Combined Tensor Fitting and TV Regularization in Diffusion Tensor Imaging Based on a Riemannian Manifold Approach

Maximilian Baust, Andreas Weinmann*, Matthias Wieczorek, Tobias Lasser, Martin Storath, and Nassir Navab

***Abstract*—In this paper, we consider combined TV denoising and diffusion tensor fitting in DTI using the affine-invariant Riemannian metric on the space of diffusion tensors. Instead of first fitting the diffusion tensors, and then denoising them, we define a suitable TV type energy functional which incorporates the measured DWIs (using an inverse problem setup) and which measures the nearness of neighboring tensors in the manifold. To approach this functional, we propose generalized forward-backward splitting algorithms which combine an explicit and several implicit steps performed on a decomposition of the functional. We validate the performance of the derived algorithms on synthetic and real DTI data. In particular, we work on real 3D data. To our knowledge, the present paper describes the first approach to TV regularization in a combined manifold and inverse problem setup.**

***Index Terms*—Combined denoising and diffusion tensor fitting, diffusion tensor imaging, generalized forward-backward algorithm, manifold-valued data, total variation minimization.**

I. Introduction

DIFFUSION tensor imaging (DTI) is an imaging modality based on nuclear magnetic resonance. It has become extremely popular during the last decades as it allows to assess the diffusional characteristics of a specimen in a non-invasive way. Applications of DTI include, but are not limited to, determination of fiber tract orientations [1], which is of great interest for planning surgeries, detection of brain ischemia [2], or investigation of neurodegenerative pathologies such as schizophrenia [3], [4], autism [5], or Huntington's disease [6].

In DTI, each pixel (or voxel), consists of a positive definite matrix, a so-called *diffusion tensor*. These positive matrices – seen as covariance matrices of multivariate normal distributions – model the diffusivity of water molecules in space, which is determined by a series of diffusion weighted images (DWIs). Each DWI measures the directional diffusivity with respect to a given gradient direction. As a consequence, the measured data consists of one real-valued vector per voxel and the task of diffusion tensor estimation is to estimate the diffusion tensor field from this data. For a detailed description of the fundamental principles of DTI we refer the interested reader to the articles of Basser *et al.* [1], Johansen *et al.* [7], Assaf *et al.* [8], and Assemlal *et al.* [9].

Unfortunately, however, DWIs typically suffer from Rician noise which make the reconstruction of the DTI volumes a difficult problem, because the positive definiteness of the reconstructed tensors is a non-linear constraint and as such hard to enforce, see Fillard *et al.* [10] for instance. As a consequence, there is a large body of literature on DTI reconstruction. In principle, one may distinguish four different types of approaches: i) fitting the tensors individually, i.e., independently per voxel, while additionally imposing (soft or hard) constraints to enforce the positive definiteness of the reconstructed tensors [11], [12], ii) denoising the input data, i.e., the DWIs, and then fitting the tensors individually [13]–[15], iii) regularizing the tensors after reconstruction [16]–[22], and iv) reconstructing and regularizing the tensors simultaneously [10], [23]–[27]. Methods related to the latter approach are more intricate, but they show the best performance with respect to reconstruction quality [10], [28]. In this work, we propose a variational approach which belongs to the latter category and thus aims for simultaneous reconstruction and regularization. It is formulated in a manifold setting using the manifold of symmetric positive definite matrices, and employs the manifold-valued version of the total variation (TV) as a regularizer. To our knowledge, this is the first combined approach in this (nonflat) Riemannian manifold setting.

Due to the amount, diversity, and variety of related work, we postpone its detailed discussion to Section II and continue with the presentation of our approach.

Manuscript received October 02, 2015; revised January 28, 2016; accepted February 02, 2016. Date of publication April 27, 2016; date of current version July 29, 2016. M. Baust and N. Navab acknowledge the support of the Collaborative Research Centre 824 (Z2) "Imaging for Selection, Monitoring and Individualization of Cancer Therapies" funded by the German Science Foundation. A. Weinmann acknowledges support by the Helmholtz Association within the young investigator group VH-NG-526 and by the DFG Grant WE 5886/3-1. M. Wieczorek was supported by the DFG Cluster of Excellence Munich-Centre for Advanced Photonics (MAP) . M. Storath was supported by the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no. 267439. A. Weinmann and M. Storath acknowledge funding within the DFG Young Researchers' Network "Mathematics for Magnetic Particle Imaging" . M. Baust and A. Weinmann are joint first authors. *Asterisk indicates corresponding author.*

M. Baust, M. Wieczorek, T. Lasser, and N. Navab are with the Chair for Computer Aided Medical Procedures, Technische Universität München, 85748 Munich, Germany (e-mail: fmaximilian.baust@tum.de).

*A. Weinmann is with the Department of Mathematics, Technische Universität München, 85748 Munich, Germany, and with Helmholtz Zentrum München, 85764 Munich, Germany, and also with the Darmstadt University of Applied Sciences, 64295 Darmstadt, Germany (e-mail: andreas.weinmanng@tum.de).

M. Storath is with the Image Analysis and Learning Group, Universität Heidelberg, 69120 Heidelberg, Germany .

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2016.2528820

A. TV Regularization in the Inverse and Manifold Setting

We explain our new approach in the abstract setup and specify the respective meaning in the DTI context. First, we need a model or signal space which, due to (hard) constraints, typically constitutes a manifold endowed with a Riemannian distance, rather than a linear space. In DTI, a suitable model space is the Riemannian symmetric space of positive matrices with the underlying affine-invariant metric. In contrast to the model space, the data space is typically a linear Euclidean space where each component of the respective vectors is a real-valued measurement. In DTI, the measured data at each voxel can be seen as a vector of diffusion weighted values, i.e., the values of the DWIs at the corresponding voxel with each component corresponding to a different weighting direction v_i. Hence, the DWI data lives in a linear Euclidean space. Thus, the distance in the model space is measured with respect to the distance in the manifold, whereas the data fidelity is measured with respect to the Euclidean distance. The measurement or modeling operator $\mathcal{A}$ maps from the manifold to this linear data space. In DTI, an example of such an operator is the Stejskal-Tanner operator (which is discussed later in detail).

On an abstract level, equipped with the measurement operator $\mathcal{A}$, and using the notation $F = (F_i)_i$ for the observed data, the inverse and TV-regularized problem in the manifold-valued setup is given by

$$\min_U \mathcal{J}_\gamma(U,F), \quad \mathcal{J}_\gamma(U,F) = \mathcal{D}(U,F) + \gamma \mathcal{TV}(U). \tag{1}$$

Here, $\mathcal{D}$ is a data term enforcing U to fit to the data F_i; in DTI, the reconstructed diffusion tensors $(U_i)_i$ are required to fit the measured DWIs F_i. An example is

$$\mathcal{D}(U,F) = \sum_{i,k} \left|(\mathcal{A}U)_i^k - F_i^k\right|^2, \tag{2}$$

where k denotes the number of the measurement, i.e., the DWI, and i denotes the voxel number; for simplicity we consider the univariate case here. In the context of DTI, the measurement operator $\mathcal{A}$ is given by (7). It is also possible to generalize (2) by replacing the squared Euclidean distance by another computationally accessible differentiable function h resulting in

$$\mathcal{D}(U,F) = \sum_{i,k} h\left((\mathcal{A}U)_i^k, F_i^k\right). \tag{3}$$

The function h employed in the context of DTI in this paper is given by (9). Furthermore,

$$\mathcal{TV}(U) = \sum_i d(U_i, U_{i+1}) \tag{4}$$

denotes the discrete total variation (TV) regularizer with the distance d measured in the manifold, and γ being a model parameter controlling the trade-off between data fidelity and regularity. For introductory purposes, we have only described the 1D case; in the detailed derivation later on we consider the multivariate case as well.

In order to give (1) a concrete meaning for DTI via (2), (3), and (4), we next specify the distance d, the function h and the imaging operator $\mathcal{A}$. The distance d is considered in Section I-B, the function h and the imaging operator $\mathcal{A}$ are specified by (7) and (9), respectively, in Section I-C.

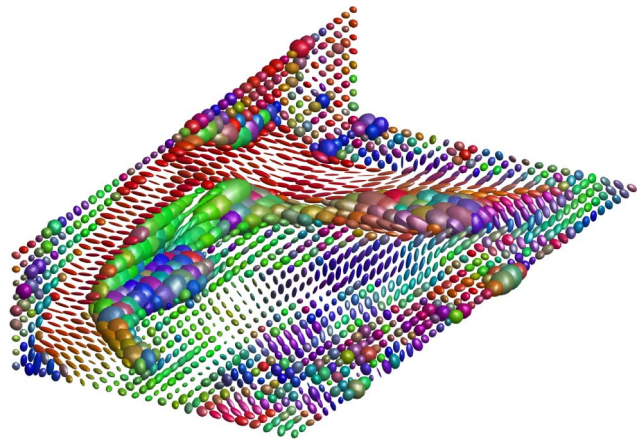


Fig. 1. Orthogonal slices of a 3D reconstruction of a diffusion tensor volume obtained from real data of the Camino project. The volume was reconstructed using the proposed Algorithm 1 with data term $\widehat{\mathcal{D}}$ for combined tensor fitting and TV denoising with model parameter $\gamma = 1.0$.

B. The Affine-Invariant Riemannian Manifold in DTI

In DTI, the tensors U_i may be viewed as elements in the Riemannian manifold of positive (definite) matrices; see [19] for instance. The underlying distance corresponds to the Fisher-Rao metric [29], which is statistically motivated as the positive matrices represent covariance matrices modeling the diffusivity of water molecules. Consequently, oriented diffusivity along fiber structures is reflected by the anisotropy of the corresponding tensors; typically, there is one large eigenvalue and the corresponding eigenvector yields the orientation of the fiber. We will denote the space of 3×3 diffusion tensors by Pos_3 and equip it with the Riemannian metric

$$g_D(W,V) = \mathrm{trace}(D^{-1/2} W D^{-1} V D^{-1/2}), \tag{5}$$

where the symmetric matrices W, V represent tangent vectors in the point $D \in \mathrm{Pos}_3$. Equipped with this metric the space of positive matrices becomes a Cartan-Hadamard manifold. This means that it is complete, simply connected, and of non-positive sectional curvature. This class of manifolds enjoys particularly nice properties not shared by general Riemannian manifolds. For example, geodesics – in the sense of curves without intrinsic acceleration when moving with constant speed – are always, not only locally, shortest paths. Furthermore, the mathematically very important notion of convexity generalizes to these spaces in a convenient way. This allows to derive global results in the context of variational methods [30]–[32]. Instead of considering Pos_3 as a Hadamard manifold with the above Riemannian metric, it is also possible to take other points of view, for example, by interpreting the positive matrices as the positive cone in the space of matrices endowed with the corresponding Euclidean distance. We discuss these issues in more detail in Section V-A1.

C. Combined Tensor Fitting and TV Regularization Using the Affine-Invariant Riemannian Manifold in DTI

In order to apply the proposed model (1) for combined tensor fitting and TV regularization, we have to specify the regularizer (4) and the data term. For the regularizer, we utilize the distance d which is induced by the Riemannian metric discussed in the previous subsection. It remains to specify the data term. One of the most popular models for DTI tensor fitting is

$$\mathcal{D}(U,F) = \sum_{i,k} \left| b\, v_k^T U_i v_k - \log\left(\frac{A_i^0}{F_i^k}\right)\right|^2 \tag{6}$$

using the imaging operator

$$\mathcal{A}(U_i) = (b\, v_k^T U_i v_k)_k \tag{7}$$

which is based on the well-known Stejskal-Tanner equation [33] (see Section III) and which corresponds to the setting in (2). Here U_i denotes the target variable representing the sought tensor at position i and $(F_i^k)_k$ is the vector-valued DWI measurement at position i. The vector v_k corresponds to the k-th DWI measurement and encodes the gradient direction according to which F_i^k has been measured. Furthermore, b and A_i^0 are normalization constants, where A_i^0 depends on the spatial position and corresponds to an unweighted measurement.

As a least-squares-type data term, (6) is matched to Gaussian noise on the logarithm of the measurements. The Gaussian assumption is not always appropriate. A frequently used model is Rician noise. According to Fillard *et al.* [10], the data term

$$\hat{\mathcal{D}}(U,F) = -\sum_{i,k} \log\left(\frac{F_i^k}{\sigma^2}\exp\left(-\frac{P_i^k(U_i)^2+(F_i^k)^2}{2\sigma^2}\right) I_0\left(\frac{P_i^k(U_i)F_i^k}{\sigma^2}\right)\right) \tag{8}$$

is particularly suited for Rician noise. Here,

$$P_i^k(U_i) = A_i^0 \exp(-b v_k^T U_i v_k)$$

denotes the predicted DWI intensity at the spatial location i, and I_0 is the modified ("cosh-like") Bessel function of the first kind of order zero. Referring to the discussion in [10], this data term is particularly suited in the context of maximum a posteriori (MAP) estimation with Rician noise of variance σ^2. We note that the data term (8) matches the form of (3) with a well-behaved differentiable function h given by

$$\begin{aligned} &h\left((\mathcal{A}U)_i^k, F_i^k\right) \\ &= -\log\left(\frac{F_i^k}{\sigma^2}\exp\left(-\frac{A_i^0\exp(-(\mathcal{A}U)_i^k)^2 + (F_i^k)^2}{2\sigma^2}\right)\right. \\ &\quad \left.\cdot\, I_0\left(\frac{A_i^0\exp(-(\mathcal{A}U)_i^k)^2 F_i^k}{\sigma^2}\right)\right), \end{aligned} \tag{9}$$

which can be differentiated analytically.

