FSP Report No. 57

Convex Regularization Methods with Non-Uniform Total Variation Penalization

Markus Grasmair

October 2007

Convex Regularization Methods with Non-Uniform Total Variation Penalization

Markus Grasmair

Institute for Computer Science
University of Innsbruck
Technikerstr. 21a/2
6020 Innsbruck
Austria
markus.grasmair@uibk.ac.at

October 10, 2007

Abstract. We consider one-dimensional regularization methods with convex fidelity term and non-uniform higher order total variation penalization term. A unique characterization of the solutions is provided using a dual formulation of the problem. Particular emphasis is put on the case of first order regularization. Here, we prove the equivalence of the continuous solution of the discretized problem with a fully discrete solution. Moreover, we indicate how the taut string algorithm, a highly efficient algorithm for the solution of L^2-total variation regularization, can be generalized to non-quadratic regularization functionals.

1 Introduction

Let $\Omega := (a, b) \subset \mathbb{R}$, and let $f \in L^2(\Omega)$ be some given data. There exist various methods for smoothing the possibly noisy data f one of which is total variation regularization defined by minimization of the functional

$$\Phi(u) := \frac{1}{2}\|u - f\|_2^2 + \alpha |Du|(\Omega) = \frac{1}{2}\int_a^b \left(u(t) - f(t)\right)^2 dt + \alpha |Du|(\Omega)$$

over the space $\mathrm{BV}(\Omega)$ of functions of bounded variation (cf. [1, 2, 7, 11, 13]). Here, $\alpha > 0$ is a regularization parameter determining the amount of smoothing, and $|Du|(\Omega)$ denotes the total variation of the function u on Ω. Minimization of Φ_α yields a function u that is close to the data in the L^2-norm while having a reduced total variation.

Depending on the exact requirements on the solution it is often necessary to consider variants of Φ for defining the regularization.

First, it makes sense not to use the L^2-norm as fidelity term, but rather a different L^p-norm with $1 \le p < \infty$. In particular, the L^1-norm has advantages, since outliers in the data can efficiently be smoothed (see [9, 12]).

Second, depending on the desired regularity of the solution, it is possible to use a higher order variation as regularization term. This leads to solutions that are piecewise polynomials of degree $k-1$, when the k-th order total variation of u is used in the definition of Φ (cf. [8, 9]).

Finally, it may be necessary to locally adjust the regularization parameter, that is, instead of a fixed parameter $\alpha > 0$ to consider a regularization function $\alpha : \Omega \to \mathbb{R}_{>0}$. In [2], for instance, the parameter α is iteratively locally adapted under the heading of local squeezing, until the residual $u - f$ resembles Gaussian noise. Similarly, in [5] several possible applications for non-constant regularization are discussed.

In this article we study more general convex regularization functionals of the form

$$\Phi(u) = \int_a^b \phi\big(t, u(t) - f(t)\big)\, dt + \int_a^b \alpha(t)\, d|D^l u|\ . \tag{1}$$

We prove existence of a minimizer of Φ under reasonable conditions on the functions ϕ and α, and provide a unique characterization of the solution set.

Special emphasis is put on the case of first order regularization $l = 1$. There exists a highly efficient algorithm, the taut string algorithm, for the solution of discrete L^2-total variation regularization (see [2]). This algorithm is based on a dual formulation of the problem of minimizing Φ (see [4, 7, 13]). In this article we show that the basic ideas of the taut string algorithm can also be used for the minimization of (1). Additionally, we show that the discrete and the continuous formulation of the functional are equivalent provided the input data f and the function α are piecewise constant. Put in other words, the continuous solution of a discretized problem is equivalent to the fully discretized solution.

The structure of the article is as follows: After introducing the basic notation we prove the main result Thm. 1 characterizing the minimizers of Φ. This is basically achieved by considering a dual functional and simultaneously characterizing the solution u of Φ and the solution ρ of the dual problem. In Thm. 2 we show that, although neither u nor ρ have to unique, at every point $t \in \Omega$ at least one of them is uniquely determined.

In Section 4 we treat first order regularization, that is, $l = 1$. In this case, one can show that with any two solutions u_1, u_2 of the minimization problem Φ and corresponding dual functions ρ_1, ρ_2, also the functions $\max\{u_1, u_2\}$ and $\min\{u_1, u_2\}$ are solutions of Φ with corresponding duals $\max\{\rho_1, \rho_2\}$ and $\min\{\rho_1, \rho_2\}$, respectively (see Thm. 3). This is the main result needed for the definition of the taut string algorithm.

Finally, we show the equivalence of the discrete and the continuous formulations (see Thm. 5), and show how the basic idea of the taut string algorithm can be carried over to the case of arbitrary convex integrands ϕ (see Thm. 6 and Alg. 1).

2 Preliminaries

Let $\Omega = (a, b) \subset \mathbb{R}$ and let $l \in \mathbb{N}$. For $u \in L^1(\Omega)$ and $(c, d) \subset \Omega$ we define the l-th order variation of u on (c, d) as

$$|D^l u|(c, d) = \sup\left\{ \int_c^d \psi^{(l)}(t)\, u(t)\, dt : \psi \in \mathcal{C}_c^l(c, d),\ \|\psi\|_\infty \leq 1 \right\}.$$

Here, $\mathcal{C}_c^l(c, d)$ denotes the space of l-times continuously differentiable functions with support compactly contained in (c, d). The space $\mathrm{BV}^l(\Omega)$ consists of all functions $u \in L^1(\Omega)$ satisfying $|D^l u|(a, b) < \infty$.

