A semi-smooth Newton method for nonlinear parameter identification problems with impulsive noise

Christian Clason* Bangti Jin†

February 28, 2011

This work is concerned with nonlinear parameter identification in partial differential equations subject to impulsive noise. To cope with the non-Gaussian nature of the noise, we consider a model with L^1 fitting. However, the non-smoothness of the problem makes its efficient solution challenging. By approximating this problem using a family of smoothed functionals, a semi-smooth Newton method becomes applicable. In particular, its super-linear convergence is proved under a second-order condition. The convergence of the solutions to the approximating problems as the smoothing parameter goes to zero is shown. A strategy for selecting the regularization parameter based on a balancing principle is suggested. The efficiency of the method is illustrated on several benchmark inverse problems of recovering coefficients in elliptic differential equations, for which one- and two-dimensional numerical examples are presented.

1. Introduction

We are interested in the nonlinear inverse problem

$$S(u) = y^\delta,$$

where $S : \mathcal{X} \to \mathcal{Y}$ is the parameter-to-observation mapping and y^δ represents experimental measurements corrupted by impulsive noise. Throughout we assume that the space $\mathcal{Y}$ compactly embeds into L^q for some $q > 2$, y^δ is bounded almost everywhere, and $\mathcal{X}$ is a Hilbert space. The spaces $\mathcal{X}$ and $\mathcal{Y}$ are defined on the bounded domains $\omega \subset \mathbb{R}^n$ and $D \subset \mathbb{R}^m$, respectively. Such models arise naturally in distributed parameter identification for differential equations, where typically $\mathcal{Y}$ is $H^1(D)$ or $H^{\frac{1}{2}}(D)$ and $\mathcal{X}$ is $L^2(\omega)$ or $H^1(\omega)$.

*Institute for Mathematics and Scientific Computing, University of Graz, Heinrichstrasse 36, A-8010 Graz, Austria, (christian.clason@uni-graz.at).

†Department of Mathematics and Institute for Applied Mathematics and Computational Science, Texas A&M University, College Station 77843-3368, Texas, USA (btjin@math.tamu.edu).

The noise model for the measured data y^δ plays a critical role in formulating and solving the problem. In practice, an additive Gaussian noise model is customarily adopted, which leads to the standard L^2 fitting. However, non-Gaussian (e.g., impulsive) noises may occur in practical applications. For instance, the noise may follow a Laplace distribution as in certain inverse problems arising in signal processing [3]. Noise models of impulsive type, e.g., salt-and-pepper or random-valued noise, are characterized by the fact that the given data contains a (possibly large) number of outliers, while other points are uncorrupted. Generally, these measurement errors are attributed to uncertainties in instrument calibration and physical limitations of the devices and experimental conditions, and occur in image processing because of malfunctioning pixels in camera sensors, faulty memory locations in hardware, or transmission in noisy channels [5].

As was demonstrated in, e.g., [12], L^2 fitting is inadequate for coping with such noise, as the use of the L^2 model can lead to inacceptable reconstruction errors in such cases. Statistically speaking, L^1 fitting is more robust to outliers than the more conventional L^2 counterpart in the sense that outliers have less influence on the solution [19], and is deemed suitable when the data possibly contains outliers. Consequently, models involving L^1 fitting have recently received considerable interest, e.g., in imaging [12, 22] as well as parameter identification [6]. These considerations motivate the adopting the model

$$(\mathcal{P}) \qquad \min_{u\in\mathcal{X}} \left\{ \mathcal{J}_\alpha(u) \equiv \|S(u) - y^\delta\|_{L^1} + \frac{\alpha}{2}\|u\|_{\mathcal{X}}^2 \right\}.$$

We are mainly interested in various structural properties of the L^1-norm fitting compared with the more conventional L^2-norm counterpart. Our main goal in this work is to resolve the computational obstacle posed by the non-differentiability of the L^1-norm, such that Newton-type methods are applicable when the operator S has the necessary differentiability properties.

Due to the practical significance of L^1 models, there has been a growing interest in analyzing their properties as well as in developing efficient minimization algorithms. A number of recent works have addressed the analytical properties of models with L^1 fitting, explaining their superior performance over the standard model for certain types of noise and elaborating the geometrical structure of the minimizers in the context of image denoising [2, 7, 14, 31], i.e., when S is the identity operator. In addition, several efficient algorithms [11–13, 29] have been developed for such problems.

However, all these works are only concerned with linear inverse problems, and their analysis and algorithms are not directly applicable to the nonlinear case of our interest. The optimality system is not differentiable in a generalized sense, and thus can not be solved directly with a (semi-smooth) Newton method. We consider a smoothed variant, and prove the convergence as the smoothing parameter tends to zero. The smoothed optimality system is solved by a semi-smooth Newton method, and its superlinear local convergence is established under a second-order condition. To the best of our knowledge, this work represents a first investigation on L^1 fitting with general nonlinear inverse problems. The applicability of the proposed approach and its numerical performance is illustrated with several benchmark problems for distributed parameter identification for elliptic partial differential equations.

The rest of this work is organized as follows. In the remainder of this section, we introduce a selection of model problems for which our approach is applicable. In Section 2,

we derive the optimality system of problem ($\mathcal{P}$) for general nonlinear inverse problems. The approximating problems, convergence as the smoothing parameter tends to zero, and their numerical solution using a semi-smooth Newton method are studied in Section 3. We also discuss the important issue of choosing suitable regularization and smoothing parameters. Finally, in Section 4, we present numerical results for our model problems.

1.1. Inverse potential problems

A first nonlinear model problem consists in recovering the potential term in an elliptic equation. Let $\Omega \subset \mathbb{R}^d$ be an open bounded domain with a Lipschitz boundary Γ. We consider the equation

$$\begin{cases} -\Delta y + uy = f & \text{in } \Omega, \\ \dfrac{\partial y}{\partial n} = 0 & \text{on } \Gamma. \end{cases} \tag{1.1}$$

The inverse problem is to recover the potential u defined on $\omega = \Omega$ from noisy observational data y^δ in the domain $D = \Omega$, i.e., S maps $u \in \mathcal{X} = L^2(\Omega)$ to the solution $y \in \mathcal{Y} = H^1(\Omega)$ of (1.1). Such problems arise in heat transfer, e.g., damping design [26] and identifying heat radiative coefficient [28]. We shall seek u in the admissible set $U = \{u \in L^\infty(\Omega) : u \geq c\} \subset \mathcal{X}$ for some fixed $c > 0$.

1.2. Inverse Robin coefficient problem

Our second example considers the recovery of a Robin boundary condition from boundary observation. Let $\Omega \subset \mathbb{R}^2$ be an open bounded domain with a Lipschitz boundary Γ consisting of two disjoint parts Γ_i and Γ_c. We consider the equation

$$\begin{cases} -\Delta y = 0 & \text{in } \Omega, \\ \dfrac{\partial y}{\partial n} = f & \text{on } \Gamma_c, \\ \dfrac{\partial y}{\partial n} + uy = 0 & \text{on } \Gamma_i. \end{cases} \tag{1.2}$$

The inverse problem consists in recovering the Robin coefficient u defined on $\omega = \Gamma_i$ from noisy observational data y^δ on the boundary $D = \Gamma_c$, i.e., S maps $u \in \mathcal{X} = L^2(\Gamma_i)$ to $y|_{\Gamma_c} \in \mathcal{Y} = H^{\frac{1}{2}}(\Gamma_c)$, where $v \mapsto v|_{\Gamma_c}$ denotes the Dirichlet trace operator and y is the solution to (1.2). This class of problems arises in corrosion detection and thermal analysis of quenching processes [6, 21]. We shall seek u in the admissible set $U = \{u \in L^\infty(\Gamma_i) : u \geq c\} \subset \mathcal{X}$ for some fixed $c > 0$.

1.3. Inverse diffusion coefficient problem

Our last example, identification of a diffusion coefficient, addresses stronger regularization for the parameter. Let $\Omega \subset \mathbb{R}^2$ be an open bounded domain with a smooth boundary Γ. We consider the equation

$$\begin{cases} -\nabla \cdot (u \nabla y) = f & \text{in } \Omega, \\ y = 0 & \text{on } \Gamma. \end{cases} \tag{1.3}$$

with $f \in L^q(\Omega)$ for some $q > 2$. The inverse problem consists in recovering the diffusion coefficient u within $\omega = \Omega$ from the noisy observational data y^δ in the domain $D = \Omega$, i.e., S maps $u \in \mathcal{X} = H^1(\Omega)$ to the solution $y \in \mathcal{Y} = W_0^{1,q}(\Omega)$, $q > 2$, of (1.3). Such problems arise in estimating the permeability of underground flow and the conductivity of heat transfer [4,9,30]. We shall seek u in the admissible set $U = \{u \in L^\infty(\Omega) : \lambda \le u \le \lambda^{-1}\} \cap \mathcal{X}$ for some fixed $\lambda \in (0, 1)$.

Remark 1.1. In this work, we assume that the true solution $u^\dagger$ of the inverse problem lies in the interior of U and do not explicitly enforce the constraint $u \in U$, in order to focus the presentation on the treatment of the non-smoothness inherent in the L^1-fitting problem. There is no fundamental difficulty in including this constraint in the optimization, however, in which case the first equality in the optimality conditions (OS) should be replaced by a variational inequality. When the domain of definition is given by box constraints (as in the model problems), the modified optimality system can still be solved using a semi-smooth Newton method after applying a Moreau–Yosida regularization, cf. [17].

These model problems share the following properties, which are verified in Appendix A and are sufficient to guarantee the applicability of our approach.

(A1) The operator S is uniformly bounded in $\mathcal{X}$ and completely continuous: If for $u \in U$, the sequence $\{u_n\} \subset U$ satisfies $u_n \rightharpoonup u$ in $\mathcal{X}$, then

$$S(u_n) \to S(u) \quad \text{in } L^2(D).$$

(A2) S is twice Fréchet differentiable.

