Relations between Higher Order TV Regularization and Support Vector Regression*

G. Steidl[1], S. Didas[2] and J. Neumann[1]

[1] Faculty of Mathematics and Computer Science, D7, 27
University of Mannheim, 68131 Mannheim, Germany
{steidl;jneumann}@uni-mannheim.de
http://www.kiwi.math.uni-mannheim.de

[2] Mathematical Image Analysis Group
Faculty of Mathematics and Computer Science, Building 27
Saarland University, 66123 Saarbrücken, Germany
didas@mia.uni-saarland.de
http://www.mia.uni-saarland.de

Abstract. We study the connection between higher order total variation (TV) regularization and support vector regression (SVR) with spline kernels in a one-dimensional discrete setting. We prove that the contact problem arising in the tube formulation of the TV minimization problem is equivalent to the SVR problem. Since the SVR problem can be solved by standard quadratic programming methods this provides us with an algorithm for the solution of the contact problem even for higher order derivatives. Our numerical experiments illustrate the approach for various orders of derivatives and show its close relation to corresponding nonlinear diffusion and diffusion–reaction equations.

1 Introduction

In this paper, we are interested in constructing a function u that minimizes the functional

$$\int_0^1 (u(x) - f(x))^2 + 2\lambda |u^{(m)}(x)| \, \mathrm{d}x. \tag{1}$$

More precisely, we are concerned with a discrete version of (1), where the functions are only considered at equispaced points.

For $m = 1$ and arbitrary space dimensions, we are in the classical Rudin–Osher–Fatemi setting [16] applied in image denoising and segmentation. Several numerical solution algorithms were proposed, see, e.g., [24] and references therein. A quite interesting method uses the tube formulation of (1). In one space dimension, the tube approach is known as a non–parametric regression model in statistics [10]. A generalization to the two-dimensional setting was proposed in

* This joint research was supported by the Deutsche Forschungsgemeinschaft within the projects WE 2602/2-2 and SCHN 457/5-2. This is gratefully acknowledged.

[7]. The heart of the tube method consists in the solution of a contact problem within a tube of width depending on the regularization parameter $\lambda > 0$. For $m = 1$, this contact problem can be solved efficiently by the so–called 'taut string algorithm' [4] in one dimension, but becomes harder in higher dimensions [7].

In recent years, there has been a growing interest in higher order variational methods [13,17,3,26,9,8]. In particular, a tube approach for $m \geq 2$ was addressed for one dimension in [10] and for higher dimensions based on Meyer's G–norm [12] in [14].

In this paper, we will see that the contact problem can be tackled by solving a simple quadratic optimization problem, namely a so–called *support vector regression* (SVR) problem. SVR methods became very popular in machine learning during the last years, see [23]. The SVR approach also approximates a given function within a tube, but by minimizing a different cost functional. The SVR solution is always contained in a previously determined reproducing kernel Hilbert space. We will prove that in our discrete setting the solution of the contact problem corresponding to (1) coincides with the SVR solution in an appropriately chosen reproducing kernel Hilbert space. This space is a discrete variant of the Sobolev space W_0^m which has a reproducing kernel determined by splines of degree $2m - 1$. We remark that similar results can be obtained by applying the dual approach of Chambolle [2] to our setting. This is discussed in [20]. In this paper, we want to emphasize the spline point of view.

Our paper is organized as follows. We start by developing the tube formulation and SVR with spline kernels in a discrete setting in Sections 2 and 3, respectively. In Section 4, we prove the equivalence of the SVR problem and the key part of the tube algorithm, the contact problem. To prepare a numerical comparison, a discretization of corresponding partial differential equations (PDEs) is provided in Section 5. Our denoising experiments in Section 6 demonstrate properties of our method for various orders of derivatives and show the relation of the variational approach to the numerical solution of corresponding diffusion and diffusion–reaction equations. The paper is concluded with Section 7.

