A note on the dual treatment of higher order regularization functionals

G. Steidl, Mannheim

April 12, 2005

Abstract

In this paper, we apply the dual approach developed by A. Chambolle for the Rudin–Osher–Fatemi model to regularization functionals with higher order derivatives. We emphasize the linear algebra point of view by consequently using matrix–vector notation. Numerical examples demonstrate the differences between various second order regularization approaches.

AMS Subject Classification:
Key words:

1 Introduction

In this paper, we are interested in constructing for a given function f a function u that minimizes

$$\frac{1}{2}\int_\Omega (u-f)^2 \, \mathrm{d}x + \lambda \int_\Omega |\mathcal{J}(u)| \, \mathrm{d}x, \tag{1}$$

where the regularization functional $\mathcal{I}(u) := \int_\Omega |\mathcal{J}(u)| \, \mathrm{d}x$ is convex and positive homogeneous of degree one, i.e., $\mathcal{I}(\alpha u) = \alpha \mathcal{I}(u)$ for every u and $\alpha > 0$. By Ω we denote an interval $[a,b]$ in the onedimensional setting and a rectangle $[a,b] \times [c,d]$ in the twodimensional case. There is a large amount of literature on applications of (1) with various, in general nonlinear, regularization functionals in image processing. Here we only refer to the books [1, 27] for an overview.

A frequently applied approach in image denoising and segmentation is the Rudin–Osher–Fatemi (ROF) model [23] with the gradient $\mathcal{J}(u) := \nabla u$.

Meanwhile there exist various solution methods for the corresponding minimization problem, see [26] and the references therein. Most of these methods introduce a small additional smoothing parameter to cope with the non differenciability of $|\nabla u|$. Legendre–Fenchel dualization techniques as proposed, e.g., in [5, 3], avoid such parameter and will be the method of choice in this paper. We remark that another wavelet inspired technique without additional smoothing parameter was presented in [28].

In recent years, there has been a growing interest in higher order variational methods. In [4], the minimizer of the functional $\int_\Omega (f-u)^2 + \lambda_1 |\nabla u - \nabla v| + \lambda_2 |\nabla^2 v| \, \mathrm{d}x$ was studied and in [24] the asymptotical case $\lambda_1 \to \infty$ was considered. In [6], a second order term (directed Laplacian) was added to the TV functional in order to reduce the staircasing effect known from TV regularization. For the same purpose, a regularization functional of the form $\int_\Omega \varphi(|\triangle u|) \, \mathrm{d}x$ with φ corresponding to the Perona–Malik diffusivity [21] was considered in [29]. In [14], second order regularization functionals were applied in magnetic resonance imaging and in [12] for denoising and convexification. Higher order regularization functionals were embedded in a scale–space context in [19].

In this paper, we will apply the dual approach developed by A. Chambolle [3] for the ROF model to regularization functionals with higher order derivatives. To be more concrete, we are only concerned with a discrete version of (1), where the functions are considered at equispaced points. We arrange the function values in corresponding vectors, where we reshape twodimensional arrays columnwise. Then, with a discretization J of $\mathcal{J}$ and the usual vector norms, we obtain

$$\frac{1}{2}\|\boldsymbol{u} - \boldsymbol{f}\|_2^2 + \lambda \|J(\boldsymbol{u})\|_1 \to \min, \tag{2}$$

where $\|J(\boldsymbol{u})\|_1$ is a lower–semicontinuous, proper convex function in $\boldsymbol{u}$ which is again one–homogeneous. We will solve this problem by considering its dual formulation. Problem (2) is equivalent to the computation of $\boldsymbol{u} = \boldsymbol{f} - \boldsymbol{v}$, where $\boldsymbol{v}$ satisfies the constrained convex optimization problem

$$\|\boldsymbol{f} - \boldsymbol{v}\|_2^2 \to \min, \qquad \text{subject to } \boldsymbol{v} \in \mathcal{V}_\lambda, \tag{3}$$

where $\mathcal{V}_\lambda := \{\boldsymbol{v} \in \mathbb{R}^N : (\boldsymbol{v}, \boldsymbol{w}) \le \lambda \|J(\boldsymbol{w})\|_1 \; \forall \boldsymbol{w} \in \mathbb{R}^N\}$, see Proposition 1 in the appendix. In the following, we apply this dual approach to various regularization functionals with higher order derivatives. We prefer to use matrix–vector notation which makes the MATLAB implementation of the corresponding algorithms very comfortable. Our operators J are in general of the form $J(\boldsymbol{w}) = g(\boldsymbol{A}\boldsymbol{w})$ with an (M, N) matrix $\boldsymbol{A}$ of rank smaller than N and with a function $g : \mathbb{R}^M \to \mathbb{R}^{\tilde{M}}$ satisfying $g(\boldsymbol{0}) = \boldsymbol{0}$. Then it is not hard to prove that

$$\mathcal{V}_\lambda = \{\boldsymbol{v} \in \mathcal{R}(\boldsymbol{A}^T) : (\boldsymbol{v}, \boldsymbol{w}) \le \lambda \|J(\boldsymbol{w})\|_1 \ \forall \boldsymbol{w} \in \mathbb{R}^N\},$$

where $\mathcal{R}(\boldsymbol{A}^T)$ denotes the range of $\boldsymbol{A}^T$, see Proposition 2 in the appendix.

This paper is organized as follows. To make the general idea more comprehensible, we start by considering the onedimensional setting in Section 2. Moreover, we explain the close relation of (3) to the support vector regression (SVR) problem with spline kernels. Section 3 deals with the twodimensional problem. First, we recapitulate A. Chambolle's approach for the ROF model using our matrix–vector notation. Then we apply the idea to various functionals with second order derivatives. Section 4 contains numerical experiments. Finally, the appendix verifies the equivalence of (2) and (3) and the restriction of $\mathcal{V}_\lambda$.

2 Onedimensional Setting

We find it useful to consider the onedimensional case with derivative operators $\mathcal{J}(u) = u^{(m)}$ of various orders $m \in \mathbb{N}$ first. As discretization of the first derivative of u, we use the forward difference $u'(jh) \approx (u((j+1)h) - u(jh))/h$, $j = 1, \ldots, N-1$ with $h := (b-a)/N$. For simplicity, we assume in the following that $h = 1$. As disrete version of $u^{(m)}$ we use its *m–th forward difference* $J(\boldsymbol{u}) := \boldsymbol{D}_{N,m}\boldsymbol{u}$ with $\boldsymbol{D}_{N,m} := \tilde{\boldsymbol{D}}_{N,m} \cdots \tilde{\boldsymbol{D}}_{N,1} \in \mathbb{R}^{N-m,N}$ and *forward difference matrices*

$$\tilde{\boldsymbol{D}}_{N,r} := \begin{pmatrix} -1 & 1 & 0 & \ldots & 0 \\ 0 & -1 & 1 & \ldots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & & -1 & 1 \end{pmatrix} \in \mathbb{R}^{N-r,N-r+1}, \qquad r < N.$$

