Zentrum für Technomathematik

Fachbereich 3 – Mathematik und Informatik

Semismooth Newton and Quasi-Newton Methods in Weighted ℓ^1−Regularization of Nonlinear Inverse Problems

Peter Maass Pham Q. Muoi

Report 11–02

Berichte aus der Technomathematik

Report 11–02 June 2011

Semismooth Newton and Quasi-Newton Methods in Weighted ℓ^1-Regularization of Nonlinear Inverse Problems

Peter Maass[†] and Pham Q. Muoi[†]

June, 2011

[†] Center for Industrial Mathematics, University of Bremen,
Bibliothekstr. 1, D-28334 Bremen, Germany
Email: pmaass@math.uni-bremen.de, pham@math.uni-bremen.de

Abstract

In this paper, we investigate the semismooth Newton and quasi-Newton methods for the minimization problem in the weighted ℓ^1-regularization of nonlinear inverse problems. We propose the conditions for obtaining the convergence of two methods. The semismooth Newton method is proven to locally converge with superlinear rate and the semismooth quasi-Newton method is proven to locally converge at least with linear rate. Two methods are presented as active set methods as well.

For using the semismooth quasi-Newton method in practice, we propose two specific cases. The first one returns to a gradient-type method with Barzilai-Borwein rule for step-sizes. The second one based on Broyden's method is proven to converge and its convergence rate is superlinear in finite dimensional spaces. Finally, the efficiency of the methods are illustrated in a parameter identification problem in elliptic equations.

Keywords : Sparsity Regularization, Nonlinear Inverse Problems, Semismooth Newton Method, Semismooth Quasi-Newton Method.

1 Introduction

In this work, we consider the minimization problem

$$\min_{u\in\mathcal{H}} F(u) + \sum_{k\in\Lambda} \omega_k |\langle u, \varphi_k\rangle|, \tag{1}$$

where $\mathcal{H}$ is a Hilbert space, $F : \mathcal{H} \to \mathbb{R}$ is a smooth functional but not necessary convex, $\{\varphi_k\}_{k\in\Lambda}$ is an orthonormal basis of $\mathcal{H}$ and $\{\omega_k\}$ is a positive sequence such that $\omega_k \geqslant \omega_{min} > 0$ for all $k \in \Lambda$. For simplicity, we here assume that $\{\varphi_k\}_{k\in\Lambda}$ is an orthonormal basis of $\mathcal{H}$, but by standard arguments, the results in this paper are still valid when $\{\varphi_k\}_{k\in\Lambda}$ is an basis or frame of $\mathcal{H}$, see, e.g. [29].

Such a problem arises from sparsity regularization for the operator equation

$$K(u) = f, \tag{2}$$

where $K : \mathcal{H}_1 \to \mathcal{H}_2$ is an ill-posed, nonlinear operator between Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$, and only noisy data f^δ with

$$\|f - f^\delta\|_{\mathcal{H}_2} \leqslant \delta \tag{3}$$

are available.

The sparsity regularization method applying to equations (2) - (3) leads to considering the minimization problem (see, e.g. [11, 15])

$$\min_{u\in\mathcal{H}_1} \frac{1}{2}\|K(u) - f^\delta\|^2_{\mathcal{H}_2} + \alpha \sum_{k\in\Lambda} \omega_k |\langle u, \varphi_k\rangle|, \tag{4}$$

or in the general form

$$\min_{u\in\mathcal{H}_1} F\big(K(u), f^\delta\big) + \alpha \sum_{k\in\Lambda} \omega_k |\langle u, \varphi_k\rangle|, \tag{5}$$

where $\{\omega_k\}$ is a positive sequence such that $\omega_k \geqslant \omega_{min} > 0$ for all $k \in \Lambda$ and $\{\varphi_k\}$ is an orthonormal basis (or frame) of the Hilbert space $\mathcal{H}_1$; $F\big(K(u), f^\delta\big)$ is a discrepancy functional measuring the difference between $K(u)$ and f^δ. It is obvious that problems (4) and (5) are specific cases of problem (1).

Sparsity regularization has been analyzed for linear and nonlinear settings over the last years [11, 24, 15]. Numerical algorithms for computing minimizers of (4) have been proposed in [11, 6, 35] for linear inverse problems and in [4, 29] for nonlinear inverse problems. Most of them are known to have a linear convergence rate in theory and to be quite slow in practice, especially for nonlinear inverse problems.

Recently, a gradient descent method and its two accelerated algorithms have proposed for problems of type (1) in [25]. Note that the gradient descent method is similar to the algorithms in [11, 6, 35, 4], but it has been speeded up by selecting stepsizes. However, the method is still slow. Moreover, its two accelerated algorithms are only proven to converge for convex minimization problems.

The motivation for the present paper originates in the results of R. Griesse and D. A. Lorenz [16]. There, they have applied the semismooth Newton method for problem (4) with a linear operator K. They have proved that the convergence rate of the method is superlinear. Therefore, it is a fast method and by our best knowledge, it is the best convergence rate in all algorithms proposed for problem (4) until now.

In this paper, we first extend the semismooth Newton method in [16] for problem (1) with a general functional F (can be non-convex), i.e. problem (1) includes regularization of nonlinear, ill-posed problem (2). Under certain conditions of F, the method is proved to locally converge with superlinear rate. Although the method converges very fast, its disadvantage is to have to compute the second derivative of F in each iterate. This is a hard work for nonlinear inverse problems. Therefore, the applicability of the method is restricted.

To overcome this shortcoming, we investigate the so-called semismooth quasi-Newton method. Under different conditions, the method is proved to converge with linear rate. Two specific cases of the method, which can be applied in practice, are proposed as well. The first one returns to a gradient-type method with Barzilai-Borwein rule for step-sizes, but it is a little different from gradient-type methods in [25, 11, 6, 35, 4]. The second one based on Broyden's method is proven to locally converge with linear rate. Furthermore, it locally converge with superlinear rate in finite dimensional spaces.

The remainder of the paper is organized as follows. In Section 2, we consider the optimality condition equation of problem (1) and prove the semismooth property of the operator in the equation. Section 3 is devoted to present the semismooth Newton method and examine its convergence. The semismooth quasi-Newton method is analyzed in Section 4. In Section 5, we propose two specific cases of the semismooth quasi-Newton method. After that the methods are represented as active set methods in Section 6. Finally, we illustrate the efficiency of the methods in a numerical example in Section 7.

2 Auxiliary results

In this section, we are going to show that the optimality necessary condition of problem (1) results in an operator equation in which the operator is not Gâteaux differentiable, but Newton differentiability. To this end, we first introduce the so-called soft shrinkage operator.

Definition 2.1 *Let $w = \{\omega_k\}_{k\in\Lambda}$ with $\omega_k \geqslant \omega_{min} > 0$ and $\{\varphi_k\}$ is an orthonormal basis of $\mathcal{H}$. The soft shrinkage operator $\mathbb{S}_w : \mathcal{H} \to \mathcal{H}$ is defined by*

$$\mathbb{S}_w(u) := \sum_{k\in\Lambda} S_{\omega_k}(u_k)\varphi_k, \tag{6}$$

where $u_k := \langle u, \varphi_k\rangle$ and $S_{\omega_k}(u_k) := \operatorname{sgn}(u_k)\max\{0, |u_k| - \omega_k\}$.

Note that because $u \in \mathcal{H}$, $\{u_k\}$ converges to zero. Therefore, the range of $\mathbb{S}_w$ is $\mathcal{H}^0 = \{v \in \mathcal{H} : v_k = 0 \text{ for almost every } k\}$. Using this operator, the optimality condition of problem (1) is given explicitly in the lollowing lemma.

Lemma 2.1 *Let $\overline{u}$ is a minimizer of problem (1). Assume that F is Fréchet differentiable at $\overline{u}$. Then, the minimizer $\overline{u}$ satisfies*

$$\overline{u} = \mathbb{S}_{\beta w}\big(\overline{u} - \beta F'(\overline{u})\big) \text{ for any fixed } \beta > 0.$$

Additionally, if F is convex then this is also the sufficient condition.

Proof. Denote $\Phi(u) = \sum_{k\in\Lambda}\omega_k|\langle u, \varphi_k\rangle|$. The necessary condition of optimality is

$$-F'(\overline{u}) \in \partial\Phi(\overline{u}).$$

Multiplying with $\beta > 0$, adding $\overline{u}$ to both sides and inverting $(I + \beta\partial\Phi)$ gives

$$\overline{u} = (I + \beta\partial\Phi)^{-1}\big(\overline{u} - \beta F'(\overline{u})\big)$$

(note that $(I + \beta\partial\Phi)^{-1}$ exists and is single-valued, see [16]). A straightforward calculation shows that

$$(I + \beta\partial\Phi)^{-1} = \mathbb{S}_{\beta\omega}.$$

■

From Lemma 2.1, the optimal condition equation of problem (1) is

$$D(u) := u - \mathbb{S}_{\beta w}(u - \beta F'(u)) = 0, \tag{7}$$

for any fixed $\beta > 0$. Now, instead of solving problem (1), we find solutions of equation (7). Note that the function $D(\cdot)$ is not Gâteaux differentiable, but it is Newton (slantly) differentiable as shown below. Therefore, the semismooth Newton and quasi-Newton methods might be applied. For details of the slant differentiability and the semismooth Newton and quasi-Newton methods, we refer to [9, 18, 33] and references therein. Here, for the convenience, we give the definition of the Newton differentiability.

