On the minimization of a Tikhonov functional with a non-convex sparsity constraint

Ronny Ramlau

Johannes Kepler University Linz (JKU), Institute of Industrial Mathematics, Altenbergerstrasse 69, A-4040 Linz, Austria

E-mail: `ronny.ramlau@jku.at`

Clemens A Zarzer

Johann Radon Institute for Computational and Applied Mathematics (RICAM), Austrian Academy of Sciences, Altenbergerstrasse 69, A-4040 Linz, Austria

E-mail: `clemens.zarzer@ricam.oeaw.ac.at`

Abstract. In this paper we present a numerical algorithm for the optimization of a Tikhonov functional with ℓ_p sparsity constraints and $p < 1$. Recently it was proven that the minimization of this functional provides a regularization method. We show that the idea used to obtain these theoretical results can also be utilized for a numerical approach. Particularly we exploit the technique of transforming the Tikhonov functional to a more viable one. In this regard we consider a surrogate functional approach and show that this technique can be applied straightforward. It is proven that at least a critical point of the transformed functional is obtained, which directly translates to the original functional. For a special case it is shown that a gradient based algorithm can be used to reconstruct the global minimizer of the transformed and the original functional respectively. At the end we present a simple numerical example and provide numerical evidence for the theoretical results and the desired sparsity promoting features of this method.

1. Introduction

In this paper we consider a Tikhonov type regularization method for solving a (generally nonlinear) ill-posed operator equation

$$\mathcal{F}(x) = y \tag{1}$$

with noisy measurements y^δ with $\|y^\delta - y\| \leq \delta$. Throughout the paper we will assume that $\mathcal{F}$ maps between sequence spaces, i.e.,

$$\mathcal{F} : \ell_p \to \ell_2 \ . \tag{2}$$

Please note that operator equations between suitable separable function spaces such as L^p, Sobolev and Besov spaces, i.e.

$$F : D(F) \subset X \to Y \ , \tag{3}$$

can be transformed to a sequence setting by using suitable basis or frames for $D(F)$ and $R(F)$: Indeed, if we assume that we are given some preassigned frames $\{\Phi^i_\lambda\}_{\lambda\in\Lambda_i, i=1,2}$, ($\Lambda_i$ countable index sets) for $D(F) \subset X$, $R(F) \subset Y$, with the associated frame operators T_1, T_2 then the operator $\mathcal{F} := T_2 F T_1^*$ maps between sequence spaces.

We are particularly interested in *sparse* reconstructions, i.e. the reconstruction of sequences with only few nonzero elements of the sequence. To this end, we want to minimize the Tikhonov functional

$$\begin{aligned} J_\alpha \,:\, \ell_p &\to \mathbb{R} \\ x &\mapsto \left\|\mathcal{F}(x) - y^\delta\right\|_2^2 + \alpha\ \|x\|_p^p \ , \end{aligned} \tag{4}$$

where $\alpha > 0$, $p \in (0,1]$ and

$$\|x\|_p^p = \sum_k |x_k|^p \ , \tag{5}$$

is the (quasi-) norm of ℓ_p. The main aim of our paper will be the development of an iterative algorithm for the minimization of (4), which is due to the non-convexity of the quasi-norm and the nonlinearity of $\mathcal{F}$ a nontrivial task.

The reconstruction of the sparsest solution of an underdetermined system has already a long history, in particular in signal processing, and more recently, in compressive sensing. Usually the problem is formulated as

$$\tilde{x} := \arg\min_{y=\Phi x} \|x\|_1 \tag{6}$$

where $y \in \mathbb{R}^m$ is given and $\Phi \in \mathbb{R}^{m,n}$ is an rank deficient matrix (i.e. $m < n$), see [1, 2]. Please note that here the minimization of the ℓ_1 - norm is used for the reconstruction of the sparsest solution of the equation $\Phi x = y$. Indeed, under certain assumptions on the matrix Φ, it can be shown that, if there is a sparse solution, (6) really recovers it [3, 4, 5, 6]. Moreover, Gribonval and Nielsen [7] showed that for special cases the minimization of (6) also recovers ℓ_p - minimizers with $0 < p < 1$. In this sense it might seem that nothing is gained by considering ℓ_p minimization with $0 < p < 1$ instead of ℓ_1 minimization, or equivalently, using an ℓ_p penalty with $0 < p < 1$ in (4). However, we have to keep in mind that we are considering a different setting as the above cited papers. First of all, we are working in an infinite dimensional setting, whereas the above mentioned Φ is a finite dimensional matrix. Additionally, properties that guarantee the above cited results as the so-called Restricted Isometry Property, which was introduced by Candes and Tao [8, 4] or the Null Space Property [9, 10] are not likely to hold even for linear infinite dimensional ill-posed problems where, e.g., the eigenvalues of the operator converge to zero, not to speak of nonlinear operators. Recently, there has also been numerical evidence from a nonlinear parameter identification problem for chemical reaction systems that an ℓ_1 penalty in (4) failed to reconstruct a desired sparse parameter whereas stronger ℓ_p penalties with $0 < p < 1$ achieved sparse reconstructions [11]. In the mentioned paper, the intention of the authors was the reconstruction of reduced chemical networks (represented by a sparse parameter) from chemical measurements. Therefore we conclude that the use of the stronger ℓ_p penalties might be necessary in infinite dimensional ill-posed problems if one wants a sparse reconstruction. In particular, algorithms for the minimization of (4) are needed.

There has been an increased interest in the investigation of the Tikhonov functional with sparsity constraints. First results were presented by Daubechies, Defriese and De Mol [12]. The authors were in particular interested in solving linear operator equations. As constraint in (4) they used a Besov semi - norm, which can be equivalently expressed by a weighted ℓ_p norm of the wavelet coefficients of the functions with $p \geq 1$. In particular the paper focuses on the analysis of a surrogate functional approach for the minimization of (4) with $p \geq 1$. It was shown that the proposed iterative method converges towards a minimizer of the Tikhonov functional under consideration. Additionally, the authors proposed a rule for the choice of the regularization parameter that guarantees the convergence of the minimizer x_α^δ of the Tikhonov functional to the solution as the data error δ converges to zero. Subsequently, many results on the regularization properties of the Tikhonov functional with sparsity constraints with $p \geq 1$ as well as on its minimization were published. In [13, 14] the surrogate functional approach for the minimization of the Tikhonov functional was generalized to nonlinear operator equations and in [15, 16] to multi-channel data, whereas in [17, 18] a conditional gradient method and in [19] a semismooth Newton method was proposed for the minimization. Further results on the topic of minimization and the respective algorithms can be found in [20, 21, 22]. The regularization properties with respect to different topologies and parameter choice rules were considered in [14, 15, 23, 24, 25, 26]. Please note again that the above cited results only consider the case $p \geq 1$. For the case $p < 1$, a first regularization result for some types of linear operators was presented in [26]. In [27] and [28] the authors recently presented general results on the regularization properties of the Tikhonov functional with a nonlinear operator and $0 < p < 1$. Concerning the minimization of (4) with $0 < p < 1$, to our knowledge no results are available in the infinite dimensional setting. In the finite dimensional setting, Daubechies et. al [10] presented an iteratively re-weighted least squares method for the solution of (6) that achieved local superlinear convergence. However, these results do not carry over to the minimization of (4), as the assumptions made in [10] (e.g., finite dimension, null space property) will not hold for general inverse problems. Other closely related results in the finite dimensional case can be found in [29, 30].

In this paper, we will present two algorithms for the minimization of (4) which are based on the surrogate functional algorithm [12, 13, 14, 23] and the TIGRA algorithm [31, 32]. Based on a technique presented in [28] and based on methods initially developed in [33], the functional (4) is nonlinearly transformed by an operator $\mathcal{N}_{p,q}$ to a new Tikhonov functional, now with an ℓ_q norm as penalty and $1 < q \leq 2$. Due to the nonlinear transformation, the new Tikhonov functional involves a nonlinear operator, even when the original problem was linear. Provided the operator $\mathcal{F}$ fulfills some properties, it is shown that the surrogate functional approach will at least reconstruct a critical point of the transformed functional. Moreover, the minimizers of the original and the transformed functional are connected by the transformation $\mathcal{N}_{p,q}$, and thus we can obtain a minimizer for the original functional. For the special case $q = 2$ we show that the TIGRA algorithm reconstructs a global minimizer if the solution fulfills a smoothness condition. For the case $\mathcal{F} = \mathcal{I}$, where $\mathcal{I}$ denotes the identity, we show that the smoothness condition is always fulfilled for sparse solutions, whereas for $\mathcal{F} = \mathcal{A}$ with linear $\mathcal{A}$ an restricted invertibility condition is needed additionally. The paper is organized as follows: In Section 2 we recall some results from [28] and introduce the transformation operator $\mathcal{N}_{p,q}$. Section 3 is concerned with some analytical properties of $\mathcal{N}_{p,q}$, whereas Section 4 investigates the operator $\mathcal{F} \circ \mathcal{N}_{p,q}$. In Section 5 we use the surrogate functional approach for the minimization of the transformed functional, and in Section 6 we introduce the TIGRA method for the reconstruction of a global minimizer. Finally we present in Section 7 numerical results for the reconstruction of a function from its convolution data that confirm our analytical results.

Whenever it is appropriate, we omit the subscripts for norms, sequences, dual pairings and so on. If not denoted otherwise we consider the particular notions in terms of Hilbert space ℓ_2 and the respective topology. Furthermore we would like to mention that the subscript k shall indicate the individual components of an element of the sequence $x = \{x_k\}_{k\in\mathbb{N}}$. Likewise the subscripts l and n shall only be used for sequences of this elements or iterates in terms of the considered algorithms, e.g. $x_n = \{x_{n,k}\}_{k\in\mathbb{N}}$.

2. A transformation of the Tikhonov functional

In [28] it was shown that (4) provides a regularization method under classic assumptions on the operator. The key idea was to transform the Tikhonov type functional by means of a superposition operator into a standard formulation. Below we give a brief summary on some results presented in [28] and consequently show additional properties of the transformation operator.

Definition 2.1. We denote by $\eta_{p,q}$ the function given by

$$\begin{aligned} \eta_{p,q} \;:\; \mathbb{R} &\to \mathbb{R} \\ r &\mapsto \operatorname{sign}(r)\,|r|^{\frac{q}{p}} \;, \end{aligned} \tag{7}$$

for $0 < p \le 1$ and $1 \le q \le 2$.

