SpezialForschungsBereich F 32

Karl–Franzens Universität Graz
Technische Universität Graz
Medizinische Universität Graz

Total generalised variation in diffusion tensor imaging

T. Valkonen F. Knoll

SFB-Report No. 2012-003 February 2012

A–8010 GRAZ, HEINRICHSTRASSE 36, AUSTRIA

Supported by the
Austrian Science Fund (FWF)

Der Wissenschaftsfonds.

SFB sponsors:

- **Austrian Science Fund** (FWF)
- **University of Graz**
- **Graz University of Technology**
- **Medical University of Graz**
- **Government of Styria**
- **City of Graz**

Total generalised variation in diffusion tensor imaging

Tuomo Valkonen* and Florian Knoll†

Abstract

We study the extension of total variation (TV) and (second-order) total generalised variation (TGV^2) to tensor fields, and the application thereof to the regularisation of noisy diffusion tensor images from medical applications.

Mathematics subject classification: 92C55, 94A08, 26B30, 49M29.

Keywords: total variation, diffusion tensor imaging, DTI, regularisation, medical imaging.

1. Introduction

Let $f \in L^1(\Omega; \mathrm{Sym}^2(\mathbb{R}^d))$ be a second-order symmetric tensor field on a domain $\Omega \subset \mathbb{R}^d$. In the following, we intend to study ROF-type (Rudin, Osher and Fatemi [29]) regularisation problems

$$\min_{u \geq 0} \frac{1}{2}\|f - u\|_{F,2}^2 + H(u). \tag{1.1}$$

for a suitable regularisation functional H, particularly total variation (TV) and second-order total generalised variation (TGV^2) [10], extended to the tensor case. Observe that we require the solution $u \in L^1(\Omega; \mathrm{Sym}^2(\mathbb{R}^d))$ to be positive semi-definite.

This kind of regularisation problems arise, for instance, in diffusion-tensor imaging (DTI) of brain tissues. As a first step towards DTI, diffusion weighted magnetic resonance imaging (DWI) is performed. It measures the anisotropic diffusion of water molecules, and provides valuable and unique in-vivo insight into the white matter structure of the brain [8, 33]. To capture the diffusion information, images have to be obtained with diffusion sensitising gradients in multiple directions. This leads to very long acquisition times, even with ultra fast sequences like echo planar imaging (EPI). Therefore, DWI is inherently a low-resolution and low-SNR method. It exhibits Rician noise [20], eddy-current distortions [33], and is very sensitive with respect to artifacts originating from patient motion [21, 2].

By taking multiple DWI images, a *diffusion-tensor* describing the probability of water diffusing in different spatial directions can be solved from the Stejskal-Tanner equation [8, 23]. These tensors can be visualised as a confidence ellipsoids, which have varying *anisotropy* – or directional dependence – depending on the probability of diffusion of water in that direction. In particular, in the brain, *grey matter* has low anisotropy – the ellipsoids are almost spheres, water having uniform diffusion. In *white matter*, which transmits messages between areas of grey matter, the anisotropy is, by contrast, high – the ellipsoids are far from spheres, and water has a single most probable direction of diffusion. Since the DWI measurements are noisy, as described in the previous paragraph, we are led to the problem of

* Institute for Mathematics and Scientific Computing, University of Graz, Austria. tuomo.valkonen@iki.fi
† Institute of Medical Engineering, Graz University of Technology, Austria. florian.knoll@tugraz.at

denoising the diffusion tensors obtained this way. The denoised tensors (symmetric matrices) should, moreover, be positive-definite, as the failure of this condition is non-physical.

Pennec and his cooperators [5, 6, 19] have studied log-Euclidean metrics in the context of diffusion-tensor imaging. Regularisation approaches based on log-Euclidean metrics facilitate maintaining the positive-definiteness of the tensors, as well as avoiding swelling, i.e., increase in volume of the ellipsoids corresponding to the tensors. Basically the log-Euclidean approach amounts to taking the logarithm of the data, to expand the positive definite cone to the whole space, applying usual Euclidean techniques (such as ROF), and transforming the solution back to the positive-define cone by taking the exponential. This approach has therefore many desirable theoretical qualities, and is also computationally tractable. We are interested in whether other regularisation approaches can provide results with other desirable qualities. We study those in the family (1.1). One of these, namely TV, however lacking the positive-definiteness constraint, has already been studied by Setzer et al. [30]. Some further related work has been conducted by Chefd'hotel et al. [15] and Tschumperlé and Deriche [34]. In these works, instead of minimising a regularisation functional, a flow of the tensor field is defined, using the theory of Riemannian manifolds (also employed by Pennec). The measured tensor field is then transported by the flow until a result of desired quality is found. Finally, Poupon et al. [28] and Coulon et al. [16] study the regularisation of the principal eigenvectors alone, not the whole tensors.

The rest of our paper is organised as follows. First in Section 2 we introduce the tensor and tensor field calculus to set up the framework in which our results are represented. Additionally we study the rotational invariance of tensor norms. Then in Section 3 we introduce and study variants of the model problem (1.1). Section 4 describes our implementation of the Chambolle-Pock algorithm [14] that we use to solve these problems. Finally, in Section 5 we study the numerical results that we have obtained, on both synthetic test data and an in-vivo brain measurement. To conclude the paper, we state our final conclusions and outlook for future research in Section 6.

2. Tensors and tensor fields

We now recall basic tensor calculus, and then move on to develop norms and differentials of tensor fields, and finally defining *tensor fields of bounded variation* and of *bounded deformation*. While diffusion tensors are simply symmetric matrices, and the whole machinery of tensor calculus is therefore not needed for talking about DTI measurements, for the development of differentials of tensor fields, employed in the regularisation approaches that we study, we do need to introduce the machinery for higher-order tensors as well. The development of total generalised variation indeed depends on higher-order tensors even for scalar fields.

2.1. Basic tensor calculus

A k-tensor $A \in \mathcal{T}^k(\mathbb{R}^m)$ is a k-linear mapping $A : \mathbb{R}^m \times \cdots \times \mathbb{R}^m \to \mathbb{R}$. A symmetric tensor $A \in \mathrm{Sym}^k(\mathbb{R}^m)$ satisfies for any permutation π of $\{1, \ldots, k\}$ that $A(c_{\pi 1}, \ldots, c_{\pi k}) = A(c_1, \ldots, c_k)$. For any $A \in \mathcal{T}^k(\mathbb{R}^m)$ we may define the symmetrisation $\|\!| A$ by

$$(\|\!| A)(c_1, \ldots, c_k) := \frac{1}{k!} \sum_{\pi} A(c_{\pi 1}, \ldots, c_{\pi k}),$$

where the sum is taken over all $k!$ permutations π of $\{1, \ldots, k\}$.

Example 2.1 (Vectors)**.** Vectors $A \in \mathbb{R}^m$ are symmetric 1-tensors: $A(x) = \langle A, x \rangle$. The symmetrisation satisfies $\|\!| A = A$.

Example 2.2 (Matrices)**.** Matrices are 2-tensors: $A(x, y) = \langle Ax, y \rangle$. Symmetric matrices $A = A^T$ are symmetric 2-tensors. The symmetrisation is given by $\|\!| A = (A + A^T)/2$. We use the notation $A \geq 0$ for positive-semidefinite A.

2.2. Tensor norms

Let $e_1, \dots, e_m$ be the standard basis of $\mathbb{R}^m$. We then define 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

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

We have $\langle A, \|\|B\rangle = \langle \|\|A, B\rangle$. Thus, in particular, for symmetric $B \in \mathrm{Sym}^k(\mathbb{R}^m)$, and general $A \in \mathrm{Sym}^k(\mathbb{R}^m)$, we have $\langle A, B\rangle = \langle \|\|A, B\rangle$.

Next we discuss the so called the largest and smallest *reasonable cross-norms*, following [17]. A word of note, however: all the results in [17] that we refer to are stated there only for $k = 2$, but readily generalise to $k > 2$. For a choice of $p \in [1, \infty]$, we thus define the *smallest reasonable cross-norm* as

$$\|A\|_\vee := \sup\{A(x_1, \dots, x_k) \mid \|x_i\|_p \le 1,\ i = 1, \dots, k\}. \tag{2.1}$$

The *largest reasonable cross-norm* is defined by duality as

$$\|A\|_\wedge := \sup\{\langle A, \xi\rangle \mid \|\xi\|_\vee \le 1\}$$

It may alternatively be written as

$$\|A\|_\wedge = \inf\{\sum_{i=1}^N \|x_1^i\|_p \cdots \|x_k^i\|_p \mid A = \sum_{i=1}^N x_1^i \otimes \cdots \otimes x_k^i\}. \tag{2.2}$$

Only the case $p = 2$ is interesting to us, as in that case these norms are rotationally invariant, as is the Frobenius norm.

Proposition 2.1. *Let $A \in \mathcal{T}^k(\mathbb{R}^m)$, and let $R \in \mathbb{R}^{m\times m}$ be a rotation matrix (i.e., orthogonal with $R^T = R^{-1}$ and $\det R = 1$). Define $A_R \in \mathcal{T}^k(\mathbb{R}^m)$ according to*

$$A_R(c_1, \dots, c_k) = A(Rc_1, \dots, Rc_k). \tag{2.3}$$

Then each of the norms $\|\cdot\|_\bullet$ for $\bullet = F, \wedge, \vee$ and $p = 2$, is rotationally invariant in the sense that $\|A_R\|_\bullet = \|A\|_\bullet$.

Proof. That $\|\cdot\|_\vee$ is rotationally invariant, is clear from the the equation (2.1) defining the norm, as seen by evaluating $A(x_1', \dots, x_k')$ with $x_i' = R^{-1}x_i$. That $\|\cdot\|_\wedge$ is rotationally invariant is clear from the formulation (2.2): If $A = \sum_{i=1}^N x_1^i \otimes \cdots \otimes x_k^i$, then $A_R = \sum_{i=1}^N R^T x_1^i \otimes \cdots \otimes R^T x_k^i$. Likewise $\|\cdot\|_F$ can be seen to be rotationally invariant by arbitrarily decomposing $A = \sum_{i=1}^N x_1^i \otimes \cdots \otimes x_k^i$. (Such a decomposition always exists; we may, for example, take $A = \sum_{p\in\{1,\dots,m\}^k} A(e_{p_1}, \dots, e_{p_k})e_{p_1} \otimes \cdots \otimes e_{p_k}$.) Then

$$\begin{aligned}\langle A_R, A_R\rangle &= \sum_{i=1}^N \langle Rx_1^i \otimes \cdots \otimes Rx_k^i, Rx_1^i \otimes \cdots \otimes Rx_k^i\rangle = \sum_{i=1}^N \prod_{j=1}^k \|Rx_j^i\|^2 = \sum_{i=1}^N \prod_{j=1}^k \|x_j^i\|^2 \\ &= \langle x_1^i \otimes \cdots \otimes x_k^i, x_1^i \otimes \cdots \otimes x_k^i\rangle = \langle A, A\rangle.\end{aligned}$$

□

Example 2.3 (Vectors)**.** The inner product is the usual inner product in $\mathbb{R}^m$, and the Frobenius norm $\|A\|_F = \|A\|_2$. We have $\|A\|_\wedge = \|A\|_p$, and $\|A\|_\wedge = \|A\|_q$, where $1/p + 1/q = 1$. Thus for $p = 2$ all the three norms are equivalent.

Example 2.4 (Matrices)**.** The inner product is $\langle A, B\rangle = \sum_{i,j} A_{ij}B_{ij}$ and $\|A\|_F$ is the matrix Frobenius norm. Denoting by $\{\lambda_i\}$ the eigenvalues of A, we find that for $p = 2$, the largest reasonable cross-norm $\|A\|_\wedge = \sum_i |\lambda_i|$ is the *nuclear norm*, and the smallest reasonable cross-norm $\|A\|_\vee = \max_i |\lambda_i|$ is the *spectral norm*.

2.3. Tensor fields

Let $u : \Omega \to \mathcal{T}^k(\mathbb{R}^m)$ for a domain $\Omega \subset \mathbb{R}^m$, and let $\|\cdot\|_\bullet$ be a norm on $\mathcal{T}^k(\mathbb{R}^m)$. We then set

$$\|u\|_{\bullet,p} := \Big(\int_\Omega \|u(x)\|_\bullet^p \, dx\Big)^{1/p} \quad (p \in [1,\infty)), \qquad \text{and} \quad \|u\|_{\bullet,\infty} := \operatorname{ess\,sup}_{x\in\Omega} \|u(x)\|_\bullet,$$

and define the spaces

$$L^p(\Omega;\mathcal{T}^k(\mathbb{R}^m)) = \{u : \Omega \to \mathcal{T}^k(\mathbb{R}^m) \mid \|u\|_{\bullet,p} < \infty\}, \quad \text{and}$$
$$L^p(\Omega;\operatorname{Sym}^k(\mathbb{R}^m)) = \{u : \Omega \to \operatorname{Sym}^k(\mathbb{R}^m) \mid \|u\|_{\bullet,p} < \infty\}, \quad (p \in [1,\infty]).$$

The choice of the finite-dimensional norm $\|\cdot\|_\bullet$ does not affect the definition of these spaces, since all finite-dimensional norms are equivalent. Hence strong and weak convergence in the spaces is unambiguously defined, with the dual spaces analogous to the case of scalar functions: for $1 \le p < \infty$, the dual of $L^p(\Omega;\mathcal{T}^k(\mathbb{R}^m))$ (resp. $L^p(\Omega;\operatorname{Sym}^k(\mathbb{R}^m))$) is $L^q(\Omega;\mathcal{T}^k(\mathbb{R}^m))$ (resp. $L^q(\Omega;\operatorname{Sym}^k(\mathbb{R}^m))$), where q is the Hölder conjugate of p, satisfying $1/p+1/q = 1$. In a standard fashion we also deduce that $\|\cdot\|_{\bullet,p}$ is lower-semicontinuous with respect to weak convergence in L^p.

A tensor field $\varphi : \Omega \to \mathcal{T}^k(\mathbb{R}^m)$ is symmetrised pointwise, that is

$$(\|\!|\varphi)(x) := \|\!|(\varphi(x)).$$

We say that φ is symmetric, if $\varphi(x) \in \operatorname{Sym}^k(\mathbb{R}^m)$ for every $x \in \Omega$.

Finally, if $u \in C^1(\Omega;\mathcal{T}^k(\mathbb{R}^m))$, $k \ge 1$, we define the divergence $\operatorname{div} u \in C(\Omega;\mathcal{T}^{k-1}(\mathbb{R}^m)$ by contraction as

$$[\operatorname{div} u(x)](e_{i_2},\dots,e_{i_k}) := \sum_{i_1=1}^m \partial_{i_1}[x \mapsto u(x)(e_{i_1},\dots,e_{i_k})] = \sum_{i_1=1}^m \langle e_{i_1}, \nabla u(\cdot)(e_{i_1},\dots,e_{i_k})\rangle. \tag{2.4}$$

Observe that if u is symmetric, then so is $\operatorname{div} u$.

Example 2.5 (Vector fields)**.** Let $u \in C^1(\Omega;\mathbb{R}^m) = C^1(\Omega;\mathcal{T}^1(\mathbb{R}^m))$. Then $\operatorname{div} u(x) = \sum_{i=1}^m \partial_i u_i(x)$ is the usual divergence.

Example 2.6 (Matrix fields)**.** Let $u \in C^1(\Omega;\mathcal{T}^2(\mathbb{R}^m))$. Then $[\operatorname{div} u(x)]_j = \sum_{i=1}^m \partial_i u_{ij}(x)$. That is, we take columnwise the divergence of a vector field. We use the notation $u \ge 0$ for pointwise a.e. positive-semidefinite u.

