Augmented Lagrangian Regularization of Matrix-Valued Images

Guy Rosman, Yu Wang, Xue-Cheng Tai, Ron Kimmel and Alfred M. Bruckstein

October 3, 2011

Abstract

Regularization of matrix-valued data is of important in medical imaging, motion analysis and scene understanding. Here we propose a novel method for efficient regularization of matrix group-valued images.

Using the augmented Lagrangian framework we separate the total-variation regularization of matrix-valued images into a regularization and projection steps, both of which are fast and parallelizable. We demonstrate the effectiveness of our method for denoising of several group-valued image types, with data in $SO(n)$, $SE(n)$, and $SPD(n)$.

1 Introduction

Matrix Lie-group data, and specifically matrix-valued images have become an integral part of computer vision and image processing. Whether in image processing [37, 21, 44, 8, 11, 27], tracking [40] and motion analysis [23], $3D$ reconstruction [7], surface segmentation [29] or more general optimization research [45], the importance of accurate and efficient handling of matrix manifolds and matrix-valued maps is obvious.

Here, we present an augmented Lagrangian method for regularizing maps from a Cartesian domain onto matrix manifolds such as $SO(n)$, $SE(n)$ and $SPD(n)$, the manifolds of special-orthogonal, special-Euclidean, and symmatric, positive-definite, matrices, respectively. Specifically, the data we regularize can be represented as matrices with constraints on their singular values or eigenvalues. The augmented Lagrangian technique allows us in such cases to separate the optimization process into a *total-variation* (TV, [30]) regularization step and an eigenvalue or singular value projection step, both of which are fast and easily parallelizable using consumer graphic processing units (GPUs).

One motivation for this exploration is that the proposed method, combined with current fast optimization methods for TV regularization on surfaces [22, 48], lends itself to fast methods for segmentation of articulated objects [29]. Yet another possible application we explore, in the case of $SPD(n)$, lies in DT-MRI signal denoising and reconstruction.

We suggest treating each constraint separately as proposed, for example, in [42], using a different treatment of the constraint. This results in a unified framework for $SO(n)$, $SE(n)$ and $SPD(n)$, as we describe in Sections 3 and 4. Section 6 suggests an additional regularizing functional for regularization of SPD matrices often encountered in medical DT-MRI data. This is followed by a short discussion of convergence

properties of the suggested algorithms in Section 5. In Section 7 we demonstrate a few results of our method, and Section 8 concludes the paper.

2 A Short Introduction to Lie-Groups

Lie-groups are groups endowed with a differentiable manifold structure, and generated in a continuous manner. Treating correctly the structure of Lie-group data in computer vision has been the subject of intense research effort, aspecially involving statistics of matrix-valued data [27], and regularizing it [38, 16], as well as describing the evolution of differential processes with Lie-group data [20]. We give a short introduction to Lie-groups in this section and refer the reader to the literature for an in-depth discussion [17, 35].

Lie-groups elements can be diffeomorphically mapped, by group action with their inverse, to the origin. The tangent space in the origin therefore defines a canonical way of parameterizing small changes of the manifold elements via a vector space. Such a vector space is known as the *Lie-algebra* corresponding to the Lie-group. Lie-algebras are equipped with an anti-symmetric bilinear operator, the *Lie-bracket*, that describes the non-commutative part of the group product. Lie-brackets are used in tracking [4], robotics, and computer vision [26], among other applications.

We deal with three Lie-groups,

The rotations group $SO(n)$ - The group $SO(n)$ describes all rotations of the n-dimensional Euclidean space. Elements of this group can be described in a matrix form

$$SO(n) = \left\{ \mathbf{R} \in \mathbb{R}_{n\times n}, \mathbf{R}^T\mathbf{R} = \mathbf{I}, \det(\mathbf{R}) = 1 \right\}, \tag{1}$$

with the group product being matrix multiplication. The Lie-algebra of this group is the space $so(n)$, which can be described by the set of skew-symmatric matrices,

$$so(n) = \left\{ \mathbf{A} \in \mathbb{R}_{n\times n}, \mathbf{A}^T = -\mathbf{A} \right\}, \tag{2}$$

The special-Euclidean group $SE(n)$ - This group represents rigid transformations of the n-dimensional Euclidean space. This group can be thought of as the product manifold of $SO(n)$ and the manifold $\mathbb{R}^n$ describing all translations of the Euclidean space. In matrix form this group can be written as

$$SE(n) = \left\{ \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{pmatrix}, \mathbf{R} \in SO(n), \mathbf{t} \in \mathbb{R}^n \right\}. \tag{3}$$

The Lie-algebra of this group can be written as

$$se(n) = \left\{ \begin{pmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{0} & 1 \end{pmatrix}, \mathbf{A} \in so(n), \mathbf{t} \in \mathbb{R}^n \right\}, \tag{4}$$

The symmetric positive definite group $SPD(n)$ - This group is the group of symmetric positive definite matrices. This group has been studied extensively in control theory [12], as well as in the context of diffusion tensor images [27], where the matrices are used to describe the diffusion coefficients along each direction. By definition, this group is given in matrix form as

$$SPD(n) = \left\{ \mathbf{A} \in \mathbb{R}_{n\times n}, \mathbf{A} \succeq 0 \right\}. \tag{5}$$

Its Lie-algebra consists of the group of symmetric matrices [13],

$$S(n) = \left\{ \mathbf{A} \in \mathbb{R}_{n\times n}, \mathbf{A} = \mathbf{A}^T \right\}. \tag{6}$$

3 An Augmented Lagrangian Regularization of Maps onto $SO(n)$,$SE(n)$

The optimization problem we consider is

$$\underset{u\in\mathcal{G}}{\operatorname{argmin}} \int \|u-u_0\|^2 + \lambda\|u^{-1}\nabla u\|dx, \tag{7}$$

where u represents an element in an embedding of the Lie-group $\mathcal{G}$ into Euclidean space, specifically for the groups $SO(n), SE(n), SPD(n)$. Elements of $SO(n)$ can be embedded into $\mathbb{R}^{n^2}$, and elements of $SE(n)$ can similarly be embedded into $\mathbb{R}^{(n+1)^2}$, or more precisely, an $n(n+1)$-dimensional linear subspace of $\mathbb{R}^{(n+1)^2}$. The elements of $SPD(n)$ can be embedded into $\mathbb{R}^{n(n+1)/2}$. We first relate our method in the context of $\mathcal{G} = SO(n)$, and then detail the differences required when $\mathcal{G} = SE(n)$ and $\mathcal{G} = SPD(n)$.

For brevity's sake, we use the same notation for representation of the Lie-group element, its matrix representation, and its embedding onto the embedding space, as specified in each case we explore.

