SpezialForschungsBereich F 32

Karl–Franzens Universität Graz
Technische Universität Graz
Medizinische Universität Graz

Total generalised variation in diffusion tensor imaging

T. Valkonen K. Bredies F. Knoll

SFB-Report No. 2012-003 February 2012

A–8010 GRAZ, HEINRICHSTRASSE 36, AUSTRIA

Supported by the
Austrian Science Fund (FWF)

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*, Kristian Bredies†, and Florian Knoll‡

Abstract. We study the extension of total variation (TV), total deformation (TD), and (second-order) total generalised variation (TGV^2) to symmetric tensor fields. We show that for a suitable choice of finite-dimensional norm, these variational semi-norms are rotation-invariant in a sense natural and well-suited for application to diffusion tensor imaging (DTI). Combined with a positive definiteness constraint, we employ these novel semi-norms as regularisers in ROF-type denoising of medical in-vivo brain images. For the numerical realisation, we employ the Chambolle-Pock algorithm, for which we develop a novel duality-based stopping criterion which guarantees error bounds with respect to the functional values. Our findings indicate that TD and TGV^2, both of which employ the symmetrised differential, provide improved results compared to other evaluated approaches.

Key words. total variation, total deformation, total generalised variation, regularisation, medical imaging, diffusion tensor imaging.

AMS subject classifications. 92C55, 94A08, 26B30, 49M29

1. Introduction.

In this paper, we propose and study novel edge-preserving regularisation functionals for positivity-constrained variational denoising problems on symmetric tensor fields, i.e., minimising functionals of the type

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

Here, $f \in L^1(\Omega;\mathrm{Sym}^2(\mathbb{R}^m))$ is a second-order symmetric tensor field on a domain $\Omega \subset \mathbb{R}^m$, $\|\cdot\|_{F,2}$ is the L^2-norm on Ω with Frobenius norm for symmetric 2-tensors as pointwise norm, and H is a regularisation functional. Observe, moreover, that we require the solution $u \in L^1(\Omega;\mathrm{Sym}^2(\mathbb{R}^m))$ to be pointwise a.e. positive semi-definite. The regularisation functionals we are introducing in this paper generalise the total-variation functional

$$H(u) = \|Du\|_{\mathcal{M}(\Omega;)},$$

extended to symmetric tensor fields of order k, which is a popular choice for denoising problems and usually results in the ROF (Rudin-Osher-Fatemi [34]) model. The generalisations are aiming at two directions. On the one hand, we utilise the symmetrised derivative Eu of u instead of the full derivative Du in order to obtain a weaker measure of gradient information. This results in the *total deformation*, which reads as

$$H(u) = \|Eu\|_{\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))}$$

and can also be seen as first-order total generalised variation TGV^1 for tensor fields. In particular, it imposes, as the total variation, regularity on the data which includes discontinuities

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

and is therefore well-suited for symmetric tensor-valued image data. However, like all Radon-type first-order regularisations, solutions tend to exhibit staircasing artefacts. In order to reduce these effects we also propose, on the other hand, a generalisation of the second-order *total generalised variation* TGV^2 [11] for symmetric tensor fields:

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

Besides providing an analysis for these functionals and the associated denoising problems, we moreover develop and discuss a numerical algorithm for the approximate solution of these problems together with a rigorous stopping criterion. The performance of the algorithm is tested on synthetic as well as in-vivo brain data.

The need for solving positivity-constrained denoising problems with tensor-valued data arises, 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, 38]. 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 [23], eddy-current distortions [38], and is very sensitive with respect to artefacts originating from patient motion [25, 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, 27]. 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 denoising the diffusion tensors obtained this way. A natural requirement is that the denoising result should be invariant with respect to rotations of the imaged object. In diffusion tensor imaging, unlike in scalar imaging, this involves individual tensors rotating as well, not only shifting to another point of the domain. Our regularisation functionals will be rotation-invariant. The denoised tensors should, moreover, be positive definite, as the failure of this condition is non-physical.

Let us shortly discuss some of the existing approaches for the denoising of DTI data and their relations to the present work. In [5, 6, 20, 21, 22] log-Euclidean metrics giving a suitable Riemannian manifold structure to $\mathrm{Sym}^2(\mathbb{R}^m)$ have been studied 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. Log-Euclidean metrics have, moreover, the advantage over earlier affine-invariant metrics [32] that they are computationally more efficient. Basically the 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

with smoothed TV regularisation term [20], or Gaussian denoising [21], and transforming the solution back to the positive definite cone by taking the exponential. This approach has therefore many desirable theoretical qualities, and is also computationally tractable. We are interested in whether our regularisation approaches can provide results with other desirable qualities. It is known [11, 28] that TGV^2 in the scalar case tends to avoid the stair-casing effect exhibited by TV, so we expect some improvements. It is also interesting to see what is the effect of the symmetrised differential employed by TD (that is to say, TGV^1) and TGV^2 in contrast to the normal differential employed by TV.

We note that TV, however lacking the positive semi-definiteness constraint, has in fact already been studied in [35]. In [24], which also takes the approach of Riemannian metrics, and [43], the positive semi-definiteness constraint is incorporated through a Cholesky or $u = LDL^T$ factorisation approach. These two works moreover incorporate the Stejskal-Tanner equation into the fidelity function, resulting in a difficult non-convex problem. The problem (1.1) is still convex, albeit constrained. Our novel algorithm, based on recent state-of-the-art non-smooth optimisation methods [15], can however handle the constraint efficiently.

A completely different approach to DTI denoising is taken in [16, 40], similar to the one in [39] for colour images. It is of the Perona-Malik or anisotropic diffusion [44] type. Instead of directly minimising a regularisation functional, a constrained gradient flow of a time-dependent tensor field $t \mapsto u_t$ is defined on a suitable manifold. The structure of this manifold incorporates any desired constraints, such as positive semi-definiteness of $u_t(x)$, or the magnitude of the eigenvalues of $u_t(x)$, if we only want to regularise the eigenvectors, as is also done in [33, 17]. The gradient flow is then used to transport the noisy measurement $u_0 = f$ until a solution u_t of desired quality is found. Choosing the cost function giving rise to the (unconstrained) gradient flow to have suitable anisotropic smoothing properties, edges should be preserved – empirically, in the discretisation. Indeed, Perona-Malik type approaches in general have the theoretical difficulty that the edges in space appear in time as shocks that cause the solution to break down. An obvious advantage of our approach (1.1) is that it also theoretically preserves edges (between white matter and grey matter), being based on rigorous formulations of L^1 gradient penalties.

The rest of our paper is organised as follows. First in Section 2 we begin by introducing tensor and tensor field calculus to set up the framework in which our results are represented. We then define variational semi-norms of tensor fields, and show that these are rotation-invariant in a sense natural to diffusion tensor imaging. In Section 3 we introduce in more detail the positivity-constrained denoising problems (1.1) with TV, TD, and TGV^2 regularisation term, and show the existence of solutions. This involves a few new proofs, because we define TGV^2 directly through the differentiation cascade formulation, which is more practical for numerical realisation than the original dual-ball formulation. After that, in Section 4, we describe our implementation of the Chambolle-Pock algorithm [15] that we use to solve these problems, and represent a novel duality-based stopping criterion. 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 in Section 6 our conclusions and outlook for future research. Additionally, in Appendix A we show that the extra primal variable of the differentiation cascade formulation of TGV^2 is bounded, and in Appendix B discuss alternative rotation-invariant tensor norms.

2. Tensors and tensor fields. The development of total generalised variation demands the machinery of differential calculus of tensor fields that we derive next. We also show that the derived *total variation* and *total deformation* semi-norms of tensor fields are rotation-invariant, in a sense to be discussed in more detail, when the base finite-dimensional norm is the Frobenius norm, or one of the alternatives discussed in Appendix B.

2.1. Basic tensor calculus. We begin by recalling basic tensor calculus. We make some simplifications as we do not need the machinery in its full differential-geometric glory, as can be found in, e.g., [9]. In particular, as we are working on the Euclidean space $\mathbb{R}^m$, and do not need tensors with simultaneous covariant ($x \in \mathbb{R}^m$) and contravariant variables ($x \in (\mathbb{R}^m)^* \sim \mathbb{R}^m$), we make no distinction between them.

Therefore, we define a k-tensor $A \in \mathcal{T}^k(\mathbb{R}^m)$ as a k-linear mapping $A : \mathbb{R}^m \times \cdots \times \mathbb{R}^m \to \mathbb{R}$. A symmetric tensor $A \in \mathrm{Sym}^k(\mathbb{R}^m)$ then 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 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 $\{1, \ldots, k\}$. For a k-tensor A and a m-tensor B, we define the $(m+k)$-tensor $A \otimes B$ by

$$(A \otimes B)(c_1, \ldots, c_m, c_{m+1}, \ldots, c_{m+k}) = A(c_1, \ldots, c_m) B(c_{m+1}, \ldots, c_{m+k}).$$

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

$$\langle A, B \rangle := \sum_{p \in \{1,\ldots,m\}^k} A(e_{p_1}, \ldots, e_{p_k}) B(e_{p_1}, \ldots, e_{p_k}), \tag{2.1}$$

and the Frobenius norm

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

We will see later in Proposition 2.2 that these are invariant with respect to orthogonal basis transformations. We moreover have $\langle A, \|\!| B \rangle = \langle \|\!| A, B \rangle$. Thus, in particular, for symmetric $B \in \mathrm{Sym}^k(\mathbb{R}^m)$, and general $A \in \mathcal{T}^k(\mathbb{R}^m)$, we have $\langle A, B \rangle = \langle \|\!| A, B \rangle$.

Example 2.1 (Vectors). Vectors $A \in \mathbb{R}^m$ can be identified with symmetric 1-tensors: $A(x) = \langle A, x \rangle$. The symmetrisation satisfies $\|\!| A = A$. The inner product is the usual inner product in $\mathbb{R}^m$, and the Frobenius norm $\|A\|_F = \|A\|_2$.

Example 2.2 (Matrices). Matrices can be identified with 2-tensors: $A(x, y) = \langle Ax, y \rangle$. Symmetric matrices $A = A^T$ can be identified with symmetric 2-tensors. The symmetrisation is given by $\|\!| A = (A + A^T)/2$. The inner product is $\langle A, B \rangle = \sum_{i,j} A_{ij} B_{ij}$ and $\|A\|_F$ is the matrix Frobenius norm. We use the notation $A \geq 0$ for positive semi-definite A.

Much of what follows in this section is stated for general tensor norms, although our numerical work will employ the Frobenius norm. Appendix B however contains a discussion of some alternative norms, the so-called largest and smallest reasonable cross-norms. The application of these could provide interesting results, provided an efficient numerical implementation. We therefore do not fix the norm in the following when not necessary.

2.2. Tensor fields. Let $\|\cdot\|_\bullet$ be a norm on $\mathcal{T}^k(\mathbb{R}^m)$. We denote its dual norm with respect to the inner product (2.1) by $\|\cdot\|_*$. For $u : \Omega \to \mathcal{T}^k(\mathbb{R}^m)$ on a domain $\Omega \subset \mathbb{R}^m$, 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 semi-continuous 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 $x \in \Omega$.

