Nonstationary Iterated Tikhonov-Morozov Method and Third Order Differential Equations for the Evaluation of Unbounded Operators

C.W. Groetsch† O. Scherzer‡

September 30, 1999

Abstract

In this paper we analyze the nonstationary iterative Tikhonov-Morozov method analytically and numerically for the stable evaluation of differential operators and for denoising images. A relationship between nonstationary iterative Tikhonov-Morozov regularization and a filtering technique based on a differential equation of third order is established and both methods are shown to be effective for denoising images and for the stable evaluation of differential operators. The theoretical results are verified numerically on model problems in ultrasound imaging and numerical differentiation.

1 Introduction

In this paper we analyze the *nonstationary iterated Tikhonov–Morozov method* for the evaluation of unbounded operators. In particular we show that in image processing this method is suitable for denoising images. Such denoising

†Department of Mathematics, University of Cincinnati, Cincinnati, Ohio 45221-0025, USA.

‡Institut für Industriemathematik, Johannes-Kepler-Universität, Altenberger Str. 69, A-4040, Linz, Austria.

techniques can be associated with filtering techniques based on parabolic partial differential equations which are, surprisingly, of *third order*. We are concerned with the following problem: let $L : \mathcal{D}(L) \subseteq H_1 \to H_2$ be a linear, closed, densely defined, unbounded operator between two real Hilbert spaces H_1 and H_2. The goal is to derive stable approximations to $L\rho^\dagger$, where $\rho^\dagger \in \mathcal{D}(L)\backslash\mathcal{N}(L)$, from given noisy data $\rho^\delta \in H_1$ satisfying

$$\|\rho^\dagger - \rho^\delta\|_{H_1} \leq \delta \ . \tag{1}$$

A basic example of evaluating unbounded operators occurs in signal and image processing, where $\mathcal{D}(L)$ is the space of images, and L is a differential operator. Typically in image processing the data signal ρ^δ is distorted by noise and therefore it is not contained in $\mathcal{D}(L)$. The operator L can be modeled in various ways. If L is a differential operator of first order, for example, the gradient, then evaluation of L can be used for edge detection since edge regions correspond to values of L having large norm. Differential operators of second order are used to determine singularities in the first derivative. Other applications in image processing concern denoising and deblurring of images.

Most differential equations considered in the literature for denoising and deblurring images are of parabolic type. The diffusive nature of these parabolic equations is exploited to smooth away noise in the signal. Stopping the diffusion before important features of the true signal are over-smoothed is a key aspect of such techniques. The history of parabolic denoising and deblurring techniques is rooted in the Marr-Hildreth-Koenderink-Witkin multiscale analysis which is based on the heat equation

$$\frac{\partial \rho}{\partial t} = \Delta \rho \ . \tag{2}$$

More generally, denoising by Perona-Malik anisotropic diffusion employs the equation

$$\frac{\partial \rho}{\partial t} = \nabla \cdot (g(|\nabla \rho|)\nabla \rho) \ , \tag{3}$$

where g is an appropriate positive function, while *mean curvature motion* diffusion uses

$$\frac{\partial \rho}{\partial t} = |\nabla \rho| \nabla \cdot \left(\frac{1}{|\nabla \rho|} \nabla \rho \right) \ . \tag{4}$$

The fundamental paper of Alvarez, Guichard, Lions and Morel [AGLM93] establishes a mathematical framework for the architecture of the multiscale image analysis. *Multiscale analysis* refers to the fact that the sequence of images $u(\cdot, t)$ is used for diagnostic purposes, each $t > 0$ corresponding to a different filtering technique. The axiomatic approach of Alvarez, Guichard, Lions and Morel is inspired by a systematic derivation of models in *continuum mechanics*.

The approach we follow in this paper is inspired by regularization theory. As we will show below, this approach leads to differential equations where the highest derivative is a *third order* derivative with respect to the space and the scale variables. Such differential equations have not, to our knowledge, previously received attention in *signal - and image processing* and in *computer vision* studies.

Let $\{\alpha_n\}_{n \in I\!N}$ be a sequence of positive parameters. With the sequence $\{\alpha_n\}_{n \in I\!N}$ we associate a sequence $\{t_n\}_{n \in I\!N_0}$ satisfying

$$t_0 = 0 \text{ and } t_n - t_{n-1} = \alpha_n \, .$$

Below we introduce *nonstationary iterative Tikhonov-Morozov regularization*: let $\rho_0 = 0$ and $\{\rho_n\}$ be the sequence of functions minimizing the functional

$$f_n(\rho) := \|\rho^\delta - \rho\|_{H_1}^2 + \alpha_n \|L(\rho - \rho_{n-1})\|_{H_2}^2 \, . \tag{5}$$

