International Journal of Computer Vision 70(3), 241–255, 2006

DOI: 10.1007/s11263-006-8066-7

Splines in Higher Order TV Regularization

GABRIELE STEIDL
University of Mannheim, Faculty of Mathematics and Computer Science , 68131 Mannheim, Germany
steidl@math.uni-mannheim.de

STEPHAN DIDAS
Mathematical Image Analysis Group, Faculty of Mathematics and Computer Science, Saarland University, 66123 Saarbrücken, Germany
didas@mia.uni-saarland.de

JULIA NEUMANN
University of Mannheim, Faculty of Mathematics and Computer Science , 68131 Mannheim, Germany
jneumann@uni-mannheim.de

Received June 2, 2005; Revised December 12, 2005; Accepted January 4, 2006

Abstract. Splines play an important role as solutions of various interpolation and approximation problems that minimize special functionals in some smoothness spaces. In this paper, we show in a strictly discrete setting that splines of degree $m-1$ solve also a minimization problem with quadratic data term and m-th order total variation (TV) regularization term. In contrast to problems with quadratic regularization terms involving m-th order derivatives, the spline knots are not known in advance but depend on the input data and the regularization parameter λ. More precisely, the spline knots are determined by the contact points of the m–th discrete antiderivative of the solution with the tube of width 2λ around the m-th discrete antiderivative of the input data. We point out that the dual formulation of our minimization problem can be considered as support vector regression problem in the discrete counterpart of the Sobolev space $W_{2,0}^m$. From this point of view, the solution of our minimization problem has a sparse representation in terms of discrete fundamental splines.

Keywords: higher order TV regularization, splines, support vector regression, Legendre-Fenchel dualization taut-string algorithm

1. Introduction

In this paper, we are interested in the solution of the minimization problem

$$\frac{1}{2}\int_0^1 (u(x)-f(x))^2+\lambda|u^{(m)}(x)|\,\mathrm{d}x \quad \rightarrow \quad \min \quad (1)$$

and some of its 2D versions involving first and second order partial derivatives. More precisely, we work in a strictly discrete setting which is appropriate for tasks in digital signal processing. For a discrete signal $u = (u(1), \ldots , u(n))^{\mathrm{T}}$, we use the m-th *forward difference*

$$\triangle^m u(j) := \sum_{k=0}^{m}(-1)^{k+m}\binom{m}{k}u(j+k), \quad j = 1, \ldots , n-m \qquad (2)$$

as discretization of the m-th derivative. Then, for given input data $\boldsymbol{f} \in \mathbb{R}^n$, we are looking for the solution of the minimization problem

$$\frac{1}{2}\sum_{j=1}^{n}(u(j) - f(j))^2 + \lambda \sum_{j=1}^{n-m} |\triangle^m u(j)| \quad \rightarrow \quad \min, \tag{3}$$

where we refer to the penalty term as m–order TV regularization. Of course, other discretizations of (1) are possible. In contrast to the solution of the well examined version of (3) with quadratic penalty term $|\triangle^m u(j)|^2$, the solution of (3) does not linearly depend on the input data. This results in some advantages over the linear solution as better edge preserving. For two dimensions and first order derivatives in the penalizer, problem (3) becomes the classical approach of Rudin, Osher and Fatemi (ROF) Rudin et al. (1992) which has many applications in digital image processing. Meanwhile there exist various solution methods for this problem, see Vogel (2002) and the references therein. Most of these methods introduce a small additional smoothing parameter to cope with the non differentiability of $|\cdot|$. There are two approaches which avoid such an additional parameter, namely a wavelet inspired technique Welk et al. (2005) and the Legendre–Fenchel dualization technique, see, e.g., Chambolle (2004); Chan et al. (1999) which is also relevant in the present considerations. We further mention that other cost functionals than the quadratic one have to come into the play when dealing, e.g., with denoising of images corrupted with other than white Gaussian noise. In this context we only refer to recent papers of Nikolova (2004); Chan et al. and the references therein.

In this paper, we are interested in the structure of the solution $\boldsymbol{u}$ even for $m > 1$. We show that $\boldsymbol{u}$ is a discrete spline of degree $m-1$, where the spline knots, in contrast to the linear problem with quadratic regularization term, depend on the input data $\boldsymbol{f}$ and on the regularization parameter λ. More precisely, the spline knots are determined by the contact points of the m–th discrete antiderivative of $\boldsymbol{u}$ with the tube of width 2λ around the m–th discrete antiderivative of $\boldsymbol{f}$. We will see that the dual formulation of our minimization problem can be considered as support vector regression (SVR) problem in the discrete counterpart of the Sobolev space $W^m_{2,0}$. The SVR problem can be solved by standard quadratic programming methods. This provides us with a sparse representation of $\boldsymbol{u}$ in terms of discrete fundamental splines. We formally extend the approach to two dimensions. Here further research has to be involved to see the relation, e.g., to classical radial basis functions.

This paper is organized as follows: since discrete approaches can be best described in matrix–vector notation, the next section introduces the basic difference operators as matrices. Section 3 shows that our minimization problem (3) is equivalent to a spline contact problem. To this end, we have to define discrete splines. Based on the dual formulation of our problem, Section 4 treats the spline contact problem as support vector regression problem and presents some denoising results. Section 5 gives future prospects to two dimensional problems. The paper is concluded with Section 6.

2. Difference Matrices

The discrete setting can be best handled using matrix-vector notation. To this end, we introduce the lower triangular $n \times n$ Toeplitz matrix

$$\boldsymbol{D}_n := \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ -1 & 1 & \dots & 0 & 0 \\ & \ddots & \ddots & & \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & -1 & 1 \end{pmatrix}.$$

By straightforward computation we see that the inverse of $\boldsymbol{D}_n$ is the addition matrix

$$\boldsymbol{A}_n := \boldsymbol{D}_n^{-1} = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 1 & 1 & \dots & 0 & 0 \\ & \ddots & \ddots & & \\ 1 & 1 & \dots & 1 & 0 \\ 1 & 1 & \dots & 1 & 1 \end{pmatrix}. \tag{4}$$

Remark 2.1. While application of $\boldsymbol{D}_n^m$ is a discrete version of m times differentiation, $\boldsymbol{A}_n^m$ realizes m–fold integration, i.e., $\boldsymbol{A}_n^m \boldsymbol{f}$ is a discrete version of the m-th antiderivative of $\boldsymbol{f}$. For example, the components of $\boldsymbol{A}_n^m \boldsymbol{f}$ are given for $m = 1, 2$ by

$m = 1$	$m = 2$
$f(1)$	$f(1)$
$f(1) + f(2)$	$2f(1) + f(2)$
$f(1) + f(2) + f(3)$	$3f(1) + 2f(2) + f(3)$
$\vdots$	$\vdots$
$f(1) + f(2) + \ldots + f(n)$	$nf(1) + (n-1)f(2) + \ldots + f(n)$

and may be considered as discrete version of $A^1 f(x) = \int_0^x f(t)\,\mathrm{d}t$ and $A^2 f(x) = \int_0^x \int_0^{t_1} f(t)\,\mathrm{d}t\,\mathrm{d}t_1$, respectively. For general m, the j–th component of $\boldsymbol{A}_n^m \boldsymbol{f}$ is $\sum_{k=1}^{j} \frac{(j+1-k)^{(m-1)}}{(m-1)!} f(k)$. Here $k^{(m)} := 1$ for $m = 0$ and $k^{(m)} := k(k+1)\dots(k+m-1)$ for $m \geq 1$ is a discrete equivalent of the m–th power function.

Let $\mathbf{0}_{n,m}$ denote the matrix consisting of $n \times m$ zeros, $\mathbf{1}_{n,m}$ the matrix consisting of $n \times m$ ones and $\boldsymbol{I}_n$ the $n \times n$ identity matrix. Then the m–th forward difference (2) can be realized by applying the m–th *forward difference matrix*

$$\boldsymbol{D}_{n,m} := \left(\mathbf{0}_{n-m,m} | \boldsymbol{I}_{n-m}\right)\, \boldsymbol{D}_n^m$$

and our minimization problem (3) can be rewritten as

$$\frac{1}{2}\|\boldsymbol{f} - \boldsymbol{u}\|_2^2 + \lambda \|\boldsymbol{D}_{n,m}\boldsymbol{u}\|_1 \to \min. \tag{5}$$

The functional in (5) is strictly convex and has therefore a unique minimizer. The matrix $\boldsymbol{D}_{n,m}$ has full rank $n-m$, i.e., $\mathcal{R}(\boldsymbol{D}_{n,m}) = \mathbb{R}^{n-m}$. Moreover, the range $\mathcal{R}(\boldsymbol{D}_{n,m}^{\mathrm{T}})$ of $\boldsymbol{D}_{n,m}^{\mathrm{T}}$ and the kernel $\mathcal{N}(\boldsymbol{D}_{n,m})$ of $\boldsymbol{D}_{n,m}$ are given by

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

see, e.g., Didas (2004). The space Π_m collects just the discrete polynomials of degree $\leq m$. Then we have the orthogonal decomposition

$$\boldsymbol{R}^n = \mathcal{R}(\boldsymbol{D}_{n,m}^{\mathrm{T}}) \oplus \mathcal{N}(\boldsymbol{D}_{n,m}). \tag{6}$$

Obviously, $\boldsymbol{D}_{n,m}$ is given by cutting of the first m rows of $\boldsymbol{D}_n^m$. The following relations between $\boldsymbol{D}_n^m$ and $\boldsymbol{D}_{n,m}$ are proved in the appendix.

