G-NORM PROPERTIES OF BOUNDED VARIATION REGULARIZATION

STANLEY OSHER * AND OTMAR SCHERZER †

Abstract. Recently Y. Meyer derived a characterization of the minimizer of the Rudin-Osher-Fatemi functional in a functional analytical framework. In statistics the discrete version of this functional is used to analyze one dimensional data and belongs to the class of nonparametric regression models. In this work we generalize the functional analytical results of Meyer and apply them to a class of regression models, such as quantile, robust, logistic regression, for the analysis of multi–dimensional data. The characterization of Y. Meyer and our generalization is based on G-norm properties of the data and the minimizer. A geometric point of view of regression minimization is provided.

1. Introduction. For given data $f : \mathbb{R}^n \to \mathbb{R}$ representing image intensity values, we consider *reconstruction (denoising/decomposition)* methods, where the reconstruction $u : \mathbb{R}^n \to R$ satisfies

$$u - f \in \mathcal{G}^s(\alpha) := \{v = \nabla\cdot\vec{v} : \||\vec{v}|_{l^s}\|_{L^\infty} \leq \alpha\} = \{v : \|v\|_{G^s} \leq \alpha\} .$$

Here

$$\|v\|_{G^s} := \inf\{\||\vec{v}|_{l^s}\|_{L^\infty} : v = \nabla\cdot\vec{v}\}$$

is called G-norm (cf. Meyer [11, p. 30]).

One motivation for this paper arises from the statistical literature where an equivalence relation between the minimizer of the discretized Rudin-Osher-Fatemi (ROF) model [19] (see also [13]) and the solution of the *taut-string* algorithm has been established (see Mammen & Geer [10] and Davies & Kovac [2]). The ROF-model consists in minimization of the functional

$$\mathcal{F}_{ROF}(u) := \frac{1}{2}\int (u-f)^2 + \alpha\|Du\| ,$$

where $\|Du\|$ denotes the total variation semi-norm of u and $\alpha > 0$. The minimizer is called the *bounded variation regularized* solution. The taut-string algorithm consists in finding a string of minimal length in a tube (with radius α) around the primitive of f. The differentiated string is the taut-string reconstruction and corresponds to the minimizer of the ROF-model. Generalizing these ideas to higher dimensions is complicated by the fact that there is no obvious analog to primitives in higher space dimensions. We overcome this difficulty by solving Laplace's equation with right hand side f (i.e. integrate twice), and differentiating. The tube with radius α around the derivative of the potential specifies all functions u which satisfy $\|u - f\|_{G^s} \leq \alpha$ (see also [21]). In this paper we show that the bounded variation regularized solutions (in any number of space dimensions) are contained in a tube of radius α. For several other regression models in statistics, such as robust, quantile, and logistic regression (reformulated in a Banach space setting for analyzing multi dimensional data) the tube property can be verified as well (cf. Subsection 3.1). Moreover, following [11] characterizations of minimizers of general regression models are derived. This work

*Mathematics Department, Los Angeles, CA 90095-1555, USA. (sjo@math.ucla.edu)

†Department of Computer Science, University of Innsbruck, Technikerstr. 25, A–6020 Innsbruck, Austria. (otmar.scherzer@uibk.ac.at)

provides some geometrical insight into the structure of bounded variation minimization and establishes a link between statistics and image processing. The results of this paper can be applied to limit the well–known *stair casing* effects in numerical minimization of the ROF-functional. We observe that this undesirable effect occurs if the regularized solution has contact with the tube. Using higher order discretization schemes in contact zones limits the effect of staircasing. This topic will be discussed in a forthcoming paper, together with numerical studies. The ideas of the present paper can also be applied to bounded variation regularization models designed for filtering of multiplicative noise (see e.g. [17, 20]), but that is not within the scope of this paper. We note that a variety of image denoising algorithms based on the G-norm have been developed recently (see e.g. [16, 23, 1, 8, 21]).

2. Prerequisites.

2.1. G-Norm. The set

$$G := \{v : \|v\|_{G^s} < \infty\}$$

associated with the norm $\|\cdot\|_{G^s}$ is a Banach space, which is the dual of the Sobolev space

$$\tilde{W}_0^{1,1} := \overline{\{w \in C_0^\infty := C_0^\infty(\mathbb{R}^n)\}}\,,$$

where the closure is taken with respect to the norm

$$\|w\|_{W^{1,1,r}} := \int |\nabla w|_{l^r} \text{ for } 1 \leq r \leq \infty$$

(see [11]). We note that $\tilde{W}_0^{1,1} = W_0^{1,1,r}$, independently of $1 \leq r \leq \infty$ and note that $\tilde{W}_0^{1,1}$ is *not* the standard Sobolev space $W_0^{1,1}$, where in its definition the closure of C_0^∞ is taken with respect to the norm

$$\|w\|_{W^{1,1}} = \|w\|_{L^1} + \|\nabla w\|_{L^1}\ .$$

In fact from the Gagliardo-Nirenberg-Sobolev inequality (cf. [6, Formula *, p 140]) it follows that

$$\|w\|_{L^{p_n}} \leq \int |\nabla w|_{l^r} \text{ for every } w \in \tilde{W}_0^{1,1}\,, \tag{2.1}$$

where

$$p_n := \frac{n}{n-1} \text{ for space dimension } n \geq 2 \text{ and } p_n := \infty \text{ for } n = 1\ . \tag{2.2}$$

Actually in [6] (2.1) is proven for $r = 2$. The proof of the general case is along the lines of the proof there: for a function $f \in C^1(\mathbb{R}^n)$ with compact support we have

$$f(x_1, \ldots, x_i, \ldots, x_n) = \int_{-\infty}^{x_i} \frac{\partial f}{\partial x_i}(x_1, \ldots, t_i, \ldots, x_n)\, dt_i\ .$$

This gives the estimate

$$|f(x)| \leq \int_{-\infty}^{\infty} \left|\frac{\partial f}{\partial x_i}(x_1, \ldots, t_i, \ldots, x_n)\right| dt_i \leq \int_{-\infty}^{\infty} |\nabla f|_{l^r}$$

(independent of r). The rest of the proof is analogous to the proof of [6, Theorem 1, p. 139].

In particular from (2.1) it follows that $\tilde{W}_0^{1,1} \subseteq L^{p_n}$.

For $v \in G$ there exists $\vec{v}$ such that

$$v = \nabla \cdot \vec{v} \text{ and } \|v\|_{G^s} = \||\vec{v}|_{l^s}\|_{L^\infty} \tag{2.3}$$

(cf. [11]) and consequently for any $w \in \tilde{W}_0^{1,1}$

$$\int vw = \int \nabla \cdot \vec{v}\, w = -\int \vec{v} \nabla w \leq \||\vec{v}|_{l^s}\|_{L^\infty} \|w\|_{W^{1,1,s_*}} ,$$

where $1/s_* + 1/s = 1$ with $1 \leq s_*, s \leq \infty$.

In the sequel we make use of the following lemma:

Lemma 2.1. *Assume that there exists $\alpha > 0$ such that for every $v \in C_0^\infty$*

$$\left| \int wv \right| \leq \alpha \int |\nabla v|_{l^{s_*}} \tag{2.4}$$

hold, then $\|w\|_{G^s} \leq \alpha$.

