

**A Note on the Dual Treatment of
Higher Order Regularization Functionals**

G. Steidl

278

2004

A note on the dual treatment of higher order regularization functionals

G. Steidl

Faculty of Mathematics and Computer Science, D7, 27
University of Mannheim, 68131 Mannheim, Germany
steidl@uni-mannheim.de
<http://www.kiwi.math.uni-mannheim.de>

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.

1 Introduction

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

$$\int_{\Omega} (u - f)^2 dx + 2\lambda \int_{\Omega} |\mathcal{J}(u)| dx, \quad (1)$$

where the regularization functional $\mathcal{I}(u) := \int_{\Omega} |\mathcal{J}(u)| dx$ 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 [1,23] for an overview. A frequently applied approach in image denoising and segmentation is the Rudin–Osher–Fatemi (ROF) model [20] with the gradient $\mathcal{J}(u) := \nabla u$.

In recent years, there has been a growing interest in higher order variational methods. In [5], the minimizer of the functional $\int_{\Omega} (f - u)^2 + \lambda_1 |\nabla u - \nabla v| + \lambda_2 |\nabla^2 v| dx$ was studied and in [21] the asymptotical case $\lambda_1 \rightarrow \infty$ was considered. In [7], 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 $\int_{\Omega} \varphi(|\Delta u|) dx$ with φ corresponding to the Perona–Malik diffusivity [18] was considered in [24]. In [13], 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 [16].

In this paper, we will apply the dual approach developed by A. Chambolle [4] for the ROF model to regularization functionals with higher order derivatives.

For dual approaches to the ROF model we also refer to [6,3]. 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

$$\|\mathbf{u} - \mathbf{f}\|_2^2 + 2\lambda\|J(\mathbf{u})\|_1 \rightarrow \min, \quad (2)$$

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

$$\|\mathbf{f} - \mathbf{v}\|_2^2 \rightarrow \min, \quad \text{subject to } \mathbf{v} \in \mathcal{V}_\lambda, \quad (3)$$

where

$$\mathcal{V}_\lambda := \{\mathbf{v} \in \mathbb{R}^N : (\mathbf{v}, \mathbf{w}) \leq \lambda\|J(\mathbf{w})\|_1 \ \forall \mathbf{w} \in \mathbb{R}^N\}, \quad (4)$$

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 $J(\mathbf{w}) = g(\mathbf{A}\mathbf{w})$ with an (M, N) matrix \mathbf{A} of rank smaller than N and with a function $g : \mathbb{R}^M \rightarrow \mathbb{R}^{\tilde{M}}$ satisfying $g(\mathbf{0}) = \mathbf{0}$. Then it is not hard to prove that

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

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

This paper is organized as follows. To make the general idea more comprehensive, we start by considering the onedimensional setting in Section 2. We will see 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 briefly explains the equivalence of (2) and (3) and verifies the above restriction of \mathcal{V}_λ .

2 Onedimensional Setting

We find it useful to consider the onedimensional case with derivative operators $\mathcal{J}(u) = u^{(m)}$ of various orders m first. As discretization of the first derivative of u , we use the forward difference $u'(jh) \approx (u((j+1)h) - u(jh))/h$, $j = 1, \dots, N-1$ with $h := (b-a)/N$. For simplicity, we assume in the following that $h = 1$.

For $r < N$, we introduce the $(N - r, N - r + 1)$ matrices $\tilde{D}_{N,r}$ and their transposed matrices by

$$\tilde{D}_{N,r} := \begin{pmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & & -1 & 1 \end{pmatrix} \quad \text{and} \quad \tilde{D}_{N,r}^T = \begin{pmatrix} -1 & 0 & \dots & 0 & 0 \\ 1 & -1 & \dots & 0 & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & \dots & 1 & -1 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}.$$

2.1 First order derivatives

For $m = 1$, we are concerned with the $(N - 1, N)$ matrix $D := \tilde{D}_{N,1}$ of rank $N - 1$. We use $J(\mathbf{u}) := D\mathbf{u}$ in (2). For this special J we want to characterize \mathcal{V}_λ in a different way. Since the columns of D^T add up to zero, we see that $\mathbf{v} \in \mathcal{R}(D^T)$ if and only if