Proposition 2.2. *The difference matrices fulfill the properties*

i) $\boldsymbol{D}_{n,m}^{T} = (-1)^m \boldsymbol{D}_n^m \begin{pmatrix} \boldsymbol{I}_{n-m} \\ \mathbf{0}_{m,n-m} \end{pmatrix}$,

ii) $\boldsymbol{D}_{n,m}\boldsymbol{D}_n^m = \boldsymbol{D}_{n+m,2m} \begin{pmatrix} \mathbf{0}_{m,n} \\ \boldsymbol{I}_n \end{pmatrix}$,

iii) $\boldsymbol{D}_{n+m,m} \begin{pmatrix} \mathbf{0}_{m,n} \\ \boldsymbol{I}_n \end{pmatrix} = \boldsymbol{D}_n^m$.

Proof:

i) Since $\boldsymbol{D}_{n,m}\boldsymbol{f} = (\Delta^m f(1), \dots, \Delta^m f(n-m))^{\mathrm{T}}$ we can rewrite $\boldsymbol{D}_{n,m}$ as

$$\begin{aligned}
\boldsymbol{D}_{m,n} &= \boldsymbol{D}_{n-(m-1),1} \cdot \ \dots \ \cdot\\
\boldsymbol{D}_{n,1} &= (\mathbf{0}_{n-m,1} | \boldsymbol{I}_{n-m})\\
&\qquad \times \boldsymbol{D}_{n-(m-1)} \cdot \ \dots \ \cdot (\mathbf{0}_{n-1,1} | \boldsymbol{I}_{n-1})\boldsymbol{D}_n
\end{aligned}$$

Using that by definition

$$\boldsymbol{D}_{n,1}^{\mathrm{T}} = \boldsymbol{D}_n^{\mathrm{T}} \begin{pmatrix} \mathbf{0}_{1,n-1} \\ \boldsymbol{I}_{n-1} \end{pmatrix} = -\boldsymbol{D}_n \begin{pmatrix} \boldsymbol{I}_{n-1} \\ \mathbf{0}_{1,n-1} \end{pmatrix}$$

we obtain for the transposed matrix

$$\begin{aligned}
\boldsymbol{D}_{n,m}^{\mathrm{T}} &= \boldsymbol{D}_{n,1}^{\mathrm{T}} \cdot \ \dots \ \cdot \boldsymbol{D}_{n-(m-1),1}^{T}\\
&= (-1)^m \boldsymbol{D}_n \begin{pmatrix} \boldsymbol{I}_{n-1} \\ \boldsymbol{O}_{1,n-1} \end{pmatrix} \cdot \ \dots \ \cdot\\
&\qquad \boldsymbol{D}_{n-(m-1),1} \begin{pmatrix} \boldsymbol{I}_{n-m} \\ \mathbf{0}_{1,n-m} \end{pmatrix}.
\end{aligned}$$

Multiplication of $\boldsymbol{f}^T$ from the left is again successive application of first order differences. Equivalently we can apply m–th order finite differences and cut off all additional components which results in assertion i).

ii) By definition of $\boldsymbol{D}_{n,m}$ we have

$$\begin{aligned}
\boldsymbol{D}_{n+m,2m}\begin{pmatrix} \mathbf{0}_{m,n} \\ \boldsymbol{I}_n \end{pmatrix} &= \left(\mathbf{0}_{n-m,2m} | \boldsymbol{I}_{n-m}\right) \boldsymbol{D}_{n+m}^{2m} \begin{pmatrix} \mathbf{0}_{m,n} \\ \boldsymbol{I}_n \end{pmatrix}\\
&= \left(\mathbf{0}_{n-m,m} | \boldsymbol{I}_{n-m}\right)\, \left(\mathbf{0}_{n,m} | \boldsymbol{I}_n\right)\\
&\qquad \times \boldsymbol{D}_{n+m}^{2m} \begin{pmatrix} \mathbf{0}_{m,n} \\ \boldsymbol{I}_n \end{pmatrix}.
\end{aligned}$$

Since the cutoff of the first m rows and columns of a Toeplitz matrix results in the same Toeplitz matrix but with m times reduced order the last equation can be rewritten as

$$\boldsymbol{D}_{n+m,2m} \begin{pmatrix} \mathbf{0}_{m,n} \\ \boldsymbol{I}_n \end{pmatrix} = \left(\mathbf{0}_{n-m,m} | \boldsymbol{I}_{n-m}\right)\ \boldsymbol{D}_n^{2m}$$

and finally, by applying again the definition of $\boldsymbol{D}_{n,m}$ as

$$\boldsymbol{D}_{n+m,2m}\begin{pmatrix}\boldsymbol{0}_{m,n}\\ \boldsymbol{I}_n\end{pmatrix} = \boldsymbol{D}_{n,m}\boldsymbol{D}_n^m.$$

iii) Using the definition of $\boldsymbol{D}_{n,m}$, we obtain

$$\boldsymbol{D}_{n+m,m}\begin{pmatrix}\boldsymbol{0}_{m,n}\\ \boldsymbol{I}_n\end{pmatrix} = \left(\boldsymbol{0}_{n,m}|\boldsymbol{I}_n\right)\ \boldsymbol{D}_{m+n}^m\begin{pmatrix}\boldsymbol{0}_{m,n}\\ \boldsymbol{I}_n\end{pmatrix} = \boldsymbol{D}_n^m.$$

This completes the proof.

□

3. Spline Contact Problem

In this section, we will see that our higher order TV problem (5) is equivalent to a discrete spline interpolation problem, where the spline knots are not known in advance but depend on the input data $\boldsymbol{f}$ and λ. For $m = 1$, the resulting spline contact problem is well examined and can be solved by the so–called 'taut string algorithm', see, e.g., Hinterberger et al. (2003).

A necessary and sufficient condition for $\boldsymbol{u}$ to be the minimizer of (5) is that the zero vector is an element of the functional's subdifferential

$$\boldsymbol{0}_{n,1} \in \boldsymbol{u} - \boldsymbol{f} + \lambda\,\partial\|\boldsymbol{D}_{n,m}\boldsymbol{u}\|_1 .$$

By (Rockafellar, 1970, Theorem 23.9) and since the subgradient of $|x|$ is given by

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

this can be rewritten as

$$\boldsymbol{u} \in \boldsymbol{f} - \lambda\,\boldsymbol{D}_{n,m}^{\mathrm{T}}\frac{\boldsymbol{D}_{n,m}\boldsymbol{u}}{|\boldsymbol{D}_{n,m}\boldsymbol{u}|},$$

where $\cdot/|\cdot|$ is taken componentwise. These inclusions in their present form are not very convenient for the computation of $\boldsymbol{u}$. However, multiplying with $\boldsymbol{A}_n^m$ and applying Proposition 2.2i) leads to

$$\boldsymbol{A}_n^m\boldsymbol{u} \in \boldsymbol{A}_n^m\boldsymbol{f} - (-1)^m\lambda\begin{pmatrix}\boldsymbol{I}_{n-m}\\ \boldsymbol{0}_{m,n-m}\end{pmatrix}\frac{\boldsymbol{D}_{n,m}\boldsymbol{u}}{|\boldsymbol{D}_{n,m}\boldsymbol{u}|}.$$

Setting

$$\begin{pmatrix}\boldsymbol{F}_I\\ \boldsymbol{F}_R\end{pmatrix} := \boldsymbol{A}_n^m\boldsymbol{f},\quad \begin{pmatrix}\boldsymbol{U}_I\\ \boldsymbol{U}_R\end{pmatrix} := \boldsymbol{A}_n^m\boldsymbol{u} \tag{7}$$

with the splitting into the inner vector $\boldsymbol{F}_I \in \mathbb{R}^{n-m}$ and the right boundary vector $\boldsymbol{F}_R \in \mathbb{R}^m$, the inclusions can be rewritten as

$$\boldsymbol{U}_I \in \boldsymbol{F}_I - (-1)^m\lambda\frac{\boldsymbol{D}_{n,m}\boldsymbol{u}}{|\boldsymbol{D}_{n,m}\boldsymbol{u}|},\qquad \boldsymbol{U}_R = \boldsymbol{F}_R.$$

It remains to replace $\boldsymbol{D}_{n,m}\boldsymbol{u}$. By (7) and (4), we see that

$$\boldsymbol{f} = \boldsymbol{D}_n^m\begin{pmatrix}\boldsymbol{F}_I\\ \boldsymbol{F}_R\end{pmatrix},\quad \boldsymbol{u} = \boldsymbol{D}_n^m\begin{pmatrix}\boldsymbol{U}_I\\ \boldsymbol{U}_R\end{pmatrix} \tag{8}$$

and further by Proposition 2.2ii) that

$$\boldsymbol{D}_{n,m}\boldsymbol{u} = \boldsymbol{D}_{n+m,2m}\begin{pmatrix}\boldsymbol{0}_{m,n}\\ \boldsymbol{I}_n\end{pmatrix}\begin{pmatrix}\boldsymbol{U}_I\\ \boldsymbol{U}_R\end{pmatrix}.$$

Introducing an artificial left boundary $\boldsymbol{U}_L := \boldsymbol{0}_{m,1}$ and extending our vector by

$$\boldsymbol{U} := (\boldsymbol{U}_L^{\mathrm{T}}, \boldsymbol{U}_I^{\mathrm{T}}, \boldsymbol{U}_R^{\mathrm{T}})^{\mathrm{T}}$$

our inclusions become finally

$$\boldsymbol{U}_I \in \boldsymbol{F}_I - (-1)^m\lambda\frac{\boldsymbol{D}_{n+m,2m}\boldsymbol{U}}{|\boldsymbol{D}_{n+m,2m}\boldsymbol{U}|},\qquad \boldsymbol{U}_R = \boldsymbol{F}_R.$$

Consequently, $\boldsymbol{U}$ is the unique solution of the following spline contact problem, where we have to explain the spline notation later.

