Combined ℓ_2 data and gradient fitting in conjunction with ℓ_1 regularization

Stephan Didas* Simon Setzer† Gabriele Steidl‡

April 13, 2007

Abstract

We are interested in minimizing functionals with ℓ_2 data and gradient fitting term and ℓ_1 regularization term with higher order derivatives in a discrete setting. We examine the structure of the solution in 1d by reformulating the original problem into a contact problem which can be solved by dual optimization techniques. The solution turns out to be a 'smooth' discrete polynomial spline whose knots coincide with the contact points while its counterpart in the contact problem is a discrete version of a spline with higher defect and contact points as knots. In 2d we modify Chambolle's algorithm to solve the minimization problem with the ℓ_1 norm of interacting second order partial derivatives as regularization term. We show that the algorithm can be implemented efficiently by applying the fast cosine transform. We demonstrate by numerical denoising examples that the ℓ_2 gradient fitting term can be used to avoid both edge blurring and staircasing effects.

Short title: ℓ_1 regularized data and gradient fitting

AMS Subject Classification: 65K10, 65F22, 65T50, 49M29

Key words: higher order ℓ_1 regularization, TV regularization, convex optimization, dual optimization methods, discrete splines, splines with defect, G–norm, fast cosine transform.

1 Introduction

In image denoising one is interested in removing noise while preserving or even enhancing important structures such as edges. While linear filters

*didas@mia.uni-saarland.de, Faculty of Mathematics and Computer Science, Saarland University, 66041 Saarbrücken, Germany

†ssetzer@rumms.uni-mannheim.de, University of Mannheim, Institute of Mathematics, D–68131 Mannheim

‡steidl@math.uni-mannheim.de, University of Mannheim, Institute of Mathematics, D–68131 Mannheim

typically smooth edges, some edge enhancing methods create artificial edges out of the continuous gray value transitions. This so-called 'staircasing effect' is one of the most prominent shortcomings of many well-established image denoising algorithms, e.g., of the Rudin–Osher–Fatemi (ROF) model. It creates an oversegmentation of the image into artificial parts. To avoid these artifacts one can include higher order derivatives into the model which would prefer not only piecewise constant, but also piecewise linear results [13, 23, 28]. Unfortunately, these higher order derivative methods also tend to introduce some blurring in the region of image edges.

In this context, we are concerned with minimizing discrete versions of the functional

$$\frac{1}{2}\|f-u\|_{L_2}^2+\frac{\alpha}{2}\,\|\nabla f-\nabla u\|_{L_2}^2+\beta\,\|\,(\sum_{|\gamma|=m} w_\gamma|u^{(\gamma)}|^2)^{1/2}\,\|_{L_1}\,, \qquad (1)$$

where $\gamma\in\mathbb{N}_0^2$ denotes the order of the derivatives in x and y directions, $|\gamma|:=\gamma_1+\gamma_2$ and w_γ are nonegative weights. For $|\gamma|=1$ and $\alpha=0$ the penalization in (1) becomes $\int|\nabla x|\,dxdy$ and we obtain the frequently applied ROF model. The main reason for introducing the additional gradient fitting term with $\alpha>0$ consists in avoiding the staircasing effect on the one hand and in preserving edges and discontinuities on the other hand. Image edges can be characterised as regions where the gradient is high. Thus the gradient fitting term is intended to force the solution to be similar to the initial image especially near edges.

A first approach that uses partial derivatives of first and second order based on the idea of inf-convolution was proposed by Chambolle and Lions in [3]. For comparisons we sketch the corresponding algorithm in the numerical part of this paper. An asymptotical case of [3] with respect to one parameter was studied in [21]. The (directed) Laplacian as second order term was added to the TV functional in [5]. Other approaches are possible, e.g., by the applying Bregman distances [19]. Further, we mention that l_1 regularization terms with decoupled partial derivatives were used, e.g., in [11, 2, 13]. These approaches might be useful for special image processing tasks and can be handled by different algorithms.

To get a better idea concerning the structure of the solution of our minimization problem, we first deal with the univariate setting

$$\frac{1}{2}\|f-u\|_{L_2}^2+\frac{\alpha}{2}\|f'-u'\|_{L_2}^2+\beta\|f^{(m)}\|_{L_1}\,. \qquad (2)$$

Again, we focus on the discrete approach with forward differences instead of derivatives. Note that in the continuous setting L_1 regularization in connection with splines was treated in [7, 14] with a careful handling of the non–reflexive space L_1. In this paper, we reformulate (2) as a contact problem which can be solved via the dual formulation of (2). In case of

an additional gradient fitting term ($\alpha > 0$) the solution can be computed efficiently by applying the fast discrete cosine transform. We prove that the solution U of the contact problem is a discrete polynomial spline of degree $2m-1$ with the contact points being related to the spline knots. For $\alpha > 0$, it can be considered as a discrete version of a spline with defect three. The solution u of (2) is directly determined by the solution U of the contact problem and turns out to be a 'smooth' discrete polynomial spline of degree $m-1$ with knots related to the contact points. We do not present numerical denoising examples in 1d since they only confirm the 2d findings. For $\alpha = 0$ and various derivatives m, denoising results are given in [24].

Having examined the structure of the solution in 1d, we turn to our original 2d denoising problem. Here the regularization term includes the ℓ_1 norm of interacting partial derivatives which ensures rotationally invariant solutions. We adapt an algorithm of Chambolle [1], which is also based on the dual version of (1), to our setting. In case of an additional gradient fitting term we can include the fast cosine transform into this algorithm.

This paper is organized as follows: We start with the 1d part in Section 2. First we provide our discrete setting in Subsection 2.1 . Then we reformulate the discrete minimization problem as a contact problem and deal with its solution via the dual formulation of the minimization problem in Subsection 2.2. Finally, we examine the structure of the solution both of the contact problem and the original minimization problem in Subsection 2.3.

Section 3 deals with 2d images, where we focus on the practically relevant regularization with at most second order derivatives in the regularization term. After introducing the discrete setting in Subsection 3.1 we turn to the dual formulation and Chambolle's algorithm in connection with the discrete cosine transform in Subsection 3.2. Finally, Subsection 3.3 presents numerical denoising results demonstrating the influence of the additional gradient fitting term.

2 Higher order ℓ_1 regularization in 1d

2.1 Discrete setting

In this section, we deal with a discrete version of (2). To this end, let

$$D_{1,N} := \begin{pmatrix} -1 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 & 0 \\ & & \ddots & \ddots & \ddots & & \\ 0 & 0 & 0 & \dots & -1 & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & -1 & 1 \end{pmatrix} \in \mathbb{R}^{N-1,N} \tag{3}$$

be the *first order forward difference matrix* and

$$D_{m,N} := D_{1,N-(m-1)} \cdot \ldots \cdot D_{1,N-1} D_{1,N} \in \mathbb{R}^{N-m,N}$$

the *m–th order forward difference matrix*. If the size N of the difference matrix $D_{m,N}$ is given in the context we skip the second index and write only D_m. Then it is well known that

$$\begin{aligned}\mathcal{R}(D_m^{\mathrm{T}}) &= \{f \in \mathbb{R}^N : \sum_{j=1}^{N} j^r f(j) = 0,\ r = 0, \dots, m-1\},\\ \mathcal{N}(D_m) &= \operatorname{span}\{(j^r)_{j=1}^N :\ r = 0, \dots, m-1\},\end{aligned}$$

i.e., the range $\mathcal{R}(D_m^{\mathrm{T}})$ of D_m^{T} consists of the vectors with m vanishing moments while the kernel $\mathcal{N}(D_m)$ of D_m is just given by the discrete polynomials of degree $\leq m-1$.

We are interested in minimizing the discrete counterpart of (2)

$$F(u) = \frac{1}{2}\|f-u\|_2^2 \ + \ \frac{\alpha}{2}\|D_1 f - D_1 u\|_2^2 \ + \ \beta\|D_m u\|_1 \tag{4}$$

which can be rewritten as

$$F(u) = \frac{1}{2}(f-u)^{\mathrm{T}}(I_N + \alpha D_1^{\mathrm{T}} D_1)(f-u) \ + \ \beta\|D_m u\|_1\,. \tag{5}$$

Clearly, for $\alpha \geq 0$, the matrix

$$A = A(\alpha) := I_N + \alpha D_1^{\mathrm{T}} D_1$$

is positive definite. Setting $B^{\mathrm{T}}B := A$ and $L := D_m$ the functional (5) becomes

$$F(u) = \frac{1}{2}\|B(f-u)\|_2^2 \ + \ \beta\|Lu\|_1\,. \tag{6}$$

The minimizer of (6) can be computed in various ways. In the next subsection, we propose to minimize (6) using its dual formulation. This is closely related to the reformulation of (6) as a contact problem and serves as our basis to gain some insight into the structure of the solution u.

