} B_{k\ell}^{(n)} C_{ijk\ell}} - S_n, \ n = 1, 2, ..., N$$

a solution to (2) minimizes $\sum_{n=1}^{N} |\Delta_n|^2$. By adding S_n to both sides and taking logarithms, we get

$$\ln(S_n + \Delta_n) = \ln(S_0) - B_{ij}^{(n)} \widehat{D}_{ij} + \frac{1}{2} B_{ij}^{(n)} B_{k\ell}^{(n)} C_{ijk\ell}.$$

If Δ_n is small compared to S_n, we get $\ln(S_n + \Delta_n) = \ln(S_n) + \ln(1 + \frac{\Delta_n}{S_n}) \approx \ln(S_n) + \frac{\Delta_n}{S_n}$. As a result, a straightforward least squares implementation will (approximately) minimize $\sum_{n=1}^{N} |\frac{\Delta_n}{S_n}|^2$. This is compensated for by multiplying the linearized equation for the nth measurement with the corresponding signal values S_n.

Appendix D.: Equivalence of QP and SDP for the unconstrained linearized equation

Here we show that (7) is equivalent to (9), namely, we first see that to find an x which minimizes $x^\top Q x + c^\top x$, where $Q \succ 0$, is equivalent to solving the problem

$$\underset{t,\, x}{\operatorname{argmin}} t, \text{ subject to } t - x^\top Q x - c^\top x \geq 0$$

because minimizing t in the inequality $t \quad x^\top Q x + c^\top x$ leads us to find the minimum value of $x^\top Q x + c^\top x$. Next, suppose P is a square matrix with $P^\top P = Q$. Then, $\begin{pmatrix} I & Px \\ x^T P^T & t - c^T x \end{pmatrix} \succeq 0$ is equivalent to the statement

$$\forall v, a, \quad (v^\top \quad a)\begin{pmatrix} I & Px \\ x^\top P^\top & t - c^\top x \end{pmatrix}\begin{pmatrix} v \\ a \end{pmatrix} \geq 0.$$

But the expression on the left hand side can be written (using $P^\top P = Q$) as

$$(v + aPx)^\top (v + aPx) + a^2(t - x^\top Q x - c^\top x). \tag{D.1}$$

But if this is nonnegative for all v and a, then (by choosing $v = -aPx$) so is $t - x^\top Q x - c^\top x$. Conversely, if $t - x^\top Q x - c^\top x \quad 0$. then so is the expression in (D.1).

Appendix E.: Checking condition (m) using SDP

Here, we describe our scheme for checking condition (m), i.e., For all v_i, u_i, $\mathbb{M}_{ijkl} v_i v_j u_k u_l \geq 0$?

By putting $v_i = (x, y, z)^\top$ and $u_i = (r, s, t)^\top$, the contraction $M_{ijkl} v_i v_j u_k u_l$ becomes a fourth order homogeneous polynomial p in the six variables x, y, z, r, s, t. Condition (m) can then be formulated as

$$\text{Is } p(x, y, z, r, s, t) \geq 0 \text{ for all } x, y, z, r, s, t \text{ ?} \tag{E.1}$$

Because of the symmetries of $\mathbb{M}_{ijkl}$ and the form of the contraction with v_i and u_i, it is possible to represent p in the following way. We start by forming the vector $V = (xr,\ xs,\ xt,\ yr,\ ys,\ yt,\ zr,\ zs,\ zt)^\top$. Then, for any symmetric 9×9 matrix M, it is clear that also $V^\top M V$ is a fourth order homogeneous polynomial in x, y, z, r, s, t, and it is not difficult to see that any $\mathbb{M}$ can be represented by such a matrix M. In fact, this representation is not unique, and by solving the equation $V^\top M V = 0$, one finds the solution space to be a nine-dimensional subspace (in the space of symmetric 9×9 matrices):

$$V^T \begin{pmatrix} 0 & 0 & 0 & 0 & l_1 & l_2 & 0 & l_4 & l_6 \\ 0 & 0 & 0 & -l_1 & 0 & l_3 & -l_4 & 0 & l_7 \\ 0 & 0 & 0 & -l_2 & -l_3 & 0 & -l_6 & -l_7 & 0 \\ 0 & -l_1 & -l_2 & 0 & 0 & 0 & 0 & l_5 & l_8 \\ l_1 & 0 & -l_3 & 0 & 0 & 0 & -l_5 & 0 & l_9 \\ l_2 & l_3 & 0 & 0 & 0 & 0 & -l_8 & -l_9 & 0 \\ 0 & -l_4 & -l_6 & 0 & -l_5 & -l_8 & 0 & 0 & 0 \\ l_4 & 0 & -l_7 & l_5 & 0 & -l_9 & 0 & 0 & 0 \\ l_6 & l_7 & 0 & l_8 & l_9 & 0 & 0 & 0 & 0 \end{pmatrix} V = 0, \, l_1, \ldots, l_9 \in \mathbf{R}.$$

We indicate this freedom by writing

$$p(x, y, z, r, s, t) = V^\mathsf{T} \mathbf{M}(\ell) V$$

where $\ell = (l_1, \ldots, l_9)$. Now if, for some value of the parameter vector ℓ the matrix $\mathbf{M}(\ell)$ is positive semi-definite, then it is clear that (E.1) (and hence also (m)) holds. This is exactly the feasibility problem (14). By diagonalizing $M(\ell)$ (if it is nonnegative) one sees that p is expressed as a sum of squared (SoS) polynomials, i.e.,

$$p(x, y, z, r, s, t) = \sum_{i=1}^{9} \alpha_i p_i^2(x, y, z, r, s, t)$$

where $\{\alpha_i\}_{i=1}^{9}$ are the (non-negative) eigenvalues of $M(\ell)$ and each p_i is a linear combination of the entries of the vector V. This is thus an example of the rich theory of SoS polynomials (Berg et al., 1976; Lasserre, 2007). We should remark that it is not strictly necessary that $M(\ell)$ is positive semi-definite for (E.1) to hold, which means that the condition $M(\ell) \succcurlyeq 0$ is slightly stronger, as there are non-negative polynomials which are not SoS. In practice, however, this drawback is compensated for by the computational convenience offered by SDP. Also, there are results which show that the set of SoS polynomials are, in a certain sense(Lasserre, 2007), dense in the set of non-negative polynomials.

Appendix F.: Imposing condition (m) using SDP

By choosing the independent variables $x_1, x_2, \ldots, x_{28}$ to encode for S_0, $\widehat{\mathbf{D}}_{ij}$ and $\mathbb{C}_{ijkl}$ as explained in Section 3, the tensor $\mathbb{M}_{ijkl} = \mathbb{C}_{ijkl} + \widehat{\mathbf{D}}_{ij} \otimes \widehat{\mathbf{D}}_{kl}$ becomes linear in $x_8, \ldots, x_{28}$ but quadratic in $x_2, \ldots, x_7$. For this reason, we have adopted the strategy that if (m) is violated (in an estimate where (d) and (c) are imposed), we fix the estimates of $x_1, \ldots, x_7$, (i.e., $\ln(S_0)$ and $\widehat{\mathbf{D}}_{ij}$) and re-estimate $x_8, \ldots, x_{28}$ using SDP while imposing both (c) and (m).