D. Contributions

In this paper, we propose an approach for simultaneous fitting and TV regularization of diffusion tensors which take the intrinsic geometry of Pos_3 into account. As discussed in Section II, tensor fitting and denoising are well-studied as separate tasks. To improve the reconstruction quality, combined approaches have been developed in vector space setups; see Section II-B. The novelties of this paper are as follows:

1) We propose a framework for simultaneous fitting and TV regularization of diffusion tensors in a (nonflat) Riemannian manifold setting. The proposed framework is flexible in the sense that it can be combined with different data terms and that it, with appropriate modifications, carries over to other manifolds as well. To our knowledge, this paper is the first one to address TV regularization for manifold valued data in an inverse problem setup.
2) We propose a generalized forward backward scheme for TV functionals of the form (1). The scheme combines an explicit and several implicit steps performed in a cyclic fashion on a decomposition of the functional. We apply this strategy to DTI and derive a concrete algorithm for the joint tensor fitting and regularization.

The choice of the name "generalized forward backward scheme" is motivated by Raguet *et al.* [34] who consider related algorithms in the vector space setting using this nomenclature. Concerning the realization of the backward steps, we employ efficient techniques based on geodesic averaging which we developed in [30].

E. Organization of the Article

In Section II we discuss closely related work. In Section III, we present our approach and derive the respective variational formulations in Section III-A. In order to solve the corresponding problems, we derive a generalized forward backward splitting in Section III-B, we provide the necessary building blocks in the Sections III-CIII-D – III-E, and we give a pseudo-algorithm in Section III-F. In Section IV, we run experiments on synthetic as well as on real 3D data. In Section V we discuss work related to the presented paper in a wider sense and point out ways to apply our framework in more advanced models beyond DTI. In Section VI we summarize the contributions of the paper and work out topics of future research.

II. Related Work

Most approaches to tensor fitting in DTI make use of the well-known Stejskal-Tanner equation

$$F_i^k = A_0 e^{-b\, v_k^T U_i v_k} \tag{10}$$

with constant $b > 0$ and unweighted measurement $A_0 > 0$, modeling the relation between the diffusion tensor U_i at position i and the DWIs F_i^k [33]. As explained in Section I-A, the actual data measured in DTI are the DWIs F_i^k, which capture the directional diffusivity in the direction v_k. We note that (10) is the simplified version of the Stejskal-Tanner equation in the sense that b consolidates several parameters, i.e., strength of the magnetic field pulse gradients, gyromagnetic ratio, etc. [9]. Typically, 6 to 30 diffusion weighted images F^k are acquired [7, 3. IV C]. A first approach to obtain the diffusion tensors consists of a (voxelwise) least squares approach based on the Stejskal-Tanner (10) using its logarithmic formulation as in (6). A central issue with naive least squares fitting is, however, that there is no notion of

spatial regularity enforcing that neighboring tensors, which belong to the same tissue type, are somewhat similar.

As the contribution of this paper is a variational approach for joint diffusion tensor estimation and total variation regularization, which takes spatial regularity into account, we focus on related aspects in the following. We first discuss non-combined approaches with emphasis on TV regularization in Section II-A. Then, we consider combined approaches for tensor fitting and regularization in Section II-B. At last, we discuss TV regularization for scalar and vector-valued data in the context of inverse problems (with a particular focus on forward-backward-type approaches) in Section II-C.

We postpone issues related to our work in a wider sense to Section V. In particular, we briefly discuss different underlying mathematical structures used in DTI in Section V-A1. Issues concerning (voxel-wise) tensor fitting are discussed in Section V-A2 later on. For a general introduction to DTI and related mathematical methods we moreover refer the interested reader to the articles of Assemlal *et al.* [9] and Lenglet *et al.* [35].

A. Non-Combined Approaches for Tensor Fitting and Regularization

One possibility of achieving spatial regularity of the reconstructed diffusion tensors is to denoise the DWIs prior to reconstruction. In this regard we mention the work of Basu *et al.* [13] who developed a method for removing Rician noise from DWIs. We also refer to the works of Bao *et al.* [14], and Luo *et al.* [15].

Another way of achieving spatial regularity is to reconstruct the tensors first and regularize the resulting tensor field afterwards. Early approaches for regularizing or denoising DTIs are descendants of the well-known Perona-Malik model [36]. One of the pioneering works in this field is the article of Tschumperlé *et al.* [16]. Their approach has been further generalized by Chefd'hotel *et al.* [17], [24] to regularizing flows for matrix-valued data, by Arsigny *et al.* [26] who translated this concept to the log-Euclidean case, and by Pennec *et al.* [19] who presented a framework for deriving regularizing partial differential equations in the context of DTI. A related approach based on local anisotropic filtering and the affine-invariant Riemannian metric was proposed by Castaño-Moraga *et al.* [21]. Another approach for DTI regularization has been proposed by Gur and Sochen [22], which is a differential geometric approach which considers the space of all possible DTI images as the space of sections of a fiber bundle.

Besides regularization strategies in the spirit of Perona and Malik [36], TV minimization is well-known for achieving edge preserving regularization. Generalizations of the TV model for manifold-valued data have gained a lot of interest in the last few years – both from theoretical viewpoint [37]–[39] and from an algorithmic one: Weinmann *et al.* [30], for instance, introduced algorithms for total variation minimization for manifold-valued data with a direct data term. They used a cyclic proximal point algorithm as well as a parallel proximal point algorithm to minimize TV functionals with ℓ^p-type data terms. These algorithms have been applied to DTI data. For the class of Cartan-Hadamard manifolds – which includes the data space in diffusion tensor imaging – convergence of these algorithms to a global minimizer was shown. Another algorithm suitable for TV minimization on Riemannian manifolds was proposed by Lellmann *et al.* [40]. This approach uses a reformulation of the original problem as a multi-label optimization problem and is closely related to approaches for circle-valued data by Strekalovkiy and Cremers [41], [42]. However, the number of labels grows rapidly with the dimension of the data space. Recently, Grohs and Sprecher [32] proposed an approach for TV minimization which is based on iteratively reweighted least squares. Moreover, [32] features convergence statements for the sphere.

B. Approaches for Combined Tensor Fitting and Regularization

Instead of two-stage approaches consisting of separate reconstruction and denoising, it is reasonable to combine these two tasks in one simultaneous task to improve the quality of the reconstructed tensor field. Typically, this results in more challenging problems. Some of the work discussed in Section II-A has been extended to this setting. For example, the constraint preserving (gradient) flows of Tschumperlé *et al.* [16] and Chefd'hotel *et al.* [24] served as a basis for the combined approach presented by Tschumperlé and Deriche in [28]. A similar approach employing a different regularizer and a projected gradient descent has also been proposed by Neji *et al.* [25]. Furthermore, the log-Euclidean framework, see [26] for instance, has been used as basis for a combined approach proposed by Fillard *et al.* [10].

Concerning the goal of combined tensor fitting and regularization as well as the choice of the employed data terms the presented approach has similarity to [10]. However, we emphasize that our method employs a different mathematical structure, i.e., the affine invariant Riemannian metric in contrast to a log-Euclidean structure; see also Section V-A1. This implies a significantly different setup in which the well known optimization strategies developed for linear spaces are not applicable anymore. We propose a generalized forward-backward splitting which deals with the non-smooth TV regularizer in the manifold setting. This also contrasts the method of [10], which employs a gradient-descent-based minimization scheme for a smooth regularizer. We consider the idea of the generalized forward-backward splitting scheme for TV regularization in a manifold setup as one of the main contributions of this article.

Another approach for combined fitting and regularization is described by Pasternak *et al.* [43], which is building on the works [27], [44], [45]. There, the product space of $\mathbb{R}^2$ and the space of positive matrices (or, synonymously, the trivial bundle over the base space $\mathbb{R}^2$ with characteristic fibre consisting of the positive matrices) is considered. For each target variable U, the graph of U is a section of this product space, which allows the authors to consider the metric on the graph induced by the product metric of the product space. Integrating the square-root of the determinant of the corresponding metric over the base space defines the employed regularizer. We note that not only one Riemannian metric is considered, but as induced metric from a product space, each target U defines its own Riemannian metric. We further note that the graph-energy employed in that

work typically yields non-zero values for constant grid functions U. The authors of [43] use gradient descent techniques to compute the flow of the functional they consider. The regularizer employed in the mentioned paper is significantly different from the TV regularizer we use here. As a simple example, a constant function always has TV energy zero. Furthermore, our regularizer is built on the affine-invariant Riemannian metric on the space of positive matrices without considering families of metrics induced on graphs. Moreover, our functional is non-differentiable. To address the non-differentiability of our regularizer, we employ proximal mappings for the corresponding computations contrasting the pure gradient descent techniques employed in [43] (for a different functional.)

C. TV Regularization for Inverse Problems in Linear Spaces

The first approach for TV regularization in image processing is the Rudin-Osher-Fatemi (ROF) functional [46] proposed for denoising scalar-valued data. Since then, many generalizations and further applications have been proposed. For a comprehensive overview on this topic we refer the interested reader to the introductory article of Chambolle *et al.* [47]. Some recent examples of such generalizations, which also consider an inverse setting, are [48], [49]. This means that the data term of the minimized functional can be written as $\|\mathcal{A}u - f\|_2^2$, where f is the measured data, $\mathcal{A}$ is a linear measurement operator, and u is the target variable.

Furthermore, total variation regularization in a vector-valued setup has recently gained a lot of attention. It is still an active topic of ongoing research; see for instance [50]–[52] and the references therein. The central property of TV functionals in the vector-valued setup is that they are well-behaved convex functionals. Hence, all the powerful methods of convex analysis, convex geometry, and convex optimization are available for their numerical treatment; see the book of Rockafellar [53], for instance. One approach for convex optimization in vector spaces is the so-called forward-backward splitting, which can be traced back to the early works of Peaceman and Rachford [54] and Douglas and Rachford [55], respectively. Based on their work, Eckstein and Bertsekas [56] (among others) then paved the way for modern splitting methods. The general idea of forward-backward splitting schemes is to split the target functional $\mathcal{F}$ into a sum $\mathcal{F} = \mathcal{F}_0 + \mathcal{F}_1$, where $\mathcal{F}_0$ is minimized by an explicit step (e.g., a gradient descent step), and where $\mathcal{F}_1$ is treated implicitly (e.g., using proximal mappings). Recently, a generalized forward-backward splitting algorithm has been considered in [34]. There, the underlying decomposition is

$$\mathcal{F} = \mathcal{F}_0 + \mathcal{F}_1 + \cdots + \mathcal{F}_N,$$

where $\mathcal{F}_0$ is treated explicitly and $\mathcal{F}_1, \ldots, \mathcal{F}_N$ are treated implicitly. More precisely, for each $\mathcal{F}_i, i = 1, \ldots, N$, the proximal mapping of $\mathcal{F}_i$ is computed. Finally, these results are averaged. As the presented approach is similar to this strategy we also consider it as a generalized forward-backward splitting.

III. The Proposed Generalized Forward-Backward Splitting Approach for Joint Tensor Fitting and TV Regularization in the Manifold Setting

In this section we develop the proposed model and the proposed generalized forward-backward algorithm which we implement for jointly fitting the tensors and regularizing them. After describing the proposed variational model in Section III-A, we explain how to split the considered functionals in Section III-B. Then we explain in detail how to compute the explicit steps in Section III-D as well as the implicit steps in Section III-E. As both of these steps rely on the exponential and inverse exponential map for Pos_3, which constitute the elementary operations on a manifold, we first discuss their computation in Section III-C.

A. Proposed Models for Combined Fitting and TV Denoising

We have already described the basic ideas and the univariate setup in the introduction. To be precise, and, in particular, to explicitly incorporate the important bivariate case, we provide a more detailed presentation. The respective quantities will be denoted by $(\cdot)_{ij}$. We stress that the proposed framework is not at all restricted with regard to the dimensionality. For instance, when we consider a 3D volume as domain, we only have to add an additional index to obtain $(\cdot)_{ijl}$, but we do not have to change the structure.

As we pointed out in Section II, the most common method for fitting a tensor to given DWIs is via the Stejskal-Tanner equation (10) by using a least-squares approach. This means that fitting the tensor U_{ij} at location (i, j) is achieved by minimizing the atomic data term

$$\mathcal{D}_{ij}(U_{ij}, F_{ij}) = \sum_k \left| b\, v_k^T U_{ij} v_k - \log\left(\frac{A_{ij}^0}{F_{ij}^k}\right)\right|^2, \quad (11)$$

with constants $b, A_{ij}^0 > 0$. As a consequence, we can write the two dimensional version of the data term in (6) as

$$\mathcal{D}(U, F) = \sum_{i,j} \mathcal{D}_{ij}(U_{ij}, F_{ij}). \quad (12)$$

From a statistical point of view, this data term assumes the logarithm of the DWI measurements to be corrupted by Gaussian noise. As we pointed out in [10], [13], a more reasonable noise model for DWIs is Rician noise. Rician noise causes a so-called shrinking effect which manifests itself in the reconstructed tensor being too small. An atomic data term which accounts for this noise model is given by

$$\hat{\mathcal{D}}_{ij}(U_{ij}, F_{ij}) = -\sum_k \log\left(\frac{F_{ij}^k}{\sigma^2} \exp\left(-\frac{P_{ij}^k(U_{ij})^2 + (F_{ij}^k)^2}{2\sigma^2}\right) I_0\left(\frac{P_{ij}^k(U_{ij}) F_{ij}^k}{\sigma^2}\right)\right), \quad (13)$$

where $P_{ij}^k(U_{ij}) = A_{ij}^0 \exp(-b v_k^T U_{ij} v_k)$; see [10], [57].