2.2 Contact problem and dual formulation

In this subsection, we focus on minimizing strictly convex functionals of the form

$$F(u) = \frac{1}{2}\|B(f-u)\|_2^2 \ + \ \beta\|Lu\|_1\,, \tag{7}$$

where $B \in \mathbb{R}^{N,N}$ and $L \in \mathbb{R}^{N-m,N}$ are arbitrary matrices of full rank. In particular, we are interested in our special setting from the previous subsection.

Decomposition related to $\mathcal{N}(L)$. Since the regularization term becomes zero if u is in $\mathcal{N}(L)$ we want to restrict ourselves to those parts of u which are in a certain sense orthogonal to $\mathcal{N}(L)$. To this end, we define an inner product on $\mathbb{R}^N$ by

$$\langle u, v\rangle_A = \langle Au, v\rangle = v^{\mathrm{T}} Au\,.$$

Corresponding to the A orthogonal decomposition

$$\mathbb{R}^N = \mathcal{N}(L) \;\oplus_A\; \mathcal{R}(A^{-1}L^{\mathrm{T}})$$

every vector $u \in \mathbb{R}^N$ has a unique decomposition as

$$u = u_0 \;+\; u_1\,, \quad u_0 \in \mathcal{N}(L)\,,\; u_1 \in \mathcal{R}(A^{-1}L^{\mathrm{T}})\,. \tag{8}$$

Using this decomposition for f and u, we obtain in (7),

$$\begin{aligned} F(u) &= \frac{1}{2}\|B(f_0 - u_0)\|_2^2 \;+\; \langle B(f_0 - u_0), B(f_1 - u_1)\rangle \;+\; \frac{1}{2}\|B(f_1 - u_1)\|_2^2 \\ &\quad + \beta\|Lu_1\|_1 \\ &= \frac{1}{2}\|B(f_0 - u_0)\|_2^2 \;+\; \frac{1}{2}\|B(f_1 - u_1)\|_2^2 \;+\; \beta\|Lu_1\|_1\,. \end{aligned}$$

It is easy to check that $f_1 = A^{-1}L^{\mathrm{T}}KLf$, where

$$K := \left(LA^{-1}L^{\mathrm{T}}\right)^{-1}.$$

Note that K exists since L has full rank. Consequently, to solve (7), we can set $u_0 := f_0 = f - f_1$ and search for $u_1 \in \mathcal{R}(A^{-1}L^{\mathrm{T}})$ minimizing

$$\frac{1}{2}\|B(f_1 - u_1)\|_2^2 \;+\; \beta\|Lu_1\|_1\,.$$

In the following, we assume that $f \in \mathcal{R}(A^{-1}L^{\mathrm{T}})$ such that $f_1 = f$ and $u_1 = u$.

Reformulation as a contact problem. For the solution u of (7) it is necessary and sufficient that 0_N is an element of the subdifferential $\partial F(u)$:

$$0_N \in A(u - f) + \beta L^{\mathrm{T}}\,\frac{Lu}{|Lu|}\,, \tag{9}$$

where the quotient is meant componentwise and

$$\frac{x}{|x|} := \begin{cases} 1 & \text{if } x > 0, \\ -1 & \text{if } x < 0, \\ [-1,1] & \text{if } x = 0. \end{cases}$$

This can be rewritten as

$$\begin{aligned} u &\in f - \beta A^{-1} L^{\mathrm{T}} \frac{Lu}{|Lu|}\,, \\ Lu &\in Lf - \beta L A^{-1} L^{\mathrm{T}} \frac{Lu}{|Lu|}\,. \end{aligned} \tag{10}$$

Since $f, u \in \mathcal{R}(A^{-1}L^{\mathrm{T}})$ there exist $F, U \in \mathbb{R}^{N-m}$ such that

$$f = A^{-1}L^{\mathrm{T}}F\,,\ u = A^{-1}L^{\mathrm{T}}U\,. \tag{11}$$

Conversely, we have that

$$F = KLf\,,\ U = KLu\,. \tag{12}$$

Multiplying (10) by K and using (12) we obtain the inclusion

$$U \in F - \beta\,\frac{K^{-1}U}{|K^{-1}U|}\,.$$

Hence (7) can be reformulated as the following **contact problem**:

Find $U \in \mathbb{R}^{N-m}$ so that

- $\|F - U\|_\infty \le \beta$.
  U lies in a tube around F of width 2β.
- if $(K^{-1}U)_j > 0$ we have a lower contact point $U_j = F_j - \beta$,
  if $(K^{-1}U)_j < 0$ we have an upper contact point $U_j = F_j + \beta$.

To get an idea concerning the structure of U and u in the next subsection let us write

$$U = K\,c\,, \tag{13}$$

so that by (11)

$$u = \tilde{K}\,c\,, \tag{14}$$

with $\tilde{K} := A^{-1}L^{\mathrm{T}}K$. Then the contact problem reads as follows:

Find $c \in \mathbb{R}^{N-m}$ so that

- $\|F - Kc\|_\infty \le \beta$.
- if $c_j > 0$ we have a lower contact point $U_j = F_j - \beta$,
  if $c_j < 0$ we have an upper contact point $U_j = F_j + \beta$.

Let

$$\Xi := \{j \in \{0, \ldots, N-m-1\} : c_j \neq 0\} \tag{15}$$

be the (sub)set of contact point indices. If $\#\Xi$ is small, then c becomes sparse and (13) (resp. (14)) are sparse representations of U (resp. u) determined

by the corresponding columns of K (resp. $\tilde{K}$). In the next subsection we will have a closer look at these columns.

In general solving the contact problem is not straightforward. Only for the special case that $B = I_N$ and $L = D_1$ there exists the so-called '*taut–string*' algorithm [6] which is based on a convex hull algorithm and requires only $\mathcal{O}(N)$ arithmetic operations. Concerning tube algorithms see also [14, 9].

We will solve the problem via the dual approach to (7).

Dual formulation. To give the dual formulation of (7) we apply that $J(u) := \|Lu\|_1$ is one–homogeneous so that its conjugate J^* is the indicator function of the convex set

$$\mathcal{S}_L := \{v \in \mathcal{R}(L^{\mathrm{T}}) : \langle v, w\rangle \le J(w) \quad \forall w \in \mathbb{R}^N\}\,. \tag{16}$$

It is easy to check that

$$\mathcal{S}_L = \{v = L^{\mathrm{T}}V : \|V\|_\infty \le 1\}\,. \tag{17}$$

Then the inclusion (9) can be rewritten as

$$\frac{1}{\beta}A(f-u) \in \partial J(u)$$

which is equivalent to

$$u \in \partial J^*\left(\frac{1}{\beta}A(f-u)\right)$$

and with $v := A(f-u)$, i.e., $u = f - A^{-1}v$ to

$$f - A^{-1}v \in \partial J^*\left(\frac{v}{\beta}\right).$$

Obviously, v fulfills this inclusion if and only if it minimizes the functional

$$\frac{1}{2}\|Bf - (B^{-1})^{\mathrm{T}}v\|_2^2 + J^*\left(\frac{v}{\beta}\right)\,. \tag{18}$$

By (17) this is the case if and only if $v = L^{\mathrm{T}}V$ and V solves the minimization problem

$$\|Bf - (B^{-1})^{\mathrm{T}}L^{\mathrm{T}}V\|_2^2 \;\to\; \min, \quad \text{s.t. } \|V\|_\infty \le \beta\,. \tag{19}$$

This is actually a quadratic optimization problem with linear constraints which can be solved by standard optimization techniques. Finally, we obtain

$$u = f - A^{-1}L^{\mathrm{T}}V\,.$$

Up to now we have not used the decomposition (8) for the solution. To see the relation to the contact problem, we assume again that $f \in \mathcal{R}(A^{-1}L^{\mathrm{T}})$. Then, using (11), we can reformulate (19) as

$$\|(B^{-1})^{\mathrm{T}}L^{\mathrm{T}}U\|_2^2 = \|U\|_{K^{-1}} \;\to\; \min, \quad \text{s.t. } \|F-U\|_\infty \le \beta \tag{20}$$

and with (13) as

$$c^{\mathrm{T}}Kc \;\to\; \min, \quad \text{s.t. } \|F-Kc\|_\infty \le \beta\,.$$

Thus, the vector U in our contact problem solves the minimization problem (20).