Proof. The linear operator

$$L : C_0^\infty \to \mathbb{R}, \quad v \to \int wv$$

can be extended to a linear bounded operator on $\tilde{W}_0^{1,1}$. Note that by (2.4) for a sequence $\{v_n\}_{n\in\mathbb{N}}$ converging to v, $\{Lv_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence and thus convergent with limit Lv. □

In order to guarantee that L is well defined, it is required that $Lv \in \mathbb{R}$ for any $v \in C_0^\infty$. This is satisfied for instance if $w \in L^p$ for some $1 \leq p \leq \infty$.

2.2. Functions of bounded variation.

Definition 2.2. *The space of* functions of bounded variation *(BV) consists of functions $u \in L^{p_n}$ satisfying* [1]

$$\|Du\|_{s_*} := \sup \left\{ \int u \nabla \cdot \vec{\varphi} \ : \vec{\varphi} \in C_0^1(\mathbb{R}^n; \mathbb{R}^n)\, , |\vec{\varphi}(x)|_{l^s} \leq 1 \text{ for } x \in \mathbb{R}^n \right\} < \infty \, .$$

Note that for $u \in \tilde{W}_0^{1,1}$

$$\|Du\|_{s_*} = \int |\nabla u|_{l^{s_*}} = \|u\|_{W^{1,1,s_*}} \, .$$

For more background on functions of bounded variation we refer to Evans & Gariepy [6].

In the sequel we frequently make use of the following results:

[1] The definition of BV differs from the standard definition where it is assumed that $u \in L^1$ and not in L^{p_n}.

Lemma 2.3. *Let $1 \le q < \infty$ and $h \in \mathtt{BV} \cap L^q$. Then there exists a sequence $\{h_l\}_{l\in\mathbb{N}}$ in C_0^∞ satisfying*

$$h_l \to h \text{ in } L^q \text{ and } \|Dh_l\|_{s_*} \to \|Dh\|_{s_*} \,.$$

Proof. The proof consists in a modification of the proof of [6, Theorem 2,p172 ff] taking into account the results on *mollification* in Section 4.2.1 of this book. For $m \in \mathbb{N}$ fixed, we follow [6, Theorem 2,p172 ff] and define the open spheres with radius $k+m$

$$B_k = B(0, k+m) \quad (k = 1, \dots) \,.$$

For $\varepsilon > 0$ we fix m such that

$$\|Dh\|_{s_*}(\mathbb{R}^n \backslash B_1) = \int_{\mathbb{R}^n \backslash B_1} |\nabla h|_{s_*} < \varepsilon \text{ and } \|h\|_{L^q(\mathbb{R}^n \backslash B_1)} < \varepsilon \,. \tag{2.5}$$

Set $B_0 = \emptyset$ and define $V_k = B_{k+1} \backslash \overline{B}_{k-1}$, $k = 1, \dots$ Let $\{\zeta_k\}_{k=1,\dots}$ be a sequence of smooth functions satisfying

$$\zeta_k \in C_0^\infty(V_k) \quad 0 \le \zeta_k \le 1 \quad (k = 1, \dots) \quad \sum_{k=1}^\infty \zeta_k = 1 \text{ in } \mathbb{R}^n \,.$$

Fix a mollifier η (as in [6]) and for each k select $\varepsilon_k > 0$ such that

$$\begin{aligned} supp(\eta_{\varepsilon_k} * (h\zeta_k)) &\subseteq V_k \\ \int |\eta_{\varepsilon_k} * (h\zeta_k) - h\zeta_k|^q \, dx &< \frac{\varepsilon}{2^k} \\ \int |\eta_{\varepsilon_k} * (h\nabla\zeta_k) - h\nabla\zeta_k| \, dx &< \frac{\varepsilon}{2^k} \,. \end{aligned} \tag{2.6}$$

Up to now, the only difference from the proof in [6, Theorem 2,p172 ff] is the second item. Since we assume that $h \in L^q$ this item follows from properties of the mollifier η [6, Theorem 1,p122 ff]. Define

$$h_\varepsilon := \sum_{k=1}^\infty \eta_{\varepsilon_k} * (h\zeta_k) \in C^\infty \text{ and } h_\varepsilon^{(N)} := \sum_{k=1}^{(N)} \eta_{\varepsilon_k} * (h\zeta_k) \in C_0^\infty \,.$$

To prove the assertion, taking into account the properties of h_ε in [6, Theorem 2,p172 ff] it is sufficient to prove that

$$\|h_\varepsilon^{(2)} - h_\varepsilon\|_{L^q} \le K\varepsilon \text{ and } \left| \int (h_\varepsilon - h_\varepsilon^{(2)}) \nabla \cdot \vec{\phi} \right| \le K\varepsilon$$

for any $\vec{\phi} \in C_0^\infty$ with $|\vec{\phi}(x)|_{l^s} \le 1$. Using that $\eta_{\varepsilon_k} * (h\zeta_k)$ has support in V_k and V_k

has only non-empty intersection with V_{k-1} and V_{k+1} we find that

$$\begin{aligned}
\|h_\varepsilon^{(2)} - h_\varepsilon\|_{L^q}^q &= \int \left| \sum_{k=3}^\infty \eta_{\varepsilon_k} * (h\zeta_k) \right|^q \\
&= \sum_{l=3}^\infty \int_{V_l} \left| \sum_{k=l-1}^{l+1} \eta_{\varepsilon_k} * (h\zeta_k) \right|^q \\
&\le 3^q \sum_{l=3}^\infty \int_{V_l} \max_{l-1\le k\le l+1} |\eta_{\varepsilon_k} * (h\zeta_k)|^q \\
&\le 3^q \sum_{l=3}^\infty \left[\left(\int_{V_l} + \int_{V_{l-1}} + \int_{V_{l+1}} \right) |\eta_{\varepsilon_k} * (h\zeta_k)|^q \right] \\
&\le 3^{q+1} \sum_{l=2}^\infty \int_{V_l} |\eta_{\varepsilon_k} * (h\zeta_k)|^q \\
&\le 2^q 3^{q+1} \sum_{l=2}^\infty \left[\int_{V_l} |h\zeta_k|^q + |h\zeta_k - \eta_{\varepsilon_k} * (h\zeta_k)|^q \right] \\
&\le 2^q 3^{q+1} \int_{\mathbb{R}^n \setminus B_1} |h|^q + 2^q 3^{q+1} \varepsilon \le 6^{q+1} \varepsilon \;.
\end{aligned}$$

Moreover, proceeding as in [6, Item 4, Theorem 2, p174] we find that

$$\begin{aligned}
\left| \int (h_\varepsilon - h_\varepsilon^{(2)}) \nabla \cdot \vec{\phi} \right| &= \left| \sum_{k=3}^\infty \int \Big(\eta_{\varepsilon_k} * (h\zeta_k) \Big) \nabla \cdot \vec{\phi} \right| \\
&\le \left| \sum_{k=3}^\infty \int h \nabla \cdot \Big(\zeta_k (\eta_{\varepsilon_k} * \vec{\phi}) \Big) \right| \\
&\quad + \left| \sum_{k=3}^\infty \int \vec{\phi} \Big(\eta_{\varepsilon_k} * (h \nabla \zeta_k) - h \nabla \zeta_k \Big) \right| \\
&\le \|Dh\|_r (\mathbb{R}^n \setminus B_1) + \varepsilon \sum_{k=3}^\infty \frac{1}{2^k} \\
&\le 2\varepsilon \;.
\end{aligned}$$

This shows the assertion □

Corollary 2.4.

