Diffusion tensor imaging with deterministic error bounds

Artur Gorokh[*]   Yury Korolev[†]   Tuomo Valkonen[‡]

January 26, 2016

Abstract

Errors in the data and the forward operator of an inverse problem can be handily modelled using partial order in Banach lattices. We present some existing results of the theory of regularisation in this novel framework, where errors are represented as bounds by means of the appropriate partial order.

We apply the theory to diffusion tensor imaging (DTI), where correct noise modelling is challenging: it involves the Rician distribution and the nonlinear Stejskal-Tanner equation. Linearisation of the latter in the statistical framework would complicate the noise model even further. We avoid this using the error bounds approach, which preserves simple error structure under monotone transformations.

1 Introduction

Often in inverse problems, we have only very rough knowledge of noise models, or the exact model is too difficult to realise in a numerical reconstruction method. The data may also contain process artefacts from black box devices [44]. Partial order in Banach lattices has therefore recently been investigated in [33, 35, 34] as a less-assuming error modelling approach for inverse problems. This framework allows the representation of errors in the data as well as in the forward operator of an inverse problem by means of order intervals (i.e., lower and upper bounds by means of appropriate partial orders). An important advantage of this approach vs. statistical noise modelling is that deterministic error bounds preserve their simple structure under monotone transformations.

[*]Faculty of Physics, Lomonosov Moscow State University, Russia. `arturgorokh@yahoo.com`

[†]School of Engineering and Materials Science, Queen Mary University of London, United Kingdom. `korolev.msu@gmail.com`

[‡]Department of Applied Mathematics and Theoretical Physics, University of Cambridge, United Kingdom. `tuomo.valkonen@iki.fi`

We apply partial order in Banach lattices to diffusion tensor imaging (DTI). We will in due course explain the diffusion tensor imaging progress, as well as the theory of inverse problems in Banach lattices, but start by introducing our model

$$\min_u \; R(u) \quad \text{subject to} \quad \begin{aligned} & u \geqslant 0, \\ & g_j^l \leqslant A_j u \leqslant g_j^u, \quad \mathcal{L}^n\text{-a.e. on } \Omega, \; (j = 1, \ldots, N). \end{aligned}$$

That is, we want to find a field of symmetric 2-tensors $u : \Omega \to \mathrm{Sym}^2(\mathbb{R}^3)$ on the domain $\Omega \subset \mathbb{R}^3$, minimising the value of the regulariser R on the feasible set. The tensor field u is our unknown image. It is subject a positivity constraint, as well as partial order constraints imposed through the operators $[A_j u](x) := -\langle b_j, u(x) b_j \rangle$, and the upper and lower bounds $g_j^l := \log(\hat{s}_j^l / \hat{s}_0^u)$ and $g_j^u := \log(\hat{s}_j^u / \hat{s}_0^l)$. These model, in terms of error intervals after logarithmic transformation, the Stejskal–Tanner equation

$$s_j(x) = s_0(x) \exp(-\langle b_j, u(x) b_j \rangle), \quad (j = 1, \ldots, N), \tag{1.1}$$

central to the diffusion tensor imaging process.

To shed more light on u and the equation (1.1), let us briefly outline the diffusion tensor imaging process. As a first step towards DTI, diffusion weighted magnetic resonance imaging (DWI) is performed. This process measures the anisotropic diffusion of water molecules. To capture the diffusion information, the magnetic resonance images have to be measured with diffusion sensitising gradients in multiple directions. These are the different b_i's in (1.1). Eventually, multiple DWI images $\{s_j\}$ are related through the Stejskal-Tanner equation (1.1) to the symmetric positive-definite diffusion-tensor field $u : \Omega \to \mathrm{Sym}^2(\mathbb{R}^3)$ [4, 30]. At each point $x \in \Omega$, the tensor $u(x)$ is the covariance matrix of a normal distribution for the probability of water diffusing in different spatial directions.

The fact that multiple b_i's are needed to recover u, leads to very long acquisition times, even with ultra fast sequences like echo planar imaging (EPI). Therefore, DTI is inherently a low-resolution and low-SNR method. In theory, the amplitude DWI images exhibit Rician noise [25]. However, as the histogram of an in vivo measurement in Figure 1 illustrates, this may not be the case for practical data sets from black-box devices. Moreover, the DWI process is prone to eddy-current distortions [53], and due to the slowness of it, it is very sensitive to patient motion [28, 1]. We therefore have to use techniques that remove these artefacts in solving for $u(x)$. We also need to ensure the positivity u, as non-positive-definite diffusion tensor are non-physical. One proposed approach for the satisfaction of this constraint is that of log-Euclidean metrics [3]. This approach has several theoretically desirable aspects, but some practical shortcomings [57]. Special Perona-Malik type constructions on Riemannian manifolds can also be used to
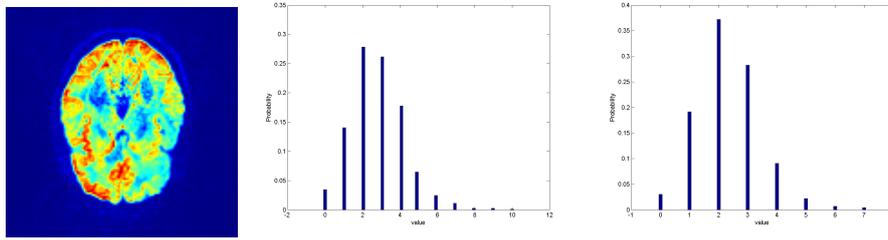
(a) Slice of a real MRI measurement (b) 50-bin histogram of noise estimated from background (c) 50-bin histogram after eddy-current correction with FSL

Figure 1: The noise in the absolute values of complex MRI data should be Rician. Here we have taken a 50-bin histogram of the noise in real data. This divides the pixels into bins of 50 different noise levels. However, we only find approximately 10 noise levels to have non-zero pixel count. As the Rician distribution is continuous, we see that the noise cannot be Rician, some bins of the 50-bin histogram being empty. The measurement setup of the data used here is described in Section 5.3.

maintain the structure of the tensor field [14, 54]. Such anisotropic diffusion is however severely ill-posed [60]. Recently manifold-valued discrete-domain total variation models have also been applied to diffusion tensor imaging [6].

Our approach is also in the total variation family, first considered for diffusion tensor imaging in [48]. Namely, we follow up on the work in [57, 59, 58, 55] on the application of total generalised variation regularisation [9] to DTI. We should note that in all of these works the fidelity function was the ROF-type [45] L^2 fidelity. This would only be correct, according to the assumption that noise of MRI measurements is Gaussian, if we had access to the original complex k-space MRI data. The noise of the inverse Fourier-transformed magnitude data s_j, that we have in practice access to, is however Rician under the Gaussian assumption on the original complex data [25]. This is not modelled by the L^2 fidelity.

Numerical implementation of Rician noise modelling has been studied in [40, 23]. As already discussed, in this work, we take the other direction. Instead of modelling the errors in a statistically accurate fashion, not assuming to know an exact noise model, we represent them by means of pointwise bounds. The details of the model are presented in Section 3. We study the practical performance in Section 5 using the numerical method presented in Section 4. First we however start with the general error modelling theory in Section 2. Readers who are not familiar with notation for Banach lattices or symmetric tensor fields are advised to start with the Appendix, where we introduce our mathematical notation and techniques.

3

2 Deterministic error modelling

2.1 Mathematical basis

We now briefly outline the theoretical framework [33] that is the basis for our approach. Consider a linear operator equation

$$Au = f, \quad u \in U, \ f \in F, \tag{2.1}$$

where U and F are Banach lattices, $A \colon U \to F$ is a regular injective operator. The inaccuracies in the right-hand side f and the operator A are represented as bounds by means of appropriate partial orders, i.e.

$$\begin{aligned}
f^l, f^u \colon & \quad f^l \leqslant_F f \leqslant_F f^u, \\
A^l, A^u \colon & \quad A^l \leqslant_{L^\sim(U,F)} A \leqslant_{L^\sim(U,F)} A^u,
\end{aligned} \tag{2.2}$$

where the symbol $\leqslant_F$ stands for the partial order in F and $\leqslant_{L^\sim(U,F)}$ for the partial order for regular operators induced by partial orders in U and F. Further, we will drop the subscripts at inequality signs where it will not cause confusion.

The exact right-hand side f and operator A are not available. Given the approximate data (f^l, f^u, A^l, A^u), we need to find an approximate solution u that converges to the exact solution $\bar{u}$ as the inaccuracies in the data diminish. This statement needs to be formalised. We consider monotone convergent sequences of lower and upper bounds

$$\begin{aligned}
f_n^l \colon & f_{n+1}^l \geqslant f_n^l, & A_n^l \colon & A_{n+1}^l \geqslant A_n^l, \\
f_n^u \colon & f_{n+1}^u \leqslant f_n^u, & A_n^u \colon & A_{n+1}^u \leqslant A_n^u, \\
& f_n^l \leqslant f \leqslant f_n^u, & & A_n^l \leqslant A \leqslant A_n^u \quad \forall n \in \mathbb{N}, \\
& \|f_n^l - f_n^u\| \to 0, & & \|A_n^l - A_n^u\| \to 0 \quad \text{as } n \to \infty.
\end{aligned} \tag{2.3}$$

We are looking for an approximate solution u_n such that $\|u_n - \bar{u}\|_U \to 0$ as $n \to \infty$.

Let us ask the following question. What are the elements $u \in U$ that could have produced data within the tolerances (2.3)? Obviously, the exact solution is one of such elements. Let us call the set containing all such elements the feasible set $U_n \subset U$.

Suppose that we know *a priori* that the exact solution is positive (by means of the appropriate partial order in U). Then it is easy to verify that the following inequalities hold for all $n \in \mathbb{N}$

$$\bar{u} \geqslant_U 0, \quad A_n^u \bar{u} \geqslant_F f_n^l, \quad A_n^l \bar{u} \leqslant_F f_n^u.$$

This observation motivates our choice of the feasible set:

$$U_n = \{ u \in U : u \geqslant_U 0, \quad A_n^u u \geqslant_F f_n^l, \quad A_n^l u \leqslant_F f_n^u \}.$$

It is clear that the exact solution $\bar{u}$ belongs to the sets U_n for all $n \in \mathbb{N}$. Our goal is to define a rule that will choose for any n an element u_n of the set U_n such that the sequence $u_n \in U_n$ will strongly converge to the exact solution $\bar{u}$. We do so by minimising an appropriate regularisation functional $R(u)$ on U_n:

$$u_n = \arg\min_{u \in U_n} R(u). \tag{2.4}$$

This method, in fact, is a lattice analogue of the well-known residual method [24, 52]. The convergence result is as follows [33].

Theorem 2.1. *Suppose that*

1. *$R(u)$ is bounded from below on U,*

2. *$R(u)$ is lower semi-continuous,*

3. *level sets $\{u\colon R(u) \leqslant C\}$ ($C = $ const) are sequentially compact in U (in the strong topology induced by the norm).*

Then the sequence defined in (2.4) strongly converges to the exact solution $\bar{u}$ and $\mathcal{R}(u_n) \to \mathcal{R}(\bar{u})$.

Examples of regularisation functionals that satisfy the conditions of Theorem 2.1 are as follows. Total Variation in $L^1(\Omega)$, where Ω is a subset of $\mathbb{R}^n$, assures strong convergence in L^1, given that the L^1-norm of the solution is bounded. The Sobolev norm $\|u\|_{W^{1,q}(\Omega)}$ yields strong convergence in the spaces $L^p(\Omega)$, where $p \geqslant 1$, $q > \frac{np}{p+n}$. The latter fact follows from the compact embedding of the corresponding Sobolev $W^{1,q}(\Omega)$ space into $L^p(\Omega)$ [17].

However, the assumption that the sets $\{u\colon R(u) \leqslant C\}$ are strong compacts in U is quite strong. It can be replaced by the assumption of weak compactness, provided that the regularisation functional possesses the so-called Radon-Riesz property.

Definition 2.1. A functional $F\colon U \to \mathbb{R}$ has the Radon-Riesz property (sometimes referred to as the H-property), if for any sequence $u_n \in U$ weak convergence $u_n \rightharpoonup u_0$ and simultaneous convergence of the values $F(u_n) \to F(u_0)$ imply strong convergence $u_n \to u_0$.

Theorem 2.2. *Suppose that*

1. *$R(u)$ is bounded from below on U,*

2. *$R(u)$ is weakly lower semi-continuous,*

3. *level sets $R(u) \leqslant C$ ($C = $ const) are weakly sequentially compact in U,*

4. *$R(u)$ possesses the Radon-Riesz property.*

5

Then the sequence defined in (2.4) strongly converges to the exact solution $\bar{u}$ and $\mathcal{R}(u_n) \to \mathcal{R}(\bar{u})$.

It is easy to verify that the norm in any Hilbert space possesses the Radon-Riesz property. Moreover, this holds for the norm in any reflexive Banach space [17].

As we explain in the Appendix, the spaces $L^p(\Omega; \mathrm{Sym}^2(\mathbb{R}^m))$ are not Banach lattices, therefore, Theorems 2.1 and 2.2 cannot be applied directly. Further theoretical work will be undertaken to extend the framework to the non-lattice case. For the moment, however, we will prove that if there are no errors in the operator A in (2.1), the requirement that the solution space U is a lattice can be dropped.