over $\mathcal{D}(L)$. From the definition of the sequence $\{t_n\}$ it follows that the minimizer ρ_n of the functional f_n satisfies

$$\rho_n - \rho^\delta + \frac{1}{t_n - t_{n-1}} L^* L(\rho_n - \rho_{n-1}) = 0 \, , \tag{6}$$

where L^* is the adjoint of the operator L with respect to the spaces H_1 and H_2. Formally this is an implicit difference scheme in time for the solution of

$$\rho - \rho^\delta + L^* L \frac{d}{dt} \rho = 0 \, , \text{ with } \rho(0) = \rho_0 = 0 \, , \tag{7}$$

which can be rewritten as

$$\frac{d}{dt} \rho + (L^* L)^\dagger \rho = (L^* L)^\dagger \rho^\delta \, , \tag{8}$$

where $(L^*L)^\dagger$ is a left inverse of L^*L.

The formal solution of the homogeneous part of the differential equation (8)

$$\frac{d}{dt}\rho + (L^*L)^\dagger \rho = 0\,, \tag{9}$$

with $\rho(0) = f$ can be characterized via semi-group theory (see e.g. [Paz83]) as

$$\rho(t) = exp(-(L^*L)^\dagger t) f\ . \tag{10}$$

Using the terminology of semi-group theory the operator $-(L^*L)^\dagger$ is called *generator* and the set of operators

$$T(t) = exp(-(L^*L)^\dagger t)$$

forms the semi-group associated with the generator $-(L^*L)^\dagger$. Typically in semi-group theory the generator is a differential operator, while in our application the generator is an integral operator.

From standard results on semi-group theory it follows that the solution of the inhomogeneous equation (8) with initial data $\rho(0) = 0$ is given by

$$\rho(t) = \int_0^t T(t-s)(L^*L)^\dagger \rho^\delta\, ds\ . \tag{11}$$

Since ρ^δ does not depend on time it follows that

$$\int_0^t T(t-s)(L^*L)^\dagger \rho^\delta\, ds = \big(I - exp(-(L^*L)^\dagger t\big)\, \rho^\delta\,,$$

and consequently

$$\rho(t) = (I - T(t))\rho^\delta\ . \tag{12}$$

In the sequel we make frequent use of this representation.

To highlight some important features of this denoising technique we set $L = \nabla$, applied to functions defined on a domain Ω, with either homogeneous Neumann or Dirichlet data on the boundary $\partial\Omega$. The Laplace transform,

$$\hat{w}(s,x) = \int_0^\infty e^{-st} w(x,t)\, dt\,,$$

satisfies

$$s\hat{w}(s,x) = \frac{\hat{dw}}{dt}(s,x) + w(0,x)\ .$$

Therefore, from (7), it follows that

$$\hat{\rho}(s,x) - s\Delta\hat{\rho}(s,x) = \hat{\rho}^\delta(x)\ . \tag{13}$$

For fixed $s > 0$ this is a linear, elliptic partial differential equation of second order. The diffusivity in the above differential equation increases with increasing s and hence the solution of the differential equations (13) becomes increasingly regular. Indeed, $\hat{\rho}(s,x)$ converges to a constant function as $s \to \infty$.

The derivation above of the denoising technique and the method for evaluating unbounded operators gives rise to new filtering and evaluation techniques by successively minimizing the functional

$$f_n(\rho) := \|\rho^\delta - \rho\|^2 + \alpha_n \int_\Omega g(\|\nabla\rho - \nabla\rho_{n-1}\|^2)\,,$$

where g is a positive, continuously differentiable function.

With this functional we associate the differential equation

$$\rho - \rho^\delta = \nabla\cdot\left(g'\left(\left\|\nabla\frac{\partial\rho}{\partial t}\right\|^2\right)\right)\ .$$

These nonlinear filtering techniques will not be studied in this paper. For the sake of simplicity of the presentation we consider only the linear equation (7) and the corresponding regularization technique (5).

The outline of the paper is the following. In Section 2 we give some physical motivation for the filtering technique (7). Architectural axioms of the filtering technique (7) are discussed in Section 3 and in Section 4 we study the nonstationary iterated Tikhonov-Morozov method as a stable implicit scheme for the numerical solution of (7). In Section 5 we present two numerical examples. The first uses nonstationary iterative regularization for denoising of an ultrasound image and the second illustrates the use of the technique for the stable evaluation of a differential operator.

2 Physical Models for Diffusion Filtering and Third Order Differential Equations for Denoising

The diffusion filtering techniques (2), (3), and (4) are motivated by *continuum* and *fluid mechanics* (see e.g. [AGLM93]); for some background on fluid mechanics we refer to [CM93]). In order to give a physical interpretation of (7) we review the derivation of the diffusion filtering techniques (2), (3), and (4).