Structure of A. In the numerical solution of (19) as well as in the solution of the 2d problem we utilize the special structure of

$$A = A(\alpha) = I_N + \alpha D_1^{\mathrm{T}} D_1$$

that allows us to determine A^{-1} analytically. Moreover, we make use of this structure when examining the form of the solutions U and u. To this end, let

$$T(a) := \begin{pmatrix} a_0 & a_1 & \cdots & a_{N-2} & a_{N-1} \\ a_1 & a_0 & \cdots & a_{N-3} & a_{N-2} \\ \vdots & \vdots & \ddots & \vdots & \\ a_{N-2} & a_{N-3} & \cdots & a_0 & a_1 \\ a_{N-1} & a_{N-2} & \cdots & a_1 & a_0 \end{pmatrix},$$

$$H(a) := \begin{pmatrix} a_0 & a_1 & \cdots & a_{N-2} & a_{N-1} \\ a_1 & a_2 & \cdots & a_{N-1} & a_{N-2} \\ \vdots & \vdots & \ddots & \vdots & \\ a_{N-2} & a_{N-1} & \cdots & a_2 & a_1 \\ a_{N-1} & a_{N-2} & \cdots & a_1 & a_0 \end{pmatrix}$$

be the symmetric Toeplitz matrix and the persymmetric Hankel matrix generated by the vector $a \in \mathbb{R}^N$, respectively. By

$$S_{N-1} := \left(\frac{2}{N}\right)^{1/2} \left(\sin \frac{jk\pi}{N}\right)_{j,k=1}^{N-1} \in \mathbb{R}^{N-1,N-1}$$

we denote the transform matrix of the sine–I transform of length $N-1$ and by

$$C_N := \left(\frac{2}{N}\right)^{1/2} \left(\varepsilon_j \cos \frac{j(2k+1)\pi}{2N}\right)_{j,k=0}^{N-1} \in \mathbb{R}^{N,N}$$

with $\varepsilon_0 := 1/\sqrt{2}$ and $\varepsilon_j := 1$ for $j \neq 0$ the matrix of the cosine–II transform of length N, cf. [20]. Both matrices are orthogonal, i.e., $S_{N-1}\,S_{N-1} = I_{N-1}$ and $C_N^{\mathrm{T}}\,C_N = I_N$. Moreover, the vector multiplication with S_{N-1} and C_N can be realized in an FFT–like manner with only $\mathcal{O}(N \log N)$ arithmetic operations.

It is well known that these transforms are strongly related to Toeplitz plus Hankel matrices in the following sense, see, e.g., [20]:

Lemma 2.1 *The following relations hold true*

$$\begin{aligned} S_{N-1}\,\mathrm{diag}\,(d_j)_{j=1}^{N-1}\,S_{N-1} &= T(a_0,\ldots,a_{N-2}) - H(a_2,\ldots,a_{N-2},0,0)\,,\\ C_N^{\mathrm{T}}\,\mathrm{diag}\,(d_j)_{j=0}^{N-1}\,C_N &= T(a_0,\ldots,a_{N-1}) + H(a_1,\ldots,a_{N-1},0)\,, \end{aligned}$$

where

$$(d_j)_{j=0}^{N-1} = 2\left(\varepsilon_k^2\cos\frac{jk\pi}{N}\right)_{j,k=0}^{N-1}(a_0,\ldots,a_{N-2},0)^{\mathrm{T}}\,.$$

In particular, it follows by Lemma 2.1 that

$$D_1^{\mathrm{T}}D_1 = T(2,-1,0_{N-2}) + H(-1,0_{N-1}) = C_N^{\mathrm{T}}\Lambda^2 C_N,$$

where

$$\Lambda^2 := \mathrm{diag}\,(\lambda_j^2)_{j=0}^{N-1}\,,\quad \lambda_j^2 := 2-2\cos\frac{j\pi}{N} = \left(2\sin\frac{j\pi}{2N}\right)^2\,.$$

and consequently

$$A(\alpha) = I_N + \alpha D_1^{\mathrm{T}}D_1 = C_N^{\mathrm{T}}(I_N+\alpha\Lambda^2)C_N\,. \tag{21}$$

2.3 Spline character of the solution

In the following, we are interested in the structure of u and U for our original setting with A defined by (21) and $L := D_m$. Based on the representations (13) and (14) of U and u it seems to be useful to have a closer look at the matrices K and $\tilde{K}$ which for our special setting are defined by

$$K = K_m(\alpha) = \left(D_m A(\alpha)^{-1}D_m^{\mathrm{T}}\right)^{-1}\,, \tag{22}$$

$$\tilde{K} = \tilde{K}_m(\alpha) = A(\alpha)^{-1}D_m^{\mathrm{T}}K_m(\alpha). \tag{23}$$

We will see that u and U are splines of different character.

Spline structure of u. A real-valued function s defined on $[a,b]$ is a *polynomial spline of order m with knots $a<x_1<\ldots<x_r<b$* if

$$s^{(m)} = \sum_{k=1}^{r} c_k\delta(\cdot - x_k)\,,$$

where δ denotes the delta-distribution. In other words, s is a polynomial of degree $\le m-1$ on each interval $[x_k,x_{k+1}]$, $k=0,\ldots,r$; $x_0:=a$, $x_{r+1}:=b$ and $s\in C^{m-2}[a,b]$. These smoothest polynomial splines are also called *splines with defect* 1 or *with knot multiplicity* 1.

Let $n := \lfloor m/2 \rfloor$. Then we can analogously define the *discrete polynomial splines* on $\{0, \ldots, N-1\}$ *of order* m *with knots* $j_1+n, \ldots, j_r+n \in \{n, \ldots, N-\lceil \frac{m}{2} \rceil\}$ as the vectors $s \in \mathbb{R}^N$ satisfying

$$D_m\, s = \sum_{k=1}^{r} c_{j_k} e_{j_k}\,,$$

where $e_j \in \mathbb{R}^{N-m}$ denotes the j–th unit vector. Material on discrete splines can be found, e.g., in [22] and in connection with optimization problems different from the one considered here in [15, 16].

Now we see by definition (23) of $\tilde{K}_m(\alpha)$ that

$$D_m \tilde{K}_m(\alpha) = I_{N-m}\,, \quad D_m u = c\,.$$

Consequently, the k-th column of $\tilde{K}_m(\alpha)$ is a discrete polynomial spline of order m with only one knot $k+n$. The solution u of (7) is a discrete polynomial spline of order m with knots $\Xi + n$, where Ξ is given by the indices of the contact points (15).

Fig. 1 illustrates the fundamental splines given by the columns of $\tilde{K}_m(\alpha)$ for various values of m and α.

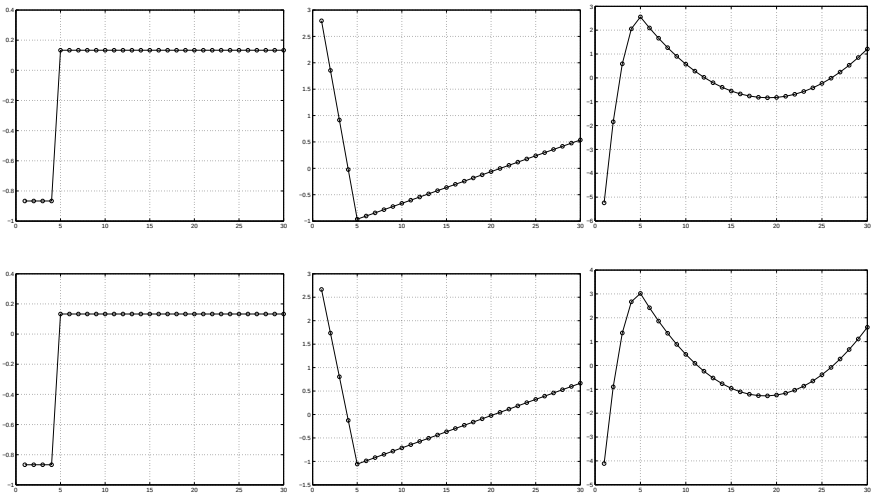

Figure 1: Fourth column of $\tilde{K}_m(\alpha)$ with knot $4, 5, 5$, resp. for $m = 1, 2, 3$ (left to right) and $\alpha = 0$ (top), $\alpha = 10$ (bottom), where $N = 30$.

