A semismooth Newton method for Tikhonov functionals with sparsity constraints

R Griesse and D A Lorenz

1 Johann Radon Institute for Computational and Applied Mathematics (RICAM), Austrian Academy of Sciences, Altenbergerstraße 69, A-4040 Linz, Austria
2 Zentrum für Technomathematik, University of Bremen, D-28334 Bremen, Germany

E-mail: roland.griesse@oeaw.ac.at and dlorenz@math.uni-bremen.de

Received 20 September 2007, in final form 4 March 2008
Published 8 April 2008
Online at stacks.iop.org/IP/24/035007

Abstract
Minimization problems in $\ell^2$ for Tikhonov functionals with sparsity constraints are considered. Sparsity of the solution is ensured by a weighted $\ell^1$ penalty term. The necessary and sufficient condition for optimality is shown to be slantly differentiable (Newton differentiable), hence a semismooth Newton method is applicable. Local superlinear convergence of this method is proved. Numerical examples are provided which show that our method compares favorably with existing approaches.

1. Introduction

In this work, we consider the optimization problem

$$\text{Minimize } \frac{1}{2} \| Ku - f \|_H^2 + \sum_{k=1}^{\infty} w_k |u_k| \quad \text{over } u \in \ell^2. \quad (1)$$

Here, $K : \ell^2 \to \mathcal{H}$ is a linear and injective operator mapping the sequence space $\ell^2$ into a Hilbert space $\mathcal{H}$, $f \in \mathcal{H}$ and $w = (w_k)$ is a sequence satisfying $w_k \geq w_0 > 0$.

One well-understood algorithm for the solution of (1) is the so-called iterated soft-thresholding for which convergence has been proven in [10] (see also [2, 9]). While the iterated soft-thresholding is very easy to implement it converges slowly in practice (in fact the method converges linearly but with a constant very close to 1 [2]). Another well-analyzed method is the iterated hard-thresholding which converges like $O(n^{-1/2})$ [3] (i.e. even slower than the iterated soft-thresholding but practically it is faster in many cases).

In this paper, we derive an algorithm for which we prove local superlinear convergence in the infinite-dimensional setting. Our algorithm is an active set, or semismooth Newton, method, and hence the analysis is based on the notion of slant differentiability [8, 16]. The semismooth Newton method is easily implementable as an active set method. Numerical
experiments show that the method is robust with respect to the choice of the initial value and that it compares favorably with existing approaches in terms of computation time.

The background for problems of type (1) is, for example, the attempt to solve the linear operator equation $Ku = f$ in an infinite-dimensional Hilbert space which models the connection between some quantity of interest $u$ and some measurements $f$. Often, the measurements $f$ contain noise which makes the direct inversion ill-posed and practically impossible. Thus, instead of considering the linear equation, a regularized problem is posed for which the solution is stable with respect to noise. A common approach is to regularize by minimizing a Tikhonov functional [10, 13, 21]. A special class of these regularizations has been of recent interest, namely of the type (1). These problems model the fact that the quantity of interest $u$ is composed of a few elements, i.e. it is sparse in some given, countable basis. To make this precise, let $A : \mathcal{H}_1 \to \mathcal{H}_2$ be a bounded operator between two Hilbert spaces and let $(\psi_k)$ be an orthonormal basis of $\mathcal{H}_1$. Denote by $B : \ell^2 \to \mathcal{H}_1$ the synthesis operator $B(u_k) = \sum_k u_k \psi_k$. Then the problem

$$\min_{u \in \mathcal{H}_1} \frac{1}{2} \|Au - f\|_{\mathcal{H}_2}^2 + \sum_{k=1}^{\infty} w_k |\langle u, \psi_k \rangle|$$

can be rephrased as

$$\min_{u \in \ell^2} \frac{1}{2} \|ABu - f\|_{\mathcal{H}_2}^2 + \sum_{k=1}^{\infty} w_k |u_k|.$$  

The sequence $w_k$ plays the role of the regularization parameter where each coefficient is regularized individually. However, for an analysis of the regularizing properties one might use $\alpha w_k$ instead and investigate $\alpha \to 0$. We refer to, e.g., [10, 18, 20] for analysis of the regularizing properties and parameter choice rules.

Recently, sparsity constraints have also appeared in the context of optimal control of PDEs [24].

The paper is organized as follows. In section 2, we derive a semismooth formulation for the minimization problem (1). Section 3 states the algorithm and local superlinear convergence is proven. Section 4 presents numerical results on the regularization of the ill-posed problems of inverse integration and deblurring and shows an application to $\ell^1$ minimization in the context of compressed sensing.

Notation. For $1 \leq p < \infty$, $\ell^p$ denotes the space of $p$-summable sequences with norm $\|u\|_p = \left( \sum_{k=1}^{\infty} |u_k|^p \right)^{1/p}$, whereas $\ell^\infty$ denotes the space of bounded sequences with norm $\|u\|_\infty = \max_{k \in \mathbb{N}} |u_k|$. Recall that these spaces satisfy $\ell^p \hookrightarrow \ell^q$ for $1 \leq p \leq q \leq \infty$ and that $|u_k| \leq \|u\|_p$ holds for any $u \in \ell^p$. In the case $p = 2$ we simply write $\|u\|_2$ and $\langle \cdot, \cdot \rangle$ denotes the inner product in $\ell^2$. With $B_\rho(u)$ we denote the open ball of radius $\rho$ with respect to the norm of $\ell^2$, centered at $u$. The operator $K^* : \mathcal{H} \to \ell^2$ is the Hilbert space adjoint of $K$ and $L(X, Y)$ is the space of bounded linear operators from $X$ to $Y$.

2. Optimality conditions

In this section, we are going to derive the necessary and sufficient optimality condition for the problem (1). It is going to be the basis for the semismooth Newton algorithm. This condition can be derived and expressed in different ways, for example by using the classical Lagrange duality, or by using subgradient calculus.

Let us first address the conditions obtained by subgradient calculus. To this end we introduce the so-called soft-thresholding function.
Definition 2.1. Let \( w = (w_k) \) with \( w_k \geq w_0 > 0 \) and \( 1 \leq p < \infty, 1 \leq q \leq \infty \). The soft-thresholding of \( u \) with the sequence \( w \) is defined as the mapping \( S_w : \ell^p \to \ell^q \) given by
\[
S_w(u)_k = S_{w_k}(u_k) = \max\{0, |u_k| - w_k\} \sgn(u_k).
\]
(2)

Remark 2.2. Since elements of \( \ell^p \) are sequences converging to zero, the range of \( S_w \) is \( \ell^0 = \{ u \in \mathbb{R}^\mathbb{N} : u_k = 0 \text{ for almost every } k \} \subset \ell^q \).