The term $\|u^{-1}\nabla u\|$ can be thought of as a regularization term placed on elements of the Lie algebra about each point (pixel). Such a regularization may be relevant for motion analysis [23], or generalized to regularization on surfaces [22, 48], leading to motion segmentation on surfaces [29].

One possibility for a fast regularization scheme is to use the augmented Lagrangian technique and a dual field v, further constrained to lie in $SO(n)$,

$$\underset{u\,\in\, SO(n)}{\operatorname{argmin}} \int \|u-u_0\|^2 + \lambda\|\nabla u\|dx. \tag{8}$$

Instead of restricting u to $SO(n)$, we add an auxiliary variable, v, at each point, such that $u = v$ and restrict v to $SO(n)$, where the equality constraint is enforced via an augmented Lagrangian term. The suggested augmented Lagrangian minimization now reads

$$\begin{aligned}&\underset{v\in SO(n)}{\operatorname{argmin}}\ \mathcal{L}(u,v) = \\ &\underset{v\in SO(n)}{\operatorname{argmin}} \int \|u-u_0\|^2 + \lambda\|\nabla u\| + \frac{r}{2}\|v-u\|^2 + \langle\mu, v-u\rangle dx.\end{aligned} \tag{9}$$

The rationale behind the different regularization term $\|\nabla u\|$ can be seen by observing the norm of each canonical directional derivative. Denote by u_i vectors in $\mathbb{R}^n$ representing the columns of the rotation matrix $u(x)$.

$$\begin{aligned}\left\|\frac{\partial}{\partial x_i}u\right\|_F^2 &= \sum_{j=1}^{n}\left\|\frac{\partial}{\partial x_i}u_j\right\|^2 + \\ \lim_{\epsilon\to 0}\sum_{j=1}^{n}&\left\{\ \tfrac{1}{\epsilon}\left\|u_j(x+\epsilon e_i)-u_j(x)\right\|^2\ \right\},\end{aligned} \tag{10}$$

Multiplying these vector differences by u^{-1} from the left does not change their norm since u is an isometry of $\mathbb{R}^n$. Thus, the extent to which this term represents regularization of maps into $SO(n)$ depends on the assumption that $u(x)$ are in $SO(n)$.

The functional shown in Equation 9 lends itself to an elegant optimization scheme. We split the optimization into an alternating minimization steps with respect to u and v. The minimization w.r.t. v is dependent solely on minimizing a squared distance expression per pixel,

$$\begin{aligned}
&\underset{v\in SO(n)}{\operatorname{argmin}} \frac{r}{2}\|v-u\|^2 + \langle \mu, v-u\rangle = \\
&\underset{v\in SO(n)}{\operatorname{argmin}} \frac{r}{2}\|v\|^2 - r\langle v,u\rangle + \langle v,\mu\rangle + \frac{r}{2}\|u\|^2 - \langle \mu,u\rangle = \\
&\underset{v\in SO(n)}{\operatorname{argmin}} \frac{r}{2}\left\|v-\left(\frac{\mu}{r}+u\right)\right\|^2 - \left\|\frac{\mu}{r}+u\right\|^2 + \frac{r}{2}\left(\|u\|^2 + \frac{2}{r}\langle \mu,u\rangle\right) = \\
&\underset{v\in SO(n)}{\operatorname{argmin}} \frac{r}{2}\left\|v-\left(\frac{\mu}{r}+u\right)\right\|^2
\end{aligned} \tag{11}$$

There are several ways of taking the closest rotation matrix to a given matrix – in our case, $\left(\frac{\mu}{r}+u\right)$. Specifically, an easy solution is often obtained by the singular value decomposition [14, 34],

$$v^{k+1} = \mathbf{U}(x)\mathbf{V}^T(x), \mathbf{U}\mathbf{S}\mathbf{V}^T = \left(\frac{\mu}{r}+u^k\right). \tag{12}$$

other possibilities include using the Euler-Rodrigues formula, or quaternions, (see for example [31]).

Minimization with respect to u is simply a vectorial TV denoising problem

$$\begin{aligned}
\underset{u\in\mathbb{R}^9}{\operatorname{argmin}} \int \|u-u_0\|^2 + \lambda\|\nabla u\| + \frac{r}{2}\|v-u\|^2 + \langle \mu, v-u\rangle dx = \\
\underset{u\in\mathbb{R}^9}{\operatorname{argmin}} \int \left(\frac{r}{2}+1\right)\left\|u - \frac{\mu+2u_0+rv}{r+2}\right\|^2 + \lambda\|\nabla u\| = \\
\underset{u\in\mathbb{R}^9}{\operatorname{argmin}} \int \tilde{\lambda}\left\|u - \frac{\mu+2u_0+rv}{r+2}\right\|^2 + \|\nabla u\|
\end{aligned} \tag{13}$$

which lends itself to fast minimization techniques. Specifically, we use the augmented-Lagrangian TV denoising algorithm suggested by Tai and Wu [36], as we now describe. In order to obtain fast optimization of the problem with respect to u, we add an auxiliary variable p, along with a constraint that $p = \nabla u$, enforcing the constraint via an augmented Lagrangian term. The complete minimization problem now reads

$$\begin{aligned}
&\underset{\substack{u\in\mathbb{R}^9\\ p\in\mathbb{R}^{18}}}{\operatorname{argmin}} \int \tilde{\lambda}\left\|u - \frac{\mu+2u_0+rv}{r+2}\right\|^2 + \|p\| + \mu_2^T(p-\nabla u) + \frac{r_2}{2}\|p-\nabla u\|^2,
\end{aligned} \tag{14}$$

with u updated, for example, in the Fourier domain. The dual field p is updated by rewriting the minimization w.r.t. p as

$$\underset{p\in\mathbb{R}^{18}}{\operatorname{argmin}} \int \|r_2 p\| + \mu_2^T p + \frac{1}{2}\|r_2 p - \nabla u\|^2, \tag{15}$$

with the solution [41, 36]

$$p = \frac{1}{r_2} max \left(1 - \frac{1}{\|w\|}, 0\right) w, w = \nabla u - \frac{\mu_2}{r_2}. \tag{16}$$

Finally, according to the augmented Lagrangian scheme [18, 28], the Lagrange multipliers μ should also be updated, according to

$$\begin{aligned} \mu^k &= \mu^{k-1} + r\left(v^k - u^k\right) \\ \mu_2^k &= \mu_2^{k-1} + r_2\left(p^k - \nabla u^k\right). \end{aligned} \tag{17}$$

An algorithmic description is given as Algorithm 1, whose convergence properties are discussed in Section 5.