It can be shown that for every function $u \in \mathrm{BV}^l(\Omega)$ there exists a signed Radon measure $D^l u$ such that

$$\int_a^b \psi^{(l)}(t)\, u(t)\, dt = (-1)^l \int_a^b \psi(t)\, dD^l u$$

for every $\psi \in \mathcal{C}_c^l(a, b)$. Moreover we have the equality $|D^l u| = (D^l u)^+ + (D^l u)^-$, where $D^l u = (D^l u)^+ - (D^l u)^-$ denotes the Jordan-Hahn decomposition of $D^l u$. If on the other hand μ is a Radon measure on Ω and p a polynomial of degree at most $l - 1$, then the sum of p and the l-fold integral of μ is an element of $\mathrm{BV}^l(\Omega)$.

Bounded sets in $\mathrm{BV}^l(\Omega)$ are sequentially compact in the sense that every sequence $\{u_k\}_{k \in \mathbb{N}} \subset \mathrm{BV}^l(\Omega)$ satisfying $\sup_k \|u_k\|_1 + |D^l u_k|(\Omega) < \infty$ has a subsequence $\{u_{k_j}\}_{j \in \mathbb{N}}$ converging in the L^1-norm to a function $u \in \mathrm{BV}^l(\Omega)$.

In this article we consider minimization of a functional

$$\Phi(u) = S(u) + R(u), \tag{2}$$

where

$$S(u) := \int_a^b \phi\big(t, u(t) - f(t)\big)\, dt, \qquad R(u) := \int_a^b \alpha(t)\, d|D^l u|. \tag{3}$$

We assume that ϕ and α satisfy the following conditions:

1. $\phi : \Omega \times \mathbb{R} \to [0, +\infty]$ is a normal and convex non-negative integrand (cf. [10]).
2. There exist $c_1 \in \mathbb{R}$ and $c_2 > 0$ such that

$$\phi(t, \xi) \geq c_1 + c_2\, |\xi| \qquad \text{for all } (t, \xi) \in \Omega \times \mathbb{R}.$$

3. $\alpha : \Omega \to (0, +\infty)$ is positive and lower semi-continuous.
4. There exists $C \geq 1$ such that

$$C^{-1} \leq \alpha(t) \leq C \qquad \text{for all } t \in \Omega.$$

Above assumptions in particular imply that the functional Φ is well-defined: First, since ϕ is normal and non-negative, it follows that the function $t \mapsto \phi\big(t, u(t) - f(t)\big)$ is integrable (with possibly infinite integral). Second, since α is bounded and lower semi-continuous, it follows that it is integrable with respect to any finite Radon measure on Ω, which in turn shows that $R(u)$ makes sense.

Additionally, the lower bounds on ϕ and α imply coercivity of Φ, whereas the normality of ϕ and lower-semicontinuity of α imply its lower semi-continuity in the L^1-norm. This proves the existence of a minimizer, provided Φ is proper, that is, there exists at least one function $u \in \mathrm{BV}^l(\Omega)$ such that $\Phi(u) < \infty$. In the following we will always assume that Φ is proper.

3 Characterization of the Minimizers and Uniqueness

In this section we will provide a characterization of the minimizers of the functional Φ.

By $\partial\phi(t, \xi)$ we denote the subgradient of ϕ with respect to ξ.

Definition 1. *Let $a \leq x < y \leq b$ and $u \in \mathrm{BV}^l(x, y)$. We define $\mathcal{Y}(u; x, y)$ as the set of all functions $\rho \in W^{l,1}(x, y)$ such that*

- *$(-1)^{l-1}\rho^{(l)}(t) \in \partial\phi\big(t, u(t) - f(t)\big)$ for almost every $t \in (x, y)$,*
- *$|\rho(t)| \leq \alpha(t)$ for every $t \in (x, y)$,*
- *$\rho(t) = +\alpha(t)$ for $(D^l u)^+$-almost every $t \in (x, y)$ and $\rho(t) = -\alpha(t)$ for $(D^l u)^-$-almost every $t \in (x, y)$.*

Note that $\mathcal{Y}(u; x, y)$ in general may be empty.

Theorem 1. *A function $u \in \mathrm{BV}^l(a, b)$ is a minimizer of Φ_α if and only if $\mathcal{Y}(u; a, b) \cap W_0^{l,1}(\Omega) \neq \emptyset$.*

Proof. Every minimizer u of Φ is characterized by the condition $0 \in \partial\Phi(u)$ or, equivalently, $-\partial R(u) \cap \partial S(u) \neq \emptyset$.