Theorem 2.3. *Let U be a Banach space, and F be a Banach lattice. Let the operator A in (2.1) be a linear, continuous and injective operator. Let f_n^l and f_n^u be sequences of lower and upper bounds for the right-hand side defined in (2.3), and suppose that there are no errors in the operator A. Let us redefine the feasible set in the following way*

$$U_n = \{u \in U: \quad f_n^l \leqslant_F Au \leqslant_F f_n^u\}.$$

Suppose also that the regulariser $R(x)$ satisfies conditions of either Theorem 2.1 or Theorem 2.2. Then the sequence defined in (2.4) strongly converges to the exact solution $\bar{u}$ and $R(u_n) \to R(\bar{u})$.

Proof. Let us define an approximate right-hand side and its approximation error in the following way

$$f_{\delta_n} = \frac{f_n^u + f_n^l}{2}, \quad \delta_n = \frac{\|f_n^u - f_n^l\|}{2}.$$

One can easily verify, that the inequality $\|f - f_{\delta_n}\| \leqslant \delta_n$ holds. Indeed, we have

$$f - f_{\delta_n} \leqslant f_n^u - f_{\delta_n} = \frac{f_n^u - f_n^l}{2},$$

$$-(f - f_{\delta_n}) \leqslant f_{\delta_n} - f_n^l = \frac{f_n^u - f_n^l}{2},$$

$$|f - f_{\delta_n}| = (f - f_{\delta_n}) \vee (-(f - f_{\delta_n})) \leqslant \frac{f_n^u - f_n^l}{2},$$

$$\|f - f_{\delta_n}\| \leqslant \frac{\|f_n^u - f_n^l\|}{2}.$$

The first two inequalities are consequences of the conditions (2.3), the third one holds by the definition of supremum and the equality $|f| = f \vee (-f)$ that holds for all $f \in F$, and the last inequality is due to the monotonicity of the norm in a Banach lattice.

Similarly, one can show that for any $u \in U_n$, we have

$$\|Au - f_{\delta_n}\| \leqslant \delta_n.$$

Therefore, the inclusion $U_n \subset \{u \colon \|Au - f_{\delta_n}\| \leqslant \delta_n\}$ holds.

Now we will proceed with the proof of convergence $\|u_n - \bar{u}\| \to 0$. Will prove it for the case when the regulariser $R(u)$ satisfies conditions of Theorem 2.1. Suppose that the sequence u_n does not converge to the exact solution $\bar{u}$. Then it contains a subsequence u_{n_k} such that $\|u_{n_k} - \bar{u}\| \geqslant \varepsilon$ for any $k \in \mathbb{N}$ and some fixed $\varepsilon > 0$.

Since the inclusion $\bar{u} \in U_n$ holds for all $n \in \mathbb{N}$, we have $R(u_n) \leqslant R(\bar{u})$ for all $n \in \mathbb{N}$. Since the level set $\{u \colon R(u) \leqslant R(\bar{u})\}$ is a compact set by assumptions of the theorem, the sequence u_{n_k} contains a strongly convergent subsequence. With no loss of generality, let us assume that $u_{n_k} \to u_0$. We will now show that $u_0 = \bar{u}$. Indeed, we have

$$\|Au_{n_k} - A\bar{u}\| \leqslant \|Au_{n_k} - f_{\delta_n}\| + \|f - f_{\delta_n}\| \leqslant 2\delta_{n_k} \to 0.$$

On the other hand, we have

$$\|Au_{n_k} - A\bar{u}\| \to \|Au_0 - A\bar{u}\|$$

due to continuity of A and $\|\cdot\|$. Therefore, $Au_0 = A\bar{u}$ and $u_0 = \bar{u}$, since A is an injective operator. By contradiction, we get $\|u_n - \bar{u}\| \to 0$.

Finally, since the regulariser $R(u)$ is lower semi-continuous, we get that $\liminf R(u_n) = R(\bar{u})$. However, for any n we have $R(u_n) \leqslant R(\bar{u})$, therefore, we get the convergence $R(u_n) \to R(\bar{u})$ as $n \to \infty$. □

2.2 Philosophical discussion and statistical interpretation

In practice, our data is discrete. So let us momentarily switch to measurements $\hat{f} = (\hat{f}^1, \ldots, \hat{f}^n) \in \mathbb{R}^n$ of a true data $f \in \mathbb{R}^n$. If we actually knew the pointwise noise model of the data, then one way to obtain potentially useful upper and lower bounds for the deterministic model is by means of statistical interval estimates: confidence intervals. Roughly, the idea is to find for each true signal f individual *random* upper and lower bounds $\hat{f}^u$ and $\hat{f}^l$ such that

$$P(\hat{f}^u \leq f \leq \hat{f}^l) = 1 - \theta.$$

If $\hat{f}^u$ and $\hat{f}^l$ are computed based on multiple experiments (i.e., multiple noisy samples $\hat{f}_1, \ldots, \hat{f}_m$, of the true data f), the interval $[\hat{f}^{u,i}, \hat{f}^{l,i}]$ will converge in probability to the true data $\hat{f}^i$, as the number of experiments m increases. Thus we obtain a probabilistic version of the convergences in (2.3).

Let us try to see, how such intervals might work in practice. For the purpose of the present discussion, assume that the noise is additive and normal-distributed with variance σ and zero mean—an assumption that does not hold in practice, as we have already seen in Figure 1, but will suffice for the next thought experiments. That is, $\hat{f}_j = f + \nu_j$ for the noise

ν_j. Let the sample mean of $\{\hat{f}_j\}_{j=1}^m$ be $\bar{f}_m = (\bar{f}_m^1, \ldots \bar{f}_m^n)$, and pointwise sample variance $\sigma = (\sigma^1, \ldots, \sigma^n)$. The product of the pointwise confidence intervals I_i with confidence $1 - \theta$ is [15, 49]

$$\prod_{i=1}^n I_i = [\bar{f}_m - k_{1-\theta/2}^* \frac{\sigma}{\sqrt{m}}, \bar{f}_m + k_{1-\theta/2}^* \frac{\sigma}{\sqrt{m}}], \quad k_{1-t}^* := \Phi^{-1}(t),$$

for Φ the cumulative normal distribution function. For $\theta = 0.05$, i.e., the 95% confidence interval, $\Phi^{-1}(0.05/2) = 1.96$. Now, the probability that I_i covers the true f^i is $1 - \theta$, e.g. 95%. If we have only a single sample $m = 1$, the intervals stay large, but the joint probability, $(1 - \theta)^n$ goes to zero as n increases. As an example, for a rather typical single 128×128 slice of a DTI measurement, the probability that exactly $\phi = 5\%$ (to the closest discrete value possible) of the $1 - \theta = 95\%$ confidence intervals do not cover the true parameter would be about 1.4%, or

$$1.4\% \approx \binom{n}{m} \theta^m (1 - \theta)^{n-m}, \quad \text{where } n = 128^2 \text{ and } m = \lceil \phi n \rceil.$$

The probability of *at least* $\phi = 5\%$ of the pointwise 95% confidence intervals not covering the true parameter is in this setting approximately 49%. This can be verified by summing the above estimates over $m = \lceil \phi n \rceil, \ldots, n$.

In summary, unless θ simultaneously goes to 1, the product intervals are very unlikely to cover the true parameter. Based on a single experiment, the deterministic approach as interpreted statistically through confidence intervals, is therefore very likely to fail to discover the true solution as the data size n increases unless the pointwise confidence is very low. But, if we let the pointwise confidences be arbitrarily high, such that the intervals are very large, the discovered solution in our applications of interest would be just a constant!

Asymptotically, the situation is more encouraging. Indeed, if we could perform more experiments to compute the confidence intervals, then for any fixed n and θ, it is easy to see that the solution of the "deterministic" error model is an asymptotically consistent and hence asymptotically unbiased estimator of the true f. That is, the estimates converge in probability to f as the experiment count m increases. Indeed, the error-bounds based estimator $\tilde{f}_m$, based on m experiments, by definition satisfies $\tilde{f}_m \in \prod_{i=1}^n I_i$. Therefore, we have

$$P(|\tilde{f}_m^i - \bar{f}_m^i| > \epsilon \text{ for some } i) = 0 \quad \text{whenever} \quad m \geq (k_{1-\theta/2}^* \sigma / \epsilon)^2.$$

Thus $f \xrightarrow{P} \bar{f}$ in probability. Since by the law of large numbers also $\bar{f}_m \xrightarrow{P} f$, this proves the claim, and to some extent justifies our approach from the statistical viewpoint.

It should be noted that this is roughly the most that has previously been known of the maximum a posteriori estimate (MAP), corresponding to the Tikhonov models

$$\min \frac{1}{2}\|\hat{f} - u\|^2 + \alpha R(u).$$

In particular, the MAP is not the Bayes estimator for the typical squared cost functional. This means that it does not minimise $\tilde{f} \mapsto \mathbb{E}[\|f - \tilde{f}\|^2]$. The minimiser in this case is the conditional mean (CM) estimate, which is why it has been preferred by Bayesian statisticians despite its increased computational cost. The MAP estimate is merely an asymptotic Bayes estimator for the uniform cost function. In a very recent work [12], it has however been proved that the MAP estimate *is* the Bayes estimator for certain Bregman distances. One possible critique of the result is that these distances are not universal and do depend on the regulariser R, unlike the squared distance for CM. The CM estimate however has other problems in the setting of total variation and its discretisation [37, 36].

3 Application to diffusion tensor imaging

We now build our model for applying the deterministic error modelling theory to diffusion tensor imaging. We start by building our forward model based on the Stejskal-Tanner equation, and then briefly introduce the regularisers we use.

3.1 The forward model

For $u : \Omega \to \mathrm{Sym}^2(\mathbb{R}^3)$, $\Omega \subset \mathbb{R}^3$, a mapping from Ω to symmetric second order tensors, let us introduce non-linear operators T_j, defined by

$$[T_j(u)](x) := s_0(x) \exp(-\langle b_j, u(x)b_j \rangle), \quad (j = 1, \dots, N).$$

Their role is to model the so-called Stejskal-Tanner equation [4]

$$s_j(x) = s_0(x) \exp(-\langle b_j, u(x)b_j \rangle), \quad (j = 1, \dots, N). \tag{3.1}$$

Each tensor $u(x)$ models the covariance of a Gaussian probability distribution at x for the diffusion of water molecules. The data $s_j \in L^2(\Omega)$, $(j = 1, \dots, N)$, are the diffusion-weighted MRI images. Each of them is obtained by performing the MRI scan with a different non-zero diffusion sensitising gradient b_j, while s_0 is obtained with a zero gradient. After correcting the original k-space data for coil sensitivities, each s_j is assumed real. As a consequence, any measurement $\hat{s}_j$ of s_j has—in theory—Rician noise distribution [25].

9

Our goal is to reconstruct u with simultaneous denoising. Following [55, 32], we consider using a suitable regulariser R the Tikhonov model

$$\min_{u \geq 0} \sum_{j=1}^{N} \frac{1}{2} \|\hat{s}_j - T_j(u)\|^2 + \alpha R(u). \tag{3.2}$$

The constraint $u \geqq 0$ is to be understood in the sense that $u(x)$ is positive semidefinite for $\mathcal{L}^n$-a.e. $x \in \Omega$ (see Appendix for more details). Due to the Rician noise of $\hat{s}_j$, the Gaussian noise model implied by the L^2-norm in (3.2) is not entirely correct. However, in some cases the L^2 model may be accurate enough, as for suitable parameters the Rician distribution is not too far from a Gaussian distribution. If one were to model the problem correctly, one should either modify the fidelity term to model Rician noise, or include the (unit magnitude complex number) coil sensitivities in the model. The Rician noise model is highly nonlinear due to the Bessel functional logarithms involved. Its approximations have been studied in [5, 23, 40] for single MR images and DTI. Coil sensitivities could be included either by knowing them in advance, or by simultaneous estimation as in [31]. Either way, significant complexity is introduced into the model, and for the present work, we are content with the simple L^2 model.

We may also consider, as is often the case, and as was done with TGV in [57], the linearised model

$$\min_{u \geq 0} \|f - u\|^2 + \alpha R(u), \tag{3.3}$$

where, for each $x \in \Omega$, $f(x)$ is solved by regression for $u(x)$ from the system of equations (3.1) with $s_j(x) = \hat{s}_j(x)$. Further, as in [58], we may also consider

$$\min_{u \geq 0} \sum_{j=1}^{N} \frac{1}{2} \|g_j - A_j u\|^2 + \alpha R(u), \tag{3.4}$$

with $[A_j u](x) := -\langle b_j, u(x) b_j \rangle$, and $g_j(x) := \log(\hat{s}_j(x)/\hat{s}_0(x))$. In both of these linearised models, the assumption of Gaussian noise is in principle even more remote from the truth than in the nonlinear model (3.2). We will employ (3.3) and (3.2) as benchmark models.

We want to further simplify the model, and forgo with accurate noise modelling. After all, we often do not know the real noise model for the data available in practice. It can be corrupted by process artefacts from black-box algorithms in the MRI devices. This problem of black box devices has been discussed extensively in [44], in the context of Computed Tomography. Moreover, as we have discussed above, even without such artefacts, the correct model may be difficult to realise numerically. So we might be best off choosing the least assuming model of all – that of error bounds.

