

Variational Methods with Higher-Order Derivatives in Image Processing

S. Setzer and G. Steidl

Abstract. This is an overview of recent research of the authors on the application of variational methods with higher-order derivatives in image processing. We focus on gray-valued and matrix-valued images and deal with a purely discrete setting. We show that regularization methods with second-order derivatives can be successfully applied to the denoising of gray-value images. In 1D the solutions of the corresponding minimization problems are discrete polynomial splines (sometimes with higher defects) and inf-convolution splines with certain knots. The proposed methods can be transferred to matrix fields. Due to the operator structure of matrices, new tasks like the preservation of positive definiteness and the meaningful coupling of the matrix components come into play.

§1. Introduction

In recent years mathematical methods from optimization theory, harmonic analysis, stochastics or partial differential equations were successfully applied in digital image processing, while conversely image processing tasks have led to interesting mathematical questions. In this paper, we restrict our attention to applications of variational methods in conjunction with higher-order derivatives in image processing. In a couple of papers, these techniques have proved to be useful for *scalar-valued images*, *vector-valued images* and *tensor-valued images*. In this paper, we are only interested in scalar- and matrix-valued images, more precisely in the denoising of gray-value images and matrix fields. Vector-valued images are for example colored images or optical flow fields, see Fig. 1 (middle). One of the authors has used higher-order regularization methods for the simultaneous estimation and decomposition of optical flows [47]. Matrix-valued data have gained significant importance in recent years, e.g., in diffusion tensor magnetic resonance imaging (DT-MRI). Here, every image pixel corresponds

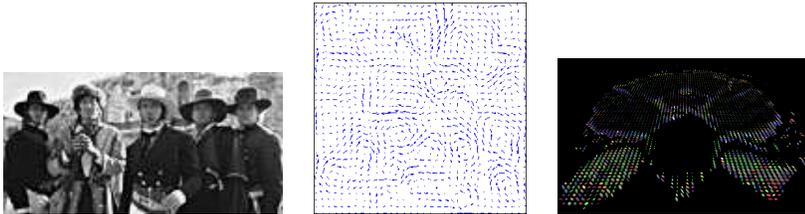


Fig. 1. Gray-value image of the battle at the Alamo in San Antonio (left), vector-valued image of an optical flow field (middle), matrix-valued image of a DT-MRI slice (right).

to a symmetric positive definite matrix A which can be visualized as the ellipsoid

$$\{x \in \mathbb{R}^3 : x^T A^{-2} x = 1\}.$$

The lengths of the axes of the ellipsoid are the eigenvalues of A and the ellipsoid illustrates the direction of the diffusion of water molecules, see Fig. 1 (right).

A well-established method for the denoising of a scalar-valued image u from a given image f degraded by white Gaussian noise consists in calculating

$$\operatorname{argmin}_u \int_{\Omega} (f - u)^2 + \alpha \Phi(|\nabla u|^2) dx dy \quad (1)$$

with a regularization parameter $\alpha > 0$ and an increasing function $\Phi : [0, \infty] \rightarrow \mathbb{R}$ in the penalizing term. The first summand encourages similarity between the restored image and the original one, while the second term rewards smoothness. Some illustrating examples are given in Fig. 2. For the straightforward choice $\Phi(s) = s^2$, the penalizing term coincides with the H^1 norm of u . The corresponding minimizer becomes too smooth at edges. The frequently applied ROF-model introduced by Rudin, Osher and Fatemi [33] with

$$\Phi(s^2) := \sqrt{s^2} = |s| \quad (2)$$

preserves sharp edges, but leads to the so-called *staircasing effect*. We will see that one way to overcome both artifacts is to use higher-order derivatives in the functional.

This paper is organized as follows: In Section 2, we start with the basic background concerning our discrete setting and positively homogeneous penalizing terms. Then, in Section 3, we deal with the denoising of gray-value images. More precisely, in Section 3.1, we start with 1D signals and verify that the solutions of the corresponding minimization problems are discrete splines whose knots correspond to the contact points of their dual counterparts with some tube boundary. In Section 3.2, we generalize

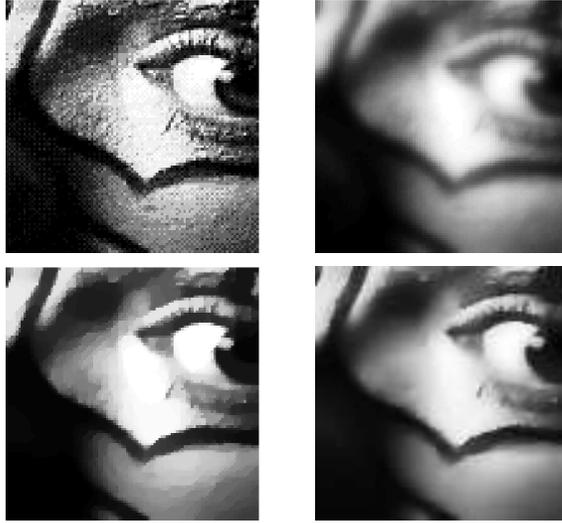


Fig. 2. Top left: Part of the MATLAB clown image. Top right: Denoised image by the H^1 -model with $\alpha = 5$. Edges are smoothed. Bottom left: Denoised image by the ROF-model with $\alpha = 10$. The staircasing effect is visible. Bottom right: Denoised image by using higher-order derivatives in the penalizing functional (Here: inf-convolution with $r = 2$ and $\alpha_1 = 10$, $\alpha_2 = 20$).

the results to images. In Section 4 we transfer the proposed successful techniques from the scalar-valued case to tensor fields. Due to the matrix structure of the data a couple of new questions appears, e.g., the meaningful coupling of the matrix components in the functionals and the preservation of positive definiteness by the minimizers. In Section 4.1 we propose a so-called 'component-based' method which directly adapts the scalar-valued approach but takes only the vector space structure of the matrices into account. The more sophisticated so-called 'operator-based' method which we introduce in Section 4.2 respects also the operator structure of matrices. Numerical examples for the different approaches with artificial as well as real-world data sets are given in Section 4.3.

§2. Preliminaries

In this paper, we will only work in a discrete setting. To this end, we approximate the derivatives by forward differences. Further, we restrict our attention to positively homogeneous penalizing terms. In the following we provide the necessary background.

Forward difference matrices. Starting with the *forward difference matrix*

$$D_{N,1} := \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 & 0 \\ & & & \ddots & & & \\ 0 & 0 & 0 & \cdots & -1 & -1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -1 & 1 \end{pmatrix} \in \mathbb{R}^{N-1,N}$$

the m -th order difference matrices $D_{N,m} \in \mathbb{R}^{N-m,N}$ can be defined by

$$D_{N,m} = D_{N-(m-1),1} \cdots D_{N,1} = D_{N-k,m-k} D_{N,k}, \quad 1 \leq k \leq m-1. \quad (3)$$

If the length N is fixed, we abbreviate $D_{N,m}$ to D_m . The image $\mathcal{R}(D_{N,m}^\top)$ of $D_{N,m}^\top$ is given by the vectors with m vanishing moments and the kernel $\mathcal{N}(D_{N,m})$ of $D_{N,m}$ by the discrete polynomials of degree $\leq m-1$, i.e.,

$$\begin{aligned} \mathcal{R}(D_{N,m}^\top) &= \{f \in \mathbb{R}^N : \sum_{j=0}^{N-1} j^r f(j) = 0, r = 0, \dots, m-1\}, \\ \mathcal{N}(D_{N,m}) &= \text{span}\{(j^r)_{j=0}^{N-1} : r = 0, \dots, m-1\}. \end{aligned}$$

We have the orthogonal decomposition $\mathbb{R}^N = \mathcal{R}(D_{N,m}^\top) \oplus \mathcal{N}(D_{N,m})$.

Positively homogeneous penalizers. For given $f \in \mathbb{R}^N$, we consider

$$\operatorname{argmin}_{u \in \mathbb{R}^N} \frac{1}{2} \|f - u\|_{\ell_2}^2 + J(u), \quad (4)$$

where J is a convex, positively homogeneous functional, i.e., $J(\lambda u) = \lambda J(u)$ for $\lambda > 0$ and all $u \in \mathbb{R}^N$. The solution \hat{u} of this problem coincides with $\hat{u} = f - \hat{v}$, where \hat{v} is the solution of the dual problem

$$\operatorname{argmin}_{v \in \mathbb{R}^N} \frac{1}{2} \|f - v\|_{\ell_2}^2 + J^*(v), \quad (5)$$

and $J^*(v) := \sup_{w \in \mathbb{R}^N} \{\langle v, w \rangle - J(w)\}$ denotes the *Legendre-Fenchel conjugate* of J . Since J is positively homogeneous, J^* is the indicator function of the convex set

$$\mathcal{C} := \{v : \langle v, w \rangle \leq J(w), \forall w \in \mathbb{R}^N\}, \quad (6)$$

i.e.,

$$J^*(v) = \delta_{\mathcal{C}}(v) := \begin{cases} 0, & \text{for } v \in \mathcal{C}, \\ \infty, & \text{for } v \notin \mathcal{C}. \end{cases}$$

Hence, (5) implies that

$$\hat{u} = f - \Pi_{\mathcal{C}} f, \quad (7)$$

where $\Pi_{\mathcal{C}}$ denotes the orthogonal projection onto \mathcal{C} .