Regarding the discrete momentum properties of the finite forward differences, we see that $\boldsymbol{v} \in \mathcal{R}(\boldsymbol{D}^T_{N,m})$ if and only if

$$\sum_{j=1}^{N} j^r v_j = 0, \qquad r = 0, \ldots, m-1. \tag{4}$$

For $r = 0$, this condition is in particular fulfilled if $\boldsymbol{v}$ is white Gaussian noise. The matrix $\boldsymbol{D}_{N,m}$ has rank $N - m$. Hence, for any $\boldsymbol{v} \in \mathcal{R}(\boldsymbol{D}^T_{N,m})$, there exists a unique $\boldsymbol{V} \in \mathbb{R}^{N-m}$ such that $\boldsymbol{v} = \boldsymbol{D}^T_{N,m}\boldsymbol{V}$ and

$$(\boldsymbol{v}, \boldsymbol{w}) = (\boldsymbol{D}^T_{N,m}\boldsymbol{V}, \boldsymbol{w}) = (\boldsymbol{V}, \boldsymbol{D}_{N,m}\boldsymbol{w}) \le \|\boldsymbol{V}\|_\infty \|\boldsymbol{D}_{N,m}\boldsymbol{w}\|_1 \quad \forall \boldsymbol{w} \in \mathbb{R}^N.$$

The inequality is sharp in the sense that there exists no constant $C < \|\boldsymbol{V}\|_\infty$ such that $(\boldsymbol{v}, \boldsymbol{w}) \le C\|\boldsymbol{D}_{N,m}\boldsymbol{w}\|_1$ holds true for all $\boldsymbol{w} \in \mathbb{R}^N$. Consequently, $\mathcal{V}_\lambda = \{\boldsymbol{v} := \boldsymbol{D}_{N,m}^T \boldsymbol{V} : \|\boldsymbol{V}\|_\infty \le \lambda,\ \boldsymbol{V} \in \mathbb{R}^{N-m}\}$ and problem (3) is equivalent to

$$\|\boldsymbol{f} - \boldsymbol{D}_{N,m}^T \boldsymbol{V}\|_2^2 \to \min, \qquad \text{subject to} \quad \|\boldsymbol{V}\|_\infty \le \lambda. \tag{5}$$

This problem which is just the Fenchel dual of (2) can be solved by quadratic programming (QP) methods. The final solution is $\boldsymbol{u} = \boldsymbol{f} - \boldsymbol{D}_{N,m}^T \boldsymbol{V}$. By (4), the first m discrete moments of $\boldsymbol{u}$ coincide with those of $\boldsymbol{f}$.

In [25] we have examined higher order TV regularization in one dimension from a different point of view, namely with respect to its relation to spline interpolation with variable knots and to SVR with discrete spline kernels. Finishing [25], we became aware of its close relation to Legendre–Fenchel dualization techniques. Since this was indeed the motivation to write this paper, we briefly want to explain the relation to [25]. By adding an appropriate last row to $\tilde{\boldsymbol{D}}_{N,1} \in \mathbb{R}^{N-1,N}$, we introduce the Toeplitz matrix

$$\boldsymbol{D}_{-1} == \begin{pmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & & & & & \vdots \\ 0 & 0 & 0 & \dots & -1 & 1 \\ 0 & 0 & 0 & \dots & 0 & -1 \end{pmatrix} \in \mathbb{R}^{N,N},$$

which has the upper triangular matrix with coefficients -1 as inverse matrix. Then the first $N-m$ rows of $\boldsymbol{D}_{-1}^m$ coincide with those of $\boldsymbol{D}_{N,m}$. Now, for any $\boldsymbol{v} \in \mathbb{R}^N$, there exists a unique $\boldsymbol{V} \in \mathbb{R}^N$ such that $\boldsymbol{v} = (\boldsymbol{D}_{-1}^m)^T \boldsymbol{V}$. Assuming that $V_{N-j} = 0$ for $j = 0, \dots, m-1$ which is equivalent to the restrictions (4) on $\boldsymbol{v}$, we obtain with $\tilde{\boldsymbol{V}} := (V_1, \dots, V_{N-m})^T$ that

$$(\boldsymbol{v}, \boldsymbol{w}) = \big((\boldsymbol{D}_{-1}^m)^T \boldsymbol{V}, \boldsymbol{w}\big) = (\tilde{\boldsymbol{V}}, \boldsymbol{D}_{N,m}\boldsymbol{w}) \le \|\boldsymbol{V}\|_\infty \|\boldsymbol{D}_{N,m}\boldsymbol{w}\|_1 \quad \forall \boldsymbol{w} \in \mathbb{R}^N,$$

where the inequality is sharp. Consequently, (5) can be rewritten as

$$\begin{gathered} \|\boldsymbol{f} - (\boldsymbol{D}_{-1}^m)^T \boldsymbol{V}\|_2^2 \to \min, \\ \text{subject to} \quad \|\boldsymbol{V}\|_\infty \le \lambda \quad \text{and} \quad V_{N-j} = 0,\ j = 0, \dots, m-1. \end{gathered}$$

Defining $\boldsymbol{F}$ and $\boldsymbol{U}$ by $\boldsymbol{f} = (\boldsymbol{D}_{-1}^m)^T \boldsymbol{F}$ and $\boldsymbol{u} = (\boldsymbol{D}_{-1}^m)^T \boldsymbol{U}$, respectively, this problem becomes

$$\begin{gathered} \|(\boldsymbol{D}_{-1}^m)^T \boldsymbol{U}\|_2^2 \to \min, \\ \text{subject to} \quad \|\boldsymbol{F} - \boldsymbol{U}\|_\infty \le \lambda \quad \text{and} \quad F_{N-j} = U_{N-j},\ j = 0, \dots, m-1. \end{gathered} \tag{6}$$

For $m = 1$, the solution of (6) can be computed by the so–called *taut–string algorithm*. This algorithm has complexity $\mathcal{O}(N)$ and is much faster than QP methods, see [8, 15]. For tube methods in more than one dimension, we refer to [11]. In [25], we have shown that for given $\boldsymbol{F} \in \mathbb{R}^N$, the solution $\boldsymbol{U}$ of (6) is a discrete spline of degree $2m - 1$ which interpolates $\boldsymbol{F} \pm \lambda$ at its spline knots. For discrete splines, we refer to [16]. On the other hand, we have verified that $\boldsymbol{U}$ can be interpreted as sparse approximation of $\boldsymbol{F}$ in the sense of [10] or as solution of a SVR problem with discrete spline kernel. To see the last relation in the context of this paper, let us consider $\boldsymbol{u} = \boldsymbol{D}_{-1}^m (\boldsymbol{D}_{-1}^m)^T \boldsymbol{\psi}$ as discrete counterpart of $u = \psi^{(2m)}$. Then $\psi = k * u$, where k is the fundamental solution of the $(2m)$–th derivative operator, i.e., the spline $k(x) = x_+^{2m-1}$. Here $(x)_+ := \max\{0, x\}$. With $\boldsymbol{K} := \left(\boldsymbol{D}_{-1}^m (\boldsymbol{D}_{-1}^m)^T\right)^{-1}$ we have that $\boldsymbol{\psi} = \boldsymbol{K}\boldsymbol{u}$. Let $U := \psi^{(m)} = k * u^{(m)}$. Its discrete version reads $\boldsymbol{U} = \boldsymbol{K}\boldsymbol{D}_{-1}^m \boldsymbol{u}$. Setting $\boldsymbol{c} := \boldsymbol{D}_{-1}^m \boldsymbol{u}$, our minimization problem becomes

$$\boldsymbol{c}^T \boldsymbol{K} \boldsymbol{c} \to \min,$$
$$\text{subject to} \quad \|\boldsymbol{F} - \boldsymbol{K}\boldsymbol{c}\|_\infty \le \lambda \quad \text{and} \quad F_{N-j} = (Kc)_{N-j},\ j = 0, \ldots, m-1.$$

This is the SVR problem with discrete spline kernel considered in [25].