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

This condition is in particular fulfilled if \mathbf{v} is white Gaussian noise. Then, for any $\mathbf{v} \in \mathcal{R}(D^T)$, there exists a unique $\mathbf{V} \in \mathbb{R}^{N-1}$ such that $\mathbf{v} = D^T \mathbf{V}$ and

$$(\mathbf{v}, \mathbf{w}) = (D^T \mathbf{V}, \mathbf{w}) = (\mathbf{V}, D\mathbf{w}) \leq \|\mathbf{V}\|_\infty \|D\mathbf{w}\|_1 \quad \forall \mathbf{w} \in \mathbb{R}^N,$$

where the inequality is sharp in the sense that there exists no constant $C < \|\mathbf{V}\|_\infty$ such that $(\mathbf{v}, \mathbf{w}) \leq C \|D\mathbf{w}\|_1$ holds true for all $\mathbf{w} \in \mathbb{R}^N$. Consequently, $\mathcal{V}_\lambda = \{\mathbf{v} := D^T \mathbf{V} : \|\mathbf{V}\|_\infty \leq \lambda, \mathbf{V} \in \mathbb{R}^{N-1}\}$ and problem (3) is equivalent to

$$\|\mathbf{f} - D^T \mathbf{V}\|_2^2 \rightarrow \min, \quad \text{subject to} \quad \|\mathbf{V}\|_\infty \leq \lambda. \quad (5)$$

The final solution $\mathbf{u} = \mathbf{f} - D^T \mathbf{V}$ satisfies $\sum_{j=1}^N u(j) = \sum_{j=1}^N f(j)$. Problem (5) can be solved by standard quadratic programming (QP) methods.

2.2 Higher order derivatives

For arbitrary $m \geq 1$, we use the $(N - m, N)$ matrix $D_{N,m} := \tilde{D}_{N,m} \cdots \tilde{D}_{N,1}$ of rank $N - m$ and $J(\mathbf{u}) := D_{N,m} \mathbf{u}$ to define a discrete version of the m -th derivative of u . Regarding the discrete momentum properties of the finite forward differences, we see that $\mathbf{v} \in \mathcal{R}(D_{N,m}^T)$ if and only if

$$\sum_{j=1}^N j^r v_j = 0, \quad r = 0, \dots, m - 1. \quad (6)$$

Then we obtain as in the previous subsection that there exists a unique $\mathbf{V} \in \mathbb{R}^{N-m}$ such that $\mathbf{v} = D_{N,m}^T \mathbf{V}$ and

$$(\mathbf{v}, \mathbf{w}) = (D_{N,m}^T \mathbf{V}, \mathbf{w}) = (\mathbf{V}, D_{N,m} \mathbf{w}) \leq \|\mathbf{V}\|_\infty \|D_{N,m} \mathbf{w}\|_1 \quad \forall \mathbf{w} \in \mathbb{R}^N,$$

where the inequality is sharp. Thus, problem (3) is equivalent to

$$\|f - D_{N,m}^T V\|_2^2 \rightarrow \min, \quad \text{subject to } \|V\|_\infty \leq \lambda. \quad (7)$$

The final solution is $u = f - D_{N,m}^T V$ and can be computed by standard QP methods.

Remark: In [22] we have examined the relation between (7) and SVR methods with discrete spline kernels. To see this relation, we introduce the invertible (N, N) Toeplitz matrix

$$D_{-1} = \begin{pmatrix} D & & & & & \\ 0 & \dots & 0 & -1 & & \\ & & & & & \end{pmatrix} = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & & & & & \vdots \\ 0 & 0 & 0 & \dots & -1 & 1 \\ 0 & 0 & 0 & \dots & 0 & -1 \end{pmatrix}.$$

Note that its inverse D_{-1}^{-1} is the upper triangular matrix with coefficients -1 . Instead of $D_{N,m}$ we want to use the m -th power D_{-1}^m of D_{-1} . The first $N - m$ rows of D_{-1}^m coincide with those of $D_{N,m}$. Now, for any $v \in \mathbb{R}^N$, there exists a unique $V \in \mathbb{R}^N$ such that $v = (D_{-1}^m)^T V$. Assuming that $V_{N-j} = 0$ for $j = 0, \dots, m - 1$ which is equivalent to the restrictions (6) on v , we obtain with $\tilde{V} := (V_1, \dots, V_{N-m})^T$ that

