Global convergence of damped semismooth Newton methods for $\ell_1$ Tikhonov regularization

Esther Hans$^1$ and Thorsten Raasch$^1$

$^1$ Institut für Mathematik, Johannes Gutenberg-Universität Mainz, 55099 Mainz, Germany
E-mail: hanse@uni-mainz.de, raasch@uni-mainz.de

Abstract. We are concerned with Tikhonov regularization of linear ill-posed problems with $\ell_1$ coefficient penalties. In [Inverse Probl. 24 (2008) 035007], Griesse and Lorenz proposed a semismooth Newton method for the efficient minimization of the corresponding Tikhonov functionals. While the convergence of semismooth Newton methods is locally superlinear in general, their application to $\ell_1$ Tikhonov regularization is particularly attractive because here, one obtains the exact Tikhonov minimizer after a finite number of iterations, given a sufficiently good initial guess. In this work, we discuss the efficient globalization of B(ouligand)-semismooth Newton methods for $\ell_1$ Tikhonov regularization by means of damping strategies and suitable descent with respect to an associated merit functional. Numerical examples are provided which show that our method compares well with existing iterative, globally convergent approaches.

1. Introduction

We are concerned with the efficient numerical solution of the minimization problem

$$\min_{u \in \ell_2} \frac{1}{2}\|Ku - f\|^2_Y + \sum_{k \in \mathcal{N}} w_k |u_k|,$$  \hspace{1cm} (1)

where $K : \ell_2 \to Y$ is a linear, bounded and injective operator mapping the sequence space $\ell_2 = \ell_2(\mathcal{N})$, $\mathcal{N} \in \{\mathbb{N}, \{1, \ldots, n\}\}$, into a separable Hilbert space $Y$, $f \in Y$ is a given datum and $w = (w_k)_{k \in \mathcal{N}}$ is a weight sequence with $w_k \geq w_0 > 0$ for all $k \in \mathcal{N}$. By the injectivity of $K$ and the coercivity of the penalty term $\sum_{k \in \mathcal{N}} w_k |u_k|$ with respect to $\| \cdot \|_{\ell_2}$, (1) has a unique minimizer $\hat{u} \in \ell_2$ which, moreover, has only a finite number of nonzero entries [16].

Such minimization problems typically arise from sparsity-constrained Tikhonov regularization of the discretization of linear ill-posed operator equations $Au = f$. Here $A : X \to Y$ is a bounded linear mapping from a separable Hilbert space $X$ into $Y$, and $f \in Y$ are given measurement data. For the stable numerical solution of such operator equations in the presence of measurement errors, regularization strategies are
Global convergence of damped semismooth Newton methods for $\ell_1$ regularization

A common approach is to regularize by minimizing a Tikhonov functional,

$$
\min_{u \in X} \frac{1}{2} \|Au - f\|_Y^2 + R(u),
$$

where the penalty term $R : X \to \mathbb{R}$ models certain a priori regularity assumptions on the unknown solution. Recently, it has become clear that sparse expansibility of the quantity of interest with respect to a given Riesz basis $\Psi = \{\psi_k\}_{k \in \mathbb{N}}$ of $X$ may serve to stabilize the recovery problem [10]. Denoting by $T : \ell_2 \to X$ the synthesis operator $T u = \sum_{k \in \mathbb{N}} u_k \psi_k$ associated with $\Psi$, the sparsity of $u \in X$ with respect to $\Psi$ can be modeled by $\ell_1$ coefficient penalties

$$
R(u) = \sum_{k \in \mathbb{N}} w_k |u_k|, \quad u = Tu,
$$

which readily induces the minimization problem (1) by setting $K := AT$.

For the numerical solution of (1), several classes of algorithms have been discussed in the literature, see also [22] for an extensive survey and numerical comparisons. Direct homotopy-type solvers like the LARS-LASSO algorithm [12, 24] follow the piecewise linear, continuous trajectory of minimizers $\hat{u} = \hat{u}(\alpha)$, $w_k = \alpha \hat{w}_k$ for all $k \in \mathbb{N}$, as $\alpha$ decreases to 0. They compute the exact Tikhonov minimizer within finite many computational operations. However, their performance is poor as the number of active coefficients $\# \mathrm{supp} \hat{u}$ of the Tikhonov minimizer increases, which drives the need for efficient iterative solvers for large-scale problems.

Iterative algorithms for the computation of Tikhonov minimizers are typically inspired by (sub)gradient descent strategies for the convex Tikhonov functional. The iterative soft thresholding algorithm (ISTA) [10] and its interpretations as proximal forward-backward operator splitting [8] and as a generalized conditional gradient method [5] are very popular in practical applications and have been studied intensively in the literature. It is well-known that the performance of gradient descent strategies hinges on a clever step size selection along the iteration, as iterative soft shrinkage with constant step size can perform arbitrarily slowly in the presence of small eigenvalues of $K^*K$.

Among the most efficient gradient descent strategies with variable step sizes, we mention the FISTA algorithm of Beck and Teboulle [2], see also [22]. By the convexity of the Tikhonov functional, the convergence of gradient descent-type algorithms is usually global. However, it is still linear [4], and the iteration does not yield the exact minimizer after finite many iterations.

As an alternative approach to gradient descent, one may use that the minimization problem (1) is equivalent to the shrinkage equation

$$
F(u) := u - S_{\gamma w}(u + \gamma K^*(f - Ku)) = 0, \quad \gamma > 0,
$$

where $\gamma$ can be chosen arbitrarily, and $S_\beta(v) := (\mathrm{sgn}(v_k)(|v_k| - \beta_k)_+)_k \in \mathbb{N}$ denotes componentwise soft thresholding of $v$ with respect to a positive weight sequence $\beta = (\beta_k)_{k \in \mathbb{N}}$. It was shown in [16] that the continuous, piecewise linear nonlinearity...
Global convergence of damped semismooth Newton methods for $\ell_1$ regularization

$\mathbf{F} = (F_j)_{j \in \mathcal{N}}$ in (2) is Newton differentiable at each $u \in \ell_2$, i.e. there exists a family of mappings $G: \ell_2 \to L(\ell_2, \ell_2)$ such that

$$\lim_{h \to 0} \frac{\|F(u + h) - F(u) - G(u + h)h\|_{\ell_2}}{\|h\|_{\ell_2}} = 0,$$

(3)

cf. [7]. The function $G$ is called a generalized derivative of $F$. Therefore, (2) can be treated by semismooth Newton methods [7,16]

$$u^{(j+1)} = u^{(j)} + d^{(j)}, \quad G(u^{(j)})d^{(j)} = -F(u^{(j)}), \quad j = 0, 1, \ldots$$

(4)

The convergence of semismooth Newton methods is locally superlinear in general [7]. However, in the case of the particular piecewise linear left-hand side $F$ from (2), one may even obtain the exact solution after a finite number of iterations, as long as the initial guess is sufficiently close to it, see also Section 2.

The global convergence properties of the semismooth Newton method, applied to Tikhonov functionals with sparsity constraints, have not yet been considered in the literature. In this work, we discuss a damping strategy which, at least in a finite-dimensional setting, yields a globally convergent variant of the semismooth Newton method from [16]. Our key idea to prove the global convergence is to work with Bouligand derivatives instead of Newton derivatives in (4), as proposed in [25]. Recall from [14,30,31] that a locally Lipschitz-continuous mapping $F: \ell_2 \to \ell_2$ is Bouligand differentiable at $u \in \ell_2$ if all directional derivatives $F'(u, d) := \lim_{h \to 0} \frac{F(u + h) - F(u) - F'(u, d)h}{h}$ exist and fulfill the approximation property $F(u + d) - F(u) - F'(u, d) = o(\|d\|_2)$, $d \to \mathbf{0}$. For a B-differentiable nonlinearity $F$, the associated B-semismooth Newton method reads as

$$u^{(j+1)} = u^{(j)} + d^{(j)}, \quad F'(u^{(j)}, d^{(j)}) = -F(u^{(j)}), \quad j = 0, 1, \ldots$$

(5)

In view of (3), the generalized derivative $G$ fulfills $G(u + d)d - F'(u, d) = o(\|d\|_2)$, as $d \to \mathbf{0}$, which is the reason why the B-semismooth Newton method (5) usually has the very same convergence properties as the semismooth Newton method (4), see also [25]. The main issue in the concrete realization of (5) is to prove that the nonlinear system

$$F'(u, d) = -F(u)$$

(6)

has a unique solution $d \in \ell_2$ for each $u \in \ell_2$. We will accomplish this by reducing (6) to a finite-dimensional linear complementarity problem [9]. What is more, we will show that there exists a particular generalized derivative $G$ of $F$ such that even $G(u)d = F'(u, d)$ for all $u, d \in \ell_2$.

As to the envisaged globalization strategy, we will focus our analysis on damping schemes. Here, the (semismooth) Newton direction $d^{(j)}$ is multiplied with a suitable damping parameter, and global convergence is enforced by monitoring the descent properties of the iterates $u^{(j)}$ with respect to a suitable merit functional $\Theta: \ell_2 \to \mathbb{R}$. In our paper, we will choose

$$\Theta_p(u) := \|F(u)\|_p = \sum_{k \in \mathcal{N}} |F_k(u)|_p, \quad u \in \ell_2,$$

(7)
where we shall restrict the discussion to the special cases $p \in \{1, 2\}$. On the one hand, $p = 2$ is a widely accepted choice in the globalization of Newton methods for smooth equations [11]. On the other hand, the case $p = 1$ is tempting in view of the piecewise linear nonlinearity $F$ at hand.

Our work is inspired by similar investigations for the numerical solution of optimal control problems [21] and of nonlinear complementarity problems [17] by semismooth Newton methods, and by the recent work [23] where the semismooth Newton iteration of [16] is globalized by interleaving it with iterative soft shrinkage steps.

The paper is organized as follows. In Section 2 we review the semismooth Newton method from [16] and give counterexamples to demonstrate that the method is not globally convergent in general. A local semismooth Newton method based on B-derivatives is treated in Section 3. In Section 4, we propose a damped version of the semismooth Newton method and prove its global convergence properties. Finally, Section 5 provides numerical experiments, ranging from inverse integration and deblurring of images to an inverse heat equation.