Concerning the TV regularizer we use a coordinate-wise discretization related to the Manhattan metric, i.e.,

$$\mathcal{TV}(U) = \sum_{i,j} d(U_{ij}, U_{(i,j)+e_1}) + d(U_{ij}, U_{(i,j)+e_2}),$$

where $e_1 = (1,0)$ and $e_2 = (0,1)$. As this discretization is known to produce so-called metrization artifacts, one can also employ a larger system of vectors describing the neighorhood. For example, one might use $e_1 = (1,0)$, $e_2 = (1,1)$, $e_3 = (0,1)$, $e_4 = (-1,1)$, in order get the more isotropic finite difference discretization

$$\mathcal{TV}(U) = \sum_l \sum_{i,j} \alpha_l d(U_{(i,j)+e_l}, U_{i,j}),$$

with weights $\alpha_1 = \alpha_3 = \sqrt{2}-1$ and $\alpha_1 = \alpha_3 = 1-\sqrt{2}/2$. We refer the interested reader to [58] for a more elaborate treatment of this topic or the more recent works [31], [59], which in particular consider finer discretizations. Using these approaches, anisotropy effects become practically invisible.

Combining the above expressions, we get the following compact notation for the proposed functionals

$$\mathcal{J}_\gamma(U,F) = \sum_{i,j} \mathcal{D}_{ij}(U_{ij}, F_{ij}) + \gamma \mathcal{TV}(U) \tag{14}$$

and

$$\hat{\mathcal{J}}_\gamma(U,F) = \sum_{i,j} \hat{\mathcal{D}}_{ij}(U_{ij}, F_{ij}) + \gamma \mathcal{TV}(U). \tag{15}$$

We note that both functionals (14) and (15) are *invariant* with respect to the parameter b which appears both in the Stejskal-Tanner equation and in the term P_{ij}^k. Saying this, we mean that solving (14) and (15) with respect to $b_1, b_2 > 0$ results in solutions which are related by $U_1 = (b_1/b_2)U_2$. To see the invariance property for the data term, we notice that the error is measured in the image space of the operator $\mathcal{A}$, and thus does not see a reparametrization of the preimage space. To see the invariance of the regularizing term, we notice that, on the space of positive matrices, we have affine invariance: this, in particular, implies invariance with regard to any way of scaling matrices by positive scalars, say b_1/b_2. We emphasize that this is a particular property of the chosen Riemannian metric which is not true for the Euclidean or the log-Euclidean metric.

B. The Proposed Splitting of the TV Functional

We consider the functionals (14) and (15). At first, we recall that $\mathcal{D}$ and $\hat{\mathcal{D}}$ can be split with respect to the corresponding atomic data terms $\mathcal{D}_{ij}$ and $\hat{\mathcal{D}}_{ij}$. More precisely, $\mathcal{D}(U,F) = \sum_{i,j} \mathcal{D}_{ij}(U_{ij}, F_{ij})$ and $\hat{\mathcal{D}}(U,F) = \sum_{i,j} \hat{\mathcal{D}}_{ij}(U_{ij}, F_{ij})$ (see (12) and (13)). Employing the notation

$$\mathcal{TV}(U) = \sum_{i,j} \mathcal{TV}_{ij}(U),$$

where

$$\mathcal{TV}_{ij}(U) = \sum_l \mathcal{TV}_{ij}^l(U) = \sum_l \alpha_l d(U_{(i,j)+e_l}, U_{i,j}), \tag{16}$$

$\mathcal{J}_\gamma$ can be written using the splitting

$$\mathcal{J}_\gamma = \mathcal{D} + \gamma \sum_{i,j,l} \mathcal{TV}_{ij}^l, \tag{17}$$

where the l corresponds to the directions of the employed differences. Analogously, we have

$$\hat{\mathcal{J}}_\gamma = \hat{\mathcal{D}} + \gamma \sum_{i,j,l} \mathcal{TV}_{ij}^l. \tag{18}$$

Here, we dropped the dependency on the arguments in order to simplify the notation. The core idea of the proposed algorithm is now to use this splitting in the following way: we treat all atomic regularization terms $\mathcal{TV}_{ij}^l$ implicitly while the differentiable data terms $\mathcal{D}$ and $\hat{\mathcal{D}}$ are treated explicitly. We observe that since these data terms are separable, the computation of the respective gradient results in computing the gradient of the corresponding atoms $\mathcal{D}_{ij}$ and $\hat{\mathcal{D}}_{ij}$. Based on this splitting, we develop our algorithms in the following.

C. Computing the Riemannian Exponential Map $\exp$ and Its Inverse $\exp^{-1}$

Let $D \in \mathrm{Pos}_3$ denote a positive matrix and $W \in T_D\mathrm{Pos}_3$ a tangent vector at D. The Riemannian exponential mapping $\exp_D$ maps W to the point on the manifold reached on the geodesic starting at D with velocity W after unit time. For the particularly nice space of positive matrices, explicit formulas are available; see, for example, [60]. We have

$$\exp_D(W) = D^{1/2} \exp(D^{-1/2} W D^{-1/2}) D^{1/2}, \tag{19}$$

where the mapping exp (without index) denotes the usual matrix exponential. Furthermore, there is also a closed form expression for the inverse of the Riemannian exponential mapping: for positive matrices $D, E \in \mathrm{Pos}_3$ it reads

$$\exp_D^{-1}(E) = D^{1/2} \log(D^{-1/2} E D^{-1/2}) D^{1/2}, \tag{20}$$

where the matrix logarithm log is well-defined since the argument is a positive matrix. After diagonalizing the argument $D^{-1/2}ED^{-1/2}$ the matrix logarithm is just the logarithm of the diagonal entries, which also applies to the matrix exponential. Hence, both matrix functions can be efficiently computed by diagonalizing the symmetric matrix under consideration and then applying the corresponding scalar functions to the eigenvalues. Finally, we need the distance between two elements D and E, which simply equals the length of the tangent vector $\exp_D^{-1}(E)$ and which explicitly reads

$$d(D,E) = \left(\sum_{l=1}^{3} \log(\kappa_l)^2 \right)^{-1/2}, \tag{21}$$

where κ_l is the l^{th} eigenvalue of the matrix $D^{-1/2}ED^{-1/2}$.

D. Computing the Explicit Gradient Descent Step for the Data Terms $\mathcal{D}$ and $\hat{\mathcal{D}}$ of the TV Functional

First, we note that Pos_n is an open subset in the linear space of symmetric matrices Sym_n. Hence, the notion of the differential in the manifold setting agrees with the differential in the (finite-dimensional) vector space setting. As noted in Section III-B, the separable structure of $\mathcal{D}$ and $\hat{\mathcal{D}}$ allows for computing differential quantities in a componentwise way, i.e., by considering

the atoms $\mathcal{D}_{ij}$ and $\hat{\mathcal{D}}_{ij}$. So let $V \in T_{U_{ij}}\mathrm{Pos}_3$ be an arbitrarily chosen element in the tangent space of Pos_3 at U_{ij}. Then the first variation of $\mathcal{D}_{ij}$ at U_{ij} in the direction V can be expressed via the Riemannian metric (5) as

$$\left.\frac{d}{dt}\right|_{t=0} \mathcal{D}_{ij}(U_{ij}+tV, F_{ij}) = g_{U_{ij}}(\nabla\mathcal{D}_{ij}, V), \qquad (22)$$

where $\nabla\mathcal{D}_{ij} \in T_{U_{ij}}\mathrm{Pos}_3$ is called the Riemannian gradient of $\mathcal{D}_{ij}$ at U_{ij}. Analogously,

$$\left.\frac{d}{dt}\right|_{t=0} \hat{\mathcal{D}}_{ij}(U_{ij}+tV, F_{ij}) = g_{U_{ij}}(\nabla\hat{\mathcal{D}}_{ij}, V), \qquad (23)$$

for the Riemannian gradient of $\hat{\mathcal{D}}_{ij}$ at U_{ij}.

In order to compute $\nabla\mathcal{D}_{ij}$, we note that, due to the symmetry of all diffusion tensors, there is a canonical isomorphism $\mathfrak{I}$ which maps any tensor W to the six entries of its upper triangular component, i.e.,

$$\mathfrak{I}(W) = (W^{(1,1)}, \ldots, W^{(1,3)}, W^{(2,2)}, \ldots, W^{(3,3)}). \qquad (24)$$

Using this isomorphism it is possible to reformulate the atomic data term $\mathcal{D}_{ij}$ in the form

$$\frac{1}{2}\|Au_{ij} - f_{ij}\|_2^2. \qquad (25)$$

Here u_{ij} is the vector in $\mathbb{R}^6$ with components from the upper triangle of U_{ij}, $A \in \mathbb{R}^{N\times 6}$ is determined by all N gradient vectors v_k (independent of i, j) and $f_{ij} \in \mathbb{R}^N$ is a vector with k-th entry given by $f_{ij}^k = \log(F_{ij}^k/A_0)$. The first variation of this least squares problem can now be represented by the standard scalar product of $\mathbb{R}^6$ as

$$\left.\frac{d}{dt}\right|_{t=0} \mathcal{D}_{ij}(u_{ij}+tv) = \langle m, v\rangle, \qquad (26)$$

where $m = A^T(Au_{ij} - f_{ij})$ for arbitrarily chosen $v \in \mathbb{R}^6$. Defining $M = \mathfrak{I}^{-1}(m)$ and $V = \mathfrak{I}^{-1}(v)$ we can rewrite this first variation as

$$\langle m, v\rangle = \mathrm{trace}(M^T V). \qquad (27)$$

In other words, M^T is the Euclidean gradient of $\mathcal{D}_{ij}$. Employing the Riesz representation theorem and using the definition of the Riemannian metric (5) we thus find

$$\mathrm{trace}\left(U_{ij}^{-1}\nabla\mathcal{D}_{ij}U_{ij}^{-1}V\right) = \mathrm{trace}(M^T V). \qquad (28)$$

As this equality holds true for all tangent vectors $V \in T_{U_{ij}}\mathrm{Pos}_3$, we find that

$$\nabla\mathcal{D}_{ij}(U_{ij}) = U_{ij}M^T U_{ij}. \qquad (29)$$

Proceeding similarly for $\hat{\mathcal{D}}_{ij}$, we find that

$$\nabla\hat{\mathcal{D}}_{ij} = U_{ij}\hat{M}^T U_{ij}, \qquad (30)$$

where

$$\hat{M}^T = \sum_k \frac{-bP_{ij}^k(U_{ij})}{\sigma^2}\left(P_{ij}^k(U_{ij}) - \frac{I_0'}{I_0}\left(\frac{P_{ij}^k(U_{ij})F_{ij}^k}{\sigma^2}\right)F_{ij}^k\right)v_k v_k^T \qquad (31)$$

Proposed generalized forward-backward scheme for a general functional $\mathcal{J}_\gamma = \mathcal{T} + \gamma\sum_{n=1}^N \mathcal{D}_n$.

Iterate w.r.t. m until convergence (step sizes $(\lambda_m)_m$):

1. Forward Step: Gradient Decent w.r.t. $\mathcal{T}$,
   $u^{(m,0)} = \exp_{u^{(m-1)}}(-\lambda_m \mathrm{grad}\ \mathcal{T}(u^{(m-1)}))$.
2. Backward Step:

   For $n = 1, \ldots, N$

   Proximal Mappings w.r.t. $\mathcal{D}_n$ (cyclic way):
   $u^{(m,n)} = \mathrm{prox}_{\lambda_m\gamma\mathcal{D}_n}(u^{(m,n-1)})$.

   Alternative:

   Proximal Mappings w.r.t. $\mathcal{D}_n$ (parallel way):
   $u^{(m,n)} = \mathrm{prox}_{\lambda_m\gamma\mathcal{D}_n}(u^{(m,0)})$.

   $u^{(m)} = \mathrm{mean}_n(u^{(m,n)})$.

Fig. 2. Concept of the Generalized Forward-Backward Splitting Algorithm (With Cyclic and, Alternatively, Parallel Backward Step) for Manifold Valued Data.

and I_0'/I_0 denotes the logarithmic derivative of the Bessel function I_0. Here, the P_{ij}^k are given as in (8) and $v_k v_k^T$ is the rank one matrix obtained from the direction v_k. Once the gradients are computed, we can perform a gradient descent step

$$U_{ij}^{m+1} = \exp_{U_{ij}^m}(-\lambda_m \nabla\mathcal{D}(U_{ij}^m)), \qquad (32)$$

using (29) (and (30) for $\nabla\hat{\mathcal{D}}$, respectively) where λ_m denotes the step size chosen at the m-th iteration step.

E. Computing the Implicit or Proximal Mapping Steps for the Atomic Regularizing Terms $\mathcal{TV}_{ij}^l$ of the TV Functional

We treat all regularization terms $\mathcal{TV}_{ij}^l(U)$ given by (16) implicitly, i.e., by proximal mappings. We do so in order to deal with the non-differentiability of these terms. In this context, we mention that applying related explicit schemes such as subgradient descent typically show oscillatory behavior. We do not observe such behavior in our implicit approach. We consider the atomic regularization terms $\mathcal{TV}_{ij}^l(U)$ which were given by

$$\mathcal{TV}_{ij}^l(U) = \alpha_l d(U_{(i,j)+e_l}, U_{i,j}).$$

Taking a closer look at these terms, we note that they basically take two positive matrices as arguments and map them to the distance between them induced by the Riemannian metric. More precisely, they are mappings from the product manifold $\mathrm{Pos}_3 \times \mathrm{Pos}_3$ to the nonnegative reals given by $(P, Q) \mapsto d(P, Q)$. Thus, we need to compute proximal mappings of the form

$$\mathrm{prox}_{\lambda d}(P, Q) = \mathrm{argmin}_{P',Q'}\lambda d(P', Q') + \frac{1}{2}d^2(Q, Q') + \frac{1}{2}d^2(P, P'),$$

where $\lambda > 0$. These proximal mappings have closed form expressions, i.e., the minimizers can be given by explicit formulas.

