SAMPLING THEORY IN SIGNAL AND IMAGE PROCESSING Vol. 3, No. 3, Sept. 2004, pp. 225-246

ISSN: 1530-6429

Regularization of Sobolev Embedding Operators and Applications Part II: Data Driven Regularization and Applications

Ronny Ramlau and Gerd Teschke
Universität Bremen, Zentrum für Technomathematik
Postfach 33 04 40
28334 Bremen
Germany
ramlau@math.uni-bremen.de, teschke@math.uni-bremen.de

Abstract

This paper is concerned with the construction of a data-driven parameter rule which is essential in the context of regularization methods for inverse problems. In particular, we consider the inversion problem of Sobolev embedding operators. For a Fourier- and wavelet-based treatment of this problem we refer the reader to our previous paper *Regularization of Sobolev Embedding Operators and Applications Part I: Fourier and Wavelet based Methods.* Here we focus on developing a related data-driven scheme for an optimal choice of the regularization parameter. We prove order optimal convergence of the resulting regularization scheme, and, moreover, show by two examples the applicability of the proposed method.

Key words and phrases : Parameter rule, Sobolev embedding operator, Tikhonov regularization, Statistical noise model

2000 AMS Mathematics Subject Classification — 65C20, 65J10, 65J20, 65J22, 94A12, 94A20

1 Introduction

We consider the inversion problem of Sobolev embedding operators, i.e., we want to solve the linear equation $ix = x$ from noisy data x^δ, where

i denotes the embedding operator $i : H^s(\Omega) \to L_2(\Omega)$ for some $s > 0$. As it is well-known that i is an ill-posed operator, we have to apply regularization methods. As suggested in [8], we will particularly focus on Tikhonov-regularization. The variational formulation is then given by

$$J_\alpha(x) = \|x^\delta - ix\|_2^2 + \alpha\|x\|_s^2 \ , \tag{1.1}$$

where the minimizing element, here expressed by means of a Fourier series expansion, has the following form

$$x_\alpha^\delta = T_\alpha(x^\delta) = \sum_k \frac{(x^\delta)_k}{1+\alpha(1+k^2)^s} e^{ik\cdot}. \tag{1.2}$$

In the classical theory, the choice of the parameter α is connected to the error level δ with $\|x - x^\delta\|_2 \leq \delta$, and is used to prove convergence rates of the approximated solution. A common assumption in this framework is that the noise level is a-priori known.

However, when it comes to any sort of practical applications, this a-priori information is not or only roughly known. This fact very often results in unstable or suboptimal numerical approximations of the true solution. To this end, a serious estimation of the noise level is required and would essentially improve the quality of numerical results. In contrary to the classical framework, we therefore suggest to switch to a statistical setting and assume that only the noise model is known (not the noise level). Under this weaker assumption we propose a data-driven principle which automatically detects optimal regularization parameters.

The main results in this paper are the construction of a noise level detector and a related parameter rule, and a proof of order optimal convergence of the resulting regularization scheme. Since the statistical setting requires an adequate number of observations, the optimality result might become somewhat biased if the number of observations decreases. The proposed machinery applied to general linear inverse and, moreover, to nonlinear problems will be discussed in forthcoming papers.

The paper is organized as follows: in Section 2 we introduce the statistical setting, and establish a noise level detector. In Section 3 we construct a data-driven Morozov discrepancy principle (parameter rule) and give a proof of order optimality of the proposed scheme. In Section 4 we suggest further parameter rules which are numerically quite effective but where no proof of convergence has come in yet. In Section 5 we finally show by two examples arising in the context of medical imaging and meteorological signal processing the capability of the proposed scheme.

The proposed techniques and proofs in this paper are throughout given in a one-dimensional framework. They carry over without any difficulties to higher dimensions.

2 Noise Level Detection

Let $t_0, \ldots, t_{N-1}$ be a regular grid on $\Omega = [0, 2\pi)$, i.e., $t_k = 2\pi k/N$, $k = 0, \ldots, N-1$. We assume that the data sampled on this grid satisfy an additive noise model

$$x^\varepsilon(t_k) = x^\varepsilon_{t_k} = x_{t_k} + \varepsilon_{t_k} \ , \tag{2.1}$$

where the random variables $\varepsilon_{t_k} : \omega \mapsto \varepsilon_{t_k}(\omega)$ are independently identically distributed with $\mathsf{E}\varepsilon_{t_k} = 0$ and $\mathsf{Cov}\varepsilon_{t_k}\bar{\varepsilon}_{t_l} = \sigma^2\delta_{kl}$. Here the standard deviation σ can be identified with the noise level δ in the classical framework.

At first, by common statistical computation we derive some properties of ε which are frequently used later on. The discrete Fourier coefficients are given by

$$\hat{\varepsilon}_k = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \varepsilon_{t_n} e^{-ikt_n} \ . \tag{2.2}$$

Hence, we have $\mathsf{E}\hat{\varepsilon}_k = 0$, and $\mathsf{Cov}\hat{\varepsilon}_k\bar{\hat{\varepsilon}}_l = \sigma^2\delta_{kl}$. On the basis of (1.2) we may define

$$\tilde{\varepsilon}_{t_l} := i(T_\alpha\varepsilon)(t_l) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \hat{\varepsilon}_k \left(1 + \alpha(1+n^2)^s\right)^{-1} e^{i2\pi nl/N} \ . \tag{2.3}$$

We observe that $\mathsf{E}\tilde{\varepsilon}_{t_l} = 0$ and

$$\mathsf{Cov}\tilde{\varepsilon}_{t_l}\bar{\tilde{\varepsilon}}_{t_k} \quad = \quad \frac{\sigma^2}{N} \sum_{n=0}^{N-1} \left(1 + \alpha(1+n^2)^s\right)^{-2} e^{i2\pi n(k-l)/N} \ . \tag{2.4}$$

In order to incorporate the statistical setup in our inverse problem we need to rewrite the Tikhonov variational functional (1.1) and to establish an estimation for the noise level σ^2. We begin by adapting the variational formulation

$$J_{\alpha,\varepsilon}(x) := \mathsf{E}\|x^\varepsilon - ix\|_2^2 + \alpha\|x\|_s^2 \quad , \tag{2.5}$$

where the norm is vector-valued. The minimizing element is given by $x_\alpha = \mathsf{E}x^\varepsilon_\alpha$. As the next step we introduce residuum random variables

$$r^\varepsilon_\alpha(t_k) := N^{-1/2}(x_\alpha(t_k) - x^\varepsilon(t_k)). \tag{2.6}$$

We note that $\mathsf{E}r^\varepsilon_\alpha(t_k) = N^{-1/2}(x_\alpha(t_k) - x(t_k))$. Consequently, we have

$$\mathsf{E}\|r^\varepsilon_\alpha - \mathsf{E}r^\varepsilon_\alpha\|_2^2 = \mathsf{E}\|N^{-1/2}\varepsilon\|_2^2 = \sigma^2. \tag{2.7}$$

On the basis of (2.7) we aim at constructing an estimator for σ^2.

In the following we prove by standard techniques a sequence of Lemmatas essential for approximating the noise level σ^2. We start with some well-known observation, see [4]: for M given observations $\{\tilde{\xi}_m\}_{1..M}$ of some vector $\{\xi_m\}_{1..M}$ of i.i.d. sample random variables one might consider the following mapping

$$\mathcal{E} : \{\tilde{\xi}_m\}_{m=1..M} \mapsto \frac{1}{M}\sum_{m=1}^{M} \tilde{\xi}_i \ . \tag{2.8}$$

It is shown that $\mathsf{E}\mathcal{E}\{\xi_m\}_{m=1..M} = \mathsf{E}\xi_m$ and $\mathsf{Var}\mathcal{E}\{\xi_m\}_{m=1..M} = M^{-1}\sigma^2$, where σ^2 here denotes the variance of ξ_m.

