# Stable evaluation of differential operators and linear and nonlinear multi–scale filtering [*]

Otmar Scherzer

### Abstract

Diffusion processes create multi–scale analyses, which enable the generation of simplified pictures, where for increasing scale the image gets sketchier. In many practical applications the "scaled image" can be characterized via a variational formulation as the solution of a minimization problem involving unbounded operators. These unbounded operators can be evaluated by regularization techniques. We show that the theory of stable evaluation of unbounded operators can be applied to efficiently solve these minimization problems.

## 1 Introduction

Morel and Solimini describe in their book [10] a *multi–scale analysis*:

> Multi–scale analysis is concerned with the generation of a sequence of simplified pictures $u_\lambda$ from a single initial picture $u_0^\delta$, where $u_\lambda$ appears to be a rougher and sketchier version of $u_0^\delta$ as $\lambda$ increases.

In the literature various possibilities for generating multi–scale analyses have been proposed. One common method to generate a multi–scale analysis is by a *diffusion process*, such as

$$\begin{array}{rcl} \frac{\partial u}{\partial t}(x,t) & = & \Delta u(x,t) \\ u(x,0) & = & u_0^\delta(x) \,. \end{array} \tag{1}$$

It is easy to see that in $\mathbb{R}^2$ convolution of the signal $u_0^\delta$ at each scale is equivalent to the solution of the heat equation with the signal as initial datum [9]. The solution of (1) for the initial datum $u_0^\delta$ with bounded quadratic norm is

$$u(x,t) = G_t \star u_0^\delta(x) = \int G_t(x-y)u_0^\delta(y)\,dy$$

where $G_\sigma(x) = \frac{1}{4\pi\sigma}\exp(-\|x\|^2/4\sigma)$ $(x \in \mathbb{R}^2)$. The sequence $u_\lambda(.) = u(.,\lambda)$ defines a multi–scale analysis, which satisfies the following properties: Let $S_\lambda$ be the mapping $u_0^\delta \to S_\lambda(u_0^\delta) := u_\lambda$

1. $u_\lambda \to u_0^\delta$ as $\lambda \to 0$

2. $S_\lambda(u_0^\delta)$ only depends on $S_{\lambda'}(u_0^\delta)$ if $\lambda > \lambda'$

3. If $A$ is an isometry, then $S_\lambda(u_0^\delta \circ A) = (S_\lambda(u_0^\delta)) \circ A$.

In image processing and image enhancement techniques one easily understands the necessity of a previous low–pass filtering, i.e., by convolution of the signal with a Gaussian kernel: If the signal $u_0^\delta$ is noisy, the gradient has a lot of irrelevant spikes, which have to be eliminated.

As we will see in Section 3 the linear multi–scale analysis based on (1) has a tendency to over smooth data. Many attempts have been made in the literature to construct diffusion processes which satisfy a *strong causality* principle, that is, the associated multi–scale analysis reconstructs edges and corners of images, while simultaneously eliminating highly oscillating noise in the image. Diffusion processes where the associated multi–scale analysis satisfies a strong causality principle have been found to be mostly nonlinear. In Section 2 we give an outline of linear and nonlinear diffusion processes and motivate associated multi–scale analysis. The multi–scale analysis can be the basis for efficient algorithms for denoising images and edge–detection. In Section 3 we will outline the concept of stable evaluation of unbounded operators and we show in Sections 3, 4 that this concept is relevant for the multi–scale analysis associated with diffusion processes. In Section 5 we will discuss the numerical implementation of a multi–scale analysis associated with a nonlinear diffusion process and we will show some numerical results.