2 Tube characterization of TV regularization functionals with higher order derivatives

In the following, we are concerned with discrete functions defined on some subsets of the integers. As a discrete version of the m–th derivative we choose the m–th finite difference

$$D^m u(j) := \sum_{k=0}^{m} (-1)^{k+m} \binom{m}{k} u\left(j - \left\lfloor \frac{m}{2} \right\rfloor + k\right), \tag{2}$$

where $\lfloor x \rfloor$ denotes the largest integer $\leq x$. Given the values $f(j)$, $j = 1, \ldots, n$, we are interested in finding a discrete function u that minimizes the functional

$$J(u) := \sum_{j=1}^{n} (u(j) - f(j))^2 + 2\lambda \sum_{j=1}^{n} |D^m u(j)|, \tag{3}$$

where we suppose the boundary conditions

$$D^m u(j) := 0, \qquad j = -\left\lfloor \frac{m-1}{2} \right\rfloor, \ldots, \left\lfloor \frac{m}{2} \right\rfloor; \, n - \left\lfloor \frac{m-1}{2} \right\rfloor, \ldots, n + \left\lfloor \frac{m}{2} \right\rfloor. \quad (4)$$

The boundary conditions for $j = 1, \ldots, \lfloor \frac{m}{2} \rfloor; n - \lfloor \frac{m-1}{2} \rfloor, \ldots, n$ imply that the second sum in (3) runs indeed only from $\lfloor \frac{m}{2} \rfloor + 1$ to $n - \lfloor \frac{m+1}{2} \rfloor$. The other boundary conditions are imposed to keep the summation index in the following derivation simple. We remark that these boundary conditions are equivalent to $D^k u(0) = D^k u(n) = 0$, $k = m, \ldots, 2m-1$.

Since the functional J is strictly convex, our problem has a unique solution. A necessary and sufficient condition for u to be the minimizer of (3) is that the zero vector is an element of the subgradient $\partial J(u)$, i.e., for $j = 1, \ldots, n$, the following inclusions must be fulfilled:

$$0 \in u(j) - f(j) + \lambda \sum_{k=0}^{m} (-1)^k \binom{m}{k} \frac{D^m u(j - \lfloor \frac{m+1}{2} \rfloor + k)}{|D^m u(j - \lfloor \frac{m+1}{2} \rfloor + k)|}, \quad (5)$$

where $y/|y| := [-1, 1]$ if $y = 0$, and where the same quotient $D^m u(\cdot)/|D^m u(\cdot)|$ in different inclusions denotes the same numbers in $[-1, 1]$. Moreover, the summands corresponding to our boundary conditions (4) are zero.

We want to find linear combinations of the right–hand sides of (5) such that most of the terms in the sum vanish. For this, we introduce a discrete equivalent to the m–th power function by $k^{(m)} := 1$ for $m = 0$ and $k^{(m)} := k(k+1)\ldots(k+m-1)$ for $m \geq 1$ and a discrete version of the m–th antiderivative of a function f by

$$F_f(j) := \sum_{k=1}^{j} \frac{(j+1-k)^{(m-1)}}{(m-1)!} f(k), \qquad j = 1, \ldots, n. \quad (6)$$

Then we obtain the following proposition which can be considered as a discrete counterpart of a result in [10].

Proposition 1 (Tube Characterization of TV Minimization)
The function u is a solution of (3) *if and only if F_u fulfills the conditions*

$$F_u(j) \in F_f(j) - (-1)^m \lambda \frac{D^m u(j + \lfloor \frac{m}{2} \rfloor)}{|D^m u(j + \lfloor \frac{m}{2} \rfloor)|}, \qquad j = 1, \ldots, n-m \quad (7)$$

and $\quad F_u(n-j) = F_f(n-j)$, $j = 0, \ldots, m-1$.

The basic idea of the proof is the following: For $j \in \{1, \ldots, n\}$ and $k = 1, \ldots, j$, we multiply the k–th inclusion in (5) by $(j+1-k)^{(m-1)}/(m-1)!$, add the corresponding j expressions and transfer $F_u(j)$ to the opposite side. By (6) and setting $F_u(j) := 0$, $j = -(m-1), \ldots, 0$, we obtain u from given F_u by

$$u(j) = \sum_{k=0}^{m} (-1)^k \binom{m}{k} F_u(j-k), \qquad j = 1, \ldots, n. \quad (8)$$

Then, by (2), the finite differences appearing in (7) can be written as

$$D^m u\left(j + \left\lfloor \frac{m}{2} \right\rfloor\right) = D^{2m} F_u(j), \qquad j = 1, \ldots, n-m.$$

Together with Proposition 1 this implies that the function F_u corresponding to the minimizer u of (3) is uniquely determined by the following *contact problem*:

(T1) $F_u(j) = 0$ for $j = -(m-1), \ldots, 0$,
(T2) $F_u(n-j) = F_f(n-j)$ for $j = 0, \ldots, m-1$.
(T3) F_u lies in a tube around F_f of width 2λ, i.e.,
$|F_f(j) - F_u(j)| \le \lambda$ for $j = 1, \ldots, n-m$.
(T4) Let $\Lambda := \{j \in \{1, \ldots, n\} : D^{2m} F_u(j) \neq 0\}$.
If $j \in \Lambda$, then $F_u(j)$ contacts the boundary of the tube, where
$D^{2m} F_u(j) > 0 \Longrightarrow F_u(j) = F_f(j) - (-1)^m \lambda$,
$D^{2m} F_u(j) < 0 \Longrightarrow F_u(j) = F_f(j) + (-1)^m \lambda$.