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

where the inequality is sharp. Consequently, (7) can be rewritten as

$$\begin{aligned} & \|f - (D_{-1}^m)^T V\|_2^2 \rightarrow \min, \\ & \text{subject to } \|V\|_\infty \leq \lambda \quad \text{and} \quad V_{n-j} = 0, \quad j = 0, \dots, m - 1. \end{aligned}$$

Defining F and U by $f = (D_{-1}^m)^T F$ and $u = (D_{-1}^m)^T U$, this problem becomes

$$\begin{aligned} & \|(D_{-1}^m)^T U\|_2^2 \rightarrow \min, \\ & \text{subject to } \|F - U\|_\infty \leq \lambda, \\ & \text{and } F_{n-j} = U_{n-j}, \quad j = 0, \dots, m - 1. \end{aligned} \quad (8)$$

Both problems can be solved by standard QP methods. However, for $m = 1$, the solution of (8) 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 [9,14].

Finally, let us consider $u = D_{-1}^m (D_{-1}^m)^T \psi$ 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 $(x)_+ := \max\{0, x\}$. In our discrete setting, we have with $K := (D_{-1}^m (D_{-1}^m)^T)^{-1}$ that $\psi = Ku$.

Let $U := \psi^{(m)} = k * u^{(m)}$. Its discrete version reads $U = \mathbf{K} \mathbf{D}_{-1}^m \mathbf{u}$. Setting $\mathbf{c} := \mathbf{D}_{-1}^m \mathbf{u}$, our minimization problem becomes

$$\begin{aligned} & \mathbf{c}^T \mathbf{K} \mathbf{c} \rightarrow \min, \\ \text{subject to } & \|\mathbf{F} - \mathbf{K} \mathbf{c}\|_\infty \leq \lambda \quad \text{and} \quad F_{N-j} = (\mathbf{K} \mathbf{c})_{N-j}, \quad j = 0, \dots, m-1. \end{aligned}$$

This is an SVR problem with discrete spline kernels, see [22].

3 Twodimensional Setting

For simplicity, we restrict our attention to quadratic (n, n) arrays and reshape them columnwise into a vector of length $N = n^2$. Further, we introduce the matrix

$$\mathbf{D}_0 := \begin{pmatrix} \mathbf{D} \\ 0 \dots 0 \end{pmatrix}.$$

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 repeat the approach of A. Chambolle [4] using our matrix-vector notation.

For $\mathbf{F} := \begin{pmatrix} \mathbf{F}^1 \\ \mathbf{F}^2 \end{pmatrix}$ with $\mathbf{F}^1, \mathbf{F}^2 \in \mathbb{R}^N$, let

$$|\mathbf{F}| := ((\mathbf{F}^1)^2 + (\mathbf{F}^2)^2)^{1/2} = (\mathbf{F}^1 \circ \mathbf{F}^1 + \mathbf{F}^2 \circ \mathbf{F}^2)^{1/2} \in \mathbb{R}^N,$$

where \circ denotes the componentwise vector product. As discrete version of \mathcal{J} we use $J(\mathbf{u}) = |\mathcal{D}\mathbf{u}|$ with the $(2N, N)$ matrix of rank $N - 1$

$$\mathcal{D} := \begin{pmatrix} \mathbf{I}_n \otimes \mathbf{D}_0 \\ \mathbf{D}_0 \otimes \mathbf{I}_n \end{pmatrix}, \quad \mathcal{D}^T = (\mathbf{I}_n \otimes \mathbf{D}_0^T, \mathbf{D}_0^T \otimes \mathbf{I}_n).$$

Here \otimes denotes the Kronecker product. Since the columns of \mathcal{D}^T add up to zero, we see that $\mathbf{v} \in \mathcal{R}(\mathcal{D}^T)$ if and only if

$$\sum_{j=1}^N v_j = 0. \tag{9}$$

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

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

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

$$(\mathbf{V}^1, (\mathbf{I}_n \otimes \mathbf{D}_0)\mathbf{w}) + (\mathbf{V}^2, (\mathbf{D}_0 \otimes \mathbf{I}_n)\mathbf{w}) \leq (|\mathbf{V}|, |\mathcal{D}\mathbf{w}|). \tag{10}$$