(A3) There exists a constant $C > 0$ such that for all $u \in U$ and $h \in \mathcal{X}$ there holds

$$\|S'(u)h\|_{L^2} \le C\|h\|_{\mathcal{X}}.$$

(A4) There exists a constant $C > 0$ such that for all $u \in U$ and $h \in \mathcal{X}$ there holds

$$\|S''(u)(h,h)\|_{L^2} \le C\|h\|_{\mathcal{X}}^2.$$

The twice differentiability of S in (A2) is required for a Newton method, see Section 3.2, and ensures strict differentiability required for the chain rule, see the proof of Theorem 2.1. The a priori estimate in (A3) is employed in analyzing the convergence of the approximate solutions, while (A4) will be used to show local superlinear convergence of the semi-smooth Newton method.

2. Optimality system

The existence of a minimizer $u_\alpha \in \mathcal{X}$ of problem ($\mathcal{P}$) for fixed $\alpha > 0$ follows from standard arguments (cf., e.g., [27]) due to the uniform boundedness and complete continuity of S. We next derive the necessary first-order optimality conditions for u_α.

Theorem 2.1. *For any local minimizer $u_\alpha \in \mathcal{X}$ of problem ($\mathcal{P}$) there exists a $p_\alpha \in \mathrm{L}^\infty(D)$ with $\|p_\alpha\|_{\mathrm{L}^\infty} \leq 1$ such that the following relations hold:*

$$\text{(OS)} \qquad \begin{cases} S'(u_\alpha)^* p_\alpha + \alpha j(u_\alpha) = 0, \\ \langle S(u_\alpha) - y^\delta, p - p_\alpha \rangle_{\mathrm{L}^2} \leq 0 \quad \text{for all } \|p\|_{\mathrm{L}^\infty} \leq 1. \end{cases}$$

Here $S'(u)^*$ denotes the adjoint of $S'(u)$ with respect to $\mathrm{L}^2(D)$, and $j : \mathcal{X} \to \mathcal{X}^*$ is the (linear) duality mapping, i.e., $j(u) = \partial(\frac{1}{2}\|u\|_{\mathcal{X}}^2)$. Note that both $S(u)$ and y^δ are in $\mathrm{L}^2(D)$, and hence the duality pairing $\langle S(u) - y^\delta, p \rangle_{\mathrm{L}^1, \mathrm{L}^\infty}$ coincides with the standard L^2-inner product.

Proof. Upon setting

$$\begin{aligned} \mathcal{F} &: \mathcal{X} \to \mathbb{R}, & u &\mapsto \frac{\alpha}{2}\|u\|_{\mathcal{X}}^2, \\ \mathcal{G} &: \mathrm{L}^1(D) \to \mathbb{R}, & v &\mapsto \|v\|_{\mathrm{L}^1}, \end{aligned}$$

we have that

$$\mathcal{J}_\alpha(u) = \mathcal{F}(u) + \mathcal{G}(S(u) - y^\delta).$$

Since the operator S is twice Fréchet differentiable ((A2), which implies strict differentiability) and $\mathcal{G}$ is real-valued and convex, the sum and chain rules for the generalized gradient [10, Thms. 2.3.3, 2.3.10] yield that for all $u \in \mathcal{X}$, the functional $\mathcal{J}_\alpha$ is Lipschitz continuous near u and the relation

$$\partial \mathcal{J}_\alpha(u) = \mathcal{F}'(u) + S'(u)^* \partial \mathcal{G}(S(u) - y^\delta)$$

holds. The necessary condition $0 \in \partial \mathcal{J}_\alpha(u_\alpha)$ for every local minimizer u_α of $\mathcal{J}_\alpha$ (cf., e.g., [10, Prop. 2.3.2]), thus implies the existence of a subgradient $p_\alpha \in \partial \mathcal{G}(S(u_\alpha) - y^\delta) \subset \mathrm{L}^\infty(D)$ such that

$$0 = \alpha j(u_\alpha) + S'(u_\alpha)^* p_\alpha$$

holds, which is the first relation of (OS). Since $\mathcal{G}$ is convex, the generalized gradient reduces to the convex subdifferential (cf. [10, Prop. 2.2.7]), and by its definition we have the equivalence

$$p_\alpha \in \partial \mathcal{G}(S(u_\alpha) - y^\delta) \;\Leftrightarrow\; S(u_\alpha) - y^\delta \in \partial \mathcal{G}^*(p_\alpha),$$

where $\mathcal{G}^*$ is the Fenchel conjugate of $\mathcal{G}$ (cf., e.g., [15, Chap. I.4]), given by the indicator function of the unit ball $B \equiv \{p \in \mathrm{L}^\infty(D) : \|p\|_{\mathrm{L}^\infty} \leq 1\}$. The subdifferential of $\mathcal{G}^*$ coincides with the normal cone to B. Consequently, we deduce that $p_\alpha \in \partial \mathcal{G}(S(u_\alpha) - y^\delta)$ if and only if

$$\langle S(u_\alpha) - y^\delta, p - p_\alpha \rangle_{\mathrm{L}^2} \leq 0$$

holds for all $p \in \mathrm{L}^\infty(D)$ with $\|p\|_{\mathrm{L}^\infty} \leq 1$, which is the second relation of (OS). □

The following structural information for a solution u_α of problem ($\mathcal{P}$) is a direct consequence of (OS) and is of independent interest.

Corollary 2.2. *Let u_α be a minimizer of problem ($\mathcal{P}$), and $p_\alpha \in \mathrm{L}^\infty(D)$ as given by Theorem 2.1. Then the following relations hold:*

$$\begin{aligned} S(u_\alpha) - y^\delta &= 0 \quad \text{a.e. on} \quad \{x \in D : |p_\alpha(x)| < 1\}, \\ S(u_\alpha) - y^\delta &\geq 0 \quad \text{a.e. on} \quad \{x \in D : |p_\alpha(x)| = 1\}, \\ S(u_\alpha) - y^\delta &\leq 0 \quad \text{a.e. on} \quad \{x \in D : |p_\alpha(x)| = -1\}. \end{aligned}$$

This can be interpreted as follows: the box constraint on the dual solution p_α is active where the data is not attained by the primal solution u_α. In particular, the dual solution p_α acts as a noise indicator.

By using a complementarity function [8, 20], we can rewrite the second relation of (OS) as

$$S(u_\alpha) - y^\delta = \max(0, S(u_\alpha) - y^\delta + c(p_\alpha - 1)) + \min(0, S(u_\alpha) - y^\delta + c(p_\alpha + 1))$$

for any $c > 0$. This can be further discriminated by pointwise inspection to the following three cases:

1. $(S(u_\alpha) - y^\delta)(x) > 0$ and $p_\alpha(x) = 1$,
2. $(S(u_\alpha) - y^\delta)(x) < 0$ and $p_\alpha(x) = -1$,
3. $(S(u_\alpha) - y^\delta)(x) = 0$ and $p_\alpha(x) \in [-1, 1]$.

Consequently, we have the following concise relation

$$p_\alpha = \operatorname{sign}(S(u_\alpha) - y^\delta),$$

from which we obtain a reduced optimality system

$$\text{(OS')} \qquad \alpha j(u_\alpha) + S'(u_\alpha)^*(\operatorname{sign}(S(u_\alpha) - y^\delta)) = 0.$$

3. Solution by semi-smooth Newton method

In view of (OS') and the lack of smoothness of the sign function, the optimality system (OS) is not differentiable even in a generalized sense, which precludes the application of Newton-type methods. Meanwhile, gradient descent methods are inefficient unless the step lengths are chosen appropriately, which, however, necessarily requires a detailed knowledge of Lipschitz constants. Therefore, we propose to approximate ($\mathcal{P}$) using a local smoothing of the L^1 norm.

3.1. Approximation

To obtain a semi-smooth Newton system, we wish to replace the sign function in (OS′) by a locally linear smoothing. We therefore consider for $\beta > 0$ the following smoothed problem:

$$(\mathcal{P}_\beta) \qquad \min_{u \in \mathcal{X}} \|S(u) - y^\delta\|_{\mathrm{L}^1_\beta} + \frac{\alpha}{2}\|u\|^2_{\mathcal{X}},$$

where $\|v\|_{\mathrm{L}^1_\beta}$ is a Huber-type smoothing of the L^1 norm:

$$\|v\|_{\mathrm{L}^1_\beta} \equiv \int_\Omega |v(x)|_\beta \, dx, \qquad |v(x)|_\beta \equiv \begin{cases} v(x) - \frac{\beta}{2} & \text{if } v(x) > \beta, \\ -v(x) - \frac{\beta}{2} & \text{if } v(x) < -\beta, \\ \frac{1}{2\beta} v(x)^2 & \text{if } |v(x)| \leq \beta. \end{cases}$$

The existence of a minimizer u_β of $(\mathcal{P}_\beta)$ follows as before. Since the mapping $\psi : \mathbb{R} \to \mathbb{R}$, $t \mapsto |t|_\beta$, is differentiable with a globally Lipschitz continuous derivative $t \mapsto \mathrm{sign}_\beta(t)$,

$$\mathrm{sign}_\beta(t) \equiv \begin{cases} 1 & \text{if } t > \beta, \\ -1 & \text{if } t < -\beta, \\ \frac{1}{\beta} t & \text{if } |t| \leq \beta, \end{cases}$$

we have that ψ defines a differentiable Nemytskii operator from $\mathrm{L}^p(D)$ to $\mathrm{L}^2(D)$ for every $p \geq 4$ (see, e.g., [27, Chap. 4.3] and references therein) with pointwise defined derivative $\mathrm{sign}_\beta(v)h$. We thus obtain the necessary optimality conditions for u_β:

$$(\mathrm{OS}_\beta) \qquad \alpha j(u_\beta) + S'(u_\beta)^*(\mathrm{sign}_\beta(S(u_\beta) - y^\delta)) = 0.$$