Then the usual tube method for solving (3) consists of the three steps

1. compute F_f from given f by (6),
2. solve the contact problem (T1) – (T4) to obtain F_u,
3. compute u by (8),

where the second step requires further explanation.

For the classical setting $m = 1$, it is well-known, see, e.g., [4], that (T1) – (T4) is equivalent to the construction of the uniquely determined taut string within the tube around F_f of width 2λ fixed at $(0,0)$ and $(n, F_f(n))$, i.e., to the solution F_u of the following optimization problem:

$$\sum_{j=0}^{n-1} \left(1 + (F_u(j+1) - F_u(j))^2\right)^{1/2} \to \min, \tag{9}$$

subject to $|F_u(j) - F_f(j)| \le \lambda$, $j = 1, \ldots, n-1$ and $F_u(0) = 0$, $F_u(n) = F_f(n)$. 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.

For $m \ge 2$, the computation of F_u is more complicated. An iterative method based on an exchange of contact knots of conjectured complexity $\mathcal{O}(n^2)$ was, e.g., proposed in [10].

Finally, we remark that discrete functions F fulfilling the property $D^{2m} F(j) = 0$ for all $j \notin \Lambda$ and some boundary conditions were introduced as *discrete splines of degree* $2m - 1$ *with spline knots* Λ by Mangasarian and Schumaker [11].

3 Support vector regression with spline kernels

The SVR method searches for approximations of functions in reproducing kernel Hilbert spaces. Among the large amount of literature on SVR we refer to [23, Chapter 11]. Well–known examples of reproducing kernel Hilbert spaces are the Sobolev spaces W_0^m of real–valued functions having a weak m–th derivative in $L_2[0,1]$ and fulfilling $F^{(j)}(0) = 0$ for $j = 0, \ldots, m-1$ with inner product

$\langle F, G\rangle_{W_0^m} := \int_0^1 F^{(m)}(x)G^{(m)}(x)\,\mathrm{d}x$. These spaces have the positive definite reproducing kernels $K(x,y) := \int_0^1 (x-t)_+^{m-1}(y-t)_+^{m-1}/((m-1)!)^2\,\mathrm{d}t$, where $(x)_+ := \max\{0,x\}$; see [25, p. 5–14].

For our purposes, we introduce discrete versions of W_0^m by the Hilbert spaces $\mathcal{W}_0^m$ of real–valued functions defined on $\{-(m-1),\ldots,n\}$ and fulfilling $F(j)=0$ for $j=-(m-1),\ldots,0$ with inner products

$$\langle F, G\rangle_{\mathcal{W}_0^m} := \sum_{j=-\lfloor \frac{m-1}{2}\rfloor}^{n-\lfloor \frac{m+1}{2}\rfloor} D^m F(j) D^m G(j). \tag{10}$$

We can prove that $\mathcal{W}_0^m$ has the reproducing kernel

$$K(i,j) := \sum_{k=0}^{\min(i,j)-1} \frac{(i-k)^{(m-1)}}{(m-1)!}\,\frac{(j-k)^{(m-1)}}{(m-1)!},$$

i.e., $\langle F, K(i,\cdot)\rangle_{\mathcal{W}_0^m} = F(i)$. Moreover, $K(i,\cdot)$ fulfills for fixed $i\in\{1,\ldots,n\}$ the properties $D^{2m}K(i,i)\neq 0$ and

$$D^{2m}K(i,j) = 0, \qquad j=1,\ldots,n; j\neq i, \tag{11}$$

$$K(i,j) = 0, \qquad j=-(m-1),\ldots,0. \tag{12}$$

Let $\boldsymbol{K} := (K(i,j))_{i,j=1}^n$ and $\boldsymbol{F} := (F(1),\ldots,F(n))'$ be given. Then we are looking for a function

