



## Enforcing necessary non-negativity constraints for common diffusion MRI models using sum of squares programming

Dela Haije, Tom; Ozarlan, Evren; Feragen, Aasa

*Published in:*  
NeuroImage

*Link to article, DOI:*  
[10.1016/j.neuroimage.2019.116405](https://doi.org/10.1016/j.neuroimage.2019.116405)

*Publication date:*  
2020

*Document Version*  
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

*Citation (APA):*  
Dela Haije, T., Ozarlan, E., & Feragen, A. (2020). Enforcing necessary non-negativity constraints for common diffusion MRI models using sum of squares programming. *NeuroImage*, 209, Article 116405. <https://doi.org/10.1016/j.neuroimage.2019.116405>

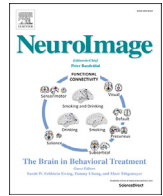
---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



# Enforcing necessary non-negativity constraints for common diffusion MRI models using sum of squares programming



Tom Dela Haije<sup>a,\*</sup>, Evren Özarslan<sup>b,c</sup>, Aasa Feragen<sup>d</sup>

<sup>a</sup> Department of Computer Science, University of Copenhagen, Copenhagen, Denmark

<sup>b</sup> Department of Biomedical Engineering, Linköping University, Linköping, Sweden

<sup>c</sup> Center for Medical Image Science and Visualization, Linköping University, Linköping, Sweden

<sup>d</sup> Department of Applied Mathematics and Computer Science, Technical University of Denmark, Lyngby, Denmark

## ARTICLE INFO

### Keywords:

Constrained optimization  
Cumulant expansion  
Diffusion MRI  
Diffusional kurtosis imaging  
Diffusion tensor imaging  
Mean apparent propagator  
Sampling scheme design  
Semidefinite programming  
Spherical deconvolution  
Sum of squares optimization  
Sum of squares polynomials

## ABSTRACT

In this work we investigate the use of sum of squares constraints for various diffusion-weighted MRI models, with a goal of enforcing strict, global non-negativity of the diffusion propagator. We formulate such constraints for the mean apparent propagator model and for spherical deconvolution, guaranteeing strict non-negativity of the corresponding diffusion propagators. For the cumulant expansion similar constraints cannot exist, and we instead derive a set of auxiliary constraints that are necessary but not sufficient to guarantee non-negativity. These constraints can all be verified and enforced at reasonable computational costs using semidefinite programming. By verifying our constraints on standard reconstructions of the different models, we show that currently used weak constraints are largely ineffective at ensuring non-negativity. We further show that if strict non-negativity is not enforced then estimated model parameters may suffer from significant errors, leading to serious inaccuracies in important derived quantities such as the main fiber orientations, mean kurtosis, etc. Finally, our experiments confirm that the observed constraint violations are mostly due to measurement noise, which is difficult to mitigate and suggests that properly constrained optimization should currently be considered the norm in many cases.

## 1. Introduction

Diffusion-weighted magnetic resonance imaging (MRI) captures local micro-structural information by observing diffusing (water) molecules probing their surroundings at a microscopic scale. In order to analyze this type of data one can either estimate parameters that describe the diffusion itself, which provides a somewhat abstract but accurate description of the observed stochastic motion, or one can use a model of the ambient structure that re-expresses the observed diffusion in terms of more intuitive structural parameters. Both cases generally rely on optimization to reconstruct the descriptive parameters from diffusion-weighted images, and in this work we introduce a specific set of basic constraints to improve such model reconstructions.

Constrained optimization guarantees that known or assumed bounds on parameters are respected, thereby improving the accuracy of a reconstruction. It also gives a sense of the data quality and the suitability of the model—in an ideal setting constrained and unconstrained results *should* be identical. In practice, large discrepancies between the two may point to problems such as artifacts or excessive noise in the data, or could

indicate that the selected model is not appropriate for the considered data. Finally, constrained optimization can be important in further analysis steps, which may (implicitly) rely on physical or mathematical properties of the model/basis expansion that could be lost in the case of unsatisfied constraints. In diffusion tensor imaging (DTI) (Basser et al., 1994a) for example, a non-positive-definite diffusion tensor can result in unphysical ‘negative’ apparent diffusion or extreme fractional anisotropy values.

The use of constraints has been prevalent in diffusion MRI since the early days of DTI, where the diffusion tensor can be restricted to be positive-definite through various means (Pennec et al., 2006; Fillard et al., 2005; Lenglet et al., 2006; Koay, 2010; Koay et al., 2006; Wang et al., 2004). Constrained optimization has also been applied in e.g. diffusional kurtosis imaging (DKI) (Jensen et al., 2005; Veraart et al., 2011) and higher order tensor models such as the generalized DTI model proposed by Özarslan and Mareci (2003), cf. (Chen et al., 2013; Qi et al., 2010; Ghosh et al., 2014; Barmpoutis and Vemuri, 2010; Barmpoutis et al., 2009, 2007, 2012), while common implementations of spherical deconvolution (SD) methods (Tournier et al., 2007) and of the mean

\* Corresponding author.

E-mail address: [tom@di.ku.dk](mailto:tom@di.ku.dk) (T. Dela Haije).

apparent propagator (MAP) method (Özarslan et al., 2013b) often rely on soft constraints or regularization (Fick et al., 2016). For these and related models, non-negativity constraints based on square root representations have also been studied extensively (Hanyga and Sereďnyška, 2012; Goh et al., 2009; Cheng et al., 2009, 2014; 2011, 2012; Vemuri et al., 2019; Sun et al., 2013). Practical square root representations impose potentially significant restrictions on the set of permitted parameter values, but they do guarantee non-negativity globally without any compromise on speed. Strict constraints have also been proposed for functions expressed in terms of spherical harmonics by Schwab et al. (2012).

Although constrained reconstruction obviously has many advantages, and its development is being actively pursued by numerous research groups, its use and availability in popular software packages is not as ubiquitous as one might expect. One reason for this is the fact that constrained optimization is (typically significantly) slower than its unconstrained counterpart, which together with the size of modern diffusion-weighted data sets leads to restrictive hardware requirements. Constrained optimization algorithms generally also require a nontrivial reformulation of the parameter estimation problem to match a specific form, complicating their implementation, and in some cases efficient algorithms may not be available.

In this paper we will derive a set of basic mathematical constraints for diffusion-weighted MRI (Section 2). These constraints are based on the non-negativity of the so-called ensemble average propagator or associated functions, but reformulated as the (relaxed) condition that these functions can be written as a sum of squared (SOS) polynomials. For many commonly used models and basis expansions these constraints take the form of a positive-definiteness condition on a matrix that is linear in the model parameters, which can thus be enforced efficiently through the use of semidefinite programming. We derive explicit constraints for MAP MRI, spherical deconvolution, and for the cumulant expansion (CE) (Liu et al., 2004; Kiselev, 2010), and show that the application of these constraints should be considered essential in many situations despite the associated computational costs (Sections 3 and 4). Next, we use the presence of unsatisfied constraints as a data quality measure to review the performance of grid- and shell-based acquisition sequences, and to assess the importance of additional corrections in the preprocessing pipeline. To emphasize the generality of the strict constraints we propose here, and to contrast them with existing soft approaches to non-negative propagator constraints, we denote constrained models with a ‘+’ suffix (i.e., MAP+, CSD+, CE+).

SOS polynomials have been considered previously in the context of diffusion-weighted MRI by several authors, specifically to enforce positive-definiteness of the (generalized) diffusion tensor (Özarslan and Mareci, 2003). The first explicit use of these constraints that we are aware of is the work by Barmpoutis et al. (2009, 2007) on the constrained reconstruction of fourth order diffusion tensors—a special case in which the SOS constraints are exactly equal to the desired non-negativity constraints. In subsequent works the authors enforce non-negativity on diffusion tensors of orders greater than four, by approximating the SOS constraints with a set of linear constraints (Barmpoutis and Vemuri, 2010; Barmpoutis et al., 2012). A formulation for general higher order tensors based on semidefinite programming was later provided by Chen et al. (2013). We are unaware of any works that investigated SOS constraints for the more general models that we discuss in this work. Parts of the current manuscript have been presented in preliminary form as conference and workshop abstracts (Dela Haije and Feragen, 2018; Dela Haije et al., 2015, 2019).

## 2. Methods

In the following we first review some background on diffusion MRI, Section 2.1. We then describe a set of general constraints and their practical implementation in Section 2.2. In Section 2.3 we discuss their application in three relevant and commonly used models/basis expansions of the diffusion-weighted signal. Sections 2.4 and 2.5 give some

details regarding implementation and cover the data used in the experiments presented in Section 3.

### 2.1. Diffusion MRI fundamentals

The normalized diffusion-weighted signal  $S$  is generated by the average molecular motion over a time  $\Delta$  within a single voxel. This voxel-averaged diffusion can be described by a (bounded and Lebesgue integrable) displacement probability density function  $P$ , where  $P(\mathbf{r})$  gives the probability of a displacement  $\mathbf{r}$  occurring within a voxel over a time  $\Delta$ . The signal  $S$  and the density function  $P$  are related by a Fourier transform (Mitra and Halperin, 1995; Novikov and Kiselev, 2010; Callaghan, 1991; Callaghan et al., 1990, 1988; Özarslan et al., 2009)

$$S(\mathbf{q}) = \int_{\mathbb{R}^3} e^{-i \mathbf{q} \cdot \mathbf{r}} P(\mathbf{r}) d\mathbf{r}, \quad (1)$$

where  $\cdot$  denotes the standard inner product,  $\mathbf{r} \in \mathbb{R}^3$  is a displacement vector, and  $\mathbf{q} \in \mathbb{R}^3$  is the gradient wave vector. The gradient wave vector is defined as  $\mathbf{q} := \gamma \delta \mathbf{G}$ , combining the gyromagnetic ratio  $\gamma$ , the gradient pulse width  $\delta$ , and the applied gradient  $\mathbf{G} \in \mathbb{R}^3$ . The measured (non-normalized) signal  $\tilde{S}$  is related to  $S$  with the normalization factor  $\tilde{S}_0 := \tilde{S}(0)$  according to  $S = \tilde{S}/\tilde{S}_0$ . We take  $P(\mathbf{r}) = P(-\mathbf{r})$  so that the normalized signal in Eq. (1) is real-valued, where we assume that (active) flow is negligible in our data (Musiel et al., 2017). The probability density function  $P$  is also referred to as the average propagator.

A typical diffusion-weighted MRI data set consists of a relatively small set of diffusion-weighted measurements acquired with different  $\mathbf{q}$  and/or different timing parameters. To aid the interpretation and further processing of these measurements we can reconstruct models or representations of the signal  $S$ , where the basic fitting procedure is as follows. Given a set of applied wave vectors  $\{\mathbf{q}_i\}_{i=1}^N$ , the corresponding measurements  $y_i$  (representing  $\tilde{S}(\mathbf{q}_i)$  or  $\log \tilde{S}(\mathbf{q}_i)$ ), and a model or basis expansion  $m_\omega$  (of  $\tilde{S}$  or  $\log \tilde{S}$ ) with a set of parameters  $\omega$ , we estimate the optimal parameters by minimizing the sum of squares of the residuals

$$\arg \min_{\omega} \sum_{i=1}^N (m_\omega(\mathbf{q}_i) - y_i)^2. \quad (2)$$

Depending on the specific techniques and algorithms used to solve the least squares problem, this formulation can be adapted.

### 2.2. Relaxed non-negativity constraints

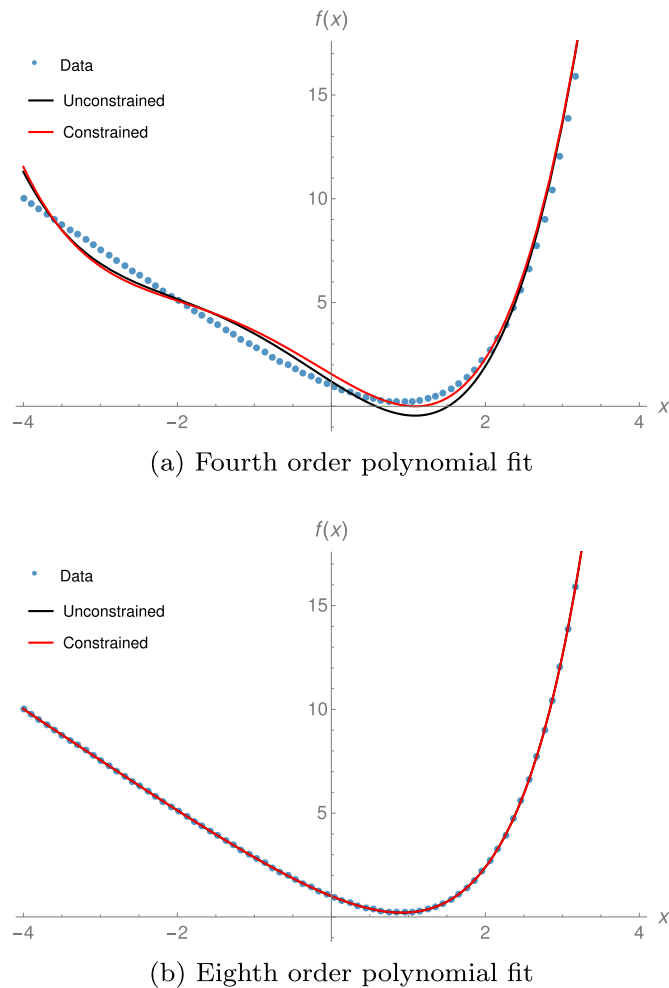
When solving the optimization problem (2) for a specific diffusion MRI signal/model/expansion, it is often desired that the result is guaranteed to satisfy certain constraints. The fundamental property from which our constraints are derived is non-negativity of the displacement probability density function  $P$ , i.e., the requirement that  $P(\mathbf{r}) \geq 0$  for all  $\mathbf{r} \in \mathbb{R}^3$ . Unlike e.g. constraints based on the physiologically expected values of a parameter, non-negativity of the propagator is essentially a universally accepted, purely mathematical constraint.