This inequality is not sharp for an arbitrary $\mathbf{V} \in \mathbb{R}^{2N}$, but becomes sharp if we restrict ourselves, e.g, to $\mathbf{V} \in \mathcal{R}(\mathcal{D})$ which does not imply further restriction on \mathbf{v} . Finally, we can estimate the right-hand side of (10) by

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

which becomes sharp, if we replace $\|\mathbf{V}\|_{\infty}$ by $\inf_{\mathbf{v}=\mathcal{D}^T\mathbf{V}} \|\mathbf{V}\|_{\infty}$, see [2]. This was denoted as G^d -norm of \mathbf{v} in [2] and can be considered as discrete version of Meyer's G -norm [15]. Recently, the G -norm was generalized to second order derivatives in [17]. This is related to the next subsection. Now problem (3) reads

$$\|\mathbf{f} - \mathcal{D}^T\mathbf{V}\|^2 \rightarrow \min, \quad \text{subject to} \quad \|\mathbf{V}\|_{\infty} \leq \lambda. \quad (11)$$

This is a quadratic problem with convex constraints. The Lagrangian of (11) is given by

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

where \mathbf{e} denotes the vector with components one and $\boldsymbol{\alpha} \in \mathbb{R}^N$ with $\alpha_j \geq 0$ componentwise. A necessary condition for \mathbf{V} to produce a minimum of (11) is that the gradient of \mathcal{L} with respect to \mathbf{V} is the zero vector, i.e.,

$$\nabla_{\mathbf{V}} \mathcal{L}(\mathbf{V}, \boldsymbol{\alpha}) = 2\mathcal{D} \mathcal{D}^T \mathbf{V} - 2\mathcal{D}\mathbf{f} + 2 \begin{pmatrix} \boldsymbol{\alpha} \circ \mathbf{V}^1 \\ \boldsymbol{\alpha} \circ \mathbf{V}^2 \end{pmatrix} = \mathbf{0}. \quad (12)$$

Let $\mathbf{W} := \mathcal{D} \mathcal{D}^T \mathbf{V} - \mathcal{D}\mathbf{f}$. If $\alpha_j > 0$, then, by the Karush-Kuhn-Tucker conditions, the j -th constraint in (11) has to be the equality $(\mathbf{V}_j^1)^2 + (\mathbf{V}_j^2)^2 = \lambda^2$. Consequently, by (12), $W_j^1 = -\alpha_j V_j^1$ and $W_j^2 = -\alpha_j V_j^2$ so that

$$(W_j^1)^2 + (W_j^2)^2 = \alpha_j^2 \lambda^2. \quad (13)$$

If $\alpha_j = 0$, then $W_j^1 = 0 = \alpha_j$ and $W_j^2 = 0 = \alpha_j$ so that we also have (13). Hence we can replace $\boldsymbol{\alpha}$ in (12) by (13) and obtain

$$\mathbf{W} + \frac{1}{\lambda} \begin{pmatrix} |\mathbf{W}| \\ |\mathbf{W}| \end{pmatrix} \circ \mathbf{V} = \mathbf{0}. \quad (14)$$

By [8, Theorem9.2-4], the Karush-Kuhn-Tucker conditions summarized in (14) are also sufficient for \mathbf{V} to provide a minimum of (11). To solve (14), A. Chambolle [4] suggested the semi-implicit gradient descent approach

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

In summary, we obtain the following algorithm:

Algorithm 1.

Input: \mathbf{f} and $\mathbf{V}^{(0)} := \mathbf{0}$.

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

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

where the inverse is taken componentwise.

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

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

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

Now $\|\mathcal{D}^T\|_2^2 = \rho(\mathcal{D}^T\mathcal{D})$, where ρ denotes the spectral radius of the matrix. The matrix $\mathcal{D}^T\mathcal{D}$ 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 $4(\sin(j\pi/(2n))^2 + \sin(k\pi/(2n))^2)$, $j, k = 0, \dots, n-1$. Thus, $\|\mathcal{D}^T\|_2^2 = 8$. However, in numerical experiments convergence was observed for $\tau \leq 1/4$.