$$U(j) := \sum_{i=1}^n c_i K(i,j) \tag{13}$$

with coefficient vector $\boldsymbol{c} := (c_1,\ldots,c_n)'$ that solves the following constrained quadratic optimization problem:

$$\begin{aligned} &\boldsymbol{c}'\boldsymbol{K}\boldsymbol{c} \to \min,\\ \text{subject to}\quad &\boldsymbol{F}-\boldsymbol{K}\boldsymbol{c} \leq \lambda\boldsymbol{e},\\ &-\boldsymbol{F}+\boldsymbol{K}\boldsymbol{c} \leq \lambda\boldsymbol{e},\\ &\textstyle\sum_{i=1}^n c_i K(i,n-j) = F(n-j) \qquad j=0,\ldots,m-1. \end{aligned} \tag{14}$$

Here $\boldsymbol{e}$ denotes the vector consisting of n components one and the inequalities are taken componentwise. This problem without the equality constraints is known as *SVR problem*. Since $\boldsymbol{K}$ is positive definite, it has a unique solution which can be computed by standard quadratic programming methods. Obviously, by (13), the inequality constraints in (14) can be rewritten as $|F(j)-U(j)|\leq\lambda$ while the equality constraints read $F(n-j)=U(n-j)$, $j=0,\ldots,m-1$. Further, the kernel property (12) implies together with (13) that $U(j)=0$ for $j=-(m-1),\ldots,0$. Based on the Karush–Kuhn–Tucker conditions and the dual formulation of (14), see [6], one can further show that $c_i\neq 0$ implies $|F(i)-U(i)|=\lambda$. The points

$i \in \{1, \ldots, n\}$ with $c_i \neq 0$ are called *support vectors*. Clearly, the function U only depends on the support vectors. If $\tilde{\Lambda}$ denotes the set of support vectors, then

$$U(j) = \sum_{i \in \tilde{\Lambda}} c_i K(i,j), \tag{15}$$

so that by (11), the support vectors j can be also characterized by $D^{2m} U(j) \neq 0$.

We summarize the properties of the SVR solution:

(S1) $U(j) = 0$ for $j = -(m-1), \ldots, 0$,
(S2) $U(n-j) = F(n-j)$ for $j = 0, \ldots, m-1$,
(S3) U lies in a tube around F of width 2λ, i.e.,
$|F(j) - U(j)| \leq \lambda$ for $j = 1, \ldots, n-m$.
(S4) Let $\tilde{\Lambda} := \{j \in \{1, \ldots, n\} : D^{2m} U(j) \neq 0\}$.
If $j \in \tilde{\Lambda}$, then $U(j)$ contacts the boundary of the tube,
where j are the support vectors obtained by solving (14).

Comparing these properties with (T1) – (T4), we see that for $F = F_f$ only the fourth condition differs.

Finally, we remark that the SVR solution U can be considered as sparse approximation of F. In particular, by [6], U (without the last m equality constraints) is also the solution of the unconstrained minimization problem

$$\|F - U\|^2_{\mathcal{W}_0^m} + 2\lambda \|\boldsymbol{c}\|_{\ell_1} \to \min .$$

4 Equivalence of tube and SVR solution

In the following, we set $F := F_f$ in (14) and show that the solution U of (14) coincides with the solution of the contact problem (T1) – (T4). Since by the reproducing kernel property $\boldsymbol{c}'\boldsymbol{K}\boldsymbol{c} = \|U\|^2_{\mathcal{W}_0^m}$, we can use (10) to rewrite (14) as

$$\mathcal{E}(U) := \sum_{j=-\lfloor \frac{m-1}{2} \rfloor}^{n-\lfloor \frac{m+1}{2} \rfloor} \left(D^m U(j)\right)^2 \to \min, \tag{16}$$

$$\text{subject to} \qquad |U(j) - F_f(j)| \leq \lambda, \quad j = 1, \ldots, n-m,$$
$$U(n-j) = F_f(n-j), \quad j = 0, \ldots, m-1.$$

In particular, for $m = 1$, we minimize just the sum of the squared lengths

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

instead of the lengths in (9). However, by the following proposition both problems are equivalent.

Proposition 2 (Equivalence of Contact and SVR Problem) *The solution of the contact problem* (T1) – (T4) *coincides with the solution of the SVR problem* (16).