Spline Contact Problem

(C1) Boundary conditions:
$\boldsymbol{U}_L = \boldsymbol{0}_{m,1}$ and $\boldsymbol{U}_R = \boldsymbol{F}_R$.

(C2) Tube condition: $\|\boldsymbol{F}_I - \boldsymbol{U}_I\|_\infty \le \lambda$
$\boldsymbol{U}_I$ lies in a tube around $\boldsymbol{F}_I$ of width 2λ.

(C3) Contact condition:
Let $\Lambda_I := \{j \in \{m+1,\dots,n-m\} : \triangle^{2m}U(j-m) \neq 0\}$.
If $j \in \Lambda_I$, then $U(j)$ contacts the boundary of the tube, where
$(-1)^m\triangle^{2m}U(j-m) > 0 \Longrightarrow$
$U(j) = F(j) - \lambda$ (lower contact),
$(-1)^m\triangle^{2m}U(j-m) < 0 \Longrightarrow$
$U(j) = F(j) + \lambda$ (upper contact).

Remark 3.1. (Continuous and Discrete Natural Splines)
We recall that a natural polynomial spline of degree $2m-1$ with knots $x_1 < \cdots < x_r$ is a function $s \in C^{2m-2}$ such that

$$s^{(2m)}(x) = 0, \quad \text{for } x \in (x_j, x_{j+1}),\ j = 1, \ldots, r-1,$$
$$s^{(m)}(x) = 0, \quad \text{for } x < x_1,\ x > x_r.$$

These splines are the solutions in W^m, the Sobolev space of $(m-1)$ times continuousely differentiable functions with m-th weak derivative in L_2, of

$$\begin{aligned} &\frac{1}{2}\,\|f^{(m)}\|_2^2 \quad \to \quad \min \\ &\text{s.t.} \quad f(x_j) = \gamma_j, \quad j = 1, \ldots, r. \end{aligned}$$

Mangasarian and Schumaker (1971, 1973) have introduced the discrete natural polynomial spline of degree $2m-1$ with knots $\Xi = \{i_1, \ldots, i_r\}$, $i_j < i_k$ for $j < k$, as a vector $\boldsymbol{s} = (s(1), \ldots, s(N))^{\mathrm{T}}$ which satisfies for $j \notin \Xi$ the relations

$$\begin{aligned} \triangle^{2m} s(j-m) &= 0, \quad j = m+1, \ldots, N-m; \\ \triangle^m s(j) &= 0, \quad j = 1, \ldots, i_1 - 1; \\ &\qquad\quad\ i_r + 1, \ldots, N-m. \end{aligned}$$

As its continuous analogue the discrete natural polynomial spline of degree $2m-1$ solves the minimization problem

$$\begin{aligned} &\frac{1}{2}\sum_{j=1}^{N-m} (\triangle^m y(j))^2 \quad \to \quad \min \\ &\text{s.t.} \quad y(i_j) = \gamma_j, \quad j = 1, \ldots, r. \end{aligned} \tag{9}$$

For relations between continuous and natural spline in the limiting process $N \to \infty$ see also Mangasarian and Schumaker (1971, 1973).

Setting $N := n + m$ and using the spline knots $\Xi = \{1, \ldots, m\} \cup \Lambda_I \cup \{n-m+1, \ldots, n\}$, we can interpret $\boldsymbol{U}$ defined by (C1) – (C3) is a discrete natural polynomial spline of degree $2m-1$. In contrast to (9), the inner spline knots Λ_I are only determined by (C3) and not known in advance. This reflects the nonlinear character of our problem solution.

We extend the discrete spline concept to splines of even degree as follows: we call $\boldsymbol{s} = (s(1), \ldots, s(n))^{\mathrm{T}}$ a *discrete spline of degree* $m-1$ *with inner knots* $\Xi = \{i_1, \ldots, i_r\} \subseteq \{\lfloor \frac{m}{2} \rfloor + 1, \ldots, n - \lfloor \frac{m+1}{2} \rfloor\}$ if

$$\triangle^m s\left(j - \left\lfloor \frac{m}{2} \right\rfloor\right) = 0, \quad j = \lfloor \frac{m}{2} \rfloor + 1, \ldots, \; n - \left\lfloor \frac{m+1}{2} \right\rfloor;\ j \notin \Xi.$$

Then the discrete interpolation problem

$$s(i_j) = \gamma_j,\ i_j \in \Xi \cup \left\{1, \ldots, \left\lfloor \frac{m}{2} \right\rfloor\right\} \cup \left\{n - \left\lfloor \frac{m+1}{2} \right\rfloor + 1, \ldots, n\right\}$$

has a unique solution. Thus, for given spline knots Λ_I, we could solve a spline interpolation problem. Unfortunately, the spline knots depend on the input data $\boldsymbol{f}$ and λ. Therefore, the solution of the spline contact problem in its present form is only convenient for $m = 1$, see Remark 3.2. For larger m and the continuous setting, an attempt to solve the contact problem is contained in Mammen and van de Geer (1997). For our discrete setting, we will see in the following section that the contact problem can be treated by simply solving a constraint quadratic minimization problem.

Remark 3.2. (Taut String Algorithm for $m = 1$)
For $m = 1$, condition (C3) means that the polygon through $\boldsymbol{U}$ is convex at upper contact points and concave at lower contact points. Thus, the construction of $\boldsymbol{U}$ satisfying (C1) – (C3) is equivalent to the construction of the uniquely determined taut string within the tube around F of width 2λ fixed at $(0, 0)$ and $(n, F(n))$. In other words, the polygon through $\boldsymbol{U}$ has minimal lengths within the tube, i.e., it minimizes

$$\sum_{j=0}^{n-1} (1 + (U(j+1) - U(j))^2)^{1/2},$$

subject to the tube and boundary conditions. An example of a taut string is shown in Figure 1. For solving this problem there exists a very efficient algorithm of complexity $\mathcal{O}(n)$, the so–called 'taut string algorithm', which is based on a convex hull algorithm, see, e.g., Davies and Kovac (2001); Mammen and van de Geer (1997).

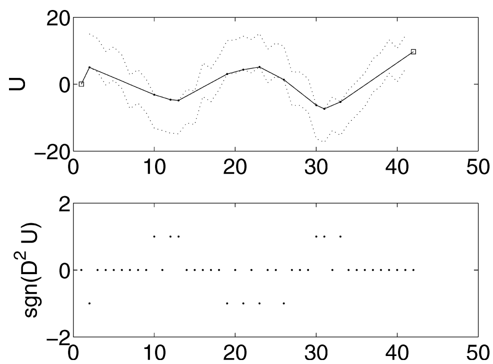

Figure 1. Solution of the spline contact problem (C1) – (C3) for a signal $\boldsymbol{F}$ of lengths $n+m$ with $n=40$ and $m=1$.

Interestingly, it was shown in Steidl et al. (2004); Yip and Park (2003) that for $m=1$ the spline knots fulfill a so–called '*tree–property*'.

Remark 3.3. (Tree Property of Spline Knots for $m=1$)
Let $\lambda_{\max}$ be the smallest regularization parameter such that $\Lambda_I = \emptyset$. It is not hard to show that $\lambda_{\max} = \|P\boldsymbol{f}\|_{W_1(\boldsymbol{D}_{n,1})'}$, where P denotes the orthogonal projection of $\boldsymbol{f}$ onto $\mathcal{R}(\boldsymbol{D}_{n,1}^{\mathrm{T}})$ and $W_1(\boldsymbol{D}_{n,1})'$ is the dual space of $W_1(\boldsymbol{D}_{n,1}) := \mathcal{R}(\boldsymbol{D}_{n,1}^{\mathrm{T}})$ equipped with the norm $\|\boldsymbol{u}\|_{W_1(\boldsymbol{D}_{n,1})} := \|\boldsymbol{D}_{n,1}\boldsymbol{u}\|_1$.

If λ moves from $\lambda_{\max}$ to 0 and $\Lambda_I(\lambda)$ denotes the corresponding set of inner spline knots, then, for $\lambda_j > \lambda_k$,

$$\begin{aligned}\emptyset = \Lambda_I(\lambda_{\max}) &\subseteq \Lambda_I(\lambda_j) \subseteq \Lambda_I(\lambda_k) \subseteq \Lambda_I(0)\\ &= \{m+1,\dots,n-m\}.\end{aligned}$$

Figure 2 shows a tree of inner spline knots. The tree property does not hold for $m \geq 2$.