This is what we propose in the reconstruction model

$$\min_u R(u) \quad \text{subject to} \quad u \geq 0, \tag{3.5}$$
$$g_j^l \leqslant A_j u \leqslant g_j^u, \quad \mathcal{L}^n\text{-a.e.,} \ (j = 1, \dots, N).$$

Here $g_j^l := \log(\hat{s}_j^l / \hat{s}_0^u)$ and $g_j^u := \log(\hat{s}_j^u / \hat{s}_0^l)$, $g_j^l, g_j^u \in L^2(\Omega)$, are our upper and lower bounds on g_j that we derive from the data.

3.2   Choice of the regulariser R

A prototypical regulariser in image processing is the total variation, first studied in this context in [45]. It can be defined for a symmetric tensor field $u \in L^1(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$ as

$$\mathrm{TV}(u) := \|Eu\|_{\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))}$$
$$:= \sup \left\{ \int_\Omega \langle \mathrm{div}\, \phi(x), u(x) \rangle \, \mathrm{d}\,x \left|\ \begin{array}{l} \phi \in C_c^\infty(\Omega; \mathrm{Sym}^{k+1}(\mathbb{R}^m)) \\ \sup_x \|\phi(x)\|_F \leq 1 \end{array} \right. \right\}.$$

Observe that for scalar or vector fields, i.e., the cases $k = 0, 1$, we have $\mathrm{Sym}^0(\mathbb{R}^m) = \mathcal{T}^0(\mathbb{R}^m) = \mathbb{R}$, and $\mathrm{Sym}^1(\mathbb{R}^m) = \mathcal{T}^1(\mathbb{R}^m) = \mathbb{R}^m$. Therefore, for scalars in particular, this gives the usual isotropic total variation

$$\mathrm{TV}(u) = \|Du\|_{\mathcal{M}(\Omega))}.$$

Total generalised variation was introduced in [9] as a higher-order extension of TV. Following [57], the second-order variant may be defined using the differentiation cascade formulation for symmetric tensor fields $u \in L^1(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$ as the marginal

$$\mathrm{TGV}^2_{(\beta,\alpha)}(u) := \min\{\Phi_{(\beta,\alpha)}(u, w) \mid w \in L^1(\Omega; \mathrm{Sym}^{k+1}(\mathbb{R}^m))\} \tag{3.6}$$

for

$$\Phi_{(\beta,\alpha)}(u, w) := \alpha \|Eu - w\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} + \beta \|Ew\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+2}(\mathbb{R}^m))}.$$

It turns out that the standard BV norm

$$\|u\|_{\mathrm{BV}(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))} := \|u\|_{L^1(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))} + \mathrm{TV}(u)$$

and the "BGV norm" [9]

$$\|u\|' := \|u\|_{L^1(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))} + \mathrm{TGV}^2_{(\beta,\alpha)}(u)$$

are topologically equivalent norms [10, 11] on $\mathrm{BV}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$, yielding the same convergence results for TGV regularisation as for TV regularisation. The geometrical regularisation behaviour is however different, and TGV tends to avoid the staircasing observed in TV regularisation.

Regarding topologies, we say that a sequence $\{u^i\}$ in $\mathrm{BV}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$ converges *weakly** to u, if $u^i \to u$ strongly in L^1, and $Eu^i \overset{*}{\rightharpoonup} Eu$ weakly* as Radon measures [2, 51, 57]. The latter is characterised as $\int_\Omega \langle \mathrm{div}\, \phi(x), u^i(x) \rangle \, \mathrm{d}\,x \to \int_\Omega \langle \mathrm{div}\, \phi(x), u(x) \rangle \, \mathrm{d}\,x$ for all $\phi \in C_c^\infty(\Omega; \mathrm{Sym}^{k+1}(\mathbb{R}^m))$.

3.3 Compact subspaces

Now, for a weak* lower semi-continuous seminorm R on $\mathrm{BV}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$, let us set

$$\mathrm{BV}_{0,R}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m)) := \mathrm{BV}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))/\ker R.$$

That is, we identify elements $u, \tilde{u} \in \mathrm{BV}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$, such that $R(u - \tilde{u}) = 0$. Now R is a norm on $\mathrm{BV}_{0,R}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$; compare, e.g., [42] for the case of $R = \mathrm{TV}$.

Suppose

$$\|u\|' := \|u\|_{L^1(\Omega)} + R(u)$$

is a norm on $\mathrm{BV}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$, equivalent to the standard norm. If also the *R-Sobolev-Korn-Poincaré inequality*

$$\inf_{R(v)=0} \|u - v\|_{L^1(\Omega)} \le CR(u) \tag{3.7}$$

holds, we may then bound

$$\inf_{R(v)=0} \|u - v\|_{\mathrm{BV}(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))} \le \inf_{R(v)=0} C'\|u - v\|'$$
$$= \inf_{R(v)=0} C'\big(\|u - v\|_{L^1(\Omega)} + R(u - v)\big) \le C'(1 + C)R(u).$$

By the weak* lower semicontinuity of the BV-norm, and the weak* compactness of the unit ball in $\mathrm{BV}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$—we refer to [2] for these and other basic properties of BV-spaces—we may thus find a representative $\tilde{u}$ in the $\mathrm{BV}_{0,R}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$ equivalence class of u, satisfying

$$\|\tilde{u}\|_{\mathrm{BV}(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))} \le C'(1 + C)R(u).$$

Again using the weak* compactness of the unit ball in $\mathrm{BV}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$, and the weak* lower semicontinuity of R, it follows that the sets

$$\mathrm{lev}_a R := \{u \in \mathrm{BV}_{0,R}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m)) \mid R(u) \le a\}, \quad (a > 0),$$

are weak* compact in $\mathrm{BV}_{0,R}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$, in the topology inherited form $\mathrm{BV}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$. Consequently, they are strongly compact subsets of $L^1(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$. This feature is crucial for the application of the regularisation theory in Banach lattices above.

On a connected domain Ω, in particular

$$\mathrm{BV}_{0,\mathrm{TV}}(\Omega) \simeq \left\{u \in \mathrm{BV}(\Omega)\,\bigg|\,\int_\Omega u\,\mathrm{d}\,x = 0\right\}.$$

That is, the space consists of zero-mean functions. Then $u \mapsto \|Du\|_{\mathcal{M}(\Omega;\mathbb{R}^m)}$ is a norm on $\mathrm{BV}_{0,\mathrm{TV}}(\Omega)$ [42], and this space is weak* compact. In particular, the sets $\mathrm{lev}_a \mathrm{TV}$ are compact in $L^1(\Omega)$.

More generally, we know from [8] that on a connected domain Ω, $\ker\mathrm{TV}$ consists of $\mathrm{Sym}^k(\mathbb{R}^m)$-valued polynomials of maximal degree k. By extension, $\ker\mathrm{TGV}^2$ consists of $\mathrm{Sym}^k(\mathbb{R}^m)$-valued polynomials of maximal degree $k+1$. In both cases, (3.7), weak* lower semicontinuity of R, and the equivalence of $\|\cdot\|'$ to $\|\cdot\|_{\mathrm{BV}(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))}$ hold by the results in [8, 11, 51]. Therefore, we have proved the following.

Lemma 3.1. *Let $\Omega\subset\mathbb{R}^m$ and $k\geq 0$. Then the sets $\mathrm{lev}_a\,\mathrm{TV}$ and $\mathrm{lev}_a\,\mathrm{TGV}^2$ are weak* compact in $\mathrm{BV}(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))$ and strongly compact in $L^1(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))$.*

Now, in the above cases, $\ker R$ is finite-dimensional, and we may write

$$\mathrm{BV}(\Omega;\mathrm{Sym}^k(\mathbb{R}^m)) \simeq \mathrm{BV}_{0,R}(\Omega;\mathrm{Sym}^k(\mathbb{R}^m)) \oplus \ker R.$$

Denoting by

$$B_X(r) := \{x\in X \mid \|x\|\leq r\},$$

the closed ball of radius r in a normed space X, we obtain by the finite-dimensionality of $\ker R$ the following result.

Proposition 3.1. *Let $\Omega\subset\mathbb{R}^m$ and $k\geq 0$. Pick $a>0$. Then the sets*

$$V := \mathrm{lev}_a\,R \oplus B_{\ker R}(a)$$

for $R=\mathrm{TV}$ and $R=\mathrm{TGV}^2$ are weak* compact in $\mathrm{BV}(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))$ and strongly compact in $L^1(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))$.

The next result summarises Theorem 2.3 and Proposition 3.1.

Theorem 3.1. *With $U = L^1(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))$, let the operator $A:U\to F$ be linear, continuous and injective. Let f_n^l and f_n^u be sequences of lower and upper bounds for the right-hand such that*

$$f_n^l\colon f_{n+1}^l\geqslant f_n^l, \qquad f_n^u\colon f_{n+1}^u\leqslant f_n^u,$$
$$f_n^l\leqslant f\leqslant f_n^u, \qquad \|f_n^l-f_n^u\|\to 0 \quad as\ n\to\infty.$$

Supposing that there are no errors in the operator A and the exact solution $\bar{u}$ exists, define the feasible set as follows

$$U_n = \{u\in U\colon \quad f_n^l\leqslant_F Au\leqslant_F f_n^u\}.$$

Decomposing $u\in U$ as $u=u_0+u^\perp$ with $u^\perp\in\ker R$, suppose

$$u\in U_n \implies \|u^\perp\|\leq a \qquad (3.8)$$

for some constant $a>0$, then for $R=\mathrm{TV}$ and $R=\mathrm{TGV}^2$, the sequence

$$u_n = \operatorname*{arg\,min}_{u\in U_n} R(u)$$

converges strongly in $L^1(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))$ to the exact solution $\bar{u}$ and $R(u_n)\to R(\bar{u})$.

Proof. With the decomposition $u_n = u_{0,n} + u_n^\perp$, where $u_n^\perp \in \ker R$, we have $u_{0,n} \in \mathrm{lev}_a R$ for suitably large $a > 0$ through

$$R(u_{0,n}) = R(u_n) = \min_{u' \in U_n} R(u') \leq R(\bar{u}).$$

The assumption (3.8) bounds $\|u_n^\perp\| \leq a$. Thus $u_n \in V$ for V as in Proposition 3.1. The proposition thus implies the necessary compactness in $U = L^1(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$ for the application of Theorem 2.3. $\square$

Remark 3.1. The condition (3.8) simply says for $R = \mathrm{TV}$ that the data has to bound the solution in mean. This is very reasonable to expect for practical data; anything else would be very non-degenerate. For $R = \mathrm{TGV}^2$ we also need that the data bounds the entire affine part of the solution. Again, this is very likely for real data. Indeed, in DTI practice, with at least 6 independent diffusion sensiting gradients, A is an invertible or even over-determined linear operator. In that typical case, the bounds f_n^l and f_n^u will be translated into U_n being a bounded set.

4 Solving the optimisation problem

4.1 The Chambolle–Pock method

The Chambolle–Pock algorithm is an inertial primal-dual backward-backward splitting method, classified in [19] as the modified primal-dual hybrid gradient method (PDHGM). It solves the minimax problem

$$\min_x \max_y \ G(x) + \langle Kx, y \rangle - F^*(y), \tag{4.1}$$

where $G : X \to \overline{\mathbb{R}}$ and $F^* : Y \to \overline{\mathbb{R}}$ are convex, proper, lower semicontinuous functionals on (finite-dimensional) Hilbert spaces X and Y. The operator $K \colon X \to Y$ is linear, although an extension of the method to nonlinear K has recently been derived [55]. The PDHGM can also be seen as a preconditioned ADMM (alternating directions method of multipliers); we refer to [19, 47, 56] for reviews of optimisation methods popular in image processing. For step sizes $\tau, \sigma > 0$, and an over-relaxation parameter $\omega > 0$, each iteration of the algorithm consists of the updates

$$u_{i+1} := (I + \tau \partial G)^{-1}(u_i - \tau K^* y_i), \tag{4.2a}$$
$$\bar{u}_{i+1} := u_{i+1} + \omega(u_{i+1} - u_i), \tag{4.2b}$$
$$y_{i+1} := (I + \sigma \partial F^*)^{-1}(y_i + \sigma K \bar{u}_{i+1}). \tag{4.2c}$$

We should remark that the order of the primal (u) and dual (y) updates here is reversed from the original presentation in [13]. The reason is that

when reordered, the updates can, as discovered in [27], be easily written in a proximal point form.

The first and last update are the backward (proximal) steps for the primal (x) and dual (y) variables, respectively, keeping the other fixed. However, the dual step includes some "inertia" or over-relaxation, as specified by the parameter ω. Usually $\omega = 1$, which is required for convergence proofs of the method. If G or F^* is uniformly convex, by smartly choosing for each iteration the step length parameters τ, σ, and the inertia ω, the method can be shown to have convergence rate $O(1/N^2)$. This is similar to Nesterov's optimal gradient method [43]. In the general case the rate is $O(1/N)$. In practice the method produces visually pleasing solutions in rather few iterations, when applied to image processing problems.