With the help of the soft-thresholding operator, we can formulate the optimality condition in a compact way.

Proposition 2.3. If \( K : \ell^2 \to \mathcal{H} \) is injective, the functional
\[
\Psi(u) = \frac{1}{2} \| Ku - f \|_\mathcal{H}^2 + \sum_{k=1}^\infty w_k |u_k|
\]
has a unique minimizer \( \bar{u} \in \ell^2 \). This minimizer is characterized by
\[
\bar{u} = S_{\gamma w} (\bar{u} - \gamma K^*(K \bar{u} - f)) \quad \text{for any } \gamma > 0.
\]
(4)

Proof. Since \( K \) is injective, \( \Psi \) is strictly convex and coercive, and hence it has a unique minimizer. This minimizer is characterized by
\[
0 \in \partial \Psi(\bar{u}),
\]
which is equivalent to
\[
-K^*(K \bar{u} - f) \in \partial F(\bar{u}),
\]
(5)
where \( F(u) = \sum_k w_k |u_k| \). Multiplying with \( \gamma > 0 \), adding \( \bar{u} \) to both sides and inverting \((I + \gamma \partial F)\) gives
\[
\bar{u} = (I + \gamma \partial F)^{-1}(\bar{u} - \gamma K^*(K \bar{u} - f)).
\]
(Note that \((I + \gamma \partial F)^{-1}\) exists and is single-valued since the subgradient \( \partial F \) is maximal monotone if \( F \) is convex and lower semicontinuous [26, proposition 32.17, corollary 32.30].)

A straightforward calculation shows that
\[
(I + \gamma \partial F)^{-1} = S_{\gamma w}.
\]
From the characterization (4) and remark 2.2 we can derive the following corollary.

Corollary 2.4. The minimizer \( \bar{u} \) of (3) is a finitely supported sequence.
The extremality conditions [12, chapter III.4] are
\[
F(u) + F^*(K^* p) - (K^* p, u) = 0 \tag{7a}
\]
\[
G(Ku) + G^*(-p) + (p, Ku) = 0. \tag{7b}
\]

The first condition (7a) yields
\[
\sum_{k=1}^{\infty} w_k |u_k| - (K^* p, u) = 0 \quad \text{and} \quad |(K^* p)_k| \leq w_k \\
\Leftrightarrow \quad u_k = 0 \quad \text{or} \quad (K^* p)_k = w_k \sign u_k = \begin{cases} w_k, & \text{if } u_k > 0 \\ -w_k, & \text{if } u_k < 0 \end{cases}
\]
and \( |(K^* p)_k| \leq w_k \).

This condition can be written as the complementarity system
\[
K^* p - w \leq 0, \quad u^* \geq 0, \quad [K^* p - w]u^* = 0
\]
\[
-K^* p - w \leq 0, \quad u^- \geq 0, \quad [K^* p + w]u^- = 0 \tag{8}
\]
in a coordinatewise sense, which is in turn equivalent to
\[
u = \max\{0, u + \gamma (K^* p - w)\} + \min\{0, u + \gamma (K^* p + w)\} \tag{9}
\]
for any \( \gamma > 0 \).

The second condition (7b) yields
\[
\frac{1}{2} \| Ku - f \|_H^2 + \frac{1}{2} \| -p \|_H^2 + (-p, f) + (p, Ku) = \frac{1}{2} \| Ku - f + p \|_H^2 = 0 \tag{10}
\]
and thus
\[
Ku - f + p = 0.
\]

By plugging (10) into (9) we end up with
\[
u - \max\{0, u - \gamma (K^* (Ku - f) + w)\} - \min\{0, u - \gamma (K^* (Ku - f) - w)\} = 0, \tag{11}
\]
which is just another way to express (4).

**Remark 2.5.** The usual characterization \( 0 \in \partial \Psi(\bar{u}) \) of the unique minimizer \( \bar{u} \) of (1) is difficult to handle for numerical algorithms because it is a nonsmooth inclusion. One attempt to tackle the problem is by interior point regularization as proposed in [17]. This, however, introduces additional nonlinearities into the problem. By contrast, our algorithm is based on the necessary and sufficient condition (11). As we shall prove in the following section, (11) is a semismooth equation in \( \ell^2 \), so that Newton’s method can be applied.

### 3. The semismooth Newton method

The previous section has shown that we can solve the minimization problem (1) by solving equation (4) or (11), or briefly
\[
F(u) = u - S_{\gamma w}(u - \gamma K^*(Ku - f)) = 0, \tag{12}
\]
for some \( \gamma > 0 \).

This is an operator equation in the space \( \ell^2 \), involving the non-differentiable max and min operations. Optimality conditions of this form also frequently occur in the context of optimal control problems for partial differential equations, in the presence of control...
constraints. Then (12) is considered in $L^p$ function spaces, and it is known that the max operation, i.e., $u \mapsto \max\{0, u\}$, is so-called Newton or slantly differentiable from $L^p$ to $L^q$ for $1 \leq q < p \leq \infty$, see [8, theorem 2.6] in view of its Lipschitz continuity. In the presence of a norm gap $1 \leq q < p \leq \infty$, the generalized derivative, or slanting function, can be chosen as an indicator function, see [16, proposition 4.1]. This allows for the interpretation of the generalized Newton method as a so-called active set method. This norm gap is made up for in the context of the partial differential equation because $K$ and $K^*$ are solution operators which provide the necessary smoothing.

It turns out that the behavior of the max and min operations is more intricate than in function space. Again, it follows from the Lipschitz continuity of $u \mapsto \max\{0, u\}$ from $\ell^p$ to $\ell^q$ that slant differentiability holds [8, theorem 2.6] for $1 \leq p, q \leq \infty$. However, we are not aware of any simple slanting function even with norm gap which can be algorithmically exploited (see remark 9). It may be surprising that nonetheless, the soft-thresholding operator $S_w$ and thus equation (12) are slantly differentiable and admit a simple slanting function between any pair of $\ell^p, \ell^q$ spaces (see proposition 3.3). This allows us to apply a generalized Newton’s method to solve (12), which takes the form of an active set method.

