Paper D

Total Variation Regularization of Matrix Valued Images *

* Preprint submitted to International Journal of Biomedical Imaging (Hindawi)

Total Variation Regularization of Matrix Valued Images.

Oddvar Christiansen, Tin-Man Lee , Johan Lie, Usha Sinha, Tony F. Chan
VERSION January 2, 2007

***Abstract*— In this work we generalize the total variation restoration model, introduced by Rudin, Osher and Fatemi in 1992, to matrix valued data. In particular to Diffusion Tensor Images (DTI). Our model is a natural extension of the color total variation model proposed by Blomgren and Chan in 1996. We treat the diffusion matrix D implicitly as the product $D = LL^T$, and work with the elements of L as variables, instead of working directly on the elements of D. This ensures positive definiteness of the tensor during the regularization flow, which is essential when regularizing DTI. We perform numerical experiments on both synthetical data and 3D human brain DTI, and measure the quantitative behavior of the proposed model.**

I. Introduction

Image processing methods using variational calculus and partial differential equations (PDEs) have been popular for a long time in the image processing community. Among popular PDE methods are the anisotropic diffusion method proposed by Perona and Malik [1], the total variation method introduced by Rudin, Osher and Fatemi [2] and various methods related to these [3]–[7]. Many of these methods were originally introduced for scalar valued (gray scale) images, and later generalized to vector valued (color) images.

During the last decade or so, a new Magnetic Resonance modality called Diffusion Tensor Imaging (DTI) has been extensively studied [8]–[13]. Using DTI, it is possible to study anatomical structures like the nerve fibers in the human brain non-invasively. The DTI images are matrix valued. From the developments in DTI a need for robust regularization methods for matrix valued images has emerged. As far as the authors are aware of, there exists no state-of-the-art method for regularization of tensor valued images, although many methods have been proposed [14]–[17]. In each voxel of the imaging

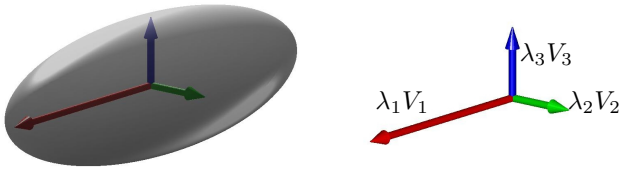


Fig. 1. The diffusion matrix D can be represented by a diffusion ellipsoid, where the semi-axes are spanned by the eigenvectors $\{V_i\}_{i=1}^3$ of D, and the length of each semi-axis is given by the eigenvalues $\{\lambda_i\}_{i=1}^3$. In this illustration the diffusion is anisotropic. The principal diffusion direction is along eigenvector V_1.

domain we construct a *diffusion tensor* (i.e. *diffusion matrix*) D based on a series of K direction-specific MR measurements $\{S_k\}_{k=1}^K$. The matrix $D \in R^{3\times 3}$ is a symmetric, positive definite matrix. This means that we can write the matrix as

$$D = V\Lambda V^{-1}, \tag{1}$$

where V is an orthogonal matrix, and Λ is a diagonal matrix with positive elements. This decomposition is called the eigenvalue decomposition of D. The elements $\{\lambda_i\}_{i=1}^3$ of Λ are called eigenvalues of D, and the columns of V are called eigenvectors of D. We may look at the diffusion matrix as a hyper-ellipse where the eigenvectors $\{V_i\}_{i=1}^3$ span the ellipsoid and the corresponding eigenvalues $\{\lambda_i\}_{i=1}^3$ determines the length of each semi-axis. It is customary to arrange the eigenvalues in decreasing order. By the diffusion tensor model we assume that the set of images $\{S_k\}_{k=1}^K$ are related to the non-weighted image S_0 by the Stejskal-Tanner equation [18], [19]

$$S_k = S_0 e^{-bg_k^T D g_k} \quad k = 1, 2, \cdots, K. \tag{2}$$

Here $g_k \in R^3$ denotes the direction associated with S_k, and $b > 0$ is a scalar which among other factors depends on the acquisition time and the strength of the magnetic field [20]. Since $D \in R^{3\times 3}$ is symmetric, it has six degrees of freedom. Thus at least six measurements S_k are required to estimate the tensor, as well as the non-weighted measurement S_0. The tensor D can be estimated from equation (2). In the case of more than six measurements S_k, we can for example use a least squares minimization. From the eigenvalue decomposition of the diffusion tensor, we can reveal properties like the dominant diffusion direction and the anisotropy of diffusing water molecules [21]. This information can be used to construct maps of the anatomy of the brain.

All measurements $\{S_k\}_{k=1}^K$ contain noise, which degrades the accuracy of the estimated tensor. Compared with conventional MR, direction sensitive acquisitions have a lower signal to noise ratio (SNR). Thus the gradient weighted images $\{S_k\}_{k=1}^K$ contain more noise than S_0. There are several ways to increase the accuracy of the estimated tensor. The most intuitive way is to make an average of a series of repeated measurements. Alternatively, we can increase the number of gradient directions. An obvious disadvantage of both these approaches is the increased scanner time. Perhaps a better way to improve the quality of the tensor is by postprocessing the data. In this paper we follow this approach, by introducing a novel regularization method for tensor valued data.

Since D models diffusion, regularization methods in DTI must preserve the positive definiteness of D. A positive definite matrix has only positive eigenvalues, which is necessary

from the physical modelling perspective. In a minimization method for regularization of the tensor data, one possible way to ensure positive definiteness would be to impose a constraint on the minimization problem. Then the constrained problem would have a solution which is on the manifold of positive definite matrices. Regularization of tensor valued data constrained to manifolds has been studied during the last couple of years, see [22]–[24]. We however follow a different strategy. Using essentially the same idea as Wang and coworkers did in a slightly different setting, we treat D implicitly by writing D as the product $D = LL^T$, where L is a lower triangular matrix. Every symmetric positive definite (SPD) matrix has a factorization on this form. We will in this work develop a novel regularization method for diffusion tensor images, by generalizing methods previously developed for scalar and vector valued images [2], [25].

Before we go into details of the proposed method, we briefly introduce the Total Variation (TV) methods for scalar and vector valued images. During the last 15 years or so, TV models have undergone extensive studies, initiated by the work of Rudin Osher and Fatemi (ROF) [2].

Define the Total Variation (TV) semi-norm for scalar valued data as

$$\mathrm{TV}[u] = \int_\Omega |\nabla u| \mathrm{d}x \tag{3}$$

Throughout this paper, ∇ denotes the spatial gradient, while $\nabla\cdot$ denotes the divergence operator. In the ROF model, the TV semi-norm with an added L_2 fidelity norm is minimized

$$\min_u \left\{ G(u,f) = \mathrm{TV}[u] + \frac{\lambda}{2}||u-f||_2^2 \right\}. \tag{4}$$

Note that we can write the functional G more abstractly as

$$G(u,f,\lambda) = R(u) + \frac{\lambda}{2}F(u,f), \tag{5}$$

where $R(u)$ is a regularization functional and $F(u,f)$ is a fidelity functional. The regularization term is a geometric functional measuring smoothness of the estimated solution. The fidelity term is a measure of fitness of the estimated solution compared to the input data. It is customary to measure the fidelity in the sense of least squares. The equation (4) has the corresponding Euler-Lagrange equation

$$\partial_u G = -\nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right) + \lambda(u-f). \tag{6}$$

