A note on the minimization of a Tikhonov functional with $\ell^1$-penalty

Fabian Hinterer$^{1,3}$, Simon Hubmer$^2$ and Ronny Ramlau$^{1,2}$

1 Johannes Kepler University Linz, Institute of Industrial Mathematics, Altenbergerstraße 69, A-4040 Linz, Austria
2 Johann Radon Institute Linz, Altenbergerstraße 69, A-4040 Linz, Austria

E-mail: fabian.hinterer@indmath.uni-linz.ac.at, simon.hubmer@ricam.oeaw.ac.at, ronny.ramlau@jku.at and ronny.ramlau@ricam.oeaw.ac.at

Received 9 January 2020, revised 14 March 2020
Accepted for publication 16 April 2020
Published 17 June 2020

Abstract
In this paper, we consider the minimization of a Tikhonov functional with an $\ell^1$ penalty for solving linear inverse problems with sparsity constraints. One of the many approaches used to solve this problem uses the Nemskii operator to transform the Tikhonov functional into one with an $\ell^2$ penalty term but a non-linear operator. The transformed problem can then be analyzed and minimized using standard methods. However, by the nature of this transform, the resulting functional is only once continuously differentiable, which prohibits the use of second order methods. Hence, in this paper, we propose a different transformation, which leads to a twice differentiable functional that can now be minimized using efficient second order methods like Newton’s method. We provide a convergence analysis of our proposed scheme, as well as a number of numerical results showing the usefulness of our proposed approach.

Keywords: inverse and ill-posed problems, Tikhonov regularization, sparsity, second-order methods, Newton’s method

(Some figures may appear in colour only in the online journal)
1. Introduction

In this paper, we consider linear operator equations of the form

\[ Ax = y, \]  

(1.1)

where \( A : \ell^2 \to \ell^2 \) is a bounded linear operator on the (infinite-dimensional) sequence space \( \ell^2 \). Note that by using a suitable basis or frame, operator equations between separable function spaces such as \( L^p \), Sobolev, or Besov spaces can all be transformed into problems of the form (1.1). We assume that only noisy data \( y^\delta \) satisfying

\[ \| y - y^\delta \|_2 \leq \delta \]  

(1.2)

are available, where \( \| \cdot \|_2 \) denotes the standard \( \ell^2 \)-norm. Problems of the form (1.1) arise in many practical applications including, but not limited to, image processing (compression, denoising, enhancement, inpainting, etc), image reconstruction, as well as medical and tomographic imaging. For example, in the case in tomography, where \( A \) is the Radon transform and \( x \) is the internal density to be reconstructed from sinogram data \( y^\delta \), the solution \( x \) can be expected to have a sparse representation in a given basis. Hence, we are particularly interested in sparse solutions of (1.1), to which end we consider the minimization of the following Tikhonov functional

\[ T_{\alpha,\delta}(x) := \| Ax - y^\delta \|_2^2 + \alpha \| x \|_1, \]  

(1.3)

where \( \| \cdot \|_1 \) denotes the standard \( \ell^1 \)-norm. This problem has already been thoroughly studied analytically (compare with section 2) as well as numerically (see section 3 for an overview of previously proposed methods). However, the efficient minimization of the Tikhonov functional \( T_{\alpha,\delta} \) still remains a field of active study, especially since the presence of the \( \ell^1 \)-norm makes the functional non-differentiable at the origin. One approach to circumvent this issue was proposed in [1], where the authors considered a transformation of the Tikhonov functional into one which is once differentiable. In this paper, we extend their transformation idea by using an approximate transformation approach in order to end up with a functional that is also twice differentiable. This then allows the application of efficient second-order iterative methods for carrying out the minimization.

This paper is organized as follows: in section 2, we review known regularization results concerning sparsity regularization via the Tikhonov functional (1.3) and in section 3, we discuss some of the existing methods for its minimization. In section 4, we consider the transformation approach presented in [2] and its extension for obtaining twice differentiable functionals, for which we provide a convergence analysis. Furthermore, in section 5, we present numerical simulations based on a tomography problem to demonstrate the usefulness of our approach. Finally, a conclusion is given in section 7.

2. Sparsity regularization

In this section, we recall some basic results [adapted from ([3], section 3.3)] concerning the regularization properties of Tikhonov regularization with sparsity constraints. For a more exten-
sive review on regularization theory for Tikhonov functionals with sparsity constraints the reader is referred to [4–6], and more recently, [7, 8].

First of all, concerning the well-definedness of minimizers of $T_{\alpha,\delta}$ and their stability with respect to the data $y^\delta$, we get the following result, which is an immediate consequence of ([3], theorem 3.48):

**Theorem 2.1.** Let $A : \ell^2 \to \ell^2$ be weakly sequentially continuous, $\alpha > 0$ and $y^\delta \in \ell^2$. Then there exists a minimizer of the functional $T_{\alpha,\delta}$ defined in (1.3). Furthermore, the minimization is weakly subsequentially stable with respect to the noisy data $y^\delta$.

Concerning the convergence of the minimizers of the Tikhonov functional, we get the following theorem, which follows directly from ([3], theorem 3.49):

**Theorem 2.2.** Let $A : \ell^2 \to \ell^2$ be weakly sequentially continuous, assume that the problem (1.1) has a solution in $\ell^1$, and let $\alpha(\delta) : (0, \infty) \to (0, \infty)$ be chosen such that

$$\alpha(\delta) \to 0, \quad \text{and} \quad \frac{\delta^2}{\alpha(\delta)} \to 0, \quad \text{as} \ \delta \to 0. \quad (2.1)$$

Moreover, assume that the sequence $\delta_k$ converges to 0, that $y_k := y^\delta_k$ satisfies the estimate $\|y - y_k\|_2 \leq \delta_k$, and that $x_k$ is a sequence of elements minimizing $T_{\alpha(\delta_k),\delta_k}$. Then there exists an $\ell^1$-minimum-norm solution $x^1$ and a subsequence $x_{k_n}$ of $x_k$ such that $\|x_k - x^1\|_2 \to 0$ as $n \to \infty$. Furthermore, if the $\ell^1$-minimum-norm solution $x^1$ is unique, then $\|x_k - x^1\|_2 \to 0$ as $k \to \infty$.

Note that typically, one only gets weak subsequential convergence of the minimizers of the Tikhonov functional to the minimum-norm solution. However, the above theorem shows that for sparsity regularization, one even gets strong subsequential convergence.

Furthermore, note that if $A$ is injective, the $\ell^1$-minimizing solution is sparse (i.e., only finitely many of its coefficients are non-zero) and satisfies a variational source condition, then it is possible to prove optimal convergence rates under the *a priori* parameter choice $\alpha(\delta) \sim \delta$, both in Bregman distance and in norm ([3], theorem 3.54).

3. Minimization of the Tikhonov functional

In this section, we review some of the previously proposed methods for the minimization of (1.3). Due to the non-differentiability of the $\ell^1$-norm in zero, this minimization problem is a non-trivial task.