In implementation of the method, it is crucial that the resolvents $(I + \tau\partial G)^{-1}$ and $(I + \sigma\partial F^*)^{-1}$ can be computed quickly. We recall that they may be written as

$$(I + \tau\partial G)^{-1}(u) = \arg\min_{u'} \left\{ \frac{\|u' - u\|^2}{2\tau} + G(u') \right\}.$$

Usually in applications, these computations turn out to be simple projections or linear operations – or the soft-thresholding operation for the L_1-norm.

As a further implementation note, since the algorithm (4.2) is formulated in Hilbert spaces (see however [29]), and we work in the Banach space $\mathrm{BV}(\Omega; \mathrm{Sym}^2(\mathbb{R}^3))$, we have to discretise our problems before application of the algorithm. We do this by simple forward-differences discretisation of the operator E with cell width $h = 1$ on a regular rectangular grid corresponding to the image voxels.

4.2 Implementation of deterministic constraints

We now reformulate the problem (3.5) of DTI imaging with deterministic error bounds in the form (4.1). Suppose we have some upper and lower bounds $s_j^l \leqslant s_j \leqslant s_j^u$ on the DWI signals s_j, $(j = 0, \ldots, N)$. Then the bounds for $g_j = \log(s_j/s_0)$ are

$$g_j^l = \log(s_j^l/s_0^u); \quad g_j^u = \log(s_j^u/s_0^l), \quad (j = 1, \ldots, N), \tag{4.3}$$

because g_j is monotone in regards to s_j. We are thus trying to solve

$$u = \arg\min_{u' \in U = \cap U^j} R(u') \tag{4.4}$$

where

$$U^j = \{u : g_j^l \leqslant A_j u \leqslant g_j^u\}$$

For the ease of notation, we write

$$g = \begin{pmatrix} g_1, & \ldots, & g_N \end{pmatrix}, \quad \text{and}$$
$$Au = \begin{pmatrix} A_1 u_1, & \ldots, & A_N u_N \end{pmatrix}, \tag{4.5}$$

so that the Stejskal-Tanner equation is satisfied with

$$Au = g.$$

The problem (4.4) may be rewritten as

$$\min_{u'} \; F_0(Au') + R(u'),$$

for

$$F_0(y) = \delta(g^l \leqslant y \leqslant g^r)$$

with $\delta(g^l \leqslant y \leqslant g^u)$ denoting the indicator function of the convex set $\{y \colon g^l \leqslant y \leqslant g^r\}$. Solving the conjugate

$$F_0^*(y) = \begin{cases} \langle g^l, y \rangle, & y < 0 \\ \langle g^u, y \rangle, & y \geq 0. \end{cases},$$

and also writing

$$R(u) = R_0(K_0 u)$$

for some R_0 and K_0, the problem can further be written in the saddle point form (4.1) with

$$G(u) = 0,$$
$$K = \begin{pmatrix} A \\ K_0 \end{pmatrix},$$
$$F^*(y, \psi) = F_0^*(y) + R_0^*(\psi).$$

To apply algorithm (4.2), we need to compute the resolvents of G_0^* and R_0^*. For details regarding R_0^* for $R = \mathrm{TGV}_{(\beta,\alpha)}^2$ and $R = \alpha \, \mathrm{TV}$ in the discretised setting, we refer to [57, 58]; here it suffices to note that for $R = \alpha \, \mathrm{TV}$, we have $K_0 = \alpha E$ and $R_0^*(\phi)$ is the indicator function of the dual ball $\{\phi \mid \sup_x \|\phi(x)\|_F \leq 1\}$. Thus the resolvent $(I + \tau \partial R_0^*)^{-1}$ becomes a projection to the dual ball. The case of $R = \mathrm{TGV}_{(\beta,\alpha)}^2$ is similar. For F_0^* we have

$$(I + \tau \partial F_0^*)^{-1}(y) = \arg\min_{y'} F_0^*(y') + \frac{|y - y'|^2}{2\tau},$$

which resolves pointwise at each $\xi \in \Omega$ into the expression

$$[(I + \tau \partial F_0^*)^{-1}(y)](\xi) = S(y(\xi))$$

for

$$S(y(\xi)) = \begin{cases} y(\xi) - g^u(\xi)\sigma, & y(\xi) \geq g^u(\xi)\sigma, \\ 0, & g^u(\xi)\sigma \leq y(\xi) \leq g^l(\xi)\sigma, \\ y(\xi) - g^l(\xi)\sigma, & y(\xi) \leq g^l(\xi)\sigma. \end{cases}$$

Finally, we note that the saddle point system (4.1) has to have a solution for the Chambolle–Pock algorithm to converge. In our setting, in particular, we need to find error bounds g^l and g^u, for which there exists a solution u to

$$g^l \leqslant Au \leqslant g^u. \tag{4.6}$$

If one uses at most six independent diffusion directions ($N = 6$), as we will, then, for any g, there in fact exists a solution to $g = Au$. The condition (4.6) becomes $g^l \leqslant g^u$, immediately guaranteed through the monotonicity of (4.3), and the trivial conditions $s_j^l \leqslant s_j^u$.

We are thus ready to apply the algorithm (4.2) to diffusion tensor imaging with deterministic error bounds. For the realisation of (4.2) for models (3.3), (3.4), and (3.2), we refer to [57, 58, 59, 55].

5 Experimental results

We now study the efficiency of the proposed reconstruction model in comparison to the different L^2-squared models, i.e., ones with Gaussian error assumption. This is based on a synthetic data, for which a ground-truth is available, as well as a real in vivo DTI data set. First we, however, have to describe in detail the procedure for obtaining upper and lower bounds for real data, when we do not know the true noise model, and are unable to perform multiple experiments as required by the theory of Section 2.2.

5.1 Estimating lower and upper bounds from real data

As we have already discussed, in practice the noise in the measurement signals $\hat{s}_j$ is not Gaussian or Rician; in fact we do not know the true noise distribution and other corruptions. Therefore, we have to estimate the noise distribution from the image background. To do this, we require a known correspondence between the measurement, the noise, and the true value. As we have no better assumptions available, the standard one that we use is that of additive noise. Continuing in the statistical setting of Section 2.2, we now describe the procedure, working on discrete images expressed as vectors $\hat{f} = s_j \in \mathbb{R}^n$ for some fixed $j \in \{0, 1, \ldots, N\}$. We use superscripts to denote the voxel indices, that is $\hat{f} = (f^1, \ldots, f^n)$.

In the i-th voxel, the measured value $\hat{f}^i$ is the sum of the true value f^i and additive noise ν^i:

$$\hat{f}^i = f^i + \nu^i.$$

All ν^i are assumed independent and identically distributed (i.i.d.), but their distribution is unknown. If we did know the true underlying cumulative distribution function F of the noise, we could choose a confidence parameter $\theta \in (0,1)$ and use the cumulative distribution function to calculate $\nu_{\theta/2}, \nu_{1-\theta/2}$ such that[1]

$$P(\nu_{\theta/2} \leqslant \nu^i \leqslant \nu_{1-\theta/2}) = 1 - \theta. \tag{5.1}$$

Let us instead proceed non-parametrically, and divide the whole image into two groups of voxels - the ones belonging to the background region and the rest. For simplicity, let the indices $i = 1, \ldots, k$, $(k < n)$, stand for the background voxels. In this region, we have $f^i = 0$ and $\hat{f}^i = \nu^i$. Therefore, the background region provides us with a number of independent samples from the unknown distribution of ν. The Dvoretzky–Kiefer–Wolfowitz inequality [18, 41, 38] states that

$$P\big(\sup_t |F(t) - F_k(t)| > \epsilon\big) \leq 2e^{-2k\epsilon^2},$$

for the empirical cumulative distribution function

$$F_k(t) := \frac{1}{k} \sum_{i=1}^{k} \chi_{(-\infty, \hat{f}^i]}(t).$$

Therefore, computing $\nu_{\theta/2}$ and $\nu_{1-\theta/2}$ such that

$$P_k(\nu_{\theta/2} \leqslant \nu \leqslant \nu_{1-\theta/2}) = 1 - \theta,$$

we also have

$$P_k(f^i + \nu_{\theta/2} \leqslant \hat{f}^i \leqslant f^i + \nu_{1-\theta/2}) = 1 - \theta.$$

We may therefore use the values

$$\hat{f}^{l,i} = \hat{f}^i - \nu_{1-\theta/2}, \quad \hat{f}^{u,i} = \hat{f}^i - \nu_{\theta/2}, \tag{5.2}$$

as lower and upper bounds for the true values f^i outside the background region.

The Dvoretzky–Kiefer–Wolfowitz inequality implies that the interval estimates converge to the true intervals, determined by (5.1), as the number of background pixels k increases with the image size n. This procedure, with large k, will therefore provide an estimate of a single-experiment ($m = 1$) confidence interval for f^i. We note that this procedure will, however, not yield the convergence of the interval estimate $[\hat{f}^{l,i}, \hat{f}^{u,i}]$ to the true data; for that we would need multiple experiments, i.e., multiple sample images ($m > 1$), not just agglomeration of the background voxels into a single noise distribution estimate. In practice, however, we can only afford a single experiment ($m = 1$), and cannot go to the limit.

---

[1]Recall that for a random variable X with a cumulative distribution function F, the quantile function F^{-1} returns a number $x_\theta = F^{-1}(\theta)$ such that $P(X \leqslant x_\theta) = \theta$.

5.2 Verification of the approach with synthetic data

To verify the effectiveness of the considered approach and to compare it to the other models, we use synthetic data. For the ground-truth tensor field $u_{g.t.}$ we take a helix region in a 3D box $100 \times 100 \times 30$, and choose the tensor in each point inside the helix in such a way that the principal eigenvector coincides with the helix direction (Figure 2). The helix region is described by the following equations:

$$
\begin{aligned}
x &= (R + r\cos(\theta))\cos(\phi), \\
y &= (R + r\cos(\theta))\sin(\phi), \\
z &= r\sin(\theta) + \phi/\phi_{\max}, \\
\phi &\in [0, \phi_{\max}], \quad r \in [0, r_{\max}], \quad \theta \in [0, 2\pi].
\end{aligned}
$$

The vector direction in every point coincides with helix direction:

$$
\vec{r} = \begin{pmatrix} -R\sin(\phi) \\ R\cos(\phi) \\ 1/\phi_{\max} \end{pmatrix}.
$$

We take the parameters $R = 0.3, \phi_{\max} = 4\pi, r_{\max} = 0.07$ in this numerical experiment.

We apply the forward operators $T_j(u_{g.t.})$, $(j = 0, \ldots, 6)$, to obtain the data $s_j(x)$. We then add Rician noise to this data $\bar{s}_j = s_j + \delta$ with $\sigma = 2$, which corresponds to PSNR ≈ 27dB.

We apply several models for solving the inverse problem of reconstructing u: the linear and non-linear L^2 approaches (3.3) and (3.2), and the constrained problem (3.5). As the regulariser we use $R = \text{TGV}^2_{(0.9\alpha,\alpha)}$, where the choice $\beta = 0.9\alpha$ was made somewhat arbitrarily, however yielding good results for all the models. This is slightly lower than the range $[1, 1.5]\alpha$ discovered in comprehensive experiments for other imaging modalities [7, 16].

For the linear and non-linear L^2 models (3.3) and (3.2), respectively, the regularisation parameter α is chosen either by a version of the discrepancy principle for inconsistent problems [52] or optimally with regard to the $\|\cdot\|_{F,2}$ distance between the solution and the ground-truth. In case of the discrepancy principle, such an α was chosen that the following equality holds:

$$
\Delta\rho(\alpha) = \sum_j \|T_j(u) - \bar{s}_j\|^2 - \tau \sum_j \|\bar{s}_j - s_j\|^2 = 0 \tag{5.3}
$$

We find α by solving this equation numerically using bisection method. We start by finding such α_1, α_2 that $\Delta\rho(\alpha_1) > 0$ and $\Delta\rho(\alpha_2) < 0$. We calculate $\Delta\rho(\alpha_3), \alpha_3 = \frac{\alpha_1 + \alpha_2}{2}$ and depending on its sign replace either α_1 or α_2 with α_3. We repeat this procedure until the stopping criteria is reached.
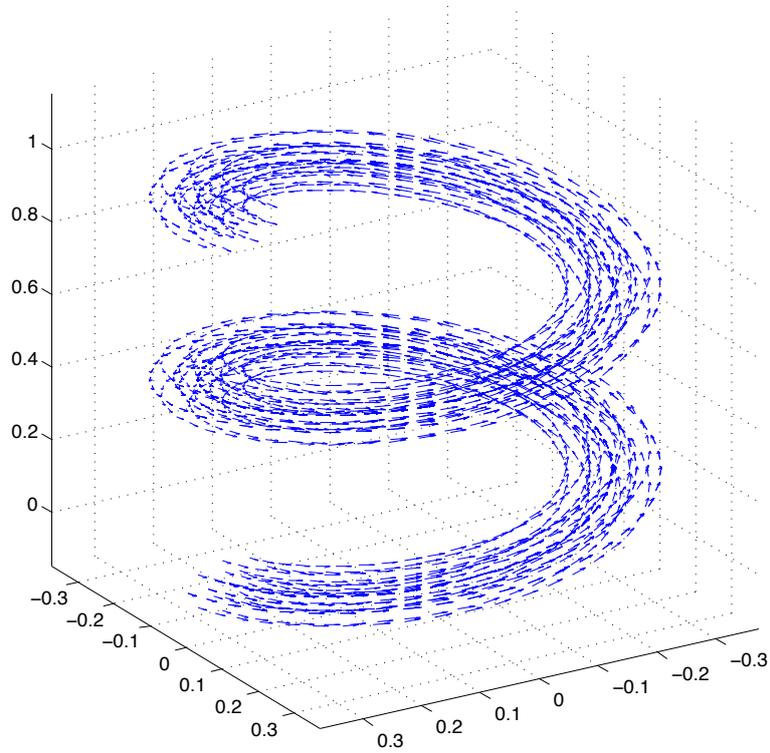
Figure 2: Helix vector field for the principal eigenvector of the ground-truth tensor field

As stopping criteria we use $|f(\alpha)| < \epsilon$. We use $\tau = 1.05$, $\epsilon = 0.01$ for linear and $\tau = 1.2$, $\epsilon = 0.0001$ for non-linear L^2 solution. A value of τ yielding a reasonable degree of smoothness has been chosen by trial and error, and is different for the non-linear model, reflecting a different non-linear objective in the discrepancy principle. For the constrained problem we calculate $\theta = 90\%$, 95%, and 99% confidence intervals to generate the upper and lower bounds. We however digress a little bit from the approach of Section 2.2. Minding that we do not know the true underlying distribution, which fails to be Rician as illustrated in Figure 1, we do not use it to calculate the confidence intervals, but use the estimation procedure described in Section 5.1. We stress that we only have a single sample of each signal s_j, so are unable to verify any asymptotic estimation properties.