Definition 2.2 *Let X and Y be Banach spaces and $U \subset X$ be an open subset. A mapping $\psi : U \to Y$ is called to be Newton (or slantly) differentiable at $u \in U$ if there exists a family of mappings $\chi : U \to L(X, Y)$ such that*

$$\lim_{h\to 0} \frac{\|\psi(u + h) - \psi(u) - \chi(u + h)h\|_Y}{\|h\|_X} = 0. \tag{8}$$

The function χ is called a generalized derivative (or a slanting function) of ψ at u.

Remark 2.1 *Note that if ψ is Newton differentiable at u, then it might have many generalized derivatives and the Fréchet differentiability of ψ implies its Newton differentiability.*

Using this definition, we prove that the function D defined by (7) is Newton differentiable. To this end, for each $u \in \mathcal{H}$, the operator $G(u) : \mathcal{H} \to \mathcal{H}$ is defined by

$$\big(G(u)v\big)_k = \begin{cases} v_k \text{ for } |u_k| > \omega_k \\ 0 \text{ for } |u_k| \leqslant \omega_k \end{cases}, \tag{9}$$

where $u_k = \langle u, \varphi_k \rangle$ and $v_k = \langle v, \varphi_k \rangle$.

It is easy to show that $G(u)(\cdot)$ is a continuous linear operator with $\|G(u)\| \leqslant 1$. The following lemma will show that G is a generalized derivative of $\mathbb{S}_w$ at u.

Lemma 2.2 *Let $\mathbb{S}_w$ be defined by (6). Then, $\mathbb{S}_w$ is Newton differentiable and $G(u)$ given by (9) is a generalized derivative of $\mathbb{S}_w$ at u.*

Proof. See the proof in [16, Proposition 3.3]. ∎

Remark 2.2 *In matrix notation, we can express the derivative $G(u)$ as*

$$G(u) = \begin{pmatrix} I_{\mathbb{A}} & 0 \\ 0 & 0 \end{pmatrix},$$

where $\mathbb{A} = \{k \in \Lambda : |u_k| > \omega_k\}$.

To calculate a generalized derivative of D in (7), we first prove the chain rule for the generalized derivative and introduce active and inactive sets.

Lemma 2.3 *Let $J : U \subset \mathcal{H} \to \mathcal{H}$ be Fréchet differentiable with Lipschitz continuous derivative in a neighborhood of $u \in U$ (U is an open set) and $\psi : \mathcal{H} \to \mathcal{H}$ be Newton differentiable at $J(u)$ with a slanting function χ. Furthermore, let $\|\chi(u)\|$ be uniformly bounded. Then, $T(u) = \psi(J(u))$ is Newton differentiable at u with a slanting function $H(u) = \chi(J(u))J'(u)$.*

Proof. Since J is Fréchet differentiable at u, it holds

$$J(u + h) = J(u) + J'(u)h + r(h)$$

and since J' is Lipschitz continuous (with a Lipschitz constant L), it holds that the remainder r fulfills the inequality $\|r(h)\| \leq L/2\|h\|^2$. We denote $k(h) = J'(u)h + r(h)$ and estimate

$$\begin{aligned} &\|T(u + h) - T(u) - H(u + h)h\| \\ &\quad = \|\psi(J(u + h)) - \psi(J(u)) - \chi(J(u + h))J'(u + h)h\| \\ &\quad = \|\psi(J(u) + k(h)) - \psi(J(u)) - \chi(J(u) + k(h))J'(u + h)h\| \\ &\quad \leq \|\psi(J(u) + k(h)) - \psi(J(u)) - \chi(J(u) + k(h))k(h)\| \\ &\qquad + \|\chi(J(u) + k(h))(J'(u + h)h - k(h))\|. \end{aligned}$$

The last term is further estimated as

$$\begin{aligned} &\|\chi(J(u) + k(h))(J'(u + h)h - k(h))\| \\ &\quad = \|\chi(J(u) + k(h))(J'(u + h)h - J'(u)h - r(h))\| \\ &\quad \leq \|\chi(J(u) + k(h))\| \left(\|J'(u + h) - J'(u)\| \, \|h\| + \|r(h)\|\right) \\ &\quad \leq \|\chi(J(u) + k(h))\| \, (L + L/2)\|h\|^2. \end{aligned}$$

Putting the above estimates together, we obtain

$$\begin{aligned}&\frac{\|T(u+h)-T(u)-H(u+h)h\|}{\|h\|}\\&\quad\leq\frac{\|\psi(J(u)+k(h))-\psi(J(u))-\chi(J(u)+k(h))k(h)\|}{\|k(h)\|}\frac{\|k(h)\|}{\|h\|}\\&\quad\quad+\|\chi(J(u)+k(h))\|\,(L+L/2)\|h\|.\end{aligned}$$

Now the claim follows since $\|k(h)\|/\|h\|$ as well as $\|\chi(J(u)+k(h))\|$ is bounded for $\|h\|\to 0$. ■

Definition 2.3 *For $u\in\mathcal{H}$, the active set $\mathbb{A}(u)$ and the inactive set $\mathbb{I}(u)$ are defined by*

$$\begin{aligned}\mathbb{A}(u)&=\{k\in\Lambda:|u-\beta F'(u)|_k>\beta\omega_k\},\\\mathbb{I}(u)&=\{k\in\Lambda:|u-\beta F'(u)|_k\leqslant\beta\omega_k\}.\end{aligned}$$

With this notation, we are now in a position to formulate a generalized derivative of D. Note that for any $u\in\mathcal{H}$, the active set $\mathbb{A}(u)$ is always finite, since $u-\beta F'(u)\in\mathcal{H}$ and thus $|(u-\beta F'(u))_k|$ converges to zero as k tends to infinity.

Theorem 2.1 *Let F be twice Fréchet differentiable in U (U is a open set) and F'' is Lipschitz continuous in a neighborhood of $u\in U$. Then, function D defined by (7) is Newton differentiable at u and a generalized derivative of D at u is given by*

$$D'(u)=I-G(u-\beta F'(u))(I-\beta F''(u)).$$

Furthermore, denote the active and inactive sets at u by $\mathbb{A}$ and $\mathbb{I}$ as in Definition 2.3 and represent the operator $F''(u)$ as

$$F''(u)=\begin{pmatrix}\mathcal{M}_{\mathbb{A}\mathbb{A}}&\mathcal{M}_{\mathbb{A}\mathbb{I}}\\\mathcal{M}_{\mathbb{I}\mathbb{A}}&\mathcal{M}_{\mathbb{I}\mathbb{I}}\end{pmatrix}.$$

Then, the generalized derivative of D at u is rewritten by

$$D'(u)=\begin{pmatrix}0&0\\0&I_{\mathbb{I}}\end{pmatrix}+\begin{pmatrix}I_{\mathbb{A}}&0\\0&0\end{pmatrix}(\beta F''(u))=\begin{pmatrix}\beta\mathcal{M}_{\mathbb{A}\mathbb{A}}&\beta\mathcal{M}_{\mathbb{A}\mathbb{I}}\\0&I_{\mathbb{I}}\end{pmatrix}.\qquad(10)$$

Proof. The theorem follows from Lemma 2.3 with $J(u)=u-\beta F'(u)$. ■

3 Semismooth Newton Method (SSN)

In this section, we present the semismooth Newton method for equation (7) and propose the sufficient conditions for obtaining the local convergence of the method. Note that the method has been considered by many authors, for example, [9, 18, 33]. Recently, it has been applied to problem (1) for $F(u)=\frac{1}{2}\|Ku-f^\delta\|^2$ with a linear operator K [16]. Here, the method is analyzed for problem (1) in a general setting.

As discussed in the previous section, instead of solving problem (1) we solve equation (7). The semismooth Newton method for the equation is the following iteration

$$u^{n+1}=u^n-D'(u^n)^{-1}D(u^n).\qquad(11)$$

In each iteration, we denote the active set and the inactive set by

$$\mathbb{A}^n=\{k\in\Lambda:|u^n-\beta F'(u^n)|_k>\beta\omega_k\},\ \mathbb{I}^n=\{k\in\Lambda:|u^n-\beta F'(u^n)|_k\leqslant\beta\omega_k\}.$$

Then by Theorem 2.1, we have

$$\begin{aligned} u^{n+1} &= u^n - \begin{pmatrix} \frac{1}{\beta}\mathcal{M}^{-1}_{\mathbb{A}^n\mathbb{A}^n} & -\mathcal{M}^{-1}_{\mathbb{A}^n\mathbb{A}^n}\mathcal{M}_{\mathbb{A}^n\mathbb{I}^n} \\ 0 & I_{\mathbb{I}^n} \end{pmatrix} \left(u^n - \mathbb{S}_{\beta w}(u^n - \beta F'(u^n))\right) \\ &= u^n - \begin{pmatrix} \frac{1}{\beta}\mathcal{M}^{-1}_{\mathbb{A}^n\mathbb{A}^n} & -\mathcal{M}^{-1}_{\mathbb{A}^n\mathbb{A}^n}\mathcal{M}_{\mathbb{A}^n\mathbb{I}^n} \\ 0 & I_{\mathbb{I}^n} \end{pmatrix} \begin{pmatrix} \beta[F'(u^n) \pm w]\big|_{\mathbb{A}^n} \\ u^n_{\mathbb{I}^n} \end{pmatrix} \\ &= \begin{pmatrix} u^n_{\mathbb{A}^n} - \mathcal{M}^{-1}_{\mathbb{A}^n\mathbb{A}^n}\left([F'(u^n) \pm w]\big|_{\mathbb{A}^n} - \mathcal{M}_{\mathbb{A}^n\mathbb{I}^n}u^n_{\mathbb{I}^n}\right) \\ 0 \end{pmatrix}. \end{aligned} \tag{12}$$

Here, we have implicitly assumed that $\mathcal{M}^{-1}_{\mathbb{A}^n\mathbb{A}^n}$ exists. The sign of w depends on the sign of $u^n - \beta F'(u^n)$. Hence, instead of calculating the Newton update, we can set $u^{n+1}_{\mathbb{I}^n} = 0$, then solve the equation $\mathcal{M}_{\mathbb{A}^n\mathbb{A}^n}\delta u_{\mathbb{A}^n} = [F'(u^n) \pm w]\big|_{\mathbb{A}^n} - \mathcal{M}_{\mathbb{A}^n\mathbb{I}^n}u^n_{\mathbb{I}^n}$ and compute $u^{n+1}_{\mathbb{A}^n} = u^n_{\mathbb{A}^n} - \delta u_{\mathbb{A}^n}$.