We can find a minimum of (4) by searching for a steady state of

$$\frac{\partial u}{\partial t} = -\partial_u G, \tag{7}$$

which is the way the ROF model was first formulated [2], or by directly attacking the zero of the Euler-Lagrange equation

$$-\partial_u G = 0, \tag{8}$$

for example by a fixed-point iteration [26]. This is in general less time consuming than solving the equation using the method of steepest descent, but more tedious to carry out numerically. When we generalize the method to matrix valued images, we solve the minimization problem by the method of steepest descent. Various methods have been proposed to generalize the ROF model to vector valued image regularization. Among the successful methods we find the vector TV model developed by Blomgren and Chan [25] and the model by Shapiro [27]. Blomgren and Chan [25] generalized the ROF model to vector (color) image regularization using a set of coupled equations

$$\left\{ \frac{\partial u_i}{\partial t} = \alpha_i \nabla \cdot \left(\frac{\nabla u_i}{|\nabla u_i|} \right) - \lambda(u_i - f_i) \quad i = 1,2,3 \right\} \tag{9}$$

with

$$\alpha_i = \frac{\mathrm{TV}[u_i]}{\mathrm{TV}[\mathbf{u}]}, \quad i = 1,2,3 \tag{10}$$

and

$$\mathrm{TV}[\mathbf{u}] = \sqrt{\sum_{i=1}^{3} \mathrm{TV}[u_i]^2}. \tag{11}$$

The weight α_i in Eq. (9) acts as a coupling between the geometric part of the three image channels. In this work we extend in a natural way the vector TV model of Blomgren and Chan to a matrix TV model. However, the method we propose is not restricted to our choice of regularization functional (TV). For a detailed treatment of TV regularization methods we refer the reader to the recent book by Chan and Shen [5].

In section II we define the minimization problem that we propose to solve, and arrive at the Euler-Lagrange Equations corresponding to this minimization problem. We perform numerical experiments in section III, before we finish the paper in section IV with a conclusion. Details on the Euler-Lagrange equation and the numerical implementation are given in the Appendix at the end of the paper.

II. Minimization Problem

In this section we introduce the functional that we minimize in order to regularize tensor valued data. Let L be a lower triangular matrix. We define D by

$$D = LL^T. \tag{12}$$

This has immediate implications on D; symmetry, positive definiteness and orthogonality of eigenvectors. These properties are required by the diffusion tensor model. Thus (12) is a natural choice. We define ℓ_{ij} as the element in the i'th row and j'th column of L. The elements d_{ij} are defined in the same manner.

Let us look at the algebraic equation expressing D as a function of ℓ_{ij}. We derive the expressions for $D \in R^{3\times 3} \subset SPD$. We explicitly write out the matrix multiplication (12)

$$D = \begin{pmatrix} \ell_{11}^2 & \ell_{11}\ell_{21} & \ell_{11}\ell_{31} \\ \ell_{11}\ell_{21} & \ell_{21}^2 + \ell_{22}^2 & \ell_{21}\ell_{31} + \ell_{22}\ell_{32} \\ \ell_{11}\ell_{31} & \ell_{21}\ell_{31} + \ell_{22}\ell_{32} & \ell_{31}^2 + \ell_{32}^2 + \ell_{33}^2 \end{pmatrix}. \tag{13}$$

In our proposed model we solve a minimization problem in terms of ℓ_{ij}. For each unique ℓ_{kl} we minimize

$$\min_{\ell_{kl}} \left\{ \sqrt{\sum_{ij} \mathrm{TV}[d_{ij}(\ell_{kl})]^2} + \frac{\lambda}{2} \sum_{ij} \|d_{ij} - \hat{d}_{ij}\|_2^2 \right\}, \tag{14}$$

where $\{kl\} \in \{11, 21, 22, 31, 32, 33\}$ and $\hat{d}_{ij}$ denotes the elements of the tensor estimated from the noisy data. Similarly as in the scalar model, the functional (14) has the abstract form (5). The scalar ROF ($TV - L_2$) functional is convex. But when we introduce the factorization (12) into the model, we can not expect the functional (14) to be convex or even quazi-convex. However, from numerical experiments where we use different (random) initial conditions we do end up with almost exactly the same solution. This means that even though we are not able to prove that the functional is convex, we have indications that it is quazi-convex. [We could include a numerical example where we use one of the synthetic fields, and run simulations where we use for example 10 different initial conditions. We could then make a table that shows the difference between the recovered solutions.]

We here note that minimizing the functional (14) is related to the functional used by Wang and coauthors [14]. Apart from the fact that they simultaneously estimate and regularize the tensor, there are fundamental differences between our proposed regularization functional and the functional proposed by Wang and others. Even though we represent the diffusion matrix on the form of a Cholesky factorization, we regularize the elements of the full diffusion tensor D, while Wang and coauthors regularize the elements of the lower triangular matrix L. Intuitively, regularizing the elements of D is more direct than regularizing the elements of L. We highlight the difference between Wang's method and our proposed method by a numerical simulation in a simplified setting in section III. In addition, the method proposed in this paper has a coupling between all elements of D in the regularization PDE, while the method proposed by Wang et. al. does not have such a coupling between the channels.

We also note that the functional (14) is chosen mainly because of the good properties of the corresponding scalar and vector valued functionals [2], [25], with edge preservation as the most prominent property. Depending on the application at hand, other functionals might be considered as alternatives. The framework developed in this paper is however by applicable for other regularization functionals as well.

A. Euler-Lagrange Equations

In this section we derive the Euler-Lagrange equations corresponding to the minimization functional (14). We first differentiate the fidelity functional

$$\begin{aligned}\frac{\partial F}{\partial \ell_{ij}} &= \frac{\partial}{\partial \ell_{ij}} \sum_{ij} \|d_{ij} - \hat{d}_{ij}\|_2^2 \\ &= 2 \sum_{ij} \left(d_{ij} - \hat{d}_{ij} \right) \frac{\partial d_{ij}}{\partial \ell_{ij}}. \end{aligned} \tag{15}$$

We differentiate D with respect to ℓ_{ij}.

$$\frac{\partial D}{\partial \ell_{11}} = \begin{pmatrix} 2\ell_{11} & \ell_{21} & \ell_{31} \\ \ell_{21} & 0 & 0 \\ \ell_{31} & 0 & 0 \end{pmatrix}, \tag{16}$$

$$\frac{\partial D}{\partial \ell_{12}} = \begin{pmatrix} 0 & \ell_{11} & 0 \\ \ell_{11} & 2\ell_{21} & \ell_{31} \\ 0 & \ell_{31} & 0 \end{pmatrix} \tag{17}$$

$$\frac{\partial D}{\partial \ell_{31}} = \begin{pmatrix} 0 & 0 & \ell_{11} \\ 0 & 0 & \ell_{21} \\ \ell_{11} & \ell_{21} & 2\ell_{31} \end{pmatrix}, \tag{18}$$

$$\frac{\partial D}{\partial \ell_{22}} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 2\ell_{22} & \ell_{32} \\ 0 & \ell_{32} & 0 \end{pmatrix}, \tag{19}$$

$$\frac{\partial D}{\partial \ell_{32}} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \ell_{22} \\ 0 & \ell_{22} & 2\ell_{32} \end{pmatrix}, \tag{20}$$

$$\frac{\partial D}{\partial \ell_{33}} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2\ell_{33} \end{pmatrix}. \tag{21}$$

Writing out (15), we find the derivative of the fidelity functional

$$\begin{aligned}\frac{\partial F}{\partial \ell_{kl}} &= 2\Big[(d_{11} - \hat{d}_{11})\frac{\partial d_{11}}{\partial \ell_{kl}} + 2(d_{21} - \hat{d}_{21})\frac{\partial d_{21}}{\partial \ell_{kl}} \\ &\quad + (d_{22} - \hat{d}_{22})\frac{\partial d_{22}}{\partial \ell_{kl}} + 2(d_{31} - \hat{d}_{31})\frac{\partial d_{31}}{\partial \ell_{kl}} \\ &\quad + 2(d_{32} - \hat{d}_{32})\frac{\partial d_{32}}{\partial \ell_{kl}} + (d_{33} - \hat{d}_{33})\frac{\partial d_{33}}{\partial \ell_{kl}}\Big], \end{aligned} \tag{22}$$

where $\{d_{ij}\}_{i=1,j=1}^{3}$ denotes the elements of the matrix D. We differentiate the regularization functional in (14). Define the total variation norm of a matrix $D \in R^3 \times R^3$ as

$$\begin{aligned}\mathrm{TV}[D] = \Big(&\mathrm{TV}[d_{11}(\ell_{ij})]^2 + 2\mathrm{TV}[d_{21}(\ell_{ij})]^2 \\ &+ \mathrm{TV}[d_{22}(\ell_{ij})]^2 + 2\mathrm{TV}[d_{31}(\ell_{ij})]^2 \\ &+ 2\mathrm{TV}[d_{32}(\ell_{ij})]^2 + \mathrm{TV}[d_{33}(\ell_{ij})]^2\Big)^{\frac{1}{2}}. \end{aligned} \tag{23}$$

This is a straightforward generalization of the total variation norm of a vector [25].