3 Twodimensional Setting

For simplicity, we restrict our attention to quadratic (n, n) arrays and reshape them columnwise into a vector of length $N = n^2$. By adding a zero row to the forward difference matrix $\tilde{\boldsymbol{D}}_{n,1} \in \mathbb{R}^{n-1,n}$, we define the matrix

$$\boldsymbol{D}_0 := \begin{pmatrix} -1 & 1 & 0 & \ldots & 0 & 0 \\ 0 & -1 & 1 & \ldots & 0 & 0 \\ \vdots & & & & & \vdots \\ 0 & 0 & 0 & \ldots & -1 & 1 \\ 0 & 0 & 0 & \ldots & 0 & 0 \end{pmatrix} \in \mathbb{R}^{n,n}.$$

3.1 First order derivatives

In this subsection, we are concerned with the ROF model $\mathcal{J}(u) := \nabla u$. Since we will apply similar ideas for regularization functionals with higher order derivatives in the next subsection, we briefly reconsider the approach of A. Chambolle [3] using our matrix–vector notation. As discrete versions of ∇u and its adjoint $\nabla^* U = -\operatorname{div} U$ we use $\boldsymbol{\mathcal{D}}\boldsymbol{u}$ and $\boldsymbol{\mathcal{D}}^T \boldsymbol{U}$, respectively, where

$$\boldsymbol{\mathcal{D}} := \begin{pmatrix} \boldsymbol{I}_n \otimes \boldsymbol{D}_0 \\ \boldsymbol{D}_0 \otimes \boldsymbol{I}_n \end{pmatrix} \in \mathbb{R}^{2N,N}, \quad \boldsymbol{\mathcal{D}}^T = \left(\ \boldsymbol{I}_n \otimes \boldsymbol{D}_0^T \ , \ \boldsymbol{D}_0^T \otimes \boldsymbol{I}_n \ \right) \in \mathbb{R}^{N,2N}$$

and where $\otimes$ denotes the Kronecker product. Both matrices have rank $N-1$. This is just the discretization considered in [3]. Now the discrete version of $|\nabla u| = (u_x^2 + u_y^2)^{1/2}$ reads $J(\boldsymbol{u}) = |\boldsymbol{\mathcal{D}}\boldsymbol{u}|$, where

$$\left|\begin{pmatrix} \boldsymbol{F}^1 \\ \boldsymbol{F}^2 \end{pmatrix}\right| := \left((\boldsymbol{F}^1)^2 + (\boldsymbol{F}^2)^2\right)^{1/2} = \left(\boldsymbol{F}^1 \circ \boldsymbol{F}^1 + \boldsymbol{F}^2 \circ \boldsymbol{F}^2\right)^{1/2} \in \mathbb{R}^N$$

and $\circ$ denotes the componentwise vector product. Since the columns of $\boldsymbol{\mathcal{D}}^T$ add up to zero, we see that $\boldsymbol{v} \in \mathcal{R}(\boldsymbol{\mathcal{D}}^T)$ if and only if

$$\sum_{j=1}^{N} v_j = 0. \tag{7}$$

Then we obtain for all $\boldsymbol{V} \in \mathbb{R}^{2N}$ with $\boldsymbol{v} = \boldsymbol{\mathcal{D}}^T\boldsymbol{V}$ that

$$(\boldsymbol{v}, \boldsymbol{w}) = (\boldsymbol{\mathcal{D}}^T\boldsymbol{V}, \boldsymbol{w}) = (\boldsymbol{V}, \boldsymbol{\mathcal{D}}\boldsymbol{w}) = \left(\boldsymbol{V}^1, (\boldsymbol{I}_n \otimes \boldsymbol{D}_0)\boldsymbol{w}\right) + \left(\boldsymbol{V}^2, (\boldsymbol{D}_0 \otimes \boldsymbol{I}_n)\boldsymbol{w}\right).$$

Applying Schwarz's inequality to the sum of corresponding components in both inner products, we get

$$(\boldsymbol{v}, \boldsymbol{w}) \le (|\boldsymbol{V}|, |\boldsymbol{\mathcal{D}}\boldsymbol{w}|) \le \|\,|\boldsymbol{V}|\,\|_\infty \|\,|\boldsymbol{\mathcal{D}}\boldsymbol{w}|\,\|_1 \quad \forall \boldsymbol{w} \in \mathbb{R}^N.$$

By [2], we have that $\mathcal{V}_\lambda = G_\lambda^d := \{\boldsymbol{v} \in \mathcal{R}(\mathcal{D}^T) : \|\boldsymbol{v}\|_{G^d} \le \lambda\}$, where d abbreviates 'discrete' and where $\|\boldsymbol{v}\|_{G^d} := \inf_{\boldsymbol{v}=\mathcal{D}^T\boldsymbol{V}} \|\,|\boldsymbol{V}|\,\|_\infty$ is the *discrete version of Meyer's G–norm* [18]. Recently, the G–norm was generalized to second order derivatives in [20] in the continuous setting. This is related to the next subsection. Instead of problem (3) we solve

$$\|\boldsymbol{f} - \boldsymbol{\mathcal{D}}^T\boldsymbol{V}\|^2 \to \min, \quad \text{subject to} \quad \|\,|\boldsymbol{V}|\,\|_\infty \le \lambda. \tag{8}$$

This is a quadratic problem with convex constraints. The Lagrangian of (8) is given by

$$\mathcal{L}(\boldsymbol{V}, \boldsymbol{\alpha}) = \boldsymbol{V}^T\boldsymbol{\mathcal{D}}\boldsymbol{\mathcal{D}}^T\boldsymbol{V} - 2\boldsymbol{f}^T\boldsymbol{\mathcal{D}}^T\boldsymbol{V} + \boldsymbol{f}^T\boldsymbol{f} - \boldsymbol{\alpha}^T\left(\lambda^2\boldsymbol{e} - (\boldsymbol{V}^1)^2 - (\boldsymbol{V}^2)^2\right),$$

where $\boldsymbol{e}$ denotes the vector with components one and $\boldsymbol{\alpha} \in \mathbb{R}^N$ with $\boldsymbol{\alpha} \ge 0$ componentwise. A necessary and sufficient condition for $\boldsymbol{V}$ to produce a minimum of (8) is that the gradient of $\mathcal{L}$ with respect to $\boldsymbol{V}$ is the zero vector, i.e.,

$$\nabla_{\boldsymbol{V}}\mathcal{L}(\boldsymbol{V}, \boldsymbol{\alpha}) = 2\,\boldsymbol{\mathcal{D}}\boldsymbol{\mathcal{D}}^T\boldsymbol{V} - 2\,\boldsymbol{\mathcal{D}}\boldsymbol{f} + 2\begin{pmatrix} \boldsymbol{\alpha} \circ \boldsymbol{V}^1 \\ \boldsymbol{\alpha} \circ \boldsymbol{V}^2 \end{pmatrix} = \boldsymbol{0}. \tag{9}$$