The numerical results are in Table 1 and Figures 3–5, with the first of the figures showing the colour-coded principal eigenvector of the reconstruction, the second showing the fractional anisotropy and principal eigenvectors, and the last one the errors in the latter two, in a colour-coded manner. All plots are masked to represent only the non-zero region. The field of

20

Table 1: Numerical results for the synthetic data. For the linear and non-linear L^2 the free parameter chosen by the parameter choice criterion is the regularisation parameter α, and for the constrained problem it is the confidence interval.

Method	Parameter choice	Frobenius PSNR	Pr. e.val. PSNR	Pr. e.vect. angle PSNR
Regression		33.90dB	25.04dB	47.86dB
Linear L^2	Discr. Principle	32.93dB	27.81dB	61.89dB
Linear L^2	Frob. Error-optimal	34.51dB	28.42dB	60.93dB
Non-linear L^2	Discr. Principle	37.33dB	27.81dB	61.89dB
Non-linear L^2	Frob. Error-optimal	37.44dB	28.03dB	61.12dB
Constraints	90%	32.28dB	28.86dB	65.65dB
Constraints	95%	30.97dB	28.14dB	64.80dB
Constrains	99%	27.86dB	24.51dB	61.41dB

fractional anisotropy is defined for a field u of 2-tensors on $\Omega \subset \mathbb{R}^m$ as

$$\text{FA}_u(x) = \left(\sum_{i=1}^{m}(\lambda_i - \bar{\lambda})^2\right)^{1/2}\left(\sum_{i=1}^{m}\lambda_i^2\right)^{-1/2} \in [0,1], \quad (x \in \Omega),$$

with $\lambda_1, \ldots, \lambda_m$ denoting the eigenvalues of $u(x)$. It measures how far the ellipsoid prescribed by the eigenvalues and eigenvectors is from a sphere, with $\text{FA}_u(x) = 1$ corresponding a full sphere, and $\text{FA}_u(x) = 0$ corresponding to a degenerate object not having full dimension.

As we can see, the non-linear approach (3.2) performs overall the best by a wide margin, in terms of the pointwise Frobenius error, i.e., error in $\|\cdot\|_{F,2}$. This is expressed as a PSNR in Table 1. What is, however, interesting, is that the constraint-based approach (3.5) has a much better reconstruction of the principal eigenvector angle, and a comparable reconstruction of its magnitude. Indeed, the 95% confidence interval in Figure 3(g) and Figure 4(g) suggests a nearly perfect reconstruction in terms of smoothness. But, the Frobenius PSNR in Table 1 for this approach is worse than the simple unregularised inversion by regression. The problem is revealed by Figure 5(f): the large white cloudy areas indicate huge fractional anisotropy errors, while at the same time, the principal eigenvector angle errors expressed in colour are much lower than for other approaches. Good reconstruction of the principal eigenvector is important for the process of tractography, i.e., the reconstruction of neural pathways in a brain. One explanation for our good results is that the regulariser completely governs the solution in areas where the error bounds are inactive due to generally low errors. This results in very smooth reconstructions, which is in the present case desirable as our synthetic tensor field is also smooth within the helix.

(a) Ground-truth (b) Regression result (c) Linear L^2, discrepancy principle (d) Linear L^2, error-optimal

(e) Non-linear L^2, discrepancy principle (f) Non-linear L^2, error-optimal (g) Constrained, 95% confidence intervals (h) Constrained, 90% confidence intervals
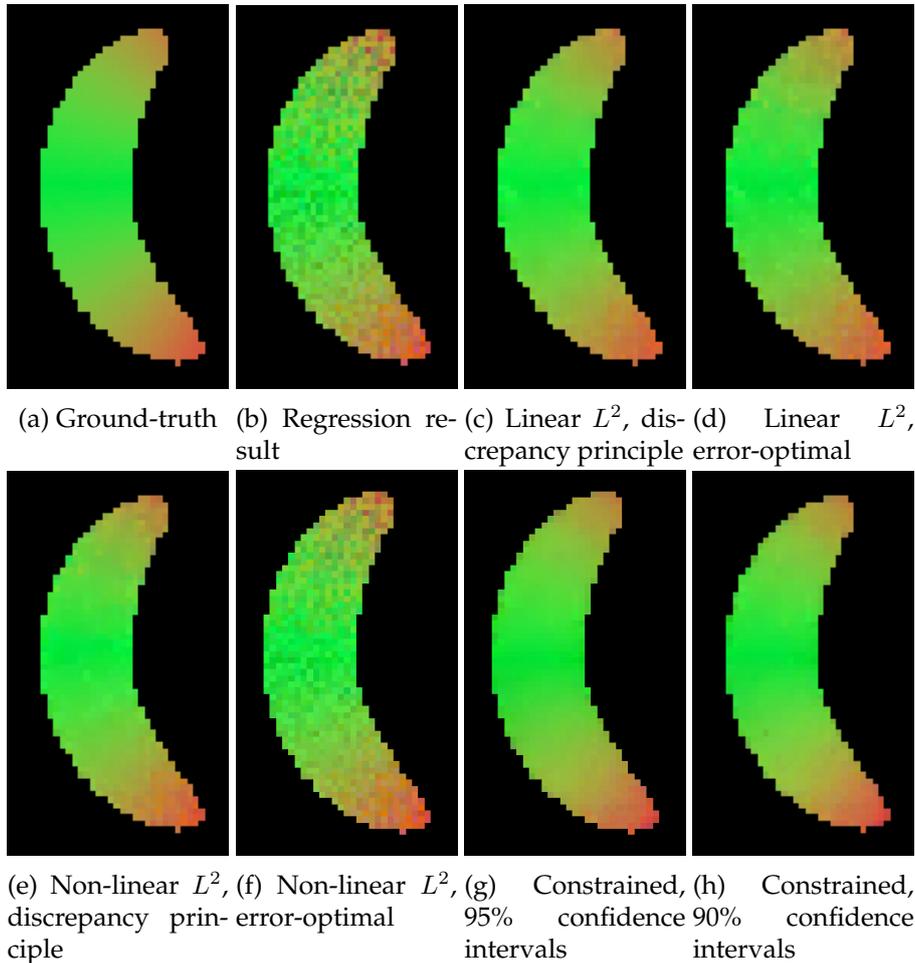
Figure 3: Synthetic test data results. (a) ground-truth plot. (b) regression result plot. (c)–(h) Plot of a slice of the solution for L^2, non-linear L^2 and constrained models. The legend on the right indicates the colour-coding of directions of the principal eigenvector plotted.
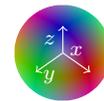
5.3 Results with in vivo brain imaging data

We now wish to study the proposed regularisation model on a real in-vivo diffusion tensor image. Our data is that of a human brain, with the measurements of a volunteer performed on a clinical 3T system (Siemens Magnetom TIM Trio, Erlangen, Germany), with a 32 channel head coil. A 2D diffusion weighted single shot EPI sequence with diffusion sensitising gradients applied in 12 independent directions ($b = 1000s/mm^2$). An additional reference scan without diffusion was used with the parameters: TR $= 7900$ms, TE $= 94$ms, flip angle $90°$. Each slice of the 3D data set

(a) Ground-truth (b) Regression re-(c) Linear L^2, dis-(d) Linear L^2, sult crepancy principle error-optimal

(e) Non-linear L^2, (f) Non-linear L^2, (g) Constrained, (h) Constrained, discrepancy prin-error-optimal 95% confidence 90% confidence ciple intervals intervals
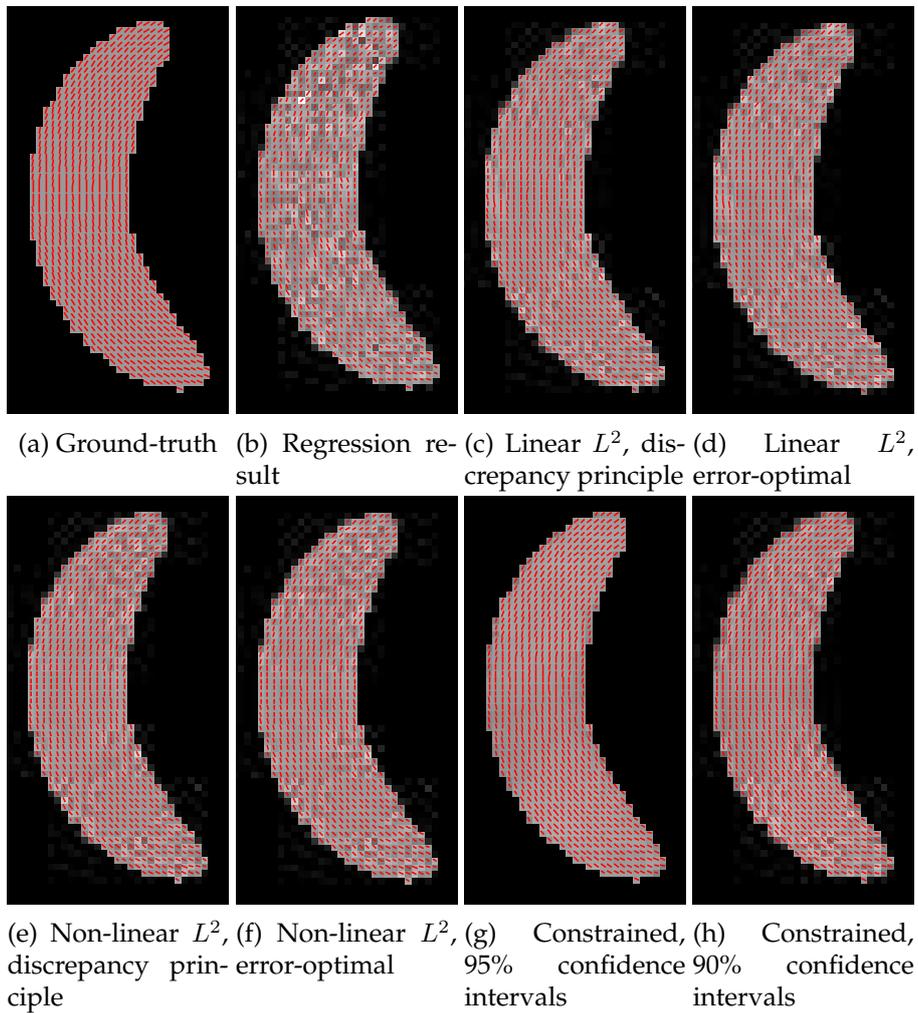
Figure 4: Fractional anisotropy in greyscale superimposed by principal eigenvector. Legend on left indicates the greyscale intensities of the fractional anisotropy.

has plane resolution 1.95mm $\times$ 1.95mm, with a total of 128×128 pixels. The total number of slices is 60 with a slice thickness of 2mm. The data set consists of 4 repeated measurements. The GRAPPA acceleration factor is 2. Prior to the reconstruction of the diffusion tensor, eddy current correction was performed with FSL [50]. Written informed consent was obtained from the volunteer before the examination.

For error bounds calculation according to the procedure of Section 5.1, to avoid systematic bias near the brain, we only use about 0.6% of the total volume near the borders, or roughly $k \approx 6000$ voxels.