Among the first and perhaps the most well-known method is the so-called *iterative shrinkage thresholding algorithm* (ISTA), proposed in [9]. Each iteration of this algorithm consists of a gradient-descent step applied to the residual functional, followed by a thresholding step, which leads to the iterative procedure

$$x_{k+1}^\delta = S_{\omega \alpha} (x_k^\delta - \omega A^* (A x_k^\delta - y^\delta)) \quad (3.1)$$

where $S_{\omega \alpha}$ denotes the component-wise thresholding (shrinkage) operator.
\( (S_{\alpha \omega}(x))_k := \text{sgn}(x_k) \max\{|x_k| - \alpha \omega, 0\} \).

It was shown that the iterates generated by ISTA converge to a minimizer of the Tikhonov functional (1.3) under suitable assumptions [9, 10]. Unfortunately, this convergence can be very slow, which motivated the introduction of fast ISTA (FISTA) in [11]. Based on Nesterov’s acceleration scheme [12], the iterates of FISTA are defined by

\[
\begin{align*}
    x_k^\delta &= S_{\alpha \omega} \left( z_k^{\delta - 1} - \omega A^* \left( A z_k^{\delta - 1} - y^\delta \right) \right), \\
    t_k &= \frac{1 + \sqrt{1 + 4 t_{k-1}^2}}{2}, \\
    z_k^{\delta} &= x_k^\delta + \left( \frac{t_{k-1} - 1}{t_k} \right) (x_k^\delta - x_{k-1}^\delta), \\
    z_0^\delta &= x_0, \\
    t_0 &= 1.
\end{align*}
\]

The convergence analysis presented in [11] as well as many numerical experiments show that the iterates of FISTA converge much faster than those of ISTA, the residual converging with a rate of \(O(1/k^2)\) for FISTA compared to \(O(1/k)\) for ISTA, hence making it more practical. This speedup also holds for a generalized version of FISTA, which is applicable to composite (convex) minimization problems [13]. Applied to problem (1.3), it has the same form as (3.2), but with the computation of \(z_k^\delta\) replaced by

\[
\begin{align*}
    z_k^\delta &= x_k^\delta + \frac{k - 1}{k + \beta - 1} (x_k^\delta - x_{k-1}^\delta),
\end{align*}
\]

where the choice of \(\beta = 3\) is common practice. The convergence of this method also for any other choice of \(\beta > 3\) was established in [13].

In the context of compressed sensing, where one tries to recover signals from incomplete and inaccurate measurements in a stable way, minimization problems of the form (1.3) have been analyzed and numerically treated in finite dimensions (see e.g. [14–16]). Also in finite dimensions, the minimization problem (1.3) has been tackled successfully by using various Krylov-subspace techniques (see e.g. [17–19]).

In infinite dimensions, a number of different minimization algorithms for (1.3) have been proposed. For example, the authors of [1, 20, 21] have proposed a surrogate functional approach, while the authors of [22–24] have proposed conditional gradient and semi-smooth Newton methods, respectively.

Of particular interest to us is the minimization approach presented in [2, 25], which we discuss in detail in section 4 below. It is based on a nonlinear transformation utilizing a Nemytskii operator, which turns the Tikhonov functional (1.3) into one with a standard \(\ell_2\)-norm penalty, but with a nonlinear operator. Since the resulting transformed functional is continuously Fréchet differentiable, one can use standard first-order iterative methods for its minimization. Unfortunately, the functional is not twice differentiable, which prohibits the use of second-order methods, known for their efficiency. Circumventing this shortcoming is the motivation for the minimization approach based on an approximate transformation presented below.

4. Transformation approach

The concept of approximating a nonsmooth operator with a convergent sequence of smooth operators has been used before, e.g., in [26] in the context of BV regularization. In the related setting where only an inexact forward operator is known, convergence of the resulting
approximate solutions as the uncertainty in the forward operator and the data decreases has been studied e.g., in [27]. As described above, the authors of [2, 25] considered a transformation approach for minimizing the Tikhonov functional (1.3). This approach is based on a nonlinear transformation of the functional using the Nemskii operator

\[ N_{pq} : (x_k)_{k \in \mathbb{N}} \mapsto \left( \eta_{pq}(x_k) \right)_{k \in \mathbb{N}}, \]  

(4.1)

where the function \( \eta_{pq} \) is defined by

\[ \eta_{pq} : \mathbb{R} \rightarrow \mathbb{R}, \quad \tau \mapsto \text{sgn}(\tau) |\tau|^q_p. \]  

(4.2)

The operator \( N_{pq} \) has for example been used in the context of maximum entropy regularization [28]. Since here we need it only for the special case \( p = 1 \) and \( q = 2 \), we now define the operator

\[ N : \ell^2 \rightarrow \ell^1, \quad x \mapsto N_{1,2}(x), \]  

(4.3)

and the function

\[ \eta : \mathbb{R} \rightarrow \mathbb{R}, \quad \tau \mapsto \eta_{1,2}(\tau). \]  

(4.4)

The operator \( N \) is continuous, bounded, bijective, and Fréchet differentiable with

\[ N'(x)h = \left( 2 \left| x_k \right| h_k \right)_{k \in \mathbb{N}^*}, \]  

(4.5)

and is used to define the following nonlinear operator

\[ F : \ell^2 \rightarrow \ell^2, \quad x \mapsto (A \circ N)(x). \]  

(4.6)

This is then used to transform the problem of minimizing (1.3) into a standard \( \ell^2 - \ell^2 \) minimization problem, as shown by the following result from [2]:

**Proposition 4.1.** The following two problems are equivalent:

(a) Find \( x^* \in \ell^1 \), such that \( x^* \) minimizes

\[ T_{\alpha,\delta}(x) = \|Ax - y^\delta\|_2^2 + \alpha \|x\|_1. \]  

(4.7)

(b) Find \( x^* = N(\tilde{x}) \), such that \( \tilde{x} \in \ell^2 \) minimizes

\[ J_{\alpha,\delta}(\tilde{x}) := \|F(x) - y^\delta\|_2^2 + \alpha \|x\|_2^2. \]  

(4.8)

Due to the above proposition, both the original and the transformed problem recover the same solution, which thus have the same sparsity properties. Note that the operator \( F \) is nonlinear even if \( A \) is linear. However, using the transformed operator has the advantage that the resulting functional \( J_{\alpha,\delta} \) is differentiable.

**Proposition 4.2.** The operator \( F \) and the functional \( J_{\alpha,\delta} \) defined in (4.6) and (4.8), respectively, are continuously Fréchet differentiable, with

\[ F'(x)h = AN'(x)h, \quad \text{and} \quad J'_{\alpha,\delta}(x)h = \langle 2F'(x)^*(F(x) - y^\delta) + 2\alpha x, h \rangle. \]
Proof. This is an immediate consequence of the definition of $J_{\alpha, \delta}$ and the fact that $A$ is linear and $N$ is differentiable. □

Due to the above result, it is now possible to apply gradient based (iterative) methods for minimizing the transformed functional $J_{\alpha, \delta}$, and thus to compute a minimizer of the functional $T_{\alpha, \delta}$, which itself is not differentiable.