2.4. Tensor-valued measures

We let now $\mu \in \mathcal{M}(\Omega;\mathcal{T}^k(\mathbb{R}^m))$ be a tensor-valued measure, that is, a family of measures $\mu(\cdot)(e_{p_1},\dots,e_{p_k}) \in \mathcal{M}(\Omega)$ over the componentwise indices $p \in \{1,\dots,m\}^k$. For $\varphi \in C_c^\infty(\Omega;\mathcal{T}^k(\mathbb{R}^m))$, we then define

$$\mu(\varphi) := \int_\Omega \langle \varphi, d\mu\rangle := \sum_{p\in\{1,\dots,m\}^k} \int_\Omega \varphi(x)(e_{p_1},\dots,e_{p_k})\, d\mu(x)(e_{p_1},\dots,e_{p_k}). \tag{2.5}$$

We symmetrise μ as

$$(\|\!|\mu)(\varphi) := \mu(\|\!|\varphi), \quad (\varphi \in \mathcal{T}^k(\mathbb{R}^m)).$$

For $\mu = f\mathcal{L}^m$, where $\mathcal{L}^m$ is the Lebesgue measure on $\mathbb{R}^m$, this gives the desired expression $(\|\!|\mu)(\varphi) = \int_\Omega \langle \varphi(x), \|\!| f(x)\rangle\, dx$, and for symmetric $\mu \in \mathcal{M}(\Omega;\operatorname{Sym}^k(\mathbb{R}^m))$ we have $\|\!|\mu = \mu$.

For each $\bullet = F, \wedge, \vee$, let us define the non-symmetric unit ball

$$V^k_{\bullet,\mathrm{ns}} := \{\varphi \in C^\infty_c(\Omega; \mathcal{T}^k(\mathbb{R}^m)) \mid \|\varphi\|_{\bullet,\infty} \le 1\}, \tag{2.6}$$

and the symmetric unit ball

$$V^k_{\bullet,\mathrm{s}} := \{\varphi \in C^\infty_c(\Omega; \mathrm{Sym}^k(\mathbb{R}^m)) \mid \|\varphi\|_{\bullet,\infty} \le 1\}.$$

Topologically the choice of the finite-dimensional norm $\|\cdot\|_\bullet$ does not of course matter, while geometrically it does.

For $\mu \in \mathcal{M}(\Omega; \mathcal{T}^k(\mathbb{R}^m))$, we now define for each $\bullet = F, \wedge, \vee$ the total variation norm

$$\|\mu\|_{\bullet,\mathcal{M}(\Omega;\mathcal{T}^k(\mathbb{R}^m))} := \sup\{\mu(\varphi) \mid \varphi \in V^k_{\bullet,\mathrm{ns}}\}.$$

For symmetric $\mu \in \mathcal{M}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$, we have

$$\|\mu\|_{\bullet,\mathcal{M}(\Omega;\mathcal{T}^k(\mathbb{R}^m))} = \sup\{\mu(\varphi) \mid \varphi \in V^k_{\bullet,\mathrm{s}}\};$$

indeed

$$\begin{aligned}\sup\{\mu(\varphi) \mid \varphi \in V^k_{\bullet,\mathrm{s}}\} &\le \sup\{\mu(\varphi) \mid \varphi \in V^k_{\bullet,\mathrm{ns}}\} = \|\mu\|_{\bullet,\mathcal{M}(\Omega;\mathcal{T}^k(\mathbb{R}^m))} \\ &= \|\|\mu\|\|_{\bullet,\mathcal{M}(\Omega;\mathcal{T}^k(\mathbb{R}^m))} = \sup\{\mu(\|\varphi) \mid \varphi \in V^k_{\bullet,\mathrm{ns}}\} \le \sup\{\mu(\varphi) \mid \varphi \in V^k_{\bullet,\mathrm{s}}\}.\end{aligned}$$

Finally, weak* convergence of $\{\mu^i\}_{i=0}^\infty \in \mathcal{M}(\Omega; \mathcal{T}^k(\mathbb{R}^m))$ to $\mu \in \mathcal{M}(\Omega; \mathcal{T}^k(\mathbb{R}^m))$ holds when

$$\mu^i(\varphi) \to \mu(\varphi) \quad \text{for all } \varphi \in C^\infty_c(\Omega; \mathcal{T}^k(\mathbb{R}^m)).$$

Observe that $\|\cdot\|_{\bullet,\mathcal{M}(\Omega;\mathcal{T}^{k+1}(\mathbb{R}^m))}$ is lower-semicontinuous with respect to weak* convergence.

2.5. Symmetric tensor fields of bounded variation and deformation

We define the distributional gradient $Du \in [C^\infty_c(\Omega; \mathcal{T}^{k+1}(\mathbb{R}^m))]^*$ of $u \in L^1(\Omega; \mathcal{T}^k(\mathbb{R}^m))$ by

$$Du(\varphi) := \int_\Omega \langle u(x), \operatorname{div} \varphi(x)\rangle \, dx, \quad (\varphi \in C^\infty_c(\Omega; \mathcal{T}^{k+1}(\mathbb{R}^m))).$$

Likewise we define the symmetrised distributional gradient $Eu \in [C^\infty_c(\Omega; \mathrm{Sym}^{k+1}(\mathbb{R}^m))]^*$ by

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

With these notions at hand, we now define the spaces of symmetric tensor fields of *bounded variation* and *bounded deformation*, respectively, as (see also [9])

$$\begin{aligned}\mathrm{BV}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m)) &:= \Big\{u \in L^1(\Omega; \mathrm{Sym}^k(\mathbb{R}^m)) \;\Big|\; \sup\nolimits_{\varphi \in V^{k+1}_{F,\mathrm{ns}}} Du(\varphi) < \infty\Big\}, \quad \text{and} \\ \mathrm{BD}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m)) &:= \Big\{u \in L^1(\Omega; \mathrm{Sym}^k(\mathbb{R}^m)) \;\Big|\; \sup\nolimits_{\varphi \in V^{k+1}_{F,\mathrm{s}}} Eu(\varphi) < \infty\Big\}.\end{aligned}$$

For functions $u \in \mathrm{BV}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$, we have $Du \in \mathcal{M}(\Omega; \mathcal{T}^{k+1}(\mathbb{R}^m))$, that is, the operator Du is a measure. Analogously, for $u \in \mathrm{BD}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$, it holds $Eu \in \mathcal{M}(\Omega; \mathrm{Sym}^{k+1}(\mathbb{R}^m))$. The proof of these facts follows as in, e.g., [18, §4.1.5].

Remark 2.1. For $k = 0$ the space $\mathrm{BV}(\Omega; \mathrm{Sym}^0(\mathbb{R}^m))$ agrees with the usual space of (scalar) functions of bounded variation. For $k = 1$ the space $\mathrm{BD}(\Omega; \mathrm{Sym}^1(\mathbb{R}^m))$ agrees with the space of *functions of bounded deformation* studied by Temam [32].

Proposition 2.2. *Let $u \in \mathrm{BV}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$. Then $u \in \mathrm{BD}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$ and $Eu = \|\!| Du$.*

Proof. Let $\varphi \in C_c^\infty(\Omega; \mathrm{Sym}^{k+1}(\mathbb{R}^m))$. Then by the very definition of Eu and Du, we have

$$Eu(\varphi) = \int_\Omega \langle u(x), \operatorname{div} \varphi(x) \rangle \, dx = Du(\varphi).$$

In particular, $\sup\{Eu(\varphi) \mid \varphi \in V_{F,\mathrm{s}}^{k+1}\} < \infty$, so $u \in \mathrm{BD}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$. Let then $\varphi \in C_c^\infty(\Omega; \mathcal{T}^{k+1}(\mathbb{R}^m))$ be possibly non-symmetric. We now obtain

$$(\|\!| Du)(\varphi) = Du(\|\!| \varphi) = Eu(\|\!| \varphi) = Eu(\varphi).$$

In the final step we have used the fact that Eu is a measure, is hence defined for non-symmetric φ, and clearly satisfies $\|\!| Eu = Eu$. □

Analogously to the case of scalar functions, we define on $\mathrm{BV}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$, and on $\mathrm{BD}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$, the norms

$$\begin{aligned} \|u\|_{\bullet,\mathrm{BV}(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))} &:= \|u\|_{\bullet,1} + \|Du\|_{\bullet,\mathcal{M}(\Omega;\mathcal{T}^{k+1}(\mathbb{R}^m))}, \quad \text{and, respectively,} \\ \|u\|_{\bullet,\mathrm{BD}(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))} &:= \|u\|_{\bullet,1} + \|Eu\|_{\bullet,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))}. \end{aligned}$$

These define strong convergence and Banach-space structure; see [9]. Weak convergence of $\{u^i\}_{i=0}^\infty \subset \mathrm{BV}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$ to $u \in \mathrm{BV}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$ is defined as

$$u^i \to u \text{ strongly in } L^1(\Omega; \mathrm{Sym}^k(\mathbb{R}^m)) \quad \text{and} \quad Du^i \overset{*}{\rightharpoonup} Du \text{ weakly* in } \mathcal{M}(\Omega; \mathcal{T}^{k+1}(\mathbb{R}^m)).$$

In $\mathrm{BD}(\Omega; \mathrm{Sym}^{k+1}(\mathbb{R}^m))$ weak convergence is defined analogously by

$$u^i \to u \text{ strongly in } L^1(\Omega; \mathrm{Sym}^k(\mathbb{R}^m)) \quad \text{and} \quad Eu^i \overset{*}{\rightharpoonup} Eu \text{ weakly* in } \mathcal{M}(\Omega; \mathrm{Sym}^{k+1}(\mathbb{R}^m)).$$

It is immediate that $\|\cdot\|_{\bullet,\mathrm{BV}(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))}$ (resp. $\|\cdot\|_{\bullet,\mathrm{BD}(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))}$) is lower-semicontinuous with respect to weak convergence in $\mathrm{BV}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$ (resp. $\mathrm{BD}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$).

We finally come to an additional justification for the choice of the finite-dimensional norms $\|\cdot\|_\bullet$. Recalling the definition of $u(x)_R$ from (2.3), clearly $\|u(\cdot)_R\|_{\bullet,p} = \|u\|_{\bullet,p}$ for $\bullet = F, \wedge, \vee$ and a rotation matrix R. The next lemma and proposition show that also the norm of the gradient is invariant when the domain $\Omega = B := B(0,1)$ (or any other ball) is rotated as well.

Lemma 2.1. *Suppose $u \in L^1(B; \mathcal{T}^k(\mathbb{R}^m))$, and that $\varphi \in C_c^\infty(B; \mathcal{T}^{k+1}(\mathbb{R}^m))$. Let R be a rotation matrix on $\mathbb{R}^m$. Define $u^R \in L^1(B; \mathcal{T}^k(\mathbb{R}^m))$ by*

$$u^R(y) := [u(R^{-1}y)]_R, \quad (y \in B)$$

and $\varphi^R \in C_c^\infty(B; \mathcal{T}^{k+1}(\mathbb{R}^m))$ analogously. Then

$$Du^R(\varphi^R) = Du(\varphi)$$

as well as

$$\|\varphi^R\|_{\bullet,p} = \|\varphi\|_{\bullet,p}, \quad (\bullet = F, \wedge, \vee;\ p \in [1,\infty]). \tag{2.7}$$

Moreover, if u is a symmetric tensor field, then so is u^R, and if φ is a symmetric tensor field, then so is φ^R, and we have

$$Eu^R(\varphi^R) = Eu(\varphi).$$

Proof. That u_R (resp. φ^R) is symmetric whenever u (resp. φ) is, is clear from the definition

$$u^R(y)(c_1,\dots,c_k) = u(R^{-1}y)(Rc_1,\dots,Rc_k).$$

Observe then similarly to the proof of Proposition 2.1 that $\langle A_{R^{-1}}, C_{R^{-1}}\rangle = \langle A, C\rangle$ for $A, C \in \mathcal{T}^k(\mathbb{R}^m)$. Thus we may now calculate for $y \in B$ that

$$\begin{aligned}
&\langle u^R(y), \operatorname{div}\varphi^R(y)\rangle = \langle [u^R(y)]_{R^{-1}}, [\operatorname{div}\varphi^R(y)]_{R^{-1}}\rangle \\
&= \sum_{p\in\{1,\dots,m\}^k} u^R(y)(R^{-1}e_{p_1},\dots,R^{-1}e_{p_k})(\operatorname{div}\varphi^R(y))(R^{-1}e_{p_1},\dots,R^{-1}e_{p_k}) \\
&= \sum_{p\in\{1,\dots,m\}^k} u(R^{-1}y)(e_{p_1},\dots,e_{p_k})\Big(\sum_{p_0=1}^m \langle e_{p_0}, \nabla_y[\varphi(R^{-1}y)(Re_{p_0}, e_{p_1},\dots,e_{p_k})]\rangle\Big) \\
&= \sum_{p\in\{1,\dots,m\}^k} u(R^{-1}y)(e_{p_1},\dots,e_{p_k})\Big(\sum_{p_0=1}^m \langle Re_{p_0}, \nabla[\varphi(\cdot)(Re_{p_0}, e_{p_1},\dots,e_{p_k})](R^{-1}y)\rangle\Big) \\
&= \sum_{p\in\{1,\dots,m\}^k} u(R^{-1}y)(e_{p_1},\dots,e_{p_k})\Big(\sum_{p_0=1}^m \langle e_{p_0}, \nabla[\varphi(\cdot)(e_{p_0}, e_{p_1},\dots,e_{p_k})](R^{-1}y)\rangle\Big) \\
&= \sum_{p\in\{1,\dots,m\}^k} u(R^{-1}y)(e_{p_1},\dots,e_{p_k})(\operatorname{div}\varphi(R^{-1}y))(e_{p_1},\dots,e_{p_k}) = \langle u(R^{-1}y), \operatorname{div}\varphi(R^{-1}y)\rangle.
\end{aligned}$$

In the next-to-last step we have employed the fact that for any matrix $A \in \mathcal{T}^2(\mathbb{R}^m)$, the trace $\operatorname{tr} A := \sum_i \langle c_i, Ac_i\rangle$ is invariant of the orthonormal basis $c_1,\dots,c_m$ of $\mathbb{R}^m$. Thus

$$\begin{aligned}
Du^R(\varphi^R) &= \int_B \langle u^R(y), \operatorname{div}\varphi^R(y)\rangle\, dy \\
&= \int_B \langle u(R^{-1}y), \operatorname{div}\varphi(R^{-1}y)\rangle\, dy = \int_B \langle u(y), \operatorname{div}\varphi(y)\rangle\, dy = Du(\varphi).
\end{aligned}$$

Analogously, when φ and then φ^R is symmetric, we get

$$Eu^R(\varphi^R) = Eu(\varphi).$$

Finally, $\|\varphi(R^{-1}y)\|_{\bullet,p} = \|[\varphi(R^{-1}y)]_R\|_{\bullet,p} = \|\varphi^R(y)\|_{\bullet,p}$ due to the rotational invariance of the finite-dimensional norm $\|\cdot\|_\bullet$, shown in Proposition 2.1. Thus (2.7) can be seen to hold. Indeed,

$$\|\varphi^R\|_{\bullet,\infty} = \sup_{y\in B}\|\varphi(R^{-1}y)\|_\bullet = \sup_{y\in B}\|\varphi(y)\|_\bullet = \|\varphi\|_{\bullet,\infty}.$$

and

$$\|\varphi^R\|_{\bullet,p} = \int_B \|\varphi(R^{-1}y)\|_\bullet\, dy = \int_B \|\varphi(y)\|_\bullet\, dy = \|\varphi\|_{\bullet,p}, \quad (p\in[1,\infty)).$$ □

Proposition 2.3. *Let $u \in \mathrm{BV}(B;\mathrm{Sym}^k(\mathbb{R}^m))$ (resp. $u \in \mathrm{BD}(B;\mathrm{Sym}^k(\mathbb{R}^m))$). Given a rotation matrix R on $\mathbb{R}^m$, and a choice $\bullet = F, \wedge, \vee$ of the norm, we then have*

$$\begin{aligned}
\|Du^R\|_{\bullet,\mathcal{M}(\Omega;\mathcal{T}^{k+1}(\mathbb{R}^m))} &= \|Du\|_{\bullet,\mathcal{M}(\Omega;\mathcal{T}^{k+1}(\mathbb{R}^m))} \\
\textit{(resp. } \|Eu^R\|_{\bullet,\mathcal{M}(\Omega;\mathcal{T}^{k+1}(\mathbb{R}^m))} &= \|Eu\|_{\bullet,\mathcal{M}(\Omega;\mathcal{T}^{k+1}(\mathbb{R}^m))} \textit{).}
\end{aligned}$$

Proof. Immediate consequence of Lemma 2.1. □

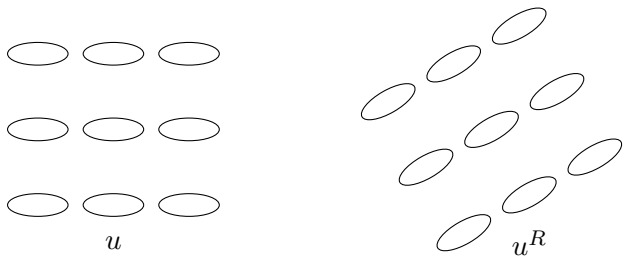

Figure 1: Illustration of the rotational invariance $\|Du^R\|_{F,\mathcal{M}(\mathbb{R}^2;\mathcal{T}^3(\mathbb{R}^2))} = \|Du\|_{F,\mathcal{M}(\mathbb{R}^2;\mathcal{T}^3(\mathbb{R}^2))}$.