Ideally we would like to directly translate the non-negativity of  $P$  to a set of constraints on the signal domain parameters  $\omega$  of a given expression using e.g. Bochner’s theorem (Rudin, 1996), but this only leads to practical constraints in exceptional cases. The best known example is that of the diffusion tensor imaging model, where positive-definiteness of the diffusion tensor implies positivity of  $P$  (Wang et al., 2004). The vast majority of models, however, allow no useful direct correspondences between non-negativity of  $P$  and the model parameters, and in these cases we must resort to increasingly relaxed modifications of the non-negativity constraint.

In order to obtain a practical (but more restrictive) method to guarantee non-negativity, let us consider the problem of constraining (real) polynomials to be non-negative. In this setting we have exact and veri-

fiable conditions for non-negativity of degree 2 polynomials of arbitrary variables, and for arbitrary degree polynomials of a single variable, but outside of these special cases it is generally an NP-hard problem to determine whether a given polynomial is non-negative (Parrilo, 2003). However, non-negativity of a degree  $2d$  polynomial  $\mathcal{P}$  is trivially guaranteed if  $\mathcal{P}$  can be written as a sum of squared polynomials, i.e., if there exists a finite set of polynomials  $g_1, g_2, \dots$  such that  $\mathcal{P}(x) = \sum_k g_k(x)^2$ , in which case we say that  $\mathcal{P}$  is SOS. We could thus optimize over the set of SOS polynomials instead of the set of non-negative polynomials, and obtain a reconstruction that satisfies non-negativity. There is of course a cost associated to this relaxation, which comes from the fact that not every non-negative polynomial can be written as a sum of squared polynomials (Hilbert, 1933). But because non-negative polynomials (on a compact domain) can be approximated to an arbitrary degree of accuracy by SOS polynomials (Berg et al., 1976; Lasserre, 2007)—the set of SOS polynomials is dense in the set of non-negative polynomials for the  $L^1$  norm—we consider a restriction to the set of sum of squares polynomials a warranted relaxation. In Fig. 1 we show an example of a sum of squares constrained polynomial fit in one dimension.

The relaxed non-negativity constraints based on SOS polynomials can be enforced in practice as described in the work of Parrilo (2003). Let  $\mathcal{P}_m$



**Fig. 1.** A one dimensional example of a polynomial fit to an artificial data set (blue dots, generated by sampling  $f(x) = e^x - 2.5x$  on the range  $[-4, 4]$ , 100 samples) under sum of squares constraints (red) and without any constraints (black). (a) Although the actual function is strictly positive, unconstrained optimization for low order polynomials can become negative where the constrained fit remains non-negative. (b) In this noiseless example an unconstrained higher order polynomial fit will also be non-negative, in which case the constrained and unconstrained fits will be the same.

be a degree  $2d$  polynomial associated to  $m$ , for example employed for representing/approximating the propagator  $P$ . If  $\mathcal{P}_m$  can be written as a sum of squared polynomials, there must exist a positive semidefinite matrix  $A$  such that

$$\mathcal{P}_m(x) = e(x)^T \cdot A \cdot e(x), \quad (3)$$

where  $e$  is a vectorized polynomial basis for polynomials in  $x$  up to degree  $d$ . The original optimization problem (2) can thus be reformulated to include the relaxed non-negativity constraint as

$$\begin{aligned} \arg \min_{A, \omega} \sum_{i=1}^m (m_\omega(q_i) - y_i)^2 \\ \text{subject to } e^T \cdot A \cdot e = \mathcal{P}_{m_\omega} \\ A \succeq 0, \end{aligned} \quad (4)$$

where  $\succeq 0$  denotes positive semidefiniteness. This is a semidefinite programming problem that can be solved efficiently using e.g. interior point methods. Furthermore, if we fix the parameters  $\omega$  in the constrained problem (4) to the solution of the unconstrained problem (2), then the interior point method will determine whether the selected parameters result in a feasible problem, i.e., if the unconstrained fit already satisfies the proposed SOS polynomial-based constraint. The remaining challenge is now to find polynomials  $\mathcal{P}_m$  for a given model  $m$  of  $S$ , such that the solution to problem (4) guarantees that the corresponding probability density function  $P$  is non-negative.

### 2.3. Constraints for common diffusion MRI models

We will consider sum of squares constraints for three models/basis expansions for the diffusion-weighted signal, which we consider representative of the many proposed models; see e.g. (Assemlal et al., 2011; Panagiotaki et al., 2012; Ghosh and Deriche, 2016) for a comprehensive overview.

#### 2.3.1. Mean apparent propagator

The first expression we consider is the mean apparent propagator model proposed by Özarslan et al. (2013b). This model expresses the signal  $S$  as a weighted sum of Hermite functions (Özarslan et al., 2013a, 2008; Abramowitz and Stegun, 1964), which endow the representation with a number of convenient properties (Özarslan et al., 2013b, Introduction). Most importantly, Hermite functions are eigenfunctions of the Fourier transform that links the signal  $S$  to the propagator  $P$ , which means that they allow an analytical expression for  $P$  given a MAP reconstruction of  $S$ .

#### 2.3.2. The order $K$ MAP model has the functional form

$$S(q) \approx \sum_{k=0,2,\dots,K; n_1+n_2+n_3=k} a_{n_1 n_2 n_3} \Phi_{n_1 n_2 n_3}(S \cdot q) \quad (5)$$

where  $a_{n_1 n_2 n_3} := a_{n_1 n_2 n_3}(S)$  are the parameters and  $n_1, n_2, n_3$  non-negative integers. The matrix  $S$  is a scaling matrix, typically defined as

$$S = \text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \sqrt{\lambda_3}) \cdot Q^T \quad (6)$$

with  $\lambda$  the eigenvalues of the cumulant tensor  $C_2$  (cf. Kiselev, 2010), see Eq. (17) for further details) and  $Q$  a matrix of its eigenvectors, so that  $S^T \cdot S = C_2$ . With this definition the scaling matrix  $S$  ensures that the lowest order term ( $k = 0$ ) in the expansion (5) corresponds to the expected Gaussian behavior described by e.g. DTI. The Hermitian basis functions  $\Phi_{n_1 n_2 n_3}$  are defined as

$$\Phi_{n_1 n_2 n_3}(q) = \frac{(-1)^{\frac{k}{2}} e^{-\frac{1}{2}\|q\|^2}}{\sqrt{2^k n_1! n_2! n_3!}} H_{n_1}(q_1) H_{n_2}(q_2) H_{n_3}(q_3), \quad (7)$$

with  $k$  as before and  $H_n$  the  $n^{\text{th}}$  order Hermite polynomial

$$H_n(q) = (-1)^n e^{q^2} \frac{d^n}{dq^n} e^{-q^2}. \quad (8)$$

The corresponding expression for the basis that describes the displacement probability density function is

$$\Psi_{n_1 n_2 n_3}(\mathbf{r}) = \frac{(2\pi)^{-\frac{3}{2}} e^{-\frac{1}{2}\|\mathbf{r}\|^2}}{\sqrt{2^k n_1! n_2! n_3!}} H_{n_1}(r_1) H_{n_2}(r_2) H_{n_3}(r_3), \quad (9)$$

so that

$$P(\mathbf{r}) = \frac{1}{|\det \mathbf{S}|} \sum_{k=0,2,\dots,K; n_1+n_2+n_3=k} a_{n_1 n_2 n_3} \Psi_{n_1 n_2 n_3}(\mathbf{S}^{-T} \cdot \mathbf{r}) \quad (10)$$

with  $\mathbf{S}^{-T}$  the transpose of  $\mathbf{S}^{-1}$ . By making use of the aforementioned fact that Hermite functions are eigenfunctions of the Fourier transform,  $\mathbf{S}$  and  $P$  can thus be described using the same coefficients  $a_{n_1 n_2 n_3}$ . The above expressions are modified somewhat from the expressions given by Özarlan et al. (2013b), since we restrict ourselves to the case  $P(\mathbf{r}) = P(-\mathbf{r})$  and use a different definition of the gradient wave vector  $\mathbf{q}$ .

Sum of squares constraints can be applied to MAP MRI fairly straightforwardly by defining the polynomial

$$\mathcal{P}(\mathbf{r}) = \sum_{k=0,2,\dots,K; n_1+n_2+n_3=k} \frac{a_{n_1 n_2 n_3}}{\sqrt{2^k n_1! n_2! n_3!}} H_{n_1}(r_1) H_{n_2}(r_2) H_{n_3}(r_3), \quad (11)$$

whose non-negativity implies non-negativity of the MAP displacement probability density function, Eq. (10). This follows from the fact that the positive factors  $(2\pi)^{-\frac{3}{2}} e^{-\frac{1}{2}\|\mathbf{r}\|^2}$  in Eq. (9) and the factor  $|\det \mathbf{S}|^{-1}$  in Eq. (10) can be factored out, leaving non-negativity of the above polynomial as the sole requirement after a change of basis.

### 2.3.3. Spherical deconvolution

The second expression for  $S$  considered here is the spherical deconvolution model (Tournier et al., 2004). This model assumes that each voxel can be viewed as a large set of separate compartments, that are all indistinguishable except for their orientation. In this scenario the signal in a voxel is generated by the sum of the contributions from these individual compartments, whose orientations are distributed according to an orientation distribution function (ODF)  $U: \mathbb{S}^2 \rightarrow [0, \infty)$ . By construction the signal can then be written as a spherical convolution with an axially symmetric convolution kernel  $\mathcal{H}$ :

$$S(\mathbf{q}) \approx \int_{\mathbb{S}^2} U(\hat{\mathbf{v}}) \mathcal{H}(\mathbf{q}, \hat{\mathbf{v}} \cdot \hat{\mathbf{q}}) d\sigma(\hat{\mathbf{v}}). \quad (12)$$

where the gradient magnitude  $q = \|\mathbf{q}\|$  is typically fixed, where  $\hat{\mathbf{q}} = \mathbf{q}/q$ , and with  $d\sigma$  the standard Lebesgue measure on the sphere. The kernel  $\mathcal{H}$  can be postulated, but is more typically estimated by combining data from different positions (Tournier et al., 2007; Tax et al., 2014) or computed from the local signal (Novikov et al., 2019).

We may approximate the function  $U$  through a weighted sum of spherical harmonics  $Y_l^m$ :

$$U(\hat{\mathbf{v}}) \approx \sum_{l=0,2,\dots,L} \sum_{m=-l}^l u_l^m Y_l^m(\hat{\mathbf{v}}), \quad (13)$$

where the order  $L$  determines the number of expansion coefficients used, and thus determines the accuracy of the approximation. If we then approximate the convolution kernel  $\mathcal{H}$  as  $\sum_{l=0,2,\dots,L} k_l^0(\mathbf{q}) Y_l^0(\hat{\mathbf{v}})$  (where  $\hat{\mathbf{q}}$  is fixed to be the north pole so that axial symmetry permits a description using a smaller number of coefficients) then Eq. (12) can be evaluated analytically and we obtain the simplified form

$$S(\mathbf{q}) \approx \sum_{l=0,2,\dots,L} \sum_{m=-l}^l \sqrt{\frac{4\pi}{2l+1}} k_l^0(\mathbf{q}) u_l^m Y_l^m(\hat{\mathbf{q}}). \quad (14)$$

In the case of spherical deconvolution we may assume that the Fourier transform of the kernel itself is non-negative, and observe that non-negativity of  $P$  is then guaranteed by non-negativity of the orientation distribution function  $U$  (Tournier et al., 2007). Non-negativity of the spherical function  $U$  can be enforced using sum of squares constraints by defining  $\mathbf{v} := v \hat{\mathbf{v}}$  ( $v \leq 1$ ) and considering

$$\mathcal{P}(\mathbf{v}) = \sum_{l=0,2,\dots,L} \sum_{m=-l}^l \sqrt{\frac{2l+1}{4\pi}} u_l^m R_l^m(\mathbf{v}), \quad (15)$$

where  $R_l^m(\mathbf{v}) := \sqrt{\frac{4\pi}{2l+1}} v^l Y_l^m(\hat{\mathbf{v}})$  are the solid harmonics (Abramowitz and Stegun, 1964). The function  $\mathcal{P}$  as defined in Eq. (15) is a polynomial in the Cartesian coordinates of  $\mathbf{v}$ , and because the restriction of this polynomial to the sphere ( $v = 1$ ) is equal to  $U$  (Eq. (13)), we find that a SOS condition on  $\mathcal{P}$  guarantees non-negativity of  $U$ . Likewise, non-negativity of  $U$  implies that  $\mathcal{P}$  is non-negative over its domain  $0 \leq v \leq 1$  by the maximum (minimum) principle for harmonic functions (Evans, 2010), and we thus find that  $U$  is non-negative if and only if  $\mathcal{P}$  is non-negative on the unit sphere. The restriction of the non-negativity constraint on  $\mathcal{P}$  to the sphere  $v \leq 1$  still allows a tractable semidefinite programming formulation, as described for example by Magnani et al. (2005) using results from Schmüdgen (1991).