Since

$$\|\mathbf{V}\|_\infty \leq \|\mathbf{V}\|_\infty \leq \sqrt{2}\|\mathbf{V}\|_\infty.$$

we also consider the problem

$$\|\mathbf{f} - \mathcal{D}^T\mathbf{V}\|^2 \rightarrow \min, \quad \text{subject to } \|\mathbf{V}\|_\infty \leq \lambda. \quad (15)$$

This problem has linear constraints and can be solved by standard QP methods. Constraints of the form (15) were also used also by W. Hinterberger, M. Hintermüller, K. Kunisch, M. von Oehsen and O. Scherzer [11]. However, motivated by the taut-string algorithm, these authors minimized a more complicated functional.

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

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

2. the Frobenius norm of the Hessian mentioned also in [10]

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

3. the modified Laplacian considered in [13]

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

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

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

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 $\mathcal{D}^T \mathcal{D}$ add up to zero, we have that $\mathbf{v} \in \mathcal{R}(\mathcal{D}_\Delta)$ if and only if (9) holds true. Then we obtain for all $\mathbf{V} \in \mathbb{R}^N$ with $\mathbf{v} = \mathcal{D}_\Delta \mathbf{V}$ that

$$(\mathbf{v}, \mathbf{w}) = (\mathcal{D}_\Delta \mathbf{V}, \mathbf{w}) = (\mathbf{V}, \mathcal{D}_\Delta \mathbf{w}) \leq \|\mathbf{V}\|_\infty \|\mathcal{D}_\Delta \mathbf{w}\|_1 \quad \forall \mathbf{w} \in \mathbb{R}^N. \quad (16)$$

Regarding that the nullspace of \mathcal{D}_Δ is given by $\{c\mathbf{e} : c \in \mathbb{R}\}$, we see that $\mathbf{v} = \mathcal{D}_\Delta \mathbf{V}$ if and only if $\mathbf{v} = \mathcal{D}_\Delta (\mathbf{V} + c\mathbf{e})$. Choosing $\mu := (V_{\min} + V_{\max})/2$ with the maximal and minimal components V_{\min} and V_{\max} of \mathbf{V} , respectively, we obtain that the components of $\tilde{\mathbf{V}} := \mathbf{V} - \mu\mathbf{e}$ fulfill $-(V_{\max} - V_{\min})/2 \leq \tilde{V}_j \leq (V_{\max} - V_{\min})/2$, where we have lower and upper equality for some components j_- and j_+ , respectively. Thus, $\|\tilde{\mathbf{V}}\|_\infty = (V_{\max} - V_{\min})/2$ and $\|\tilde{\mathbf{V}} + c\mathbf{e}\|_\infty = \|\tilde{\mathbf{V}}\|_\infty + |c|$. Consequently, $\min_{\mathbf{v}=\mathcal{D}_\Delta \mathbf{V}} \|\mathbf{V}\|_\infty = \|\tilde{\mathbf{V}}\|_\infty$. Choosing $\mathcal{D}_\Delta \mathbf{w}$ as vector consisting of zeros except for $(\mathcal{D}_\Delta \mathbf{w})_{j_-} := -1$ and $(\mathcal{D}_\Delta \mathbf{w})_{j_+} := 1$, we obtain in (16) the equality $(\mathbf{v}, \mathbf{w}) = (\tilde{\mathbf{V}}, \mathcal{D}_\Delta \mathbf{w}) = 2\|\tilde{\mathbf{V}}\|_\infty = \|\tilde{\mathbf{V}}\|_\infty \|\mathcal{D}_\Delta \mathbf{w}\|_1$.

Finally, we solve

$$\|\mathbf{f} - \mathcal{D}_\Delta \mathbf{V}\|_2^2 \rightarrow \min, \quad \text{subject to} \quad \|\mathbf{V}\|_\infty \leq \lambda. \quad (17)$$

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

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

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

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

We can just repeat the arguments from the previous subsection. Again, \mathcal{D}_H^T has rank $N - 1$ and its columns sum up to zeros. Therefore we have that $\mathbf{v} \in \mathcal{R}(\mathcal{D}_H^T)$ if and only if (9) holds true. Then we obtain for $\mathbf{V} \in \mathbb{R}^{4N}$ with $\mathbf{v} = \mathcal{D}_H^T \mathbf{V}$ that

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

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

$$\|\mathbf{f} - \mathcal{D}_H \mathbf{V}\|_2^2 \rightarrow \min, \quad \text{subject to} \quad \|\mathbf{V}\|_\infty \leq \lambda. \quad (18)$$

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

Algorithm 2.

Input: \mathbf{f} and $\mathbf{V}^{(0)} := \mathbf{0}$.