Remark 3.1. This Huber-type smoothing is equivalent to an L^2-penalization of the dual variable $p \in \mathrm{L}^\infty(D)$ in (OS). To see this, we consider (OS) as the optimality conditions of the primal-dual saddle point problem

$$\min_{u \in \mathrm{L}^2} \max_{\|p\|_{\mathrm{L}^\infty} \leq 1} \langle S(u) - y^\delta, p \rangle_{\mathrm{L}^2} + \frac{\alpha}{2} \|u\|^2_{\mathcal{X}},$$

which makes use of the dual representation of the L^1-norm. We now introduce for $\beta > 0$ the penalized saddle point problem

$$\min_{u \in \mathrm{L}^2} \left(\max_{\|p\|_{\mathrm{L}^\infty} \leq 1} \langle S(u) - y^\delta, p \rangle_{\mathrm{L}^2} - \frac{\beta}{2} \|p\|^2_{\mathrm{L}^2} \right) + \frac{\alpha}{2} \|u\|^2_{\mathcal{X}}.$$

The corresponding optimality conditions are given by

$$(3.1) \qquad \begin{cases} S'(u_\beta)^* p_\beta + \alpha j(u_\beta) = 0, \\ \langle S(u_\beta) - y^\delta - \beta p_\beta, p - p_\beta \rangle_{\mathrm{L}^2} \leq 0. \end{cases}$$

for all $p \in \mathrm{L}^\infty(D)$ with $\|p\|_{\mathrm{L}^\infty} \leq 1$. By expressing the variational inequality again using a complementarity function with $c = \beta$, we obtain by pointwise inspection that

$$p_\beta = \mathrm{sign}_\beta(S(u_\beta) - y^\delta).$$

Inserting this expression into the first relation of (3.1) yields precisely (OS_β).

We next show the convergence of solutions to the approximating problems ($\mathcal{P}_\beta$) to a solution to problem ($\mathcal{P}$).

Theorem 3.2. *Let $\beta_n \to 0$. Then the sequence of minimizers $\{u_{\beta_n}\}$ of ($\mathcal{P}_\beta$) contains a subsequence converging in $\mathcal{X}$ to a minimizer of ($\mathcal{P}$).*

Proof. Let $u_n \equiv u_{\beta_n}$. Note that for any $\beta > 0$, there holds $|v(x)|_\beta \leq |v(x)|$, and consequently

$$\|S(u_\alpha) - y^\delta\|_{\mathrm{L}^1_{\beta_n}} \leq \|S(u_\alpha) - y^\delta\|_{\mathrm{L}^1}.$$

Now the minimizing property of u_n implies

$$\|S(u_n) - y^\delta\|_{\mathrm{L}^1_{\beta_n}} + \frac{\alpha}{2}\|u_n\|^2_{\mathcal{X}} \leq \|S(u_\alpha) - y^\delta\|_{\mathrm{L}^1_{\beta_n}} + \frac{\alpha}{2}\|u_\alpha\|^2_{\mathcal{X}}, \tag{3.2}$$

from which it follows that the sequence $\{u_n\}$ is uniformly bounded in $\mathcal{X}$. Therefore, there exists a subsequence, also denoted by $\{u_n\}$, and some $u^* \in \mathcal{X}$ such that $u_n \rightharpoonup u^*$ in $\mathcal{X}$. By the strong continuity of S, c.f. (A1), we have $S(u_n) \to S(u^*)$ in L^2, and this convergence is pointwise almost everywhere after possibly passing to a further subsequence [16]. In addition, since $|t|_\beta \to |t|$ as $\beta \to 0$ for every $t \in \mathbb{R}$, we have that $|S(u_\alpha) - y^\delta|_{\beta_n}$ converges pointwise to $|S(u_\alpha) - y^\delta|$. Fatou's Lemma then implies

$$\|S(u^*) - y^\delta\|_{\mathrm{L}^1} = \int_D \lim_{n\to\infty} |S(u_n) - y^\delta|_{\beta_n}\, dx \leq \liminf_{n\to\infty} \|S(u_n) - y^\delta\|_{\mathrm{L}^1{}_{\beta_n}}. \tag{3.3}$$

Meanwhile, by virtue of Lebesgue's dominated convergence theorem [16], we deduce

$$\lim_{n\to\infty} \|S(u_\alpha) - y^\delta\|_{\mathrm{L}^1_{\beta_n}} = \|S(u_\alpha) - y^\delta\|_{\mathrm{L}^1}.$$

These three relations together with the weak lower semicontinuity of norms indicate

$$\|S(u^*) - y^\delta\|_{\mathrm{L}^1} + \frac{\alpha}{2}\|u^*\|^2_{\mathcal{X}} \leq \|S(u_\alpha) - y^\delta\|_{\mathrm{L}^1} + \frac{\alpha}{2}\|u_\alpha\|^2_{\mathcal{X}}.$$

This together with the minimizing property of u_α implies that u^* is a minimizer of ($\mathcal{P}$).

To conclude the proof, it suffices to show that $\limsup_{n\to\infty}\|u_n\|_{\mathcal{X}} \leq \|u^*\|_{\mathcal{X}}$ holds. To this end, we assume the contrary, i.e. that there exists a subsequence of $\{u_n\}$, also denoted by $\{u_n\}$, satisfying $u_n \rightharpoonup u^*$ in $\mathcal{X}$ and $\lim_{n\to\infty}\|u_n\|_{\mathcal{X}} \equiv c > \|u^*\|_{\mathcal{X}}$. Letting $u_\alpha = u^*$ and $n \to \infty$ in (3.2), we arrive at

$$\limsup_{n\to\infty} \|S(u_n) - y^\delta\|_{\mathrm{L}^1{}_{\beta_n}} + \frac{\alpha}{2}c^2 \leq \|S(u^*) - y^\delta\|_{\mathrm{L}^1} + \frac{\alpha}{2}\|u^*\|^2_{\mathcal{X}},$$

i.e., $\limsup_{n\to\infty}\|S(u_n) - y^\delta\|_{\mathrm{L}^1{}_{\beta_n}} < \|S(u^*) - y^\delta\|_{\mathrm{L}^1}$, which is in contradiction with the weak lower semicontinuity in (3.3). This concludes the proof. □

3.2. Semi-smooth Newton method

To solve the optimality system (OS_β) with a semi-smooth Newton method, we consider it as an operator equation $F(u) = 0$ for $F : \mathcal{X} \to \mathcal{X}^*$,

$$F(u) = \alpha j(u) + S'(u)^*(\mathrm{sign}_\beta(S(u) - y^\delta)),$$

where $\mathrm{sign}_\beta : \mathcal{Y} \to \mathrm{L}^2(D)$ can be expressed using the pointwise max and min functions as

$$\mathrm{sign}_\beta(v) = \beta^{-1}\left(v - \max(0, v - \beta) - \min(0, v + \beta)\right).$$

We now argue the Newton differentiability of F. We know (e.g., from [25]) that the function $z \mapsto \max(0, z)$ is Newton differentiable from L^p to L^q for any $p > q \geq 1$ with its Newton derivative given pointwise by

$$D_N \max(0, z)(x) = \begin{cases} 1, & \text{if } z(x) > 0, \\ 0, & \text{if } z(x) \leq 0. \end{cases}$$

An analogous statement holds for the min function. This yields Newton differentiability of sign_β from $\mathcal{Y} \hookrightarrow \mathrm{L}^q(D)$, $q > 2$, to $\mathrm{L}^2(D)$. By the chain rule and the Fréchet differentiability of S, it follows that $P : \mathcal{X} \to \mathrm{L}^2(D)$,

$$P(u) = \mathrm{sign}_\beta(S(u) - y^\delta),$$

is Newton differentiable as well, and its Newton derivative acting on a direction $v \in \mathcal{X}$ is given as

$$D_N P(u)v = \beta^{-1}\chi_{\mathcal{I}}(S'(u)v).$$

Here, $\chi_{\mathcal{I}}$ is defined pointwise for $x \in D$ by

$$\chi_{\mathcal{I}}(x) = \begin{cases} 1, & \text{if } |(S(u) - y^\delta)(x)| \leq \beta, \\ 0, & \text{else.} \end{cases}$$

For a given u^k, one Newton step consists in solving for the increment $\delta u \in \mathcal{X}$ in

$$\alpha j'(u^k)\delta u + (S''(u^k)\delta u)^* P(u^k) + \beta^{-1} S'(u^k)^*(\chi_{\mathcal{I}^k} S'(u^k)\delta u) = -F(u^k) \tag{3.4}$$

and setting $u^{k+1} = u^k + \delta u$. Given a way to compute the action of the derivatives $S'(u)v$, $S'(u)^*v$ and $[S''(u)v]^*p$ for given u, p and v (given in Appendix A for the model problems), system (3.4) can be solved iteratively, e.g., using a Krylov method.

It remains to show the uniform well-posedness of system (3.4), from which superlinear convergence of the semi-smooth Newton method follows by standard arguments. Since the operator S is nonlinear and the functional is possibly non-convex, we assume the following condition at a minimizer u_β: There exists a constant $\gamma > 0$ such that

$$\langle S''(u_\beta)(h, h), P(u_\beta)\rangle_{\mathrm{L}^2} + \alpha\|h\|^2_{\mathcal{X}} \geq \gamma\|h\|^2_{\mathcal{X}} \tag{3.5}$$

holds for all $h \in \mathcal{X}$. This is related to standard second-order sufficient optimality conditions in PDE-constrained optimization (cf., e.g., [27, Chap. 4.10]). The condition is satisfied for either large α or sparse residual $S(u_\beta) - y^\delta$, since

$$\langle S''(u_\beta)(h,h), P(u_\beta)\rangle_{L^2} + \alpha\|h\|_{\mathcal{X}}^2 \geq (\alpha - C\|P(u_\beta)\|_{L^2})\|h\|_{\mathcal{X}}^2 \tag{3.6}$$

holds by the a priori estimate on S'' (A4). In the context of parameter identification problems, this is a reasonable assumption, since for a large noise level, α would take a large value, while a small α is chosen only for small noise levels (which, given the impulsive nature of the noise, is equivalent to strong sparsity of the residual). In the latter case, we observe that $P(u_\beta) = \operatorname{sign}_\beta(S(u_\beta) - y^\delta)$ can be expected to be small due to the L^2 smoothing of $\operatorname{sign}_\beta$ (cf. Remark 3.1 and note that $P(u_\beta) = p_\beta$). Condition (3.5) is thus satisfied if either α or β is sufficiently large. However, this property depends on β, which together with Theorem 3.2 motivates the use of a continuation strategy in β, see section 3.3. We remind that in general it is not possible to check such conditions a priori even for quadratic functionals.