Now let $J = J_1 \square \cdots \square J_r$ be the *infimal convolution* (inf-convolution) of the proper convex functionals J_1, \dots, J_r defined by

$$(J_1 \square \cdots \square J_r)(u) := \inf_{u_1, \dots, u_r} \{J_1(u_1) + \cdots + J_r(u_r) : u_1 + \cdots + u_r = u\}.$$

Then $J_1 \square \cdots \square J_r$ is also a proper convex functional. It is not hard to show that for lower semi-continuous proper functionals J_k with $J_k(-u) = J_k(u)$, the solution \hat{u} of (4) is given by $\hat{u} = \hat{u}_1 + \cdots + \hat{u}_r$, where $\hat{u}_k, k = 1, \dots, r$, are determined by

$$\operatorname{argmin}_{u_1, \dots, u_r} \frac{1}{2} \|f - u_1 - \cdots - u_r\|_2^2 + J_1(u_1) + \cdots + J_r(u_r).$$

The Legendre-Fenchel conjugate of the inf-convolution is given by

$$(J_1 \square \cdots \square J_r)^* = J_1^* + \cdots + J_r^*.$$

If $J_k, k = 1, \dots, r$, are positively homogeneous, then $J_k^* = \delta_{\mathcal{C}_k}$ for appropriate convex sets \mathcal{C}_k in (6) and the minimizer of (4) reads

$$\hat{u} = f - \Pi_{\cap \mathcal{C}_k} f. \quad (8)$$

§3. Denoising of Gray-Value Images

Before turning to images, we consider 1D signals and characterize the minimizers of our functionals as discrete splines with certain knots. We note that minimization problems of the form

$$\operatorname{argmin}_u \frac{1}{2} \sum_{i \in I} (f(x_i) - u(x_i))^2 + \lambda \|Lu\|_{L^p}^p, \quad (9)$$

where L denotes a general differential operator and their relations to splines are well-examined in approximation theory. For $p = 2$, one may start, e.g., with I. J. Schoenberg's and C. deBoor's papers [35, 13] in 1D, consider the results of G. Wahba [43] and in 2D of J. Duchon [14]. For $p \in (1, \infty)$, a good overview can be found in the book [4] (see also [11]), while the paper of S. D. Fisher and J. W. Jerome [15] deals with the non-reflexive setting $p = 1$. For the discrete setting, we refer to the papers of O. L. Mangasarian and L. L. Schumaker [29, 30].

3.1. Higher-Order Regularization in 1D

We start with a basic approach and consider useful generalizations afterwards.

Basic approach. For given $f \in \mathbb{R}^N$, we are interested in

$$\operatorname{argmin}_u \frac{1}{2} \|f - u\|_{\ell_2}^2 + \alpha \|D_m u\|_{\ell_1}. \quad (10)$$

By (6) and (7), the solution of (10) is given by

$$\begin{aligned} \hat{u} &= f - \Pi_{\mathcal{C}} f, \\ \mathcal{C} &= \{v : \langle v, w \rangle \leq \alpha \|D_m w\|_{\ell_1} \ \forall w \in \mathbb{R}^N\}. \end{aligned}$$

It is easy to check that $v \in \mathcal{C}$ implies that $v \in \mathcal{R}(D_m^T)$. Since $D_m^T \in \mathbb{R}^{N, N-m}$ has full rank, we have that for every $v \in \mathcal{R}(D_m^T)$ there exists a uniquely determined $V \in \mathbb{R}^{N-m}$ such that $v = D_m^T V$, and conversely $V = (D_m D_m^T)^{-1} D_m v = D_m^\dagger v$ with the Moore–Penrose inverse D_m^\dagger . Furthermore, simple estimates give

$$\sup_{\substack{w \in \mathcal{R}(D_m^T) \\ w \neq 0}} \frac{|\langle v, w \rangle|}{\|D_m w\|_{\ell_1}} \leq \|V\|_{\ell_\infty},$$

so that $\mathcal{C} = \{v := D_m^T V : \|V\|_{\ell_\infty} \leq \alpha\}$. Thus, $\hat{u} = f - D_m^T \hat{V}$, where \hat{V} is the minimizer of

$$\|f - D_m^T V\|_{\ell_2}^2 \rightarrow \min, \quad \text{subject to} \quad \|V\|_{\ell_\infty} \leq \alpha. \quad (11)$$

This is a quadratic minimization problem with linear constraints, and can be solved by standard methods.

Assume that $f \in \mathcal{R}(D_m^T)$, i.e., $f = D_m^T F$ and set $U := F - V$. Then we see that $\hat{u} = D_m^T \hat{U}$, where \hat{U} is the solution of

$$\|D_m^T U\|_{\ell_2}^2 \rightarrow \min, \quad \text{subject to} \quad \|F - U\|_{\ell_\infty} \leq \alpha.$$

It can be shown that \hat{U} is also the solution of the following *contact problem* [37]:

1. U lies in a tube around F of width 2α , i.e., $\|F - U\|_{\ell_\infty} \leq \alpha$.
2. Let $\Xi := \{j \in \{0, \dots, N - m - 1\} : (D_{N+m, 2m} \tilde{U})_j \neq 0\}$, where $\tilde{U} := (0_m, U, 0_m)$. If $j \in \Xi$, then $U(j) = F(j) - (-1)^m \alpha$, i.e., $U(j)$ contacts the boundary of the tube.

In general, solving this contact problem is not straightforward. Only for the special case $m = 1$ there exists the so-called ‘*taut-string*’ algorithm [12] which is based on a convex hull algorithm and requires only $\mathcal{O}(N)$ arithmetic operations. Concerning tube algorithms, see also [28, 20].

By the second contact condition, we see that U and u are indeed polynomial splines of degree $2m - 1$ and $m - 1$, resp., in the following sense: Recall that a real-valued function s defined on $[a, b]$ is a *polynomial spline*

of degree $m - 1$ with knots $\Theta := \{x_1, \dots, x_q\}$, $a < x_1 < \dots < x_q < b$, if $s \in C^{m-2}[a, b]$ and s is a polynomial of degree $\leq m - 1$ on each interval $[x_k, x_{k+1}]$, $k = 0, \dots, q$; $x_0 := a$, $x_{q+1} := b$, i.e.,

$$s^{(m)}(x) = 0, \quad \text{for } x \in (a, b), \quad x \notin \Theta.$$

These smoothest polynomial splines are also called *splines with defect 1* or *with knot multiplicity 1*. Similarly, we say that $u \in \mathbb{R}^N$ is a *discrete polynomial spline of degree $m - 1$ with knots $\Theta = \Xi + \lfloor \frac{m}{2} \rfloor$* if

$$(D_{N,m}u)(j) = 0, \quad j \in \{0, \dots, N - m - 1\}, \quad j \notin \Xi.$$

By this definition \tilde{U} is a *discrete polynomial spline of degree $2m - 1$ with knots $\Theta = \Xi + m$* . Now $D_m^T D_m$ is the restriction of $D_{N+m,2m}$ to its middle $N - m$ columns. Thus, $D_{N+m,2m}\tilde{U} = D_m^T D_m U$, and since $u = D_m^T U$ we see that

$$(D_m u)(j) = 0, \quad \text{except for } j \in \Xi.$$

Hence, \hat{u} is a discrete polynomial spline of degree $m - 1$ with knots $\Xi + \lfloor \frac{m}{2} \rfloor$. Material on discrete splines can be found, e.g., in [36] and in connection with optimization problems different from the one considered here in [29, 30]. The relation of $(D_m^T D_m)^{-1}$ to G. Wahba's reproducing kernels in the reproducing kernel Hilbert spaces $W_{2,0}^m$ is explained in [37].

Generalizations. The approach (10) can be generalized in various directions, e.g.:

- by introducing additional data fitting terms to encourage the similarity between derivatives of f and u ,
- by using an inf-convolution penalizing term,
- by using other matrices L than difference matrices in the penalizing term.