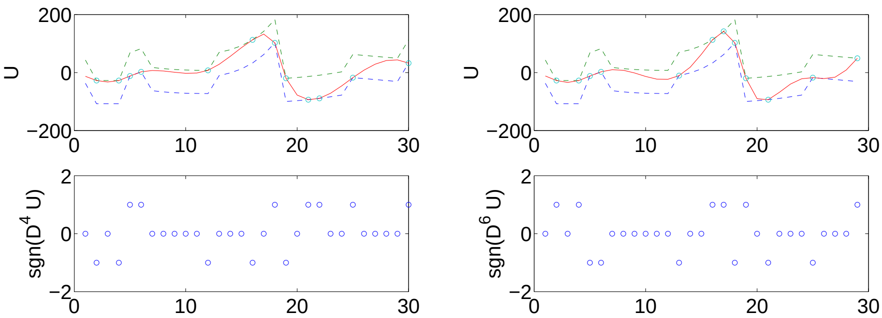

Fig. 1. Property (T4) of U computed by (16) for $m = 2$ (left) and $m = 3$ (right). Top: U (solid line) with tube (dashed line) and contact points. Bottom: Sign of $D^{2m}U$.

Proof. Let U be the solution of (16). Assume that $j \in \tilde{\Lambda}$ is an upper contact point that does not fulfill (T4), i.e., $(-1)^m D^{2m} U(j) > 0$. (Similar arguments can be used for lower contact points.) By (2), this means that

$$U(j) > W(j) := -(-1)^m \sum_{\substack{k=0 \\ k \neq m}}^{2m} \binom{2m}{k} (-1)^k U(j-m+k) / \binom{2m}{m}.$$

By definition $W(j)$ lies on the discrete spline of degree $2m-1$ through $U(j \pm k)$, $k = 1, \ldots, m$. Now we define a function V which is equal to U except at j, where

$$V(j) := \begin{cases} W(j), & \text{if } W(j) > U(j) - 2\,\lambda, \\ U(j) - 2\,\lambda & \text{otherwise.} \end{cases}$$

Obviously, V fulfills the constraints of (16) and

$$U(j) > V(j) \geq W(j). \tag{17}$$

We show that $\mathcal{E}(V) < \mathcal{E}(U)$. This contradicts the choice of U as minimizer of (16) and we are done. Replacing D^m in $\mathcal{E}$ by (2) and regarding that $V(i) = U(i)$ for $i \neq j$, we obtain after some technical computations that

$$\mathcal{E}(U) - \mathcal{E}(V) = \binom{2m}{m} \big(U(j) - V(j)\big) \big(U(j) + V(j) - 2\,W(j)\big).$$

Now we have by (17) that $\mathcal{E}(U) - \mathcal{E}(V) > 0$. This completes the proof. $\square$

Fig. 1 demonstrates property (T4) for the solution U of (16).

5 Parabolic PDEs with higher order derivatives

Regularization methods are closely related to parabolic PDEs by the Euler–Lagrange equation, see, e.g.,[19]. To allow for a comparison of our tube–SVR method with PDE approaches we shortly describe their relations. For this, we consider the slightly modified version of (1) suggested in [1,3]

$$\int_0^1 (u(x) - f(x))^2 + 2\lambda\,\varphi\left((u^{(m)})^2\right)\,\mathrm{d}x$$

with $\varphi(s^2) := \left(\varepsilon^2 + s^2\right)^{\frac{1}{2}}$. A minimizer u of this functional necessarily satisfies the Euler–Lagrange equation

$$\frac{u - f}{\lambda} = (-1)^{m+1} \frac{\partial^m}{\partial x^m} \left(2\,\varphi'\left((u^{(m)})^2\right) u^{(m)}\right) \tag{18}$$

with natural boundary conditions $u^{(k)}(0) = u^{(k)}(1) = 0$, $k = m, \ldots, 2m-1$, see [5]. These boundary conditions are in agreement with our boundary conditions (4). Introducing an additional time variable t, the left–hand side of equation (18) can be understood as fully implicit time discretization of the *diffusion equation*

$$\frac{\partial u}{\partial t} = (-1)^{m+1} \frac{\partial^m}{\partial x^m} \left(\frac{u^{(m)}}{\sqrt{(u^{(m)})^2 + \varepsilon^2}}\right) \tag{19}$$

with natural boundary conditions, initial value f and stopping time λ. To solve (19) we use finite differences for the derivatives in space and an explicit Euler scheme in time. This leads to the following iterative scheme:

$$\begin{aligned} u^0(j) &:= f(j), \\ v^k(j) &:= \frac{D^m u^k(j)}{\sqrt{\left(D^m u^k(j)\right)^2 + \varepsilon^2}}, \\ u^{k+1}(j) &:= u^k(j) - \tau \sum_{l=0}^{m} (-1)^l \binom{m}{l}\; v^k\left(j - \left\lfloor \frac{m+1}{2} \right\rfloor + l\right), \end{aligned}$$

where we set $D^m u^k(j) := 0$ for $j = 0, \ldots, \lfloor \frac{m}{2} \rfloor; n - \lfloor \frac{m-1}{2} \rfloor, \ldots, n$. This scheme satisfies stability in the Euclidean norm if the time step size τ is chosen sufficiently small, namely $\tau \le \frac{\varepsilon}{2^{2m-1}}$. In comparisons with regularization methods we use the regularization parameter λ as stopping time, i.e., we iterate until $k = \frac{\lambda}{\tau}$.

Alternatively, we can also approximate a solution of (18) by solving the *diffusion–reaction equation*

$$\frac{\partial u}{\partial t} = \frac{u - f}{\lambda} + (-1)^m \frac{\partial^m}{\partial x^m} \left(2\,\varphi'\left((u^{(m)})^2\right) u^{(m)}\right)\,. \tag{20}$$

A discretization of this equation can be obtained in a similar way to the one for the diffusion equation. It should be noted that the steady state of (20) for $t \to \infty$ yields a solution of (18) while the diffusion approach (19) leads in general only to an approximation of the solution. Only for the classical setting $m = 1$ without additional ε–regularization, it was shown in [21,15] that the analytical solution of the space discrete diffusion equation (19) is equivalent to the solution of the optimization problem (1). For a space continuous version we refer to [22]. Even for first order derivatives this is a very special property of the TV regularization function φ.

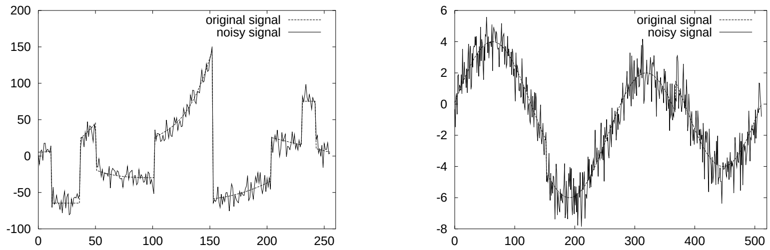

Fig. 2. Test signals. Left: Piecewise polynomial signal, 256 pixels, SNR 14.74dB. Right: Piecewise sine signal, 512 pixels, SNR 10.25dB.

	Polynomial signal		Sine signal	
Order m	λ	SNR (dB)	λ	SNR (dB)
1	15	21.34	3	20.03
2	5	18.45	16	21.96
3	2	17.59	174	21.91

Table 1. Optimal parameters λ and SNR values for tube–SVR denoising.

6 Denoising experiments

In this section, we show by denoising experiments that our tube–SVR approach works well even in comparison with corresponding PDE methods and demonstrate the influence of higher order derivatives.

As examples we have used the signals shown in Fig. 2. The first signal is piecewise polynomial, and Gaussian noise with standard deviation 10 was added. The other one consists of piecewise sine signals and the noise standard deviation is 1. First, we have determined the optimal parameters λ for our tube–SVR denoising method 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. We have applied the tube–SVR method described at the end of Section 2, where the contact problem was solved by applying the Matlab quadratic programming routine to (16). This routine is based on an active set method. The results are contained in Tab. 1.

Then we compared the quality of the results obtained by our tube–SVR approach and by the PDE methods for various orders of derivatives m. In our PDE experiments we have used a regularization parameter $\varepsilon = 10^{-4}$ and for each order the maximal time step size. One should be aware of the influence of the parameter ε for both PDE methods and the number of iterations for the diffusion–reaction method. For smaller values of ε one could even obtain better results at the cost of a higher number of iterations. Figs. 3 and 4 show the denoising results. Since one can visually not distinguish between the tube–SVR and the diffusion–reaction results we have only plotted the diffusion results in the PDE part. However, the diffusion results look also very similar except for slight smoothing effects for $m = 2$. To affirm this impression numerically, Tab. 2 shows the maximal absolute differences between the results of our tube–SVR method and the diffusion/diffusion–reaction methods.