Unfortunately, the transformed functional $J_{\alpha, \delta}$ is not twice differentiable, due to the fact that $N$ is not twice differentiable (at zero). This prohibits the use of second order methods like Newton’s method, which are known to be very efficient in terms of iteration numbers. Hence, we propose to approximate $N$ by a sequence of operators $N_\varepsilon$ which are twice continuously differentiable, and to minimize, instead of $J_{\alpha, \delta}$, the functional

$$J_{\alpha, \delta}^\varepsilon(x) := \|F_\varepsilon(x) - y\|^2_2 + \alpha\|x\|^2_2,$$

where we define the operator $F_\varepsilon$ by

$$F_\varepsilon : \ell^2 \to \ell^2, \quad x \mapsto (A \circ N_\varepsilon)(x),$$

for a suitable approximation $N_\varepsilon$ of the operator $N$. This approximation is based on suitable approximations $\eta_\varepsilon$ of the functions $\eta$, which we introduce in the following (figure 1).

**Definition 4.1.** For $\varepsilon > 0$ we define functions $\eta_\varepsilon : \mathbb{R} \to \mathbb{R}$ by

$$\eta_\varepsilon(\tau) := \begin{cases} -\frac{\tau^2}{3} - \frac{1}{3}\varepsilon^2, & \tau \in (-\infty, -\varepsilon), \\ \frac{1}{3}\varepsilon\tau^3 + \varepsilon\tau, & \tau \in [-\varepsilon, \varepsilon], \\ \tau^2 + \frac{1}{3}\varepsilon^2, & \tau \in (\varepsilon, \infty). \end{cases}$$

(4.11)

Obviously, $\eta_\varepsilon \to \eta$ as $\varepsilon \to 0$ and furthermore, we get the following

**Lemma 4.3.** The functions $\eta_\varepsilon$ defined by (4.11) are twice continuously differentiable.
Proof. It follows from its definition that $\eta_\varepsilon$ is everywhere continuous and that

$$
\eta'_\varepsilon(\tau) := \begin{cases} 
-2\tau, & \tau \in (-\infty, -\varepsilon), \\
\frac{1}{\varepsilon^2} + \varepsilon, & \tau \in [-\varepsilon, \varepsilon], \\
2\tau, & \tau \in (\varepsilon, \infty).
\end{cases}
$$

Again it follows that $\eta'_\varepsilon$ is everywhere continuous and that

$$
\eta''_\varepsilon(\tau) := \begin{cases} 
-2, & \tau \in (-\infty, -\varepsilon), \\
\frac{2}{\varepsilon}, & \tau \in [-\varepsilon, \varepsilon], \\
2, & \tau \in (\varepsilon, \infty),
\end{cases}
$$

which is again continuous everywhere, which concludes the proof. \qed

We now use the functions $\eta_\varepsilon$ to build the operators $N_\varepsilon$ via the following

**Definition 4.2.** For all $\varepsilon > 0$ we define the operators $N_\varepsilon : \ell^2 \to \ell^2$, $\eta_\varepsilon(x_k)_{k \in \mathbb{N}} \mapsto (\eta_\varepsilon(x_k))_{k \in \mathbb{N}}$. (4.12)

Concerning the well-defined and boundedness of $N_\varepsilon$, we have the following

**Lemma 4.4.** The operators $N_\varepsilon$ defined by (4.12) satisfy

$$
\|N_\varepsilon(x)\|_2 \leq \|x\|_2 \sqrt{\frac{16}{9} \varepsilon^2 + 2 \|x\|_2^2},
$$

and are therefore well-defined as operators from $\ell^2 \to \ell^2$.

**Proof.** Let $\varepsilon > 0$ be arbitrary but fixed and take $x = (x_k)_{k \in \mathbb{N}} \in \ell^2$. We have that

$$
|\eta_\varepsilon(x_k)| = \begin{cases} 
|x_k|^2 + \frac{1}{3} \varepsilon^2, & |x_k| > \varepsilon, \\
\frac{1}{3\varepsilon} |x_k|^3 + \varepsilon |x_k|, & |x_k| \leq \varepsilon,
\end{cases}
$$

which is again continuous everywhere, which concludes the proof.
from which we derive that
\[
\|N(x)\|_2^2 \leq \frac{16}{9} \varepsilon^2 \sum_{k=1}^{\infty} |x_k|^2 + 2 \sum_{k=1}^{\infty} |x_k|^4 \\
= \frac{16}{9} \varepsilon^2 \|x\|_2^2 + 2 \|x\|_4^4 \leq \left( \frac{16}{9} \varepsilon^2 + 2 \|x\|_2^2 \right) \|x\|_2^2 ,
\]
which immediately yields the assertion.

\[\square\]

The operators \( N_{\varepsilon} \) are also continuous, as we see in the following

**Proposition 4.5.** The operators \( N_{\varepsilon} : \ell^2 \to \ell^2 \) defined by (4.12) are continuous.

**Proof.** Let \( \varepsilon > 0 \) and \( x = (x_k)_{k \in \mathbb{N}} \in \ell^2 \) be arbitrary but fixed, and consider a sequence \( x^n = (x^n_k)_{k \in \mathbb{N}} \in \ell^2 \) converging to \( x \). It follows that the norm of \( x^n \) is uniformly bounded, i.e., there exists a constant \( c > 0 \) such that \( \|x^n\| \leq c \) for all \( n \), from which it also follows that \( |x^n_k| \leq c \) for all \( k \) and \( n \). Furthermore, since the function \( \eta_{\varepsilon} \) is continuously differentiable, it follows that it is Lipschitz continuous on bounded sets. This implies that there exists a Lipschitz constant \( L > 0 \) such that

\[
|\eta_{\varepsilon}(x^n_k) - \eta_{\varepsilon}(x_k)| \leq L |x^n_k - x_k| .
\]

Hence, we get that

\[
\|N_{\varepsilon}(x^n) - N_{\varepsilon}(x)\|_2^2 = \sum_{k=1}^{\infty} |\eta_{\varepsilon}(x^n_k) - \eta_{\varepsilon}(x_k)|^2 \leq L^2 \sum_{k=1}^{\infty} |x^n_k - x_k|^2 = L^2 \|x^n - x\|_2^2 ,
\]
and therefore,

\[
\|N_{\varepsilon}(x^n) - N_{\varepsilon}(x)\|_2 \leq L \|x^n - x\|_2 \to 0 \text{ as } n \to \infty ,
\]

which shows the continuity of \( N_{\varepsilon} \) and concludes the proof. \[\square\]

By their construction, the operators \( N_{\varepsilon} \) are also twice differentiable, as we see in