Remark 2.2. To every tensor $A \in \mathcal{T}^2(\mathbb{R}^2)$, there corresponds an ellipse, with major and minor axes with magnitude and direction corresponding to the eigenvalues and eigenvectors of A. What the proposition above roughly says, is that the norms of the gradients don't change when we rotate the image consisting of these ellipses; see Figure 1 for an illustration. That is, it is not sufficient to rotate the domain or the tensors alone to obtain invariance, they both have to be rotated.

Remark 2.3. As the final remark of this section, we note that it can be argued that from the point of view of general theory, the space

$$\mathrm{BV}(\Omega;\mathcal{T}^k(\mathbb{R}^m)) := \Big\{ u \in L^1(\Omega;\mathcal{T}^k(\mathbb{R}^m)) \;\Big|\; \sup\nolimits_{\varphi\in V^{k+1}_{F,\mathrm{ns}}} Du(\varphi) < \infty \Big\}.$$

is a more natural one than $\mathrm{BV}(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))$. Indeed, in the former case neither u nor Du is restricted to have symmetric values, while in the latter u alone is. The symmetrised derivative Eu, by contrast, has symmetric values, so the restricting $u \in \mathrm{BD}(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))$ to have symmetric values, can be argued to be very natural. In our forthcoming applications we are, however, interested in symmetric tensors, so our results above are stated for $\mathrm{BV}(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))$. They nevertheless hold for $\mathrm{BV}(\Omega;\mathcal{T}^k(\mathbb{R}^m))$ and the analogously defined $\mathrm{BD}(\Omega;\mathcal{T}^k(\mathbb{R}^m))$ as well.

3. Regularisation of tensor fields

We now begin the study of regularisation models for tensor fields. We concentrate on models in the class (1.1), reminiscent of the Rudin-Osher-Fatemi (ROF) regularisation model for scalar fields. We therefore begin by recalling this model.

3.1. For recollection: TV and ROF for scalar fields

Let $\Omega \subset \mathbb{R}^m$ be a domain and $u \in L^1(\Omega)$. We write the *total variation* of u as

$$\mathrm{TV}(u) := \sup_{\varphi\in V} \int_\Omega u(x) \operatorname{div}\varphi(x)\,dx = \|Du\|_{\mathcal{M}(\Omega)},$$

where

$$V := \{\varphi \in C_c^\infty(\Omega;\mathbb{R}^m) \mid \|\varphi\|_\infty \le 1\}.$$

Given a regularisation parameter $\alpha > 0$, the ROF regularisation of $f \in L^1(\Omega)$ is then given by the solution $\hat{u}$ of the problem

$$\min_{u\in\mathrm{BV}(\Omega)} \frac{1}{2}\|f-u\|^2_{L^2(\Omega)} + \alpha\mathrm{TV}(u).$$

3.2. Total variation and ROF for tensor fields

We now extend ROF to tensor fields. Given a domain $\Omega \subset \mathbb{R}^m$ and $u \in L^1(\Omega; \mathcal{T}^k(\mathbb{R}^m))$, we write

$$\mathrm{TV}(u) := \sup_{\varphi \in V^{k+1}_{F,\mathrm{ns}}} \int_\Omega \langle u(x), \operatorname{div} \varphi(x) \rangle \, dx = \|Du\|_{F,\mathcal{M}(\Omega;\mathcal{T}^{k+1}(\mathbb{R}^m))}.$$

Here we recall the defining equation (2.6) of

$$V^{k+1}_{F,\mathrm{ns}} := \{\varphi \in C^\infty_c(\Omega; \mathcal{T}^{k+1}(\mathbb{R}^m)) \mid \|\varphi\|_{F,\infty} \leq 1\}.$$

Observe that we bound φ pointwise by the Frobenius norm. The reason for this is that we desire the rotation-invariance detailed in Proposition 2.3. Alternatively we could take the largest or smallest reasonable cross-norm, but these norms are computationally demanding for tensors of degree greater than two; compare to the related tensor decompositions in [26]. In our application $\varphi(x)$ has degree three, so we concentrate on the Frobenius norm.

For $\alpha > 0$, a *positive semi-definite* ROF-type regularisation of $f \in L^1(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$ is now given by the problem

$$\min_{0 \leq u \in L^1(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))} \frac{1}{2}\|f - u\|^2_{F,2} + \alpha \mathrm{TV}(u). \tag{P-TV}$$

Although derived by other means, and superficially different, it turns out that the "component-based regularisation" of [30] for $\Psi = \sqrt{\cdot}$, is very similar to (P-TV). The difference is that the former lacks the positive semi-definiteness constraint. If the data f is positive semi-definite, then the constraint can indeed be shown to be superfluous, and the two problems equivalent.

Denoting by

$$\delta_A(x) := \begin{cases} 0, & x \in A, \\ \infty, & x \notin A, \end{cases}$$

the indicator function of a set A in the sense of convex analysis, and particularly by

$$\delta_{\geq 0}(u) := \begin{cases} 0, & u(x) \text{ is positive semi-definite for a.e. } x \in \Omega, \\ \infty, & \text{otherwise}, \end{cases}$$

the indicator function of the pointwise positive semi-definite cone, the problem (P-TV) may also be given the saddle-point formulation

$$\min_{u \in L^1(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))} \max_{\varphi \in C^\infty_c(\Omega;\mathcal{T}^{k+1}(\mathbb{R}^m))} \Big(\frac{1}{2}\|f - u\|^2_{F,2} + \delta_{\geq 0}(u) + \langle u, K^*\varphi \rangle - \delta_{\alpha V^{k+1}_{F,\mathrm{ns}}}(\varphi) \Big) \tag{S-TV}$$

where $K^*\varphi := -\operatorname{div} \varphi$, and the indicator function $\delta_{\alpha V^{k+1}_{F,\mathrm{ns}}}$ takes the role of the constraint $\varphi \in \alpha V^{k+1}_{F,\mathrm{ns}}$ on the dual variable in the definition of $\mathrm{TV}(u)$. The conjugate-like notation K^* will be justified in Section 4, where we study the numerical solution of (P-TV) through the formulation (S-TV).

3.3. Total deformation regularisation of tensor fields

We may also restrict φ to be symmetric, that is

$$\varphi \in V^{k+1}_{F,\mathrm{s}} \subset V^{k+1}_{F,\mathrm{ns}},$$

yielding a symmetrised TV variant, called *total deformation* and defined as

$$\mathrm{TD}(u) := \sup_{\varphi \in V^{k+1}_{F,\mathrm{s}}} \int_\Omega \langle u(x), \operatorname{div} \varphi(x) \rangle \, dx = \|Eu\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))}.$$

With then readily arrive at the regularisation problem

$$\min_{0\leq u\in L^1(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))} \frac{1}{2}\|f-u\|_{F,2}^2 + \alpha \mathrm{TD}(u), \tag{P-TD}$$

and the equivalent saddle-point problem

$$\min_{u\in L^1(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))} \max_{\varphi\in C_c^\infty(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} \Big(\frac{1}{2}\|f-u\|_{F,2}^2 + \delta_{\geq 0}(u) + \langle u, K^*\varphi\rangle - \delta_{\alpha V_{F,\mathrm{s}}^{k+1}}(\varphi)\Big), \tag{S-TD}$$

where again $K^*\varphi := -\operatorname{div}\varphi$. The difference to (S-TV) is that φ is constrained to be symmetric. Indeed, observe that a symmetric $\varphi(x)$ has significantly less degrees of freedom than a non-symmetric one; in $\mathbb{R}^2$, for example, a generic $(k+1)$-tensor requires 2^{k+1} scalars to represent, while a symmetric $(k+1)$-tensor can be represented by $k+2$ scalars. Already for $k=2$ this translates to 8 versus 4. This has an influence on numerical efficiency.

3.4. Second-order total generalised variation (TGV^2) for tensor fields

Total generalised variation was introduced in [10] as a higher-order extension of TV. The application to magnetic resonance imaging, in particular, is studied in [24], yielding improved results in comparison to TV.

For a scalar field $u\in L^1(\Omega)$, second-order TGV may according to [11, 12] be written as the "differentiation cascade"

$$\mathrm{TGV}^2_{(\beta,\alpha)}(u) := \min_{w\in L^1(\Omega;\mathrm{Sym}^1(\mathbb{R}^m))} \alpha\|Eu-w\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^1(\mathbb{R}^m))} + \beta\|Ew\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^2(\mathbb{R}^m))}, \tag{3.1}$$

where the parameters $\alpha,\beta>0$. Observe that $Du=Eu$ for scalar fields $u\in L^1(\Omega)=L^1(\Omega;\mathcal{T}^0(\mathbb{R}^m))$.

Remark 3.1. The reason for writing the parametrisation in the order (β,α) instead of (α,β), is that the differentiation cascade formulation in some sense reverses the parameters in comparison to the original dual-ball formulation

$$\mathrm{TGV}^2_{(\beta,\alpha)}(u) := \sup\Big\{\int_\Omega u\operatorname{div}^2\varphi\,dx \;\Big|\; \varphi\in C_c^2(\Omega;\mathrm{Sym}^2(\mathbb{R}^m)),\ \|v\|_{F,\infty}\leq\beta,\ \|\operatorname{div} v\|_{F,\infty}\leq\alpha\Big\}. \tag{3.2}$$

Remark 3.2 (A quick word aside on topology)**.** It turns out that $\|u\|_{L^1}+\mathrm{TV}(u)$ and $\|u\|_{L^1}+\mathrm{TGV}^2_{(\beta,\alpha)}(u)$ are topologically equivalent norms [11, 12]. Geometrically they are very different however: compare the unit balls for the 1, 2, and ∞ norms on $\mathbb{R}^2$. For one, regularisation by TGV^2 tends to avoid the stair-casing that TV exhibits [10]. For another, the kernel of TGV^2 is larger, including the so-called infinitesimal rigid displacements $A=-A^T$.

Readily the above definition extends to $u\in L^1(\Omega;\mathcal{T}^k(\mathbb{R}^m))$ as

$$\mathrm{TGV}^2_{(\beta,\alpha)}(u) := \min_{w\in L^1(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} \alpha\|Eu-w\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} + \beta\|Ew\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+2}(\mathbb{R}^m))}. \tag{3.3}$$

A positive semi-definite TGV^2-regularisation of a tensor field $f\in L^1(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))$ may now be defined as a solution of

$$\min_{0\leq u\in L^1(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))} \frac{1}{2}\|f-u\|_{F,2}^2 + \mathrm{TGV}^2_{(\beta,\alpha)}(u). \tag{P-TGV^2}$$

This can be further written as the saddle-point problem

$$\min_{u,w}\max_{\varphi,\psi}\Big(\frac{1}{2}\|f-u\|_{F,2}^2 + \delta_{\geq 0}(u) + \langle (u,w), K^*(\varphi,\psi)\rangle - \delta_{\alpha V_{F,\mathrm{s}}^{k+1}}(\varphi) - \delta_{\beta V_{F,\mathrm{s}}^{k+2}}(\psi)\Big), \tag{S-TGV^2}$$

where the unknowns $u\in L^1(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))$, $w\in L^1(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))$, $\varphi\in C_c^\infty(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))$, and $\psi\in C_c^\infty(\Omega;\mathrm{Sym}^{k+2}(\mathbb{R}^m))$, while the operator K^* is defined by $K^*(\varphi,\psi):=(-\operatorname{div}\varphi, -\varphi-\operatorname{div}\psi)$.

Remark 3.3. It is also possible to define "non-symmetric TGV^2", bearing similar differences to the "symmetric TGV^2" above, as TV bears to TD. We however do not do that, because it is computationally much heavier than symmetric TGV^2, due to far greater degree of freedom. A numerical study of the effects of symmetry versus non-symmetry will be provided by the comparison between TD and TV.

3.5. Existence of solutions

We now show that the minimisation problems discussed above admit solutions, as stated by the following theorem.

Theorem 3.1. *Let $\Omega \subset \mathbb{R}^m$ be a bounded Lipschitz domain, $k \geq 0$, and $f \in L^2(\Omega; \mathcal{T}^k(\mathbb{R}^m))$. Then* (P-TV) *admits a solution $0 \leq \hat{u} \in \mathrm{BV}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$, and* (P-TD)*, and* (P-$\mathrm{TGV}^2$) *admit solutions $0 \leq \hat{u} \in \mathrm{BD}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$.*

Proof. The proof for (P-TV) and (P-TD) is quite standard; cf., e.g., [22]. Indeed, let $\{u^i \geq 0\}_{i=0}^{\infty} \subset L^1(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$ be a minimising sequence for (P-TD). Observe that

$$\sup_i \|u^i\|_{F,2} + \|Eu^i\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} < \infty. \tag{3.4}$$

Therefore, as in the case of scalar functions, we deduce that there exists a subsequence, unrelabelled, and $0 \leq \hat{u} \in \mathrm{BD}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$, such that $u^i \rightharpoonup \hat{u}$ weakly in $L^2(\Omega; \mathcal{T}^k(\mathbb{R}^m))$, and $Eu^i \overset{*}{\rightharpoonup} E\hat{u}$ weakly* in $\mathcal{M}(\Omega; \mathrm{Sym}^{k+1}(\mathbb{R}^m))$. Lower-semicontinuity of $\|\cdot\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))}$ and of $\|\cdot\|_{F,2}$ now shows that $\hat{u}$ is a solution to (P-TD).

The claim for (P-TV) follows analogously in $\mathrm{BV}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$, with with Eu replaced by Du above.

It remains to show the existence of a solution $0 \leq \hat{u} \in \mathrm{BD}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$, $\hat{u} \geq 0$, to (P-TGV^2). As in the scalar case $k = 0$, considered in [12], we need to show that for some $c > 0$ it holds

$$c\big(\|u\|_{F,1} + \|Eu\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))}\big) \leq \|u\|_{F,1} + \mathrm{TGV}^2_{(\beta,\alpha)}(u), \quad (u \in L^1(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))), \tag{3.5}$$

and that $\mathrm{TGV}^2_{(\beta,\alpha)}$ is lower-semicontinuous with respect to weak* convergence of Eu^i to Eu. These results are contained in the following two lemmas. We let then $\{u^i \geq 0\}_{i=1}^{\infty} \subset L^1(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$ be a minimising sequence for (P-TGV^2). We want to show that (3.4) holds. Indeed, $\sup_i \|u^i\|_{F,1} < \infty$ thanks to Ω being bounded and

$$\sup_i \|f - u^i\|_{F,2} + \mathrm{TGV}^2_{(\beta,\alpha)}(u^i) < \infty, \tag{3.6}$$

the latter following from $\{u^i\}$ being a minimising sequence for (P-TGV^2). It now follows from (3.5) that $\sup_i \|Eu^i\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} < \infty$. This and (3.6) lead to (3.4). The same arguments as above now provide $0 \leq \hat{u} \in \mathrm{BD}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$, and lower-semicontinuity establishes that it solves (P-TGV^2). □

Lemma 3.1. *Let $\Omega \subset \mathbb{R}^m$ be a bounded Lipschitz domain and $k \geq 0$. Then there exist constants $c, C > 0$, dependent on Ω, k, m, such that for all $u \in L^1(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$ it holds*

$$c\|u\|_{\mathrm{BD}(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))} \leq \|u\|_{F,1} + \mathrm{TGV}^2_{(\beta,\alpha)}(u) \leq C\|u\|_{\mathrm{BD}(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))}. \tag{3.7}$$

Proof. The proof is a straightforward extension of the the equivalence proof for $k = 0$ in [12], employing the tensor Sobolev-Korn estimates from [9]. Indeed, by definition

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

for all $w \in L^1(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))$, so setting $w = 0$ gives

$$\|u\|_{F,1} + \mathrm{TGV}^2_{(\beta,\alpha)}(u) \leq \|u\|_{F,1} + \alpha\|Eu\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} \leq C\|u\|_{\mathrm{BD}(\Omega;\mathrm{Sym}^{k}(\mathbb{R}^m))}$$

for $C = \max\{1, \alpha\}$. Thus the second inequality of (3.7) holds.