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

$$[\operatorname{div} u(x)](e_{i_2},\dots,e_{i_k}) := \sum_{i_1=1}^m \partial_{i_1}[x \mapsto u(x)(e_{i_1},\dots,e_{i_k})] = \sum_{i_1=1}^m \langle e_{i_1}, \nabla u(\cdot)(e_{i_1},\dots,e_{i_k})\rangle. \quad (2.2)$$

Observe that if u is symmetric, then so is $\operatorname{div} u$. This definition is also the reason why have assumed the dimension m of the space $\mathbb{R}^m \supset \Omega \ni x$ and of the parameters of $u(x) \in \mathcal{T}^k(\mathbb{R}^m)$ to agree. The definition of divergence does not as such apply to tensor fields $u : \mathbb{R}^K \to \mathcal{T}^k(\mathbb{R}^m)$ for $K \ne m$. As our regularisation functionals will be based on calculating divergences (2.2), this has the implication that when we want to denoise 3D diffusion tensors, we should in principle have a full 3D volume Ω of data, not a 2D slice! We will return to this topic later on.

Example 2.3 (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 2.4 (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 semi-definite u.

Preparing to define tensor-valued measures next, we define the non-symmetric unit ball

$$V^k_{*,\mathrm{ns}} := \{\varphi \in C_c^\infty(\Omega;\mathcal{T}^k(\mathbb{R}^m)) \mid \|\varphi\|_{*,\infty} \le 1\}, \quad (2.3)$$

and the symmetric unit ball

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

2.3. Tensor-valued measures. We let now $\mu \in \mathcal{M}(\Omega; \mathcal{T}^k(\mathbb{R}^m))$ be a tensor-valued measure, which it suffices for our purposes to define to mean that μ is a bounded linear functional on $C_c^\infty(\Omega; \mathcal{T}^k(\mathbb{R}^m))$, namely $\mu \in [C_c^\infty(\Omega; \mathcal{T}^k(\mathbb{R}^m))]^*$, the topological dual of $C_c^\infty(\Omega; \mathcal{T}^k(\mathbb{R}^m))$, and

$$\sup\{\mu(\varphi) \mid \varphi \in V^k_{*,\mathrm{ns}}\} < \infty.$$

For a justification of this definition, we refer to [19]. Other possible definitions also exist, along the lines of that taken for vector-valued measures in [4]. The space $\mathcal{M}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$ of symmetric tensor-valued measures is defined analogously as bounded linear functionals on $C_c^\infty(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$.

We extend $\mu \in \mathcal{M}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$ to $\varphi \in C_c^\infty(\Omega; \mathcal{T}^k(\mathbb{R}^m))$ by symmetrising, $\mu(\varphi) := \mu(\|\varphi)$. Conversely, we symmetrise $\mu \in \mathcal{M}(\Omega; \mathcal{T}^k(\mathbb{R}^m))$ by

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

For $\mu(\varphi) = \int_\Omega \langle f(x), \varphi(x)\rangle \, d\mathcal{L}^m$, where $\mathcal{L}^m$ is the Lebesgue measure on $\mathbb{R}^m$, we then have the desired expression

$$(\|\mu)(\varphi) = \int_\Omega \langle \varphi(x), \|f(x)\rangle \, dx.$$

For symmetric $\mu \in \mathcal{M}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$ clearly $\|\mu = \mu$.

For $\mu \in \mathcal{M}(\Omega; \mathcal{T}^k(\mathbb{R}^m))$, and choice of norm $\|\cdot\|_\bullet$ on $\mathcal{T}^k(\mathbb{R}^m)$, we now define the total variation norm

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

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

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

indeed

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

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

$$\mu^i(\varphi) \to \mu(\varphi), \quad (\varphi \in C_c^\infty(\Omega; \mathcal{T}^k(\mathbb{R}^m))).$$

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

2.4. Symmetric tensor fields of bounded variation and deformation. We define the distributional gradient $Du \in [C_c^\infty(\Omega; \mathcal{T}^{k+1}(\mathbb{R}^m))]^*$ of $u \in L^1(\Omega; \mathcal{T}^k(\mathbb{R}^m))$ by

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

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

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

With these notions at hand, we now define the spaces of symmetric tensor fields of *bounded variation* and *bounded deformation*, respectively, as (see also [10])

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

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 $Eu \in \mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))$.

Remark 2.1 (Equivalences and terminology). For $k = 0$, i.e., in the case of scalar fields, the space $\mathrm{BV}(\Omega;\mathrm{Sym}^0(\mathbb{R}^m)) = \mathrm{BD}(\Omega;\mathrm{Sym}^0(\mathbb{R}^m))$ agrees with the usual space of (scalar) functions of bounded variation. For $k = 1$, i.e., the case of vector fields, the space $\mathrm{BD}(\Omega;\mathrm{Sym}^1(\mathbb{R}^m))$ agrees with the space of *functions of bounded deformation* studied in [37]. This is our motivation for the term *tensor fields of bounded deformation*. The space $\mathrm{BD}(\Omega;\mathrm{Sym}^1(\mathbb{R}^m))$ does not agree with the space $\mathrm{BV}(\Omega;\mathrm{Sym}^1(\mathbb{R}^m))$ for $m > 1$, as the kernel of E includes the *infinitesimal rigid displacements* $u(x) = u_0 + Ax$, where A is skew-symmetric, while the kernel of D includes only constants.

Remark 2.2 (Symmetric tensor fields). It can be argued that from the point of view of general theory, the space

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

of general *tensor fields of bounded variation* is more natural than the space $\mathrm{BV}(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))$ of *symmetric tensor fields of bounded variation*. 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 restriction of $u \in \mathrm{BD}(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))$ to have symmetric values, can be argued to be very natural. In our forthcoming applications we are, however, interested in symmetric tensors fields u only, so our results in this section are stated for $\mathrm{BV}(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))$. They nevertheless hold likewise for $\mathrm{BV}(\Omega;\mathcal{T}^k(\mathbb{R}^m))$ and the analogously defined space $\mathrm{BD}(\Omega;\mathcal{T}^k(\mathbb{R}^m))$

Proposition 2.1. *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, $\sup\{Eu(\varphi) \mid \varphi \in V^{k+1}_{F,\mathrm{s}}\} < \infty$, 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

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

For the first equality we have used the definition of the symmetrisation $\|\!|Du$, and for the final equality the definition $\mu(\varphi) := \mu(\|\!|\varphi)$ for measures $\mu \in \mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))$. ∎

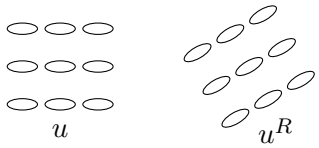


Figure 2.1. *Illustration of the rotation-invariance* $\|Du^R\|_{F,\mathcal{M}(\mathbb{R}^2;\mathcal{T}^3(\mathbb{R}^2))} = \|Du\|_{F,\mathcal{M}(\mathbb{R}^2;\mathcal{T}^3(\mathbb{R}^2))}$*, for* $u^R(y)(c_1,c_2) = u(R^{-1}y)(Rc_1,Rc_2)$ *and* R *a rotation matrix on* $\mathbb{R}^2$*.*

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 [10]. 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 semi-continuous with respect to weak convergence in $\mathrm{BV}(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))$ (resp. $\mathrm{BD}(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))$).

2.5. Rotation-invariance. We now study the rotation-invariance of the total variation semi-norm $\|Du\|_{\bullet,\mathcal{M}(\Omega;\mathcal{T}^{k+1}(\mathbb{R}^m))}$ and the total deformation semi-norm $\|Eu\|_{\bullet,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))}$. It is crucial for the DTI denoising application that these norms are invariant of rotations in a suitable sense, explained in the next example, because the denoising results should not depend on the imaged object having rotated.

Example 2.5. We may draw a tensor $u(x) \in \mathrm{Sym}^2(\mathbb{R}^2)$ as an ellipse whose major and minor axes have magnitude and direction corresponding to the eigenvalues and eigenvectors of $u(x)$. What our main rotation-invariance result, Proposition 2.4, roughly says is that $\|Du\|_{\bullet,\mathcal{M}(\Omega;\mathcal{T}^{k+1}(\mathbb{R}^m))}$ does not change when we rotate the image consisting of these ellipses, as illustrated in Figure 2.1. That is, it is not sufficient to rotate the domain or the tensors alone to obtain invariance, both the domain and the tensors have to be rotated.

We begin by showing orthogonal invariance of the Frobenius norm. After that we study the rotation-invariance of the variational semi-norms, culminating in Proposition 2.4.

Proposition 2.2. *Let* $A \in \mathcal{T}^k(\mathbb{R}^m)$*, and let* $R \in \mathbb{R}^{m\times m}$ *be an orthogonal matrix (i.e.,* $R^T = R^{-1}$*). Define* $A_R \in \mathcal{T}^k(\mathbb{R}^m)$ *according to*

$$A_R(c_1,\ldots,c_k) = A(Rc_1,\ldots,Rc_k). \tag{2.4}$$

Then the Frobenius norm $\|\cdot\|_F$ *is orthogonally invariant in the sense that* $\|A_R\|_F = \|A\|_F$*. In fact, the inner product* (2.1) *is orthogonally invariant in the sense that* $\langle A_R, B_R\rangle = \langle A, B\rangle$ *for* $A, B \in \mathcal{T}^k(\mathbb{R}^m)$*.*

Proof. It obviously suffices to prove the claim for the inner product, from which the result for the Frobenius norm follows. We begin by observing that that we may decompose

$$A = \sum_{i=1}^{N} \alpha_i x_1^i \otimes \cdots \otimes x_k^i \quad \text{and} \quad B = \sum_{i=1}^{N} \beta_i y_1^i \otimes \cdots \otimes y_k^i$$

for some $N \geq 0$, $x_j^i, y_j^i \in \mathbb{R}^m$, and $\alpha_i, \beta_i \in \mathbb{R}$, $(j = 1, \ldots, k; i = 1, \ldots, N)$, where

$$\langle x_1^i \otimes \cdots \otimes x_k^i, x_1^\ell \otimes \cdots \otimes x_k^\ell \rangle = \langle y_1^i \otimes \cdots \otimes y_k^i, y_1^\ell \otimes \cdots \otimes y_k^\ell \rangle = 0, \quad (i \neq \ell).$$

For example, we may take $A = \sum_{p \in \{1,\ldots,m\}^k} A(e_{p_1}, \ldots, e_{p_k}) e_{p_1} \otimes \cdots \otimes e_{p_k}$, and likewise for B. Then also

$$A_R = \sum_{i=1}^{N} \alpha_i R x_1^i \otimes \cdots \otimes R x_k^i \quad \text{and} \quad B_R = \sum_{i=1}^{N} \beta_i R y_1^i \otimes \cdots \otimes R y_k^i,$$

with

$$\langle Rx_1^i \otimes \cdots \otimes Rx_k^i, Rx_1^\ell \otimes \cdots \otimes Rx_k^\ell \rangle = \langle Ry_1^i \otimes \cdots \otimes Ry_k^i, Ry_1^\ell \otimes \cdots \otimes Ry_k^\ell \rangle = 0, \quad (i \neq \ell).$$

Thus

$$\begin{aligned}
\langle A_R, B_R \rangle &= \sum_{i=1}^{N} \sum_{\ell=1}^{N} \alpha_i \beta_\ell \langle Rx_1^i \otimes \cdots \otimes Rx_k^i, Ry_1^\ell \otimes \cdots \otimes Ry_k^\ell \rangle \\
&= \sum_{i=1}^{N} \sum_{\ell=1}^{N} \alpha_i \beta_\ell \prod_{j=1}^{k} \langle Rx_j^i, Ry_j^\ell \rangle = \sum_{i=1}^{N} \sum_{\ell=1}^{N} \alpha_i \beta_\ell \prod_{j=1}^{k} \langle x_j^i, y_j^\ell \rangle \\
&= \sum_{i=1}^{N} \sum_{\ell=1}^{N} \alpha_i \beta_\ell \langle x_1^i \otimes \cdots \otimes x_k^i, y_1^\ell \otimes \cdots \otimes y_k^\ell \rangle = \langle A, B \rangle.
\end{aligned}$$

This proves the claim. ■

We let now $\|\cdot\|_\bullet$ be a generic norm on $\mathcal{T}^k(\mathbb{R}^m)$, satisfying the orthogonal invariance conclusion of Proposition 2.2. and $R \in \mathbb{R}^{m \times m}$ be a rotation matrix, i.e., orthogonal with $\det(R) = 1$. We let $u \in L^1(\mathbb{R}^m; \mathcal{T}^k(\mathbb{R}^m))$, and $\varphi \in C_c^\infty(\mathbb{R}^m; \mathcal{T}^k(\mathbb{R}^m))$. We define $u^R \in L^1(\mathbb{R}^m; \mathcal{T}^k(\mathbb{R}^m))$ by

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

and φ^R analogously. Then an application of the area formula shows that

$$\|u^R\|_{\bullet,p} = \|u\|_{\bullet,p},$$

as well as

$$\int_{\mathbb{R}^m} \langle u^R(y), \varphi^R(y) \rangle \, dy = \int_{\mathbb{R}^m} \langle u(y), \varphi(y) \rangle \, dy. \tag{2.5}$$

The following lemma and proposition show that also the norm of the gradient is invariant.