Using the chain rule, we find the derivatives of the regularization functional

$$\frac{\partial R}{\partial \ell_{kl}} = -\sum_{ij} \alpha_{ij} \nabla \cdot \left(\frac{\nabla d_{ij}}{|\nabla d_{ij}|} \right) \frac{\partial d_{ij}}{\partial \ell_{kl}}, \tag{24}$$

with

$$\alpha_{ij} = \frac{\mathrm{TV}[d_{ij}]}{\mathrm{TV}[D]}. \tag{25}$$

Note here that this derivative is essentially similar to the derivative in the color TV model of Blomgren and Chan [25], but with the important difference that we represent the diffusion matrix by its Cholesky factors.

In the next section, we perform numerical simulations using the method proposed in this paper. We give more details on the Euler-Lagrange equations in the Appendix, which also contains some details on the numerical implementation of the model.

III. Numerical Experiments

In this section we perform numerical experiments on synthetically constructed tensor fields and real tensor fields from a human brain. The numerical implementation of the method is briefly discussed in the Appendix. For the synthetical fields, we have constructed clean tensor fields which are degraded

with noise with a prior known distribution. Thus we are able to measure how well the method performs on synthetical data. For the real human brain DTI the "true" solution is of course not known in advance. In this case we measure the performance of the method in terms of a reference solution where a large series of acquisitions are averaged. This is explained in detail in section III-C. For the numerical implementation of the method and some of the visualizations we have used Matlab [28].

A. Synthetical Tensor Fields

In the first numerical experiment, displayed in Fig. 2, we test the performance of the proposed method on a simple tensor field. This field is mapping a square domain $\Omega \subset R^2$, with four distinct regions, to $R^{2\times 2}$. We construct the clean tensor valued data by prescribing the eigenvalues and corresponding eigenvectors. The values of each element of L is in the range $[0, 1]$. Then we add normally distributed noise $\eta(\sigma)$ to each element of the Cholesky factorization of the matrix, that is $\hat{L} = L + \eta(\sigma)$. Finally, the degraded diffusion tensor is constructed by $\hat{D} = \hat{L}\hat{L}^T$. The noise levels in the simulations in Figs. 2 and 4 are given by $\sigma = 0.35$ and $\sigma = 0.25$ respectively. The time step is $\Delta t = 0.001$. Notice that the discontinuities in the data are preserved in the solution, i.e. the edge preserving property of scalar and vector valued TV flow is kept in the proposed matrix valued flow. In the first example, the diffusion is anisotropic in the whole domain. To show how the proposed method differentiate between isotropic and anisotropic regions, we show a similar example, where one of the four regions is interchanged with an isotropic region. The isotropic region is constructed by considering the orthogonal matrix Q from the QR factorization of a random 2×2 matrix. The columns of the matrix Q are considered to be the eigenvectors of the diffusion tensor. We specify two random numbers in the range $[0, 1]$ as the eigenvalues of the tensor. Thus the diffusion is random in the isotropic region. As we can observe from these two numerical examples on synthetical data edges are preserved in the regularized images. We observe that even when the noise level is as high as 35% we are able to reconstruct an image which is close to the true noise-free image.

From these numerical experiments on synthetical data we see that the proposed method gives encouraging results. Similarly as in the scalar and vector valued setting, edges are well preserved. We furhter investigate the edge preservation in the next experiment.

B. Qualitative experiments

To highlight the qualitative differences between regularizing the elements of the tensor D and the elements of the Cholesky factors L, we have constructed a simple numerical example in 1D. We have removed the fidelity measure from the model, thus the method is in this setting merely a diffusion filter. Thus we have simplified the model in such a way that we can study the qualitative behaviour of the two regularization filters in the same setting. From this example we clearly see that when we regularize D the edges are better preserved than when we regularize L. Note that Wang et.al regularize the Cholesky factors [14].

We also present a numerical example in 2D where we solve the PDEs first as an uncoupled system, i.e. by employing the weighting factors $\alpha_{ij} = 1$, and then as a coupled system where we use the weighting factors from equation 10. We denote the clean field by D, the noisy field by $\hat{D}$, the field regularized with the uncoupled system by D^u and the field regularized with the coupled system by D^c. In Figure 5 (a)-(d) we show the element D_{11},$\hat{D}_{11}$, D^u_{11} and D^c_{11} respectively. Subindexes denotes the elements of the matrix field. Figure 5(e)-(h) shows the element D_{12},$\hat{D}_{12}$, D^u_{12} and D^c_{12}, while Figure (i)-(l) shows the element D_{22},$\hat{D}_{22}$, D^u_{22} and D^c_{22}. From Figure (a)-(e) we observe that the uncoupled system does not discriminate the noise from the weak signal. The coupled system on the other hand better discriminates the noise from the weak signal. A similar 1D example is shown by Blomgren and Chan using the color TV model [25].

In the next section we go one step further, and process real human brain DTMRI.

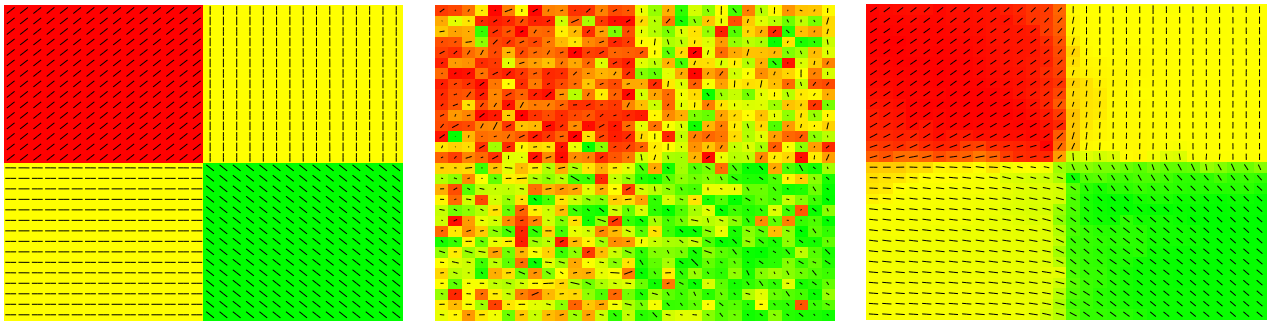
C. Human Brain DTMRI

We also perform numerical experiments on DTMRI acquisitions of a healthy human brain from a volunteer. The human subject data is acquired using a 3.0 T scanner (Magnetom Trio®, Siemens Medical Solutions, Erlangen, Germany) with a 8-element head coil array and a gradient subsystem with the maximum gradient strength of 40 mT/m and maximum slew rate of 200 mT/m/ms. The DTI data is based on spin-echo single shot EPI acquired utilizing generalized auto calibrating partially parallel acquisitions (GRAPPA) technique with acceleration factor of 2 and 64 reference lines. The DTI acquisition consists of one baseline EPI and six diffusion weighted images (b-factor of 1000 s/mm2) along following gradient directions: $G_1 = 1/\sqrt{2}[1, 0, 1]^T$, $G_2 = 1/\sqrt{2}[-1, 0, 1]^T$, $G_3 = 1/\sqrt{2}[0, 1, 1]^T$, $G_4 = 1/\sqrt{2}[0, 1, -1]^T$, $G_5 = 1/\sqrt{2}[1, 1, 0]^T$, $G_6 = 1/\sqrt{2}[1, 1, 0]^T$. Each acquisition has the following parameters: TE/TR/averages is 91 ms/10000 ms/2, FOV is 256 mm x 256 mm, slice thickness/gap is 2 mm/0 mm, acquisition matrix is 192 x 192 pixels and partial Fourier encoding is 75%.