Let therefore $\psi \in -\partial R(u) \cap \partial S(u)$. Using [10, Cor. 3E] it follows that $\psi(t) \in \partial\phi\big(t, u(t) - f(t)\big)$ for almost every $t \in \Omega$. Now define

$$\rho(x) := (-1)^{l-1} \int_a^x \int_a^{t_l} \cdots \int_a^{t_2} \psi(t_1)\, dt_1\, \ldots\, dt_{l-1}\, dt_l\,,$$

Since by assumption $-\psi \in \partial R(u)$ we have

$$-\int_a^b \big(v(t) - u(t)\big)\psi(t)\, dt \leq \int_a^b \alpha(t)\, d\big(|D^l v| - |D^l u|\big) \tag{4}$$

for all $v \in \mathrm{BV}^l(a, b)$.

Let first $v = u + p$, where p is polynomial of degree at most $l-1$ on Ω. Then $D^l u = D^l v$, which implies that

$$0 \geq -\int_a^b \big(v(t) - u(t)\big)\psi(t)\,dt$$

$$= (-1)^l \int_a^b p(t)\rho^{(l)}(t)\,dt = \sum_{j=1}^l (-1)^{l-j}\rho^{(l-j)}(b)\,p^{(j-1)}(b)\,.$$

Since this inequality holds for every polynomial p of degree at most $l-1$, it follows that $\rho \in W_0^{l,1}(\Omega)$.

Using integration by parts it follows from inequality (4) that

$$\int_a^b \rho(t)\,dD^l v - \int_a^b \alpha(t)\,d|D^l v| \leq \int_a^b \rho(t)\,dD^l u - \int_a^b \alpha(t)\,d|D^l u| \qquad (5)$$

for all $v \in \mathrm{BV}^l(\Omega)$. Since the right hand side of (5) is finite and R is positively homogeneous, it follows that

$$\int_a^b \rho(t)\,dD^l v \leq \int_a^b \alpha(t)\,d|D^l v|$$

for all $v \in \mathrm{BV}^l(\Omega)$. Since every Radon measure may occur as l-th order total variation, this implies that $|\rho(t)| \leq \alpha(t)$ for all $t \in \Omega$.

As a consequence we obtain from (4) by choosing $v = 0$ that

$$\int_a^b \rho(t)\,dD^l u = \int_a^b \alpha(t)\,d|D^l u|\,. \qquad (6)$$

Since $|\rho(t)| \leq \alpha(t)$ for all $t \in (\Omega)$, this directly implies that the last item in Definition 1 is satisfied.

Now assume that $\rho \in \mathcal{Y}(u; a, b) \cap W_0^{l,1}(\Omega)$. Then $\psi := (-1)^{l-1}\rho^{(l)} \in \partial S(u)$. Thus it is sufficient to prove that $-\psi \in \partial R(u)$. We therefore have to show that (4) holds for every $v \in \mathrm{BV}^l(\Omega)$. Since by assumption $\rho \in W_0^{l,1}(\Omega)$, this is equivalent to (5). Since $\rho \in \mathcal{Y}(u; a, b)$, it follows that the right hand side of (5) is zero. Thus it is enough to show that

$$\int_a^b \rho(t)\,dD^l v \leq \int_a^b \alpha(t)\,d|D^l v|$$

for all $v \in \mathrm{BV}^l(\Omega)$. This inequality, however, follows from the fact that $|\rho(t)| \leq \alpha(t)$ for all $t \in (\Omega)$.

Remark 1. Since the function ρ is continuous, it follows that the last item in the definition of $\mathcal{Y}(u; x, y)$ is satisfied if and only if the function $u^{(l-1)}$ is monotoneously increasing in a neighbourhood of every point $x \in (a, b)$ with $\rho(x) > -\alpha(x)$, and monotoneously decreasing in a neighbourhood of every $x \in (a, b)$ with $\rho(x) < +\alpha(x)$.

In general neither the minimizer u of Φ nor the corresponding function $\rho \in \mathcal{Y}(u; a, b)$ have to be unique. The next result, however, states that locally at least one of them is unique regardless of differentiability or strict convexity of ϕ.

Theorem 2. *Let* u, $v \in \mathrm{BV}^l(x, y)$, *let* $\rho_u \in \mathcal{Y}(u; x, y)$ *and* $\rho_v \in \mathcal{Y}(v; x, y)$. *Assume moreover that* $\rho_u - \rho_v \in W_0^{l,1}(x, y)$. *Then for almost every* $t \in (x, y)$ *at least one of the equations* $u(t) = v(t)$ *or* $\rho_u(t) = \rho_v(t)$ *holds. If* ϕ *is differentiable, then* $\rho_u = \rho_v$. *If* ϕ *is strictly convex, then* $u = v$.

Proof. Denote $S^+ := \{t : \rho_u(t) > \rho_v(t)\}$ and $S^- := \{t : \rho_u(t) < \rho_v(t)\}$. Since $\|\rho_u\|_\infty \leq \alpha$ and $\|\rho_v\|_\infty \leq \alpha$, it follows that

$$\alpha(t) \geq \rho_u(t) > \rho_v(t) \geq -\alpha(t), \quad t \in S^+,$$
$$-\alpha(t) \leq \rho_u(t) < \rho_v(t) \leq +\alpha(t), \quad t \in S^-.$$