In the remainder of the section, we consider the local convergence of the semismooth Newton method. To this end, we need some assumptions on F and the existence of solution of equation (7). We collect them in Assumption 3.1.

Assumption 3.1 *We assume that*

1) *Equation (7) has a solution $u^* \in U$, where U is an open set.*

2) *F is twice Fréchet differentiable. Both F' and F'' are Lipschitz continuous in U.*

3) *For each finite index set $\mathbb{A} \subset \Lambda$ and $\mathbb{I} = \Lambda \backslash \mathbb{A}$, we represent $F''(u)$ by*

$$F''(u) = \begin{pmatrix} \mathcal{M}_{\mathbb{A}\mathbb{A}} & \mathcal{M}_{\mathbb{A}\mathbb{I}} \\ \mathcal{M}_{\mathbb{I}\mathbb{A}} & \mathcal{M}_{\mathbb{I}\mathbb{I}} \end{pmatrix}.$$

Then, there exists $\rho > 0$ such that $\mathcal{M}^{-1}_{\mathbb{A}\mathbb{A}}$ exists and uniformly bounded on $B_\rho(u^*) \subset U$.

Remark 3.1 *The reasons for requiring Assumption 3.1 are as follows.*

1. *Since we are going to solve equation (7), Condition 1) is clear to be fulfilled.*
2. *The Lipschitz continuity of F'' is required so that we can apply Theorem 2.1.*
3. *The Lipschitz continuity of F' and Condition 3) are needed for the existence and uniformly boundedness of $D'(u^n)^{-1}$ in iterations (11) or (12), see Theorem 3.1 below.*

The following examples show that Assumption 3.1 is satisfied in some situations.

Example 3.1 *We consider the Tikhonov functionals in linear inverse problems, i.e, $F(u) = \|Ku - f^\delta\|^2$ with linear operators K. It is easy to check condition 1) and 2) of Assumption 3.1. Furthermore, if K satisfies the finite basis injectivity property (FBI) [5], the functional F satisfies the condition 3) of Assumption 3.1, see [16, Remark 3.15].*

Example 3.2 *Let F be twice Fréchet differentiable; F', F'' Lipschitz in a neighborhood of a solution u^* and F'' be the operator such that*

$$\tau_1\|h\|^2 \leqslant \langle F''(u)h, h\rangle \leqslant \tau_2\|h\|^2, 0 < \tau_1, \tau_2 < \infty$$

for all u in a neighborhood of u^*. Then, F satisfies the properties 2) and 3).

With Assumption 3.1, we shall show that the semismooth Newton method locally converges and its convergence rate is supperlinear. To this end, following the outline of [16], we need some auxiliary results in the following lemmas.

Lemma 3.1 *If F' is Lipschitz continuous with the Lipschitz constant L in a neighborhood of u^* then there exist $k_0 \in \Lambda$ and $\rho > 0$ such that the condition $\|u - u^*\| < \rho$ implies the inclusion*

$$\mathbb{A}(u) \subset [1, k^0].$$

Moreover, k^0 and ρ depend on β, u^*, L and ω_{min}.

Proof. The triangle inequality implies

$$|u_k - \beta F'(u)_k| \leqslant |u_k^* - \beta F'(u^*)_k| + |u_k - u_k^* - \beta(F'(u)_k - F'(u^*)_k)|. \tag{13}$$

The first term converges to zero as k tends to infinity because u^* and $F'(u^*)$ are in $\mathcal{H}$. In particular, there exists k_0, depending only on u^* and β, such that

$$|u_k^* - \beta F'(u^*)_k| < \beta\omega_{min}/2 \text{ for all } k \geqslant k_0. \tag{14}$$

The second term can be estimated as follows

$$\begin{aligned} |u_k - u_k^* - \beta(F'(u)_k - F'(u^*)_k)| &\leqslant |u_k - u_k^*| + \beta|F'(u)_k - F'(u^*)_k)| \\ &\leqslant (1 + \beta L)\|u - u^*\|, \end{aligned}$$

where L is the Lipschitz constant of F'. Thus there exists $\rho > 0$ depending only the named quantities such that

$$|u_k - u_k^* - \beta(F'(u)_k - F'(u^*)_k)| \leqslant \beta\omega_{min}/2 \text{ for all } k \in \Lambda. \tag{15}$$

The proof of the lemma follows from (13)-(15). ∎

Lemma 3.2 *If $F''(u) = \begin{pmatrix} \mathcal{M}_{\mathbb{A}\mathbb{A}} & \mathcal{M}_{\mathbb{A}\mathbb{I}} \\ \mathcal{M}_{\mathbb{I}\mathbb{A}} & \mathcal{M}_{\mathbb{I}\mathbb{I}} \end{pmatrix}$ and $\mathcal{M}_{\mathbb{A}\mathbb{A}}$ is injective, then $D'(u) : \mathcal{H} \to \mathcal{H}$ is bounded invertible and*

$$\|D'(u)^{-1}\| \leqslant \|\mathcal{M}_{\mathbb{A}\mathbb{A}}^{-1}\|\big(\frac{1}{\beta} + \|\mathcal{M}_{\mathbb{A}\mathbb{I}}\|\big) + 1,$$

where $\mathbb{A}$ and $\mathbb{I}$ are the active and inactive sets at u.

Proof. Let $r \in \mathcal{H}$ we consider the equation $D'(u)\delta u = r$, i.e. the equation

$$\begin{pmatrix} \beta\mathcal{M}_{\mathbb{A}\mathbb{A}} & \beta\mathcal{M}_{\mathbb{A}\mathbb{I}} \\ 0 & I_{\mathbb{I}} \end{pmatrix} \begin{pmatrix} \delta u_{\mathbb{A}} \\ \delta u_{\mathbb{I}} \end{pmatrix} = \begin{pmatrix} r_{\mathbb{A}} \\ r_{\mathbb{I}} \end{pmatrix}.$$

This equation is equivalent to $\delta u_{\mathbb{I}} = r_{\mathbb{I}}$ and

$$\beta\mathcal{M}_{\mathbb{A}\mathbb{A}}\delta u_{\mathbb{A}} = r_{\mathbb{A}} - \beta\mathcal{M}_{\mathbb{A}\mathbb{I}} r_{\mathbb{I}}. \tag{16}$$

On the other hand, the active set $\mathbb{A}$ is finite and thus $\mathcal{M}_{\mathbb{A}\mathbb{A}}$ is an injective operator on a finite dimensional space. Therefore, it is also surjective. We conclude that (16) has a unique solution and thus $D'(u)^{-1}$ exists.

On the other hand, we have the estimation

$$\begin{aligned}\|D'(u)^{-1}r\| &= \left\| \begin{pmatrix} \frac{1}{\beta}\mathcal{M}_{\mathbb{A}\mathbb{A}}^{-1} & \mathcal{M}_{\mathbb{A}\mathbb{A}}^{-1}\mathcal{M}_{\mathbb{A}\mathbb{I}} \\ 0 & I_{\mathbb{I}} \end{pmatrix} \begin{pmatrix} r_{\mathbb{A}} \\ r_{\mathbb{I}} \end{pmatrix} \right\| \\ &\leqslant \frac{1}{\beta}\|\mathcal{M}_{\mathbb{A}\mathbb{A}}^{-1}\|\|r_{\mathbb{A}}\| + \|\mathcal{M}_{\mathbb{A}\mathbb{A}}^{-1}\|\|\mathcal{M}_{\mathbb{A}\mathbb{I}}\|\|r_{\mathbb{I}}\| + \|r_{\mathbb{I}}\| \\ &\leqslant \big(\frac{1}{\beta}\|\mathcal{M}_{\mathbb{A}\mathbb{A}}^{-1}\| + \|\mathcal{M}_{\mathbb{A}\mathbb{A}}^{-1}\|\|\mathcal{M}_{\mathbb{A}\mathbb{I}}\| + 1\big)\|r\|.\end{aligned}$$

■

We are now in a position to consider the local convergence of the semismooth Newton method.

Theorem 3.1 *Assume that Assumption 3.1 holds. Then, there exists a radius* $r > 0$ *such that the inequality* $\|u^0 - u^*\| < r$ *implies that all sequence* $\{u^n\}$ *defined by (11) satisfies* $\|u^n - u^*\| < r$*, and* $u^n \to u^*$ *superlinearly.*

Proof. Let $\rho > 0$ be a number such that Assumption 3.1 and Lemma 3.1 are satisfied in $B_\rho(u^*)$. By Lemma 3.1, the active set satisfies $\mathbb{A}(u) \subset [0, k_0]$. We shall show that $D'^{-1}(u)$ depends only on k_0. Indeed, we define

$$c(k_0) = \max_{\emptyset \neq \mathbb{A} \subset [0,k_0]} \sup_{u \in B_\rho(u^*)} \|\mathcal{M}_{\mathbb{A}\mathbb{A}}^{-1}\| > 0.$$

Note that, by Assumption 3.1, for every $\mathbb{A} \subset [0, k_0], \mathbb{A} \neq \emptyset, \sup_{u \in B_\rho(u^*)} \|\mathcal{M}_{\mathbb{A}\mathbb{A}}^{-1}\|$ is finite, hence $c(k_0)$ is the maximum of finitely many positive numbers. On the other hand, by Assumption 3.1 (the Lipschitz continuity implies the uniformly boundedness) $\|\mathcal{M}_{\mathbb{A}\mathbb{I}}\| \leqslant \|F''(u)\| \leqslant \tau$ for all choices of $\mathbb{A}$ and $\mathbb{I}$ with τ is a positive constant.