For validation of the proposed regularization method on real data, we construct a reference solution D^* by averaging 18 replications. Each replication consists of six direction weighted and one non-weighted acquisition. This reference solution is compared to solutions where averages of two, four and six replications are post-processed with the proposed regularization method. As a measure of the distance between the reference solution and the processed solution, we use the following metric

$$\begin{aligned} m(D, D^*) &= \Big([d_{11} - d^*_{11}]^2 + 2[d_{12} - D^*_{12}]^2 \\ &+ 2[d_{13} - d^*_{13}]^2 + [d_{22} - d^*_{22}]^2 \\ &+ 2[d_{23} - d^*_{23}]^2 + [d_{33} - d^*_{33}]^2 \Big)^{1/2} \end{aligned} \quad (26)$$

For each simulation, we report the regularization parameter λ, and the metric $m(\cdot, \cdot)$ in Table I, and in Fig. III-C. We display the result before and after applying the proposed method on



(a) (b) (c)

Fig. 2. A synthetically produced purely anisotropic tensor field with four distinct regions is degraded with normally distributed noise. The noisy field is then processed with our proposed method. (a) The clean vector field D_0. (b) The noisy field $\hat{D} = D_0 + \eta$. (c) The recovered field D.

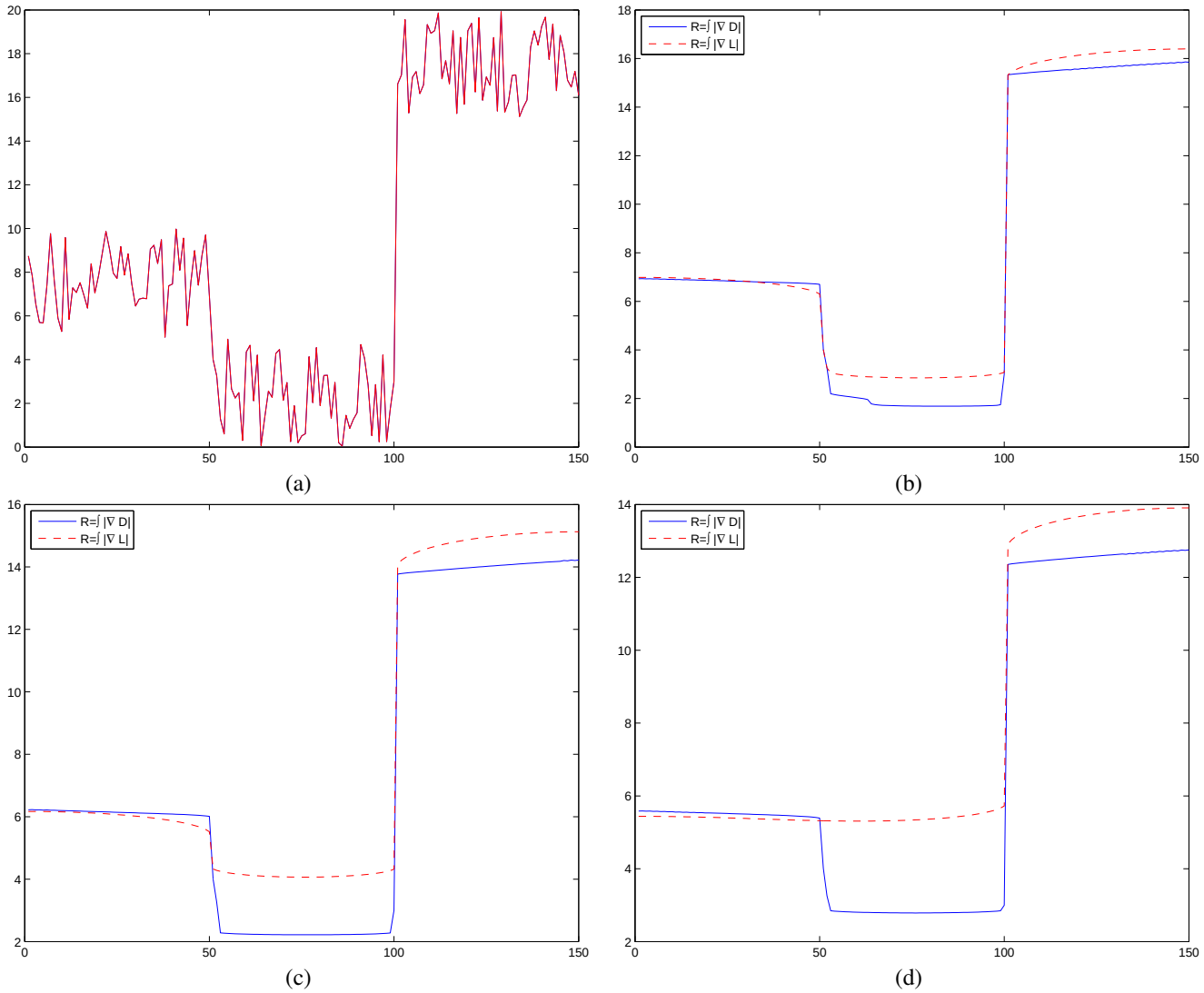
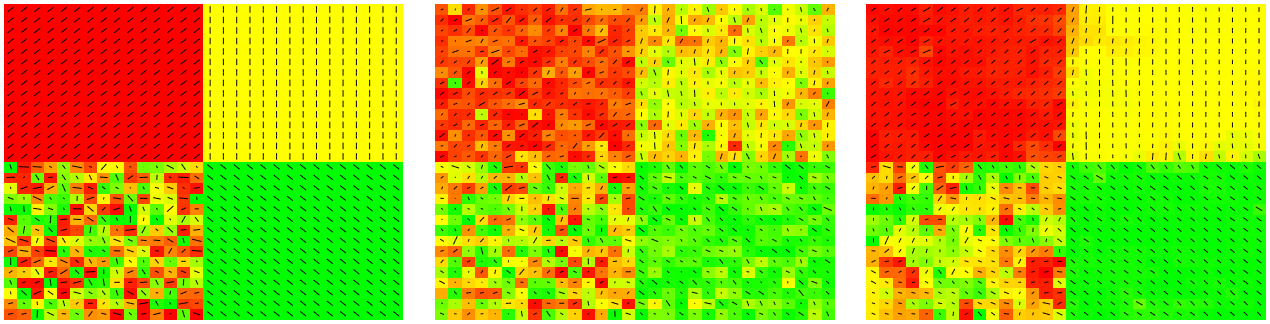


Fig. 3. A simple 1D example showing the qualitative behaviour of the model for regularizers $\int_\Omega |\nabla L|$ and $\int_\Omega |\nabla D|$. The noisy signal in (a) is processed with both flows. Figures (b), (c) and (d) are snapshots during the flow at the three times $t = 8, t = 16$ and $t = 24$.



(a) (b) (c)

Fig. 4. Visualization of the true vector field (a), the noisy field (b), and the recovered field (c). In this example, the tensor field is isotropic in the lower left corner, anisotropic in the other parts.