Proposition 3.3. *Let $\beta > 0$ be given. If condition (3.5) holds, then for each $u \in U$ sufficiently close to a solution u_β of (OS_β), the mapping $D_N F : \mathcal{X} \to \mathcal{X}^*$,*

$$D_N F(u) = \alpha j'(u) + S''(u)^* P(u) + \beta^{-1} S'(u)^* \chi_I S'(u)$$

is invertible, and there exists a constant $C > 0$ independent of u such that

$$\|(D_N F)^{-1}\|_{L(\mathcal{X}^*,\mathcal{X})} \leq C.$$

Proof. For given $w \in \mathcal{X}^*$, we need to find $\delta u \in \mathcal{X}$ satisfying

$$\langle \alpha j'(u)\delta u + (S''(u)\delta u)^* P(u) + \beta^{-1} S'(u)^* \chi_I S'(u)\delta u, v\rangle_{\mathcal{X}^*,\mathcal{X}} = \langle w, v\rangle_{\mathcal{X}^*,\mathcal{X}}$$

for all $v \in \mathcal{X}$. Letting $v = \delta u$ and observing that $\langle j'(u)v, v\rangle_{\mathcal{X}^*,\mathcal{X}} = \|v\|_{\mathcal{X}}^2$ (since $\mathcal{X}$ is a Hilbert space), we obtain

$$\alpha\,\|\delta u\|_{\mathcal{X}}^2 + \langle S''(u)(\delta u, \delta u), P(u)\rangle_{L^2} + \beta^{-1}\left\|\chi_I S'(u)\delta u\right\|_{L^2}^2 = \langle w, \delta u\rangle_{\mathcal{X}^*,\mathcal{X}}.$$

Now the pointwise contraction property of the min and the max function implies

$$\begin{aligned}\|P(u_\beta) - P(u)\|_{L^2} &\leq \beta^{-1}\|S(u_\beta) - S(u)\|_{L^2} \\ &\quad + \beta^{-1}\|\max(0, S(u_\beta) - y^\delta - \beta) - \max(0, S(u) - y^\delta - \beta)\|_{L^2} \\ &\quad + \beta^{-1}\|\min(0, S(u_\beta) - y^\delta + \beta) - \min(0, S(u) - y^\delta + \beta)\|_{L^2} \\ &\leq 3\beta^{-1}\|S(u_\beta) - S(u)\|_{L^2}.\end{aligned}$$

Consequently, by the continuity of the mapping S, for sufficiently small $\|u_\beta - u\|_{\mathcal{X}}$, we have small $\|P(u_\beta) - P(u)\|_{L^2}$ as well. Thus, by condition (3.5) and the locally uniform boundedness of S'', c.f. (A4), there exists an $\varepsilon > 0$ such that

$$\begin{aligned}&\alpha\,\|\delta u\|_{\mathcal{X}}^2 + \langle S''(u)(\delta u, \delta u), P(u)\rangle_{L^2} \\ &\quad = \alpha\,\|\delta u\|_{\mathcal{X}}^2 + \langle S''(u)(\delta u, \delta u), P(u_\beta)\rangle_{L^2} + \langle S''(u)(\delta u, \delta u), P(u) - P(u_\beta)\rangle_{L^2} \\ &\quad \geq \gamma\,\|\delta u\|_{\mathcal{X}}^2 - C\varepsilon\,\|\delta u\|_{\mathcal{X}}^2 \geq \frac{\gamma}{2}\,\|\delta u\|_{\mathcal{X}}^2\end{aligned}$$

holds for all u with $\|u - u_\beta\|_{\mathcal{X}} \leq \varepsilon$ if ε is sufficiently small.

Finally, we deduce by the Cauchy-Schwarz inequality that

$$\frac{\gamma}{4} \|\delta u\|_{\mathcal{X}}^2 \leq \|w\|_{\mathcal{X}^*} \|\delta u\|_{\mathcal{X}}.$$

This implies the claim. □

Thus, system (3.4) is semi-smooth, and from standard results (e.g., [25], [20, Th. 8.5]) we deduce the following convergence result for the semi-smooth Newton method.

Theorem 3.4. *Let $\beta > 0$ and condition (3.5) hold. Then the sequence of iterates u^k in the semi-smooth Newton method (3.4) converge superlinearly to a solution u_β of (OS_β), provided that u^0 is sufficiently close to u_β.*

3.3. Parameter choice

The regularized formulation ($\mathcal{P}$) of the parameter identification problem $S(u) = y^\delta$ requires specifying the regularization parameter α, whose correct choice is crucial in practice. Usually, it is determined using knowledge of the noise level δ by, e.g., discrepancy principle. However, in practice, the noise level δ may be unknown, rendering such rules inapplicable. To circumvent this issue, we propose a heuristic choice rule based on the following balancing principle [12]: Choose α such that

$$(\sigma - 1)\|S(u_\alpha) - y^\delta\|_{L^1} - \frac{\alpha}{2}\|u_\alpha\|_{\mathcal{X}}^2 = 0 \tag{3.7}$$

is satisfied. The underlying idea of the principle is to balance the data fitting term with the penalty term, and the weight $\sigma > 1$ controls the trade-off between them. This weight depends on the relative smoothness of residual and parameter, but not on the data realization. The principle does not require a knowledge of the noise level, and has been successfully applied to linear inverse problems with L^1 data fitting [11, 12].

We compute a solution α^* to the balancing equation (3.7) by the following simple fixed point algorithm proposed in [11]:

$$\alpha_{k+1} = (\sigma - 1)\frac{\|S(u_{\alpha_k}) - y^\delta\|_{L^1}}{\frac{1}{2}\|u_{\alpha_k}\|_{\mathcal{X}}^2}. \tag{3.8}$$

This fixed point algorithm can be derived formally from the model function approach [12]. The convergence can be proven similar to [11], by observing that the proof given there does not depend on the linearity of the forward operator.

Theorem 3.5. *If the initial guess α_0 satisfies $(\sigma - 1)\|S(u_{\alpha_0}) - y^\delta\|_{L^1} - \frac{\alpha_0}{2}\|u_{\alpha_0}\|_{\mathcal{X}}^2 < 0$, then the sequence $\{\alpha_k\}$ generated by the fixed point algorithm is monotonically decreasing and converges to a solution to (3.7).*

Of similar importance is the proper choice of the smoothing parameter β. If β is too large, the desirable structural property of the L^1 model will be lost. However, the second-order condition (3.6) depends on β and cannot be expected to hold for arbitrarily small β. In particular, the convergence basin for the semi-smooth Newton method is likely to shrink

Algorithm 1 Path-following semi-smooth Newton method.

Choose β_0, $q < 1$, $\beta_{\min} > 0$, $u_0 \in U$, $k^* > 0$, set $n = 0$
repeat
  Set $u^0 = u_n$, $k = 0$
  repeat
    Compute $y^k = S(u^k)$
    Compute active and inactive sets

$$\mathcal{A}_+^k = \left\{ x \in D : (y^k - y^\delta)(x) > \beta \right\}$$
$$\mathcal{A}_-^k = \left\{ x \in D : (y^k - y^\delta)(x) < -\beta \right\}$$
$$\mathcal{I}^k = \left\{ x \in D : |(y^k - y^\delta)(x)| \leq \beta \right\}$$

    Compute $p^k = \operatorname{sign}_\beta(y^k - y^\delta)$
    Compute $F(u^k) = \alpha j(u^k) + S'(u^k)^*(p^k)$
    Compute update δu by solving

$$\alpha j'(u^k)\delta u + S''(u^k)^*(p^k)\delta u + \beta^{-1} S'(u^k)^* \chi_{I^k} S'(u^k)\delta u = -F(u^k))$$

    Set $u^{k+1} = u^k + \delta u$, $k \leftarrow k + 1$.
  until ($\mathcal{A}_+^k = \mathcal{A}_+^{k-1}$ and $\mathcal{A}_-^k = \mathcal{A}_-^{k-1}$ and $\|r^k\| \leq \text{tol}$) or $k = k^*$
  if $k < k^*$ **then**
    Set $n \leftarrow n + 1$, $u_n = u^k$, $\beta_n = q\beta_{n-1}$
  end if
until $k = k^*$ or $\beta_n < \beta_{\min}$

as β decreases to zero. This motivates the following continuation strategy: Starting with a large β_0 and setting $\beta_{n+1} = q\beta_n$ for some $q \in (0, 1)$, we compute the solution u_{β_n} of (OS_β) using the previous solution u_{β_n} as an initial guess.

A crucial issue is then selecting an appropriate stopping criterion for the continuation. Since we are most interested in the L^1 structure of the problem, we base our stopping rule on the following finite termination property of the linear L^1 fitting problem [12, Prop. 3.6]: If the active sets coincide for two consecutive iterations of the semi-smooth Newton method, the semi-smooth optimality system is solved exactly. In addition, the convergence is usually very fast due to the continuation strategy, and the required number of iterations is independent of the mesh size (this property is well-known as *mesh independence* [18]). Hence, if the active sets (cf. $\mathcal{A}_+^k$ and $\mathcal{A}_-^k$ in Algorithm 1) are still changing after a fixed number of iterations, we deduce that the semi-smoothness of the operator $F(u)$ might be lost and return the last feasible solution $u_{\beta_{n-1}}$ as the desired approximation. In practice, we also check for smallness of the norm of the gradient to take into account the nonlinearity of S, and safeguard termination of the algorithm by stopping the continuation if a given very small value $\beta_{\min}$ is reached.