Let $\boldsymbol{W} := \boldsymbol{\mathcal{D}}\boldsymbol{\mathcal{D}}^T\boldsymbol{V} - \boldsymbol{\mathcal{D}}\boldsymbol{f}$. If $\alpha_j > 0$, then, by the Karush–Kuhn–Tucker conditions, the j–th constraint in (8) has to be the equality $(\boldsymbol{V}_j^1)^2 + (\boldsymbol{V}_j^2)^2 = \lambda^2$. Consequently, by (9), $W_j^1 = -\alpha_j V_j^1$ and $W_j^2 = -\alpha_j V_j^2$ so that

$$\left(W_j^1\right)^2 + \left(W_j^2\right)^2 = \alpha_j^2 \lambda^2. \tag{10}$$

If $\alpha_j = 0$, then (10) holds obviousely true. Hence we can replace $\boldsymbol{\alpha}$ in (9) by (10) and obtain

$$\boldsymbol{W} + \frac{1}{\lambda}\begin{pmatrix} |\boldsymbol{W}| \\ |\boldsymbol{W}| \end{pmatrix} \circ \boldsymbol{V} = \boldsymbol{0}. \tag{11}$$

By [7, Theorem 9.2-4], the Karush–Kuhn–Tucker conditions summarized in (11) are also sufficient for $\boldsymbol{V}$ to provide a minimum of (8). To solve (11), A. Chambolle [3] suggested the semi–implicit gradient descent approach

$$\boldsymbol{V}^{(k+1)} = \boldsymbol{V}^{(k)} - \tau\left(\boldsymbol{W}^{(k)} + \frac{1}{\lambda}\begin{pmatrix} |\boldsymbol{W}^{(k)}| \\ |\boldsymbol{W}^{(k)}| \end{pmatrix} \circ \boldsymbol{V}^{(k+1)}\right).$$

In summary, we obtain the following algorithm:

Algorithm 1.
Input: $\boldsymbol{f}$ and $\boldsymbol{V}^{(0)} := \boldsymbol{0}$.

Repeat for $k = 0$ until a stopping criterion is reached

$$\begin{aligned} \boldsymbol{W}^{(k)} &:= \boldsymbol{\mathcal{D}}\boldsymbol{\mathcal{D}}^T\boldsymbol{V}^{(k)} - \boldsymbol{\mathcal{D}}\boldsymbol{f}, \\ \boldsymbol{V}^{(k+1)} &:= \left(\boldsymbol{1} + \frac{\tau}{\lambda}\begin{pmatrix} |\boldsymbol{W}^{(k)}| \\ |\boldsymbol{W}^{(k)}| \end{pmatrix}\right)^{-1} \circ \left(\boldsymbol{V}^{(k)} - \tau\boldsymbol{W}^{(k)}\right), \\ k &:= k+1, \end{aligned}$$

where the inverse is taken componentwise.

Output: $\boldsymbol{u} := \boldsymbol{f} - \boldsymbol{\mathcal{D}}^T\boldsymbol{V}^{(k)}$.

A. Chambolle proved that $\boldsymbol{\mathcal{D}}^T\boldsymbol{V}^{(k)}$ converges for $k \to \infty$ to the solution $\boldsymbol{v}$ of (3) if

$$\tau \leq 1/\|\boldsymbol{\mathcal{D}}^T\|_2^2.$$

Now $\|\boldsymbol{\mathcal{D}}^T\|_2^2 = \rho(\boldsymbol{\mathcal{D}}^T\boldsymbol{\mathcal{D}})$, where ρ denotes the spectral radius of the matrix. The matrix $\boldsymbol{\mathcal{D}}^T\boldsymbol{\mathcal{D}}$ is well–known from the five point finite difference discretization of the Laplacian with Neumann boundary conditions. The eigenvalues of this matrix are given by $4(\sin(j\pi/(2n))^2 + \sin(k\pi/(2n))^2)$, $j,k = 0,\ldots,n-1$. Thus, $\|\boldsymbol{\mathcal{D}}^T\|_2^2 = 8$. However, in numerical experiments convergence was observed for $\tau \leq 1/4$.

3.2 Second order derivatives

Starting with the Hessian $\nabla^2 u := \begin{pmatrix} u_{xx} & u_{xy} \\ u_{yx} & u_{yy} \end{pmatrix}$ of u, we consider the following functionals:

1. the trace of the Hessian, i.e., the Laplacian

$$\mathcal{J}(u) := \triangle u = u_{xx} + u_{yy},$$

2. the Frobenius norm of the Hessian mentioned also in [9]

$$\mathcal{J}(u) := \left(u_{xx}^2 + u_{yy}^2 + u_{xy}^2 + u_{yx}^2\right)^{1/2},$$

3. the modified Laplacian considered in [14]

$$\mathcal{J}(u) := |u_{xx}| + |u_{yy}|.$$

The Laplacian and the Frobenius norm of the Hessian are the most straightforward functionals with second order derivatives that are rotationally invariant. Although the modified Laplacian lacks rotation invariance, we just include it for comparisons. We will see that our discrete version of the Frobenius norm of the Hessian can be handled as in the previous subsection by a semi–implicit gradient descent method while the Laplacian and the modified Laplacian lead to QP.

1. The Laplacian. As discretization of the Laplacian we use $J(\boldsymbol{u}) := \boldsymbol{\mathcal{D}}_\triangle \boldsymbol{u}$, where

$$\boldsymbol{\mathcal{D}}_\triangle := \boldsymbol{\mathcal{D}}^T \boldsymbol{\mathcal{D}} = \boldsymbol{I}_n \otimes \boldsymbol{D}_0^T \boldsymbol{D}_0 + \boldsymbol{D}_0^T \boldsymbol{D}_0 \otimes \boldsymbol{I}_n$$

denotes the symmetric matrix of rank $N-1$ arising from the five point finite difference discretization of the Laplacian with Neumann boundary conditions. Since the columns of $\boldsymbol{\mathcal{D}}^T \boldsymbol{\mathcal{D}}$ add up to zero, we have that $\boldsymbol{v} \in \mathcal{R}(\boldsymbol{\mathcal{D}}_\triangle)$ if and only if (7) holds true. Then we obtain for all $\boldsymbol{V} \in \mathbb{R}^N$ with $\boldsymbol{v} = \boldsymbol{\mathcal{D}}_\triangle \boldsymbol{V}$ that

$$(\boldsymbol{v}, \boldsymbol{w}) = (\boldsymbol{\mathcal{D}}_\triangle \boldsymbol{V}, \boldsymbol{w}) = (\boldsymbol{V}, \boldsymbol{\mathcal{D}}_\triangle \boldsymbol{w}) \le \|\boldsymbol{V}\|_\infty \|\boldsymbol{\mathcal{D}}_\triangle \boldsymbol{w}\|_1 \quad \forall \boldsymbol{w} \in \mathbb{R}^N. \tag{12}$$

Regarding that the nullspace of $\boldsymbol{\mathcal{D}}_\triangle$ is given by $\{c\boldsymbol{e} : c \in \mathbb{R}\}$, we see that $\boldsymbol{v} = \boldsymbol{\mathcal{D}}_\triangle \boldsymbol{V}$ if and only if $\boldsymbol{v} = \boldsymbol{\mathcal{D}}_\triangle(\boldsymbol{V} + c\boldsymbol{e})$. Choosing $\mu := (V_{\min} + V_{\max})/2$ with the maximal and minimal components $V_{\min}$ and $V_{\max}$ of $\boldsymbol{V}$, respectively, we obtain that the components of $\tilde{\boldsymbol{V}} := \boldsymbol{V} - \mu \boldsymbol{e}$ fulfill $-(V_{\max} - V_{\min})/2 \le \tilde{V}_j \le$