Definition 2.2. We denote by $\mathcal{N}_{p,q}$ the superposition operator given by

$$\mathcal{N}_{p,q} \;:\; x \mapsto \{\eta_{p,q}(x_k)\}_{k\in\mathbb{N}} \;, \tag{8}$$

where $x \in \ell_q$, $0 < p \le 1$ and $1 \le q \le 2$.

Proposition 2.3. *For all $0 < p \le 1$, $1 \le q \le 2$, $x \in \ell_q$ and $\mathcal{N}_{p,q}$ as in definition 2.2 holds $\mathcal{N}_{p,q}(x) \in \ell_p$, and the operator $\mathcal{N}_{p,q} : \ell_q \to \ell_p$ is bounded, continuous and bijective.*

Using the concatenation operator:

$$\begin{aligned} \mathcal{G} \;:\; \ell_q &\to \ell_2 \\ x &\mapsto \mathcal{F} \circ \mathcal{N}_{p,q}(x) \;. \end{aligned} \tag{9}$$

one obtains the following two equivalent minimization problems.

Problem 1. *Let y^δ be an approximation of the right hand side of (1) with $\|y - y^\delta\| \le \delta$ and $\alpha > 0$. Let $0 < p \le 1$. Minimize*

$$\left\|\mathcal{F}(x_s) - y^\delta\right\|_2^2 + \alpha \;\|x_s\|_p^p \;, \tag{10}$$

subject to $x_s \in \ell_p$.

Problem 2. *Let y^δ be an approximation of the right hand side of (1) with $\|y - y^\delta\| \le \delta$ and $\alpha > 0$. Determine $x_s = \mathcal{N}_{p,q}(x)$, $0 < p \le 1$, $1 \le q \le 2$, where x minimizes*

$$\left\|\mathcal{G}(x) - y^\delta\right\|_2^2 + \alpha \;\|x\|_q^q \;, \tag{11}$$

subject to $x \in \ell_q$.

Proposition 2.4. *Problem 1 and problem 2 are equivalent.*

[28] provides classical results on existence of minimizers, stability and convergence for the particular Tikhonov approach considered here. These results are obtained via the observation of weak (sequential) continuity of the transformation operator.

3. Properties of the operator $\mathcal{N}_{p,q}$

Let us start with an analysis of the operator $\mathcal{N}_{p,q}$.The following proposition was given in [28]. We restate the proof as we directly use it for the following proposition.

Proposition 3.1. *The operator $\mathcal{N}_{p,q} : \ell_q \to \ell_q$ is weakly (sequentially) continuous for $0 < p \leq 1$ and $1 < q \leq 2$, i.e.,*

$$x_n \stackrel{\ell_q}{\rightharpoonup} x \Longrightarrow \mathcal{N}_{p,q}(x_n) \stackrel{\ell_q}{\rightharpoonup} \mathcal{N}_{p,q}(x) \ . \tag{12}$$

Proof. We set $r = q/p + 1$ and observe $r \geq 2$. A sequence in ℓ_q is weakly convergent if and only if the coefficients converge and the sequence is norm bounded. Thus we conclude from the weak convergence of x_n that $\|x_n\|_q \leq C$ and $x_{n,k} \to x_k$. As $r \geq q$, we have a continuous embedding of ℓ_r into ℓ_q, i.e.,

$$\|x_n\|_r \leq \|x_n\|_q \leq C \ ,$$

which shows that also

$$x_n \stackrel{\ell_r}{\rightharpoonup} x$$

holds. The operator $(\mathcal{N}_{p,q}(x))_k = \operatorname{sgn}(x_k)|x_k|^{r-1}$ is the derivative of the function

$$f(x) = r^{-1} \cdot \|x\|_r^r \ ,$$

or, in other words, $\mathcal{N}_{p,q}(x)$ is the *duality mapping* on ℓ_r with respect to the weight function

$$\varphi(t) = t^{r-1}$$

(for more details on duality mappings we refer to [34]). Now it is a well known result that every duality mapping on ℓ_r is weakly (sequentially) continuous, see, e.g. [34], Prop. 4.14. Thus we obtain

$$x_n \stackrel{\ell_r}{\rightharpoonup} x \Longrightarrow \mathcal{N}_{p,q}(x_n) \stackrel{\ell_r}{\rightharpoonup} \mathcal{N}_{p,q}(x) \ .$$

Again, as $\mathcal{N}_{p,q}(x_n)$ is weakly convergent, we have $\{\mathcal{N}_{p,q}(x_n)\}_k \to \{\mathcal{N}_{p,q}(x)\}_k$. For for $p \leq 1$, $q \geq 1$ holds $q \leq q^2/p$ and thus we have $\|x\|_{q^2/p} \leq \|x\|_q$. It follows

$$\|\mathcal{N}_{p,q}(x_n)\|_q^q = \sum_k |x_{n,k}|^{q^2/p} = \|x_n\|_{q^2/p}^{q^2/p} \leq \|x_n\|_q^{q^2/p} \leq C^{q^2/p} \ ,$$

i.e., $\mathcal{N}_{p,q}(x_n)$ is also uniformly bounded with respect to ℓ_q and thus also weakly convergent. □

The same result holds with respect to weak ℓ_2-convergence:

Proposition 3.2. *The operator $\mathcal{N}_{p,q} : \ell_2 \to \ell_2$ is weakly (sequentially) continuous w.r.t. ℓ_2 for $0 < p \leq 1$ and $1 < q \leq 2$, i.e.,*

$$x_n \stackrel{\ell_2}{\rightharpoonup} x \Longrightarrow \mathcal{N}_{p,q}(x_n) \stackrel{\ell_2}{\rightharpoonup} \mathcal{N}_{p,q}(x) \ . \tag{13}$$

Proof. First we have for $x \in \ell_2$ with $2q/p \geq 2$

$$\|\mathcal{N}_{p,q}(x)\|_2^2 = \sum_k |x_k|^{2q/p} = \|x\|_{2q/p}^{2q/p} \leq \|x\|_2^{2q/p} < \infty \ ,$$

i.e. $\mathcal{N}_{p,q}(x) \in \ell_2$ for $x \in \ell_2$. Setting again $r = q/p + 1$, the remainder of the proof follows the lines of the previous one, with $\|\cdot\|_q$ replaced by $\|\cdot\|_2$. □

Next, we want to investigate the Fréchet derivative of $\mathcal{N}_{p,q}$. Beforehand we need the following Lemma:

Lemma 3.3. *The map* $x \mapsto \operatorname{sgn}(x)\,|x|^\alpha\,,\ x \in \mathbb{R}$ *is Hölder continuous with exponent* α*, for* $\alpha \in (0,1]$*. Moreover we have locally for* $\alpha > 1$ *and globally for* $\alpha \in (0,1]$*:*

$$|\operatorname{sgn}(x)\,|x|^\alpha - \operatorname{sgn}(y)\,|y|^\alpha| \leq \kappa\ |x-y|^\beta \ , \tag{14}$$

where $\beta = \min(\alpha, 1)$*.*

Proof. As the problem is symmetric with respect to x and y, we assume $x \geq y$ w.l.o.g. Moreover we introduce the parameter γ s.t.: $\gamma\,|y| = |x|$. For $\gamma \in [1,\infty)$ and $\alpha \in (0,1]$ we have

$$(\gamma^\alpha - 1) \leq (\gamma - 1)^\alpha \ , \tag{15}$$

which can be obtained by comparing the derivatives of $(\gamma^\alpha - 1)$ and $(\gamma - 1)^\alpha$, and the fact that we have equality for $\gamma = 1$. Additionally we have for $\gamma \in [0,\infty)$ and $\alpha \in (0,1]$

$$(\gamma^\alpha + 1) \leq 2\,(\gamma + 1)^\alpha \ . \tag{16}$$

As it will be crucial that the constant in inequality (16) is independent of γ, we derive the factor two. The ratio

$$\frac{(\gamma^\alpha + 1)}{(\gamma + 1)^\alpha} \ ,$$

is monotonously increasing for $\gamma \in (0,1]$ and monotonously decreasing for $\gamma \in (1,\infty)$, which can be easily seen from its derivative. Hence the maximum is attained at $\gamma = 1$ and given by $2^{1-\alpha}$, which yields,

$$\frac{(\gamma^\alpha + 1)}{(\gamma + 1)^\alpha} \leq 2^{1-\alpha} \leq 2 \ .$$

Consequently we can conclude in the case of $x \cdot y > 0$ that $\gamma \geq 1$ and further

$$\begin{aligned} |\operatorname{sgn}(x)\,|x|^\alpha - \operatorname{sgn}(y)\,|y|^\alpha| &= |\gamma^\alpha\,|y|^\alpha - |y|^\alpha| = |(\gamma^\alpha - 1)|y|^\alpha| \\ &\overset{(15)}{\leq} |(\gamma - 1)^\alpha\,|y|^\alpha| = |x-y|^\alpha \ , \end{aligned}$$

and for $x \cdot y < 0$ we have:

$$\begin{aligned} |\operatorname{sgn}(x)\,|x|^\alpha - \operatorname{sgn}(y)\,|y|^\alpha| &= |\gamma^\alpha\,|y|^\alpha + |y|^\alpha| = |(\gamma^\alpha + 1)|y|^\alpha| \\ &\overset{(16)}{\leq} 2\ |(\gamma + 1)^\alpha\,|y|^\alpha| = 2\ |x-y|^\alpha \ . \end{aligned}$$

□

Remark 3.4. Subsequently Lemma 3.3 will be used to uniformly estimate the remainder of a Taylor series. As shown in the proof this immedeatly holds true for $\alpha \in (0, 1]$. In the case of the Lipschitz estimate this is valid only locally. However as all sequences in the following Proposition 3.5 are bounded and we are only interested in a local estimate, Lemma 3.3 can be directly applied.