1. Assume $n = 1$ and $1 \le q < \infty$. Then for every $h \in \mathtt{BV} \cap L^q$

$$\|h\|_{L^{p_n}} \le \|Dh\|_{s_*} \;. \tag{2.7}$$

2. Assume $n \ge 2$. Then for every $h \in \mathtt{BV}$ (2.7) holds.

Proof.

1. From Lemma 2.3 it follows that there exists a sequence $\{h_l\}_{l \in \mathbb{N}}$ in C_0^∞ satisfying

$$h_l \to h \text{ in } L^q \text{ and } \|Dh_l\|_{s_*} \to \|Dh\|_{s_*} \;.$$

From the Gagliardo-Nirenberg-Sobolev inequality it follows that

$$\|h_l\|_{L^\infty} \le \int |\nabla h_l|_{l^{s_*}} = \|Dh_l\|_{s_*} \;.$$

Since L^∞ is isomorphic to the dual of L^1 and C_0^∞ is dense in L^1 we have

$$\begin{aligned} \|h\|_{L^\infty} &= \sup_{\{v\in L^1:\|v\|_{L^1}\le 1\}} \int vh \\ &= \sup_{\{v\in C_0^\infty:\|v\|_{L^1}\le 1\}} \int vh \\ &= \sup_{\{v\in C_0^\infty:\|v\|_{L^1}\le 1\}} \lim_{l\to\infty} \int vh_l \\ &\le \liminf_{l\to\infty} \|h_l\|_{L^\infty} \\ &\le \lim \|Dh_l\|_{s_*} \\ &= \|Dh\|_{s_*}. \end{aligned}$$

Note, that in order to prove the third identity we have used the fact that any function $v \in C_0^\infty$ is in L^{q_*}.

2. For $n \ge 2$, since $p_n < \infty$ the proof follows with $q = p_n$ by using the Gagliardo-Nirenberg-Sobolev inequality.

□

LEMMA 2.5. *Assume $1 \le q < \infty$ and $w \in L^{q_*}$ with $\|w\|_{G_s} \le \alpha$, where $1/q_* + 1/q = 1$ (for $q = 1$, $q_* = \infty$). Then for any $h \in \mathtt{BV} \cap L^q$*

$$\left|\int wh\right| \le \alpha \|Dh\|_{s_*} \,. \tag{2.8}$$

Proof. For $h \in C_0^\infty$ and $w = \nabla\cdot\vec{w}$ satisfying $\||\vec{w}|_{l^s}\|_{L^\infty} = \|w\|_{G_s}$ it follows that

$$\left|\int wh\right| = \left|\int \nabla\cdot\vec{w}h\right| = \left|\int \vec{w}\nabla h\right| \le \|w\|_{G^s}\|Dh\|_{s_*} \,. \tag{2.9}$$

Let $h \in \mathtt{BV} \cap L^q$, then there exists a sequence $\{h_l\}_{l\in\mathbb{N}}$ in C_0^∞ such that $h_l \to h$ in L^q and $\|Dh_l\|_{s_*} \to \|Dh\|_{s_*}$. Consequently, from (2.9) it follows that

$$\begin{aligned} \left|\int wh\right| &= \lim_{l\to\infty} \left|\int wh_l\right| \\ &\le \liminf_{l\to\infty} \|w\|_{G^s}\|Dh_l\|_{s_*} \\ &= \|w\|_{G^s}\|Dh\|_{s_*} \,. \end{aligned}$$

□

3. Generalized Rudin-Osher-Fatemi Model. Total variation minimization was introduced into image restoration in [19, 18] and for multiplicative denoising/deblurring in [17]. The original model consists in minimization of the following functional

$$\mathcal{F}_{ROF}(u) := \frac{1}{2}\int (u-f)^2 + \alpha\|Du\|_{s_*} \text{ with } s_* = 2 \,.$$

More general equivalent norms, including the case $s_* \ne 2$, have been considered in [15]. The unique minimizer u_α of this functional is called *bounded variation regularized solution*. In this work we consider minimization of functionals

$$\mathcal{F}_S(u) := \int S(x, u(x))\, dx \;+\; \alpha\|Du\|_{s_*} \tag{3.1}$$

over $\mathtt{BV}$. We denote by

$$X_S := \{u \in \mathtt{BV} : \mathcal{F}_S(u) < \infty\}$$

the domain of definition of the operator $\mathcal{F}_S$.

Several *regression models* from *statistics* can be embedded in this context (see e.g. [4]). We impose the following assumption on S:

ASSUMPTION 3.1. *$S : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}_{\geq 0}$ satisfies*

1. *$S(x, \cdot)$ is convex for almost every x,*
2. *$S(\cdot, v)$ is measurable for each $v \in \mathbb{R}$, and*
3. *$S(x, 0) \in L^1$.*

Note that Item 3 guarantees that $X_S \neq \emptyset$ and thus $\mathcal{F}_S$ is proper. Since a convex function is continuous and Items 1, 2 are satisfied S is a *Caratheodory function* (see [5, p234]) and therefore *normal* (cf. [5, p234, Proposition 1.1]). If S is normal and non–negative, the functional

$$\mathcal{S} : u \to \int S(x, u(x))\, dx$$

is lower semi-continuous on L^β for every $1 \leq \beta \leq \infty$ (cf. [5, Corollary 1.2, p.239]). By [5, Corollary 2.2, p.11] every convex, lower semi-continuous operator is weakly lower semi continuous on L^β. This is called the *compensated compactness theorem.*

Occasionally we impose the additional assumption on S that there exist $\underline{c} > 0$, $1 \leq p_0 < \infty$ and $k \in L^1$ such that

$$\underline{c}|v|^{p_0} - k(x) \leq S(x, v)\ . \tag{3.2}$$

THEOREM 3.2. *Let $S : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}_{\geq 0}$ satisfy Assumption 3.1.*

1. *Assume $n \geq 2$. Then there exists a minimizer $u_\alpha \in \mathtt{BV}$ of $\mathcal{F}_S$.*
2. *Let additionally* (3.2) *be satisfied, then $u_\alpha \in \mathtt{BV} \cap L^p$ for all $p \in [p_{\min} := \min\{p_0, p_n\}, p_{\max} := \max\{p_0, p_n\}]$.*

Proof. To prove that the functional $\mathcal{F}_S$ attains a minimizer, we take a sequence $\{u_l\}_{l \in \mathbb{N}}$ in $\mathtt{BV}$ satisfying

$$\mathcal{F}_S(u_l) \to \inf \mathcal{F}_S\ .$$

From Assumption 3.1 it follows that

$$0 \leq \inf \mathcal{F}_S \leq \mathcal{F}_S(0) = \int S(x, 0)\, dx\ < \infty\,,$$

and therefore $\{\|Du_l\|_{s_*}\}_{l \in \mathbb{N}}$ is uniformly bounded.