### 2.3.4. Cumulant expansion

The final expression for  $S$  considered here is the cumulant expansion (Liu et al., 2004; Kiselev, 2010), which can be obtained from a Taylor series of  $\log S$ . To ensure that this series exists, we assume that  $S$  has a well-defined moment generating function

$$\mathcal{M}(\mathbf{q}) = \int_{\mathbb{R}^3} e^{\mathbf{q} \cdot \mathbf{r}} P(\mathbf{r}) d\mathbf{r}. \quad (16)$$

Together with the properties mentioned in Section 2.1, this implies that the domain of  $S$  can be extended to complex-valued arguments (Hörmander, 1990), so that we can write for example  $\mathcal{M}(\mathbf{q}) = S(i\mathbf{q})$ . The expansion coefficients  $C_{i_1 \dots i_k}$  form the fully symmetric positive-definite  $k$ -tensors  $\mathbf{C}_k := \mathbf{C}_k(\Delta)$  (the cumulant tensors), and the logarithm of the signal is then given by

$$\log S(\mathbf{q}) \approx \sum_{k=2,4,\dots,K} \frac{(-1)^{\frac{k}{2}}}{k!} \sum_{i_1, \dots, i_k=1}^3 C_{i_1 \dots i_k} q_{i_1} \dots q_{i_k}, \quad (17)$$

where the diffusion time  $\Delta$  is fixed and with the order  $K$  a positive even integer. The cases  $K = 2$  and  $K = 4$  are functionally identical to diffusion tensor imaging (Basser et al., 1994a, b) and diffusional kurtosis imaging (Jensen et al., 2005), where for  $K = 2$  Eq. (1) can be evaluated analytically to obtain

$$P(\mathbf{r}) = \frac{(2\pi)^{-\frac{3}{2}}}{\sqrt{|\det \mathbf{C}_2|}} \exp \left[ -\frac{1}{2} \mathbf{r}^T \cdot \mathbf{C}_2^{-1} \cdot \mathbf{r} \right]. \quad (18)$$

The cumulant expansion converges in a bounded set containing the origin (Kiselev, 2010), so this model is typically considered for data sets acquired with fairly weak gradients ( $b \lesssim 2.5 \text{ms}/\mu\text{m}^2$ ). In addition to having a finite convergence radius, it holds that truncated cumulant expansions with non-zero higher order ( $k > 2$ ) coefficients *never* correspond to valid propagators. Combined, this means that the cumulant expansion cannot describe general  $P$  accurately in their entirety, and as a result it is impossible to impose our relaxed non-negativity constraint as is.

In lieu of enforcing non-negativity on  $P$  directly we therefore proceed with defining a set of necessary but not sufficient ancillary constraints, which must be satisfied in the case of non-negative  $P$  but will not in



general *guarantee* non-negativity. These constraints are based on the convexity property of the cumulant generating function  $\mathcal{C}(\mathbf{q}) := \log S(\mathbf{i}\mathbf{q})$  that follows from non-negativity of  $P$ , and which asserts that  $\mathcal{C}(\alpha \mathbf{q}_1 + (1-\alpha) \mathbf{q}_2) \leq \alpha \mathcal{C}(\mathbf{q}_1) + (1-\alpha) \mathcal{C}(\mathbf{q}_2)$  for all  $\mathbf{q}_1, \mathbf{q}_2 \in \mathbb{R}^3$  and all  $0 \leq \alpha \leq 1$ . Convexity can be verified as follows starting from Eq. (1) and using Hölder's inequality, which applies only in the case of non-negative  $P$ :

$$\begin{aligned} \mathcal{C}(\alpha \mathbf{q}_1 + (1-\alpha) \mathbf{q}_2) &= \\ \log S(\mathbf{i}(\alpha \mathbf{q}_1 + (1-\alpha) \mathbf{q}_2)) &= \\ \log \left[ \int_{\mathbb{R}^3} (e^{\mathbf{q}_1 \cdot \mathbf{r}})^\alpha (e^{\mathbf{q}_2 \cdot \mathbf{r}})^{1-\alpha} P(\mathbf{r}) d\mathbf{r} \right]^{\text{Hölder}} &\leq \\ \log \left[ \left( \int_{\mathbb{R}^3} e^{\mathbf{q}_1 \cdot \mathbf{r}} P(\mathbf{r}) d\mathbf{r} \right)^\alpha \left( \int_{\mathbb{R}^3} e^{\mathbf{q}_2 \cdot \mathbf{r}} P(\mathbf{r}) d\mathbf{r} \right)^{1-\alpha} \right] &= \\ \alpha \mathcal{C}(\mathbf{q}_1) + (1-\alpha) \mathcal{C}(\mathbf{q}_2). \end{aligned} \quad (19)$$

It then follows that every reconstructed model/expansion that corresponds to a non-negative displacement probability density function, must also correspond to a convex cumulant generating function. The presented argument can be directly extended to include convexity of the moment generating function  $\mathcal{M}(\mathbf{q}) = S(\mathbf{i}\mathbf{q})$ .

Unlike the non-negativity constraint, convexity of  $\mathcal{C}$  or  $\mathcal{M}$  is naturally expressed in the signal domain, which obviates the need for a practical expression of the propagator. Noting that convexity of a (continuous, twice differentiable) function  $\mathcal{C}$  on  $\mathbb{R}^3$  is equivalent to the requirement that

$$h(\mathbf{x}, \mathbf{s}) := \mathbf{s}^T \cdot \mathbf{H}_{\mathcal{C}}(\mathbf{x}) \cdot \mathbf{s} \geq 0 \quad (20)$$

for all  $\mathbf{x}, \mathbf{s} \in \mathbb{R}^3$ , where  $\mathbf{H}_{\mathcal{C}}$  is the Hessian matrix containing the second order partial derivatives of  $\mathcal{C}$ ,  $h$  is again a polynomial of degree  $2d$  and  $\mathbf{s}$  is a dummy variable of the same dimensions as  $\mathbf{x}$ , we find that a relaxed convexity constraint (Magnani et al., 2005) can be imposed as in (4) by defining a polynomial  $\mathcal{C} \approx \mathcal{M}$  (moment generating function) or  $\mathcal{C} \approx \mathcal{C}$  (cumulant generating function) and setting  $\mathcal{P}_{m_\omega} = h$ . The relaxed convexity constraint can thus be imposed on the cumulant expansion by applying sum of squares constraints on  $h$  as given in Eq. (20), with

$$\mathcal{C}(\mathbf{q}) = \mathcal{C}(\mathbf{q}) = \sum_{k=2,4,\dots,K} \frac{1}{k!} \sum_{i_1, \dots, i_k=1}^3 C_{i_1 \dots i_k} q_{i_1} \dots q_{i_k}. \quad (21)$$

Note finally that for the cases  $K=2$  and  $K=4$  the resulting relaxed constraints are in fact exact and coincide with the usual constraints, e.g. positive-definiteness of the diffusion tensor in the case of DTI.

## 2.4. Implementation

In our experiments we always take  $m_\omega$  in Eqs. (2) and (4) to be linear in  $\omega$  so that we can use linear least squares. In the case of the cumulant expansion this requires linearizing both the model and the measurements, as well as the introduction of weights in Eq. (2) to stabilize the measurement variances (Salvador et al., 2005). An initial fit is computed using (iteratively reweighted) linear least squares, and constrained optimization is only performed if the initial fit does not already satisfy the constraints, cf. Section 2.2.

The unconstrained optimum  $\omega^*$ —i.e., the vector of parameters that solves Eq. (2)—is used to reduce the complexity of the corresponding constrained optimization problem as described in Appendix A. We further simplify the constrained optimization problems by merging the constraints in Eq. (4) as follows. For a given polynomial  $\mathcal{P}_{m_\omega}$  (Section 2.3) and a given basis  $\mathbf{e}$ , we parameterize the space of potential  $\mathbf{A}$  in Eq. (4) as

$$\mathbf{A}(\omega, \alpha) = \mathbf{H}(\omega) + \mathbf{L}(\alpha), \quad (22)$$

where  $\mathbf{H}$  is any symmetric matrix satisfying  $\mathbf{e}^T \cdot \mathbf{H} \cdot \mathbf{e} = \mathcal{P}_m$ , and  $\mathbf{L}(\alpha)$  is any parameterization of the linear space  $\mathbf{e}^T \cdot \mathbf{L} \cdot \mathbf{e} = 0$  with parameters  $\alpha$ .

The matrix  $\mathbf{L}(\alpha)$  is typically very sparse, with its sparsity pattern depending on the choice for the basis  $\mathbf{e}$ . Although it is possible to use for example a Hermite polynomial basis for  $\mathbf{e}$ , we always use a monomial basis (Lofberg, 2009) in the experiments presented here.

The resulting semidefinite programs are solved using MOSEK 8.1.0.72 (MOSEK ApS, 2018). We compare the proposed semidefinite programming reconstruction of the MAP model, MAP+, to the weakly constrained reconstruction implemented in TORTOISE 3.1.3 (Irfanoglu et al., 2017; Pierpaoli et al., 2010), which enforces non-negativity of the propagator in a large set of individual points in the neighborhood of the origin as described in the original MAP paper (Özarslan et al., 2013b). The spherical deconvolution (CSD+) experiments are compared to the standard constrained spherical deconvolution (CSD) (Tournier et al., 2007) implementation provided by MRtrix 3, which pushes for non-negativity through soft constraints as described in the work of Tournier et al. (2007). The convolution kernel  $\mathcal{N}$  (the single fiber response function) is estimated with the tournier method implemented in MRtrix (Tournier et al., 2013; Tax et al., 2014). The sum of squares constrained reconstructions for CSD+ are computed using the same single fiber response functions as the weakly constrained SD implementation, but in the case of MAP+ we had to recompute the scaling tensor because this information was not provided directly by TORTOISE. The constrained cumulant expansion (CE+) reconstructions are only compared to unconstrained cumulant expansion fits.

## 2.5. Data

The MAP MRI experiments presented here make use of a subject in the MGH Human Connectome Project (HCP) adult diffusion database. This data set has 1.5 mm isotropic voxels, and  $b = \{1, 3, 5, 10\}$  ms/ $\mu\text{m}^2$  shells with  $\{64, 64, 128, 256\}$  measurements respectively, as well as 10  $b = 0$  ms/ $\mu\text{m}^2$  images. Further details can be found in the references (Setsonpop et al., 2013).

For the spherical deconvolution experiments we use images of a single subject in the WU-Minn HCP (Van Essen et al., 2013). The data has an isotropic voxel size of 1.25 mm, with 90 approximately uniformly distributed orientations acquired for each of the three shells,  $b = \{1, 2, 3\}$  ms/ $\mu\text{m}^2$  and 18 baseline images. We will only make use of the  $b = 3$  ms/ $\mu\text{m}^2$  shell.

For the cumulant expansion experiments we rely on the MASSIVE data set made available by Froeling et al. (2017), which consists of a large number of single-subject measurements distributed over five shells and two Cartesian grids. We generate randomly downsampled data sets for both the Cartesian and the spherically sampled subsets. For a given order  $K$ , random samples are generated from the shells with  $b = 1, 2, \dots, \frac{K}{2}$  ms/ $\mu\text{m}^2$ , and in the Cartesian case we pick random samples such that  $b < \frac{K}{2}$  ms/ $\mu\text{m}^2$  always. The number of samples is a multiple (the ‘sampling rate’) of the number of degrees of freedom  $\sum_{i=0}^{\frac{K}{2}} (i+1)(2i+1)$ , excluding four baseline images.

To get an idea of the importance of an optimized preprocessing pipeline, we use artificial data sets created by adding Gaussian noise to the function values estimated from maximum  $K$  MAP+, CSD+, and CE+ fits. For the CE+ experiments we use a shell-based scheme with a sampling rate of 10, and for the others we simulate data using the same acquisition schemes as described above. The resulting data sets have ideal noise characteristics, so can be viewed as the output of a perfect bias correction and artifact removal algorithm. Furthermore, these data sets can act as ground truth data sets for validation purposes, under the assumption that the considered model and its reconstruction are approximately valid for the measured data. We set the standard deviation of the artificial Gaussian noise based on a robust estimate of the standard deviation in the residuals for each diffusion-weighted image (1.4826 times the median absolute deviation, cf. Rousseeuw and Croux, 1993)).

### 3. Results

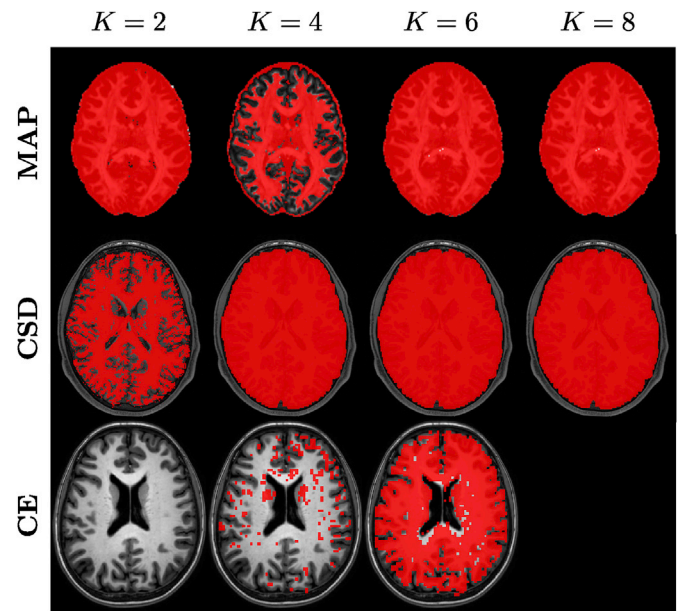
#### 3.1. The prevalence of constraint violations

Sum of squares optimization can be used to enforce non-negativity constraints on the propagator for a number of common models (Section 2.3), but comes at significant computational costs (Ahmadi and Majumdar, 2019). In this first set of experiments we investigate the necessity of these constraints in the generic case, i.e., in the case where we have no application-specific arguments to favor speed over reconstruction accuracy. The general methodology that we follow is similar to the work of Veraart et al. (2013): we use a high quality data set suitable for the specific model under consideration, perform an unconstrained or weakly constrained reconstruction, and then determine in what percentage of voxels in this data set we can prove that all constraints are satisfied.