Algorithm 1 Fast TV regularization of matrix-valued data

1: **for** $k = 1, 2, \ldots,$ until convergence **do**
2: Update $u^k(x), p^k(x)$, according to Equations (13,16).
3: Update $v^k(x)$, by projection onto $SO(n)$, according to Equation (12), by projection onto $SE(n)$ using Equation 19, or by projection onto $SPD(n)$, using the eigenvalue decomposition Equation 20.
4: Update $\mu^k(x), \mu_2^k(x)$, according to Equation (17).
5: **end for**

3.1 Regularization of maps unto $SE(n)$

We note that the same method can easily be used to regularize images with values in $SE(n)$. Using the same notation as in Equation 9, our minimization problem is

$$\underset{v \in SE(n)}{\operatorname{argmin}} \mathcal{L}(u, v) \tag{18}$$

Optimization w.r.t. u will be the same is in Equation 13. The projection step w.r.t. v applies only hold for the n^2 elements of v describing the rotation matrix, leaving the translation component unchanged.

Specifically, let $v = (v_R, v_t), v_R \in \mathbb{R}^{n^2}, v_t \in \mathbb{R}^n$ denotes the rotation and translation parts of the current solution. Updating v assumes the form

$$\begin{aligned} v_R^{k+1} &= \mathbf{U}(x)\mathbf{V}^T(x), \mathbf{U}\mathbf{S}\mathbf{V}^T = \left(\frac{\mu_R}{r} + u_R^k\right) \\ v_t^{k+1} &= \left(\frac{\mu_t}{r} + u_t^k\right). \end{aligned} \tag{19}$$

4 Fast TV prior for SPD matrices

Yet another usage of the technique described above can be found for symmetric positive-definite matrices. These can be decomposed using the eigenvalue decomposition,

$$A \in SPD(n) \leftrightarrow A = U(\operatorname{diag}(\boldsymbol{\lambda}))U^T, \boldsymbol{\lambda} \geq 0,$$

where the matrix U is a unitary one, representing the eigenvectors of the matrix, and the eigenvalues are all-positive. $SPD(n)$ matrices can be found, for example, in the

diffusion tensors in DT-MRI, for $n = 3$. There have been several attempts for fast DT-MRI regularization. Tschumperlé et al. [10] formulated matrix-valued data regularization and showed its usefulness for DT-MRI denoising. Gur and Sochen [15] suggested to use the Iwasawa coordinates to formulate a simple minimizing flow for group-valued maps. For DT-MRI regularized reconstruction, Wang et al. [43] used a different representation based on Cholseky decomposition of the tensors, and a general limited-memory quasi-Newton solver, and Steidl et al. [32] proposed a scheme based on SOCP. Many papers dealing with the the analysis of DT-MRI rely on the eigenvalue decomposition of the tensor as well, i.e. for tractography [9], fractional anisotrophy measurements [46], and so forth. We also use this decomposition to project the constrained data onto the matrix manifold.

One possibility is to denoise the matrix valued image in a similar fashion to the one shown in Equation 8.

$$\underset{\substack{u \in \mathbb{R}^{3\times 3}, u = v \\ v \in SPD(n)}}{\operatorname{argmin}} \int \|u - u_0\|^2 + \|\nabla u\|.$$

Using an augmented Lagrangian approach, the minimization problem is

$$\underset{v \in SPD(n)}{\operatorname{argmin}} \quad \mathcal{L}(u, v), \tag{20}$$

where optimization is done as proposed in Algorithm 1, but while replacing step (3) with a projection of the eigenvalues of v into the positive domain [19, 6, 3]. Furthermore, the optimization w.r.t. u, v is now over the domain $\mathbb{R}^9 \times SPD(n)$, and the cost function is convex, resulting in a convex optimization problem. An example of using the proposed method for DT-MRI denoising is shown in Section 7.

The convex domain of optimization allows us to formulate a convergence proof for the algorithm, as shown in 5.2.

4.1 Regularized DTI Reconstruction

Another possible use of the proposed fast regularization is to use the prior directly for DT-MRI reconstruction. Instead of adding a fidelity term as in Equation 20, we add a term for fitting the Stejskal-Tanner equations [33], based on a set of measurements describing the diffusion in specific directions, and reconstruction the full diffusion tensor at each voxel. The fitting term can be written as

$$\sum_i \left\| b_i \mathbf{g}_i^T u \mathbf{g}_i - \log\left(\frac{S_i}{S_0}\right) \right\|^2,$$

where b_i and $\mathbf{g}_i$ are the b-values and gradient vectors, u is the diffusion tensor reconstructed at each voxel, and $\frac{S_i}{S_0}$ define the relative signal ratio for each direction at each voxel. The complete minimization problem reads

$$\underset{v \in SPD(n)}{\operatorname{argmin}} \int \sum_i \left\| b_i \mathbf{g}_i^T u \mathbf{g}_i - \log\left(\frac{S_i}{S_0}\right) \right\|^2 + \lambda \|\nabla u\| + \frac{r}{2}\|v - u\|^2 + \langle \mu, v - u \rangle dx. \tag{21}$$

While the memory requirements seems to hinder fast optimization, looking closely at the quadratic penalty data term, we see it can be expressed by looking at a fitting term for the Stejskal-Tanner equations ,

$$\sum_i \left\| b_i \mathbf{g}_i^T u \mathbf{g}_i - \log\left(\frac{S_i}{S_0}\right) \right\|^2 = u^T \mathbf{A} u + \mathbf{b}^T u + c, \tag{22}$$

where $\mathbf{A}$ is a constant matrix over the whole volume,

$$\mathbf{A} = \sum_i b_i^2 \begin{pmatrix} g_1^4 & 2g_1^3 g_2 & 2g_1^3 g_3 & g_1^2 g_2^2 & 2g_1^2 g_2 g_3 & g_1^2 g_3^2 \\ 2g_1^3 g_2 & 4g_1^2 g_2^2 & 4g_1^2 g_2 g_3 & 2g_1 g_2^3 & 4g_1 g_2^2 g_3 & 2g_1 g_2 g_3^2 \\ 2g_1^3 g_3 & 4g_1^2 g_2 g_3 & 4g_1^2 g_3^2 & 2g_1 g_2^2 g_3 & 4g_1 g_2 g_3^2 & 2g_1 g_3^3 \\ g_1^2 g_2^2 & 2g_1 g_2^3 & 2g_1 g_2^2 g_3 & g_2^4 & 2g_2^3 g_3 & g_2^2 g_3^2 \\ 2g_1^2 g_2 g_3 & 4g_1 g_2^2 g_3 & 4g_1 g_2 g_3^2 & 2g_2^3 g_3 & 4g_2^2 g_3^2 & 2g_2 g_3^3 \\ g_1^2 g_3^2 & 2g_1 g_2 g_3^2 & 2g_1 g_3^3 & g_2^2 g_3^2 & 2g_2 g_3^3 & g_3^4 \end{pmatrix} \tag{23}$$

and $\mathbf{b}$ is the vector

$$\mathbf{b} = \sum_i b_i \log\left(\frac{S_i}{S_0}\right) \begin{pmatrix} 2g_1^2 & 4g_1 g_2 & 4g_1 g_3 & 2g_2^2 & 4g_2 g_3 & 2g_3^2 \end{pmatrix}^T, \tag{24}$$

and c is the scalar image

$$c = \sum_i \left(\log\left(\frac{S_i}{S_0}\right)\right)^2. \tag{25}$$

We note that, unlike the denoising case, in the reconstruction case it is the data term that couples together the elements of the tensor together. Care must be taken so as to handle this coupled data term.