Since $\rho_u \in \mathcal{Y}(u; x, y)$ and $\rho_v \in \mathcal{Y}(v; x, y)$, it follows that $D^l(u - v) \llcorner S^+$ and $-D^l(u - v) \llcorner S^-$ are positive Radon measures. In particular,

$$\int_x^y \big(\rho_u(t) - \rho_v(t)\big)\, dD^l(u - v) \geq 0. \tag{7}$$

With integration by parts it follows that

$$\int_x^y (-1)^l \big(\rho_u^{(l)}(t) - \rho_v^{(l)}(t)\big)\big(u(t) - v(t)\big)\, dt \geq 0. \tag{8}$$

Now recall that $\rho_u^{(l)}(t) \in \partial\phi\big(t, u(t) - f(t)\big)$ and $\rho_v^{(l)} \in \partial\phi\big(t, v(t) - f(t)\big)$ for almost every t. Since ϕ is a convex integrand and thus its subgradient increasing, this in particular implies that $(-1)^{l-1}\rho_u^{(l)}(t) \geq (-1)^{l-1}\rho_v^{(l)}(t)$ whenever $u(t) > v(t)$, and $(-1)^{l-1}\rho_u^{(l)}(t) \leq (-1)^{l-1}\rho_v^{(l)}(t)$ whenever $u(t) < v(t)$. Thus, the integral in (8) is non-positive, and equals zero if and only if for almost every $t \in (x, y)$ either $u(t) = v(t)$ or $\rho_u^{(l)}(t) = \rho_v^{(l)}(t)$. Additionally it follows that the integral in (7) equals zero, which implies that $\rho_u(t) = \rho_v(t)$ for $|D^l(u - v)|$-almost every $t \in (x, y)$.

Now let $t \in (x, y)$ be such that $\rho_u(t) \neq \rho_v(t)$. Then there exists an open neighbourhood U of t such that $|D^l(u - v)|(U) = 0$, and therefore $u - v$ is a polynomial of degree at most $l - 1$ on U. In case $u(s) \neq v(s)$ for some $s \in U$, it follows that $u(s) \neq v(s)$ for almost every $s \in U$, and thus $\rho_u^{(l)} - \rho_v^{(l)} = 0$ on U, which implies that $\rho_u - \rho_v$ is a polynomial on U, too.

Now assume that $u(t) \neq v(t)$. Then there exists a maximal interval $(c, d) \subset (x, y)$ such that $\rho_u - \rho_v$ and $u - v$ are polynomials of degree at most $l - 1$ on (c, d). From the argumentation above and the maximality of (c, d) it follows that $\rho_u(c) = \rho_v(c)$ and $\rho_u(d) = \rho_v(d)$.

Since $\rho_u - \rho_v$ is a polynomial of degree at most $l - 1$ on (c, d), it follows that $\rho_u - \rho_v$ can have at most $l - 1$ zeros in $[c, d]$. Since $\rho_u - \rho_v \in W_0^{l,1}(x, y)$, this implies that $x < c < d < y$.

Now assume that $|\rho_u(c)| = \alpha(c)$. Since $\rho_u \in W^{l,1}(x,y)$ and $|\rho_u| \leq \alpha$ everywhere on (x,y), this implies that $\rho_u^{(k)}(c) = 0$ for all $1 \leq k \leq l-1$. Since additionally $\rho_v(c) = \rho_u(c)$, we obtain that $\rho_u^{(k)}(c) = \rho_v^{(k)}(c)$ for all $0 \leq k \leq l-1$, that is, c is an l-fold zero of $\rho_u - \rho_v$ on $[c,d]$, a contradiction to the fact that $\rho_u - \rho_v$ is a non-trivial polynomial of degree at most $l-1$. Thus $|\rho_u(c)| < \alpha$ and $|\rho_v(c)| < \alpha$.

This shows that $|D^l u|(c) = |D^l v|(c) = 0$, which shows that $u^{(l-1)}$ and $v^{(l-1)}$ are continuous at c. Since $u - v$ is a non-trivial polynomial of degree at most $l-1$ on (c,d), it follows that $u^{(k)}(c) \neq v^{(k)}(c)$ for some $0 \leq k \leq l-1$. Thus there exists $\varepsilon > 0$ such that $u(s) \neq v(s)$ for $s \in (c-\varepsilon, c)$. This implies that $\rho_u - \rho_v$ is a polynomial of degree at most $l-1$ on $(c-\varepsilon, d)$. Consequently, there exists $\delta > 0$ such that $\rho_u(s) \neq \rho_v(s)$ on $(c-\delta, c)$, which in turn shows that $u - v$ is a polynomial of degree at most $l-1$ on $(c-\delta, c)$. Since $D^l u(c) = D^l v(c) = 0$, it follows that in fact $u - v$ is a polynomial of degree at most $l-1$ on $(c-\delta, d)$, which is a contradiction to the maximality of the interval (c,d).

This proves that $u(t) = v(t)$ whenever $\rho_u(t) \neq \rho_v(t)$.

Now assume that ϕ is differentiable. Then the l-th order derivatives of ρ_u and ρ_v are $\rho_u^{(l)} = \phi'(u-f)$ and $\rho_v^{(l)} = \phi'(v-f)$. Since $u = v$ whenever $\rho_u^{(l)} \neq \rho_v^{(l)}$, this implies that $\rho_u^{(l)} = \rho_v^{(l)}$ almost everywhere. Since $\rho_u - \rho_v \in W_0^{l,1}(x,y)$, this shows that $\rho_u = \rho_v$.