4. Support Vector Regression with Spline Kernels

In this section we want to show the relation of the discrete spline contact problem with discrete SVR. We start by a brief introduction to SVR in the continuous setting, where we emphasize the role of splines in the solution of the SVR problem in Sobolev spaces. Then we switch to the discrete context to explain the solution of (5) from the SVR point of view.

4.1. Support Vector Regression - Continuous Approach

The SVR method searches for approximations of functions in reproducing kernel Hilbert spaces (RKHS) and plays an important role, e.g., in Learning Theory. Among the large amount of literature on SVR we refer to (Vapnik, 1998, Chapter 11). SVR can be briefly explained as follows: Let $H \subset L_2(\mathbb{R}^d)$ be a Hilbert space with inner product $(\cdot,\cdot)_H$ having the property that the point evaluation functional is continuous. Then H possesses a so–called reproducing kernel $K \in L_2(\mathbb{R}^d \times \mathbb{R}^d)$ with reproducing property $(F, K(\cdot, x_j))_H = F(x_j)$ for all $F \in H$ and is called a *reproducing kernel Hilbert space* (RKHS). Given some function values $F(x_j)$, $j=1,\dots,p$, the *soft margin SVR* problem consists in finding a function $U \in H$ which minimizes

$$\mu \sum_{j=1}^{p} V_\lambda(F(x_j) - U(x_j)) + \frac{1}{2}\|U\|_H^2,$$

where $V_\lambda(x) := \max\{0, |x| - \lambda\}$ denotes Vapnik's λ-insensitive loss function. In other words, Vapnik's cost functional penalizes those $U(x_j)$ lying not in a λ neighbourhood of $F(x_j)$. If μ tends to infinity, then our cost functional must become zero and we obtain the *hard margin SVR* problem

$$\begin{aligned}&\frac{1}{2}\|U\|_H^2 \quad \to \quad \min \\ &\text{s.t.}\ \ |F(x_j) - U(x_j)|_\infty \leq \lambda, \quad j=1,\dots,p.\end{aligned} \tag{10}$$

By the Representer Theorem of Kimmeldorf and Wahba (1971), the solution of (10) has the form

$$U(x) = \sum_{k=1}^{p} c(k) K(x_k, x),$$

i.e., only the given knots x_k are involved into the representation. Then (10) can be rewritten as

$$\begin{aligned}&\frac{1}{2}\boldsymbol{c}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{c} \quad \to \quad \min \\ &\text{s.t.}\ \ \|\boldsymbol{F} - \boldsymbol{K}\boldsymbol{c}\|_\infty \leq \lambda\end{aligned} \tag{11}$$

with $\boldsymbol{F} := (F(x_j))_{j=1}^p$, $\boldsymbol{c} := (c(k))_{k=1}^p$ and $\boldsymbol{K} := \big(K(x_j, x_k)\big)_{j,k=1}^p$. This is the usual hard margin SVR formulation.

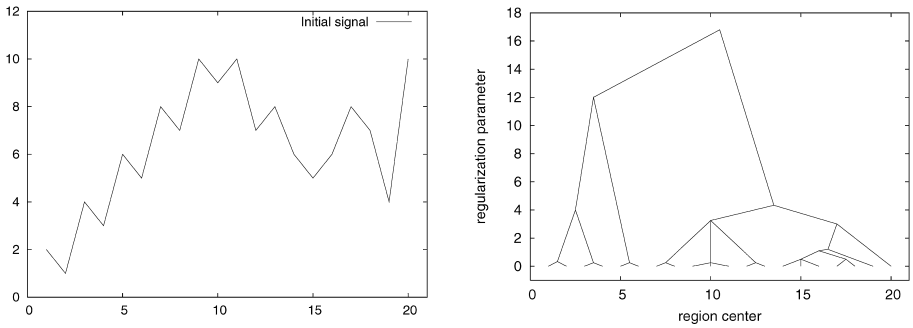

Figure 2. Original signal $\boldsymbol{f}$ (left), tree of spline knots with increasing regularization parameter λ from leaves to root (right).

Based on the Karush – Kuhn – Tucker conditions it follows that $c(k) \neq 0$ implies $|F(x_k) - U(x_k)| = \lambda$. Let

$$\Lambda := \{k \in \{1, \dots, p\} : c(k) \neq 0\}.$$

Then the solution U can be rewritten as

$$U(x) = \sum_{k \in \Lambda} c(k) K(x_k, x). \tag{12}$$

The functions $K(x_k, x)$ with $k \in \Lambda$ are called *support vectors*. Obviousely, U depends only on these support vectors and has a sparse representation in terms of the support vectors if $|\Lambda|$ is small compared to p. In the image processing context, SVR regression is mainly applied in high dimensional function spaces ($d \gg 1$), where often the Gaussian is involved as reproducing kernel.

For our purposes we will consider other well–known reproducing kernel Hilbert spaces, namely the Sobolev spaces $H = W_{2,0}^m$ of real–valued functions on $\mathbb{R}$ having a weak m–th derivative in $L_2\,[0, 1]$ and fulfilling $F^{(r)}(0) = 0$ for $r = 0, \dots, m-1$ with inner product

$$\langle F, G \rangle_{W_{2,0}^m} := \int_0^1 F^{(m)}(x) G^{(m)}(x)\,\mathrm{d}x.$$

These RKHS were for example considered in (Wahba, 1990, p. 5–14). The reproducing kernel in $W_{2,0}^m$ is

$$K(x, y) := \int_0^1 (x - t)_+^{m-1}\,(y - t)_+^{m-1}/((m-1)!)^2\,\mathrm{d}t, \tag{13}$$

where $(x)_+ := \max\{0, x\}$. For fixed y, the functions $K(\cdot, y)$ are splines fulfilling $K(\cdot, y) \in C^{2m-2}$, $K(\cdot, y) \in \Pi_{2m-1}$ in $[0, y]$ and $K(\cdot, y) \in \Pi_{m-1}$ in $[y, 1]$.

In this context we mention that another minimization problem having so–called smoothing splines as solutions was considered the literature, see, e.g., Wahba (1990); Unser and Blu (2000): find $U \in W_{2,0}^m$ such that

$$\frac{1}{2} \sum_{j=1}^{p} (F(x_j) - U(x_j))^2 + \lambda \|U\|_{W_{2,0}^m}^2 \quad \to \quad \min.$$

Again by the Representer Theorem, this problem has a solution of the form $U = \sum_{k=1}^{p} c(k)\,K(\cdot, x_k)$. Consequently, U is a continuous spline of degree $2m-1$ with knots x_k, $k = 1, \dots, p$. However, in contrast to the solution (12) of (10), all coefficients $c(k)$ are in general $\neq 0$ and we obtain no sparse representation.

4.2. Support Vector Regression – Discrete Approach

To see the relation between our spline contact problem and SVR methods, we consider the dual formulation of problem (5).

Proposition 4.1. *The solution* $\boldsymbol{u}$ *of* (5) *is given by* $\boldsymbol{u} = \boldsymbol{f} - \boldsymbol{D}_{n,m}^{\mathrm{T}} \boldsymbol{V}_I$*, where* $\boldsymbol{V}_I$ *is the unique solution of the minimization problem*

$$\begin{aligned} &\frac{1}{2}\,\|\boldsymbol{f} - \boldsymbol{D}_{n,m}^{\mathrm{T}} \boldsymbol{V}_I\|_2^2 \quad \to \quad \min \\ \text{s.t.} \quad &\|\boldsymbol{V}_I\|_\infty \leq \lambda. \end{aligned} \tag{14}$$

For a proof see, e.g., Steidl (2006).

By (8) and Proposition 2.2 i) and iii) we obtain that

$$\begin{aligned} \|\boldsymbol{f} - \boldsymbol{D}_{n,m}^{\mathrm{T}} \boldsymbol{V}_I\|_2 &= \|\boldsymbol{D}_n^m \begin{pmatrix} \boldsymbol{F}_I \\ \boldsymbol{F}_R \end{pmatrix} \\ &\qquad - (-1)^m \boldsymbol{D}_n^m \begin{pmatrix} \boldsymbol{I}_{n-m} \\ \boldsymbol{0}_{m,n-m} \end{pmatrix} \boldsymbol{V}_I\|_2 \\ &= \|\boldsymbol{D}_{n+m,m}(\boldsymbol{F} - (-1)^m \boldsymbol{V})\|_2, \end{aligned}$$

where $\boldsymbol{V} := (\boldsymbol{0}_{m,1}^{\mathrm{T}}, \boldsymbol{V}_I^{\mathrm{T}}, \boldsymbol{0}_{m,1}^{\mathrm{T}})^{\mathrm{T}}$. Setting $\boldsymbol{U} := \boldsymbol{F} - (-1)^m \boldsymbol{V}$, problem (14) can be rewritten as

$$\begin{aligned} &\frac{1}{2}\|\boldsymbol{D}_{n+m,m}\boldsymbol{U}\|_2^2 \quad \to \quad \min \qquad (15)\\ \text{s.t.}\quad &\|\boldsymbol{F}_I - \boldsymbol{U}_I\|_\infty \le \lambda, \quad \boldsymbol{U}_R = \boldsymbol{F}_R. \end{aligned}$$

The unique solution $\boldsymbol{U}$ of this problem which can be computed by standard quadratic programming (QP) methods is also the unique solution of our spline contact problem. Figure 3 illustrates the solution for $m = 3$.