2. The semismooth Newton method and its basic properties

In this section, we are going to review the semismooth Newton method, applied to the numerical solution to the shrinkage equation (2).

2.1. The semismooth Newton method

One of the key observations of [16] was that the nonlinearity $F$ of the shrinkage equation is Newton differentiable on the full sequence space $\ell_2$, cf. (3). We cite the following theorem from [16].

**Theorem 2.1.** The mapping $F : \ell_2 \to \ell_2$ from (2) is Newton differentiable in each $u \in \ell_2$ with (nonunique) generalized derivative

$$G(u) = \begin{pmatrix} \gamma(K^*K)_{\mathcal{A}(u),\mathcal{A}} & \gamma(K^*K)_{\mathcal{A}(u),\mathcal{I}} \\ 0 & I_{\mathcal{I}} \end{pmatrix} \in L(\ell_2)$$

which is a block matrix with respect to the active set

$$\mathcal{A} = \mathcal{A}(u) := \{ k \in \mathbb{N} : \left| (u + \gamma K^* (f - K u))_k \right| > \gamma w_k \}$$

and the inactive set $\mathcal{I} = \mathcal{I}(u) := \mathcal{N} \setminus \mathcal{A}(u)$.

Note that by $w_k \geq w_0 > 0$ and $K \in L(\ell_2, Y)$, the active sets $\mathcal{A}(u)$ are finite for each $u \in \ell_2$, even in the case $\mathcal{N} = \mathbb{N}$. Therefore, $G(u)$ is boundedly invertible for each $u \in \ell_2$. What is more, $G(u)^{-1}$ is uniformly bounded in the operator norm in a sufficiently small neighborhood of the Tikhonov minimizer $\hat{u}$, see the following lemma from [16].
Lemma 2.2. There exist $\rho, C > 0$ such that $\|G(u)^{-1}\|_{\mathcal{L}(\ell_2)} \leq C$ for all $u \in \ell_2$ with $\|u - \hat{u}\|_{\ell_2} \leq \rho$.

The iteration (4) is well-defined by the bounded invertibility of $G$ from (8) on $\ell_2$. Moreover, by using the local uniform boundedness of $G(u)^{-1}$ and by applying the generic results from [7], it was proved in [16] that (4) is locally superlinearly convergent. Moreover, as pointed out in loc. cit., the proof of Theorem 2.1 reveals that the numerator of (3) for $F$ from (2), i.e., $F(u + h) - F(u) - G(u + h)h$, already vanishes for sufficiently small $\|h\|_{\ell_2}$, $u \in \ell_2$ being fixed. Therefore, if the current iterate $u^{(j)}$ is sufficiently close to the exact minimizer $\hat{u}$, it holds that

$$F(u^{(j)}) - F(\hat{u}) - G(u^{(j)})(u^{(j)} - \hat{u}) = F(u^{(j)}) - G(u^{(j)})(u^{(j)} - \hat{u}) = 0,$$

so that the next step of the semismooth Newton iteration (4) jumps into $\hat{u}$,

$$u^{(j+1)} = u^{(j)} - G(u^{(j)})^{-1}F(u^{(j)}) = \hat{u}.$$  

2.2. Cycling effect

One of the well-known pitfalls of the classical Newton method is the cycling effect. Here, after a certain number of iterations, the iterates start to oscillate between two or more distinct points. We will show now that the semismooth Newton iteration (4) may run into the same situation, at least if the parameter $\gamma$ is not chosen large enough.

2.2.1. One-dimensional case  
Assume that $\mathcal{N} = \{1\}$, $Y = \mathbb{R}$, $f \in \mathbb{R}$, $K \in \mathbb{R} \setminus \{0\}$ and $w > 0$. The unique minimizer of the Tikhonov functional

$$J(u) = \frac{1}{2}(Ku - f)^2 + w|u|, \quad u \in \mathbb{R}$$  

is given by

$$\hat{u} = S_{\frac{w}{K^2}}(\frac{f}{K}) = \begin{cases} \frac{f}{K} - \frac{w}{K^2}, & \frac{f}{K} > \frac{w}{K^2} \\ 0, & |\frac{f}{K}| \leq \frac{w}{K^2} \\ \frac{f}{K} + \frac{w}{K^2}, & \frac{f}{K} < -\frac{w}{K^2} \end{cases}.$$  

Let us analyze the convergence behavior of the semismooth Newton iteration (4), (8) in detail. Using the notation $r(u) := K(f - Ku)$, the nonlinearity $F$ from the shrinkage equation (2) has the explicit form

$$F(u) = \begin{cases} \gamma(K^2u - Kf + w), & u + \gamma r(u) > \gamma w \\ u, & |u + \gamma r(u)| \leq \gamma w \\ \gamma(K^2u - Kf - w), & u + \gamma r(u) < -\gamma w \end{cases}.$$  

The shape of the graph of $F$ depends on the size of the parameter $\gamma > 0$. In the outer region $|u + \gamma r(u)| = |(1 - \gamma K^2)u + \gamma Kf| > \gamma w$, the slope of $F$ is given by $\gamma K^2$. For small parameters $0 < \gamma < \frac{1}{K^2}$, the graph of $F$ is hence changing from convex to concave.
as \( u \) is growing, and vice versa for large parameters \( \gamma > \frac{1}{K^2} \). We refer to Figure 1 for the six possible cases concerning the location of the zero \( \hat{u} \) of \( F \). The generalized derivative (8) of \( F \) at \( u \in \mathbb{R} \) is

\[
G(u) = \begin{cases} 
\gamma K^2, & |u + \gamma r(u)| > \gamma w, \\
1, & |u + \gamma r(u)| \leq \gamma w,
\end{cases}
\] (13)

so that one semismooth Newton step maps the current iterate \( u^{(j)} \in \mathbb{R} \) to

\[
u^{(j+1)} = u^{(j)} - \frac{F(u^{(j)})}{G(u^{(j)})} = \begin{cases} 
\frac{f}{K} - \frac{w}{K^2}, & u^{(j)} + \gamma r(u^{(j)}) > \gamma w \\
0, & |u^{(j)} + \gamma r(u^{(j)})| \leq \gamma w, \\
\frac{f}{K} + \frac{w}{K^2}, & u^{(j)} + \gamma r(u^{(j)}) < -\gamma w
\end{cases}.
\] (14)

By taking all possible configurations into account, the complete behavior of the semismooth Newton iteration (4), (8) in the one-dimensional case can be summarized as in Table 1, compare also Figure 1.

If \( \hat{u} \neq 0 \) and for any \( \gamma > 0 \), the algorithm obviously terminates after at most three iterations with the exact minimizer \( \hat{u} \). In case that \( \hat{u} = 0 \), however, the behavior of the semismooth Newton iteration strongly depends on the size of \( \gamma \).

On the one hand, if \( \hat{u} = 0 \) and \( \gamma > \frac{1}{K^2} \), then \( u^{(j+2)} = 0 = \hat{u} \) regardless of the choice of \( u^{(j)} \), because one of the cases for \( u^{(j+2)} \) in the second row of Table 1 never occurs. As an example, if \( u^{(j)} + \gamma r(u^{(j)}) > \gamma w \) and hence \( u^{(j+1)} = \frac{f}{K} - \frac{w}{K^2} \), it holds

**Figure 1.** Qualitative behavior of \( F \) and the location of \( \hat{u} \) in the one-dimensional case

\[
\begin{align*}
\text{(a) } & \gamma < \frac{1}{K^2}, \quad \hat{u} = \frac{f}{K} + \frac{w}{K^2} < 0 \\
\text{(b) } & \gamma < \frac{1}{K^2}, \quad \hat{u} = 0 \\
\text{(c) } & \gamma < \frac{1}{K^2}, \quad \hat{u} = \frac{f}{K} - \frac{w}{K^2} > 0 \\
\text{(d) } & \gamma > \frac{1}{K^2}, \quad \hat{u} = \frac{f}{K} + \frac{w}{K^2} < 0 \\
\text{(e) } & \gamma > \frac{1}{K^2}, \quad \hat{u} = 0 \\
\text{(f) } & \gamma > \frac{1}{K^2}, \quad \hat{u} = \frac{f}{K} - \frac{w}{K^2} > 0
\end{align*}
\]
Global convergence of damped semismooth Newton methods for \( \ell_1 \) regularization

<table>
<thead>
<tr>
<th>( \tilde{u} )</th>
<th>( u^{(j)} + \gamma r(u^{(j)}) &gt; \gamma w )</th>
<th>Location of ( u^{(j)} )</th>
<th>( u^{(j)} + \gamma r(u^{(j)}) &lt; -\gamma w )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \frac{f}{K} + \frac{w}{K^2} &lt; 0 )</td>
<td>( u^{(j+1)} = \frac{f}{K} - \frac{w}{K^2} + \gamma w ) ( u^{(j+2)} = \left{ \begin{array}{l} \tilde{u}, \quad \frac{f}{K} - \frac{w}{K^2} &lt; -2\gamma w \ 0, \quad \frac{f}{K} - \frac{w}{K^2} \geq -2\gamma w \end{array} \right. )</td>
<td>( u^{(j+1)} = 0 ) ( u^{(j+2)} = \tilde{u} )</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>( u^{(j+1)} = \frac{f}{K} - \frac{w}{K^2} ) ( u^{(j+2)} = \left{ \begin{array}{l} \frac{f}{K} - \frac{w}{K^2}, \quad \frac{f}{K} - \frac{w}{K^2} &lt; -\gamma w \ \frac{f}{K} - \frac{w}{K^2}, \quad \frac{f}{K} - \frac{w}{K^2} \geq -\gamma w \end{array} \right. )</td>
<td>( u^{(j+1)} = \tilde{u} ) ( u^{(j+2)} = \left{ \begin{array}{l} 0, \quad \frac{f}{K} + \frac{w}{K^2} \leq 2\gamma w \ \tilde{u}, \quad \frac{f}{K} + \frac{w}{K^2} &gt; 2\gamma w \end{array} \right. )</td>
<td></td>
</tr>
<tr>
<td>( \frac{f}{K} - \frac{w}{K^2} &gt; 0 )</td>
<td>( u^{(j+1)} = \tilde{u} ) ( u^{(j+2)} = \tilde{u} )</td>
<td>( u^{(j+1)} = 0 ) ( u^{(j+2)} = \tilde{u} )</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Behavior of the semismooth Newton iteration in the one-dimensional case.

that \( u^{(j+1)} + \gamma r(u^{(j+1)}) = \frac{f}{K} - \frac{w}{K^2} + \gamma w \). \( \tilde{u} = 0 \) and hence \( |\frac{f}{K}| \leq \frac{w}{K^2} \) imply that \( u^{(j+1)} + \gamma r(u^{(j+1)}) \leq \gamma w \). Finally, \( \gamma > \frac{1}{K^2} \) yields \( u^{(j+1)} + \gamma r(u^{(j+1)}) \geq -\frac{w}{K^2} + \gamma w > -\gamma w \) and hence \( u^{(j+2)} = 0 = \tilde{u} \). The same argument works in the case \( u^{(j)} + \gamma r(u^{(j)}) < -\gamma w \).