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

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

where the inverse is taken componentwise.

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

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

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

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

We have that $\mathbf{v} \in \mathcal{R}(\mathcal{D}_{\Delta,1}^T)$ if and only if (9) is fulfilled. Using our standard arguments, we arrive at the problem

$$\|\mathbf{f} - \mathcal{D}_{\Delta,1}^T \mathbf{V}\|^2 \rightarrow \min, \quad \text{subject to } \|\mathbf{V}\|_{\infty} \leq \lambda \quad (19)$$

4 Numerical experiments

For the onedimensional setting, numerical experiments are already contained in [22]. Therefore, we restrict our attention to two dimensions. We apply the various algorithms to the part (50 : 150, 100 : 200) of the clown image in Fig. 1.

All programs are written in MATLAB. For the solution of the standard QP problems we have used the CPLEX package whose QP routine is based on a barrier optimizer.

Fig. 2 shows the results of Alg. 1 versus (15). Both algorithms show segmentations. However, the image transformed by (15) emphasises horizontal and vertical directions and introduces harder staircasing effects.

The following figures show results for regularization functionals with second order derivatives. In general, these images look smoother than those in Fig. 2. Fig. 3 shows the results for the Laplacian. Both images have visible artifacts

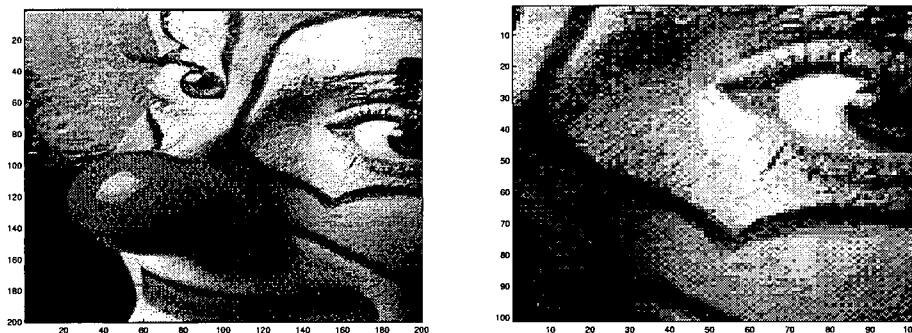


Fig. 1. Left: Original clown image. Right: Part of the clown image.

(white pixels). Finally, Fig. 4 presents the results obtained for the Frobenius norm of the Hessian by Alg. 2 and for the modified Laplacian by (19). For the Hessian the smoothing effect is larger than for the modified Laplacian.

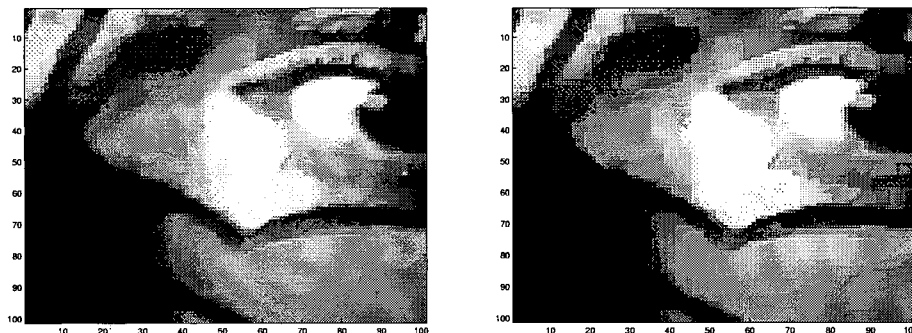


Fig. 2. Left: Clown image (part) transformed by Alg. 1 with $\lambda = 10$. Right: Clown image (part) transformed by (15) with $\lambda = 10$.

Appendix

Let us briefly derive the equivalence of (2) and (3). We follow mainly the lines of [2,4].