Lemma 2.3. *Suppose $u \in L^1(\mathbb{R}^m; \mathcal{T}^k(\mathbb{R}^m))$, and that $\varphi \in C_c^\infty(\mathbb{R}^m; \mathcal{T}^{k+1}(\mathbb{R}^m))$. Let $R \in \mathbb{R}^{m\times m}$ be a rotation matrix. Then*

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

as well as

$$\|\varphi^R\|_{\bullet,p} = \|\varphi\|_{\bullet,p}, \quad (p \in [1,\infty]). \tag{2.6}$$

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

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

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

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

By Proposition 2.2 we have $\langle A_{R^{-1}}, C_{R^{-1}}\rangle = \langle A, C\rangle$ for $A, C \in \mathcal{T}^k(\mathbb{R}^m)$. Thus we may now calculate for $y \in \mathbb{R}^m$ 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 does not depend on the choice of the orthonormal basis $c_1,\dots,c_m$ of $\mathbb{R}^m$. Thus, by application of the area formula

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

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

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

Finally, $\|\varphi(R^{-1}y)\|_{\bullet,p} = \|[\varphi(R^{-1}y)]_R\|_{\bullet,p} = \|\varphi^R(y)\|_{\bullet,p}$ due to the assumed orthogonal invariance of the finite-dimensional norm $\|\cdot\|_\bullet$. Thus (2.6) 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

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

This concludes the proof. ■

The following is our main result on rotation-invariance.

Proposition 2.4. *Let $\|\cdot\|_\bullet$ be a norm on $\mathcal{T}^k(\mathbb{R}^m)$, satisfying the conclusion of Proposition 2.2. Let $u \in \mathrm{BV}(\mathbb{R}^m; \mathrm{Sym}^k(\mathbb{R}^m))$ (resp. $u \in \mathrm{BD}(\mathbb{R}^m; \mathrm{Sym}^k(\mathbb{R}^m))$). Given an rotation matrix $R \in \mathbb{R}^{m\times m}$, i.e. orthogonal with $\det(R) = 1$, we then have*

$$\|Du^R\|_{\bullet,\mathcal{M}(\mathbb{R}^m;\mathcal{T}^{k+1}(\mathbb{R}^m))} = \|Du\|_{\bullet,\mathcal{M}(\mathbb{R}^m;\mathcal{T}^{k+1}(\mathbb{R}^m))}$$
$$(\text{resp. } \|Eu^R\|_{\bullet,\mathcal{M}(\mathbb{R}^m;\mathcal{T}^{k+1}(\mathbb{R}^m))} = \|Eu\|_{\bullet,\mathcal{M}(\mathbb{R}^m;\mathcal{T}^{k+1}(\mathbb{R}^m))}).$$

Proof. Immediate consequence of Lemma 2.3. ■

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 regularisation of tensor fields. We now extend total variation 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 (2.3) of

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

Observe that we bound φ pointwise by the Frobenius norm. The reason for this is that we desire the rotation-invariance detailed in Proposition 2.4. Alternatively, it would be interesting to use the largest or smallest reasonable cross-norm, discussed in Appendix B, but these norms are computationally demanding for tensors of degree greater than two; compare to the related tensor decompositions in [30]. In our application $\varphi(x)$ has degree three.

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 [35] 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 inf-sup formulation

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

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

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

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

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

With then readily arrive at the regularisation problem

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

and the equivalent inf-sup formulation

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

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 [11] as a higher-order extension of TV. The application to magnetic resonance imaging, in particular, is studied in [28], yielding improved results in comparison to TV.

For a scalar field $u \in L^1(\Omega)$, second-order TGV may according to [12, 13] 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 $u \in L^1(\Omega) = L^1(\Omega;\mathcal{T}^0(\mathbb{R}^m))$.

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

$$\mathrm{TGV}^2_{(\beta,\alpha)}(u) := \min_{w \in L^1(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} \alpha\|Eu - w\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} + \beta\|Ew\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+2}(\mathbb{R}^m))}. \tag{3.2}$$

Observe that also this semi-norm is rotation-invariant (on $\Omega = \mathbb{R}^m$), as follows from Lemma 2.3 together with (2.5).

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

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

This can again be written in the inf-sup form

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

where $u\in L^1(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))$, $w\in L^1(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))$, $\varphi\in C_c^\infty(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))$, 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 3.1 (Symmetric differentials). It can be argued that TGV^2 defined above should be called TGD^2, for *total generalised deformation*, due to the use of symmetrised differentials, and the fact that TD, not TV, is the first-order equivalent of TGV^2. The reason for calling (3.2) TGV is historical: TGV^2 in the scalar case (3.1) already employs symmetric differentials. Of course, for scalar fields $Eu=Du$, and so TV and TD also agree.

It is also possible to define "non-symmetric TGV^2", bearing similar differences to "symmetric TGV^2" as TV bears to TD. We however do not do that, because such a regulariser would be computationally much heavier, due to far greater degree of freedom, and because we have found the symmetric TD to offer numerically better results than 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-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., [26]. 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

$$\sup_i\|u^i\|_{F,2}+\|Eu^i\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))}<\infty. \tag{3.3}$$

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 semi-continuity 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 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 [13], we need to show that for some $c>0$ it holds

$$c\big(\|u\|_{F,1}+\|Eu\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))}\big)\leq\|u\|_{F,1}+\mathrm{TGV}^2_{(\beta,\alpha)}(u),\quad(u\in L^1(\Omega;\mathrm{Sym}^k(\mathbb{R}^m)),\tag{3.4}$$

and that $\mathrm{TGV}^2_{(\beta,\alpha)}$ is lower semi-continuous 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 (3.3) holds. Indeed, $\sup_i\|u^i\|_{F,1}<\infty$ thanks to Ω being bounded and

$$\sup_i\|f-u^i\|_{F,2}+\mathrm{TGV}^2_{(\beta,\alpha)}(u^i)<\infty, \tag{3.5}$$

the latter following from $\{u^i\}$ being a minimising sequence for (P-TGV^2). It now follows from (3.4) that $\sup_i \|Eu^i\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} < \infty$. This and (3.5) lead to (3.3). The same arguments as above now provide $0 \leq \hat{u} \in \mathrm{BD}(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))$, and lower semi-continuity establishes that it solves (P-TGV^2). ■

Lemma 3.2. *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*

$$c\|u\|_{\mathrm{BD}(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))} \leq \|u\|_{F,1} + \mathrm{TGV}^2_{(\beta,\alpha)}(u) \leq C\|u\|_{\mathrm{BD}(\Omega;\mathrm{Sym}^k(\mathbb{R}^m))}. \tag{3.6}$$

Proof. The proof is a straightforward extension of the equivalence proof for $k = 0$ in [13], employing the tensor Sobolev-Korn estimates from [10]. 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 (3.6) holds.

For the first inequality of (3.6), 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

$$\|Eu\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} \leq C_1\big(\|u\|_{F,1} + \|Eu - \bar{w}\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))}\big) \tag{3.7}$$

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 [10]. Indeed, suppose that (3.7) 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 $i = 1, 2, 3, \ldots$, it holds

$$\|Eu^i\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} = 1 \quad \text{and} \quad \|u^i\|_{F,1} + \|Eu^i - \bar{w}^i\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} \leq 1/i, \tag{3.8}$$

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 (3.8) 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 (3.8) 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 (3.8). Hence (3.7) holds.

Now we employ from [10] 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 $w \in \mathrm{BD}(\Omega; \mathrm{Sym}^{k+1}(\mathbb{R}^m))$ 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 (3.7) 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 $w \in \mathrm{BD}(\Omega; \mathrm{Sym}^{k+1}(\mathbb{R}^m))$ and $u \in \mathrm{BD}(\Omega; \mathrm{Sym}^k(\mathbb{R}^m))$. Hence the first inequality of (3.6) holds with $c = C_4^{-1}$. ■

Lemma 3.3. *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 semi-continuous 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 [10, 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 semi-continuity 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 semi-continuity. Since $\sup_i F(\mu^i) + \|\mu^i\|_{F,\mathcal{M}(\Omega;\mathrm{Sym}^{k+1}(\mathbb{R}^m))} < \infty$, we deduce that

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

Hence some subsequence of $\{\hat{w}^i\}_{i=0}^\infty$ 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 semi-continuity 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 semi-continuity of F. ■

Remark 3.2 (Dual-ball formulation). If we extended to the tensor case the equivalence proof [12, 13] of the differentiation cascade formulation (3.1) of $\mathrm{TGV}^2_{(\beta,\alpha)}$, and the original dual-ball formulation

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

then, following the original proof in [11], we could almost trivially obtain lower semi-continuity of $\mathrm{TGV}^2_{(\beta,\alpha)}$ with respect to convergence in L^p. This would imply weak lower semi-continuity in L^1, and could be used to replace Lemma 3.3 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

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

for proper convex lower semi-continuous G, F^*. This suggests that the Chambolle-Pock algorithm [15] 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 $K^* : C_c^\infty(\Omega; \mathcal{T}^{k+1}(\mathbb{R}^m)) \to C_c^\infty(\Omega; \mathcal{T}^k(\mathbb{R}^m))$. 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^*.

For TD and TV the function

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

is uniformly convex. We therefore describe the accelerated version of the Chambolle-Pock algorithm. It has the following assumptions.

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 semi-continuous, with F^* the conjugate of a convex lower semi-continuous function F. Let, moreover, $\gamma \geq 0$ be 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 [15], 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.

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

For TD and TV we have $\gamma = 1$. For TGV^2 we should in take $\gamma = 0$, because $G(u, w) = G_0(u)$ does not depend on w, and is thus not uniformly convex. In this case the above algorithm reduces to the unaccelerated version that does not update θ_i and σ_i. In numerical practise $\gamma = 1$ works often better, but at other times does not converge.

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 primary case of interest with $k = 2$, it turns out that they reduce to easily calculable projections, as follows.