For their derivation we refer to the authors' paper [30]. Considering the geodesic c starting at a point $c(0) = P \in \mathrm{Pos}_3$ and ending at another point $c(1) = Q \in \mathrm{Pos}_3$ we denote the point reached after time t (for constant speed parametrization) by $[P, Q]_t$. Using this notation, the explicit formula for the proximal mappings reads

$$\begin{aligned}\mathrm{prox}_{\lambda d}(P, Q) &= \left([P, Q]_{\min(\lambda, 1/2)}, [Q, P]_{\min(\lambda, 1/2)}\right) \\ &= \left(\exp_P \left(\min\left(\lambda, \frac{1}{2}\right) \exp_P^{-1} Q \right), \right. \\ &\qquad \left. \exp_Q \left(\min\left(\lambda, \frac{1}{2}\right) \exp_Q^{-1} P \right) \right) \end{aligned} \tag{33}$$

which can be computed using (19) and (20).

At this point we make the following remarks. First, the intuition behind these proximal mappings is that we solve the initial problem locally while penalizing the deviation from the previous iterate, where the parameter $\lambda > 0$ controls how much we penalize the deviation. This is obviously of implicit nature since the conditions are imposed on the sought solution. Second, the strength of the proposed method is that the proximal mappings for the atomic regularization terms can be computed explicitly. In contrast to this, we strongly believe that there is no explicit formula for the proximal mapping of the TV regularizer as a whole. Certainly, at the present time, there is no such formula available in the literature. Furthermore, no analogues of the well-known dual or primal-dual formulations are presently available in the manifold setting. Finally, we note that the proximal mappings are possibly multivalued in the general manifold-setting since the uniqueness of the minimizers cannot always be guaranteed. However, in the scenario we consider, this is a rather mathematical pathology which we did not observe in practice. In fact, such situations only occur on negligible sets of data; we refer to the discussion in [30].

F. Proposed Generalized Forward-Backward Algorithm

As we have discussed all building blocks, we can now explain how to combine them within a generalized forward-backward algorithm for manifold-valued data; we refer to Fig. 2 where the concept of the algorithm is presented.

Starting from the decomposition (17) of $\mathcal{J}_\gamma$, we first apply the gradient descent for the data term $\mathcal{D}$; this results in performing the gradient steps (32) for all atoms of the data term. Then, we proceed with the backward step, where we apply the proximal mappings for the atomic terms $\mathcal{TV}_{ij}^l$ in a cyclic fashion. More formally, we impose a linear ordering on the three indices i, j, l, i.e., we consider a mapping $(i, j, l) \mapsto n \in \{1, \ldots, N\}$. We denote the inverse of this mapping by $n \mapsto (i_n, j_n, l_n)$ and so obtain a series of intermediate iterates

$$U^{(m,n+1)} = \mathrm{prox}_{\lambda_m \gamma \mathcal{TV}_{i_n,j_n}^{l_n}}(U^{(m,n)}) \tag{34}$$

at the "macro"-step m. We point out that we use the notation $U^{(m,0)}$ for the outcome of the previous gradient descent step. This approach amounts to the cyclic variant in Fig. 2. It is important to note that the ordering of the indices i, j, and l in this cyclic variant provides a considerable degree of freedom which can effectively be used to exploit peculiarities of the computing architecture, for example, for partially parallel implementations in a GPU, or the employed storage scheme. Instantiating the gradient of the data term as described in Section III-D and the atoms of the TV term as in Section III-E, we obtain generalized forward-backward algorithms for the joint fitting and regularization of diffusion tensor fields; see Algorithm 1.

Algorithm 1: Proposed generalized forward backward algorithm (with cyclic backward step) for the TV problems (17) and (18) for joint diffusion tensor fitting and denoising.

input : DWI images $\mathbf{F}^k = (F_{ij}^k)$, parameter γ, number of steps M
output: Signal $\mathbf{U}$ (solution of (14) and (15), resp.)

```
Initialize U;
for m ← 1 to S do
    //compute stepsize;
    λ_m ← CompLambda(m)
    //Riemannian gradient step
    for i, j ← 1 to I, J do
        // Riem. grad. of (11) or (13)
        G_ij ← CompGradient(U_ij,F_ij);
        //adapt step length
        λ_m ←
        min(searchLine(G_ij,U_ij,F_ij),λ_m);
    end
    for i, j ← 1 to I, J do
        //update according to (32)
        U_{i,j} ← exp_{U_{i,j}}(−λ_m G_{i,j});
    end
    //proximal mappings of the atomic
    regularizers (w.r.t. numbering
    n = 1, .., N)
    for n ← 1 to N do
        // determine t
        t ← min(λ_m γ α_{l_n}, 1/2);
        // cf. definition of [·,·]_t in (33)
        Ū_{i_n,j_n} ← [U_{i_n,j_n}, U_{(i_n,j_n)+v_{l_n}}]_t;
        Ū_{(i_n,j_n)+v_{l_n}} ← [U_{(i_n,j_n)+v_{l_n}}, U_{i_n,j_n}]_t;
        // update
        U_{i_n,j_n} ← Ū_{i_n,j_n};
        U_{(i_n,j_n)+v_{l_n}} ← Ū_{(i_n,j_n)+v_{l_n}};
    end
end
```

Besides being suited to Riemannian data, the difference of our algorithm to the generalized forward-backward splitting of Raguet *et al.* [34] is that we apply the proximal mappings in a cyclic fashion whereas Raguet *et al.* apply them in a parallel fashion. Such a parallelization, however, requires the outputs of the individual proximal mappings to be averaged. This can be done in a Riemannian manifold using the well-known concept of intrinsic means which are frequently called Riemannian

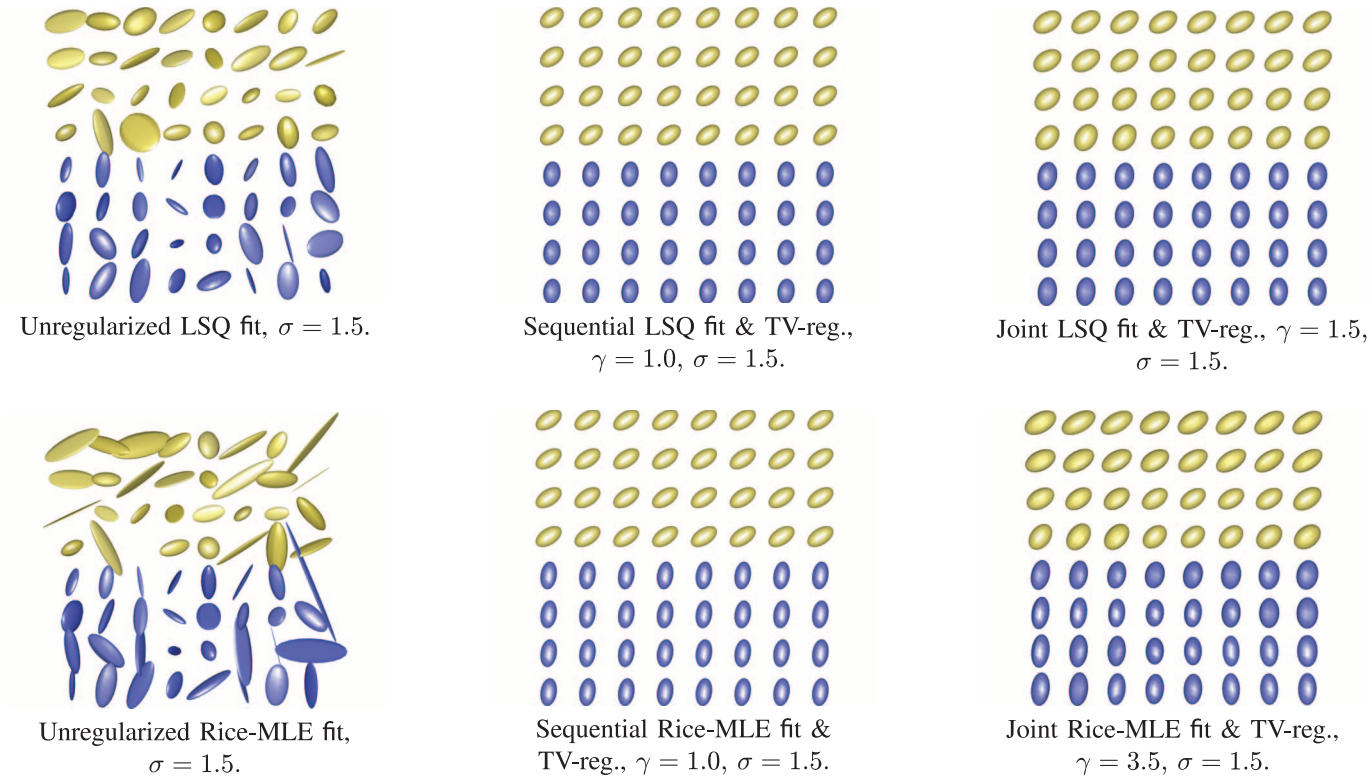


Fig. 3. Synthetic Experiments (Examples). The unregularized tensor fitting is heavily affected by the noise which makes regularization necessary. All TV based regularization methods yield edge-preserving regularizations. We clearly observe an undesired shrinking effect for the schemes based on the LSQ data term.

centers of mass, Karcher and Frechet means. For a related discussion in the context of proximal point algorithms (including faster approximative variants) we refer to [30]. Using the notation mean for the intrinsic mean, the parallel variant of our generalized forward-backward algorithm reads

$$U^m = \text{mean}_n \left(\text{prox}_{\lambda_m \gamma \mathcal{T} \mathcal{V}^{l_n}_{i_n, j_n}} (U^{(m,0)}) \right), \qquad (35)$$

where U^m is the iterate after m complete forward-backward steps. We note that (35) comes with the additional cost of computing the mean.

IV. Experiments

We evaluate the proposed method on synthetic data, see Section IV-A, as well as real data from the UCL Camino Diffusion MRI Toolkit [61], see Section IV-B. The algorithm is implemented in C++. The operations which are necessary for computing matrix roots, logarithms, and exponentials have been implemented using Eigen v.3.2.4[1]. The parameter λ_s in Algorithm 1 was chosen as $\lambda_m = Cm^{-1/2}$, with $C = 100$. For the visualization, we used a modified version of the fanDTasia ToolBox by Angelos Barmpoutis [62].

A. Experiments on Synthetic Data

We perform a quantitative evaluation using synthetic test data. We generated synthetic data similar to the one in Fillard *et al.* [10]: we use the diffusion tensors

$$\begin{pmatrix} 0.970 & 0.0 & 0.0 \\ 0.0 & 1.751 & 0.0 \\ 0.0 & 0.0 & 0.842 \end{pmatrix} \text{ and } \begin{pmatrix} 1.556 & 0.338 & 0.0 \\ 0.338 & 1.165 & 0.0 \\ 0.0 & 0.0 & 0.842 \end{pmatrix}$$

[1]Available at http://eigen.tuxfamily.org.

to generate a volume of size $16 \times 16 \times 16$. This ground truth volume has two "phases" of tensors, i.e., one half of the volume consists of copies of the first tensor and one half consists of copies of the second tensor; see Fig. 3. Next, we use the Stejskal-Tanner equation to create ten ground truth DWIs F^k_{GT}, $k = 1, \ldots, 10$. Here, we have chosen $A^0 = 10$ for all tensors as well as $b = 1$, and the ten gradient directions

$$\begin{aligned} &v_1 = (1,0,0)^T, v_2 = (0.267, -0.535, 0.802)^T, \\ &v_3 = (0.667, 0.333, -0.667)^T, v_4 = (0, -1, 0)^T, \\ &v_5 = (0,0,1)^T, v_6 = (0.743, -0.557, -0.371)^T, \\ &v_7 = (-0.577, -0.577, -0.577)^T, v_8 = (0.707, 0.707, 0)^T, \\ &v_9 = (0, -0.894, 0.447)^T, v_{10} = (-0.801, -0.267, 0.535)^T. \end{aligned}$$

Then we impose Rician noise on the produced DWIs. This means that we replace the true intensity $F^k_{GT,ij}$ of the k-th DWI at location (i, j) by

$$F^k_{N,ij} = \sqrt{(F^k_{GT,ij} + X)^2 + Y^2}, \qquad (36)$$

where real and imaginary parts $X, Y \sim N(0, \sigma^2)$ are Gaussian distributed. We use three different noise levels $\sigma = 0.5$, $\sigma = 1.0$, and $\sigma = 1.5$.