In fluid mechanics *conservation of mass* is expressed by the differential equation

$$\frac{\partial \rho}{\partial t} = \nabla \cdot (\rho \vec{v}) \,, \tag{14}$$

where ρ denotes the *density* and $\vec{v}$ denotes the *velocity field* of the fluid. For fixed time t we consider a *level curve* of the density, i.e., let $(x(\tau), y(\tau))$ be a curve in Ω such that

$$\rho(x(\tau), y(\tau)) = \text{ constant } .$$

Differentiating this equation we find that

$$\frac{\partial \rho}{\partial x}(x(\tau), y(\tau))\dot{x}(\tau) + \frac{\partial \rho}{\partial y}(x(\tau), y(\tau))\dot{y}(\tau) = 0 \,. \tag{15}$$

Now, we note that $(x(\tau), y(\tau))$ is the parameterization of the level curve, and thus the direction of the tangent at the level curve at the point $(x(\tau), y(\tau))$ is $(\dot{x}(\tau), \dot{y}(\tau))$. From (15) it follows that the direction orthogonal to the tangent is the vector $\left(\frac{\partial \rho}{\partial x}, \frac{\partial \rho}{\partial x}\right)(x(\tau), y(\tau))$. Diffusion moves the density in a direction which is orthogonal to the level curves. Then, from (15), it follows that the velocity field $\vec{v}$, describing the movement of the density, is orthogonal to the tangent vector $(\dot{x}(\tau), \dot{y}(\tau))$, i.e.,

$$\vec{v} := C \left(\frac{\partial \rho}{\partial x}, \frac{\partial \rho}{\partial y} \right) \,.$$

There are various ways to model the multiplier C:

1. If $C = 1/\rho$, then we get the diffusion filtering technique (2). This choice of C represents the fact that a moving object in a fluid of higher density moves slower than an object in a fluid of lower density.

2. If $C = 1/(\rho\|\nabla\rho\|)$, then we get the diffusion filtering technique (3) with $g(s) = 1/s$. The choice of C represents the fact that a moving object in a fluid moves faster at smooth portions of level curves. This model has been used in studies of droplet formation (see e.g. Osher and Sethian [OS88], see also e.g. [HO98]).

In image processing the grey value of an image can be interpreted as the density of a fluid, and hence we can use this analogy to derive similar interpretations for images.

Let $L = \nabla$. Formal differentiation of (7) with respect to t gives

$$\frac{d\rho}{dt} - \frac{d\rho^\delta}{dt} - \Delta\frac{d^2\rho}{dt^2} = 0 \; .$$

Since ρ^δ is independent of t we get

$$\frac{d\rho}{dt} = \Delta\frac{d^2\rho}{dt^2} \; . \tag{16}$$

Interpreting ρ again as the density of the fluid: let $C = \frac{1}{\rho}$ and assume that velocity is

$$\vec{v} = C\left(\frac{\partial^3\rho}{\partial x\partial t^2}, \frac{\partial^3\rho}{\partial y\partial t^2}\right) \; .$$

Then a solution of (16) satisfies (14) with this particular velocity field $\vec{v}$. In the model derived above, the velocity is assumed to involve the derivative with respect to time. That is, we take into account the *history* of the process, corresponding to fluids with *memory*.

3 Architectural Structure of Image Processing

Alvarez, Guichard, Lions, Morel [AGLM93] established architectural axioms for a good image processing method. In the following we show that the denoising technique (7) satisfies architectural axioms of image processing such as *causality* and *regularity*.

First we make note of a couple of basic properties of the semigroup $T(t)$. For $t \to 0$ we have

$$(I - T(t)) = I - \exp(-(L^*L)t) \to 0$$

and thus in turn

$$\rho(0) = 0 \ .$$

For $t \to \infty$ we have

$$(I - T(t)) = I - \exp(-(L^*L)t) \to I$$

and thus in turn

$$\rho(\infty) = \rho^\delta \ .$$

We now show that the filtering technique (7) satisfies a *causality axiom*. The original meaning of causality is that for any $s > 0$, $\rho(t+s)$ can be calculated from $\rho(t+\tau)$ where $\tau < s$, without using information on $\rho(\psi)$ with $\psi < t$.

In our setting we have

$$\begin{aligned} \rho(t+s) &= \big(I - \exp(-(L^*L)^\dagger(t+s))\big)\,\rho^\delta \\ &= \big(I - \exp(-(L^*L)^\dagger t)\exp(-(L^*L)^\dagger s)\big)\,\rho^\delta \\ &= \exp(-(L^*L)^\dagger s)\,\big(I - \exp(-(L^*L)^\dagger t)\big)\,\rho^\delta \\ &\quad - \big(\exp(-(L^*L)^\dagger s) - I\big)\,\rho^\delta \ . \end{aligned}$$

Thus for small values of s we have (independently of t)

$$\rho(t+s) = \exp(-(L^*L)^\dagger s)\rho(t) + o(1) \ .$$

Consequently in our situation the diffusion process is "almost causal".