To estimate errors for the all the considered reconstruction models, for
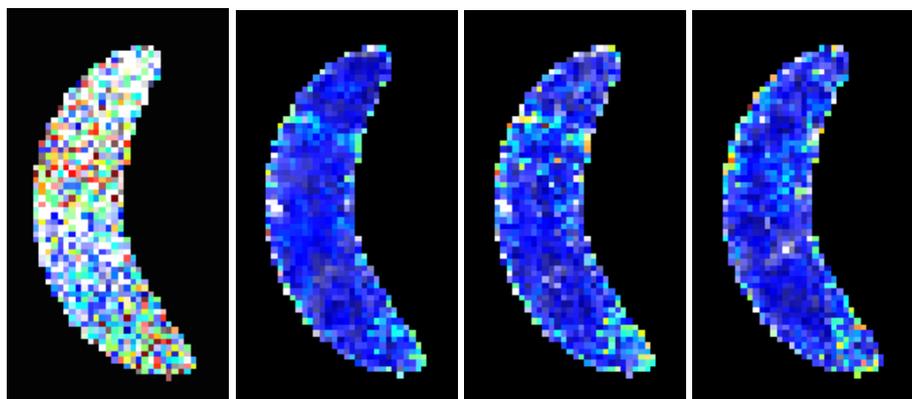
(a) Regression re-
sult

(b) Linear L^2, dis-
crepancy principle

(c) Linear L^2,
error-optimal

(d) Non-linear L^2,
discrepancy prin-
ciple
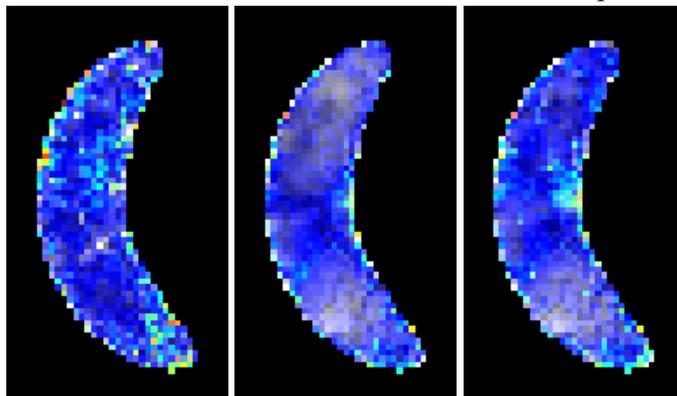
(e) Non-linear L^2,
error-optimal

(f) Constrained,
95% confidence
intervals

(g) Constrained,
90% confidence
intervals

Figure 5: Colour-coded errors of fractional anisotropy and
principal eigenvector for the computations on the synthetic
test data. Legend on the right indicates the colour-coding
of errors between u and g_0 as functions of the principal
eigenvector angle error $\theta = \cos^{-1}(\langle \hat{v}_u, \hat{v}_{g_0} \rangle)$ in terms of the
hue, and the fractional anisotropy error $e_{\text{FA}} = |\text{FA}_u - \text{FA}_{g_0}|$
in terms of whiteness.
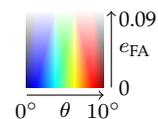
each gradient direction b_i we use only one out of the four duplicate mea-
surements. We then calculate the errors using a somewhat less than ideal
pseudo-ground-truth, which is the linear regression reconstruction from all
the available measurements.

The results are in Table 2 and Figures 6–8, again with the first of the
figures showing the colour-coded principal eigenvector of the reconstruc-

Table 2: Numerical results for the in-vivo brain data data. For the L^2 and non-linear L^2 reconstruction models the free parameter chosen by the parameter choice criterion is the regularisation parameter α, and for the constrained problem it is the confidence interval.

Method	Parameter choice	Frobenius PSNR	Pr. e.val. PSNR	Pr. e.vect. angle PSNR
Regression		32.35dB	33.67dB	28.56dB
Linear L^2	Discr. Principle	34.80dB	36.35dB	24.81dB
Linear L^2	Frob. Error-optimal	34.81dB	36.32dB	24.97dB
Non-linear L^2	Discr. Principle	33.53dB	35.87dB	27.12dB
Non-linear L^2	Frob. Error-optimal	33.57dB	36.03dB	27.58dB
Constraints	90%	33.71dB	34.93dB	27.00dB
Constraints	95%	33.70dB	34.97dB	26.91dB
Constraints	99%	33.67dB	34.89dB	26.88dB

tion, the second showing the fractional anisotropy and principal eigenvectors, and the last one the errors in the latter two, in a colour-coded manner. Again, all plots are masked to represent only the non-zero region. In the figures, we concentrate on error bounds based on 95% confidence intervals, as the results for the 90% and 99% cases do not differ significantly according to Table 2.

This time, the linear L^2 approach (3.3) has best overall reconstruction (Frobenius PSNR), while the nonlinear L^2 approach (3.2) has clearly the best principal eigenvector angle reconstruction besides the regression, which does not seem entirely reliable regarding our regression-based pseudo-ground-truth. The constraints based approach (3.5), with 95% confidence intervals is, however, not far behind in terms of numbers. More detailed study of the *corpus callosum* in Figure 8 (small picture in picture) and Figure 7 however indicates a better reconstruction of this important region by the nonlinear approach. The constrained approach has some very short vectors there in the white region. Naturally, however, these results on the in vivo data should be taken with a grain of salt, as we have only a somewhat unreliable pseudo-ground-truth available for comparison purposes.

5.4 Conclusions from the numerical experiments

Our conclusion is that the error bounds based approach is a feasible alternative to standard modelling with incorrect Gaussian assumptions. It can produce good reconstructions, although the non-linear L^2 approach of [55] is possibly slightly more reliable. The latter does, however, in principle depend on a good initialisation of the optimisation method, unlike the convex bounds based approach.

Further theoretical work will be undertaken to extend the partial-order-based approach to modelling errors in linear operators to the non-lattice

(a) Pseudo-ground-truth

(b) Regression result

(c) Linear L^2, discrepancy principle

(d) Linear L^2, error-optimal

(e) Non-linear L^2, discrepancy principle

(f) Non-linear L^2, error-optimal

(g) Constrained, 95% confidence intervals
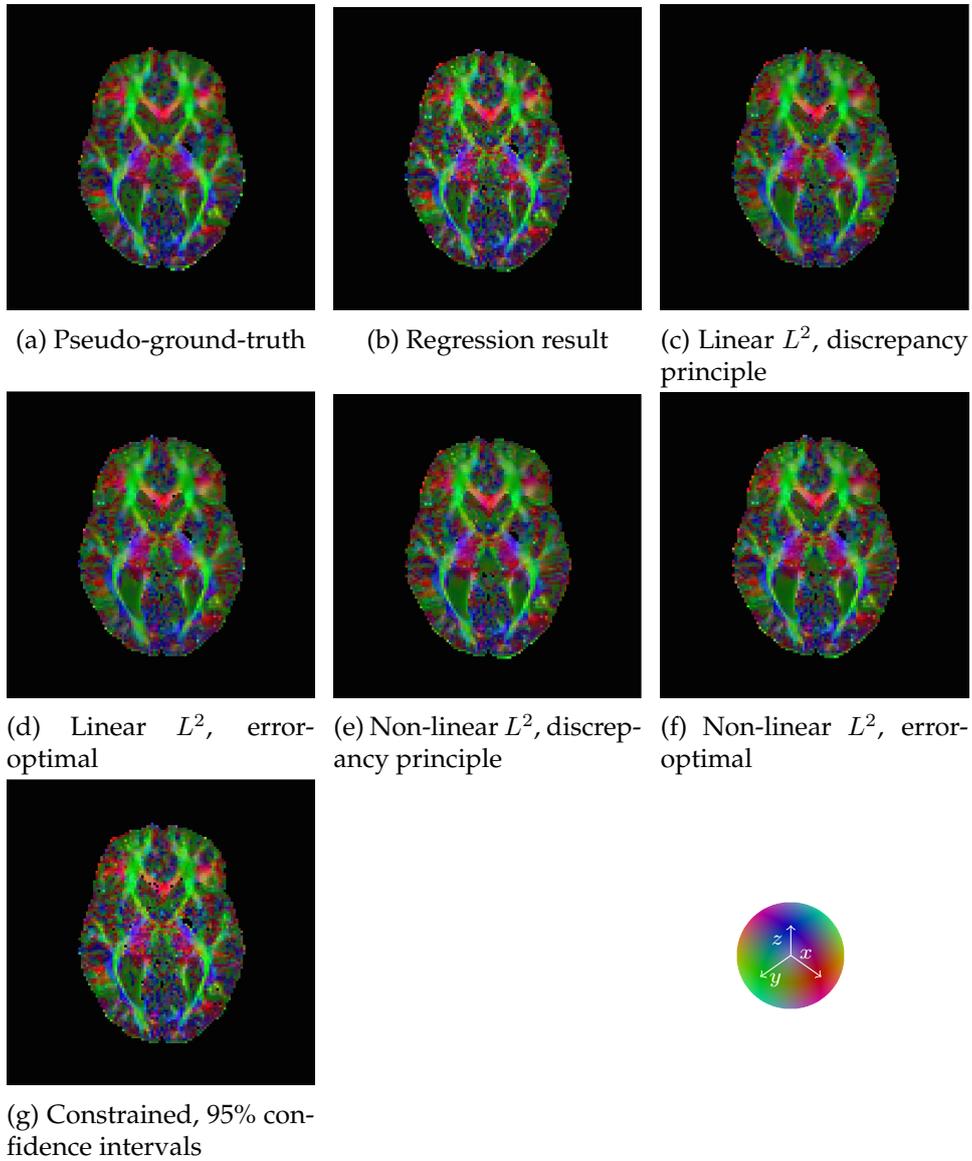
Figure 6: Reconstruction results on the in vivo brain data. (a) Pseudo-ground-truth plot. (b) regression result. (c)–(g) Plot of a slice of the solution for L^2, non-linear L^2 and constrained problem approach. The legend on the bottom-right indicates the colour-coding of directions of the principal eigenvector plotted.

case of the semidefinite partial order for symmetric matrices, which will allow us to consider problems of diffusion MRI with errors in the forward operator.

It shall also be investigated whether the error bounds approach needs

(a) Pseudo-ground-truth (b) Linear L^2, discrepancy principle (c) Linear L^2, error-optimal

(d) Non-linear L^2, discrepancy principle (e) Non-linear L^2, error-optimal (f) Constrained, 95% confidence intervals
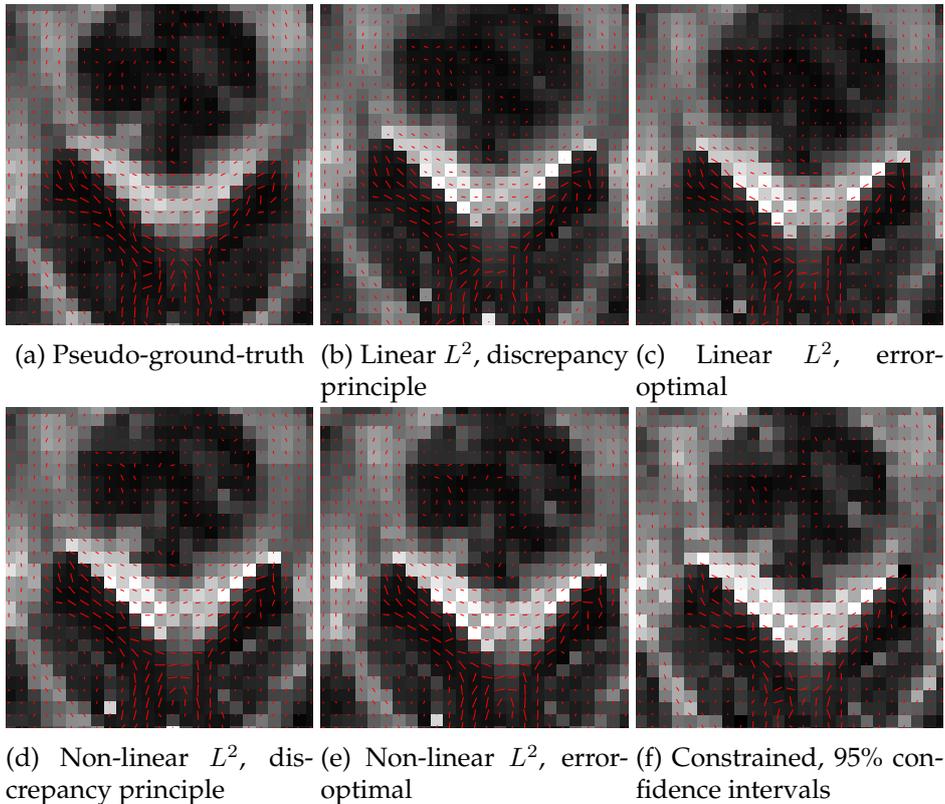
Figure 7: Fractional anisotropy of the *corpus callosum* in greyscale superimposed by principal eigenvector. Legend on the right indicates the greyscale intensities of the fractional anisotropy.

to be combined with an alternative, novel, regulariser that would ameliorate the fractional anisotropy errors that the approach exhibits. It is important to note, however, that from the practical point of view, of using the reconstruction tensor field for basic tractography methods based solely on principal eigenvectors, these are not that critical. As pointed out by one of the reviewers, the situation could differ with more recent geodesic tractography methods [26, 22, 21] employing the full tensor. We provide basic principal eigenvector tractography results for reference in Figure 9, without attempting to extensively interpret the results. It suffices to say that the results look comparable. With this in sight, the error bounds approach produces a very good reconstruction of the direction of the principal eigenvectors, although we saw some problems with the magnitude within the *corpus callosum*.