A complete description of this approach, hereafter denoted by *path-following semi-smooth Newton method*, is given in Algorithm 1.

4. Numerical examples

We now present some numerical results for several benchmark parameter identification problems with one- and two-dimensional elliptic differential equations to illustrate the features of the proposed approach. In each case, the forward operator was discretized using finite elements on a uniform grid (triangular, in the case of two dimensions). We denote by P_0 the space of piecewise constant functions (on each element), while P_1 is the space of piecewise linear functions. Unless otherwise stated, the number N of grid points are 1001 in 1d and 128×128 in 2d.

We implemented the semi-smooth Newton (SSN) method as given in Algorithm 1. The iteration was terminated if the active sets did not change and the norm of the gradient fell below 1.00×10^{-6}, or if 20 iterations were reached. Given the true solution $u^\dagger$, we generate the noisy data y^δ by modifying the exact data $y^\dagger = S(u^\dagger)$ pointwise as follows

$$y^\delta = \begin{cases} y^\dagger + \|y^\dagger\|_{L^\infty}\xi, & \text{with probability } r, \\ y^\dagger, & \text{otherwise,} \end{cases}$$

where the random variable ξ follows the standard normal distribution and $r \in (0,1)$ is the percentage of corrupted data points. Unless otherwise noted, we take $r = 0.3$. The exact noise level δ is defined by $\delta = \|y^\delta - y^\dagger\|_{L^1}$. The Newton system (3.4) is solved iteratively using BiCGstab (with tolerance 1.00×10^{-6} and maximum number of iterations 100). The reduction rate q is set to $\frac{1}{2}$.

All timing tests were performed with MATLAB (R2010b) on a single core of a 2.8 GHz workstation with 24 GByte of RAM. The MATLAB codes of our implementation can be downloaded from `http://www.uni-graz.at/~clason/codes/l1nonlinfit.zip`. To keep the presentation concise, all tables are collected in Appendix B.

4.1. Inverse potential problem

This example is concerned with determining the potential $u \in L^2(\Omega)$ in (1.1) from noisy measurements of the state $y \in H^1(\Omega)$ in the domain Ω. The discretized operator S_h maps $u_h \in U_h = P_0$ to $y_h \in Y_h = P_1$ which satisfies

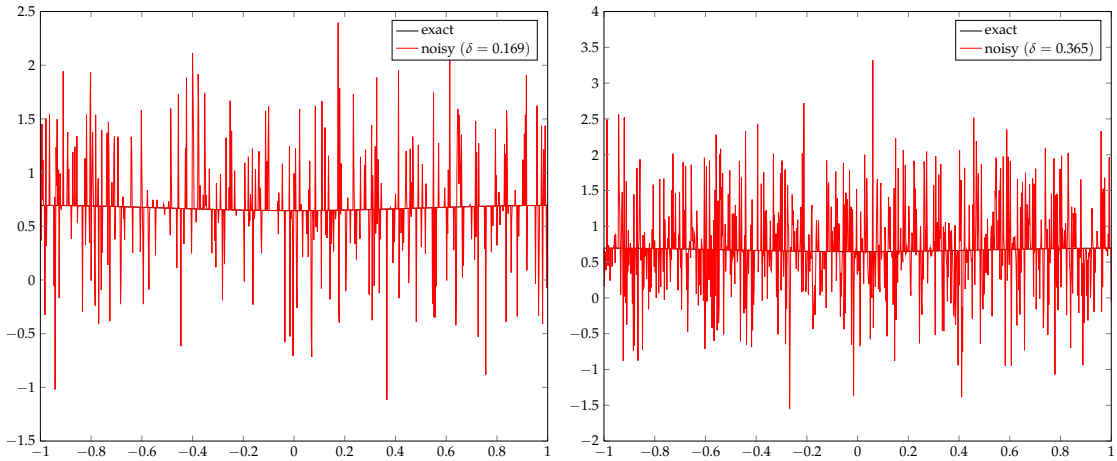
$$\langle \nabla y_h, \nabla v_h \rangle_{L^2} + \langle u_h y_h, v_h \rangle_{L^2} = \langle f, v_h \rangle_{L^2} \qquad \text{for all } v_h \in Y_h.$$

For the automatic parameter choice using the balancing principle, we have set the weight σ to 1.03 and the initial guess α_0 to 1.

One-dimensional example. Here, we take $\Omega = [-1,1]$, $f(x) = 1$ and

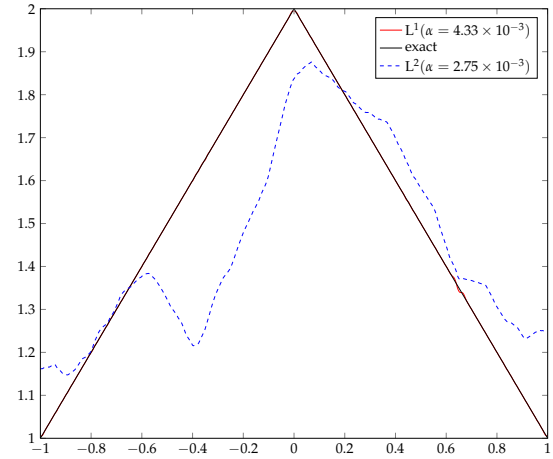
$$u^\dagger(x) = 2 - |x| \geq 1.$$

A typical realization of noisy data is displayed in Fig. 1a for $r = 0.3$ and Fig. 1b for $r = 0.6$. The fixed-point iteration (3.8) converged after 3 (4) iterations for $r = 0.3$ ($r = 0.6$), and yielded the values 4.33×10^{-3} (9.39×10^{-3}) for the regularization parameter α. The respective reconstructions u_α, shown in Figs. 1c and 1d, are nearly indistinguishable from the true solution $u^\dagger$. To measure the accuracy of the solution u_α quantitatively, we compute

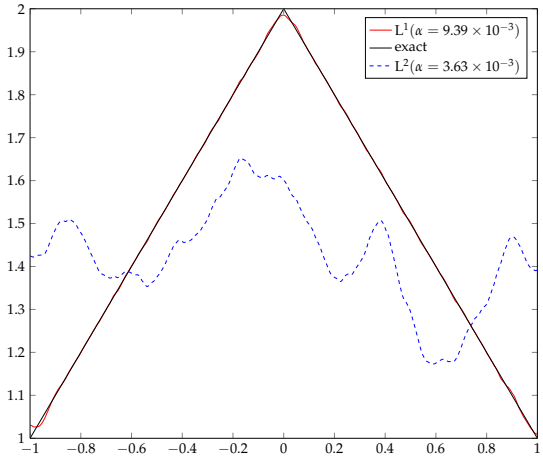


(a) exact and noisy data ($r = 0.3$)

(b) exact and noisy data ($r = 0.6$)



(c) comparison of L^1 and L^2 reconstruction



(d) comparison of L^1 and L^2 reconstruction

Figure 1: Results for 1d inverse potential problem. Left: $r = 0.3$, right: $r = 0.6$.

the L^2-error $e = \|u_\alpha - u^\dagger\|_{L^2}$, which is 8.65×10^{-4} for $r = 0.3$ and 3.32×10^{-3} for $r = 0.6$. For comparison, we also show the solution by the L^2 data fitting problem (solved by a standard Newton method), where the parameter α has been chosen to give the smallest L^2 error. We observe that the L^2 reconstructions are clearly inacceptable compared to their L^1 counterparts, which illustrates the importance of the correct choice of noise model, and especially the suitability of L^1 fitting for impulsive noise.

The performance of the balancing principle is further illustrated in Table 1 (see Appendix B), where we compare the balancing parameter α_b with the "optimal", sampling-based parameter α_o for different noise levels. This parameter is obtained by sampling each interval $[0.1\alpha_b, \alpha_b]$ and $[\alpha_b, 10\alpha_b]$ uniformly with 50 parameters and taking as α_o the one with smallest L^2-error $e_o \equiv \|u_{\alpha_o} - u^\dagger\|_{L^2}$. We observe that both the regularization parameters and the reconstruction errors obtained from the two approaches are comparable. This shows the feasibility of the balancing principle for choosing an appropriate regularization

parameter in nonlinear L^1 models. Table 1 also illustrates the fundamentally different nature of impulsive noise and L^1 fitting compared with Gaussian models, since the L^2-error does not depend linearly on the noise level or the percentage r of corrupted data. This can be attributed to the fact that the structure properties of the noise (e.g., clustering of corrupted data points, which is increasingly likely for $r \geq 0.5$) is more important than the noise percentage itself.

Next we study the convergence behavior of the path-following SSN method. First, the convergence behavior in the smoothing parameter β is illustrated in Table 2 by showing for each step in the continuation procedure the value of β, the required number of SSN iterations and the L^2-error e. The required number of SSN iterations is relatively independent of the value of β provided it is sufficiently large. Then the semi-smoothness of the optimality system (OS_β) is gradually lost after the β value drops below 1.00×10^{-7}, and more and more iterations are required for the Krylov method to solve the Newton system (3.4) to the prescribed accuracy. Nonetheless, the reconstruction already represents a very reasonable approximation (in terms of the L^2-error e) at $\beta = 1.19 \times 10^{-7}$. Second, we illustrate the superlinear convergence of the SSN method by solving the optimality system (3.1) with fixed $r = 0.3$, $\alpha = 4.00 \times 10^{-3}$ and $\beta = 1.00 \times 10^{-1}$. Table 3 shows the number of elements that changed between active and inactive sets and the residual norm $\|F(u^k)\|_{L^2}$ after the kth iteration for several problem sizes N. The superlinear convergence as well as the mesh independence can be observed.