Proposition 3.5. *The Fréchet derivative of $\mathcal{N}_{p,q} : \ell_q \to \ell_q$, $0 < p \le 1$, $1 < q \le 2$ is given by the sequence*

$$\mathcal{N}'_{p,q}(x)h = \left\{ \frac{q}{p} |x_k|^{(q-p)/p} \cdot h_k \right\} . \tag{17}$$

Proof. Let $w := \min\left(\frac{q}{p} - 1, 1\right) > 0$. The derivative of the function $\eta_{p,q}(t) = |t|^{q/p} \operatorname{sgn}(t)$ is given by $\eta'_{p,q}(t) = \frac{q}{p} |t|^{(q-p)/p}$. Let

$$\eta_{p,q}(t+\tau) - \eta_{p,q}(t) - \eta'_{p,q}(t)\,\tau \ := r(t,\tau) \ ,$$

then the absolute value of $r(t,\tau)$ may be represented as follows:

$$\begin{aligned} |r(t,\tau)| &= \left| \int_t^{t+\tau} \frac{q}{p} \frac{q-p}{p} (t+\tau-s)\, |s|^{\frac{q}{p}-2} \, \mathrm{d}s \right| \\ &= \left| \left[\frac{q}{p} (t+\tau-s)\, |s|^{q/p-1} \operatorname{sgn}(s) \right]_t^{t+\tau} + \int_t^{t+\tau} \frac{q}{p} \, |t|^{q/p-1} \operatorname{sgn}(t) \, \mathrm{d}s \right| \\ &= \left| \frac{q}{p} \tau \left(|\xi|^{q/p-1} \operatorname{sgn}(\xi) - |t|^{q/p-1} \operatorname{sgn}(t) \right) \right| \overset{(14)}{\le} \kappa \frac{q}{p} |\tau|^{w+1} \ , \end{aligned}$$

with $\xi \in (t, t+\tau)$ and by using Lemma 3.3 with $\alpha = q/p - 1$. Hence we may write for $\|h\| = \|\{h_k\}\|$ sufficiently small

$$\begin{aligned} \left\| \mathcal{N}_{p,q}(x+h) - \mathcal{N}_{p,q}(x) - \mathcal{N}'_{p,q}(x)h \right\|_q^q = \|\{r(x_k,h_k)\}\|_q^q &= \sum_k |r(x_k,h_k)|^q \\ &\le \sum_k \left(\frac{\kappa\, q}{p} \right)^q |h_k|^{q(w+1)} \\ &\le \left(\frac{\kappa\, q}{p} \right)^q \max\left(\{|h_k|^{qw}\}\right) \sum_k |h_k|^q \ , \end{aligned}$$

where the estimate holds uniformly with respect to κ (see Remark 3.4). Hence we conclude $\|\{r(x_k,h_k)\}\|_q \,/\, \|h\|_q \to 0$ for $\|h\|_q \to 0$ and obtain for the derivative $\mathcal{N}'_{p,q}(x)h = \left\{\eta'_{p,q}(x_k)h_k\right\}$. □

Remark 3.6. Please note that the result of Proposition 3.5 also holds in the case of the operator $\mathcal{N}_{p,q} : \ell_2 \to \ell_2$, as one can immediately see from the proof.

Lemma 3.7. *The operator $\mathcal{N}'_{p,q}(x)$ is self-adjoint with respect to ℓ_2.*

Proof. We have $\langle \mathcal{N}'_{p,q}(x)h, z\rangle = \frac{q}{p} \sum |x_k|^{(q-p)/p} h_k z_k = \langle h, \mathcal{N}'_{p,q}(x)z\rangle$. □

Please note that the Fréchet derivative of the operator $\mathcal{N}_{p,q}$ and its adjoint can be understand as (infinite dimensional) diagonal matrices, that is

$$\mathcal{N}'_{p,q}(x) = \operatorname{diag}\left(\frac{q}{p} |x_k|^{(q-p)/p} \right) \ ,$$

and $\mathcal{N}'_{p,q}(x)h$ is then a matrix - vector multiplication.

4. Properties of the concatenation operator $\mathcal{G}$

The convergence of the surrogate functional approach, applied to the transformed Tikhonov functional (11), relies mainly on some mapping properties of the operator $\mathcal{G} = \mathcal{F} \circ \mathcal{N}_{p,q}$. In the following, we will assume that the operator $\mathcal{F}$ is Fréchet differentiable and $\mathcal{F}, \mathcal{F}'$ fulfill the following conditions:

$$x_n \rightharpoonup x \Longrightarrow \quad \mathcal{F}(x_n) \to \mathcal{F}(x) \text{ for } n \to \infty \tag{18}$$

$$x_n \rightharpoonup x \Longrightarrow \quad \mathcal{F}'(x_n)^* z \to \mathcal{F}'(x)^* z \text{ for } n \to \infty \text{ and all } z \tag{19}$$

$$\|\mathcal{F}'(x) - \mathcal{F}'(x')\| \leq L\|x - x'\| \ \text{ locally } . \tag{20}$$

Convergence and weak convergence in (18),(19) has to be understand with respect to ℓ_2. The main goal of this Section is to show that the concatenation operator $\mathcal{G}$ is Fréchet differentiable and that these operators also fulfill the conditions given above. First we obtain

Proposition 4.1. *Let $\mathcal{F} : \ell_q \to \ell_2$ be strongly continuous w.r.t. ℓ_q, i.e.*

$$x_n \stackrel{\ell_q}{\rightharpoonup} x \Longrightarrow \mathcal{F}(x_n) \stackrel{\ell_q}{\to} \mathcal{F}(x). \tag{21}$$

Then $\mathcal{F}(\mathcal{N}_{p,q})$ is also strongly continuous w.r.t. ℓ_q. If $\mathcal{F} : \ell_2 \to \ell_2$ is strongly continuous w.r.t. ℓ_2, then $\mathcal{F}(\mathcal{N}_{p,q})$ is also strongly continuous w.r.t. ℓ_2.

Proof. If $x_n \stackrel{\ell_q}{\rightharpoonup} x$, then, by Proposition 3.1, also $\mathcal{N}_{p,q}(x_n) \stackrel{\ell_q}{\rightharpoonup} \mathcal{N}_{p,q}(x)$, and due to the strong continuity of $\mathcal{F}$ follows $\mathcal{F}(\mathcal{N}_{p,q}(x_n)) \to \mathcal{F}(\mathcal{N}_{p,q}(x))$. The second part of the Propositions follows in the same way by Proposition 3.2. □

By the chain rule follows immediately

Lemma 4.2. *Let $\mathcal{F} : \ell_q \to \ell_2$ be Fréchet differentiable. Then*

$$(\mathcal{F} \circ \mathcal{N}_{p,q})'(x) = \mathcal{F}'(\mathcal{N}_{p,q}(x)) \cdot \mathcal{N}'_{p,q}(x) \ , \tag{22}$$

where the multiplication has to be understand as a matrix product. The adjoint (with respect to ℓ_2) of the Fréchet derivative is given by

$$\left((\mathcal{F} \circ \mathcal{N}_{p,q})'(x)\right)^* = \mathcal{N}'_{p,q}(x) \cdot \mathcal{F}'(\mathcal{N}_{p,q}(x))^* \ . \tag{23}$$

Proof. It is

$$\begin{aligned}
\langle ((\mathcal{F} \circ \mathcal{N}_{p,q})'(x))\, u, z\rangle &= \langle \mathcal{F}'(\mathcal{N}_{p,q}(x)) \cdot \mathcal{N}'_{p,q}(x) \cdot u, z\rangle \\
&= \langle \mathcal{N}'_{p,q}(x) \cdot u, \mathcal{F}'(\mathcal{N}_{p,q}(x))^* \cdot z\rangle \\
&= \langle u, \mathcal{N}'_{p,q}(x) \cdot \mathcal{F}'(\mathcal{N}_{p,q}(x))^* z\rangle \ ,
\end{aligned}$$

as $\mathcal{N}'_{p,q}(x)$ is self adjoint. □

We further need the following result:

Lemma 4.3. *Let $\mathcal{B} : \ell_q \to \ell_q$ be a (infinite dimensional) diagonal matrix with diagonal elements $b = \{b_k\}$. Then*

$$\|\mathcal{B}\| \leq \|b\|_q \tag{24}$$

Proof. The assertion follows by

$$\|\mathcal{B}\|^q = \sup_{\|u\|\leq 1} \|\mathcal{B}u\|^q = \sup_{\|u\|\leq 1} \sum_k |b_k \cdot u_k|^q \leq \sum_k |b_k|^q \ .$$

□

Hence we may identify the operator $\mathcal{N}'_{p,q}(x_n)$ with its sequence and vice versa. Now we can conclude

Proposition 4.4. *Let $x_n \rightharpoonup x$ with respect to ℓ_2, $z \in \ell_2$ and let q and p be such that $(q-p)/p \geq 1$. Assume that*

$$(\mathcal{F}'(x_n))^* z \to (\mathcal{F}'(x))^* z \tag{25}$$

holds for any weakly convergent sequence. Then also

$$\left((\mathcal{F} \circ \mathcal{N}_{p,q})'(x_n)\right)^* z \to \left((\mathcal{F} \circ \mathcal{N}_{p,q})'(x)\right)^* z \ . \tag{26}$$

Proof. As $x_n \stackrel{\ell_2}{\rightharpoonup} x$, we have in particular $x_{n,k} \to x_k$ for fixed k. We conclude by Proposition 3.2

$$\mathcal{N}_{p,q}(x_n) \stackrel{\ell_2}{\rightharpoonup} \mathcal{N}_{p,q}(x). \tag{27}$$

The sequence $\mathcal{N}'_{p,q}(x_n)$ is given elementwise by

$$\frac{q}{p}|x_{n,k}|^{(q-p)/p} \to \frac{q}{p}|x_k|^{(q-p)/p} \ ,$$

and thus the coefficients of $\mathcal{N}'_{p,q}(x_n)$ converge to the coefficients of $\mathcal{N}'_{p,q}(x)$. In order to show weak convergence of the sequences, it remains to show that $\{\frac{q}{p}|x_{n,k}|^{(q-p)/p}\}$ stays uniformly bounded: We have

$$\begin{aligned} \|\mathcal{N}'_{p,q}(x_n)\|_2^2 &= \frac{q^2}{p} \sum \left(|x_{n,k}|^{(q-p)/p}\right)^2 \\ &\leq \frac{q^2}{p} \sum_k |x_{n,k}|^{2(q-p)/p} \end{aligned}$$