Consider now the case where ϕ is strictly convex. Then the subgradient of ϕ is strictly increasing. Since $\rho_u^{(l)} \in \partial S(u)$ and $\rho_v^{(l)} \in \partial S(v)$, the equality $\rho_u^{(l)} = \rho_v^{(l)}$ therefore implies that $u = v$.

Remark 2. Every function $\rho \in W^{l,1}(x,y)$ has a unique continuous representative. Similarly, every $u \in \mathrm{BV}^l(x,y)$ has a unique representative that is continuous from the left (continuous in the case $l > 1$). Using these representatives, the almost everywhere in Thm. 2 can be replaced by everywhere, that is, for every $t \in (x,y)$ either $u(t) = v(t)$ or $\rho_u(t) = \rho_v(t)$.

4 First Order Regularization

In the following we consider the case $l = 1$, that is, $R(u) = \int_a^b \alpha(t)\, d|Du|$. We now adapt the definition of $\mathcal{Y}(u; x, y)$ to the first order regularization case. This reformulation allows us to define $\mathcal{Y}(u; x, y)$ also when u is an arbitrary function defined everywhere on (x,y) and taking values in $\bar{\mathbb{R}} := [-\infty, +\infty]$. This extension will be necessary for some results later on.

Definition 2. *Let $u : (x,y) \to \bar{\mathbb{R}}$. We define $\mathcal{Y}(u; x, y)$ as the set of all $\rho \in W^{1,1}(x,y)$ such that*

- $\rho'(t) \in \partial\phi\big(t, u(t) - f(t)\big)$ *for almost every $t \in (x,y)$,*
- $|\rho(t)| \leq \alpha(t)$ *for every $t \in (x,y)$,*
- *u is increasing on every interval $(c,d) \subset (x,y)$ where $\rho > -\alpha$, and decreasing on every interval (c,d) where $\rho < +\alpha$.*

In this definition we set $\partial\phi(+\infty) = \phi^\infty(1)$, and $\partial\phi(-\infty) = -\phi^\infty(-1)$, where $\phi^\infty : \mathbb{R} \to \bar{\mathbb{R}}$ denotes the recession function of ϕ, that is,

$$\phi^\infty(t) := \lim_{s \to +\infty} \frac{\phi(st)}{s} .$$

Theorem 3. *Let $u_i \in \mathrm{BV}(a, y)$, $i \in I$, for an arbitrary index set I, and assume that $\rho_i \in \mathcal{Y}(u_i; a, y)$ with $\rho_i(a) = 0$. Define*

$$\rho^+ := \sup_i \rho_i \quad u^+ := \sup_i u_i ,$$
$$\rho^- := \inf_i \rho_i \quad u^- := \inf_i u_i .$$

If $\rho^\pm \in W^{1,1}(a, y)$, then $\rho^+ \in \mathcal{Y}(u^+; a, y)$ and $\rho^- \in \mathcal{Y}(u^-; a, y)$.

Proof. It is sufficient to show that $\rho^+ \in \mathcal{Y}(u^+; a, y)$, the other inclusion then follows by symmetry.

Assume first that $I = \{1, 2\}$ consists of only two elements. Note that the condition $|\rho^+(t)| \leq \alpha(t)$ is trivially satisfied.

Denote

$$x := \max\{t \in [a, y] : \rho_1(t) = \rho_2(t)\} .$$

Without loss of generality assume that $\rho_1(t) > \rho_2(t)$ for every $t > x$.

In case $x < y$ it follows that for every $\varepsilon > 0$

$$\mathcal{L}^1\big(\{t \in (x, x + \varepsilon) : \rho_1'(t) > \rho_1'(t)\}\big) > 0 .$$

Since $\rho_1(t) \in \partial\phi\big(t, u_1(t) - f(t)\big)$ and $\rho_2(t) \in \partial\phi\big(t, u_2(t) - f(t)\big)$ for almost every t, and the subgradient of a convex function is increasing, this proves that for every $\varepsilon > 0$

$$\mathcal{L}^1\big(\{t \in (x, x + \varepsilon) : u_1(t) \geq u_2(t)\}\big) > 0 .$$

Since by assumption $\alpha(t) \geq \rho_1(t) > \rho_2(t) \geq -\alpha(t)$ for $t \in (x, y)$, it follows that u_1 is increasing and u_2 is decreasing on (x, y). Consequently $u_1 \geq u_2$ on (x, y).

Let $t \in (a, y)$ be such that $\rho^+(t) > -\alpha(t)$. We have to show that u^+ is increasing in a neighbourhood of t.

Assume first that $\rho_1(t) = \rho_2(t) > -\alpha(t)$. Then both u_1 and u_2 are increasing near t, which implies that also $u^+ = \max\{u_1, u_2\}$ is increasing near t.

Now assume that $\rho_1(t) \neq \rho_2(t)$. In case $t < x$ it follows from Thm. 2 that $u_1 = u_2$ in a neighbourhood of t. Since at least one of the functions ρ_i is larger than $-\alpha$ near t, this implies that $u^+ = u_1 = u_2$ are both increasing near t. If, on the other hand, $t > x$, then $\rho^+(t) = \rho_1(t)$ and $u^+(t) = u_1(t)$, which again implies that u^+ is increasing near t.