For MAP MRI and spherical deconvolution we compare three implementations: an unconstrained least squares fit of the parameters, the commonly used weakly constrained quadratic programming reconstructions (Özarslan et al., 2013b; Tournier et al., 2007) that enforce non-negativity on a finite set of points, and the here proposed semi-definite programming reconstruction. Fits produced by the newly proposed method always satisfy the constraints. For the cumulant expansion we compare only fully constrained and fully unconstrained reconstructions for different orders. An overview of the percentages of voxels with violated constraints is presented in Table 1, and Fig. 2 shows some examples of the spatial distribution of voxels with unsatisfied constraints for the different models. These results show quite clearly that violated constraints are commonplace in all models and orders except for the  $K = 2$  cumulant expansion (DTI). The commonly used  $K = 8$  MAP and CSD reconstructions produce violated constraints in essentially all voxels.

#### 3.2. Impact on applications

Although the results of the previous subsection are quite alarming, a noteworthy caveat in their interpretation is that they do not reflect the severity of the violated constraints. It might be that the non-negative parts of the propagator have a very small magnitude, or that violations occur in regions that are not necessarily of interest for a specific



**Fig. 2.** The spatial distribution of voxels where constraints are not satisfied for different orders  $K$ , for a standard MAP reconstruction (top row), a standard CSD reconstruction (middle row), and an unconstrained CE reconstruction (bottom row). The cumulant expansion reconstruction was computed for a subsampled version (sampling rate 10) of the shell data in the MASSIVE data set. Voxels are colored red if they lie inside the brain mask and if the constraints are not satisfied in the stated reconstruction, and in the background we show fractional anisotropy (MAP) or T1-weighted images. Our results generally follow the expected trend where the number of voxels with violated constraints generally increases with  $K$  for each model, although MAP at  $K = 4$  has a surprisingly low number of violated constraints in the more isotropic gray matter. Overall, voxels with violated constraints appear to be concentrated in more anisotropic regions and near tissue interfaces.

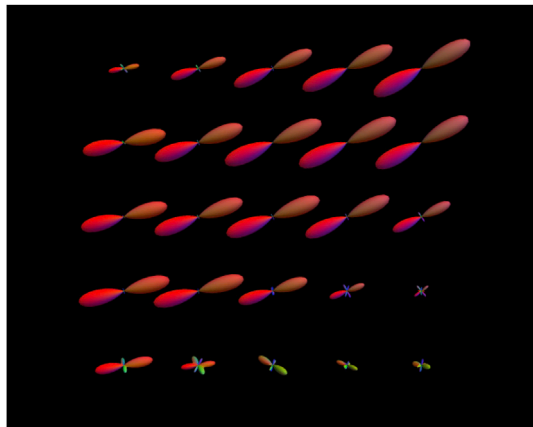
application. In the case of spherical deconvolution one might for example be interested only in the directions of the ODF maxima, and based on e.g. a visual comparison of ODFs (Fig. 3) one could be tempted to argue that these directions could be obtained reliably under weaker constraints. If it

**Table 1**

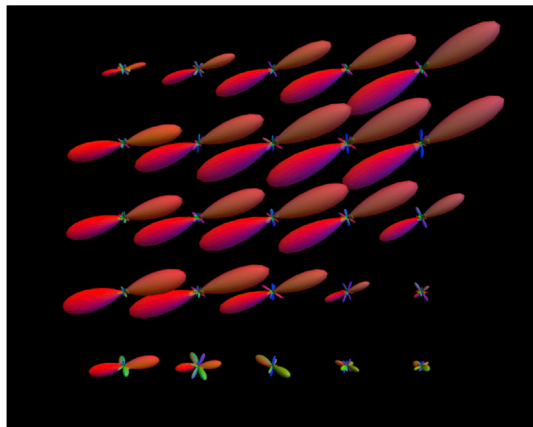
The percentage of voxels in the brain where the constraints derived in Section 2 are satisfied for different reconstruction methods. Percentages are computed relative to the total number of voxels inside the mask (439 248 for MGH HCP, 936 256 for WU-Minn HCP, 81 423 for MASSIVE). All cumulant expansion computations were performed on a shell-based acquisition (of the appropriate  $b$ -values) in the subsampled MASSIVE data (sampling rate 10), corresponding approximately to a high quality clinical acquisition. The standard MAP and CSD results were obtained by verifying the constraints in reconstructions generated by TORTOISE and MRtrix as described in Section 2.4, and the unconstrained reconstructions for all models were computed with custom code. (The unconstrained and standard MAP are based on different scaling tensors, while all SD experiments are performed with the same response function. TORTOISE failed in a number of voxels; these are not counted as violations in this table.) The '+' postfix is used to indicate a reconstruction that imposes sum of squares constraints. This method by definition does not generate results with violated constraints. We have highlighted a number of commonly used configurations: MAP order 8, CSD order 8, and CE orders 2 (DTI) and 4 (DKI). (We limit  $K$  to values that we consider to currently be of interest in the literature, but in principle the proposed techniques can be applied to models of any order.)

Method	Data	$K = 2$	$K = 4$	$K = 6$	$K = 8$
Unconstrained MAP	MGH HCP	1.9%	55.4%	0.0%	0.2%
Standard MAP		0.7%	42.9%	0.2%	0.2%
MAP+		100.0%	100.0%	100.0%	100.0%
Unconstrained SD	WU-Minn HCP	42.8%	0.0%	0.0%	0.0%
Standard CSD		43.1%	0.0%	0.0%	0.0%
CSD+		100.0%	100.0%	100.0%	100.0%
Unconstrained CE	MASSIVE	100.0%	83.9%	21.1%	
CE+		100.0%	100.0%	100.0%	

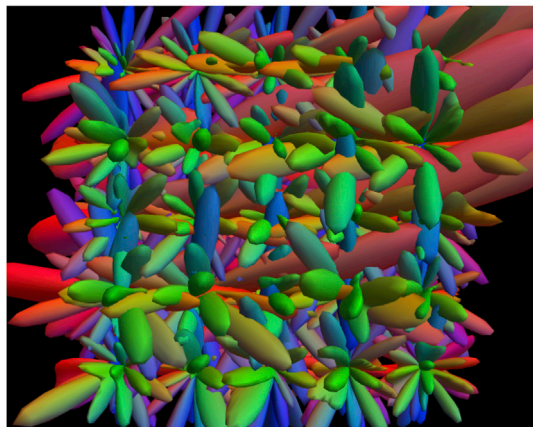
is a priori known that (mild) violations of the constraints do not cause problems in a specific pipeline, then it could be acceptable (but not recommended!) to use unconstrained optimization. To illustrate how the



(a) CSD+ (strict constraints)



(b) CSD (soft constraints)



(c) SD (unconstrained)

**Fig. 3.** A visualization of ODFs in a small region in the corpus callosum, obtained with  $K = 8$  spherical deconvolution under different constraints. (a) In the proposed method, the small negative lobes that can typically be seen near the center of each glyph are entirely absent, and thus the problem of spurious peaks is greatly diminished. (b) Other than the clearly visible spurious peaks, the differences between the two constrained reconstructions are small. The dominant directions of the ODFs in particular seem to be well-preserved under soft constraints. (c) The unconstrained reconstruction, included here only for reference, is entirely unusable.

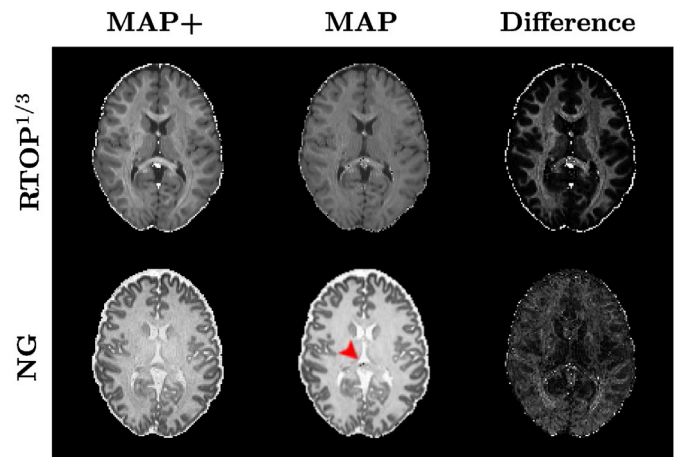
effects of violated constraints can propagate in a given analysis pipeline, we examine this propagation for two typical applications of our models.

For the cumulant expansion and for MAP MRI we consider the effect of violated constraints on various scalar measures, which are increasingly being used for example as summary statistics in population studies. For MAP MRI these measures are the return-to-origin probability (RTOP) and the non-Gaussianity measure (NG), and for the cumulant expansion we look at mean kurtosis (MK;  $K = 4$ ). Common scalar measures for the  $K = 2$  cumulant expansion (DTI), like fractional anisotropy or mean diffusivity, are not considered because the unconstrained reconstruction for  $K = 2$  is practically identical to the constrained reconstruction (recall Table 1). The precise definitions of these scalar indices can be found in the references (Özarslan et al., 2013b; Basser and Pierpaoli, 1996; Jensen et al., 2005; Jensen and Helpert, 2010).

The differences between strictly and weakly constrained reconstructions can already be appreciated visually in the case of MAP MRI, cf. Fig. 4. The impact of a constrained reconstruction varies between scalar indices, with larger deviations materializing in the RTOP measure than in the NG measure. The differences in RTOP are also much more clearly confined to a specific tissue, whereas the differences observed in NG are more homogeneous.

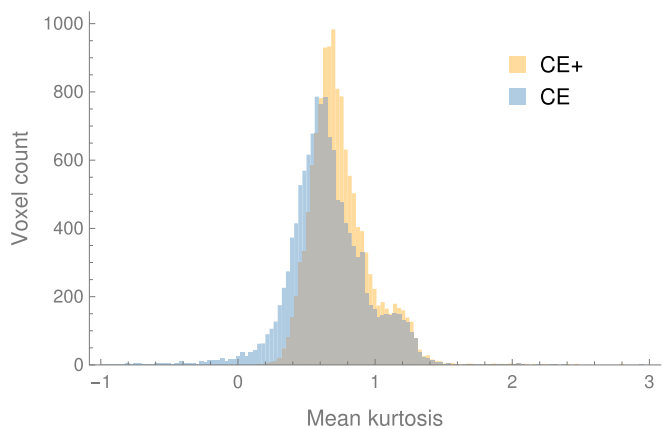
For the cumulant expansion the visual differences between constrained and unconstrained fits are less pronounced, simply because the fits are the same in most voxels. We therefore consider only the 13 106 voxels in our data set where constraints are violated, Fig. 5. Again we observe significant differences between the constrained and unconstrained fits, and, remarkably, we find that all MK values derived from the constrained fit lie in the expected biological range of  $[0, 3]$ . Note that this improvement over the unconstrained case is really a reflection of the improved accuracy associated with properly constrained fits; the algorithm does not actually enforce the MK to lie within this range.

As a second application we consider whether constrained reconstruction is necessary if one wishes to do tractography (Jbabdi and Johansen-Berg, 2011), which is commonly based on spherical



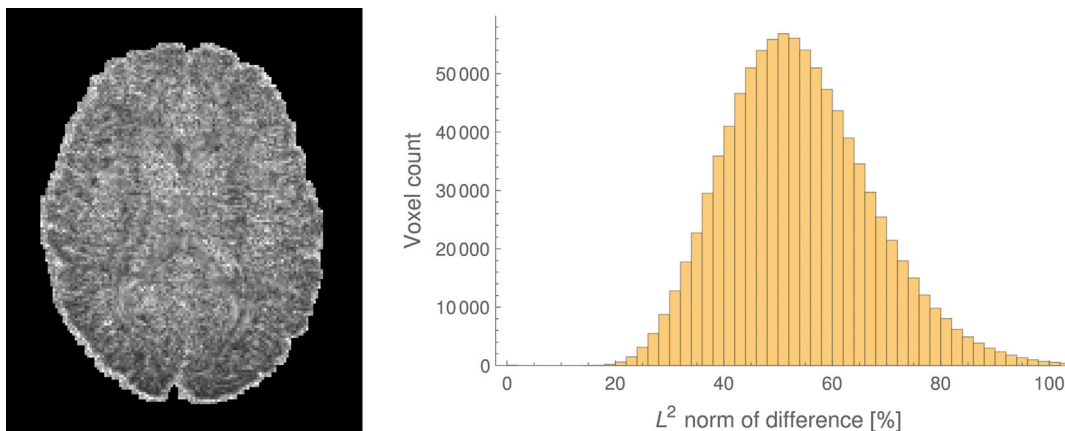
**Fig. 4.** A comparison between scalar measures computed from strictly constrained (left) and weakly constrained reconstructions (middle), together with their relative differences (right). The return-to-origin probability plots are scaled so that white corresponds to an  $RTOP^{1/3}$  value of  $265\text{mm}^{-1}$ , and in the non-Gaussianity plots white corresponds to an NG of 0.76. (The acquisition parameters are not tailored to examine regions with very large diffusivity, thus localized overfitting to Rician noise gives rise to abnormally high NG values in the cerebrospinal fluid.) The difference plots show the absolute differences between the figures in the left and middle column, relative to the strictly constrained values shown on the left (black is 0, white is 0.5). RTOP appears to be more sensitive to constraint violations than NG, with NG differences distributed rather evenly compared to the RTOP differences that are focused mainly in white matter. The red arrowhead points to a region where unphysical non-Gaussianity values in the unconstrained map result in black voxels.