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, $G = G_0$, for which we solve

$$[(I + \tau\partial G_0)^{-1}(v)](x) = P_{\geq 0}\left(\frac{v(x) + f(x)\tau}{1 + \tau} \right), \quad (x \in \Omega). \tag{4.2}$$

The pointwise projection

$$P_{\geq 0}(A) := \min_{0 \leq X \in \mathrm{Sym}^2(\mathbb{R}^m)} \|A - X\|_F^2, \quad (A \in \mathrm{Sym}^2(\mathbb{R}^m)),$$

can be performed by projecting each eigenvalue of A to $\mathbb{R}^+$. (This can be seen from the structure of the normal cone $N_{\geq 0}(x')$, spelled out in, e.g., [42, Lemma 3.1]. See also [31] for related eigenvalue projection results that are useful for dealing with the nuclear and spectral norms.)

Regarding (S-TD), still $G = G_0$, so we get the resolvent (4.2). Also for F^* 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

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

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, $\delta_{B(0,M)}(\|w\|_{F,1})$ vanishes in Algorithm 4.1, 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 $x \in M_{i+1}U$. 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.

Remark 4.1 (Computations on 2D slices of 3D data). The mathematical theory on tensor fields above, in particular the definition (2.2) 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. Therefore, it is not directly possible to do computations on 2D slices of 3D tensor fields $f : \Omega \subset \mathbb{R}^2 \to \mathrm{Sym}^2(\mathbb{R}^3)$. For reasons of computational efficiency, working on 2D slices of data instead of the full 3D volume can sometimes however be desirable. A solution is to assume that the full image is constant in the z direction, and work on the extension $f_{3\mathrm{D}}(x, y, z) = f(x, y)$. At the level of the numerical implementation this extension can be reduced to performing 3D calculations on the slice f, taking the differentials in the z direction as zero.

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 discussing in detail how the computational algorithm is parametrised. We then describe how our synthetic test data is constructed, and, and finally represent and analyse the results for both this synthetic test data and an in-vivo brain measurement.

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 $\hat{\lambda}_f(x)$ of $f(x)$. As the final geometrical error measure we take the weighted L^2 norm

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

of the differences of normalised principal eigenvectors $\hat{v}_u(x)$ of $u(x)$ and $\hat{v}_f(x)$ of $f(x)$. 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 for $\delta = 0.005$ the thresholding function

$$\nu_{f,u}(x) = \min\{\max\{0, \mathrm{FA}_u(x) - \delta, \mathrm{FA}_f(x) - \delta\}, \delta\}/\delta$$

to ignore or put less weight on points x where the either $u(x)$ or $f(x)$ is almost anisotropic, namely either $\mathrm{FA}_u(x)$ or $\mathrm{FA}_f(x)$ is less than $2\delta = 0.01$. This thresholding will also be used in our error visualisations, discussed in the next subsection.

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. [38]) of mapping the (x, y, z) components of the 3D principal eigenvector ($z = 0$ for 2D data) directly to the $(\mathrm{red}, \mathrm{green}, \mathrm{blue})$ 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 5.1 (2D) and Figure 5.3 (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], a good choice of visualisation is to plot the angle $\cos^{-1}(\langle \hat{v}_u, \hat{v}_{f_0} \rangle)$ between the principal eigenvectors of f_0 and u. However, as discussed in the context of the error

measures in the previous subsection, this error is meaningless for fully anisotropic tensors. Therefore we use the same thresholding $\nu_{f,u}$ as in the definition of d_v to put less weight on such points, and plot

$$\theta = \nu_{f_0,u} \cos^{-1}(\langle \hat{v}_u, \hat{v}_{f_0} \rangle).$$

using the "jet" colour map of Matlab, spanning from blue through cyan and yellow to red.

This colour 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 5.2 and 5.4.

5.3. 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 log-Euclidean [5, 6, 20] regularisation

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

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 $\log f(x)$.

In addition to the models (P-TD), (P-TV), (P-TGV^2), and (P-logTV), for numerical experiments on in-vivo data we evaluate for comparison the more conventional approach of denoising each DWI image separately [1]. Following [29], 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.

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 [14].

5.4. Choice of regularisation parameter. The choice of the regularisation parameters α and β highly affects the denoising results, and care has to be taken to make the comparison of different denoising models fair. One possibility would be to find the "best" parameters for each model, and compare these results. The choice of "best" is, however, not a trivial one, especially as in clinical data features that are difficult to quantify can be more important than simple error measures.

Moreover, we find that it is important to compare the sensitivity of the models to parameter choice, and to study the range of denoising results achievable with each model. We therefore report denoising results of each model for two different choices of the regularisation parameters. We call them the "large parameter choice" and the "small parameter choice". In each test case, we pick for each regularisation model H a parameter $\alpha_0(H)$ by trial and error. For the large parameter choice we then set for each model $\alpha = \alpha_0(H)$ and, for TGV^2, $\beta = 10\alpha_0(\mathrm{TGV}^2)$. For the small parameter choice we set $\alpha = 0.4\alpha_0(H)$, and for TGV^2, $\beta = 0.6\alpha_0(\mathrm{TGV}^2)$. Observe that in this case $\alpha + \beta = \alpha_0(\mathrm{TGV}^2)$, which is a heuristic we want to test for TGV^2 against TV and TD.

Indeed, we pick $\alpha_0(\mathrm{TV}) = \alpha_0(\mathrm{TD}) = \alpha_0(\mathrm{TGV}^2)$. For the large parameter choice this is clearly fair: When β approaches infinity, the only real difference between the functionals is their kernel. For TV it is constant functions, for TD it is "infinitesimal rigid displacements" [10], i.e., functions u such that $Eu = 0$, and for TGV^2 it is functions u such that the absolutely continuous part w of Eu satisfies $Ew = 0$. For the small parameter choice, we will justify the choice of same $\alpha_0(H)$ numerically.

For $H = \mathrm{logTV}, \mathrm{DWITV}$ we pick by trial and error another $\alpha_0(H)$ that seems to provide results comparable to TV, TD and TGV^2 for both the large and small parameter choices. For our tests with in-vivo brain data, for $H = \mathrm{TV}, \mathrm{TD}, \mathrm{TGV}^2$ we end up in practise picking

$$\alpha_0(H) = \alpha_0\mu, \quad \mu := \|\max_i \lambda_i(f_0(x))\|_{L^\infty(\Omega)},$$

for a single "base regularisation parameter" α_0 found by trial and error, and the maximum eigenvalue μ of f_0. For logTV we have then often found

$$\alpha_0(\mathrm{logTV}) = \alpha_0|\log\mu|, \tag{5.1}$$

to be a good choice, and for DWITV

$$\alpha_0(\mathrm{DWITV}) = \alpha_0\|\operatorname{mean}_b A_b(x)\|_{L^\infty(\Omega)}.$$

These rather simple choices are somewhat justified by the fact that the factor between α_0 and $\alpha_0(H)$ is in each case computed by an analogous rule on the data, minding that logTV does TV regularisation on $\log(f_0)$. For our experiments on synthetic test data, we however have ended up choosing

$$\alpha_0(\mathrm{logTV}) = 2\alpha_0|\log\mu|, \tag{5.2}$$

which we have found by trial and error to be fairer than (5.1), although the scale of the latter choice is also right, and the computational results not foo far from TV, TD, and logTV.

Although it is difficult to do complete justice to all the models, we believe to have chosen the parameters reasonably fairly. Given that we report results of all models for two parameter choices, we are able to observe trends that are relatively independent of parameter choice.

5.5. 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 5.1(e) contains an illustration. The origin $(0,0)$ of the image is in the lower-right corner.

In the rectangle $(64,128) \times (64,128)$ 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 lower-left rectangle $(0,64) \times (0,64)$ 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 $(0,64) \times (64,128)$ 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 $(64,128) \times (0,64)$, 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 $\frac{3}{4} e_1 \otimes e_1 + \frac{1}{2} e_2 \otimes e_2$ by an angle between $[0, \pi/2]$ as the x-coordinate varies.

5.6. 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

$$A_b = A_0 \exp(-\langle b, D \rangle). \tag{5.3}$$

Here the b matrix parametrises a *diffusion gradient*, and A_b is the DWI measurement corresponding to b. At least $K \geq 6$ independent non-zero diffusion gradients are needed to solve for 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., [23]. 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 is a Matlab function that applies random Rician noise.

5.7. Results on synthetic data. To the synthetic noise-free test data f_0, we apply Rician noise with the parameter $\sigma = 0.15$, yielding the noisy test data f with PSNR $\approx 34.4\,\mathrm{dB}$ (in the DWI domain with exponential relationship to the tensor field). The maximal eigenvalue of the test data f_0, affecting the regularisation parameters, is $\mu \approx 2.3$. We have chosen the base regularisation parameter $\alpha_0 = 0.25$ by trial and error and the rule (5.2) for logTV. The results of applying the denoising models described above on f are depicted in Figure 5.1, and the error maps in Figure 5.2.

Our principal observations are the following. Firstly, in the top-left region, where u belongs to $(\ker E)^c$, TGV^2 is best at restoring the diagonal line between the red and green-coloured regions for the large parameter choice. TV and logTV display a quite prominent S-shape: the diagonal bends downwards on the right and upwards on the left. For TD the reconstruction of the diagonal is quite noisy. We, however, note that the fact these effects can be seen is a result of the non-linearities 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. Although for the small parameter choice not so much differences can be seen along the diagonal in Figure 5.1, studying Figure 5.2 we observe that the reconstruction error is for both parameter choices quite noisy around the diagonal for the first-order models, while for TGV^2 it is relatively flat.

Secondly, in the top-right region, where u belongs to $\ker D$, we observe, by contrast, that for the small parameter choice TV is best at restoring the flatness of this region. Presumably the stair-casing effect helps here.

In the bottom-left region, where u belongs to $\ker E \setminus \ker D$, the differences between TGV^2, TD, and TV are again minor; studying Figure 5.2, we see, however, that TV exhibits slightly more noisy errors, while logTV exhibits in general more noise in this region.

Indeed, in the nonlinear bottom-right region, we observe that logTV is extremely noisy for both parameter choices. For the smaller parameter choice all the methods are somewhat noisy in this region. Studying Figure 5.2, we see that it has large errors in fractional anisotropy. Looking at the error score d_A superimposed on Figure 5.1, we see indeed that logTV has high errors in fractional anisotropy, a trend that we will also see in tests with in-vivo brain data. It could be argued that the regularisation parameter should be higher. We tried to increase it by a factor of two, but the S-shape become even more prominent and d_F increased, with the errors disappearing. We also expected some stair-casing effect for the first order methods in this region, but it is not noticeable. Only in the error plot of Figure 5.2 do we see it.

Concerning iteration counts (also superimposed on Figure 5.1), we observe TGV^2 for the large parameter choice has required 14 times as much iterations as the other models, and for the smaller parameter choice also 1.5 times as many. It however provides clearly the best results for the large parameter choice, by the error scores as well (superimposed on Figure 5.1). For the small parameter choice TV can be said to provide better results, but neither is so good as TGV^2 for the large parameter choice.

Finally, we note that the error measures superimposed on Figure 5.1(b_3) are comparable

to those on (a_1) and (a_4). This suggests that while visually (b_3) is worse, from the point of view of error measures the heuristic choice of $\alpha + \beta$ for TGV^2 equal to α for TV or TD, as discussed in Subsection 5.4, seems at least as reasonable as choosing the same α for all the models and a large β. This suspicion is however not confirmed by the following tests on in-vivo brain data, and indeed the choice of α seems more important than β.

5.8. In-vivo brain data. 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 ($b = 1000\mathrm{s/mm}^2$) 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 [36]. Given the four averages of 12 independent diffusion gradients, plus the zero gradient, we have altogether 52 different DWI measurements.