We shall return to the (un-constricted) scenario formulated as a quadratic programming problem. We seek

$$\underset{x}{\operatorname{argmin}} x^{\mathsf{T}} \mathbf{Q} x + c^{\mathsf{T}} x \tag{F.1}$$

where $x = (x_1, \ldots, x_{28})^{\mathsf{T}}$ and $\mathbf{Q}$ is a symmetric matrix of size 28×28. Next, we decompose $x = (\check{x}^{\mathsf{T}}, \tilde{x}^{\mathsf{T}})^{\mathsf{T}}$, where $\check{x} = (x_1, \ldots, x_7)^{\mathsf{T}}$ is fixed, while $\tilde{x} = (x_8, \ldots, x_{28})^{\mathsf{T}}$ contains our remaining free parameters. To match this, we decompose $\mathbf{Q}$ and c into block matrices in the following fashion:

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} \end{pmatrix}, \text{ and } c = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

where the sizes of $\mathbf{Q}_{11}, \mathbf{Q}_{12}, \mathbf{Q}_{21}, \mathbf{Q}_{22}$ are 7×7, 7×21, 21×7, and 21×21, respectively, and c_1, c_2 are vectors with 7 and 21 elements. Then $x^{\mathsf{T}} \mathbf{Q} x + c^{\mathsf{T}} x$ becomes

$$(\check{x}^{\mathsf{T}}, \tilde{x}^{\mathsf{T}}) \begin{pmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} \end{pmatrix} \begin{pmatrix} \check{x} \\ \tilde{x} \end{pmatrix} + (c_1^{\mathsf{T}} \quad c_2^{\mathsf{T}}) \begin{pmatrix} \check{x} \\ \tilde{x} \end{pmatrix}$$

which simplifies to

$$\tilde{x}^{\mathsf{T}} \widetilde{\mathbf{Q}} \tilde{x} + \tilde{c}^{\mathsf{T}} \tilde{x} + c_0 \tag{F.2}$$

where

$$\widetilde{\mathbf{Q}} = \mathbf{Q}_{22}$$
$$\tilde{c}^{\mathsf{T}} = 2\check{x}\mathbf{Q}_{12} + c_2^{\mathsf{T}} \text{ and}$$
$$c_0 = \check{x}^{\mathsf{T}} \mathbf{Q}_{11} \check{x} + c_1^{\mathsf{T}} \check{x}.$$

With $\widetilde{\mathbf{P}}$ such that $\widetilde{\mathbf{P}}^{\mathsf{T}} \widetilde{\mathbf{P}} = \widetilde{\mathbf{Q}}$ (and remembering that c_0 is just a constant that does not affect the minimizing argument $\tilde{x}$), (F.2) is if the form which admits a SDP formulation. Disregarding c_0, this is found in the upper left blocks of the matrix in (15). Since all the variables x_8, $\ldots$, x_{28} are still free, and since they form $\mathbf{C}_{\alpha\beta}$, the positivity condition (c) remains the same: $\mathbf{C}_{\alpha\beta} \succcurlyeq 0$. Finally, with $\widehat{\mathbf{D}}_{ij}$ fixed, and since $\mathbb{M}_{ijk\ell} = \mathbb{C}_{ijk\ell} + \widehat{\mathbf{D}}_{ij} \otimes \widehat{\mathbf{D}}_{k\ell}$, all entries of the components $\mathbb{M}_{ijk\ell}$ will be first order polynomials in $x_8, \ldots, x_{28}$ and hence can be cast into a 9×9 symmetric matrix $\mathbf{M}$ as described in Appendix E. By adding the freedom in terms of the parameters $\ell_1, \ell_2, \ldots, \ell_5$ as also described in the previous section, we get the matrix $\mathbf{M}(\boldsymbol{\ell}$, (which could also be written $\mathbf{M}(\ell, \tilde{x})$) and by this formulation, both conditions (c) and (m) will be imposed simultaneously in the formulation (15).

Appendix G.: Scalar measures in the rank deficient case

As mentioned in Section 3, protocols that use measurements of only type (i) and (iii), i.e., LTE and STE, will produce matrices $\mathbf{A}$, which are rank deficient. In this case, the (maximum) rank will be 23 instead of 28, and as a result, there is a (five parameter worth)

family of covariance tensors $\mathbb{C}_{ijk\ell}$ compatible with the (fit of the) measurements. Here, we will describe this freedom and also indicate why the scalars measures FA, MD, C_M, C_μ, μFA, C_MD, and C_c are unaffected by this non-uniqueness in the estimates of $\mathbb{C}_{ijk\ell}$.

First we note that for measurements of type (i), i.e., of type LTE, such a measurement picks up only the content of the completely symmetric part $\mathbb{C}_{(ijk\ell)}$ of $\mathbb{C}_{ijk\ell}$. Namely, since any measurement tensor $\mathbf{B}_{ij}$ of type (i) is symmetric, positive semi-definite and has rank one, we can write $B_{ij} = v_i v_j$ for some vector v_i. It is then clear that $C_{ijkl} B_{ij} B_{kl} = C_{ijkl} v_i v_j v_k v_\ell$ $= C_{(ijkl)} v_i v_j v_k v_\ell$ since $v_i v_j v_k v_\ell$ itself is completely symmetric. As a result, the difference $\mathbb{K}_{ijk\ell} = \mathbb{C}_{ijk\ell} - \mathbb{C}_{(ijk\ell)}$ satisfies $K_{ijkl} v_i v_j v_k v_\ell = 0$ for all vectors v_i. Moreover, since $\mathbb{C}_{ijk\ell}$ has 21 independent components while $\mathbb{C}_{(ijk\ell)}$ has 15, this means that $\mathbb{K}_{ijk\ell}$ has six degrees of freedom.

Next, we express the tensor $\mathbb{K}_{ijk\ell}$ as a 6×6 matrix using the Voigt notation, yielding

$$\mathbf{K}_{\alpha\beta} = \begin{pmatrix} 0 & a+d-w & w-d & 0 & 0 & e \\ a+d-w & 0 & w-a & 0 & c & 0 \\ w-d & w-a & 0 & b & 0 & 0 \\ 0 & 0 & b & -a-d+w & -\dfrac{e}{\sqrt{2}} & -\dfrac{c}{\sqrt{2}} \\ 0 & c & 0 & -\dfrac{e}{\sqrt{2}} & d-w & -\dfrac{b}{\sqrt{2}} \\ e & 0 & 0 & -\dfrac{c}{\sqrt{2}} & -\dfrac{b}{\sqrt{2}} & a-w \end{pmatrix},$$

where a, b, c, d, e, and w are the six free parameters. Because $\mathbf{K}_{\alpha\beta}$ is symmetric, the corresponding fourth order tensor $\mathbb{K}_{ijk\ell}$ has the correct symmetries. Moreover, for any vector $v_i = (x, y, z)^\top$, the tensor $v_i v_j^\top$ can be expressed as $V_\alpha = \left(x^2, y^2, z^2, \sqrt{2}xy, \sqrt{2}xz, \sqrt{2}yz \right)^\top$ through the Voigt notation. Using this, it can be verified that $K_{\alpha\beta} V_\alpha V_\beta = 0$ and hence $K_{ijkl} v_i v_j v_k v_\ell = 0$ for all vectors v_i.

To proceed, we now refer to (Westin et al., 2016), in which the definitions of the scalars can be found. The key observation is then that all the scalar measures involves $\mathbb{C}_{ijk\ell}$ (or, equivalently, $\mathbf{C}_{\alpha\beta}$) in such a way that inner products are taken with linear combinations of the fourth order tensors $\mathbb{E}^{\mathrm{iso}}_{ijk\ell}$ and $\mathbb{E}^{\mathrm{bulk}}_{ijk\ell}$, whose 6×6 representations are given by

$$\mathbf{E}_{\alpha\beta}^{\text{bulk}} = \frac{1}{9}\begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \text{ and }$$

$$\mathbf{E}_{\alpha\beta}^{\text{iso}} = \frac{1}{3}\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

where in particular $\mathbb{E}^{\text{bulk}}$ is related to measurements of type (iii), i.e., STE. The contributions from $\mathbb{K}_{ijk\ell}$ to such inner products are

$$K_{ijk\ell}E_{ijk\ell}^{\text{bulk}} = \frac{2w}{9},$$
$$K_{ijk\ell}E_{ijk\ell}^{\text{iso}} = -\frac{w}{3} = -\frac{3}{2}K_{ijk\ell}E_{ijk\ell}^{\text{bulk}}.$$

This implies that the set of scalar measures considered here are sensitive to $\mathbb{K}_{ijk\ell}$ only through w, which can be obtained via an STE measurement in addition to a proper set of LTE acquisitions that enable the estimation of $\mathbb{C}_{(ijk\ell)}$.