1. We consider the case $n \geq 2$. Since the embedding of $\mathtt{BV}$ into L^{p_n} is bounded (cf. Corollary 2.4), $\{u_l\}_{l \in \mathbb{N}}$ is uniformly bounded in L^{p_n} as well. Then there exists a weakly convergent subsequence, which is again denoted by $\{u_l\}_{l \in \mathbb{N}}$ and the weak limit is denoted by u. Since S is convex with respect to the second component, it follows from the compensated compactness theorem that

$$\mathcal{S}(u) \leq \liminf \mathcal{S}(u_l) < \infty\ . \tag{3.3}$$

   The weak lower semi-continuity of the $\mathtt{BV}$-semi-norm gives $\|Du\|_{s_*} \leq \liminf \|Du_l\|_{s_*}$. This, together with (3.3), shows that $\mathcal{F}(u) = \inf \mathcal{F}_S$. Or in other words $u_\alpha = u$.

2. From (3.2) we additionally find that

$$\mathcal{F}_S(u_l) \geq \underline{c} \int |u_l|^{p_0} \; - \; \int k \; + \; \alpha \|Du_l\|_{s_*} \,,$$

showing that $\int |u_l|^{p_0}$ is uniformly bounded as well. From the compensated compactness theorem it follows that $u_\alpha \in \mathtt{BV} \cap L^{p_n} \cap L^{p_0}$.

□

EXAMPLE 3.3.

1. *Let $f \in L^1$. For $S(x,v) = |v - f(x)|$ the functional $\mathcal{F}_S$ attains a minimizer in $X_S = L^1 \cap \mathtt{BV}$. This model is called* robust regression.
2. *Let $f \in L^1$. For* quantile regression *(see [9]) we have*

$$S(x,v) := \begin{cases} (1-\beta)(f(x)-v) & \text{for} \quad f(x) \geq v \\ \beta(v - f(x)) & \text{for} \quad f(x) \leq v \end{cases}$$

where $0 < \beta < 1$. The associated functional $\mathcal{F}_S$ attains a minimizer in $X_S = L^1 \cap \mathtt{BV}$.

3. *Let $f \in L^\infty$ satisfy $0 \leq f \leq 1$. For $\beta \geq 0$ and an open bounded set B the* generalized logistic regression *is*

$$S(x,v) = \frac{\beta}{2}(v - f(x))^2 + \begin{cases} \ln(1 + \exp(v)) - vf(x) & \text{for } x \in B\,, \\ 0 & \text{else.} \end{cases}$$

$\beta = 0$ is the standard logistic regression; actually the restriction to the bounded set B is never considered explicitly in statistical papers, but has to be implemented in a functional analytical context. S is non-negative; $S(x,\cdot)$ is convex for almost all x; $S(\cdot,v)$ is measurable for each $v \in \mathbb{R}$; and $S(\cdot,0) = ln(2)\chi_B + \frac{\beta}{2}f^2 \in L^1$, if $\beta = 0$ or $\beta > 0$ and $f \in L^2$. According to Theorem 3.2, the associated functional $\mathcal{F}_S$ attains a minimizer in
   (a) *$X_S \subseteq \mathtt{BV}$ if $\beta = 0$ and $n \geq 2$;*
   (b) *$X_S \subseteq \mathtt{BV} \cap L^2$ if $\beta > 0$ (for $n = 1, 2, \ldots$).*
4. *Let $f \in L^2$. For $S(x,v) = \frac{1}{2}|v - f(x)|^2$, the Rudin-Osher-Fatemi functional $\mathcal{F}_S$ attains a minimizer in $X_S = L^2 \cap \mathtt{BV}$.*

We summarize the results on smoothness of the regression models in the following table:

Method	n	$u_\alpha \in L^p$
Robust Regression	1	$1 \leq p \leq \infty$
+	2	$1 \leq p \leq 2$
Quantile Regression	3	$1 \leq p \leq 3/2$
Logistic Regression	1	$(2 \leq p \leq \infty)$, $p = \infty$
$\beta > 0$	2	$p = 2$,
$\beta = 0, n \geq 2$	3	$3/2 \leq p \leq 2$ $(p = 3/2)$
ROF-Regularization	1	$2 \leq p \leq \infty$
	2	$p = 2$
	3	$3/2 \leq p \leq 2$

For the ROF-regularization and generalized logistic regression model ($\beta > 0$) the functional $\mathcal{F}_S$ is *strictly convex* and thus there exists a unique minimizer. For robust and quantile regression the according functional $\mathcal{F}_S$ is convex and thus in general there *cannot* be expected a unique minimizer.

3.1. Characterization of Minimizers. Since S is convex with respect to v

$$S_v(x,\cdot) := \frac{\partial S}{\partial v}(x,\cdot) \in \mathtt{BV_{loc}} \text{ for almost every } x\ .$$

In the following we differ between the two cases, that S_v is either continuous or it is not. The later situation is more involved since the function S_v has to be considered set-valued.

3.2. S_v is continuous. In order to derive characterizations for the minimizers of $\mathcal{F}_S$ we impose the following assumptions on S:

ASSUMPTION 3.4.

1. *S satisfies Assumption 3.1.*
2. *For some $1 \le q < \infty$, $C_0^\infty \subseteq X_S \subseteq L^q$.*

We assume that S_v satisfies

ASSUMPTION 3.5.

1. *$S_v(x,\cdot)$ is continuous for almost every x,*
2. *$S_v(\cdot,v)$ is measurable for each $v \in \mathbb{R}$, and*
3. *for every $\psi \in X_S$, $S_v(\cdot,\psi) \in L^{q_*}$.*

Again, the first two items guarantee that S_v is normal.

THEOREM 3.6. *Let S, S_v satisfy Assumptions 3.4, 3.5, respectively. Moreover, we assume that for every $v, h \in X_S$*

$$\eta(t,v,h) := \int \{S(x,v(x)+th(x)) - S(x,v(x)) - tS_v(x,v(x))h(x)\}\ dx$$

satisfies

$$\frac{\eta(t,v,h)}{t} \to 0 \text{ as } t \to 0\ . \tag{3.4}$$

Then, $\|S_v(\cdot,0)\|_{G^s} \le \alpha$ if and only if u_α is zero.

Proof. In the first part we derive general properties of u_α which are used later on as well.