## 2   Variational formulations of diffusion processes

### 2.1   The Hildreth–Marr theory

The *energy functional* associated with the differential equation (1) is

$$E(u) = \int |\nabla u|^2 .$$

This functional gives an evaluation of the amount of information in $u$. The energy semi norm is the right functional to be associated with (1) since the heat equation is the descent method associated with this energy. If we discretize (1) we end up with

$$u(x,t) - u_0^\delta(x) = t\Delta u(x,t)$$

which is the solution of a variational problem: roughly speaking, numerically solving the heat equation is therefore equivalent to minimizing the following functional iteratively with $t = \lambda^2$ small:

$$E_\lambda(u) = \lambda^2 \int_\Omega |\nabla u|^2 + \int_\Omega (u - u_0^\delta)^2 .$$

In this formulation the scaling parameter appears explicitly as the multiplier of the smoothing term. The solution $u_\lambda$ of this problem can be defined as the "analysis of $u$ at scale $\lambda$" as well as the solution of the heat equation at time $t = \lambda^2$.

## 2.2 The Perona and Malik theory

For an edge–detection filtering technique Perona and Malik (see e.g. [10]) suggested the *nonlinear* diffusion process

$$\begin{aligned}
\tfrac{\partial u}{\partial t}(x, t) &= \nabla.\left(f(|\nabla u|)\nabla u\right)(x, t) \\
u(x, 0) &= u_0^\delta(x) ,
\end{aligned} \qquad (2)$$

where $f$ is a smooth non-increasing function satisfying

$$f(0) = 1, \quad f(s) \geq 0 \quad \text{and} \quad \lim_{s \to \infty} f(s) \to 0 .$$

The idea of using this type of nonlinear filtering technique is that the smoothing process is conditional: if $\nabla u(x)$ is large, then the diffusion will be low and therefore the exact location of the edges in the image are preserved. If $\nabla u(x)$ is small, then the diffusion will tend to smooth around $x$. Similarly to the Hildreth–Marr theory one can derive a variational formulation of the Perona and Malik model. In fact, if we set $\tilde{f}(a) = f(\sqrt{a})$ then equation (2) can be rewritten as

$$\begin{aligned}
\tfrac{\partial u}{\partial t}(x, t) &= \nabla.\left(\tilde{f}(|\nabla u|^2)\nabla u\right)(x, t) \\
u(x, 0) &= u_0^\delta(x) ,
\end{aligned}$$

which yields a multi–scale analysis

$$E_\lambda(u) = \lambda^2 \int_\Omega \tilde{F}(|\nabla u|^2) + \int_\Omega (u - u_0^\delta)^2 , \qquad (3)$$

where $\tilde{F}(t) = \int_0^t \tilde{f}(s)$ .

### 2.3   A nonlinear filtering technique studied by Osher and Rudin

Based on the Perona and Malik theory Osher and Rudin [12] considered a non-linear diffusion process for edge detection:

$$\begin{array}{rcl} \frac{\partial u}{\partial t}(x,t) & = & \nabla.\left(\frac{\nabla u}{|\nabla u|}\right)(x,t) \\ u(x,0) & = & u_0^\delta(x) \ . \end{array} \tag{4}$$

The energy functional associated with this diffusion process is

$$E(u) = \int_\Omega |\nabla u| \ .$$

As above an associated multi–scale analysis can be motivated in terms of the functional:

$$E_\lambda(u) = \lambda^2 \int_\Omega |\nabla u| \ + \ \int_\Omega (u - u_0^\delta)^2 \ . \tag{5}$$

### 2.4   Higher order nonlinear filtering techniques

Diffusion filters of higher order play an important role in parameter estimation problems [14]. An appropriate filtering technique for enhancing jumps in derivatives of signals is

$$\frac{\partial u}{\partial t}(x,t) = H.\left(\frac{Hu}{|Hu|}\right)(x,t) \ . \tag{6}$$

Here $Hu$ denotes the Hessian of $u$ and $H. = (\nabla.\nabla.)$. The multi–scale analysis associated with this diffusion process is

$$E_\lambda(u) = \lambda^2 \int_\Omega |Hu| \ + \ \int_\Omega (u - u_0^\delta)^2 \ .$$