**Fig. 5.** Histograms of the mean kurtosis values observed in voxels where the auxiliary constraints described in Section 2.3 are violated. The distribution of MK values derived from the unconstrained reconstruction appear to be biased towards lower values. The unconstrained fit produces a number of MK values outside of the biologically plausible range of [0, 3], which disappear entirely if the reconstruction is properly constrained.

deconvolution. For now, we are solely interested in seeing whether there will be any significant effect at all when applying more stringent constraints, so we consider first what happens with the orientation distribution function as a whole (relevant for e.g. probabilistic streamline (Friman et al., 2006; Lazar and Alexander, 2005; Berman et al., 2008; Sherbondy et al., 2008; Parker and Alexander, 2005; Hosey et al., 2005), geodesic (Schober et al., 2014; Fuster et al., 2016; O’Donnell et al., 2002; Pichon et al., 2005; Melonakos et al., 2008; Péchaud et al., 2009; Sepasian et al., 2014, 2012), and (to a lesser extent) global tractography approaches (Jbabdi et al., 2007; Christiaens et al., 2015; Reisert et al., 2011, 2014)), as well as what happens with the orientation of its peaks (relevant for e.g. deterministic streamline tractography (Conturo et al., 1999; Wedeen et al., 2008; Basser et al., 2000; Mori et al., 1999)). These results are shown in Figs. 6 and 7, and highlight the severe difference in function values and main orientation that result from the small changes in the fitted parameters (see Fig. 8 for a histogram comparing the sum of squared residuals between CSD+ and CSD). These large differences are very likely to have a significant effect on the outcome of common tractography algorithms, in particular because local deviations typically accumulate.



**Fig. 6.** A plot showing the  $L^2$  norm of the differences between the ODFs computed with weakly constrained spherical deconvolution and those computed with the proposed semidefinite programming reconstruction for the same slice as shown in Fig. 2, together with a histogram of the differences over the whole data set. The differences are shown relative to the  $L^2$  norm of the semidefinite programming reconstruction, with white pixels corresponding to a difference of 100%. Although there is some structure visible in the difference map, significant differences can be observed quite randomly throughout the brain. The large differences visualized here can be explained to some extent by overall size differences between the strict and weak constrained reconstructions, recall e.g. Fig. 3.

### 3.3. Sampling and preprocessing

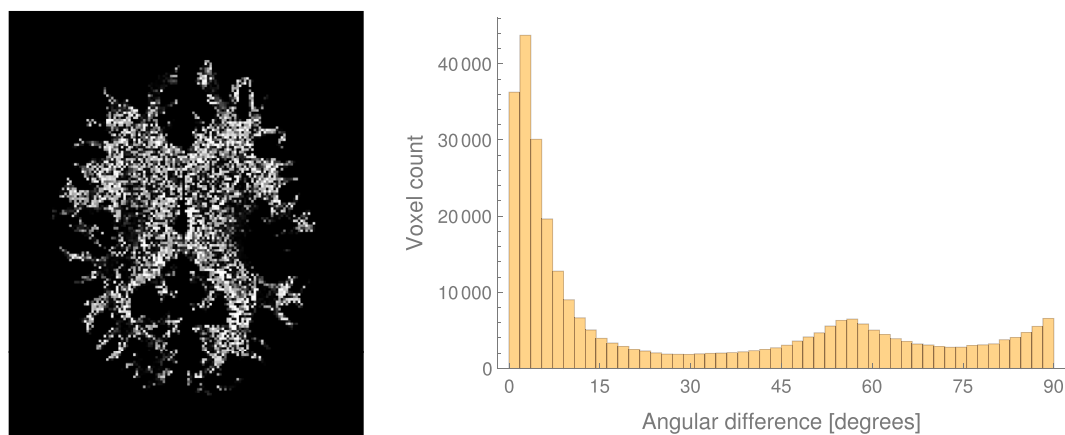
In the previous subsections we have used minimally processed data, as is typically used in standard analysis pipeline. In this section we build on these results to evaluate the importance of improved preprocessing (bias correction, Gibbs ringing correction, etc.) and the effect of different acquisition design choices (sampling rate and shell-versus grid-based acquisitions). For the former we rely on the artificial data described in Section 2.5, and again look at the percentage of voxels where the unconstrained reconstruction results in constraint violations. To do the latter we perform an unconstrained cumulant expansion reconstruction for different subsampled versions of the MASSIVE data set, and then check in what percentage of the voxels in the brain mask the constraints are not satisfied.

The percentages of violated constraints in the artificial data are shown in Table 2. Except for some unexpected differences in the MAP percentages obtained with TORTOISE, these results do not change much compared to the results on in vivo data (Table 1). Overall we find that reconstructions using artificial data are slightly worse than the reconstructions based on the original data, which is likely due to a slight overestimation of the noise standard deviation. As we have access to the ground truth parameters in these artificial data sets, it is possible to demonstrate how much fitting accuracy is actually improved by the proposed constraints, see e.g. Fig. 9. Likewise, we get closer to the true values of quantities derived from the estimated reconstructions, such as the orientations of the dominant peaks from CSD (Fig. 10).

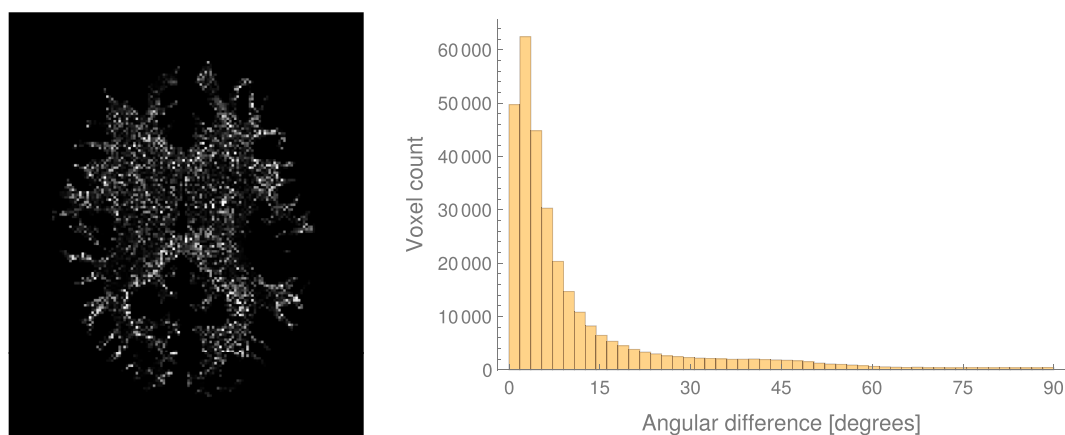
Fig. 11 shows graphs of the number of constraint violations observed in CE reconstructions for different numbers of measurements, where data points are computed as the mean numbers of voxels with unsatisfied constraints in 10 randomly generated data sets, recall Section 2.5. As expected, we observe a decreasing trend when increasing the sampling rate, and it turns out that this decrease is significantly more rapid in the case of spherical sampling. Voxels where the constraints are not satisfied appear to be concentrated near interfaces between tissue types (white matter/gray matter/cerebrospinal fluid) in all cases (Fig. 12).

### 3.4. Computational cost

The sum of squares formulation has a lot of advantages, the foremost being guaranteed non-negativity (Lofberg, 2009) and a wide applicability. Its main *disadvantage* however is that it scales quite badly with model complexity, see e.g. discussions in the work by Ahmadi and Majumdar (2019). This can be seen clearly in Table 3, which contains the



(a) Dominant peak difference



(b) Neighboring peak difference

**Fig. 7.** A comparison between the peaks estimated from standard CSD and from CSD+ in white matter. (a) A plot and histogram of the angular differences between the dominant peak orientations. Because dominant peaks are defined as the orientations where the ODF is maximal, secondary peaks occasionally present as dominant due to noise. This results in the relatively high number of voxels with angular differences around  $90^\circ$  for example. (b) The difference with the neighboring peak found through a local maximum search in the weakly constrained reconstruction, initialized at the location of the dominant peak of the strictly constrained reconstruction. The typical angular difference between the two types of constrained reconstructions is only a few degrees, but there are *many* cases where the angular difference is much larger—differences greater than  $5^\circ$  can be observed in about half of the voxels inside the mask. We can reasonably expect that less prominent (secondary) peaks will suffer from larger deviations than those observed here. Because local errors accumulate in most common tractography methods, a lack of strict constraints will surely have a significant impact in this application.

wall clock running times for the reconstructions used in the previous subsections. Note that the number of parameters, the number of data points, the number of iterations needed to stabilize the measurement variance, the exact formulation of the constraints, and the number of voxels that require constrained optimization all have an impact on the timings. Although timings increase quickly with model order, we can conclude based on these timings that current techniques are sufficient for state-of-the-art acquisitions, even using our current implementation prototype which relies on suboptimal MOSEK calls generated through an interface with Wolfram Mathematica (Wolfram, 2003).

## 4. Discussion

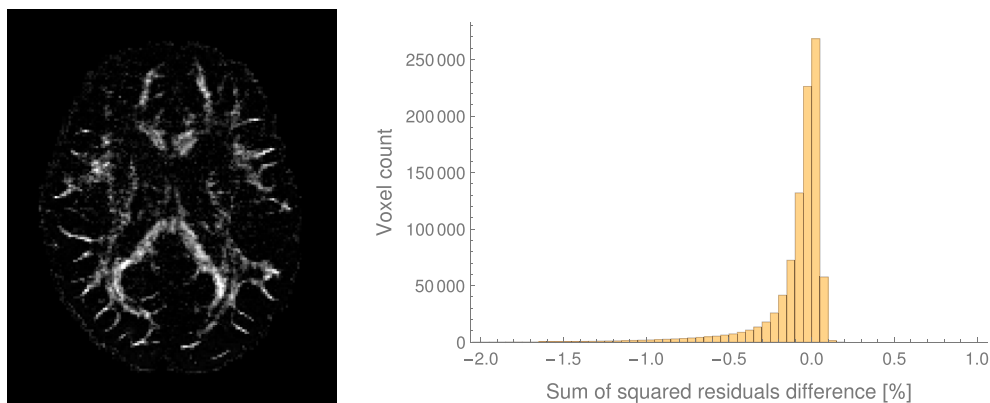
### 4.1. The necessity of strict constraints

Table 1 clearly highlights how crucial constrained reconstruction is under typical experimental conditions—only for the Gaussian diffusion model ( $K = 2$  cumulant expansion, equivalent to DTI) do we obtain an acceptable percentage of voxels with satisfied constraints. For higher order cumulant expansion reconstructions, as well as for MAP MRI and

spherical deconvolution, violated constraints are extremely prevalent. In these cases, regions of high anisotropy in the white matter as well as tissue interfaces seem particularly susceptible (Fig. 2). As the brain white matter is the main object of study in for example tractography, these results strongly imply that non-negativity ought to be enforced through strict constraints to ensure that subsequent data analyses are reliable.

We also found that the number of constraint violations in real data remains largely the same if we consider artificial data with ideal noise characteristics (Table 2). From this we conclude that the observed violations are not likely to be caused to any significant extent by errors in our implicit modeling assumptions (Gaussian noise, no artifacts, etc.), but are instead effected predominantly by measurement noise. However, this is partly a reflection of the quality of the public data that we have used in our experiments, and we can assume that in noisy or corrupted data further constraint violations will likely be present.

In the case of appropriately processed, high quality data, it follows that the only effective strategies to decrease the number of constraint violations must necessarily reduce the effective power of measurement noise, and we can see in Fig. 11 that optimally distributed samples and an increased number of measurements are both advantageous in this regard.

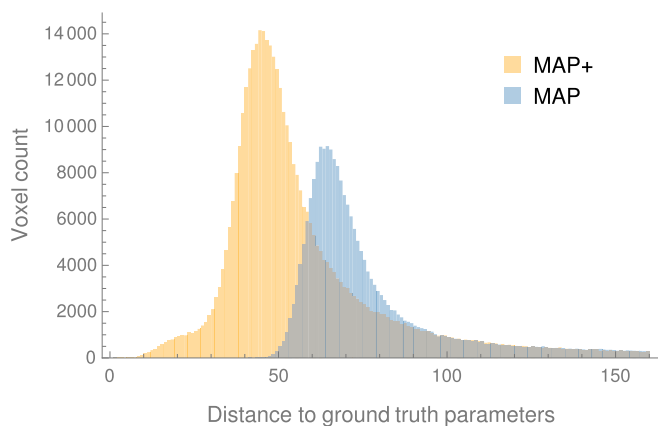


**Fig. 8.** A plot and a histogram of the differences between the sum of squared residuals for CSD+ and standard CSD, shown as percentages relative to the sum of squared residuals of CSD+. Negative values correspond to residuals where CSD+ outperforms CSD, and this is the case in the vast majority of the voxels. This trend was actually unexpected to us, since the SOS constraints are more strict than the constraints employed in CSD, but can be reasonably explained by implementation-specific issues unrelated to the theoretical performance of the methods. Unconstrained reconstructions have slightly lower residuals still compared to the SOS constrained fits (typically less than 1% relative difference). Based on these observations, we can be confident that our fitting quality is acceptable. In the plot on the left, white corresponds to an absolute difference of 2%, and black corresponds to 0%, and the slice is again the same as in Fig. 2. In the case of CSD shown here we clearly find that the largest deviations occur in the highly anisotropic white matter regions.

**Table 2**

The percentage of voxels in the brain where constraints are satisfied for different reconstruction methods under ideal noise conditions. The artificial data used to compute these results is based on the maximum  $K$  MAP+, CSD+, and CE+ reconstructions obtained in Table 1, and generated by adding Gaussian noise to the function values computed for those constrained reconstructions using the gradient schemes described in Section 2.5. The standard deviation of the artificial noise is estimated from the residuals in the in vivo data. We have again highlighted commonly used configurations in the table: MAP order 8, CSD order 8, and CE orders 2 (DTI) and 4 (DKI).