The *regularity axiom* formulated in Alvarez, Guichard, Lions, Morel [AGLM93] is that for all $h, t \in [0,1]$ and smooth functions f, g

$$\|\rho(t)(f + hg) - \rho(t)f - hg\|_{L^\infty} \leq Cht \,,$$

where C is a constant independent of h and t, but dependent on f and g. The notation $\rho(t)(f)$ indicates that ρ is generated with initial data f.

For the diffusion filtering technique (7) we have

$$\begin{aligned} \rho(t)(f+hg) - \rho(t)f - hg &= (I - \exp(-(L^*L)^\dagger t))hg - hg \\ &= \exp(-(L^*L)^\dagger t)hg \; . \end{aligned}$$

Let us assume that the eigenvalues of $(L^*L)^\dagger$ are in $]0,1[$. Then

$$\exp(-(L^*L)^\dagger t)g = \int_0^1 \exp(-\lambda t) dE_\lambda g \; .$$

Consequently,

$$\| \exp(-(L^*L)^\dagger t)g \| = \int_0^1 \exp(-2\lambda t) d\|E_\lambda g\|^2$$

is monotonically decreasing in t and

$$\lim_{t\to 0} \| \exp(-(L^*L)^\dagger t)g\| = \|g\|^2 \text{ and } \lim_{t\to\infty} \| \exp(-(L^*L)^\dagger t)g\| = 0 \; .$$

This shows that the diffusion filtering technique satisfies the regularity condition at $t = \infty$.

4 Theoretical Results on Nonstationary Tikhonov Regularization

We now treat the time–discretized system of the diffusion filtering technique (7). Note that for a numerical solution of (7) the diffusion process (7) has to be discretized in time. If we solve the time discretized systems of (7) at time $0, t_1, \ldots,$ then this is equivalent to using nonstationary iterative regularization with regularization parameters $\alpha_i = \frac{1}{t_i - t_{i-1}}$. Thus nonstationary iterative regularization can be considered as a numerical algorithm for the solution of (7). Consequently, nonstationary iterative regularization can be considered as a semi-implicit approximation of the diffusion filtering technique (5), where the regularization parameter α_n serves as the discretization parameter in (7). In the following we analyze properties of nonstationary iterative regularization.

In the following we require some notation: Let

$$\tilde{L} := L^* L \,, \hat{L} := L L^*$$

and let $\{\alpha_n\}$ be a sequence of positive parameters which are all less than 1. Moreover, let

$$A_n = I - (I + \alpha_n \hat{L})^{-1} \,, B_n = I - (I + \alpha_n \tilde{L})^{-1} \text{ and } C_n = L(I + \alpha_n \tilde{L})^{-1} \,.$$

In the following we summarize some well–known results from operator theory. The operators $(I + \alpha_n \tilde{L})^{-1}$ and $(I + \alpha_n \hat{L})^{-1}$ are bounded, self-adjoint and satisfy

$$0 \leq (I + \alpha_n \tilde{L})^{-1} \leq I, \text{ and } 0 \leq (I + \alpha_n \hat{L})^{-1} \leq I \,.$$

C_n is a bounded linear operator with

$$\|C_n\| \leq \alpha_n^{-1/2}$$

and for $z \in \mathcal{D}(L)$

$$C_n z = (I + \alpha_n \hat{L})^{-1} L z \,.$$

Moreover,

$$\|A_n\| \leq 1 \text{ and } \|B_n\| \leq 1 \,.$$

We now discuss convergence of the iterative regularization technique (5). Here and in the sequel we use the following convention: ρ_n^δ denotes the minimizer of the functional f_n and ρ_n denotes the minimizer of the functional f_n where ρ^δ is replaced by $\rho^\dagger$.

From the definition of (5) it follows that

$$\rho_n = (I + \alpha_n \tilde{L})^{-1} \rho^\dagger + \alpha_n (I + \alpha_n \tilde{L})^{-1} \tilde{L} \rho_{n-1} \,,$$

and

$$L\rho_n - L\rho^\dagger = A_n (L\rho_{n-1} - L\rho) = -\left(\prod_{i=1}^n A_i \right) L\rho^\dagger \,. \tag{17}$$

Since

$$\left\| \prod_{i=1}^n A_i \right\| \leq 1 \text{ for all } n \in \mathbb{N}$$

and since for all $\lambda \geq 0$

$$\prod_{j=1}^{n} \left(1 - \frac{1}{1+\alpha_j \lambda}\right) \to 0 \text{ as } n \to \infty\,,$$

it follows that

$$L\rho_n \to L\rho^\dagger \text{ as } n \to \infty\ .$$

Moreover, we have

$$\begin{aligned} L\rho_n &= C_n\rho^\dagger + \alpha_n \hat{L}(I+\alpha_n\hat{L})^{-1}L\rho_{n-1} \\ &= C_n\rho^\dagger + (I-(I+\alpha_n\hat{L})^{-1})L\rho_{n-1}\ . \end{aligned}$$