3.1. Semismoothness of the optimality condition

The concept of slant or Newton differentiability is closely related to the notion of semismoothness [8, 16, 25], and we will use the terms interchangeably.

Definition 3.1. Let $X$ and $Y$ be Banach spaces and $D \subset X$ be an open subset. A mapping $F : D \to Y$ is called Newton (or slantly) differentiable in $x \in D$ if there exists a family of mappings $G : D \to L(X, Y)$ such that
\[
\lim_{h \to 0} \frac{\|F(x + h) - F(x) - G(x + h)h\|_Y}{\|h\|_X} = 0.
\] (13)

The function $G$ is called a generalized derivative (or slanting function) for $F$ in $x$.

It is shown in [8] that any Lipschitz continuous function is Newton differentiable. However, this is only of little help algorithmically unless there is a generalized derivative $G(u)$ of (12) which is easily invertible.

Remark 3.2. A natural candidate for a generalized derivative $G$ of the function $F(u) = \max\{0, u\}$ is
\[
G(u)(h)_k = \begin{cases} h_k, & u_k > 0 \\ \delta h_k, & u_k = 0 \\ 0, & u_k < 0 \end{cases}
\]
for any $\delta \in \mathbb{R}$.

We are going to show that this $G$ cannot serve as a generalized derivative of $F : \ell^p \to \ell^q$ for any $p \in [1, \infty]$ and $1 \leq q \leq \infty$. We consider a point $u \in \ell^p$ for which the set $\{n | u_n \neq 0\}$ is infinite and take a special sequence $h^n \in \ell^p$, namely
\[
h^n_k = \begin{cases} 0 & \text{for } k \neq n \\ -2u_k & \text{for } k = n. \end{cases}
\]

Hence, we have $\|h^n\|_p = 2|u_n| \to 0$ for $n \to \infty$. It is an easy calculation to see that
\[
\frac{\|\max\{u + h^n, 0\} - \max\{u, 0\} - G(u + h^n)h^n\|_q}{\|h^n\|_p} = \frac{1}{2}
\]
for all $n$ with $u_n \neq 0$. 

5
The following proposition shows that the thresholding operator (2) is Newton differentiable and that a function similar to $G$ serves as a generalized derivative.

**Proposition 3.3.** The mapping $S_w : \ell^p \to \ell^q$ from definition 2.1 is Newton differentiable for any $1 \leq p < \infty$, $1 \leq q \leq \infty$. A generalized derivative is given by

$$ (G(u)v)_k = \begin{cases} v_k & \text{for } |u_k| > w_k \\ 0 & \text{for } |u_k| \leq w_k. \end{cases} $$

**Proof.** Without loss of generality we may assume $\|h\|_p < \frac{w_k}{2}$ and hence $|h_k| < \frac{w_k}{2}$. Since $u \in \ell^p$ with $p < \infty$ there exists $k_0$ such that $|u_k| < \frac{w_k}{2}$ for $k > k_0$. We estimate

$$ \|S_w(u + h) - S_w(u) - G(u + h)(h)\|_q $$

$$ = \sum_{k=1}^{\infty} |S_{w_k}(u_k + h_k) - S_{w_k}(u_k) - G(u + h)(h)_k|^q $$

$$ = \sum_{k \leq k_0 \atop |u_k| > w_k} |S_{w_k}(u_k + h_k) - S_{w_k}(u_k) - G(u + h)(h)_k|^q. $$

It is easy to check that the above sum is zero for $\|h\|_p < \min\{|u_k| - w_k : k \leq k_0 \text{ and } |u_k| \neq w_k\}$ because $|h_k| < \|h\|_p$ holds. It follows that

$$ \|S_w(u + h) - S_w(u) - G(u + h)(h)\|_q = 0 $$

for $\|h\|_p$ small enough, which proves Newton differentiability. \qed

**Remark 3.4.** In matrix notation we can express the generalized derivative $G(u)$ as

$$ G(u) = \begin{pmatrix} I_A & 0 \\ 0 & 0 \end{pmatrix}, $$

where $A = \{k \in \mathbb{N} : |u_k| > w_k\}$. 

To calculate a generalized derivative for the mapping $F$ in (12), we prove a chain rule for the generalized derivative.

**Lemma 3.5.** Let $S : X \to Y$ be Newton differentiable, $A \in L(X, X)$ and $y \in X$. Let furthermore $\bar{G}$ be a generalized derivative of $S$. Define $T(u) = S(Au + y)$. Then $H(u) = \bar{G}(Au + y)A$ is a generalized derivative of $T$.

**Proof.** It holds

$$ \|T(u + h) - T(u) - H(u + h)h\| $$

$$ = \frac{\|S(Au + Ah + y) - S(Au + y) - \bar{G}(Au + Ah + y)Ah\|}{\|Ah\|} \cdot \frac{\|Ah\|}{\|h\|}. $$

The right-hand side converges to zero because $\bar{G}$ is a generalized derivative of $S$ in $Au + y$ in the direction $Ah$, and $A$ is bounded. \qed

In order to specify a generalized derivative of $F$, we introduce the active and the inactive sets. For the sake of simplicity we will restrict ourself to the case $F : \ell^2 \to \ell^2$ in the following:
**Definition 3.6.** For \( u \in \ell^2 \), the active set \( A(u) \) and the inactive set \( I(u) \) are given by

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

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

Whenever the active and inactive sets correspond to an iterate \( u^n \), we will denote them by \( A_n \) and \( I_n \), respectively. We will drop the subscript or the argument if no ambiguity can occur.

We are now in the position to calculate a generalized derivative of \( F \).

**Proposition 3.7.** The mapping \( F : \ell^2 \rightarrow \ell^2 \),

\[
F(u) = u - S_{\gamma w}(u - \gamma K^*(Ku - f))
\]

is Newton differentiable. Denote the active and inactive set \( A \) and \( I \) as in definition 3.6 and split the operator \( K^*K \) according to

\[
K^*K = \begin{pmatrix}
M_{AA} & M_{AI} \\
M_{IA} & M_{II}
\end{pmatrix}.
\]

Then a generalized derivative is given by

\[
G(u) = \begin{pmatrix}
0 & 0 \\
0 & I_2
\end{pmatrix} + \begin{pmatrix}
I_A & 0 \\
0 & 0
\end{pmatrix}(\gamma K^*K) = \begin{pmatrix}
\gamma M_{AA} & \gamma M_{AI} \\
0 & I_2
\end{pmatrix}. \tag{14}
\]

**Proof.** The claim follows from the sum rule for the generalized derivative and from proposition 3.3 and lemma 3.5 with \( S = S_{\gamma w}, A = I - \gamma K^*K, y = \gamma K^*f \).