Remark 4.2. Regarding Remark 3.2, we see that for $m = 1$ the minimization problems

$$\sum_{j=1}^{n} \left(1 + (U(j+1) - U(j))^2\right)^{1/2} \quad \to \quad \min,$$

and

$$\|\boldsymbol{D}_{n+1,1}\boldsymbol{U}\|_2^2 = \sum_{j=1}^{n} (U(j+1) - U(j))^2 \quad \to \quad \min$$

subject to the tube and boundary constraints lead to the same solution.

We will see that problem (15) can be considered as a *hard margin SVR problem*. To this end, we only have to define the appropriate RKHS. Let $\mathcal{W}_{2,0}^m := \{\boldsymbol{F} \in \mathbb{R}^{n+m} : F(j) = 0,\ j = 1, \ldots, m\}$ equipped with the inner product

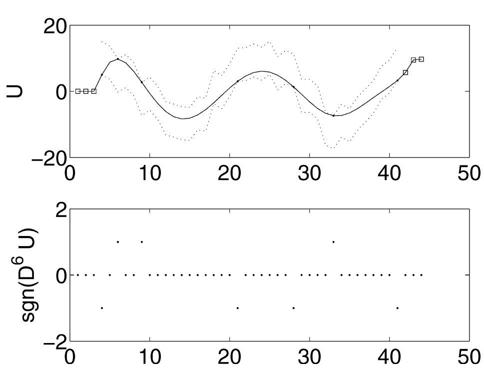

Figure 3. Solution of the spline contact problem (C1) – (C3) for a signal $\boldsymbol{F}$ of lengths $n + m$ with $n = 40$ and $m = 3$.

$$\begin{aligned} \langle \boldsymbol{F}, \boldsymbol{G}\rangle_{\mathcal{W}_{2,0}^m} &:= \sum_{j=1}^{n} \triangle^m F(j) \triangle^m G(j)\\ &:= \langle \boldsymbol{D}_{n+m,m}\boldsymbol{F}, \boldsymbol{D}_{n+m,m}\boldsymbol{G}\rangle\\ &= \left\langle \boldsymbol{D}_n^m \begin{pmatrix} \boldsymbol{F}_I\\ \boldsymbol{F}_R \end{pmatrix}, \boldsymbol{D}_n^m \begin{pmatrix} \boldsymbol{G}_I\\ \boldsymbol{G}_R \end{pmatrix} \right\rangle. \end{aligned}$$

Then the minimization term in (15) is just the norm of $\boldsymbol{U}$ in $\mathcal{W}_{2,0}^m$. Now we can straightforwardly determine the reproducing kernel in $\mathcal{W}_{2,0}^m$. Setting

$$\boldsymbol{K} := \left((\boldsymbol{D}_n^m)^{\mathrm{T}} \boldsymbol{D}_n^m\right)^{-1} = \boldsymbol{A}_n^m (\boldsymbol{A}_n^m)^{\mathrm{T}}, \qquad (16)$$

we see that the columns $\boldsymbol{K}_{0,k}$ of

$$\boldsymbol{K}_0 := \left(\boldsymbol{0}_{n,m} | \boldsymbol{K}\right)^{\mathrm{T}} \in \mathbb{R}^{n+m,n}$$

form a special basis of $\mathcal{W}_{2,0}^m$, namely with reproducing property $\langle \boldsymbol{F}, \boldsymbol{K}_{0,j}\rangle_{\mathcal{W}_{2,0}^m} = F(j+m)$. Let us have a closer look at the structure of $\boldsymbol{K}$. Straightforward computation shows that the components of our discrete kernel are given by the discrete counterpart of (13), namely

$$K(j,k) = \sum_{r=0}^{\min(j,k)-1} (j-r)^{(m-1)}\,(k-r)^{(m-1)}/((m-1)!)^2,$$

with (m) defined as in Remark 2. By Proposition 2 ii) and i) we obtain that

$$\begin{aligned} \boldsymbol{D}_{n+m,2m}\boldsymbol{K}_0 &= \boldsymbol{D}_{n+m,2m} \begin{pmatrix} \boldsymbol{0}_{m,n}\\ \boldsymbol{I}_n \end{pmatrix} \boldsymbol{A}_n^m (\boldsymbol{A}_n^m)^{\mathrm{T}}\\ &= \boldsymbol{D}_{n,m}\boldsymbol{D}_n^m \boldsymbol{A}_n^m (\boldsymbol{A}_n^m)^{\mathrm{T}}\\ &= (-1)^m \left(\boldsymbol{I}_{n-m}, \boldsymbol{0}_{n-m,m}\right). \end{aligned}$$

In other words, we have for $j = m+1, \ldots, n-m$ that

$$\begin{aligned} &\triangle^{2m} K_{0,k}(j-m) = 0, \quad k = 1, \ldots, n-m;\ j \ne k,\\ &\triangle^{2m} K_{0,k}(k-m) = (-1)^m,\ k = 1, \ldots, n-m,\\ &\triangle^{2m} K_{0,k}(j-m) = 0, \quad k = n-m+1, \ldots, n, \end{aligned} \qquad (17)$$

i.e., $\boldsymbol{K}_{0,k}$ is a discrete spline of degree $2m-1$ with one inner knot $k+m$ for $k = 1, \ldots, n-m$ and a discrete polynomial in Π_{2m-1} for $k = n-m+1, \ldots, n$. For $n = 32$ and $m = 1, 2$, some columns of $\boldsymbol{K}_0$ are depicted in Figure 4.

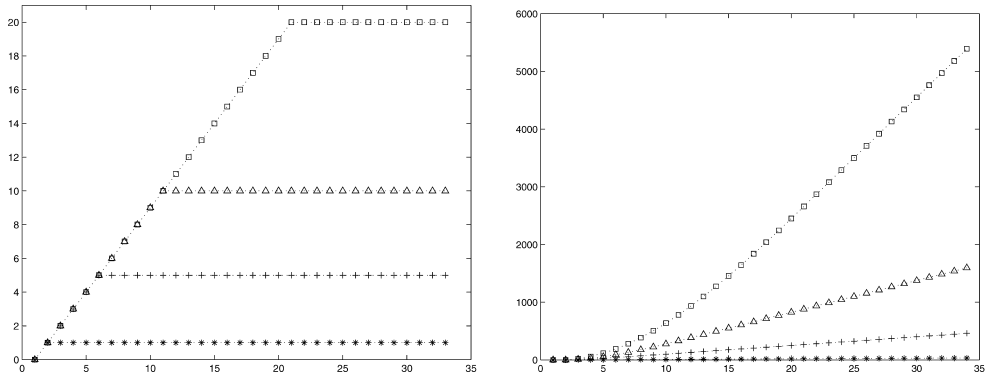

Figure 4. Discrete splines $\boldsymbol{K}_{0,k}$, $k = 1, 5, 10, 20$, for $n = 32$ and $m = 1$ (left), $m = 2$ (right).

For every $\boldsymbol{U} \in \mathcal{W}_{2,0}^m$, there exists a uniquely determined $\boldsymbol{c} \in \mathbb{R}^n$ such that $\boldsymbol{U} = \boldsymbol{K}_0\boldsymbol{c}$ and by the reproducing property of $\boldsymbol{K}_0$, problem (15) can be rewritten as

$$\frac{1}{2}\boldsymbol{c}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{c} \quad \to \quad \min \tag{18}$$
$$\text{s.t.} \quad \|\boldsymbol{F}_I - (\boldsymbol{K}\boldsymbol{c})_I\|_\infty \leq \lambda, \quad (\boldsymbol{K}\boldsymbol{c})_R = \boldsymbol{F}_R.$$

This is the usual form (11) of a hard margin SVR problem. Let $\boldsymbol{c}$ be the solution of (18) and let

$$\tilde{\Lambda}_I := \{j \in \{m+1, \ldots, n\} : c(j-m) \neq 0\}$$

so that

$$\boldsymbol{U} = \sum_{j \in \tilde{\Lambda}_I} c(j-m)\boldsymbol{K}_{0,j-m} + \sum_{j=n-m+1}^{n} c(j)\boldsymbol{K}_{0,j}. \tag{19}$$

The vectors $\boldsymbol{K}_{0,j-m}$, $j \in \tilde{\Lambda}_I$ are called (inner) *support vectors*. By (19) and property (17) of $\boldsymbol{K}_0$ they are related to the spline knots as follows:

Proposition 4.3. *The support vector indices $\tilde{\Lambda}_I$ of the solution $\boldsymbol{U}$ in* (19) *of the SVR problem are exactly the spline knots Λ_I, i.e.,*

$$\triangle^{2m} U(j-m) \neq 0 \quad \Longleftrightarrow \quad j \in \tilde{\Lambda}_I.$$

If the number of contact points $|\Lambda_I|$ is small compared to n, then $\boldsymbol{c}$ has only a small number of nonzero coefficients and (19) provides us with a *sparse representation* of $\boldsymbol{U}$. This can also be seen by noting that our SVR problem (18) means to find $\boldsymbol{U} = \boldsymbol{K}_0\boldsymbol{c}$ such that the equality constraints are fulfilled and

$$\frac{1}{2}\|\boldsymbol{F} - \boldsymbol{U}\|^2_{\mathcal{W}_{2,0}^m} + \lambda\|\boldsymbol{c}\|_1 \quad \to \min.$$

Compare with Girosi (1998) in a general SVR context. In contrast to the 2–norm, the 1–norm of $\boldsymbol{c}$ in the penalty term implies for sufficiently large λ that some of the coefficients $c(j)$ are 0. This implies a sparse representation of $\boldsymbol{U}$ from another point of view.