We start out by reconstructing the tensor field using the developed joint reconstruction and TV regularization methods. In Table I, we employ the least squares (LSQ) data term $\mathcal{D}$ as well as the data term $\hat{\mathcal{D}}$ based on the maximum likelihood estimator (Rice-MLE) particularly suited to Rician noise. We compare these methods under different noise levels, i.e., $\sigma = 0.5$, $\sigma = 1.0$, and $\sigma = 1.5$, and under different choices of the regularization parameter, i.e., $\gamma = 0$, $\gamma = 1$, $\gamma = 2$, $\gamma = 5$, and γ

TABLE I
LSQ-BASED VS. RICE-MLE-BASED ESTIMATION

		Joint LSQ fit & TV-reg.			Joint Rice-MLE fit & TV-reg.		
γ	σ	MSE	ΔSNR	trace	MSE	ΔSNR	trace
0	0.50	0.002	1.32	95 %	0.002	1.94	96 %
0	1.00	0.010	1.18	94 %	0.009	2.93	99 %
0	1.50	0.016	1.08	93 %	0.010	2.02	108 %
1	0.50	0.000	3.88	96 %	0.000	3.59	96 %
1	1.00	0.000	3.52	94 %	0.000	7.44	98 %
1	1.50	0.000	3.24	93 %	0.000	8.99	105 %
2	0.50	0.000	2.52	96 %	0.000	4.75	96 %
2	1.00	0.001	2.93	94 %	0.000	7.50	98 %
2	1.50	0.002	2.81	93 %	0.000	9.34	105 %
5	0.50	0.002	1.13	96 %	0.000	5.24	97 %
5	1.00	0.003	2.30	94 %	0.000	7.19	97 %
5	1.50	0.003	2.35	93 %	0.001	8.77	104 %
10	0.50	0.002	1.13	96 %	0.000	5.04	97 %
10	1.00	0.003	2.30	94 %	0.001	6.29	97 %
10	1.50	0.003	2.35	93 %	0.002	7.89	104 %

We use the least squares (LSQ) data term as well as the data term based on the maximum likelihood estimator (Rice-MLE) particularly suited to Rician noise. The Rice-MLE-based estimation performs better w.r.t. tensor reconstruction (measured by MSE) as well as in terms of the DWIs derived from the reconstructed tensors (measures by SNR). Furthermore, the Rice-MLE-based estimation also much better preserves the trace value.

$= 10$. The quantitative evaluation in Table I uses the following means of comparison:

1) MSE on the DTIs: we compute the mean squared error between the reconstructed tensors and the ground truth tensors using the Riemannian metric (5).
2) ΔSNR on the DWIs: we compute the signal-to-noise ratio improvement ΔSNR, see [66]. It is given by

$$\Delta\text{SNR} = 10 \log_{10} \left(\frac{\sum_{ijk} \left| F^k_{GT,ij} - F^k_{N,ij} \right|^2}{\sum_{ijk} \left| F^k_{GT,ij} - F^k_{R,ij} \right|^2} \right), \tag{37}$$

where $F^k_{R,ij}$ denotes the k-th DWI reconstructed from the fitted tensors at position (i, j).
3) trace: we compute the average percentage of the traces of the fitted tensors U_{ij} in comparison to the ground truth tensors $U_{GT,ij}$, i.e.,

$$100 \cdot \text{mean} \left(\frac{\text{trace } U_{ij}}{\text{trace } U_{GT,ij}} \right). \tag{38}$$

We observe in Table I that the joint estimation based on the Rice-MLE data term performs better with respect to MSE as well as with respect to ΔSNR. Furthermore, the Rice-MLE-based combined fitting and TV regularization preserves the trace much better.

Focusing on the Rice-MLE based data term, we compare the proposed combined method with the corresponding (uncombined) sequential baseline method which works as follows: we first fit the tensors using the data term $\hat{\mathcal{D}}$ without any regularization. Then, in a subsequent step, we perform TV-regularization on the fitted tensors using the method described in [30]. For the experiment in Table II, we consider different noise levels, i.e., $\sigma = 0.5$, $\sigma = 1.0$, $\sigma = 1.5$, and $\sigma = 2.0$. For each noise level, we consider the regularizers $\gamma = 0, \gamma = 0.2, \ldots, \gamma = 11$ and compute the MSE and ΔSNR for both the proposed combined fit and TV regularization and its sequential counterpart. In Table II, the best value of the respective quality measure, i.e., MSE and ΔSNR, for the respective scheme, i.e., combined and sequential fit and TV regularization, is shown. The approaches which combine diffusion tensor fitting with TV denoising perform better with respect to mean squared error (MSE) on the DTIs as well as with respect to ΔSNR on the DWIs. Hence, the combined approach yields a measurable gain.

TABLE II
SEQUENTIAL VS. JOINT RICE-MLE-BASED ESTIMATION

Sequential Rice-MLE fit & TV-reg.				
σ	0.5	1.0	1.5	2.0
MSE	0.000	0.000	0.001	0.024
ΔSNR	9.64	8.74	6.81	4.28

Joint Rice-MLE fit & TV-reg. (proposed method)				
σ	0.5	1.0	1.5	2.0
MSE	0.000	0.000	0.001	0.002
ΔSNR	10.40	10.25	10.13	10.10

For the data term based on the maximum likelihood estimator for Rician noise (Rice-MLE), we compare the sequential approach of first fitting the tensors and then denoising with our proposed joint approach. The joint approach is better w.r.t. MSE (measured by the DTIs directly) for higher noise level; for lower noise level, both approaches yield very small MSE. With respect to SNR (measured on the DWIs derived from the reconstructed tensors), the proposed combined approach performs better throughout.

For illustration, we plotted some examples in Fig. 3. For completeness, we here also include the sequential method of first tensor fitting using the data term $\mathcal{D}$, corresponding to a least squares fit, followed by TV denoising. As this approach showed no superior performance than the joint least squares fitting and TV regularization, we omitted the detailed quantitative comparison in the paper. Fig. 3 in particular illustrates the shrinking effect when using the LSQ data term for higher levels of Rician noise.

In our synthetic experiments, we make the following observations: (i) The combined fitting and TV denoising approach using the Rice-MLE data term $\hat{\mathcal{D}}$ yields better results than the combined LSQ-based approach. (ii) The combined fitting and TV denoising approach using the Rice-MLE data term $\hat{\mathcal{D}}$ yields better results than the corresponding baseline approach which first fits the tensors using $\hat{\mathcal{D}}$ and subsequently performs TV regularization. (iii) The data term $\hat{\mathcal{D}}$ prevents the shrinking effect [10] caused by Rician noise. Hence, the proposed method of joint fitting and TV regularization using the data term $\hat{\mathcal{D}}$ shows the best performance in reconstructing the ground truth in our synthetic experiments which indicates the potential of the proposed method.

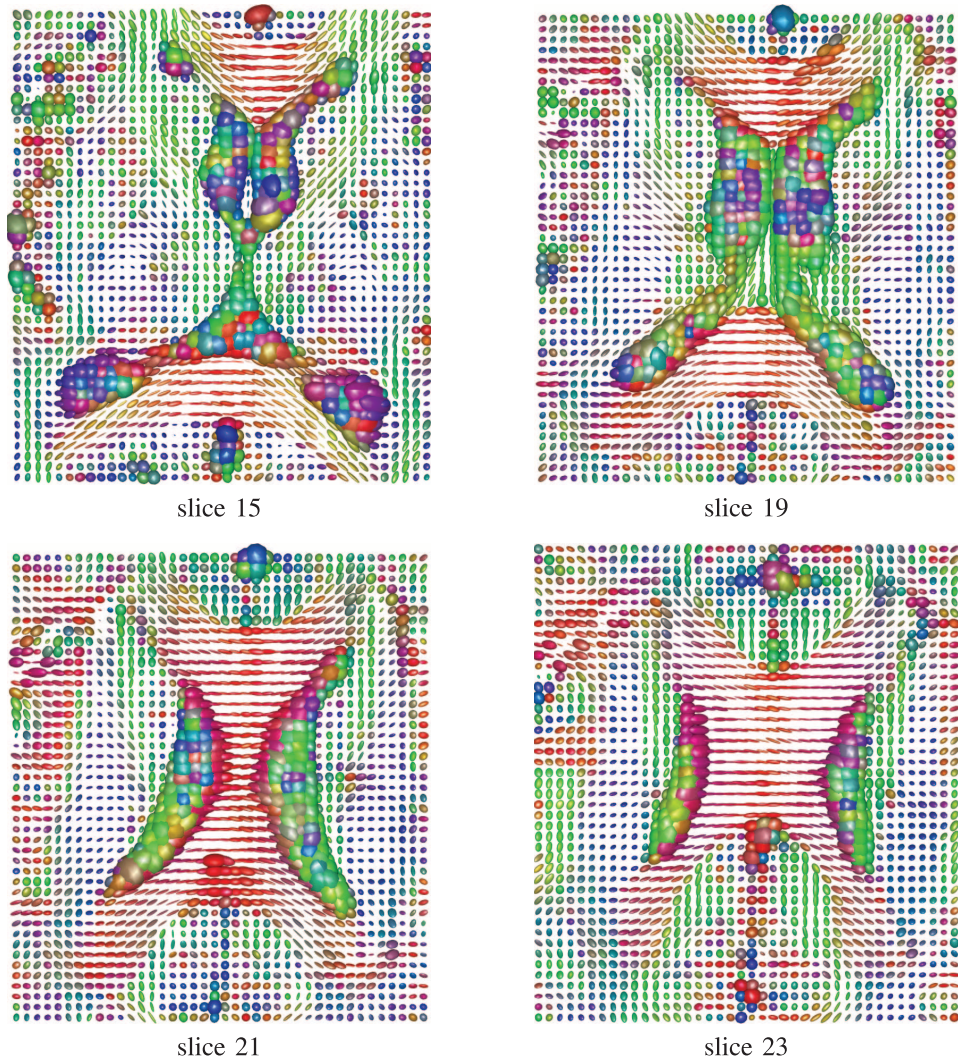


Fig. 4. Joint Fitting and TV Regularization (Proposed Approach) on Real Data. We display four axial sections of the 3D volume reconstructed from the Camino dataset using the proposed approach with the Rice-MLE data term. All tensors were color-coded according to the principle direction: the color red encodes left-right principal orientation, green corresponds to anterior-posterior, and blue corresponds to superior-inferior. We observe that the proposed approach constructs a regular tensor volume and that it preserves sharp boundaries between oriented structures at the same time.

B. Experiments on Real Data

In order to demonstrate the potential of the proposed method on real data, we applied it to diffusion weighted data of the human brain from the Camino project [61]. This data is freely available. Thus it is particularly suitable as a test data set since it allows researchers to compare their algorithms on the same data set. We reconstructed the tensors in three dimensions with the proposed approach using the joint tensor fitting and TV-regularization using the Rice-MLE data term, where we performed 1000 steps ($S = 1000$). We estimated a Rician noise level of $\sigma_r = 16777.5$ using the method proposed by Koay and Basser in [64]. Since the maximum DWI magnitude is of order 10^5, and the magnitude of the unweighted measurements is of order 10^6 the noise level can be considered as low to moderate. Based on the experiences gained during the synthetic experiments, shown in Table I we considered different results with γ being chosen in the range $0.5, 0.6 \ldots, 2.0$ and found $\gamma = 1.0$ to be a reasonable choice for the regularization parameter. The results are shown in Fig. 4 where we show four axial sections of the reconstructed 3D diffusion tensor volume; see also Fig. 1 for a 3D impression.

For comparison, we computed the DTI volumes using the Camino software; see Fig. 5. We furthermore compare with the sequential approach of plain fitting using the Rice-MLE data term followed by regularizing by the TV method proposed in [30]. Sections of the resulting volume are displayed in Fig. 6. We determined the regularization parameter $\gamma = 0.2$ for the method of [30] empirically (by exploring results for γ in the range $0.1, 0.2, \ldots, 1.0$). Comparing Figs. 4 and 5, we observe that the tensors obtained by the Camino software are slightly smaller than the ones obtained by our method. This effect might be explained by using a least squares fit on data corrupted by Rician noise. We avoid this effect by using the Rician MLE data term. Comparing Figs. 4 and 6, we notice that the proposed method performs considerably better than the sequential approach. In particular, the sequential approach exhibits severe outliers which are not present in the combined approach. We remark that we used twice the number of iterations for the regularization step of the sequential approach, than for the combined

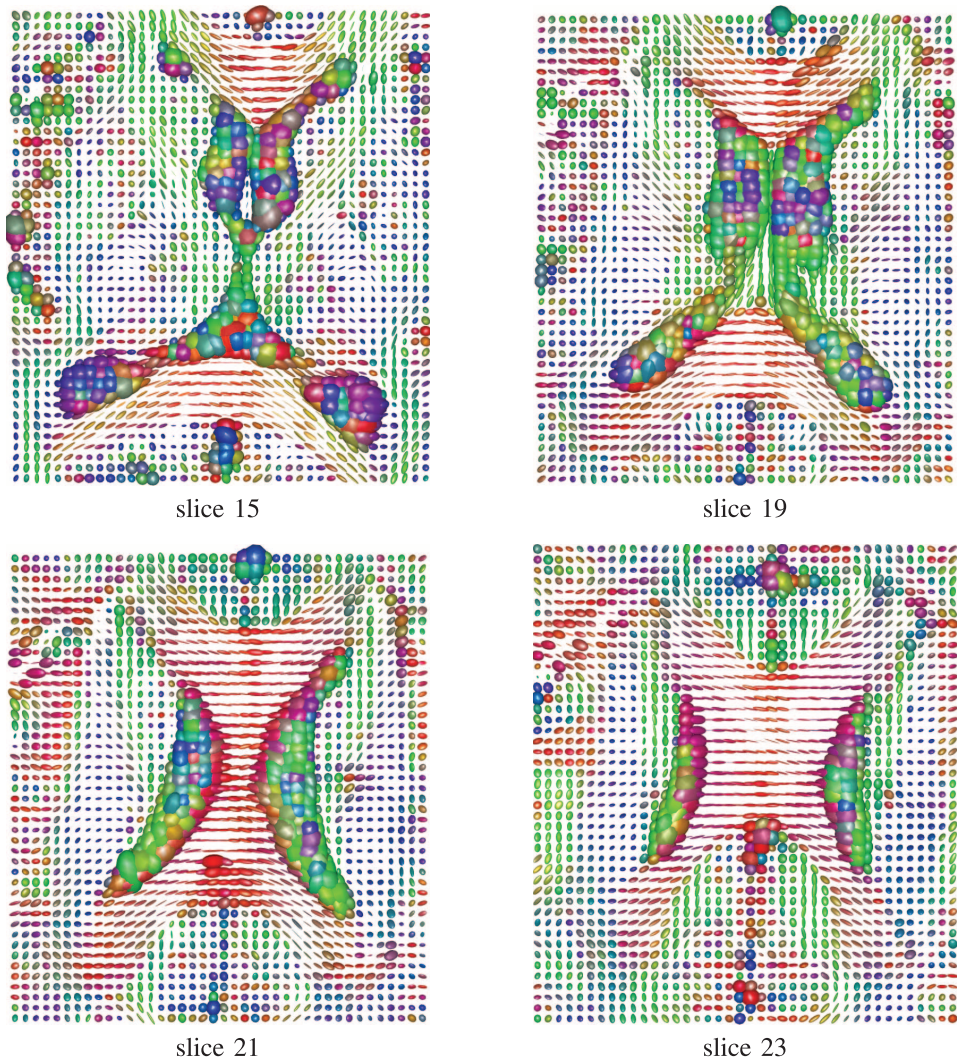


Fig. 5. Reconstruction using the Camino Software. To provide the reader with a reference for comparison, we show results obtained by the Camino software (http://cmic.cs.ucl.ac.uk/camino/) on the same dataset as used for the joint approach (Fig. 4). We note that the reconstructed tensors are slightly smaller than the tensors obtained by our approach in Fig. 4. This might be explained by the effects of least squares fitting in the presence of Rician noise which we avoid using the Rician MLE data term.

method (where we used $S = 1000$ iterations.) This was done to ensure that, in the experiment, the computational costs for the proposed method do not exceed those of the sequential approach (actually only the regularization step). This illustrates the gain of the proposed method when bounding runtime by that of the baseline approach.