When acquiring diffusion weighted MRI data, there is always a tradeoff between SNR, the imaged field of view, spatial resolution and measurement time. From a clinical point of view, it would be extremely desirable to be able to obtain DTI data sets with whole brain coverage and an isotropic spatial resolution of approximately $1\mathrm{mm}^3$. However, clinical scan protocols are limited in measurement time to approximately 5 minutes for a diffusion weighted acquisition [38]. The reason for this is patient comfort, the required patient throughput in a clinical facility, as well as the inability to hold completely still for longer periods of time. This is of special importance for diffusion weighted imaging, as it is a technique that is prone to artifacts from patient movement.

Our data set is of rather high fidelity at the expense of low resolution. It is therefore suitable as a "ground truth" or "gold standard" to which additional noise is applied, and against which the results of denoising the noisy data are compared. 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 $\mathrm{Sym}^2(\mathbb{R}^3)$ tensors, to which the considerations in Remark 4.1 apply. We apply Rician noise with both the parameters $\sigma = 10$ ("low-noise case", PSNR $\approx 29.0\,\mathrm{dB}$) and $\sigma = 50$ ("high-noise case", PSNR $\approx 14.5\,\mathrm{dB}$) on the original DWI measurements, before the diffusion tensors are extracted from the Stejskal-Tanner equation (5.3). Regarding regularisation parameter, the maximal eigenvalue of the data $\mu \approx 0.0042$, and we have by trial and error chosen the "base regularisation parameter" $\alpha_0 = 0.05$ for the low-noise case and $\alpha_0 = 0.15$ for the high-noise case. The regularisation parameter for logTV is chosen as (5.1).

The computational results for the low-noise case are presented in Figure 5.3, and the error maps in Figure 5.4. The results for the high-noise case are presented in Figure 5.5 and the error maps in Figure 5.6. 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 masked out. Additionally, for the high-noise case, we include in Figure

5.7 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 data, containing the corpus callosum; the region is marked with a rectangle in Figure 5.5. For the low-noise case this plot is not included, because any differences between most of the methods are hard to observe.

5.9. Analysis of the low-noise case. We first study the low-noise case. From Figure 5.3 it is difficult to observe any significant differences between the models, except that TV and logTV exhibit stair-casing. Another noticeable aspect is 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: Compared to the noisy data f, the error measures (superimposed on Figure 5.3) 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 consistently had the best values for the error measures (superimposed on Figure 5.3) in comparison to TV, TD, 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, for which the large parameter choice can be argued to slightly over-regularise. That is, however, not the case for the small parameter choice. 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.

More interesting information is provided by the error plot in Figure 5.4. Regarding the error in the direction of principal eigenvector, plotted with colours, we observe no significant difference between the models. In comparison to f, we observe that for the large parameter choice, all the models introduce significant *local* 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 small parameter choice, similar conclusions hold, but the differences are not so easy to detect by visual observation. For the large parameter choice, TV, logTV and DWITV exhibit higher local errors than TGV^2 and TD. For latter the difference between the large and the small parameter choice is the smallest, also confirmed by inspection of Figure 5.3. This indicates that these models, both using the symmetrised differential, are less sensitive towards parameter choice than TV and logTV.

5.10. Analysis of the high-noise case. We now turn to the high-noise case, for which the computational results are depicted in Figure 5.5. Here the reconstruction by all the models is already significantly poorer than the original image f_0. For the large parameter choice TGV^2 and TD provide nevertheless visually acceptable results, given the high noise levels, while both TV and logTV exhibit significant stair-casing. For the small parameter choice, none of the methods manage to remove significant amounts of noise although logTV exhibits stair-casing already then. Indeed, we note that logTV does not fare well in this test case, introducing large errors and noise in fractional anisotropy construction, that are visually observable already from this result plot for both parameter choice. However, with regard to the fractional anisotropy error measure d_A, DWITV performs even worse. TV has the best fractional anisotropy reconstruction by error, with TD and TGV^2 not far behind. Inspecting

the error plot in Figure 5.5, TV however has the highest local errors errors in fractional anisotropy for the large parameter choice.

Studying the zoomed-in Figure 5.7, we confirm the noise that logTV introduces in fractional anisotropy. For the small parameter choice we note none of 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 5.5) is comparable for all the models. For the large parameter choice logTV seems to be over-regularising the directions, while TV has apparently the best reconstruction of the principal eigenvectors here. That is, 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.

6. Some final remarks and outlook. From our numerical studies, we conclude that TGV^2 and TD regularisation appear to be very reasonable novel approaches for the denoising of diffusion tensors. Both are incidentally based on the symmetric differential. While DWITV or TV sometimes perform better by some of the error measures, TGV^2 and TD perform consistently good. DWITV, in particular, suffers from errors in fractional anisotropy. So does logTV, which we had expected 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 5.3, of negative eigenvalues in the noisy data, which the model does not directly allow.

TGV^2 never performs noticeable worse than TD, and sometimes better, but is computationally significantly more expensive. The choice between these two regularisers, according to our research so far, is therefore a matter of computational resources. 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 (5.3) 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, 41].

Acknowledgements. The authors would like to thank Karl Koschutnig (Department of Psychology, University of Graz) for providing the in-vivo brain DWI data.

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

Appendix 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*

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

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 3.2, 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

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

whence

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

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

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

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

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

This concludes the proof. ■

Appendix B. Alternative tensor norms.

In this appendix, we discuss the so called the largest and smallest *reasonable cross-norms*, following [18]. A word of note, however: all the results in [18] 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

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

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

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

It may alternatively be written as

$$\|A\|_\wedge = \inf\{\sum_{i=1}^N \|x_1^i\|_p \cdots \|x_k^i\|_p \mid A = \sum_{i=1}^N x_1^i \otimes \cdots \otimes x_k^i\}. \tag{B.2}$$

Example B.1 (Vectors). We have $\|A\|_\wedge = \|A\|_p$, and $\|A\|_\vee = \|A\|_q$, where $1/p + 1/q = 1$. Thus for $p = 2$ both norms are the same.

Example B.2 (Matrices). 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*.

Only the case $p = 2$ is of interest to us, as in that case these norms are orthogonally invariant, as is the Frobenius norm. Indeed, the following proposition is the counterpart to Proposition 2.2. As a consequence, Proposition 2.4 holds for these norms.

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

$$A_R(c_1, \dots, c_k) = A(Rc_1, \dots, Rc_k).$$

Then the norms $\|\cdot\|_\bullet$ for $\bullet = \vee, \wedge$ and $p = 2$, are orthogonally invariant in the sense that $\|A_R\|_\bullet = \|A\|_\bullet$.

Proof. That $\|\cdot\|_\vee$ is orthogonally invariant, is clear from the the equation (B.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 orthogonally invariant is clear from the formulation (B.2): If $A = \sum_{i=1}^N x_1^i \otimes \cdots \otimes x_k^i$, then $A_R = \sum_{i=1}^N R^T x_1^i \otimes \cdots \otimes R^T x_k^i$. ∎

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), pp. 1389–1403.

[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), pp. 366–378.

[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), pp. 1138–1150.

[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.*, Tech. Report 5584, INRIA, 2005.

[6] ———, *Log-euclidean metrics for fast and simple calculus on diffusion tensors*, Magnetic Resonance in Medicine, 56 (2006), pp. 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), pp. 90–102.

[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), pp. 456–467.

[9] R. L. Bishop and S. I. Goldberg, *Tensor Analysis on Manifolds*, Dover Publications, Dover ed., 1980.

[10] K. Bredies, *Symmetric tensor fields of bounded deformation*, Ann. Mat. Pura Appl., (2012). Accepted.

[11] K. Bredies, K. Kunisch, and T. Pock, *Total generalized variation*, SIAM J. Imaging Sci., 3 (2011), pp. 492–526.

[12] 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.

[13] 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.

[14] A. Chambolle, *An algorithm for total variation minimization and applications*, J. Math. Imaging Vision, 20 (2004), pp. 89–97.

[15] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision, 40 (2011), pp. 120–145.

[16] C. Chefd'hotel, D. Tschumperlé, R. Deriche, and O. Faugeras, *Regularizing flows for constrained matrix-valued images*, J. Math. Imaging Vision, 20 (2004), pp. 147–162.

[17] O. Coulon, D. Alexander, and S. Arridge, *Diffusion tensor magnetic resonance image regularization*, Medical Image Anal., 8 (2004), pp. 47–67.

[18] J. Diestel, J. H. Fourie, and J. Swart, *The metric theory of tensor products: Grothendieck's résumé revisited*, American Mathematical Society, 2008.

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

[20] 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), pp. 1472–1482.

[21] L. Florack and L. Astola, *A multi-resolution framework for diffusion tensor images*, in Computer Vision and Pattern Recognition Workshops, 2008. CVPRW '08. IEEE Computer Society Conference on, june 2008, pp. 1 –7.

[22] L. Florack and H. van Assen, *Multiplicative calculus in biomedical image analysis*, J. Math. Imaging Vision, 42 (2012), pp. 64–75.

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

[24] Y. Gur, O. Pasternak, and N. Sochen, *Fast $GL(n)$-invariant framework for tensors regularization*, Int. J. Comput. Vis., 85 (2009), pp. 211–222.

[25] 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).

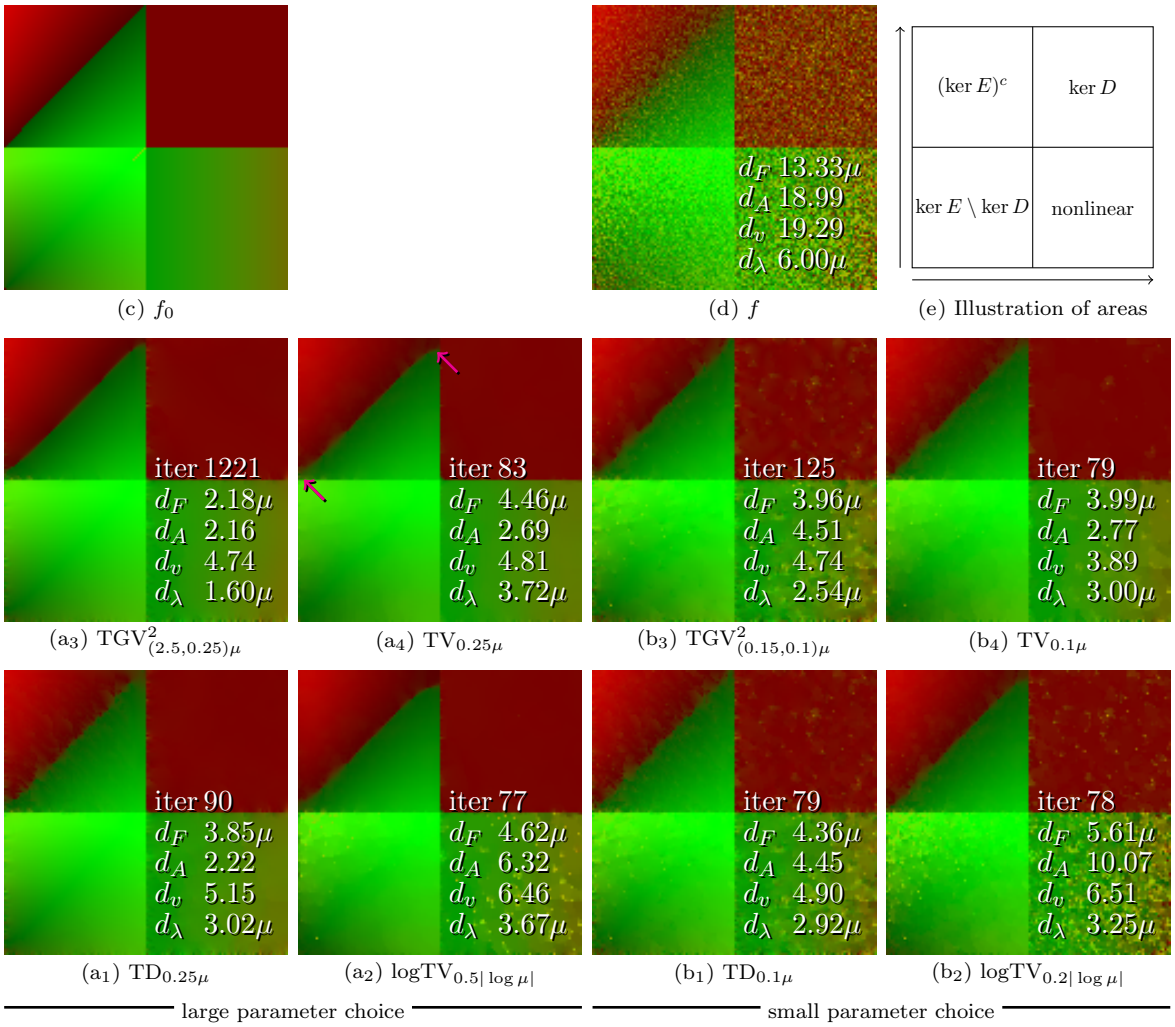
[26] B. Hofmann, B. Kaltenbacher, C. Pöschl, and O. Scherzer, *A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators*, Inverse Problems, 23 (2007), p. 987.