Mapping (2.8) shall be used for our purpose to approximate $\mathsf{E}x^\varepsilon_\alpha$ and $\mathsf{E}r^\varepsilon_\alpha$. To this end, assume we are given data sample random variables $\{x^\varepsilon_m\}_{1..M}$. Then the minimizing random variables, derived by (1.2), are given by $\{x^\varepsilon_{\alpha,m}\}_{m=1..M}$.

Lemma 2.1 *Assume that the minimizing random variables are given by* $\{x^\varepsilon_{\alpha,m}\}_{m=1..M}$ *and that the regularization is given by* T_α. *Then*

1. $\mathcal{E}$ *is unbiased with respect to the expectation* x_α, *i.e.,*

$$\mathsf{E}\mathcal{E}\{x^\varepsilon_{\alpha,m}\}_{m=1..M} = \mathsf{E}x^\varepsilon_{\alpha,m} = x_\alpha \ ,$$

2. *the approximation* $\mathcal{E}(\cdot)$ *converges with rate* M^{-1}

$$\mathsf{Var}\mathcal{E}\{x^\varepsilon_{\alpha,m}\}_{m=1..M} = M^{-1}\sigma^2|||T_\alpha|||^2 \ .$$

Proof The first assertion is clear. To prove the second assertion we assume that T_α denotes the regularization given by (1.2), then we obtain

$$\mathsf{Var}\mathcal{E}\{x^\varepsilon_{\alpha,m}\}_{m=1..M} \quad = \quad M^{-2}\mathsf{E}|\sum_{m=1}^{M} x^\varepsilon_{\alpha,m} - \mathsf{E}\sum_{m=1}^{M} x^\varepsilon_{\alpha,m}|^2$$

$$
\begin{aligned}
&= M^{-2}\mathsf{E}|\sum_{m=1}^{M} T_\alpha \varepsilon_m|^2 \\
&= M^{-1}\sigma^2 |||T_\alpha|||^2 \ ,
\end{aligned}
$$

where $|||T_\alpha|||^2 = N^{-1}\sum_{n=0}^{N-1}(1+\alpha(1+n^2)^s)^{-2}$. ■

To approximate (2.6) we define residuum random sample variables

$$
r^\varepsilon_{\alpha,m} := N^{-1/2}(\mathcal{E}\{x^\varepsilon_{\alpha,m}\}_{m=1..M} - x^\varepsilon_m). \tag{2.9}
$$

Lemma 2.2 *Assume that the residuum random variables are given by* $\{r^\varepsilon_{\alpha,m}\}_{m=1..M}$ *and that the regularization is given by* T_α*. Then*

1. $\mathcal{E}$ *is unbiased with respect to the expectation* $\mathsf{E}r^\varepsilon_\alpha$*, i.e.,*

$$
\mathsf{E}\mathcal{E}\{r^\varepsilon_{\alpha,m}\}_{m=1..M} = \mathsf{E}r^\varepsilon_{\alpha,m} \ ,
$$

2. *the approximation* $\mathcal{E}(\cdot)$ *converges with rate* $(NM)^{-1}$

$$
\mathsf{Var}\mathcal{E}\{r^\varepsilon_{\alpha,m}\}_{m=1..M} = \frac{\sigma^2}{NM}|||Id - T_\alpha|||^2 \ .
$$

Proof Assertion *1* is clear by

$$
\begin{aligned}
\mathsf{E}\mathcal{E}\{r^\varepsilon_{\alpha,m}\}_{1..M} &= M^{-1}\sum_{m=1}^{M} N^{-1/2}\left(\mathsf{E}\mathcal{E}\{x^\varepsilon_{\alpha,m}\}_{1..M} - \mathsf{E}x^\varepsilon_m\right) \\
&= M^{-1}\sum_{m=1}^{M} N^{-1/2}\left(\mathsf{E}x^\varepsilon_{\alpha,m} - \mathsf{E}x^\varepsilon_m\right) = \mathsf{E}r^\varepsilon_{\alpha,m} \ .
\end{aligned}
$$

In order to prove assertion *2* we observe that

$$
\begin{aligned}
\mathsf{Var}\mathcal{E}\{r^\varepsilon_{\alpha,m}\}_{1..M} &= \mathsf{E}|\mathcal{E}\{r^\varepsilon_{\alpha,m}\}_{1..M}|^2 - |\mathsf{E}\mathcal{E}\{r^\varepsilon_{\alpha,m}\}_{1..M}|^2 \\
&= \mathsf{E}|\mathcal{E}\{r^\varepsilon_{\alpha,m}\}_{1..M}|^2 - N^{-1}|x_\alpha - x|^2 \ ,
\end{aligned}
$$

where

$$
\begin{aligned}
N\mathsf{E}|\mathcal{E}\{r^\varepsilon_{\alpha,m}\}_{1..M}|^2 &= \mathsf{E}|\mathcal{E}\{x^\varepsilon_{\alpha,m}\}_{1..M}|^2 - \mathsf{E}\mathcal{E}\{x^\varepsilon_m\}_{1..M}\overline{\mathcal{E}\{x^\varepsilon_{\alpha,m}\}}_{1..M} \\
&\quad - \mathsf{E}\mathcal{E}\{x^\varepsilon_{\alpha,m}\}_{1..M}\overline{\mathcal{E}\{x^\varepsilon_m\}}_{1..M} + \mathsf{E}|\mathcal{E}\{x^\varepsilon_m\}_{1..M}|^2 \ .
\end{aligned}
$$

Setting $a_k := (1+\alpha(1+k^2)^s)^{-1}$ we obtain

$$\begin{aligned}
\mathsf{E}|\mathcal{E}\{x^\varepsilon_{\alpha,m}\}_{1..M}|^2 &= M^{-2}\sum_{j,l=1}^{M}\{|x_\alpha|^2+\frac{1}{N}\sum_{k=0}^{N-1}\sigma^2\delta_{j,l}a_k^2\} \ , \\
\mathsf{E}|\mathcal{E}\{x^\varepsilon_{m}\}_{1..M}|^2 &= M^{-2}\sum_{j,l=1}^{M}\{|x|^2+\frac{1}{N}\sum_{k=0}^{N-1}\sigma^2\delta_{j,l}\} \ , \\
\mathsf{E}\mathcal{E}\{x^\varepsilon_{\alpha,m}\}_{1..M}\overline{\mathcal{E}\{x^\varepsilon_{m}\}}_{1..M} &= M^{-2}\sum_{j,l=1}^{M}\{x_\alpha\bar{x}+\frac{1}{N}\sum_{k=0}^{N-1}\sigma^2\delta_{j,l}a_k\} \ , \text{ and} \\
\mathsf{E}\mathcal{E}\{x^\varepsilon_{m}\}_{1..M}\overline{\mathcal{E}\{x^\varepsilon_{\alpha,m}\}}_{1..M} &= M^{-2}\sum_{j,l=1}^{M}\{x\bar{x}_\alpha+\frac{1}{N}\sum_{k=0}^{N-1}\sigma^2\delta_{j,l}a_k\} \ .
\end{aligned}$$

Consequently, we obtain

$$\mathsf{Var}\mathcal{E}\{r^\varepsilon_{\alpha,m}\}_{1..M} = \frac{\sigma^2}{NM}\frac{1}{N}\sum_{k=0}^{N-1}(1-a_k)^2 \ .$$

The estimate

$$\|x-T_\alpha x\|_2^2 \le \|x\|_2^2\frac{1}{N}\sum_{k=0}^{N-1}(1-a_k)^2$$

completes the proof. ■

As the final step we define a mapping $\mathcal{S}$ by

$$\mathcal{S}^2 : \{\tilde{r}^\varepsilon_{\alpha,m}\}_m \mapsto \frac{1}{(M-1)}\sum_{m=1}^{M}\|\tilde{r}^\varepsilon_{\alpha,m}-\mathcal{E}\{\tilde{r}^\varepsilon_{\alpha,j}\}_{j=1..M}\|_2^2 \ . \tag{2.10}$$

This mapping $\mathcal{S}^2$ is by its definition a natural candidate to estimate the variance σ^2, see formula (2.7).