Replications	λ	reg $m(D, D^*)$	non reg $m(D, D^*)$
1	9	136,1	208,3
2	13	113,5	154
3	19	84,8	105,6

TABLE I

THE DISTANCE $m(D, D^*)$ OF THE REGULARIZED AND THE NON-REGULARIZED TENSOR FIELDS FROM THE NUMERICAL EXAMPLES SHOWN IN FIGS. 7 AND 8.

real DTMRI data in Figs. 7 and 8. In the figures we display a 2D slice of a (RGB direction encoded) fractional anisotropy (FA) measure defined by

$$FA = \sqrt{\frac{3}{2}\frac{(\bar{\lambda}-\lambda_1)^2+(\bar{\lambda}-\lambda_2)^2+(\bar{\lambda}-\lambda_3)^2}{\lambda_1^2+\lambda_2^2+\lambda_3^2}}, \qquad (27)$$

where $\bar{\lambda} = (\lambda_1 + \lambda_2 + \lambda_3)/3$. The FA measure is direction encoded as described by Pajevic and Pierpaoli [29]. We use the DTMRI software DTIStudio to construct the visualizations [30]. In the figures we show the color coded FA.

The noise level is different for each simulation due to the varying number of acquisitions. Consequently, the regularization parameter λ is different for each simulation in this paper. However, for clinical applications, the regularization parameter is estimated once for each imaging protocol. When this is done, the same regularization parameter can be used for subsequent applications of the same imaging protocol.

D. Human Brain ROI study

Since our algorithm regularizes the tensor field, we focus on the evaluation of the tensor field, and the derived scalar FA map. However, we note that from the prosessed tensor field we may reconstruct the corresponding diffusion weighted images $\{S_i\}_{i=1}^6$ by equation 2. There are obvious visual improvements in the processed diffusion weighted images compared to the noisy diffusion weighted images. Edges are preserved and noise is suppressed. Quantitatively the mean and standard deviation at certain homogeneous regions of interests (ROI) show significant improvements. We will now assess the visual and quantitative improvements in terms of the denoised tensors.

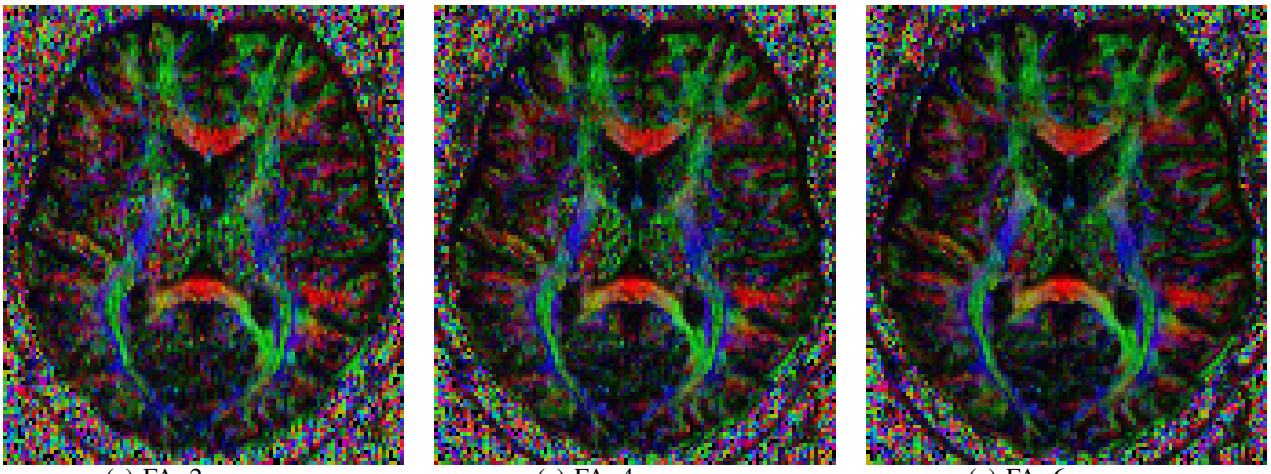
For qualitative evaluation, we select three regions-of-interest (ROI) from one slice of the aquired images, with a 15 by 15 voxel size. We plot the 2D projection of the eigenvector corresponding to the major eigenvalue in Figure 9. From Figure 9 we can clearly see that our method preserves discontinuities (edges) in the tensor field, while it smooths the tensor field in homogenous regions. The denoised tensor field from the 2-average aquisition is close to the tensor field obtained from the 18-average aquisition.

For quantitative measures, we use the average deviation angle (ADA) index of Chen and Hsu to measure if the tensor field undergo gradual changes or sharp turns [17]. The PDE filtering is performed after the tensors are computed, so we use the angle deviation in adjacent voxels as a measure of its performance instead of normalized magnitude of diffusion tensor error (NMTE) index [17]. Denote the eigenvector corresponding to the largest eigenvalue by V^*. Define the ADA by

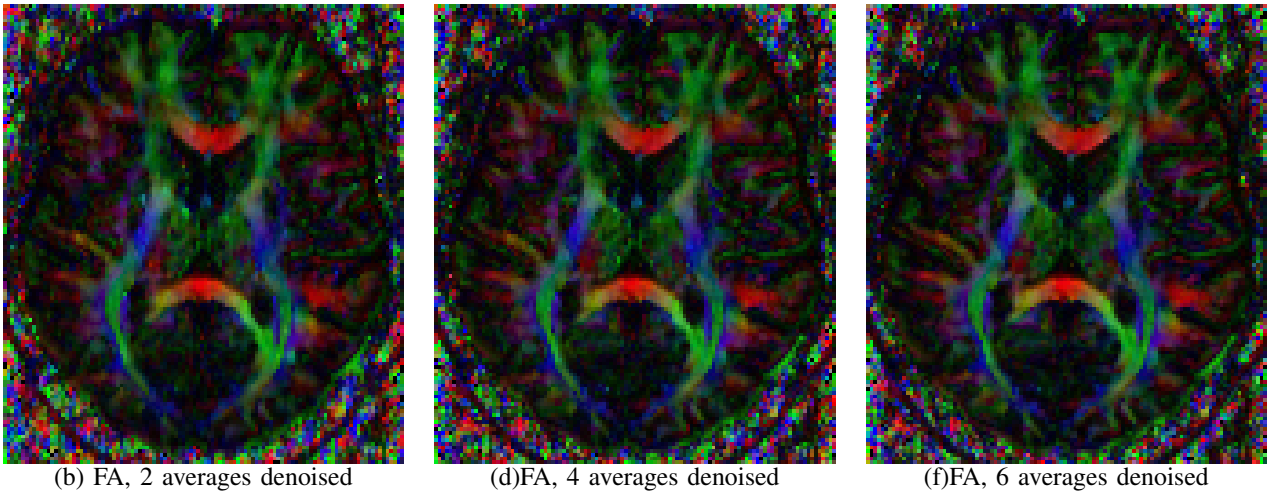
$$ADA = \frac{\Delta\alpha_{i-1}+\Delta\alpha_{i+1}+\Delta\alpha_{j-1}+\Delta\alpha_{j+1}+\Delta\alpha_{k-1}+\Delta\alpha_{k+1}}{6}, \qquad (28)$$

where e.g. $\Delta\alpha_{i-1} = \cos^{-1}\left(|(V^*_{ijk}, V^*_{i-1jk})|\right)$. We note that we use the absolute value of the innerproduct $(\cdot,\cdot)$ to accommodate anti-sense directional vectors. A small change in direction from one voxel to its neighbour gives a small ADA, while a large change in direction gives a large ADA.