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

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

$$\mathbf{0} \in \mathbf{u} - \mathbf{f} + \lambda \partial I(\mathbf{u}), \quad \text{i.e.,} \quad \frac{\mathbf{f} - \mathbf{u}}{\lambda} \in \partial I(\mathbf{u}), \quad (20)$$

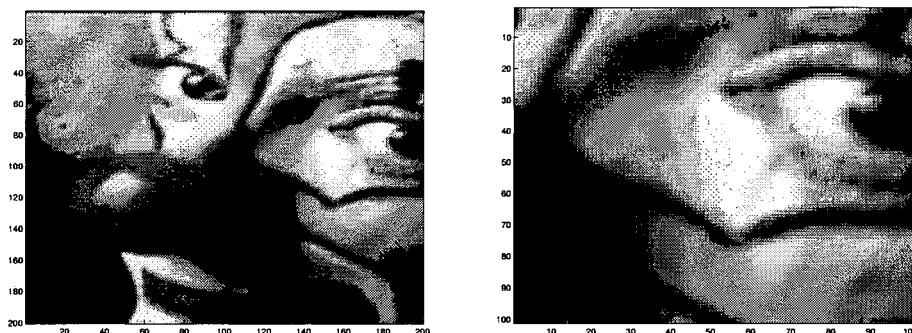


Fig. 3. Clown image and its part transformed by (17) with $\lambda = 10$.

where ∂I denotes the subgradient of I . By [19, Theorem 23.5], condition (20) is equivalent to

$$\mathbf{u} \in \partial I^* \left(\frac{\mathbf{f} - \mathbf{u}}{\lambda} \right) \quad \text{i.e.,} \quad \mathbf{f} - \lambda \tilde{\mathbf{v}} \in \partial I^*(\tilde{\mathbf{v}}), \quad (21)$$

where $\lambda \tilde{\mathbf{v}} := \mathbf{f} - \mathbf{u}$ and where I^* denotes the Legendre–Fenchel conjugate of I . Now $\tilde{\mathbf{v}}$ fulfills inclusion (21) if and only if $\tilde{\mathbf{v}}$ minimizes the functional

$$\|\mathbf{f}/\lambda - \tilde{\mathbf{v}}\|_2^2 + (2/\lambda) I^*(\tilde{\mathbf{v}}). \quad (22)$$

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

$$I^*(\tilde{\mathbf{v}}) := \sup_{\mathbf{w} \in \mathbb{R}^N} \{(\tilde{\mathbf{v}}, \mathbf{w}) - I(\mathbf{w})\} = \sup_{\mathbf{w} \in \mathbb{R}^N} \{(\tilde{\mathbf{v}}, \mathbf{w}) - \lambda I(\mathbf{w}/\lambda)\} = \lambda I^*(\tilde{\mathbf{v}}). \quad (23)$$

Therefore and since I^* is proper, either $I^*(\tilde{\mathbf{v}}) = 0$ or $I^*(\tilde{\mathbf{v}}) = \infty$ holds true. In the second case, the vector $\tilde{\mathbf{v}}$ cannot become a minimum of (22). Consequently, problem (22) can be rewritten as

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

Setting $\mathbf{v} := \lambda \tilde{\mathbf{v}}$, we see by (23) that this problem is equivalent to (3). \square

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

Proposition 2. Let $J(\mathbf{w}) := g(\mathbf{A}\mathbf{w})$ with an (M, N) matrix \mathbf{A} of rank smaller than N and a function $g : \mathbb{R}^M \rightarrow \mathbb{R}^M$ satisfying $g(\mathbf{0}) = \mathbf{0}$. Let \mathcal{V}_λ be defined by (4). Then $\mathbf{v} \in \mathcal{V}_\lambda$ implies that $\mathbf{v} \in \mathcal{R}(\mathbf{A}^T)$.