Reconstruction with the new data term can be computed using several techniques.

- Freezing all elements of the tensor but one, we obtain from the Euler-Lagrange equations pertaining to Equation 21 an update rule for the image, to be computed in the Fourier domain. While the coupling between the tensor elements (expressed via the non-diagonal matrix $\mathbf{A}$) prevents us from treating each tensor element separately, the optimization w.r.t. each of the elements converges quite rapidly.
- Another possibility is to take a block Gauss-Seidel approach, and optimize each tensor separately, going over all the voxels one-by-one.
- Yet another possibility is to further decouple the TV and data term, using separate variables and constraining them using an augmented Lagrangian approach.

5 Convergence Properties of the Algorithm

We now turn to discuss the local convergence of Algorithm 1.

5.1 Local Convergence for $SO(n)$,$SE(n)$ Regularization

Looking at regularization of maps onto $SO(n)$,$SE(n)$, the non-convex nature of the optimization domain in equations 11,18 makes it difficult to prove global convergence. Furthermore, the nature of the projection operator into $SO(n)$ and $SE(n)$, makes it difficult to ascertain that at some point the sequence of iterants will converge. While showing there exists a converging subsequence of iterants is easy due to the boundedness of the sublevel-sets [39], the discontinuous nature of the projection unto non-convex spaces may cause the algorithm to oscilate, although this behaviour does not appear in practice. In order to avoid such a possibility and allow for an easy proof of convergence, we take a proximal step approach, and slightly modify our algorithm, as suggested by Attouch et al. [1], changing the first two steps of the algorithm into the minimization problems

$$\begin{aligned} u^k &= \underset{u}{\operatorname{argmin}}\, L(u, v^{k-1}, \mu) + \frac{1}{\theta_k}\|u - u_{k-1}\|^2 \\ v^k &= \underset{v\in\mathcal{G}}{\operatorname{argmin}}\, L(u^k, v, \mu) + \frac{1}{\theta_k}\|v - v_{k-1}\|^2. \end{aligned} \tag{26}$$

The proof of convergence become quite easy, as shown by Attouch et al. [1, Lemma 5]. Since $L(u, v) > -\infty$ and $\{L(u^k, v^k)\}$ is non-increasing, we have that $L(u^k, v^k)$ converges to some finite value. Furthermore, using induction and the fact that

$$\begin{aligned} &L(u^k, v^k) + \frac{1}{\theta_k}\|u - u_{k-1}\|^2 + \frac{1}{\theta_k}\|v - v_{k-1}\|^2 \leq \\ &L(u^{k-1}, v^k) + \frac{1}{\theta_k}\|v - v_{k-1}\|^2 \leq L(u^{k-1}, v^{k-1}), \end{aligned} \tag{27}$$

one case see that

$$\sum_k^\infty \frac{1}{\theta_k}(\|u - u_{k-1}\|^2 + \|v - v_{k-1}\|^2) < \infty. \tag{28}$$

Taking θ_k to be constant, we see that u_k,v_k converge, since $SO(n)$ is compact.

Furthermore, using the same lemma [1, Lemma 5, iii], the following can be shown: denote by $\tilde{L}(u, v)$ the unconstrained Lagrangian, where we incorporate the indicator function of the group $\mathcal{G}$,

$$\begin{aligned} \tilde{L}(u, v) &= L(u, v) + i_{\mathcal{G}}(v), \\ i_{\mathcal{G}}(v) &= \begin{cases} 0, & v \in \mathcal{G} \\ \infty, & otherwise \end{cases}, \end{aligned} \tag{29}$$

then 0 converges in the limit to the Fréchet subdifferential of the Lagrangian, $\tilde{L}(u_k, v_k)$, as the algorithm converges,

$$d(0, \partial\tilde{L}(u_k, v_k)) \to 0. \tag{30}$$

The optimization steps in our algorithm remain a projection step and total-variation denoising, but with a change in their parameters. For example, the optimal update rule

for v becomes

$$\begin{aligned}
&\underset{v\in SO(n)}{\operatorname{argmin}} \frac{r}{2}\|v-u\|^2 + \langle \mu, v-u\rangle + \frac{1}{2\theta_k}\|v-v_{k-1}\|^2 + = \\
&\underset{v\in SO(n)}{\operatorname{argmin}} \left(\frac{r}{2}+\frac{1}{2\theta_k}\right)\|v\|^2 - \langle v, ru+\mu+\frac{v_{k-1}}{\theta_k}\rangle + \frac{r}{2}\|u\|^2 - \langle \mu, u\rangle + \frac{1}{2\theta_k}\|v_{k-1}\|^2 = \\
&\underset{v\in SO(n)}{\operatorname{argmin}} \left(\frac{r}{2}+\frac{1}{2\theta_k}\right)\left\|v - \frac{ru+\mu+\frac{v_{k-1}}{\theta_k}}{r+\frac{1}{\theta_k}}\right\|^2,
\end{aligned}$$

where $\frac{1}{2\theta_k}$ denotes the coupling between each iterant and its previous value. We stress, however, that in practice the algorithm converges without the above modification quite well.

5.2 Global Convergence for $SPD(n)$ Regularization

For $SPD(n)$ regularization we basically do a coordinate descent on a convex domain [39] and therefore can show global convergence of our method. At each step of the inner iteration, we do a full minimization with respect to the selected variables block - u or v. Using the notation provided by [39], we can write our function as

$$\mathcal{L}_{\mu,r}(u,v) = f_0(u,v) + f_1(u) + f_2(v), \tag{31}$$

where

1. f_0 is a convex, smooth, function.
2. f_1, f_2 are convex, lower-semiconinuous, continuous in their effective domain.

This defines our minimization process in the same manner as in [39], and therefore shows convergence of the iterants into a global minimizer (see Proposition 1 in [39]).