Method	Data	$K = 2$	$K = 4$	$K = 6$	$K = 8$
Unconstrained MAP	Artificial MGH HCP	1.3%	51.3%	0.0%	0.2%
Standard MAP		17.1%	0.1%	0.1%	0.1%
Unconstrained SD	Artificial WU-Minn HCP	46.8%	0.3%	0.0%	0.0%
Standard CSD		47.2%	0.4%	0.0%	0.0%
Unconstrained CE	Artificial MASSIVE	99.9%	73.7%	13.2%	



**Fig. 9.** A histogram of the Euclidean norm of the difference between the ground truth parameter vectors and the parameter vectors obtained through strictly constrained and unconstrained  $K = 8$  MAP MRI reconstructions, computed for an artificial data set. This simple plot clearly illustrates the improved accuracy achieved by incorporating prior knowledge through constraints.

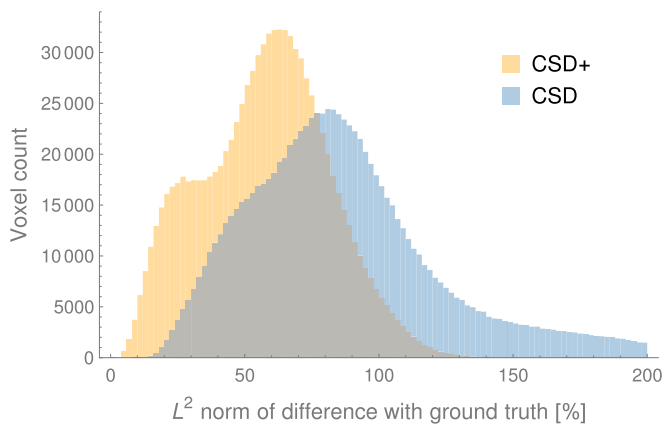
However, if we extrapolate the plotted trends, it becomes clear that the number of measurements required to obviate the need of a constrained reconstruction quickly becomes untenable. For a  $K = 4$  cumulant

expansion (DKI) we would need hundreds of measurements—not unreasonable in a modern preclinical setting although infeasible in a clinical one—but for  $K = 6$  this number already runs in the thousands. Considering Table 1 we can expect other non-Gaussian models to require similarly large numbers of measurements, and it therefore seems likely that constrained reconstruction will be required in many cases even taking into account possibly significant future gains in acquisition speed and quality.

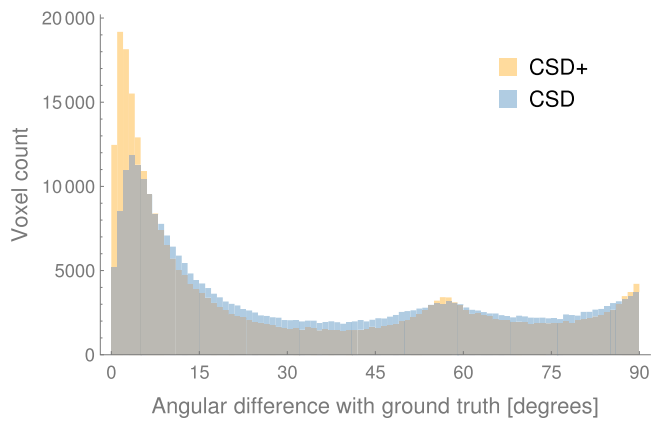
#### 4.2. Consequences for applications

While constraint violations should typically be avoided, their occurrence is not necessarily fatal in practical applications. We therefore considered a number of example scenarios, and the proposed sum of squares constrained reconstruction method proved to offer considerable practical benefits in all cases. Parameters obtained through constrained optimization differ significantly from the standard weakly constrained and unconstrained reconstructions, with the estimation accuracy improving noticeably in the constrained case (Figs. 5, 9, and 10). Correspondingly, the constrained fits produce more accurate estimates of derived scalar measures, orientation distribution functions, and dominant fiber orientations.

In the case of spherical deconvolution the favorable effects resulting from this accuracy gain can be appreciated immediately from the reduced incidence of spurious peaks (Fig. 3, also reported by e.g. Cheng



(a) Function value difference



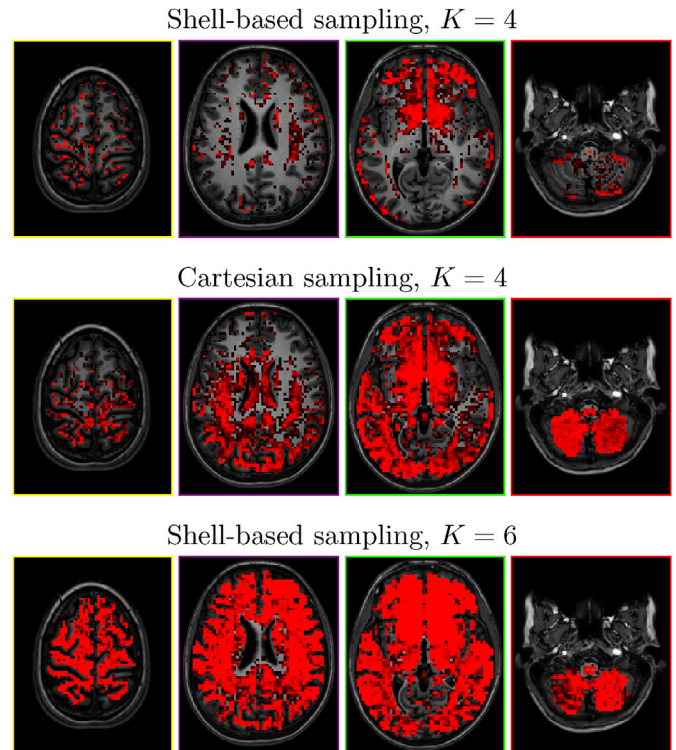
(b) Dominant peak difference

**Fig. 10.** (a) A histogram of the  $L^2$  norm of the differences between constrained reconstructions and ground truth functions based on artificial data, computed for CSD+ and CSD reconstructions and shown relative to the  $L^2$  norm of the ground truth functions. (b) A histogram of the angular differences between the reconstructed dominant peak and the ground truth in artificial data, computed for CSD+ and CSD reconstructions in white matter. In both cases we see clearly that the strict constraints imposed in CSD + contribute to more accurate estimates of function values and peak orientations.

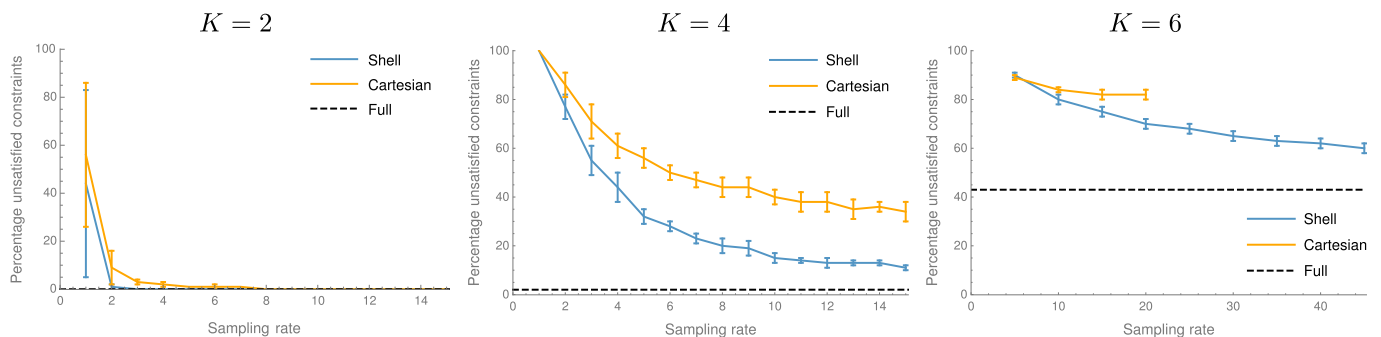
et al. (2014) for other strict constraints). As spurious peaks and other inaccuracies in local orientation estimates are detrimental to the performance of for example tractography and connectivity algorithms, we

expect that enforcing these constraints will have a significant effect on such applications.

For MAP MRI and for the cumulant expansion, our constraints turn out to provide a more natural solution than explicit biologically motivated constraints (e.g. the assumed absence of negative mean kurtosis, cf. Veraart et al., 2011) to cases of problematic ‘black voxels’ that are often seen in scalar maps, see for example Fig. 4. Furthermore, we found that although these scalar values clearly still deviate from their correct values when computed under weaker constraints, they are often not obviously



**Fig. 12.** The spatial distribution of voxels where constraints are not satisfied when using an unconstrained reconstruction of the cumulant expansion. The figures are generated using subsampled data sets with a sampling rate of 15 for  $K = 4$  and 45 for  $K = 6$ , with the foreground color indicating the number of data sets where the reconstruction failed (varying from black to red, with red indicating that none of the reconstructions satisfied the non-negativity constraints). The orders and data sets used in generating each figure are indicated in the titles. A T1-weighted image is shown in the background for anatomical reference.



**Fig. 11.** These plots show the percentage of voxels where the reconstructed model parameters (cumulants up to order  $K$ ) do not satisfy the constraints derived in Section 2.2. The parameters used here are computed directly from the subsampled MASSIVE data set described in Section 2.5, with no additional preprocessing performed. This percentage decreases if more samples are included, with space-filling (Cartesian) sampling schemes consistently producing more ‘incorrect’ voxels than spherically sampled data sets. Sampling rates of 15/10/6 respectively correspond approximately to WU-Minn HCP quality data. The error bars show the estimated standard deviations.



**Table 3**

Approximate wall clock running times for constrained reconstructions based on different models. All computations were done with an implementation prototype on an Intel Core i7-7600U CPU @ 2.80 GHz  $\times$  4 laptop. The cumulant expansion computations were performed on a shell-based acquisition in the subsampled MASSIVE data as before (sampling rate 10 for all  $K$ ), but it should be noted that because of the result obtained in Appendix A the effect of the number of data points to these timings is relatively minor—it is instead the increasing complexity of the constraints that leads to very large semidefinite programs. The time spent on computing the response function in the case of CSD+ and of the scaling matrix in the case of MAP+ is not included in the reported values, but the unconstrained reconstructions that are used to initialize all ‘+’ reconstructions and the checks for constraint violations are counted. The total number of voxels inside the mask for each data set is as follows: 439 248 for MGH HCP, 936 256 for WU-Minn HCP, 81 423 for MASSIVE. We include the timings of our custom unconstrained implementation for comparison.

Method	Data	$K = 2$	$K = 4$	$K = 6$	$K = 8$
MAP	MGH HCP	1 m 48 s	4 m 52 s	13 m 32 s	1 h 5 m 33 s
		0.25 ms/voxel	0.67 ms/voxel	1.85 ms/voxel	8.95 ms/voxel
MAP+	MGH HCP	13 m 2 s	37 m 55 s	4 h 25 m 21 s	21 h 10 m 15 s
		1.78 ms/voxel	5.18 ms/voxel	36.25 ms/voxel	173.51 ms/voxel
		(7 parameters)	(22 parameters)	(50 parameters)	(95 parameters)
SD	WU-Minn HCP	5 m 25 s	9 m 42 s	20 m 56 s	41 m 6 s
		0.35 ms/voxel	0.62 ms/voxel	1.34 ms/voxel	2.63 ms/voxel
CSD+	WU-Minn HCP	26 m 7 s	1 h 35 m 54 s	6 h 20 m 31 s	27 h 59 m 47 s
		1.67 ms/voxel	6.15 ms/voxel	24.39 ms/voxel	107.65 ms/voxel
		(6 parameters)	(15 parameters)	(28 parameters)	(45 parameters)
CE	MASSIVE	20 s	3 m 24 s	19 m 35 s	
		0.25 ms/voxel	2.50 ms/voxel	14.43 ms/voxel	
CE+	MASSIVE	20 s	6 m 43 s	1 h 57 m 52 s	
		0.25 ms/voxel	4.95 ms/voxel	86.85 ms/voxel	
		(7 parameters)	(22 parameters)	(50 parameters)	

wrong—i.e., they typically do not lie outside of the range of physically plausible values (Fig. 5). It would thus be difficult to detect these deviations without a proper constrained reconstruction, which highlights the fact that simpler constraints such as the linear/quadratic programming constraints considered by e.g. Veraart et al. (2011) cannot be considered viable alternatives. Note that if there are reasons to enforce these constraints in a specific application, they could still be incorporated straightforwardly in the proposed framework as additional constraints.

#### 4.3. Future work

The constrained optimization approach we proposed here extends straightforwardly to a large number of other (linearizable) models, including generalized DTI (as previously considered in related works (Barmpoutis et al., 2009, 2007; Chen et al., 2013)) and multi-shell multi-tissue type spherical deconvolution approaches (Jeurissen et al., 2014), and we expect that the inclusion of the proposed constraints will offer similar advantages in those cases. Sum of squares constraints can also be used for other constraints than non-negativity, as we have seen in the CE case where we essentially enforce convexity, although identifying other useful constraints may prove challenging. Finally, it is at least theoretically possible to extend our approach to problems with non-linear objective functions. This includes sum of squares constrained non-linear least squares reconstructions, for example of many common multi-compartment models (Panagiotaki et al., 2012), as well as maximum likelihood estimation based on a Rician noise model. This may be especially valuable for models that rely on inherently noisier large gradient strength acquisitions, although the running time will likely increase significantly. There is on the other hand an ongoing effort to improve the scalability of sum of squares optimization techniques (Permenter and Parrilo, 2019; Lofberg and Parrilo, 2004; Ahmadi and Majumdar, 2019), and we generally expect that the benefits of constrained optimization will continue to outweigh the computational costs. For specific applications, we can also investigate whether existing methods capable of strictly enforcing non-negativity (Cheng et al., 2014; Qi et al., 2010; Chen et al., 2013; Schwab et al., 2012; Barmpoutis and Vemuri, 2010) can produce comparable constrained reconstructions at a

reduced computational cost. Square root representations (Cheng et al., 2014) and linear SOS approximations (Barmpoutis and Vemuri, 2010) for example, can generally be expected to be faster than the proposed approach, although care must be taken to ensure that these methods can accurately represent all possible data.