As an example of the **first approach** we consider

$$\operatorname{argmin}_u \frac{1}{2} \|f - u\|_{\ell_2}^2 + \frac{\beta}{2} \|D_1 f - D_1 u\|_{\ell_2}^2 + \alpha \|D_m u\|_{\ell_1}. \quad (12)$$

Let $A := B^T B = I_N + \beta D_1^T D_1 = C_N^T (I_N + \beta \Lambda^2) C_N$, where C_N denotes the matrix of the DCT-II transform, and $\Lambda := \operatorname{diag} \left(2 \sin \frac{j\pi}{2N} \right)_{j=0}^{N-1}$. Let $f = f_0 + f_1$ be the A -orthogonal decomposition of f related to $\mathbb{R}^N = \mathcal{N}(D_m) \oplus_A \mathcal{R}(A^{-1} D_m^T)$, where orthogonality is meant with respect to $\langle u, v \rangle_A = v^T A u$. Then the solution of (12) is given by $\hat{u} = f_0 + \hat{u}_1$, where $\hat{u}_1 = \hat{f}_1 - A^{-1} D_m^T \hat{V}$ and \hat{V} solves

$$\frac{1}{2} \|B f - (B^{-1})^T D_m^T V\|_{\ell_2}^2 \rightarrow \min, \quad \text{subject to } \|V\|_{\ell_\infty} \leq \alpha.$$

Using the spectral decomposition of A and the fact that the DCT-II can be computed with $\mathcal{O}(N \log N)$ arithmetic operations, the solution of this quadratic problem can be computed efficiently. Finally, for $f \in \mathcal{R}(A^{-1}D_m^\top)$ and $f = A^{-1}D_m^\top F$, the dual problem can be rewritten as $\hat{u} = A^{-1}D_m^\top \hat{U}$, where \hat{U} solves

$$\|(B^{-1})^\top D_m^\top U\|_{\ell_2}^2 \rightarrow \min, \quad \text{subject to } \|F - U\|_{\ell_\infty} \leq \alpha.$$

It turns out that \hat{U} is a discrete counterpart of a polynomial spline of degree $2m - 1$ *with defect 3*, while \hat{u} is again a discrete polynomial spline of degree $m - 1$ with defect 1, see [38].

As an example of the **second approach** we consider the inf-convolution penalizing term

$$\operatorname{argmin}_u \frac{1}{2} \|f - u\|_{\ell_2}^2 + (J_1 \square \cdots \square J_m)(u)$$

with $J_k(u) := \alpha_k \|D_k u\|_{\ell_1}$, $k = 1, \dots, m$. By (8) we obtain the solution $\hat{u} = f - \hat{v}$, where \hat{v} solves

$$\begin{aligned} \frac{1}{2} \|f - v\|_{\ell_2}^2 \rightarrow \min \quad \text{subject to} \quad v = D_1^\top V_1 = \cdots = D_m^\top V_m, \quad (13) \\ \|V_k\|_{\ell_\infty} \leq \alpha_k, \quad k = 1, \dots, m. \end{aligned}$$

By (3) we have

$$v = D_m^\top V_m = D_{N,k}^\top D_{N-k,m-k}^\top V_m,$$

and since V_k is uniquely determined, this implies that $V_k = D_{N-k,m-k}^\top V_m$. Hence (13) can be reformulated as $\hat{v} = D_m^\top \hat{V}$, where \hat{V} is the solution of

$$\frac{1}{2} \|f - D_m^\top V\|_{\ell_2}^2 \rightarrow \min, \quad \text{subject to } \|D_{N-k,m-k}^\top V\|_{\ell_\infty} \leq \alpha_k, \quad k = 1, \dots, m.$$

Here \hat{u} is the sum of discrete polynomial splines of degree $k - 1$, $k = 1, \dots, m$ with different knots. Note that so-called 'inf-convolution splines' related to the data fitting term in (9) were introduced by P. J. Laurent [24, 25].

The **third approach** leads to discrete L -splines as minimizers. There exists a rich literature on L -splines; for an overview see [36]. Recently, L -splines were studied in signal processing problems by M. Unser et al. [42].

3.2. Higher-Order Regularization in 2D

Higher-order regularization methods and PDE-based methods were considered in a different setting than proposed in this paper in [34, 10, 46, 27, 21, 22]. The inf-convolution technique was first applied in image processing by A. Chambolle and P.-L. Lions [9]. As in 1D, we start with a general approach, where we restrict our attention to second-order derivatives.

Basic approach. For the sake of simplicity, we consider quadratic images $\mathbf{f} \in \mathbb{R}^{n,n}$. We transform \mathbf{f} into a vector $f \in \mathbb{R}^N$ with $N = n^2$ in the following way:

$$\text{vec } \mathbf{f} := \begin{pmatrix} f_0 \\ \vdots \\ f_{n-1} \end{pmatrix},$$

where f_j denotes the j -th column of \mathbf{f} . Let $D := \begin{pmatrix} D_{n,1} \\ 0_n^T \end{pmatrix}$. Then we approximate the partial derivative operators by

$$\begin{aligned} D_x &:= I_n \otimes D, & D_{xx} &:= I_n \otimes D^T D, \\ D_y &:= D \otimes I_n, & D_{yy} &:= D^T D \otimes I_n, \end{aligned}$$

where $A \otimes B$ denotes the *Kronecker product* of A and B . Further, we set

$$\mathcal{D}_1 := \begin{pmatrix} D_x \\ D_y \end{pmatrix}, \quad \mathcal{D}_2 := \begin{pmatrix} D_{xx} \\ D_{yy} \end{pmatrix}.$$

As an alternative to \mathcal{D}_2 , we can also use $\mathcal{D}_{2,H} := (D_{xx}^T, D_{yy}^T, D_{xy}^T, D_{yx}^T)^T$ with an appropriate matrix D_{xy} for the mixed derivatives. Now we consider the problem

$$\underset{u \in \mathbb{R}^N}{\text{argmin}} \frac{1}{2} \|f - u\|_{\ell_2}^2 + \alpha \| |Lu| \|_{\ell_1}, \quad (14)$$

where $L \in \{\mathcal{D}_2, \mathcal{D}_{2,H}\}$ and $|W|(j) := \left(\sum_{k=0}^{p-1} W(j + kN)^2 \right)^{1/2}$ for $W \in \mathbb{R}^{pN}$ and $j = 1, \dots, N$. For $L = \mathcal{D}_1$, problem (14) is a discrete version of the ROF-model (1), (2). By (6) and (7), the solution \hat{u} of (14) is given by

$$\begin{aligned} \hat{u} &= f - \Pi_{\mathcal{C}} f, \\ \mathcal{C} &= \{v : \langle v, w \rangle \leq \alpha \| |Lw| \|_{\ell_1}, \forall w \in \mathbb{R}^N\}. \end{aligned}$$

Again we see that $v \in \mathcal{C}$ implies that $v \in \mathcal{R}(L^T)$, i.e., there exist vectors $V \in \mathbb{R}^{pN}$ such that $v = L^T V$. In contrast to the 1D case, the function V is not uniquely determined. However, we can prove that, see, e.g., [38],

$$\sup_{\substack{w \in \mathcal{R}(L^T) \\ w \neq 0}} \frac{|\langle v, w \rangle|}{\| |Lw| \|_{\ell_1}} \leq \min_{v=L^T V} \| |V| \|_{\ell_\infty}.$$

Thus, $\hat{u} = f - L^T \hat{V}$, where \hat{V} is the minimizer of

$$\|f - L^T V\|_{\ell_2}^2 \rightarrow \min, \quad \text{subject to} \quad \| |V| \|_{\ell_\infty} \leq \alpha. \quad (15)$$

This is a quadratic minimization problem with quadratic constraints (if squared). We propose to solve this problem by one of the following two techniques:

- Chambolle's descent algorithm [8] which is simple to implement, including modifications,
- second-order cone programming (SOCP) [17, 26] which is based on a primal/dual interior point method. There exists sophisticated software for SOCP, e.g., the packages SeDuMi [40] or MOSEK [1].

In our numerical experiments we have applied SOCP.

Generalizations. The generalizations considered in 1D carry over to images. Concerning the **first approach**, we may consider

$$\operatorname{argmin}_u \frac{1}{2} \|f - u\|_2^2 + \frac{\beta}{2} \|\mathcal{D}_1 f - \mathcal{D}_1 u\|_{\ell_2}^2 + \alpha \| |Lu| \|_{\ell_1}.$$

Then the dual formulation becomes

$$\|Bf - (B^{-1})^\top L^\top V\|_{\ell_2}^2 \rightarrow \min, \quad \text{subject to } \| |V| \|_{\ell_\infty} \leq \alpha, \quad (16)$$

where

$$A := B^\top B = (C_n \otimes C_n)^\top (I_N + \alpha \Lambda_2^2) (C_n \otimes C_n)$$

and $\Lambda_2^2 = \Lambda^2 \otimes I_n + I_n \otimes \Lambda^2$.