$(V_{\max}-V_{\min})/2$, where we have lower and upper equality for some components j_- and j_+, respectively. Thus, $\|\tilde{\boldsymbol{V}}\|_\infty = (V_{\max} - V_{\min})/2$ and $\|\tilde{\boldsymbol{V}} + c\boldsymbol{e}\|_\infty = \|\tilde{\boldsymbol{V}}\|_\infty + |c|$. Consequently, $\min\limits_{\boldsymbol{v}=\boldsymbol{\mathcal{D}}_\triangle \boldsymbol{V}} \|\boldsymbol{V}\|_\infty = \|\tilde{\boldsymbol{V}}\|_\infty$. Choosing $\boldsymbol{\mathcal{D}}_\triangle \boldsymbol{w}$ as vector consisting of zeros except for $(\mathcal{D}_\triangle w)_{j_-} := -1$ and $(\mathcal{D}_\triangle w)_{j_+} := 1$, we obtain in (12) the equality $(\boldsymbol{v}, \boldsymbol{w}) = (\tilde{\boldsymbol{V}}, \boldsymbol{\mathcal{D}}_\triangle \boldsymbol{w}) = 2\|\tilde{\boldsymbol{V}}\|_\infty = \|\tilde{\boldsymbol{V}}\|_\infty \|\boldsymbol{\mathcal{D}}_\triangle \boldsymbol{w}\|_1$.

Finally, we solve

$$\|\boldsymbol{f} - \boldsymbol{\mathcal{D}}_\triangle \boldsymbol{V}\|_2^2 \to \min, \quad \text{subject to} \quad \|\boldsymbol{V}\|_\infty \le \lambda. \tag{13}$$

2. The Frobenius norm of the Hessian. We discretize the Frobenius norm of the Hessian by $J(\boldsymbol{u}) := |\boldsymbol{\mathcal{D}}_H \boldsymbol{u}|$, where

$$\boldsymbol{\mathcal{D}}_H := \begin{pmatrix} (\boldsymbol{I}_n \otimes \boldsymbol{D}_0^T)(\boldsymbol{I}_n \otimes \boldsymbol{D}_0) \\ (\boldsymbol{D}_0^T \otimes \boldsymbol{I}_n)(\boldsymbol{D}_0 \otimes \boldsymbol{I}_n) \\ (\boldsymbol{I}_n \otimes \boldsymbol{D}_0^T)(\boldsymbol{D}_0 \otimes \boldsymbol{I}_n) \\ (\boldsymbol{D}_0 \otimes \boldsymbol{I}_n)(\boldsymbol{I}_n \otimes \boldsymbol{D}_0^T) \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}_n \otimes \boldsymbol{D}_0^T \boldsymbol{D}_0 \\ \boldsymbol{D}_0^T \boldsymbol{D}_0 \otimes \boldsymbol{I}_n \\ \boldsymbol{D}_0^T \otimes \boldsymbol{D}_0 \\ \boldsymbol{D}_0 \otimes \boldsymbol{D}_0^T \end{pmatrix}$$

and where for $\boldsymbol{F} := ((\boldsymbol{F}^1)^T, (\boldsymbol{F}^2)^T, (\boldsymbol{F}^3)^T, (\boldsymbol{F}^4)^T)^T$ with $\boldsymbol{F}^i \in \mathbb{R}^N$, $i = 1, \ldots, 4$,

$$|\boldsymbol{F}| := \left((\boldsymbol{F}^1)^2 + (\boldsymbol{F}^2)^2 + (\boldsymbol{F}^3)^2 + (\boldsymbol{F}^4)^2\right)^{1/2} \in \mathbb{R}^N.$$

We can just repeat the arguments from the previous subsection. Again, $\boldsymbol{\mathcal{D}}_H^T$ has rank $N-1$ and its columns sum up to zeros. Therefore we have that $\boldsymbol{v} \in \mathcal{R}(\boldsymbol{\mathcal{D}}_H^T)$ if and only if (7) holds true. Then we obtain for $\boldsymbol{V} \in \mathbb{R}^{4N}$ with $\boldsymbol{v} = \boldsymbol{\mathcal{D}}_H^T \boldsymbol{V}$ that

$$(\boldsymbol{v}, \boldsymbol{w}) = (\boldsymbol{\mathcal{D}}_H^T \boldsymbol{V}, \boldsymbol{w}) = (\boldsymbol{V}, \boldsymbol{\mathcal{D}}_H \boldsymbol{w}) = \sum_{i=1}^4 \left(\boldsymbol{V}^i, (\boldsymbol{\mathcal{D}}_H \boldsymbol{w})^i\right).$$

Applying Schwarz's inequality to the sum of the corresponding components in the four inner products, we obtain $(\boldsymbol{v}, \boldsymbol{w}) \le (|\boldsymbol{V}|, |\boldsymbol{\mathcal{D}}_H \boldsymbol{w}|)$ and further $(\boldsymbol{v}, \boldsymbol{w}) \le \|\,|\boldsymbol{V}|\,\|_\infty \|\,|\boldsymbol{\mathcal{D}}_H \boldsymbol{w}|\,\|_1 \ \forall \boldsymbol{w} \in \mathbb{R}^N$. We solve the problem

$$\|\boldsymbol{f} - \boldsymbol{\mathcal{D}}_H \boldsymbol{V}\|^2 \to \min, \quad \text{subject to} \quad \|\,|\boldsymbol{V}|\,\|_\infty \le \lambda. \tag{14}$$

by the following algorithm which can be deduced in the same way as Alg. 1 in the previous subsection:

Algorithm 2.
Input: $\boldsymbol{f}$ and $\boldsymbol{V}^{(0)} := \boldsymbol{0}$.
Repeat for $k = 0$ until a stopping criterion is reached

$$\begin{aligned}
\boldsymbol{W}^{(k)} &:= \boldsymbol{\mathcal{D}}_H \boldsymbol{\mathcal{D}}_H^T \boldsymbol{V}^{(k)} - \boldsymbol{\mathcal{D}}_H \boldsymbol{f}, \\
\boldsymbol{V}^{(k+1)} &:= \left(\boldsymbol{1} + \frac{\tau}{\lambda} \left(|\boldsymbol{W}^{(k)}|, |\boldsymbol{W}^{(k)}|, |\boldsymbol{W}^{(k)}|, |\boldsymbol{W}^{(k)}| \right)^T \right)^{-1} \circ \left(\boldsymbol{V}^{(k)} - \tau \boldsymbol{W}^{(k)} \right), \\
k &:= k + 1,
\end{aligned}$$

where the inverse is taken componentwise.
Output: $\boldsymbol{u} := \boldsymbol{f} - \boldsymbol{\mathcal{D}}_H^T \boldsymbol{V}^{(k)}$.

Similarly as for Alg. 1 it can be proved that the iterative process converges for step sizes $\tau \le 1/\|\boldsymbol{\mathcal{D}}_H^T\|_2^2 = 1/\rho(\boldsymbol{\mathcal{D}}_H^T \boldsymbol{\mathcal{D}}_H)$. Having a closer look at the special structure of $\boldsymbol{\mathcal{D}}_H^T \boldsymbol{\mathcal{D}}_H$, we conclude by Gerschgorin's circle theorem that the eigenvalues of this matrix lie in a circle around 20 with radius 44. Thus, $\rho(\boldsymbol{\mathcal{D}}_H^T \boldsymbol{\mathcal{D}}_H) \le 64$ and we can prove convergence for $\tau \le 1/64$. However, in numerical experiments, convergence can be observed for $\tau \le 1/32$.