Appendix H.: The mean and covariance of the Wishart distribution

Deriving the first two moments of the non-central Wishart distribution directly from the probability distribution function is a bit involved. However, using (16), we can find these by matching this expression to our model (1). In essence, we want to find $\widehat{\mathbf{D}}_{ij}$ and $\mathbb{C}_{ijk\ell}$ so that

$$|\mathbf{I} + \mathbf{\Sigma B}|^{-p}\exp(-\text{Tr}[\mathbf{B}(\mathbf{I} + \mathbf{\Sigma B})^{-1}\mathbf{\Omega}]) = e^{-B_{ij}\widehat{D}_{ij} + \frac{1}{2}B_{ij}B_{kl}C_{ijk\ell} + O((B_{ij})^3)}.$$

By taking logarithms and introducing a scale parameter x, we demand that for each fixed (symmetric, positive semi-definite) $\mathbf{B}_{ij}$, it holds that

$$\begin{aligned}-p\ln|\mathbf{I} + x\mathbf{\Sigma B}| &- \text{Tr}[x\mathbf{B}(\mathbf{I} + x\mathbf{\Sigma B})^{-1}\mathbf{\Omega}] \\ &= -x\boldsymbol{B}_{ij}\widehat{\boldsymbol{D}}_{ij} + \frac{x^2}{2}\boldsymbol{B}_{ij}\boldsymbol{B}_{kl}\boldsymbol{C}_{ijk\ell} + O(x^3).\end{aligned} \tag{H.1}$$

It follows from the definition of the (3×3) determinant that

$$\begin{aligned}|\mathbf{I} + x\mathbf{\Sigma B}| &= 1 + x\text{Tr}[\mathbf{\Sigma B}] + \frac{x^2}{2}(\text{Tr}[\mathbf{\Sigma B}]^2 - \text{Tr}[(\mathbf{\Sigma B})^2]) \\ &\qquad + O(x^3)\end{aligned}$$

and

$$\ln |\mathbf{I} + x\mathbf{\Sigma B}| = x\mathrm{Tr}[\mathbf{\Sigma B}] - \frac{x^2}{2}\mathrm{Tr}[(\mathbf{\Sigma B})^2] + O(x^3)$$

so that

$$-p\ln |\mathbf{I} + x\mathbf{\Sigma B}| = -px\Sigma_{ij}B_{ji} + p\frac{x^2}{2}\Sigma_{ij}B_{jk}\Sigma_{k\ell}B_{\ell i} \\ + O(x^3). \tag{H.2}$$

Next, since $(\mathbf{I} + x\mathbf{\Sigma B})^{-1} = \mathbf{I} - x\mathbf{\Sigma B} + O(x^2)$,

$$-\mathrm{Tr}[x\mathbf{B}(\mathbf{I} + x\mathbf{\Sigma B})^{-1}\mathbf{\Omega}] = -xB_{ik}(I_{kn} - x\Sigma_{km}B_{mn})\Omega_{ni} + O(x^2). \tag{H.3}$$

Inserting (H.2) and (H.3) into (H.1) and identifying terms w.r.t. x, we find that, for each symmetric $\mathbf{B}_{ij}$,

$$(\widehat{D}_{ij} - p\Sigma_{ij} - \Omega_{ij})B_{ij} = 0 \tag{H.4}$$

$$(C_{ijk\ell} - p\Sigma_{i\ell}\Sigma_{jk} - 2\Sigma_{jk}\Omega_{\ell i})B_{ij}B_{k\ell} = 0. \tag{H.5}$$

Note that the first equation does not imply that $\widehat{D}_{ij} = p\Sigma_{ij} + \Omega_{ij}$ unless we take the symmetry of $\widehat{D}_{ij}$ into account since $A_{ij}B_{ij} = 0$ for any anti-symmetric matrix A_{ij} (B_{ij} being symmetric). On the other hand, with $\widehat{D}_{ij}$ symmetric, it is necessary that $\widehat{D}_{ij} = p\Sigma_{ij} + \Omega_{ij}$, since if for a symmetric matrix A_{ij}, $A_{ij}B_{ij} = 0$ for all symmetric positive definite matrices B_{ij}, then $A_{ij} = 0$. For (H.5), the terms $\Sigma_{i\ell}\Sigma_{jk}$ and $\Sigma_{jk}\Omega_{\ell i}$ do not have the symmetries of $\mathbb{C}_{ijk\ell}$. On the other hand, using the symmetry of B_{ij}, one can check that, for all B_{ij}, $\Sigma_{i\ell}\Sigma_{jk}B_{ij}B_{k\ell} = \frac{1}{2}(\Sigma_{i\ell}\Sigma_{jk} + \Sigma_{ik}\Sigma_{j\ell})B_{ij}B_{k\ell}$ and $2\Sigma_{jk}\Omega_{\ell i}B_{ij}B_{k\ell} = \frac{1}{2}(\Sigma_{jk}\Omega_{i\ell} + \Sigma_{i\ell}\Omega_{jk} + \Sigma_{j\ell}\Omega_{ik} + \Sigma_{ik}\Omega_{j\ell})B_{ij}B_{k\ell}$, so that we can replace (H.5) by

$$A_{ijk\ell}B_{ij}B_{k\ell} = 0$$

for all symmetric positive semidefinite matrices B_{ij}, where $A_{ijk\ell} = C_{ijk\ell} - \frac{p}{2}(\Sigma_{i\ell}\Sigma_{jk} + \Sigma_{ik}\Sigma_{j\ell}) = \frac{1}{2}(\Sigma_{jk}\Omega_{i\ell} + \Sigma_{i\ell}\Omega_{jk} + \Sigma_{j\ell}\Omega_{ik} + \Sigma_{ik}\Omega_{j\ell})$ has the same symmetries as $\mathbb{C}_{ijk\ell}$. But this forces A_{ijkl} to be zero since we know (c.f. the discussion in (3.9)) that with general measurement tensors B_{ij}, we can produce tensors with components $B_{ij}B_{kl}$, which together determine A_{ijkl} above. This proves (18).

References

Afzali M, Knutsson H, Özarslan E, Jones DK, 2020. Computing the orientational-average of diffusion-weighted MRI signals: a comparison of different techniques. bioRxiv (10.1101/2020.11.18.388272).

Afzali M, Yolcu C, Özarslan E, 2015. Characterizing Diffusion Anisotropy for Molecules under the Influence of a Parabolic Potential: A Plausible Alternative to Dti. In: Proc Intl Soc Mag Reson Med, Vol. 23, p. 2795.

Andersen KW, Lasic S, Lundell H, Nilsson M, Topgaard D, Sellebjerg F, Szczepankiewicz F, Siebner HR, Blinkenberg M, Dyrby TB, 2020. Disentangling white-matter damage from physiological fibre orientation dispersion in multiple sclerosis. Brain Commun. 2 (2), fcaa077. doi:10.1093/braincomms/fcaa077. [PubMed: 32954329]

Avram AV, Sarlls JE, Hutchinson E, Basser PJ, 2018. Efficient experimental designs for isotropic generalized diffusion tensor MRI (IGDTI). Magn. Reson. Med 79 (1), 180–194. doi:10.1002/mrm.26656. [PubMed: 28480613]

Barmpoutis A, Ho J, Vemuri BC, 2012. Approximating symmetric positive semidefinite tensors of even order. SIAM J. Imaging Sci 5 (1), 434–464. doi:10.1137/100801664. [PubMed: 23285313]

Barmpoutis A, Hwang MS, Howland D, Forder JR, Vemuri BC, 2009. Regularized positive-definite fourth order tensor field estimation from DW-MRI. Neuroimage 45 (1). doi:10.1016/j.neuroimage.2008.10.056. S153–S162 [PubMed: 19063978]

Basser PJ, Mattiello J, LeBihan D, 1994. Estimation of the effective self-diffusion tensor from the NMR spin echo. J. Magn. Reson. B 103 (3), 247–254. [PubMed: 8019776]

Basser PJ, Mattiello J, LeBihan D, 1994. MR Diffusion tensor spectroscopy and imaging. Biophys. J 66 (1), 259–267. [PubMed: 8130344]

Basser PJ, Pajevic S, 2003. A normal distribution for tensor-valued random variables: applications to diffusion tensor MRI. IEEE Trans. Med. Imaging 22 (7), 785–794. doi:10.1109/TMI.2003.815059. [PubMed: 12906233]

Berg C, Christensen JPR, Ressel P, 1976. Positive definite functions on abelian semigroups. Math. Ann 223, 253–274.

Bevington PR, Robinson DK, 2003. Data reduction and error analysis for the physical sciences, 3rd McGraw-Hill.

Boito D, Yolcu C, Özarslan E, 2020. Compartment-specific Diffusivity: A New Dimension in Multidimensional Diffusion MRI? In: Proc Intl Soc Mag Reson Med, Vol. 28, p. 4442.