Lemma 2.3 *The mapping $\mathcal{S}^2$ defined by (2.10) is unbiased with respect to σ^2*

$$\mathsf{E}\mathcal{S}^2\{r^\varepsilon_{\alpha,m}\}_{m=1..M} = \sigma^2 \ .$$

Proof Let $a_k = (1+\alpha(1+k^2)^s)$. For residuum random variables $r^\varepsilon_{\alpha,m}$, $m = 1,\dots,M$, we have

$$\mathsf{E}\langle r^\varepsilon_{\alpha,l}, r^\varepsilon_{\alpha,p}\rangle = \frac{1}{N}\|(T_\alpha - Id)x\|_2^2$$

$$+\frac{\sigma^2}{N}\left(M^{-1}\sum a_k^2 - 2M^{-1}\sum_j \delta_{m,j}\sum_k a_k + \delta_{m,l}\sum_k 1\right) .$$

This leads to

$$\begin{aligned}
\mathsf{E}\mathcal{S}^2(r_\alpha^\varepsilon) &= \frac{\sigma^2}{(M-1)N}\left(-\sum_k a_k^2 + 2\sum_k a_k - \sum_k - \sum_k a_k^2 + 2\sum_k a_k\right. \\
&\quad -\sum_k + \frac{1}{M}\sum_{m=1}^{M}\sum_k a_k^2 - \frac{2}{M^3}\sum_{m,n=1}^{M}\sum_{l,j=1}^{M}\delta_{l,j}\sum_k a_k \\
&\quad \left. +\frac{1}{M^3}\sum_{m,l,n=1}^{M}\sum_k + \sum_k a_k^2 - \frac{2}{M}\sum_{m,j=1}^{M}\delta_{m,j}\sum_k a_k + \sum_{m=1}^{M}\sum_k\right) \\
&= \frac{\sigma^2}{(M-1)N}(M-1)\sum_k = \sigma^2 .
\end{aligned}$$

■

Remark 2.1 Note that the analysis presented here is limited to the expectation (bias) of $\mathcal{S}^2$. Usually one aims to analyze the mean squared error $MSE_\theta(\mathcal{S}^2) := \mathsf{E}(\mathcal{S}^2-\theta)^2 = \mathsf{Var}(\mathcal{S}^2) + (\mathsf{E}\mathcal{S}^2-\theta)^2$. Since this would require the computation of higher moments we have abstained from the full analysis.

In order to obtain a reasonable estimate for σ^2 the number M of observation must be sufficiently large. In case of having only one measurement available the established method would fail since the estimator $\mathcal{S}^2$ would diverge. However, in order to overcome this deficiency we might assume that there exists a small neighborhood $U_\varrho(t)$ of each t where the residua r_m^ε are stationary. Then, the mean value can be approximated by taking an average over $U_\varrho(t)$.

3 Data-driven Morozov Principle and Order Optimal Convergence

Once we have established a data-driven estimator for σ^2 we may ask how to construct a parameter rule. The basic idea is to use the estimator in

Morozov's discrepancy principle. We begin by establishing some required estimates.

Proposition 3.1 *Let the noisy data $x^\varepsilon \in L_2(\Omega)$ be given and let x_* be a solution of the inverse problem. Moreover, we assume that there exists $w \in H^s(\Omega)$ such that $x_* = i^*w$ and that x_α minimizes* (2.5)*. Then the estimates*

$$\begin{aligned} \|x_\alpha - x_*\|_s &\leq \frac{\sigma + \alpha\|w\|}{\sqrt{\alpha}} , && (3.1)\\ \mathsf{E}\|x_\alpha - x^\varepsilon\|_2 &\leq \sigma + 2\alpha\|w\| && (3.2)\end{aligned}$$

hold.

Proof Since x_* is a solution we have $i(x_*) = x$. Consequently, we obtain

$$J_{\alpha,\varepsilon}(x_\alpha) \leq \mathsf{E}\|\varepsilon\|^2 + \alpha\|x_*\|^2 = \sigma^2 + \alpha\|x_*\|^2. \tag{3.3}$$

By the Ljapunov inequality, i.e., by $(\mathsf{E}|X|^r)^{1/r} \leq (\mathsf{E}|X|^s)^{1/s}$ for $1 \leq r \leq s$ and a random variable X, we have

$$\mathsf{E}\|x^\varepsilon - x\| \leq (\mathsf{E}\|\varepsilon\|^2)^{1/2} = \sigma . \tag{3.4}$$

Estimates (3.3) and (3.4) and the made assumptions lead to

$$\begin{aligned} \mathsf{E}\|x^\varepsilon - x_\alpha\|^2 + \alpha\|x_* - x_\alpha\|^2 &\leq \sigma^2 + \alpha\mathsf{E}\{\|x_*\|^2 - \|x_\alpha\|^2 + \|x_* - x_\alpha\|^2\}\\ &\leq \sigma^2 + 2\alpha\mathsf{E}\langle w, x - x_\alpha + x^\varepsilon - x^\varepsilon\rangle\\ &\leq \sigma^2 + 2\alpha\|w\|(\sigma + \mathsf{E}\|x^\varepsilon - x_\alpha\|)\\ &\leq \sigma^2 + 2\alpha\|w\|(\sigma + (\mathsf{E}\|x^\varepsilon - x_\alpha\|^2)^{1/2}).\end{aligned}$$

The latter inequality can be rewritten as

$$\alpha\|x_* - x_\alpha\|^2 \leq \sigma^2 - \mathsf{E}\|x^\varepsilon - x_\alpha\|^2 + 2\alpha\|w\|(\sigma + (\mathsf{E}\|x^\varepsilon - x_\alpha\|^2)^{1/2}) . \tag{3.5}$$

By

$$\begin{aligned}\Big(\mathsf{E}\|x^\varepsilon - x_\alpha\|^2)^{1/2} - \alpha\|w\|\Big)^2 &= -\big(-\mathsf{E}\|x^\varepsilon - x_\alpha\|^2\\ &\qquad +2\alpha\|w\|(\mathsf{E}\|x^\varepsilon - x_\alpha\|^2)^{1/2}\Big) + \alpha^2\|w\|^2\end{aligned}$$

we finally obtain

$$\Big((\mathsf{E}\|x^\varepsilon - x_\alpha\|^2)^{1/2} - \alpha\|w\|\Big)^2 + \alpha\mathsf{E}\|x_* - x_\alpha\|^2 \leq (\sigma + \alpha\|w\|)^2 .$$

Stressing Ljapunov's inequality again, we complete the proof

$$\|x_* - x_\alpha\| \leq \frac{\sigma + \alpha\|w\|}{\sqrt{\alpha}} \quad \text{and} \quad \mathsf{E}\|x^\varepsilon - x_\alpha\| \leq \sigma + 2\alpha\|w\| \ .$$

■

We wish to remark that the above given result is based on a convergence rate proof for Tikhonov regularization, first given in a deterministic setting by Engl, Kunisch and Neubauer [2, 3].