In the **second approach**, we restrict our attention to

$$\operatorname{argmin}_{u \in \mathbb{R}^N} \frac{1}{2} \|f - u\|_{\ell_2}^2 + (J_1 \square J_2)(u),$$

where

$$J_1(u) := \alpha_1 \| |\mathcal{D}_1 u| \|_{\ell_1}, \quad \text{and} \quad J_2(u) := \alpha_2 \| |\mathcal{D}_2 u| \|_{\ell_1},$$

Consequently, by (8), we obtain that $\hat{u} = f - \hat{v}$, where \hat{v} is the solution of

$$\|f - v\|_2^2 \rightarrow \min, \quad \text{subject to} \quad v = \mathcal{D}_1^\top V_1 = \mathcal{D}_2^\top V_2, \quad (17)$$

$$\| |V_1| \|_{\ell_\infty} \leq \alpha_1, \quad \| |V_2| \|_{\ell_\infty} \leq \alpha_2.$$

Now we have that

$$\mathcal{D}_2^\top = \mathcal{D}_1^\top \begin{pmatrix} D_x & 0 \\ 0 & D_y \end{pmatrix}.$$

Assuming that $V_1 = \begin{pmatrix} D_x & 0 \\ 0 & D_y \end{pmatrix} V_2$, which is in general not true, we modify (17) as

$$\|f - \mathcal{D}_2^\top V\|_2^2 \rightarrow \min, \quad \text{subject to} \quad \left\| \begin{pmatrix} D_x V^1 \\ D_y V^2 \end{pmatrix} \right\|_{\ell_\infty} \leq \alpha_1, \quad (18)$$

$$\| |V| \|_{\ell_\infty} \leq \alpha_2.$$

Note that this minimization problem is similar but not equivalent to (17). The solutions of (17) and (18) can be computed by SOCP.

A numerical comparison of the different methods, namely, (15) with $L = \mathcal{D}_1$, (15) with $L = \mathcal{D}_2$, (16) with $L = \mathcal{D}_2$, and (18) is given in Fig. 3. The parameters are chosen with respect to the best SNR. Note that the images with best SNR do not in general have the best visual quality. For a comparison of images with good visual quality see [38]. Fig. 3 demonstrates that using second order derivatives in (15) we can reduce the staircasing effect and preserve edges. By incorporating the gradient fitting term as in (16), the image quality can be further improved. Finally, the modified inf-convolution approach (18) gives the best results in our example.

Improved finite difference discretizations. Finally, we want to give some remarks on the discretization of derivatives by finite differences. For the applications at hand, the standard forward differences work well. However, e.g., for optical flow estimation/decomposition of non-rigid motion, we have to ensure that the integral identities fulfilled by the continuous operators are still correct in the discrete setting. This is not possible by using just one grid. A remedy consists in applying the finite mimetic difference method [23] which was done, e.g., in [47].

For images with distinguished directions as in Fig. 4 the results with simple forward or central differences can be improved by using, e.g., the following Haar wavelet inspired discretization of the gradient proposed in [45]: We discretize the squared gradient magnitude $|\nabla u|^2$ at cell midpoints $(i + \frac{1}{2}, j + \frac{1}{2})$ in a twofold way. To simplify notation, let us fix $i = j = 1$ and consider $U := (u(i, j))_{i,j=1}^2$. First, we can approximate u_x and u_y by arithmetic means of central difference approximations

$$\begin{aligned} u_x \left(\frac{3}{2}, \frac{3}{2} \right) &\approx (u(2, 2) - u(2, 1) + u(1, 2) - u(1, 1))/2 = -\hat{u}(1, 2), \\ u_y \left(\frac{3}{2}, \frac{3}{2} \right) &\approx (u(2, 2) + u(2, 1) - u(1, 2) - u(1, 1))/2 = -\hat{u}(2, 1), \end{aligned}$$

where $\hat{U} := HUH$ and $H := \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$. This leads to

$$\begin{aligned} |\nabla u|^2 &\approx \frac{1}{2}((u(2, 2) - u(1, 1))^2 + (u(2, 1) - u(1, 2))^2) \\ &= \hat{u}(1, 2)^2 + \hat{u}(2, 1)^2, \end{aligned} \tag{19}$$

and may also be interpreted as squared gradient magnitude with respect to the coordinates $\frac{1}{\sqrt{2}}(1, 1)^T$ and $\frac{1}{\sqrt{2}}(1, -1)^T$. On the other hand, we can

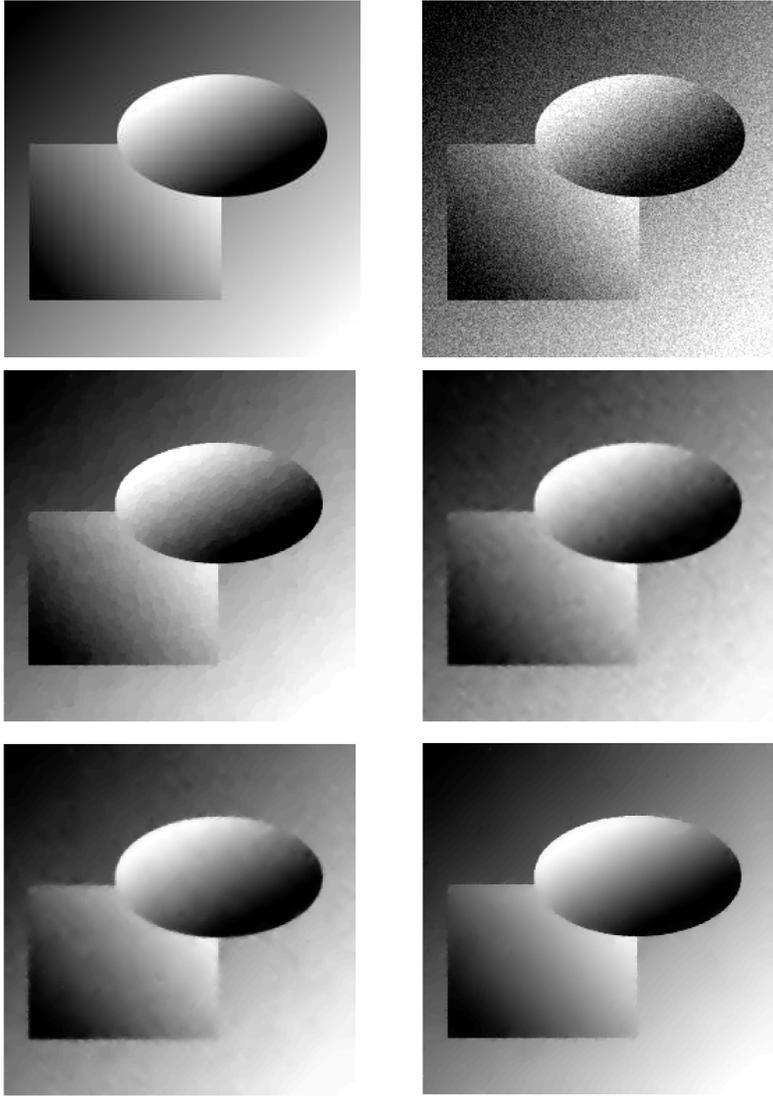


Fig. 3. Denoising experiment in 2D. Top left: Original image (size 256×256). Top right: Image with additive white Gaussian noise, SNR 11.16. Middle left: Denoised image with ROF model $\alpha = 21$, SNR=24.04. The staircasing effect is visible. Middle right: Denoised image with second order model $\alpha = 15$, SNR = 22.45. The staircasing becomes less visible. Bottom left: Denoised image with second order model (16) $\beta = 0.5$, $\alpha = 33$, SNR= 22.83. The background becomes smoother while edges are preserved. Bottom right: Denoised image with simplified inf-convolution model (18) and $\alpha_1 = 21$, $\alpha_2 = 79$. SNR = 25.80. This is the best result for our example.