**Proposition 4.6.** The operators \( N_{\varepsilon} : \ell^2 \to \ell^2 \) defined by (4.12) are twice continuously Fréchet differentiable, with

\[
N_{\varepsilon}'(x)h = \left( \eta_{\varepsilon}'(x_k)h_k \right)_{k \in \mathbb{N}} , \text{ and } N_{\varepsilon}''(x)(h, w) = \left( \eta_{\varepsilon}''(x_k)h_k w_k \right)_{k \in \mathbb{N}} .
\]

**Proof.** This follows from the definition of \( N_{\varepsilon} \) together with lemma 4.3. \[\square\]

The approximation properties of the operators \( N_{\varepsilon} \) are studied in the following

**Proposition 4.7.** For \( N_{\varepsilon} \) be defined by (4.3) and (4.12), respectively, it holds that

\[
\|N(x) - N_{\varepsilon}(x)\|_2 \leq \frac{7}{3} \varepsilon \|x\|_2 .
\]
Proof. Let $\varepsilon > 0$ and $x \in \ell^2$ be arbitrary but fixed. Then it holds that

$$\eta_{\varepsilon}(x_k) - \eta(x_k) = \begin{cases} 
-\frac{1}{3}\varepsilon^2, & x_k \in (-\infty, -\varepsilon), \\
\frac{1}{3}\varepsilon^2 + \varepsilon x_k + x_k^2, & x_k \in [-\varepsilon, 0], \\
\frac{1}{3}\varepsilon^2 + \varepsilon x_k - x_k^2, & x_k \in [0, \varepsilon], \\
\frac{1}{3}\varepsilon^2, & x_k \in (\varepsilon, \infty),
\end{cases}$$

from which it follows that

$$|\eta_{\varepsilon}(x_k) - \eta(x_k)| = \begin{cases} 
\frac{1}{3}\varepsilon^2, & |x_k| > \varepsilon, \\
\frac{1}{3}\varepsilon|x_k|^3 + \varepsilon|x_k| - |x_k|^2, & |x_k| \leq \varepsilon.
\end{cases}$$

and therefore

$$|\eta_{\varepsilon}(x_k) - \eta(x_k)| \leq \frac{7}{3}\varepsilon|x_k|.$$ 

This now implies that

$$\|N_{\varepsilon}(x) - N(x)\|_2 = \sum_{k=1}^{\infty} |\eta_{\varepsilon}(x_k) - \eta(x_k)|^2 = \left(\frac{7}{3}\varepsilon\right)^2 \sum_{k=1}^{\infty} |x_k|^2 = \left(\frac{7}{3}\varepsilon\right)^2 \|x\|_2^2,$$

from which the statement immediately follows.

The above result immediately implies an approximation result for the operators $F_{\varepsilon}$.

**Corollary 4.8.** Let $A : \ell^2 \to \ell^2$ be a bounded and linear operator and let $F$ and $F_{\varepsilon}$ be defined by (4.6) and (4.10), respectively. Then it holds that

$$\|F(x) - F_{\varepsilon}(x)\|_2 \leq \frac{7}{3}\varepsilon \|A\| \|x\|_2.$$ 

**Proof.** By the definition of $F$ and $F_{\varepsilon}$, we have that

$$\|F(x) - F_{\varepsilon}(x)\|_2 = \|(A \circ N)(x) - (A \circ N_{\varepsilon})(x)\|_2 \leq \|A\| \|N(x) - N_{\varepsilon}(x)\|_2,$$

which, together with proposition 4.7 now yields the assertion.

Other important properties of the operators $F$ and $F_{\varepsilon}$ are collected in the following

**Proposition 4.9.** Let $A : \ell^2 \to \ell^2$ be a bounded linear operator. Then the operators $F$ and $F_{\varepsilon}$ defined by (4.6) and (4.10), respectively, are continuous and weakly sequentially closed.

**Proof.** Since $A$ and, due to proposition 4.5, $N_{\varepsilon}$ are continuous, by its definition also $F_{\varepsilon}$ is continuous. In order to show the weak sequential closedness of $F_{\varepsilon}$, note that since its definition
space is the whole of \( \ell^2 \), it suffices to show that \( F_\varepsilon \) is weakly continuous. For this, take an arbitrary sequence \( x^n \in \ell^2 \) converging weakly to some element \( x \in \ell^2 \). Since in \( \ell^2 \) a sequence converges weakly if and only if it converges componentwise and its norm is bounded [29], it follows from the continuity and boundedness of \( N_\varepsilon \) (lemma 4.4 and proposition 4.5) that \( N_\varepsilon (x^n) \) converges weakly to \( N_\varepsilon (x) \). Now, as a bounded linear operator, \( A \) is also weakly sequentially continuous. Hence, since \( F_\varepsilon = A \circ N_\varepsilon \), it follows that \( F_\varepsilon (x^n) \) converges weakly to \( F_\varepsilon (x) \), which establishes its weak sequential continuity and consequentially also its weak sequential closedness. For the operator \( F \), these result have already been shown in [2]. However, noting that lemma 4.4 and proposition 4.5 also hold for the limit case \( \varepsilon = 0 \), they also follow the same way as above.

Furthermore, the differentiability of \( N_\varepsilon \) immediately translates into the following

**Proposition 4.10.** The operators \( F_\varepsilon \) and thus the functionals \( J_{\alpha,\delta}^\varepsilon \) defined in (4.10) and (4.9), respectively, are twice continuously Fréchet differentiable, where

\[
F_\varepsilon'(x)h = AN_\varepsilon'(x)h, \quad F_\varepsilon''(x)(h, w) = AN_\varepsilon''(x)(h, w),
\]

\[
J_{\alpha,\delta}^\varepsilon'(x)h = 2 \langle F_\varepsilon'(x) + \alpha x, h \rangle,
\]

\[
J_{\alpha,\delta}^\varepsilon''(x)(h, w) = 2 \langle F_\varepsilon'(x) - y^\varepsilon, F_\varepsilon''(x)(h, w) \rangle + 2 \langle F_\varepsilon'(x)F_\varepsilon''(x)(w + \alpha w, h) \rangle.
\]

**Proof.** This follows from the definition of \( F_\varepsilon \) and \( J_{\alpha,\delta}^\varepsilon \) together with proposition 4.6. \( \square \)

We now consider the problem of minimizing the Tikhonov functional \( J_{\alpha,\delta}^\varepsilon \), whose minimizers we denote by \( x^{\varepsilon}_{\alpha,\delta} \). Due to the above results, the classical analysis of Tikhonov regularization for nonlinear operators is applicable (see for example [30, 31]), and we immediately get the following

**Theorem 4.11.** Let \( A : \ell^2 \to \ell^2 \) be a bounded, linear operator and let \( F_\varepsilon \) be defined by (4.10). Then for each \( \alpha > 0 \), a minimizer \( x^{\varepsilon}_{\alpha,\delta} \) of the functional \( J_{\alpha,\delta}^\varepsilon \) defined in (4.9) exists. Furthermore, the minimization of \( J_{\alpha,\delta}^\varepsilon \) is stable under perturbations of \( y^\varepsilon \).