**Remark 3.8.** Note that for any \( u \in \ell^2 \), the active set \( A \) is always finite, since \( u - \gamma K^*(Ku - f) \in \ell^2 \) holds and thus \( |u - \gamma K^*(Ku - f)|_k \rightarrow 0 \) for \( k \rightarrow \infty \).

### 3.2. Semismooth Newton method

The semismooth or generalized Newton method for the solution of (12) can be stated as the iteration

\[
u^{n+1} = u^n - G(u^n)^{-1}F(u^n), \tag{15}
\]

where \( G \) is a generalized derivative of \( F \). We use the generalized derivative \( G \) given by (14) and state the method as algorithm 1. Naturally, the semismooth Newton method can also be interpreted as an active set method, and we state it as algorithm 2.

**Remark 3.9.**

(i) Algorithm 1 is the generalized Newton method (15). The unique solvability in step 8 is shown in proposition 3.11.

(ii) Given an initial iterate \( u^0 \in \ell^2 \), the algorithm is well defined, and all iterates remain in \( \ell^2 \). We refer again to proposition 3.11.

(iii) At the end of step 10, the iterate \( u^{n+1} \) satisfies \( u^{n+1}_I = 0 \). Note that \( r^n_L = u^n_L \) holds which implies \( \delta u^n_L = -u^n_L \).

Note that (iii) implies that all iterates \( u^n (n \geq 1) \) of algorithm 1 are finitely supported sequences. However, \( K^*(Ku^n - f) \) is in general not finitely supported, and hence in a practical implementation, this term will be truncated after a number of entries.
Algorithm 1. Semismooth Newton method for the solution of (12).

1: Initialize $u^0$, choose $\gamma > 0$, set $n := 0$ and not done := false
2: while $n < n_{\text{max}}$ and not done do
3: Calculate the active and inactive sets:
   \[ A = \{ k \in \mathbb{N} : |u^n_k - \gamma K^*(Ku^n - f)_k| > \gamma w_k \} \]
   \[ I = \{ k \in \mathbb{N} : |u^n_k - \gamma K^*(Ku^n - f)_k| \leq \gamma w_k \} \]
4: Compute the residual
   \[ r^n = F(u^n) = u^n - S_{\gamma w}(u^n - \gamma K^*(Ku^n - f)) \]
5: if $||r^n|| \leq \varepsilon$ then
6:   done := true
7: else
8: Calculate the Newton update by solving
   \[ \begin{pmatrix} \gamma M_A A & \gamma M_A I \\ 0 & I_T \end{pmatrix} \begin{pmatrix} \delta u_A \\ \delta u_I \end{pmatrix} = - \begin{pmatrix} r^n_A \\ r^n_I \end{pmatrix} \]
9: Update $u^{n+1} := u^n + \delta u$
10: Set $n := n + 1$
11: end if
12: end while

3.3. Active set method

One may set up algorithm 1 equivalently as an active set method. This can be seen by a closer analysis of the Newton step (steps 8 and 9 in algorithm 1),

\[ u^{n+1} = u^n - \left( \gamma M^{-1}_A A + \gamma M^{-1}_A I \right) \left( u^n - S_{\gamma w}(u^n - \gamma K^*(Ku^n - f)) \right) \]
\[ = u^n - \left( \gamma M^{-1}_A A + \gamma M^{-1}_A I \right) \left( \gamma [K^*(Ku^n - f)]_A \pm w_A \right) \]
\[ = \left( u^n_A - M^{-1}_A (K^*(Ku^n - f)_A \pm w_A - M_A I u^n_I) \right) \]
\[ = \left( M^{-1}_A (K^* f \pm w)_A \right) \]

The sign of $w$ depends of the sign of $u^n - \gamma K^*(Ku^n - f)$. Hence, instead of calculating the Newton update in step 8, one may set $u^{n+1}_I := 0$ and solve $M_A u^{n+1}_A = (K^* f \pm w)_A$. This shows that the subsequent iterate $u^{n+1}$ depends on the current iterate $u^n$ solely through the active set $A$. As a consequence, differing values of $u^n$ can lead to the same next iterate $u^{n+1}$.

For completeness, we state the active set method as algorithm 2. Note that the algorithm is initialized with active sets $A^+$ and $A^-$ instead of $u^0$. 

8
Algorithm 2. Active set method for the solution of (12).

1: Initialize $A_0^+, A_0^-$, choose $\gamma > 0$, set $n := 0$ and done := false
2: Set $A_0 = A_0^+ \cup A_0^-$, $I_0 = \mathbb{N} \setminus A_0$
3: Set the signs of the weights:
   \[
   s_k^0 = \begin{cases} 
   1, & k \in A_0^+ \\
   0, & k \in I_0 \\
   -1, & k \in A_0^-
   \end{cases}
   \]
4: while $n < n_{\text{max}}$ and not done do
5: Set $u_n^I = 0$ and calculate $u_n^A$ by solving
   \[
   M_{A_n^+} u_n^{A_n} = (K^* f - s^0 w)_{A_n}
   \]
6: Calculate the new active sets:
   \[
   A_{n+1}^+ = \{k \in \mathbb{N} : |u^n - \gamma K^* (Ku^n - f)_k| \geq \gamma w_k\}
   
   A_{n+1}^- = \{k \in \mathbb{N} : |u^n - \gamma K^* (Ku^n - f)_k| \leq \gamma w_k\}
   
   I_{n+1} = \{k \in \mathbb{N} : |u^n - \gamma K^* (Ku^n - f)_k| \leq \gamma w_k\}.
   \]
7: Set the signs of the weights:
   \[
   s_{n+1}^k = \begin{cases} 
   1, & k \in A_{n+1}^+ \\
   0, & k \in I_{n+1} \\
   -1, & k \in A_{n+1}^-
   \end{cases}
   \]
8: if $s_{n+1} = s^n$ then
done := true
9: end if
10: Set $n := n + 1$
11: end while

In this setting, the stopping criterion is coincidence of the active sets in consecutive iterations—other choices are also possible. In the numerical examples in section 4 we chose the norm of the residual because a sudden drop of the residual norm occurred before the minimizer was identified.