As $(q-p)/p \geq 1$ and $\|x\|_r \leq \|x\|_q$ for $q \leq r$ we conclude with $r = 2(q-p)/p \geq 2$

$$\|\mathcal{N}'_{p,q}(x_n)\|^2 \leq \frac{q^2}{p}\|x_n\|_r^r \leq \frac{q^2}{p}\|x_n\|_2^r \leq C \ , \tag{28}$$

as weakly convergent sequences are uniformly bounded. Thus we conclude

$$\mathcal{N}'_{p,q}(x_n) \rightharpoonup \mathcal{N}'_{p,q}(x) \ .$$

With the same arguments we get for fixed z

$$\mathcal{N}'_{p,q}(x_n)z \rightharpoonup \mathcal{N}'_{p,q}(x)z \ .$$

The convergence of this sequence holds also in the strong sense. For this, it is sufficient to show that $\lim_{n\to\infty} \|\mathcal{N}'_{p,q}(x_n)z\| = \|\mathcal{N}'_{p,q}(x)z\|$ holds: As x_n is weakly convergent, the sequence is also uniformly bounded, i.e. $\|x_n\|_{\ell_2} \le \tilde{C}$, and thus $|x_{n,k}| \le \tilde{C}$. Setting $\gamma_k := \tilde{C}^{2(q-p)/p} z_k^2$ we observe $|x_{n,k}|^{2(q-p)/p} \cdot z_k^2 \le \gamma_k$ and

$$\frac{q^2}{p} \sum_k |x_{n,k}|^{2(q-p)/p} \cdot z_k^2 \le \frac{q^2}{p} \sum_k \gamma_k = \tilde{C}^{2(q-p)/p} \frac{q^2}{p} \sum_k z_k^2 = \frac{q^2}{p} \gamma \|z\|_2^2 < \infty \ .$$

Therefore, by the dominated convergence theorem, we can interchange limes and summation, i.e.

$$\begin{aligned} \lim_{n\to\infty} \|\mathcal{N}'_{p,q}(x_n) z\|^2 &= \lim_{n\to\infty} \frac{q^2}{p} \sum_k |x_{n,k}|^{2(q-p)/p} \cdot z_k^2 \\ &= \frac{q^2}{p} \sum_k \lim_{n\to\infty} |x_{n,k}|^{2(q-p)/p} \cdot z_k^2 \\ &= \frac{q^2}{p} \sum_k |x_k|^{2(q-p)/p} \cdot z_k^2 = \frac{q^2}{p} \|\mathcal{N}'_{p,q}(x) z\|^2 \ , \end{aligned}$$

and thus

$$\mathcal{N}'_{p,q}(x_n) z \xrightarrow{\ell_2} \mathcal{N}'_{p,q}(x) z \ . \tag{29}$$

We further conclude

$$\begin{aligned} &\| ((\mathcal{F} \circ \mathcal{N}_{p,q})'(x_n))^* z - ((\mathcal{F} \circ \mathcal{N}_{p,q})'(x))^* z \| \\ &\qquad = \|\mathcal{N}'_{p,q}(x_n) \mathcal{F}'(\mathcal{N}_{p,q}(x_n))^* z - \mathcal{N}'_{p,q}(x) \mathcal{F}'(\mathcal{N}_{p,q}(x))^* z\| \\ &\qquad \le \underbrace{\|\mathcal{N}'_{p,q}(x_n) \mathcal{F}'(\mathcal{N}_{p,q}(x_n))^* z - \mathcal{N}'_{p,q}(x_n) \mathcal{F}'(\mathcal{N}_{p,q}(x))^* z\|}_{D_1} \\ &\qquad\quad + \underbrace{\|\mathcal{N}'_{p,q}(x_n) \mathcal{F}'(\mathcal{N}_{p,q}(x))^* z - \mathcal{N}'_{p,q}(x) \mathcal{F}'(\mathcal{N}_{p,q}(x))^* z\|}_{D_2} \ . \end{aligned}$$

The two terms can be estimated as follows:

$$D_1 \le \underbrace{\|\mathcal{N}'_{p,q}(x_n)\|}_{\stackrel{(28)}{\le} C} \, \underbrace{\| \mathcal{F}'(\mathcal{N}_{p,q}(x_n))^* z - \mathcal{F}'(\mathcal{N}_{p,q}(x))^* z\|}_{\stackrel{(25),(27)}{\longrightarrow} 0}$$

and therefore $D_1 \to 0$. For D_2 we get with $\tilde{z} := \mathcal{F}'(\mathcal{N}_{p,q}(x))^* z$

$$D_2 = \|\mathcal{N}'_{p,q}(x_n) \tilde{z} - \mathcal{N}'_{p,q}(x) \tilde{z}\| \xrightarrow{(29)} 0 \ ,$$

which concludes the proof. □

In the final step of this section we will show the Lipschitz continuity of the derivative.

Proposition 4.5. *Assume that $\mathcal{F}'(x)$ is (locally) Lipschitz continuous with constant L. Then $(\mathcal{F} \circ \mathcal{N}_{p,q})'(x)$ is locally Lipschitz for $p < 1$ and $1 \le q \le 2$ such that $2p < q$.*

Proof. The function $f(t) = |t|^s$ with $s > 1$ is locally Lipschitz continuous, hence we have on a bounded interval:

$$|f(t) - f(\tilde{t})| \le s \max |t|^{s-1} |t - \tilde{t}| \ . \tag{30}$$

Assume $x \in B_\rho(x_0)$, then $\|x\|_2 \le \|x - x_0\|_2 + \|x_0\|_2 \le \rho + \|x_0\|_2$, and therefore

$$\sup_{x \in B_\rho(0)} \|x\|_\infty \le \rho + \|x_0\|_2 =: \tilde{\rho} \ .$$

We have $s := (q-p)/p \ge 1$, and therefore (30) holds. $\mathcal{N}'_{p,q}(x)$ is a diagonal matrix, thus we obtain with Lemma 4.3 for $x, \tilde{x} \in B_\rho(x_0)$

$$\begin{aligned} \|\mathcal{N}'_{p,q}(x) - \mathcal{N}'_{p,q}(\tilde{x})\|^2 = & \sum_k \left(|x_k|^{(q-p)/p} - |\tilde{x}_k|^{(q-p)/p} \right)^2 \\ \stackrel{(30)}{\le} & \left(\frac{q-p}{p} \tilde{\rho}^{(q-2p)/p} \right)^2 \sum_k |x_k - \tilde{x}_k|^2 \\ \le & \left(\frac{q-p}{p} \tilde{\rho}^{(q-2p)/p} \right)^2 \|x - \tilde{x}\|_2^2 \ . \end{aligned}$$

With the same arguments we show that $\mathcal{N}_{p,q}$ is Lipschitz,

$$\|\mathcal{N}_{p,q}(x) - \mathcal{N}_{p,q}(\tilde{x})\|_2 \le \frac{q}{p} \tilde{\rho}^{(q-p)/p} \|x - \tilde{x}\| \ .$$

The assertion now follows from

$$\begin{aligned} & \|\mathcal{F}'(\mathcal{N}_{p,q}(x))\mathcal{N}'_{p,q}(x) - \mathcal{F}'(\mathcal{N}_{p,q}(\tilde{x}))\mathcal{N}'_{p,q}(\tilde{x})\| \\ & \qquad \le \| \left(\mathcal{F}'(\mathcal{N}_{p,q}(x)) - \mathcal{F}'(\mathcal{N}_{p,q}(\tilde{x}))\right) \mathcal{N}'_{p,q}(x)\| \\ & \qquad\qquad + \|\mathcal{F}'(\mathcal{N}_{p,q}(\tilde{x})) \left(\mathcal{N}'_{p,q}(x) - \mathcal{N}'_{p,q}(\tilde{x})\right) \| \\ & \qquad \le L \|\mathcal{N}_{p,q}(x) - \mathcal{N}_{p,q}(\tilde{x})\| \|\mathcal{N}'_{p,q}(x)\| \\ & \qquad\qquad + \|\mathcal{F}'(\mathcal{N}'_{p,q}(\tilde{x}))\| \|\mathcal{N}'_{p,q}(x) - \mathcal{N}'_{p,q}(\tilde{x})\| \\ & \qquad \le \tilde{L} \|x - \tilde{x}\| \ , \end{aligned}$$

with

$$\begin{aligned} \tilde{L} = 2 \max \Big(& L \cdot \max_{x \in B_\rho} \|\mathcal{N}'_{p,q}(x)\| \cdot \frac{q}{p} \tilde{\rho}^{(q-p)/p}, \\ & \max_{x \in B_\rho} \|\mathcal{F}'(\mathcal{N}_{p,q}(x))\| \cdot \frac{q-p}{p} \tilde{\rho}^{(q-2p)/p} \Big) \ . \end{aligned}$$

□

Combining the results of Lemma 4.2, Proposition 4.1, 4.4 and 4.5, we get

Proposition 4.6. *Assume that the operator $\mathcal{F} : \ell_2 \to \ell_2$ is Fréchet differentiable and fulfills conditions* (18)-(20). *Then $\mathcal{G} = \mathcal{F} \circ \mathcal{N}_{p,q}$ is also Fréchet differentiable. If the parameters $0 < p < 1$ and $1 < q \le 2$ fulfill the relation $2p < q$, then*

$$x_n \rightharpoonup x \Longrightarrow \quad \mathcal{G}(x_n) \to \mathcal{G}(x) \text{ for } n \to \infty \tag{31}$$

$$x_n \rightharpoonup x \Longrightarrow \quad \mathcal{G}'(x_n)^* z \to \mathcal{G}'(x)^* z \text{ for } n \to \infty \text{ and all } z \tag{32}$$

$$\|\mathcal{G}'(x) - \mathcal{G}'(x')\| \le L \|x - x'\| \quad \text{locally} \tag{33}$$

holds.