Fig. 4. Top left: Original image with values in $[0,1]$. Top right: Noisy image with white Gaussian noise of standard deviation 0.4. Bottom left: Denoised image by the ROF-model and forward difference discretization. Bottom right: Denoised image by the ROF-model and Haar filter discretization.

also average the squared derivatives

$$\begin{aligned}
 u_x^2 \left(\frac{3}{2}, \frac{3}{2} \right) &\approx \frac{1}{2} \left((u(2,2) - u(2,1))^2 + (u(1,2) - u(1,1))^2 \right), \\
 u_y^2 \left(\frac{3}{2}, \frac{3}{2} \right) &\approx \frac{1}{2} \left((u(2,2) - u(1,2))^2 + (u(2,1) - u(1,1))^2 \right),
 \end{aligned}$$

and obtain

$$\begin{aligned}
 |\nabla u|^2 &\approx \frac{1}{2} \left((u(2,2) - u(1,2))^2 + (u(2,1) - u(1,1))^2 \right) \\
 &\quad + (u(2,2) - u(2,1))^2 + (u(1,2) - u(1,1))^2 \\
 &= \hat{u}(1,2)^2 + \hat{u}(1,2)^2 + 2\hat{u}(2,2)^2.
 \end{aligned} \tag{20}$$

Now each convex combination of (19) and (20) can be used as an approximation of the squared gradient magnitude. Here we restrict our attention to their average

$$|\nabla u|^2 \approx \hat{u}(1,2)^2 + \hat{u}(1,2)^2 + \hat{u}(2,2)^2. \tag{21}$$

Transferring this approximation to the whole image, we obtain as a discrete version of (1) with function (2) the minimization problem (14) with $L := ((H_1 \otimes H_0)^T, (H_0 \otimes H_1)^T, (H_1 \otimes H_1)^T)^T$ and

$$H_0 := \frac{1}{\sqrt{2}} \begin{pmatrix} \sqrt{2} & 0 & \cdots & 0 & 0 \\ 1 & 1 & \cdots & 0 & 0 \\ & \ddots & \ddots & & \\ 0 & 0 & \vdots & 1 & 1 \\ 0 & 0 & \vdots & 0 & \sqrt{2} \end{pmatrix}, \quad H_1 := \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & -1 & \cdots & 0 & 0 \\ & \ddots & \ddots & & \\ 0 & 0 & \vdots & 1 & -1 \\ 0 & 0 & \vdots & 0 & 0 \end{pmatrix}.$$

Note that $H_0 \in \mathbb{R}^{n+1,n}$ and $H_1 \in \mathbb{R}^{n+1,n}$ are related to Haar filters with the corresponding modifications at the boundary so that $\frac{1}{2}(H_0^T H_0 + H_1^T H_1) = I_n$. Fig. 4 demonstrates the improvement by using the proposed discretization (21) of the gradient. Note that with central differences we get checkerboard patterns, not shown in the figure.

§4. Denoising of Tensor-Valued Images

The following variety of applications make it worthwhile to develop appropriate tools for the restoration and processing of matrix-valued data: *First*, *diffusion tensor magnetic resonance imaging (DT-MRI)* [3] is a modern medical imaging technique that measures a 3×3 positive semidefinite matrix-field. A so-called diffusion tensor is assigned to each voxel. This diffusion tensor describes the diffusive property of water molecules. Since water diffuses preferably along ordered tissue such as nerve fibers, this matrix gives valuable information about the geometry and organization of the tissue under examination. Hence this matrix field plays a very important role for the diagnosis of multiple sclerosis and stroke. For detailed information about the acquisition of this type of data, the reader is referred to [2] and the literature cited therein. *Second*, in the field of technical sciences such as civil engineering and solid mechanics or geology, anisotropic behaviour is often described satisfactorily by inertia, diffusion, stress, and permittivity tensors. *Third*, tensors have been recognized as a useful concept in image analysis itself [18]. The *structure tensor* [16], for instance, has been employed not only for corner detection [19], but also for texture analysis [32] and motion estimation [5].

In the following, we want to transfer the techniques from the previous section to matrix fields. When designing filters for these fields, treating the channels independently is a simple though not recommended strategy. Any relation between the different matrix channels is ignored which leads to serious shortcomings. In a straightforward approach, which we call *component-based*, we relate the matrix components by the Frobenius norm which takes only the vector space structure of matrices into account. We will see that for this approach, the denoising methods from the previous

section can be directly applied. However, specific questions appear, e.g., whether these methods preserve positive definiteness. A second approach which we call *operator-based* is more adapted to the operator structure of the matrices. The first 'operator-based' method in the context of PDEs can be found in [7].

4.1. Component-based regularization

In the following, let $F : \mathbb{R}^2 \supset \Omega \rightarrow \text{Sym}_m(\mathbb{R})$ be a matrix field, where $\text{Sym}_m(\mathbb{R})$ is the vector space of real symmetric $m \times m$ matrices. This space can be treated as a Euclidian vector space with respect to the trace inner product $\langle A, B \rangle := \text{tr} AB = (\text{vec}A, \text{vec}B)$, where (\cdot, \cdot) on the right-hand side denotes the Euclidian inner vector product. Then $\langle A, A \rangle = \text{tr} A^2 = \|A\|_F^2 = \|\text{vec}A\|_{\ell_2^2}$ is the squared *Frobenius norm* of A . In $\text{Sym}_m(\mathbb{R})$, the positive semi-definite matrices $\text{Sym}_m^+(\mathbb{R})$ form a closed convex set whose interior consists of the positive definite matrices. More precisely, $\text{Sym}_m^+(\mathbb{R})$ is a cone with a base.

Analogously to (1), we consider

$$\underset{U}{\text{argmin}} \int_{\Omega} \|F - U\|_F^2 + \alpha \Phi(\text{tr}(U_x^2 + U_y^2)) \, dx dy, \quad (22)$$

where the partial derivatives are taken componentwise. The penalizing term $J(U)$ in (22) was introduced by Deriche and D. Tschumperlé [41]. Rewriting this term as

$$J(U) = \int_{\Omega} \Phi(\|U_x\|_F^2 + \|U_y\|_F^2) \, dx dy = \int_{\Omega} \Phi\left(\sum_{j,k=1}^m \nabla u_{jk}^T \nabla u_{jk}\right) \, dx dy, \quad (23)$$

we see its component-based structure implied by the Frobenius norm. However, due to the sum on the right-hand side, Φ is indeed applied to coupled matrix coefficients. By [6], the Euler-Lagrange equation of (23) is given by

$$0 = F - U + \alpha (\partial_x(\Phi'(\text{tr}(U_x^2 + U_y^2))U_x + \partial_y(\Phi'(\text{tr}(U_x^2 + U_y^2))U_y)). \quad (24)$$

Again, we are only interested in the ROF-function Φ given by (2). More precisely, since Φ in (2) is not differentiable at zero, we have to use its modified version

$$\Phi(s^2) = \sqrt{s^2 + \varepsilon^2}, \quad (25)$$

with a small additional parameter ε .

For computations we consider the discrete counterpart of (22), where we once again replace the derivative operators by simple forward difference

operators

$$\operatorname{argmin}_U \sum_{i,j=0}^{N-1} \frac{1}{2} \|F(i,j) - U(i,j)\|_F^2 + \alpha J(U), \quad (26)$$

$$J(U) := \sum_{i,j=0}^{N-1} (\|U(i,j) - U(i-1,j)\|_F^2 + \|U(i,j) - U(i,j-1)\|_F^2)^{1/2}$$

with $U(-1,j) = U(i,-1) = 0$. The functional in (26) is strictly convex and coercive, and thus has a unique minimizer.

We say that the discrete matrix field $F : \mathbb{Z}_n^2 \rightarrow \operatorname{Sym}_m^+(\mathbb{R})$ has all eigenvalues in an interval \mathcal{I} if all the eigenvalues of every matrix $F(i,j)$ of the field lie in \mathcal{I} . By the following proposition, the minimizer of (26) preserves positive definiteness. The proof is based on Courant's Min-Max principle and the projection theorem for convex sets, and can be found in [39].

Proposition 1. *Let all eigenvalues of $F : \mathbb{Z}_n^2 \rightarrow \operatorname{Sym}_m^+(\mathbb{R})$ be contained in the interval $[\lambda_{\min}, \lambda_{\max}]$. Then the minimizer \hat{U} of (26) has all eigenvalues in $[\lambda_{\min}, \lambda_{\max}]$.*

To see how the methods from previous section carry over to matrix fields, we rewrite (26) in matrix-vector form. To this end, let $N = n^2$ and $M := m(m+1)/2$. We reshape $F : \mathbb{Z}_n^2 \rightarrow \operatorname{Sym}_m(\mathbb{R})$ into the vector

$$f := \begin{pmatrix} \varepsilon_{1,1} & \operatorname{vec}(F_{1,1}) \\ \vdots & \\ \varepsilon_{1,m} & \operatorname{vec}(F_{1,m}) \\ \varepsilon_{2,2} & \operatorname{vec}(F_{2,2}) \\ \vdots & \\ \varepsilon_{2,m} & \operatorname{vec}(F_{2,m}) \\ \vdots & \\ \varepsilon_{m,m} & \operatorname{vec}(F_{m,m}) \end{pmatrix} \in \mathbb{R}^{MN},$$

where $F_{k,l} := (F_{k,l}(i,j))_{i,j=0}^{n-1}$ and $\varepsilon_{k,l} := \begin{cases} \sqrt{2}, & \text{for } k \neq l \\ 1, & \text{otherwise.} \end{cases}$

Then (26) becomes

$$\operatorname{argmin}_{u \in \mathbb{R}^{MN}} \frac{1}{2} \|f - u\|_{\ell_2}^2 + \alpha \| (I_M \otimes \mathcal{D}_1) u \|_{\ell_1}.$$

This problem has the structure of (14) with $L := I_M \otimes \mathcal{D}_1 \in \mathbb{R}^{2MN, MN}$ and $p = 2M$. Thus it can be solved by applying SOCP to its dual given by (15).