## 3   Unbounded operators

In this section we consider the problem of computing values of an unbounded operator $L$. Here and in the following of this section, we will always denote by $L : \mathcal{D}(L) \subseteq H_1 \to H_2$ a closed densely defined unbounded linear operator between two Hilbert spaces $H_1$ and $H_2$. The problem of computing values $y = Lu_0$, for $u_0 \in D(L)$, is then ill–posed in the sense that small perturbations in $u_0$ may lead to data $u_0^\delta$ satisfying $\|u_0 - u_0^\delta\| \le \delta$, but $u_0^\delta \notin \mathcal{D}(L)$, or even if $u_0^\delta \in \mathcal{D}(L)$, it may happen that $Lu_0^\delta \not\to Lu_0$ as $\delta \to 0$. Morozov has studied a stable method for approximating the value $Lu_0$ when only approximate data $u_0^\delta$ is available (see [11] for information on Morozov's work). This method takes as an approximation to $y = Lu_0$ the vector

$$y_{\lambda^2}^\delta = Lz_{\lambda^2}^\delta,$$

where $z_{\lambda^2}^{\delta}$ minimizes the functional

$$\|z - u_0^{\delta}\|^2 + \lambda^2\|Lz\|^2 \quad (\lambda^2 > 0)$$

over $D(L)$. For $L = \nabla$ the approximate $z_{\lambda^2}^{\delta}$ of $u_0$ is the solution of the Hildreth–Marr problem at level $\lambda^2$. In [7] we discussed the *optimal* choice of the scaling parameter $\lambda^2$ in relation to the noise level $\delta$ and smoothness assumptions on the solution and the numerical realization. Of particular interest for image reconstruction and image enhancement techniques is the choice of the scaling parameter. The theory developed for the stable evaluation of unbounded operators to optimally choose the *regularization* parameter is **directly** applicable to optimally choosing the *scaling* parameter in the Hildreth–Marr theory. In order to clarify the choice of the scaling parameter we cite a result developed in [7].

**Theorem 3.1** *If $u_0 \in \mathcal{D}(L^*L)$ and $u_0 \notin \mathcal{N}(L)$, then $\|y_{\lambda^2(\delta)}^{\delta} - Lu_0\| = O(\sqrt{\delta})$, when $\lambda^2(\delta)$ is chosen such that*

$$\|u_{\lambda^2(\delta)}^{\delta} - u_0\| = O(\sqrt{\delta}) \, . \tag{7}$$

Roughly speaking in the particular case $L = \nabla$ the assumption $u_0 \in \mathcal{D}(L^*L)$ means $u_0 \in H^2$, and the derivative of $u_0$ can be reconstructed with a rate $O(\sqrt{\delta})$ if $\lambda^2(\delta)$ is chosen according to the discrepancy principle (7).

Other strategies for optimally choosing the regularization parameters when calculating unbounded operators can be found in [7]. For more background on the stable evaluation of unbounded operators we refer to [6].

From a simple example, by taking $L = \nabla$, $\mathcal{D}(L) = H^1(\mathbb{R})$, the Sobolev space of functions possessing a weak derivative in $L^2(\mathbb{R})$, it follows that $z_{\lambda^2}^{\delta} \in H^1(\mathbb{R})$. This shows that at any scale discontinuities of the original image are smeared out, and thus the method is not strongly causal. Nonlinear filtering techniques have turned out to be strongly causal and therefore, we study below numerical implementation of nonlinear filtering techniques based on the nonlinear diffusion process developed by Rudin, Osher, and Fatemi [13].

# 4   Unbounded operators in nonlinear multi–scale analysis

The multi–scale analysis associated with the nonlinear diffusion process developed by Osher and Rudin is sometimes called *regularization with functions of bounded variation*, since the scale–energy functional (5) is finite for any function $u \in \mathrm{BV}$, that is, the space of all functions which satisfy

$$\int_{\Omega} |\nabla u| := \sup\left\{ \int_{\Omega} u \nabla . v dx \, : \, v \in C_0^{\infty}(\Omega, \mathbb{R}^d), |v(x)| \leq 1, x \in \Omega \right\} < \infty \, .$$

In practical applications, since the functional (5) is not differentiable, the minimizer of (5) is frequently approximated by the solution of the (nonlinear) *least squares problem* (see e.g. [4, 2, 3])

$$\min_{u \in L^2(\Omega)} \left\{ \frac{1}{2} \|u - u_0^\delta\|_{L^2(\Omega)}^2 + \lambda^2 J_\beta(u) \right\}, \tag{8}$$

where

$$J_\beta(u) := \int_\Omega \sqrt{|\nabla u|^2 + \beta^2} \ ; \tag{9}$$

here $\Omega \subseteq \mathbb{R}^d$. The term $J_\beta(u)$ is used for stabilization and is therefore called the *regularizing term*.