From Lemma 3.2, it follows that

$$\|D^{-1}(u)\| \leqslant c(k_0)(\frac{1}{\beta} + \tau) + 1.$$

Therefore, the inverse of the generalized derivative $D^{-1}(u)$ is uniformly bounded in $B_\rho(u^*)$. The result is then a standard conclusion of generalized Newton methods, see [9, Remark 2.7] or [18, Theorem 1.1]. ■

Remark 3.2 *Theorem 3.1 shows that the semismooth Newton method for solving (7) locally converges with superlinear rate. Thus, it is a fast algorithm. However, its disadvantage is to have to compute the second derivative of* F*, which is often difficult in practical problems. This restricts the applicability of the method. Note that for smooth operator equations, the quasi-Newton method is good candidates instead of the Newton method [28, 31, 19, 26]. The next section will generalize the quasi-Newton method for the nonsmooth equation (7), which will be called the semismooth quasi-Newton method.*

4 Semismooth quasi-Newton Method (SSQN)

In the semismooth Newton method, the computation of the second derivative $F''(u)$ is a hard work. Therefore, one often computes its approximations. We denote $C(u)$ by an approximation of $F''(u)$. Then, $D'(u)$ is approximated by

$$D_1(u) := I - G(u - \beta F'(u))[I - \beta C(u)].$$

In this section, we consider the following semismooth quasi-Newton method

$$u^{n+1} = u^n - D_1(u^n)^{-1}D(u^n). \tag{17}$$

It is clear that if $C(u^n) = F''(u^n)$, then the semismooth quasi-Newton method becomes the semismooth Newton method (11). Similar to the semismooth Newton method, if in each iteration we split the operator $C(u^n)$ as

$$C(u^n) = \begin{pmatrix} \mathcal{M}_{\mathbb{A}^n\mathbb{A}^n} & \mathcal{M}_{\mathbb{A}^n\mathbb{I}^n} \\ \mathcal{M}_{\mathbb{I}^n\mathbb{A}^n} & \mathcal{M}_{\mathbb{I}^n\mathbb{I}^n} \end{pmatrix},$$

then the semismooth quasi-Newton method can be rewritten by

$$u^{n+1} = \begin{pmatrix} u^n_{\mathbb{A}^n} - \mathcal{M}^{-1}_{\mathbb{A}^n\mathbb{A}^n}\left([F'(u^n) \pm w]\big|_{\mathbb{A}^n} - \mathcal{M}_{\mathbb{A}^n\mathbb{I}^n}u^n_{\mathbb{I}^n}\right) \\ 0 \end{pmatrix}. \tag{18}$$

In the remainder of this section, we will consider the local linear convergence of the semismooth quasi-Newton method (17) or (18). The following theorems will prove that the semismooth quasi-Newton method converges with linear rate under some conditions on both operator F and C. These results and their proofs follows the ideas of Sun and Han in [32]. There, the authors have investigated the Newton and quasi-Newton methods for a class of nonsmooth equations in finite dimensional spaces.

Theorem 4.1 *Let F be twice Fréchet differentiable with Lipschitzian F' and $u^* \in U$ a solution of (7) (U is an open set). Let $C(u)$ be an approximation of $F''(u)$ in which $C(u)^{-1}$ exists, and both $C(u)$ and $C(u)^{-1}$ are uniformly bounded in a small enough neighborhood of u^*. Furthermore, suppose that there exist positive constants ϵ, Δ such that if $u^0 \in U, \|u^0 - u^*\| \leqslant \epsilon$ and*

$$\|C(u^n) - F''(u^n)\| \leqslant \Delta,$$

then the sequence of points generated by (17) is well defined and converges linearly to u^* in a neighborhood of u^*.

Proof. By the hypothesis, we assume that $\|C(u)\| \leqslant \tau, \|C^{-1}(u)\| \leqslant \gamma$ for $u \in B_\rho(u^*) \subset U$. Define $\theta = \gamma(\frac{1}{\beta} + \tau) + 1$. Proving as in Lemmas 3.2 and Theorem 3.1 for D_1, we get $D_1^{-1}(u) \leqslant \theta$ for $u \in B_\rho(u^*)$. Choose $\Delta > 0$ such that

$$(\beta + 1)\theta\Delta < 1. \tag{19}$$

Since D is Newton differentiable at u^*, we can choose a positive ϵ small enough such that for any $u \in B_\epsilon(u^*) \subset B_\rho(u^*)$, we have

$$\|D(u) - D(u^*) - D'(u)(u - u^*)\| \leqslant \Delta\|u - u^*\|.$$

Noting that $\|D_1(u^n) - D'(u^n)\| \leqslant \beta\|C(u^n) - F''(u^n)\| \leqslant \beta\Delta$ for $u^n \in B_\epsilon(u^*)$, we have

$$\begin{aligned} \|u^{n+1} - u^*\| &= \|u^n - D_1^{-1}(u^n)D(u^n) - u^*\| \\ &\leqslant \|D_1^{-1}(u^n)\|\|D(u^n) - D(u^*) - D_1(u^n)(u^n - u^*)\| \\ &\leqslant \|D_1^{-1}(u^n)\|\big[\|D(u^n) - D(u^*) - D'(u^n)(u^n - u^*)\| \\ &\quad + \|D_1(u^n) - D'(u^n)\|\|u^n - u^*\|\big] \\ &\leqslant \theta\big[\Delta\|u^n - u^*\| + \beta\Delta\|u^n - u^*\|\big] \\ &\leqslant \theta(\beta + 1)\Delta\|u^n - u^*\|. \end{aligned}$$

This shows that the sequence of points generated by (17) is well defined and converges linearly to u^* in a neighborhood of u^*. ∎

Remark 4.1 *1. We do not require the condition 3) of Assumption 3.1. It implies that F'' does not need to be invertable.*

2. The hypothesis of the theorem does not require $\|C(u^n) - F''(u^n)\|$ to converge to zero, it only need

$$\|C(u^n) - F''(u^n)\| \leqslant \Delta,$$

for some Δ small enough. Therefore, it is possible that $C(u^n)$ is invertable, but $F''(u^n)$ maybe not. The invertibiity and uniformly boundedness of $C(u^n)$ can ensure by approximation methods, e.g. the following case.

3. If we take $C(u^n) = \frac{1}{\beta}I$, then the semismooth quasi-Newton method becomes a gradient-type method, see [6, 5, 4]. Note that in general if $C(u^n) = s^n I$, then two methods are not the same.

The following theorem gives the other conditions for the convergence of the semismooth quasi-Newton method.

Theorem 4.2 *Assume that Assumption 3.1 is satisfied and there exist positive constants ϵ, Δ such that if $u^0 \in U, \|u^0 - u^*\| \leqslant \epsilon$ and*

$$\|C(u^n) - F''(u^n)\| \leqslant \Delta.$$

Then, the sequence of points generated by (17) is well defined and converges linearly to u^* in a neighborhood of u^*.

Proof. From the proof of Theorem 3.1, there exists $\rho > 0$ such that $\|D'^{-1}(u)\| \leqslant \theta$ for $u \in B_\rho(u^*)$. Choose $\Delta > 0$ such that

$$6\theta\beta\Delta < 1.$$

By the hypothesis and the Newton differentiability of D, there exist $\epsilon > 0$ such that

$$\|D(u) - D(u^*) - D'(u)(u - u^*)\| \leqslant \frac{3}{2}\beta\Delta\|u - u^*\|, u \in B_\epsilon(u^*) \subset B_\rho(u^*).$$

Since $\|D_1(u^n) - D'(u^n)\| \leqslant \beta\|C(u^n) - F''(u^n)\| \leqslant \beta\Delta$, by virtue of Theorem 2.3.2 of Ortega and Rheinboldt [27] $D_1(u^n)$ is invertible and

$$\|D_1^{-1}(u^n)\| \leqslant \frac{\|D'^{-1}(u^n)\|}{1 - \|D^{-1}(u^n)[D'(u^n) - D_1(u^n)]\|} \leqslant \frac{6\theta}{5}.$$

Then for $u^n \in B_\epsilon(u^*)$, we have

$$\begin{aligned}
\|u^{n+1} - u^*\| &= \|u^n - D_1^{-1}(u^n)D(u^n) - u^*\| \\
&\leqslant \|D_1^{-1}(u^n)\|\|D(u^n) - D(u^*) - D_1(u^n)(u^n - u^*)\| \\
&\leqslant \|D_1^{-1}(u^n)\|\big[\|D(u^n) - D(u^*) - D'(u^n)(u^n - u^*)\| \\
&\quad + \|D_1(u^n) - D'(u^n)\|\|u^n - u^*\|\big] \\
&\leqslant \frac{6\theta}{5}\big[\frac{3}{2}\beta\Delta\|u^n - u^*\| + \beta\Delta\|u^n - u^*\|\big] \\
&\leqslant 3\theta\beta\Delta\|u^n - u^*\| \\
&\leqslant \frac{1}{2}\|u^n - u^*\|.
\end{aligned}$$

This shows that the sequence of points generated by (17) is well defined and converges linearly to u^* in a neighborhood of u^*. ∎

Remark 4.2 *Although the results in Theorem 4.1 and Theorem 4.2 are the same, they are slightly different in the condition on F. Theorem 4.2 requires F to satisfy Assumption 3.1. However, Theorem 4.1 does not require the condition 3) of Assumption 3.1.*

5 Two specific cases of SSQN

In the previous section, we have present the semismooth quasi-Newton method. The convergence and linear rate of the method are obtained under difference conditions. It is clear that in order to implement this method, we need a specific strategy for computing $C(u^n)$, which is an approximation of F''. Note that computing an approximation of F'' has attracted many authors, specially when they aim at solving the smooth minimization problems using the quasi-Newton method. For those problems, there have been some methods proposed, e.g. [28, 31, 19, 26]. In this section, we present two specific cases of $SSQN$ respecting to the methods for approximating F''. Firstly, we approximate F'' by $C = s^n I$ and suggest a formula for computing s^n. In this case, the semismooth quasi-Newton method ($SSQN$) looks like a gradient descent method, but it is not the same (see Remark 4.1). The second one for approximating F'' is Broyden's method. This method is very well-known in smooth minimization problems. For more detail about Broyden's method, we refer to [28, 31] and references therein.