Finally, we demonstrate the scalability of the proposed approach. Table 4 summarizes the computing time for one run of the path-following SSN method and for the full fixed point iteration. Since the computing time depends on the α value, we present the results with the final value of α as obtained from the fixed-point iteration (3.8). The presented results are the mean and standard deviation over ten noise realizations. We observe that both the fixed point iteration and the path-following SSN method scale very well with the problem size N, which corroborates the mesh independence of the SSN method [18]. We point out that the computational cost of calculating the balancing parameter is only two to three times that of solving the L^1 model with one fixed regularization parameter. Therefore, the balancing principle is also computationally inexpensive.

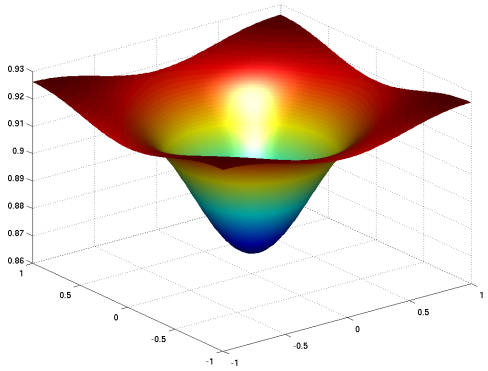
Two-dimensional example. Here, we take $\Omega = [-1,1]^2$, $f(x_1, x_2) = 1$ and

$$u^\dagger(x_1, x_2) = 1 + \cos(\pi x_1)\cos(\pi x_2)\chi_{\{|(x_1,x_2)|_\infty < 1/2\}} \geq 1,$$

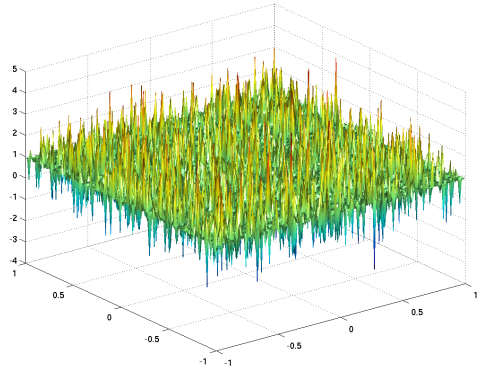
see Fig. 2c. The exact and noisy data (with $r = 0.3$) are given in Figs. 2a and 2b, respectively. The fixed point algorithm (3.8) converged within two iterations to the value $\alpha_b = 1.06 \times 10^{-2}$. The solution (with an L^2-error $e = 5.28 \times 10^{-3}$), shown in Fig. 2d, accurately captures the shape as well as the magnitude of the potential $u^\dagger$, and thus represents a good approximation. The reconstruction by the L^2 model is again far from the true solution, and thus not shown here.

4.2. Inverse Robin coefficient problem

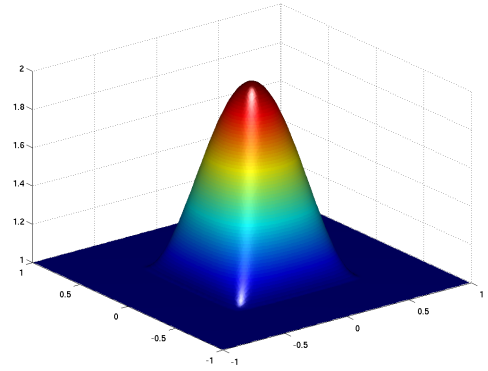
This example, meant to illustrate coefficient recovery from boundary data, concerns reconstructing the Robin coefficient $u \in L^2(\Gamma_i)$ in (1.2) from noisy measurements of the Dirichlet



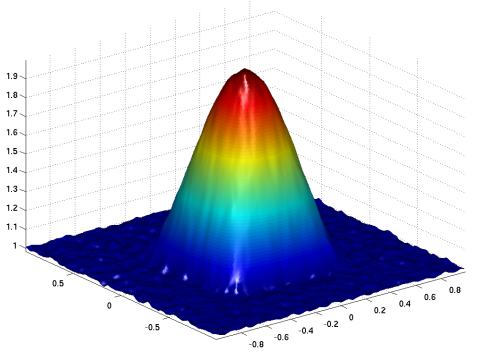
(a) exact data



(b) noisy data ($r = 0.3$)



(c) true solution



(d) L^1 reconstruction ($\alpha = 1.06 \times 10^{-2}$)

Figure 2: Results for the 2d inverse potential problem with $r = 0.3$ ($\delta = 2.24 \times 10^{-1}$).

trace of $y \in H^1(\Omega)$ on the boundary Γ_c. The discretization S_h of the forward operator S thus maps $u_h \in U_h = P_0(\Gamma_i)$ to the restriction of $y_h \in Y_h = P_1$ to the nodes on Γ_c, where y_h satisfies

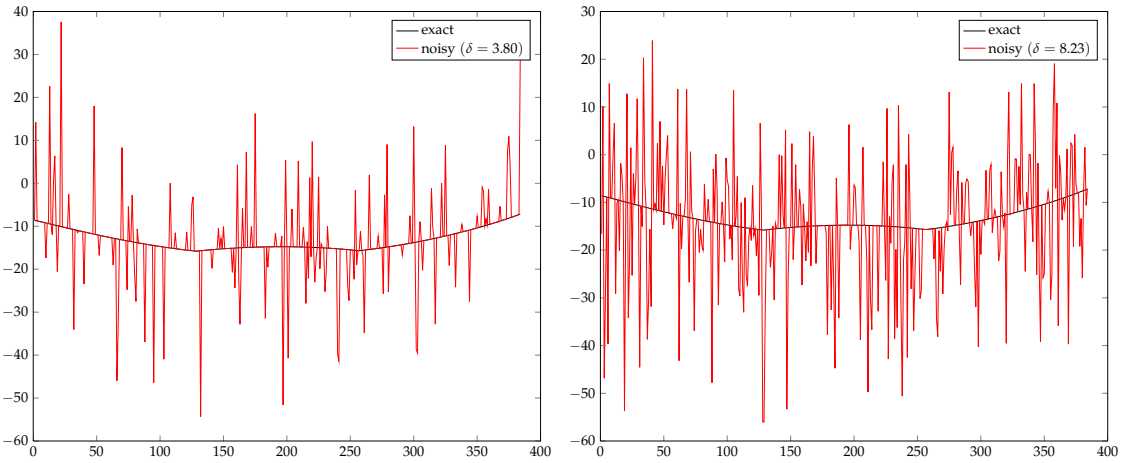
$$\langle \nabla y_h, \nabla v_h \rangle_{L^2} + \langle u_h y_h, v_h \rangle_{L^2(\Gamma_i)} = \langle f, v_h \rangle_{L^2(\Gamma_c)} \qquad \text{for all } v_h \in Y_h.$$

Here, we take the domain $\Omega = [0,1]^2$, inaccessible boundary $\Gamma_i = \{(x_1, x_2) \in \partial\Omega : x_1 = 1\}$ and accessible (contact) boundary $\Gamma_c = \partial\Omega \setminus \Gamma_i$. Further, we set $f(x_1, x_2) = -4 + x_1$ and

$$u^\dagger(x_2) = 1 + x_2 \geq 1.$$

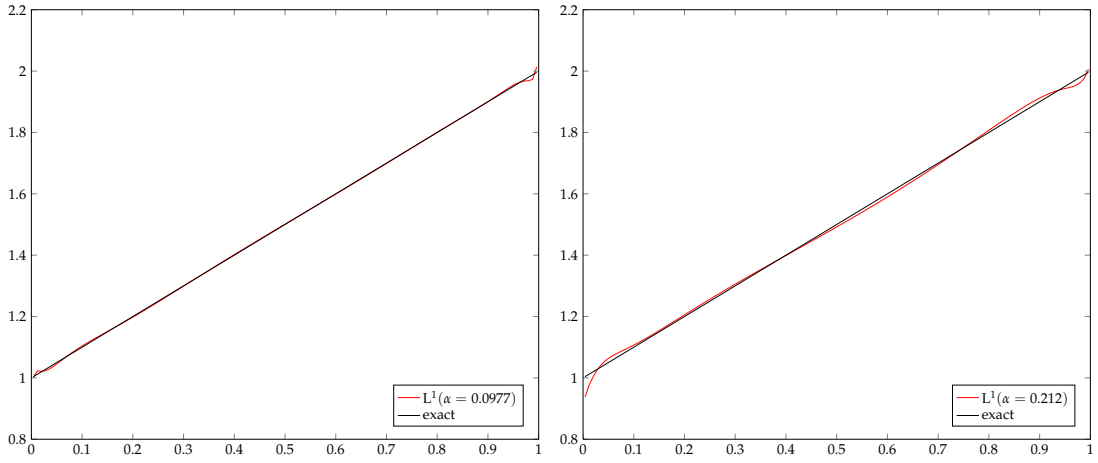
For the automatic parameter choice using the balancing principle, we have set the weight σ to 1.03 and the initial guess α_0 to 1 as before.

The noisy data for $r = 0.3$ and $r = 0.6$ are displayed in Figs. 3a and 3b, respectively. The fixed point algorithm (3.8) converged after two iterations in both cases, giving a value 9.77×10^{-2} ($r = 0.3$) and 2.12×10^{-1} ($r = 0.6$) for the regularization parameter α. The



(a) exact and noisy data ($r = 0.3$)

(b) exact and noisy data ($r = 0.6$)



(c) L^1 reconstruction

(d) L^1 reconstruction

Figure 3: Results for the inverse Robin coefficient problem. Left: $r = 0.3$, right: $r = 0.6$.

corresponding reconstructions u_α, with respective L^2-error 3.13×10^{-3} and 1.05×10^{-2}, are shown in Figs. 3c and 3d. Overall, the approximate solutions agree well with the true coefficient, except around the two end points, where the reconstructions suffer from pronounced boundary effect, especially in case of $r = 0.6$. Again, the reconstruction by the L^2 model (with optimal choice of α) is not acceptable, and is thus not shown. A comparison of the balancing principle with the optimal choice based on sampling is given in Table 5. The results by these two approaches are very close to each other. From the table, we also observe the non-monotonicity of the error as a function of r, where the reconstruction error e shows a noticeable jump after $r = 0.5$.