[27] P. Kingsley, *Introduction to diffusion tensor imaging mathematics: Parts I-III*, Concepts in Magnetic Resonance Part A, 28 (2006), pp. 101–179. 10.1002/cmr.a.20049, 10.1002/cmr.a.20050.

[28] F. Knoll, K. Bredies, T. Pock, and R. Stollberger, *Second order total generalized variation (TGV) for MRI.*, Magnetic Resonance in Medicine, 65 (2011), pp. 480–491.

[29] 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, p. 5088.

[30] T. Kolda and B. Bader, *Tensor decompositions and applications*, SIAM Rev., 51 (2009), pp. 455–500.
[31] A. Lewis, *The convex analysis of unitarily invariant matrix functions*, J. Convex Anal., 2 (1995), pp. 173–183.
[32] X. Pennec, P. Fillard, and N. Ayache, *A Riemannian Framework for Tensor Computing*, Int. J. Comput. Vis., 66 (2006), p. 41–66.
[33] 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 Interventation–MICCAI98, Springer, 1998, pp. 489–498.
[34] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D, 60 (1992), pp. 259–268.
[35] S. Setzer, G. Steidl, B. Popilka, and B. Burgeth, *Variational methods for denoising matrix fields*, in Visualization and Processing of Tensor Fields, J. Weickert and H. Hagen, eds., Springer, 2009, pp. 341–360.
[36] 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), pp. S208–S219.
[37] R. Temam, *Mathematical problems in plasticity*, Gauthier-Villars, 1985.
[38] J.-D. Tournier, S. Mori, and A. Leemans, *Diffusion tensor imaging and beyond*, Magnetic Resonance in Medicine, 65 (2011), pp. 1532–1556.
[39] D. Tschumperle and R. Deriche, *Vector-valued image regularization with PDEs: a common framework for different applications*, IEEE TPAMI, 27 (2005), pp. 506 –517.
[40] D. Tschumperlé and R. Deriche, *Orthonormal vector sets regularization with PDE's and applications*, J. Math. Imaging Vision, 50 (2002), pp. 237–252.
[41] 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), pp. 1342–1348.
[42] T. Valkonen, *Diff-convex combinations of Euclidean distances: a search for optima*, no. 99 in Jyväskylä Studies in Computing, University of Jyväskylä, 2008. Ph.D Thesis.
[43] Z. Wang, B. Vemuri, Y. Chen, and T. Mareci, *A constrained variational principle for direct estimation and smoothing of the diffusion tensor field from complex DWI*, IEEE Trans. Medical Imaging, 23 (2004), pp. 930 –939.
[44] J. Weickert, *Anisotropic Diffusion in Image Processing*, ECMI series, B.G. Teubner Stuttgart, 1998.

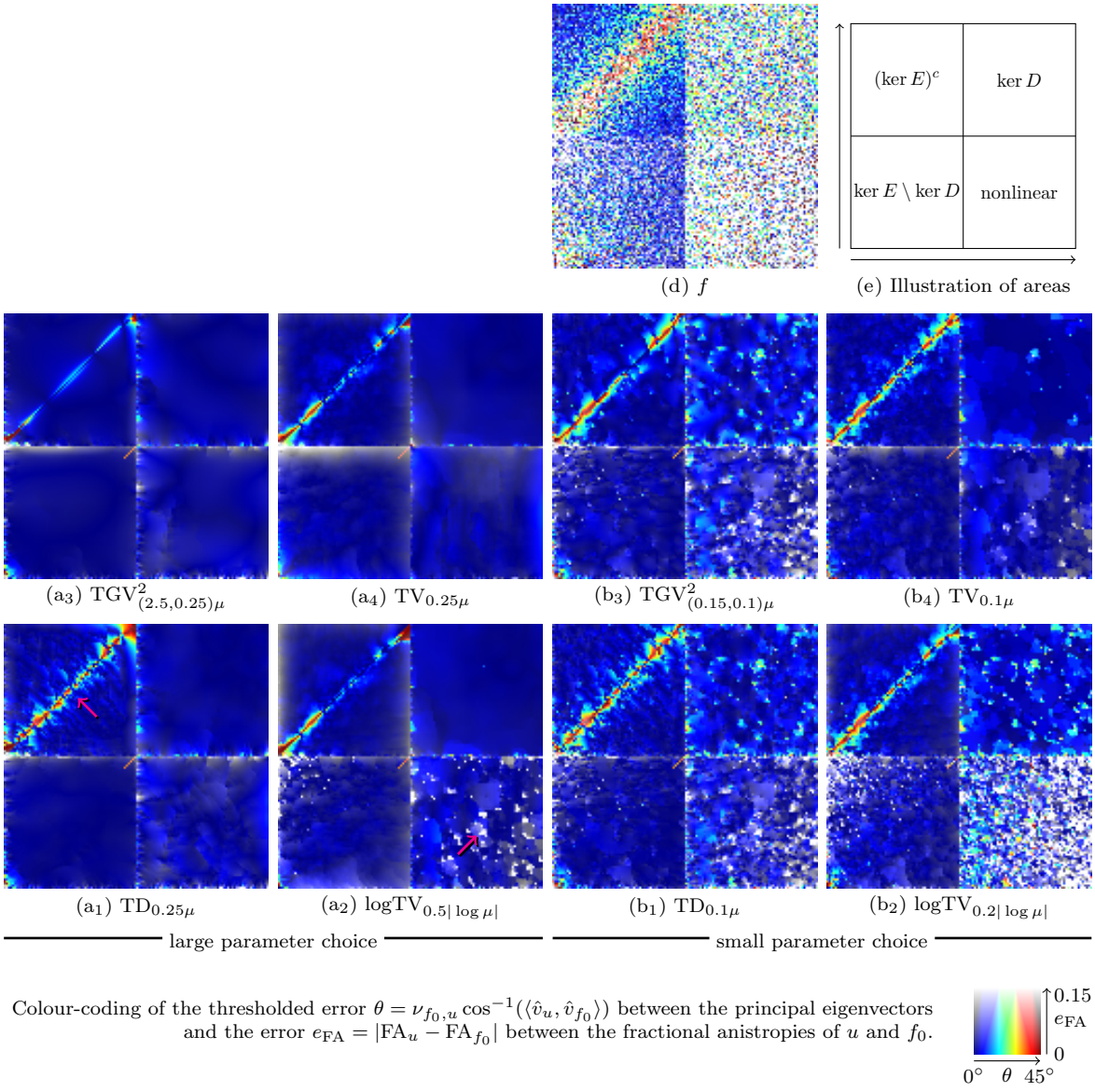


(c) f_0 (d) f (e) Illustration of areas

(a_3) $\mathrm{TGV}^2_{(2.5,0.25)\mu}$ (a_4) $\mathrm{TV}_{0.25\mu}$ (b_3) $\mathrm{TGV}^2_{(0.15,0.1)\mu}$ (b_4) $\mathrm{TV}_{0.1\mu}$

(a_1) $\mathrm{TD}_{0.25\mu}$ (a_2) $\mathrm{logTV}_{0.5|\log\mu|}$ (b_1) $\mathrm{TD}_{0.1\mu}$ (b_2) $\mathrm{logTV}_{0.2|\log\mu|}$

Figure 5.1. *Computational results for* TD, logTV, TV, *and* TGV^2 *on the synthetic test data (c) with Rician noise of parameter* $\sigma = 0.15$ *(DWI domain PSNR* $\approx$ *34.4 dB) applied (d). The visualisation is as in the legend above. The "base regularisation parameter" found by trial and error is* $\alpha_0 = 0.25$*, and the parameters for each model, displayed under the plots, are chosen as in Section 5.4. The results of the "large parameter choice" are displayed in* (a_1)*–*(a_4)*, while the results for the "small parameter choice" are displayed in* (b_1)*–*(b_4)*. The iteration counts and error measures discussed in Subsection 5.1 have been superimposed on the results. The areas of the synthetic data are illustrated in (e). Observe in particular that while* TGV^2 *in* (a_3) *restores the diagonal in the top-left region well,* TV *and* logTV *in* (a_4)*,*(a_2) *exhibit a quite prominent S-shape (indicated by the arrows in the* TV *plot), while* TD *in* (a_1) *exhibits blurring at the ends of the diagonal.*



Colour-coding of the thresholded error $\theta = \nu_{f_0,u} \cos^{-1}(\langle \hat{v}_u, \hat{v}_{f_0} \rangle)$ between the principal eigenvectors and the error $e_{\mathrm{FA}} = |\mathrm{FA}_u - \mathrm{FA}_{f_0}|$ between the fractional anistropies of u and f_0.