Spline structure of U. For $\alpha = 0$, we have shown in [24] that

$$K_m^{\mathrm{ext}}(0) := \begin{pmatrix} 0_{m,N-1} \\ K_m(0) \\ 0_{m,N-1} \end{pmatrix}$$

is also a discrete polynomial spline of order $2m$ with knots $\Xi+m$. Here $0_{m,N}$ denotes the $m \times N$ matrix with zero entries. However, for $\alpha > 0$, things become slightly more complicated. We need the following two auxiliary lemmas.

Lemma 2.2 *Let $b_k := (-1)^k \binom{2m}{m-k}$, $k = 0, \ldots, m$ be the coefficients of the $2m$–th binomial filter multiplied by $(-1)^k$. Let $\Lambda := \operatorname{diag}(\lambda_j)_{j=0}^{N-1}$, $\tilde{\Lambda} := \operatorname{diag}(\lambda_j)_{j=1}^{N-1}$, where $\lambda_j := 2\sin\frac{j\pi}{2N}$. By*

$$R_n := (0_{N-m,n}, I_{N-m}, 0_{N-m,n}) \in \mathbb{R}^{N-m,N-m+2n}\,.$$

we denote the cutoff matrix that cancels the first and last n vector components. Then the following relations hold true:

$$\text{i)} \qquad \begin{aligned} S_{N-1}\tilde{\Lambda}^{2m} S_{N-1} &= T(b_0, \ldots, b_m, 0_{N-m-2}) - H(b_2, \ldots, b_m, 0_{N-m})\,, \\ C_N^{\mathrm{T}} \Lambda^{2m} C_N &= T(b_0, \ldots, b_m, 0_{N-m-1}) + H(b_1, \ldots, b_m, 0_{N-m})\,. \end{aligned}$$

$$\text{ii)} \qquad \begin{aligned} R_n S_{N-1}\tilde{\Lambda}^{2m} S_{N-1} R_n^{\mathrm{T}} &= T(b_0, \ldots, b_m, 0_{N-2m-1})\,, \quad \textit{for } m = 2n+1\,, \\ R_n C_N^{\mathrm{T}} \Lambda^{2m} C_N R_n^{\mathrm{T}} &= T(b_0, \ldots, b_m, 0_{N-2m-1})\,, \quad \textit{for } m = 2n\,. \end{aligned}$$

Proof. i) By Lemma 2.1 we have that

$$T(b_0, \ldots, b_m, 0_{N-m-2}) - H(b_2, \ldots, b_m, 0_{N-m}) = S_{N-1} \operatorname{diag}(d_j)_{j=1}^{N-1} S_{N-1}$$

where $d_j := b_0 + 2\sum_{k=1}^{m} b_k \cos\frac{jk\pi}{N}$. It remains to show that

$$b_0 + 2\sum_{k=1}^{m} b_k \cos\frac{jk\pi}{N} = \left(2 - 2\cos\frac{jk\pi}{N}\right)^m\,, \quad j = 1, \ldots, N-1$$

i.e., that

$$b_0 + \sum_{k=1}^{m} b_k (e^{ix} + e^{-ix}) = \left(2 - e^{ix} - e^{-ix}\right)^m\,, \quad x := \frac{j\pi}{N}\,.$$

This can easily be verified by induction on m. The second assertion of i) follows in a similar way.
ii) By i) the Hankel matrix summand influences only the first and last n rows and columns of $S_{N-1}\tilde{\Lambda}^{2m} S_{N-1}$ and $C_N^{\mathrm{T}} \Lambda^{2m} C_N$, respectively. Thus we obtain ii). □

Lemma 2.3 *Let Λ and $\tilde{\Lambda}$ be defined as in Lemma 2.2. Then the kernel $K_m^{-1}(\alpha)$ can be written as*

$$K_m^{-1}(\alpha) = \begin{cases} R_n S_{N-1} \dfrac{\tilde{\Lambda}^{2m}}{I_{N-1} + \alpha\tilde{\Lambda}^2} S_{N-1} R_n^{\mathrm{T}} & \text{for } m = 2n+1, \\ R_n C_N^{\mathrm{T}} \dfrac{\Lambda^{2m}}{I_N + \alpha\Lambda^2} C_N R_n^{\mathrm{T}} & \text{for } m = 2n\,, \end{cases}$$

where the quotient is defined componentwise.

Proof. By Lemma 2.2i) it is easy to check that

$$D_{2n,N-1} = (-1)^n\, R_n S_{N-1} \tilde{\Lambda}^{2n} S_{N-1}, \tag{24}$$
$$D_{2n,N} = (-1)^n\, R_n C_N^{\mathrm{T}} \Lambda^{2n} C_N. \tag{25}$$

First, let $m = 2n$. Then we obtain by (21) and (25) that

$$\begin{aligned} K_m^{-1}(\alpha) &= D_m\, C_N^{\mathrm{T}}\, (I_N + \alpha\Lambda^2)^{-1}\, C_N\, D_m^{\mathrm{T}} \\ &= R_n\, C_N^{\mathrm{T}}\, \frac{\Lambda^{2m}}{I_N + \alpha\Lambda^2}\, C_N\, R_n^{\mathrm{T}}\,. \end{aligned}$$

Assume now that $m = 2n + 1$. By (21) we have that

$$K_1^{-1}(\alpha) = D_1\, C_N^{\mathrm{T}}\, (I_N + \alpha\Lambda^2)^{-1}\, C_N\, D_1^{\mathrm{T}}.$$

Straightforward computation gives

$$\begin{aligned} D_1\, C_N^{\mathrm{T}} &= \left(\frac{2}{N}\right)^{1/2} \left(\varepsilon_j \left(\cos\, \frac{j(2k+3)\pi}{2N} - \cos\, \frac{j(2k+1)\pi}{2N}\right)\right)_{k,j=0}^{N-2,N-1} \\ &= \left(\frac{2}{N}\right)^{1/2} \left(-2\varepsilon_j\, \sin\, \frac{j(k+1)\pi}{N}\, \sin\, \frac{j\pi}{2N}\right)_{k,j=0}^{N-2,N-1} \\ &= -\,(0_{N-1,1}, S_{N-1})\, \Lambda \end{aligned}$$

and consequently

$$\begin{aligned} K_1^{-1}(\alpha) &= (0_{N-1,1}, S_{N-1})\, \Lambda\, (I_N + \alpha\Lambda^2)^{-1}\, \Lambda\, (0_{N-1,1}, S_{N-1})^{\mathrm{T}} \\ &= S_{N-1}\, \frac{\tilde{\Lambda}^2}{I_{N-1} + \alpha\tilde{\Lambda}^2}\, S_{N-1}\,. \end{aligned}$$

Using this relation and (24) we obtain

$$\begin{aligned} K_m^{-1}(\alpha) &= D_{m-1,N-1}\, D_1\, (I_N + \alpha\Lambda^2)^{-1}\, D_1^{\mathrm{T}}\, D_{m-1,N-1}^{\mathrm{T}} \\ &= R_n S_{N-1}\, \frac{\tilde{\Lambda}^{2m}}{I_{N-1} + \alpha\tilde{\Lambda}^2}\, S_{N-1} R_n^{\mathrm{T}}. \end{aligned}$$

This completes the proof. □

Now we are ready to describe the structure of $K_m(\alpha)$.