Summarizing, we can conclude that the semismooth Newton iteration (4), (8) converges in at most three steps if \( \gamma \) is chosen larger than \( \frac{1}{K^2} \).

On the other hand, if \( \tilde{u} = 0 \), i.e., \( |\frac{f}{K}| \leq \frac{w}{K^2} \), and \( 0 < \gamma < \frac{1}{K^2} \) is sufficiently small, the semismooth Newton iteration might start to oscillate between \( \tilde{u} \). We obtain a cycle, e.g., if \( |\frac{f}{K}| < \frac{w}{K^2} \), \( 0 < \gamma < \frac{1}{K^2} \) and \( u^{(j)} = \frac{f}{K} + \frac{w}{K^2} \neq 0 = \tilde{u} \). Here \( u^{(j)} + \gamma r(u^{(j)}) = \frac{f}{K} + \frac{w}{K^2} - \gamma w \geq \frac{w}{K^2} - |\frac{f}{K}| - \gamma w > \gamma w \) and hence \( u^{(j+1)} = \frac{f}{K} - \frac{w}{K^2} + \gamma w < -\gamma w \) and hence \( u^{(j+2)} = \frac{f}{K} + \frac{w}{K^2} = u^{(j)} \).

### 2.2.2. Finite-dimensional case

The cycling effect can be observed in higher dimensions as well. In the two-dimensional case, choosing

\[
K = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}, \quad f = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad w = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad \gamma = \frac{3}{2}, \quad u^{(0)} = \begin{pmatrix} -6 \\ 12 \end{pmatrix},
\]

the next iterates are

\[
u^{(1)} = \begin{pmatrix} 10 \\ -12 \end{pmatrix}, \quad u^{(2)} = \begin{pmatrix} -6 \\ 12 \end{pmatrix} = u^{(0)},
\]

i.e. a cycle arises. Choosing \( \gamma = 3 \) instead, the zero of \( F \) would be found within two steps.

In the four-dimensional case, the choice

\[
K = \frac{1}{4} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}, \quad f = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad w = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad \gamma = 2, \quad u^{(0)} = \begin{pmatrix} 36 \\ -\frac{112}{3} \\ 0 \\ 0 \end{pmatrix},
\]
Global convergence of damped semismooth Newton methods for \(\ell_1\) regularization

leads to the iterates

\[
\mathbf{u}^{(1)} = \begin{pmatrix} -28 \\ \frac{112}{3} \\ 3 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{u}^{(2)} = \begin{pmatrix} 36 \\ -\frac{112}{3} \\ 3 \\ 0 \\ 0 \end{pmatrix} = \mathbf{u}^{(0)}.
\]

These examples show that the unmodified semismooth Newton method (4), (8) is not globally convergent in general and may even lead to cycles, motivating the analysis of suitable globalization strategies.

3. The local B-semismooth Newton method

Instead of (4), we are going to study the B-semismooth Newton method (5) which is based on the B(ouligand)-derivative \(F'\). The B-Newton directions \(d^{(j)}\) will be chosen by solving the generalized Newton equation (6), cf. Subsection 3.2.

3.1. Directional differentiability of the nonlinearity \(F\)

Let us first analyze the directional differentiability of the nonlinearity \(F = (F_k)_{k \in \mathbb{N}}\), i.e., the existence and concrete shape of \(F'(\mathbf{u}, \mathbf{d}) = \lim_{h \searrow 0} \frac{F(\mathbf{u} + hd) - F(\mathbf{u})}{h}\), where \(\mathbf{u}, \mathbf{d} \in \ell_2\). To this end, we shall use the following straightforward componentwise representation of \(F\):

\[
F_k(\mathbf{u}) = \begin{cases} 
\gamma(K^*(K\mathbf{u} - f))_k + \gamma w_k, & k \in A^+(\mathbf{u}) \\
\gamma(K^*(K\mathbf{u} - f))_k - \gamma w_k, & k \in A^-(\mathbf{u}) \\
u_k, & k \in I(\mathbf{u})
\end{cases}
\]

where we have splitted the active set into \(A(\mathbf{u}) = A^+(\mathbf{u}) \cup A^-(\mathbf{u})\) with

\[
A^+(\mathbf{u}) := \{k \in \mathbb{N} : (\mathbf{u} + \gamma K^*(f - K\mathbf{u}))_k > \gamma w_k\},
\]

\[
A^-(\mathbf{u}) := \{k \in \mathbb{N} : (\mathbf{u} + \gamma K^*(f - K\mathbf{u}))_k < -\gamma w_k\}.
\]

Let us also split the inactive set into \(I(\mathbf{u}) = I^+(\mathbf{u}) \cup I^0(\mathbf{u}) \cup I^-(\mathbf{u})\), where

\[
I^+(\mathbf{u}) := \{k \in \mathbb{N} : (\mathbf{u} + \gamma K^*(f - K\mathbf{u}))_k = \gamma w_k\},
\]

\[
I^-(\mathbf{u}) := \{k \in \mathbb{N} : (\mathbf{u} + \gamma K^*(f - K\mathbf{u}))_k = -\gamma w_k\},
\]

\[
I^0(\mathbf{u}) := \{k \in \mathbb{N} : |(\mathbf{u} + \gamma K^*(f - K\mathbf{u}))_k| < \gamma w_k\}.
\]

In the following we drop the argument \(\mathbf{u}\) of the active and inactive sets if there is no risk of confusion. With this notation at hand, we shall now prove the directional differentiability of \(F\). While the directional differentiability of a single component \(F_k\) and the corresponding formula for the directional derivative \(F'_k(\mathbf{u}, \mathbf{d}) = \lim_{h \searrow 0} \frac{F_k(\mathbf{u} + hd) - F_k(\mathbf{u})}{h}\) are almost obvious, the directional differentiability of the full operator \(F\) is nontrivial, at least in the case \(\mathcal{N} = \mathbb{N}\). For this reason, we will prove the stronger property that the residual \(F(\mathbf{u} + hd) - F(\mathbf{u}) - hF'(\mathbf{u}, \mathbf{d})\) vanishes for sufficiently small \(h\).
Lemma 3.1. Let $u, d \in \ell_2$ be arbitrary. There exists $h_0 = h_0(u, d) > 0$, so that

\begin{equation}
\frac{F_k(u + hd) - F_k(u)}{h} = \begin{cases}
\gamma(K^*Kd)_k, & k \in A(u) \\
d_k, & k \in T^*(u), \\
\min\{d_k, \gamma(K^*Kd)_k\}, & k \in T^+(u), \\
\max\{d_k, \gamma(K^*Kd)_k\}, & k \in T^-(u)
\end{cases}, \quad 0 < h \leq h_0. \tag{21}
\end{equation}

In particular, $F = (F_k)_{k \in \mathcal{N}}$ and its components $F_k$ are directionally differentiable at $u$ in the direction $d$, with directional derivative $F'(u, d) = (F'(u, d))_{k \in \mathcal{N}} \in \ell_2$ given componentwise by the right-hand side of (21).

Proof. Let $u, d \in \ell_2$, $k \in \mathcal{N}$ and $h > 0$ be arbitrary.

If $k \in T^+(u)$, (15) and (18) tell us that

$$F_k(u) = u_k = \gamma(K^*(Ku - f))_k + \gamma w_k.$$ 

If, additionally, $((I - \gamma K^*K)d)_k > 0$, then $k \in A^+(u + hd)$, so that an application of (15) yields $F_k(u + hd) - F_k(u) = h\gamma(K^*Kd)_k$, showing (21). The same argument works for $k \in T^-(u)$ and $((I - \gamma K^*K)d)_k < 0$.

If $k \in T^+(u)$ and $((I - \gamma K^*K)d)_k \leq 0$, we choose $h_0 > 0$ so small that $h_0 \|((I - \gamma K^*K)d)\|_{\ell_2} \leq \gamma w_0$. It follows that for all $0 < h \leq h_0$,

$$(u + hd + \gamma K^*(f - K(u + hd))) = \gamma w_k + h((I - \gamma K^*K)d)_k \in (-\gamma w_k, \gamma w_k]$$

and hence $k \in T(u + hd)$. We conclude that $F_k(u + hd) - F_k(u) = hd_k$ for all $0 < h \leq h_0$, showing (21). The same argument works for $k \in T^-(u)$ and $((I - \gamma K^*K)d)_k \geq 0$.

Moreover, due to the fact that $A(u)$ is a finite set, we can find $h_0 > 0$ such that $A(u) \subset A(u + hd)$ for all $0 < h \leq h_0$ with

$$h_0 \leq \min_{l \in A(u)} \inf \left\{ h > 0 : \|(u + hd + \gamma K^*(f - K(u + hd)))_l\| > \gamma w_l \right\}.$$ 

For $0 < h \leq h_0$ and $k \in A(u) \subset A(u + hd)$, (15) yields $F_k(u + hd) - F_k(u) = h\gamma(K^*Kd)_k$ and hence (21).