Proof. Proposition 4.1 yields (31). According to Lemma 4.2, $\mathcal{G}$ is differentiable. If $q > 2p$ then the conditions of Proposition 4.4 hold and thus (32). Morover, the condition $q > 2p$ is equivalent to $(q-p)/p > 1$, i.e., Proposition 4.5 holds and therefore (33). □

5. Minimization by surrogate functionals

In order to compute a minimizer of the Tikhonov functional (4), we can either use algorithms that minimize (4) directly or, alternatively, we can try to minimize (10). It turns out that the transformed functional, with an ℓ_q-norm and $q > 1$ as penalty, can be minimized more effectively by the proposed or other standard algorithms. The main drawback of the transformed functional is that, due to the transformation, we have to deal with a nonlinear operator, even if the original operator $\mathcal{F}$ is linear.
A well investigated algorithm for the minimization of the Tikhonov functional with ℓ_q penalty that works for all $1 \le q \le 2$ is the minimization via surrogate functionals. The method was introduced by Daubechies, Defrise and De Mol [12] for penalties with $q \ge 1$ and linear operator $\mathcal{F}$. Later on, the method was generalized in [13, 14, 23] to nonlinear operators $\mathcal{G} = \mathcal{F} \circ \mathcal{N}_{p,q}$ also. The method works as follows: For given iterate x_n, we consider the surrogate functional

$$J_\alpha^s(x, x_n) = \|y^\delta - \mathcal{G}(x)\|^2 + \alpha \|x\|_q^q + C\|x - x_n\|_2^2 - \|\mathcal{G}(x) - \mathcal{G}(x_n)\|_2^2 \tag{34}$$

and determine the new iterate as

$$x_{n+1} = \arg\min_x J_\alpha^s(x, x_n) \ . \tag{35}$$

The constant C in the definition of the surrogate functional has to be chosen large enough, for more details see [13, 23]. Now it turns out that the functional $J_\alpha^s(x, x_n)$ can be easily minimized by means of a fixed point iteration. For fixed x_n, the functional is minimized by the limit of the fixed point iteration

$$x_{n,l+1} = \Phi_q^{-1}\left(\frac{1}{C}\mathcal{G}'(x_{n,l})^* \left(y^\delta - \mathcal{G}(x_n)\right) + x_n\right) \ , \tag{36}$$

$x_{n,0} = x_n$ and $x_{n+1} = \lim_{l\to\infty} x_{n,l}$. For $q > 1$, the map Φ_q is defined pointwise on the coefficients of a sequence by

$$\Phi_q(x_k) = x_k + \frac{\alpha \cdot q}{C}|x_k|^{q-1}\operatorname{sgn}(x_k) \ , \tag{37}$$

i.e. in order to compute the new iterate $x_{n,l+1}$ we have to solve the equation

$$\Phi_q\left(\{x_{n,l+1}\}_k\right) = \left\{\frac{1}{C}\mathcal{G}'(x_{n,l})^* \left(y^\delta - \mathcal{G}(x_n)\right) + x_n\right\}_k \tag{38}$$

for each $k \in \mathbb{N}$. It has been shown that the fixed point iteration converges to the unique minimizer of the surrogate functional $J_\alpha^s(x, x_n)$, provided the constant C is chosen large enough and the operator fulfills the requirements (18)-(20), for full details we refer the reader to [13, 23]. Moreover, it was also shown that the outer iteration (35) converges at least to a critical point of the Tikhonov functional

$$J_\alpha(x) = \|y^\delta - \mathcal{G}(x)\|^2 + \alpha\|x\|_q^q \ , \tag{39}$$

provided that the operator $\mathcal{G}$ fulfills the conditions (31)-(33)

Based on the results of Section 2, we can now formulate our main result:

Theorem 5.1. *Let $\mathcal{F} : \ell_2 \to \ell_2$ be a weakly (sequentially) closed operator fulfilling the conditions* (18) - (20)*, and choose $q > 1$ s.t. $2p < q$, with $0 < p < 1$. Then the operator $\mathcal{G}(x) = \mathcal{F} \circ \mathcal{N}_{p,q}$ is also Fréchet - differentiable and fulfills also the conditions* (31) - (33)*. The iterates x_n, computed by the surrogate functional algorithm* (35) *converge at least to a critical point of the functional*

$$J_{\alpha,q}(x) = \|y^\delta - \mathcal{G}(x)\|^2 + \alpha \|x\|_q^q \ . \tag{40}$$

If the limit of the iteration, $x_\alpha^\delta := \lim_{n\to\infty} x_n$, is a global minimizer of (40)*, , then $x_{s,\alpha}^\delta := \mathcal{N}_{p,q}(x_\alpha^\delta)$ is a global minimizer of*

$$\|y^\delta - \mathcal{F}(x)\|^2 + \alpha \|x\|_p^p \ . \tag{41}$$

Proof. According to Proposition 4.6, the operator $\mathcal{G}$ fulfills the properties necessary for the convergence of the iterates to a critical point of the functional (40), see [23], Proposition 4.7. If x_α^δ is a global minimizer of (40), then, according to Proposition 2.4, $x_{s,\alpha}^\delta$ is a minimizer of (4). □

One may notice, that the main result in Theorem 5.1 is stated with respect to the transformed functional. Only in the case of a global minimizer as the limit of the iteration the result is also interpreted in terms of the original functional. In fact this can be slightly generalized. Assuming that the limit of the iteration is no saddle point, i.e., we obtain a local minimizer or a stationary point where the objective function is locally constant, we can directly translate this result to the original functional. Let x_α^δ be the limit of the iteration and assume there exists a neighborhood $U_\epsilon(x_\alpha^\delta)$ such that:

$$\forall x \in U_\epsilon(x_\alpha^\delta) \quad : \quad \|y^\delta - \mathcal{G}(x)\|^2 + \alpha\|x\|_q^q \geq \|y^\delta - \mathcal{G}(x_\alpha^\delta)\|^2 + \alpha \|x_\alpha^\delta\|_q^q \ . \tag{42}$$

Let $M := \{x_s : \mathcal{N}_{p,q}^{-1}(x_s) \in U_\epsilon(x_\alpha^\delta)\}$ and $x_{s,\alpha}^\delta := \mathcal{N}_{p,q}(x_\alpha^\delta)$, then we can derive that:

$$\forall x_s \in M \quad : \quad \|y^\delta - \mathcal{F}(x_s)\|^2 + \alpha\|x_s\|_p^p \geq \|y^\delta - \mathcal{F}(x_{s,\alpha}^\delta)\|^2 + \alpha \|x_{s,\alpha}^\delta\|_p^p \ . \tag{43}$$

Since $\mathcal{N}_{p,q}$ and $\mathcal{N}_{p,q}^{-1}$ are continuous there exists a neighborhood U_{ϵ_s} around the solution for the original functional $x_{s,\alpha}^\delta$, such that $U_{\epsilon_s}(x_{s,\alpha}^\delta) \subseteq M$. Consequently also stationary points and local minima of the transformed functional translate to the original functional.

6. A global minimization strategy for the transformed Tikhonov functional: the case $q = 2$

The minimization by surrogate functionals, presented in Section 5, usually guarantees the reconstruction of a critical point of the transformed functional. If we have not found the global minimizer of the transformed functional, then this also implies that we have not reconstructed the global minimizer for the original functional. In general this is not desirable. However there exists an algorithm that will - under some restrictions - guarantee the reconstruction of a global minimizer. This algorithm works in the case of $q = 2$ only, i.e., we are looking for a global minimizer of the standard Tikhonov functional

$$J_\alpha(x) = \|y^\delta - \mathcal{G}(x)\|^2 + \alpha \|x\|_2^2 \tag{44}$$

with $\mathcal{G}(x) = \mathcal{F}(\mathcal{N}_{p,2}(x))$. For the minimization of the functional, we want to use the TIGRA method [31, 32]. The main ingredient of the algorithm is a standard gradient method for the minimization of (44), i.e., the iteration is given by

$$x_{n+1} = x_n + \beta_n \left(\mathcal{G}'(x_n)^*(y^\delta - \mathcal{G}(x_n)) - \alpha x_n \right). \tag{45}$$

The following arguments are taken out of [32], where the reader finds all the proofs and further details. If the operator $\mathcal{G}$ is twice Fréchet differentiable, its first derivative is Lipschitz continuous, and a solution $x^\dagger$ of $\mathcal{G}(x) = y$ fulfills a smoothness condition

$$x^\dagger = \mathcal{G}'(x^\dagger)^*\omega , \tag{46}$$

then it has been shown that (44) is locally convex around a global minimizer x_α^δ, and the area of convexity, $K_{r(\alpha)}(x_\alpha^\delta)$, depends in particular on the Lipschitz constant, α and $\|w\|$, for details we refer to [32], equation (3.25). Now if we know an initial iterate $x_0 \in K_{r(\alpha)}(x_\alpha^\delta)$ then the scaling parameter β_n can be chosen s.t. all iterates stay within the area of convexity and $x_n \to x_\alpha^\delta$ as $n \to \infty$. The main drawback of this result is that the area of convexity shrinks to zero if $\alpha \to 0$. Thus we would need a very good initial iterate for small α, which is not reasonable. Fortunately, it can also be shown that the area of convexity grows to infinity if $\alpha \to \infty$. Based on that observation it turns out that, if we know e.g. an estimate for $\|x^\dagger - x_0\|$, we can find $\alpha_0 > \alpha$ large enough s.t. $x_0 \in K_{r(\alpha_0)}(x_{\alpha_0}^\delta)$, where $x_{\alpha_0}^\delta$ denotes a global minimizer of (44) with regularization parameter α_0. As $x_0 \in K_{r(\alpha_0)}(x_{\alpha_0}^\delta)$ we can use the gradient method (45) to reconstruct the minimizer $x_{\alpha_0}^\delta$. Choosing $s < 1$ large enough (for quantitative estimates, see [32]) and setting $\alpha_n = \max\{\alpha, s^n\alpha_0\}$ it has been found that $x_{\alpha_{n-1}}^\delta \in K_{r(\alpha_n)}(x_{\alpha_n}^\delta)$, and thus (45) with initial iterate $x_{\alpha_{n-1}}^\delta$ (or a sufficient close element) and α replaced by α_n converges to a global minimizer $x_{\alpha_n}^\delta$ of $J_{\alpha_n}(x)$. After a finite number of iterations (w.r.t the regularization parameter) we have $\alpha_{n^*} = \alpha$ and find by the iteration (45) with initial value $x_{\alpha_{n^*-1}}^\delta$ the minimizer x_α^δ we were looking for. The final algorithm looks as follows:

Input: x_0, α

Step 1: Choose $0 < s < 1$, $\alpha_0 > \alpha$ and n^* with $\alpha_{n^*} = \alpha$.

Step 2: For $n = 1, \cdots, n^*$

- If $n > 1$, set $x_0 = x_{\alpha_{n-1}}^\delta$
- Minimize $J_{\alpha_n}(x)$ by the gradient method (45) and initial value x_0.

End

In order to use the TIGRA algorithm, it remains to show that the operator $\mathcal{G} = \mathcal{F}(\mathcal{N}_{p,2})$ is twice differentiable and that the solution $x^\dagger$ of the equation $\mathcal{G}(x) = y$ fulfills the smoothness condition (46). We will restrict our attention for now to two important cases, namely where $\mathcal{F}$ is the identity (i.e. the problem of data denoising) or a linear operator.