For the first inequality of (3.7), we may assume that $Eu \in \mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))$, since otherwise $\|Eu - w\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} = \infty$ for all $w \in L^1(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))$, and the inequality holds trivially. We want to show that there exists $C_1 > 0$ such that

$$\|Eu\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} \leq C_1\big(\|u\|_{F,1} + \|Eu - \bar{w}\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))}\big) \tag{3.8}$$

for every $u \in \mathrm{BD}(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))$ and $\bar{w} \in L^1(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))$ satisfying $E\bar{w} = 0$, i.e., $\bar{w} \in \ker E$. For the proof of the fact that E has a non-trivial finite-dimensional kernel, see [9]. Indeed, suppose that (3.8) does not hold. Then there exist sequences $\{u^i\}_{i=0}^\infty \subset \mathrm{BD}(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))$ and $\{\bar{w}^i\}_{i=0}^\infty \subset L^1(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m)) \cap \ker E$ such that for $i = 0, 1, 2, \ldots$, it holds

$$\|Eu^i\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} = 1 \quad \text{and} \quad \|u^i\|_{F,1} + \|Eu^i - \bar{w}^i\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} \leq 1/i, \tag{3.9}$$

It follows that $u^i \to 0$ strongly in $L^1(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))$. Consequently $Eu^i \overset{*}{\rightharpoonup} 0$ weakly* in the space $\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))$. Employing the first half of (3.9) in the second, it moreover follows that $\sup_i \|\bar{w}^i\|_{F,1} < \infty$. Since $\ker E$ is finite-dimensional, we deduce that there exists a convergent subsequence, unrelabelled, and $\bar{w} \in L^1(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m)) \cap \ker E$, such that $\bar{w}^i \to \bar{w}$ strongly in $L^1(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))$. It hence follows from (3.9) that $Eu^i \to \bar{w}$ strongly in $\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))$. But $Eu^i \overset{*}{\rightharpoonup} 0$, so $\bar{w} = 0$. This means that $\|Eu^i\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} \to 0$ in contradiction to (3.9). Hence (3.8) holds.

Now we employ from [9] the following Sobolev-Korn estimate: there exists $C_2 > 0$ such that for all $w \in \mathrm{BD}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))$ there exists $\bar{w} \in L^1(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m)) \cap \ker E$ satisfying

$$\|w - \bar{w}\|_{F,1} \leq C_2\|Ew\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+2}(\mathbb{R}^m))}.$$

For this choice of $\bar{w}$, we deduce the existence of $C_3 > 0$ such that for all $u \in \mathrm{BD}(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))$ and $w \in L^1(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))$ it holds

$$\begin{aligned}\|Eu - \bar{w}\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} &\leq \|Eu - w\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} + \|w - \bar{w}\|_{F,1} \\ &\leq C_3\big(\alpha\|Eu - w\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} + \beta\|Ew\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+2}(\mathbb{R}^m))}\big).\end{aligned}$$

Employing this estimate in (3.8) yields for some $C_4 > 0$ the estimate

$$\|u\|_{\mathrm{BD}(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))} \leq C_4\big(\|u\|_{F,1} + \alpha\|Eu - w\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} + \beta\|Ew\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+2}(\mathbb{R}^m))}\big),$$

which holds for all $w \in L^1(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))$ and $u \in \mathrm{BD}(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))$. Hence the first inequality of (3.7) holds with $c = C_4^{-1}$. □

Lemma 3.2. *Let $\Omega \subset \mathbb{R}^m$ be a bounded Lipschitz domain and $k \geq 0$. Then the function*

$$F(\mu) := \min_{w \in L^1(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} \alpha\|\mu - w\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} + \beta\|Ew\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+2}(\mathbb{R}^m))},$$

where $\mu \in \mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))$, is lower-semicontinuous with respect to weak convergence.*

Proof. Let $\mu^i \overset{*}{\rightharpoonup} \mu$ weakly* in $\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))$. Observe that by the Banach-Steinhaus theorem, $\sup_i \|\mu^i\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} < \infty$. Consequently also $\sup_i F(\mu^i) < \infty$.

We first establish that $F(\mu^i)$ admits a minimiser $\hat{w}^i \in \mathrm{BD}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))$. Indeed, let $\{v^j\}_{j=0}^\infty \subset \mathrm{BD}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))$ be a minimising sequence for $F(\mu^i)$. The sequence is obviously bounded in $\mathrm{BD}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))$. Thus [9, Theorem 4.17], establishes that there exists a subsequence, unrelabelled, convergent strongly in $L^1(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))$ to some $v \in L^1(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))$. A standard argument (cf., e.g, [4, Proposition 3.13]) establishes that $Ev^i \overset{*}{\rightharpoonup} Ev$ weakly* in $\mathcal{M}(\Omega;\mathrm{Sym}^{k+2}(\mathbb{R}^m))$. Finally, lower-semicontinuity yields that $\hat{w}^i := v$ minimises $F(\mu^i)$.

Knowing that $F(\mu^i)$ admits a minimiser for each $i = 0, 1, 2, \ldots$, we now establish lower-semicontinuity. Since $\sup_i F(\mu^i) + \|\mu^i\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} < \infty$, we deduce that

$$\sup_i \|\hat{w}^i\|_{F,1} + \|E\hat{w}^i\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} < \infty.$$

Hence $\{\hat{w}^i\}_{i=0}^\infty$ converges weakly in $\mathrm{BD}(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))$ to some $\hat{w} \in \mathrm{BD}(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))$. By lower-semicontinuity of norms we deduce that

$$\begin{aligned} F(\mu) &\leq \alpha\|\mu - \hat{w}\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} + \beta\|E\hat{w}\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+2}(\mathbb{R}^m))}, \\ &\leq \liminf_{i\to\infty} \alpha\|\mu^i - \hat{w}^i\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} + \beta\|E\hat{w}^i\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+2}(\mathbb{R}^m))} = \liminf_{i\to\infty} F(\mu^i). \end{aligned}$$

This establishes the lower semicontinuity of F. □

Remark 3.4. If we extended to the tensor case the equivalence proof [11, 12] of the differentiation cascade formulation (3.1) of $\mathrm{TGV}^2_{(\beta,\alpha)}$, and the original dual-ball formulation (3.2), then, following the original proof in [10], we could almost trivially obtain lower-semicontinuity of $\mathrm{TGV}^2_{(\beta,\alpha)}$ with respect to convergence in L^p. This would imply weak lower-semicontinuity in L^1, and could be used to replace Lemma 3.2. in the proof of Theorem 3.1. The equivalence proof is, however, very long, and not our focus here, so we do not provide the extension, and choose to work entirely with the differentiation cascade formulation, that is more practical in the numerical methods of our choosing.

4. Algorithmic aspects

We now move on to discuss the algorithmic aspects of the solution of the regularisation problems above. We do this through the saddle-point formulations.

4.1. Discretisation and the algorithm

The problems (S-TV), (S-TD), and (S-TGV^2) are of the form

$$\min_x \max_y G(x) + \langle x, K^* y\rangle - F^*(y)$$

for proper convex lower-semicontinuous G, F^*. This suggests that the Chambolle-Pock algorithm [14] could be applied. A problem with the original infinite-dimensional problems is, however, that a (pre)conjugate of K^* cannot easily be defined, as the spaces involved are not reflexive; in the case of TV, in particular $K^* : C_c^\infty(\Omega;\mathcal{T}^3(\mathbb{R}^m)) \to C_c^\infty(\Omega;\mathcal{T}^2(\mathbb{R}^m))$. In practise the algorithm is applied on finite dimensional discretisations, however, and this problem does not surface when the discretisations are chosen suitably. We choose to represent each tensor field f, u, w, φ and ψ by values on an uniform grid, and discretise K^* by forward differences, yielding the operator K_h^*. We then take $K_h := (K_h^*)^*$ as the discrete conjugate of K_h^*.

Minding that G is uniformly convex, we may and do apply the accelerated version of the Chambolle-Pock algorithm, stated below.

Assumption 4.1. Consider the problem

$$\min_x \max_y G(x) + \langle x, K_h^* y\rangle - F^*(y),$$

where $K_h : X \to Y$ is a continuous linear operator between the finite-dimensional Hilbert-spaces X and Y, and $G : X \to [0, +\infty]$ and $F^* : Y \to [0, +\infty]$ are proper, convex and lower-semicontinuous, with F^* the conjugate of a convex lower-semicontinuous function F. Assume, moreover, that G is uniformly convex in the following sense: there exists $\gamma > 0$ such that for any $x \in \operatorname{dom} \partial G$ it holds

$$G(x') - G(x) \geq \langle z, x' - x\rangle + \frac{\gamma}{2}\|x - x'\|^2 \quad \text{for all } z \in \partial G(x),\ x' \in X.$$

Algorithm 4.1. Suppose Assumption 4.1 is satisfied. Following [14], perform the steps:

1. Pick $\tau_0, \sigma_0 > 0$ satisfying $\tau_0\sigma_0\|K_h\|^2 \leq 1$, as well as $(x^0, y^0) \in X \times Y$. Set $\bar{x}^0 = x^0$.
2. For $i = 0, 1, 2, \ldots$, repeat the following updates until a stopping criterion is satisfied.

$$\begin{aligned} y^{i+1} &:= (I + \sigma_i \partial F^*)^{-1}(y^i + \sigma_i K_h \bar{x}^i) \\ x^{i+1} &:= (I + \tau_i \partial G)^{-1}(x^i - \tau_i K_h^* y^{i+1}) \\ \theta_i &:= (1 + 2\gamma\tau_i)^{-1/2},\ \tau_{i+1} := \theta_i\tau_i,\ \sigma_{i+1} := \sigma_i/\theta_i \\ \bar{x}^{i+1} &:= x^{i+1} + \theta_i(x^{i+1} - x^i). \end{aligned}$$

The resolvent operators that need to be calculated to obtain x^{i+1} and y^{i+1}, may be written

$$(I + \tau\partial G)^{-1}(x) = \arg\min_y \left\{ \frac{\|x - y\|^2}{2\tau} + G(y) \right\}.$$

The efficient realisation of Algorithm 4.1 depends on the efficient realisation of these minimisation problems. In our situation, it turns out that they reduce to easily calculable projections:

We begin by considering (S-TV). First, for $F^*(\varphi) = \delta_{\alpha V^{k+1}_{F,\mathrm{s}}}(\varphi)$, the resolvent is

$$(I + \sigma\partial F^*)^{-1}(v) = \arg\min_\varphi \left\{ \frac{\|v - \varphi\|^2_{F,2}}{2\sigma} + \delta_{\alpha V^{k+1}_{F,\mathrm{s}}}(\varphi) \right\}.$$

This reduces to a pointwise projection

$$\varphi(x) = P_{\|\cdot\|_F \leq \alpha}(v(x)) = v(x)\min\{1, \alpha/\|v(x)\|_F\} \tag{4.1}$$

for all $x \in \Omega$. Secondly, we have

$$G(u) = G_0(u) := \frac{1}{2}\|f - u\|^2_{F,2} + \delta_{\geq 0}(u),$$

for which we solve

$$[(I + \tau\partial G_0)^{-1}(v)](x) = P_{\geq 0}\left(\frac{v(x) + f(x)\tau}{1 + \tau} \right), \quad (x \in \Omega).$$

The projection $P_{\geq 0}$ to the positive-definite cone can be performed by projecting each eigenvalue to $\mathbb{R}^+$. (This can be seen from the structure of the normal cone $N_{\geq 0}(x')$, spelled out in, e.g., [36, Lemma 3.1]. See also [27] for related eigenvalue projection results that are useful for dealing with the nuclear and spectral norms.)

Regarding (S-TD), still $G = G_0$, and also the resolvent

$$(I + \sigma\partial F^*)^{-1}(v) = \arg\min_\varphi \left\{ \frac{\|v - \varphi\|^2_{F,2}}{2\sigma} + \delta_{\alpha V^{k+1}_{F,\mathrm{ns}}}(\varphi) \right\}$$

has the same solution (4.1) as in the case of (S-TV).

It remains to consider the resolvents for (S-TGV^2). Minding that $G(u,w) = G_0(u)$, we find that

$$(I + \tau\partial G)^{-1}(v,q) = ((I + \tau\partial G_0)^{-1}(v), q).$$

Likewise, from the expression

$$F^*(\varphi,\psi) = \delta_{\alpha V^{k+1}_{F,\mathrm{s}}}(\varphi) + \delta_{\beta V^{k+2}_{F,\mathrm{s}}}(\psi),$$

we immediately deduce that

$$(I + \sigma\partial F^*)^{-1}(v,q) = (\varphi,\psi),$$

with the projection (4.1) applied on φ and ψ separately; more precisely $\varphi(x) = P_{\|\cdot\|_F\leq\alpha}(v(x))$ and $\psi(x) = P_{\|\cdot\|_F\leq\beta}(q(x))$.

4.2. Duality gap as stopping criterion

For (S-TV) and (S-TD) it poses no difficulty to calculate the duality gap

$$F(K_h u) + G(u) + G^*(-K_h^*\varphi) + F^*(\varphi),$$

and to use the reduction of the duality gap beyond a certain threshold as a stopping criterion. Regarding (S-TGV^2), the variable w from the expression $\mathrm{TGV}^2(u) = \min_w \alpha\|Eu - w\| + \beta\|Ew\|$ does not appear in $G(u,w) = G_0(u)$, yielding $G^*(a,b) = G_0^*(a) + \delta_{\{0\}}(b)$. The result is that in practise $G^*(a,b) = \infty$ in the algorithm, so the duality gap as such is not useful for a stopping criterion.

If we had an a priori bound M on $\|\hat{w}\|_{F,1}$, at an optimal solution $(\hat{u},\hat{w})$, then we could add the term $\delta_{B(0,M)}(\|w\|_{F,1})$ to G, resulting in practical G^* and duality gap. It can be shown (see Proposition A.1 in the Appendix) that such a bound indeed exists. Unfortunately, however, due to the nature of the proof, we only know the existence, but not the exact magnitude.

Fortunately it turns out that we can also pick M a posteriori, because if M is large enough, it vanishes in the Chambolle-Pock algorithm, and therefore, if the duality gap becomes infinite, we can simply increase M, *which is only used to calculate the duality gap/stopping criterion.* This suggests to employ the following generic algorithm, where for TGV^2 we have $x^i = (u^i, w^i)$, $y^i = (\varphi^i, \psi^i)$, and

$$U := \{(u,w) \in L^1(\Omega; \mathrm{Sym}^k(\mathbb{R}^m)) \times L^1(\Omega; \mathrm{Sym}^{k+1}(\mathbb{R}^m)) \mid \|w\|_{F,1} \leq 1\}.$$

The idea is that we always decrease the duality gap of the modified problem, where G is replaced by $G(x) + \delta_{MU}(x)$, for *some* $M > 0$, *unknown a priori,* by a given fraction ρ, chosen a priori.

Algorithm 4.2. Suppose Assumption 4.1 holds, and that $U \subset X$ has non-empty interior. In each step of the Algorithm 4.1, perform the following additional operations.