Theorem 2.4 *Let z^{T} denote the n–th row of $C_N^{\mathrm{T}}\, \frac{\Lambda^{2m}}{I_N+\alpha\Lambda^2}\, C_N\, R_n^{\mathrm{T}}\, K_m(\alpha)$ if $m = 2n$ and the n–th row of $-S_{N-1}\, \frac{\tilde{\Lambda}^{2m}}{I_{N-1}+\alpha\tilde{\Lambda}^2}\, S_{N-1} R_n^{\mathrm{T}}\, K_m(\alpha)$ if $m =$*

$2n+1$. *By z_{rev} we denote the vector obtained by reversing the order of the components of z. Then our kernels*

$$K_m^{\rm ext}(\alpha) := \begin{pmatrix} \alpha z^{\rm T} \\ 0_{m-1,N-m} \\ K_m(\alpha) \\ 0_{m-1,N-m} \\ \alpha z_{rev}^{\rm T} \end{pmatrix}$$

fulfill

$$D_{2m,N+m}\, K_m^{\rm ext}(\alpha) = (-1)^m\, T(1+2\alpha, -\alpha, 0_{N-m-2}). \tag{26}$$

Proof. We restrict our attention to even $m = 2n$. The proof for odd $m = 2n+1$ follows the same lines. By Lemma 2.2ii) we have that

$$\begin{aligned} T(b_0,\ldots,b_m,0_{N-2m-1}) &= R_n C_N^{\rm T} \Lambda^{2m} C_N R_n^{\rm T} \\ &= R_n\, C_N^{\rm T}(I_N+\alpha\Lambda^2)\, C_N\, C_N^{\rm T}\, \frac{\Lambda^{2m}}{I_N+\alpha\Lambda^2}\, C_N\, R_n^{\rm T}. \end{aligned}$$

Using Lemma 2.3 and considering the tridiagonal structure of $C_N^{\rm T}(I_N + \alpha\Lambda^2)\, C_N$ this can be rewritten as

$$\begin{aligned} &T(b_0,\ldots,b_m,0_{N-2m-1}) \\ &\quad = \; T(1+2\alpha,-\alpha,0_{N-m-2})\, K_m^{-1}(\alpha) \;-\; \alpha \begin{pmatrix} a^{\rm T} \\ 0_{N-m-2,N-m} \\ a_{rev}^{\rm T} \end{pmatrix}, \end{aligned}$$

where $a^{\rm T}$ denotes the n–th row of $C_N^{\rm T}\, \frac{\Lambda^{2m}}{I_N+\alpha\Lambda^2}\, C_N\, R_n^{\rm T}$. Multiplication with $K_m(\alpha)$ results in

$$\begin{aligned} &T(b_0,\ldots,b_m,0_{N-2m-1})K_m(\alpha) \;+\; \alpha \begin{pmatrix} z^{\rm T} \\ 0_{N-m-2,N-m} \\ z_{rev}^{\rm T} \end{pmatrix} \\ &\qquad = \quad T(1+2\alpha,-\alpha,0_{N-m-2})\,. \end{aligned}$$

Now we can enlarge $(-1)^m T(b_0,\ldots,b_m,0_{N-2m-1})$ by $2m$ rows and columns to $(-1)^m D_{2m,N+m}$ to obtain

$$\begin{aligned} &(-1)^m\, D_{2m,N+m} \begin{pmatrix} 0_{m,N-m} \\ K_m(\alpha) \\ 0_{m,N-m} \end{pmatrix} \;+\; \alpha \begin{pmatrix} z^{\rm T} \\ 0_{N-m-2,N-m} \\ z_{rev}^{\rm T} \end{pmatrix} \\ &\qquad = \quad T(1+2\alpha,-\alpha,0_{N-m-2}). \end{aligned}$$

Since the first coefficient in the Toeplitz matrix $(-1)^m\, D_{2m,N+m}$ is equal to 1, this can be rewritten in the form (26). □

To see the relation of (26) to splines, let us repeat that a real-valued function s defined on $[a,b]$ is a *polynomial spline of order m with knots $a < x_1 < \ldots < x_r < b$ and defect* (*knot multiplicity*) 3 if

$$s^{(m)} = \sum_{k=1}^{r} c_k \delta(\cdot - x_k) + c'_k \delta'(\cdot - x_k) + c''_k \delta''(\cdot - x_k)\,.$$

In other words, s is of lower smoothness, namely $s \in C^{m-4}[a,b]$. Here we may restrict our interest to the splines with defect 3 satisfying

$$s^{(m)} = \sum_{k=1}^{r} c_k (\delta(\cdot - x_k) + \alpha \delta''(\cdot - x_k))\,. \tag{27}$$

Using

$$\begin{array}{rcl} e''_0 & := & (2,-1,0_{N-m-2})^{\mathrm{T}}, \quad e''_{N-m-1} := (0_{N-m-2},-1,2)^{\mathrm{T}}, \\ e''_j & := & (0_{j-1},-1,2,-1,0_{N-m-2-j})^{\mathrm{T}}, \quad j = 1,\ldots,N-m-2 \end{array}$$

as discrete counterparts of $\delta''(\cdot - k)$, we see by (26)that the k–th column K_k of $K_m^{\mathrm{ext}}(\alpha)$ is a discrete version of (27) with only one knot $k+m$:

$$D_{2m}\, K_k = (-1)^m (e_k + \alpha e''_k)\,.$$

Consequently, by (13),

$$D_{2m}\, U^{\mathrm{ext}} = (-1)^m \sum_{k \in \Xi}^{r} c_k (e_k + \alpha e''_k)\,,$$

so that $U^{\mathrm{ext}} := (\alpha z^{\mathrm{T}} c, 0_{m-1}, U^{\mathrm{T}}, 0_{m-1}, \alpha z_{rev}^{\mathrm{T}} c^{\mathrm{T}})^{\mathrm{T}}$ can be considered as discrete version of a polynomial spline of order $2m$ with defect 3 and knots $\Xi + m$.

The columns of the kernel $K_m(\alpha)$ are illustrated in Fig. 2.

3 Second order ℓ_1 regularization in 2d

3.1 Discrete setting

For simplicity, we restrict our attention to quadratic (n,n) images and reshape them column by column into a vector f of length $N := n^2$. As a discrete counterpart of (1) we are interested in minimizing strictly convex functionals of the form

$$F(u) = \frac{1}{2}\|f-u\|_2^2 \;+\; \frac{\alpha}{2}\|\mathcal{D}_1 f - \mathcal{D}_1 u\|_2^2 \;+\; \beta \| \, |\mathcal{D}_m u| \, \|_1\,. \tag{28}$$

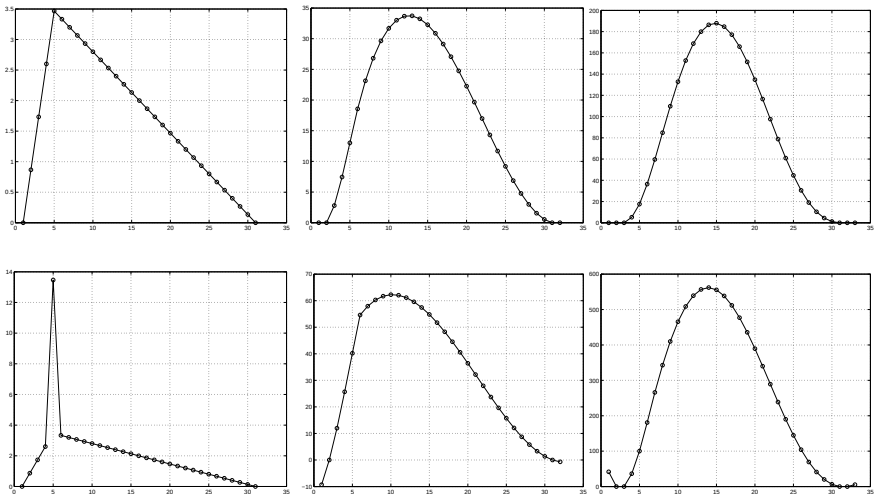

Figure 2: Fourth column of $K_m^{\mathrm{ext}}(\alpha)$ with knot $5, 6, 6$ for $m = 1, 2, 3$, resp. (left to right) and $\alpha = 0$ (top), $\alpha = 10$ (bottom), where $N = 30$.

Here $\mathcal{D}_m : \mathbb{R}^N \to \mathbb{R}^{pN}$, $p = p(m) \geq 2$ is a (weighted) discrete partial derivative operator of order m and

$$|V| := \left(\|(V_{j+rN})_{r=0}^{p-1}\|_2 \right)_{j=0}^{N-1}, \quad V \in \mathbb{R}^{pN}.$$

Note that

$$\| \, |V| \, \|_1 := \sum_{j=0}^{N-1} |V|_j \quad \text{and} \quad \| \, |V| \, \|_\infty := \max_{j=0,\ldots,N-1} |V|_j \tag{29}$$

are dual norms on $\mathbb{R}^{pN}$.