Proposition 6.1. *The operator* $\mathcal{N}_{p,2}(x)$, $0 < p < 1$ *is twice continuous differentiable, and therefore also the operator* $\mathcal{A}\mathcal{N}_{p,2}(x)$ *with continuous and linear* $\mathcal{A}$.

Proof. The proof is completely analogous to the one of Proposition 3.5, when considering $\frac{2}{p} \geq 2$ and $r(0,h) = \frac{2}{p}\,|h|^{\frac{2}{p}-1}$. Using the Taylor expansion of the function $\eta_{p,2}(t) = |t|^{2/p}\operatorname{sgn}(t)$ for $t \neq 0$,

$$\eta_{p,2}(t+h) = \eta_{p,2}(t) + \eta_{p,2}'(t)h + \eta_{p,2}''(t)h^2 + O(|h|^3)$$

with continuous

$$\eta''_{p,2}(t) = \frac{2(2-p)}{p^2}\,\mathrm{sgn}(t)|t|^{2(1-p)/p}\ ,$$

we conclude

$$\left\|\mathcal{N}'_{p,2}(x+h)\bar{h} - \mathcal{N}'_{p,2}(x)\bar{h} - \mathcal{N}''_{p,2}(x)(\bar{h},h)\right\| / \|h\| \to 0$$

for $\|h\| \to 0$. Thus we have $\mathcal{N}''_{p,2}(x)(\bar{h},h) = \left\{\eta''_{p,q}(x_k)\bar{h}_k h_k\right\}$. The twice differentiability of $\mathcal{A}\mathcal{N}_{p,2}(x)$ follows from the linearity of $\mathcal{A}$. □

Now let us turn to the source condition (46).

Proposition 6.2. *Let $\mathcal{F} = \mathcal{I}$. Then $x^\dagger \in \ell_2$ fulfills source condition* (46) *iff it is sparse.*

Proof. As $\mathcal{I} = \mathcal{I}^*$ in ℓ_2, we have $\mathcal{F}'(\mathcal{N}_{p,2}(x^\dagger))^* = \mathcal{I}$, and it follows from (23) that

$$\left(\mathcal{F}(\mathcal{N}_{p,2}(x))'\right)^* = \mathcal{N}'_{p,2}(x)\ .$$

Therefore, the source condition (46) reads coefficient-wise as

$$\frac{2}{p}|x^\dagger_k|^{(2-p)/p}\omega_k = x^\dagger_k$$

or

$$\omega_k = \frac{2}{p}\,\mathrm{sgn}(x^\dagger_k)|x^\dagger_k|^{(2p-2)/p}\ ,$$

for $x_k \neq 0$, for $x_k = 0$ we can choose w_k freely. As $\omega_k,\ x^\dagger \in \ell_2$ and $2p-2 < 0$ this can only hold if $x^\dagger$ has only a finite number of nonzero elements. □

The case of $\mathcal{F} = \mathcal{A}$ is a little bit more complicated. In particular, we need $\mathcal{A}$ to fulfill the finite basis injectivity (FBI) property which was introduced by Bredies and Lorenz [35]. Let $\mathcal{T}$ be a finite index set, and let $\#\mathcal{T}$ be the number of elements in $\mathcal{T}$. We say that $u \in \ell_2(\mathcal{T})$ iff $u_k = 0$ for all $k \in \mathbb{N}\setminus\mathcal{T}$. The FBI property states that whenever $u, v \in \ell_2(\mathcal{T})$ with $\mathcal{A}u = \mathcal{A}v$ it follows $u = v$.This is equivalent to

$$\mathcal{A}_{|\ell_2(\mathcal{T})}u = 0 \Longrightarrow u = 0\ , \tag{47}$$

where $\mathcal{A}_{|\ell_2(\mathcal{T})}$ is the restriction of $\mathcal{A}$ to $\ell_2(\mathcal{T})$. For simplicity, we will set $\mathcal{A}_{|\ell_2(\mathcal{T})} = \mathcal{A}_{\mathcal{T}}$.

Proposition 6.3. *Assume that $x^\dagger$ is sparse, $\mathcal{T} = \{k : x^\dagger_k \neq 0\}$, and assume that $\mathcal{A}$ admits the FBI property. If*

$$\mathcal{A}^* = \mathcal{A}^T \tag{48}$$

holds, then $x^\dagger$ fulfills the source condition (46)*.*

Proof. Let. As $x^\dagger$ is sparse, $\mathcal{T}$ is finite. By $x_{\mathcal{T}}$ we denote the (finite) vector that contains only those elements of x with indices out of $\mathcal{T}$. Due to the sparsity structure of $x^\dagger$ we observe

$$\mathcal{N}'_{p,2}(x^\dagger) : \ell_2 \to \ell_2(\mathcal{T})$$

and therefore also

$$\mathcal{A}\mathcal{N}'_{p,2}(x^\dagger)\ = \mathcal{A}_{\mathcal{T}}\mathcal{N}'_{p,2}(x^\dagger) \tag{49}$$

$$\mathcal{N}'_{p,2}(x^\dagger)\mathcal{A}^* = \mathcal{N}'_{p,2}(x^\dagger)\mathcal{A}^*_{\mathcal{T}} \overset{(48)}{=} \mathcal{N}'_{p,2}(x^\dagger)\mathcal{A}^T_{\mathcal{T}}\ , \tag{50}$$

where we have used the fact that $\mathcal{N}'_{p,2}(x^\dagger)$ is self adjoint.

With $\mathcal{F} = \mathcal{A}$, (46) reads as

$$x^\dagger = \mathcal{N}'_{p,2}(x^\dagger)\mathcal{A}^*\omega = \mathcal{N}'_{p,2}(x^\dagger)\mathcal{A}^T_{\mathcal{T}}\omega \ . \tag{51}$$

The operator $\mathcal{N}'_{p,2}(x^\dagger)^{-1}$ is well defined on $\ell_2(\mathcal{T})$, and as $\ell_2(\mathcal{T}) = \mathcal{D}(\mathcal{A}_{\mathcal{T}}) = \mathcal{R}(\mathcal{A}^T_{\mathcal{T}})$, we get

$$\mathcal{A}^T_{\mathcal{T}}\omega = \mathcal{N}'_{p,2}(x^\dagger)^{-1}x^\dagger \ .$$

Now we have by the FBI property $\mathcal{N}(\mathcal{A}_{\mathcal{T}}) = \{0\}$, and therefore

$$\ell_2(\mathcal{T}) = \mathcal{N}(\mathcal{A}_{\mathcal{T}})^\perp = \overline{\mathcal{R}(\mathcal{A}^*_{\mathcal{T}})} = \overline{\mathcal{R}(\mathcal{A}^T_{\mathcal{T}})}$$

As $dim(\ell_2(\mathcal{T})) = \#\mathcal{T} < \infty$, $\mathcal{R}(\mathcal{A}^T_{\mathcal{T}}) = \ell_2(\mathcal{T})$ and therefore the generalized inverse of $\mathcal{A}^T_{\mathcal{T}}$ exists and is bounded. We finally get

$$\omega = \left(\mathcal{A}^T_{\mathcal{T}}\right)^\dagger \mathcal{N}'_{p,2}(x^\dagger)^{-1}x^\dagger \tag{52}$$

and

$$\|\omega\| \leq \| \left(\mathcal{A}^T_{\mathcal{T}}\right)^\dagger \| \|\mathcal{N}'_{p,2}(x^\dagger)^{-1}\| \|x^\dagger\| \ . \tag{53}$$

□

The condition $\mathcal{A}^* = \mathcal{A}^T$ is in particular guaranteed if $\mathcal{A}$ is considered as operator between the space ℓ_2. Please note that a similar result can be obtained for twice continuous differentiable nonlinear operators $\mathcal{F}$ if we additionally assume that $\mathcal{F}'(\mathcal{N}_{p,2}(x^\dagger))$ admits the FBI condition. Propositions 6.1-6.3 show that the TIGRA algorithm can be applied in principle to the minimization of the transformed Tikhonov functional for the case $q = 2$. Please remark that the surrogate functional approach can also be applied to the case $q < 2$. This is in particular important for the numerical realization, as we will see in the following Section.

7. Numerical Results

In this Section we will present some numerical results on the reconstruction of L_2 functions from convolution data. Considering the non-standard approach for the minimization of the Tikhonov functional, numerical tests are indispensable for assessing its benefits and drawbacks. Although the analytic properties of the nonlinear transformation are well understood, the impacts of a numerical realization are unknown. To this end we consider the simple test example of reconstructing a function from its convolution data. We define the convolution operator A by

$$y(\tau) = (Au)(\tau) = \int_{-\pi}^{\pi} r(\tau - t)u(t)\,dt =: (r * u)(\tau) \tag{54}$$

where u, r and Au are 2π-periodic functions belonging to $L_2((-\pi,\pi))$. The operator A is defined between function spaces. In order to obtain a numerical realization in accordance with the present notation we have to transform this operator, see also Section 1. For this purpose we interpret all quantities in terms of the Fourier basis or their Fourier coefficients respectively. A periodic function on $[-\pi, \pi]$ can be either expressed via the orthonormal bases formed by

$$\left\{ \frac{1}{\sqrt{2\pi}} e^{ikt} \right\}_{k \in \mathbb{Z}} \qquad \text{or} \qquad \left\{ \frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}} \cos(kt), \frac{1}{\sqrt{\pi}} \sin(kt) \right\}_{k \in \mathbb{N}} . \tag{55}$$

Naturally these representations provide also the appropriate discretization of the (linear) operator. By means of the convolution theorem a discretization via the exponential basis would lead to a diagonal system matrix. However, this also yields a complex valued matrix and complex valued vectors, a setting which is not covered by our theory. Therefore we use the trigonometrical basis from now on. This implies, that we obtain a non-diagonal linear operator (see [24] for details on the deconvolution problem and the operator). Hence by using the Fourier convolution theorem for the exponential basis and transformation formulas between the exponential and trigonometrical bases, we obtain a formulation in terms of the considered sequence spaces.