Chang LC, Walker L, Pierpaoli C, 2012. Informed "restore" : a method for robust estimation of diffusion tensor from low redundancy datasets in the presence of physiological noise artifacts. Magn. Reson. Med 68 (5), 1654–1663. doi:10.1002/mrm.24173. [PubMed: 22287298]

Chen Y, Dai Y, Han D, Sun W, 2013. Positive semidefinite generalized diffusion tensor imaging via quadratic semidefinite programming. SIAM J. Imaging Sci 6 (3), 1531–1552. doi:10.1137/110843526.

Cheng Y, Cory DG, 1999. Multiple scattering by NMR. J. Am. Chem. Soc 121, 7935–7936.

Cory DG, Garroway AN, Miller JB, 1990. Applications of spin transport as a probe of local geometry. Polym Preprints 31, 149.

De Swiet TM, Mitra PP, 1996. Possible Systematic Errors in Single-Shot Measurements of the Trace of the Diffusion Tensor., J Magn Reson B 111, 15–22. [PubMed: 8661259]

Dela Haije T, Özarslan E, Feragen A, 2020. Enforcing necessary non-negativity constraints for common diffusion MRI models using sum of squares programming. Neuroimage 209, 116405. [PubMed: 31846758]

Ghosh A, Milne T, Deriche R, 2014. Constrained diffusion kurtosis imaging using ternary quartics & MLE. Magn. Reson. Med 71 (4), 1581–1591. doi:10.1002/mrm.24781. [PubMed: 23821241]

Grant M, Boyd S, 2008. Graph Implementations for Nonsmooth Convex Programs. In: Blondel V, Boyd S, Kimura H (Eds.), Recent Advances in Learning and Control. Springer-Verlag Limited, pp. 95–110. Lecture Notes in Control and Information Sciences; http://stanford.edu/~boyd/graph_dcp.html

Grant M, Boyd S, 2014. CVX: Matlab software for disciplined convex programming, version 2.1 http://cvxr.com/cvx.

Gudbjartsson H, Patz S, 1995. The rician distribution of noisy MRI data. Magn. Reson. Med 34 (6), 910–914. [PubMed: 8598820]

Henriques RN, Jespersen SN, Shemesh N, 2020. Correlation tensor magnetic resonance imaging. Neuroimage 211, 116605. doi:10.1016/j.neuroimage.2020.116605. [PubMed: 32044435]

Herberthson M, Yolcu C, Knutsson H, Westin CF, Özarslan E, 2019. Orientationally-averaged diffusion-attenuated magnetic resonance signal for locally-anisotropic diffusion. Sci. Rep 9 (1), 4899. doi:10.1038/s41598-019-41317-8. [PubMed: 30894611]

Ianus A, Shemesh N, Alexander DC, Drobnjak I, 2017. Measuring Microscopic Anisotropy with Diffusion Magnetic Resonance: From Material Science to Biomedical Imaging. In: Schultz T, Özarslan E, Hotz I (Eds.), Modeling, Analysis, and Visualization of Anisotropy. Springer International Publishing; Mathematics and Visualization, pp. 229–255.

Jensen JH, Helpern JA, Ramani A, Lu H, Kaczynski K, 2005. Diffusional kurtosis imaging: the quantification of non-gaussian water diffusion by means of magnetic resonance imaging. Magn. Reson. Med 53, 1432–1440. [PubMed: 15906300]

Jespersen SN, Olesen JL, Ianucs A, Shemesh N, 2019. Effects of nongaussian diffusion on "isotropic diffusion" measurements: an ex-vivo microimaging and simulation study. J. Magn. Reson 300, 84–94. doi:10.1016/j.jmr.2019.01.007. [PubMed: 30711786]

Jeurissen B, Westin CF, Sijbers J, Szczepankiewicz F, 2019. Improved precision and accuracy in q-space trajectory imaging by model-based super-resolution reconstruction. Proc intl Soc Mag Reson Med. 27.

Jian B, Vemuri B, Özarslan E, 2009. A Mixture of Wisharts (MOW) Model for Multi–fiber Reconstruction. In: Laidlaw D, Weickert J (Eds.), Visualization and Processing of Tensor Fields. Springer-Verlag, pp. 39–55.

Jian B, Vemuri BC, Özarslan E, Carney PR, Mareci TH, 2007. A novel tensor distribution model for the diffusion-weighted MR signal. Neuroimage 37 (1), 164–176. doi:10.1016/j.neuroimage.2007.03.074. [PubMed: 17570683]

Kamiya K, Kamagata K, Ogaki K, Hatano T, Ogawa T, Takeshige-Amano H, Murata S, Andica C, Murata K, Feiweier T, et al. , 2020. Brain white-matter degeneration due to aging and parkinson disease as revealed by double diffusion encoding. Front Neurosci. 14, 584510. [PubMed: 33177985]

Koay CG, 2010. Least Squares Approaches to Diffusion Tensor Estimation. In: Jones DK (Ed.), Diffusion MRI. Oxford University Press, pp. 272–284.

Koay CG, Chang LC, Carew JD, Pierpaoli C, Basser PJ, 2006. A unifying theoretical and algorithmic framework for least squares methods of estimation in diffusion tensor imaging. J. Magn. Reson 182 (1), 115–125. doi:10.1016/j.jmr.2006.06.020. [PubMed: 16828568]

Koay CG, Özarslan E, Basser PJ, 2009. A signal transformational framework for breaking the noise floor and its applications in MRI. J. Magn. Reson 197 (2), 108–119. doi:10.1016/j.jmr.2008.11.015. [PubMed: 19138540]

Lampinen B, Zampeli A, Björkman-Burtscher IM, Szczepankiewicz F, Källén K, Compagno Strandberg M, Nilsson M, 2020. Tensor-valued diffusion MRI differentiates cortex and white matter in malformations of cortical development associated with epilepsy. Epilepsia doi:10.1111/epi.16605.

Lasi S, Szczepankiewicz F, Eriksson S, Nilsson M, Topgaard D, 2014. Microanisotropy imaging: quantification of microscopic diffusion anisotropy and orientational order parameter by diffusion MRI with magic-angle spinning of the q-vector. Front. Phys 2, 11.

Lasserre JB, 2007. A sum of squares approximation of nonnegative polynomials. SIAM Rev. 49, 651–669.

Lawrenz M, Koch MA, Finsterbusch J, 2010. A tensor model and measures of microscopic anisotropy for double-wave-vector diffusion-weighting experiments with long mixing times. J. Magn. Reson 202 (1), 43–56. doi:10.1016/j.jmr.2009.09.015. [PubMed: 19854085]

Lenglet C, Rousson M, Deriche R, Faugeras O, 2006. Statistics on the manifold of multivariate normal distributions: theory and application to diffusion tensor MRI processing. J. Math. Imaging Vis 25 (3), 423–444. doi:10.1007/s10851-006-6897-z.

Letac G, Massam H, 1998. Quadratic and inverse regressions for wishart distributions. Ann. Stat 26 (2), 573–595.

Liu C, Bammer R, Acar B, Moseley ME, 2004. Characterizing non-gaussian diffusion by using generalized diffusion tensors. Magn. Reson. Med 51 (5), 924–937. doi:10.1002/mrm.20071. [PubMed: 15122674]

Liu C, Özarslan E, 2019. Multimodal integration of diffusion MRI for better characterization of tissue biology. NMR Biomed. 32 (4), e3939. doi:10.1002/nbm.3939. [PubMed: 30011138]

Magdoom KN, Pajevic S, Dario G, Basser PJ, 2021. A new framework for MR diffusion tensor distribution. Sci. Rep 11 (1), 2766. doi:10.1038/s41598-021-81264-x. [PubMed: 33531530]

Mattiello J, Basser PJ, LeBihan D, 1994. Analytical expressions for the b-matrix in NMR diffusion imaging and spectroscopy. J. Magn. Reson. A 108 (2), 131–141.

Maximov II, Alnaes D, Westlye LT, 2019. Towards an optimised processing pipeline for diffusion magnetic resonance imaging data: effects of artefact corrections on diffusion metrics and their age associations in UK biobank. Hum Brain Mapp 40 (14), 4146–4162. doi:10.1002/hbm.24691. [PubMed: 31173439]

Maximov II, Grinberg F, Shah NJ, 2011. Robust tensor estimation in diffusion tensor imaging. J. Magn. Reson 213 (1), 136–144. doi:10.1016/j.jmr.2011.09.035. [PubMed: 21993763]

Mayerhofer E, 2013. On the existence of non-central wishart distributions. J. Multivariate Anal 114, 448–456.