The final step is to formulate Morozov's discrepancy and to apply this principle to Proposition 3.1 to obtain the convergence result.

Definition 3.1 *We define the data-driven Morozov discrepancy principle by*

$$\mathcal{S}^2(r_\alpha^\varepsilon) \leq \mathcal{E}\{||r_\alpha^\varepsilon||_2^2\}_m \leq C^2 \cdot \mathcal{S}^2(r_\alpha^\varepsilon) \ , \tag{3.6}$$

for some constant $C \geq 1$.

Theorem 3.1 *Let the assumption of Propostion 3.1 hold. If the regularization parameter α is chosen such that* (3.6) *hold for some $C \geq 1$, we obtain the error estimate*

$$\|x_\alpha - x_*\| \leq \sqrt{2\|w\|(C+1)}\sigma^{1/2} \ , \tag{3.7}$$

i.e., the data driven Morozov discrepancy principle is order optimal for $\nu = 1$.

Proof To proof this theorem we have to switch from realizations $\tilde{r}_\alpha^\varepsilon$ to random sample variables r_α^ε

$$\mathcal{S}^2(r_\alpha^\varepsilon) \leq \mathcal{E}\{||r_\alpha^\varepsilon||_2^2\}_m \ \leq C \cdot \mathcal{S}^2(r_\alpha^\varepsilon) \ .$$

Taking the expectation we obtain by Lemma 2.3 and (2.6)

$$\sigma^2 \leq \mathsf{E}||T_\alpha x^\varepsilon - x^\varepsilon||_2^2 \leq C \cdot \sigma^2 \ .$$

Hence, we have the inequalities

$$\sigma^2 - \mathsf{E}||T_\alpha x^\varepsilon - x^\varepsilon||_2^2 \leq 0 \quad \text{and} \quad \left(\mathsf{E}||T_\alpha x^\varepsilon - x^\varepsilon||_2^2\right)^{1/2} + \sigma \leq (C+1)\sigma \ .$$

Consequently, we can deduce by inequality (3.5) that

$$\|x_\alpha - x_*\|^2 \le 2\|w\|(C+1)\sigma \ .$$

The last estimate shows order optimal convergence rate for $\nu = 1$, see [8]. ∎

We finish this subsection with considering some numerical aspects. When it comes to real implementations we have to deal with a finite number of observations. To this end, we might consider

$$J^M_{\alpha,\varepsilon}(x) := \mathcal{E}\{\|x^\varepsilon_m - ix\|_2^2\}_m + \alpha\|x\|_s^2 \ . \tag{3.8}$$

In order to analyze the rate of convergence we have to derive the bounds (3.1) and (3.2). Setting

$$\mathcal{S}_1^2 := \mathcal{E}\{\|x^\varepsilon_{m,*} - x_*\|_2^2\}_m \ , \quad \mathcal{S}_2^2 := \mathcal{E}\{\|x - x^\varepsilon_m\|_2^2\}_m \quad \text{and} \quad \mathcal{S}_3^2 := \mathcal{S}^2(r^\varepsilon_\alpha)$$

one can show that

$$\|x_* - x_\alpha\|_2^2 \le \frac{(\mathcal{S}_1^2 + \alpha\|w\|)^2}{\alpha} + 2\alpha\|w\|(\mathcal{S}_2 - \mathcal{S}_1)$$

and

$$\mathcal{E}\{\|x^\varepsilon_m - x^\varepsilon_\alpha\|_2^2\}_m \le \sqrt{(\mathcal{S}_1^2 + \alpha\|w\|)^2 + 2\alpha\|w\|(\mathcal{S}_2 - \mathcal{S}_1)} + \alpha\|w\| \ .$$

In accordance with Theorem 3.1 we finally obtain

$$\|x_* - x_\alpha\|_2^2 \le \frac{\mathcal{S}_1^2 - \mathcal{S}_3^2}{\alpha} + 2\|w\|(\mathcal{S}_2 + C\mathcal{S}_3) \ ,$$

i.e., if M become large we asymptotically attain order optimality. For some finite (and small) M we have to deal with a somewhat systematic biased Morozov principle and hence, with biased error estimates.

4 Further Parameter Rules

Beside Morozov's discrepancy principle we suggest the use of two rules based on the auto-covariance function. We define the auto-covariance function of a random variable $\xi(t)$ in the common way

$$C[\xi](\mu) := \mathsf{E}\langle \xi, S_\mu \xi\rangle_{L_2(\Omega)} \ , \tag{4.1}$$

where $(S_\mu \xi)(t) = \xi(t+\mu)$. The inner product in $L_2(\Omega)$ results in

$$\langle \xi, S_\mu \xi \rangle_{L_2(\Omega)} = \sum_k \sum_l \hat{\xi}_k \bar{\hat{\xi}}_l e^{-i\mu m} \int_0^{2\pi} e^{it(k-l)} dt = \sum_k \hat{\xi}_k \bar{\hat{\xi}}_k e^{-i\mu k} . \quad (4.2)$$

Consequently, by normalizing $C[\xi]$ by the variance we obtain the so-called auto-correlation function

$$A[\xi](\mu) := C[\xi](\mu) \cdot (C[\xi](0))^{-1} = \sum_k \mathsf{E} \hat{\xi}_k \bar{\hat{\xi}}_k e^{-i\mu k} \cdot \left(\sum_{k=0} \mathsf{E} \hat{\xi}_k \bar{\hat{\xi}}_k \right)^{-1} . \quad (4.3)$$

Note that by definition (4.1) one can write

$$C[T_\alpha x^\varepsilon - x^\varepsilon](0) = \mathsf{E} \| T_\alpha x^\varepsilon - x^\varepsilon \|_2^2 . \quad (4.4)$$

The basic ingredient of the suggested parameter rules rest on the heuristic idea of reconstructing nearly independent residue, i.e., we choose $\alpha > 0$ such that the values of the auto-correlation function (4.3) in a small neighborhood of zero become small. We must admit that analytical results of any sort of convergence of the resulting scheme have not come in yet. However, in numerical experiments and in comparison with other parameter rules we have found that the scheme is quite stable and yields reasonable regularization parameters.

The basic technique is to approximate $A[\xi]$ by its Taylor expansion at some grid point t_0

$$A[\xi](t) = A[\xi](t_0) + (t - t_0) A'[\xi](t_0) + \frac{(t-t_0)^2}{2} A''[\xi](t_0) + \rho(t, t_0) . \quad (4.5)$$

The first rule rests on a second order approximation of $A[\xi]$ at $t_0 = 0$. Thus, we obtain a parabola approximating $A[\xi]$ close to zero, i.e.,

$$A[T_\alpha x^\varepsilon - x^\varepsilon](t) \doteq 1 - \frac{t^2}{2} \frac{\sum_k a_{k,s} \mathsf{E} \hat{x}_k^\varepsilon \bar{\hat{x}}_k^\varepsilon k^2}{\mathsf{E} \| T_\alpha x^\varepsilon - x^\varepsilon \|^2} .$$

Minimizing the range of dependence can be obtained by minimizing the distance between zero and the roots of the parabola. The roots are given by

$$t^2 = 2 \frac{\mathsf{E} \| T_\alpha x^\varepsilon - x^\varepsilon \|^2}{\sum_k a_{k,s} \mathsf{E} \hat{x}_k^\varepsilon \bar{\hat{x}}_k^\varepsilon k^2} .$$