1. Pick $\rho \in (0,1)$ and $M_0 \geq 0$. Define

$$G_i(x) := G(x) + \delta_{M_i U}(x).$$

2. Update the variables as in Algorithm 4.1 with $G = G_i$. Pick $M_{i+1} \geq M_i$ large enough that $x \in M_{i+1}E$. Calculate the initial and current pseudo-duality gaps

$$d_0^{i+1} := F(K_h x^0) + G_{i+1}(x^0) + G_{i+1}^*(-K_h^* y^0) + F^*(y^0),$$
$$d^{i+1} := F(K_h x^{i+1}) + G_{i+1}(x^{i+1}) + G_{i+1}^*(-K_h^* y^{i+1}) + F^*(y^{i+1}).$$

   If $d^{i+1} < \rho d_0^{i+1}$, finish execution of the algorithm, with the solution (x^{i+1}, y^{i+1}). Otherwise continue iteration.

4.3. Computations on 2D slices of 3D data

Our mathematical work on tensor fields above, in particular the definition (2.4) of the tensor divergence, only applies to tensors fields $f : \Omega \to \mathrm{Sym}^k(\mathbb{R}^m)$, where the domain $\Omega \subset \mathbb{R}^m$ has the same dimension m as the tensor parameters. Thus the semi-norm $\|Df\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))}$ is also only defined for tensor fields of this type. This limitation has the consequence that our regularisation models cannot directly be applied to 2D slices of 3D diffusion tensor measurements $f : \Omega \to \mathcal{T}^2(\mathbb{R}^3)$ for $\Omega \subset \mathbb{R}^2$. We, nevertheless, want to perform computations on such slices. There are two reasons for that. For one, often (but not always) the resolution of DTI measurements in the z direction is so low, that no "true" 3D voxels exist. For example, an in-vivo brain measurement may have dimensions $128 \times 128 \times 32$, consisting essentially of 32 unconnected 128×128 slices. In fact, even in high-resolution data sets there may be "slice gaps" in the data sets in order to reduce signal leakage. This means that, for example, one slice is 2mm thick, then there is a 1mm gap, and then another 2mm thick slice. The other reason for wanting to perform computations on 2D slices is that 3D computations are much more demanding of computational resources.

A discrete $N \times M \times 1$ data slice can, of course, be seen as a 3D volume. This view, however, introduces the difficulty of defining the discretised differential in the z direction, there being no additional slices to compare to. Simple forward or backward differences with zero boundary conditions – as is required from the dual variable φ (and ψ) – would give as the differential $-\varphi$ or $+\varphi$, while central differences would give the vastly different differential 0. What we choose to do, is to simply choose the differential in the z direction as zero, irrespective of any discrete differentiation scheme. This approach is consistent with the measurement voxels having low resolution in the z direction in comparison to the x and y directions: If we were to subdivide the low-resolution voxels in the z direction into L parts, the $N \times M \times L$ problem would then for optimal solutions have essentially zero differentials on the central slices in the z direction, and therefore split into distinct $N \times M \times 1$ sub-problems.

In summary, when we have a $N \times M \times 1$ slice of 3D tensors, we perform calculations otherwise as if we had a full 3D volume of data, but calculate the differentials only in the x and y directions.

5. Numerical results

We will now numerically study the performance of the different regularisation models on a synthetic test data, as well as an in-vivo brain measurement. We first discuss how the results are reported, followed by the construction of the synthetic test data. We then detail how the computational algorithm is parametrised, and finally represent and analyse the results.

5.1. Error measures

We would like to have a numerical value for the quality of the solution, compared to noise-free test data. An obvious candidate is, of course, the Frobenius-L^2 norm

$$d_F(f, u) := \|f - u\|_{F,2}.$$

This distance is, however, difficult to interpret in geometric terms directly related to f and u. Therefore we introduce three other error measures that measure different geometrical aspects of the tensors.

As the first geometrical error measure we have the L^2 norm

$$d_A(f, u) := \|\mathrm{FA}_f - \mathrm{FA}_u\|_{L^2(\Omega)}$$

of the differences of the fractional anisotropies, defined by

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

where $\lambda_1, \ldots, \lambda_m$ are the eigenvalues of $u(x)$, and $\bar{\lambda} = \sum_{i=1}^m \lambda_i/m$. The second geometrical error measure is the L^2 norm

$$d_\lambda(f,u) := \|\hat{\lambda}_u - \hat{\lambda}_f\|_{L^2(\Omega)}$$

of the differences of the principal eigenvalues $\hat{\lambda}_u(x)$ of $u(x)$ and $f(x)$. As the final geometrical error measure we take the L^2 norm

$$d_v(f,u) := \|\chi_U(1 - |\langle \hat{v}_u, \hat{v}_f \rangle|)\|_{L^2(\Omega)}$$

of the differences of the normalised principal eigenvectors $\hat{v}_u(x)$ of $u(x)$ and $f(x)$. (We also have the expression $d_v(f,u) = \min 2\|\chi_U(\hat{v}_u \pm \hat{v}_f)\|_{L^2(\Omega;\mathbb{R}^m)}$ if the minimum is taken pointwise over the sign.) Since for a fully isotropic tensor the direction of the principal eigenvector is completely undetermined, being able to lie anywhere on the unit sphere, it is not meaningful to compare the principal eigenvectors of such tensors. We therefore employ

$$U := \{x \in \Omega \mid \mathrm{FA}_u(x) \geq 0.01 \text{ or } \mathrm{FA}_f(x) \geq 0.01\}$$

in the above expression to filter out such points.

The distance d_λ measures the error in the size of the tensors, i.e., the lengths of the principal axes of the corresponding ellipsoids. The distance d_v measures the error in the directions of the principal axes, and, finally, d_A measures the error in the shape of the ellipsoids.

5.2. Visualisation

The visualisation of tensor images, each point consisting of multiple values, is not trivial, and the choice of visualisation highly affects what we can learn. The most obvious approach (for 2D slices) is to plot the ellipses or ellipsoids corresponding to the tensor, but in practise such a visualisation is very difficult to read except for very small images. Fractional anisotropy, as a scalar quantity, can easily be studied, and provides useful information, but is not sufficient by itself. Superimposing the principal eigenvector on the fractional anisotropy is a relatively common means of visualisation, but can also become difficult to read for high-resolution data.

Colour-coding of different tensor quantities is another common means of visualisation. This is what we have chosen to principally use in the present paper, as suitable colour-coding allows to easily observe differences between the different regularisation models. Particularly, for plotting the computational results and source data, we have chosen to use as a basis the standard coding (cf. [33]) of mapping the (x, y, z) components of the 3D principal eigenvector ($z = 0$ for 2D data) directly to the (red, green, blue) channels of the RGB colour model. Additionally, to increase the information available in the plots, we have modulated this unit vector by a function of the fractional anisotropy. Namely, we have

$$\mathrm{RGB} = \hat{v}_u \cdot \min\{1, \mathrm{FA}_u + 1/3\}.$$

The effect of the nonlinearities in the fractional anisotropy modulation is to stop features of interest in highly isotropic areas from disappearing, while still making them significantly darker than highly anisotropic areas. Explanatory plots of this colour-coding are included in Figure 2 (2D) and Figure 4 (3D).

It turns out, however, that analysing the error between the original noise-free data f_0 and the computational result u is more useful for comparing the different computational models. Following,

e.g., [2], we plot the angle $\theta = \cos^{-1}(\langle \hat{v}_u, \hat{v}_{f_0} \rangle)$ between the principal eigenvectors of f_0 and u, using the "jet" colour map of Matlab, spanning from blue through cyan and yellow to red. This coding does not yet describe errors $e_{\mathrm{FA}} = |\mathrm{FA}_u - \mathrm{FA}_{f_0}|$ in reconstruction of fractional anisotropy, and therefore we have chosen to plot that as well, using shades of grey, so that the resulting RGB (red, green, blue) value of an image pixel is the componentwise maximum of the two colours,

$$\mathrm{RGB} = \max\{\mathrm{jet}(\min\{1, 2\theta/\pi\}), \min\{1, e_{\mathrm{FA}}/0.15\}\}.$$

The minimum expressions serve to compress the colours for high errors. A clarifying plot of the colour-coding is included in Figures 3 and 5.

5.3. Synthetic data

Our noise-free synthetic test data consists of a 128×128 image of $\mathrm{Sym}^2(\mathbb{R}^2)$ tensors, divided into four smaller rectangles each having a tensor field with different properties; Figure 2 contains an illustration.

In the rectangle $(65, 128) \times (65, 128)$ we have the constant tensors

$$f_0(x, y) = \begin{bmatrix} 1.1 & 0 \\ 0 & 0.9 \end{bmatrix}.$$

(We avoid the identity tensor due to ambiguity of a principal eigenvector.) This region is in the kernel of both differential operators D and E. Next there is a smaller lower-left rectangle $(1, 64) \times (1, 64)$ consisting of an affine tensor field

$$f_0(x, y) = I + 0.005 \left(\begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix} (x - 1) + \begin{bmatrix} -2 & -1 \\ -1 & 0 \end{bmatrix} (y - 1) \right).$$

In this region $Ef_0 = 0$ but $Df_0 \neq 0$. Then, we have the upper-left rectangle $(1, 64) \times (65, 128)$ that has an affine tensor field

$$f_0(x, y) = I + 0.02 \left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} (x - 1) + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} (y - 65) \right).$$

In this region both $Df_0 \neq 0$ and $Ef_0 \neq 0$. Finally, in the lower-right rectangle $(65, 128) \times (1, 64)$, we have the non-linear tensor field

$$f_0(x, y) = R_x \begin{bmatrix} 3/4 & 0 \\ 0 & 1/2 \end{bmatrix} R_x^T, \quad R_x := \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cos\left(\frac{\pi}{2} \frac{x - 65}{64} \right) + \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \sin\left(\frac{\pi}{2} \frac{x - 65}{64} \right),$$

rotating the tensor $\left[\begin{smallmatrix} 3/4 & 0 \\ 0 & 1/2 \end{smallmatrix}\right]$ by an angle between $[0, \pi/2]$ as the x-coordinate varies.

5.4. The Stejskal-Tanner equation and Rician noise

A diffusion tensor $D \in \mathrm{Sym}^2(\mathbb{R}^3)$ at a given voxel of a diffusion tensor image produced through DWI measurements is governed by the Stejskal-Tanner equation

$$A_b = A_0 \exp(-b : D). \tag{5.1}$$

Here the b matrix parametrises a *diffusion gradient*, while A_b is the DWI measurement corresponding to this diffusion gradient, and $b : D := \sum_{i,j} b_{ij} D_{ij}$. At least $K \geq 6$ independent non-zero diffusion gradients are needed to solve D by regression from the measurements $\{A_0, A_{b^1}, \ldots, A_{b^K}\}$; see, e.g., [8].

The noise in the DWI measurements A_{b^i} is Rician; see, e.g., [20]. We wish to apply the same noise model on our synthetic test data as well. By choosing the b-matrices suitably, we can extract each individual component of D, and therefore apply Rician noise on each component separately

$$\widetilde{D}_{ij} = \log(\mathrm{ricernd}(\exp(D_{ij}), \sigma)),$$

where σ parametrises the Rice distribution, and ricernd applies random Rician noise. To the synthetic noise-free test data f_0, we apply Rician noise with the parameter $\sigma = 0.05$, yielding the noisy test data f.

5.5. The evaluated models

We evaluate the models (P-TD), (P-TV), and (P-TGV^2), applying algorithm Algorithm 4.2 to solve each of them, as discussed in the preceding sections. Moreover, we study the Pennec-like regularisation [5, 6, 19]

$$\min_{u\geq 0} \frac{1}{2}\|\log f - \log u\|^2 + \alpha \mathrm{TV}(\log u). \tag{P-logTV}$$

In practise this is implemented by taking $\hat{u} = \exp \hat{v}$ where $\hat{v}$ solves $\min_v \frac{1}{2}\|\log f - v\|^2 + \alpha \mathrm{TV}(v)$. This can be calculated with Algorithm 4.2 again, just like normal TV, but without the positive-definiteness constraint. However, we have the minor problem that $\log f(x) = \sum_{i=1}^m \lambda_i (v_i \otimes v_i)$ has complex values when $f(x)$ is not positive definite, i.e, has a non-positive eigenvalue λ_i. In principle we could calculate the distance $|\log f(x) - \log u(x)|$ in the complex sense. But, by the theory of log-Euclidean metrics, the boundary of the positive-definite cone should be infinitely far from any positive-definite tensor. Therefore, in practise, when such data occurs, we replace λ_i by a small positive number $\epsilon > 0$ for the calculation of $\log f(x)$.

For each evaluated model we pick Algorithm 4.2 as the numerical method. As the stopping criterion we use the normalised duality gap $\rho = 0.001$, and additionally limit the number of iterations to at most 5000. The initial iterates are always $x^0 = 0$, $y^0 = 0$.

For (P-TGV^2) we take $\tau_0 = \sigma_0 = 0.95/\sqrt{L_1'}$, where $L_h' = (16 + h^2 + \sqrt{h^4 + 32h^2})/(2h^2)$ is a bound on the squared norm $\|K_h^*\|_2^2$ of the discretisation K_h^* of $K^*(\varphi, \psi) := (-\operatorname{div}\varphi, -\varphi - \operatorname{div}\psi)$ on a grid of cell width h (for domains $\Omega \subset \mathbb{R}^2$!). At each iteration we update $M_{i+1} := \|w^{i+1}\|_{F,1}$.

For (P-TV), (P-TD), and (P-logTV) we take, likewise, $\tau_0 = \sigma_0 = 0.95/\sqrt{L_1}$, where this time $L_h = 8/h^2$ bounds $\|\nabla\|_2^2$ on a grid of cell width h on domains $\Omega \subset \mathbb{R}^2$; see [13].

5.6. Results on synthetic data

The results of running the algorithms described above on the synthetic test data with Rician noise of parameter $\sigma = 0.05$ are depicted in Figure 2, and the errors maps in Figure 3. Denoting the maximal eigenvalue of the test data f_0 by $\mu \approx 2.3$, and setting $\alpha_0 = 0.15$, the figure depicts the result of solving, as described above, (P-TGV^2), (P-TD), and (P-TV) for the parameters $(\beta, \alpha) = (10, 1)\alpha_0\mu$ (left half of the figure) and $(\beta, \alpha) = (0.6, 0.4)\alpha_0\mu$ (right half). For (P-logTV) we use $\alpha = \alpha_0|\log\mu|$ (left half) and $\alpha = 0.4\alpha_0|\log\mu|$ (right half). The scaling by $|\log\mu|$ instead of μ is a heuristic choice that seems to provide results more comparable to the other models than the same parameter. The parameter choice $(\beta, \alpha) = (0.6, 0.4)\alpha_0\mu$ for TGV^2 is a heuristic to compare the model against TD and TV with $\alpha = \alpha_0\mu = (0.6 + 0.4)\alpha_0\mu$; the results for TD and TV for $\alpha = 0.4\alpha_0\mu$ are also provided as a further comparison.