**Proof.** Since by proposition 4.9, the operator \( F_\varepsilon \) is continuous and weakly sequentially closed, this follows immediately from ([30], theorem 10.2).

Next, we are interested in the behaviour of the minimizers \( x^{\varepsilon}_{\alpha,\delta} \) as \( \varepsilon \to 0 \). Given a suitable coupling of the noise level \( \delta \) and the parameter \( \varepsilon \), we get the following

**Theorem 4.12.** Assume that \( F(x) = y \) has a solution and let \( \alpha(\delta) \) and \( \varepsilon(\delta) \) satisfy

\[
\alpha(\delta) \to 0, \quad \varepsilon(\delta) \to 0, \quad \frac{\delta^2}{\alpha(\delta)} \to 0, \quad \frac{\varepsilon^2}{\alpha(\delta)} \to 0, \quad \text{as } \delta \to 0. \quad (4.20)
\]

Then \( x^{\varepsilon}_{\alpha(\delta),\delta(\delta)} \) has a convergent subsequence. Moreover, the limit of every convergent subsequence is a minimum-norm solution of \( F(x) = y \). Furthermore, if the minimum-norm solution \( x^1 \) is unique, then

\[
\lim_{\delta \to 0} x^{\varepsilon}_{\alpha(\delta),\delta(\delta)} = x^1. \quad (4.21)
\]

**Proof.** The proof of this theorem follows the same lines as the classical proof of convergence of Tikhonov regularization [30] and the proof for the case that the operator is approximated by a series of finite dimensional operators [32, 33] (in which case a slightly stronger condition
than what we can derive from proposition 4.7 was used). Hence, we here only indicate the main differences in the proof.

Note first due to proposition 4.7, it follows that

\[ \|F_\varepsilon(x) - F(x)\|_2 \leq \|A\|\|N_\varepsilon(x) - N(x)\|_2 \leq \frac{7}{3}\|A\|\|x\|_2. \quad (4.22) \]

This, together with \(x_{\alpha,\varepsilon}\) being a minimizer of \(J_{\alpha,\varepsilon}^\dagger\) implies that

\[ \|F_\varepsilon(x_{\alpha,\varepsilon}) - y\|_2^2 + |x_{\alpha,\varepsilon}\|_2^2 \leq \|F_\varepsilon(x^\dagger) - y\|_2^2 + |x^\dagger\|_2^2 \leq \left(\frac{7}{3}\|A\|\|x^\dagger\|_2 + \delta\right)^2 + |x^\dagger\|_2^2. \quad (4.23) \]

Together with (4.20), this implies the boundedness of \(x_{\alpha,\varepsilon}\) and

\[ \lim_{\delta \to 0} \|F_\varepsilon(x_{\alpha,\varepsilon}) - y\|_2 = 0. \]

Hence, since then there holds

\[ \|F(x_{\alpha,\varepsilon}) - y\|_2 \leq \|F_\varepsilon(x_{\alpha,\varepsilon}) - y\|_2 + \|F_\varepsilon(x_{\alpha,\varepsilon}) - F(x_{\alpha,\varepsilon})\|_2 + \|y - y\|_2 \leq \|F_\varepsilon(x_{\alpha,\varepsilon}) - y\|_2 + \delta + \frac{7}{3}\|A\|\|x_{\alpha,\varepsilon}\|_2 \varepsilon \to 0, \]

the weak sequential closedness of \(F\) implies the convergence of a subsequence of \(x_{\alpha,\varepsilon}\) to a solution of \(F(x) = y\). The remainder of the proof then follows analogously to the one of ([30], theorem 10.3) and is therefore omitted here.

The above result shows that minimizing \(J_{\alpha,\varepsilon}^\dagger\) instead of \(J_{\alpha,\varepsilon}\) to approximate the solution of \(F(x) = y\) makes sense if \(\varepsilon\) and the noise level \(\delta\) are suitably coupled, for example via \(\varepsilon \sim \delta\). Furthermore, the assumption that \(F(x) = y\) solvable, is for example satisfied if \(Ax = y\) has a solution belonging not only to \(l^2\) but also to \(l^1\), i.e., is sparse.

**Remark 4.1.** Following the line of the proofs of classical Tikhonov regularization results, it is also possible to derive convergence rate results under standard assumptions. Furthermore, the above analysis also holds for nonlinear operators \(A\) which are Lipschitz continuous, since then corollary 4.8 also holds.

### 5. Minimization methods for the Tikhonov functional

In the previous section, we established existence, stability, and convergence of the minimizers of \(J_{\alpha,\varepsilon}\) and \(J_{\alpha,\varepsilon}^\dagger\) under standard assumptions. However, there still remains the question of how to actually compute those minimizers in an efficient way.

One way to do this is to interpret the minimization of \(J_{\alpha,\varepsilon}\) and \(J_{\alpha,\varepsilon}^\dagger\) as Tikhonov regularization for the nonlinear operator equations \(F(x) = y\) and \(F_\varepsilon(x) = y\), respectively, and to use iterative regularization methods for their solution. Since both the operator \(F\) and \(F_\varepsilon\) are continuously Fréchet differentiable, iterative regularization methods like Landweber iteration [34], TIGRA [35], the Levenberg–Marquardt method [36, 37] or iteratively regularized Gauss–Newton [38, 39] are applicable. Of course, as all of those methods only require a
once differentiable operator, it makes sense in terms of accuracy to apply them for the operator $F$ and not for the approximated operator $F_\varepsilon$. Another way is to use standard iterative optimization methods for the (well-posed) problem of minimizing $J_{\alpha,\delta}$ or $J_{\alpha,\delta}^{\varepsilon}$. In particular, since we have derived in the previous section that $J_{\alpha,\delta}^{\varepsilon}$ is twice continuously Fréchet differentiable, efficient second order methods like Newton’s method are applicable for its minimization.

In this section, we introduce and discuss some details of the minimization methods used to obtain the numerical results presented in section 6 below.

5.1. Gradient descent, ISTA and FISTA

We have seen that the Tikhonov functional $J_{\alpha,\delta}$ defined in (4.8) is continuously Fréchet differentiable. Hence, it is possible to apply gradient descent for its minimization.

For this, note first that since $N'(x)h$ is a linear operator, it can be written as

$$N'(x)h = G(x)h,$$

(5.1)

where $G(x)$ is the infinite dimensional ‘matrix’ representation of $N'(x)$ given by

$$G(x) := \text{diag}(2|x_k|)_{k\in\mathbb{N}},$$

which is called the gradient of $N$. Similarly, there is an (infinite-dimensional) matrix representation of $J_{\alpha,\delta}'(x)$, i.e., the gradient $\nabla J_{\alpha,\delta}(x)$ of $J_{\alpha,\delta}(x)$, which is given by

$$\nabla J_{\alpha,\delta}(x) := 2G(x)A^T(AN(x) - y) + 2\alpha x,$$

where, with a small abuse of notation, $A$ denotes the (infinite-dimensional) matrix representation of the linear operator $A$, and $A^T$ denotes its transpose.