4.3. Inverse diffusion coefficient problem

Finally, we consider the problem of determining the diffusion coefficient $u \in \mathrm{H}^1(\Omega)$ in (1.3) from noisy measurements of the solution $y \in \mathrm{H}^1_0(\Omega)$. Here we take $U_h = P_1$ and $Y_h = P_1 \cap \mathrm{H}^1_0(\Omega)$ and consider the discrete operator S_h as mapping $u_h \in U_h$ to $y_h \in Y_h$ satisfying

$$\langle u_h \nabla y_h, \nabla v_h \rangle_{\mathrm{L}^2} = \langle f, v_h \rangle_{\mathrm{L}^2} \qquad \text{for all } v_h \in Y_h.$$

To accelerate the convergence of the Krylov solver, we precondition the Newton system with the inverse Helmholtz operator $(-\Delta + I)^{-1}$, i.e., the gradient $\alpha(-\Delta u + u) - \nabla y \cdot \nabla p$ is replaced by

$$\alpha u - (-\Delta + I)^{-1}(\nabla y \cdot \nabla p),$$

and similarly the action of the Hessian on δu is computed as

$$\alpha u - (-\Delta + I)^{-1}(\nabla \delta y \cdot \nabla p + \nabla y \cdot \nabla \delta p).$$

For the automatic parameter choice using the balancing principle, we have set the weight σ to 1.001 and the initial guess α_0 to 0.1. As noted, the different weight is chosen according to the stronger smoothness assumption on u (H^1 instead of L^2 regularization).

One-dimensional example. Here, we take the domain $\Omega = [0, 1]$ and $f(x) = 1$. The exact solution $u^\dagger$ is given by

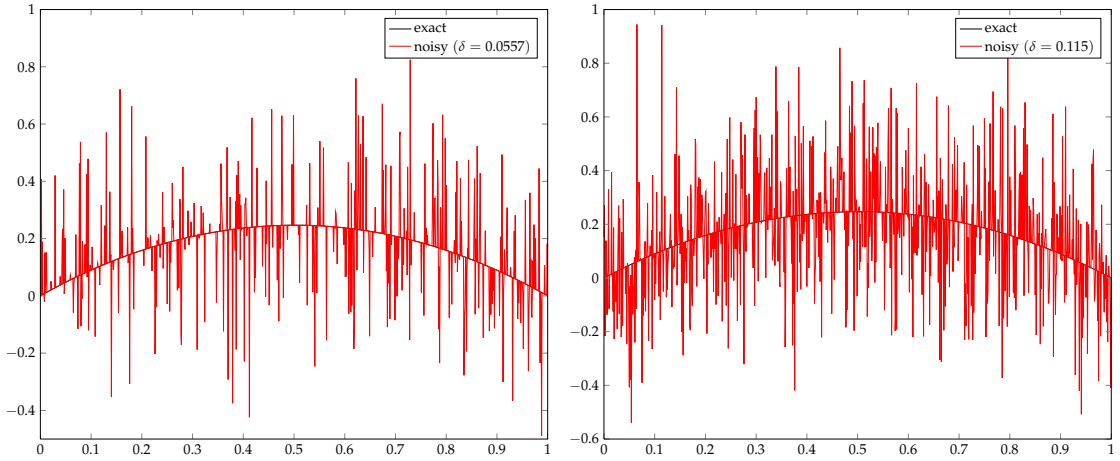
$$u^\dagger(x) = \frac{1}{2} + x^2(1 - x)^4 \geq \frac{1}{2}.$$

Noisy data with $r = 0.3$ and $r = 0.6$ and the reconstructions ($\alpha = 3.85 \times 10^{-4}$, L^2-error 2.77×10^{-5} and $\alpha = 6.90 \times 10^{-4}$, L^2-error 3.86×10^{-5}) are shown in Fig. 4. In both cases, the fixed point iteration (3.8) converged within two iterations. The convergence of the path-following method and the SSN method are similar to the inverse potential problem. A comparison of the balancing principle with the optimal choice based on sampling is given in Table 6. The results by these two approaches are very close to each other. From the table, we also observe the non-monotonicity of the error as a function of r, where the reconstruction error e remains almost constant for $r \leq 0.5$ and then increases quickly. Again the proposed SSN method scales very well with the problem size, as shown in Table 7.

Two-dimensional example. Here, we take $\Omega = [0, 1]^2$, $f(x_1, x_2) = 1$ and

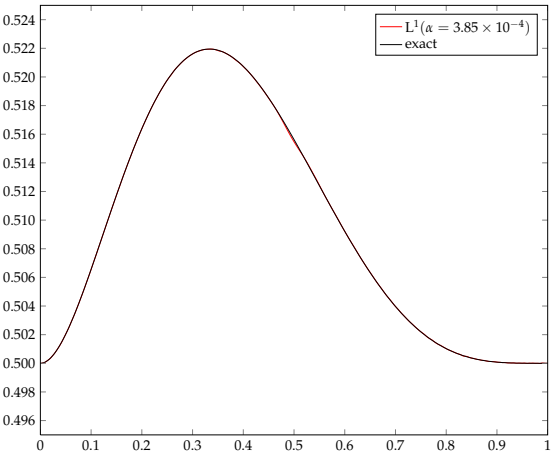
$$u^\dagger(x_1, x_2) = \frac{1}{2} + x_1^2 x_2^2 \geq \frac{1}{2}.$$

see Fig. 5c. The exact and noisy data ($r = 0.3$) are given in Figs. 5a and 5b, respectively. The fixed point algorithm converged in seven iterations to the value $\alpha = 5.14 \times 10^{-5}$. The reconstruction, shown in Fig. 5d, agrees well with the true solution (the L^2-error being 5.63×10^{-3}). The less accurate approximation around the corner might be attributed to the fact that the true solution does not satisfy the homogeneous Neumann conditions imposed by the Newton step. Again, we remark that the L^2 reconstruction (not presented) is far from the true solution.

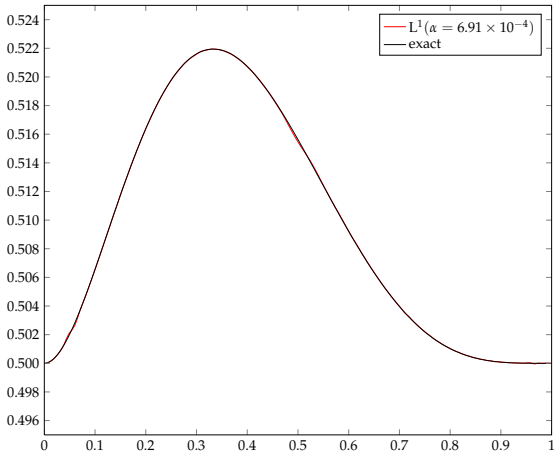


(a) exact and noisy data ($r = 0.3$)

(b) exact and noisy data ($r = 0.6$)



(c) L^1 reconstruction



(d) L^1 reconstruction

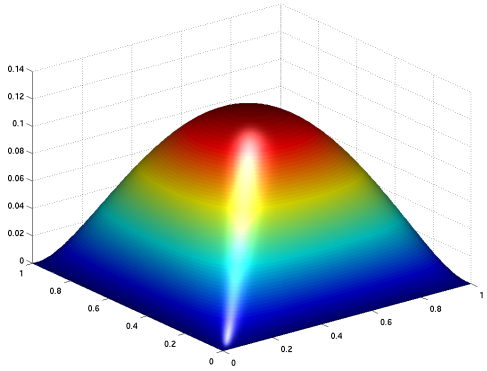
Figure 4: Results for the 1d inverse diffusion coefficient problem. Left: $r = 0.3$, right: $r = 0.6$.

5. Conclusion

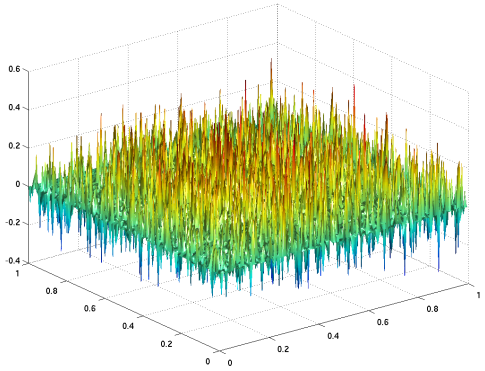
In this paper we have presented a path-following semi-smooth Newton method for the efficient numerical solution of nonlinear parameter identification problems with impulsive noise. The method is based on a Huber-type smoothing of the L^1 fitting functional, and its superlinear convergence is proved and demonstrated numerically. Furthermore, mesh independence of the method can be observed. Several model examples for elliptic differential equation illustrate the efficiency of this approach.

The balancing principle is shown to be an effective parameter choice method, which required little a priori information such as the noise level, while adding only a small amount of computational overhead over the solution of one single minimization problem.

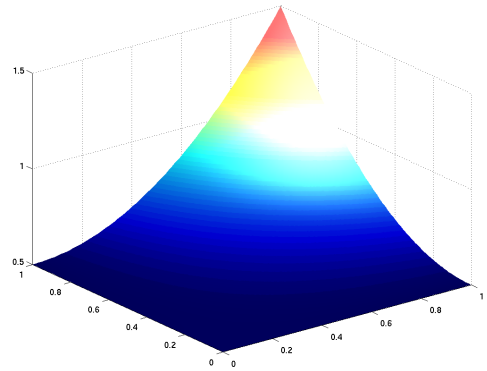
The presented approach can be extended in several directions. As noted in Remark 1.1,



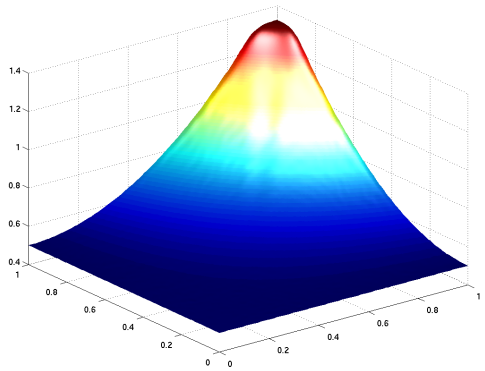
(a) exact data



(b) noisy data ($r = 0.3$)



(c) true solution



(d) L^1 reconstruction ($\alpha = 5.14 \times 10^{-5}$)

Figure 5: Results for the 2d inverse diffusion coefficient problem with $r = 0.3$ ($\delta = 3.06 \times 10^{-2}$).

including constraints on the solution would be a natural progression. The extension to time-dependent problems would be straightforward, but pose interesting challenges for the efficient implementation.