Our principal observations are the following. Firstly, in the top-left region, where u belongs to $(\ker E)^c$, TGV^2 well restores the diagonal line between the red and green-coloured regions, especially for the smaller parameters (right half of the figure), while TD is a bit blurry at the ends. TV and logTV display an S-shape: the diagonal bends downwards on the right and upwards on the left. Especially this is true for the larger parameters (left half of the figure). We, however, note that the fact that this shape can be seen is a result of sharp discontinuities in our visualisation; if the principal eigenvector were modulated by the fractional anisotropy directly, i.e., if $\hat{v}_u \mathrm{FA}_u$ were plotted, this S-shape could not be seen, as on the diagonal the tensors are fully isotropic, and the region around the diagonal would blend to black without the jump. What happens in our test data across the diagonal is that the principal eigenvector switches from pointing horizontally to pointing vertically. It appears that TGV^2 restores such *extremely sensitive* differences better than the other regularisation models. Studying

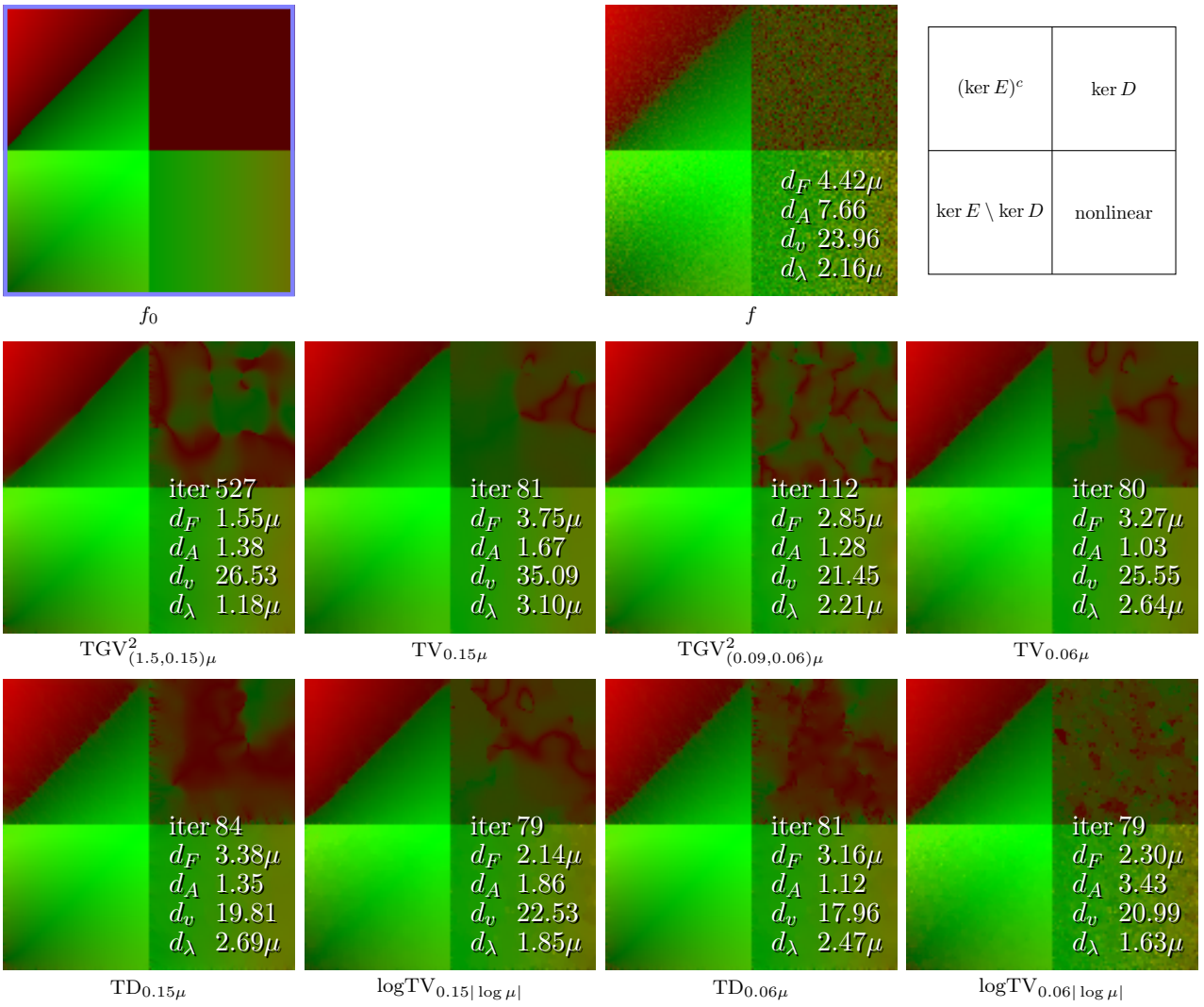

Figure 2: Results for the synthetic test data f_0 with Rician noise applied (f).

Figure 3 we observe that the reconstruction error is also noisy around the diagonal for the first-order models.

Secondly, in the top-right region, where u belongs to $\ker D$, we observe, by contrast, that TGV^2 and TD, both employing the symmetrised differential, are worse at restoring the flatness of this region, the results containing more blots of different colours than TV and logTV, employing the full differential. However, TGV^2 and TD restore the direction of the principal eigenvector better than TV and logTV, there being more shades of red in the visualisation of Figure 2 (blue in Figure 3) in this region, matching f_0.

In the bottom-left region, where u belongs to $\ker E \setminus \ker D$, the differences between TGV^2 and TD are again minor; studying Figure 3, we see that TD exhibits slightly less errors in fractional anisotropy along the interface to the top-left region. This interface is also restored poorly by TV, while logTV generally fails to restore fractional anisotropy well.

Finally, in the nonlinear bottom-right region, we observe stair-casing in the error plot of Figure 3 for all other models except TGV^2. Fractional anisotropy is again restored poorly by logTV. Still

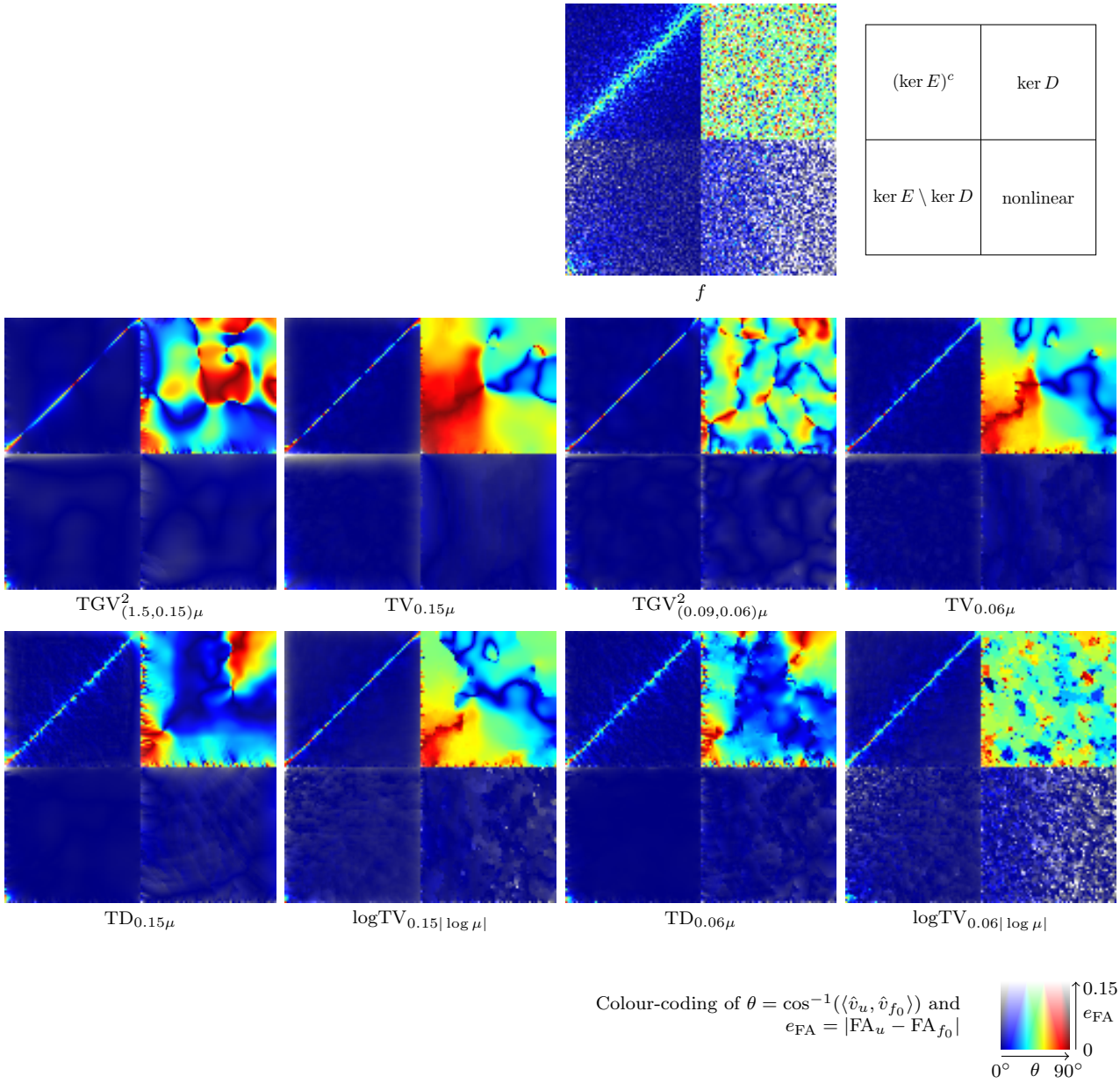

Figure 3: Error plots of computational results and noisy test data f in comparison to the noise-free synthetic test data f_0.

regarding fractional anisotropy, we observe that TGV^2 performs better than TD in this region; TD however achieves overall lower error d_A (superimposed on Figure 2) in the fractional anisotropy.

Concerning iteration counts (also superimposed on Figure 2), we observe TGV^2 for the large parameter value has required 7 times as much iterations as the other models, and for the smaller parameter also 1.5 times as many. It however provides consistently some of the best values for the error measures d_F and d_λ (again superimposed on Figure 2), while it is beaten by TD on d_A and d_v. This indicates that although TD suffers from the aforementioned blurring on the diagonal on the small scale, on the large scale, in the sense of the L^2 norm, it restores the direction of the principal eigenvector better than TGV^2. The principal eigenvalue, by contrast, is restored best by logTV and TGV^2, as seen from the values of d_λ. Minding that log-Euclidean metrics should avoid the swelling of the tensors (nonlinear interpolation of the determinant), good d_λ for logTV should also be matched by good d_A. For the larger parameter value this is the case. For the smaller parameter, by contrast, the error d_A is approximately three times as large as for the other models, while at the same time d_λ is significantly

smaller than what is achieved by the other models.

Studying the error measures we also observe that the parameter-sum heuristic $\alpha(\text{TGV})+\beta(\text{TGV}) = \alpha(\text{TD})$ for TGV^2 versus TD seems to be at least equally reasonable as choosing the same α for both models and a large β: the results for TGV^2 in the right half of the figure are comparable to TD (and TV) in the left half of the figure.

5.7. In-vivo brain measurement

Finally, we apply the regularisation models to a clinical in-vivo diffusion tensor image of a human brain. The measurements for our test data set were performed on a clinical 3T system (Siemens Magnetom TIM Trio, Erlangen, Germany), using a 32 channel head coil. Written informed consent was obtained from all volunteers before the examination. A 2D diffusion weighted single shot EPI sequence with diffusion sensitising gradients applied in 12 independent directions ($b = 1000\text{s/mm}^2$) and an additional reference scan without diffusion was used with the following sequence parameters: TR = 7900ms, TE = 94ms, flip angle 90°, matrix size 128×128, 60 slices with a slice thickness of 2mm, in plane resolution 1.95mm × 1.95mm, 4 averages, GRAPPA acceleration factor 2. Prior to the reconstruction of the diffusion tensor, eddy current correction was performed with FSL [31].

In addition to the models (P-TD), (P-TV), (P-TGV^2), and (P-logTV), we evaluate for comparison the more conventional approach of denoising each DWI image separately [1]. Following [25], we perform this by total variation regularisation, i.e., solving (P-TV) for $k = 0$ by Algorithm 4.2 for each of the DWI measurements. The reported iteration count for this regularisation model, denoted DWITV, will be the maximum over the different DWI measurements. Since in the present data set there are four averages of 12 independent diffusion gradients, plus the zero gradient, this amounts to 52 different DWI measurements. The regularisation parameter for this model is chosen as $\alpha = \alpha_0 \| \operatorname{mean}_b A_b \|_{L^\infty(\Omega)}$, where $A_b \in L^1(\Omega)$ is the signal corresponding to the diffusion gradient b. This choice appears approximately comparable to the other methods, and the rule is almost analogous; the replacement of mean by maximum would in principle be closer, but in practise yields too high parameters due to noise. Although a simplistic choice – to do complete justice to all the methods, we should in some sense find the optimal parameters for all of them – by studying all the different choices of α_0 and σ, we are nevertheless able to observe some differences in the models without too serious injustice.

For reasons of visualisational and computational practicality, we perform computations only on a single slice of the data set. The data f_0 is thus a 128×128 image of $\text{Sym}^2(\mathbb{R}^3)$ tensors, to which the considerations in Subsection 4.3 apply. As the data set is of relatively high fidelity, we are justified in applying some additional Rician noise. This is performed for both the parameters $\sigma = 10$ ("low-noise case") and $\sigma = 50$ ("high-noise case") on the original DWI measurements, before the diffusion tensors are solved from the Stejskal-Tanner equation (5.1).

The parameters α and β for (P-TD), (P-TV), (P-TGV^2), and (P-logTV), are chosen analogously to the test case with synthetic data, however, for the present data the maximal eigenvalue $\mu \approx 0.0042$, and we choose the "base regularisation parameter" $\alpha_0 = 0.05$ for the low-noise case and $\alpha_0 = 0.15$ for the high-noise case.

The computational results for the low-noise case are presented in Figure 4, and the error maps in Figure 5. The results for the high-noise case are presented in Figure 6 and the error maps in Figure 7. The results have been zoomed-in on the area $(16, 112) \times (16, 112)$, and points where the average DWI signal intensity is less than 10% of the mean over the whole image, have been mask out. Additionally, for the high-noise case, we include in Figure 8 a plot of the fractional anisotropy superimposed with the projections of the principal eigenvectors to the (x, y)-plane, for region $(55, 73) \times (74, 92)$ of the

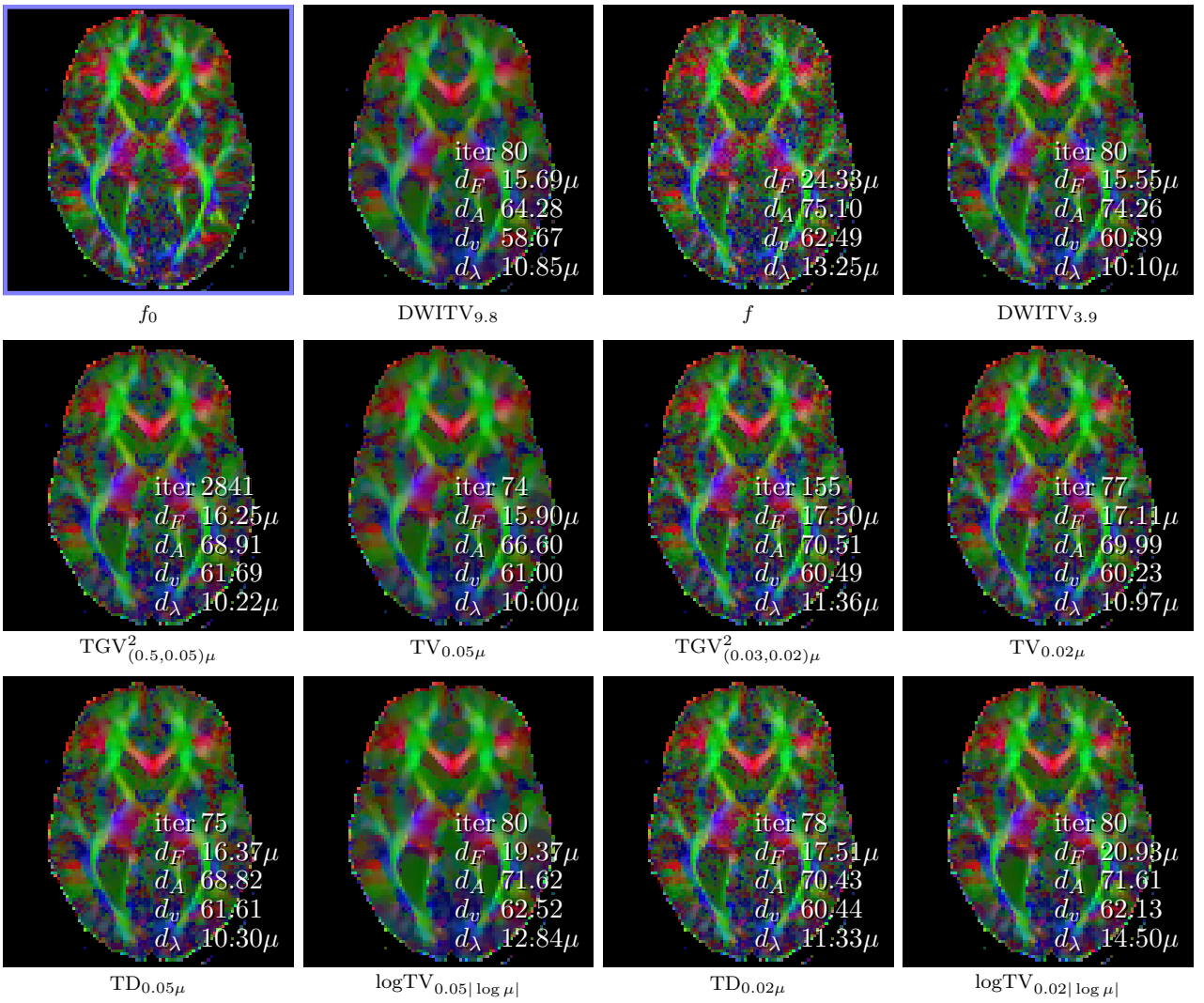

Figure 4: Results for the in-vivo brain data f_0 with additional Rician noise (f) of parameter $\sigma = 10$.

data, containing the corpus callosum; the region is marked with a rectangle in Figure 6. For the low-noise case this plot is not included, because any differences between most of the methods are hard to observe.