1. From the definition of a minimizer u_α of $\mathcal{F}_S$ it follows that for every $h \in X_S$, $\varepsilon \neq 0$

$$\begin{aligned} &\int S(x,u_\alpha(x))\,dx \;+\; \alpha\|Du_\alpha\|_{s_*} \\ \le &\int S(x,u_\alpha(x)+\varepsilon h(x))\,dx \;+\; \alpha\|D(u_\alpha+\varepsilon h)\|_{s_*} \\ \le &\int S(x,u_\alpha(x)+\varepsilon h(x))\,dx \;+\; \alpha\left(\|Du_\alpha\|_{s_*} + |\varepsilon|\|Dh\|_{s_*}\right) \\ \le &\int \{S(x,u_\alpha(x)) + \varepsilon S_v(x,u_\alpha(x))h(x)\}\ dx \;+\; \eta(\varepsilon,u_\alpha,h) \\ &+\alpha\left(\|Du_\alpha\|_{s_*} + |\varepsilon|\|Dh\|_{s_*}\right)\ . \end{aligned} \tag{3.5}$$

Consequently, it follows by dividing the terms in the inequality by $|\varepsilon|$ and taking $\varepsilon \to 0^\pm$ afterwards that

$$\left|\int S_v(x,u_\alpha)h(x)\,dx\right| \le \alpha\|Dh\|_{s_*} \text{ for } h \in X_S\ . \tag{3.6}$$

2. The definition of u_α implies that $u_\alpha \equiv 0$ if and only if for every function $h \in \mathtt{BV}$

$$\mathcal{F}_S(h) \geq \mathcal{F}_S(0) = \int S(x,0)\,dx\ . \tag{3.7}$$

If $u_\alpha \equiv 0$, then, from (3.6) it follows that for every $h \in X_S$

$$\left| \int S_v(x,0)h(x)\,dx \right| \leq \alpha \|Dh\|_{s_*}\ . \tag{3.8}$$

Since $S(x,\cdot)$ is convex and $S_v(x,\cdot)$ is continuous, for every $v, h \in \mathbb{R}$ and almost every x we have

$$S(x, v+h) - S(x,v) - S_v(x,v)h \geq 0\ . \tag{3.9}$$

Therefore, from (3.8) it follows that for every $h \in X_S$

$$\begin{aligned} &\int (S(x,h(x)) - S(x,0))\,dx \ + \alpha\|Dh\|_{s_*} \\ \geq &\int S_v(x,0)h(x)\,dx \ + \alpha\|Dh\|_{s_*} \\ \geq &0\ . \end{aligned}$$

Since $\mathcal{F}_S(h) = \infty$ for $h \notin X_S$ and $\mathcal{F}_S(0) < \infty$, we can write

$$\int (S(x,h(x)) - S(x,0))\,dx \ + \alpha\|Dh\|_{s_*} \geq 0 \text{ for every } h \in \mathtt{BV}\ .$$

In summary, we have shown that $u_\alpha \equiv 0$ if and only if (3.8) is satisfied for any $h \in X_S$.

3. Let $u_\alpha \equiv 0$, then from (3.8), the assumption $S_v(\cdot,0) \in L^{q_*}$, and Lemma 2.1 it follows that $\|S_v(\cdot,0)\|_{G^s} \leq \alpha$.
4. Let $\|S_v(\cdot,0)\|_{G^s} \leq \alpha$. Since by assumption $S_v(\cdot,0) \in L^{q_*}$ it follows from Lemma 2.5 that for any $h \in X_S \subseteq L^q$

$$\int S_v(x,0)h(x)\,dx \leq \leq \alpha\|Dh\|_{s_*}\ .$$

Thus (3.8) holds for any $h \in X_S$. Consequently $u_\alpha \equiv 0$.

□

We note that if (3.4) holds, then $S_v(\cdot, v(\cdot)h(\cdot)$ is the *directional derivative* of S at $v \in X_S$ in direction $h \in X_S$. For instance if the function S is twice Fréchet-differentiable with respect to v, with uniformly bounded second derivative, then we have $\eta(t,v,h) \leq t^2h^2$ and thus (3.4) holds if $X_S \subseteq L^2$. This is for instance utilized in Example 3.8 below.

THEOREM 3.7. *Let S, S_v satisfy Assumptions 3.4, 3.5, respectively. Moreover, we assume that $\|S_v(.,0)\|_{G^s} > \alpha$.*

Then $u = u_\alpha$ minimizes $\mathcal{F}_S$ if and only if

1. $u \in X_S$,

2.

$$\|S_v(\cdot, u(\cdot))\|_{G^s} = \alpha\,, \tag{3.10}$$

3. and

$$-\int S_v(x,u(x))u(x)\,dx = \alpha\|Du\|_{s_*} . \tag{3.11}$$

Proof. From the assumption $\|S_v(.,0)\|_{G^s} > \alpha$ it follows from Theorem 3.6 that $u_\alpha \neq 0$. From Theorem 3.2 it is evident that $u_\alpha \in X_S$.

1. From the definition of a minimizer u_α of $\mathcal{F}_S$ it follows that for every $0 \neq |\varepsilon| < 1$

$$\begin{aligned}&\int S(x,u_\alpha(x))\,dx \;+\; \alpha\|Du_\alpha\|_{s_*}\\ \leq&\int S(x,(1+\varepsilon)u_\alpha(x))\,dx \;+\; \alpha(1+\varepsilon)\|Du_\alpha\|_{s_*}\\ \leq&\int \{S(x,u_\alpha(x))+\varepsilon S_v(x,u_\alpha(x))u_\alpha(x)\}\,dx \;+\; \eta(\varepsilon,u_\alpha,u_\alpha)\\ &+\alpha(1+\varepsilon)\|Du_\alpha\|_{s_*} .\end{aligned}$$

Showing that

$$-\varepsilon\int S_v(x,u_\alpha(x))u_\alpha(x) \leq \alpha\varepsilon\|Du_\alpha\|_{s_*} + \eta(\varepsilon,u_\alpha,u_\alpha) .$$

Dividing the inequality by $|\varepsilon|$ and taking $\varepsilon \to 0\pm$ shows (3.11). Since $\|Du_\alpha\|_{s_*} \neq 0$, it follows from (3.6) that $\|S_v(\cdot,u_\alpha(\cdot))\|_{G^s} = \alpha$.

2. To prove the converse direction we note that for $u \in X_S$ satisfying (3.10) it follows from Lemma 2.5 that for any function $h \in X_S$

$$\|D(u+h)\|_{s_*} \geq -\frac{1}{\alpha}\int (u(x)+h(x))S_v(x,u(x))\,dx . \tag{3.12}$$

From (3.9), (3.12), and (3.11) it follows that for any function $h \in X_S$

$$\begin{aligned}&\int S(x,u(x)+h(x))\,dx \;+\; \alpha\|D(u+h)\|_{s_*}\\ \geq&\int S(x,u(x))\,dx \;+\int S_v(x,u(x))h(x)\,dx\\ &-\int (u(x)+h(x))S_v(x,u(x))\,dx\\ =&\int S(x,u(x))\,dx + \alpha\|Du\|_{s_*} .\end{aligned}$$

For $h \notin X_S$, we have $\mathcal{F}(u+h) = \infty$ and $\mathcal{F}(u) < \infty$, which finally shows that for all $h \in \mathtt{BV}$

$$\int S(x,u(x)+h(x))\,dx \;+\; \alpha\|D(u+h)\|_{s_*} \geq \int S(x,u(x))\,dx + \alpha\|Du\|_{s_*} ,$$

and u is a global minimizer.

□

EXAMPLE 3.8.