5.1 SSQN with $C(u^n) = s^n I$

In Theorem 4.1, we assume that F'' is approximated by $C(u^n)$, where

$$C = s^n I \text{ with } s_n \in [\underline{s}, \overline{s}],\ 0 < \underline{s} \leqslant \overline{s} < \infty.$$

Here, I is the identity operator. With this method, the convergence of the semismooth quasi-Newton method depends on the choices of s^n. In the next step, we are going to give an approximation of s^n.

In each iteration, s^n should be chosen such that

$$\|C(u^n) - F''(u^n)\| = \inf_s \|sI - F''(u^n)\|.$$

This problem is not easy to exactly solve and thus we will solve it approximately. To this end, we first note that

$$\begin{aligned} \|C(u^n) - F''(u^n)\| &= \inf_s \|sI - F''(u^n)\| \\ &\approx \inf_s \sup_{\vartheta \neq 0} \frac{|s\langle \vartheta, \vartheta\rangle - \langle F''(u^n)\vartheta, \vartheta\rangle|}{\|\vartheta\|} \\ &\approx \inf_s \sup_{\vartheta \neq 0} \frac{|s\langle \vartheta, \vartheta\rangle - \langle F'(u^n + \vartheta) - F'(u^n), \vartheta\rangle|}{\|\vartheta\|} \\ &\geqslant \inf_s \frac{|s\langle \vartheta, \vartheta\rangle - \langle F'(u^n + \vartheta) - F'(u^n), \vartheta\rangle|}{\|\vartheta\|}, \vartheta \neq 0. \end{aligned} \tag{20}$$

Therefore, we shall choose s^n as the minimizer of the problem in the right hand side of (20) with $\vartheta = u^{n-1} - u^n$, i.e.

$$s^n = \frac{\langle F'(u^{n-1}) - F'(u^n), u^{n-1} - u^n\rangle}{\|u^{n-1} - u^n\|^2}. \tag{21}$$

Together with the condition $s_n \in [\underline{s}, \overline{s}]$, we choose s^n by

$$s^n = P_{[\underline{s},\overline{s}]} \frac{\langle F'(u^{n-1}) - F'(u^n), u^{n-1} - u^n \rangle}{\|u^{n-1} - u^n\|^2}, \tag{22}$$

with P is the orthogonal projection on $[\underline{s}, \overline{s}]$. For $n = 0$, we can take any value in the interval $[\underline{s}, \overline{s}]$, e.g. $s^0 = 1$.

Remark 5.1 *For finite dimensional spaces $\mathcal{H}$, s^n computed by formula (21) is Barzilai-Borwein's stepsize [3]. In this case, SSQN looks like a gradient method with Barzilai-Borwein stepsizes, but it is not the same.*

5.2 SSQN with $C(u^n)$ computed by Broyden's Method

Broyden's method have been used for the quasi-Newton method and the smoothing quasi-Newton method [28, 8, 32, 31]. Here, we are going to apply it for the semismooth quasi-Newton method in a Hilbert space setting. To this end, we combine the ideas of the authors in [32, 31]. Before presenting the method in detail, we first introduce the rank one operator $u \otimes v$ defined by

$$v \otimes u : \mathcal{H}_1 \to \mathcal{H}_2,\ v \otimes u(x) = \langle u, x \rangle v,$$

where $v \in \mathcal{H}_2, u \in \mathcal{H}_1$ ($\mathcal{H}_1, \mathcal{H}_2$ are Hilbert spaces).

Some properties of this operator are proposed in [31]. For convenience, we give them here.

Lemma 5.1 *[31, Lemma 3.2] Let $\mathcal{H}_1, \mathcal{H}_2$ be Hilbert spaces and $u_i \in \mathcal{H}_1, v_i \in \mathcal{H}_2, i = 1, 2, T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ with the adjoint operator T^*. Then, the following properties hold:*

1. $(v_1 \otimes u_1)^* = u_1 \otimes v_1 \in \mathcal{L}(\mathcal{H}_2, \mathcal{H}_1)$,
2. $(v_1 \otimes u_1)(u_2 \otimes v_2) = \langle u_1, u_2 \rangle (v_1 \otimes v_2) \in \mathcal{L}(\mathcal{H}_2, \mathcal{H}_2)$,
3. $T(u_1 \otimes v_1) = Tu_1 \otimes v_1 \in \mathcal{L}(\mathcal{H}_2, \mathcal{H}_2)$,
4. $(v_1 \otimes u_1)T^* = v_1 \otimes Tu_1 \in \mathcal{L}(\mathcal{H}_2, \mathcal{H}_2)$,
5. $\|v_1 \otimes u_1\| = \|v_1\| \|u_1\|$.

We now consider the convergence of the semismooth quasi-Newton method with F'' being approximated by Broyden's method. Using this method, the semismooth quasi-Newton method is presented by Algorithm 5.1.

Algorithm 5.1 SSQN for Broyden's case

Input: Initial guess $u^0 \in U,\ C^0 \in L(\mathcal{H})$.

for $n = 0, 1, 2, \ldots$ **do**
  $D_1(u^n) \leftarrow I - G(u^n - \beta F'(u^n))[I - \beta C^n]$
  $u^{n+1} \leftarrow u^n - D_1(u^n)^{-1} D(u^n)$
  $p^n \leftarrow u^{n+1} - u^n;\ y^n \leftarrow F'(u^{n+1}) - F'(u^n)$.
  $C^{n+1} \leftarrow C^n + \frac{1}{\langle p^n, p^n \rangle}(y^n - C^n p^n) \otimes p^n$.
end for

Output: $u = \lim u^n$.

Lemma 5.2 *Let F be twice Fréchet differentiable and F'' satisfy*

$$\|F''(u) - F''(u^*)\| \leqslant L'\|u - u^*\|, \text{ for all } u \in U.$$

Then, we have

$$\|C^{n+1} - F''(u^*)\| \leqslant \|C^n - F''(u^*)\| + \frac{L'}{2}(\|u^{n+1} - u^*\| + \|u^n - u^*\|).$$

Proof. The definition of the update implies that

$$C^{n+1} - F''(u^*) = (C^n - F''(u^*))\big(I - \frac{p^n \otimes p^n}{\langle p^n, p^n\rangle}\big) + \frac{(y^n - F''(u^*)p^n) \otimes p^n}{\langle p^n, p^n\rangle}.$$

Since $I - \frac{p^n \otimes p^n}{\langle p^n, p^n\rangle}$ is an orthogonal projection,

$$\|I - \frac{p^n \otimes p^n}{\langle p^n, p^n\rangle}\| = 1.$$

On the other hand, by the hypothesis of the lemma, we have

$$\|y^n - F''(u^*)p^n\| \leqslant \frac{L'}{2}(\|u^{n+1} - u^*\| + \|u^n - u^*\|)\|p^n\|.$$

Therefore, the lemma is proved. ∎

The linear convergence rate is obtained in the following theorem

Theorem 5.1 *Assume that Assumption 3.1 is satisfied and there exist positive constants ϵ, δ such that if $u^0 \in U, \|u^0 - u^*\| \leqslant \epsilon$ and*

$$\|C^0 - F''(u^*)\| \leqslant \delta.$$

Then, the sequence of points generated by Algorithm 5.1 is well defined and converges to u^* linearly in a neighborhood of u^* and

$$\|C^n - F''(u^n)\| \leqslant \Delta \text{ for all } n \in \mathbb{N},$$

where Δ is a positive number.

Proof. Choose ϵ and Δ as in the proof of Theorem 4.2 and restrict ϵ to be small enough such that for any $u \in B_\epsilon(u^*)$, we have

$$\|F''(u) - F''(u^*)\| \leqslant L'\|u - u^*\|, \tag{23}$$

$$3\epsilon L' \leqslant \Delta, \tag{24}$$

where L' is the Lipschitz constant of F'' in U.