Finally, in order to show (21) for $k \in T^0(u)$, we decompose $T^0(u) = T_1(u) \cup T_2(u)$, setting

$$T_1(u) := \{ k \in T^0(u) : \|(u + \gamma(K^*(f - K'u)))_k\| > \frac{\gamma w_0}{2} \},$$
$$T_2(u) := \{ k \in T^0(u) : \|(u + \gamma(K^*(f - K'u)))_k\| \leq \frac{\gamma w_0}{2} \}.$$

Note that $T_1(u)$ is a finite set, due to $u + \gamma K^*(f - K'u) \in \ell_2$. We may therefore choose $h_0 > 0$ even so small that $T_1(u) \subset T(u + hd)$ for all $0 < h \leq h_0$, e.g., by additionally requiring that

$$h_0 \leq \min_{l \in T_1(u)} \inf \left\{ h > 0 : \|(u + hd + \gamma K^*(f - K(u + hd)))_l\| < \gamma w_l \right\}.$$
Additionally, we choose \( h_0 > 0 \) so small that \( h_0 \|(I - \gamma K^*K)d\|_{\ell_2} \leq \frac{\gamma w_0}{2} \). We conclude that for all \( 0 < h \leq h_0 \) and \( k \in I_2(u) \),
\[
\|(u + hd + \gamma K^*(f - K(u + hd)))_k\| \leq \|(u + \gamma K^*(f - Ku))_k\| + \|(I - \gamma K^*K)d\|_{\ell_2} \leq \gamma w_0 \leq \gamma w_k,
\]
so that also \( I_2(u) \subset I(u + hd) \) for all \( 0 < h \leq h_0 \). (15) implies that \( F_k(u + hd) - F_k(u) = hd_k \) for all \( k \in I^o(u) \) and \( 0 < h \leq h_0 \). Note, that \( h_0 = h_0(u, d) \) is independent of \( k \). \( \square \)

As already mentioned, our globalization strategy in section 4 will choose damping parameters \( t_j \) in such a way that sufficient descent of the merit functional \( \Theta_p \) from (7) is guaranteed. To this end, we compute the directional derivatives of \( \Theta_p \).

**Lemma 3.2.** Let \( u, d \in \ell_2 \) be arbitrary. Then for \( p \in \{1, 2\} \), \( \Theta_p \) is directionally differentiable at \( u \) in the direction \( d \) with directional derivatives
\[
\Theta'_p(u, d) = \begin{cases} 
\sum_{F_k(u) > 0} F'_k(u, d) + \sum_{F_k(u) = 0} |F'_k(u, d)| - \sum_{F_k(u) < 0} F'_k(u, d), & p = 1 \\
2 \sum_{k \in \mathcal{N}} F'_k(u, d) F_k(u), & p = 2
\end{cases} \tag{22}
\]

**Proof.** (22) immediately follows from the chain rule for directionally differentiable mappings, see [29, Theorem 3.1.1]. \( \square \)

### 3.2. The feasibility of the local B-semismooth Newton iteration

In the sequel, for a given \( u \in \ell_2 \), we will discuss the existence and uniqueness of a solution \( d \in \ell_2 \) to the nonlinear problem (6)

\[
F'(u, d) = -F(u).
\]

Note that in view of Lemma 3.2, a solution \( d \) to (6) has the particularly interesting properties that

\[
\Theta'_1(u, d) = -\sum_{k \in \mathcal{N}} |F_k(u)| = -\Theta_1(u),
\]

\[
\Theta'_2(u, d) = 2\langle F'(u, d), F(u) \rangle = -2\Theta_2(u).
\]

(23)

It will turn out that a solution \( d \) to (6) coincides with the semismooth Newton direction from [16] almost everywhere, see Lemma 3.4. The following auxiliary results ensure the solvability of (6).

**Lemma 3.3.** Let \( u \in \ell_2 \), \( \gamma > 0 \), \( M := K^*K \) and
\[
N := \gamma \begin{pmatrix} M_{I+, I+} - M_{I+, A} M_{A, A}^{-1} M_{A, I+} M_{I+, I+} & M_{I+, A} M_{A, A}^{-1} M_{A, I+} - M_{I+, I+} \\ M_{I-, A} M_{A, A}^{-1} M_{A, I+} - M_{I-, I+} & M_{I-, I-} - M_{I-, A} M_{A, A}^{-1} M_{A, I-} \end{pmatrix} \tag{24}
\]

Then the finite-dimensional matrix \( N \) is symmetric and positive definite.
Proof. Without loss of generality, let $\gamma = 1$. The symmetry of $N$ readily follows from that of $M$. Concerning the positive definiteness of $N$, note that each real $3 \times 3$ block matrix with symmetric off-diagonal blocks and symmetric invertible diagonal block $F$ can be decomposed as

$$
\begin{pmatrix}
A & B & C \\
B^T & D & E \\
C^T & E^T & F
\end{pmatrix} = X \begin{pmatrix}
A - CF^{-1}C^T & CF^{-1}E^T - B & 0 \\
EF^{-1}C^T - B^T & D - EF^{-1}E^T & 0 \\
0 & 0 & F
\end{pmatrix} X^T,
$$

with

$$
X = \begin{pmatrix}
-I & 0 & CF^{-1} \\
0 & I & EF^{-1} \\
0 & 0 & I
\end{pmatrix}.
$$

Setting $A := M_{I^+, I^+}$, $B := M_{I^+, I^-}$, $C := M_{I^+, A}$, $D := M_{I^-, I^-}$, $E := M_{I^-, A}$ and $F := M_{A, A}$, (25) tells us that $N$ is the upper left $2 \times 2$ diagonal block of $X^{-1}M_{I^+, I^+ - I^+, A} + M_{I^-, I^- - I^-, A}X^{-T}$ and hence positive definite. $\square$

The solvability of (6) hinges on the solvability of a particular linear complementarity problem.

Lemma 3.4. Let $u \in \ell_2$. Then with $M := K^* K$, $d \in \ell_2$ solves (6) if and only if

$$
\begin{align*}
\gamma(Md)_A &= -F(u)_A, \\
d_{I^+} &= -u_{I^+},
\end{align*}
$$

and

$$
\begin{align*}
x := \begin{pmatrix} d_{I^+} + u_{I^+} \\ -d_{I^-} - u_{I^-} \end{pmatrix}, \\
y := \begin{pmatrix} \gamma(Md)_{I^+} + u_{I^+} \\ -\gamma(Md)_{I^-} - u_{I^-} \end{pmatrix},
\end{align*}
$$

solve the linear complementarity problem

$$
x, y \succeq 0, \quad y = Nx + z, \quad \langle x, y \rangle = 0,
$$

where the data $N = N(u)$ and $z = z(u)$ are given by (24) and

$$
z = \begin{pmatrix}
\gamma(M_{I^+, A}M_{A, I^+}^{-1}A_{I^+, I^+} - M_{I^+, I^+}u_{I^+}) - M_{I^+, A}F(u)_A + u_{I^+} \\
\gamma(M_{I^-, A}M_{A, I^-}^{-1}A_{I^-, I^-} - M_{I^-, I^-}u_{I^-}) + M_{I^-, A}F(u)_A - u_{I^-}
\end{pmatrix}
$$

and

$$
\begin{align*}
&= -\gamma
\begin{pmatrix}
M_{I^+, I^+ - I^+, A}M_{A, I^+}^{-1}M_{A, I^+} - M_{I^+, I^+} \\
M_{I^-, I^- - I^-, A}M_{A, I^-}^{-1}M_{A, I^-} - M_{I^-, I^-}
\end{pmatrix}
\begin{pmatrix}
u_{I^+} \\
u_{I^-}
\end{pmatrix}.
\end{align*}
$$

Proof. By componentwise application of (15) and (21) to the indices from $A, I^o, I^+$ and $I^-$, we observe first that (6) is equivalent to (26), (27) and the conditions

$$
(d_k + u_k > 0 \land \gamma(Md)_k + u_k = 0) \lor (d_k + u_k = 0 \land \gamma(Md)_k + u_k \geq 0), \quad k \in I^+,
$$

$$
(d_k + u_k < 0 \land \gamma(Md)_k + u_k = 0) \lor (d_k + u_k = 0 \land \gamma(Md)_k + u_k \leq 0), \quad k \in I^-,
$$

respectively.
respectively. The last two conditions are actually equivalent to
\[
\begin{align*}
d_{I^+} + u_{I^+} & \geq 0, \quad u_{I^+} + \gamma (Md)_{I^+} \geq 0, \quad \langle d_{I^+} + u_{I^+}, u_{I^+} + \gamma (Md)_{I^+} \rangle = 0 \quad (31)
\end{align*}
\]
and
\[
\begin{align*}
d_{I^-} + u_{I^-} & \leq 0, \quad u_{I^-} + \gamma (Md)_{I^-} \leq 0, \quad \langle d_{I^-} + u_{I^-}, u_{I^-} + \gamma (Md)_{I^-} \rangle = 0. \quad (32)
\end{align*}
\]
Now assume that \( d \in \ell_2 \) solves (6) and hence (26), (27), (31) and (32) hold. By defining \( x, y \) as in (28), (31) and (32) tell us that \( x \geq 0, y \geq 0 \) and \( \langle x, y \rangle = 0 \). Moreover, (26) and (27) imply that
\[
d_A = \frac{1}{\gamma} M^{-1}_{A,A} (\gamma M_{A,I^o} u_{I^o} - F(u)_A - \gamma M_{A,I^+} d_{I^+} - \gamma M_{A,I^-} d_{I^-}). \quad (33)
\]
Inserting (33) into the definition of \( y \) from (28), we obtain
\[
y = \begin{pmatrix}
\gamma (Md)_{I^+} + u_{I^+} \\
-\gamma (Md)_{I^-} - u_{I^-}
\end{pmatrix}
= \begin{pmatrix}
\gamma M_{I^+,A} d_A + \gamma M_{I^+,I^+} d_{I^+} + \gamma M_{I^+,I^-} d_{I^-} - \gamma M_{I^+,I^o} u_{I^o} + u_{I^+} \\
-\gamma M_{I^-,A} d_A - \gamma M_{I^-,I^+} d_{I^+} - \gamma M_{I^-,I^-} d_{I^-} + \gamma M_{I^-,I^o} u_{I^o} - u_{I^-}
\end{pmatrix}
= N \begin{pmatrix}
d_{I^+} \\
-d_{I^-}
\end{pmatrix}
+ \begin{pmatrix}
\gamma (M_{I^+,A} M^{-1}_{A,A} M_{A,I^o} - M_{I^+,I^o}) u_{I^o} - M_{I^+,A} M^{-1}_{A,A} F(u)_A + u_{I^+} \\
\gamma (M_{I^-,I^o} - M_{I^-,A} M^{-1}_{A,A} M_{A,I^o}) u_{I^o} + M_{I^-,A} M^{-1}_{A,A} F(u)_A - u_{I^-}
\end{pmatrix}
= Nx + z,
\]
with \( N \) from (24) and \( z \) from (30), showing (29).