In this paper, we focus on first and second order partial derivatives with the following operators $\mathcal{D}_1$ and $\mathcal{D}_2$, respectively. Using the difference matrix $D_1 = D_{1,n}$ defined by (3), we set

$$\mathcal{D}_1 := \begin{pmatrix} I_n \otimes \tilde{D}_1 \\ \tilde{D}_1 \otimes I_n \end{pmatrix} \in \mathbb{R}^{2N,N}, \quad \tilde{D}_1 := \begin{pmatrix} D_1 \\ 0_{1,n} \end{pmatrix}.$$

The multiplication with $\mathcal{D}_1$ mimics a discrete gradient operator, where the upper N rows correspond to the derivation in x direction and the lower N rows to the derivation in y direction. Moreover, $|\mathcal{D}_1 u|$ is a discrete version of the absolute value of the gradient $|\nabla u| = (u_x^2 + u_y^2)^{1/2}$. For a more sophisticated discretization of $|\nabla u|$ see, e.g., [26]. Further, let

$$\mathcal{D}_2 := \begin{pmatrix} I_n \otimes D_1^{\mathrm{T}} D_1 \\ D_1^{\mathrm{T}} D_1 \otimes I_n \\ \tilde{D}_1^{\mathrm{T}} \otimes \tilde{D}_1 \\ \tilde{D}_1 \otimes \tilde{D}_1^{\mathrm{T}} \end{pmatrix}$$

be our discrete version of the second order partial derivative operators ∂_{xx}, ∂_{yy}, ∂_{yx}, ∂_{xy} (top rows to bottom rows). Note that $D_1^{\mathrm{T}} D_1 = \tilde{D}_1^{\mathrm{T}} \tilde{D}_1$. Then we see that $|\mathcal{D}_2 u|$ is the discrete version of the Frobenius norm of the Hessian

$$\nabla^2 u := \begin{pmatrix} u_{xx} & u_{xy} \\ u_{yx} & u_{yy} \end{pmatrix}.$$

For a variational method including the Hessian see also [10]. Of course other discretizations of second order derivatives are possible and sometimes also necessary, for example if integral identities have to be preserved, see, e.g., [29]. Moreover, other coupled versions of the second order derivatives, e.g. $(u_{xx}^2 + u_{yy}^2)^{1/2}$ are possible.

Now the functional (28) can be rewritten for $m = 1, 2$ as

$$F(u) = \frac{1}{2}\|B(f-u)\|_2^2 \; + \; \beta \| \, |Lu| \, \|_1 \tag{30}$$

with $L \in \{\mathcal{D}_1, \mathcal{D}_2\}$ and $A = B^{\mathrm{T}} B$,

$$A = A(\alpha) := I_N + \alpha \mathcal{D}_1^{\mathrm{T}} \mathcal{D}_1.$$

The matrix $\mathcal{D}_1^{\mathrm{T}} \mathcal{D}_1$ is just the central difference discretization of the Laplacian with Neumann boundary conditions which can be diagonalized by Kronecker products of the cosine transform matrices C_n. More precisely, we obtain that

$$A(\alpha) = (C_n \otimes C_n)^{\mathrm{T}} (I_N + \alpha \Lambda_2^2)(C_n \otimes C_n) \tag{31}$$

with $\Lambda_2^2 = \Lambda^2 \otimes I_n + I_n \otimes \Lambda^2$, $\Lambda := \operatorname{diag}\left(2 \sin \frac{j\pi}{2n}\right)_{j=0}^{n-1}$.

3.2 Dual formulation

Since $J(u) := \| \, |Lu| \, \|_1$ is one–homogeneous the functional (30) can be minimized as in 1d by switching to the dual minimization problem

$$\frac{1}{2}\|Bf - (B^{-1})^{\mathrm{T}} v\|_2^2 + J^* \begin{pmatrix} v \\ \beta \end{pmatrix} \; \to \min, \tag{32}$$

where J^* is again the indicator function of the set

$$\mathcal{S}_L := \{ v \in \mathcal{R}(L^{\mathrm{T}}) : \langle v, w \rangle \le J(w) \quad \forall w \in \mathbb{R}^N \}, \tag{33}$$

cf. (16) and u is related to v by $u = f - A^{-1} v$. By Lemma A.1 this set is also given by

$$\mathcal{S}_L = \{ v \in \mathcal{R}(L^{\mathrm{T}}) : \min_{v = L^{\mathrm{T}} V} \| \, |V| \, \|_\infty \le 1 \}. \tag{34}$$

For $L := \mathcal{D}_1$, the norm $\|v\|_G := \min\limits_{v = L^{\mathrm{T}} V} \| \, |V| \, \|_\infty$ is just a discrete version of Meyer's G–norm which is known as dual norm of the BV norm on the

closed subspace $\mathcal{BV}$ of functions of bounded variation with gradient in L_1. Concerning higher order derivatives and G–norms see also [18].

With $v := L^{\mathrm{T}}V$ problem (32) is equivalent to

$$\|Bf - (B^{-1})^{\mathrm{T}}L^{\mathrm{T}}V\|_2^2 \;\to\; \min, \quad \text{s.t. } \| \, |V| \, \|_\infty \le \beta \, .$$

This is a quadratic minimization problem with quadratic constraints (if squared). In our numerical examples we have solved this problem by adapting an algorithm by Chambolle [1]. This algorithm allows us to incorporate fast cosine transforms arising from the special structure of A in (31) in a simple way. Other solution methods by including auxiliary variables, see e.g. [25] or in connection with multiplicative half-quadratic minimization [8, 17, 4] or by using second order cone programming [27] are possible but not superior in the application at hand.

Applying (31), the adapted Chambolle algorithm reads as follows:

Algorithm 3.1 *Input:* $u^{(0)} := f \in \mathbb{R}^{n,n}$ *and* $V_r^{(0)} := 0_{n,n}$, $r = 1, \ldots, 4$.

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

$$\begin{aligned}
W_1^{(k)} &:= D_1^{\mathrm{T}} D_1 u^{(k)} \\
W_2^{(k)} &:= u^{(k)} D_1^{\mathrm{T}} D_1 \\
W_3^{(k)} &:= \tilde{D}_1 u^{(k)} \tilde{D}_1 \\
W_4^{(k)} &:= \tilde{D}_1^{\mathrm{T}} u^{(k)} \tilde{D}_1^{\mathrm{T}} \\
|W^{(k)}| &:= \left(\sum_{r=1}^{4} (W_r^{(k)})^2 \right)^{1/2} \quad \text{componentwise} \\
V_r^{(k+1)} &:= \frac{V_r^{(k)} + \tau W_r^{(k)}}{1_{n,n} + \frac{\tau}{\beta} |W^{(k)}|}, \quad r = 1, \ldots, 4 \\
x^{(k+1)} &:= D_1^{\mathrm{T}} D_1 V_1^{(k+1)} + V_2^{(k+1)} D_1^{\mathrm{T}} D_1 + \tilde{D}_1^{\mathrm{T}} V_3^{(k+1)} \tilde{D}_1^{\mathrm{T}} + \tilde{D}_1 V_4^{(k+1)} \tilde{D}_1 \\
u^{(k+1)} &:= f - C_n^{\mathrm{T}} \left(\frac{C_n x^{(k+1)} C_n^{\mathrm{T}}}{M} \right) C_n
\end{aligned}$$

where M *is the matrix reshaped version of* $I_N + \alpha \Lambda_2^2$ *and the quotients are taken componentwise.*
Output: $u := u^{(k+1)}$.

Since the difference matrices are sparse and the vector multiplication with C_n can be performed in $\mathcal{O}(n \log n)$ arithmetic operations, one step of the algorithm requires only $\mathcal{O}(n^2 \log n)$ arithmetic operations.

Chambolle proved that $u^{(k)}$ converges to the solution u if

$$\tau \le 1/\|(B^{-1})^{\mathrm{T}} L^{\mathrm{T}}\|_2^2 = 1/\|LA^{-1}L^{\mathrm{T}}\|_2.$$

Obviously we have that $\|A^{-1}\|_2 \leq 1$ and $\|\mathcal{D}_1\|_2^2 = 8$. Further, we see by applying Gerschgorin's theorem that $\|\mathcal{D}_2\|_2^2 \leq 64$. Hence we have to choose

$$\tau \leq \begin{cases} 1/8 & \text{for } L = \mathcal{D}_1, \\ 1/64 & \text{for } L = \mathcal{D}_2. \end{cases}$$

3.3 Numerical examples

In this section, we present numerical examples for the denoising of grey value images in 2d. Since for a human observer as well as for some computer vision systems edges are a very important source of information in an image, one of the major goals of denoising algorithms is to preserve or even enhance edges. The main reason for introducing our model with the additional ℓ_2 gradient fitting term was to avoid both the staircasing effect and the blurring of edges.