This leads to the following cost functional

$$G^2_{x^\varepsilon}(\alpha) := \frac{\mathsf{E}\|T_\alpha x^\varepsilon - x^\varepsilon\|^2}{\sum_k a_{k,s}\mathsf{E}\hat{x}^\varepsilon_k \bar{\hat{x}}^\varepsilon_k k^2} \ . \tag{4.6}$$

A second construction is very similar. We consider the linear Taylor approximation of $A\,[\xi]$ at t_0. But instead of $t_0 = 0$ we choose t_0 very close to zero. By the same arguments as before we can define the following cost functional

$$G^1_{x^\varepsilon}(\alpha) := -\frac{A\,[T_\alpha x^\varepsilon - x^\varepsilon]\,(t_0)}{A'\,[T_\alpha x^\varepsilon - x^\varepsilon]\,(t_0)} \ . \tag{4.7}$$

The numerical treatment is given by means of the FFT. An adequate choice for t_0 in (4.7) is given by $t_0 = 2\pi/N$, where N is the number of grid points. By extensive numerical experiments we have observed two effects. Both the cost functional (4.6) and (4.7) yield reasonable α's with $\alpha(\varepsilon, x^\varepsilon) \to 0$ for $\varepsilon \to 0$. Moreover, for all simulations we have found that they are upper bounds for the lower optimal α chosen in the data driven Morozov's discrepancy principle (3.6).

5 Numerical Experiments and Applications

This section is concerned with testing the numerical capabilities of the proposed methods. At first, we test the methods by reconstructing simulated data. Secondly, we apply the procedures to real data sets.

5.1 Simulation Experiments

The setting is the following: $N = 256$, $k = 0, \dots, N-1$, $t_k = k/N$ and $x(t) = 1 + \sin(4t) + \cos(6t)$, moreover, for $\varepsilon_t \sim \mathsf{N}(0,1)$ and $c = 0.5$ the data are given by

$$x^\varepsilon(t_k) = x(t_k) + c \cdot \varepsilon_{t_k} \ ,$$

see Figure 1. We have sampled the data 20 times, i.e., 20 versions of x^ε are available. On the basis of this underlying data sets we have derived solutions of the inverse embedding problem by the data-driven Morozov principle and by minimizing the autocorrelation based cost functional (4.6). The estimation of σ^2 is based on the estimation of a mean value at each grid point with respect to the residua, see Figure 1. This curve is the basis for deriving the variances. The true averaged variance is

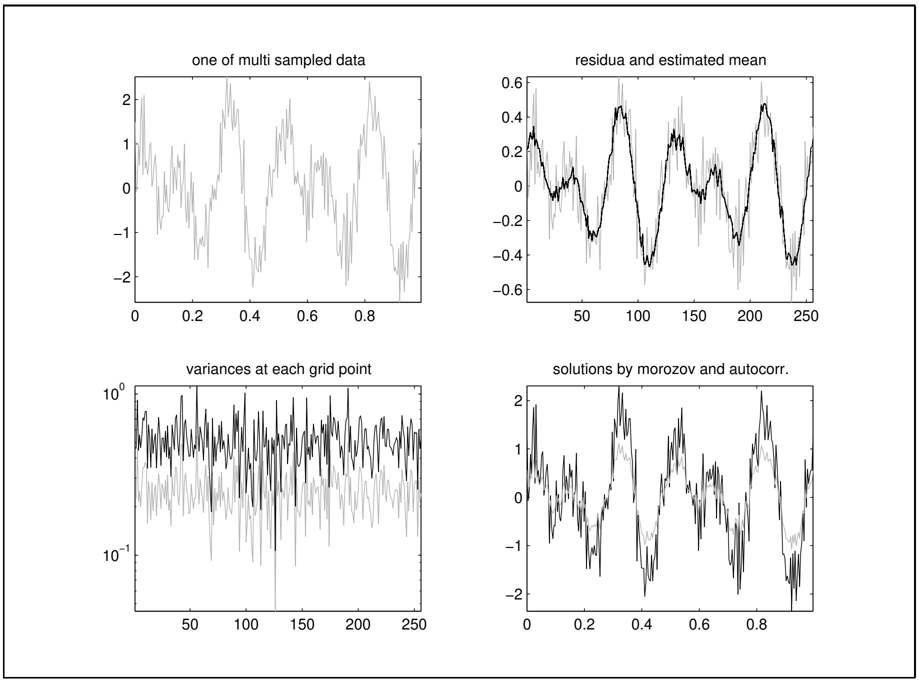

Figure 1: u.l.: one sample of the function $x^{\varepsilon}(t_k)$; u.r.: residua $T_{\alpha}x^{\varepsilon} - x^{\varepsilon}$ (gray) and the estimated mean values (black); l.l.: true variances (gray) and estimated variances (black); l.r.: solutions of the inverse embedding problem - by morozov (red) and by the autocorrelation cost functional (blue).

equal to 0.2494 whereas the average of the estimated variances is equal to 0.4639. The difference can be explained with the necessity of an 'educated' guess for α. In our specific example we obtain the following regularization parameters $\alpha_{Morozov} = 0.007139$ and $\alpha_{Autocorr} = 0.0293$. In general, we have observed that under unknown noise level σ^2 the data driven Morozov principle and the autocorrelation based cost functionals yield suitable regularization parameters α.

For a somewhat different example we have visualized the results obtained by the classical Morozov (upper row) and the data driven Morozov principle (lower row), see Figure 2. We have simulated data sets with different values for σ^2 (increasing sequence). For these different data sets we have computed the optimal parameters α (left column on the x-

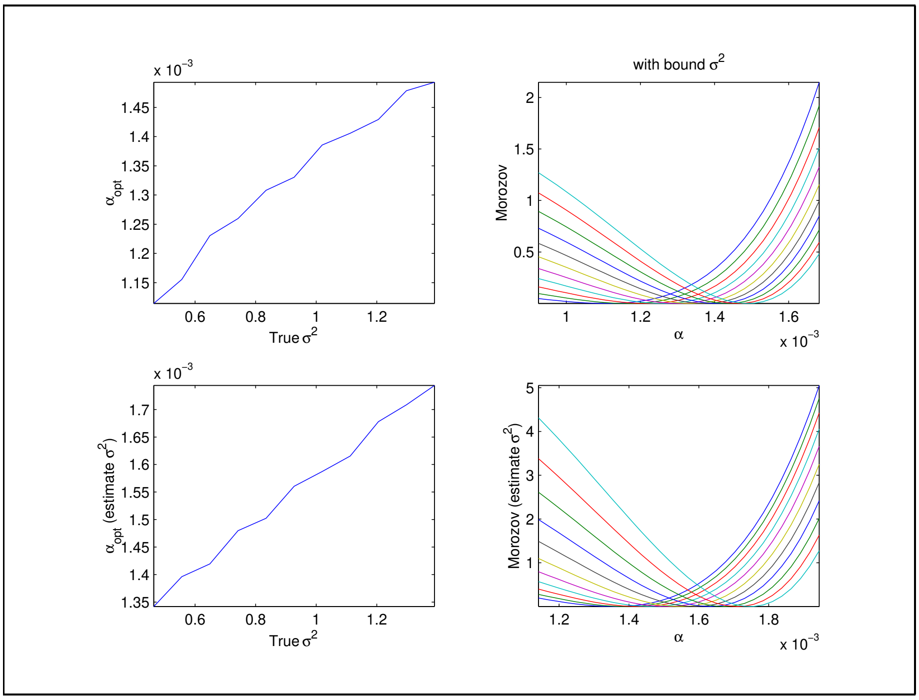

Figure 2: u.l. optimal α vs. different known σ^2 computed via the Morozov discrepancy principle ($= \sigma^2$); u.r. values of the related Morozov functional $\sigma^2 - \mathsf{E}\|T_\alpha x^\varepsilon - x^\varepsilon\|^2$; l.l. and l.r. are similar but now using the data-driven rules.

axis). This example clearly shows that the data driven Morozov scheme overestimates the variances (compared to the classical Morozov scheme). The right column shows the behavior of the Morozov functionals for the sequence of given σ^2's.