The proof that u^+ is decreasing in a neighbourhood of every point with $\rho^+ < \alpha$ is analogous.

It remains to show that $\rho^{+'}(t) \in \partial\phi\big(t, u^+(t) - f(t)\big)$ for almost every $t \in (x, y)$. This, however, is a consequence of Thm. 2 and the fact that $\rho^+ = \rho_1$ and $u^+ = u_1$ on (x, y).

Now assume that I is an arbitrary index set. Then there exists a sequence $\{i_j\}_{j \in \mathbb{N}} \subset I$ such that $\sup_j \rho_{i_j}(t) = \rho^+(t)$ and $\sup_j u_{i_j}(t) = u^+(t)$ for almost every t. Denote

$$\hat{\rho}_k := \max\{\rho_{i_1}, \ldots, \rho_{i_k}\}, \qquad \hat{u}_k := \max\{u_{i_1}, \ldots, u_{i_k}\}.$$

Then $\{\hat{\rho}_k\}_{k \in \mathbb{N}}$ and $\{\hat{u}_k\}_{k \in \mathbb{N}}$ are increasing sequences converging to ρ^+ and u^+, respectively. Since $\lim_{k \to \infty} \hat{\rho}_k = \rho^+ \in W^{1,1}(a, y)$, we additionally have that $\hat{\rho}'_k$ converges to $\rho^{+'}$ pointwise almost everywhere.

Note first that, since $|\hat{\rho}_k(t)| \leq \alpha(t)$ for every t and k, it follows that $|\rho^+(t)| \leq \alpha(t)$ for every t.

Since $\hat{\rho}_{k+1} = \max\{\hat{\rho}_k, \rho_{i_{k+1}}\}$ and $\hat{u}_{k+1} = \max\{\hat{u}_k, u_{i_{k+1}}\}$, it follows by induction that $\hat{\rho}_k \in \mathcal{Y}(u_k; a, y)$ for every $k \in \mathbb{N}$. In particular, $\hat{\rho}'_k(t) \in \partial\phi\big(t, \hat{u}_k(t) - f(t)\big)$ for almost every $t \in (a, y)$ and every $k \in \mathbb{N}$. This shows that also $\rho^{+'}(t) \in \partial\phi\big(t, u^+(t) - f(t)\big)$ for almost every t.

Now assume that $(c, d) \subset (a, y)$ is such that $\rho^+ > -\alpha$ on (c, d). Then also $\hat{\rho}_k > -\alpha$ on (c, d) for large enough k, which implies that $\hat{u}_k$ is increasing on (c, d) for k large enough. Thus u^+ is on (c, d) the limit of increasing functions and thus itself increasing. Similarly, u^+ is decreasing on every interval (c, d) with $\rho^+ < \alpha$, which proves that $\rho^+ \in \mathcal{Y}(u^+; a, y)$.

Theorem 4. *Assume that for almost every $t \in (a, b)$ zero is the unique minimizer of $\phi(t, \cdot)$. Let u be any minimizer of Φ. Let $x \in (a, b)$ be such that f and α are constant in a neighbourhood of x. Then also u is constant near x.*

Proof. Let $\rho \in \mathcal{Y}(u; a, b) \cap W_0^{1,1}(a, b)$. Assume without loss of generality that $\rho(x) > -\alpha(x)$. Then it follows from Thm. 1 that u is increasing in a neighbourhood of x. Thus it remains to show that u is also decreasing near x.

In case $\rho(x) < \alpha(x)$, this follows from Thm. 1. We may therefore assume without loss of generality that $\rho(x) = \alpha(x)$.

Let U be a neighbourhood of x such that u is increasing and f and α are constant on U. Since by assumption $\rho(x) = \alpha(x)$, it follows that for every $t \in U$ we have $\rho(t) \leq \rho(x)$. Thus for all $\varepsilon > 0$ we have

$$\mathcal{L}^1\big(\{t \in (x, x + \varepsilon) : \rho'_u(t) \leq 0\}\big) > 0.$$

Since 0 is the unique minimizer of ϕ and $\rho'_u(t) \in \partial\phi\big(t, u(t) - f(t)\big)$ for almost every t, it follows that $\rho'_u(t) > 0$ whenever $u(t) > f(t)$. Thus,

$$\mathcal{L}^1\big(\{t \in (x, x + \varepsilon) : u(t) \leq f(t)\}\big) > 0.$$

Since f is constant near x and u is increasing near x, this implies that $u \leq f$ on some interval $(x - \varepsilon_1, x + \varepsilon_1)$. One can show in a similar manner that $u \geq f$ on an interval $(x - \varepsilon_2, x + \varepsilon_2)$, which proves the assertion.

Corollary 1. *Let zero be the unique minimizer of $\phi(t, \cdot)$ for almost every t. Assume that there exist $a = x_0 < \ldots < x_n = b$ such that f and α are constant on each interval (x_{i-1}, x_i), $1 \leq i \leq n$. Then also every minimizer u of Φ is constant on each interval (x_{i-1}, x_i).*

5 Nonlinear Taut String Algorithm

In the following we assume that zero is the unique minimizer of the function $\phi(t, \cdot)$ for almost every $t \in \Omega$. Additionally we assume that for almost every $t \in \Omega$ the function $\phi(t, \cdot)$ is strictly convex and differentiable.