Therefore,

$$\begin{aligned} L\rho_n &= C_n\rho^\dagger + A_n L\rho_{n-1} \\ &= C_n\rho^\dagger + A_n C_{n-1}\rho^\dagger + A_n A_{n-1} L\rho_{n-2} \\ &. \\ &. \\ &= C_n\rho^\dagger + A_n A_{n-1} C_{n-2}\rho^\dagger + \cdots (A_n....A_2)C_1\rho^\dagger \end{aligned}$$

and a similar representation holds for $L\rho_n^\delta$. From the estimates $\|A_j\| \leq 1$ and $\|C_j\| \leq \alpha_j^{-1}$ it follows that

$$\|L\rho_n^\delta - L\rho_n\| \leq \delta\sigma_n \text{ where } \sigma_n = \sum_{k=1}^{n} \alpha_k^{-1}\ . \tag{18}$$

Similarly to (17) we find that

$$\rho_n - \rho^\dagger = -\left(\prod_{j=1}^{n} B_i\right)\rho^\dagger \to 0 \text{ as } n \to \infty\ . \tag{19}$$

We now summarize the above considerations:

If $n = n(\delta) \to \infty$ as $\delta \to 0$, while $\delta\sigma_{n(\delta)} \to 0$, then

$$\rho_{n(\delta)}^\delta \to \rho\ .$$

We note that the condition $\delta\sigma_{n(\delta)} \to 0$ can be weakened to $\delta \sum_{j=1}^{n(\delta)} \alpha_j^{-1/2} \to 0$. This condition states that an approximation to $\rho^\dagger$ can be calculated with (7)and a reasonable approximation to $\rho^\dagger$ is gained for a not too large scale time (which depends on the noise level δ).

In the following we are concerned with asymptotically estimating the optimal number of discretizations in the nonstationary iterated Tikhonov–Morozov method. This corresponds to finding the optimal number of time discretizations for the differential equation (7).

In the following we assume that

$$\|\rho^\delta\| > \tau\delta, \text{ for some } \tau > 1 \ . \tag{20}$$

That is, we assume that the signal–to–noise ratio is strictly greater than 1. This assumption means that the signal contains more information than noise. If (20) is violated, then we consider the signal as totally destroyed by noise, and there is insufficient information available to recover the original signal.

From the representation

$$\rho_n^\delta - \rho^\delta = B_n(\rho_{n-1}^\delta - \rho^\delta)\,,$$

and the fact that $\|B_n\| \leq 1$ it follows that the function

$$f^\delta(n) := \|\rho_n^\delta - \rho^\delta\| \text{ decreases with } n\ .$$

In the following we denote by $n = n(\delta)$ the first index for which

$$\|\rho_{n(\delta)}^\delta - \rho^\delta\| \leq \tau\delta\ . \tag{21}$$

From (19) it follows that

$$\begin{aligned}
\rho_{n-1} - \rho^\dagger &= \rho_{n-1}^\delta - \rho^\delta + \rho_{n-1} - \rho_{n-1}^\delta - (\rho - \rho^\delta) \\
&= \rho_{n-1}^\delta - \rho^\delta - \left(\prod_{j=1}^{n-1} B_j\right)(\rho^\dagger - \rho^\delta)\ .
\end{aligned}$$

Hence

$$\begin{aligned}
\|\rho_{n(\delta)-1} - \rho^\dagger\| &\geq \|\rho_{n(\delta)-1}^\delta - \rho^\dagger\| - \left\|\left(\prod_{j=1}^{n(\delta)-1}\right)(\rho^\delta - \rho^\dagger)\right\| \\
&\geq (\tau - 1)\delta.
\end{aligned} \tag{22}$$

Suppose now that there exists a $w \in H_1$ and a $\nu > 0$ such that

$$\rho^\dagger = (I + \tilde{L})^{-\nu} w \, . \tag{23}$$

That is, assume that $\rho^\dagger$ satisfies a source wise representation. In many applications (23) implies that $\rho^\dagger$ satisfies some smoothness constraints. From (23) it follows that

$$\rho_m - \rho^\dagger = -\left(\prod_{j=1}^{m} B_j\right)(I + \tilde{L})^{-\nu} w \, .$$