For the numerical implementation we divided the interval $[-\pi, \pi]$ into 2^{12} equidistant intervals, leading to a discretization of the convolution operator as a $2^{12} \times 2^{12}$ matrix. The convolution kernel r was defined by its Fourier coefficients with

$$\begin{aligned} a_0^r &= 0 \\ a_k^r &= (-1)^k \cdot k^{-2} \\ b_k^r &= (-1)^{k+1} \cdot k^{-2} , \end{aligned} \tag{56}$$

where

$$r(t) = a_0^r + \sum_{k \in \mathbb{N}} a_k^r \cos(kt) + b_k^r \sin(kt) . \tag{57}$$

In the numerical tests two convolution data sets were used based on two (sparse) solutions with 14 and 22 non zero components (see figure 2 and figure 5). The added noise was normally distributed and scaled with respect to the relative noise level. In all numerical tests the regularization parameter was chosen based on the quasi optimality principle (cf. [36, 37]). The chosen strategy provides an easy and attractive method of estimating the regularization parameter. It may be not the optimal choice for the considered approach, however since the exact solution was known prior to the numerical experiments we could confirm the obtained values for regularization parameter in each case.

A minimal requirement for the algorithm is its stable convergence, which we found in all our experiments. In [28] it was shown, that (4) provides a regularization method. Additionally a result on convergence rates was given, stating that the convergence of the accuracy error is at least in to order of $\sqrt{\delta}$ with respect to the Hilbert space topology, if the regularization parameter is suitably chosen a priori (cf. [38]):

$$\left\| x^* - x_\alpha^\delta \right\|_2 = \mathcal{O}(\sqrt{\delta}) . \tag{58}$$

Using the considered approach and algorithm respectively, we could observe this particular result on the convergence rates also in our numerical tests. Figure 1 shows the observed rates of convergence for decreasing noise levels δ. One may notice that the convergence slows down for very small values of δ. This behavior is well known and is caused by a finite accuracy in the numerical implementation. Another

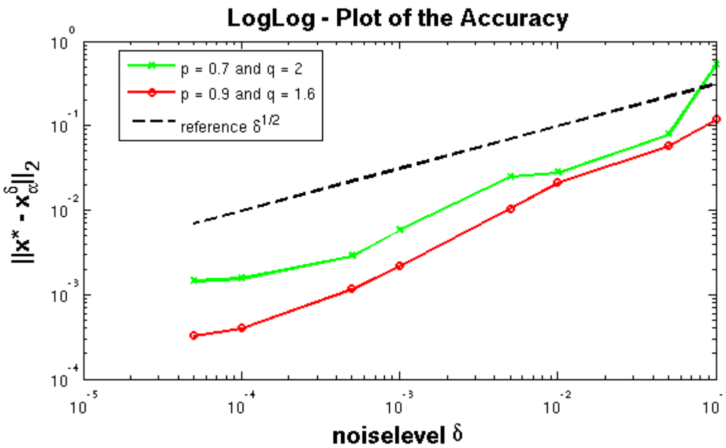

Figure 1. The plot shows the numerically observed rate of convergence for two different settings of p and q, compared to a reference line.

difficulty could be the numerical inaccuracy of the inner iterations (36), which might cause a stagnation of the iteration. In fact the observed loss of accuracy was very pronounced for the considered algorithm. Nevertheless this seems comprehensible when considering the transformation operator and the fact that all vector components are potentiated by a factor of q/p. This is also in accordance with our findings of an increased computational effort and the necessity of a sufficiently high numerical precision for decreasingly smaller values of p. These requirements were met by an efficient implementation with very accurate error criteria in all internal routines or iterations respectively. In particular high iteration numbers in the order of 10^5 and error tolerances in the order of 10^{-12} for the relative error of the inner iteration, i.e.

$$\|x_{n,l+1} - x_{n,l}\| = \mathcal{O}(10^{-12}) \,, \tag{59}$$

were used to ensure reliable and accurate results. Moreover we would like to point out that we expected a strictly decreasing objective value in every iteration step and observed this in all our numerical tests.

Example 1: We consider the reconstruction of an exact solution having 14 non zero Fourier coefficients from convolution data with approximately two percent relative Gaussian noise. In figure 2 the exact solution and the obtained reconstructions are plotted. Figure 3 shows the Fourier coefficients of the true solution and the reconstructed one in the case of $q = 2$ and $p = 0.4$. The number of non-zero coefficients of the reconstructed solution is clearly smaller than the one of the true solution, indicating the strong sparsity promoting effect ($p = 0.4$). Nevertheless the reconstructed solution provides a reasonable fit (see figure 2).

The central goal of the presented algorithm is its sparsity promoting property. Figure 4 shows the identification of sparse solutions for the considered deconvolution problem in Example 1. The threshold was chosen to be 10^{-4} and 10^{-5}, since a transformation with respect to $q = 2$, i.e. an ℓ_2-penalty term has no shrinkage and would lead to zero coefficients only in the limit. For a more detailed discussion of this issue see figure 7.

The decreasing number of non-zero coefficients indicates an increased promotion

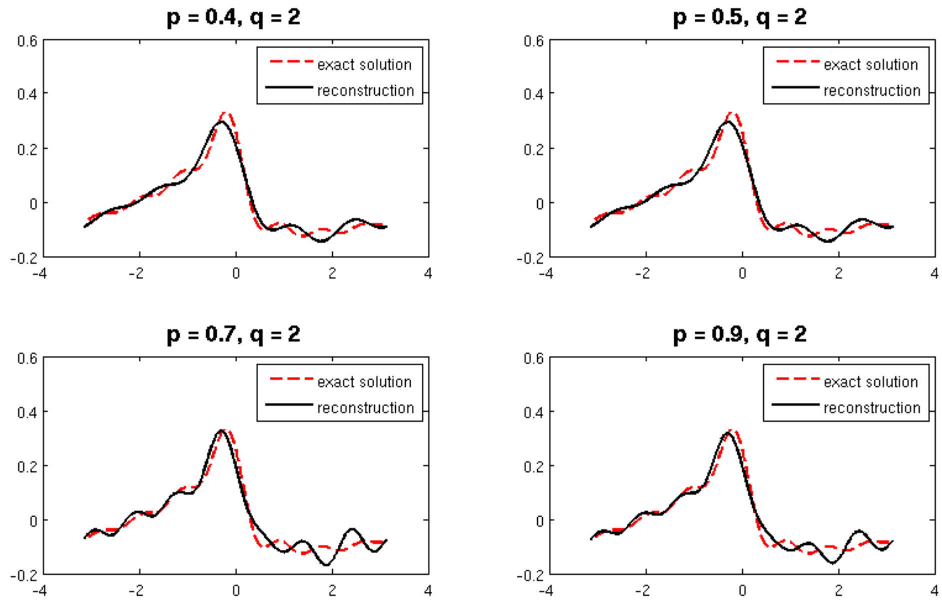

Figure 2. Example 1: The exact solution of the deconvolution problem and the reconstructions obtained for the four different values of p are plotted.

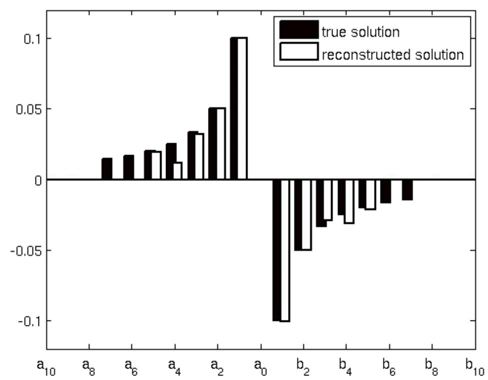

Figure 3. Non-zero Fourier coefficients of the true solution and the corresponding reconstructed solution for $q = 2$ and $p = 0.4$.

of sparse coefficient vectors for smaller values of p. This is especially worth mentioning since already for $p = 0.9$ the number of non-zero coefficients was underestimated. Further one may notice that although the number of reconstructed non zero elements is always smaller than for the exact solution the quality of the reconstruction is still very good. Moreover we want to emphasize that the reconstructed coefficients are among the true ones, i.e. the support of the true solution was identified or slightly underestimated.

Example 2: We consider the reconstruction of an exact solution with 22 non zero Fourier coefficients from convolution data with approximately seven percent relative Gaussian noise. Figure 5 shows the corresponding solution and reconstructions.

Figure 6 shows (analogously to figure 4) the number of non zero entries in the

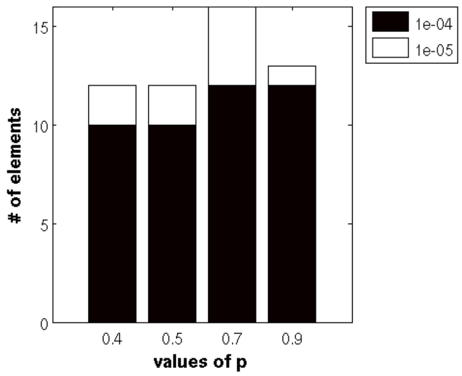

Figure 4. Example 1: The graphic shows the number of non zero elements (i.e. elements above the threshold of 10^{-4} / 10^{-5}) in the solution for the case of $q = 2$. The exact solution was designed to have 14 non zero entries.

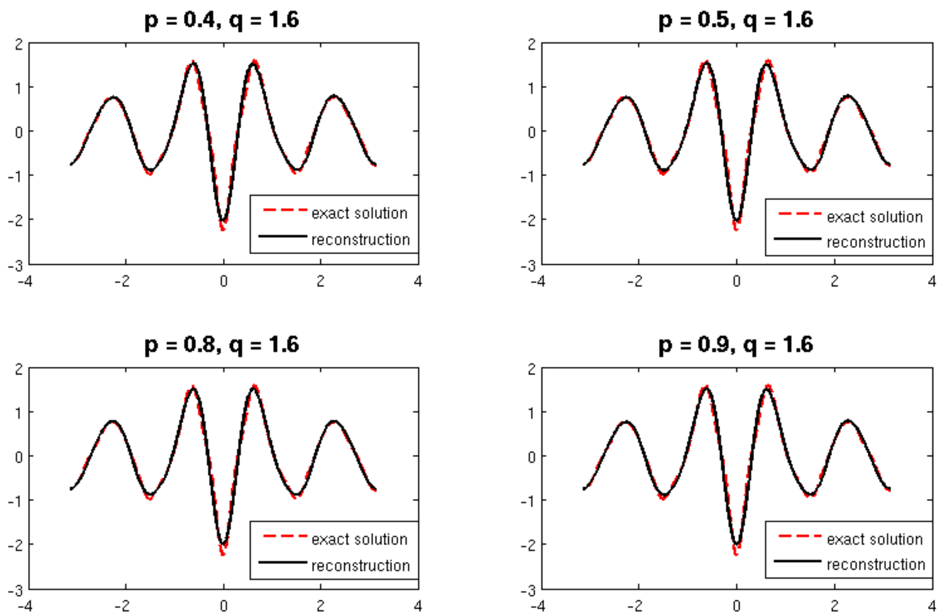

Figure 5. Example 2: The exact solution of the deconvolution problem and the reconstructions obtained for the four different values of p are plotted in the case of $q = 1.6$.

obtained reconstructions. Contrary to the first case were the exact solution had 14 non zero entries, the present setting leads to a reconstruction with significantly less non zero entries compared to the exact solution with 22 non zero entries. Nevertheless the quality of the reconstruction is again very good. However one may observe that for Example 2 also further decreased values of p do not lead to sparser solutions, since very sparse solutions have already been obtained with $p = 0.9$.