Finally, we see by (16) and (8) that

$$\boldsymbol{u} = \boldsymbol{D}_n^m \boldsymbol{A}_n^m (\boldsymbol{A}_n^m)^{\mathrm{T}}\boldsymbol{c} = (\boldsymbol{A}_n^m)^{\mathrm{T}}\boldsymbol{c} \tag{20}$$

is the corresponding sparse representation of our original solution $\boldsymbol{u}$. By Proposition 2.2i) we have that $\boldsymbol{D}_{m,n}(\boldsymbol{A}_n^m)^{\mathrm{T}} = (-1)^m(\boldsymbol{I}_{n-m}|\boldsymbol{0}_{n-m,m})$ so that the first $n-m$ columns of $(\boldsymbol{A}_n^m)^{\mathrm{T}}$ are splines of degree $m-1$ with one inner knot and the last m columns are polynomials in Π_{m-1}. For $m = 1$ and 2 some columns of $(\boldsymbol{A}_n^m)^{\mathrm{T}}$ are illustrated in Figure 5. In the context of sparse representation, the following observation is interesting: by (20), (8) and Proposition 2.2i) and iii), our original problem (5) can be rewritten as

$$\frac{1}{2}\|\boldsymbol{f} - (\boldsymbol{A}_n^m)^{\mathrm{T}}\boldsymbol{c}\|_2^2 + \lambda\|(\boldsymbol{I}_{n-m}|\boldsymbol{0}_{n-m,m})\boldsymbol{c}\|_1 \quad \to \quad \min. \tag{21}$$

Remark 4.4. Finally, let us mention that a continuous version of our considerations reads as follows: For a function $u := \Phi_u^{(2m)}$ we have that $\Phi_u = k * u$, where k is the causal fundamental solution of the $2m$–th derivative operator, i.e., the spline $k(x) = x_+^{2m-1}$. If $\boldsymbol{u}$ plays the discrete role of u then our discrete

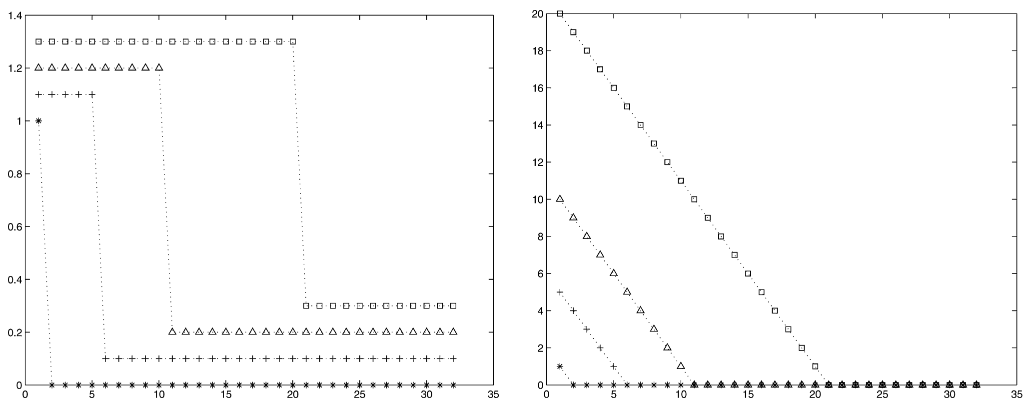

Figure 5. Discrete splines $(A_n^m)_k^{\mathrm{T}}$, $k = 1, 5, 10, 20$, for $n = 32$ and $m = 1$ (left), $m = 2$ (right). For $m = 1$, we have added 0.1, 0.2 and 0.3 to the last columns to better visualize the discrete step functions.

function $(\boldsymbol{U}_I^{\mathrm{T}}, \boldsymbol{U}_R^{\mathrm{T}})^{\mathrm{T}} = \boldsymbol{A}_n^m \boldsymbol{u} = \boldsymbol{K} \boldsymbol{D}_n^m \boldsymbol{u}$ plays the role of $U := \Phi_u^{(m)} = k * u^{(m)}$.

4.3. Denoising Example

In this section, we show the performance of our approach (5) and (15) by a denoising example. We are mainly interested in the behaviour for various differentiation orders m. Our aim is to demonstrate the spline interpolation with variable knots for various m and not to create an optimal denoising method. To this end, we have used the signal shown in Figure 6 (top, left) and have added white Gaussian noise. First, we have determined the optimal parameters λ with respect to the maximal signal–to–noise–ratio (SNR) defined by $\mathrm{SNR}(g, u) := 10 \log_{10} \left(\frac{\|g\|_2^2}{\|g-u\|_2^2} \right)$ with original signal g. For the solution of the quadratic problem (15) we have applied the Matlab quadratic programming routine which is based on an active set method. Then we compared the quality of the results obtained for various m. The following table contains the results for λ, the SNR and the peak signal–to–noise–ratio (PSNR) defined by $\mathrm{PSNR}(g, u) := 10 \log_{10} \left(\frac{n\|g\|_\infty^2}{\|g-u\|_2^2} \right)$, where n denotes the number of pixels. The noisy signal in Figure 6 (top, right) has SNR 6.94 and PSNR 10.72.

m	λ	SNR	PSNR
1	20.2	16.00	19.78
2	57.8	18.41	22.18
3	275.0	17.97	21.69
4	1453.1	17.22	20.99

The corresponding signal plots are given in Figure 6. For this signal the methods with orders $m \geq 2$ perform better than the usual method with $m = 1$ where the the linear method ($m = 2$) achieves the best restoration. In general higher order methods with l_1 regularization term neglect the staircasing effect appearing in the piecewise constant approximation with $m = 1$ and preserve on the other hand local singularities better than linear methods with quadratic regularization term. Various other examples for the denoising of signals by solving (5) were presented in Steidl et al. (2005).

5. Generalization to Two Dimensions

In this section, we briefly consider a possible generalization of our concept to two dimensions. This may be considered as starting point for future research.

Concerning *first order derivatives*, we consider the ROF model

$$\frac{1}{2} \int_\Omega (u(x) - f(x))^2 + \lambda |\nabla u| \, \mathrm{d}x \to \min \tag{22}$$

and the model

$$\frac{1}{2} \int_\Omega (u(x) - f(x))^2 + \lambda(|u_x| + |u_y|) \, \mathrm{d}x \to \min \tag{23}$$

treated, e.g., in Hintermüller and Kunisch (2004). Of course the second model is not rotationally invariant.

In the following, we restrict our attention for simplicity to quadratic $n \times n$ images and reshape them columnwise into a vector of length $N = n^2$. We

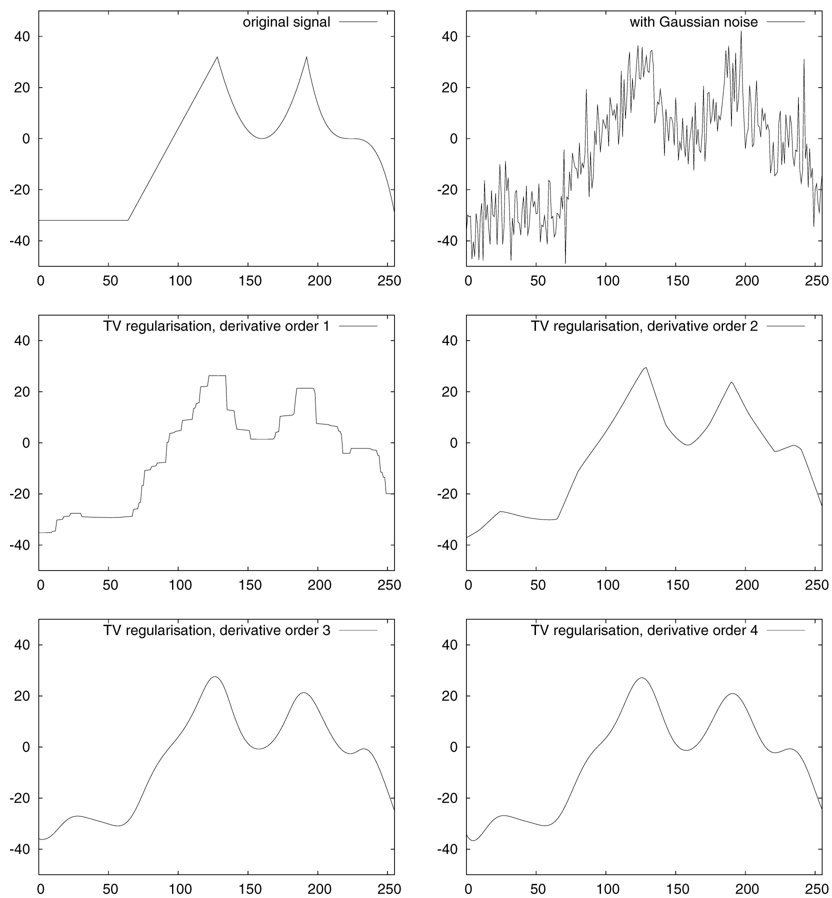

Figure 6. Denoising results with (5). Top left: original signal. Top right: noisy signal. Middle left: denoised signal for $m = 1$. Middle right: denoised signal for $m = 2$. Bottom left: denoised signal for $m = 3$. Bottom right: denoised signal for $m = 4$.