Let $a = x_0 < \ldots < x_n = b$ be a discretization of Ω. Let moreover $\mathbf{f} := (f_1, \ldots, f_n) \in \mathbb{R}^n$ and $\boldsymbol{\alpha} := (\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n$ satisfy $\alpha_i > 0$ for all i. Define the lower semi-continuous extension of $\boldsymbol{\alpha}$ to Ω:

$$
\alpha(t) := \begin{cases} \alpha_i, & \text{if } t \in (x_{i-1}, x_i)\,, \\ \min\{\alpha_{i+1}, \alpha_i\}, & \text{if } t = x_i\,. \end{cases}
$$

Define moreover $\tilde{\alpha}_i := \min\{\alpha_{i+1}, \alpha_i\} = \alpha(x_i)$ for $1 \leq i \leq n-1$.

We first consider the semidiscrete problem of minimizing the functional

$$
\widehat{\Phi}(u) = \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} \phi\big(t, u(t) - f_i\big)\, dt + \int_{\Omega} \alpha(t)\, d|Du|\,.
$$

Theorem 5. *Minimizing $\widehat{\Phi}$ is equivalent to minimizing the fully discrete functional*

$$
\Psi(\mathbf{u}) = \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} \phi\big(t, u_i - f_i\big)\, dt + \sum_{i=1}^{n-1} \tilde{\alpha}_i \,|u_{i+1} - u_i|
$$

over $\mathbb{R}^n$, that is, u is a minimizer of $\widehat{\Phi}$, if and only if $u(t) = u_i$ is constant on each (x_{i-1}, x_i) and $\mathbf{u} = (u_1, \ldots, u_n)$ is a minimizer of Ψ.

Proof. This is a direct consequence of Cor. 1.

This last result allows us to apply the results of the previous sections to the discrete functional Ψ.

For $\mathbf{v} = (v_1, \ldots, v_k) \in \bar{\mathbb{R}}^k$, $1 \leq k \leq n$, we define $\boldsymbol{\rho}^{\mathbf{v}} = (\rho_0^{\mathbf{v}}, \ldots, \rho_k^{\mathbf{v}}) \in \mathbb{R}^{k+1}$ setting

$$
\rho_i^{\mathbf{v}} := \sum_{j=1}^{i} \int_{x_{j-1}}^{x_j} \phi'(t, v_i - f_i)\, dt\,.
$$

For $1 \leq k \leq n$ we define $\mathbf{u}^{(\pm, k)} \in \bar{\mathbb{R}}^k$ setting

$$
\mathbf{u}^{(+,k)} = \sup\big\{ \mathbf{v} \in \mathbb{R}^k : |\rho_i^{\mathbf{v}}| \leq \tilde{\alpha}_i,\ v_{i+1} \geq v_i \text{ if } \rho_i^{\mathbf{v}} \leq \tilde{\alpha}_i,\ v_{i+1} \leq v_i \text{ if } \rho_i^{\mathbf{v}} \geq -\tilde{\alpha}_i \big\}\,,
$$

$$
\mathbf{u}^{(-,k)} = \inf\big\{ \mathbf{v} \in \mathbb{R}^k : |\rho_i^{\mathbf{v}}| \leq \tilde{\alpha}_i,\ v_{i+1} \geq v_i \text{ if } \rho_i^{\mathbf{v}} \leq \tilde{\alpha}_i,\ v_{i+1} \leq v_i \text{ if } \rho_i^{\mathbf{v}} \geq -\tilde{\alpha}_i \big\}\,.
$$

Here we define $\tilde{\alpha}_n := 0$.

Theorem 6. *The vectors $\mathbf{u}^{(\pm,k)}$ are well-defined and $\mathbf{u} := \mathbf{u}^{(+,n)} = \mathbf{u}^{(-,n)}$ is the unique minimizer of Ψ.*

For a fixed $1 \leq i \leq n$ the sequence $\{u_i^{(+,k)}\}_{k \geq i}$ is decreasing, $\{u_i^{(-,k)}\}_{k \geq i}$ is increasing, and $u_i^{(+,k)} \geq u_i^{(-,k)}$ for all $k \geq i$. If in particular $u_i^{(+,k)} = u_i^{(-,k)}$ for some $k \geq i$, then also $u_i = u_i^{(\pm,k)}$.

Proof. The first part is a consequence of Thms. 5, 3, and 1. The second part then directly follows from the definitions of $\mathbf{u}^{(\pm,k)}$.

From Thm. 2 it follows that whenever $u_i^{(+,k)} = u_i^{(-,k)}$ for some $i \leq k$, then also $u_j^{(\pm,k)} = u_j$ for all $j \leq i$. This shows that for every $1 \leq k \leq n$ there exists N_k such that $u_i^{(\pm,k)} = u_i$ for $i \leq N_k$, and $u_i^{(+,k)} > u_i^{(-,k)}$ for $i > N_k$.