Mori S, van Zijl PC, 1995. Diffusion weighting by the trace of the diffusion tensor within a single scan. Magn. Reson. Med 33 (1), 41–52. [PubMed: 7891534]

Nilsson M, Szczepankiewicz F, Brabec J, Taylor M, Westin CF, Golby A, van Westen D, Sundgren PC, 2020. Tensor-valued diffusion MRI in under 3 minutes: an initial survey of microscopic anisotropy and tissue heterogeneity in intracranial tumors. Magn. Reson. Med 83 (2), 608–620. doi:10.1002/mrm.27959. [PubMed: 31517401]

Nocedal J, Wright SJ, 2006. Numerical optimization, 2nd Springer-Verlag, Berlin.

Özarslan E, 2009. Compartment shape anisotropy (CSA) revealed by double pulsed field gradient MR. J. Magn. Reson 199 (1), 56–67. doi:10.1016/j.jmr.2009.04.002. [PubMed: 19398210]

Özarslan E, Koay CG, Shepherd TM, Komlosh ME, rfanoglu MO, Pierpaoli C, Basser PJ, 2013. Mean apparent propagator (MAP) MRI: a novel diffusion imaging method for mapping tissue microstructure. Neuroimage 78, 16–32. doi:10.1016/j.neuroimage.2013.04.016. [PubMed: 23587694]

Özarslan E, Mareci TH, 2003. Generalized diffusion tensor imaging and analytical relationships between diffusion tensor imaging and high angular resolution diffusion imaging. Magn. Reson. Med 50 (5), 955–965. doi:10.1002/mrm.10596. [PubMed: 14587006]

Özarslan E, Shemesh N, Koay CG, Cohen Y, Basser PJ, 2011. Nuclear magnetic resonance characterization of general compartment size distributions. New J. Phys 13, 15010. doi:10.1088/1367-2630/13/1/015010.

Özarslan E, Westin CF, Mareci TH, 2015. Characterizing magnetic resonance signal decay due to gaussian diffusion: the path integral approach and a convenient computational method. Concepts Magn. Reson. Part A 44, 203–213.

Özarslan E, Yolcu C, Herberthson M, Knutsson H, Westin CF, 2018. Influence of the size and curvedness of neural projections on the orientationally averaged diffusion MR signal. Front. Phys 6, 17. [PubMed: 29675413]

Özarslan E, Yolcu C, Herberthson M, Westin CF, Knutsson H, 2017. Effective potential for magnetic resonance measurements of restricted diffusion. Front. Phys 5, 68. [PubMed: 29629371]

Paulsen JL, Özarslan E, Komlosh ME, Basser PJ, Song YQ, 2015. Detecting compartmental non-gaussian diffusion with symmetrized double-PFG MRI. NMR Biomed. 28 (11), 1550–1556. doi:10.1002/nbm.3363. [PubMed: 26434812]

Pennec X, Fillard P, Ayache N, 2006. A riemannian framework for tensor computing. Int. J. Comput. Vis 66 (1), 41–66.

Qi L, Yu G, Wu EX, 2010. Higher order positive semidefinite diffusion tensor imaging. SIAM J. Imaging Sci 3 (3), 416–433. doi:10.1137/090755138.

Reymbaut A, Mezzani P, de Almeida MJ, Topgaard D, 2020. Accuracy and precision of statistical descriptors obtained from multidimensional diffusion signal inversion algorithms. NMR Biomed. 33, e4267. [PubMed: 32067322]

Röding M, Bernin D, Jonasson J, Särkkä A, Topgaard D, Rudemo M, Nydén M, 2012. The gamma distribution model for pulsed-field gradient NMR studies of molecular-weight distributions of polymers. J. Magn. Reson 222, 105–111. doi:10.1016/j.jmr.2012.07.005. [PubMed: 22864268]

Sen PN, Hürlimann MD, de Swiet TM, 1995. Debye-porod law of diffraction for diffusion in porous media. Phys. Rev. B 51 (1), 601–604.

Shakya S, Batool N, Özarslan E, Knutsson H, 2017. Multi-fiber Reconstruction Using Probabilistic Mixture Models for Diffusion MRI Examinations of the Brain. In: Schultz T, Özarslan E, Hotz I (Eds.), Modeling, Analysis, and Visualization of Anisotropy. Springer International Publishing, Cham, pp. 283–308.

Stejskal EO, Tanner JE, 1965. Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. J. Chem. Phys 42 (1), 288–292.

de Swiet TM, Mitra PP, 1996. Possible systematic errors in single-shot measurements of the trace of the diffusion tensor. J. Magn. Reson. B 111 (1), 15–22. doi:10.1006/jmrb.1996.0055. [PubMed: 8661259]

Szczepankiewicz F, Hoge S, Westin CF, 2019. Linear, planar and spherical tensor-valued diffusion MRI data by free waveform encoding in healthy brain, water, oil and liquid crystals. Data Brief 25, 104208. doi:10.1016/j.dib.2019.104208. [PubMed: 31338402]

Szczepankiewicz F, Lasi S, van Westen D, Sundgren PC, Englund E, Westin CF, Ståhlberg F, Lätt J, Topgaard D, Nilsson M, 2015. Quantification of microscopic diffusion anisotropy disentangles effects of orientation dispersion from microstructure: applications in healthy volunteers and in brain tumors. Neuroimage 104, 241–252. doi:10.1016/j.neuroimage.2014.09.057. [PubMed: 25284306]

Szczepankiewicz F, van Westen D, Englund E, Westin CF, Ståhlberg F, Lätt J, Sundgren PC, Nilsson M, 2016. The link between diffusion MRI and tumor heterogeneity: mapping cell eccentricity and density by diffusional variance decomposition (DIVIDE). Neuroimage 142, 522–532. doi:10.1016/j.neuroimage.2016.07.038. [PubMed: 27450666]

Tax CMW, Otte WM, Viergever MA, Dijkhuizen RM, Leemans A, 2015. "REKINDLE": Robust extraction of kurtosis indices with linear estimation. Magn. Reson. Med 73 (2), 794–808. doi:10.1002/mrm.25165. [PubMed: 24687400]

Topgaard D, 2019. Diffusion tensor distribution imaging. NMR Biomed. 32 (5), e4066. doi:10.1002/nbm.4066. [PubMed: 30730586]

Tournier JD, Calamante F, Connelly A, 2007. Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution. Neuroimage 35 (4), 1459–1472. doi:10.1016/j.neuroimage.2007.02.016. [PubMed: 17379540]

Tournier JD, Calamante F, Gadian DG, Connelly A, 2004. Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. Neuroimage 23, 1176–1185. [PubMed: 15528117]

Veraart J, Van Hecke W, Sijbers J, 2011. Constrained maximum likelihood estimation of the diffusion kurtosis tensor using a rician noise model. Magn. Reson. Med 66 (3), 678–686. doi:10.1002/mrm.22835. [PubMed: 21416503]

Wang Z, Vemuri BC, Chen Y, Mareci TH, 2004. A constrained variational principle for tensor field restoration from complex-valued DWI. IEEE Trans. on Medical Imaging 23 (8), 930–939. [PubMed: 15338727]

Westin CF, Knutsson H, Pasternak O, Szczepankiewicz F, Özarslan E, van Westen D, Mattisson C, Bogren M, O'Donnell LJ, Kubicki M, Topgaard D, Nilsson M, 2016. Q-Space trajectory imaging for multidimensional diffusion MRI of the human brain. Neuroimage 135, 345–362. doi:10.1016/j.neuroimage.2016.02.039. [PubMed: 26923372]

Wong EC, Cox RW, Song AW, 1995. Optimized isotropic diffusion weighting. Magn. Reson. Med 34 (2), 139–143. [PubMed: 7476070]

Yolcu C, Memiç M, im ek K, Westin CF, Özarslan E, 2016. NMR Signal for particles diffusing under potentials: from path integrals and numerical methods to a model of diffusion anisotropy. Phys. Rev. E 93, 052602. [PubMed: 27300946]