3.4. Local convergence of the semismooth Newton method

The local superlinear convergence of the semismooth Newton method (algorithm 1) hinges upon the uniform boundedness of $G(u^n)^{-1}$ during the iteration.

**Proposition 3.10.** There exists $k_0 \in \mathbb{N}$ and $\rho > 0$ such that $\|u - \overline{u}\|_2 < \rho$ implies that
   \[
   A(u) \subset [1, k_0].
   \]
Moreover, $k_0$ and $\rho$ depend only on $\gamma, \overline{u}, \|K^* K\|, \|K^* f\|$ and $w_0$.

**Proof.** The triangle inequality implies
   \[
   |u - \gamma K^* (Ku - f)|_k \leq |\overline{u} - \gamma K^* (K\overline{u} - f)|_k + |u - \overline{u} - \gamma K^* (K(u - \overline{u}))|_k.
   \]
The first term can be estimated by
   \[
   |\overline{u} - \gamma K^* (K\overline{u} - f)|_k \leq |\overline{u}|_k + \gamma |K^* K|_k + \gamma |K^* f|_k.
   \]
Since \( \pi, K^*K \pi \) and \( K^*f \) are in \( \ell^2 \), the right-hand side converges to 0 as \( k \to \infty \). In particular, there exists \( k_0 \), depending only on the named quantities, such that
\[
|\bar{u} - \gamma K^*(K \bar{u} - f)|_k \leq \gamma w_0 / 2 \quad \text{for all} \quad k \geq k_0. \tag{17}
\]
The second term in (16) can be estimated by
\[
|u - \bar{u} - \gamma K^*K(u - \bar{u})|_k \leq |u - \bar{u}|_k + \gamma |K^*K(u - \bar{u})|_k \\
\leq \|u - \bar{u}\| \gamma \|K^*K(u - \bar{u})\| \leq (1 + \gamma \|K^*K\|)\|u - \bar{u}\|.
\]
Hence there exists \( \rho > 0 \), depending only on the named quantities, such that
\[
|u - \bar{u} - \gamma K^*K(u - \bar{u})|_k \leq \gamma w_0 / 2 \quad \text{for all} \quad k \in \mathbb{N}. \tag{18}
\]
Combining (16)–(18) proves the claim.

At this point, we cannot yet conclude that the active sets remain uniformly bounded during the iteration of algorithm 1, since it is not evident whether the iterates will remain in a suitable \( \rho \)-neighborhood of \( \bar{u} \).

**Proposition 3.11.** The generalized derivative \( G \), given by (14), is boundedly invertible from \( \ell^2 \) into \( \ell^2 \). Moreover, the norm of \( G(u)^{-1} \) can be estimated by
\[
\|G(u)^{-1}\| \leq \|M_A^{-1}\| \left( \frac{1}{\gamma} + \|M_I\| \right) + 1,
\]
where \( A \) and \( I \) are the active and inactive sets at \( u \), see definition 3.6.

**Proof.** Let \( u, r \in \ell^2 \) and consider the equation \( G(u) \delta u = r \), i.e.,
\[
\begin{pmatrix} \gamma M_{AA} & \gamma M_{AI} \\ 0 & I_I \end{pmatrix} \begin{pmatrix} \delta u_A \\ \delta u_I \end{pmatrix} = \begin{pmatrix} r_A \\ r_I \end{pmatrix}.
\]
Necessarily, \( \delta u_I = r_I \) holds, which implies \( \delta u_I \in \ell^2 \). It remains to solve
\[
\gamma M_{AA} \delta u_A = r_A - \gamma M_{AI} \delta u_I. \tag{19}
\]
The right-hand side is an element of \( \ell^2 \). Moreover, \( M_{AA} \) is injective. (We rewrite \( M_{AA} = P_A K^* K P_A = (K P_A)^* K P_A \), where \( P_A \) is the projection of \( \ell^2 \) onto the active set. Then \( M_{AA} u_A = 0 \) implies \( \|K P_A u_A\|^2 = \langle u, M_{AA} u \rangle = 0 \), and hence \( u_A = 0 \) since \( K \) is injective.) By remark 3.8, the active set is finite, and thus \( M_{AA} \) is an injective operator on a finite-dimensional space, hence it is also surjective. We conclude that (19) has a unique solution \( \delta u_A \in \ell^2 \), hence \( G(u)^{-1} : \ell^2 \to \ell^2 \) exists.

The norm estimate follows from
\[
\|G(u)^{-1}r\| = \left\| \begin{pmatrix} \frac{1}{\gamma} M_A^{-1} & - M_A^{-1} M_I \\ 0 & I_I \end{pmatrix} \begin{pmatrix} r_A \\ r_I \end{pmatrix} \right\| \\
\leq \frac{1}{\gamma} \|M_A^{-1}\| \|r_A\| + \|M_A^{-1}\| \|M_I\| \|r_I\| + \|r_I\| \\
\leq \left( \|M_A^{-1}\| \left( \frac{1}{\gamma} \|M_I\| + 1 \right) \right) \|r\|. \tag{20}
\]

**Corollary 3.12.** Let \( k_0 \in \mathbb{N} \) and \( \rho > 0 \) be as in proposition 3.10. Then \( G(u)^{-1} \) is uniformly bounded on \( B_\rho(\bar{u}) \).

**Proof.** Let \( u \in \ell^2 \) such that \( \|u - \bar{u}\| < \rho \). By proposition 3.10, the active set satisfies \( A(u) \subset [1, k_0] \). Our plan is to show that \( \|G(u)^{-1}\| \) indeed depends only on \( k_0 \). Indeed, we
define
\[ C(k_0) := \max_{\emptyset \neq A \subset [1, k_0]} \| \mathcal{M}_{AA}^{-1} \| > 0. \]

Note that for every \( A \subset [1, k_0], A \neq \emptyset, \) \( \mathcal{M}_{AA} \) is boundedly invertible, hence \( C(k_0) \) is the maximum of finitely many positive numbers. Moreover, \( \mathcal{M}_{AI} \) is obtained from \( K^*K \) by restriction and extension, hence \( \| \mathcal{M}_{AI} \| \leq \| K^*K \| \) holds, for all choices of \( A \) and \( I \). From proposition 3.11, we conclude that
\[ \| G(u)^{-1} \| \leq C(k_0) \left( \frac{1}{\gamma} + \| K^*K \| \right) + 1. \]

We may now combine the results above to argue the local superlinear convergence of algorithm 1.