5.2 Application to Radar Wind Profiler Data

The radar wind profiler is a special application of Doppler technology and is currently used to routinely probe the vertical profile of the mean horizontal wind in the earth's atmosphere. In principle, a radar device transmits electromagnetic waves in a certain beam direction and sample the backscattered signal discretely in time. The transmitted signal h is a pulse train, i.e., a sum of ΔT time shifted pulses $h(t) = e^{i\omega_c t}\eta(t)$ with

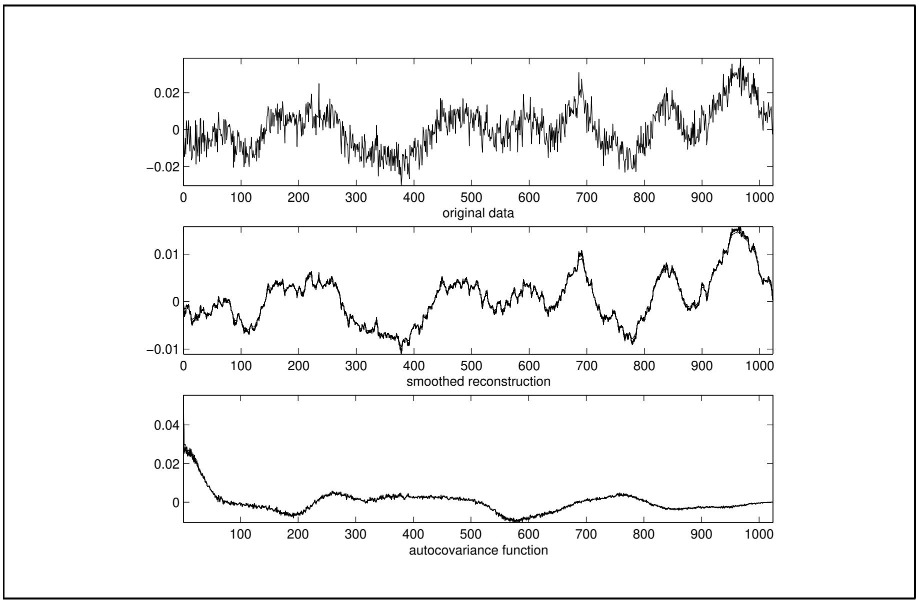

Figure 3: This figure shows a original in-phase I of radar measurements, the reconstructed long-wave component, and the related autocorrelation function.

$\eta(t) = \sum_{k=0}^{2^N-1} \mu(t - k\Delta T)$, where μ is a single pulse, e.g., a characteristic function which is contained in $[0, \Delta T]$, and ω_c denotes the carrier frequency. By extending a simple narrowband model, see [1], the received echo f can be described by

$$f(t) = \sum_{k=0}^{2^N-1} \{\mu_p * D_p(t,\cdot) + \mu_r * D_r(t,\cdot)\}(t - k\Delta T) \ ,$$

where the subscripted characters r, p denote the regressive and progressive part of the pulses and the reflectivity densities. The radar device measures real-valued voltages $r(t)$ only. But by meteorological reasons it can be assumed that f can be obtained by means of the Hilbert transform H

$$f(t) = \{r(t) + i(Hr)(t)\}e^{-i\omega_c t} = I(t) + iQ(t) \ . \tag{5.1}$$

One important goal in radar wind profiler signal processing is to predict the Doppler frequency (velocity of wind) on the basis of (5.1). The usual problem is that the data f may be heavily contaminated by echoes

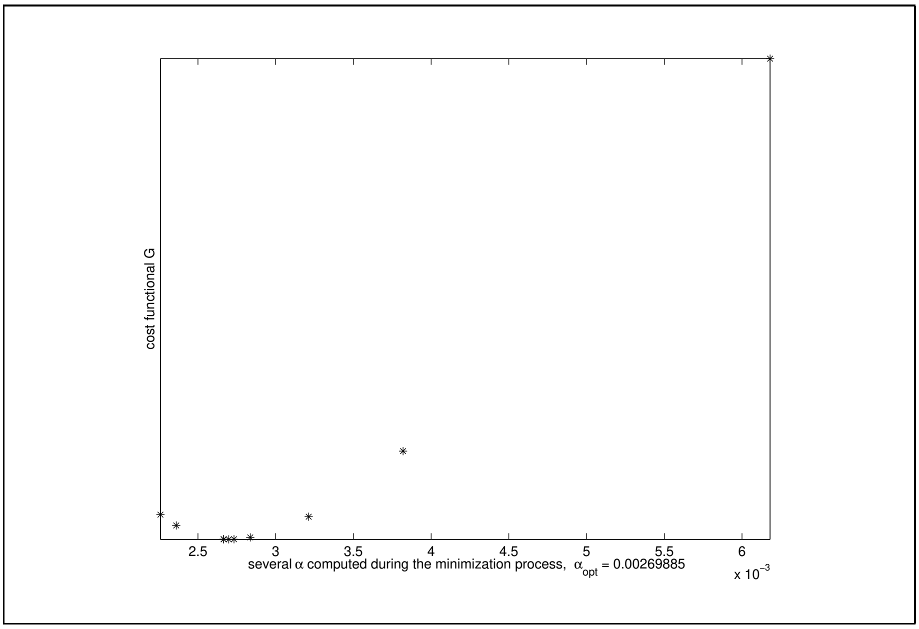

Figure 4: This figure shows the sequence of computed α's - the optimal $\alpha = 0.00269885$.

returned from ground surroundings, from targets like airplanes or birds or from external radio transmissions. Here we show how the proposed inversion scheme for Sobolev embedding operators can be applied to data sets $\{I, Q\}$ in order to overcome the contamination problems. To this end, we note that echoes from ground surroundings are supposed to have a Doppler shift nearby zero (an averaged frequency close to zero), i.e., we aim at removing a long-wave component $\{I_g, Q_g\}$ not originating from any scattering process. The idea is to reconstruct $\{I_g, Q_g\}$ via the proposed inversion scheme for Sobolev embeddings and to identify the residuum $\{I, Q\} - \{I_g, Q_g\}$ as the atmospheric echo.

In this application we minimize the autocorrelation cost functional G^1, see formula (4.7). Moreover, by meteorological heuristics we may assume that searched ground clutter component has regularity $s \approx 0.5$ (what is indeed heuristically presumed). In Figure 3 one specific data set (real part I), the reconstructed long-wave component, and the related autocorrelation function is visualized whereas Figure 4 shows the sequence of computed parameters α during the minimization process. In comparison with other common used methods in this field (peak detection, Gaussian fits etc.) for detecting and removing echoes from ground surroundings we may conclude that our method is quite competitive.