Vogel and Oman [15] introduced a fixed point iteration method to minimize (8), (9): The corresponding (formally derived) *Euler–Lagrange* equation is

$$u - \lambda^2 \nabla . \left( \frac{1}{\sqrt{|\nabla u|^2 + \beta^2}} \nabla u \right) = u_0^\delta \ ,$$

which can be expressed in operator notation as

$$(I + A_{\lambda^2 \beta}(u))u = u_0^\delta, \tag{10}$$

where

$$A_{\lambda^2 \beta}(u)v := -\lambda^2 \nabla . \left( \frac{1}{\sqrt{|\nabla u|^2 + \beta^2}} \nabla v \right) \ .$$

A fixed point iteration for (10) can be defined by

$$u^{n+1} = (I + A_{\lambda^2 \beta}(u^n))^{-1} u_0^\delta =: F(u^n), \quad n = 0, 1, ...$$

In actual computations the minimization of (8) is difficult to handle, since it involves evaluations of the *unbounded* operator $\nabla$. To overcome this difficulty we proposed in [3] to approximate the values of the unbounded operator $\nabla$ and use the minimizing element of

$$\min_{u \in L^2(\Omega)} \left\{ \frac{1}{2} \|u - u_0^\delta\|_{L^2(\Omega)}^2 + \lambda^2 J_{\beta\gamma}(u) \right\}, \tag{11}$$

where

$$J_{\beta\gamma}(u) := \int_\Omega \sqrt{|L_\gamma u|^2 + \beta^2}, \tag{12}$$

as an approximation for the desired image. In practical applications the operator $L_\gamma$ might be, for instance, a finite difference scheme, where the parameter $\gamma$ represents the grid-size. Denoting by $L_\gamma^*$ the adjoint of $L_\gamma$, the Euler–Lagrange equation for a minimizer of (11) is

$$u + \lambda^2 L_\gamma^* \left( \frac{1}{\sqrt{|L_\gamma u|^2 + \beta^2}} L_\gamma u \right) = u_0^\delta \ .$$

Extending the ideas of Vogel and Oman [15] and Dobson and Vogel [4] we studied (in [3]) the iterative scheme

$$u^{n+1} = (I + A_{\lambda^2 \beta \gamma}(u^n))^{-1} u_0^\delta =: F(u^n), \tag{13}$$

where

$$A_{\lambda^2 \beta \gamma}(u)v := \frac{\lambda^2}{\beta} L_\gamma^* \left( \frac{1}{\sqrt{\frac{|L_\gamma u|^2}{\beta^2} + 1}} L_\gamma v \right). \tag{14}$$

As long as the parameters $\lambda^2, \beta, \gamma$ are chosen appropriately (13),(14) is convergent. This setting reflects computational relevance since in numerical simulations the operator $\nabla$ is (almost) never calculated exactly, but approximately evaluated, e.g. by some finite difference scheme.

In numerical calculations (see e.g. [2, 1]) regularization with functions of bounded variation turned out to be very efficient to determine discontinuities in signals and images. This can be motivated both by arguing that the associated diffusion process (4) is designed to optimally enhance edges in images (see e.g. [12, 10]) or from arguing for the least squares functional (8) (respectively (5)) directly. Namely, efficient regularizing schemes have to satisfy the following two properties:

1. The regularizing term evaluated at the function to be recovered is finite; in signal and image processing, the images of most interest are piecewise constant and thus have finite bounded variation (see e.g. [5, 16]), i.e.,

$$\int_\Omega |\nabla u_0| := \sup \left\{ \int_\Omega u_0 \nabla . v dx \, : \, v \in C_0^\infty(\Omega, \mathbb{R}^d), |v(x)| \le 1, x \in \Omega \right\} < \infty .$$

2. The recovery of images from noisy measurements is ill–posed, in the sense that small perturbations in the data $u_0^\delta$ may lead to significant distortions of the picture (mathematically the ill-posedness is due to the fact that the bounded variation semi norm – which represents a natural norm in measuring the distortion of the picture – does not depend continuously on the $L^2$ – norm – which represents the natural norm for noise in the data), the regularizing term has to have appropriate stabilizing properties.