Conversely, assume that \( x, y \) solve (29) and define \( d \in \ell_2 \) blockwise via
\[
\begin{pmatrix}
d_{I^+} \\
-d_{I^-}
\end{pmatrix} := x + \begin{pmatrix}
-u_{I^+} \\
u_{I^-}
\end{pmatrix}
\]
and the system (26), (27) for the remaining entries from \( A \cup I^o \). It remains to show that \( d \) solves (6). As we have already seen above, (26) and (27) are equivalent to (6) holding in the entries corresponding to \( A \cup I^o \). Moreover, as above, (26) and (27) imply (33), which yields
\[
y = Nx + z = \begin{pmatrix}
\gamma (Md)_{I^+} + u_{I^+} \\
-\gamma (Md)_{I^-} - u_{I^-}
\end{pmatrix},
\]
tracing back the previous computation. By consequence, (28) holds. (29) implies (31) and (32), and thus (6) holds in the remaining entries corresponding to \( I^+ \cup I^- \). \( \square \)

We are now in the position to prove the following theorem on the solvability of (6).

**Theorem 3.5.** Let \( x, y \) be the unique solution of (29) for an iterate \( u^{(j)} \) and let \( I = I^o \cup I^+ \cup I^- \). Then the local B-semismooth Newton iteration (5) is well-defined,
with
\[
\begin{align*}
 d^{(j)}_{I^c} &= -u^{(j)}, \\
 d^{(j)}_{I^+} &= x_{I^+} - u^{(j)}_{I^+}, \\
 d^{(j)}_{I^-} &= -x_{I^-} - u^{(j)}_{I^-}, \\
 d^{(j)}_{A} &= \frac{1}{\gamma}M_{A,A}^{-1}(-\gamma M_{A,x}d^{(j)}_I - F(u^{(j)}_A)).
\end{align*}
\]

**Proof.** The finite-dimensional linear complementarity problem (29) is uniquely solvable
if \( N \) is a P-matrix, i.e. if all of its principal minors are positive [9, Definition 3.3.1, Theorem 3.3.7]. According to Lemma 3.3, \( N \) is symmetric and positive definite for every iterate \( u^{(j)} \in \ell_2 \). Hence, \( N \) is a P-Matrix and (5), (34) are well-defined.

Concerning the convergence properties of the local B-semismooth Newton method, we can rely on the following result from [27, Corollary 3.3].

**Corollary 3.6.** Let \( u^* \) be the unique solution to \( F(u) = 0 \). Then, the iterates of the local B-semismooth Newton iteration (5) converge to \( u^* \) superlinearly in a neighborhood of \( u^* \).

### 4. The global B-semismooth Newton iteration

In this section, we will analyze the following damped B-semismooth Newton method for the computation of \( \ell_1 \) Tikhonov minimizers
\[
\begin{align*}
 u^{(j+1)} &= u^{(j)} + t_j d^{(j)}, \\
 F'(u^{(j)}, d^{(j)}) &= -F(u^{(j)}), \quad j = 0, 1, \ldots
\end{align*}
\]
and its global convergence properties in a finite-dimensional setting \( N = \{1, \ldots, n\} \).

Inspired by the works [17, 21, 25, 27], we shall choose the damping parameters \( t_j \) by an Armijo-type rule. The overall algorithm can be found in Algorithm 1.

Let us first show that damping parameters \( t_j \) are well-defined.

**Algorithm 1** Damped B-semismooth Newton method

Choose starting value \( u^{(0)} \) and parameters \( p \in \{1, 2\}, \beta \in (0, 1), \sigma \in \left(0, \frac{1}{p}\right) \) and a tolerance \( tol \)

while \( \Theta_p(u^{(j)}) \geq tol \) do

Compute the Newton update \( d^{(j)} \) according to (34)
\( t_j = 1 \)

while \( \Theta_p(u^{(j)} + t_j d^{(j)}) > (1 - p\sigma t_j)\Theta_p(u^{(j)}) \) do
\( t_j = t_j \beta \)
end while

\( u^{(j+1)} = u^{(j)} + t_j d^{(j)} \)
\( j = j + 1 \)
end while
Global convergence of damped semismooth Newton methods for $\ell_1$ regularization

**Proposition 4.1.** Let $p \in \{1, 2\}, \beta \in (0, 1), \sigma \in (0, \frac{1}{p})$. Let $u^{(j)}$ with $\Theta_p(u^{(j)}) > 0$ be an iterate in Algorithm 1 and let $d^{(j)} = d(u^{(j)})$ be chosen according to (34). Then there exists a finite index $l \in \mathbb{N}$ with

$$\Theta_p(u^{(j)} + \beta^l d^{(j)}) \leq (1 - p\sigma^\beta^l)\Theta_p(u^{(j)}).$$  \hfill (36)

**Proof.** According to (23), it holds $\Theta'_p(u^{(j)}, d^{(j)}) = -p\Theta_p(u^{(j)}) < 0$. Assuming that there exists no such index $l \in \mathbb{N}$, it holds for all $l \in \mathbb{N}$

$$\frac{\Theta_p(u^{(j)} + \beta^l d^{(j)}) - \Theta_p(u^{(j)})}{\beta^l} > -p\sigma\Theta_p(u^{(j)}).$$

For $l \to \infty$, it follows that $\Theta'_p(u^{(j)}d^{(j)}) \geq -p\sigma\Theta_p(u^{(j)})$, which is a contradiction to $\Theta'_p(u^{(j)}d^{(j)}) = -p\Theta_p(u^{(j)}) < -p\sigma\Theta_p(u^{(j)})$.  \hfill $\square$

**Remark 4.2.** For the particular B-Newton direction from (34), inequality (36) is equivalent to the ordinary Armijo rule, because it holds

$$\Theta_p(u^{(j)} + t_j d^{(j)}) \leq (1 - p\sigma t_j)\Theta_p(u^{(j)}) = \Theta_p(u^{(j)}) - p\sigma t_j\Theta_p(u^{(j)})$$

$$= \Theta_p(u^{(j)}) + \sigma t_j\Theta'_p(u^{(j)}, d^{(j)})$$

with $\sigma \in (0, \frac{1}{p})$.

As an important ingredient of our convergence analysis, we will show that $\|F(\cdot)\|_p$, $p \in \{1, 2\}$ is coercive with respect to the Euclidean norm, i.e., $\lim_{\|u\|_2 \to \infty} \|F(u)\|_p = \infty$. For the proof, we need the following estimates for soft shrinkage in finite-dimensional spaces.

**Lemma 4.3.** Let $0 \leq w \in \mathbb{R}^n$. Then it holds that

$$\|u - S_w(u)\|_p \leq n^{\frac{1}{p}}\|w\|_\infty, \quad \text{for all } u \in \mathbb{R}^n.$$  \hfill (37)

**Proof.** For each $u \in \mathbb{R}^n$, (37) follows from

$$\|u - S_w(u)\|_p^p = \sum_{|u_k| \leq w_k} |u_k|^p + \sum_{|u_k| > w_k} w_k^p \leq \sum_{k=1}^n w_k^p \leq n\|w\|_\infty^p.$$

$\square$

The coercivity of $\|F(\cdot)\|_p$ with respect to the Euclidean norm for each $\gamma > 0$ immediately follows from the following lemma, because $I - \gamma K^*K \in \mathbb{R}^{n \times n}$ is symmetric and 1 is not an eigenvalue of this matrix because $K$ is injective.

**Lemma 4.4.** Let $0 \leq w \in \mathbb{R}^n, c \in \mathbb{R}^n, B \in \mathbb{R}^{n \times n}$ be symmetric, and assume that 1 is not an eigenvalue of $B$. Then there exist $\alpha > 0$ and $\beta \in \mathbb{R}$ with

$$\|u - S_w(Bu + c)\|_p \geq \alpha\|u\|_2 + \beta, \quad \text{for all } u \in \mathbb{R}^n.$$  \hfill (38)
Global convergence of damped semismooth Newton methods for $\ell_1$ regularization

Proof. By assumption on the spectral properties of $B$, there exist $B$-invariant subspaces $U_1, U_2$ and constants $0 < \rho < 1 < \eta$ with

$$\|Bu_1\|_2 \leq \rho \|u_1\|_2, \quad \text{for all } u_1 \in U_1,$$

and

$$\|Bu_2\|_2 \geq \eta \|u_2\|_2, \quad \text{for all } u_2 \in U_2,$$

such that the splitting $\mathbb{R}^n = U_1 \oplus U_2$ is orthogonal. For an arbitrary $u \in \mathbb{R}^n$ and its orthogonal projections of $u$ onto $U_j, 1 \leq j \leq 2$, a double application of Pythagoras' theorem implies that

$$\|(I-B)u\|_2 = \sqrt{\|(I-B)u_1\|_2^2 + \|(I-B)u_2\|_2^2} \geq \min\{1-\rho, \eta-1\} \|u\|_2. \quad (39)$$

By combining (39), (37) and the nonexpansivity of the soft shrinkage operator $S_w$, we obtain that

$$\|u - S_w(Bu + c)\|_2 \geq \|(I-B)(u)\|_2 - \|Bu - S_w(Bu)\|_2 - \|S_w(Bu) - S_w(Bu + c)\|_2 \geq \min\{1-\rho, \eta-1\} \|u\|_2 - \sqrt{n}\|w\|_{\infty} - \|c\|_2,$$

proving the claim for $p = 2$. The claim for $p = 1$ follows from

$$\|u - S_w(Bu + c)\|_1 \geq \|u - S_w(Bu + c)\|_2.$$

We proceed by verifying the compactness of the level sets of $\|F(\cdot)\|_p$.