6 A prior for SPD matrices based on directional diffusion

TODO: : either implement, or remove this section..

Yet another possibility is to denoise the image by denoising the orthogonal matrices and eigenvalues separately, by looking at the functional

$$\underset{l(x)>0, U(x)\in SO(n)}{\operatorname{argmin}} \int \|U diag(l) U^T - A_0\|^2 + \|\nabla l\| + \|\nabla U\| dx.$$

In this case, U is a matrix function from the spatial domain into $SO(n)$. l is a vector representing the eigenvalues of the matrices, assumed to be positive. In order to obtain a simple scheme we can perform optimization where U is represented by a 9-dimensional vector.

Minimization w.r.t. l is easy because at feasible solutions $U(x)$ is orthogonal. Hence, we have

$$\underset{l>0, U(x)\in SO(n)}{\operatorname{argmin}} \int \|U diag(l) U^T - A_0\|_F + \|\nabla l\| =$$

$$\underset{l>0, U(x)\in SO(n)}{\operatorname{argmin}} \int \|diag(l) - U^T A_0 U\|_F + \|\nabla l\|. \tag{32}$$

With respect to U, the optimization is also easy because

$$\begin{aligned}
& \|U \operatorname{diag}(l) U^T - A_0\|_F^2 \\
= \;& tr\left\{(U \operatorname{diag}(l) U^T - A_0)^T (U(w) \operatorname{diag}(l) U^T - A_0)\right\} = \\
= \;& tr\left\{(U \operatorname{diag}(l) U^T (U \operatorname{diag}(l) U^T + A_0^T A_0 - 2(U \operatorname{diag}(l) U^T A_0)\right\} + \\
= \;& tr\left\{(U(w) \operatorname{diag}(l^2) U(w)^T + A_0^T A_0 - 2(U(w) \operatorname{diag}(l) U(w)^T A_0)\right\} \quad (33) \\
= \;& tr\left\{\operatorname{diag}(l)\left[(U(w) U(w)^T (\operatorname{diag}(l) - 2A_0)\right] + A_0^T A_0\right\} \quad (34)
\end{aligned}$$

The minimization w.r.t. U would give us a projection operator, also possible via an eigenvalue decomposition. Note that in this case, the representation of u should be by 6 elements only, forcing symmetry.

7 Preliminary Results

We now demonstrate several results using the functionals described above.

7.1 SO(n) regularization

In Figure 1 we demonstrate TV denoising of an image $I : (x, y) \to SO(n)$. The color image shows the inner product of the signal at each point with the matrices

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix}.$$

Each of the quadrants in the image contains a noisy version of the group elements

$$\begin{pmatrix} 0.8572 & 0.5143 & 0 \\ -0.5143 & 0.8576 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 0.9806 & 0 & -0.1961 \\ 0 & 0 & 1 \\ 0.1961 & 0 & 0.9806 \end{pmatrix},$$

$$\begin{pmatrix} 0.9285 & 0 & -0.3714 \\ 0 & 1 & 0 \\ 0.3714 & 0 & 0.9285 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{35}$$

An element-wise, i.i.d noise with $\sigma = 0.5$ is added to the image. The results, in Figure 1 clearly demonstrate the accuracy of the method, with a reduction of MSE from 0.501 to 0.0381.

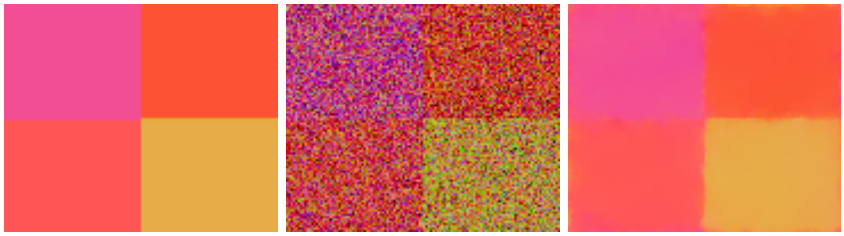

Figure 1: TV regularization of $SO(n)$ data. Left-to-right: the original image, the noisy image, the denoised image.

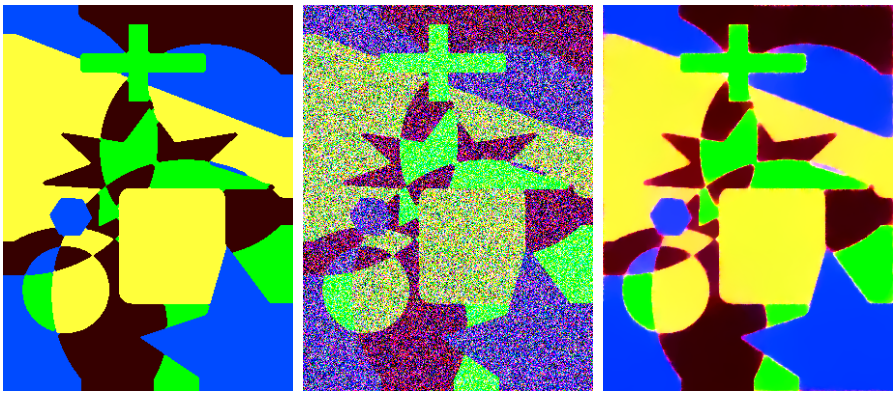

Figure 2: TV regularization of $SE(n)$ data. Left-to-right: the original image, the noisy image, the denoised image.

7.2 SE(n) regularization

In Figure 2 we demonstrate a smoothing of synthetic images with data in SE(n). Each of the colored regions in the visualization contains one of the following elements, upto additive, element-wise noise,

$$\left\{ \begin{pmatrix} 0.9285 & 0 & -0.3714 & 0.3516 \\ 0 & 1.0000 & 0 & -0.6965 \\ 0.3714 & 0 & 0.9285 & 1.6961 \end{pmatrix}, \begin{pmatrix} 0.9806 & 0 & 0.1961 & 0.8717 \\ 0 & 1.0000 & 0 & -1.4462 \\ -0.1961 & 0 & 0.9806 & -0.7012 \end{pmatrix}, \right.$$
$$\left. \begin{pmatrix} 1.0000 & 0 & 0 & 1.1650 \\ 0 & 1.0000 & 0 & 0.6268 \\ 0 & 0 & 1.0000 & 0.0751 \end{pmatrix}, \begin{pmatrix} 0.8575 & -0.5145 & 0 & 0.0591 \\ 0.5145 & 0.8575 & 0 & 1.7971 \\ 0 & 0 & 1.0000 & 0.2641 \end{pmatrix} \right\}.$$

The visualization we used projects the elements unto some representative vector in $\mathbb{R}^{12}$. It can be seen that the denoising for this synthetic example is quite good, as affirmed by reduction in the MSE (from 0.4998 to 0.0612).