Figure 5.2. *Error visualisation of the computational results for* TD, logTV, TV, *and* TGV^2 *on the synthetic test data (c) with Rician noise of parameter* $\sigma = 0.15$ *(DWI domain PSNR* $\approx$ *34.4 dB) applied (d). The visualisation is as in the legend above. The "base regularisation parameter" found by trial and error is* $\alpha_0 = 0.25$*, and the parameters for each model, displayed under the plots, are chosen as in Section 5.4. The results of the "large parameter choice" are displayed in* (a_1)*–*(a_4)*, while the results for the "small parameter choice" are displayed in* (b_1)*–*(b_4)*. The areas of the synthetic data are illustrated in (e). Observe in particular that while* TGV^2 *in* (a_3) *has small errors close to the diagonal in the top-left region, all* TD TV, logTV (a_1),(a_2), *and* (a_4),, *exhibit noisy error around the diagonal, particularly for TD, as indicated by the arrow. Moreover,* logTV *exhibits high errors in fractional anisotropy in both of the bottom rectangles, as indicated by the arrow.*

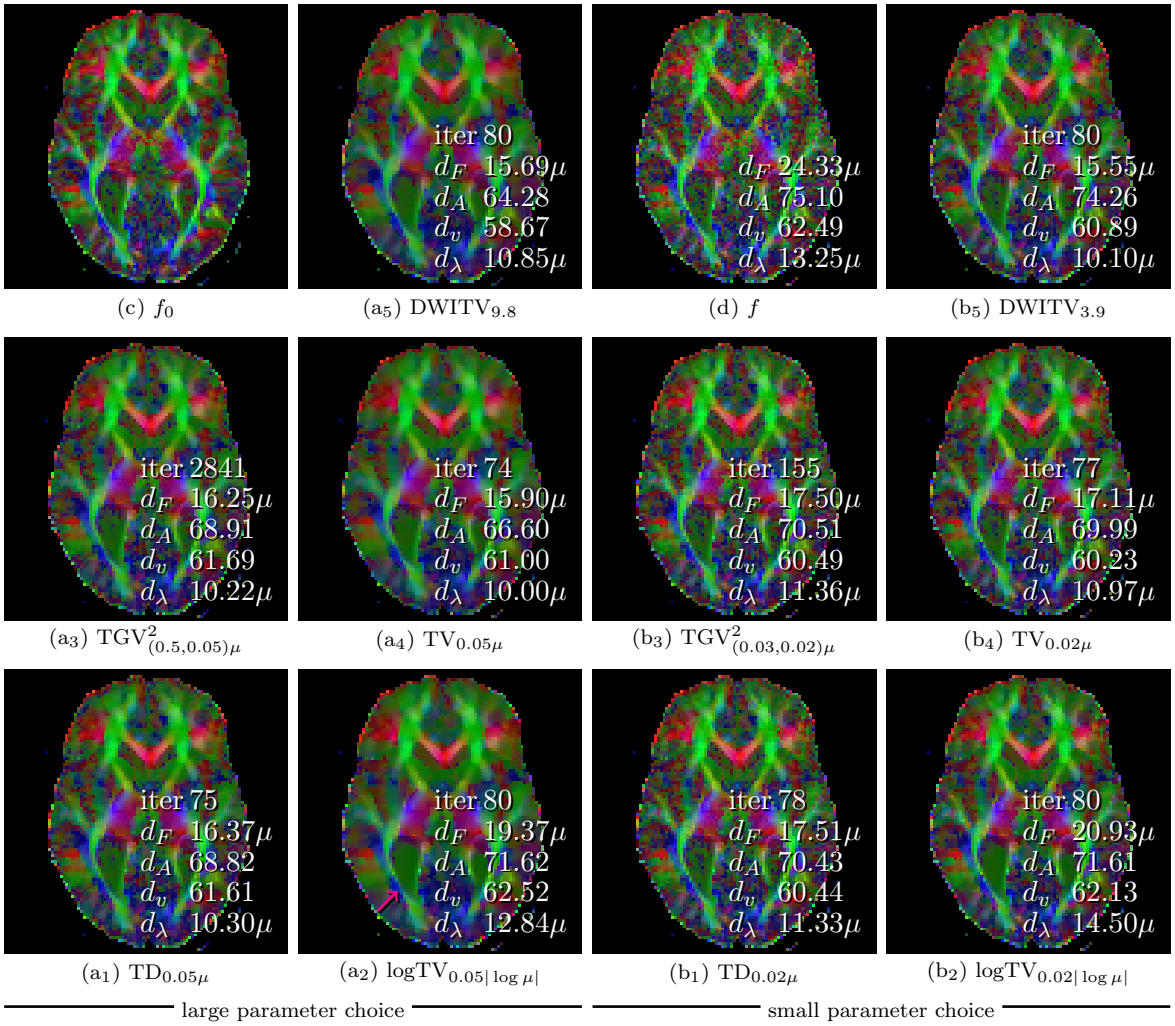


Figure 5.3. *Computational results for* TD, logTV, TV, TGV^2, *and* DWITV *on the in-vivo brain data (c) with Rician noise of parameter* $\sigma = 10$ *(DWI domain PSNR* $\approx$ *29.0 dB) applied (d). The visualisation is as in the legend above. The "base regularisation parameter" found by trial and error is* $\alpha_0 = 0.05$*, and the parameters for each model, displayed under the plots, are chosen as in Section 5.4. The results of the "large parameter choice" are displayed in* (a_1)*–*(a_4)*, while the results for the "small parameter choice" are displayed in* (b_1)*–*(b_4)*. The iteration counts and error measures discussed in Subsection 5.1 have been superimposed on the results. Observe in particular the stair-casing of* TV *and* logTV *particularly in the green region on lower-left indicated by the arrow in the* logTV *plot.*

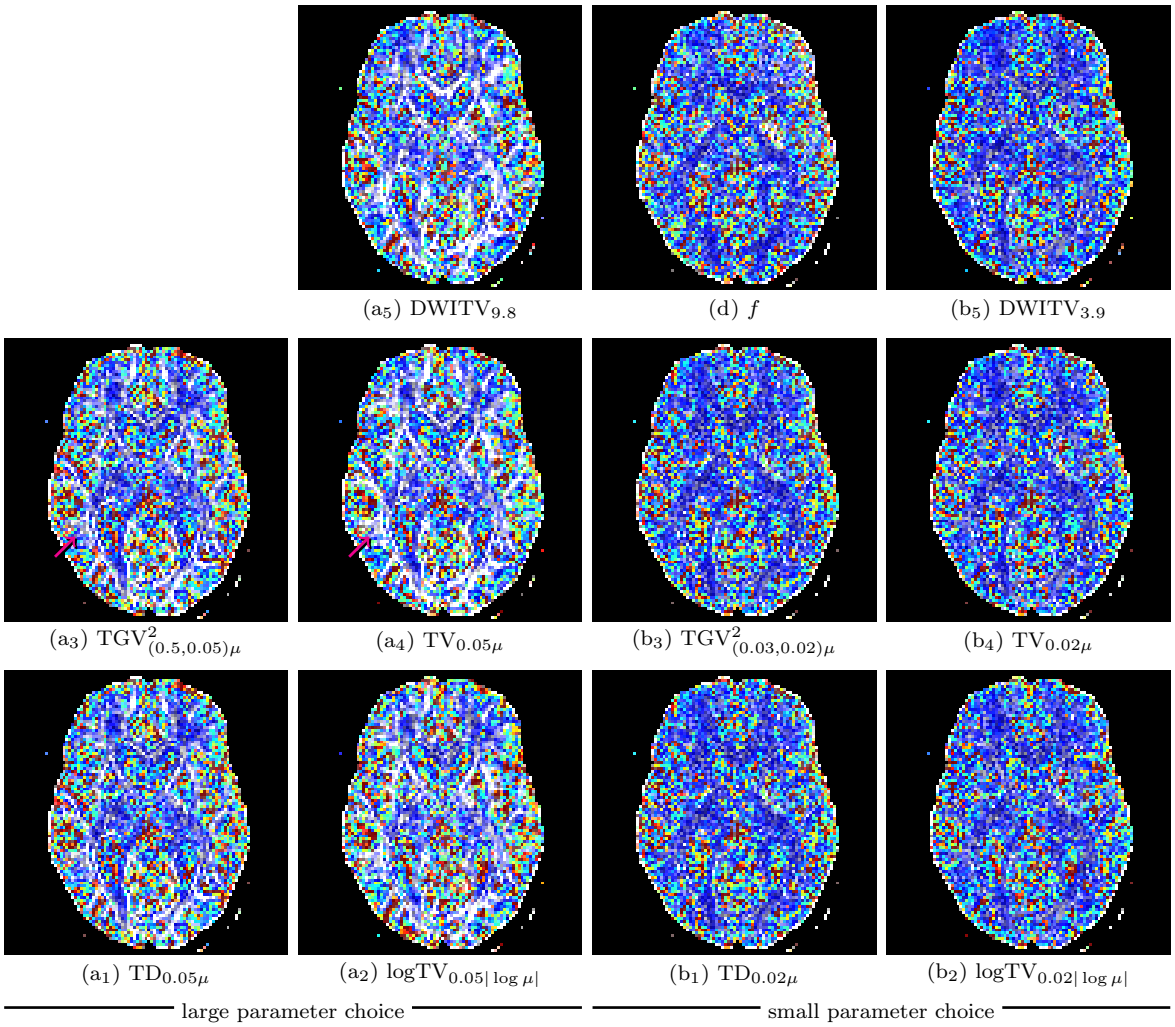


Figure 5.4. *Error visualisation of the computational results for* TD, logTV, TV, TGV^2, *and* DWITV *on the in-vivo brain data (c) with Rician noise of parameter* $\sigma = 10$ *(DWI domain PSNR* ≈ 29.0 *dB) applied (d). The visualisation is as in the legend above. The "base regularisation parameter" found by trial and error is* $\alpha_0 = 0.05$*, and the parameters for each model, displayed under the plots, are chosen as in Section 5.4. The results of the "large parameter choice" are displayed in* (a_1)*–*(a_4)*, while the results for the "small parameter choice" are displayed in* (b_1)*–*(b_4)*. Observe in particular how* TGV^2 *(and* TD*) restore the fractional anisotropy better than* TV *(or* logTV*), with less intense white areas, as indicated by the arrows.*