## 5. Conclusion

In this work we have proposed to use sum of squares optimization to enforce non-negativity of the displacement probability density function in diffusion MRI model reconstruction problems. This versatile approach was developed for three commonly used and quite different classes of models, and can likely be extended to several others. To show the practical benefit and necessity of our approach we investigated how prevalent constraint violations are in different scenarios, and how these violations impact different applications. Based on these results we can draw three conclusions. First, diffusion-weighted MRI parameter estimation schemes that do not use constraints or that only enforce weak constraints, will routinely produce physically impossible parameter values. A large segment of current approaches to (constrained) model reconstruction are thus insufficient in practice, and should only be trusted in very specific cases. Second, these unphysical fits are likely to have a significant effect in practice, as they occur in clinically relevant regions and can introduce large deviations from the more reliable constrained result. Finally, we note that these constraint violations appear for almost all tested models (the exception being the DTI model), and that improvements to the acquisition pipeline are unlikely to resolve this issue sufficiently. Weaker constraints, bias correction, artifact removal, additional measurements (i.e. noise reduction), and even localized lower model orders—none of these steps were found to offer adequate, practicable solutions. Considering this, it would be our recommendation that the weak constraints that are used in essentially all current implementations be replaced with the strict constraints proposed here, even taking into account that this will inevitably add some computational costs and complexity to those implementations. We note that computational costs will be relatively small for models where constraint violations are relatively infrequent, and that these costs are rendered insignificant by the absolute necessity

for proper constraints in other cases.

## Acknowledgements

Tom Dela Haije and Aasa Feragen were supported by the Center for Stochastic Geometry and Advanced Bioimaging and by a block stipendium, both funded by the Villum Foundation (Denmark). Evren Özarslan was supported by the Linköping University Center for Industrial Information Technology (CENIIT, Sweden) and by VINNOVA/ITEA3 17021 IMPACT (Sweden/Netherlands). The authors would also like to acknowledge Schloss Dagstuhl and the organizers of Dagstuhl seminars 16142 and 18442 for facilitating discussions that improved the quality of this paper.

Anton Mallasto and Line Kühnel are acknowledged for their help in formalizing the complexity reduction technique presented in [Appendix A](#). The authors thank Chantal Tax and Alard Roebroek for their suggestions regarding the alternative preprocessing pipeline and ground

truth experiments, and Andrea Fuster and Luc Florack are acknowledged for their contributions to preliminary works focused on the cumulant expansion. The authors further thank M. Okan Irfanoglu for his help with the TORTOISE software.

Data were provided in part by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University, and by the Human Connectome Project, MGH-USC Consortium (Principal Investigators: Bruce R. Rosen, Arthur W. Toga and Van Wedeen; U01MH093765) funded by the NIH Blueprint Initiative for Neuroscience Research grant; the National Institutes of Health grant P41EB015896; and the Instrumentation Grants S10RR023043, 1S10RR023401, 1S10RR019307. The MASSIVE database was designed and acquired by Martijn Froeling, Alexander Leemans, Chantal Tax, and Sjoerd Vos at the PROVIDI-lab (UMC Utrecht, the Netherlands).

## Appendix A. Reducing the complexity of constrained linear least squares algorithms

We can mostly remove the dependency of constrained linear least squares algorithms on the number of measurements as follows. Let us start by rewriting Eq. (4) in the standard form (with weights omitted for simplicity)

$$\arg \min_{x \in \Omega} \|M \cdot x - y\|^2, \quad (\text{A.1})$$

where  $M$  is the design matrix formed by the  $m_\omega(q_i)$  for a given linear model  $m_\omega$ ,  $x$  and  $y$  are the vectors of representing the (unknown) fitting parameters  $\omega$  and the given measurements  $y_i$  respectively, and with  $\Omega$  the abstract set of all  $x$  that satisfy the model-specific constraints. Let  $x^*$  be the solution to the unconstrained problem (i.e., the vector representation of  $\omega^*$ ), which can be expressed analytically in terms of the pseudo-inverse of  $M$ . We then note that the objective function can be effectively rewritten as

$$\|M \cdot x - y^* + y^* - y\|^2 = \|M \cdot x - y^*\|^2 + \|y^* - y\|^2 + 2(M \cdot x - y^*)^T \cdot (y^* - y), \quad (\text{A.2})$$

where we define  $y^*$  to be the orthogonal projection of  $y$  onto the image of  $M$  given by  $y^* = M \cdot x^*$ . With this definition of  $y^*$  the final term in Eq. (A.2) disappears, and since  $\|y^* - y\|$  is constant we find that Eq. (A.1) is equivalent to

$$\arg \min_{x \in \Omega} \|M \cdot x - M \cdot x^*\|^2. \quad (\text{A.3})$$

Although this reformulation as such does not reduce the complexity of the original problem at all, we can now rewrite the objective function according to

$$\begin{aligned} \|M \cdot (x - x^*)\|^2 &= (x - x^*)^T \cdot M^T \cdot M \cdot (x - x^*) \\ &= (x - x^*)^T \cdot V \cdot V^T \cdot (x - x^*) \\ &= \|V^T \cdot x - V^T \cdot x^*\|^2, \end{aligned} \quad (\text{A.4})$$

where  $V$  is the lower triangular matrix obtained by a Cholesky decomposition of the positive-definite matrix  $M^T \cdot M$ . We are thus free to replace  $M$  with  $V^T$  in Eq. (A.3), and this effectively reduces the number of (artificial) measurements  $V^T \cdot x^*$  used in the constrained reconstruction to the minimum, i.e., to the number of fitting parameters. The practical effect of this dimensionality reduction in the considered case of semidefinite programming is quite dramatic, with speed-up factors of 50 or more observed in the typical setting where the number of measurements is large relative to the number of parameters, while the associated overhead is negligible.

## References

- Abramowitz, M., Stegun, I.A., 1964. Handbook of Mathematical Functions: with Formulas, Graphs and Mathematical Tables, 9 edition. Dover, New York.
- Ahmadi, A.A., Majumdar, A., 2019. DSOS and SDSOS optimization: more tractable alternatives to sum of squares and semidefinite optimization. *SIAM J. Appl. Algebra Geometry* 3, 193–230.
- Assemlal, H.E., Tschumperlé, D., Brun, L., Siddiqi, K., 2011. Recent advances in diffusion MRI modeling: angular and radial reconstruction. *Med. Image Anal.* 15, 369–396.
- Barmpoutis, A., Vemuri, B.C., 2010. A unified framework for estimating diffusion tensors of any order with symmetric positive-definite constraints. In: 2010 IEEE International Symposium on Biomedical Imaging: from Nano to Macro. IEEE, Rotterdam, Netherlands, pp. 1385–1388.
- Barmpoutis, A., Jian, B., Vemuri, B.C., Shepherd, T.M., 2007. Symmetric positive 4th order tensors & their estimation from diffusion weighted MRI. In: Karssemeijer, N., Lelieveldt, B. (Eds.), *Information Processing in Medical Imaging*, vol. 4584. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 308–319.
- Barmpoutis, A., Hwang, M.S., Howland, D., Forder, J.R., Vemuri, B.C., 2009. Regularized positive-definite fourth order tensor field estimation from DW-MRI. *Neuroimage* 45, S153–S162.
- Barmpoutis, A., Ho, J., Vemuri, B.C., 2012. Approximating symmetric positive semidefinite tensors of even order. *SIAM J. Imaging Sci.* 5, 434–464.
- Basser, P.J., Pierpaoli, C., 1996. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *J. Magn. Reson. Ser. B* 111, 209–219.
- Basser, P.J., Mattiello, J., Le Bihan, D., 1994a. Estimation of the effective self-diffusion tensor from the NMR spin echo. *J. Magn. Reson., Ser. B* 103, 247–254.