Similarly, we can transfer the inf-convolution approach to the matrix-valued setting. Obviously, we have to compute

$$\operatorname{argmin}_u \frac{1}{2} \|f - u\|_{\ell_2}^2 + (J_1 \square J_2)(u),$$

with $J_1(u) := \alpha_1 \| | (I_M \otimes \mathcal{D}_1) u \|_{\ell_1}$ and $J_2(u) := \alpha_2 \| | (I_M \otimes \mathcal{D}_2) u \|_{\ell_1}$. In our numerical examples we solve the corresponding modified dual problem (18), which reads

$$\begin{aligned} \|f - (I_M \otimes \mathcal{D}_2^T) V\|_{\ell_2}^2 &\rightarrow \min, \text{ subject to} \\ \| | \left(I_M \otimes \begin{pmatrix} D_x & 0 \\ 0 & D_y \end{pmatrix} \right) V \|_{\infty} &\leq \alpha_1, \\ \| | V \|_{\infty} &\leq \alpha_2. \end{aligned} \quad (27)$$

4.2. Operator-based regularization

In this subsection, we introduce a regularization term that emphasizes the operator structure of matrices. In addition to their vector space structure, matrices can be multiplied. Unfortunately, the original matrix multiplication does not preserve the symmetry of the matrices. The *Jordan-product* of matrices $A, B \in \operatorname{Sym}_m(\mathbb{R})$ defined by

$$A \bullet B := \frac{1}{2}(AB + BA)$$

preserves the symmetry of the matrices but not the positive semi-definiteness. For $A \in \operatorname{Sym}_m(\mathbb{R})$ with eigenvalue decomposition $A = Q\Lambda Q^T$, let $\Phi(A) = Q\Phi(\Lambda)Q^T$, where $\Lambda := \operatorname{diag}(\lambda_1, \dots, \lambda_m)$ and $\Phi(\Lambda) := \operatorname{diag}(\Phi(\lambda_1), \dots, \Phi(\lambda_m))$. We consider the following minimization problem

$$\operatorname{argmin}_U \int_{\Omega} \|F - U\|_F^2 + \alpha \operatorname{tr}(\Phi(U_x^2 + U_y^2)) \, dx dy. \quad (28)$$

In contrast to (22), the trace is taken after applying Φ to the matrix $U_x^2 + U_y^2$.

The next proposition, which can be found in [39], shows that the functional (28) has an interesting Gâteaux derivative.

Proposition 2. *Let Φ be a differentiable function. Then the Euler-Lagrange equations for minimizing the functional (28) are given by*

$$0 = F - U + \alpha (\partial_x (\Phi'(U_x^2 + U_y^2) \bullet U_x) + \partial_y (\Phi'(U_x^2 + U_y^2) \bullet U_y)). \quad (29)$$

In contrast to (24), the Jordan product of matrices appears in (29), and the function Φ' is applied to matrices. Note that in [44] an anisotropic

diffusion concept for matrix fields was presented, where the function Φ was also applied to a matrix.

We apply Proposition 2 to compute a minimizer of (28) by solving the corresponding reaction–diffusion equation for $t \rightarrow \infty$ by a difference method:

$$U_t = F - U + \alpha (\partial_x (\Phi'(U_x^2 + U_y^2) \bullet U_x) + \partial_y (\Phi'(U_x^2 + U_y^2) \bullet U_y))$$

with Φ as in (25), homogeneous Neumann boundary conditions and initial value F . More precisely, we use the iterative scheme

$$U^{(k+1)} = (1 - \tau)U^{(k)} + \tau F + \tau \alpha \left(\partial_x \left(G^{(k)} \bullet U_x^{(k)} \right) + \partial_y \left(G^{(k)} \bullet U_y^{(k)} \right) \right) \quad (30)$$

with sufficiently small time step size τ and $G^{(k)} := \Phi'((U_x^{(k)})^2 + (U_y^{(k)})^2)$. The inner derivatives including those in G are approximated by forward differences, and the outer derivatives by backward differences, so that the penalizing term becomes

$$\begin{aligned} & \frac{1}{h_1} \left(G(i, j) \bullet \frac{U(i+1, j) - U(i, j)}{h_1} - G(i-1, j) \bullet \frac{U(i, j) - U(i-1, j)}{h_1} \right) + \\ & \frac{1}{h_2} \left(G(i, j) \bullet \frac{U(i, j+1) - U(i, j)}{h_2} - G(i, j-1) \bullet \frac{U(i, j) - U(i, j-1)}{h_2} \right), \end{aligned}$$

where h_i , $i = 1, 2$ denote the pixel distances in x and y -direction. Alternatively, we have also worked with symmetric differences for the derivatives. In this case we have to replace, e.g., $G(i, j)$ in the first summand by $(\tilde{G}(i+1, j) + \tilde{G}(i, j))/2$, and \tilde{G} is now computed with symmetric differences.

4.3. Numerical Results

Finally, we present numerical results demonstrating the performance of the different methods for matrix-valued data. All algorithms were implemented in MATLAB. Moreover, we have used the software package MOSEK for SOCP and an OpenGL-based routine for visualizing the ellipsoids.

SOCP amounts to minimizing a linear objective function subject to the constraints that several affine functions of the variables have to lie in a *second-order cone* $\mathcal{C}^{n+1} \subset \mathbb{R}^{n+1}$ defined by

$$\mathcal{C}^{n+1} = \left\{ \begin{pmatrix} x \\ \bar{x}_{n+1} \end{pmatrix} = (x_1, \dots, x_n, \bar{x}_{n+1})^T : \|x\|_2 \leq \bar{x}_{n+1} \right\}.$$

With this notation, the general form of a SOCP is given by

$$\inf_{x \in \mathbb{R}^n} f^T x \quad \text{s.t.} \quad \begin{pmatrix} A_i x + b_i \\ c_i^T x + d_i \end{pmatrix} \in \mathcal{C}^{n+1}, \quad i = 1, \dots, r. \quad (31)$$

Alternatively, one can also use the rotated version of the standard cone

$$\mathcal{K}^{n+2} := \left\{ (x, \bar{x}_{n+1}, \bar{x}_{n+2})^\top \in \mathbb{R}^{n+2} : \|x\|_2^2 \leq 2\bar{x}_{n+1}\bar{x}_{n+2} \right\},$$

which allows us to incorporate quadratic constraints. Problem (31) is a convex program for which efficient, large scale solvers are available [31]. For rewriting our minimization problems as a SOCP see [39].

We start by considering the matrix-valued function $F: \mathbb{Z}_{32}^2 \rightarrow \text{Sym}_2^+(\mathbb{R})$ in Fig. 5. The 2×2 matrices are visualized by the corresponding ellipses. The components of the original data lie in the interval $[0, 2]$. We have added white Gaussian noise with standard deviation 0.6 to all components. We compare the minimizer of the component-based approach (22) resp. (26) with those of the operator-based approach (28). For computing the minimizer of the first functional, we applied SOCP while the minimizer of the second one was computed via the reaction-diffusion equation (30) with time step size $\tau = 0.00025$. The iterations were stopped when the relative error in the ℓ_2 -norm between two consecutive iterations became smaller than 10^{-8} (approximately 20000 iterations) although the result becomes visually static much earlier. The middle row of the figure contains the error plots for both methods. The actual minima w.r.t. the Frobenius norm are given for (26) by 12.19 at $\alpha = 1.75$ and for (28) by 10.79 at $\alpha = 1.2$. Hence, with respect to the computed errors, the operator-based method outperforms the component-based one. The corresponding denoised images are shown at the bottom of the figure.