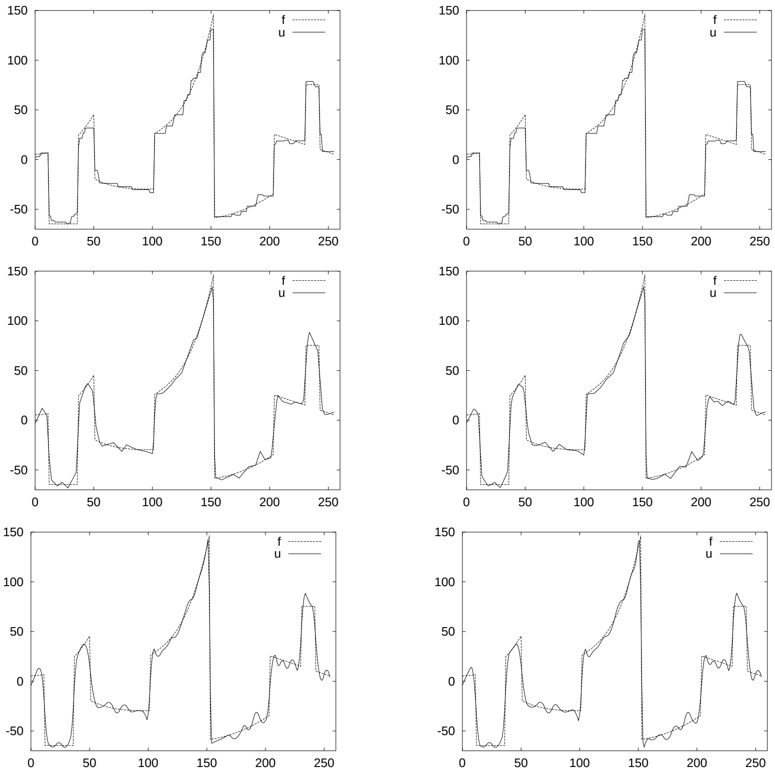

Fig. 3. Denoising results for the piecewise polynomial signal with $\lambda = 15$. Left: Tube–SVR method. Right: Diffusion method. Top: First order. Middle: Second order. Bottom: Third order.

7 Conclusions

We have proved that the contact problem arising in the tube formulation of the minimization problem with ℓ_2 data term and TV regularization term with higher order derivatives can be formulated as SVR problem with discrete spline kernels. Therefore the problem is closely related to spline interpolation with variable knots. The results can also be considered from a different point of view, namely by applying Chambolle's dual approach to our setting, see [20]. This will be the basis for handling higher space dimensions. In our denoising experiments we have also incorporated corresponding nonlinear diffusion and diffusion–reaction equations with higher order derivatives which lead to similar results.

References

1. R. Acar and C. R. Vogel. Analysis of bounded variation penalty methods for ill–posed problems. *Inverse Problems*, 10:1217–1229, 1994.
2. A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20:89–97, 2004.

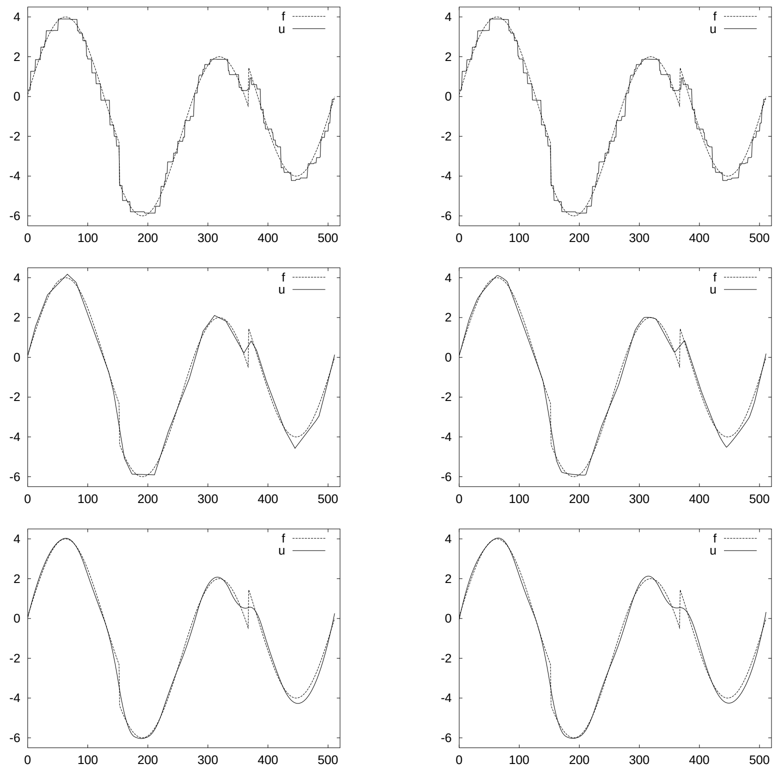

Fig. 4. Denoising results for the piecewise sine signal. Left: Tube–SVR method. Right: Diffusion method. Top: First order, $\lambda = 3$. Middle: Second order, $\lambda = 16$. Bottom: Third order, $\lambda = 174$.