7.3 DT-MRI regularization

In Figure 3 we demonstrate a smoothing of DT-MRI data from [24], based on the scheme suggested in Section 4. We show an axial view of the brain, with color coding based on the principle eigenvalue [49], and a fractional anisotropy visualization. Figure 4 demonstrates the convergence rate for the regularization. MSE of the matrix representation was 0.0489 in the corrupted image and 0.0240 in the regularized image.

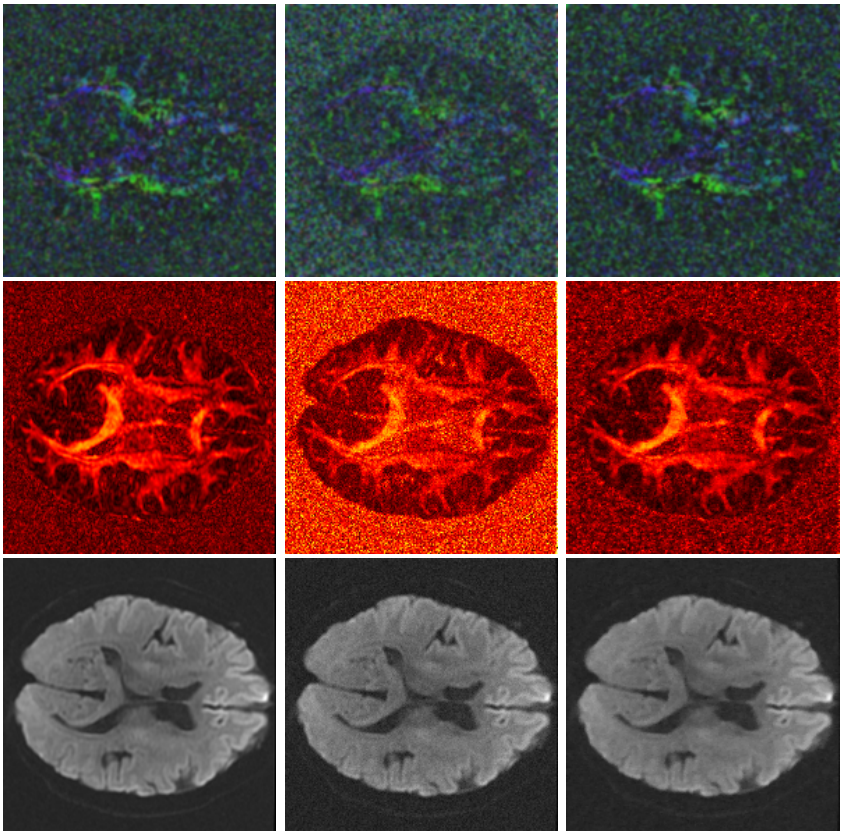

Figure 3: TV denoising of images with diffusion tensor data. Left-to-right: the original image, an image with added component-wise Gaussian noise, and the denoised image. The top row visualizes the directions, the middle row contains fractional anisotropy, and the bottom row shows the mean diffusivity. Noise was of standard deviation 0.02, $\lambda = 70$.

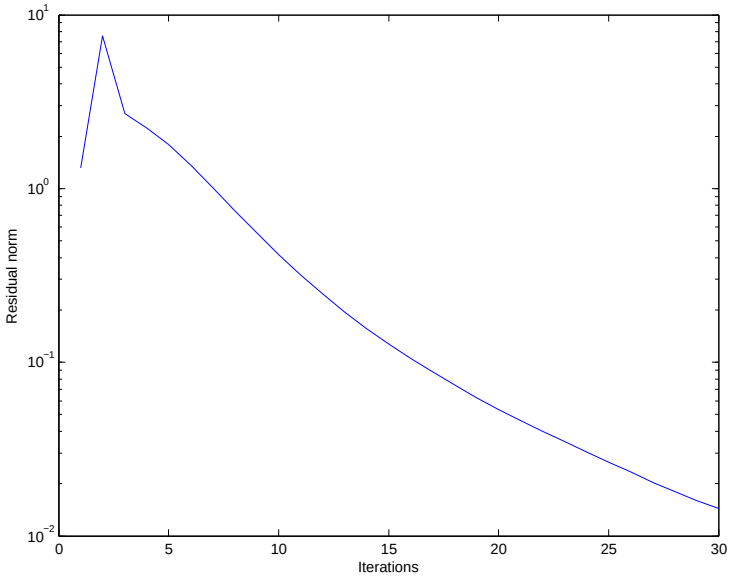

Figure 4: A residual plot for DT-MRI denoising as demonstrated in Figure 3.

In Figure 5 we demonstrate reconstruction of the DT-MRI tensors, again based data from Lundervold et al. [24], using a set of 30 directional measurements. The measure ratios $\log\left(\frac{S_i}{S_0}\right)$ were added a Gaussian additive noise of standard deviation 100.

8 Conclusions

In this paper we demonstrate the effectiveness of augmented Lagrangian regularization for regularization of matrix valued maps. Specifically, we have shown the efficiency and effectiveness of the resulting total-variation regularization of images with matrix-valued data taken from $SO(n)$, $SE(n)$, and $SPD(n)$. For the case of $SPD(n)$ we have shown the method's usefulness for denoising and regularized reconstruction of DTI data, as well as noted the convexity of the resulting optimization problem.

In future work we intend to explore the various ways of handling the matrix-valued regularization problem and the coupling between matrix elements, as well as extend our work into non-Euclidean domains.

TODO: Another question, highlighted in Xavier Pennec's work (see for example [27]), is that of choosing the right diffusion operator (with respect to the data manifold).

TODO: Regarding data - do we have more DT-MRI data to work on? Is my interpretation of the data as $\log(\frac{S_i}{S_0})$ correct?

TODO: Further relate to "Alternating Direction Augmented Lagrangian Methods for SDP" by Zaiwen Wen, Donald Goldfarb, Wotao Yin [45].