3. The modified Laplacian. Here we use the discretization $J(\boldsymbol{u}) := \boldsymbol{\mathcal{D}}_{\triangle,1} \boldsymbol{u}$, where

$$\boldsymbol{\mathcal{D}}_{\triangle,1} := \begin{pmatrix} (\boldsymbol{I}_n \otimes \boldsymbol{D}_0^T)(\boldsymbol{I}_n \otimes \boldsymbol{D}_0) \\ (\boldsymbol{D}_0^T \otimes \boldsymbol{I}_n)(\boldsymbol{D}_0 \otimes \boldsymbol{I}_n) \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}_n \otimes \boldsymbol{D}_0^T \boldsymbol{D}_0 \\ \boldsymbol{D}_0^T \boldsymbol{D}_0 \otimes \boldsymbol{I}_n \end{pmatrix}.$$

We have that $\boldsymbol{v} \in \mathcal{R}(\boldsymbol{\mathcal{D}}_{\triangle,1}^T)$ if and only if (7) is fulfilled. Using our standard arguments, we arrive at the problem

$$\|\boldsymbol{f} - \boldsymbol{\mathcal{D}}_{\triangle,1}^T \boldsymbol{V}\|^2 \to \min, \quad \text{subject to} \quad \|\boldsymbol{V}\|_\infty \le \lambda \tag{15}$$

4 Numerical experiments

For the onedimensional setting, numerical experiments are already contained in [25]. Therefore, we restrict our attention to two dimensions. All programs are written in MATLAB.

For the solution of the QP problems arising for the Laplacian and the modified Laplacian we have used the ILOG CPLEX Barrier Optimizer version 7.5. This routine uses a modification of the primal–dual predictor–corrector interior point algorithm described in [17]. The algorithms terminates if the relative complementary gap is smaller than 10^{-5}. The main reason for using this solver instead of, e.g., the MATLAB 'quadprog' routine was that CPLEX supports sparse matrix operations. Of course other

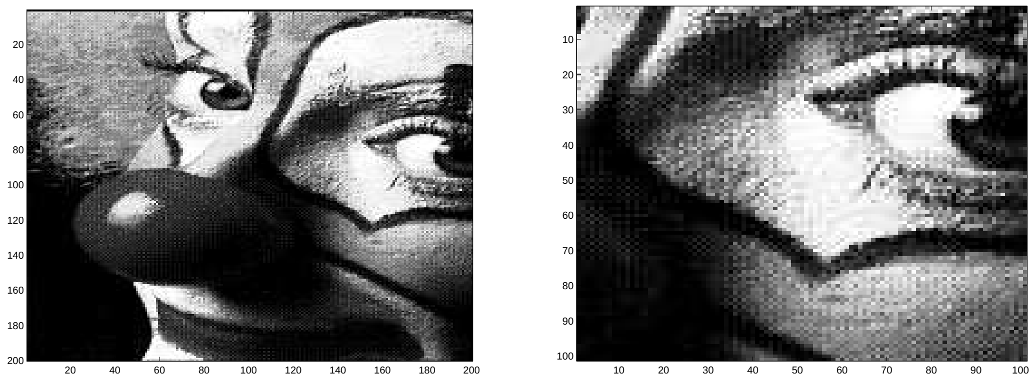

Figure 1: Left: Original clown image. Right: Part of the clown image.

QP techniques may be applied, e.g., an adaptation of the recently developed semi–smooth Newton method (primal–dual active set method) [13] to our setting with higher order derivatives. However, we will see in our experiments that the Frobenius norm of the Hessian seems to be superior to both the Laplacian and the modified Laplacian. Therefore, we will not focus our attention on the best QP method for the later problems.

In case of the ROF functional, we have applied Alg. 1 with step size $\tau = 1/4$. For the Frobenius norm of the Hessian, we have used Alg. 2 with step size $\tau = 1/64$.

We applied the four algorithms to the part $(50 : 150, 100 : 200)$ of the clown image in Fig. 1, where we stop the iterations in Alg. 1 and Alg. 2 if the relative error fulfills $\|V - V_{old}\|_2 / \|V\|_2 < 10^{-3}$. The required number of iterations for Alg. 1 and 2 is given by the following table.

λ	5	10	20
Alg. 1 (ROF)	131	159	192
Alg. 2 (Hessian)	272	347	432

The images transformed by our four algorithms with regularization parameter $\lambda = 10$ are shown in Fig. 2. Our findings can be summarized as follows:

- The images corresponding to higher order regularization functionals look smoother than the image related to the ROF functional. The later shows the typical staircasing effects.

- The image belonging to the Laplacian contains visible artefacts in form of white points. These artefacts also appear for other regularization parameters. Therefore we cannot recommend this method at least not with the current discretization.

- The images corresponding to the Frobenius norm of the Hessian and to the modified Laplacian are very similar. However, the second method is not rotationally invariant. This behaviour is demonstrated in Fig. 3.

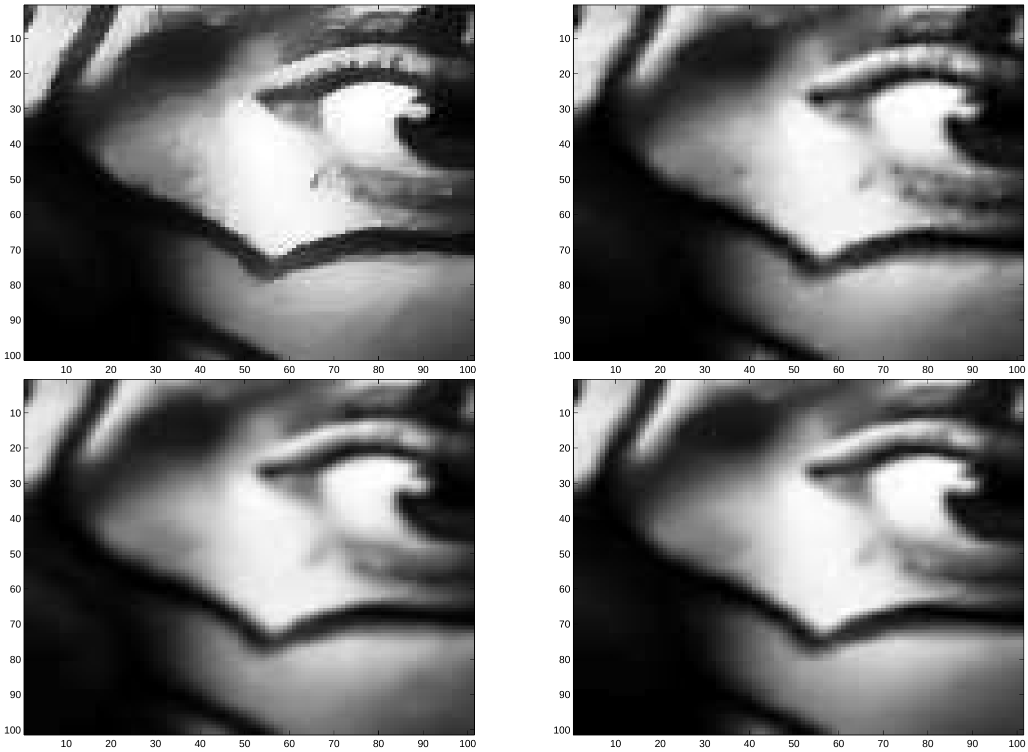

Figure 2: Transformed clown image (part) for $\lambda = 5$. Top left: Alg. 1 (ROF). Top right: QP (13) (Laplacian). Bottom left: Alg. 2 (Frobenius norm of the Hessian). Bottom right: QP (15) (modifiid Laplacian).