In our experiments we assume an additive noise model: Let $f \in \mathbb{R}^{n\times n}$ be a noisy version of the initial image $g \in \mathbb{R}^{n\times n}$, degraded with additive white Gaussian noise η, i.e., $f_{ij} = g_{ij} + \eta_{ij}$. As quality measures for the denoised image u we use the ℓ_1 norm of $g - u$ and the Signal-to-Noise Ratio (SNR) defined as

$$\mathrm{SNR}(g,u) := 10\log_{10}\left(\frac{\sum_{i,j=1}^n (g_{ij}-\mu)^2}{\sum_{i,j=1}^n (u_{ij}-g_{ij})^2}\right),$$

where $\mu := \frac{1}{n^2}\sum_{i,j=1}^n g_{ij}$ denotes the mean value of g. The SNR is a widely used measure in image processing and essentially gives the same information as an ℓ_2 distance. Roughly speaking, the larger the SNR the better is the expected image quality.

Our adapted Chambolle algorithm 3.1 was implemented in MATLAB. In addition to our approach we performed various experiments with the inf-convolution model proposed in [3]. Using our notation, the authors in [3] find the denoised image $u = u_1 + u_2$ by solving

$$\arg\min_{u_1,u_2} \frac{1}{2}\|f - u_1 - u_2\|_2^2 + \alpha\| \, |\mathcal{D}_1 u_1| \, \|_1 + \beta\| \, |\mathcal{D}_2 u_2| \, \|_1. \tag{35}$$

Originally the parameters are slightly coupled, i.e., $\beta = \alpha\lambda$ for some $\lambda \geq 0$. To solve (35) numerically we use an alternating minimization procedure that finds successively

$$\arg\min_{u_2} \frac{1}{2}\|f - u_1 - u_2\|^2 + \beta\| \, |\mathcal{D}_2 u_2| \, \|_1$$

and

$$\arg\min_{u_1} \frac{1}{2}\|f - u_1 - u_2\|^2 + \alpha\| \, |\mathcal{D}_1 u_1| \, \|_1.$$

The convergence of this procedure to the minimizer of (35) follows as in [12].

	m	α	β	SNR	$\|u-g\|_1 \cdot 10^{-5}$
Noisy image	-	-	-	11.16	10.00
ROF	1	0	50	22.74	2.08
Alg. 3.1	2	0	50	20.00	1.88
Alg. 3.1	2	1.2	50	23.13	1.64
Alg. (35)	2	20	50	22.54	1.65

Table 1: Denoising experiment in 2d: Parameters and error measures.

The top of Fig. 3 contains our test image and its noisy version. Table 1 shows the parameters and the error measures of the resulting denoised images. The error measures show that adding a gradient fitting term to variational denoising with second-order derivatives enables us to improve the results. While the ROF model leads to staircasing artifacts, the pure second order method suffers from blurred edges. The additional ℓ_2 gradient fitting can help to avoid both types of problems. One alternating minimization iteration of the inf-convolution approach (35) leads to similar results but with higher numerical effort. With a high expense of over 100 alternating minimization iterations one can further slightly improve the result. However, note that we have not optimized the parameters α and β with respect to our error measures but have chosen them to give a visually good impression. The results were confirmed by the images in Fig. 3 and their sections in Fig. 4. The staircasing effect that appears if we use only first order coupled partial derivatives is clearly visible. On the other hand, one can see the blurring of edges when using purely second order partial derivatives. The image quality can be improved by additional ℓ_2 gradient fitting.

A Appendix

By the following lemma, we see that the sets in (33) and (34) are equivalent.

Lemma A.1 *Let $L \in \mathbb{R}^{pN,N}$. Then*

$$\sup_{|Lw|\neq 0} \frac{|\langle V', Lw\rangle|}{\| \, |Lw| \, \|_1} = \min_{L^{\mathrm{T}}U = L^{\mathrm{T}}V'} \| \, |U| \, \|_\infty \, .$$

Proof. Let $\nu := \sup_{|Lw|\neq 0} \frac{|\langle V', Lw\rangle|}{\| \, |Lw| \, \|_1}$. By applying the Schwarz inequality to $(U_j^r)_{r=1}^p$ and $((Lw)_j^r)_{r=1}^p$ for $j = 0, \ldots, N-1$, we obtain

$$|\langle U, Lw\rangle| \leq \langle |U|, |Lw|\rangle \leq \| \, |U| \, \|_\infty \, \| \, |Lw| \, \|_1 .$$

Since $|\langle V', Lw\rangle| = |\langle L^{\mathrm{T}}V', w\rangle| = |\langle U, Lw\rangle|$ for all $U \in \mathbb{R}^{pN}$ with $L^{\mathrm{T}}V' = L^{\mathrm{T}}U$ and all $w \in \mathbb{R}^N$ we obtain that

$$\nu \leq \min_{L^{\mathrm{T}}U = L^{\mathrm{T}}V'} \| \, |U| \, \|_\infty \, . \tag{36}$$

To show the reverse direction we consider the subspace $\mathcal{B} := \mathcal{R}(L)$ of $\mathbb{R}^{pN}$ equipped with the norm $\|\,|\cdot|\,\|_1$. The mapping $l_{V'}(Lw) := \langle V', Lw\rangle$ is a linear functional on $\mathcal{B}$ which has exactly the norm ν. By the Hahn–Banach Theorem this functional can be extended to a linear functional l on $(\mathbb{R}^{pN}, \|\,|\cdot|\,\|_1)$ with $\|l\| = \|l_{V'}\|$. Consequently, there exists $\tilde{V} \in \mathbb{R}^{pN}$ such that $l(V) = \langle \tilde{V}, V\rangle$ for all $V \in \mathbb{R}^{pN}$ and

$$\langle \tilde{V}, Lw\rangle = \langle V', Lw\rangle \quad \forall w \in \mathbb{R}^N\,.$$

This can be rewritten as

$$\langle L^{\mathrm{T}}\tilde{V}, w\rangle = \langle L^{\mathrm{T}}V', w\rangle \quad \forall w \in \mathbb{R}^N$$

so that the vector $\tilde{V}$ must fulfil $L^{\mathrm{T}}\tilde{V} = L^{\mathrm{T}}V'$. Since the norms $\|\,|\cdot|\,\|_1$ and $\|\,|\cdot|\,\|_\infty$ are dual norms on $\mathbb{R}^{pN}$ we see that

$$\|l_{V'}\| = \|l\| = \|\,|\tilde{V}|\,\|_\infty.$$

Together with (36) this yields the assertion. □

Acknowledgements. This work has been partially funded by the *Deutsche Forschungsgemeinschaft* (DFG).

References

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

[2] A. Chambolle. Total variation minimization and a class of binary MRF models. In *EMMCVPR*, pages 136–152, 2005.

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

[4] R. H. Chan and M. Nikolova. Half-quadratic minimization of regularized objectives as the fixed point iteration of the linearized gradient. *Preprint*, 2005.

[5] 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.

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

[7] 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.

[8] D. Geman and G. Reynolds. Constrained restoration and the recovery of discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:367–383, 1992.

[9] 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.

[10] 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.

[11] W. Hintermüller and K. Kunisch. Total bounded variation regularization as a bilaterally constrained optimization problem. *SIAM Journal on Applied Mathematics*, 64(4):1311–1333, May 2004.

[12] L. B.-F. J. F. Aujol, G. Aubert and A. Chambolle. Image decomposition into a bounded variation component and an oscillating component. *Journal of Mathematical Imaging and Vision*, 22:71–88, Oct. 2005.

[13] 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. *IEEE Transactions on Image Processing*, 12(12):1579 – 1590, 2003.

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

[15] O. L. Mangasarian and L. L. Schumaker. Discrete splines via mathematical programming. *SIAM Journal on Control*, 9(2):174–183, 1971.

[16] 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.

[17] M. Nikolova and M. K. Ng. Analysis of half-quadratic minimization methods for signal and image recovery. *SIAM Journal on Scientific Computing*, 27(3):937–966, 2005.

[18] 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.

[19] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for the total variation based image restoration. *Multiscale Modeling and Simulation*, 4:460–489, 2005.