(a) Linear L^2, discrepancy principle (b) Linear L^2, error-optimal (c) Non-linear L^2, discrepancy principle

(d) Non-linear L^2, error-optimal (e) Constrained, 95% confidence intervals
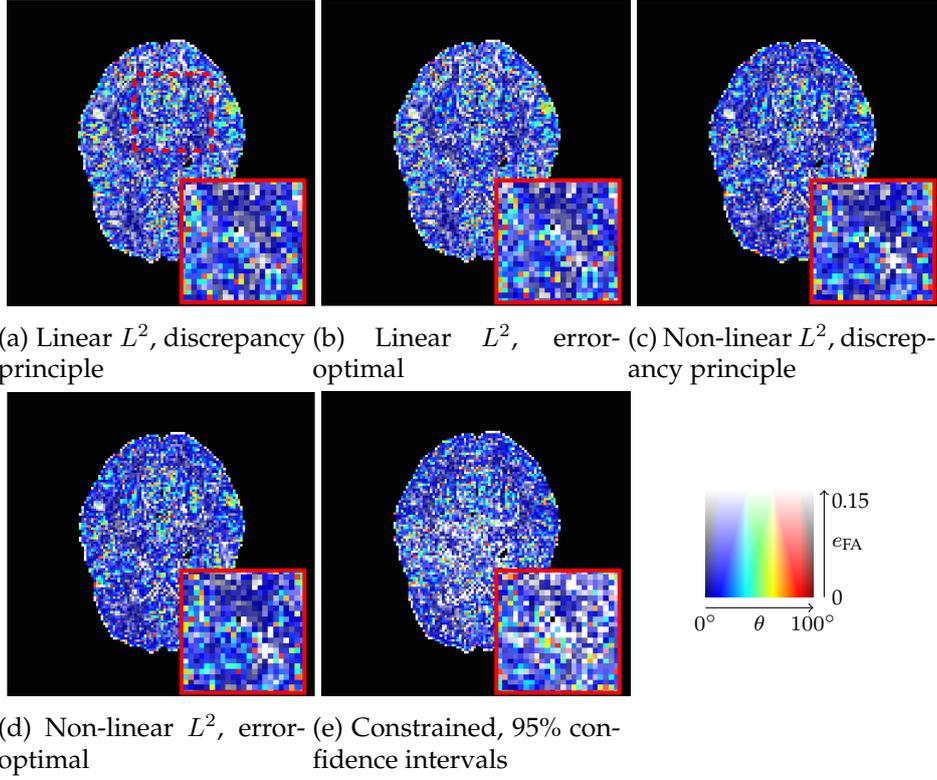
Figure 8: Colour-coded errors of fractional anisotropy and principal eigenvector for the computations on the synthetic test data. Legend on the right indicates the colour-coding of errors between u and g_0 as functions of the principal eigenvector angle error $\theta = \cos^{-1}(\langle \hat{v}_u, \hat{v}_{g_0} \rangle)$ in terms of the hue, and the fractional anisotropy error $e_{\text{FA}} = |\text{FA}_u - \text{FA}_{g_0}|$ in terms of whiteness.

Acknowledgements

(a) Pseudo-ground-truth (b) Regression result (c) Linear L^2, discrepancy principle

(d) Non-linear L^2, discrepancy principle (e) Non-linear L^2, error-optimal (f) Constrained, 95% confidence intervals

Figure 9: Tractography visualization results on the in vivo brain data. (a) Pseudo-ground-truth plot. (b) regression result. (c)–(f) Plot of a slice of the solution for L^2, non-linear L^2 and constrained problem approach.

ages, and Florian Knoll for many inspiring discussions.

A data statement for the EPSRC

The MRI scans used for this publication have been used by the courtesy of the producer. Data that is similar for all intents and purposes, is widespread, and can be easily produced by making a measurement of a human subject with an MRI scanner. Our source codes are archived at `https://www.repository.cam.ac.uk/handle/1810/253422`.

Appendix: Notation and techniques

We recall some basic, not completely standard, mathematical notation and concepts in this appendix. We begin with partially ordered vector spaces, following the book [46]. Then we proceed to tensor calculus and *tensor fields of bounded variation* and of *bounded deformation*. These are also covered in more detail for the diffusion tensor imaging application in [57].

Banach lattices

A linear space X, endowed with a partial order relation $\leqslant$ is called an ordered vector space if the partial order agrees with linear operations in the following way:

$$x \leqslant y \quad \implies \quad x + z \leqslant y + z \qquad \forall x, y, z \in X,$$
$$x \leqslant y \quad \implies \quad \lambda x \leqslant \lambda y \qquad \forall x, y \in X \text{ and } \lambda \in \mathbb{R}_+.$$

An ordered vector space is called a vector lattice if each pair of elements $x, y \in X$ have a supremum $x \vee y \in X$ and infimum $x \wedge y \in X$. Supremum of two elements x, y of a Banach lattice X is the element $z = x \vee y$ with the following properties: $z \geqslant x$, $z \geqslant y$ and $\forall \tilde{z} \in X$ such that $\tilde{z} \geqslant x$ and $\tilde{z} \geqslant y$ we have $\tilde{z} \geqslant z$.

For any $x \in X$, the element $x_+ = x \vee 0$ is called its positive part, the element $x_- = (-x) \vee 0 = (-x)_+$ is called its negative part, the element $|x| = x_+ + x_-$ is its absolute value. The equalities $x = x_+ - x_-$ and $|x| = x \vee (-x)$ hold for any $x \in X$.

It is obvious that suprema and infima exist for any finite number of elements of a vector lattice. A vector lattice X is said to be order complete if *any* bounded from above set in X has a supremum.

Let X and Y be ordered vector spaces. A linear operator $U \colon X \to Y$ is called positive, if $x \geqslant_X 0$ implies $Ux \geqslant_Y 0$. An operator U is called regular, if it can be written as $U = U_1 - U_2$, where U_1 and U_2 are positive operators.

Denote the linear space of all regular operators $X \to Y$ by $L^\sim(X, Y)$. A partial order can be introduced in $L^\sim(X, Y)$ in the following way: $U_1 \geqslant U_2$, if $U_1 - U_2$ is a positive operator. If X and Y are vector lattices and Y is order complete, then $L^\sim(X, Y)$ is also an order complete vector lattice.

A norm $\| \cdot \|$ defined in a vector lattice X is called monotone if $|x| \leqslant |y|$ implies $\|x\| \leqslant \|y\|$. A vector lattice endowed with a monotone norm is called a Banach lattice if it is norm complete. If X and Y are Banach lattices, then all operators in $L^\sim(X, Y)$ are continuous.

Let us list some examples of Banach lattices. The space of continuous functions $C(\Omega)$, where $\Omega \subset \mathbb{R}^n$, is a Banach lattice under the natural pointwise ordering: $f \geqslant_C g$ if and only if $f(x) \geqslant g(x)$ for all $x \in \Omega$. The spaces $L^p(\Omega)$, $1 \leqslant p \leqslant \infty$, are also Banach lattices under the following partial ordering: $f \geqslant_{L^p} g$ if and only if $f(x) \geqslant g(x)$ almost everywhere in Ω. With this partial order, $L^p(\Omega)$, $1 \leqslant p \leqslant \infty$, are order complete Banach lattices. The Banach lattice of continuous functions $C(\Omega)$ is not order complete.

Tensors in the Euclidean setting

We call a k-linear mapping $A : \mathbb{R}^m \times \cdots \times \mathbb{R}^m \to \mathbb{R}$ a k-tensor, denoted $A \in \mathcal{T}^k(\mathbb{R}^m)$. This is a simplification from the full differential-geometric definition, sufficient for our finite-dimensional setting. We say that A is

symmetric, denoted $A \in \mathrm{Sym}^k(\mathbb{R}^m)$, if it satisfies for any permutation π of $\{1, \dots, k\}$ that

$$A(c_{\pi 1}, \dots, c_{\pi k}) = A(c_1, \dots, c_k).$$

With $e_1, \dots, e_m$ the standard basis of $\mathbb{R}^m$, we define on $\mathcal{T}^k(\mathbb{R}^m)$ the inner product

$$\langle A, B \rangle := \sum_{p \in \{1, \dots, m\}^k} A(e_{p_1}, \dots, e_{p_k}) B(e_{p_1}, \dots, e_{p_k}),$$

and the Frobenius norm

$$\|A\|_F := \sqrt{\langle A, A \rangle}.$$

The Frobenius norm is rotationally invariant in a sense crucial for DTI. We refer to [57] for a detailed discussion of this, as well of alternative rotationally invariant norms.

Example .1 (Vectors). Vectors $A \in \mathbb{R}^m$ are of course symmetric 1-tensors, The inner product is the usual inner product in $\mathbb{R}^m$, and the Frobenius norm is the two-norm, $\|A\|_F = \|A\|_2$.

Example .2 (Matrices). Matrices are 2-tensors: $A(x, y) = \langle Ax, y \rangle$, while symmetric matrices $A = A^T$ are symmetric 2-tensors. The inner product is $\langle A, B \rangle = \sum_{i,j} A_{ij} B_{ij}$ and $\|A\|_F$ is the matrix Frobenius norm.

We use the notation $A \geq 0$ for positive-semidefinite matrices A. One can verify that this relation indeed defines a partial order in the space of symmetric matrices:

$$A \geq B \quad \text{iff } A - B \text{ is positive semidefinite}. \tag{.4}$$

With this partial order, the space of all symmetric matrices becomes an ordered vector space, but not a vector lattice. However, it enjoys some properties similar to those of vector lattices: for example, any directed upwards subset[2] has a supremum [39, Ch.8].

Symmetric tensor fields of bounded deformation

Let $u : \Omega \to \mathrm{Sym}^k(\mathbb{R}^m)$ for a domain $\Omega \subset \mathbb{R}^m$. We set

$$\|u\|_{F,p} := \left(\int_\Omega \|u(x)\|_F^p \, \mathrm{d}\, x \right)^{1/p} \quad (p \in [1, \infty)),$$

and

$$\|u\|_{F,\infty} := \mathrm{ess\,sup}_{x \in \Omega} \|u(x)\|_F,$$

---

[2]Recall that an indexed subset $\{x_\tau : \tau \in \{\tau\}\}$ of an ordered vector space X is called directed upwards if for any pair $\tau_1, \tau_2 \in \{\tau\}$ there exists $\tau_3 \in \{\tau\}$ such that $x_{\tau_3} \geqslant x_{\tau_1}$ and $x_{\tau_3} \geqslant x_{\tau_2}$.

The spaces $L^p(\Omega; \operatorname{Sym}^k(\mathbb{R}^m))$ are defined in the natural way using these norms, and clearly satisfy all the usual properties of L^p spaces.

In the particular case of matrices ($k = 2$), partial order can be introduced in the space $L^p(\Omega; \operatorname{Sym}^2(\mathbb{R}^m))$ in the following way:

$$u \geqq v \quad \text{iff } u(x) \geq v(x) \text{ almost everywhere in } \Omega. \tag{.5}$$

Recall that the symbol $\geq$ stands for the positive semidefinite order (.4) in the space of symmetric matrices. Since the positive semidefinite order is not a lattice, neither is the order (.5).

If $u \in C^1(\Omega; \operatorname{Sym}^k(\mathbb{R}^m))$, $k \geq 1$, we define by contraction the divergence $\operatorname{div} u \in C(\Omega; \operatorname{Sym}^{k-1}(\mathbb{R}^m)$ as

$$[\operatorname{div} u(\cdot)](e_{i_2}, \ldots, e_{i_k}) := \sum_{i_1=1}^{m} \langle e_{i_1}, \nabla u(\cdot)(e_{i_1}, \ldots, e_{i_k}) \rangle. \tag{.6}$$

It is easily verified that $\operatorname{div} u(x)$ is indeed symmetric. Given a tensor field $u \in L^1(\Omega; \mathcal{T}^k(\mathbb{R}^m))$ we then define the *symmetrised distributional gradient* $Eu \in [C_c^\infty(\Omega; \operatorname{Sym}^{k+1}(\mathbb{R}^m))]^*$ by

$$Eu(\varphi) := -\int_\Omega \langle u(x), \operatorname{div} \varphi(x) \rangle \, \mathrm{d}\, x, \quad (\varphi \in C_c^\infty(\Omega; \operatorname{Sym}^{k+1}(\mathbb{R}^m))).$$

With these notions at hand, we now define the spaces of *symmetric tensor fields of bounded deformation* as (see also [57, 8])

$$\operatorname{BD}(\Omega; \operatorname{Sym}^k(\mathbb{R}^m)) := \left\{ u \in L^1(\Omega; \operatorname{Sym}^k(\mathbb{R}^m)) \ \Big| \ \sup_{\varphi \in V_{F,\mathrm{s}}^{k+1}} Eu(\varphi) < \infty \right\},$$

where
$$V_{F,\mathrm{s}}^k := \{ \varphi \in C_c^\infty(\Omega; \operatorname{Sym}^k(\mathbb{R}^m)) \mid \|\varphi\|_{F,\infty} \leq 1 \}.$$

For $u \in \operatorname{BD}(\Omega; \operatorname{Sym}^k(\mathbb{R}^m))$, the symmetrised gradient Eu is a Radon measure, $Eu \in \mathcal{M}(\Omega; \operatorname{Sym}^{k+1}(\mathbb{R}^m))$. For the proof of this fact we refer to [20, §4.1.5].