Using some estimates from [HG98] we see that

$$\begin{array}{rcl} \max_{\lambda \geq 0}(1+\lambda)^{-\nu} \prod_{j=1}^{m}(1+(1+\alpha_j\lambda)^{-1}) & = & \max_{\lambda\geq 0}(1+\lambda)^{-\nu}\lambda^{\nu}\prod_{j=1}^{m}\frac{\alpha_j}{\lambda+\alpha_j} \\ & \leq & \max_{\lambda\geq 0}\lambda^{\nu}\prod_{j=1}^{m}\frac{\alpha_j}{\lambda+\alpha_j} \\ & = & O(\sigma_m^{-\nu}) \, . \end{array}$$

Therefore, we have

$$\left\|\left(\prod_{j=1}^{m} B_j\right)(I+\tilde{L})^{-\nu} w\right\| \leq \max_{\lambda\geq 0}(1+\lambda)^{-\nu}\prod_{j=1}^{m}(1+(1+\alpha_j\lambda)^{-1})\|w\| = O(\sigma_m^{-\nu}) \, .$$

Thus it follows from (18) that

$$(\tau - 1)\delta \leq \|\rho_{n(\delta)-1} - \rho^\dagger\| = O(\sigma_m^{-\nu}) \, ,$$

which in turn gives

$$\sigma_{n(\delta)} = O(\delta^{-1/\nu}) \, .$$

Then from (22) it follows that

$$\|L\rho_{n(\delta)}^{\delta} - L\rho_{n(\delta)}\| \leq \delta\sigma_{n(\delta)} = O(\delta^{1-1/\nu}) \, . \tag{24}$$

Now, we derive estimates for $\|L\rho_{n(\delta)}^{\delta} - L\rho^\dagger\|$, which together with (24) gives an estimate for $\|L\rho_{n(\delta)} - L\rho^\dagger\|$. From (19) we have

$$\begin{array}{rcl} \|L\rho_n - L\rho^\dagger\| & = & \left\|\left(\prod_{j=1}^{n} A_j\right) L\rho^\dagger\right\| \\ & = & \left\|(I+\hat{L})^{-\nu}\left(\prod_{j=1}^{n} A_j\right) Lw\right\| \, . \end{array} \tag{25}$$

Let $z = \left(\prod_{j=1}^{n} A_j\right) Lw$. From the fact that $\|A_j\| \le 1$, it follows that (note that z is dependent on n)

$$\|z\| \le \|L\rho^\dagger\| \text{ for all } n \in I\!N \ .$$

Using an interpolation inequality it follows that

$$\|L\rho_n - L\rho^\dagger\| \le \|z\|^{1/\nu} \left\|(I+\hat{L})^{-\nu^2/(\nu-1)} z\right\|^{\nu-1/\nu} \ .$$

From (19) we find that

$$\begin{aligned}(I+\hat{L})^{-\nu^2/(\nu-1)} z &= L(I+\tilde{L})^{-1}[(I+\tilde{L})^{-1}]^{1/(\nu-1)} \left(\prod_{j=1}^{n} B_j\right) (I+\tilde{L})^{-\nu} w \\ &= L(I+\tilde{L})^{-1}[(I+\tilde{L})^{-1}]^{1/(\nu-1)} \left(\prod_{j=1}^{n} B_j\right) \rho^\dagger \\ &= -L(I+\tilde{L})^{-1}[(I+\tilde{L})^{-1}]^{1/(\nu-1)} (\rho_n - \rho^\dagger) \ .\end{aligned}$$

Since both $L(I+\tilde{L})^{-1}$ and $(I+\tilde{L})^{-1}$ are bounded we get

$$\|L\rho_n - L\rho^\dagger\| = O\left(\|\rho_n - \rho^\dagger\|^{1-1/\nu}\right) \ .$$

Thus from (19) we obtain

$$\begin{aligned}\rho^\delta_{n(\delta)} - \rho^\dagger &= \rho^\delta_{n(\delta)} - \rho^\delta + \rho_{n(\delta)} - \rho^\delta_{n(\delta)} - (\rho^\dagger - \rho^\delta) \\ &= \rho^\delta_{n(\delta)} - \rho^\delta - \left(\prod_{j=1}^{n(\delta)} B_j\right) (\rho - \rho^\delta) \ .\end{aligned}$$

We then have

$$\|\rho^\delta_{n(\delta)} - \rho^\dagger\| \le (\tau + 1)\delta \ .$$

We summarize the above calculations in the following theorem: Let $n(\delta)$ be choose according to (21) and let $\rho^\dagger$ satisfy (23). Then

$$\|L\rho^\delta_{n(\delta)} - L\rho^\dagger\| = O(\delta^{1-1/\nu}) \ .$$

We note that this theorem establishes that it is in principle possible to stably approximate $L\rho^\dagger$ to an accuracy that is arbitrarily close to the accuracy of the data.