Finally we want to point out, that the choice of $q = 2$ (Example 1) directly affects the algorithm. In [13] it was shown that the solution to (38) is known in advance. Hence the computational effort is reduced significantly at the expense of numerical artifacts in the case of $q = 2$. Figure 7 shows the number of (non zero) entries above a certain threshold in the cases of $q = 1.1, 1.4$ and $q = 2$. As one expects from the theory the choice of $q \in (1, 2)$ seems to have no effect on the solution and no small

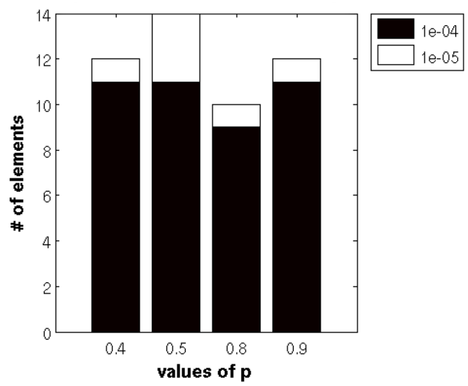

Figure 6. Example 2: The bar chart shows the number of non zero elements (above the threshold of 10^{-4} / 10^{-5}) in the solutions for the case of $q = 1.6$ and four different values of p. The exact solution was designed to have 22 non zero entries.

non zero entries occur. For $q = 2$ structurally the same solution is obtained, however numerous smaller entries occur due to numerical artefacts. Thus there are already 27 elements above the threshold of 10^{-7}, 169 above 10^{-9} and 3055 elements above the threshold of 10^{-11}, compared to 11 non zero elements for all thresholds in the case of $q = 1.1$ and $q = 1.4$.

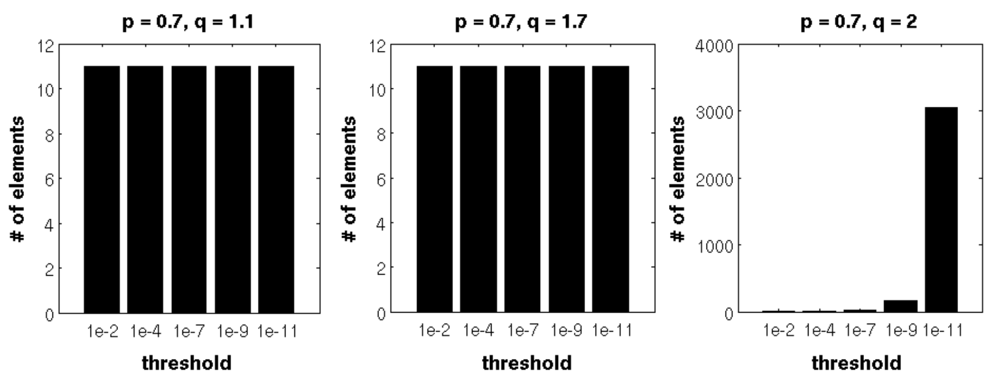

Figure 7. The bar charts show the number of (non zero) entries above the thresholds $10^{-2}, 10^{-4}, 10^{-7}, 10^{-9}$ and 10^{-11}, for $q = 1.1, 1.4, 2$ and $p = 0.7$.

Summarizing the results of example 1 and 2 we could confirm our analytical findings. As the reconstructed solutions were always close to the true solution, we might conclude that the algorithm reconstructed the global minimizer fitting the constructed data and thus providing high quality reconstructions. Moreover the strong sparsity promoting feature of the considered regularization functionals and the principle idea of exploiting the transformation operator in numerical algorithms could be confirmed.

Acknowledgments

R. Ramlau acknowledges support by the FWF grant P19496-N18 and P20237-N14. C. A. Zarzer thanks H. W. Engl for his support and supervision and acknowledges support by the WWTF Grant MA07-030.

References

[1] D. Donoho and P. Stark. Uncertainity principles and signal recovery. *SIAM J. Appl. Math.*, 49(3):906–931, 1989.

[2] D. Donoho and J. Tanner. Sparse nonnegative solutions of underdetermined linear equations by linear programming. *Proc. Nat. Acad. Sci.*, 102(27):9446–9451, 2005.

[3] E. Candes, J. Romberg, and T Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math*, 59(8):1207–1223, 2006.

[4] E. Candes and T. Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51(12):4203–4215, 2005.

[5] D. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.

[6] D. Donoho. High-dimensional centrally symmetric polytypes with neighborliness proportional to dimension. *Discrete Comput. Geom.*, 35(4):617–652, 2006.

[7] R. Gribonval and M. Nielsen. Highly sparse representations ffrom dictionaries are unique and independent of the sparseness measure. *Appl. Comput. Harmon. Anal.*, 22(3):335–355, 2007.

[8] E. Candes and T. Tao. Near optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inform. Theory*, 52(12):5406–5425, 2006.

[9] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k-term approximation. *J. Amer.Math. Soc.*, 22(1):211–231, 2009.

[10] I. Daubechies, R. Devore, M. Fornasier, and S. Güntürk. Iteratively re-weighted least squares minimization for sparse recovery. *Report*, 2009.

[11] P. Kügler, E. Gaubitzer, and S. Müller. Parameter identification for chemical reaction systems using sparstiy enforcing regularization: a case study fot the chlorite-iodide reaction. *to appear in J. Phys. Chem.*, 2009.

[12] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math.*, 51:1413–1541, 2004.

[13] R. Ramlau and G. Teschke. Tikhonov replacement functionals for iteratively solving nonlinear operator equations. *Inverse Problems*, 21(5):1571–1592, 2005.

[14] R. Ramlau and G. Teschke. A thresholding iteration for nonlinear operator equations with sparsity constraints. *Numer. Math.*, 104:177–203, 2006.

[15] R. Ramlau and G. Teschke. An iterative algorithm for nonlinear inverse problems with joint sparsity constraints in vector-valued regimes and an application to color image inpainting. *Inverse Problems*, 23(5):1851–1870, 2007.

[16] M. Fornasier, R. Ramlau, and G. Teschke. The application of joint sparsity and total variation minimization algorithms to a real-life art restoration problem. *To appear in Advances in Computational Mathematics.*, 2009.

[17] K. Bredies, D. Lorenz, and P. Maass. A generalized conditional gradient method and its connection to an iterative shrinkage method. *Comput. Optim. Appl.*, 42(2):173–193, 2008.

[18] Thomas Bonesky, Kristian Bredies, Dirk A. Lorenz, and Peter Maass. A generalized conditional gradient method for nonlinear operator equations with sparsity constraints. *Inverse Problems*, 23(5):2041–2058, 2007.

[19] R. Griesse and D. Lorenz. A semismooth newton method for tikhonov functionals with sparsity constraints. *Inverse Problems*, 24:035007, 2008.

[20] J. Bect, L. Blanc Feraud, G. Aubert, and A. Chambolle. A/1-unified variational framework for image restoration. pages Vol IV: 1–13, 2004.

[21] Patrick L. Combettes and Valérie R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.*, 4(4):1168–1200 (electronic), 2005.

[22] Kristian Bredies and Dirk A. Lorenz. Iterated hard shrinkage for minimization problems with sparsity constraints. *SIAM J. Sci. Comput.*, 30(2):657–683, 2008.

[23] R. Ramlau. Regularization properties of tikhonov regularization with sparsity constraints. *ETNA*, 30:54–74, 2008.

[24] Ronny Ramlau and Elena Resmerita. Convergence rates for regularization with sparsity constraints. Submitted.

[25] Markus Grasmair, Markus Haltmeier, and Otmar Scherzer. Sparse regularization with lq penalty term. *Inverse Problems*, 24(5):055020 (13pp), 2008.

[26] D. A. Lorenz. Convergence rates and source conditions for Tikhonov regularization with sparsity constraints. *J. Inverse Ill-Posed Probl.*, 16(5):463–478, 2008.

[27] Markus Grasmair. Well-posedness and convergence rates for sparse regularization with sublinear l^q penalty term. FSP 092: Joint Research Program of Industrial Geometry.

[28] Clemens A Zarzer. On Tikhonov regularization with non-convex sparsity constraints. *Inverse Problems*, 25:025006, 2009.

[29] Mila Nikolova. Markovian reconstruction in computed imaging and fourier synthesis. In *ICIP (2)*, pages 690–694, 1994.

[30] Radosław Pytlak. *Conjugate gradient algorithms in nonconvex optimization*, volume 89 of *Nonconvex Optimization and its Applications*. Springer-Verlag, Berlin, 2009.

[31] R. Ramlau. A steepest descent algorithm for the global minimization of the Tikhonov–functional. *Inverse Problems*, 18(2):381–405, 2002.

[32] R. Ramlau. TIGRA–an iterative algorithm for regularizing nonlinear ill–posed problems. *Inverse Problems*, 19(2):433–467, 2003.

[33] Heinz W. Engl and Gerhard Landl. Convergence rates for maximum entropy regularization. *SIAM J. Numer. Anal.*, 30(5):1509–1536, 1993.

[34] I. Cioranescu. *Geometry of Banach spaces, duality mappings and nonlinear problems.* Kluwer, Dordrecht, 1990.

[35] K. Bredies and D. Lorenz. Linear convergence of iterated soft-tresholding. *J. Fourier Anal. Appl.*, 14(5-6):813–837, 2008.

[36] Frank Bauer and Stefan Kindermann. Recent results on the quasi-optimality principle. Submitted.

[37] Frank Bauer and Stefan Kindermann. The quasi-optimality criterion for classical inverse problems. *Inverse Problems*, 24(3):035002, 20, 2008.

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