Using the above representations, we can now write the gradient descent algorithm for minimizing $J_{\alpha,\delta}$ in the well-known form

$$x_{\delta}^{n+1} = x_{\delta}^{n} - \omega_n \nabla J_{\alpha,\delta}(x_{\delta}^{n}),$$

(5.2)

where $\omega_n$ is a sequence of stepsizes. If the stepsizes are chosen in a suitable way, for example via the Armijo rule [40], the iterates converge to a stationary point of $J_{\alpha,\delta}$ [see e.g. ([40], theorem 2.2)]. In order to stop the iteration, we employ the well-known discrepancy principle, i.e., the iteration is terminated with index $n^* = n^*(\delta, y^\delta)$, when for the first time

$$\|F(x_{n^*}^{\delta}) - y^\delta\|_2 \leq \tau \delta,$$

(5.3)

where $\tau > 1$ is fixed. Note that since the Tikhonov functional may have several (local and global) minima, convergence to a global minimum is only guaranteed if a sufficiently good initial guess is chosen.

The (infinite-dimensional) matrix representations introduced above can also be used to rewrite ISTA (3.1) into the following form.
\[ x_{n+1}^{\delta} = S_{\alpha n} (x_n^{\delta} - \omega 2 G(x_n^{\delta}) A^T (AN(x_n^{\delta}) - y)) , \]

which immediately also translates to a similar rewriting of FISTA defined in (3.2).

5.2. The Levenberg–Marquardt method

It is well-known that gradient based methods like gradient descent or ISTA are quite slow with respect to convergence speed. Although it is possible to speed them up by using suitable stepsizes (see for example [41, 42]) or acceleration schemes like FISTA, it is often advantageous to use second-order methods instead. One such method is the Levenberg–Marquardt method [36, 37], which is given by

\[ x_{n+1}^{\delta} = x_n^{\delta} + (F'(x_n^{\delta})^* F'(x_n^{\delta}) + \alpha_n I)^{-1} F'(x_n^{\delta})^* (y - F(x_n^{\delta})) . \] (5.4)

Although this is a second-order method, it only requires the operator \( F \) to be once continuously Fréchet differentiable. Using again the (infinite-dimensional) matrix representation of \( N(x) h \) from (5.1), the method can be rewritten into the following form

\[ x_{n+1}^{\delta} = x_n^{\delta} + (G(x_n^{\delta}) A^T A G(x_n^{\delta}) + \alpha_n I)^{-1} G(x_n^{\delta}) A^T (y - F(x_n^{\delta})) . \]

In order to obtain convergence of this method, one needs, among other things, a suitably chosen sequence \( \alpha_n \) converging to 0 as well as a sufficiently good initial guess [36]. As a stopping rule, one usually also employs the discrepancy principle (5.3).

The Levenberg–Marquardt method typically requires only very few iterations to satisfy the discrepancy principle. However, in each iteration step the linear operator \( (F'(x_n^{\delta})^* F'(x_n^{\delta}) + \alpha_n I) \) has to be inverted, which might be costly for some applications. This can be circumvented, though, via approximating the result of this inversion by the application of number of iterations of the conjugate gradient method.

It is possible to add an additional regularization term to the Levenberg–Marquardt method, thereby ending up with the so-called iteratively-regularized Gauss–Newton method [38, 39]. Typically behaving very similar in practice, this method can be proven to converge under slightly weaker assumptions than the Levenberg–Marquardt method.

5.3. Newton’s method

In contrast to \( J_{\alpha, \delta} \), the functional \( J_{\alpha, \delta}^\varepsilon \) is twice continuously Fréchet differentiable. The information contained in this second derivative can be used to design efficient methods for its minimization. One such method, based on Newton’s method, is considered here.

Note that the first-order optimality condition for minimizing \( J_{\alpha, \delta}^\varepsilon \) is given by

\[ J_{\alpha, \delta}^\varepsilon (\varepsilon_n) h = 0 \quad \forall \ h \in \ell^2 . \] (5.5)

Using Taylor approximation in the above equation yields
\[ J_{\alpha,\delta}'(x + \tau)(h) = J_{\alpha,\delta}'(x)h + J_{\alpha,\delta}''(x)(\tau, h) \quad \forall \ h \in \ell^2, \]

which, for the special choice of \( x = x_n \) and \( \tau = (x_{n+1} - x_n) \), becomes

\[ J_{\alpha,\delta}'(x_n)(h) + J_{\alpha,\delta}''(x_n)(x_{n+1} - x_n, h) = 0 \quad \forall \ h \in \ell^2. \tag{5.6} \]

This implicitly defines an iterative procedure, which is nothing else than Newton’s method applied to the optimality condition (5.5). Since \( J_{\alpha,\delta}''(\cdot, h) \) is continuously invertible around the global minimizer, this method is (locally) well-defined and \( q \)-superlinearly convergent [see for example ([40], corollary 2.1)].

We can again use an (infinite-dimensional) matrix representation to rewrite this iterative procedure into a more familiar form. For this, we first define the ‘matrices’

\[ G_{\varepsilon}(x) := \text{diag}(\eta_\varepsilon'(x_k)h_k)_{k \in \mathbb{N}}, \quad H_{\varepsilon}(x, w) := \text{diag}(\eta_\varepsilon''(x_k)w_k)_{k \in \mathbb{N}}, \tag{5.7} \]

which correspond to the gradient and the Hessian matrix of \( N_{\varepsilon}(x) \), and use this to write

\[ N_{\varepsilon}'(x)h = G_{\varepsilon}(x)h, \quad N_{\varepsilon}''(x)(w, h) = H_{\varepsilon}(x, w)h. \tag{5.8} \]

This allows the following matrix representation of the functionals \( J_{\alpha,\delta}'(x) \) and \( J_{\alpha,\delta}''(x) \)

\[ \nabla J_{\alpha,\delta}'(x) := 2G_{\varepsilon}(x)A^T(AN_{\varepsilon}(x) - y^\delta) + 2\alpha x, \]

\[ \nabla^2 J_{\alpha,\delta}'(x) := 2H_{\varepsilon}(x, A^T(AN_{\varepsilon}(x) - y^\delta)) + 2G_{\varepsilon}(x)A^TAG_{\varepsilon}(x) + 2\alpha I, \]

where \( I \) denotes the identity matrix, and \( \nabla J_{\alpha,\delta}'(x) \) and \( \nabla^2 J_{\alpha,\delta}'(x) \) can be seen as the gradient and the Hessian matrix of the functional \( J_{\alpha,\delta}' \), respectively. Using these representations, the iterative procedure (5.6) can be rewritten into the more familiar form

\[ \nabla J_{\alpha,\delta}'(x_n) + \nabla^2 J_{\alpha,\delta}'(x_n)(x_{n+1} - x_n) = 0. \]

which is an infinite-dimensional matrix-vector system for the update \( (x_{n+1} - x_n) \).