- Basser, P.J., Mattiello, J., Le Bihan, D., 1994b. MR diffusion tensor spectroscopy and imaging. *Biophys. J.* 66, 259–267. PMID: 8130344.
- Basser, P.J., Pajevic, S., Pierpaoli, C., Duda, J., Aldroubi, A., 2000. In vivo fiber tractography using DT-MRI data. *Magn. Reson. Med.* 44, 625–632.
- Berg, C., Christensen, J.P.R., Ressel, P., 1976. Positive definite functions on Abelian semigroups. *Math. Ann.* 223, 253–274.
- Berman, J.L., Chung, S., Mukherjee, P., Hess, C.P., Han, E.T., Henry, R.G., 2008. Probabilistic streamline q-ball tractography using the residual bootstrap. *Neuroimage* 39, 215–222.
- Callaghan, P.T., 1991. *Principles of Nuclear Magnetic Resonance Microscopy*. Clarendon Press, Oxford.
- Callaghan, P.T., Eccles, C.D., Xia, Y., 1988. NMR microscopy of dynamic displacements: k-space and q-space imaging. *J. Phys. E Sci. Instrum.* 21, 820.
- Callaghan, P.T., MacGowan, D., Packer, K.J., Zelaya, F.O., 1990. High-resolution q-space imaging in porous structures. *J. Magn. Reson.* 90, 177–182.
- Chen, Y., Dai, Y., Han, D., Sun, W., 2013. Positive semidefinite generalized diffusion tensor imaging via quadratic semidefinite programming. *SIAM J. Imaging Sci.* 6, 1531–1552.
- Cheng, J., Ghosh, A., Jiang, T., Deriche, R., 2009. A Riemannian framework for orientation distribution function computing. In: Yang, G.Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009*, vol. 5761. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 911–918.
- Cheng, J., Ghosh, A., Jiang, T., Deriche, R., 2011. Diffeomorphism invariant Riemannian framework for ensemble average propagator computing. In: Fichtinger, G., Martel, A., Peters, T. (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011*, vol. 6892. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 98–106.
- Cheng, J., Jiang, T., Deriche, R., 2012. Nonnegative definite EAP and ODF estimation via a unified multi-shell HARDI reconstruction. In: Hutchison, D., Kanade, T., Kitzler, J., Kleinberg, J.M., Mattern, F., Mitchell, J.C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., Vardi, M.Y., Weikum, G., Ayache, N., Delingette, H., Golland, P., Mori, K. (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012*, vol. 7511. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 313–321.
- Cheng, J., Deriche, R., Jiang, T., Shen, D., Yap, P.T., 2014. Non-negative spherical deconvolution (NNSD) for estimation of fiber orientation distribution function in single-/multi-shell diffusion MRI. *Neuroimage* 101, 750–764.
- Christiaens, D., Reiser, M., Dhollander, T., Sunaert, S., Suetens, P., Maes, F., 2015. Global tractography of multi-shell diffusion-weighted imaging data using a multi-tissue model. *Neuroimage* 123, 89–101.
- Conturo, T.E., Lori, N.F., Cull, T.S., Akbudak, E., Snyder, A.Z., Shimony, J.S., McKinstry, R.C., Burton, H., Raichle, M.E., 1999. Tracking neuronal fiber pathways in the living human brain. *Proc. Natl. Acad. Sci.* 96, 10422–10427.
- Dela Haije, T.C.J., Feragen, A., 2018. The importance of constraints and spherical sampling in diffusion MRI. In: *Proceedings of the 26th Annual Meeting of the ISMRM*, Paris, p. 5335.
- Dela Haije, T.C.J., Fuster, A., Florack, L.M.J., 2015. Reconstruction of convex polynomial diffusion MRI models using semi-definite programming. In: *Proceedings of the 23rd Annual Meeting of the ISMRM*, Toronto, p. 2821.
- Dela Haije, T.C.J., Özarslan, E., Feragen, A., 2019. Non-negative mean apparent propagators using sum-of-squares optimization: MAP+. In: *Proceedings of the 27th Annual Meeting of the ISMRM*, Montréal, p. 6900.
- Evans, L.C., 2010. *Partial Differential Equations*. American Mathematical Society, Providence, RI.
- Fick, R.H., Wassermann, D., Caruyer, E., Deriche, R., 2016. MAPL: tissue microstructure estimation using Laplacian-regularized MAP-MRI and its application to HCP data. *Neuroimage* 134, 365–385.
- Fillard, P., Arsigny, V., Ayache, N., Pennec, X., 2005. A Riemannian framework for the processing of tensor-valued images. In: *Deep Structure, Singularities, and Computer Vision*. Springer, pp. 112–123.
- Friman, O., Farneback, G., Westin, C.F., 2006. A Bayesian approach for stochastic white matter tractography. *IEEE Trans. Med. Imaging* 25, 965–978.
- Froeling, M., Tax, C.M., Vos, S.B., Luijten, P.R., Leemans, A., 2017. “MASSIVE” brain dataset: multiple acquisitions for standardization of structural imaging validation and evaluation: MASSIVE Brain Dataset. *Magn. Reson. Med.* 77, 1797–1809.
- Fuster, A., Dela Haije, T., Tristán-Vega, A., Plantinga, B., Westin, C.F., Florack, L., 2016. Adjugate diffusion tensors for geodesic tractography in white matter. *J. Math. Imaging Vis.* 54, 1–14.
- Ghosh, A., Deriche, R., 2016. A survey of current trends in diffusion MRI for structural brain connectivity. *J. Neural Eng.* 13, 011001.
- Ghosh, A., Milne, T., Deriche, R., 2014. Constrained diffusion kurtosis imaging using ternary quartics & MLE. *Magn. Reson. Med.* 71, 1581–1591.
- Goh, A., Lenglet, C., Thompson, P.M., Vidal, R., 2009. A nonparametric Riemannian framework for processing high angular resolution diffusion images (HARDI). In: *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 2496–2503.
- Hanyga, A., Sereďnyška, M., 2012. Anisotropy in high-resolution diffusion-weighted MRI and anomalous diffusion. *J. Magn. Reson.* 220, 85–93.
- Hilbert, D., 1933. Über die Darstellung definiter Formen als Summe von Formenquadraten. In: *Algebra · Invariantentheorie · Geometrie*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 154–161.
- Hörmander, L., 1990. *Distribution Theory and Fourier Analysis*. Number 1 in the Analysis of Linear Partial Differential Operators, 2 edition. Springer, Berlin.
- Hosey, T., Williams, G., Ansorge, R., 2005. Inference of multiple fiber orientations in high angular resolution diffusion imaging. *Magn. Reson. Med.* 54, 1480–1489.
- Irfanoglu, M.O., Nayak, A., Jenkins, J., Pierpaoli, C., 2017. TORTOISE v3: improvements and new features of the NIH diffusion MRI processing pipeline. In: *Proceedings of the 25th Annual Meeting of the ISMRM*, Honolulu, HI, p. 3540.
- Jbabdi, S., Johansen-Berg, H., 2011. Tractography: where do we go from here? *Brain Connect.* 1, 169–183.
- Jbabdi, S., Woolrich, M.W., Andersson, J.L.R., Behrens, T.E.J., 2007. A Bayesian framework for global tractography. *Neuroimage* 37, 116–129.
- Jensen, J.H., Helpert, J.A., 2010. MRI quantification of non-Gaussian water diffusion by kurtosis analysis. *NMR Biomed.* 23, 698–710.
- Jensen, J.H., Helpert, J.A., 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.
- Jeurissen, B., Tournier, J.D., Dhollander, T., Connelly, A., Sijbers, J., 2014. Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data. *Neuroimage* 103, 411–426.
- Kiselev, V.G., 2010. The cumulant expansion: an overarching mathematical framework for understanding diffusion NMR. In: Jones, D.K. (Ed.), *Diffusion MRI*. Oxford University Press, pp. 152–168.
- Koay, C.G., 2010. Least squares approaches to diffusion tensor estimation. In: Jones, D.K. (Ed.), *Diffusion MRI*. Oxford University Press, pp. 272–284.
- Koay, C.G., Chang, L.C., Carew, J.D., Pierpaoli, C., Basser, P.J., 2006. A unifying theoretical and algorithmic framework for least squares methods of estimation in diffusion tensor imaging. *J. Magn. Reson.* 182, 115–125.
- Lasserre, J.B., 2007. A sum of squares approximation of nonnegative polynomials. *SIAM Rev.* 49, 651–669.
- Lazar, M., Alexander, A.L., 2005. Bootstrap white matter tractography (BOOT-TRAC). *Neuroimage* 24, 524–532.
- 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, 423–444.
- Liu, C., Bammer, R., Acar, B., Moseley, M.E., 2004. Characterizing non-Gaussian diffusion by using generalized diffusion tensors. *Magn. Reson. Med.* 51, 924–937.
- Lofberg, J., 2009. Pre- and post-processing sum-of-squares programs in practice. *IEEE Trans. Autom. Control* 54, 1007–1011.
- Lofberg, J., Parrilo, P., 2004. From coefficients to samples: a new approach to SOS optimization. In: *2004 43rd IEEE Conference on Decision and Control (CDC) (IEEE Cat. No.04CH37601)*. IEEE, Nassau, pp. 3154–3159.
- Magnani, A., Lall, S., Boyd, S., 2005. Tractable fitting with convex polynomials via sum-of-squares. In: *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC'05. 44th IEEE Conference on*, IEEE, pp. 1672–1677.
- Melonakos, J., Pichon, E., Angenent, S., Tannenbaum, A., 2008. Finsler active contours. *IEEE Trans. Pattern Anal. Mach. Intell.* 30, 412–423.
- Mitra, P.P., Halperin, B.I., 1995. Effects of finite gradient-pulse widths in pulsed-field-gradient diffusion measurements. *J. Magn. Reson., Ser. A* 113, 94–101.
- Mori, S., Crain, B.J., Chacko, V.P., Van Zijl, P.C.M., 1999. Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Ann. Neurol.* 45, 265–269.
- MOSEK ApS, 2018. MOSEK Optimizer API for C.
- Novikov, D.S., Kiselev, V.G., 2010. Effective medium theory of a diffusion-weighted signal. *NMR Biomed.* 23, 682–697.
- Mussel, M., Inzelt, L., Nevo, U., 2017. Insignificance of active flow for neural diffusion weighted imaging: a negative result. *Magn. Reson. Med.* 78 (2), 746–753. <https://doi.org/10.1002/mrm.26375>.
- Novikov, D.S., Fieremans, E., Jespersen, S.N., Kiselev, V.G., 2019. Quantifying brain microstructure with diffusion MRI: theory and parameter estimation. *NMR Biomed.* 32, e3998.
- Özarslan, E., Mareci, T.H., 2003. Generalized diffusion tensor imaging and analytical relationships between diffusion tensor imaging and high angular resolution diffusion imaging. *Magn. Reson. Med.* 50, 955–965.
- Özarslan, E., Koay, C.G., Basser, P.J., 2008. Simple harmonic oscillator based estimation and reconstruction for one-dimensional q-space MR. In: *Proceedings of the 16th Annual Meeting of the ISMRM*, Toronto, p. 35.
- Özarslan, E., Koay, C.G., Basser, P.J., 2009. Remarks on q-space MR propagator in partially restricted, axially-symmetric, and isotropic environments. *Magn. Reson. Imag.* 27, 834–844.
- Özarslan, E., Koay, C.G., Basser, P.J., 2013a. Simple harmonic oscillator based reconstruction and estimation for one-dimensional q-space magnetic resonance (1D-SHORE). In: *Excursions in Harmonic Analysis*, ume 2. Springer, pp. 373–399.
- Özarslan, E., Koay, C.G., Shepherd, T.M., Komlos, M.E., İrfanoğlu, M.O., Pierpaoli, C., Basser, P.J., 2013b. Mean apparent propagator (MAP) MRI: a novel diffusion imaging method for mapping tissue microstructure. *Neuroimage* 78, 16–32.
- O’Donnell, L., Haker, S., Westin, C.F., 2002. New approaches to estimation of white matter connectivity in diffusion tensor MRI: elliptic PDEs and geodesics in a tensor-warped space. In: Dohi, T., Kikinis, R. (Eds.), *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2002*, vol. 2488. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 459–466.
- Panagiotaki, E., Schneider, T., Siow, B., Hall, M.G., Lythgoe, M.F., Alexander, D.C., 2012. Compartment models of the diffusion MR signal in brain white matter: a taxonomy and comparison. *Neuroimage* 59, 2241–2254.
- Parker, G.J.M., Alexander, D.C., 2005. Probabilistic anatomical connectivity derived from the microscopic persistent angular structure of cerebral tissue. *Philos. Trans. R. Soc. Biol. Sci.* 360, 893–902.
- Parrilo, P.A., 2003. Semidefinite programming relaxations for semialgebraic problems. *Math. Program.* 96, 293–320.
- Péchaud, M., Descoteaux, M., Keriven, R., 2009. Brain connectivity using geodesics in HARDI. *Med. Image Comput. Comput. Assist. Interv. MICCAI 2009* 482–489.

- Pennec, X., Fillard, P., Ayache, N., 2006. A Riemannian framework for tensor computing. *Int. J. Comput. Vis.* 66, 41–66.
- Permenter, F., Parrilo, P.A., 2019. Dimension Reduction for Semidefinite Programs via Jordan Algebras. *Mathematical Programming*.
- Pichon, E., Westin, C.F., Tannenbaum, A.R., 2005. A Hamilton-Jacobi-bellman approach to high angular resolution diffusion tractography. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 180–187.
- Pierpaoli, C., Walker, L., Irfanoglu, M.O., Barnett, A., Bassar, P., Chang, L.C., Koay, C., Pajevic, S., Rohde, G., Sarlis, J., Wu, M., 2010. TORTOISE: an integrated software package for processing of diffusion MRI data. In: *Proceedings of the 18th Annual Meeting of the ISMRM, Stockholm*, p. 1597.
- Qi, L., Yu, G., Wu, E.X., 2010. Higher order positive semidefinite diffusion tensor imaging. *SIAM J. Imaging Sci.* 3, 416–433.
- Reisert, M., Mader, I., Anastasopoulos, C., Weigel, M., Schnell, S., Kiselev, V., 2011. Global fiber reconstruction becomes practical. *Neuroimage* 54, 955–962.
- Reisert, M., Kiselev, V.G., Dihtal, B., Kellner, E., Novikov, D.S., 2014. MesoFT: unifying diffusion modelling and fiber tracking. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 201–208.
- Rousseeuw, P.J., Croux, C., 1993. Alternatives to the median absolute deviation. *J. Am. Stat. Assoc.* 88, 1273.
- Rudin, W., 1996. *Fourier Analysis on Groups*, nachdr edition. Wiley Classics Library, Wiley-Interscience, New York, NY, p. 254881604. OCLC.
- Salvador, R., Peña, A., Menon, D.K., Carpenter, T.A., Pickard, J.D., Bullmore, E.T., 2005. Formal characterization and extension of the linearized diffusion tensor model. *Hum. Brain Mapp.* 24, 144–155.
- Schmüdgen, K., 1991. The K-moment problem for compact semi-algebraic sets. *Math. Ann.* 289, 203–206.
- Schober, M., Kasenburg, N., Feragen, A., Hennig, P., Hauberg, S., 2014. Probabilistic shortest path tractography in DTI using Gaussian process ODE solvers. In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2014*, Nagoya.
- Schwab, E., Afsari, B., Vidal, R., 2012. Estimation of non-negative ODFs using the eigenvalue distribution of spherical functions. In: *Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J.M., Mattern, F., Mitchell, J.C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., Vardi, M.Y., Weikum, G., Ayache, N., Delingette, H., Golland, P., Mori, K. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012*, ume 7511. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 322–330.
- Sepasian, N., ten Thije Boonkkamp, J.H.M., Ter Haar Romeny, B.M., Vilanova, A., 2012. Multivalued geodesic ray-tracing for computing brain connections using diffusion tensor imaging. *SIAM J. Imaging Sci.* 5, 483–504.
- Sepasian, N., ten Thije Boonkkamp, J.H., Florack, L.M., Romeny, B.M.T.H., Vilanova, A., 2014. Riemann-finsler multi-valued geodesic tractography for HARDI. In: *Visualization and Processing of Tensors and Higher Order Descriptors for Multi-Valued Data*. Springer, pp. 209–225.
- Setsovpop, K., Kimmlingen, R., Eberlein, E., Witzel, T., Cohen-Adad, J., McNab, J., Keil, B., Tisdall, M., Hoecht, P., Dietz, P., Cauley, S., Tountcheva, V., Matschl, V., Lenz, V., Heberlein, K., Potthast, A., Thein, H., Van Horn, J., Toga, A., Schmitt, F., Lehne, D., Rosen, B., Wedeen, V., Wald, L., 2013. Pushing the limits of in vivo diffusion MRI for the human connectome Project. *Neuroimage* 80, 220–233.
- Sherbondy, A.J., Dougherty, R.F., Ben-Shachar, M., Napel, S., Wandell, B.A., 2008. ConTrack: finding the most likely pathways between brain regions using diffusion tractography. *J. Vis.* 8, 15–15.
- Sun, J., Xie, Y., Ye, W., Ho, J., Entezari, A., Blackband, S.J., Vemuri, B.C., 2013. Dictionary learning on the manifold of square root densities and application to reconstruction of diffusion propagator fields. In: *Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J.M., Mattern, F., Mitchell, J.C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., Vardi, M.Y., Weikum, G., Gee, J.C., Joshi, S., Pohl, K.M., Wells, W.M., Zöllei, L. (Eds.), Information Processing in Medical Imaging*, vol. 7917. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 619–631.
- Tax, C.M.W., Jeurissen, B., Vos, S.B., Viergever, M.A., Leemans, A., 2014. Recursive calibration of the fiber response function for spherical deconvolution of diffusion MRI data. *Neuroimage* 86, 67–80.
- Tournier, J.D., Calamante, F., Gadian, D.G., Connelly, A., 2004. Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *Neuroimage* 23, 1176–1185.
- Tournier, J.D., Calamante, F., Connelly, A., 2007. Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution. *Neuroimage* 35, 1459–1472.
- Tournier, J.D., Calamante, F., Connelly, A., 2013. Determination of the appropriate b-value and number of gradient directions for high-angular-resolution diffusion-weighted imaging. *NMR Biomed.* 26, 1775–1786.
- Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K., 2013. The Wu-Minn human connectome Project: an overview. *Neuroimage* 80, 62–79.
- Vemuri, B.C., Sun, J., Banerjee, M., Pan, Z., Turner, S.M., Fuller, D.D., Forder, J.R., Entezari, A., 2019. A geometric framework for ensemble average propagator reconstruction from diffusion MRI. *Med. Image Anal.* 57, 89–105.
- 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, 678–686.
- Veraart, J., Sijbers, J., Sunaert, S., Leemans, A., Jeurissen, B., 2013. Weighted linear least squares estimation of diffusion MRI parameters: strengths, limitations, and pitfalls. *Neuroimage* 81, 335–346.
- Wang, Z., Vemuri, B., Chen, Y., Mareci, T., 2004. A constrained variational principle for direct estimation and smoothing of the diffusion tensor field from complex DWI. *IEEE Trans. Med. Imaging* 23, 930–939.
- Wedeen, V.J., Wang, R.P., Schmahmann, J.D., Benner, T., Tseng, W.Y.I., Dai, G., Pandya, D.N., Hagmann, P., D’Arceuil, H., de Crespigny, A.J., 2008. Diffusion spectrum magnetic resonance imaging (DSI) tractography of crossing fibers. *Neuroimage* 41, 1267–1277.
- Wolfram, S., 2003. *The Mathematica Book*, 5 edition. Wolfram Media, Champaign, IL.