Define $\delta := \Delta/2$. The proof of local linear convergence consists of showing by induction that

$$\|C^n - F''(u^*)\| \leqslant (2 - 2^{-n})\delta, \tag{25}$$

$$\|C^n - F''(u^n)\| \leqslant \Delta. \tag{26}$$

For $n = 0$, it is easy to show that (25) and (26) hold. Assume that (25) and (26) are satisfied for $n = 0, 1, \ldots, i$. From the proof of Theorem 4.2, for $n = 0, 1, \ldots, i$, we have (setting $e^n = u^n - u^*$)

$$\|e^{n+1}\| \leqslant \frac{1}{2}\|e^n\|. \tag{27}$$

For $n = i + 1$, by Lemma 5.2 and the induction hypothesis, we have

$$\|C^{n+1} - F''(u^*)\| \leqslant \|C^n - F''(u^*)\| + \frac{L'}{2}(\|e^{n+1}\| + \|e^n\|)$$
$$(2 - 2^{-n})\delta + \frac{3L'}{4}\|e^n\|. \tag{28}$$

By (27) and $\|e^0\| \leqslant \epsilon$ it follows that

$$\|e^n\| \leqslant 2^{-n}\|e^0\| \leqslant 2^{-n}\epsilon.$$

Substituting this into (28) and using (24) gives

$$\begin{aligned}\|C^{n+1} - F''(u^*)\| &\leqslant (2 - 2^{-n})\delta + \frac{3L'}{4}2^{-n}\epsilon \\ &\leqslant (2 - 2^{-n} + 2^{-(n+1)})\delta = (2 - 2^{-(n+1)})\delta.\end{aligned}$$

To complete the induction, we verify (26). We have

$$\begin{aligned}\|C^{n+1} - F''(u^{n+1})\| &\leqslant \|C^{n+1} - F''(u^*)\| + \|F''(u^{n+1}) - F''(u^*)\| \\ &\leqslant (2 - 2^{-(n+1)})\delta + 2^{-(n+1)}L'\epsilon \\ &\leqslant (2 - 2^{-(n+1)})\frac{\Delta}{2} + \frac{1}{3}2^{-(n+1)}\Delta \\ &< \Delta.\end{aligned}$$

So (26) is proved. Therefore, the local linear convergence follows from Theorem 4.2. ■

Remark 5.2 *1. For finite dimensional spaces $\mathcal{H}$, we can prove that Algorithm 5.1 converges superlinearly. The proof is similar to that of [28, Theorem 8.2.2] or [32, Corollary 4.1]. In general Hilbert spaces $\mathcal{H}$, Algorithm 5.1 can be proved to converge superlinearly under additional conditions as in [31].*

2. Similar to Broyden's method, some other methods for approximating F'' might be applied, e.g. the formulas in [12].

6 SSN and SSQN as Active Set Methods

In previous sections, we discussed about the semismooth Newton and quasi-Newton methods. They can be represented as the iteration

$$u^{n+1} = u^n - D_1^{-1}(u^n)D(u^n), \tag{29}$$

where $D_1(u) = I - G(u - \beta F'(u))[I - \beta C(u)]$. If $C(u^n) = F''(u^n)$ for all n, then iteration (29) is the semismooth Newton method, otherwise it is the semismooth quasi-Newton method. Naturally, two methods can also be interpreted as active set methods that are stated in Algorithm 6.1.

Remark 6.1 *1. Algorithm 6.1 is very efficient because we only solve a small linear system in Step 14 for each iteration. Note that Step 14 requires the invertibility of operators $\mathcal{M}_{\mathbb{A}^n\mathbb{A}^n}$ in each iteration. Their sufficient conditions are given in Theorem 3.1, Theorem 4.1, Theorem 4.2 and Theorem 5.1.*

Algorithm 6.1 SSN and SSQN as active set methods

Input: Initial guess $u^0 \in U$, choose β, set $n := 0$ and done:=false.

1: **while** $n < n_{max}$ and not done **do**
2: Calculate the active set and inactive set
3: $\mathbb{A}_+^n \leftarrow \{k \in \Lambda : [u^n - \beta F'(u^n)]_k > \beta\omega_k\}$,
4: $\mathbb{A}_-^n \leftarrow \{k \in \Lambda : [u^n - \beta F'(u^n)]_k < -\beta\omega_k\}$,
5: $\mathbb{A}^n \leftarrow \mathbb{A}_+^n \cup \mathbb{A}_-^n; \mathbb{I}^n \leftarrow \Lambda \backslash \mathbb{A}^n$.
6: Compute the residual
7: $r^n := D(u^n) \leftarrow u^n - \mathbb{S}_{\beta w}(u^n - \beta F'(u^n))$.
8: **if** $\|r^n\| \leqslant \epsilon$ **then**
9: *done* $\leftarrow$ *true*.
10: **else**
11: Compute $C(u^n)$ and represent in the form
12: $C(u^n) \leftarrow \begin{pmatrix} \mathcal{M}_{\mathbb{A}^n\mathbb{A}^n} & \mathcal{M}_{\mathbb{A}^n\mathbb{I}^n} \\ \mathcal{M}_{\mathbb{I}^n\mathbb{A}^n} & \mathcal{M}_{\mathbb{I}^n\mathbb{I}^n} \end{pmatrix}$.
13: Set $u_{\mathbb{I}^n}^{n+1} \leftarrow 0$ and solve the equation
14: $\mathcal{M}_{\mathbb{A}^n\mathbb{A}^n}\delta u_{\mathbb{A}^n} = \begin{pmatrix} [F'(u^n) + w]\big|_{\mathbb{A}_+^n} \\ [F'(u^n) - w]\big|_{\mathbb{A}_-^n} \end{pmatrix} - \mathcal{M}_{\mathbb{A}^n\mathbb{I}^n} u_{\mathbb{I}^n}^n$
15: Compute $u_{\mathbb{A}^n}^{n+1} \leftarrow u_{\mathbb{A}^n}^n - \delta u_{\mathbb{A}^n}$.
16: Set $n \leftarrow n + 1$
17: **end if**
18: **end while**

Output: $u = u^n$.

2. *In the case $\mathcal{M}_{\mathbb{A}^n\mathbb{A}^n}$ are bad-conditioned (e.g. non-invertible), instead of Step 14, we solve the following linear system*

$$\big(\mathcal{M}_{\mathbb{A}^n\mathbb{A}^n}^t \mathcal{M}_{\mathbb{A}^n\mathbb{A}^n} + \nu^n I\big)\delta u_{\mathbb{A}^n} = \mathcal{M}_{\mathbb{A}^n\mathbb{A}^n}^t \Big(\begin{pmatrix} [F'(u^n) + w]\big|_{\mathbb{A}_+^n} \\ [F'(u^n) - w]\big|_{\mathbb{A}_-^n} \end{pmatrix} - \mathcal{M}_{\mathbb{A}^n\mathbb{I}^n} u_{\mathbb{I}^n}^n \Big), \tag{30}$$

where ν^n are enough small positive numbers and $\mathcal{M}_{\mathbb{A}^n\mathbb{A}^n}^t$ is the transpose matrix of $\mathcal{M}_{\mathbb{A}^n\mathbb{A}^n}$. This technique is used in Tikhonov regularization for linear inverse problems, see e.g. [13].

7 Numerical example

We now apply the algorithms to the following parameter identification problem of elliptic equations: estimate the coefficient σ from a measurement of the solution ϕ in the elliptic boundary problem

$$\begin{aligned} -\operatorname{div}(\sigma\nabla\phi) &= y \text{ in } \Omega \subset \mathbb{R}^2, \\ \phi &= 0 \text{ on } \partial\Omega, \end{aligned} \tag{31}$$

where $y \in \mathbf{L}^2(\Omega)$.

As data we assume, that ϕ^δ is given, where ϕ^δ is the solution of the elliptic equation with parameter a^* but perturbed right hand side y^δ with $\|y - y^\delta\|_{L_2} \leq \delta$. Hence, the available data satisfies $\|\phi^* - \phi^\delta\|_{H^1(\Omega)} \leq C\delta$, where C is a positive constant. Our task is to determine an approximation of a^* from ϕ^δ.

A number of papers, such as [17, 36, 21, 14, 22, 30, 23, 10, 34, 1, 7], have examined this problem or variations of it, see also [13] and [2].

We denote

$$\mathbb{A} = \{\sigma \in \mathbf{L}^\infty(\Omega) : 0 < \lambda \leqslant \sigma \leqslant \lambda^{-1},\ \mathrm{supp}(\sigma - \sigma^0) \subset\subset \Omega\}.$$

and define $F_D : \mathbb{A} \subset L^\infty(\Omega) \to \mathbf{H}_0^1(\Omega), \sigma \mapsto \phi$, the solution of (31). The parameter identification problem regularized by sparsity constraints leads to the following minimization problem

$$\min_{\sigma \in \mathbf{L}^2(\Omega)} \Theta(\sigma) = \int_\Omega \sigma |\nabla F_D(\sigma) - \nabla \phi^\delta|^2 dx + \alpha \sum_{k \in \Lambda} \omega_k |\langle \sigma - \sigma^0, \varphi_k \rangle|, \tag{32}$$

where $\{\varphi_k\}$ is the basis consisting the finite linear elements.

It is known that $F(\sigma) = \int_\Omega \sigma |\nabla F_D(\sigma) - \nabla \phi^\delta|^2 dx$ is convex, twice Fréchet continuous differentiable; F' is Lipschitz continuous; F' and F'' are given by (see, e.g. [17, 36, 20])

$$F'(\sigma)\vartheta = -\int_\Omega \vartheta (|\nabla F_D(\sigma)|^2 - |\nabla \phi^\delta|^2) dx,\ F''(\sigma)(\vartheta, \vartheta) = \int_\Omega \sigma \nabla |F_D'(\sigma)\vartheta|^2 dx.$$

For illustrating our algorithms, we assume that Ω is the unit disk and

$$\sigma^*(x_1, x_2) = \begin{cases} 4, (x, y) \in B_{0.4}(0, 0.3) \\ 1, otherwise \end{cases},\ y(x_1, x_2) = 4\sigma^*.$$

where $B_r(x_1, x_2)$ is the disk with the center at (x_1, x_2) and the radius r.