Next, we compare the above methods on the DWI level. This means that we show the DWIs induced by the corresponding tensors fields; see Fig. 7. First, we observe that all constructed DWIs contain less noise than the actually measured DWIs. Furthermore, the sequential method yields DWIs with more homogeneous regions than the Camino software and the proposed method again yields results with even more homogeneous regions than the sequential method. We emphasize that the erroneously large tensors in Fig. 6 correspond to the unnaturally dark pixels in the corpus callosum in the fourth row of Fig. 7.

Finally, in order to indicate the relevance of the proposed method in a practical application, we conducted a qualitative comparison regarding the suitability of the proposed reconstruction technique for fiber tracking. The fiber tracking has been conducted with 3D Slicer (www.slicer.org). For all experiments we used the following settings: 1.0 seed spacing (in voxel), 0.3 min seed FA, 0.3 stopping min FA, 0.1 mm integration step, 0 mm minimum path length, 2000 mm maximum path length. Panel (a) in Fig. 8 shows the results of a weighted least squares reconstruction. Panel (b) in Fig. 8 shows the fiber tracking results obtained with the UKFT-module (Slicer plugin) which implements the method proposed by Baumgartner *et al.* [65]. Panel (c) shows the result obtained with the proposed reconstruction algorithm. Note the improved reconstruction of the (vertical) fibers corresponding to the cingulum (indicated by white arrows).

V. Discussion

The purpose of this section is two-fold. First, we discuss previous work related to the present paper in the wider sense in Section V-A complementing the discussion in Section II. In particular, we here review different mathematical structures which have been used in DTI in Section V-A1. We address this to motivate the use of the affine-invariant Riemannian metric which

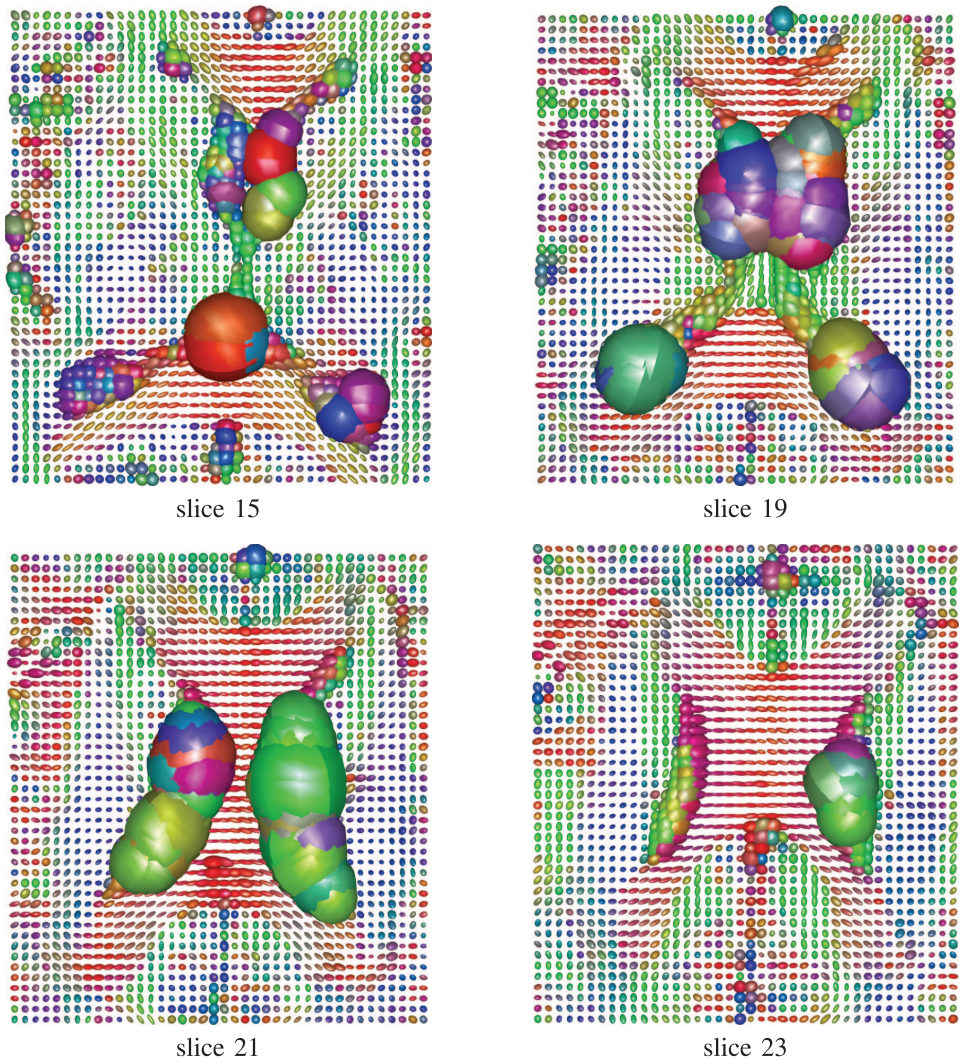


Fig. 6. Sequential Fitting and TV Regularization. We here visualize the result of first reconstructing the tensors individually using the Rice-MLE data term followed by TV regularization using the ℓ_1-TV approach for manifold valued data proposed in [30]. We use the same 3D dataset as used for the joint approach and we display the same slices (Fig. 4). We observe several very large tensors which are most probably caused by erroneous fitting. This assumption is supported by the fact that the mean diffusivity of these tensors is more than two times higher than the one of water at 35 degree Celsius, i.e., $2.895 \cdot 10^{-9}\ \mathrm{m^2 s^{-1}}$, as reported in [63]. Note that no outlier rejection was performed in this example in order to show that a posteriori regularization is not sufficient for removing such outliers.

is basic for the proposed algorithmic scheme. In Section V-A2, we discuss further work on plain voxel-wise tensor fitting. The second major purpose of this section (Section V-B) is to discuss the limitations of DTI and to point out the applicability of the proposed method to more recent imaging setups such as Q-ball imaging which overcome certain limitations of DTI.

A. Discussion of Work Related in a Wider Sense

1) Different Mathematical Structures for DTI: As we pointed out in Section I-A, it is reasonable – in the context of DTI – to endow Pos_3 with the Riemannian metric in (5). This has been done for various tasks in DTI, such as by Pennec *et al.* [19] who presented a computational framework for interpolation, basic statistics, diffusion filtering, etc. Another work in this regard is the one of Fletcher [67] for performing statistical computations, such as principal component analysis on for instance. A Riemannian setting has also been employed in the context of segmentation by Cheng *et al.* [68] who implemented a Chan-Vese model for DTI data based on the previous works of Wang and Vemuri [69], [70] which are based on information theoretic (but non-Riemannian) distance measures. We emphasize that using the Riemannian metric (5) makes the powerful machinery of manifolds available for which a lot of theory and methodology has been developed. Examples include, but are not limited to, wavelet-type multiscale transforms [71]–[73], manifold-valued partial differential equations [74], and statistics on Riemannian manifolds [60], [75]–[79].

Instead of endowing Pos_3 with the Riemannian metric (5) and considering it as a manifold, there are also approaches which take a different point of view and impose different underlying mathematical structures. One approach is to consider the positive matrices as the positive cone in the space of matrices and to equip it with the corresponding Euclidean distance. This approach has been taken in a lot of related work; as examples we refer to [80], [81] on median and related filtering, to [82] on segmentation and smoothing of DT-MRI fields, and also to [83] and the references therein. Methods based on this Euclidean concept typically have to

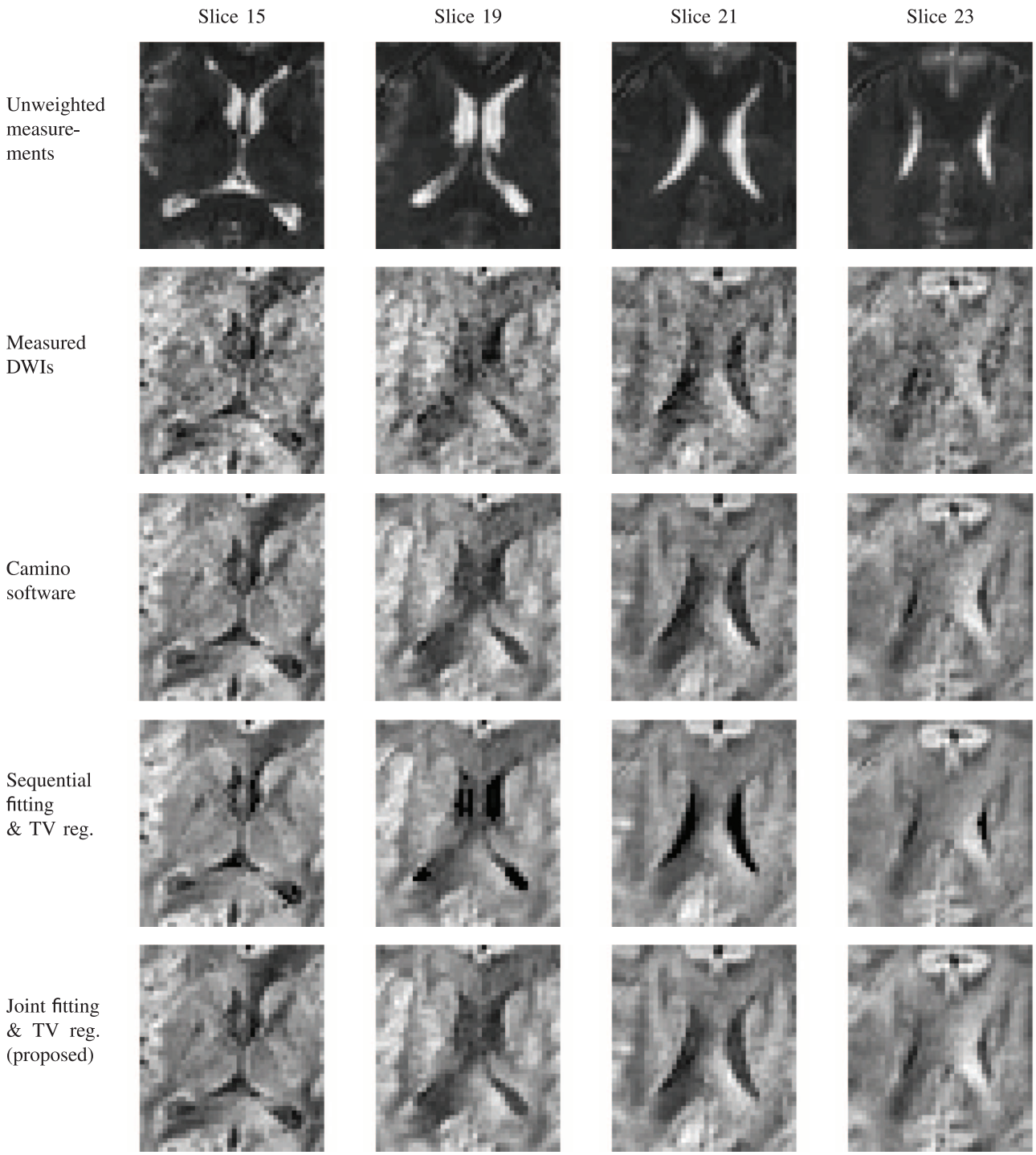
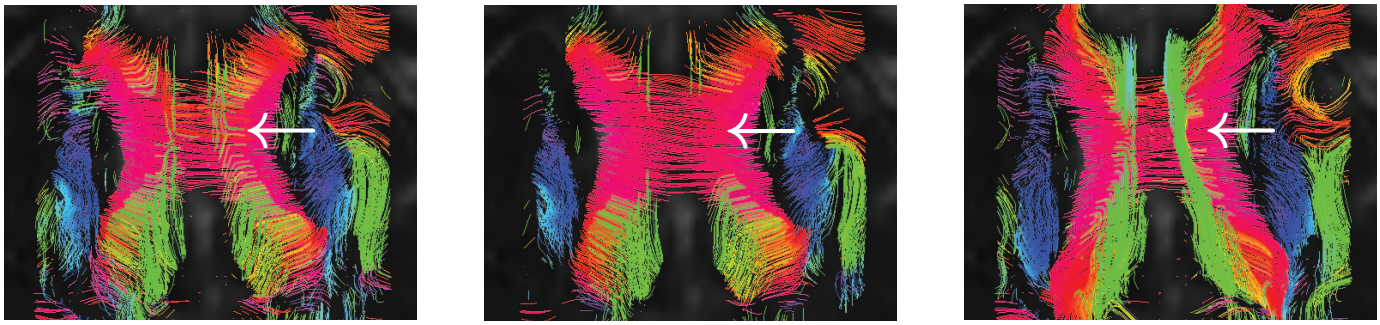


Fig. 7. Regularization Effect on the Reconstructed DWIs. We display DWIs constructed from the tensors computed with the Camino software (see Fig. 5), with the sequential fitting and TV regularization (see Fig. 6) and with the proposed joint fitting and TV regularization approach (see Fig. 4). We further display the actually measured DWIs (second row) and the unweighted acquisitions (first row). Visually, we observe an increasing regularity of the DWIs from top to bottom; the average Rician noise level σ_r estimated using the method Koay and Basser [64] equals $1.7 \cdot 10^4$ for the original DWIs, $1.6 \cdot 10^4$ in case of the Camino software, and $\sigma_r = 1.5 \cdot 10^4$ for both the sequential and the joint fitting. We noted that the sequential fitting produces erroneous tensors which are indicated by the unnaturally dark areas in the corpus callosum; see also Fig. 6.

ensure that the computations done in the ambient space do not leave the cone of positive matrices. This is usually achieved by projection. We note, however, that projections are a somewhat problematic concept in this context, since the positive matrices form an open set which means that they have no clearly defined boundary onto which one could project. Also, using the Euclidean metric for DTI data, a so-called swelling effect has been reported in [16], [18], [24]. This means that the dispersion of the corresponding covariance matrices, i.e., the determinant of the corresponding matrices, tends to be larger than the original ones [18] when reconstructing noisy data with known ground truth. Also in the context of TV regularization in the Euclidean (non-manifold) setup, a slight swelling can be observed [30].