TODO: add efficiency examples / running times /etc

TODO: compare PSNR of reconstruction / denoising - although we do not claim this is better than nonlocal methods

TODO: read David Tschumperle's thesis and see more about the justification for vectorial TV for SPD, relate to it

TODO: add algorithmic descriptions

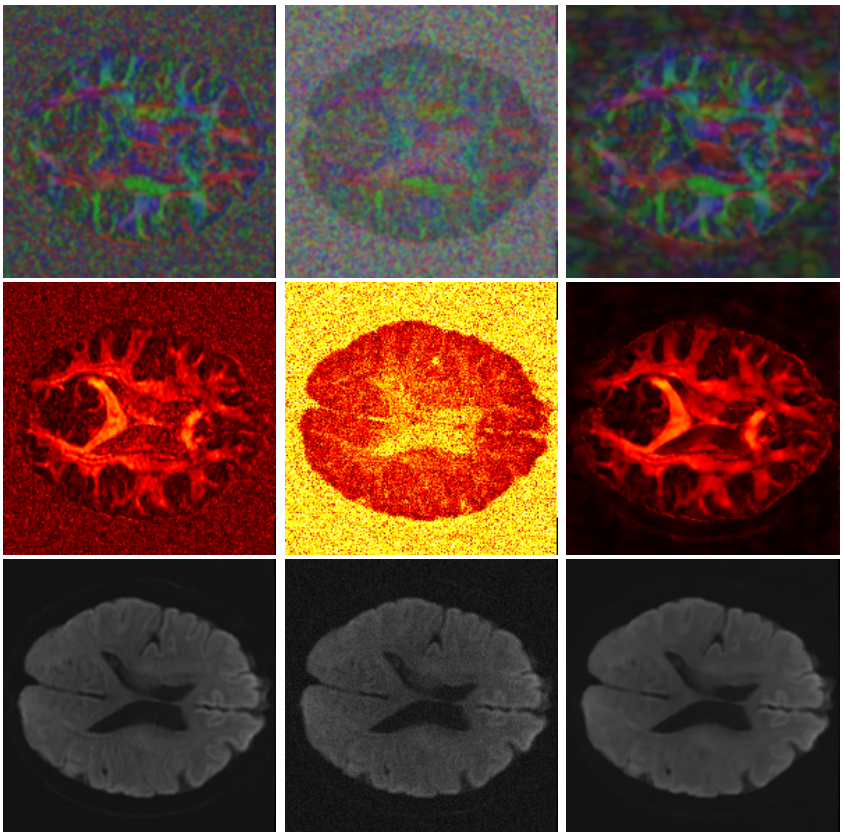

Figure 5: TV-regularized reconstruction of diffusion tensor data. Left-to-right: the original reconstruction without noise, the noisy least-squares fitting solution (used as initialization), and the regularized reconstruction result. Top-to-bottom: a visualization of the principal directions, the fractional anisotropy, and the mean diffusivity. The noise added to the field ratio logarithm was of strength 100, $\lambda = 5 \times 10^{-6}$.

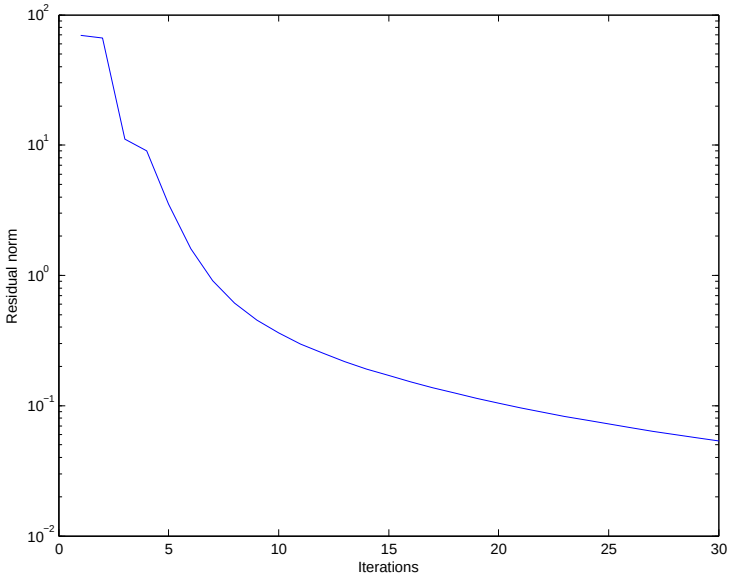

Figure 6: A residual plot for DT-MRI denoising as demonstrated in Figure 5.

TODO: relate to Bregman distance, and add a convergence proof
TODO: section 5 - try numerically, but
TODO: cite [5], look at their qualitative measurements, [44], [25], NLM methods such as [47], [2]
TODO: quote perona,
TODO: quote A Generic Approach to Diffusion Filtering of Matrix-Fields,
TODO: fix the image ordering TODO: compare to NLM or some other denoising method

References

[1] Hédy Attouch, Jérôme Bolte, Patrick Redont, and Antoine Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Lojasiewicz inequality. *Math. Oper. Res.*, 35:438–457, May 2010.

[2] Saurav Basu, P. Thomas Fletcher, and Ross T. Whitaker. Rician noise removal in diffusion tensor MRI. In Rasmus Larsen, Mads Nielsen, and Jon Sporring, editors, *MICCAI (1)*, volume 4190 of *Lecture Notes in Computer Science*, pages 117–125. Springer, 2006.

[3] Heinz H. Bauschke and Patrick L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces.* CMS Books in Mathematics, 2011.

[4] Eduardo Bayro-Corrochano and Jaime Ortegón-Aguilar. Lie algebra approach for tracking and 3D motion estimation using monocular vision. *Image Vision Comput.*, 25:907–921, June 2007.

[5] Ørjan Bergmann, Oddvar Christiansen, Johan Lie, and Arvid Lundervold. Shape-adaptive DCT for denoising of 3d scalar and tensor valued images. *J. Digital Imaging*, 22(3):297–308, 2009.

[6] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[7] Alessio Del Bue, Xavier Jo ao, Lourdes Agapito, and Marco Paladini. Bilinear factorization via augmented lagrange multipliers. In *ECCV*, pages 283–296, Berlin, Heidelberg, 2010. Springer-Verlag.

[8] Bernhard Burgeth, Stephan Didas, Luc Florack, and Joachim Weickert. A generic approach to the filtering of matrix fields with singular pdes. In *SSVM*, pages 556–567, 2007.

[9] Rachid Deriche, David Tschumperle, and Christophe Lenglet. 2004. In *IEEE International Symposium on Biomedical Imaging: Nano to Macro*, pages 9–12, 2004.

[10] Rachid Deriche, David Tschumperlé, Christophe Lenglet, and Mikael Rousson. Variational approaches to the estimation, regularization and segmentation of diffusion tensor images. In Faugeras Paragios, Chen, editor, *Mathematical Models in Computer Vision: The Handbook*. Springer, 2005 edition, 2005.

[11] Remco Duits and Bernhard Burgeth. Scale spaces on lie groups. In *SSVM*, pages 300–312, Berlin, Heidelberg, 2007. Springer-Verlag.

[12] R Fletcher. Semi-definite matrix constraints in optimization. *SIAM Journal on Control and Optimization*, 23(4):493–513, 1985.