To obtain ϕ^* and ϕ^δ, we solve (31) by the finite element method on a mesh with 1272 triangles. The solution of (31) and parameter σ is represented by piecewise linear finite elements. Note that for stabilization, before solving (31) we have cut off values of σ^n that below $\sigma^0 = 1$ in each iteration. This technique bases on the prior information of the parameter that its values are not below σ^0. Another technique is Sobolev-gradient [21], but we do not use it here.

For the numerical example we set $\omega_k = 1,\ \forall k$ and $\alpha := 5 \cdot 10^{-5}$ and the algorithms are used under the following setting

1. Algorithm 6.1 (SSQN.I) with $C^0 = I$ and $C^n = s^n I$ where s^n is computed by (22) and $[\underline{s}, \overline{s}] := [5.10^{-2}, 5.10^2]$.
2. Algorithm 6.1 (SSQN.B) with $C^0 = I$ and C^n computed by Broyden's method, where Step 14 in Algorithm 6.1 is replaced by (30) with $\nu^n := 10^{-3}$.

We measure the convergence of the computed minimizers to the true parameter σ^* by considering the mean square error sequence

$$MSE(\sigma^n) = \int_\Omega (\sigma^n - \sigma^*)^2 dx.$$

We first discuss results without noise, i.e. $\phi^\delta = \phi^*$. Figure 1 shows that $\{\|D(\sigma^n)\|_{\mathbf{L}^2}\}$ and $\Theta(\sigma^n)$ in two algorithms do not decrease monotonically. However, they show that the sequences $\{\sigma^n\}$ in two algorithms converge to the minimizer of the functional Θ and the convergence rate in SSQN.B is lightly faster than that in SSQN.I. By the decrease of $MSE(\sigma^n)$, the the sequences $\{\sigma^n\}$ also converge to σ^*. Here, the sequence $\{\sigma^n\}$ in SSQN.B converges to σ^* faster than that in SSQN.I. This agrees with the theory results, which show that SSQN.I converges with linear rate and SSQN.B converges with superlinear rate.

Figure 2 illustrates a^* and a^n with $n = 300$ in these algorithms. It shows that the algorithms have reconstructed the parameter a^* very accurately.

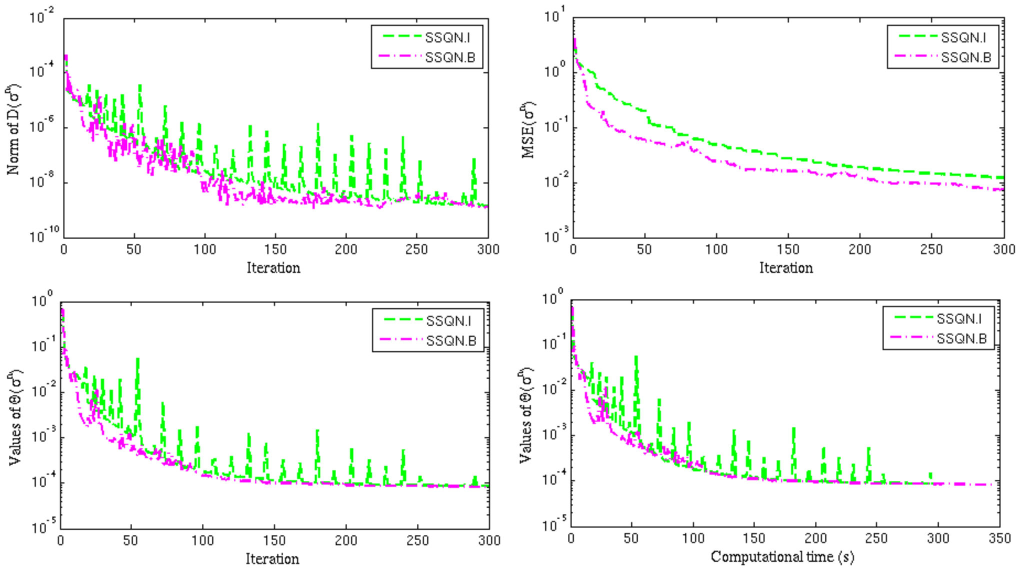

Figure 1: The values of $\|D(\sigma^n)\|_{\mathbf{L}^2(\Omega)}$, $MSE(\sigma^n)$ and $\Theta(\sigma^n)$ in the algorithms.

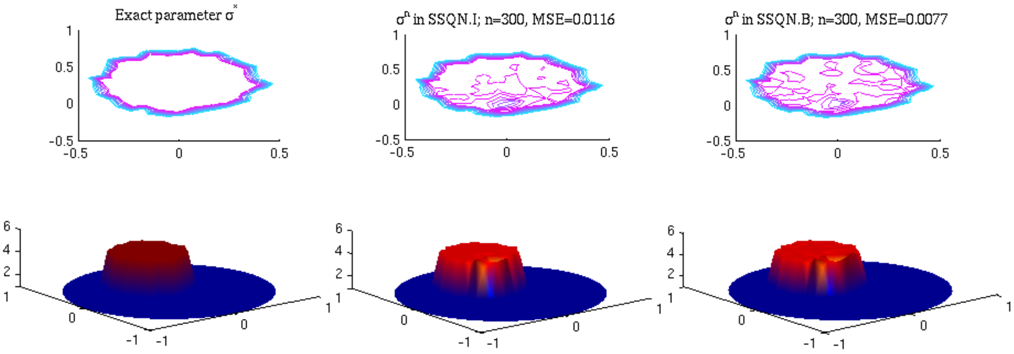

Figure 2: 3D-plots and contour plots of σ^*, σ^n.

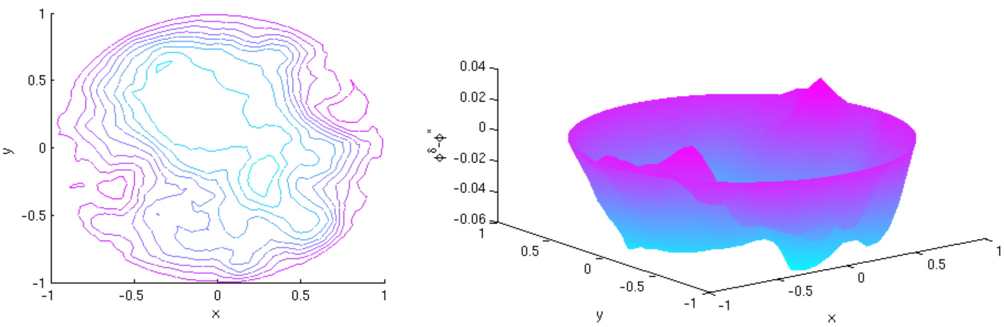

Figure 3: 3D-plot and contour plot of $\phi^\delta - \phi^*$ with $\|\phi^\delta - \phi^*\|_{\mathbf{H}^1(\Omega)} = 9.85\%$.

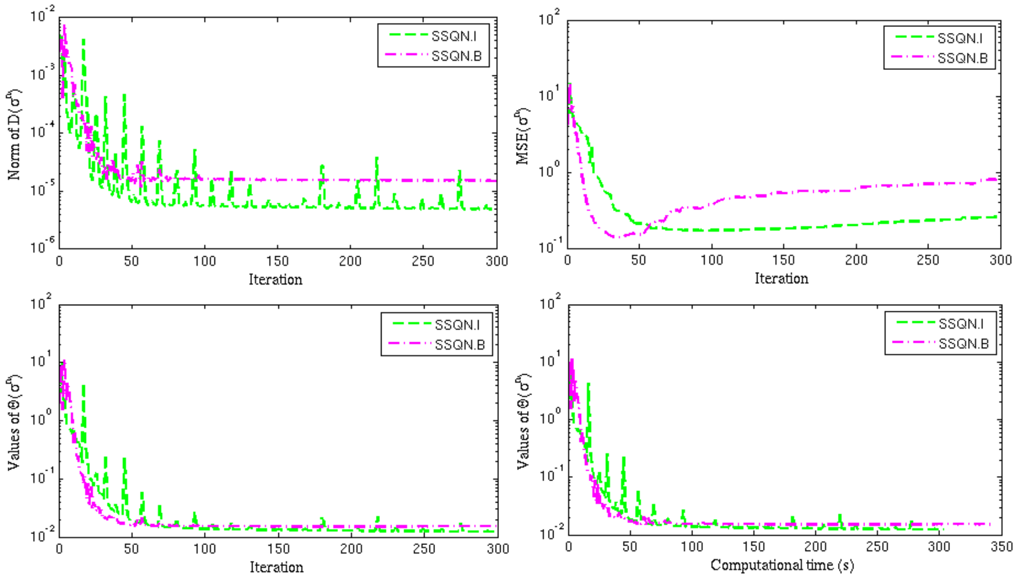

Figure 4: The values of $\|D(\sigma^n)\|_{\mathbf{L}^2(\Omega)}$, $MSE(\sigma^n)$ and $\Theta(\sigma^n)$ in the algorithms.

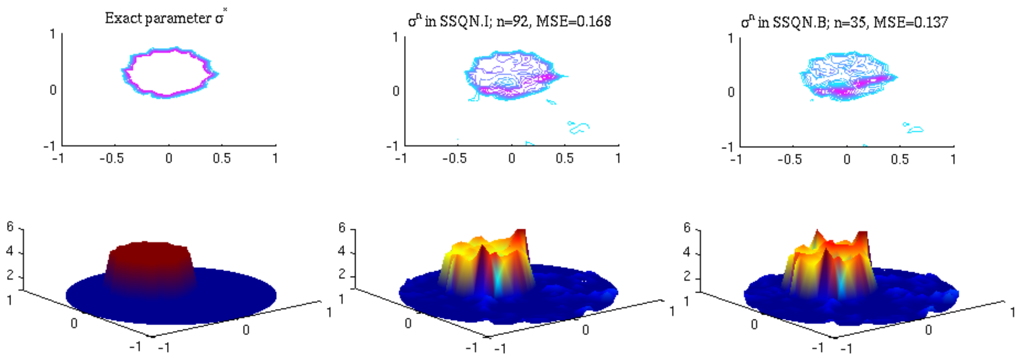

Figure 5: 3D-plots and contour plots of σ^*, σ^n.