After masking the background, we compute the average ADA within the brain, and call it the global ADA. From Table II, we see that the global ADA of the data is reduced from 12.31 to 6.27 by the denoising algorithm, whereas the 18-average clean data has an ADA of 6.65. With a higher data fidelity requirement (when λ is larger, e.g. 20), the smoothing is not very aggressive and the ADA is not as close as when $\lambda = 14$. When λ is less than 14 (data not shown here), the smoothing is excessive and the ADA values fall well below the ADA of the 18-average data. From this information we conclude that for the current aquisition data, $\lambda = 14$ is the best choice. The ADA at selected ROIs is larger than the globel ADA because in those regions, there are obvious edges that contributed to the relatively large ADA values. Compared with the noisy 2-average data, the denoised data show significant improvements. Using the regularization parameter $\lambda = 14$, the

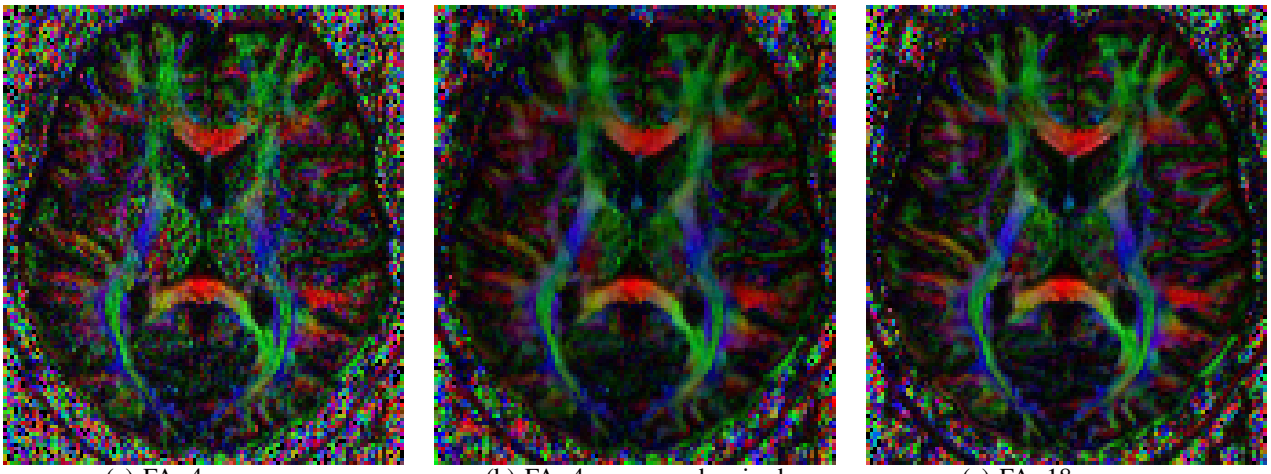


(a) FA, 2 averages (c) FA, 4 averages (e) FA, 6 averages



(b) FA, 2 averages denoised (d)FA, 4 averages denoised (f)FA, 6 averages denoised

Fig. 7. Color coded fractional anisotropy (FA) maps constructed from averages of two (a), four (c) and six (e) acquisitions, and the corresponding denoised maps (b), (d) and (f).



(a) FA, 4 averages (b) FA, 4 averages denoised (c) FA, 18 averages

Fig. 8. The noisy 4 average acquisition (a) is compared with the denoised acquisition (b) and a reference solution at 18 averages.

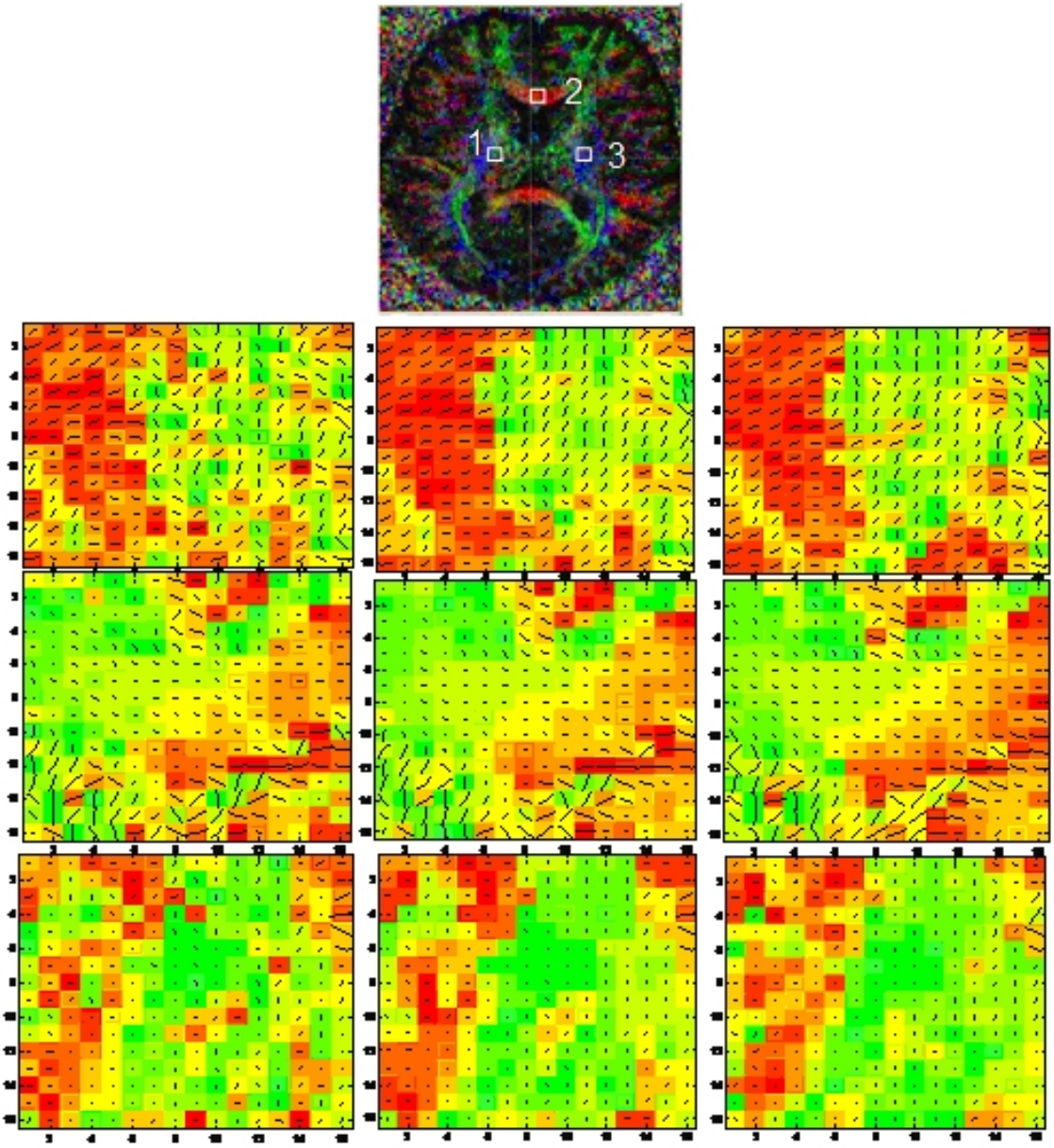


Fig. 9. ROI study. Top image shows the ROIs that we use. The second row from left to right: the noisy (2-average) data, denoised data with $\lambda = 14$, and clean (18-average) data of ROI 1. The third row from left to right: the noisy (2-average) data, denoised data with $\lambda = 14$, and clean (18-average) data of ROI 2. The fourth row from left to right: the noisy (2-average) data, denoised data with $\lambda = 14$, and clean (18-average) data of ROI 3.

ADA is close to the ADA of the 18 average data. The ADAs of all the ROIs are however reduced compared to the noisy data.

IV. Conclusion

In this work we have generalized the color TV regularization method of Blomgren and Chan [25] to yield a novel structure preserving regularization method for matrix valued images. We have shown that the proposed method performs well as a regularization method with the important property of preserving both edges in the data and positive definiteness of the diffusion tensor. Numerical experiments on synthetically produced data and real data from DTI of a human brain of a healthy volunteer indicate good performance of the proposed method.