6. Numerical examples

In this section, we demonstrate the usefulness of our proposed approximation approach on a numerical example problem based on computerized tomography (CT). In particular, we focus on how the Newton approach for the minimization of \( J_{\alpha,\delta}' \) introduced in section 5.3 above performs in comparison to the other methods presented in section 3.

In the medical imaging problem of CT, one aims to reconstruct the density function \( f \) inside an object from measurements of the intensity loss of an x-ray beam sent through it. In the 2D case, for example if one scans a cross-section of the human body, the relationship between the intensity \( I_0 \) of the beam at the emitter position and the intensity \( I_L \) at the detector position is given by [43]
\[ \log I_l(s,w) - \log I_0(s,w) = - \int_R f(sw + tw^\perp) \, dt. \] (6.1)

Thus, if one defines the well-known Radon transform operator
\[ Rf(s,w) := \int_R f(sw + tw^\perp) \, dt, \]
the reconstruction problem (6.1) can be written in the standard form
\[ Rf = g. \]

Expressing \( f \) in terms of some basis or frame, and noting that typically one considers objects whose density is equal to 0 on large subparts, the above problem precisely fits into the framework of \( \ell^1 \) sparsity regularization considered in this paper.

6.1. Discretization and implementation

In order to obtain a discretized version of problem (6.1), we make use of the toolbox AIR TOOLS II by Hansen and Jorgensen [44]. Therein, the density function \( f \) is considered as a piecewise constant function on an \( m \times m \) pixel grid (see figure 4 for examples). With this, equation (6.1) can be written in the discretized form
\[ y_i : = \left( \log I_{L_i}^0 - \log I_{0_i}^0 \right) = \sum_{j=1}^{m^2} a_{ij} x_j \] (6.2)

where the \( x_j \) denote the value of \( f \) at the \( j \)th pixel, \( I_{L_i}^0 \) and \( I_{0_i}^0 \) denote the emitted and detected intensity of the \( i \)th ray, respectively, and \( a_{ij} \) denotes the length of the path which it travels through within the \( j \)th pixel cell. Note that since any given ray only travels through relatively few cells, most of the coefficients \( a_{ij} \) are equal to 0 and thus the matrix \( A \) is sparse. Collecting the coefficients \( a_{ij} \) into a matrix \( A \), equation (6.2) can be written as a matrix-vector equation of the form
\[ Ax = y. \]

Specifying all required parameters as well as the exact solution which one wants to reconstruct, the toolbox provides both the matrix \( A \) and the right-hand side vector \( y \). For our purposes, we used the toolbox function \texttt{parallelctomo}, creating a parallel beam tomography problem with (the suggested default values of) 180 angles and 70 parallel beams for each of them. For the number of pixels we used \( m^2 = 50^2 \), which altogether leads to the dimension 12,600 \times 2500 for the matrix \( A \). The exact solution (the Shepp–Logan phantom) is depicted in figure 4. In order to obtain noisy data, we used \( y' : = y + \delta \| y \|_2 r \), where \( r \) is a randomly generated, normed vector, and \( \delta \) denotes the relative noise level.

The implementation of the methods introduced in section 3 was done in a straightforward way by using their infinite-dimensional matrix representations but for the now finite dimensional matrices. The iterations were stopped using the discrepancy principle (5.3) with the choice \( \tau = 1.1 \) for all methods. For the approximation parameter \( \varepsilon \) in the definition of \( J_{\varepsilon,\alpha} \), we have used the choice \( \varepsilon = 10^{-4} \delta \), which is conforming with the theory developed above. The stepsize \( \omega \) in ISTA and FISTA was chosen as a constant based on the norm of \( A \), and for the gradient descent method (5.2), the stepsizes \( \omega_n \) were chosen via the Armijo rule.
In the Levenberg–Marquardt method (5.4), we chose $\alpha_n = 0.6^n \delta$, which is a sequence tending to 0 in accordance with the convergence theory. All computations were carried out in Matlab on a desktop computer with an Intel Xeon E5-1650 processor with 3.20 GHz and 16 GB RAM.

6.2. Numerical results

In this section, we present the results of applying the iterative methods introduced in section 3 to the tomography problem described above.

In the following, we present reconstruction results for different noise levels $\bar{\delta}$, which is directly related to the signal-to-noise ratio (SNR) by

$$\bar{\delta} = \frac{\|y - y^4\|}{\|y\|} \approx \frac{\|y - y^\delta\|}{\|y^\delta\|} = \text{SNR}^{-1}.$$
The first results, which are related to the computational efficiency of the different methods, are presented in figure 2. One can clearly see that regardless of the noise level $\delta$, the Newton method and the Levenberg–Marquardt method outperform the gradient based methods, both in terms of computation time and number of iterations $n^*$ required to meet the discrepancy principle. Furthermore, as was to be expected, FISTA also performs much better than both ISTA and the gradient descent method. Note also that with the Levenberg–Marquardt and the Newton method, one can satisfy the discrepancy principle also for very small noise levels, which becomes infeasible for the other methods due to the too large runtime which would be required for that.

The results depicted in figure 3 show that not only do the Levenberg–Marquardt and the Newton method require less iterations and computation time to satisfy the discrepancy principle, the resulting approximations also have a comparable and even somewhat smaller relative error than for the gradient based methods. This is of course partly due to the fact that each iteration step of those methods is ‘larger’ than in the other methods, which nevertheless turns out to be an advantage in our case. The resulting approximate solutions for 10% relative noise are shown in figure 4. The higher quality of the solutions obtained by the Levenberg–Marquardt and the Newton method is apparent.

### 7. Conclusion

In this paper, we presented a minimization approach for a Tikhonov functional with $\ell^1$ penalty for the solution of linear inverse problems with sparsity constraints. The employed approximate transformation approach based on a Nemskii operator was mathematically analyzed within the framework of ill-posed problems, and the fact that the resulting transformed functional is twice continuously Fréchet differentiable served as a basis for the construction of an effective minimization algorithm using Newton’s method. Numerical example problems based
on the medical imaging problem of computerized tomography demonstrated the usefulness of the proposed approach.

Acknowledgments

The authors were funded by the Austrian Science Fund (FWF): F6805-N36.