- *We consider the Rudin-Osher-Fatemi model.* $S : (x,v) \mapsto \frac{1}{2}(v-f(x))^2$ *satisfies:*

1. $S(x,\cdot)$ *is convex with respect to* v *for almost every* x,
2. $S(\cdot,v)$ *is measurable for each* $v \in \mathbb{R}$,
3. $S(x,0) = \frac{1}{2}f^2(x) \in L^1$ *if* $f \in L^2$ *and*
4. $$\frac{v^2}{4} - \frac{f^2(x)}{2} \le S(x,v) \le v^2 + f^2(x)\ .$$

   Therefore (3.2), *with* $p_0 = 2$ *is satisfied, if* $f \in L^2$.
5. $C_0^\infty \subseteq X_S = L^2 \cap \mathtt{BV} \subset L^2$.

$S_v(x,v) = v - f(x)$ *satisfies:*

1. $S_v(x,\cdot)$ *is continuous for almost every* x,
2. $S_v(\cdot,v)$ *is measurable for each* $v \in \mathbb{R}$.
3. *For* $\psi \in X_S$, $S_v(\cdot,\psi(\cdot)) = \psi(\cdot) - f(\cdot) \in L^2$.
4. *For* $v,h \in X_S$ *we have*
   $$\eta(t,v,h) = |t|^2 \int |h|^2\,,$$

   and thus (3.4) *holds.*

From Theorem 3.2 it follows that $u_\alpha \in X_S = \mathtt{BV} \cap L^2$. *Theorem 3.6 shows that* $u_\alpha \equiv 0$ *if and only if* $\|f\|_{G^s} \le \alpha$. *Theorem 3.7 shows that for* $\|f\|_{G^s} > \alpha$, u_α *is characterized by* $\|u_\alpha - f\|_{G^s} = \alpha$ *and* $\int (f-u_\alpha)u_\alpha = \alpha \|Du_\alpha\|_{s_*}$. *For* $s_* = 2$ *this result is stated in [11].*

- *For logistic regression with* $\beta > 0$, *let* $f \in L^\infty \cap L^2$ *satisfy* $0 \le f \le 1$.

$$S(x,v) = \frac{\beta}{2}(v - f(x))^2 + \begin{cases} \ln(1+\exp(v)) - vf(x) & \text{for } x \in B\,, \\ 0 & \text{else} \end{cases}$$

satisfies:

1. $S(x,v) \ge 0$,
2. $S(x,\cdot)$ *is convex for almost every* x,
3. $S(\cdot,v)$ *is measurable for each* $v \in \mathbb{R}$ *and*
4. $S(\cdot,0) = \frac{\beta}{2}f^2(\cdot) + \ln(2)\chi_B(\cdot) \in L^1$.
5. $C_0^\infty \subseteq X_S \subseteq L^2$.

$S_v(x,v) = \beta(v - f(x)) + \left(\frac{\exp(v)}{1+\exp(v)} - f(x)\right)\chi_B$ *satisfies:*

1. $S_v(x,\cdot)$ *is continuous for almost every* x,
2. $S_v(\cdot,v)$ *is measurable for each* $v \in \mathbb{R}$,
3. *For* $\psi \in X_S$ $S_v(\cdot,\psi(\cdot)) \in L^2$.
4. *For* $v,h \in X_S$ *we have*
   $$\eta(t,v,h) \le \left(\frac{1}{2} + \beta\right)|t|^2 \int |h|^2\ .$$

Theorem 3.6 shows that $u_\alpha \equiv 0$ *if and only if* $\|S_v(\cdot,0)\|_{G^s} \le \alpha$. *Theorem 3.7 shows that for* $\|S_v(\cdot,0)\|_{G^s} > \alpha$, u_α *is characterized by* $\|S_v(\cdot,u_\alpha(\cdot))\|_{G^s} = \alpha$ *and* $-\int S_v(x,u_\alpha(x))u_\alpha(x)\,dx = \alpha\|Du_\alpha\|_{s_*}$.

The preceding results allow a geometrical interpretation of bounded variation minimization and logistic regression.

ROF-model: Let Φ be measurable and satisfy $\Delta\Phi = f$ with $F_f := \nabla\Phi \in L^\infty_{\text{loc}}$. [2] By definition $\|\rho - f\|_{G^s} \le \alpha$ if and only if $\rho - f = \nabla\cdot\vec{v}$ and $\||\vec{v}|_{l^s}\|_{L^\infty} \le \alpha$. This is equivalent to

$$\rho = \nabla\cdot(\vec{v} + F_f) \text{ and } \||\vec{v}|_{l^s}\|_{L^\infty} \le \alpha \ .$$

Or in other words, ρ is the divergence of a vector valued function $\vec{\rho}$ which is in a tube around the "primitive" (to be precise, we solve Laplacian's equation and differentiate) of f. The tube is a subset of $\mathbb{R}^{2n}$ around the vector valued function F_f. We recall that u_α is the divergence of a vector valued function $\vec{u}_\alpha$ and the distance between $\vec{u}_\alpha$ and F_f is less than α, i.e., $\||F_f - \vec{u}_\alpha|_{l^s}\|_{L^\infty} \le \alpha$. Note that the tube geometry varies with s and has an impact on the solution (cf. [15]). For $s = 2$ the tube has a cylindrical shape and for $s = 1, \infty$ the tube is a slot.

The following geometric interpretations of the bounded variation regularized solutions u_α are immediate: the associated vector field $\vec{u}_\alpha$ does *not* have contact with the tube if and only if $\|f\|_{G^s} \le \alpha$.

Logistic Regression: Here we have

$$S_v(\cdot, 0) = -(\beta + \chi_B) f(x) + \frac{1}{2}\chi_B$$

If $\|S_v(\cdot, 0)\|_{G^s} \le \alpha$, i.e., it is in a tube around the 0 manifold, then $u_\alpha \equiv 0$. In all other situations $S_v(\cdot, u_\alpha(\cdot))$ has a contact with the tube of radius α.

3.3. S_v **is not differentiable.** This case is more involved, since S_v has to be considered *set-valued*. The situation is even worse since the function $h \in \mathtt{BV}$ may be discontinuous and has to be considered set-valued as well. In this situation a basic assumption that $S(\cdot, v)$ is measurable for each v is satisfied if for instance

$$S(x, v) = \tilde{S}(v - f(x)) \,, \tag{3.13}$$

where $\tilde{S}$ is convex. Note that the function S is convex for almost every x. For each open set O the set $\tilde{S}^{-1}(O)$ is open (any convex function is continuous) and thus since f is measurable

$$\{x : S(x, v) \in O\} = \{x : f(x) \in v - \tilde{S}^{-1}(O)\} = f^{-1}\left(v - \tilde{S}^{-1}(O)\right)$$

is measurable.