Acknowledgements

We would like to thank Dr. Siamak Ardekani at UCLA for kindly providing the DTMRI data used in this project. [INCLUDE DETAILS ON FUNDING]

References

[1] 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, 1990.

[2] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithm," *Physica D.*, vol. 60, pp. 259–268, 1992.

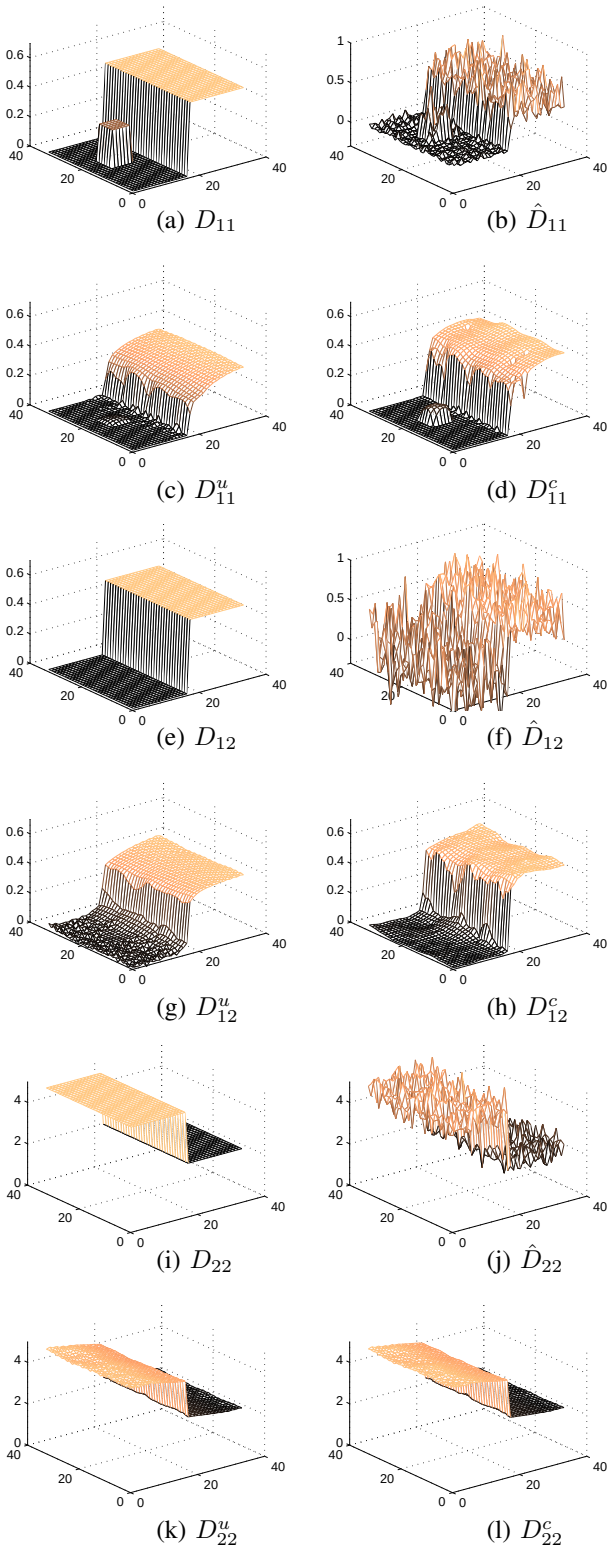


Fig. 5. A noisy 2D tensor field is regularized. In this example, the smallest parts of the signal is not easily discriminated from the noise.

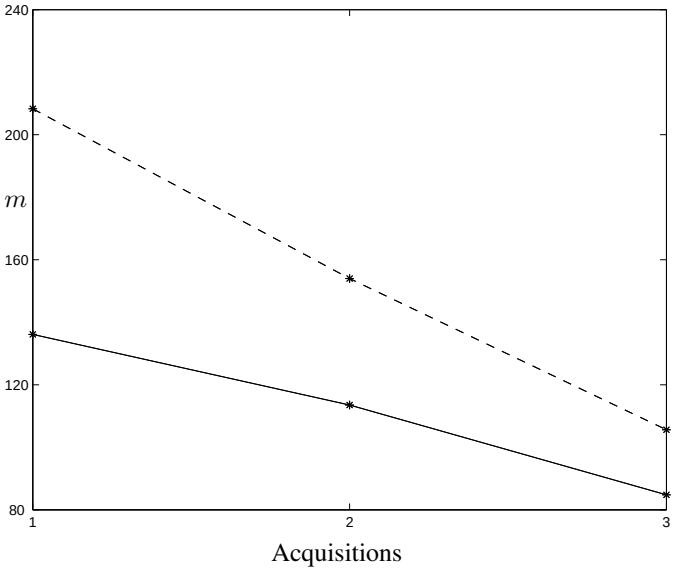


Fig. 6. Comparison of $m(D, D^*)$ for the original tensors (dashed) and the regularized tensors (solid), versus the number of averaged acquisitions.

Data(↓) ADA (→)	global	ROI 1	ROI 2	ROI 3
Noisy (2-avg.)	12.32	32.92	41.02	42.87
Denoised,$\lambda = 14$	6.27	11.77	31.50	25.27
Denoised $\lambda = 20$	7.58	13.34	32.88	28.86
Clean image (18-avg.)	6.65	18.23	24.80	24.80

TABLE II

THE AVERAGE DEVIATION ANGLE (ADA) OF THE NOISY DATA, THE PROCESSED DATA (TWO DIFFERENT REGULARIZATION PARAMETERS) AND THE REFERENCE DATA.

[3] M. Lysaker, S. Osher, and X.-C. Tai, "Noise removal using smoothed normals and surface fitting," *IEEE Trans. Image Processing*, vol. 13, no. 10, pp. 1345–1357, 2004.

[4] T. F. Chan and S. Esedoglu, "Aspects of total variation regularized L_1 function approximation," *SIAM Journal of Applied Mathematics*, vol. 65, no. 5, pp. 1817–1837, 2005.

[5] T. F. Chan and J. Shen, *Image Processing And Analysis: Variational, PDE, Wavelet, And Stochastic Methods*. SIAM, 2005.

[6] J. Weickert, "A review of nonlinear diffusion filtering," *Lecure Notes in Computer Science*, vol. 1252, pp. 3 – 28, 1997, proceedings of the First International Conference on Scale-Space Theory in Computer Vision.

[7] J. Weickert and T. Brox, "Diffusion and regularization of vector- and matrix-valued images," Universitat des Saarlandes, Fachrichtung 6.1 Mathematik, Preprint No. 58, 2002.

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

[9] D. L. Bihan, J.-F. Mangin, C. Poupon, C. A. Clark, S. Pappata, N. Molko, and H. Chabriat, "Diffusion tensor imaging: Concepts and applications," *Journal of Magnetic Resonance Imaging*, vol. 13, no. 4, pp. 534 – 546, 2001.

[10] C. Westin, S. Maier, H. Mamata, A. Nabavi, F. Jolesz, and R. Kikinis, "Processing and visualization for diffusion tensor mri," *Medical Image Analysis*, vol. 6, pp. 93–108, 2002.

[11] S. Mori and P. B. Barker, "Diffusion magnetic resonance imaging: Its principle and applications," *The Anatomical Record (NEW ANAT.)*, vol. 257, pp. 102–109, 1999, feature Article.

[12] S. Mori and P. C. van Zijl, "Fiber tracking: principles and strategies - a techical review," *NMR Biomed*, vol. 15, pp. 468–480, 2002.

[13] R. Bammer, "Basic principles of diffusion-weighted imaging," *European Journal of Radiology*, vol. 45, pp. 169–184, 2003.

[14] Z. Wang, B. Vemuri, and Y. C. T. Mareci, "A constrained variational principle for direct estimation and smoothing of the diffusion tensor field

from complex dwi," *IEEE Transactions on Medical Imaging*, vol. 23, no. 8, pp. 930 – 939, 2004.