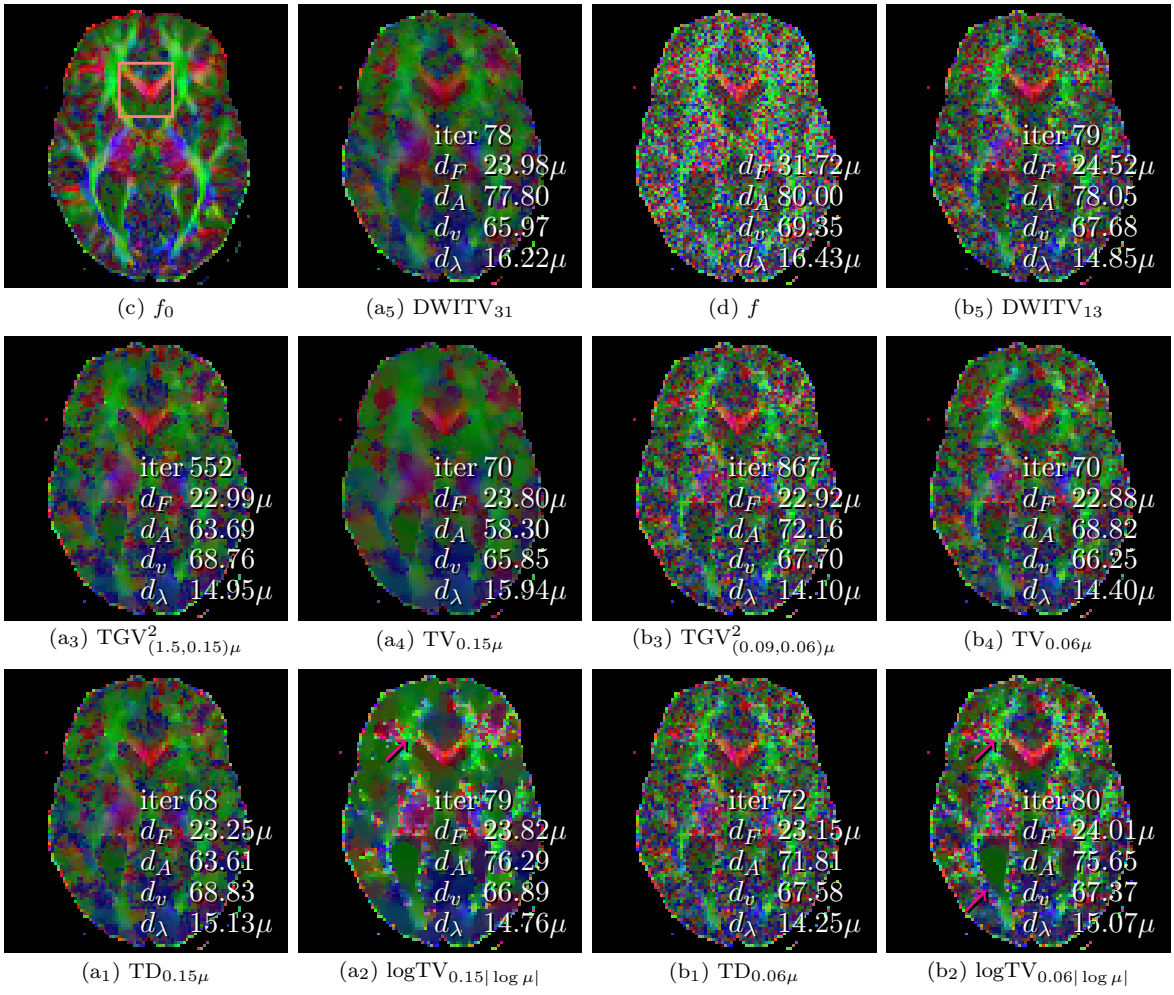


Figure 5.5. *Computational results for* TD, logTV, TV, TGV^2, *and* DWITV *on the in-vivo brain data (c) with Rician noise of parameter* $\sigma = 50$ *(DWI domain PSNR* $\approx$ *14.5 dB) applied (d). The visualisation is as in the legend above. The "base regularisation parameter" found by trial and error is* $\alpha_0 = 0.15$*, and the parameters for each model, displayed under the plots, are chosen as in Section 5.4. The results of the "large parameter choice" are displayed in* (a_1)*–*(a_4)*, while the results for the "small parameter choice" are displayed in* (b_1)*–*(b_4)*. The iteration counts and error measures discussed in Subsection 5.1 have been superimposed on the results. The rectangle in (c) indicates the corpus callosum displayed in more detail in Figure 5.7. Observe in particular the small extremely bright spots (high fractional anisotropy) exhibited by* logTV*, indicated by the arrow, and the extreme stair-casing of* TV *for the large parameter choice. At the same time,* TGV^2 *and* TD *perform reasonably well for the large parameter choice, while all methods leave significant amounts of noise for the small parameter choice. And, yet,* logTV *exhibits stair-casing already for the small parameter choice, as indicated by an arrow.*

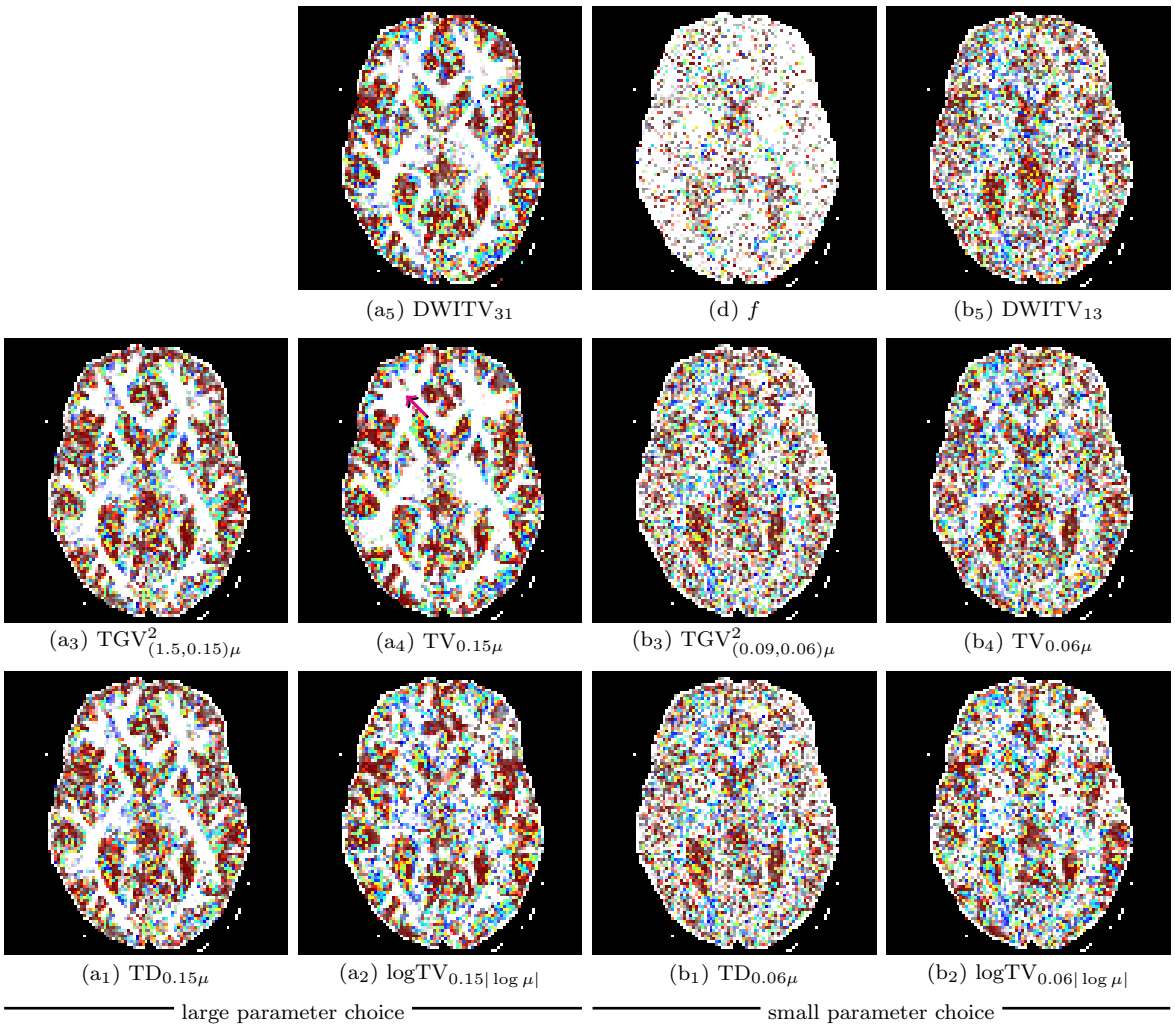


Figure 5.6. *Error visualisation of the computational results for* TD, logTV, TV, TGV^2, *and* DWITV *on the in-vivo brain data (c) with Rician noise of parameter* $\sigma = 50$ *(DWI domain PSNR* $\approx$ *14.5 dB) applied (d). The visualisation is as in the legend above. The "base regularisation parameter" found by trial and error is* $\alpha_0 = 0.15$*, and the parameters for each model, displayed under the plots, are chosen as in Section 5.4. The results of the "large parameter choice" are displayed in* (a_1)*–*(a_4)*, while the results for the "small parameter choice" are displayed in* (b_1)*–*(b_4)*. Observe the high errors of fractional anisotropy exhibited by* TV*, as indicated by the arrow.*

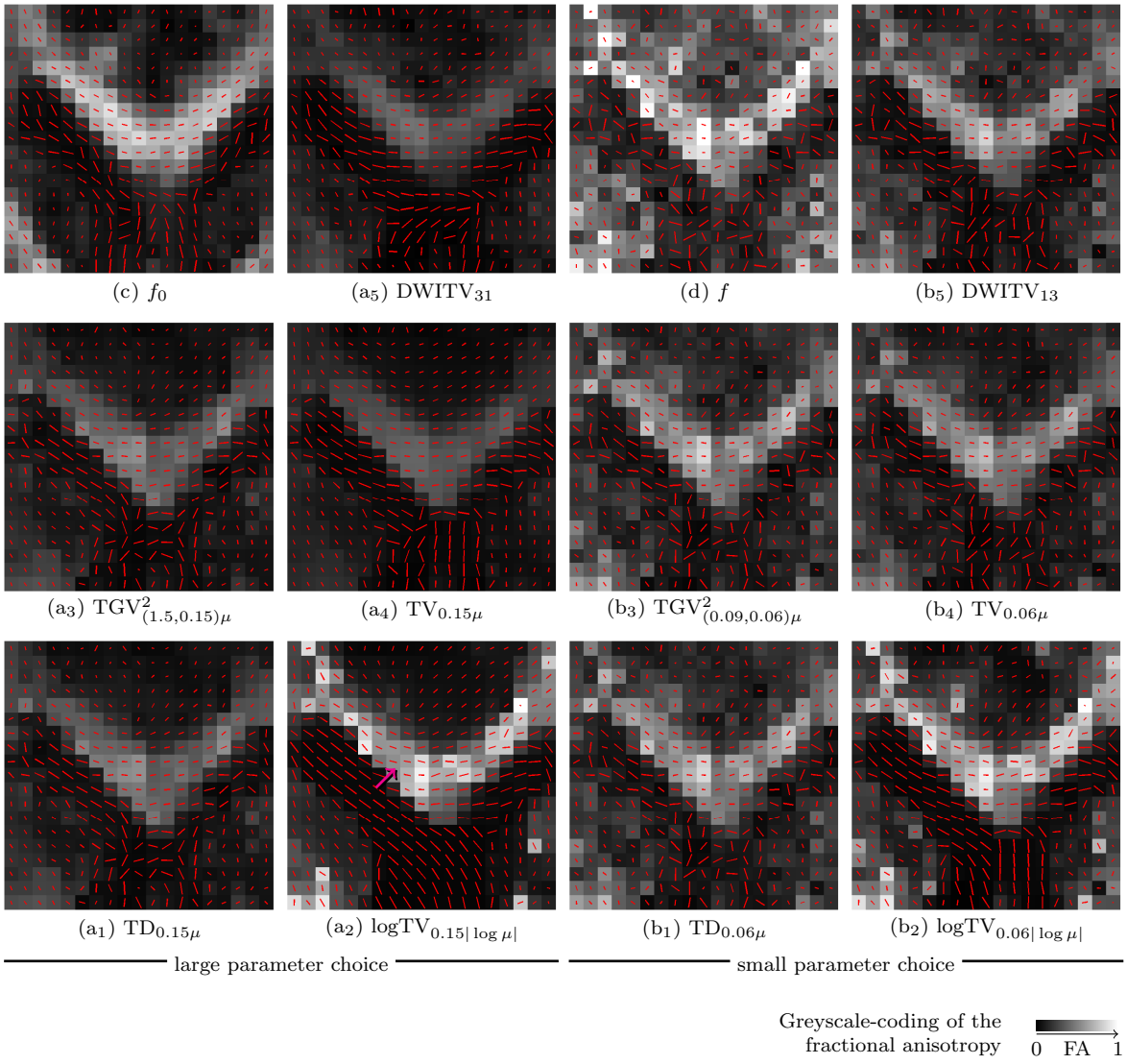


Figure 5.7. *Fractional anisotropy (see legend above) and planar projection of the principal eigenvector for the region containing the corpus callosum of the in-vivo brain data (c). Rician noise of parameter $\sigma = 50$ (DWI domain PSNR $\approx$ 14.5 dB) has been applied to yield the noisy data (d). The "base regularisation parameter" found by trial and error is $\alpha_0 = 0.15$, and the parameters for each model, displayed under the plots, are chosen as in Section 5.4. Denoising results of the the noisy data for the "large parameter choice" with for* TD, logTV, TV, TGV^2, *and* DWITV *are displayed in (a_1)–(a_5), while the denoising results for the "small parameter choice" are displayed in (b_1)–(b_5). Observe the noisy reconstruction of fractional anisotropy by* logTV, *as indicated by the arrow.*