Fig. 6 shows a function $F: \mathbb{Z}_{12}^2 \rightarrow \text{Sym}_3(\mathbb{R})$, where the matrix components lie in the interval $[-0.5, 0.5]$. We have added white Gaussian noise of standard deviation 0.06 to all components. The denoising results are displayed in the last two rows of Fig. 6. We have computed the minimizers of the component-based method (26) by SOCP. The smallest error, measured in the Frobenius-norm, is 1.102, and was obtained for the regularization parameter $\alpha = 0.11$. The minimizer of the inf-convolution approach (27) is depicted at the bottom of the figure. Here the optimal regularization parameters are $\alpha_1 = 0.14$ and $\alpha_2 = 0.08$. The corresponding Frobenius-norm error is 0.918. We see that the inf-convolution approach is also suited for matrix-valued data. For the operator-based approach which is not illustrated in the figure, we obtain as smallest Frobenius-norm error 1.0706 at $\alpha = 0.12$. This lies between the error of the approach (26) and the error of the inf-convolution method.

In our final experiment, we applied the two component-based methods (26) and (27) to a real world data set. Fig. 7 shows the original data and the minimizers of (26) and (27). The components of the original data lie in $[-4000, 7000]$, and we have used the regularization parameters $\alpha = 600$ for (26) and $\alpha_1 = 500, \alpha_2 = 600$ for (27), respectively.

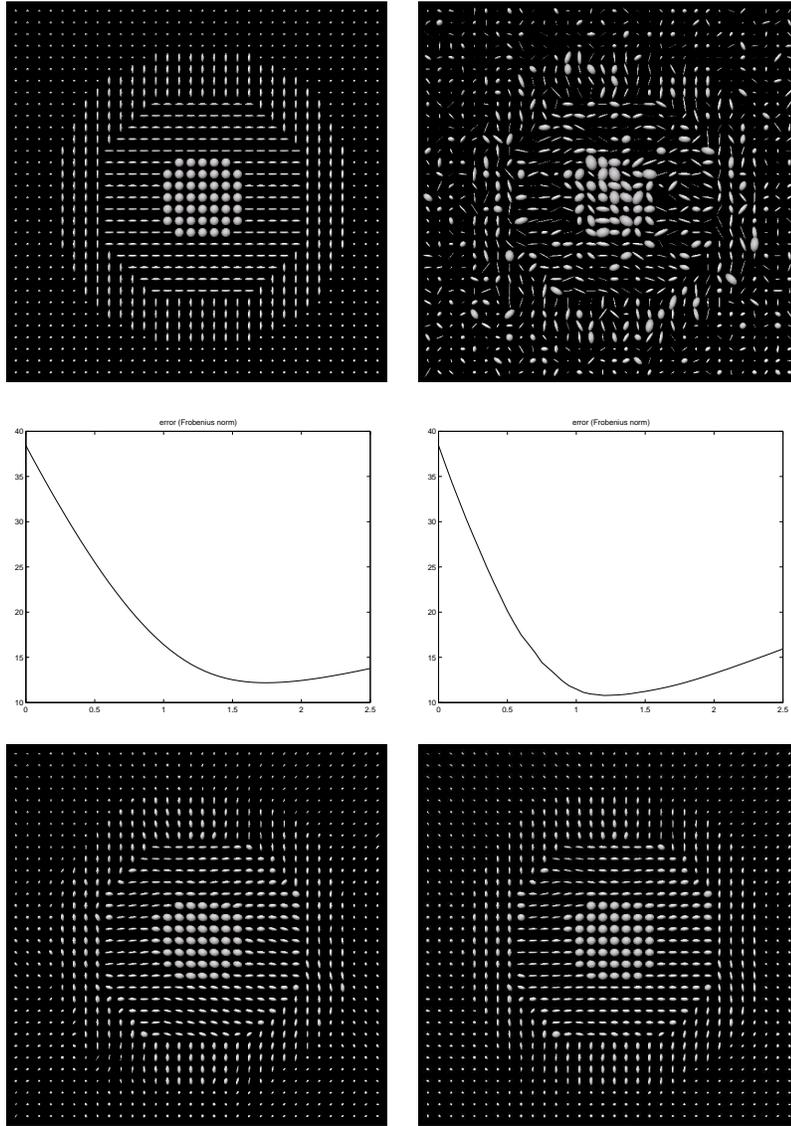


Fig. 5. Denoising of a $\text{Sym}_2(\mathbb{R})$ -valued image. Top: Original image (left), noisy image (right). Middle: Error of the Frobenius norm as a function of the regularization parameter α for the minimizers of the component-based functional (26) (left) and the operator-based functional (28) (right). Bottom: Denoised image for α corresponding to the smallest error in the Frobenius norm for the component-based functional (left) and the operator-based functional (right).

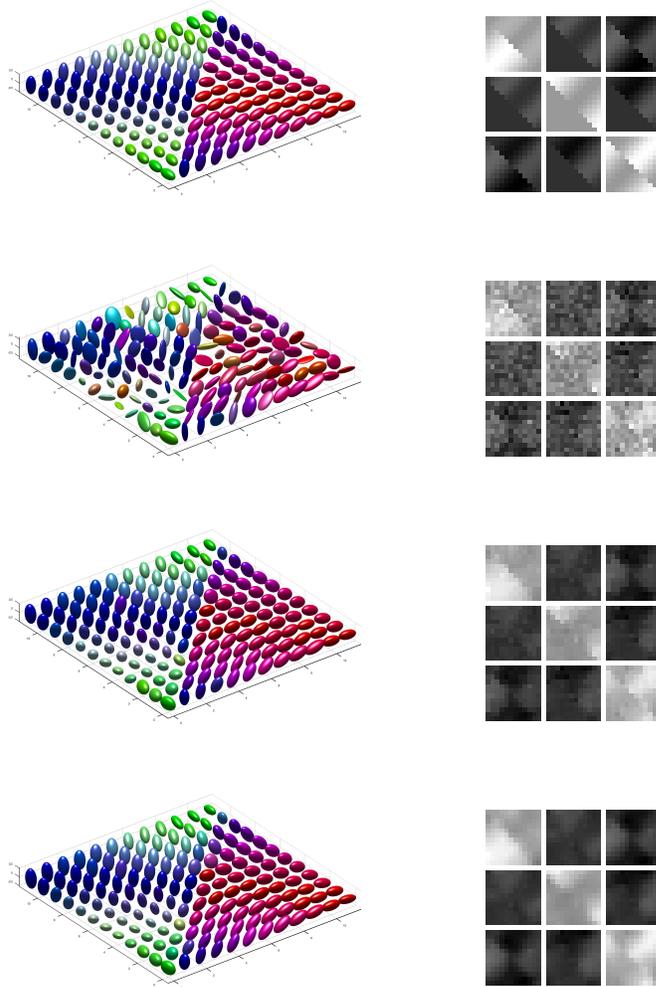


Fig. 6. Denoising of a $\text{Sym}_3(\mathbb{R})$ -valued image. Top to Bottom: Original image, noisy image, minimizer of the component-based method (26) for $\alpha = 0.11$, minimizer of the component-based inf-convolution approach (27) with parameters $\alpha_1 = 0.14$, $\alpha_2 = 0.08$. Visualization: ellipsoids (left), components of the matrix-valued data (right). The color of the ellipsoid associated with a matrix A is chosen with respect to the normalized eigenvector corresponding to the largest eigenvalue of A .

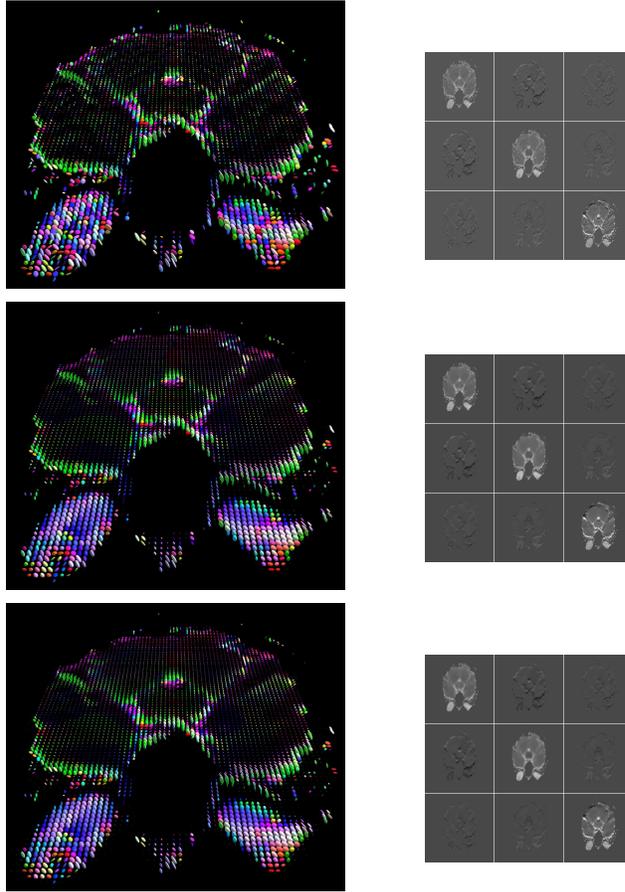


Fig. 7. Denoising of a real-world DT-MRI matrix field with values in $\text{Sym}_3(\mathbb{R})$. Top: Original image. Middle: Minimizer of the component-based method (26) for $\alpha = 600$. Bottom: Minimizer of the inf-convolution approach (27) for $\alpha_1 = 500$, $\alpha_2 = 600$.