Zucchelli M, Afzali M, Yolcu C, Westin CF, Menegaz G, Özarslan E, 2016. The Confinement Tensor Model Improves Characterization of Diffusion-weighted Magnetic Resonance Data with Varied Timing Parameters. Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on. IEEE. 1093–1096

Zwiers MP, 2010. Patching cardiac and head motion artefacts in diffusion-weighted images. Neuroimage 53 (2), 565–575. doi:10.1016/j.neuroimage.2010.06.014. [PubMed: 20600997]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript
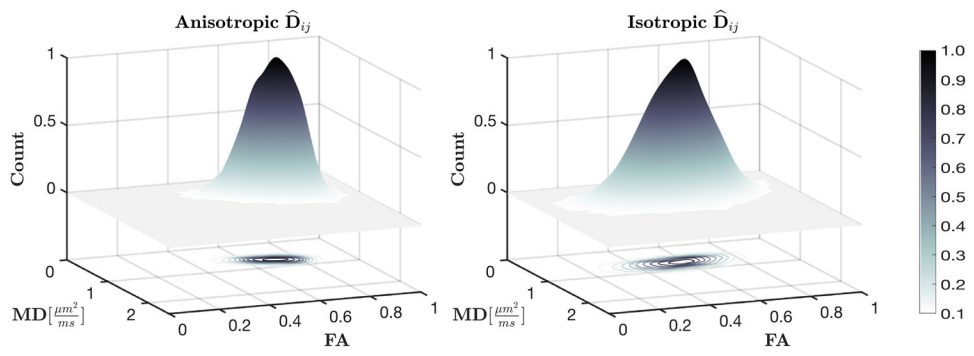
Fig. 1.
Joint distribution of MD and FA values of the diffusion tensors in the simulated DTDs. Left: DTD with anisotropic mean. Right: DTD with isotropic mean.
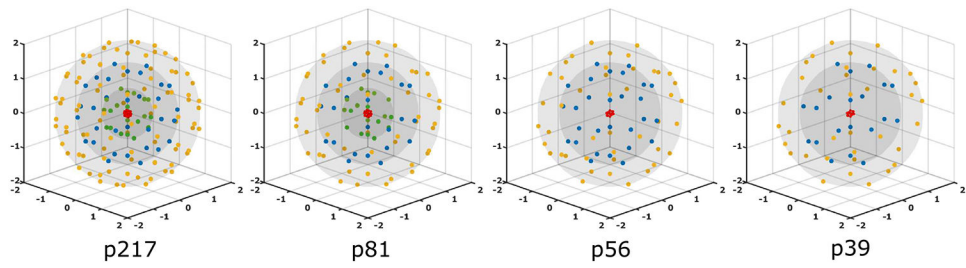
Fig. 2.
The four protocols considered in this study for the analysis of the experimental data. From left to right: p217, p81, p56, and p39 refer to the protocols having 217, 81, 56, and 39 volumes, respectively. The colored dots show the initial direction of each diffusion gradient waveform. Red, green, blue, and yellow dots indicate such directions for samples at b-values of 0.1, 0.7, 1.4, and 2.0 ms/μm^2, respectively.
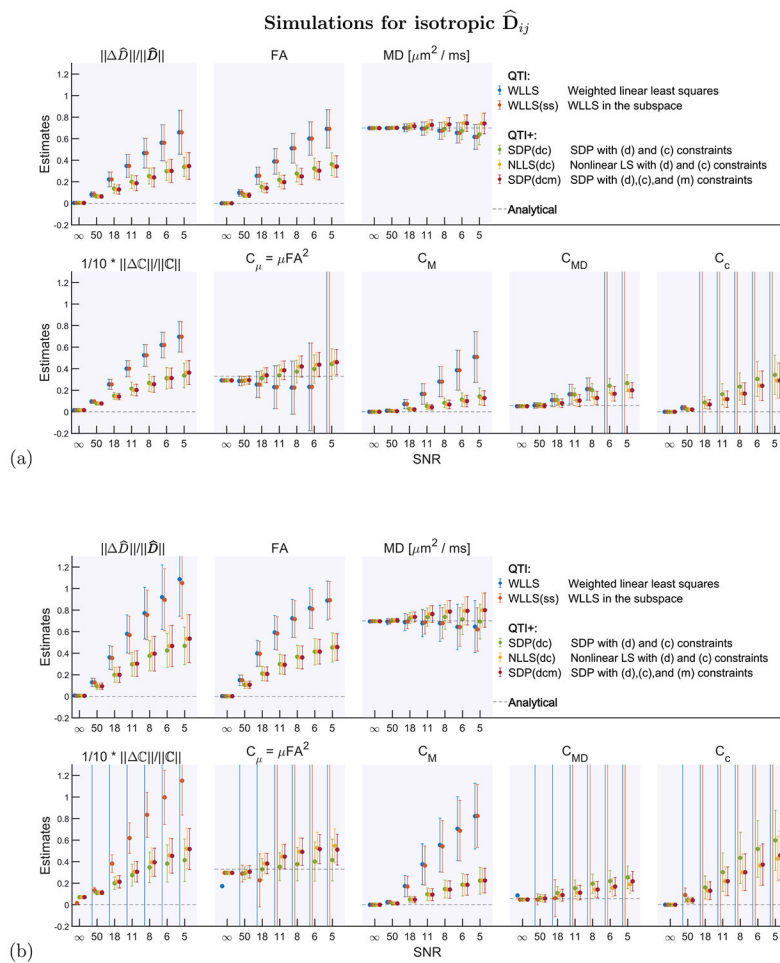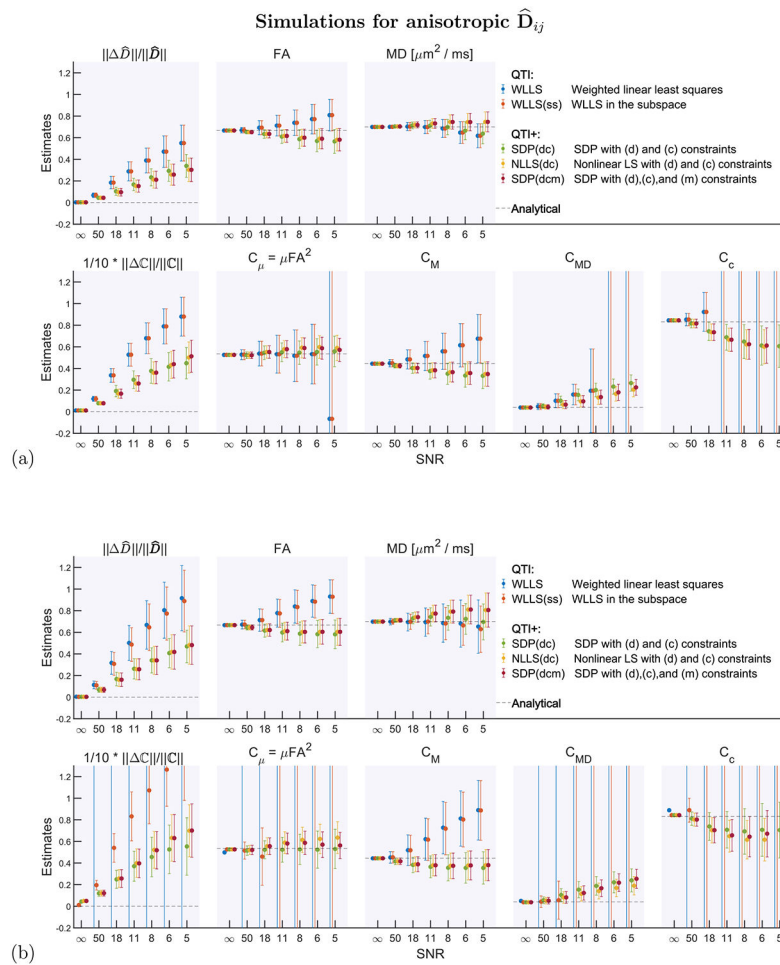
Fig. 3.
Simulations for isotropic $\widehat{\mathbf{D}}_{ij}$, and Rician distributed noisy signals. Frobenius norms (indicated by $\|\cdot\|$) and the estimated measures under varying noise levels for the estimation methods considered. $\mathbb{C}$ stands for $\mathbb{C}_{ijkl}$ while $\widehat{\mathbf{D}}$ stands for $\widehat{\mathbf{D}}_{ij}$. (a) protocol p217. (b) protocol p56s.