We first study the low-noise case. From Figure 4 it is difficult to observe any significant differences between the models, except that TGV^2 requires significantly more iterations to reach the target normalised duality gap than the other models. Nevertheless, all the models manage to remove the additional Rician noise to some degree: The error measures (superimposed on Figure 4) compared to f are reduced significantly by all models, with the exception of the error d_v in the directions of the principal eigenvectors. In contrast to the situation with the synthetic data, where TGV^2 had consistently the best values for the error measures (superimposed on Figure 4) in comparison to TV, TD, and TGV^2, this time TV appears to perform the best of these four models. Clearly, however, DWITV, i.e., individual denoising of the DWI data, has the best error scores, with one exception: the error in fractional anisotropy for the smaller regularisation parameter is significantly higher than for the other models – we shall return to this topic. Next we observe that logTV performs somewhat worse than the other models here. The larger parameter (left half of the figure) can be argued to

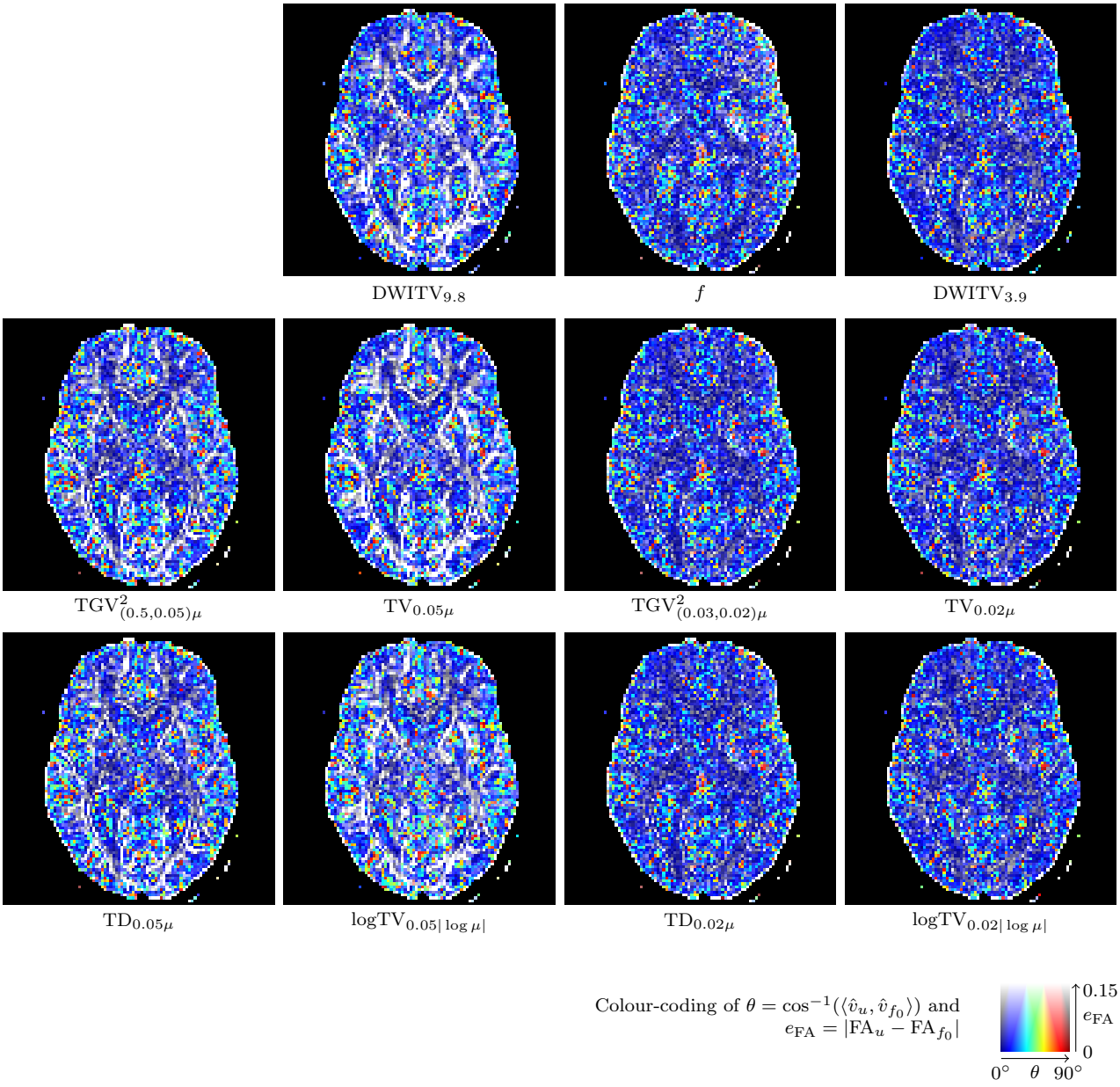

Figure 5: Error plots of computational results and the test data f with additional Rician noise ($\sigma = 10$) in comparison to the in-vivo brain data f_0.

over-regularise. That is however not necessarily the case, we argue, for the smaller parameter (right half). There d_A and d_v for logTV are comparable to the other models, but d_λ and consequently also d_F significantly worse. It therefore appears that logTV does not restore the magnitude of the tensors so well; this can to a degree be confirmed by studying principal eigenvector plots of zoomed-in regions.

More interesting information is provided by the error plot in Figure 5. Regarding the error in the direction of principal eigenvector, plotted with colours, we observe no significant difference between the models. However, from the areas of white we learn that for large parameter values (left half), the fractional anisotropy is not restored so well as well as for small parameter values (right half). This can be argued to be a case of slight over-regularisation. Nevertheless, we observe that for both parameter choices, and especially the larger one, TGV^2 and TD perform the best at restoring the fractional anisotropy – which we already observed for synthetic data – and also have smaller differences between the larger and smaller parameter choices. This indicates that these models (both using the symmetrised differential) are less sensitive towards parameter choice than TV and logTV. DWITV can again be

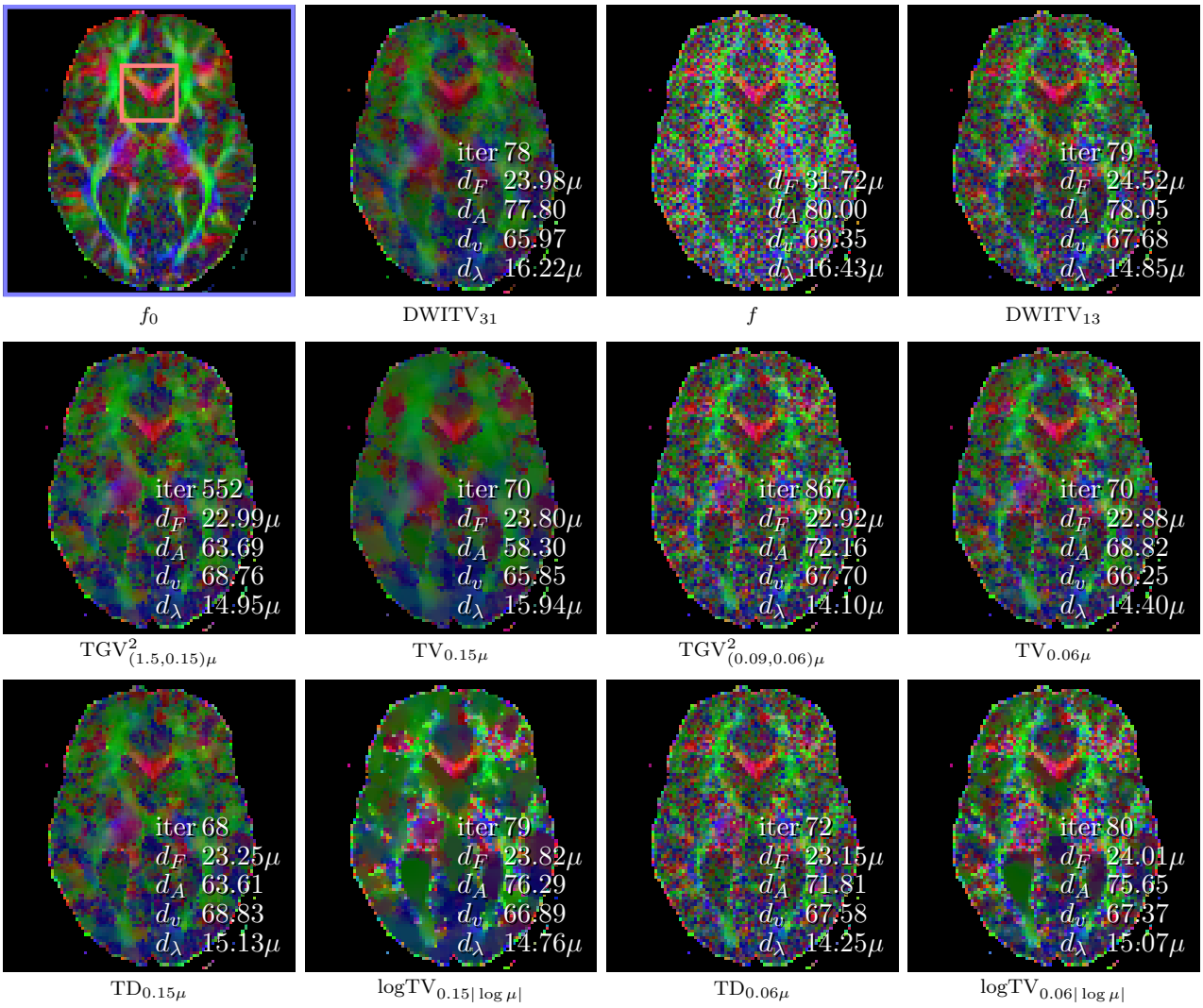

Figure 6: Results for the in-vivo brain data f_0 with additional Rician noise (f) of parameter $\sigma = 50$. The area containing the corpus callosum, presented more closely in Figure 8, is marked with the red rectangle.

seen to restore fractional anisotropy especially poorly, also for the larger parameter. In comparison to f, we observe that for the larger parameter values, all the models introduce significant errors to fractional anisotropy, where there originally was none, and only TGV^2 and TD obviously reduce the errors elsewhere (areas of dark blue at the top left of the brain). For the smaller parameter value, similar conclusions hold, but the differences are not so easy to detect by visual observation.

Based on the observations above, our choice of regularisation model for low-noise and large parameter would be DWITV, and for small parameter it would be TD or TGV^2.

We now turn to the high-noise case, for which the computational results are depicted in Figure 6. Here the reconstruction by all the models is already significantly poorer than the original image f_0, but we clearly observe that logTV is clearly out of the game, introducing large errors in fractional anisotropy, that are visually observable already from this result plot. However, with regard to the fractional anisotropy error measure d_A, DWITV performs even worse. TV has this time the best fractional anisotropy reconstruction by error measure, with TD and TGV^2 not far behind. However,

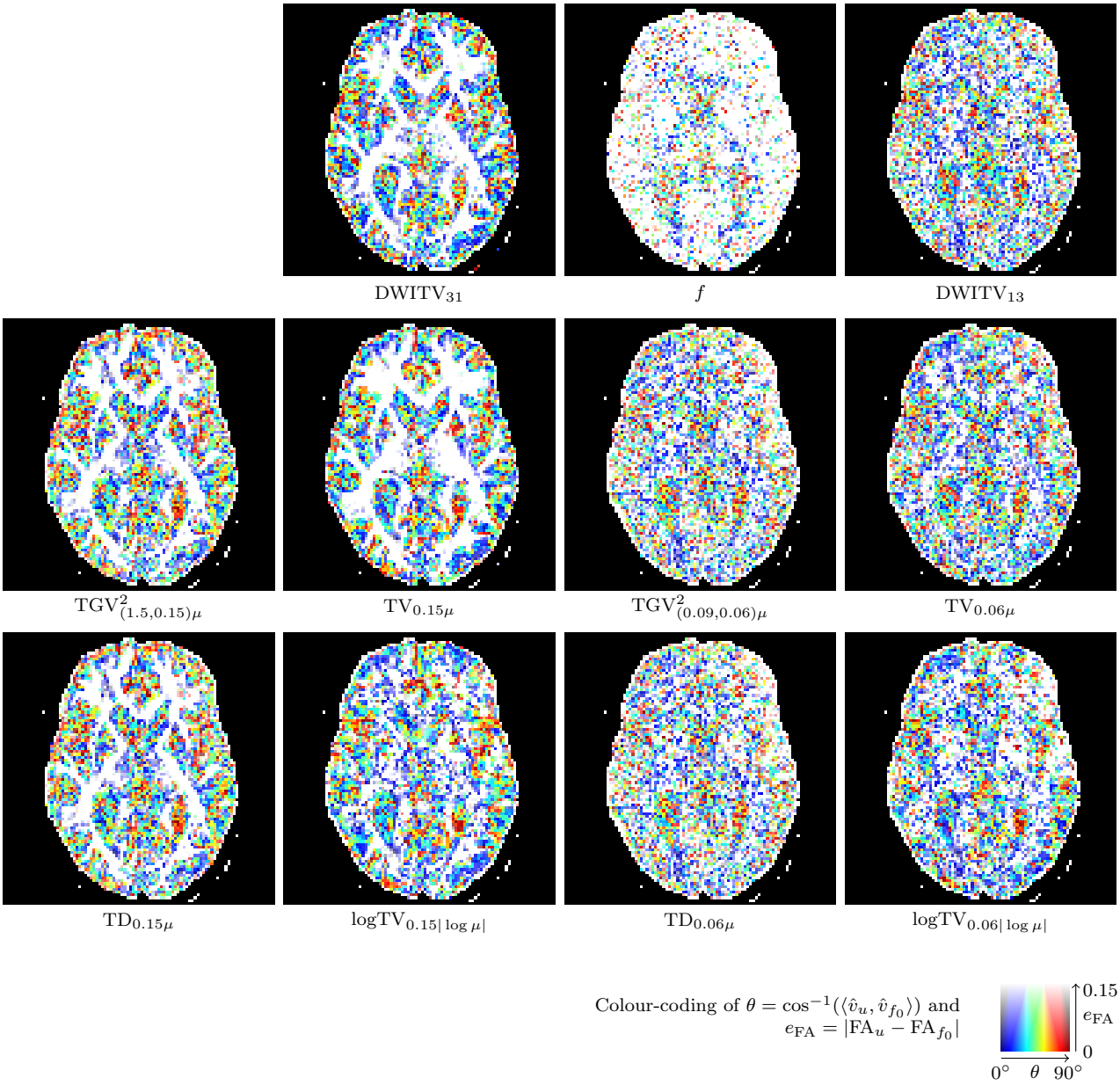

Figure 7: Error plots of computational results and the test data f with additional Rician noise ($\sigma = 50$) in comparison to the in-vivo brain data f_0.

observing the error plot in Figure 7, clearly the roles are reversed for the larger parameter value (left half), and TD and TGV^2 perform the best. For the smaller parameter value (right half), it is not possible to conclusively say that this would be the case.

Studying the zoomed-in Figure 8, we confirm the errors that logTV introduces in fractional anisotropy. For the smaller regularisation parameter (right half) we note none of the the models except logTV manage to well reconstruct the directions of the principal eigenvectors. The largest errors are, however, in the area of low anisotropy, which is reflected in the fact that d_v (superimposed on Figure 6) is comparable for all the models. For the larger regularisation parameter (left half), logTV seems to be over-regularising the directions, while TV has apparently the best reconstruction of the principal eigenvectors here. Also, although logTV had in general large errors in fractional anisotropy, in this region it is, in the sense of overall brightness, much closer to f_0 than the other models.