Now denote $\rho^{(\pm,k)} := \rho^{\mathbf{u}^{(\pm,k)}}$. Since by assumption the function $\phi(t, \cdot)$ is strictly convex, it follows that $\rho_i^{(+,k)} > \rho_i^{(-,k)}$ for $i > N_k$, which in turn implies that $u_{i+1}^{(+,k)} \geq u_i^{(+,k)}$ and $u_{i+1}^{(-,k)} \leq u_i^{(-,k)}$ for every $i > N_k$.

These observations lead to the following algorithm.

Algorithm 1 (Nonlinear Taut String Algorithm)

1. *Initialize $N_0 = 0$, $\rho_0 = 0$ and $\mathbf{u}$ is an empty vector.*
2. *For $k = 1, \ldots, n$ do:*
   (a) *Let $\mathbf{v}^\pm \in \bar{\mathbb{R}}^{k-N_k}$ be the maximal (minimal) increasing (decreasing) vectors satisfying*

$$\sum_{i=N_{k-1}+1}^{k} \int_{x_{i-1}}^{x_i} \phi'(t, v_{i-N_{k-1}}^+ - f_i)\, dt \leq +\tilde{\alpha}_k - \rho_k,$$

$$\sum_{i=N_{k-1}+1}^{k} \int_{x_{i-1}}^{x_i} \phi'(t, v_{i-N_{k-1}}^- - f_i)\, dt \geq -\tilde{\alpha}_k - \rho_{k-1}.$$

   (b) *Let j be the maximal index such that $v_j^+ = v_j^-$. Set $j = 0$ if no such index exists.*
   (c) *Define $N_k := N_{k-1} + j$ and*

$$\rho_k := \rho_{k-1} + \sum_{i=N_{k-1}+1}^{N_k} \int_{x_{i-1}}^{x_i} \phi'(t, v_{i-N_{k-1}}^+ - f_i)\, dt.$$

   (d) *Append $(v_1^+, \ldots, v_j^+)$ to the vector $\mathbf{u}$.*
3. *The final vector $\mathbf{u}$ is the minimizer of Ψ.*

11

6 Conclusion

We have given a characterization of the minimizers of convex regularization functionals with normal and convex integrands as fidelity term and non-constant higher order total variation as penalty term. One direct consequence of this characterization is that, in the case of a constant regularization parameter $\alpha > 0$, the solution is locally either a polynomial of order $l - 1$ or equals the data.

In the case of first order regularization the corresponding results are stronger. Here we have shown that for piecewise constant data f and piecewise constant regularization parameter α, also the solution u is piecewise constant *with the same jump locations*. For higher order regularization this is not the case, as examples in [9] show. In particular, this result implies the equivalence of discrete and semi-discrete solution of the problem.

Finally, we have generalized the taut string algorithm for the solution of the discretized first order regularization problem. Although this algorithm is not as efficient as in the quadratic case (there it is of complexity $O(n)$), it still has some advantages. In particular the fact that the solution **u** is computed incrementally makes the algorithm suited for online smoothing of incoming data, and also eases a possible parallelization for the handling of large data sets.

The main open questions deal with higher order regularization. Assuming that f and α are splines of order $l - 1$, it is not clear that the minimizer u is a spline as well (that is, u has only finitely many nodal points), and the nodal points of u are not as easily determined as in the first order case. Moreover, the definition of a taut string like algorithm for the solution of higher order problems is completely open.

Acknowledgement

References

1. R. Acar and C.R. Vogel. Analysis of bounded variation penalty methods for ill–posed problems. *Inverse Probl.*, 10:1217–1229, 1994.
2. P. L. Davies and A. Kovac. Local extremes, runs, strings and multiresolution. *Ann. Statist.*, 29:1–65, 2001.
3. A. Dold and B. Eckmann, editors. *Nonlinear Operators and the Calculus of Variations, Bruxelles 1975*. Springer Verlag, Berlin, Heidelberg, New York, 1976.
4. M. Grasmair. The equivalence of the taut string algorithm and bv-regularization. *J. Math. Imaging Vis.*, 27(1):59–66, 2007.
5. M. Grasmair and A. Obereder. Generalizations of the taut string method. submitted, 2007.
6. R. Klette, R. Kozera, L. Noakes, and J. Weickert, editors. *Geometric Properties of Incomplete Data*. Springer Verlag, New York, 2005. Computational Imaging and Vision, Vol. 31.

7. E. Mammen and S. van de Geer. Locally adaptive regression splines. *Ann. Statist.*, 25:387–413, 1997.

8. A. Obereder, S. Osher, and O. Scherzer. On the use of dual norms in bounded variation type regularization. In *[6]*, pages 373–390, 2005.

9. C. Pöschl and O. Scherzer. Characterization of minimizers of convex regularization functionals. *Contemporary Mathematics*, 2007. accepted.

10. R. T. Rockafellar. Integral functionals, normal integrands and measurable selections. In *[3]*, pages 157–207, 1976.

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

12. O. Scherzer, W. Yin, and S. Osher. Slope and G-set characterization of set-valued functions and applications to non-differentiable optimization problems. *Comm. Math. Sci.*, 3:479–492, 2005.

13. G. Steidl, S. Didas, and J. Neumann. Relations between higher order tv regularization and support vector regression. pages 515–527, 2005. Lecture Notes in Computer Science 3459.