discretize the first order derivatives as proposed by Chambolle in Chambolle (2004). To this end, we introduce the gradient matrix

$$\mathcal{D} := \begin{pmatrix} \boldsymbol{I}_n \otimes \boldsymbol{D}_n^0 \\ \boldsymbol{D}_n^0 \otimes \boldsymbol{I}_n \end{pmatrix} \in \mathbb{R}^{2N,N} \quad \text{with} \quad \boldsymbol{D}_n^0 := \begin{pmatrix} \boldsymbol{D}_{n,1} \\ \boldsymbol{0}_{1,n} \end{pmatrix}$$

and the Kronecker product $\otimes$. The matrix $\mathcal{D}$ has rank $N-1$ and $\mathcal{D}^{\mathrm{T}}$ plays the role of $-\mathrm{div} = \nabla^*$. Further, we have that $\triangle_N := \mathcal{D}^{\mathrm{T}}\mathcal{D}$ is the finite difference discretization of the Laplace operator with the five point scheme and Neumann boundary conditions and that

$$\begin{aligned} \mathcal{R}(\mathcal{D}^{\mathrm{T}}) = \mathcal{R}(\triangle_N) &= \{\boldsymbol{f} \in \mathbb{R}^N : \sum_{j=1}^{N} f(j) = 0\}, \\ \mathcal{N}(\mathcal{D}) = \mathcal{N}(\triangle_N) &= \{\mu\, \mathbf{1}_{N,1} : \mu \in \mathbb{R}\} = \Pi_0. \end{aligned} \tag{24}$$

Finally, the discrete version of $|\nabla u| = (u_x^2 + u_y^2)^{1/2}$ reads $|\mathcal{D}\boldsymbol{u}|$, where

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

and $\circ$ denotes the componentwise vector product. Now we can discretize (22) and (23) by

$$\frac{1}{2}\, \|\boldsymbol{f} - \boldsymbol{u}\|_2^2 + \lambda \|\, |\mathcal{D}\boldsymbol{u}|\, \|_1 \tag{25}$$

and

$$\frac{1}{2}\, \|\boldsymbol{f} - \boldsymbol{u}\|_2^2 + \lambda \|\mathcal{D}\boldsymbol{u}\|_1, \tag{26}$$

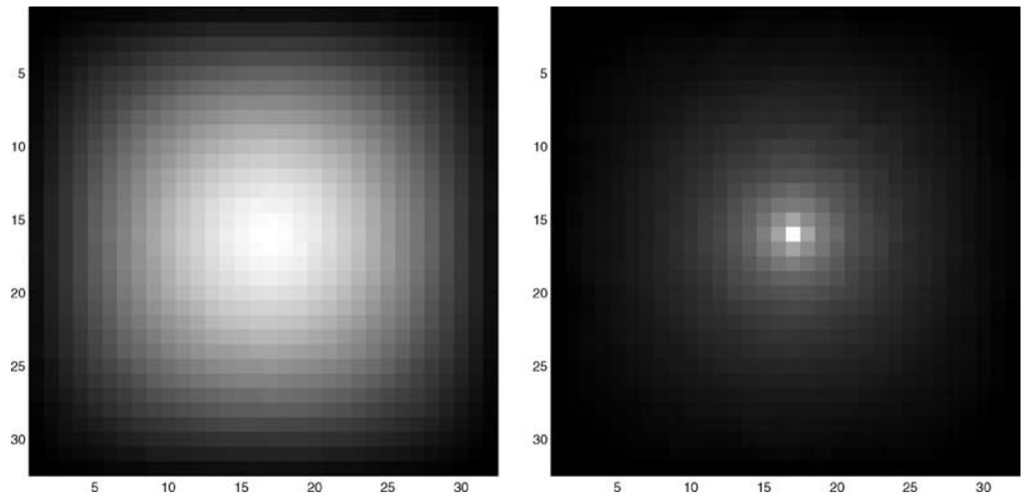

Figure 7. Column 528 of $\triangle_D^{-2}$ (left) and of $\triangle_D^{-1}$ (right) for $n = 32$.

respectively. Then, by the dual approach, see, e.g. Chambolle (2004); Steidl (2006), we obtain that $\boldsymbol{u} = \boldsymbol{f} - \boldsymbol{\mathcal{D}}^{\mathrm{T}}\boldsymbol{V}$, where $\boldsymbol{V}$ is the solution of

$$\begin{aligned} &\frac{1}{2}\|\boldsymbol{f} - \boldsymbol{\mathcal{D}}^{\mathrm{T}}\boldsymbol{V}\|_2^2 \quad \to \quad \min \\ \text{s.t.} \quad &\|\,|\boldsymbol{V}|\,\|_\infty \le \lambda, \quad \text{in case (25)}, \qquad (27) \\ \text{s.t.} \quad &\|\boldsymbol{V}\|_\infty \le \lambda, \quad \text{in case (26)}. \qquad (28) \end{aligned}$$

The first minimization problem can be solved for example by using Chambolle's semi–implicit gradient descent algorithm Chambolle (2004), while the second problem can be solved by standard QP methods. An example for the solution of both problems is presented at the bottom of Fig. 8. By the absence of rotation invariance, the solution of the second problem shows harder segmentation effects in x and y directions.

In the following, we assume that $\boldsymbol{f} \in \mathcal{R}(\boldsymbol{\mathcal{D}}^{\mathrm{T}})$, i.e., $\boldsymbol{f} = \boldsymbol{\mathcal{D}}^{\mathrm{T}}\boldsymbol{F}$ for some $\boldsymbol{F} \in \mathbb{R}^{2N}$. Otherwise we consider $\boldsymbol{f} - \mathrm{mean}(\boldsymbol{f})\mathbf{1}_{N,1}$. Then, since $\boldsymbol{\mathcal{D}u} = \boldsymbol{\mathcal{D}u}_{\mathcal{R}}$, and

$$\frac{1}{2}\|\boldsymbol{f} - \boldsymbol{u}\|_2^2 = \frac{1}{2}\|\boldsymbol{f} - \boldsymbol{u}_{\mathcal{R}}\|_2^2 + \frac{1}{2}\|\boldsymbol{u}_{\mathcal{N}}\|_2^2,$$

where $\boldsymbol{u}_{\mathcal{R}}$ is the orthogonal projection onto $\mathcal{R}(\boldsymbol{\mathcal{D}}^{\mathrm{T}})$ and $\boldsymbol{u}_{\mathcal{N}}$ the orthogonal projection onto $\mathcal{N}(\boldsymbol{\mathcal{D}}^{\mathrm{T}})$, it follows that the minimizer $\boldsymbol{u}$ of (25) and (26) is also in $\mathcal{R}(\boldsymbol{\mathcal{D}}^{\mathrm{T}})$. Now $\boldsymbol{U} = \boldsymbol{F} - \boldsymbol{V}$ solves the problem

$$\begin{aligned} &\frac{1}{2}\|\boldsymbol{\mathcal{D}}^{\mathrm{T}}\boldsymbol{U}\|_2^2 \quad \to \quad \min \\ \text{s.t.} \quad &\|\,|\boldsymbol{F} - \boldsymbol{U}|\,\|_\infty \le \lambda, \quad \text{in case (25)}, \\ \text{s.t.} \quad &\|\boldsymbol{F} - \boldsymbol{U}\|_\infty \le \lambda, \quad \text{in case (26)}. \end{aligned}$$

With respect to Remark 3 we note that the discrete G–norm defined for $\boldsymbol{v} \in \mathcal{R}(\boldsymbol{\mathcal{D}}^{\mathrm{T}})$ by $\|\boldsymbol{v}\|_G := \inf_{\boldsymbol{v}=\boldsymbol{\mathcal{D}}^{\mathrm{T}}\boldsymbol{V}} \|\,|\boldsymbol{V}|\,\|_\infty$ plays the role of the $\mathcal{W}_1(\boldsymbol{D}_{n,1})'$ norm.

For *higher order derivatives* even the choice of an appropriate disretization which preserves the basic integral identities satisfied by the continuous differential operators is a nontrivial question, see, e.g., Hyman and Shashkov (1997). However, operators of higher order were considered in image processing, e.g., in Chan et al. (2000); Chambolle and Lions (1997); Hinterberger and Scherzer (2003); Lysaker et al. (2003); Nielsen et al. (1997); Schnörr (1998); You and Kaveh (2000); Steidl (2006). Here we restrict our attention to

$$\frac{1}{2}\int_\Omega (u(x) - f(x))^2 + \lambda|\Delta u|\,\mathrm{d}x \quad \to \quad \min.$$

As discretization we choose

$$\frac{1}{2}\|\boldsymbol{f} - \boldsymbol{u}\|_2^2 + \lambda\|\triangle_D \boldsymbol{u}\|_1 \quad \to \quad \min \qquad (29)$$

where $\triangle_D$ denotes the finite difference discretization of the Laplace operator with the five point scheme and Dirichlet boundary conditions. Then $\triangle_D$ is invertible.