Appendix

We briefly derive the equivalence of (2) and (3) following mainly the lines of [2, 3].

Proposition 1. The problems (2) and (3) are equivalent.

Proof. Set $I(\boldsymbol{u}) := \|J(\boldsymbol{u})\|_1$. Since (2) is a convex functional, its minimizer has to fulfill the necessary and sufficient condition

$$\mathbf{0} \in \boldsymbol{u} - \boldsymbol{f} + \lambda \partial I(\boldsymbol{u}), \quad \text{i.e.,} \quad \frac{\boldsymbol{f} - \boldsymbol{u}}{\lambda} \in \partial I(\boldsymbol{u}), \tag{16}$$

where ∂I denotes the subgradient of I. By [22, Theorem 23.5], condition (16) is equivalent to

$$\boldsymbol{u} \in \partial I^* \left(\frac{\boldsymbol{f} - \boldsymbol{u}}{\lambda} \right) \quad \text{i.e.,} \quad \boldsymbol{f} - \lambda \tilde{\boldsymbol{v}} \in \partial I^*(\tilde{\boldsymbol{v}}), \tag{17}$$

where $\lambda \tilde{\boldsymbol{v}} := \boldsymbol{f} - \boldsymbol{u}$ and where I^* denotes the Legendre–Fenchel conjugate of I. Now $\tilde{\boldsymbol{v}}$ fulfills inclusion (17) if and only if $\tilde{\boldsymbol{v}}$ minimizes the functional

$$\frac{1}{2} \| \frac{\boldsymbol{f}}{\lambda} - \tilde{\boldsymbol{v}} \|_2^2 + \frac{1}{\lambda} I^*(\tilde{\boldsymbol{v}}). \tag{18}$$

By definition of the conjugate function and since I is one–homogeneous, we have for arbitrary $\lambda > 0$ that

$$I^*(\tilde{\boldsymbol{v}}) := \sup_{\boldsymbol{w} \in \mathbb{R}^N} \{ (\tilde{\boldsymbol{v}}, \boldsymbol{w}) - I(\boldsymbol{w}) \} = \sup_{\boldsymbol{w} \in \mathbb{R}^N} \{ (\tilde{\boldsymbol{v}}, \boldsymbol{w}) - \lambda I(\boldsymbol{w}/\lambda) \} = \lambda I^*(\tilde{\boldsymbol{v}}). \tag{19}$$

Therefore and since I^* is proper, either $I^*(\tilde{\boldsymbol{v}}) = 0$ or $I^*(\tilde{\boldsymbol{v}}) = \infty$ holds true. In the second case, the vector $\tilde{\boldsymbol{v}}$ cannot become a minimum of (18). Consequently, problem (18) can be rewritten as

$$\| \boldsymbol{f} - \lambda \tilde{\boldsymbol{v}} \|_2^2 \to \min, \quad \text{subject to} \quad I^*(\tilde{\boldsymbol{v}}) = 0.$$

Setting $\boldsymbol{v} := \lambda \tilde{\boldsymbol{v}}$, we see by (19) that this problem is equivalent to (3). $\square$

For special I, the set $\mathcal{V}_\lambda$ can be further restricted as follows:

Proposition 2. Let $J(\boldsymbol{w}) := g(\boldsymbol{A}\boldsymbol{w})$ with an (M, N) matrix $\boldsymbol{A}$ of rank smaller than N and a function $g : \mathbb{R}^M \to \mathbb{R}^{\tilde{M}}$ satisfying $g(\boldsymbol{0}) = \boldsymbol{0}$. Let $\mathcal{V}_\lambda := \{ \boldsymbol{v} \in \mathbb{R}^N : (\boldsymbol{v}, \boldsymbol{w}) \le \lambda \| J(\boldsymbol{w}) \|_1 \; \forall \boldsymbol{w} \in \mathbb{R}^N \}$. Then $\boldsymbol{v} \in \mathcal{V}_\lambda$ implies that $\boldsymbol{v} \in \mathcal{R}(\boldsymbol{A}^T)$.

Proof. Assume that there exists $\boldsymbol{v} \in \mathcal{V}_\lambda$ with $\boldsymbol{v} \notin \mathcal{R}(\boldsymbol{A}^T)$. Since $\mathbb{R}^N = \mathcal{R}(\boldsymbol{A}^T) \oplus \mathcal{N}(\boldsymbol{A})$, where $\mathcal{N}(\boldsymbol{A})$ denotes the nullspace of $\boldsymbol{A}$ and $\oplus$ the orthogonal sum, the vector $\boldsymbol{v}$ can be written as $\boldsymbol{v} = \boldsymbol{v}_0 + \boldsymbol{A}^T \boldsymbol{V}$ with $\boldsymbol{v}_0 \in \mathcal{N}(\boldsymbol{A})$, $\boldsymbol{v}_0 \neq \boldsymbol{0}$. Then we obtain for $\boldsymbol{w} := \boldsymbol{v}_0$ that

$$(\boldsymbol{v}, \boldsymbol{w}) = (\boldsymbol{v}_0 + \boldsymbol{A}^T \boldsymbol{V}, \boldsymbol{v}_0) = \| \boldsymbol{v}_0 \|_2^2 + (\boldsymbol{V}, \boldsymbol{A}\boldsymbol{v}_0) = \| \boldsymbol{v}_0 \|_2^2 > 0.$$

On the other hand, we have that

$$\| J(\boldsymbol{w}) \|_1 = \| g(\boldsymbol{A}\boldsymbol{v}_0) \|_1 = \| g(\boldsymbol{0}) \|_1 = 0,$$

so that we conclude by definition of $\mathcal{V}_\lambda$ that $\boldsymbol{v} \notin \mathcal{V}_\lambda$. This contradicts our assumption. $\square$

Acknowledgement: I like to thank S. Didas (University of Saarbrücken) and J. Neumann (University of Mannheim) for fruitful discussions.

References

[1] Aubert, G., Kornprobst, P.: Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations. New York: Springer 2002.

[2] Aujol, J.-F., Aubert, G., Blanc-Féraud, L., Chambolle A.: Image decomposition application to SAR images. In: Griffin, L. D., Lillholm, M. (eds.): Scale-Space Methods in Computer Vision, Lecture Notes in Computer Science 2695, pp. 297–312, Berlin: Springer 2003.

[3] Chambolle, A.: An algorithm for total variation minimization and applications. Journal of Mathematical Imaging and Vision 20, 89–97 (2004).

[4] Chambolle, A, Lions, P.-L.: Image recovery via total variation minimization and related problems. Numerische Mathematik 76, 167–188 (1997).

[5] Chan, T. F., Golub, G. H., Mulet, P.: A nonlinear primal–dual method for total-variation based image restoration. SIAM Journal on Scientific Computing 20(6), 1964–1977 (1999).

[6] Chan T. F., Marquina, A., Mulet, P.: High-order total variation-based image restoration. SIAM Journal on Scientific Computing 22(2), 503–516 (2000).