(a) (b) (c)

Fig. 8. Impact on Fiber Tracking. Panel (a) shows the fiber tracking results for a weighted least squares reconstruction (as implemented in Slicer). Panel (b) shows the tracking results obtained with the UKFT-module (Slicer plugin) [65]. Panel (c) shows the fiber tracking based on the proposed reconstruction algorithm. All results are displayed using mean orientation color coding. We apply the fiber tracking implemented in Slicer to the result obtained with the respective reconstruction method. We note that the proposed method leads to an improved reconstruction of the (vertical) fibers corresponding to the cingulum. Note that fiber tracking has only been performed for a rectangular region of interest. (a) Weighted least squares (WLS). (b) UKFT module. (c) Proposed.

Another suitable mathematical structure for DTI is the so-called log-Euclidean framework as employed by Arsigny *et al.* [18], [26] and Fillard *et al.* [10]. By using a Log-Euclidean metric, one essentially works in the tangent space of the identity matrix and then solves the present problem in this linear space. Fillard *et al.* [10] perform estimation, smoothing, and fiber tracking in the log-Euclidean setup and demonstrate the potential of this approach for both synthetic and real clinical data. One of the main advantages of the log-Euclidean approach is that it is computationally very efficient. This advantage comes at the cost of not being affine invariant and the drawback of being restricted to a particular base point, i.e., the identity matrix. Hence, this approach of going to a tangent space at the identity matrix yields good results for nearby points, but one loses quality when they are not in a vicinity of the identity.

2) Voxel-Wise Tensor Fitting: While the (voxelwise) least squares approach based on the Stejskal-Tanner (10) ensures the symmetry of the fitted tensors by restricting the sought tensor coefficients to the upper triangular part of U_i, it does not guarantee that the resulting objects are positive (semi-) definite. Besides imposing more sophisticated mathematical structures, it is possible to enforce this constraint (voxel-wise) by projecting the eigenvalues of the respective tensor to the positive real axis, for example, by letting negative eigenvalues equal zero [16]. Alternatively, one might use a so-called Cholesky parameterization of the tensors which leads to non-linear least squares problems as suggested by Koay *et al.* [11] and Wang *et al.* [12]. It has been shown by Fillard *et al.* in [10] that the log-Euclidean framework has advantages over the Cholesky parametrization. For example, in the latter approaches, it is possible to reach the zero tensor via geodesic shooting which might be an undesireable feature.

B. Limitations of DTI and the Proposed Framework in the Context of More Advanced Models

The major limitation of classical DTI appears in the representation of intravoxel crossings of fibers [84], [85]. In such cases the diffusion process is no longer well described by a single tensor which leads to falsely fitted plate shaped or even sphere shaped tensors.

In order to deal with such crossings, several approaches have been proposed; see for instance [84], [86]–[88]. These approaches can be classified into two groups. In the first group the single-tensor-model is replaced by a multi-tensor one. This means that instead of one tensor in each voxel one uses multiple ones; see for instance [85], [89], [90]. The second group is based on a model-free approach. The major assumption is that the diffusion process can be described by a orientation distribution function (ODF) on the 3D unit sphere; see for example [86], [88], [91]–[97]. This imaging approach is also called Q-ball imaging. For a review on both types of approaches we refer to the article of Alexander [92].

We note that – similar to the standard DTI case – the model space in the Q-ball setup can be interpreted as Riemannian manifold, see Goh *et al.* [98] for instance. Hence, the framework proposed in the present paper can be applied to the corresponding manifold as well. In this context, we refer the reader to the recent paper by Weinmann *et al.* [31] where the authors describe an a posteriori regularization method in the Q-ball setup. In [31], the needed differential-geometric operations are described so that the implementation of the framework proposed in the present paper for the manifold in Q-ball imaging is already principally described. The concrete implementation of the proposed algorithm for the the Q-ball setting is, however, a topic of future work.

Furthermore, the approach proposed here is also flexible enough to be applied in multi-tensor setups. For example, it is possible to combine our TV regularizer with the data term in [43] to deal with free-water elimination. This is another topic of future research.

VI. Conclusion and Future Research

In this paper, we have proposed a novel approach for combined tensor fitting and edge-preserving regularization. To this end, we have introduced an energy functional consisting of a data fidelity term adapted to Rician noise on the DWIs and, as regularizer, the total variation on the tensors with respect to the Riemannian manifold of positive matrices. As minimization strategy, we have developed a generalized forward-backward scheme which employs implicit steps based on geodesic averaging on the manifold and explicit gradient steps for the data

fidelity term. We validated the performance of the derived algorithms on synthetic and real DTI data. To our knowledge, the present work is the first one to propose TV regularization in a combined (non-flat) manifold and inverse problem setup.

We saw that the employed techniques are fairly general. In particular, to make the approach work for another manifold, we basically need an implementation of the gradient of the measurement operator as well as methods to compute the Riemannian exponential mapping and its inverse.

As such, our framework is also applicable to more recent models which overcome limitations of DTI. For example, the above operations are available for Q-ball imaging as indicated in Section V-B which makes the implementation of the proposed framework for Q-ball imaging a first concrete topic of future research. A second topic is the application to multi-tensor setup as also pointed out in Section V-B.

Further directions of future research are validating the parallel approach, considering further manifolds (such as rotation groups or shape spaces) as well as different inverse and modeling problems, respectively.

ACKNOWLEDGMENT

The authors gratefully acknowledge the comments of O. Pasternak on the selection of the employed fiber tracking method as well as the clinical interpretation of the obtained results by A. Ahmadi and A. Plate. Moreover, the authors acknowledge the constructive feedback of all anonymous reviewers which helped to greatly improve the manuscript.

REFERENCES

[1] P. Basser, J. Mattiello, and D. LeBihan, "MR diffusion tensor spectroscopy and imaging," *Biophys. J.*, vol. 66, no. 1, pp. 259–267, 1994.

[2] D. Le Bihan *et al.*, "Diffusion tensor imaging: Concepts and applications," *J. Magn. Reson. Imaging.*, vol. 13, pp. 534–546, 2001.

[3] J. Foong *et al.*, "Neuropathological abnormalities of the corpus callosum in schizophrenia: A diffusion tensor imaging study," *J. Neurol., Neurosurg. Psychiatry*, vol. 68, no. 2, pp. 242–244, 2000.

[4] M. Kubicki *et al.*, "A review of diffusion tensor imaging studies in schizophrenia," *J. Psychiatric Res.*, vol. 41, no. 1, pp. 15–30, 2007.

[5] A. Alexander *et al.*, "Diffusion tensor imaging of the corpus callosum in autism," *NeuroImage*, vol. 34, no. 1, pp. 61–73, 2007.

[6] H. Rosas *et al.*, "Altered white matter microstructure in the corpus callosum in Huntington's disease: Implications for cortical disconnection," *NeuroImage*, vol. 49, no. 4, pp. 2995–3004, 2010.

[7] H. Johansen-Berg and T. Behrens, *Diffusion MRI: From Quantitative Measurement to In-Vivo Neuroanatomy*. London, U.K.: Academic, 2009.

[8] Y. Assaf and O. Pasternak, "Diffusion tensor imaging (DTI)-based white matter mapping in brain research: A review," *J. Molecular Neurosci.*, vol. 34, no. 1, pp. 51–61, 2008.

[9] H.-E. Assemlal, D. Tschumperlé, L. Brun, and K. Siddiqi, "Recent advances in diffusion MRI modeling: Angular and radial reconstruction," *Med. Image Anal.*, vol. 15, no. 4, pp. 369–396, 2011.

[10] P. Fillard, X. Pennec, V. Arsigny, and N. Ayache, "Clinical DT-MRI estimation, smoothing, and fiber tracking with log-Euclidean metrics," *IEEE Trans. Med. Imag.*, vol. 26, no. 11, pp. 1472–1482, Nov. 2007.

[11] C. Koay, L.-C. Chang, J. Carew, C. Pierpaoli, and P. Basser, "A unifying theoretical and algorithmic framework for least squares methods of estimation in diffusion tensor imaging," *J. Magn. Reson.*, vol. 182, no. 1, pp. 115–125, 2006.

[12] Z. Wang, B. C. 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. Med. Imag.*, vol. 23, no. 8, pp. 930–939, Aug. 2004.

[13] S. Basu, T. Fletcher, and R. Whitaker, "Rician noise removal in diffusion tensor MRI," in *Proc. MICCAI*, 2006, vol. 4190, pp. 117–125.

[14] L. Bao *et al.*, "Denoising human cardiac diffusion tensor magnetic resonance images using sparse representation combined with segmentation," *Phys. Med. Biol.*, vol. 54, no. 6, p. 1435, 2009.

[15] J. Luo, Y. Zhu, and I. Magnin, "Denoising by averaging reconstructed images: Application to magnetic resonance images," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 3, pp. 666–674, Mar. 2009.

[16] D. Tschumperlé and R. Deriche, "Diffusion tensor regularization with constraints preservation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2001, pp. I948–I953.

[17] C. Chefd'hotel, D. Tschumperlé, R. Deriche, and O. Faugeras, "Regularizing flows for constrained matrix-valued images," *J. Math. Imag. Vis.*, vol. 20, no. 1-2, pp. 147–162, 2004.

[18] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, "Fast and simple calculus on tensors in the log-Euclidean framework," in *Proc. MICCAI*, 2005, pp. 115–122.

[19] X. Pennec, P. Fillard, and N. Ayache, "A Riemannian framework for tensor computing," *Int. J. Comput. Vis.*, vol. 66, no. 1, pp. 41–66, 2006.

[20] C. A. Castaño Moraga, C.-F. Westin, and J. Ruiz-Alzola, "Homomorphic filtering of DT-MRI fields," in *Medical Image Comput. and Computer-Assisted Intervent. (MICCAI)*, R. E. Ellis and T. M. Peters, Eds., Berlin, Germany, 2003, vol. 2879, Lecture Notes in Computer Science, pp. 990–991.

[21] C. A. Castaño Moraga, C. Lenglet, R. Deriche, and J. Ruiz-Alzola, "A Riemannian approach to anisotropic filtering of tensor fields," *Signal Process.*, vol. 87, no. 2, pp. 263–276, 2007.

[22] Y. Gur and N. Sochen, "Fast invariant Riemannian DT-MRI regularization," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2007, pp. 1–7.

[23] D. Tschumperlé and R. Deriche, "Tensor field visualization with PDE's and application to DT-MRI fiber visualization," in *Workshop Variational Level Set Methods*, 2003.

[24] C. Chefd'hotel, D. Tschumperlé, R. Deriche, and O. Faugeras, "Constrained flows of matrix-valued functions: Application to diffusion tensor regularization," in *Proc. Eur. Conf. Comput. Vis.*, 2002, pp. 251–265.

[25] R. Neji, N. Azzabou, N. Paragios, and G. Fleury, "A convex semi-definite positive framework for DTI estimation and regularization," in *Lecture Notes in Computer Science*. New York: Springer, 2007, vol. 4841, pp. 220–229.

[26] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, "Log-Euclidean metrics for fast and simple calculus on diffusion tensors," *Magn. Reson. Med.*, vol. 56, no. 2, pp. 411–421, 2006.

[27] O. Pasternak, N. Sochen, and Y. Assaf, "Variational regularization of multiple diffusion tensor fields," in *Visualization and Processing of Tensor Fields*, J. Weickert and H. Hagen, Eds. Berlin, Germany: Springer, 2006, Math.Visualizat., pp. 165–176.

[28] D. Tschumperlé and R. Deriche, "Variational frameworks for DT-MRI estimation, regularization and visualization," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2003, pp. 116–121.

[29] C. Rao, "Information and accuracy attainable in the estimation of statistical parameters," *Bull. Calcutta Math. Soc.*, vol. 37, no. 3, pp. 81–91, 1945.