ORCID iDs

Fabian Hinterer https://orcid.org/0000-0002-5862-6327
Simon Hubmer https://orcid.org/0000-0002-8494-5188
Ronny Ramlau https://orcid.org/0000-0002-0277-9615

References

[1] Ramlau R and Teschke G 2006 A Tikhonov-based projection iteration for nonlinear ill-posed problems with sparsity constraints Numer. Math. 104 177–203
[2] Ramlau R and Zarzer C 2012 On the minimization of a Tikhonov functional with a non-convex sparsity constraint Electron. Trans. Numer. Anal. 39 467–505
[3] Scherzer O, Grasmair M, Grossauer H, Haltmeier M and Lenzen F 2009 Variational Methods in Imaging 167 (Applied Mathematical Sciences) (New York, NY: Springer)
[4] Jin B and Maass P 2012 Sparsity regularization for parameter identification problems Inverse Problems 28 123001
[5] Ramlau R and Resmerita E 2010 Convergence rates for regularization with sparsity constraints Electron. Trans. Numer. Anal. 37 87–104
[6] Resmerita E 2005 Regularization of ill-posed problems in Banach spaces: convergence rates Inverse Problems 21 08
[7] Jin B, Maass P and Scherzer O 2017 Sparsity regularization in inverse problems Inverse Problems 33 060301
[8] Weidling F, Sprung B and Hohage T 2020 Optimal convergence rates for Tikhonov regularization in Besov spaces SIAM J. Numer. Anal. 58 21–47
[9] Daubechies I, Defrise M and De Mol C 2004 An iterative thresholding algorithm for linear inverse problems with a sparsity constraint Commun. Pure Appl. Math. 57 1413–57
[10] Bredies K and Lorenz D 2008 Linear convergence of iterative soft-thresholding J. Fourier Anal. Appl. 14 813–37
[11] Beck A and Teboulle M 2009 A fast iterative shrinkage-thresholding algorithm for linear inverse problems SIAM J. Imaging Sci. 2 183–202
[12] Nesterov Y 1983 A method of solving a convex programming problem with convergence rate $O(1/k^2)$ Dokl. Akad. Nauk SSSR 269 543–47
[13] Attouch H and Peyroutouq J 2016 The rate of convergence of Nesterov’s accelerated forward–backward method is actually faster than $O(1/k^2)$ SIAM J. Optim. 26 1824–34
[14] Candès E, Romberg J and Tao T 2006 Stable signal recovery from incomplete and inaccurate measurements Commun. Pure Appl. Math. 59 1207–23
[15] Daubechies I, DeVore R, Fornasier M and Güntürk C 2010 Iteratively reweighted least squares minimization for sparse recovery Commun. Pure Appl. Math. 63 1–38
[16] Donoho D and Tanner J 2005 Sparse nonnegative solution of underdetermined linear equations by linear programming Proc. Natl Acad. Sci. USA 102 9446–51
[17] Bucci A and Reichel L 2019 An $l^1 – p$ regularization method for large discrete ill-posed problems J. Sci. Comput. 78 1526–49
[18] Huang G, Lanza A, Morigi S, Reichel L and Sgallari F 2017 Majorization–minimization generalized Krylov subspace methods for $l^0 – l^q$ optimization applied to image restoration BIT Numer. Math. 57 351–78
[19] Lanza A, Morigi S, Reichel L and Sgallari F 2015 A generalized Krylov subspace method for $l^0 – l^q$ minimization SIAM J. Sci. Comput. 37 30–50
[20] Ramlau R and Teschke G 2005 Tikhonov replacement functionals for iteratively solving nonlinear operator equations Inverse Problems 21 1571–92
[21] Ramlau R and Teschke G 2010 Sparse recovery in inverse problems Theoretical Foundations and Numerical Methods for Sparse Recovery (Radon Series on Computational and Applied Mathematics) ed M Fornasier (Berlin: De Gruyter) vol 9 pp 201–62
[22] Bonesky T, Bredies K, Lorenz D A and Maass P 2007 A generalized conditional gradient method for nonlinear operator equations with sparsity constraints Inverse Problems 23 2041–58
[23] Bredies K, Lorenz D A and Maass P 2009 A generalized conditional gradient method and its connection to an iterative shrinkage method Comput. Optim. Appl. 42 173–93
[24] Griese R and Lorenz D A 2008 A semismooth Newton method for Tikhonov functionals with sparsity constraints Inverse Problems 24 035007
[25] Zarzer C 2009 On Tikhonov regularization with non-convex sparsity constraints Inverse Problems 25 025006
[26] Acar R and Vogel C 1994 Analysis of bounded variation penalty methods for ill-posed problems Inverse Problems 10 1217–29
[27] Korolev Y and Letlman J 2018 Image reconstruction with imperfect forward models and applications in deblurring SIAM J. Imaging Sci. 11 197–218
[28] Engl H W and Landl G 1993 Convergence rates for maximum entropy regularization SIAM J. Numer. Anal. 30 1509–36
[29] Conway J B 1994 A Course in Functional Analysis (Graduate Texts in Mathematics) (New York, NY: Springer)
[30] Engl H W, Hanke M and Neubauer A 1996 Regularization of Inverse Problems (Dordrecht: Kluwer)
[31] Engl H W and Ramlau R 2015 Regularization of Inverse Problems (Encyclopedia of Applied and Computational Mathematics) ed B Engquist (Berlin: Springer) DOI: 10.1007/978-3-540-70529-1_52
[32] Neubauer A 1989 Tikhonov regularization for nonlinear ill-posed problems: optimal convergence and finite-dimensional approximation Inverse Probl. 5 541–57
[33] Pöschl C, Resmerita E and Scherzer O 2010 Discretization of variational regularization in Banach spaces Inverse Problems 26 105017
[34] Kaltenbacher B, Neubauer A and Scherzer O 2008 Iterative Regularization Methods for Nonlinear Ill-Posed Problems (Berlin: de Gruyter) DOI: 10.1515/9783110208276
[35] Ramlau R 2003 TIGRA—an iterative algorithm for regularizing nonlinear ill-posed problems Inverse Problems 19 433
[36] Hanke M 1997 A regularizing Levenberg–Marquardt scheme, with applications to inverse groundwater filtration problems Inverse Problems 13 79
[37] Jin Q 2010 On a regularized Levenberg–Marquardt method for solving nonlinear inverse problems Numer. Math. 115 229–59
[38] Blaschke B, Neubauer A and Scherzer O 1997 On convergence rates for the iteratively regularized Gauss–Newton method IMA J. Numer. Anal. 17 421
[39] Jin Q and Tautenhahn U 2009 On the discrepancy principle for some Newton type methods for solving nonlinear inverse problems Numer. Math. 111 509–58
[40] Hinze M, Pinnau R, Ulbrich M and Ulbrich S 2009 Optimization with PDE Constraints vol 23 p xii+270 (New York: Springer)
[41] Neubauer A 2017 A new gradient method for ill-posed problems Numer. Funct. Anal. Optim. 0 1–26
[42] Saxenhuber D 2016 Gradient-based reconstruction algorithms for atmospheric tomography in adaptive optics systems for extremely large telescopes PhD Thesis Johannes Kepler University Linz
[43] Natterer F 2001 The Mathematics of Computerized Tomography (Philadelphia, PA: Society for Industrial and Applied Mathematics) DOI: 10.1137/1.9780898719284
[44] Hansen P C and Jorgensen J 2017 Air tools: ii: algebraic iterative reconstruction methods, improved implementation Numer. Algorithms 79 11