Example .3. The space $\operatorname{BD}(\Omega; \operatorname{Sym}^0(\mathbb{R}^m))$ agrees with the space $\operatorname{BV}(\Omega)$ of functions of bounded variation. The space $\operatorname{BD}(\Omega; \operatorname{Sym}^1(\mathbb{R}^m)) = \operatorname{BD}(\Omega)$ is the space of *functions of bounded deformation* of [51].

References

[1] M. Aksoy, C. Forman, M. Straka, S. Skare, S. Holdsworth, J. Hornegger and R. Bammer, *Real-time optical motion correction for diffusion tensor imaging.*, Magnetic Resonance in Medicine **66** (2011), 366–378, doi:10.1002/mrm.22787.

[2] L. Ambrosio, N. Fusco and D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford University Press, 2000.

[3] V. Arsigny, P. Fillard, X. Pennec and N. Ayache, *Log-euclidean metrics for fast and simple calculus on diffusion tensors*, Magnetic Resonance in Medicine **56** (2006), 411–421.

[4] P. J. Basser and D. K. Jones, *Diffusion-tensor MRI: theory, experimental design and data analysis – a technical review.*, NMR in Biomedicine **15** (2002), 456–467, doi:10.1002/nbm.783.

[5] S. Basu, T. Fletcher and R. Whitaker, *Rician noise removal in diffusion tensor mri*, in: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006*, *Lecture Notes in Computer Science*, volume 4190, Edited by R. Larsen, M. Nielsen and J. Sporring, Springer Berlin Heidelberg, 2006, 117–125, doi:10.1007/11866565_15.

[6] M. Bačák, R. Bergmann, G. Steidl and A. Weinmann, *A Second Order Non-Smooth Variational Model for Restoring Manifold-Valued Images* (2015), accepted, arXiv:1506.02409.

[7] M. Benning, L. Gladden, D. Holland, C.-B. Schönlieb and T. Valkonen, *Phase reconstruction from velocity-encoded MRI measurements – A survey of sparsity-promoting variational approaches*, Journal of Magnetic Resonance **238** (2014), 26–43, doi:10.1016/j.jmr.2013.10.003.

[8] K. Bredies, *Symmetric tensor fields of bounded deformation*, Annali di Matematica Pura ed Applicata **192** (2013), 815–851, doi:10.1007/s10231-011-0248-4.

[9] K. Bredies, K. Kunisch and T. Pock, *Total generalized variation*, SIAM Journal on Imaging Sciences **3** (2011), 492–526, doi:10.1137/090769521.

[10] K. Bredies, K. Kunisch and T. Valkonen, *Properties of L^1-TGV^2: The one-dimensional case*, Journal of Mathematical Analysis and Applications **398** (2013), 438–454, doi:10.1016/j.jmaa.2012.08.053.

[11] K. Bredies and T. Valkonen, *Inverse problems with second-order total generalized variation constraints*, in: *Proceedings of the 9th International Conference on Sampling Theory and Applications (SampTA) 2011, Singapore*, 2011.

[12] M. Burger and F. Lucka, *Maximum a posteriori estimates in linear inverse problems with log-concave priors are proper bayes estimators*, Inverse Problems **30** (2014), 114004.

[13] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, Journal of Mathematical Imaging and Vision **40** (2011), 120–145, doi:10.1007/s10851-010-0251-1.

[14] C. Chefd'hotel, D. Tschumperlé, R. Deriche and O. Faugeras, *Constrained flows of matrix-valued functions: Application to diffusion tensor regularization*, in: *Computer Vision – ECCV 2002*, *Lecture Notes in Computer Science*, volume 2350, Edited by A. Heyden, G. Sparr, M. Nielsen and P. Johansen, Springer Berlin / Heidelberg, 2002, 251–265, doi:10.1007/3-540-47969-4_17.

[15] D. Cox and D. Hinkley, *Theoretical Statistics*, Taylor & Francis, 1979.

[16] J. C. de Los Reyes, C.-B. Schönlieb and T. Valkonen, *Bilevel parameter learning for higher-order total variation regularisation models* (2015), submitted, arXiv:1508.07243.

[17] N. Dunford and J. T. Schwartz, *Linear Operators, Part I General Theory*, Interscience Publishers, Hoboken, New Jersey, 1958.

[18] A. Dvoretzky, J. Kiefer and J. Wolfowitz, *Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator*, The Annals of Mathematical Statistics **27** (1956), 642–669, doi:10.1214/aoms/1177728174.

[19] E. Esser, X. Zhang and T. F. Chan, *A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science*, SIAM Journal on Imaging Sciences **3** (2010), 1015–1046, doi:10.1137/09076934X.

[20] H. Federer, *Geometric Measure Theory*, Springer, 1969.

[21] A. Fuster, T. Dela Haije, A. Tristán-Vega, B. Plantinga, C.-F. Westin and L. Florack, *Adjugate diffusion tensors for geodesic tractography in white matter*, Journal of Mathematical Imaging and Vision **54** (2016), 1–14, doi:10.1007/s10851-015-0586-8.

[22] A. Fuster, A. Tristan-Vega, T. Haije, C.-F. Westin and L. Florack, *A novel riemannian metric for geodesic tractography in dti*, in: *Computational Diffusion MRI and Brain Connectivity*, Edited by T. Schultz, G. Nedjati-Gilani, A. Venkataraman, L. O'Donnell and E. Panagiotaki, Mathematics and Visualization, Springer International Publishing, 2014, 97–104, doi:10.1007/978-3-319-02475-2_9.

[23] P. Getreuer, M. Tong and L. A. Vese, *A variational model for the restoration of MR images corrupted by blur and Rician noise*, in: *Advances in Visual Computing*, *Lecture Notes in Computer Science*, volume 6938, Springer Berlin Heidelberg, 2011, 686–698, doi:10.1007/978-3-642-24028-7_63.

[24] M. Grasmair, M. Haltmeier and O. Scherzer, *The residual method for regularizing ill-posed problems*, Applied Mathematics and Computation **218** (2011), 2693 – 2710, doi:j.amc.2011.08.009.

[25] H. Gudbjartsson and S. Patz, *The Rician distribution of noisy MRI data*, Magnetic Resonance in Medicine **34** (1995), 910–914.

[26] X. Hao, R. Whitaker and P. Fletcher, *Adaptive riemannian metrics for improved geodesic tracking of white matter*, in: *Information Processing in Medical Imaging*, *Lecture Notes in Computer Science*, volume 6801, Edited by G. Székely and H. K. Hahn, Springer Berlin Heidelberg, 2011, 13–24, doi:10.1007/978-3-642-22092-0_2.

[27] B. He and X. Yuan, *Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective*, SIAM Journal on Imaging Sciences **5** (2012), 119–149, doi:10.1137/100814494.

[28] M. Herbst, J. Maclaren, M. Weigel, J. Korvink, J. Hennig and M. Zaitsev, *Prospective motion correction with continuous gradient updates in diffusion weighted imaging*, Magnetic Resonance in Medicine (2011), doi:10.1002/mrm.23230.

[29] T. Hohage and C. Homann, *A generalization of the Chambolle-Pock algorithm to Banach spaces with applications to inverse problems* (2014), preprint, arXiv:1412.0126.

[30] P. Kingsley, *Introduction to diffusion tensor imaging mathematics: Parts I-III*, Concepts in Magnetic Resonance Part A **28** (2006), 101–179, doi:10.1002/cmr.a.20048.

[31] F. Knoll, C. Clason, K. Bredies, M. Uecker and R. Stollberger, *Parallel imaging with nonlinear reconstruction using variational penalties*, Magnetic Resonance in Medicine **67** (2012), 34–41.

[32] F. Knoll, J. G. Raya, R. O. Halloran, S. Baete, E. Sigmund, R. Bammer, T. Block, R. Otazo and D. K. Sodickson, *A model-based reconstruction for undersampled radial spin-echo dti with variational penalties on the diffusion tensor*, NMR in Biomedicine **28** (2015), 353–366, doi:10.1002/nbm.3258.

[33] Y. Korolev, *Making use of a partial order in solving inverse problems: II.*, Inverse Problems **30** (2014), 085003.

[34] Y. Korolev and A. Yagola, *On inverse problems in partially ordered spaces with a priori information*, Journal of Inverse and Ill-posed problems **20** (2012).

[35] Y. Korolev and A. Yagola, *Making use of a partial order in solving inverse problems*, Inverse Problems **29** (2013), 095012.

[36] M. Lassas, E. Saksman and S. Siltanen, *Discretization-invariant bayesian inversion and Besov space priors*, Inverse Problems and Imaging **3** (2009), 87–122, doi:10.3934/ipi.2009.3.87.

[37] M. Lassas and S. Siltanen, *Can one use total variation prior for edge-preserving bayesian inversion?*, Inverse Problems **20** (2004), 1537, doi:10.1088/0266-5611/20/5/013.

[38] E. Lehmann and J. Romano, *Testing Statistical Hypotheses*, Springer Texts in Statistics, Springer New York, 2008.

[39] W. Luxemburg and A. Zaanen, *Riesz Spaces*, North-Holland Publishing Company, Amsterdam, 1971.

[40] A. Martín and E. Schiavi, *Automatic total generalized variation-based DTI Rician denoising*, in: *Image Analysis and Recognition*, *Lecture Notes in Computer Science*, volume 7950, Springer Berlin Heidelberg, 2013, 581–588, doi:10.1007/978-3-642-39094-4_66.

[41] P. Massart, *The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality*, The Annals of Probability **18** (1990), 1269–1283, doi:10.1214/aop/1176990746.

[42] Y. Meyer, *Oscillating patterns in image processing and nonlinear evolution equations*, American Mathematical Society, Boston, MA, USA, 2001.

[43] Y. Nesterov, *A method of solving a convex programming problem with convergence rate $O(1/k^2)$*, Soviet Mathematics Doklady **27** (1983), 372–376.

[44] X. Pan, E. Y. Sidky and M. Vannier, *Why do commercial ct scanners still employ traditional, filtered back-projection for image reconstruction?*, Inverse Problems **25** (2009), 123009, doi:10.1088/0266-5611/25/12/123009.

[45] L. Rudin, S. Osher and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D **60** (1992), 259–268.

[46] H. Schaefer, *Banach Lattices and Positive Operators*, Springer Verlag, Berlin, New York, 1974.

[47] S. Setzer, *Operator splittings, bregman methods and frame shrinkage in image processing*, International Journal of Computer Vision **92** (2011), 265–280, doi:10.1007/s11263-010-0357-3.

[48] S. Setzer, G. Steidl, B. Popilka and B. Burgeth, *Variational methods for denoising matrix fields*, in: *Visualization and Processing of Tensor Fields*, Edited by J. Weickert and H. Hagen, Springer, 2009, 341–360.

[49] A. N. Shiriaev, *Probability*, Graduate Texts in Mathematics, Springer, 1996.

[50] S. M. Smith, M. Jenkinson, M. W. Woolrich, C. F. Beckmann, T. E. J. Behrens, H. Johansen-Berg, P. R. Bannister, M. D. Luca, I. Drobnjak, D. E. Flitney, R. K. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. D. Stefano, J. M. Brady and P. M. Matthews, *Advances in functional and structural MR image analysis and implementation as FSL.*, Neuroimage **23 Suppl 1** (2004), S208–S219, doi:10.1016/j.neuroimage.2004.07.051.

[51] R. Temam, *Mathematical problems in plasticity*, Gauthier-Villars, 1985.

[52] A. N. Tikhonov, A. V. Goncharsky, V. V. Stepanov and A. G. Yagola, *Numerical Methods for the Solution of Ill-Posed Problems*, Kluwer, Dordrecht, 1995.

[53] J.-D. Tournier, S. Mori and A. Leemans, *Diffusion tensor imaging and beyond*, Magnetic Resonance in Medicine **65** (2011), 1532–1556, doi:10.1002/mrm.22924.

[54] D. Tschumperlé and R. Deriche, *Diffusion tensor regularization with constraints preservation*, in: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2001, 948–953.

[55] T. Valkonen, *A primal-dual hybrid gradient method for non-linear operators with applications to MRI*, Inverse Problems **30** (2014), 055012, doi:10.1088/0266-5611/30/5/055012, arXiv:1309.5032.

[56] T. Valkonen, *Big images*, in: *Big Data Optimization: Recent Developments and Challenges*, Edited by A. Emrouznejad, Studies in Big Data, Springer-Verlag, 2015, accepted.

[57] T. Valkonen, K. Bredies and F. Knoll, *Total generalised variation in diffusion tensor imaging*, SIAM Journal on Imaging Sciences **6** (2013), 487–525, doi:10.1137/120867172.

[58] T. Valkonen, F. Knoll and K. Bredies, *TGV for diffusion tensors: A comparison of fidelity functions*, Journal of Inverse and Ill-Posed Problems **21** (2013), 355–377, doi:10.1515/jip-2013-0005, Special issue for IP:M&S 2012, Antalya, Turkey.

[59] T. Valkonen and M. Liebmann, *GPU-accelerated regularisation of large diffusion tensor volumes*, Computing **95** (2013), 771–784, doi:10.1007/s00607-012-0277-x, Special issue for ESCO 2012, Pilsen, Czech Republic.

[60] J. Weickert, *Anisotropic diffusion in image processing*, volume 1, Teubner Stuttgart, 1998.