5 Numerical Results

We now present some numerical experiments. The goal of this section is twofold. First of all we show that (7) is an effective method for denoising images. For this we present a scale of filtered images which were calculated by using nonstationary iterative regularization, and which provide an approximation of the solution of (7). Secondly we use nonstationary iterative Tikhonov regularization to calculate derivatives of a noisy function.

5.1 Denoising of images

First we consider an ultrasound data set. We implemented nonstationary iterative Tikhonov regularization. The scale of images which correspond to the solution of (7) at different times are represented in Figure 1. One sees that the process starts with zero initial data and picks up more and more details until it finally reaches the input data ρ^δ as a stationary case. This scale of images has been generated without using the discrepancy principle for termination. The discrepancy principle enables us to effectively determine the gradient in a stable way (see the subsection below) and it is not primarily designed to effectively denoise images.

5.2 Stable calculation of a derivative

We now take up the stable numerical calculation of a derivative from given sampled data. We implemented nonstationary iterative Tikhonov regularization and calculated numerical the derivative by a finite difference approximation. The results of nonstationary iterative regularization are plotted on the left side of Figure 2. The corresponding numerical derivatives are plotted on the right hand side of Figure 2. The iteration process was stopped when the discrepancy principle became active. From the numerical results it is evident that for the calculation for the derivative it is essential to stop the iteration appropriately. The numerical results also show that smooth functions and their derivatives can be recovered more accurately.

Acknowledgements

The numerical results were implement using the software packages DIFFPACK [DT97] and MATLAB. The authors thank Walter Hinterberger (University of Linz) for his support in the implementation using DIFFPACK. The work of O.S. is supported by the Austrian Fonds zur Förderung der Wissenschaftlichen Forschung (FWF), F 1310. C.W.G. was supported in part by the National Science Foundation.

References

[AGLM93] L. Alvarez, F. Guichard, P.-L. Lions, and J.-M. Morel. Axioms and fundamental equations of image processing. *Arch. Ration. Mech. Anal.*, 123:199–257, 1993.

[CM93] A. J. Chorin and J. E. Marsden. *A Mathematical Introduction to Fluid Mechanics.* Springer-Verlag, New York, 1993. 3. ed.

[DT97] M. Daehlen and A. Tveito. *Numerical Methods and Software Tools in Industrial Mathematics.* Birkhäuser, Boston, 1997.

[HG98] M. Hanke and C.W. Groetsch. Nonstationary iterated Tikhonov regularization. *J. Optim. Theory and Applications*, 98:37–53, 1998.

[HO98] E. Harabetian and S. Osher. Regularization of ill–posed problems via the level set approach. *SIAM J. Appl. Math*, 58:1689–1706, 1998.

[OS88] S. Osher and J. A. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations. *J. Comput. Phys.*, 79:12–49, 1988.

[Paz83] A. Pazy. *Semigroups of Linear Operators and Applications to Partial Differential Equations.* Springer–Verlag, New York, 1983.