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

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

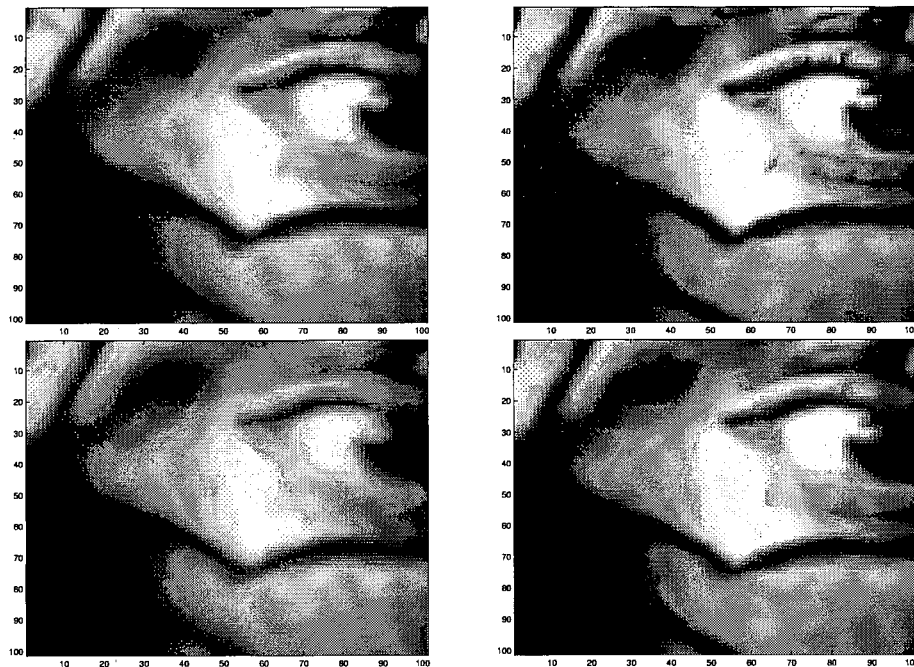


Fig. 4. Top left: Clown image (part) transformed by Alg. 2 with $\lambda = 5$. Top right: Clown image (part) transformed by (19) with $\lambda = 5$. Bottom left: Clown image (part) transformed by Alg. 2 with $\lambda = 10$. Bottom right: Clown image (part) transformed by (19) with $\lambda = 10$.

On the other hand, we have that

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

so that we conclude by definition of \mathcal{V}_λ that $\mathbf{v} \notin \mathcal{V}_\lambda$. 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. G. Aubert and P. Kornprobst. *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*, volume 147 of *Applied Mathematical Sciences*. Springer, New York, 2002.
2. J.-F. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle. Image decomposition application to SAR images. In L. D. Griffin and M. Lillholm, editors, *Scale-Space Methods in Computer Vision*, volume 2695 of *Lecture Notes in Computer Science*, pages 297–312, Berlin, 2003. Springer.

3. J. L. Carter. Dual methods for total variation-based image restoration. Ph.D. thesis, UCLA, 2001.
4. A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, (20):89–97, 2004.
5. A. Chambolle and P.-L. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76:167–188, 1997.
6. T. F. Chan, G. H. Golub, and P. Mulet. A nonlinear primal–dual method for total-variation based image restoration. *SIAM Journal on Scientific Computing*, 20(6):1964–1977, 1999.
7. 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.
8. P. G. Ciarlet. *Introduction to Numerical Linear Algebra and Optimization*. Cambridge University Press, Cambridge, 1989.
9. P. L. Davies and A. Kovac. Local extremes, runs, strings and multiresolution. *Annals of Statistics*, 29:1–65, 2001.
10. S. Didas. Higher order variational methods for noise removal in signals and images. Diplomarbeit, Universität des Saarlandes, 2003.
11. 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.
12. 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.
13. M. Lysaker, A. Lundervold, and X.-C. 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.
14. E. Mammen and S. van de Geer. Locally adaptive regression splines. *Annals of Statistics*, 25(1):387–413, 1997.
15. Y. Meyer. *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*, volume 22 of *University Lecture Series*. AMS, Providence, 2001.
16. M. Nielsen, L. Florack, and R. Deriche. Regularization, scale-space and edge detection filters. *Journal of Mathematical Imaging and Vision*, 7:291–307, 1997.
17. A. Obereder, S. Osher, and O. Scherzer. On the use of dual norms in bounded variation type regularization. Technical report, Department of Computer Science, University of Innsbruck, Austria, 2004.
18. P. Perona and J. Malik. Scale space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:629–639, 1990.
19. R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, 1970.
20. L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.
21. O. Scherzer. Denoising with higher order derivatives of bounded variation and an application to parameter estimation. *Computing*, 60:1–27, 1998.
22. G. Steidl, S. Didas, and J. Neumann. Relations between higher order TV regularization and support vector regression. Technical report, Department of Mathematics and Computer Science, University of Mannheim, Germany, 2004.
23. J. Weickert. *Anisotropic Diffusion in Image Processing*. Teubner, Stuttgart, 1998.
24. Y.-L. You and M. Kaveh. Fourth-order partial differential equations for noise removal. *IEEE Transactions on Image Processing*, 9(10):1723–1730, 2000.