In practical applications (such as in non–destructive testing) frequently discontinuities in derivatives have to be recovered instead of the function itself. Examples include electrical conductivity problems in $\mathbb{R}^1, \mathbb{R}^2$, where the idealized model is the following: Given some domain $\Omega$ in $\mathbb{R}^1$ or $\mathbb{R}^2$; the voltage potential $u_0$ inside of $\Omega$ satisfies

$$\nabla.(\sigma \nabla u_0) = 0 \quad \text{in } \Omega. \tag{15}$$

If the electrical conductivity $\sigma$ is discontinuous, then by jump relations (see e.g. [8]) the resulting solution of (15), the voltage potential $u_0$, reveals discontinuities in the derivatives of $u_0$. Therefore, discontinuities in the material (i.e., discontinuities in $\sigma$) can be detected from measurement data $u_0^\delta$ of $u_0$ and by *denoising* $u_0^\delta$ and detecting discontinuities in derivatives of $u_0$.

For estimating discontinuities in electrical conductivity problems by denoising the data $u_0^\delta$ the two demands on an efficient regularization scheme suggest the use a regularizing (semi) norm which satisfies the two properties:

1. For functions with jumps in derivatives the regularizing norm is finite

2. The stabilizing term has to cope with the ill–posedness.

A reasonable semi norm satisfying both properties is

$$|Hu| = \sum_{i=1}^{d} \sum_{j=1}^{d} \left| \frac{\partial^2 u}{\partial x_i \partial x_j} \right| ;$$

here $Hu$ denotes the *Hessian* of $u$, i.e.,

$$H = \begin{pmatrix} \frac{\partial^2 u}{\partial x_1^2} & \cdots & \frac{\partial^2 u}{\partial x_1 \partial x_d} \\ \cdots & \cdots & \cdots \\ \frac{\partial^2 u}{\partial x_d \partial x_1} & \cdots & \frac{\partial^2 u}{\partial x_d^2} \end{pmatrix} .$$

For the numerical solution of the parameter estimation problem described in (15) we propose to use a *higher order bounded variation denoising* algorithm

$$\min_{u \in L^2(\Omega)} \frac{1}{2} \|u - u_0^\delta\|_{L^2(\Omega)}^2 + \lambda^2 J_{\beta\gamma}(u), \tag{16}$$

where

$$J_{\beta\gamma}(u) := \int_\Omega \sqrt{|H_\gamma u|^2 + \beta^2} \, ;$$

here the meaning of $H_\gamma$ is a stable approximation of the Hessian in $\mathbb{R}^d$.

Formally (16) looks like (12), and actually the analysis of the fixed point iteration is formally the same (see [14]).

# 5   Numerical results: Higher order bounded variation regularization in comparison to a linear multi–scale algorithm

In this section we present some numerical experiments to estimate discontinuities in derivatives of a one-dimensional signal from its noisy measurements. We

solved the denoising problem both with regularization with *higher order bounded variations* and with a *linear multi–scale algorithm* and compared the numerical results.

In the numerical experiments presented below, when functions with derivatives of bounded variation were used for regularization, we applied the iterative scheme (16) in a discrete setting.