Moreover, we assume that

$$\underline{c}|\rho| \le \tilde{S}(\rho) \le \overline{c}|\rho| \ . \tag{3.14}$$

From this it follows that

$$\underline{c}\{|v| - |f(x)|\} \le S(x, v) = \tilde{S}(v - f(x)) \le \overline{c}|v - f(x)| \le \overline{c}\{|v| + |f(x)|\} \ .$$

Thus for $f \in L^1$, S satisfies Assumption 3.1 and (3.2) (with $p_0 = 1$). In particular

$$\mathcal{S}(u) := \int \tilde{S}(u(x) - f(x))\, dx$$

[2] All along this paper we have been considering data filtering on $\mathbb{R}^n$. If we consider data smoothing on a bounded, smooth domain Ω, the existence of a solution of Laplace's equation $\Delta\Phi = f$ with Neumann boundary data is guaranteed if $\int f = 0$. For $\mathbb{R}^n$ we assume the existence of a solution of this equation, which imposes further requirements on the data f.

is bounded and strongly continuous by the Nemytskii Theorem (see e.g. [22, Theorem 3.2]). Theorem 3.2 shows that there exists a minimizer in $X_S = \mathtt{BV} \cap L^1$.

We denote by $\partial\tilde{S}(\rho)$ the *subdifferential* of $\tilde{S}(\rho)$ at $\rho \in \mathbb{R}$ and by $\partial\tilde{s}(\rho)$ a single element of $\partial\tilde{S}(\rho)$. Under this assumptions, from (3.14) and the convexity of $\tilde{S}$ it follows that

$$|\partial\tilde{s}(\rho)| \leq \bar{c}\,, \tag{3.15}$$

for any $\partial\tilde{s}(\rho) \in \partial\tilde{S}(\rho)$ and $\rho \in \mathbb{R}$. Moreover, we assume that $\tilde{s}$ has only finitely many singularities

$$\rho_1 < \rho_2 < \rho_3 < \ldots < \rho_m\,. \tag{3.16}$$

To represent this assumption, we write

$$\partial\tilde{s}(\rho) = \tilde{s}'(\rho) \text{ for } \rho \in (\rho_i, \rho_{i+1})\,, \quad i = 0, 1, \ldots, m\,,$$

where we set $\rho_0 = -\infty$, $\rho_{m+1} = \infty$. For $v \in X_S$ we introduce the measurable sets

$$\Omega_i(v) := (v-f)^{-1}(\rho_i, \rho_{i+1})\,,\ i = 0, 1, \ldots, m\,, \quad \Omega(v) := \bigcup_{i=0}^{m} \Omega_i(v)\,,$$

$$\Gamma_i(v) := (v-f)^{-1}(\rho_i)\,,\ i = 1, 2, \ldots, m\,, \quad \Gamma(v) = \bigcup_{i=1}^{m} \Gamma_i(v)\,.$$

Note that $B := (\rho_i, \rho_{i+1})$ is open and for a measurable function $\tilde{v}$, $\tilde{v}^{-1}(B)$ is measurable (cf. [6]).

In the sequel we impose the following assumptions:

ASSUMPTION 3.9.

1. *For $v \in X_S$ we define and assume*

$$\Psi_v := \Big\{\psi \in L^\infty : \psi(x) = \tilde{s}'(v(x) - f(x)) \textit{ for } x \textit{ in } \Omega(v) \textit{ and}$$
$$\psi(x) \in \partial\tilde{s}(\rho_i) \textit{ for } x \in \Gamma_i(v)\Big\} \neq \emptyset\,.$$

   Any $\psi \in \Psi_v$ is measurable, since by definition it is in L^∞ (and thus in particular measurable). $\Psi_v \neq \emptyset$ follows from some abstract results in Deimling [3, Prop 3.2., p 22 ff]. In order to apply this result several assumptions have to be verified. For our applications it is more convenient to assume the existence of a function since at later stage we have to use one particular element of Ψ_v where its measurability is obvious.

2. *Moreover, we assume that for every $v \in X_S$ there exists $\psi_v \in \Psi_v$ such that for every $h \in X_S$*

$$\eta(t, v, h) := \int \Big(\tilde{S}(v(x) + th(x) - f(x)) - \tilde{S}(v(x) - f(x)) - t\psi_v(x)h(x)\Big)\, dx \tag{3.17}$$

   satisfies (3.4).

Here again, as in the continuous case, the existence of a directional derivative. Note, that at locations $x \in \Gamma(v)$ we choose one element ψ_v of the subgradient. For $x \in \Omega(v)$, $\psi_v(x)$ is single valued.

We recall that since $\tilde{S}(\cdot)$ is convex, for every $\psi \in \Psi_v$ and $h \in X_S$

$$\tilde{S}(v(x) + h(x) - f(x)) - \tilde{S}(v(x) - f(x)) - \psi(x)h(x) \geq 0 \,. \tag{3.18}$$

THEOREM 3.10. *Assume that $\tilde{S}$ is convex, satisfies* (3.14), *the subgradient has only finitely many singularities, and for $v \in X_S$, there exists $\psi_v \in \Psi_v$ such that η satisfies* (3.4).

Let $\psi_0 \in \Psi_0$ satisfy (3.18). *Then*

1. $\|\psi_0\|_{G^s} \leq \alpha$ *if and only if* $u_\alpha \equiv 0$.
2. *If* $\|\psi_0\|_{G^s} > \alpha$, *then*

$$\|\psi_{u_\alpha}\|_{G^s} = \alpha \text{ and } -\int \psi_{u_\alpha}(x) u_\alpha(x)\, dx \;=\; \alpha \|Du_\alpha\|_{s_*} \,.$$

The proof is along the lines of proof of Theorems 3.6 and 3.7 and thus omitted.

EXAMPLE 3.11. *Let $f \in L^1$.*

- *For the robust regression model $\tilde{S} = |\cdot|$ and $S(x,v) = |v - f(x)|$. By Theorem 3.2 $u_\alpha \in X_S = \mathrm{BV} \cap L^1$.*
  $\tilde{S}$ is convex, satisfies (3.14) *with $\underline{c} = \overline{c} = 1$, the subgradient has just one singularity at* 0.
  We have

  $$\tilde{s}'(z) = 1 \text{ if } z = v - f(x) > 0 \,,$$
  $$\tilde{s}'(z) = -1 \text{ if } z = v - f(x) < 0 \,,$$

  and set $\psi_v(z) = 0$ if $v = f(x)$.
  For $f \in L^1$ and $v \in X_S$ the function $\psi_v(v(x) - f(x))$ is measurable. Note that ψ_v attains only three values $-1, 0, 1$ and thus for any open set O and $J = O \cap \{0, \pm 1\}$

  $$\{x : \psi_v(v(x) - f(x)) \in O\} = \bigcup_{j \in J} \{x : \psi_v(v(x) - f(x)) = j\}.$$

  We have

  $$\{x : \psi_v(v(x) - f(x)) = 1\} = \{x : v(x) - f(x) > 0\} \,,$$
  $$\{x : \psi_v(v(x) - f(x)) = -1\} = \{x : v(x) - f(x) < 0\} \,,$$
  $$\{x : \psi_v(v(x) - f(x)) = 0\} = \{x : v(x) - f(x) = 0\} \,,$$

  Therefore, since v and f are measurable, so are the sets $\{x : \psi_v(v(x) - f(x)) = j\}, j = 0, \pm 1$, and therefore the finite union is measurable as well. This shows that ψ_v is measurable, and by its definition $\psi_v \in L^\infty$.
  Moreover, we have

  $$\begin{aligned} \frac{|\eta(t,v,h)|}{|t|} &\leq 2 \int_{0 < |v-f| \leq |th|} |h(x)|\, dx + \int_{0 = |v-f|} \underbrace{|\psi_v(x)|}_{=0} |h(x)|\, dx \\ &= 2 \int |h(x)| \chi_{0 < |v-f| \leq |th|}(x)\, dx \,. \end{aligned}$$