Secondly, we work with noisy data. Here, we have created the data ϕ^δ with $\|\phi^\delta - \phi^*\|_{\mathbf{H}^1(\Omega)} = 9.85\%$. The difference $\phi^\delta - \phi^*$ is illustrated in Figure 3.

In Figure 4 the decrease of $\Theta(\sigma^n)$ and $\|D(\sigma^n)\|_{\mathbf{L}^2}$ show that the sequences $\{\sigma^n\}$ generated in the algorithms converge to the minimizer of Θ. But the appearance of noise makes the minimizing sequences disconverge to the true parameter σ^*. The sequences $MSE(\sigma^n)$ in the algorithms show that $\{\sigma^n\}$ tend to σ^* in the first steps, but after that they disconverges. It is evident because the sequences $\{\sigma^n\}$ converge to the minimizer of Θ, which is different from σ^*. Therefore, in the noisy data case, σ^n with n large might be not a good approximation of σ^* and one stopping criterion is needed, which ensures that σ^n with a certain n is a good approximation of σ^*. The sequence $\{\sigma^n\}$ in SSQN.B tends to σ^* faster than that in SSQN.I in first iterations and the minimum value of $\{MSE(\sigma^n)\}$ in SSQN.B is smaller than that in SSQN.I. Therefore, with a suitable stopping criterion SSQN.B obtains a better approximation of σ^* than SSQN.I. Moreover, this approximation can not obtain by SSQN.I.

Figure 5 illustrates σ^* and σ^n, where n is taken with respect to the minimum values of $MSE(\sigma^n)$ in SSQN.I and SSQN.B. Here, $MSE(\sigma^n)$ in SSQN.B is smaller than that in SSQN.I.

8 Conclusion and future works

We have considered the minimization problem

$$\min_{u\in\mathcal{H}} F(u) + \sum_{k\in\Lambda} \omega_k |\langle u, \varphi_k\rangle|,$$

where $\mathcal{H}$ is a Hilbert space, $F : Dom(F) \subset \mathcal{H} \to \mathbb{R}$ is Fréchet differentiable and $\{\varphi_k\}_{k\in\Lambda}$ is an orthonormal basis (or frame) of $\mathcal{H}$.

Instead of solving the above problem, we aim at solving its optimality condition equation given by

$$D(u) := u - \mathbb{S}_{\beta w}(u - \beta F'(u)) = 0.$$

We have proved that $D(\cdot)$ is Newton differential and proposed the conditions for the convergence of the semismooth Newton and quasi-Newton methods. The convergence rates of two methods have also analyzed. The methods can represent as the following iteration

$$u^{n+1} = u^n - D_1^{-1}(u^n)D(u^n),$$

where $D_1(u) = I - G(u - \beta F'(u))[I - \beta C(u)]$. If $C = F''$, then the iteration is the semismooth Newton method, otherwise it is the quasi-Newton method.

The advantage of the methods are not only their fast convergence but also they can represented as the active set methods. Thus, in each iteration, the methods only need to solve a small linear system. We have also proposed two specific cases of the semismooth quasi-Newton method for implementing in practice. The theory as well as the numerical example have showed that SSQN.B converges faster than the gradient-type method (SSQN.I).

For further work, in order to obtain the global convergence of these algorithms, the above algorithms might be modified as follows

$$u^{n+1} = u^n - t^n D_1^{-1}(u^n)D(u^n), \tag{33}$$

with some choice of $t^n > 0$. Similar as the quasi-Newton method for smooth minimization problems, this iteration might be proved to globally converge under some conditions on F, C and t^n.

Furthermore, other methods for approximating of F in the quasi-Newton method might be used, e.g. the methods in the DFP-algorithm [19] and BFGS-algorithm [26].

References

[1] R. Acar and C. R. Vogel. Analysis of bounded variation penalty methods for ill-posed problems. *Inverse Problems*, 10:1217–1229, 1996.

[2] H. T. Banks and K. Kunisch. *Estimation Techniques for Distributed Parameter Systems.* Systems and Control: Foundations and Applications Series. Birkhauser, Boston, 1989.

[3] J. Barzilai and J. M. Borwein. Two-point step size gradient methods. *IMA J. Numer. Anal.*, 8:141–148, 1988.

[4] T. Bonesky, K. Bredies, D. A. Lorenz, and P. Maass. A generalized conditional gradient method for nonlinear operator equations with sparsity constraints. *Inverse Problems*, 33:2041–2058, 2007.

[5] K. Bredies and D. A. Lorenz. Linear convergence of iterative soft thresholding. *Fourier Anal. Appl.*, 14:813–837, 2008.

[6] K. Bredies, D. A. Lorenz, and P. Maass. A generalized conditional gradient method and its connection to an iterative shrinkage method. *Computational Optimization and Application*, 42(2):173–193, 2009.

[7] T. F. Chan and X. Tai. Level set and total variation regularization for elliptic inverse problems with discontinuous coefficients. *Journal of Computational Physics*, 193:40–66, 2003.

[8] X. Chen. Superlinear convergence of smoothing quasi-Newton methods for nonsmooth equations. *J. Comp. Appl. Math*, 80:105–126, 1996.

[9] X. Chen. Superlinear convergence and smoothing quasi-newton methods for nonsmooth equations. *Comput. Appl. Math.*, 80:105–26, 1997.

[10] Z. Chen and J. Zou. An augmented Lagrangian method for identifying discontinuous parameters in elliptic systems. *SIAM Journal of Control and Optimization*, 37:892–910, 1999.

[11] I. Daubechies, M. Defrise, and C. Demol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math*, 57:1413–1541, 2004.

[12] J. E. Dennis and J. J. Moré. A characterization of superlinear convergence and its application to quasi-Newton methods. *Inverse Problems*, 28(126):549–560, 1974.

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

[14] R. Falk. Error estimates for the numerical identification of a variable coefficient. *Mathematics of Computation*, 40(3):537–546, July 1983.

[15] M. Grasmair, M. Haltmeier, and O. Scherer. Sparsity regularization with l^q penalty term. *Inverse Problems*, 24:055020, 2008.

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

[17] D. N. Hào and T. N. T. Quyen. Convergence rates for Tikhonov regularization of coefficient identification problems in Laplace-type equation. *Inverse Problems*, 26:125014, 2010.

[18] M. Hintermüller, K. Ito, and K. Kunisch. The primal-dual active set strategy as a semismooth Newton method. *SIAM J. Optim.*, 13(3):865–88, 2003.

[19] L. B. Horwitz and P. E. Sarachik. Davidon's methods in Hilbert space. *SIAM Journal on Applied Mathematics*, 16:676–695, 1968.

[20] I. Knowles. A variational algorithm for electrical impedance tomography. *Inverse Problems*, 14:1513–1525, 1998.

[21] I. Knowles. Parameter identification for elliptic problems. *Journal of Computational and Applied Mathematics*, 131:175–194, 2001.

[22] R. V. Kohn and B. Lowe. A variational method for parameter identification. *Mathematical Modeling and Numerical Analysis*, 22:119–158, 1988.

[23] C. Kravaris and J. H. Seinfeld. Identification of parameters in distributed parameter systems by regularization. *SIAM Journal of Control and Optimization*, 23:217–241, 1985.

[24] D. A. Lorenz. Convergence rates and source conditions for Tikhonov regularization with sparsity constraints. *J. Inv. Ill-posed problems*, 16:463–478, 2008.

[25] D. A. Lorenz, P. Maass, and P. Q. Muoi. Gradient descent methods based on quadratic approximations of tikhonov functionals with sparsity constraints: theory and numerical comparison of stepsize rules. preprint, 2011.

[26] R. V. Mayorga and V. H. Quintana. A family of variable metric methods in function space, without exact line searches. *Journal of Optimization Theory and Applications*, 31:303–329, 1980.

[27] J. M. Ortega and W. C. Rheinboldt. *Iteration Solution of Nonlinear Equations in Several Variables*. Academic Press, New York, 1970.

[28] J. M. Ortega and W. C. Rheinboldt. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Englewood Cliffs, NJ, 1983.

[29] R. Ramlau and G. Teschke. A Tikhonov-based projection iteration for nonlinear ill-posed problems with sparsity constraints. *Numer. Math.*, 104:177–203, 2006.

[30] G. R. Richter. Numerical identification of a spatially varying diffusion coefficient. *Mathematics of Computation*, 36:375–386, 1981.

[31] E. Sachs. Broyden's method in Hilbert space. *Math. Programming*, 35:71–82, 1986.

[32] D. Sun and J. Han. Newton and quasi-Newton methods for a class of nonsmooth equations and related problems. *SIAM J. Optim.*, 7(2):463–480, 1997.

[33] M. Ulbrich. Semismooth Newton methods for operator equations in function spaces. *SIAM J. Control Optim.*, 13(3):805–842, 2003.

[34] L. W. White and J. Zhou. Continuity and uniqueness of regularized output least squares optimal estimators. *Journal of Mathematical Analysis and Applications*, 196:53–83, 1994.

[35] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo. Sparse reconstruction by separable approximation. *Trans. Sig. Proc.*, 57:2479–2493, 2009.

[36] J. Zou. Numerical methods for elliptic inverse problems. *International Journal of Computer Mathematics*, 70:211–232, 1998.