3. 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.
4. P. L. Davies and A. Kovac. Local extremes, runs, strings and multiresolution. *Annals of Statistics*, 29:1–65, 2001.
5. S. Didas. Higher order variational methods for noise removal in signals and images. Diploma Thesis, Dept. of Math., Saarland University, Saarbrücken, Germany, 2004.
6. F. Girosi. An equivalence between sparse approximation and support vector machines. *Neural Computation*, 10(6):1455–1480, 1998.
7. 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.
8. M. Loog. Support blob machines: The sparsification of linear scale-space. In T. Pajdla and J. Matas, eds., *Computer Vision – ECCV 2004, Part IV*, volume 3024 of *Lecture Notes in Computer Science*, 14–24, Springer, Berlin, 2004.
9. M. Lysaker, A. Lundervold, and X.-C. Tai. Noise removal using fourth-order partial differential equations with applications to medical magnetic resonance images in space and time. Technical Report CAM-02-44, Department of Mathematics, University of California, 2002.

Order m	Polynomial signal			Sine signal		
	Diffusion–reaction		Diffusion	Diffusion–reaction		Diffusion
	Iterations	l_∞–norm	l_∞–norm	Iterations	l_∞–norm	l_∞–norm
1	10^7	$6.2 \cdot 10^{-4}$	$1.2 \cdot 10^{-2}$	10^7	$9.0 \cdot 10^{-4}$	$8.5 \cdot 10^{-3}$
2	10^8	$8.2 \cdot 10^{-4}$	7.2	10^8	$1.3 \cdot 10^{-2}$	$1.9 \cdot 10^{-1}$
3	10^8	$6.0 \cdot 10^{-4}$	5.1	$5 \cdot 10^8$	$1.1 \cdot 10^{-1}$	$1.0 \cdot 10^{-1}$

Table 2. Difference between tube–SVR method and diffusion/diffusion–reaction approach.

10. E. Mammen and S. van de Geer. Locally adaptive regression splines. *Annals of Statistics*, 25(1):387–413, 1997.
11. 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.
12. Y. Meyer. *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*, volume 22 of *University Lecture Series*. AMS, Providence, 2001.
13. M. Nielsen, L. Florack, and R. Deriche. Regularization, scale-space and edge detection filters. *Journal of Mathematical Imaging and Vision*, 7:291–307, 1997.
14. 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, 2004.
15. I. Pollak, A. S. Willsky and Y. Huang. Nonlinear evolution equations as fast and exact solvers of estimation problems. *IEEE Trans. on Signal Processing*, in print.
16. L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.
17. O. Scherzer. Denoising with higher order derivatives of bounded variation and an application to parameter estimation. *Computing*, 60:1–27, 1998.
18. O. Scherzer. Taut string algorithm and regularization programs with G-norm. *Journal of Mathematical Imaging and Vision*, in print.
19. O. Scherzer and J. Weickert. Relations between regularization and diffusion filtering. *Journal of Mathematical Imaging and Vision*, 12:43–63, 2000.
20. G. Steidl. A note on the dual treatment of higher order regularization functionals. Technical Report 278, Institute of Mathematics, University of Mannheim, 2004.
21. G. Steidl, J. Weickert, T. Brox, P. Mrázek, and M. Welk. On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and SIDEs. *SIAM Journal on Numerical Analysis*, 42(2):686–713, 2004.
22. D. M. Strong. *Adaptive Total Variation Minimizing Image Restoration*. PhD thesis, Department of Mathematics, University of California, 1997.
23. V. N. Vapnik. *Statistical Learning Theory*. John Wiley and Sons, Inc., 1998.
24. C. R. Vogel. *Computational Methods for Inverse Problems*. SIAM, Philadelphia, 2002.
25. G. Wahba. *Spline Models for Observational Data*. SIAM, Philadelphia, 1990.
26. Y.-L. You and M. Kaveh. Fourth-order partial differential equations for noise removal. *IEEE Transactions on Image Processing*, 9(10):1723–1730, 2000.