The family of functions $g_{|t|}(x) := |h(x)|\chi_{|v-f|\leq|t||h|}(x)$ *is monotonically decreasing in* $|t|$ *and thus by the monotone convergence theorem*

$$\begin{aligned}\lim_{|t|\to 0} g_{|t|}(x)\,dx &= \int |h(x)| \lim_{|t|\to 0} \chi_{0<|v-f|\leq|th|}(x)\,dx \\ &= \int |h(x)|\chi_{M_0}(x)\,dx \\ &= 0\,,\end{aligned}$$

where M_0 *is a set of measure* 0.
Theorem 3.10 shows that $u_\alpha \equiv 0$ *if and only if* $\|\psi_0\|_{G^s} \leq \alpha$. *Moreover, if* $\|\psi_0\|_{G^s} > \alpha$, u_α *is characterized by* $\|\psi_{u_\alpha}\|_{G^s} = \alpha$ *and* $-\int \psi_{u_\alpha} u_\alpha(x)\,dx = \alpha\|Du_\alpha\|_{s_*}$.

- *For quantile regression the argument is similar. In this case we have*

$$\begin{aligned}\tilde{s}'(z) &= \beta \text{ if } z = v - f(x) > 0\,, \\ \tilde{s}'(z) &= \beta - 1 \text{ if } z = v - f(x) < 0\,,\end{aligned}$$

and set $\psi_v(z) = 0$ *if* $v = f(x)$.

The geometrical interpretation of robust regression is rather different from the ROF-model. Here, we have $\psi_0 = \chi_{f<0} - \chi_{f>0}$. For the sake of simplicity of illustration we assume that f is one-dimensional with $\{f < 0\} = (-a, a)$ and $\{f > 0\} = \emptyset$. A primitive of ψ_0 is continuous, constant with value c in $(-\infty, -a)$, linear in $(-a, a)$, and constant with value $c + 2a$ in (a, ∞). Therefore $\|\psi_0\|_G = a$. The general results of this section show that $u_\alpha \equiv 0$ if and only if $\|\psi_0\|_G = a \leq \alpha$. Thus robust regression is capable of removing isolated sources of width 2α. Note that the ROF-model does not remove isolated sources for small parameter values α, but dampens the amplitude. This property of robust regression is well documented in statistics for filtering of one-dimensional data. Here the result is in a functional analytic context and applies to multi-dimensional data.

4. Conclusion. In this paper we have investigated G-norm properties of a class of bounded variation filtering of multi-dimensional data in a functional analytical context. One motivation for studying these filtering methods is due to their success in statistics for analyzing one-dimensional data. Here we considered them in a multi-dimensional space setting, thinking of possible applications to imaging. For characterization of the minimizers of regression models we followed the analysis of Y. Meyer [11] for the ROF-model. We made extensive use of the G-norm, for which we gave a geometrical interpretation.

Acknowledgment. The work of O.S. has been supported by the FWF (Österreichischer Fonds zur Förderung der wissenschaftlichen Forschung), grant Y-123 INF-N04 and P-15617-N04. The work of S.O. was supported by NSF grants ACI-0321917 and DMS 0312222. The authors want to thank Markus Grasmair for his proof-reading of the manuscript. The authors want to thank a referee for pointing out that in the Gagliardo-Nirenberg-Sobolev inequality no constant on the right hand is required.

REFERENCES

[1] J.-F. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle. Image decomposition application to SAR images. In *[7]*, pages 297–312, 2003.

[2] P. L. Davies and A. Kovac. Local extremes, runs, strings and multiresolution. *Ann. Statist.*, 29:1–65, 2001. With discussion and rejoinder by the authors.

[3] K. Deimling. *Multivalued differential equations.* Walter de Gruyter & Co., Berlin, 1992.

[4] L. Dümbgen and A. Kovac. Extensions of smoothing via taut strings preprint.

[5] I. Ekeland and R. Temam. *Convex Analysis and Variational Problems.* North Holland, Amsterdam, 1976.

[6] L.C. Evans and R.F. Gariepy. *Measure Theory and Fine Properties of Functions.* CRC–Press, Boca Raton, 1992.

[7] L.D. Griffin and M. Lillholm, editors. *Scale-Space Methods in Computer Vision.* Lecture Notes in Computer Science Vol. 2695, Springer Verlag, 2003. Proceedings of the 4th International Conference, Scale-Space 2003, Isle of Skye, UK, June 2003.

[8] W. Hinterberger, M. Hintermüller, K. Kunisch, M. von Oehsen, and O. Scherzer. Tube methods for BV regularization. *JMIV*, 19:223–238, 2003.

[9] R. Koenker and G. Bassett. Regression quantiles. *Econometrica*, 46:33–50, 1978.

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

[11] Y. Meyer. *Oscillating patterns in image processing and nonlinear evolution equations*, volume 22 of *University Lecture Series.* American Mathematical Society, Providence, RI, 2001.

[12] M.Z. Nashed and O. Scherzer, editors. *Interactions on Inverse Problems and Imaging*, volume 313. AMS, 2002. Contemporary Mathematics.

[13] S. Osher and R. Fedkiw. *Level set methods and dynamic implicit surfaces*, volume 153 of *Applied Mathematical Sciences.* Springer-Verlag, New York, 2003.

[14] S. Osher and N. Paragios. *Geometric level set methods in Imaging, Vision, and Graphics.* Springer-Verlag, New York, 2003.

[15] S. Osher and S. Esedoglu. decomposition of images by the anisotropic Rudin-Osher- Fatemi model. Technical report, 2003. UCLA CAM report 03-34, to appear Comm. Pure Appl. Math.

[16] S. Osher and Sole A. and Vese, L. Image decomposition and restoration using total variation minimization and the H^{-1}-norm. *SIAM Multiscale Model Simul.*, 1:349–370, 2003.

[17] L.I. Rudin, P.-L. Lions, and S. Osher. Multiplicative denoising and deblurring: theory and applications. In *[14]*, pages 103–119, 2003.

[18] L.I. Rudin and S. Osher. Total variation based image restoration with free local constraints. *Proc. ICIP IEEE Int. Conf. on Image Processing, Austin TX*, pages 31–35, 1994.

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

[20] O. Scherzer. Explicit versus implicit relative error regularization on the space of functions of bounded variation. In *[12]*, pages 171–198, 2002.

[21] O. Scherzer. Taut-string algorithm and regularization programs with G-norm data fit. *Institutsbericht 13, Fachbereich Mathematik-Informatik*, 2003. accepted for publication in Journal of Mathematical Imaging and Vision.

[22] R.E. Showalter. *Monotone Operators in Banach Spaces and Nonlinear Partial Differential Equations.* American Mathematical Society, Providence, Rhode Island, 1997.

[23] L. Vese and S. Osher. Modelling textures with total variation minimization and oscillating pattern in image processing. *SIAM, J. Sci. Comput.*, 19: 553-572, 2003.