**Theorem 3.13.** There exists a radius \( r \in (0, \rho] \) such that \( \| u^0 - \varpi \| < r \) implies that all iterates of algorithm 1 satisfy \( \| u^n - \varpi \| < r, \) and \( u^n \to \varpi \) superlinearly.

**Proof.** By corollary 3.12, the inverse of the generalized derivative, \( G(u)^{-1} \), remains uniformly bounded in \( B_{\rho}(\varpi) \). The result is then a standard conclusion for generalized Newton methods (see [7, remark 2.7] or [16, theorem 1.1]).

**Remark 3.14.**
(i) The neighborhood in which superlinear convergence occurs is unknown and may be small. The global convergence behavior of the algorithm thus deserves further investigation. The numerical experiments in the following section suggest that the choice of \( \gamma \) is essential in achieving convergence from a bad initial guess. For a related problem in Hilbert spaces with a standard Tikhonov regularization term \( \| u \|_2 \), global convergence without rates was proved in [22].

(ii) The proof of proposition 3.3 together with the chain rule (lemma 3.5) shows that the remainder
\[ F(u^n) - F(\varpi) - G(\varpi)(u^n - \varpi) \]

is exactly zero for sufficiently small \( \| u^n - \varpi \| \). Hence we expect convergence in one step sufficiently close to the solution, which is confirmed by the numerical results in the following section.

**Remark 3.15.** The assumption on the injectivity of \( K \) may be relaxed. The proof of corollary 3.12 shows that we only need that all submatrices \( \mathcal{M}_{AA} \) for \( A \subset [1, k_0] \) are invertible. Hence, local superlinear convergence can also be proved when the \( K \) satisfies the finite basis injectivity (FBI) property [2]. The FBI property states that the operator \( K \) when restricted to any finite number of coefficients is injective. The FBI property is related to the so-called restricted isometry property (RIP) (see, e.g., [1]), which plays an important role in the analysis of minimizers of \( \ell^1 \) constrained problems in the theory of compressed sensing [5].

### 4. Numerical results

In this section, we present results of numerical experiments illustrating the performance of the semismooth Newton (SSN) method. We implemented the SSN method in MATLAB® and made experiments on a desktop PC with an AMD Athlon® 64 X2. Moreover, we are going to compare the SSN method to other state-of-the-art methods for the minimization of \( \ell^1 \) constrained problems, namely the GPSR methods [14] and the \( 11 \_\_ \_ \_s \) toolbox [17] where we
used the freely available MATLAB® implementations of these methods. The GPSR method is based on the gradient projection method with Barzilai–Borwein stepsizes and is known to converge r-linearly. The ℓ₁ ls method is a truncated Newton interior point method which is applied directly to the objective functional (note that we apply a Newton method to a reformulated optimality condition). In addition, we included the widely used iterative soft-thresholding from [10] in our comparison. Note that both the GPSR and the ℓ₁ ls methods are set up and analyzed in a finite-dimensional setting while our analysis on the SSN as well as the analysis for the iterative soft-thresholding is infinite dimensional.

4.1. Inverse integration

The problem under consideration is the classical ill-posed problem of inverse integration (or differentiation [3, 15, 23]), i.e. the operator $K : L^2([0, 1]) \rightarrow L^2([0, 1])$ given by

$$Ku(t) = \int_0^t u(s) \, ds, \quad t \in [0, 1].$$

The data $f$ are given as $(f(t_k))_{k=1,...,N}$ with $t_k = \frac{k}{N}$. We discretized the operator $K$ by the matrix

$$K = \frac{1}{N} \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ 1 & \cdots & \cdots & 1 \end{pmatrix}, \quad K : \mathbb{R}^N \rightarrow \mathbb{R}^N.$$

The minimization problem reads

$$\min_{u \in \mathbb{R}^N} \frac{1}{2} \sum_{i=1}^N ((Ku)_i - f^\delta_i)^2 + \sum_{k=1}^N w_k|u_k|. \quad (20)$$

One can check easily that the SSN method is also applicable in finite dimensions, and hence this minimization problem can be treated by the SSN method. The discussion of the SSN method in infinite dimension provides us with results which are independent of the dimension $N$, i.e. the algorithm scales well.

The true solution $\bar{u}$ is given by small plateaus and hence the data $f^\delta = K\bar{u} + \delta$ are a noisy function with steep linear ramps. Figure 1 shows our sample data and the result of the ℓ₁ minimization with the SSN method. Table 1 shows how the SSN method performed in this specific example. It can be observed that the residual is not decaying monotonically and it
Table 1. Illustration of the performance of the SSN method for the inverse integration problem. The second column shows the decay of the function value $\Psi$ while the third column shows the norm of the residual. The data are the same as in figure 1.

| $n$ | $\Psi(u^n)$ | $\|r^n\|$ |
|-----|-------------|------------|
| 1   | 1.3249e+01 | 6.9764e+05 |
| 2   | 1.0461e+01 | 2.1698e+02 |
| 3   | 5.3849e+00 | 2.9586e+02 |
| 4   | 4.8393e+00 | 8.3922e+02 |
| 5   | 4.2488e+00 | 1.9864e+02 |
| 6   | 3.0433e+00 | 1.5474e+02 |
| 7   | 2.8758e+00 | 3.9127e+01 |
| 8   | 2.6237e+00 | 3.5658e+01 |
| 9   | 2.7365e+00 | 2.9485e+01 |
| 10  | 2.5984e+00 | 7.7932e+00 |
| 11  | 2.5518e+00 | 1.7423e+10 |

descends slowly in the beginning while it drops significantly in the last step. Moreover, we observed in many examples that the algorithm shows a similar performance for a broad range of starting values $u^0$. Another important observation is that the performance of the algorithm depends on the value of $\gamma$. For too small as well as for too large values of $\gamma$ the algorithm does only converge when started very close to the solution. As a rule of thumb one could take $\gamma$ close to the reciprocal of the smallest singular value of the (in practice unknown) matrix $\mathcal{M}_{\mathcal{A},\mathcal{A}}$ where $\mathcal{A}$ is the sparsity pattern of the solution.