[30] A. Weinmann, L. Demaret, and M. Storath, "Total variation regularization for manifold-valued data," *SIAM J. Imag. Sci.*, vol. 7, no. 4, pp. 2226–2257, 2014.

[31] A. Weinmann, L. Demaret, and M. Storath, Mumford-Shah and Potts regularization for manifold-valued data with applications to DTI and Q-ball imaging 2015 [Online]. Available: arXiv:1410.1699, to be published

[32] P. Grohs and M. Sprecher, "Total variation regularization by iteratively reweighted least squares on Hadamard spaces and the sphere," *IMA J. Inf. Inference*, 2015, to be published.

[33] E. Stejskal and J. Tanner, "Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient," *J. Chem. Phys.*, vol. 42, no. 1, pp. 288–292, 1965.

[34] H. Raguet, J. Fadili, and G. Peyré, "A generalized forward-backward splitting," *SIAM J. Imag. Sci.*, vol. 6, no. 3, pp. 1199–1226, 2013.

[35] C. Lenglet *et al.*, "Mathematical methods for diffusion MRI processing," *NeuroImage*, vol. 45, no. 1, pp. S111–S122, 2009.

[36] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, Jul. 1990.

[37] M. Giaquinta and D. Mucci, "The BV-energy of maps into a manifold: Relaxation and density results," *Annali della Scuola Normale Superiore di Pisa – Classe di Scienze*, vol. 5, no. 4, pp. 483–548, 2006.

[38] M. Giaquinta and D. Mucci, "Maps of bounded variation with values into a manifold: Total variation and relaxed energy," *Pure Appl. Math. Quart.*, vol. 3, no. 2, pp. 513–538, 2007.

[39] M. Giaquinta, G. Modica, and J. Souček, "Variational problems for maps of bounded variation with values in S^1," *Calculus Variat. Partial Differential Equ.*, vol. 1, no. 1, pp. 87–121, 1993.

[40] J. Lellmann, E. Strekalovskiy, S. Koetter, and D. Cremers, "Total variation regularization for functions with values in a manifold," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 2944–2951.

[41] E. Strekalovskiy and D. Cremers, "Total variation for cyclic structures: Convex relaxation and efficient minimization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 1905–1911.

[42] D. Cremers and E. Strekalovskiy, "Total cyclic variation and generalizations," *J. Math. Imag. Vis.*, vol. 47, no. 3, pp. 258–277, 2013.

[43] O. Pasternak, N. Sochen, Y. Gur, N. Intrator, and Y. Assaf, "Free water elimination and mapping from diffusion MRI," *Magn. Reson. Med.*, vol. 62, no. 3, pp. 717–730, 2009.

[44] R. Kimmel, N. Sochen, and R. Malladi, "From high energy physics to low level vision," in *Scale-Space Theory in Computer Vision*. New York: Springer, 1997, pp. 236–247.

[45] Y. Gur, O. Pasternak, and N. Sochen, "Fast $\mathrm{GL}(n)$-invariant framework for tensors regularization," *Int. J. Comput. Vis.*, vol. 85, no. 3, pp. 211–222, 2009.

[46] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1, pp. 259–268, 1992.

[47] A. Chambolle, V. Caselles, D. Cremers, M. Novaga, and T. Pock, "An introduction to total variation for image analysis," in *Theoretical Foundations and Numerical Methods for Sparse Recovery*. Berlin, Germany: De Gruyter, 2010.

[48] T. Chan and J. Shen, "Local inpainting models and TV inpainting," *SIAM J. Appl. Math.*, vol. 62, no. 3, pp. 1019–1043, 2001.

[49] X. Bresson and T. Chan, "Fast dual minimization of the vectorial total variation norm and applications to color image processing," *Inverse Problems Imag.*, vol. 2, no. 4, pp. 455–484, 2008.

[50] P. Blomgren and T. Chan, "Color TV: Total variation methods for restoration of vector-valued images," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 304–309, Mar. 1998.

[51] L. Rakêt, L. Roholm, M. Nielsen, and F. Lauze, "TV-L^1 optical flow for vector valued images," in *Proc. Int. Conf. Energy Minimizat. Methods Comput. Vis. Pattern Recognit.*, 2011, pp. 329–343.

[52] T. Möllenhoff, E. Strekalovskiy, M. Möller, and D. Cremers, "Low rank priors for color image regularization," in *Energy Minimization Methods in Computer Vision and Pattern Recognition*. New York: Springer, pp. 126–140.

[53] R. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton Univ. Press, 1970.

[54] D. Peaceman and H. Rachford, "The numerical solution of parabolic and elliptic differential equations," *J. Soc. Ind. Appl. Math.*, vol. 3, no. 1, pp. 28–41, 1955.

[55] J. Douglas and H. Rachford, "On the numerical solution of heat conduction problems in two and three space variables," *Trans. Am. Math. Soc.*, vol. 82, pp. 421–439, 1956.

[56] J. Eckstein and D. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 55, no. 1–3, pp. 293–318, 1992.

[57] J. Sijbers, A. den Dekker, P. Scheunders, and D. Van Dyck, "Maximum-likelihood estimation of Rician distribution parameters," *IEEE Trans. Med. Imag.*, vol. 17, no. 3, pp. 357–361, Jun. 1998.

[58] A. Chambolle, "Finite-differences discretizations of the Mumford-Shah functional," *ESAIM: Math. Model. Numerical Anal.*, vol. 33, no. 2, pp. 261–288, 1999.

[59] M. Storath, A. Weinmann, J. Frikel, and M. Unser, "Joint image reconstruction and segmentation using the Potts model," *Inverse Problems*, vol. 31, no. 2, p. 025003, 2015.

[60] X. Pennec, "Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements," *J. Math. Imag. Vis.*, vol. 25, no. 1, pp. 127–154, 2006.

[61] P. Cook *et al.*, "Camino: Open-source diffusion-MRI reconstruction and processing," in *14th Sci. Meet. Int. Soc. Magn. Reson. Med.*, 2006, p. 2759.

[62] A. Barmpoutis, B. Vemuri, T. Shepherd, and J. Forder, "Tensor splines for interpolation and approximation of DT-MRI with applications to segmentation of isolated rat hippocampi," *IEEE Trans. Med. Imag.*, vol. 26, no. 11, pp. 1537–1546, Nov. 2007.

[63] M. Holz, S. Heil, and A. Sacco, "Temperature-dependent self-diffusion coefficients of water and six selected molecular liquids for calibration in accurate 1 h NMR PFG measurements," *Phys. Chemistry Chemical Phys.*, vol. 2, no. 20, pp. 4740–4742, 2000.

[64] C. Koay and P. Basser, "Analytically exact correction scheme for signal extraction from noisy magnitude MR signals," *J. Magn. Reson.*, vol. 179, no. 2, pp. 317–322, 2006.

[65] C. Baumgartner *et al.*, "A unified tractography framework for comparing diffusion models on clinical scans," in *Comput. Diffusion MRI Workshop MICCAI*, Nice, 2012, pp. 27–32.

[66] M. Unser and P. Tafti, *An Introduction to Sparse Stochastic Processes*. Cambridge, U.K.: Cambridge Univ. Press, 2014.

[67] P. Fletcher and S. Joshi, "Riemannian geometry for the statistical analysis of diffusion tensor data," *Signal Process.*, vol. 87, no. 2, pp. 250–262, 2007.

[68] G. Cheng, H. Salehian, and B. Vemuri, "Efficient recursive algorithms for computing the mean diffusion tensor and applications to DTI segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 390–401.

[69] Z. Wang and B. Vemuri, "DTI segmentation using an information theoretic tensor dissimilarity measure," *IEEE Trans. Med. Imag.*, vol. 24, no. 10, pp. 1267–1277, Oct. 2005.

[70] Z. Wang and B. Vemuri, "An affine invariant tensor dissimilarity measure and its applications to tensor-valued image segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2004, pp. I228–I233.

[71] I. Ur Rahman, I. Drori, V. Stodden, D. Donoho, and P. Schröder, "Multiscale representations for manifold-valued data," *Multiscale Model. Simulat.*, vol. 4, no. 4, pp. 1201–1232, 2005.

[72] P. Grohs and J. Wallner, "Interpolatory wavelets for manifold-valued data," *Appl. Computat. Harmonic Anal.*, vol. 27, pp. 325–333, 2009.

[73] A. Weinmann, "Interpolatory multiscale representation for functions between manifolds," *SIAM J. Math. Anal.*, vol. 44, pp. 162–191, 2012.

[74] P. Grohs, H. Hardering, and O. Sander, "Optimal a priori discretization error bounds for geodesic finite elements," *Foundat. Computat. Math.*, 2015, to be published.

[75] P. Fletcher, "Geodesic regression and the theory of least squares on Riemannian manifolds," *Int. J. Comput. Vis.*, vol. 105, pp. 171–185, 2013.

[76] P. Fletcher, C. Lu, S. Pizer, and S. Joshi, "Principal geodesic analysis for the study of nonlinear statistics of shape," *IEEE Trans. Med. Imag.*, vol. 23, no. 8, pp. 995–1005, Aug. 2004.

[77] J. Oller and J. Corcuera, "Intrinsic analysis of statistical estimation," *Ann. Stat.*, pp. 1562–1581, 1995.

[78] R. Bhattacharya and V. Patrangenaru, "Large sample theory of intrinsic and extrinsic sample means on manifolds I," *Ann. Stat.*, pp. 1–29, 2003.

[79] R. Bhattacharya and V. Patrangenaru, "Large sample theory of intrinsic and extrinsic sample means on manifolds II," *Ann. Stat.*, pp. 1225–1259, 2005.

[80] M. Welk *et al.*, "Median and related local filters for tensor-valued images," *Signal Process.*, vol. 87, no. 2, pp. 291–308, 2007.

[81] K. Kwon *et al.*, "Regularization of DT-MR images using a successive Fermat median filtering method," *Phys. Med. Biol.*, vol. 53, no. 10, p. 2523, 2008.

[82] T. Schultz, B. Burgeth, and J. Weickert, "Flexible segmentation and smoothing of DT-MRI fields through a customizable structure tensor," in *Advances in Visual Computing*. New York: Springerr, 2006, pp. 455–464.

[83] J. Weickert and H. Hagen, *Visualization and Processing of Tensor Fields*. New York: Springer, 2006.

[84] D. C. Alexander, G. J. Barker, and S. R. Arridge, "Detection and modeling of non-Gaussian apparent diffusion coefficient profiles in human brain data," *Magn. Reson. Med.*, vol. 48, no. 2, pp. 331–340, 2002.

[85] D. S. Tuch *et al.*, "High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity," *Magn. Reson. Med.*, vol. 48, no. 4, pp. 577–582, 2002.

[86] M. Descoteaux, E. Angelino, S. Fitzgibbons, and R. Deriche, "Regularized, fast, and robust analytical Q-ball imaging," *Magn. Reson. Med.*, vol. 58, no. 3, pp. 497–510, 2007.

[87] L. R. Frank, "Characterization of anisotropy in high angular resolution diffusion-weighted MRI," *Magn. Reson. Med.*, vol. 47, no. 6, pp. 1083–1099, 2002.

[88] E. Özarslan and T. H. Mareci, "Generalized diffusion tensor imaging and analytical relationships between diffusion tensor imaging and high angular resolution diffusion imaging," *Magn. Reson. Med.*, vol. 50, no. 5, pp. 955–965, 2003.

[89] A. L. Alexander, K. M. Hasan, M. Lazar, J. S. Tsuruda, and D. L. Parker, "Analysis of partial volume effects in diffusion-tensor MRI," *Magn. Reson. Med.*, vol. 45, no. 5, pp. 770–780, 2001.

[90] O. Pasternak, Y. Assaf, N. Intrator, and N. Sochen, "Variational multiple-tensor fitting of fiber-ambiguous diffusion-weighted magnetic resonance imaging voxel," *Magn. Reson. Imag.*, vol. 26, no. 8, pp. 1133–1144, 2008.

[91] D. S. Tuch, "Q-ball imaging," *Magn. Reson. Med.*, vol. 52, no. 6, pp. 1358–1372, 2004.

[92] D. C. Alexander, "Multiple-fiber reconstruction algorithms for diffusion MRI," *Ann. NY Acad. Sci.*, vol. 1064, no. 1, pp. 113–133, 2005.

[93] M. Descoteaux, E. Angelino, S. Fitzgibbons, and R. Deriche, "A fast and robust ODF estimation algorithm in Q-ball imaging," in *Proc. IEEE Int. Symp. Biomed. Imag., Macro to Nano.*, 2006, pp. 81–84.

[94] C. P. Hess, P. Mukherjee, E. T. Han, D. Xu, and D. B. Vigneron, "Q-ball reconstruction of multimodal fiber orientations using the spherical harmonic basis," *Magn. Reson. Med.*, vol. 56, no. 1, pp. 104–117, 2006.

[95] D. S. Tuch, T. G. Reese, M. R. Wiegell, and V. J. Wedeen, "Diffusion MRI of complex neural architecture," *Neuron*, vol. 40, no. 5, pp. 885–895, 2003.

[96] J. D. Tournier, F. Calamante, D. G. Gadian, and A. Connelly, "Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution," *NeuroImage*, vol. 23, no. 3, pp. 1176–1185, 2004.

[97] K. M. Jansons and D. C. Alexander, "Persistent angular structure: New insights from diffusion magnetic resonance imaging data," *Inverse Problems*, vol. 19, no. 5, pp. 1031–1046, 2003.

[98] A. Goh, C. Lenglet, P. M. Thompson, and R. Vidal, "A nonparametric Riemannian framework for processing high angular resolution diffusion images and its applications to ODF-based morphometry," *NeuroImage*, vol. 56, no. 3, pp. 1181–1201, 2011.