Proposition 4.5. Let $u^{(0)}$ be an arbitrary starting vector of Algorithm 1. Then the level set $L(u^{(0)}) := \{u : \|F(u)\|_p \leq \|F(u^{(0)})\|_p\}, p \in \{1,2\}$ is compact.

Proof. This follows immediately from the coercivity (38) of the function $\|F(\cdot)\|_p$. □

In the following proposition, we will show that the $B$-semismooth Newton directions from (34) are bounded with respect to the current residual norms.

Proposition 4.6. Let $u \in \mathbb{R}^n, p \in \{1,2\}$, and let $d = d(u)$ be chosen according to (34). There exists a constant $C > 0$, independent of $u$, with

$$\|d\|_p \leq C \|F(u)\|_p.$$

Proof. Let $N, z, x, y$ be as in (24), (29), (30). It suffices to consider $p = 2$. It holds

$$\|d\|_2^2 = \|d_A\|_2^2 + \|d_{I^*}\|_2^2 + \|d_{I^+}\|_2^2 + \|d_{I^-}\|_2^2$$

and

$$\|d_{I^+}\|_2 = \|x_{I^+} - u_{I^+}\|_2 \leq \|x\|_2 + \|F(u)\|_2 \|d_{I^-}\|_2 = \|x_{I^-} + u_{I^-}\|_2 \leq \|x\|_2 + \|F(u)\|_2 \|d_{I^*}\|_2 = \|F(u)_{I^*}\|_2 \leq \|F(u)\|_2.$$
Consider first $\|x\|_2$. Since $N$ is symmetric and positive definite for all $u \in \mathbb{R}^n$ (cf. Lemma 3.3), $N$ is a P-Matrix, cf. proof of Theorem 3.5. Corresponding to [9, p.478], we define

$$c(N) := \min_{\|z\|_\infty = 1} \{ \max_{1 \leq i \leq n} z_i(Nz_i) \}.$$

According to [9, Lemma 7.3.10], it holds for arbitrary $z, z' \in \mathbb{R}^n$

$$\|x - x'|_\infty \leq c(N)^{-1} \|z - z'|_\infty,$$

where $x$ is the unique solution of

$$x \geq 0, \quad Nx + z \geq 0, \quad \langle x, Nx + z \rangle = 0$$

and $x'$ is the unique solution of

$$x' \geq 0, \quad Nx' + z' \geq 0, \quad \langle x', Nx' + z' \rangle = 0.$$

Choosing $z' = 0$, it follows $\langle x', Nx' \rangle = 0$. Since $N$ is positive definite, $x' = 0$ and

$$\|x\|_\infty \leq c(N)^{-1} \|z\|_\infty \leq \max_{A, I^0, I^+, I^-} c(N)^{-1} \|z\|_\infty.$$

Note that there exist only $4^n$ possibilities for the decomposition of $\{1, \ldots, n\}$ into the sets $A, I^0, I^+, I^-$. Therefore, there exists a constant $c_1 = c_1(n) > 0$ with $\|x\|_2 \leq c_1 \|z\|_2$. Moreover, (30) implies that

$$\|z\|_2 \leq \left( \gamma \|M_{I^+, A}M_{A, I^0}^{-1}M_{A, I^0} - M_{I^+, I^0}\|_2 + \|M_{I^+, A}M_{A, A}^{-1}\|_2 + 1 \right)^2 \|F(u)\|_2^2$$

$$+ \left( \gamma \|M_{I^-, A}M_{A, I^0}^{-1}M_{A, I^0} - M_{I^-, I^0}\|_2 + \|M_{I^-, A}M_{A, A}^{-1}\|_2 + 1 \right)^2 \|F(u)\|_2^2 \frac{1}{2}$$

$$+ \|N\|_2 \|F(u)\|_2$$

$$\leq c_2 \|F(u)\|_2,$$

where

$$c_2 := \max_{A, I^0, I^+, I^-} \left[ \left( \gamma \|M_{I^+, A}M_{A, A}^{-1}M_{A, I^0} - M_{I^+, I^0}\|_2 + \|M_{I^+, A}M_{A, A}^{-1}\|_2 + 1 \right)^2 \right.\left. + \left( \gamma \|M_{I^-, A}M_{A, A}^{-1}M_{A, I^0} - M_{I^-, I^0}\|_2 + \|M_{I^-, A}M_{A, A}^{-1}\|_2 + 1 \right)^2 \right].$$

Moreover,

$$\|d_A\|_2 = \|\gamma^{-1}(-M_{A, A}^{-1}F(u), A - M_{A, A}^{-1}M_{A, I}d_I)\|_2$$

$$\leq \gamma^{-1}\|M_{A, A}^{-1}\|_2 \|F(u)\|_2 + \gamma^{-1}\|M_{A, A}^{-1}M_{A, I}\|_2 \|d_I\|_2,$$

where $I = I^+ \cup I^- \cup I^0$, and

$$\|d_I\|_2^2 = \|d_{I^+}\|_2^2 + \|d_{I^-}\|_2^2 + \|d_{I^0}\|_2^2$$

$$\leq 2(\|x\|_2 + \|F(u)\|_2^2 + \|F(u)\|_2^2$$

$$\leq (2(c_1c_2 + 1)^2 + 1) \|F(u)\|_2^2.$$

Global convergence of damped semismooth Newton methods for $\ell_1$ regularization
Therefore,
\[ \|d_A\|_2 \leq c_3 \|F(u)\|_2 \]
with
\[ c_3 := \max_{A,I^+,I^-} \gamma^{-1} \|M^{-1}_{A,A}\|_2 + \gamma^{-1} \|M^{-1}_{A,A} M_{A,I}\|_2 (2(c_1 c_2 + 1)^2 + 1)^{\frac{1}{2}}. \]

Altogether we obtain
\[ \|d\|_2 \leq C \|F(u)\|_2 \]
with \( C := (c_3^2 + 2(c_1 c_2 + 1)^2 + 1)^{\frac{1}{2}}, \) independent of \( u. \)

In our convergence proof, we shall need the continuity of \( \Theta' \).

**Lemma 4.7.** Let \((u^{(j)}, d^{(j)}) \to (u^*, \bar{d}), j \to \infty, \Theta := \Theta_p, p \in \{1, 2\}.\) If \( \Theta \) from (7) fulfills the condition
\[ \lim_{(u,v) \to (u^*,u^*)} \frac{\Theta(u) - \Theta(v) - \Theta'(u^*,u - v)}{\|u - v\|_p} = 0, \quad (40) \]
then the directional derivative \( \Theta'(u, d) \) is continuous at \((u^*, \bar{d})\), as a function of \((u, d)\), i.e.
\[ \lim_{j \to \infty} \Theta'(u^{(j)}, d^{(j)}) = \Theta'(u^*, \bar{d}). \]

**Proof.** Since \( \Theta \) from (7) is locally Lipschitz continuous and directionally differentiable in the level set \( L(u^{(0)}) \) from Proposition 4.5, it follows for \( u \in L \) that \( \Theta'(u, \cdot) \) is a locally Lipschitz continuous function with the same Lipschitz constant as \( \Theta \), cf. [29, Theorem 3.1.2], i.e. it holds
\[ \lim_{j \to \infty} \Theta'(u^*, d^{(j)}) = \Theta'(u^*, \bar{d}). \]

Because of condition (40) of \( \Theta \) at \( u^* \), the double limit
\[ \lim_{j \to \infty, t \searrow 0} \frac{\Theta(u^{(j)} + t d^{(j)}) - \Theta(u^{(j)})}{t} \]
exists and is equal to \( \Theta'(u^*, \bar{d}). \) Additionally it holds that
\[ \lim_{t \searrow 0} \frac{\Theta(u^{(j)} + t d^{(j)}) - \Theta(u^{(j)})}{t} = \Theta'(u^{(j)}, d^{(j)}) \]
for every fixed \( j \in \mathbb{N}. \) Therefore, we conclude that
\[ \lim_{j \to \infty} \Theta'(u^{(j)}, d^{(j)}) = \lim_{j \to \infty} \left( \lim_{t \searrow 0} \frac{\Theta(u^{(j)} + t d^{(j)}) - \Theta(u^{(j)})}{t} \right) \]
\[ = \lim_{j \to \infty, t \searrow 0} \frac{\Theta(u^{(j)} + t d^{(j)}) - \Theta(u^{(j)})}{t} \]
\[ = \Theta'(u^*, \bar{d}). \]
We may now prove our main result on the global convergence of Algorithm 1.

**Theorem 4.8.** Consider the damped B-semismooth Newton method for the finite-dimensional shrinkage equation (2). Let \( \{u^{(j)}\}_j \) be a sequence of iterates produced by Algorithm 1, \( \{t_j\}_j \) the chosen stepsizes and \( p \in \{1, 2\} \), \( \Theta_p := \Theta \).

(i) If \( \limsup_{j \to \infty} t_j > 0 \), then \( u^{(j)} \to u^* \), \( j \to \infty \) with \( \Theta(u^*) = 0 \).

(ii) If \( \limsup_{j \to \infty} t_j = 0 \) and if \( u^* \) is an accumulation point of \( \{u^{(j)}\}_j \) where condition (40) holds at \( u^* \), then \( u^{(j)} \to u^* \), \( j \to \infty \) with \( \Theta(u^*) = 0 \).

**Proof.** We proceed as in [17,21,25,27]. The B-Newton equation (6) has a unique solution due to Proposition 4.1 ensures that the Armijo rule outputs well-defined stepsizes in each step, according to Theorem 3.5. Moreover, Proposition 4.1 ensures that the sequence \( \{ \Theta(u^{(j)}) \}_j \) is bounded below by 0. Therefore, the sequence \( \{ \Theta(u^{(j)}) \}_j \) is convergent and the sequence of differences \( \{ \Theta(u^{(j+1)}) - \Theta(u^{(j)}) \}_j \) tends to 0. The Armijo rule (36) implies

\[
t_j \Theta(u^{(j)}) \leq \frac{\Theta(u^{(j)}) - \Theta(u^{(j+1)})}{p\sigma} \to 0, \quad j \to \infty.
\]

If the chosen stepsizes are bounded away from 0, i.e. \( \limsup_{j \to \infty} t_j > 0 \), it follows that \( u^{(j)} \to u^* \), \( j \to \infty \) with \( \Theta(u^*) = 0 \), proving (i).