We made experiments to see how the SSN method depends on the noise level and the regularization parameter. First, we fixed the regularization sequence and changed the noise level. Hence, we solved the problem (20) for fixed $N = 500$, fixed $w_k = 10^{-5}$ and varied the noise level $\delta$. Table 2 reports the results. Basically, a higher noise level leads to a smaller number of iterations but longer CPU-time (this is, because the active sets are larger during the iteration). Second, we coupled the noise level and the regularizing sequence $w_k$. Since it is shown in [10, 18] that a parameter choice $w_k \propto \delta$ provides a regularization, we used $w_k = \delta$. Hence, we solved the problem (20) for fixed $N = 500$ and different noise levels $\delta$, see table 3 for the results. Basically, the algorithm behaves similar for different noise levels, especially the CPU-time is always comparably small.

Moreover, we made a simple experiment to assess how the computational cost grow with the size of the problem. We considered the inverse integration problem with problem size $N$ between 100 and 5000. We kept all parameters, as well as the data and the noise level fixed and only refined the discretization of the problem. We stopped the algorithms when a required

Table 2. Behavior of the SSN method for different noise levels with fixed $w_k = 10^{-5}$. The problem under consideration is the inverse integration, the problem size is $N = 500$ with $\gamma = 5 \times 10^5$ throughout. The rightmost column shows the residual norm at convergence.

| $\|\delta\|$ | #iter | CPU-time (s) | $\|r\|$ |
|-------------|-------|-------------|------|
| 1.0e+00     | 5     | 6.66e-01    | 6.18e-10 |
| 1.0e+01     | 8     | 7.63e-01    | 2.19e-09 |
| 1.0e+02     | 12    | 3.98e-01    | 7.85e-10 |
| 1.0e+03     | 10    | 3.11e-01    | 1.29e-09 |
| 1.0e+04     | 11    | 3.22e-01    | 9.85e-10 |
| 1.0e+05     | 11    | 3.18e-01    | 2.49e-09 |
As a second example of an ill-posed problem we consider a blurring operator \( A : L^2([0, 1]) \rightarrow L^2([0, 1]) \) given by \( Au = k * u \) with the kernel \( k(x) = c(1 + x^2/\lambda^2)^{-1} \) with \( \lambda = 0.01 \). We choose \( c \) such that \( \int k \, dx = 1 \) and consider \( u \) to be extended periodically to \( \mathbb{R} \) in order to evaluate the convolution integral.

In this example, we work with a synthesis operator \( B : \ell^2 \rightarrow L^2([0, 1]) \) mapping coefficients \( (c_k) \) to a function \( u = \sum c_k \psi_k \) where \( (\psi_k) \) form the orthonormal Haar wavelet basis [19]. Hence, the operator under consideration \( K = AB \) is a blurring after a Haar

### 4.2. Deblurring in a Haar basis

As a second example of an ill-posed problem we consider a blurring operator \( A : L^2([0, 1]) \rightarrow L^2([0, 1]) \) given by \( Au = k * u \) with the kernel \( k(x) = c(1 + x^2/\lambda^2)^{-1} \) with \( \lambda = 0.01 \). We choose \( c \) such that \( \int k \, dx = 1 \) and consider \( u \) to be extended periodically to \( \mathbb{R} \) in order to evaluate the convolution integral.

In this example, we work with a synthesis operator \( B : \ell^2 \rightarrow L^2([0, 1]) \) mapping coefficients \( (c_k) \) to a function \( u = \sum c_k \psi_k \) where \( (\psi_k) \) form the orthonormal Haar wavelet basis [19]. Hence, the operator under consideration \( K = AB \) is a blurring after a Haar
Figure 2. The empirical growth of the computational cost for the different algorithms.

Table 5. Illustration of the performance of the SSN method for deblurring in a Haar basis. The second column shows the decay of the function value $\Psi$ while the third column shows the norm of the residual. The data are the same as in figure 3.

| $n$  | $\Psi(u^n)$     | $\|r^n\|$ |
|------|-----------------|----------|
| 1    | 3.3920e+001     | 2.9676e+006 |
| 2    | 1.3905e+002     | 3.2499e+004 |
| 3    | 1.3326e+001     | 6.2647e+005 |
| 4    | 7.9347e+000     | 1.7517e+004 |
| 5    | 6.0006e+000     | 8.0510e+002 |
| 6    | 8.9823e+000     | 1.5424e-009 |

wavelet synthesis, see [6, 10] for discussions of $\ell^1$ penalty terms in combination with wavelet expansions.

We start with a given function $u$ which is piecewise constant. The data $f$ are computed as $f = Au + \text{noise}$ such that we have 25% relative error, i.e. $\|f - Au\|/\|f\| = 0.25$. The Haar coefficients of $u$ have been reconstructed by minimizing (1). As an illustration of $\ell^1$ penalties in contrast to classical $\ell^2$ regularization we also show the results of the minimization of

$$
\frac{1}{2}\|Kc - f\|^2 + \sum_{k=1}^{\infty} w_k \|c_k\|^2.
$$

Figure 3 and table 5 show the results of both $\ell^1$ and the above $\ell^2$ regularization where we discretized the problem to 1024 Haar wavelets. The parameters $w_k$ are independent of $k$ and have been tuned by hand to produce optimal results. Since the original data are quite sparse in the Haar wavelet basis, the $\ell^1$-reconstruction leads to much better results, as expected from the model. It also turned out that the algorithm is robust with respect to different initial values $u^0$. We tested several initial values (starting at zero, at $K^*f$ or at a random position) and the observed convergence behavior was very similar in all cases.
The SSN method converged in six iterations and in 0.3 s (for comparison: the GPSR method takes 0.5 s and $l_1$ ls converged in 5.3 s).

4.3. Compressive sampling

In our last example, we illustrate the applicability of the SSN method to the decoding problem in compressive sampling alias compressed sensing (CS). In CS one aims at reconstructing a signal from very few linear measurements, see [4, 11] for an introduction to CS. A popular way of decoding a signal from data $f$ which was measured by the observation operator $K$ is to minimize a functional of type (1), see [5]. Our example on compressive sampling is taken from [14]. We obtain an observation operator $K \in \mathbb{R}^{K \times N}$ by first filling it with independent samples of a standard Gaussian distribution and then orthonormalizing the rows. Hence, the operator is not injective but it possesses the so-called restricted isometry property (see [1]) which means that all submatrices consisting of a small number of columns have singular values close to one. Especially, submatrices made of a small number of columns are injective. Hence, the SSN method works as long as the active sets are small enough.