Let us now summarise our findings from Figures 6, 7 and 8. By visual inspection alone, for the

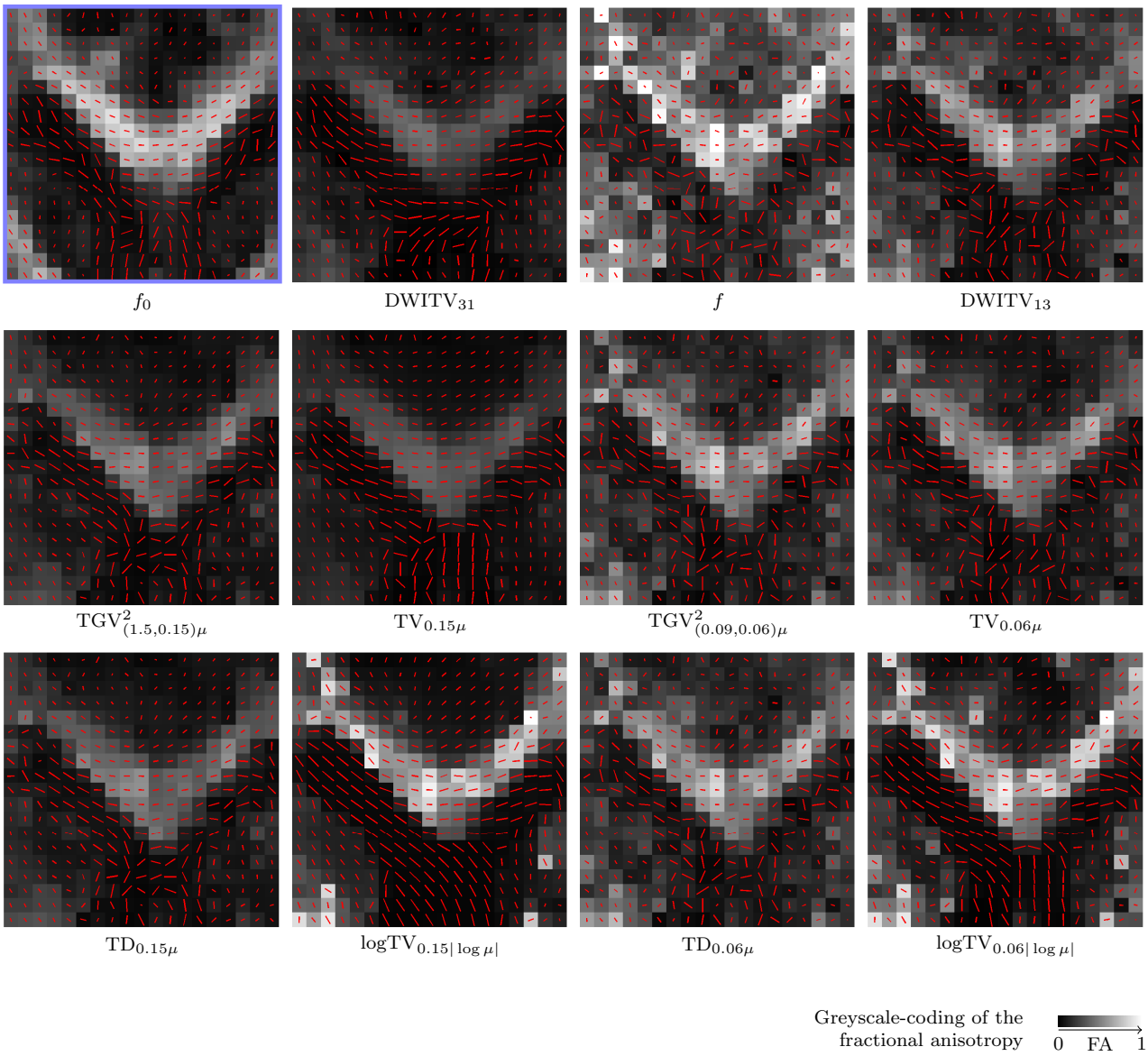

Figure 8: Fractional anisotropy and planar projection of the principal eigenvector for the region containing the corpus callosum of the in-vivo brain data f_0 with additional Rician noise (f) of parameter $\sigma = 50$.

larger regularisation parameter (left half of the figures), all TGV^2, TD, and DWITV clearly provide a reasonable approximation of the noise-free data f_0, minding the extremely noisy data f. logTV has the aforementioned problems with fractional anisotropy, while TV seems to over-regularise, although some of the errors are smallest among all the models. Minding that DWITV has large errors in fractional anisotropy, our choice of model would again be TD or TGV^2. For the smaller parameter value, not all noise is removed by any of the models. TV seems to provide the least noisy result by visual inspection, and also has the smallest values for many of the error measures. It would therefore be our choice.

6. Some final remarks and outlook

From our numerical studies, we conclude that TGV^2 and TD regularisations appear to be very reasonable novel approaches for the denoising of diffusion tensors. While DWITV or TV sometimes perform

better, at least by some of the measures, TGV^2 and TD perform consistently good. DWITV, in particular, suffers from errors in fractional anisotropy. We had expected logTV to perform better than it does. Naturally, as already remarked, it is difficult to choose the regularisation parameters such that complete justice is done to all the methods. Possibly some of the problems are also due to the processing, as discussed in Subsection 5.5, of negative eigenvalues in the noisy data, which the model does not directly allow.

Regarding computational effort, TD requires less iterations of Algorithm 4.2 than TGV^2. It is also per iteration the lightest of all the models, with the caveat that DWITV is not directly comparable, iteration cost depending on the amount of DWI data. Indeed, TGV^2 has for 2D data $16 = 3+4+4+5$ (u, φ, w, ψ) and for 3D data $42 = 6+10+10+16$ parameters per pixel, while TD has $7 = 3+4$ (u, φ) parameters per pixel for 2D data, and $16 = 6+10$ parameters for 3D data. TV also has $11 = 3+8$ (u, non-symmetric φ) parameters per pixel for 2D data and $33 = 6+27$ for 3D data – not much less than TGV^2.

An obvious philosophical defect in all our regularisation models, except the direct regularisation of the individual DWI data, is that we are already losing critical information by reconstruction the tensors by linear regression through the Stejskal-Tanner equation (5.1) before regularisation. But individual denoising of the DWI data for each diffusion gradient also introduces the problem of matching the regularisations of each signal – consider the case of a small shift in the data. In principle eddy current correction removes such matching errors, but may not do it completely, and ideally even eddy current correction would indeed be part of an overall regularisation model from signal to DTI. Future work will therefore be directed towards the integration of direct tensor regularisation in an extended reconstruction model that allows direct reconstruction of the diffusion tensors from raw MR data. The main advantage of such an approach is that it also allows to correct errors due to phase variations that arise from head motion or unavoidable brain pulsations, and to include additional coil-sensitivity information during the reconstruction [7, 3, 35].

Acknowledgements

The authors would like to thank Karl Koschutnig (Department of Psychology, University of Graz) for providing the in-vivo brain DWI data, and Kristian Bredies (Department of Mathematics and Scientific Computing, University of Graz) for many fruitful discussions.

This work has been financially supported by the SFB research program "Mathematical Optimization and Applications in Biomedical Sciences" of the Austrian Science Fund (FWF).

A. L^1 bounds for TGV^2 regularisation

Proposition A.1. *Let $\Omega \subset \mathbb{R}^m$ be a bounded Lipschitz domain. Suppose $f \in L^2(\Omega; \mathcal{T}^k(\mathbb{R}^m))$, and let $\hat{u} \in \mathrm{BD}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$ and $\hat{w} \in \mathrm{BD}(\Omega; \mathrm{Sym}^{k+1}(\mathbb{R}^m))$ obtain the minimum in the problem* (P-TGV^2)*, that is*

$$(\hat{u}, \hat{w}) = \operatorname*{arg\,min}_{(u,w)} \frac{1}{2}\|f-u\|_{F,2}^2 + \alpha\|Eu-w\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} + \beta\|Ew\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+2}(\mathbb{R}^m))}. \tag{A.1}$$

Then there exists a constant $C = C(\Omega, k, m, \alpha, \beta) < \infty$, such that

$$\|\hat{u}\|_{F,1} + \|\hat{w}\|_{F,1} \leq C(\|f\|_{F,2} + \|f\|_{F,2}^2).$$

Proof. By Lemma 3.1, there exists a constant $C_0 < \infty$, dependent on Ω, k, m, such that

$$\|u\|_{F,1} + \|Eu\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} \leq C_0[\|u\|_{F,1} + \mathrm{TGV}^2_{(\beta,\alpha)}(u)], \quad (u \in \mathrm{BD}(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))).$$

Setting $u = 0$, $w = 0$ in (A.1), we get

$$\frac{1}{2}\|f - \hat{u}\|^2_{F,2} \leq \frac{1}{2}\|f - \hat{u}\|^2_{F,2} + \mathrm{TGV}^2_{(\beta,\alpha)}(\hat{u}) \leq \frac{1}{2}\|f\|^2_{F,2}, \tag{A.2}$$

whence

$$\|E\hat{u}\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} \leq C_0[\|\hat{u}\|_{F,1} + \mathrm{TGV}^2_{(\beta,\alpha)}(\hat{u})] \leq C_0\big[\|\hat{u}\|_{F,1} + \|f\|^2_{F,2}\big].$$

The estimate (A.2) and the boundedness of Ω also give

$$\|\hat{u}\|_{F,1} \leq \mathcal{L}^m(\Omega)^{1/2}\|\hat{u}\|_{F,2} \leq \mathcal{L}^m(\Omega)^{1/2}(\|f - \hat{u}\|_{F,2} + \|f\|_{F,2}) \leq 2\mathcal{L}^m(\Omega)^{1/2}\|f\|_{F,2}$$

This gives the desired bound on $\|\hat{u}\|_{F,1}$, and the two preceding estimates together give

$$\|E\hat{u}\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} \leq C_0\big[2\mathcal{L}^m(\Omega)^{1/2}\|f\|_{F,2} + \|f\|^2_{F,2}\big]. \tag{A.3}$$

Setting $w = 0$ and $u = \hat{u}$ in (A.1) gives

$$\alpha\|E\hat{u} - \hat{w}\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} + \beta\|E\hat{w}\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+2}(\mathbb{R}^m))} \leq \alpha\|E\hat{u}\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))},$$

whence

$$\|\hat{w}\|_{F,1} \leq \|E\hat{u} - \hat{w}\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} + \|E\hat{u}\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} \leq 2\|E\hat{u}\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))}.$$

Recalling (A.3), we obtain for $\hat{w}$ the desired bound

$$\|\hat{w}\|_{F,1} \leq 2C_0\big[2\mathcal{L}^m(\Omega)^{1/2}\|f\|_{F,2} + \|f\|^2_{F,2}\big].$$

This concludes the proof. □

References

1. S. Aja-Fernández, M. Niethammer, M. Kubicki, M. Shenton and C. Westin, *Restoration of DWI data using a Rician LMMSE estimator*, IEEE Trans. Medical Imaging **27** (2008), 1389–1403, doi:10.1109/TMI.2008.920609.
2. 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.
3. M. Aksoy, C. Liu, M. E. Moseley and R. Bammer, *Single-step nonlinear diffusion tensor estimation in the presence of microscopic and macroscopic motion.*, Magnetic Resonance in Medicine **59** (2008), 1138–1150, doi:10.1002/mrm.21558.
4. L. Ambrosio, N. Fusco and D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford University Press, 2000.
5. V. Arsigny, P. Fillard, X. Pennec and N. Ayache, *Fast and simple computations on tensors with log-euclidean metrics.*, Technical Report 5584, INRIA (2005).
6. ———, *Log-euclidean metrics for fast and simple calculus on diffusion tensors*, Magnetic Resonance in Medicine **56** (2006), 411–421.
7. R. Bammer, M. Aksoy and C. Liu, *Augmented generalized sense reconstruction to correct for rigid body motion.*, Magnetic Resonance in Medicine **57** (2007), 90–102, doi:10.1002/mrm.21106.
8. 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.
9. K. Bredies, *Symmetric tensor fields of bounded deformation*, Ann. Mat. Pura Appl. (2012). Accepted.
10. K. Bredies, K. Kunisch and T. Pock, *Total generalized variation*, SIAM J. Imaging Sci. **3** (2011), 492–526.

11. K. Bredies, K. Kunisch and T. Valkonen, *Properties of L^1-TGV^2: The one-dimensional case*, SFB-Report 2011-006, Karl-Franzens University of Graz (2011).
URL `http://math.uni-graz.at/mobis/publications/SFB-Report-2011-006.pdf`
12. K. Bredies and T. Valkonen, *Inverse problems with second-order total generalized variation constraints*, in: *Proceedings of SampTA 2011 – 9th International Conference on Sampling Theory and Applications, Singapore*, 2011.
13. A. Chambolle, *An algorithm for total variation minimization and applications*, J. Math. Imaging Vision **20** (2004), 89–97.
14. A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision (2010), 1–26.
15. 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.
16. O. Coulon, D. Alexander and S. Arridge, *Diffusion tensor magnetic resonance image regularization*, Medical Image Anal. **8** (2004), 47–67.
17. J. Diestel, J. H. Fourie and J. Swart, *The metric theory of tensor products: Grothendieck's résumé revisited*, American Mathematical Society, 2008.
18. H. Federer, *Geometric Measure Theory*, Springer, 1969.
19. P. Fillard, X. Pennec, V. Arsigny and N. Ayache, *Clinical DT-MRI estimation, smoothing, and fiber tracking with log-Euclidean metrics*, IEEE Trans. Medical Imaging **26** (2007), 1472–1482.
20. H. Gudbjartsson and S. Patz, *The Rician distribution of noisy MRI data*, Magnetic Resonance in Medicine **34** (1995), 910–914.
21. 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.
22. B. Hofmann, B. Kaltenbacher, C. Pöschl and O. Scherzer, *A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators*, Inverse Problems **23** (2007), 987.
23. P. Kingsley, *Introduction to diffusion tensor imaging mathematics: Parts I-III*, Concepts in Magnetic Resonance Part A **28** (2006), 101–179.
24. F. Knoll, K. Bredies, T. Pock and R. Stollberger, *Second order total generalized variation (TGV) for MRI.*, Magnetic Resonance in Medicine **65** (2011), 480–491, doi:10.1002/mrm.22595.
25. F. Knoll, Y. Dong, M. Hintermüller and R. Stollberger, *Total variation denoising with spatially dependent regularization*, in: *Proceedings of the 18th Scientific Meeting and Exhibition of ISMRM, Stockholm, Sweden*, 2010, 5088.
26. T. Kolda and B. Bader, *Tensor decompositions and applications*, SIAM Rev. **51** (2009), 455–500.
27. A. Lewis, *The convex analysis of unitarily invariant matrix functions*, J. Convex Anal. **2** (1995), 173–183.
28. C. Poupon, J. Mangin, V. Frouin, J. Régis, F. Poupon, M. Pachot-Clouard, D. Le Bihan and I. Bloch, *Regularization of MR diffusion tensor maps for tracking brain white matter bundles*, in: *Medical Image Computing and Computer-Assisted Intervention–MICCAI98*, Springer, 1998, 489–498.
29. L. Rudin, S. Osher and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D **60** (1992), 259–268.
30. 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.
31. 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.
32. R. Temam, *Mathematical problems in plasticity*, Gauthier-Villars, 1985.
33. 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.
34. D. Tschumperlé and R. Deriche, *Diffusion tensor regularization with constraints preservation*, in: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, IEEE, 2001, volume 1, 948–953.
35. M. Uecker, A. Karaus and J. Frahm, *Inverse reconstruction method for segmented multishot diffusion-weighted MRI with multiple coils.*, Magnetic Resonance in Medicine **62** (2009), 1342–1348, doi:10.1002/mrm.22126.
36. T. Valkonen, *Diff-convex combinations of Euclidean distances: a search for optima*, number 99 in

Jyväskylä Studies in Computing, University of Jyväskylä, 2008. Ph.D Thesis.