Fig. 4.

Simulations for anisotropic $\widehat{\mathbf{D}}_{ij}$, and Rician distributed noisy signals. Frobenius norms (indicated by $\|\cdot\|$) and the estimated measures under varying noise levels for the estimation methods considered. $\mathbb{C}$ stands for $\mathbb{C}_{ijkl}$ while $\widehat{\mathbf{D}}$ stands for $\widehat{\mathbf{D}}_{ij}$. (a) protocol p217. (b) protocol p56s.
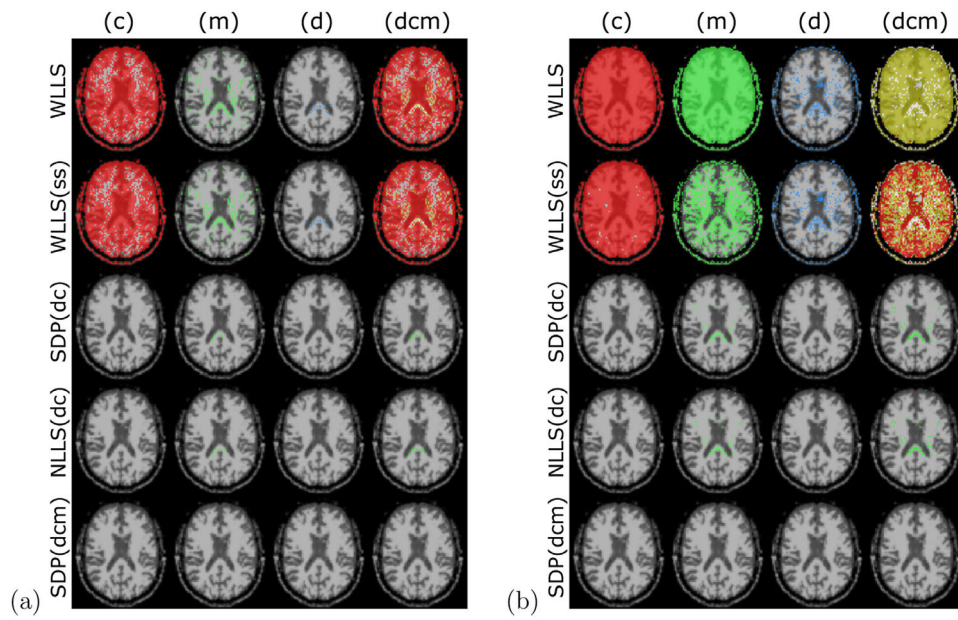
Fig. 5.
The violations of the three constraints are color encoded and depicted on brain images. Red, green, and blue indicate violation of conditions (c), (m), and (d) respectively. The last column indicates the constraints together where yellow indicates violations of (c) and (m), magenta (c) and (d), and cyan (d) and (m). All three conditions are violated in white pixels. (a) protocol p217. (b) protocol p56.
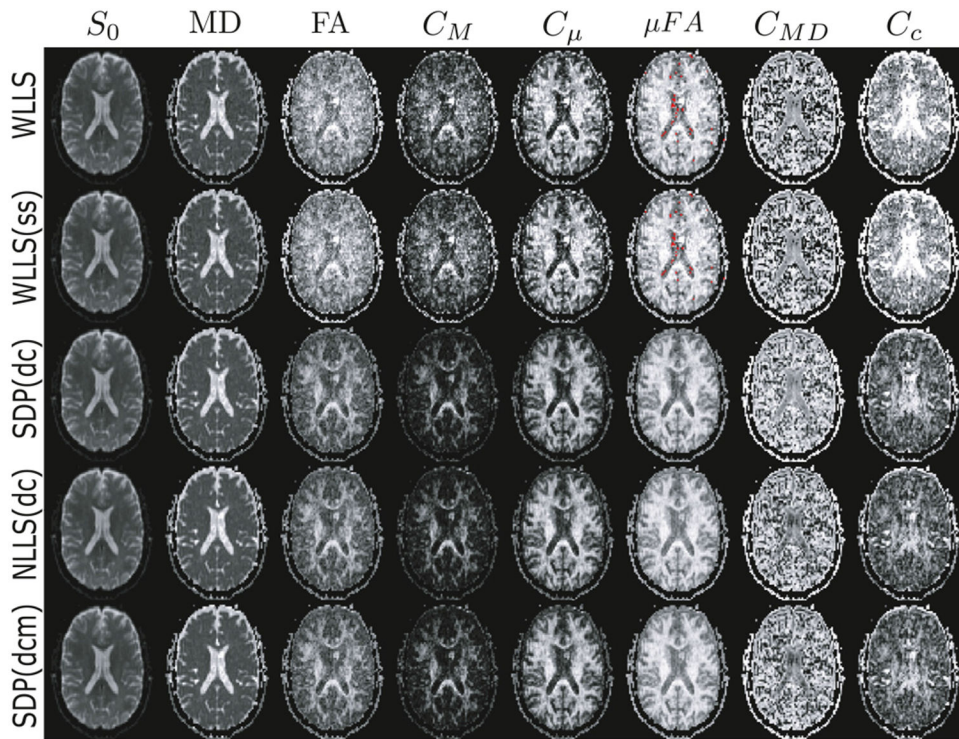
Fig. 6.
Maps estimated through various methods from data involving 56 volumes. Red voxels on the μFA maps indicate imaginary values. Despite the voxel-by-voxel analysis, the QTI+ maps (last three rows) are visibly smoother than the QTI maps employing weighted linear estimations (first two rows).
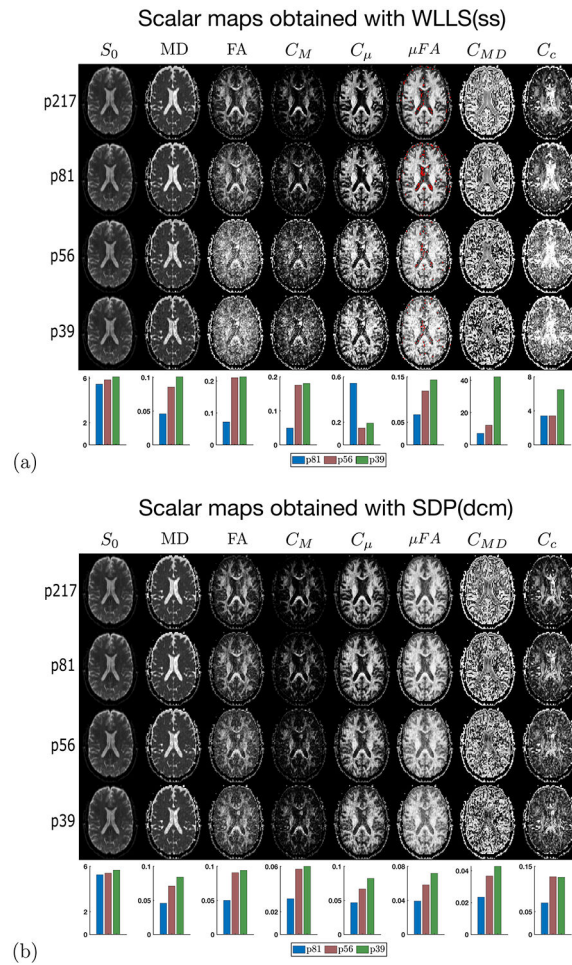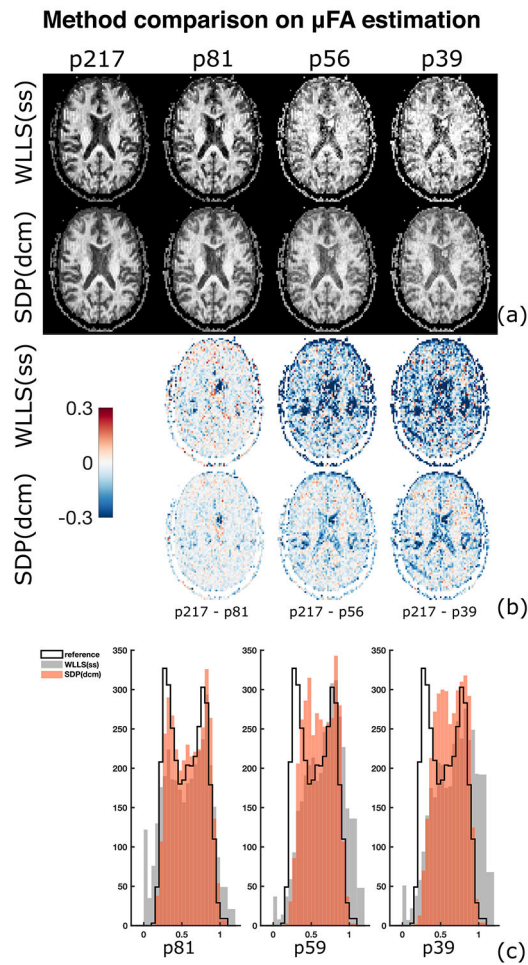
Fig. 7.
Scalar maps obtained by employing WLLS(ss) and SDP(dcm) estimation schemes on p217, p81, p56, and p39 protocols. Red pixels on the μFA maps indicate complex values. The bars on the last rows of panels (a) and (b) show the mean absolute deviation of the respective parameters due to downsampling the p217 protocol. Here, S_0 has arbitrary units while MD is expressed in $\mu m^2/ms$. (a) WLLS(ss), (b) SDP(dcm).