In this example we chose $N = 8192$, $K = 512$, and the signal $u$ contained 64 randomly placed $\pm 1$ spikes. The observation $f$ was generated by $f = Ku + \text{noise}$ such that we have 5\% relative error. The minimization of (1) with $w = 0.05$ was done with the SSN method with parameter $\gamma = 5 \times 10^4$. The SSN method converged in approximately 1.2 s in six iterations and the active sets stayed very small during the iteration, see figure 4 and table 6. Hence, the SSN method is a promising candidate for the decoding problem in CS.
5. Conclusion

We have shown that the semismooth Newton method applied to Tikhonov functionals with sparsity constraints is a fast algorithm which is easy to implement as an active set method. Each iteration involves the solution of a system of linear equations on the active coefficients only. Our numerical experiments show that these systems stay reasonably small during the iteration and are also very well conditioned. In addition, the experiments indicate that the SSN method compares favorably with existing state-of-the-art methods when applied to ill-posed problems. While we investigated only the local convergence behavior, the numerical experiments indicate that our method is robust with respect to the initial value of the iteration. However, the convergence is slow as long as the iterates are far from the minimizer and it gets faster when the solution is approached. The global convergence properties are not yet explained by our theory and need further investigation. Another direction for further research is globalization of the method, e.g., by the use of an appropriate merit function, and line search or trust region methods.
We define for $u \in \ell^2$ and $h \in H$

$$F(u) = \sum_{k=1}^{\infty} w_k |u_k|, \quad G(h) = \frac{1}{2} \|h - f\|_H^2$$

and calculate their conjugate (polar) functions, see [12, chapter I.4]. We have

$$F^*(p) = \sup_{u \in \ell^2} \left( \langle p, u \rangle - F(u) \right) = \sup_{u} \left( \langle p, u \rangle - \sum_{k=1}^{\infty} w_k |u_k| \right)$$

$$= \sup_{u} \left( \sum_{k=1}^{\infty} (p_k - w_k \text{sign } u_k) u_k \right) = \begin{cases} 0, & \text{if } |p_k| \leq w_k \text{ for all } k \\ \infty, & \text{otherwise.} \end{cases}$$

For $G$, we obtain

$$G^*(p) = \sup_{h \in H} \left( \langle p, h \rangle - G(h) \right) = \sup_{h} \left( \langle p, h \rangle - \frac{1}{2} \|h - f\|_H^2 \right) = \frac{1}{2} \|p\|_H^2 + \langle p, f \rangle,$$

since the supremum is attained at $h = p + f$.

References

[1] Baraniuk R, Davenport M, DeVore R and Wakin M 2008 A simple proof of the restricted isometry property for random matrices Construct. Approx. at press
[2] Bredies K and Lorenz D A 2007 Iterative soft-thresholding converges linearly Preprint 0709.1598 at press
[3] Bredies K and Lorenz D A 2008 Iterated hard shrinkage for minimization problems with sparsity constraints SIAM J. Sci. Comput. 30 657–83
[4] Candès E J 2006 Compressive sampling Proc. International Congress of Mathematics pp 1433–52
[5] Candès E J and Tao T 2005 Decoding by linear programming IEEE Trans. Inf. Theory 51 4203–15
[6] Chambolle A, DeVore R A, Lee N and Lucier B J 1998 Nonlinear wavelet image processing: variational problems, compression and noise removal through wavelet shrinkage IEEE Trans. Image Process. 7 319–35
[7] Chen X 1997 Superlinear convergence and smoothing quasi-Newton methods for nonsmooth equations J. Comput. Appl. Math. 80 105–26
[8] Chen X, Nazhdad Z and Qi L 2000 Smoothing methods and semismooth methods for nondifferentiable operator equations SIAM J. Numer. Anal. 38 1200–16
[9] Combettes P L and Wajs V R 2005 Signal recovery by proximal forward-backward splitting Multiscale Model. Simul. 4 1168–200
[10] 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
[11] Donoho D 2006 Compressed sensing IEEE Trans. Inf. Theory 52 1289–306
[12] Ekeland I and Temam R 1976 Convex Analysis and Variational Problems (Amsterdam: North-Holland)
[13] Engl H W, Hanke M and Neubauer A 2000 Regularization of Inverse Problems (Mathematics and its Applications vol 375) (Dordrecht: Kluwer)
[14] Figueredo M A T, Nowak R D and Wright S J 2007 Gradient projection for sparse reconstruction: applications to compressed sensing and other inverse problems IEEE J. Sel. Top. Signal Process. 1 (1) 586–97
[15] Hanke M and Scherzer O 2001 Inverse problems light: numerical differentiation Am. Math. Month. 108 512–21
[16] Hintermüller M, Ito K and Kunisch K 2002 The primal-dual active set strategy as a semismooth Newton method SIAM J. Optim. 13 865–88
[17] Kim S-J, Koh K, Lustig M, Boyd S and Gorinevsky D 2008 A method for large-scale $\ell_1$-regularized least squares problems with applications in signal processing and statistics IEEE J Sel. Top. Signal Process. 4 (1) 606–17
[18] Lorenz D A 2008 Convergence rates and source conditions for Tikhonov regularization with sparsity constraints Preprint 0801.1774 at press
[19] Louis A K, Maass P and Rieder A 1997 Wavelets: Theory and Application (Chichester: Wiley)
[20] Ramlau R and Teschke G 2006 A Tikhonov-based projection iteration for nonlinear ill-posed problems with sparsity constraints Numer. Math. 104 177–203
[21] Resmerita E 2005 Regularization of ill-posed problems in Banach spaces: convergence rates Inverse Problems 21 1303–14
[22] Rösch A and Kunisch K 2002 A primal-dual active set strategy for a general class of constrained optimal control problems SIAM J. Optim. 13 321–34
[23] Schöpfer F, Louis A K and Schuster T 2006 Nonlinear iterative methods for linear ill-posed problems in Banach spaces Inverse Problems 22 311–29
[24] Stadler G 2008 Elliptic optimal control problems with $l^1$-control cost and applications for the placement of control devices Comput. Optim. Appl. at press
[25] Ulbrich M 2003 Semismooth Newton methods for operator equations in function spaces SIAM J. Control Optim. 13 805–42
[26] Zeidler E 1990 Nonlinear Functional Analysis and its Applications II/B: Nonlinear Monotone Operators (Berlin: Springer)