[20] D. Potts and G. Steidl. Optimal trigonometric preconditioners for nonsymmetric Toeplitz systems. *Linear Algebra and its Applications*, 281:265–292, 1998.

[21] 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.

[22] L. L. Schumaker. *Spline Functions: Basic Theory.* Wiley and Sons, New York, 1981.

[23] G. Steidl. A note on the dual treatment of higher order regularization functionals. *Computing*, 76:135–148, 2005.

[24] G. Steidl, S. Didas, and J. Neumann. Splines in higher order tv regularization. *International Journal of Computer Vision*, to appear, 2006.

[25] C. R. Vogel. *Computational Methods for Inverse Problems.* SIAM, Philadelphia, 2002.

[26] M. Welk, G. Steidl, and J. Weickert. Locally analytic schemes: A link between diffusion filtering and wavelet shrinkage. Technical report, IMA Preprint, University of Minnesota, 2006.

[27] W. Yin, D. Goldfarb, and S. Osher. Image cartoon-texture decomposition and feature selection using the total variation regularized l^1 functional. Technical Report 05-45, UCLA Report, USA, 2005.

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

[29] J. Yuan, R. Ruhnau, E. Mémin, and C. Schnörr. Discrete orthogonal decomposition and variational fluid flow estimation. In *Scale Space and PDE Methods in Computer Vision*, volume 3459 of *LNCS*, pages 267–278. Springer, 2005.

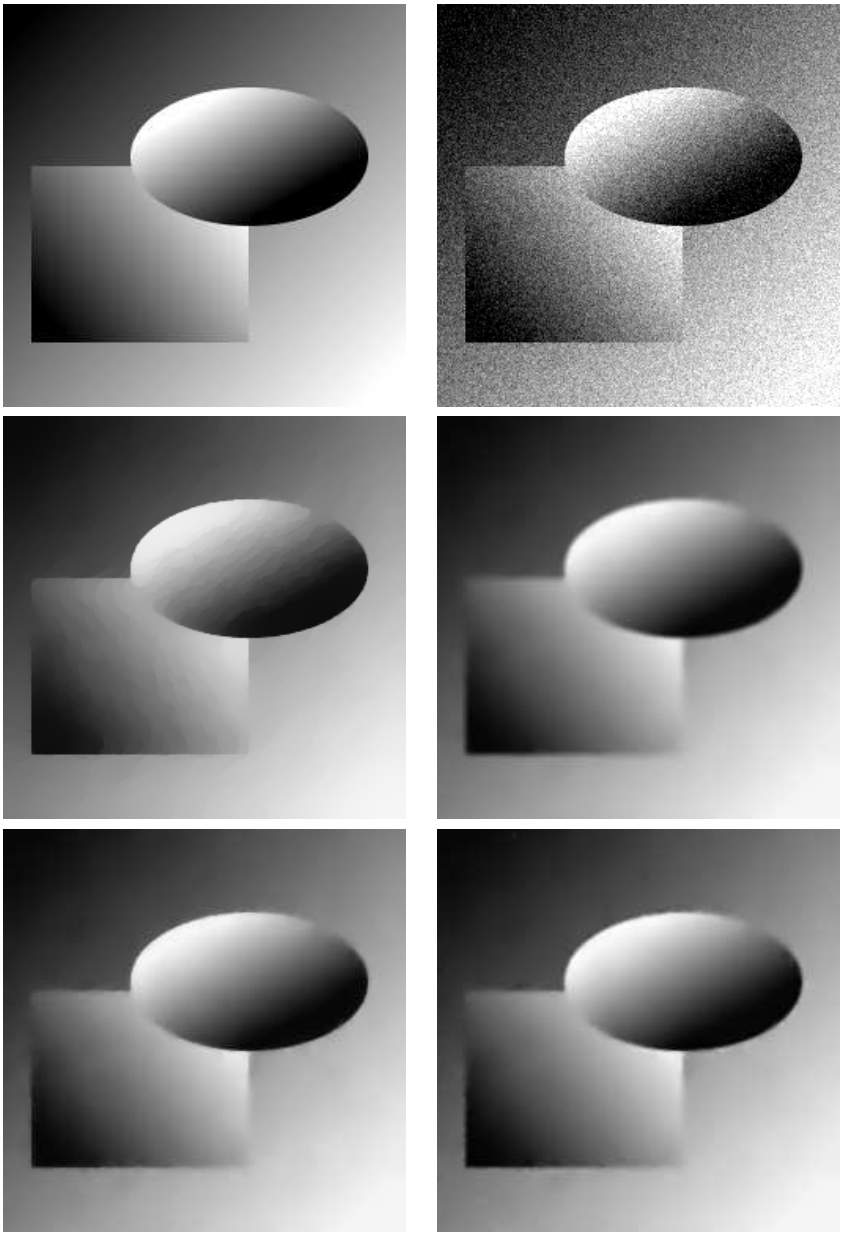

Figure 3: Denoising experiment in 2d. Top left: Original image (size 256×256). Top right: Image with additive noise, SNR 11.16. Middle left: Denoised image with ROF model $m = 1$, $\alpha = 0$, $\beta = 50$. The starcaising effect is visible. Middle right: Denoised image with second order model $m = 2$, $\alpha = 0$, $\beta = 50$. Edges are blurred. Bottom left: Denoised image with second order model (28) $m = 2$, $\alpha = 1.2$, $\beta = 50$. Edges become sharper again by adding the ℓ_2 gradient fitting term. Bottom right: Denoised image with inf-convolution model (35) and $\alpha = 20$, $\beta = 50$. Edges become sharper. Sections of the images can be seen in Fig. 4.

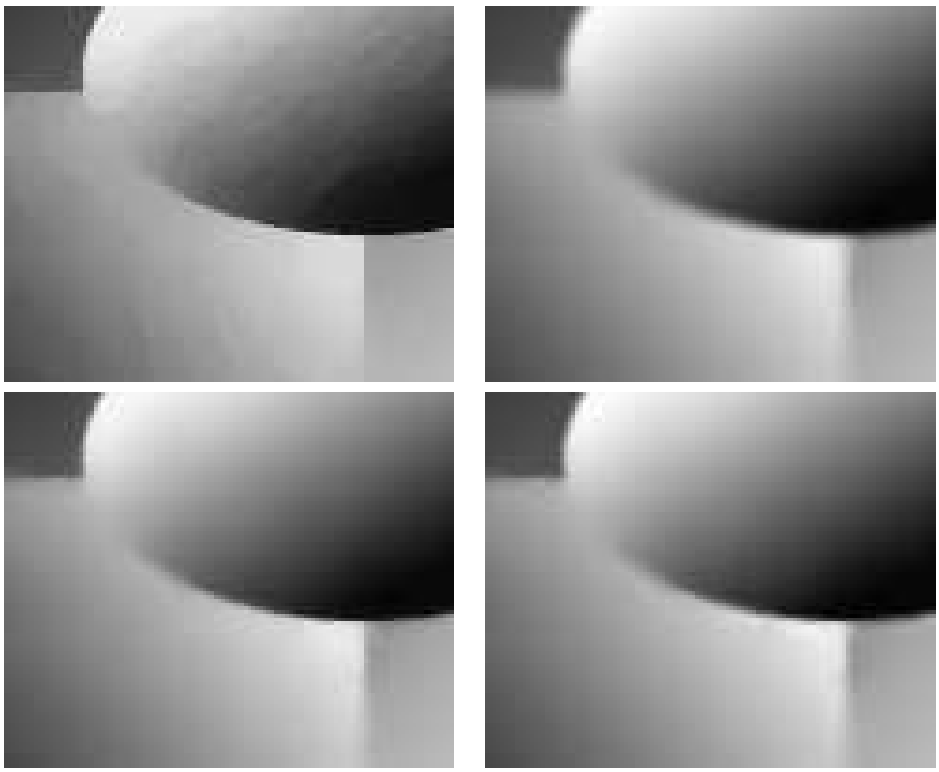

Figure 4: Denoising experiment in 2d. 100×120 sections of the denoised images in Fig. 3. Top left: Part of the denoised image with ROF model $m = 1$, $\alpha = 0$, $\beta = 50$. Top right: Part of the denoised image with second order model $m = 2$, $\alpha = 0$, $\beta = 50$. Bottom left: Part of the denoised image with second order model (28) $m = 2$, $\alpha = 1.2$, $\beta = 50$. Bottom right: Part of the denoised image with inf-convolution model (35) $\alpha = 20$, $\beta = 50$.