Method comparison on FA estimation
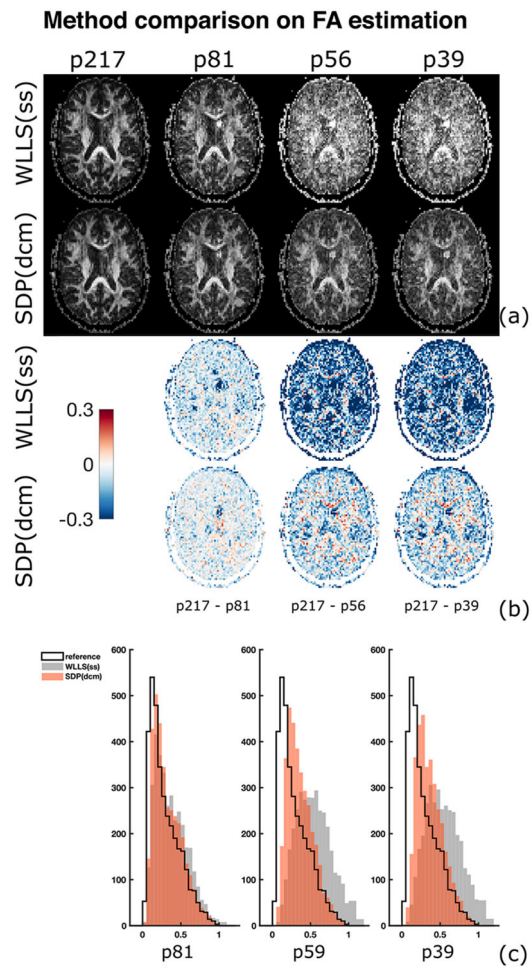
Fig. 8.
(a) FA maps obtained by fitting the QTI model with WLLS(ss) and SDP(dcm) on the four protocols. The fits performed with both methods on the p217 protocol are used as reference. (b) Difference between the reference FA maps and those estimated with both methods on the three downsampled protocols. (c) Histograms showing the distribution of FA values for the three protocols.

Fig. 9.
(a) μFA maps obtained by fitting the QTI model with WLLS(ss) and SDP(dcm) on the four protocols. The fits performed with both methods on the p217 protocol are used as reference. (b) Difference between the reference μFA maps and those estimated with both methods on the three downsampled protocols. (c) Histograms showing the distribution of μFA values for the three protocols.
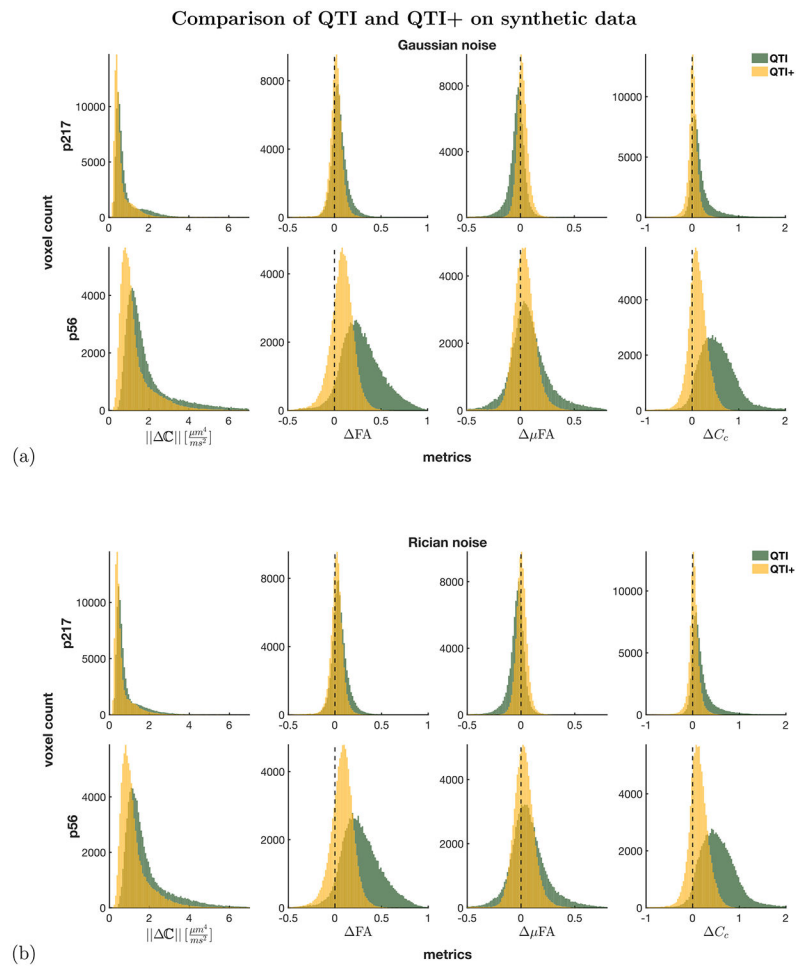
Fig. 10.

Comparison of QTI and QTI+ on synthetic data. Frobenius norm of the difference between the ground truth and estimated $\mathbb{C}_{ijkl}$ (indicated by $||\Delta\mathbb{C}||$) and differences between the estimated and ground truth metrics. Positive and negative values in the difference plots indicate parameters over- and under-estimated, respectively. (a) Gaussian noise. (b) Rician noise.

Table 1

The five protocols considered in this study. The protocol p217 contains thirteen non-diffusion weighted images while the others have one such image.

Protocol	Encoding	b-values [ms/m^2]	Samples per shell
p217	LTE	0.1, 0.7, 1.4, 2.0	10,10,16,46
p217	PTE	0.1, 0.7, 1.4, 2.0	10,10,16,46
p217	STE	0.1, 0.7, 1.4, 2.0	10,10,10,10
p81	LTE	0.1, 0.7, 1.4, 2.0	6,6,10,21
p81	STE	0.1, 0.7, 1.4, 2.0	6,6,10,15
p56	LTE	0.1, 1.4, 2.0	4,10,15
p56	STE	0.1, 1.4, 2.0	6,10,10
p56s	LTE	0.1, 1.0, 2.0	4,10,15
p56s	STE	0.1, 1.0, 2.0	6,10,10
p39	LTE	0.1, 1.4, 2.0	4,10,15
p39	STE	0.1, 1.4, 2.0	3,3,3

Table 2

Run times for the protocols with 217 and 56 volumes.

Protocol	Fitting Routine	Run Times
p217	SDP(dc), single voxel	43 min
p217	SDP(dc), multi voxel	6 min
p217	NLLS(dc), single voxel	11 min
p217	m-check + SDP(dcm), single voxel	37 min
p217	m-check + SDP(dcm), multi voxel	5 min
p217	SDP(dcm), multi voxel	14 min
p217	WLLS(ss)	4 s
p217	WLLS	2 s
p56	SDP(dc), single voxel	50 min
p56	SDP(dc), multi voxel	10 min
p56	NLLS(dc), single voxel	15 min
p56	m-check + SDP(dcm), single voxel	38 min
p56	m-check + SDP(dcm), multi voxel	10 min
p56	SDP(dcm), multi voxel	14 min
p56	WLLS(ss)	2 s
p56	WLLS	5 s