[7] Ciarlet, P. G.: Introduction to Numerical Linear Algebra and Optimization. Cambridge: Cambridge University Press 1989.

[8] Davies, P. L., Kovac, A.: Local extremes, runs, strings and multiresolution. *Annals of Statistics*, 29:1–65, 2001.

[9] Didas, S: Higher order variational methods for noise removal in signals and images. Diplomarbeit, Universität des Saarlandes, 2003.

[10] Girosi. F.: An equivalence between sparse approximation and support vector machines. Neural Computation 10(6), 1455–1480 (1998).

[11] Hinterberger, W., Hintermüller, M., Kunisch, K., von Oehsen, M., Scherzer, O.: Tube methods for BV regularization. Journal of Mathematical Imaging and Vision 19, 223 – 238 (2003).

[12] Hinterberger, W., Scherzer, O.: Variational methods on the space of functions of bounded Hessian for convexification and denoising. Technical Report, University of Innsbruck, Austria, 2003.

[13] Hintermüller, W., Kunisch, K.: Total bounded variation regularization as a bilaterally constrained optimization problem. SIAM J. Appl. Math. 64(4), 1311–1333 (2004).

[14] Lysaker, M., Lundervold, A., Tai, X.-C.: 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.

[15] Mammen, E., van de Geer, S.: Locally adaptive regression splines. Annals of Statistics 25(1), 387–413 (1997).

[16] Mangasarian, O. L., Schumaker, L. L.: Best summation formulae and discrete splines via mathematical programming. SIAM Journal on Numerical Analysis 10(3), 448–459 (1973).

[17] Mehrotra, S.: On the implementation of a primal–dual interior point method. SIAM Journal on Optimization 2(4), 575–601 (1992).

[18] Meyer, Y.: Oscillating Patterns in Image Processing and Nonlinear Evolution Equations, volume 22 of University Lecture Series. Providence: AMS 2001.

[19] Nielsen, M., Florack, L., Deriche, R.: Regularization, scale-space and edge detection filters. Journal of Mathematical Imaging and Vision 7, 291–307 (1997).

[20] Obereder, A., Osher, S., Scherzer, O.: On the use of dual norms in bounded variation type regularization. Technical report, Department of Computer Science, University of Innsbruck, Austria, 2004.

[21] Perona, P., Malik, J.: Scale space and edge detection using anisotropic diffusion. IEEE Transactions on Pattern Analysis and Machine Intelligence 12, 629–639 (1990).

[22] Rockafellar, R. T. . Convex Analysis. Princeton: Princeton University Press 1970.

[23] Rudin, L. I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D 60, 259–268 (1992).

[24] Scherzer, O.: Denoising with higher order derivatives of bounded variation and an application to parameter estimation. Computing 60, 1–27 (1998).

[25] Steidl, G., Didas, S., Neumann, J.: Relations between higher order TV regularization and support vector regression. In: Kimmel, R., Sochen, N., Weickert, J. (eds.): Scale-Space and PDE Methods in Computer Vision, Lecture Notes in Computer Science 3459, pp. 515–527, Berlin: Springer 2005.

[26] Vogel, C. R.: Computational Methods for Inverse Problems. Philadelphia: SIAM 2002.

[27] Weickert, J. : Anisotropic Diffusion in Image Processing. Stuttgart: Teubner 1998.

[28] Welk, M., Weickert, J., Steidl, G.: A four-pixel scheme for singular differential equations. In: Kimmel, R., Sochen, N., Weickert, J. (eds.): Scale-Space and PDE Methods in Computer Vision, Lecture Notes in Computer Science 3459, pp. 610–621, Berlin: Springer 2005.

[29] You, Y.-L., Kaveh, M.: Fourth-order partial differential equations for noise removal. IEEE Transactions on Image Processing 9(10), 1723–1730 (2000).

Gabriele Steidl
Faculty of Mathematics and Computer Science
University of Mannheim
A5
68131 Mannheim
Germany
steidl@uni-mannheim.de

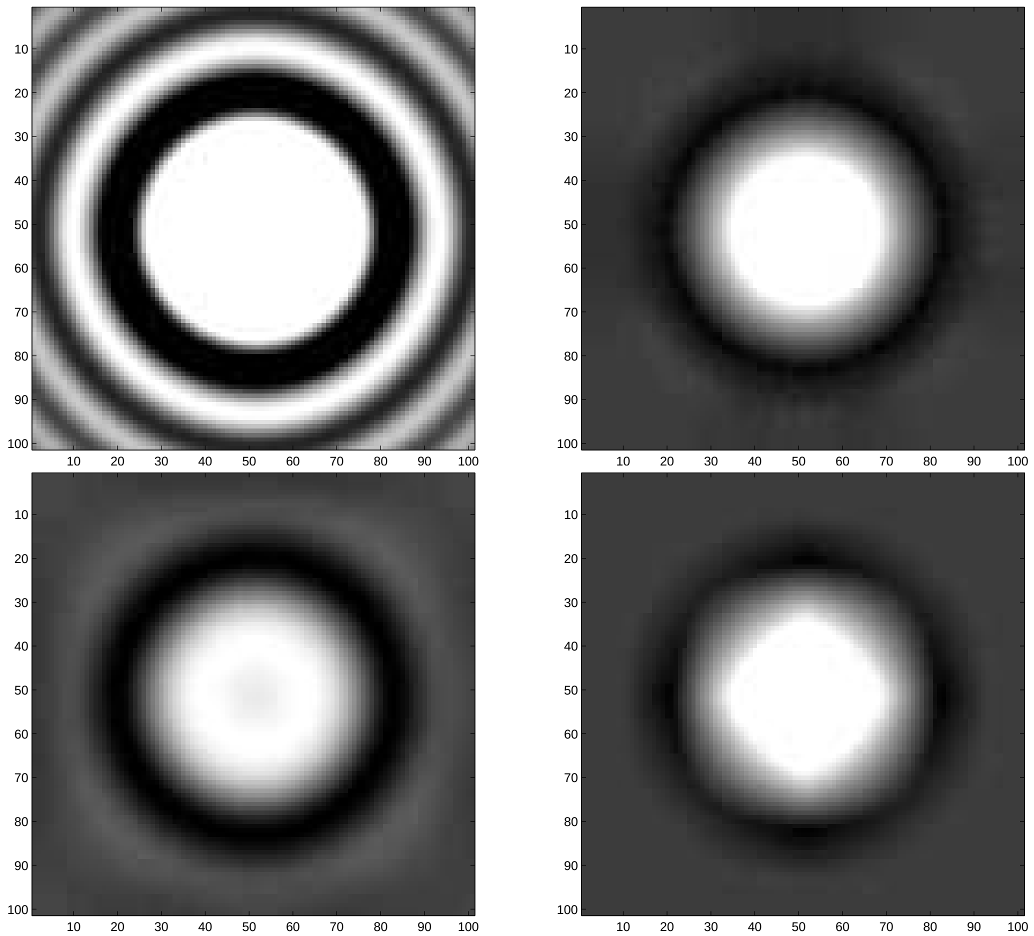

Figure 3: Top left: Original image consisting of concentric circles. Top right: Transformed image by (13) (Laplacian). Bottom left: Transformed image by Alg. 2 (Frobenius norm of the Hessian). Bottom right: Transformed image by (15) (modified Laplacian). This method is not rotationally invariant. In all transforms we have used $\lambda = 1500$.