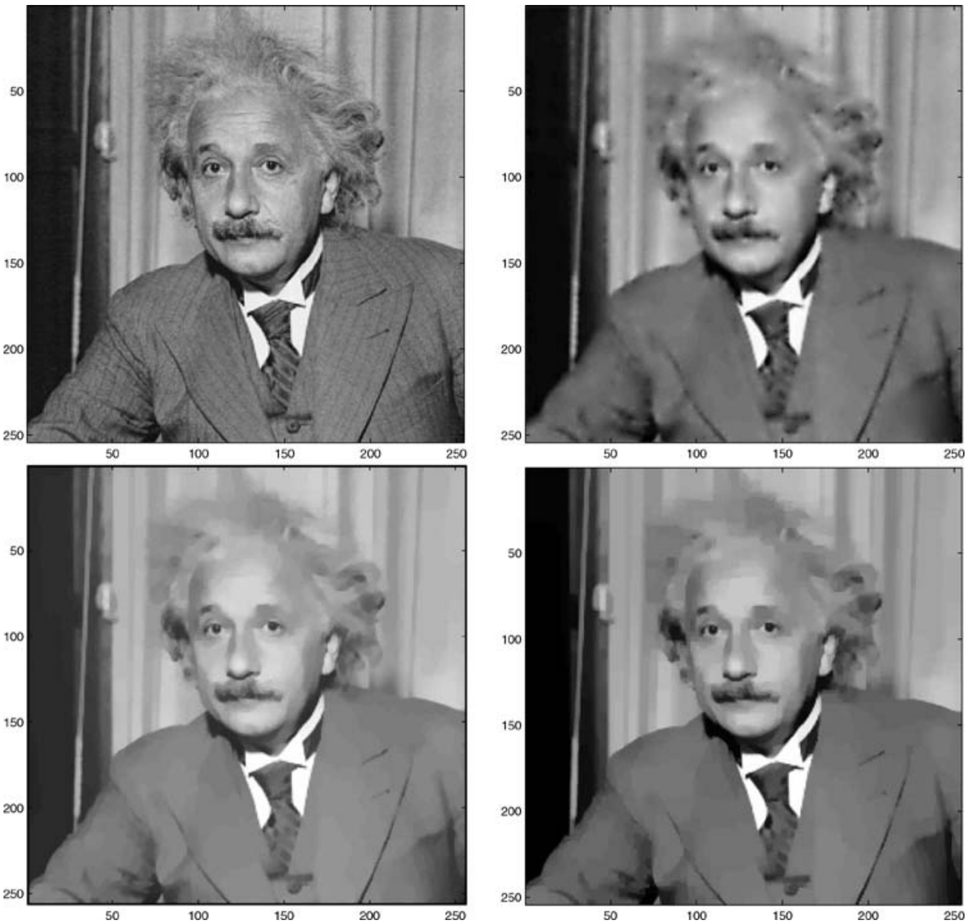

Figure 8. Top: Original 256 × 256 image (left). Solution of (30) (right). The image involves artefacts (white points). Bottom: Solution of (27) (left). Solution of (28) (right). The right-hand image shows a stronger segmentation in x and y direction. All problem were solved with $\lambda = 10$. For problem (27) we have used the semi–implicit gradient descent algorithm Chambolle (2004). Problems (30) and (28) were computed by the ILOG CPLEX Barrier Optimizer version 7.5. This routine uses a modification of the primal–dual predictor–corrector interior point algorithm described in Mehrotra, S. (1992).

The dual approach to (29) leads with $\boldsymbol{f} = \triangle_D \boldsymbol{F}$ and $\boldsymbol{u} = \triangle_D \boldsymbol{U}$ to the contact problem

$$\frac{1}{2}\|\triangle_D \boldsymbol{U}\|_2^2 \quad \rightarrow \quad \min \qquad (30)$$
$$\text{s.t.} \quad \|\boldsymbol{F} - \boldsymbol{U}\|_\infty \leq \lambda,$$

which can be solved by standard QP methods. An example for the solution of this problem in shown at the top of Fig. 8. The solution contains some artefacts in form of white points which were also mentioned in You and Kaveh (2000). Therefore the approach (29) seems to be not suited for applications in image processing. Obviously, $\triangle_D^{-2}$ is a reproducing kernel in $\mathbb{R}^N$ equipped with the norm given by the minimization term and $\boldsymbol{U} = \triangle_D^{-2}\boldsymbol{c}$ and $\boldsymbol{u} = \triangle_D^{-1}\boldsymbol{c}$ are in general sparse representations. The images corresponding to a central row of $\triangle_D^{-2}$ and $\triangle_D^{-1}$ are depicted in Figure 7.

With respect to the kernel $\triangle_D^{-2}$ let us finally note the following remark.

Remark 5.1. (Thin Plate Splines)
The so–called thin plate spline Duchon (1997) $K(x) := \frac{1}{8\pi}|x|^2 \ln |x|$ is the fundamental solution of the biharmonical operator $\triangle^2$. For appropriately chosen x_j the solution of

$$\frac{1}{2}\sum_{j=1}^{N}(f(x_j) - u(x_j))^2 + \lambda \int_\Omega u_{xx}^2 + 2u_{xy}^2 + u_{yy}^2 \,\mathrm{d}x \rightarrow \min$$

has the form $u(x) = \sum_{j=1}^{N} c_j K(x - x_j) + a_0 + a_1 x + a_2 y$.

6. Conclusions

We have shown the equivalence of the following problems in a discrete 1D setting:

i) minimzation of a functional with quadratic data term and TV regularization term with higher order derivatives,
ii) spline interpolation with variable knots depending on the input data and the regularization parameter,
iii) hard margin SVR in the discrete counterpart of the Sobolev space $W_{2,0}^m$,
iv) sparse representation in terms of fundamental splines with penalization the of l_1 norm of the coefficients.

Based on (6) a slightly different approach which handles the boundary conditions in advance (as done in 2D) is possible. Moreover, more general spline concepts as those of exponential splines, see, e.g., Unser and Blu (2005) and other data terms incorporating only few knots or related to other than Gaussian white noise can be considered in a similar way. Finally, the 2D setting deserves stronger investigation.

References

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

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

Chan, R. H., Ho, C. W., and Nikolova, M. Salt-and-pepper noise removal by median noise detectors and detail preserving regularization. *IEEE Transactions on Image Processing*, page to appear.

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

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

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

Didas, S. 2004. Higher order variational methods for noise removal in signals and images. Diplomarbeit, Universität des Saarlandes.

Duchon, J. 1997. Splines minimizing rotation-invariant seminorms in sobolev spaces. In *Constructive Theory of Functions of Several Variables*, pages 85–100, Berlin, Springer–Verlag.

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

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

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

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

Hyman, J. M. and Shashkov, M. J. 1997. Natural discretizations for the divergence, gradient, and curl on logically rectangular grids. *Comput. Math. Appl.*, 33(4):81–104.

Kimmeldorf, G. S. and Wahba, G. 1971. Some results on Tchebycheffian spline functions. *J. Anal. Appl.*, 33:82–95.

Lysaker, M., Lundervold, A., and Tai, X. 2003. 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.

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

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

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

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

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

Nikolova, M. 2004. A variational approach to remove outliers and impulse noise. *Journal of Mathematical Imaging and Vision*, 20:99–120.

Rockafellar, R. T. 1970. *Convex Analysis*. Princeton University Press, Princeton.

Rudin, L. I., Osher S., and Fatemi, E. 1992. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268.

Schnörr, C. 1998. A study of a convex variational diffusion approach for image segmentation and feature extraction. *Journal of Mathematical Imaging and Vision*, 8(3):271–292.

Steidl, G. 2006. A note on the dual treatment of higher order regularization functionals. *Computing*, 76:135–148.

Steidl, G., Didas, S., and Neumann, J. 2005. Relations between higher order TV regularization and support vector regression. In R. Kimmel, N. Sochen, and J. Weickert, editors, *Scale-Space and PDE Methods in Computer Vision*, volume 3459 of *Lecture Notes in Computer Science*, pages 515–527. Springer, Berlin.

Steidl, G., Weickert, J., Brox, T., Mrázek, P., and Welk, M. 2004. On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and SIDEs. *SIAM Journal on Numerical Analysis*, 42(2):686–713.

Unser, M. and Blu, T. 2005. Generalized smoothing splines and the optimal discretization of the Wiener filter. *IEEE Transactions on Signal Processing*, 53(6):2146–2159.

Vapnik, V. N. 1998. *Statistical Learning Theory*. John Wiley and Sons, Inc.

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

Wahba, G. 1990. *Spline Models for Observational Data*. SIAM, Philadelphia.

Welk, M., Weickert, J., and Steidl, G. 2005. A four-pixel scheme for singular differential equations. In R. Kimmel, N. Sochen, and J. Weickert, editors, *Scale-Space and PDE Methods in Computer Vision*, Lecture Notes in Computer Science. Springer, Berlin, to appear.

Yip, A. M. and Park, F. 2003. Solution dynamics, causality, and critical behaviour of the regularization parameter in total variation denoising problems. UCLA Report.

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