5.3 Application to Medical Imaging

In Computerized Tomography, the density f of a two dimensional object is reconstructed from its line integrals; the measured data is given by the Radon transform:

$$y(r,\omega) = (Rf)(r,\omega) = \int f(r\omega + t\omega^{\perp})\, dt \ , r \in \mathbb{R}, \omega \in S^1 \ .$$

We define the Sobolev space $H^s(Z)$ with Z the unit cylinder in $\mathbb{R}^2$ by

$$\|g\|_{H^s(Z)} = \int_{S^1} \int_{\mathbb{R}} (1 + |\rho|^2)^s |\hat{g}(\rho,\omega)|\, d\rho\, d\omega \ , \tag{5.2}$$

where the Fourier-transform of $g(r,\omega)$ is taken with respect to r. Then it is a well known result that for each s and $f \in C_0^\infty(\Omega)$ the estimate

$$c(s)\|f\|_{H^s(\Omega)} \leq \|Rf\|_{H^{s+1/2}(Z)} \leq C(s)\|f\|_{H^s(\Omega)}$$

holds, i.e. the Radon transform is continuously invertible for $f \in C_0^\infty(\Omega)$ (cf. [6]). Moreover, Natterer [5] showed that, for $s > 1/2$, the Radon transform is a bounded operator from $H_0^s(Z) \to H^{s+1/2}(Z)$. For medical applications, the density function f belongs to H_0^s with s close to $1/2$. As $C_0^\infty\Omega$ is dense in H_0^s, it is a reasonable assumption that the noise-free data $y = Rf$ belongs to the Sobolev space $H^{\tilde{s}}(Z)$ with $s \in [1/2, 1]$. This is of course not the case for the measured, noisy data y^δ, which might belong to $L_2(Z)$ only. For $g \in H^s(Z)$ we observe from definition (5.2) that the function $g_\omega = g(\cdot,\omega)$ belongs to the Sobolev space $H_0^s(\mathbb{R})$ for almost all unit vectors ω. Thus we propose the following algorithm for the inversion of the Radon transform: In a first step, the measured data $y^\delta(r,\omega)$ is smoothed for every unit vector ω *separately* by the regularization of the embedding operator $i : H_0^s(\mathbb{R}) \to L_2(\mathbb{R})$. As regularization parameter choice we use the data driven discrepancy principle (3.6) or the rules (4.6), (4.7). We wish to remark that we allow a different regularization parameter $\alpha = \alpha(\omega)$ for every unit vector ω, which takes into account that the noise within the data $g(\cdot,\omega)$ might be different for every

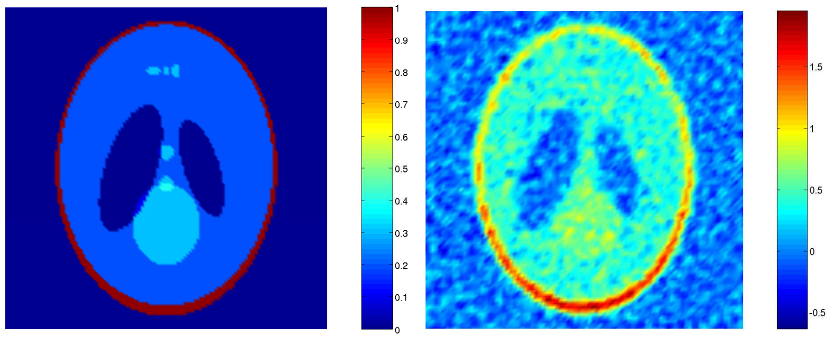

Figure 5: Shepp–Logan head phantom (left), reconstruction of the phantom from noisy data (5%), using filtered backprojection with Hamming filter (right).

ω. Finally, we apply the inverse Radon transform to the smoothed data.

In comparison to other classical regularization methods, e.g., Tikhonov regularization for the full operator equation $y = Rf$, the main advantage of our method is that the computation of a minimizer of the Tikhonov–functional for the embedding operator is fast. Moreover, no information on the noise or the noise level is required.

We have performed test computations both with simulated and real data. First, we used the so-called Shepp–Logan head phantom. The data was generated by the *radon* function in $Matlab$; we have used 120 angles over the full circle and 128 samples. The data was blurred with approximately 5% noise. For comparison, the data was inverted by the *iradon* function of $Matlab$; this function performs a filtered backprojection with a Hamming filter, cf. Figure 5. The denoised data was calculated for the smoothness index $s = 0.7$, i.e., it was assumed that the data belonged to $H_0^{0.7}$. We have used the data-driven Morozov's discrepancy principle (3.6), which means that the regularization parameter was chosen automatically and without knowledge of the error level. Figure (6) shows the data and the result of the denoising. We have then applied the *iradon* function of $Matlab$ to the denoised data, cf. Figure 7. Obviously, the reconstruction of the head has a much better quality as the reconstruction in Figure 5.

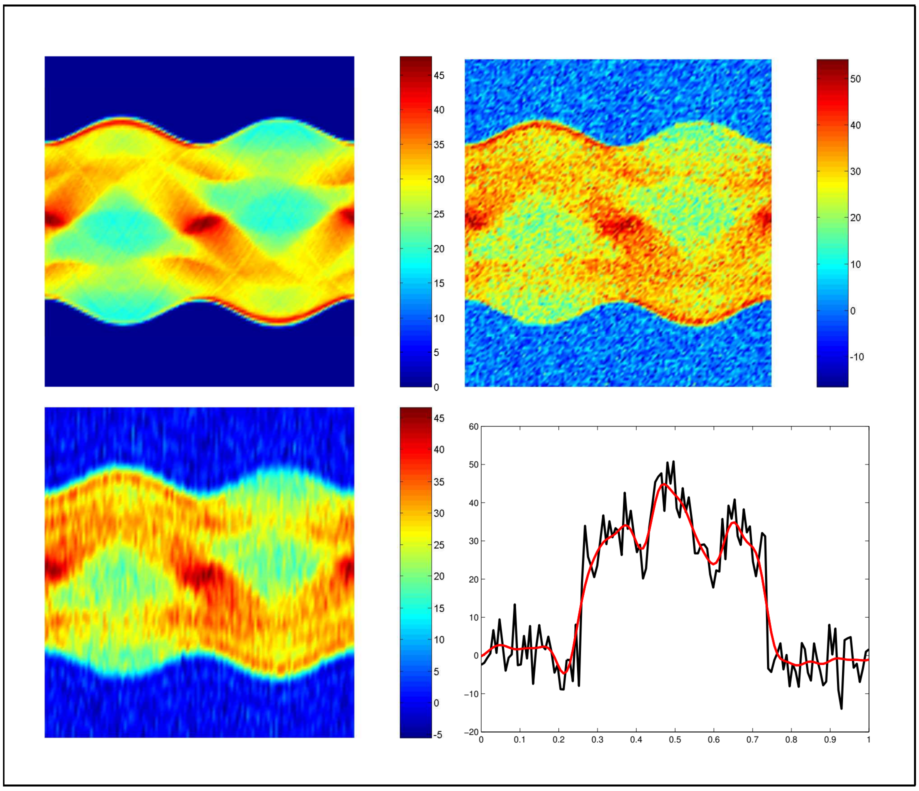

Figure 6: Sinogram data y of the head phantom (u.l.), noisy data y^δ (u.r.), denoised data y^δ_α (l.l) and a plot of $y^\delta(r,\omega)$ and $y^\delta_\alpha(r,\omega)$ for $\omega = 177°$ (l.r).