**Example 5.1 Regularization with functions with derivatives of higher order bounded variations:** The first example is to determine the discontinuity in the derivative of the piecewise continuously differentiable function

$$f : [0,1] \rightarrow \mathbb{R} \, . \\ x \rightarrow \tfrac{1}{2} - \left| \tfrac{1}{2} - x \right| \tag{17}$$

In Figure 1 the unperturbed signal and its derivative are plotted. We added to 97 uniformly distributed measurement points random noise to the data (see Figure 2) to simulate noisy measurements of the signal. Figures 3, 4 show recon-
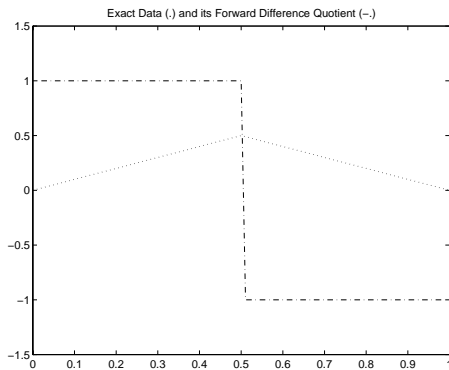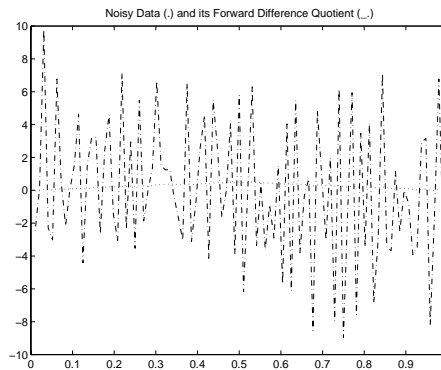


Figure 1:                          Figure 2:

structions with regularization with *higher order derivatives of bounded variation*. In each of the Figures 3, 4 the dotted line represents the reconstructed function, while the dashed dotted line represents its derivative, which was calculated by a forward difference scheme. In the bottom line of these figures the actual parameter setting of $\lambda^2, \beta, \gamma$ is plotted.

**Example 5.2 A linear multi–scale algorithm:** Here the same denoising problem as in Example 5.1 is studied, but as resolving algorithm we used a linear multi–scale algorithm. In Figures 5, 6 we have plotted the Tikhonov regularized solutions, i.e., the minimizing elements of

$$\min \| u - u_0^\delta \|_{L^2}^2 + \lambda^2 \| u_{xx} \|_{L^2}^2 \tag{18}$$

for different values of the regularization parameter $\lambda^2$. The reconstruction in Figures 5 - 6 the $\star$–line shows reconstructions with a linear multi–scale algorithm
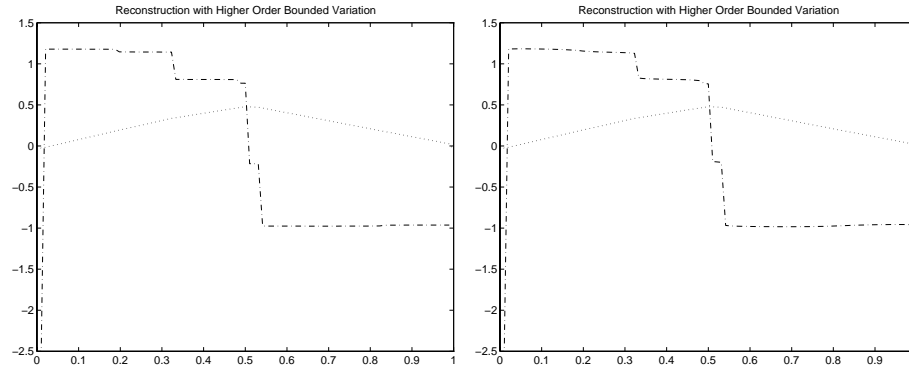
Figure 3: $\lambda^2 = 1.e - 4$, $\beta = 1.e - 2$, $\gamma = 1/m$, $m = 96$