[15] D. Tschumperle and R. Deriche, "Variational frameworks for DT-MRI estimation,regularization and visualization," in *Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV 2003)*, 2003.

[16] C.-F. Westin and H. Knutsson, "Tensor field regularization using normalized convolution," in *Computer Aided Systems Theory (EUROCAST'03), LNCS 2809*, R. M. Diaz and A. Q. Arencibia, Eds. Las Palmas de Gran Canaria, Spain: Springer Verlag, February 24–28 2003, pp. 564–572.

[17] B. Chen and E. W. Hsu, "Noise removal in magnetic resonance diffusion tensor imaging," *Magnetic Resonance in Medicine*, vol. 54, pp. 393–407, 2005.

[18] E. O. Stejskal and J. E. Tanner, "Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient," *The Journal of Chemical Physics*, vol. 42, no. 1, pp. 288–292, 1965.

[19] E. O. Stejskal, "Use of spin echoes in a pulsed magnetic-field gradient to study anisotropic, restricted diffusion and flow," *The Journal of Chemical Physics*, vol. 43, no. 10, pp. 3597–3603, 1965.

[20] P. Tofts, Ed., *Quantitative MRI of the Brain*. John Wiley and Sons, 2005.

[21] S. E. and Maier, H. Mamata, A. Nabavi, and and, "Processing and visualization for diffusion tensor mri," *Medical Image Analysis*, vol. 6, pp. 93–108, 2002.

[22] C. Chefd'hotel, D. Tchumperlé, R. Deriche, and O. Faugeras, "Regularizing flows for constrained matrix-valued images," *Journal of Mathematical Imaging and Vision*, vol. 20, pp. 147–162, 2004.

[23] Y. Gur and N. Sochen, "Denoising tensors via Lie group flow," in *Variational, Geometric, and Level Set Methods in Computer Vision: Third International Workshop, VLSM 2005, Beijing, China, October 16, 2005. Proceedings*, ser. Lecture Notes in Computer Science, vol. 3752. Springer, 2005, pp. 13–24.

[24] D. Groisser, "Some differential-geometric remarks on a method for minimizing constrained functionals of matrix-valued functions," *Journal of Mathematical Imaging and Vision*, 2006.

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

[26] T. F. Chan, G. H. Golub, and P. Mulet, "A nonlinear primal-dual method for total variation-based image restoration," *SIAM J. Sci. Comput.*, vol. 20, no. 6, 1999.

[27] G. Sapiro, "Color snakes," Technical Report HPL-95-113, Hewlett Packard Computer Peripherals Laboratory, 1995.

[28] The Mathworks, "MatLab, The Language of Technical Computing." [Online]. Available: www.mathworks.com/matlab

[29] S. Pajevic and C. Pierpaoli, "Color schemes to represent the orientation of anisotropic tissues from diffusion tensor data: Application to white matter fiber tract mapping in the human brain," *Magnetic Resonance in Medicine*, vol. 42, pp. 526–540, 1999.

[30] Mori and coworkers, "Dti-studio." [Online]. Available: http://cmrm.med.jhmi.edu/

APPENDIX: EULER-LAGRANGE EQUATION, AND NUMERICAL IMPLEMENTATION

In this appendix we explicitly write out the Euler-Lagrange equations corresponding to the minimization functional 14. In addition, the numerical scheme used in the simulations in section III of this paper is briefly discussed.
Using the short hand notation

$$p(x_{ij}) = \alpha_{ij}\nabla \cdot \left(\frac{\nabla x_{ij}}{|\nabla x_{ij}|}\right), \tag{29}$$

we can write out the derivatives of R and F as

$$\begin{aligned}
\frac{\partial R}{\partial \ell_{11}} &= 2\Big(\ell_{11}p(\ell_{11}^2) + \ell_{21}p(\ell_{11}\ell_{21}) + \ell_{31}p(\ell_{11}\ell_{31})\Big) && (30)\\
\frac{\partial R}{\partial \ell_{21}} &= 2\Big(\ell_{11}p(\ell_{11}\ell_{21}) + \ell_{21}p(\ell_{21}^2 + \ell_{22}^2) + \ell_{31}p(\ell_{21}\ell_{31} + \ell_{22}\ell_{32})\Big) && (31)\\
\frac{\partial R}{\partial \ell_{22}} &= 2\Big(\ell_{22}p(\ell_{21}^2 + \ell_{22}^2) + \ell_{32}p(\ell_{21}\ell_{31} + \ell_{22}\ell_{32})\Big) && (32)\\
\frac{\partial R}{\partial \ell_{31}} &= 2\Big(\ell_{11}p(\ell_{11}\ell_{31}) + \ell_{21}p(\ell_{21}\ell_{31} + \ell_{22}\ell_{32}) + \ell_{31}p(\ell_{31}^2 + \ell_{32}^2 + \ell_{33}^2)\Big) && (33)\\
\frac{\partial R}{\partial \ell_{32}} &= 2\Big(\ell_{22}p(\ell_{21}\ell_{31} + \ell_{22}\ell_{32}) + \ell_{32}p(\ell_{31}^2 + \ell_{32}^2 + \ell_{33}^2)\Big) && (34)\\
\frac{\partial R}{\partial \ell_{33}} &= \Big(2\ell_{33}p(\ell_{31}^2 + \ell_{32}^2 + \ell_{33}^2)\Big) && (35)\\
\frac{\partial F}{\partial \ell_{11}} &= 4\Big[(d_{11} - \hat{d}_{11})\ell_{11} + (d_{21} - \hat{d}_{21})\ell_{21} + (d_{31} - \hat{d}_{31})\ell_{31}\Big] && (36)\\
\frac{\partial F}{\partial \ell_{21}} &= 4\Big[(d_{21} - \hat{d}_{21})\ell_{11} + (d_{22} - \hat{d}_{22})\ell_{21} + (d_{32} - \hat{d}_{32})\ell_{31}\Big] && (37)\\
\frac{\partial F}{\partial \ell_{22}} &= 4\Big[(d_{22} - \hat{d}_{22})\ell_{22} + (d_{32} - \hat{d}_{32})\ell_{32}\Big] && (38)\\
\frac{\partial F}{\partial \ell_{32}} &= 4\Big[(d_{32} - \hat{d}_{32})\ell_{22} + (d_{33} - \hat{d}_{33})\ell_{32}\Big] && (39)\\
\frac{\partial F}{\partial \ell_{31}} &= 4\Big[(d_{31} - \hat{d}_{31})\ell_{11} + (d_{32} - \hat{d}_{32})\ell_{21} + (d_{33} - \hat{d}_{33})\ell_{31}\Big] && (40)\\
\frac{\partial F}{\partial \ell_{33}} &= 4\Big[(d_{33} - \hat{d}_{33})\ell_{33}\Big] && (41)
\end{aligned}$$

By combining each of the equations

$$\frac{\partial G}{\partial \ell_{ij}} = \frac{\partial R}{\partial \ell_{ij}} + \frac{\partial F}{\partial \ell_{ij}}, \qquad \{ij\} \in \{11, 21, 22, 31, 32, 33\}, \tag{42}$$

we arrive at the Euler-Lagrange equations corresponding to the minimization problem (14). In the numerical simulations, we use the steepest descent method with a fixed time step. Thus we have the six equations

$$d_{ij}^{n+1} = d_{ij}^{n} - \Delta t \frac{\partial G^n}{\partial \ell_{ij}}, \qquad \{kl\} \in \{11, 21, 22, 31, 32, 33\}, \tag{43}$$

where n is the iteration index, and Δt is the time step parameter. We approximate the gradient $\frac{\partial G}{\partial \ell_{ij}}$ by standard finite difference schemes, see e.g. [4]. We here note that each iteration of the form (43) is performed sequentially. Thus the equations are solved as a coupled system of six PDEs.