Suppose now that \( \limsup_{j \to \infty} t_j = 0 \), which implies that \( \lim_{j \to \infty} t_j = 0 \). Let \( \{ \Theta(u^{(j)}) \}_{j \in J} \) be a subsequence converging to \( u^* \) with \( \Theta(u^*) > 0 \). Let \( t_j = \beta_j \) denote the chosen stepsizes in Algorithm 1. We define \( \tau_j := \beta_j^{j-1} \). The Armijo rule (36) yields

\[
\frac{\Theta(u^{(j)}) + \tau_j d^{(j)} - \Theta(u^{(j)})}{\tau_j} > -p\sigma \Theta(u^{(j)}), \quad j = 0, 1, \ldots
\]

(41)

According to Proposition 4.6, the sequence \( \{d^{(j)}\}_j \) is bounded. Without loss of generality we may suppose that \( d^{(j)} \to \bar{d} \), \( j \to \infty, j \in J \). Passing to the limit in (41), assumption (40) of \( \Theta \) at \( u^* \) yields with Lemma 4.7

\[
\Theta'(u^*, \bar{d}) \geq -p\sigma \Theta(u^*).
\]

Due to (40), the directional derivative \( \Theta' \) as a function of \( (u, v) \) is continuous at \( (u^*, \bar{d}) \), see Lemma 4.7. Therefore,

\[
\Theta'(u^*, \bar{d}) = \lim_{j \to \infty, j \in J} \Theta'(u^{(j)}, d^{(j)}) = \lim_{j \to \infty, j \in J} -p\Theta(u^{(j)}) = -p\Theta(u^*).
\]

 Altogether, we obtain

\[
-p\Theta(u^*) \geq -p\sigma \Theta(u^*),
\]

which is a contradiction to \( \sigma < \frac{1}{p} \leq 1 \) and hence finishes the proof. \( \square \)
Remark 4.9. Assumption (40) on $\Theta$ is only needed in case (ii) of Theorem 4.8. Pang proved in [26, Proposition 1] that if the merit function $\Theta_2$ has no strong $F$-derivative at the accumulation point $u^*$, which is a sufficient condition for (40), then $u^*$ is not a zero of $\Theta_2$. Therefore, this assumption in Theorem 4.8 is necessary.

Concerning the speed of convergence of our B-semismooth Newton scheme, we cite the following corollary from [27, Theorem 4.3, Corollary 4.4].

Corollary 4.10. Let $\{u^{(j)}\}$ be any sequence generated by Algorithm 1 with $p = 2$. Assume that $F(u^{(j)}) \neq 0$ for all $j$. Suppose that $u^*$ is an accumulation point of $\{u^{(j)}\}$. Then $u^*$ is the zero of $F$ if and only if the sequence $\{u^{(j)}\}$ converges to $u^*$ superlinearly and $t_j$ eventually becomes 1. On the other hand, $u^*$ is not the zero of $F$ if and only if either the sequence $\{u^{(j)}\}$ diverges or $\lim_{j \to \infty} t_j = 0$.

The following remark clarifies the naming of our B-semismooth Newton method, by interpreting it as a semismooth Newton method in the sense of [7].

Remark 4.11. Since $F$ is locally Lipschitz continuous, there exists a generalized derivative $G$ with

$$ F'(u, d) = G(u)d $$

(42)

such that the Newton direction from (4) coincides with the B-Newton direction (34), see for example [28, Lemma 2.2, Proposition 3.4]. Here, $G(u)$ is given by the block matrix

$$ G(u) = \begin{pmatrix} \gamma K^*K_{B,B} & \gamma K^*K_{B,C} \\ 0 & I_C \end{pmatrix}, $$

(43)

where

$$ B = B(u) := A(u) \cup \{k \in I^+: x_k > 0 \} $$

$$ C = C(u) := C^o(u) \cup \{k \in I^+: x_k = 0 \} = N \setminus B $$

and $x$ is the solution to the linear complementarity problem (29). With this choice, (42) holds. If $I^+(u) = \emptyset$, then the generalized derivatives $G$ from (8) and (43) coincide. (43) is indeed a generalized derivative of $F$ because

$$ \lim_{h \to 0} \frac{\|F(u + h) - F(u) - G(u + h)h\|_p}{\|h\|_p} \leq \lim_{h \to 0} \frac{\|F(u + h) - F(u) - F'(u, h)\|_p}{\|h\|_p} + \lim_{h \to 0} \frac{\|F'(u, h) - F'(u + h, h)\|_p}{\|h\|_p}. $$

The first limit is equal to 0 because of the B-differentiability of $F$. The second limit is 0 since $F$ is semismooth and locally Lipschitz continuous, see [28, Theorem 2.3].
5. Numerical results

In this section, we discuss numerical results on the application of the damped B-semismooth Newton method from Algorithm 1 to $\ell_1$ Tikhonov regularization of different problems. We shall consider the inverse integration problem [16], the inverse heat equation and deblurring of images [18–20]. In particular, we will analyze the influence of the parameter $\gamma > 0$ on the performance of Algorithm 1. For our computations, we use MATLAB® 2013a and the package regtools from [20]. In the Armijo rule (36), we choose the parameter values $\sigma = 0.01$ and $\beta = 0.5$.

We will work with constant weights $w_k = w_0$, for all $k \in \mathcal{N}$, and $w_0$ is adapted to the measurement errors a posteriori by the discrepancy principle, see also [1, 3]. If not otherwise specified, $p = 2$ is used to define the merit function $\Theta_p$. In fact, our algorithm yields similar results independent of the choice $p \in \{1, 2\}$. In the following examples, the index sets $\mathcal{I}^+, \mathcal{I}^-$ were empty in each iteration. If all stepsizes are chosen equal to 1 and if $\mathcal{I}^+ \cup \mathcal{I}^- = \emptyset$ in each step, then Algorithm 1 coincides with the semismooth Newton iteration (4), (8) from [16]. Because of the sparsity of the solution, the starting vector was chosen as the zero vector in all following computations, but our algorithm performs well also for arbitrary starting vectors. In all examples, the stopping criterion was a residual norm smaller than a tolerance of $10^{-7}$.

5.1. Inverse integration

Here, analogous to [16], we consider the inverse integration problem. Our aim is to find $u(x)$ from measurements $f(x)$ with $Ku(x) = f(x)$ on $[0, 1]$. The operator $K: L^2([0, 1]) \to L^2([0, 1])$ is defined as

$$Ku(x) = \int_0^x u(t) dt.$$ 

$K$ is discretized by

$$K = \frac{1}{N} \begin{pmatrix} 1 & 0 & \ldots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 0 & \vdots \\ 1 & \cdots & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{N \times N}$$

and the data $f = (f(x_k))_{k=1,\ldots,N}$ are given on a grid $\Delta = \{x_k = \frac{k}{N} : k = 1, \ldots, N\}$.

Figure 2 illustrates the reconstruction of $u = (u(x_k))_{k=1,\ldots,N}$ from 5% noisy data $f^\delta$ with the damped B-semismooth Newton method for $p = 1$. Here, the space dimension was $N = 500$, $\gamma = 10^4$ and $w_k = 10^{-2}$ for all $k \in \{1, \ldots, 500\}$. The residual norms in the original semismooth Newton method (4) from [16] might not decrease monotonically, see Table 1 of loc. cit. However, the damped Newton steps in Algorithm 1 effect a strictly monotonic decreasing of the residual norm. Close to the zero of the merit functional, the stepsizes were chosen equal to 1. This fits well the locally superlinear convergence of the B-semismooth Newton method, see Corollary 3.6. One can often observe a remarkable
Global convergence of damped semismooth Newton methods for \( \ell_1 \) regularization

Figure 2. Inverse integration problem with \( N = 500, \gamma = 10^4, w_k = 10^{-2} \) and \( p = 1 \). From left to right and top to bottom: noisy data containing 5% of noise, reconstruction, residual norm, chosen stepsizes.

Table 2. Performance of Algorithm 1, depending on the choice of \( \gamma \) for space dimension \( N = 500, w_k = 10^{-2} \) for all \( k \) and with 5% noisy data \( \mathbf{f}^\delta \) for the example of inverse integration.

<table>
<thead>
<tr>
<th>( \gamma )</th>
<th># iterations</th>
<th>( J(u^{(m)}) )</th>
<th>( |F(u^{(m)})|_2 )</th>
<th>#( { j : t_j = 1 } )</th>
<th>( |F(u^{(0)})|_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 10^2 )</td>
<td>209</td>
<td>7.5758</td>
<td>4.3812e−14</td>
<td>4</td>
<td>7.9459e + 02</td>
</tr>
<tr>
<td>( 10^3 )</td>
<td>122</td>
<td>7.5758</td>
<td>3.1096e−13</td>
<td>7</td>
<td>7.9459e + 03</td>
</tr>
<tr>
<td>( 10^4 )</td>
<td>27</td>
<td>7.5758</td>
<td>2.8093e−12</td>
<td>5</td>
<td>7.9459e + 04</td>
</tr>
<tr>
<td>( 10^5 )</td>
<td>12</td>
<td>7.5758</td>
<td>4.4143e−11</td>
<td>8</td>
<td>7.9459e + 05</td>
</tr>
<tr>
<td>( 10^6 )</td>
<td>15</td>
<td>7.5758</td>
<td>2.9147e−10</td>
<td>7</td>
<td>7.9459e + 06</td>
</tr>
<tr>
<td>( 10^7 )</td>
<td>21</td>
<td>7.5758</td>
<td>4.0261e−09</td>
<td>7</td>
<td>7.9459e + 07</td>
</tr>
<tr>
<td>( 10^8 )</td>
<td>28</td>
<td>7.5758</td>
<td>3.5129e−08</td>
<td>7</td>
<td>7.9459e + 08</td>
</tr>
</tbody>
</table>

decrease of the residual norm in the last step, as mentioned for the semismooth Newton method in [16] and in Section 2.1. If \( p \) is chosen equal to 2 in this example, 12 steps need to be computed to achieve a similar reconstruction if \( \gamma \) is chosen equal to \( 10^5 \), see Table 2.