Figure 4: $\lambda^2 = 1.e - 3$, $\beta = 1.e - 2$, $\gamma = 1/m$, $m = 96$

if it is implemented with noisy boundary data, and the "$\times$"–line shows its derivative; lines $-\cdot$ and $--$ denote the reconstruction and its derivative if "a–priori" information on exact boundary data is available and implemented.
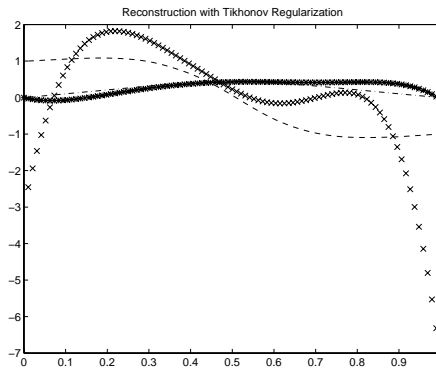


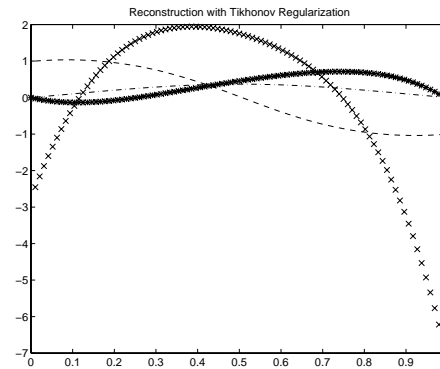Figure 5: $\lambda^2 = 1.e - 4$, $m = 96$



Figure 6: $\lambda^2 = 1.e - 3$, $m = 96$

It is apparent from the numerical experiments that regularization with higher order bounded variation resolves discontinuities in derivatives efficiently; in comparison with a linear multi–scale algorithm the location of the discontinuities can be sharply estimated. Of course there is a price to pay: a linear multi–scale algorithm requires only the solution of one sparse matrix problem. For regularization with functions of higher order bounded variation a sparse matrix problem (of the same dimension as in a linear multi–scale algorithm) has to be solved in each iteration.

# References

[1] A. Chambolle and P.-L. Lions. Image recovery via total variation minimization and related problems. *Numer. Math.*, 72:168-188, 1997.

[2] D. Dobson and F. Santosa. An image enhancement technique for electrical impedance tomography. *Inverse Problems*, 10:317–334, 1994.

[3] D. Dobson and O. Scherzer. Analysis of regularized total variation penalty methods for denoising. *Inverse Problems*, 12:601–617, 1996.

[4] D. Dobson and C.R. Vogel. Convergence of an iterative method for total variation denoising. *SIAM J. on Numerical Analysis*. to appear.

[5] L. C. Evans and R. F. Gariepy. *Measure Theory and Fine Properties of Functions*. CRC Press, Ann Arbor, 1995.

[6] C. W. Groetsch. Spectral methods for linear inverse problems with unbounded operators. *J. Approx. Th.*, 70:16–28, 1992.

[7] C. W. Groetsch and O. Scherzer. Optimal order of convergence for stable evaluation of differential operators. *Electronic Journal of Differential Equations*, 1993 No.4:1–10, 1993.

[8] V. Isakov. *Inverse Source Problems*. Amer. Math. Soc., Providence, 1990.

[9] J. J. Koenderink. The structure of images. *Biol. Cybern.*, 50:363–370, 1984.

[10] J.-M. Morel and S. Solimini. *Variational Methods in Image Segmentation*. Birkhäuser, Basel, 1995.

[11] V. A. Morozov. *Methods for solving incorrectly posed problems*. Springer Verlag, New York, Berlin, Heidelberg, 1984.

[12] S. Osher and L. Rudin. Feature oriented image enhancement using shock filters. *SIAM J. on Numerical Analysis*, 27:919–940, 1990.

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

[14] O. Scherzer. Denoising with higher order derivatives of bounded variation and an application to parameter estimation. Accepted for publication in Computing.

[15] C. Vogel and M. Oman. Iterative methods for total variation denoising. *SIAM J. Stat. Sci. Comput*, 17:227–238, 1996.

[16] W. P. Ziemer. *Weakly Differentiable Functions*. Springer Verlag, New York, Berlin, Heidelberg, 1980.

Otmar Scherzer
Institut für Industriemathematik
Universität Linz
Altenberger Str. 69, A–4040 Linz, Austria
E-Mail address: scherzer@indmath.uni-linz.ac.at