Acknowledgement: The authors acknowledge the fruitful cooperation with Ch. Schnörr and J. Yuan (University of Mannheim) within the DFG-project SCHN 457/9-1 and with J. Weickert, D. Didas and B. Burgeth (MIA-group, Saarland University) within the DFG-project WE 2602/1-3. Moreover, we thank the MIA group for providing us the test images in Figs. 5 and 7.

References

1. *The MOSEK Optimization Toolbox*. <http://www.mosek.com>.
2. P. J. Basser. Inferring microstructural features and the physical state of tissues from diffusion-weighted images. *Nuclear Magnetic Resonance in Biomedicine*, 8:333–334, 1995.
3. P. J. Basser, J. Mattiello, and D. LeBihan. MR diffusion tensor spectroscopy and imaging. *Biophysical Journal*, 66:259–267, 1994.
4. A. Y. Bezhaev and V. A. Vasilenko. *Variational Theory of Splines*. Kluwer, New York, 2001.
5. J. Bigün, G. H. Granlund, and J. Wiklund. Multidimensional orientation estimation with applications to texture analysis and optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(8):775–790, Aug. 1991.
6. T. Brox, J. Weickert, B. Burgeth, and P. Mrázek. Nonlinear structure tensors. *Image and Vision Computing*, 24(1):41–55, 2006.
7. B. Burgeth, S. Didas, L. Florack, and J. Weickert. Singular pdes for the processing of matrix-valued data. *Lecture Notes in Computer Science*. Springer, Berlin, to appear.
8. A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, (20):89–97, 2004.
9. A. Chambolle and P.-L. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76:167–188, 1997.
10. T. F. Chan, A. Marquina, and P. Mulet. High-order total variation-based image restoration. *SIAM Journal on Scientific Computing*, 22(2):503–516, 2000.
11. P. Copley and L. L. Schumaker. On pLg-splines. *Journal of Approximation Theory*, 23:1–28, 1978.
12. P. L. Davies and A. Kovac. Local extremes, runs, strings and multiresolution. *Annals of Statistics*, 29:1–65, 2001.
13. C. de Boor and R. E. Lynch. On splines and their minimum properties. *Journal of Mathematics and Mechanics*, (13):953–968, 1966.
14. J. Duchon. Splines minimizing rotation-invariant seminorms in sobolev spaces. In *Constructive Theory of Functions of Several Variables*, pages 85–100, Berlin, 1997. Springer-Verlag.
15. S. D. Fisher and J. W. Jerome. Spline solutions to l_1 extremal problems in one and several variables. *Journal of Approximation Theory*, 13:73–83, 1975.

16. W. Förstner and E. Gülch. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In *Proc. ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data*, pages 281–305, Interlaken, Switzerland, June 1987.
17. D. Goldfarb and W. Yin. Second order cone programming methods for total variation–based image restoration. *SIAM J. Scientific Computing*, 2(27):622–645, 2005.
18. G. H. Granlund and H. Knutsson. *Signal Processing for Computer Vision*. Kluwer, Dordrecht, 1995.
19. C. G. Harris and M. Stephens. A combined corner and edge detector. In *Proc. Fourth Alvey Vision Conference*, pages 147–152, Manchester, England, Aug. 1988.
20. W. Hinterberger, M. Hintermüller, K. Kunisch, M. von Oehsen, and O. Scherzer. Tube methods for BV regularization. *Journal of Mathematical Imaging and Vision*, 19:223 – 238, 2003.
21. W. Hinterberger and O. Scherzer. Variational methods on the space of functions of bounded Hessian for convexification and denoising. Technical report, University of Innsbruck, Austria, 2003.
22. W. Hintermüller and W. Kunisch. Total bounded variation regularization as a bilaterally constrained optimization problem. *SIAM J. Appl. Math.*, 4(64):1311–1333, 2004.
23. J. M. Hyman and M. J. Shashkov. Natural discretizations for the divergence, gradient, and curl on logically rectangular grids. *Comput. Math. Appl.*, 33(4):81–104, 1997.
24. J. P. Laurent. Quadratic convex analysis and splines. volume 76 of *Internat. Series of Numerical Math.*, pages 17–43. Birkhäuser, Basel, 1986.
25. P. J. Laurent. Inf-convolution splines. *Constructive Approximation*, 7:469–484, 1991.
26. M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebet. Applications of second–order cone programming. *Linear Algebra and its Applications*, 284:193–228, 1998.
27. M. Lysaker, A. Lundervold, and X. Tai. Noise removal using fourth-order partial differential equations with applications to medical magnetic resonance images in space and time. Technical Report CAM-02-44, Department of Mathematics, University of California at Los Angeles, CA, U.S.A., 2002.
28. E. Mammen and S. van de Geer. Locally adaptive regression splines. *Annals of Statistics*, 25(1):387–413, 1997.
29. O. L. Mangasarian and L. L. Schumaker. Discrete splines via mathematical programming. *SIAM Journal on Control*, 9(2):174–183, 1971.

30. O. L. Mangasarian and L. L. Schumaker. Best summation formulae and discrete splines via mathematical programming. *SIAM Journal on Numerical Analysis*, 10(3):448–459, 1973.
31. H. Mittelmann. An independent benchmarking of SDP and SOCP solvers. *Mathematical Programming Series B*, 95(2):407–430, 2003.
32. A. R. Rao and B. G. Schunck. Computing oriented texture fields. *CVGIP: Graphical Models and Image Processing*, 53:157–185, 1991.
33. L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica A*, 60:259–268, 1992.
34. C. Schnörr. A study of a convex variational diffusion approach for image segmentation and feature extraction. *Journal of Mathematical Imaging and Vision*, 8(3):271–292, 1998.
35. I. Schoenberg. Contribution to the problem of approximation of equidistant data by analytic functions. *Quarterly of Applied Mathematics*, 4:45–99 and 112–141, 1946.
36. L. L. Schumaker. *Spline Functions: Basic Theory*. Wiley and Sons, New York, 1981.
37. G. Steidl, S. Didas, and J. Neumann. Splines in higher order TV regularization. *International Journal of Computer Vision*, 70(3):241–255, 2006.
38. G. Steidl, S. Setzer, and D. Didas. Combined ℓ_2 data and gradient fitting in conjunction with ℓ_1 regularization. *Adv. Comput. Math.*, to appear.
39. G. Steidl, S. Setzer, B. Popilka, and B. Burgeth. Restoration of matrix fields by second order cone programming. *Computing*, to appear.
40. J. S. Sturm. *Using SeDuMi 1.02, A Matlab toolbox for optimization over symmetric cones*. <http://fewcal.kub.nl/sturm>, 2001.
41. D. Tschumperlé and R. Deriche. Diffusion tensor regularization with constraints preservation. In *Proc. 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 948–953, Kauai, HI, 2001. IEEE Computer Science Press.
42. M. Unser and T. Blu. Cardinal exponential splines: part i -theory and filtering algorithms. *tsp*, 53(4):1425–1438, 2005.
43. G. Wahba. *Spline Models for Observational Data*. SIAM, Philadelphia, 1990.
44. J. Weickert and T. Brox. Diffusion and regularization of vector- and matrix-valued images. In M. Z. Nashed and O. Scherzer, editors, *Inverse Problems, Image Analysis, and Medical Imaging*, volume 313 of *Contemporary Mathematics*, pages 251–268. AMS, Providence, 2002.
45. M. Welk, G. Steidl, and J. Weickert. Locally analytic schemes: a link between diffusion filtering and wavelet shrinkage. *Applied and Compu-*

tational Harmonic Analysis, to appear.

46. Y.-L. You and M. Kaveh. Fourth-order partial differential equations for noise removal. *IEEE Transactions on Image Processing*, 9(10):1723–1730, 2000.
47. J. Yuan, C. Schnörr, and G. Steidl. Simultaneous higher order optical flow estimation and decomposition. *SIAM J. Sci. Comput.*, to appear.

Gabriele Steidl and Simon Setzer
University of Mannheim
A5
68131 Mannheim, Germany
steidl@math.uni-mannheim.de
<http://www.kiwi.uni-mannheim.de>