Next, we will present some results with real data. The data was obtained from a CT scan of a Jaszczak torso phantom on a Picker P2000 SPECT camera. We wish to remark that the Jaszczak torso phantom is mainly used in Single Photon Emission Tomography (SPECT), cf. e.g. [9, 7] To improve the image quality of SPECT, most cameras perform CT scan of the object simultaneously. Naturally, SPECT cameras are not optimized for CT scans, and thus the reconstructions from the CT data are usually poor. As above, we have used the *iradon* function of *Matlab* for a first reconstruction. In a second step, the data was again denoised by our method, using again $s = 0.7$ and the data driven discrepancy principle. Again, the reconstruction quality is much better by using the

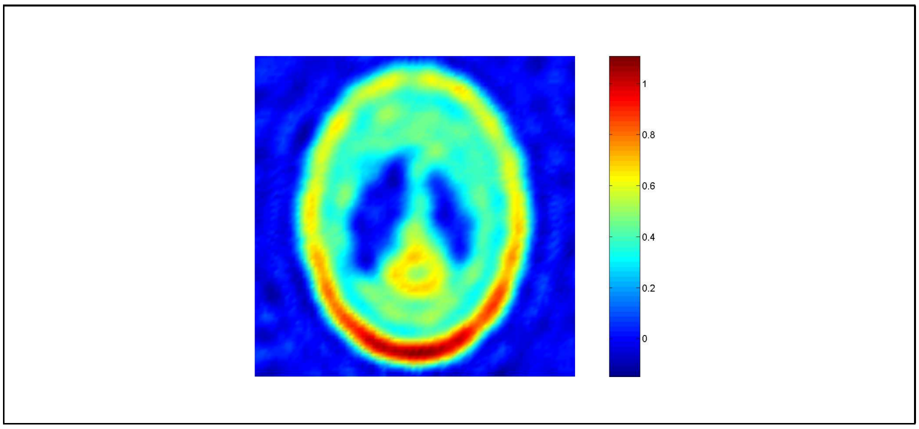

Figure 7: Reconstruction of the head phantom with denoised data.

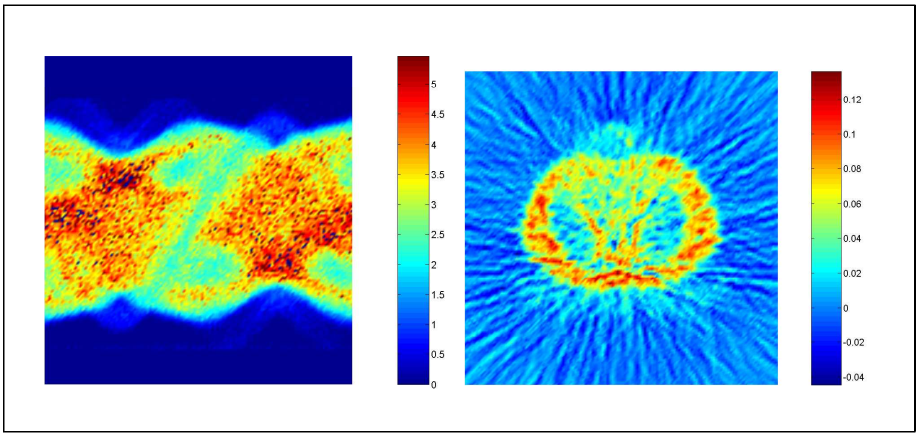

Figure 8: Sinogram data y^δ of the Jasczcak torso phantom (l.), and the reconstruction from these data by using the filtered backprojection.

preprocessed data.

Acknowledgement

The authors gratefully acknowledge support by DFG grants Te 354/1-2 and Ma 1657/7-2.

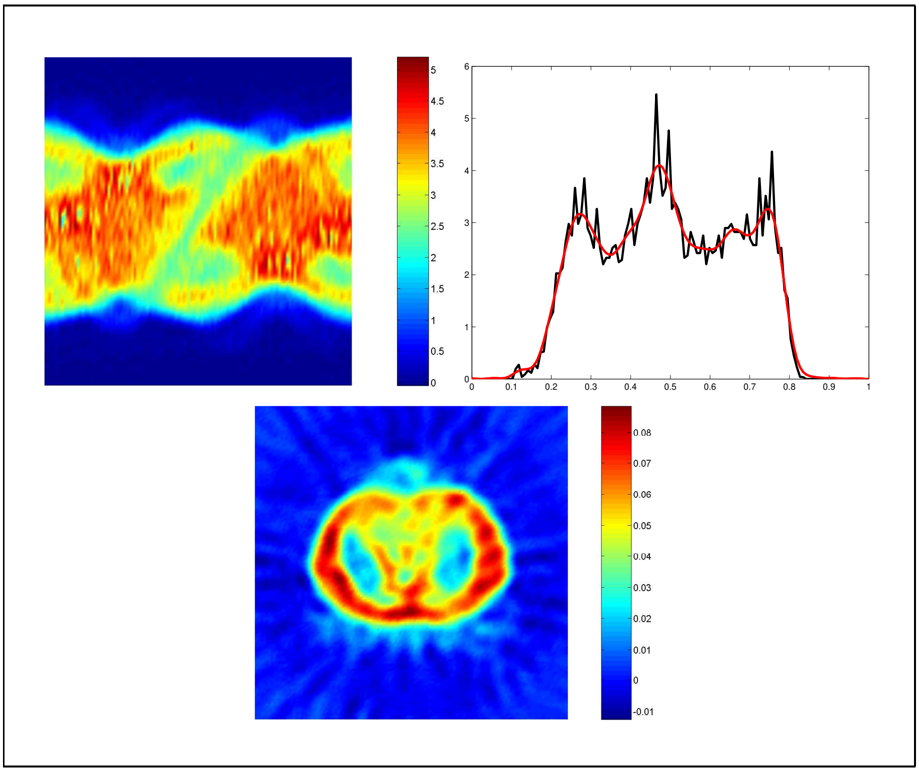

Figure 9: Denoised sinogram data y_α^δ of the Jaszczak torso phantom (u.l.), a plot of $y^\delta(r,\omega)$ and $y_\alpha^\delta(r,\omega)$ for $\omega = 192°$ (u.r), reconstruction from denoised data

References

[1] S. Dahlke, V. Lehmann and G. Teschke, Application of wavelet methods to the analysis of meteorological radar data - an overview, *invited paper in Arabian Journal of Science and Engineering*, **28**(1C), 3-44, 2003.

[2] H.W. Engl, K. Kunisch and A. Neubauer, Convergence rates for Tikhonov regularization of nonlinear ill-posed problems, *Inverse Problems*, **5**, 523-540, 1989.

[3] H. W. Engl and M. Hanke and A. Neubauer, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996.

[4] E.W. Kamen and J.K.Su, *Introduction to optimal estimation*, Springer, London, 1999.

[5] F. Natterer, Sobolev space analysis of picture reconstruction, *SIAM J. Appl. Math.*, **39**, 402-411, 1980.

[6] F. Natterer, *The Mathematics of Computerized Tomography*, Teubner & Wiley, Stuttgart, 1986.

[7] R. Ramlau, *Regularization of Nonlinear Ill–Posed Operator Equations: Methods and Applications*, Habilitation thesis, University of Bremen, Germany, 2003.

[8] R. Ramlau and G. Teschke, Regularization of Sobolev Embedding Operators and Applications Part I: Fourier and Wavelet based Methods, *Sampling Theory in Signal and Image Processing*, **3**(2), 175-196), 2004.

[9] J. A. Terry, B. M. W. Tsui, J. R. Perry, J. L. Hendricks and G. T. Gullberg, The design of a mathematical phantom of the upper human torso for use in 3-D SPECT imaging research, *Proc. 1990 Fall Meeting Biomed. Eng. Soc. (Blacksburg, VA)*, 1467-74, New York University Press, 1990.