The influence of the parameter \( \gamma \) on the performance of Algorithm 1 is illustrated in Table 2. The space dimension \( N = 500 \) was fixed and the number \( m \) of iterations,
Table 3. Performance of Algorithm 1, depending on the problem size $N$ for the example of inverse integration, where $\gamma = 10^5$, $w_k = 10^{-2}$ for all $k$. The noisy data $f^\delta$ contained 5\% of noise.

<table>
<thead>
<tr>
<th>$N$</th>
<th># iterations</th>
<th>$|F(u^{(m)})|_2$</th>
<th>$|F(u^{(0)})|_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>9</td>
<td>2.0748e-11</td>
<td>3.6052e+05</td>
</tr>
<tr>
<td>250</td>
<td>10</td>
<td>1.3074e-11</td>
<td>6.7566e+05</td>
</tr>
<tr>
<td>500</td>
<td>12</td>
<td>4.4143e-11</td>
<td>7.9459e+05</td>
</tr>
<tr>
<td>1000</td>
<td>14</td>
<td>3.2354e-11</td>
<td>1.1188e+06</td>
</tr>
<tr>
<td>2500</td>
<td>32</td>
<td>5.0218e-11</td>
<td>1.7675e+06</td>
</tr>
<tr>
<td>5000</td>
<td>38</td>
<td>7.9742e-11</td>
<td>2.4968e+06</td>
</tr>
</tbody>
</table>

Table 4. Influence of the noise level on the performance of Algorithm 1 for the example of inverse integration. Here, $N = 500$ and $\gamma = 10^5$.

<table>
<thead>
<tr>
<th>noise level</th>
<th>$w_k$</th>
<th># iterations</th>
<th>$|F(u^{(m)})|_2$</th>
<th>$|F(u^{(0)})|_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1%</td>
<td>$10^{-4}$</td>
<td>10</td>
<td>2.4803e-11</td>
<td>8.1114e+05</td>
</tr>
<tr>
<td>0.5%</td>
<td>$10^{-3}$</td>
<td>12</td>
<td>4.2488e-11</td>
<td>8.0960e+05</td>
</tr>
<tr>
<td>1%</td>
<td>$10^{-3}$</td>
<td>12</td>
<td>2.2392e-11</td>
<td>8.0996e+05</td>
</tr>
<tr>
<td>2%</td>
<td>$10^{-3}$</td>
<td>12</td>
<td>2.6868e-11</td>
<td>8.1067e+05</td>
</tr>
<tr>
<td>5%</td>
<td>$10^{-2}$</td>
<td>12</td>
<td>4.4143e-11</td>
<td>7.9459e+05</td>
</tr>
<tr>
<td>7.5%</td>
<td>$10^{-2}$</td>
<td>12</td>
<td>2.0516e-11</td>
<td>7.9637e+05</td>
</tr>
<tr>
<td>10%</td>
<td>$10^{-2}$</td>
<td>13</td>
<td>2.3249e-11</td>
<td>7.9815e+05</td>
</tr>
</tbody>
</table>

the functional value at convergence

$$J(u^{(m)}) = \frac{1}{2} \|Ku^{(m)} - f\|_2^2 + \sum_{k=1}^N w_k |u_k^{(m)}|,$$

the residual norm $\|F(u^{(m)})\|_2$, the number of steps with stepsize equal to 1 as well as the residual norm at the initial guess are listed depending on the choice of $\gamma$. For small values of $\gamma$, many iterations need to be computed. Table 2 shows that the choice of $\gamma$ is crucial for the performance of Algorithm 1. In contrast to the semismooth Newton method from [16], see also Section 2, the damped B-semismooth Newton method is guaranteed to converge for an arbitrary choice of $\gamma > 0$.

However, the computational work does not only depend on the number of iterations, but also on the space dimension $N$, the number of active indices of each iterate, the chosen stepsizes and the presence of degenerated indices $k \in I^+(u^{(j)}) \cup I^-(u^{(j)})$ during the iteration. Degenerated indices did not appear in our numerical experiments.

Table 3 demonstrates the influence of the problem size $N$ on the performance of Algorithm 1. We chose $\gamma = 10^5$, $w_k = 10^{-2}$ for all $k \in \{1, \ldots, N\}$ and a noise level of 5\%. The number of iterations only increases very slowly with the problem size.

Moreover, the number $m$ of iterations is almost independent of the noise level for the example of inverse integration, as is demonstrated in Table 4. Here, the problem
size was chosen as $N = 500, \gamma = 10^5$ and the regularization parameters were chosen again by the discrepancy principle.

5.2. Deblurring of images

The next example treats the deblurring of digital images which are degraded by atmospheric turbulence blur, see [18,19]. The mathematical model is a Fredholm integral equation with Gaussian point-spread kernel

$$h(x, y) = \frac{1}{2\pi \rho^2} e^{-\frac{x^2 + y^2}{2\rho^2}}.$$ 

The problem is discretized by a symmetric doubly Toeplitz matrix

$$K = \frac{1}{2\pi \rho^2} T \otimes T \in \mathbb{R}^{N \times N},$$

where $T$ is a symmetric banded Toeplitz matrix with bandwidth 5

$$T = \begin{pmatrix}
1 & e^{-\frac{1}{2\rho^2}} & e^{-\frac{4}{2\rho^2}} \\
&e^{-\frac{1}{2\rho^2}} & \cdots & \cdots & \cdots \\
e^{-\frac{4}{2\rho^2}} & \cdots & \cdots & \cdots & \cdots \\
&\cdots & \cdots & \cdots & \cdots & e^{-\frac{4}{2\rho^2}} \\
e^{-\frac{1}{2\rho^2}} & e^{-\frac{4}{2\rho^2}} & 1
\end{pmatrix} \in \mathbb{R}^{n \times n}$$

with $N = n^2$, see [19].

Figure 3 shows the reconstruction of a test image from [20], from 2% noisy data $f^\delta$. The image contained $N = 200^2$ pixels and we chose $\rho = 0.7, \gamma = 100$ and $w_k = 10^{-1}$ for all $k$. The algorithm terminated after 5 steps with a residual norm of $5.5548 \cdot 10^{-12}$. The residual norm of the starting vector was $1.9122 \cdot 10^4$. All stepsizes were chosen equal.
Global convergence of damped semismooth Newton methods for $\ell_1$ regularization

Figure 4. Solution of the inverse heat equation with 1% noisy data. From left to right: Exact solution, reconstruction for $\kappa = 5$ and $N = 100$, reconstruction for $\kappa = 1$ and $N = 20$.

to one and the index sets $I^\pm$ did not appear throughout the iteration. Therefore, the algorithm coincided with the semismooth Newton method from [16] for this very choice of $\gamma$. For larger parameter values $\gamma$, however, the output of our Algorithm 1 deviates from that of the semismooth Newton method from [16] because the Armijo rule will select stepsizes $t_j < 1$ at least for early iterates.

5.3. Inverse heat equation

The third example treats the inverse heat equation [6,19]. Here, we consider a Volterra integral equation of the first kind on the interval $[0,1]$ with kernel

$$K(s,t) = \frac{(s-t)^{\frac{3}{2}}}{2\kappa\sqrt{\pi}}e^{-\frac{1}{4\kappa^2(s-t)}}.$$ 

For $\kappa = 1$, the inverse heat equation is severely ill-posed. The equation is less ill-posed for larger $\kappa$. The problem is discretized as in [20].

Figure 4 shows the results of Algorithm 1 for the inverse heat equation problem with 1% noisy data. First, we chose $\kappa = 5$. With $N = 100$, $w_k = 10^{-3}$ and $\gamma = 10^4$, the reconstruction was found after 4 steps with a residual norm of $6.2653 \cdot 10^{-13}$. Second, we considered $\kappa = 1$ and $N = 20$. The regularization parameters were chosen as $w_k = 10^{-4}$ by the discrepancy principle. With $\gamma = 10^4$, we obtain a residual norm of $4.2114 \cdot 10^{-14}$ after 6 steps. In these two examples, the B-semismooth Newton iteration was consistent with the semismooth Newton iteration from [16].

6. Discussion

In [16], the locally superlinearly convergent semismooth Newton method for Tikhonov functionals with sparsity constraints was presented in the infinite-dimensional setting. We have discussed a globalized variant of this very method, by introducing damping parameters which ensure the proper descent of the semismooth Newton iterates with respect to a suitable merit functional. The consideration of a Bouligand direction was crucial for our proof of global convergence. Interestingly, the Newton direction
of [16] always coincided with the Bouligand direction in our numerical experiments. More precisely, our method can be interpreted as a semismooth Newton method in the sense of [7], with a generalized derivative that coincides with the generalized derivative from [16] almost everywhere.

Several properties of the proposed B-semismooth Newton method are advantageous: our scheme combines the convergence speed of the semismooth Newton method from [16] with the global convergence property of other state-of-the-art, but still linearly convergent, methods. Additionally, our method yields the exact minimizer of the Tikhonov functional within finite many iterations.

Further research will focus on the analysis of the infinite-dimensional setting, which may shed some light on the robustness with respect to the spatial dimension and on the option to accelerate the convergence even more, e.g., by multigrid strategies. Moreover, the analysis of matrix-free inexact B-semismooth Newton methods will be of interest in the context of $\ell_1$ Tikhonov regularization of large-scale discrete ill-posed problems.

References

[1] Anzengruber and S W Ramlau R 2010 Morozov’s discrepancy principle for Tikhonov-type functionals with nonlinear operators Inverse Problems 26(2) 025001
Global convergence of damped semismooth Newton methods for $\ell_1$ regularization


[22] Loris I 2009 On the performance of algorithms for the minimization of $\ell_1$-penalized functionals *Inverse Probl.* 25(3) 035008