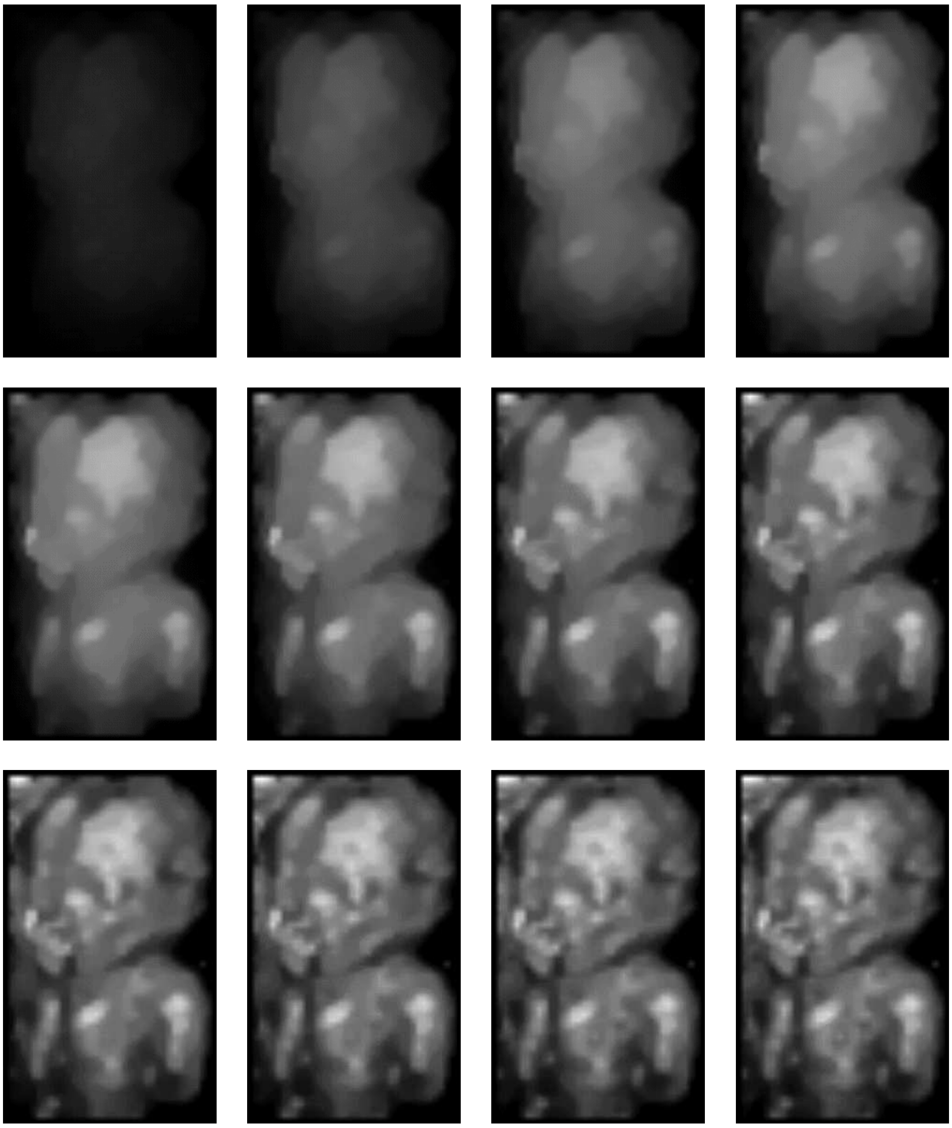
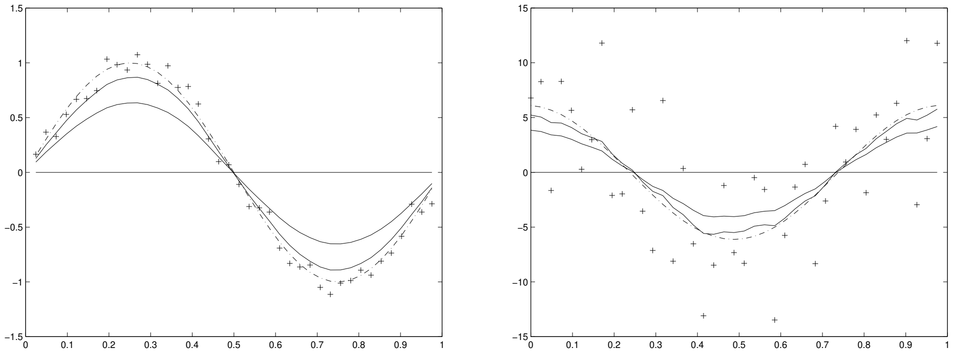
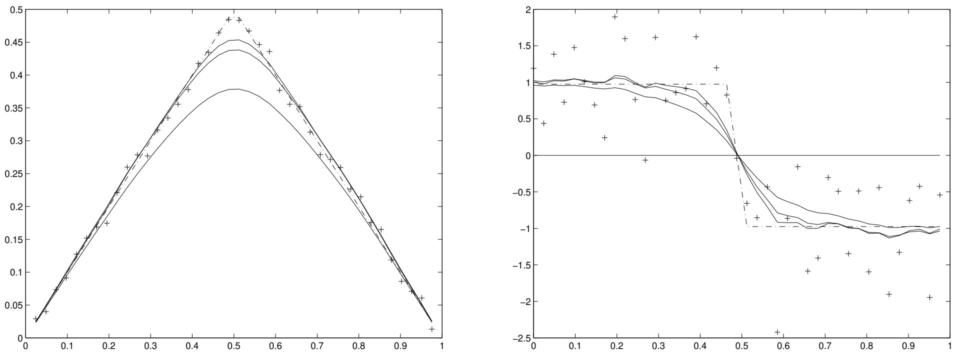


Figure 1: The result of nonstationary regularization with regularization parameters $\alpha_1 = 100000$, $\alpha_2 = 50000$, $\alpha_3 = 25000$, $\alpha_4 = 12500$, $\alpha_5 = 6250$, $\alpha_6 = 3125$, $\alpha_7 = 1500$, $\alpha_8 = 750$, $\alpha_9 = 300$, $\alpha_{10} = 150$, $\alpha_{11} = 75$, and $\alpha_{12} = 30$



Nonstationary iterative regularization, $\alpha_i = 1000$; stopping criterion: discrepancy principle $\tau = 1.5$



Nonstationary iterative regularization, $\alpha_i = 1000$; stopping criterion: discrepancy principle $\tau = 1.5$

Figure 2: $+$ the sampled data and its numerically calculated derivate, $-.$ the exact data, $-$ several iterations with nonstationary iterative regularization