[13] Jean Gallier. *Geometric methods and applications: for computer science and engineering*. Springer-Verlag, London, UK, 2000.

[14] W. Gibson. On the least-squares orthogonalization of an oblique transformation. *Psychometrika*, 27:193–195, 1962. 10.1007/BF02289637.

[15] Yaniv Gur and Nir Sochen. Fast invariant riemannian DT-MRI regularization. In *ICCV*, pages 1–7, 2007.

[16] Yaniv Gur and Nir A. Sochen. Regularizing flows over lie groups. *JMIV*, 33(2):195–208, 2009.

[17] Brian C. Hall. *Lie Groups, Lie Algebras,and Representations, An Elementary Introduction*. Springer, 2004.

[18] Magnus R. Hesteness. Multipliers and gradient methods. *Journal of Optimization Theory and Applications*, 4:303–320, 1969.

[19] Nicholas J. Higham. Matrix nearness problems and applications. In M. J. C. Grover and S. Barnett, editors, *Applications of Matrix Theory*, pages 1–27. Oxford University Press, Oxford, 1989.

[20] Arieh Iserles, Hans Z. Munthe-kaas, Syvert P. Nrsett, and Antonella Zanna. Lie group methods. *Acta Numerica*, pages 215–365, 2000.

[21] Ron Kimmel, , and Nir Sochen. Orientation diffusion or how to comb a porcupine. *special issue on PDEs in Image Processing, Computer Vision, and Computer Graphics, Journal of Visual Communication and Image Representation*, 13:238–248, 2002.

[22] Rongjie Lai and Tony F. Chan. A framework for intrinsic image processing on surfaces. CAM report, UCLA, 2010.

[23] D. Lin, W.E.L. Grimson, and J. Fisher. Learning visual flows: A lie algebraic approach. In *CVPR*, pages 747–754, 2009.

[24] Arvid Lundervold. On consciousness, resting state fMRI, and neurodynamics. *Nonlinear Biomed Phys*, 4 Suppl 1, 2010.

[25] Maryam Moazeni, Alex Bui, and Majid Sarrafzadeh. Accelerating total variation regularization for matrix-valued images on gpus. In *Proceedings of the 6th ACM conference on Computing frontiers*, CF '09, pages 137–146, New York, NY, USA, 2009. ACM.

[26] Lionel Moisan. Perspective invariant movie analysis for depth recovery. *Proc. SPIE*, 2567:84–94, 1995.

[27] Xavier Pennec, Pierre Fillard, and Nicholas Ayache. A Riemannian framework for tensor computing. *IJCV*, 66(1):41–66, 2006.

[28] Michael J.D. Powell. *Optimization*, chapter A method for nonlinear constraints in minimization problems, pages 283–298. Academic Press, 1969.

[29] Guy Rosman, Alexander M. Bronstein, Michael M. Bronstein, Alon Wolf, and Ron Kimmel. Group-valued regularization for motion segmentation of dynamic non-rigid shapes. In *SSVM*, 2011.

[30] Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D Letters*, 60:259–268, 1992.

[31] S W Shepperd. Quaternion from rotation matrix. *Eng. Notes*, pages 223–224, 1978.

[32] Gabriele Steidl, Simon Setzer, B. Popilka, and Bernhard Burgeth. Restoration of matrix fields by second-order cone programming. *Computing*, 81(2-3):161–178, 2007.

[33] E. O. Stejskal and J. E. Tanner. Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient. *Journal of Chemical Physics*, 42:288–292, 1965.

[34] Michael A. Stephens. Vector correlation. *Biometrika*, 66:41–48, 1979.

[35] John Stillwell. *Naive Lie Theory*. Undergraduate Texts in Mathematics. Springer, New York, NY, 2008.

[36] Xue-Cheng Tai and Chunlin Wu. Augmented Lagrangian method, dual methods and split Bregman iteration for ROF model. In *SSVM*, pages 502–513, 2009.

[37] Bei Tang, Guillermo Sapiro, and Vicent Caselles. Diffusion of general data on non-flat manifolds via harmonic maps theory: The direction diffusion case. *IJCV*, 36:149–161, February 2000.

[38] David Tschumperle and Rachid Deriche. Vector-valued image regularization with pdes: A common framework for different applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:506–517, 2005.

[39] Paul Tseng. Coordinate ascent for maximizing nondifferentiable concave functions. LIDS-P 1940, MIT, 1988.

[40] Oncel Tuzel, Fatih Porikli, and Peter Meer. Learning on Lie-groups for invariant detection and tracking. In *CVPR*, pages 1–8, 2008.

[41] Yilun Wang, Junfeng Yang, Wotao Yin, and Yin Zhang. A new alternating minimization algorithm for total variation image reconstruction. *SIAM J. Imag. Sci.*, 1(3):248–272, 2008.

[42] Yu Wang, Xue-Cheng Tai, and ? Saddle-point problems for minimizing p-harmonic energy. unpublished, 2011.

[43] Zhizhou Wang and Baba C. Vemuri. DTI segmentation using an information theoretic tensor dissimilarity measure. *IEEE Trans. Med. Imaging*, 24(10):1267–1277, 2005.

[44] Joachim Weickert and Thomas Brox. Diffusion and regularization of vector- and matrix-valued images. volume 313 of *Inverse problems, image analysis, and medical imaging*, 2002.

[45] Zaiwen Wen, Donald Goldfarb, and Wotao Yin. Alternating direction augmented lagrangian methods for semidefinite programming. RICE CAAM TR09-42, Rice university, 2009.

[46] C.-F. Westin, S. Peled, H. Gudbjartsson, R. Kikinis, and F. A. Jolesz. Geometrical diffusion measures for MRI from tensor basis analysis. In *ISMRM '97*, page 1742, Vancouver Canada, April 1997.

[47] Nicolas Wiest-Daesslé, Sylvain Prima, Pierrick Coupé, Sean Patrick Morrissey, and Christian Barillot. Non-local means variants for denoising of diffusion-weighted and diffusion tensor MRI. In *Proceedings of the 10th international conference on Medical image computing and computer-assisted intervention*, MICCAI'07, pages 344–351, Berlin, Heidelberg, 2007. Springer-Verlag.

[48] Chunlin Wu, Juyong Zhang, Yuping Duan, and Xue-Cheng Tai. Augmented lagrangian method for total variation based image restoration and segmentation over triangulated surfaces. *Journal of Scientific Computing*, pages 1–22, 2011. 10.1007/s10915-011-9477-3.

[49] Song Zhang, David S. Laidlaw, and Gordon Kindlmann. Diffusion tensor MRI visualization. In C. Johnson and C. Hansen, editors, *Visualization Handbook*, 2004.