Acknowledgments

The work of the first author was supported by the Austrian Science Fund (FWF) under grant SFB F32 (SFB "Mathematical Optimization and Applications in Biomedical Sciences"), and that of the second author was supported by Award No. KUS-C1-016-04, made by King Abdullah University of Science and Technology (KAUST).

A. Verification of properties for model problems

For completeness, we collect in this section some results which verify the continuity and differentiability properties (A1)–(A4) for our model problems. Throughout, we shall denote by C a generic constant, which is independent of $u \in U$.

A.1. Elliptic potential problem

For this model problem, S maps $u \in \mathcal{X} = L^2(\Omega)$ to the solution $y \in \mathcal{Y} = H^1(\Omega)$ of (1.1), and we take $U = \{u \in L^\infty(\Omega) : u \geq c\}$ for some fixed $c > 0$. The verification of properties (A1)–(A4) is analogous to [23]. We therefore only give, for the sake of completeness, the explicit form of the derivatives required for the solution of the Newton system (3.4) using a Krylov subspace method.

For given $u \in L^2(\Omega)$, $F(u)$ is computed by the following steps:

1: Solve for $y \in H^1(\Omega)$ in

$$\langle \nabla y, \nabla v \rangle_{L^2} + \langle uy, v \rangle_{L^2} = \langle f, v \rangle_{L^2} \qquad \text{for all } v \in H^1(\Omega).$$

2: Solve for $p \in H^1(\Omega)$ in

$$\langle \nabla p, \nabla v \rangle_{L^2} + \langle up, v \rangle_{L^2} = -\langle \operatorname{sign}_\beta(y - y^\delta), v \rangle_{L^2} \qquad \text{for all } v \in H^1(\Omega).$$

3: Set $F(u) = \alpha u + yp$.

For given $\delta u \in L^2(\Omega)$, the application of $D_N F(u)$ on δu is computed by:

1: Solve for $\delta y \in H^1(\Omega)$ in

$$\langle \nabla \delta y, \nabla v \rangle_{L^2} + \langle u\delta y, v \rangle_{L^2} = -\langle y\delta u, v \rangle_{L^2} \qquad \text{for all } v \in H^1(\Omega).$$

2: Solve for $\delta p \in H^1(\Omega)$ in

$$\langle \nabla \delta p, \nabla v \rangle_{L^2} + \langle up, v \rangle_{L^2} = -\langle \tfrac{1}{\beta}\chi_{\mathcal{I}}\delta y + p\delta u, v \rangle_{L^2} \qquad \text{for all } v \in H^1(\Omega).$$

3: Set $D_N F(u) = \alpha\delta u + p\delta y + y\delta p$.

A.2. Robin coefficient problem

Here, S maps $u \in \mathcal{X} = L^2(\Gamma_i)$ to $y|_{\Gamma_c} \in \mathcal{Y} = H^{\frac{1}{2}}(\Gamma_c)$, where y is the solution to (1.2). Set $U = \{u \in L^\infty(\Gamma_i) : u \geq c\}$ for some fixed $c > 0$. We shall denote the mapping of $u \in U$ to the solution $y \in H^1(\Omega)$ of (1.2) by $y(u)$. The following a priori estimate follows directly from the Lax-Milgram theorem.

Lemma A.1. *For any $u \in U$, problem (1.2) has a unique solution $y \in H^1(\Omega)$ which satisfies*

$$\|y\|_{H^1} \leq C \|f\|_{H^{-\frac{1}{2}}(\Gamma_c)}.$$

Since $f \in H^{-\frac{1}{2}}(\Gamma_c)$ is fixed, the uniform boundedness of S follows from the continuity of the trace operator. We next address the complete continuity of S.

Lemma A.2. *Let $\{u_n\} \subset U$ be a sequence converging weakly in $L^2(\Gamma_i)$ to $u^* \in U$. Then*

$$S(u_n) \to S(u^*) \quad \text{in } L^2(\Gamma_c).$$

Proof. For $u_n \in U$, set $y_n = y(u_n) \in H^1(\Omega)$. By the a priori estimate from Lemma A.1, the sequence $\{y_n\}$ is uniformly bounded in H^1 and has a convergent subsequence, also denoted by $\{y_n\}$, such that there exists $y^* \in H^1(\Omega)$ with

$$y_n \rightharpoonup y^* \text{ in } H^1(\Omega).$$

The trace theorem and the Sobolev embedding theorem [1] imply

$$y_n \to y^* \text{ in } L^p(\Gamma_c)$$

for any $p < +\infty$. In particular, we will take $p = 4$. Then we have

$$|\langle u_n(y_n - y^*), v\rangle_{L^2(\Gamma_i)}| \le \|u_n\|_{L^2(\Gamma_i)}\|y_n - y^*\|_{L^4(\Gamma_i)}\|v\|_{L^4(\Gamma_i)} \to 0$$

by the weak convergence of $\{u^n\}$ in $L^2(\Gamma_i)$ and the strong convergence of $\{y_n\}$ in $L^4(\Gamma_i)$. Therefore, we have

$$\lim_{n\to\infty} \langle u_n y_n, v\rangle_{L^2(\Gamma_i)} = \lim_{n\to\infty} \left(\langle u_n(y_n - y^*), v\rangle_{L^2(\Gamma_i)} + \langle u_n y^*, v\rangle_{L^2(\Gamma_i)} \right) = \langle u^* y^*, v\rangle_{L^2(\Gamma_i)}.$$

Now passing to the limit in the weak formulation indicates that y^* satisfies

$$\langle \nabla y^*, \nabla v\rangle_{L^2} + \langle u^* y^*, v\rangle_{L^2(\Gamma_i)} = \langle f, v\rangle_{L^2(\Gamma_c)} \quad \text{for all } v \in H^1(\Omega),$$

i.e., $y^* = y(u^*)$. Since every subsequence has itself a subsequence converging weakly in H^1 to $y(u^*)$, the whole sequence converges weakly. The continuity of $S : u \mapsto y(u)|_{\Gamma_c}$ then follows from the trace theorem and Sobolev embedding theorem for $p = 2$. □

The above two statements imply that property (A1) holds. We next address the remaining properties.

Lemma A.3. *The mapping $u \mapsto y(u)$ is twice Fréchet differentiable from U to $H^1(\Omega)$, and for every $u \in U$ and all directions $h_1, h_2 \in L^2(\Gamma_i)$, the derivatives are given by*

(i) $y'(u)h_1 \in H^1(\Omega)$ is the solution z of

$$\langle \nabla z, \nabla v\rangle_{L^2} + \langle uz, v\rangle_{L^2(\Gamma_i)} = -\langle h_1 y(u), v\rangle_{L^2(\Gamma_i)} \quad \text{for all } v \in H^1(\Omega),$$

and the following estimate holds

$$\|y'(u)h_1\|_{H^1} \le C\|h_1\|_{L^2(\Gamma_i)}.$$

(ii) $y''(u)(h_1, h_2) \in H^1(\Omega)$ is the solution z of

$$\langle \nabla z, \nabla v\rangle_{L^2} + \langle uz, v\rangle_{L^2(\Gamma_i)} = -\langle h_1 y'(u)h_2, v\rangle_{L^2(\Gamma_i)} - \langle h_2 y'(u)h_1, v\rangle_{L^2(\Gamma_i)} \quad \text{for all } v \in H^1(\Omega),$$

and the following estimate holds

$$\|y''(u)(h_1, h_2)\|_{H^1} \le C\|h_1\|_{L^2(\Gamma_i)}\|h_2\|_{L^2(\Gamma_i)}.$$

Proof. The characterization of the derivatives follows from direct calculation. It remains to show boundedness and continuity. By setting $v = y'(u)h_1$ in the weak formulation, Hölder's inequality, the trace theorem and the a priori estimate in Lemma A.1, we have

$$\begin{aligned}\left\|y'(u)h_1\right\|_{\mathrm{H}^1}^2 &\leq C \left\|y'(u)h_1\right\|_{\mathrm{L}^4(\Gamma_i)} \|h_1\|_{\mathrm{L}^2(\Gamma_i)} \|y(u)\|_{\mathrm{L}^4(\Gamma_i)} \\ &\leq C \left\|y'(u)h_1\right\|_{\mathrm{H}^1(\Omega)} \|h_1\|_{\mathrm{L}^2(\Gamma_i)} \|y(u)\|_{\mathrm{H}^1(\Omega)} \\ &\leq C \left\|y'(u)h_1\right\|_{\mathrm{H}^1(\Omega)} \|h_1\|_{\mathrm{L}^2(\Gamma_i)},\end{aligned}$$

from which the first estimate follows. Analogously we deduce that

$$\|y(u+h_1) - y(u)\|_{\mathrm{H}^1} \leq C \|h_1\|_{\mathrm{L}^2(\Gamma_i)}.$$

Next let $w = y(u+h_1) - y(u) - y'(u)h_1$, which satisfies

$$\langle \nabla w, \nabla v\rangle_{\mathrm{L}^2} + \langle uw, v\rangle_{\mathrm{L}^2(\Gamma_i)} = -\langle h_1(y(u+h_1) - y(u)), v\rangle_{\mathrm{L}^2(\Gamma_i)} \qquad \text{for all } v \in \mathrm{H}^1(\Omega).$$

By repeating the proof of the preceding estimate, we deduce that

$$\|w\|_{\mathrm{H}^1} \leq C \|h_1\|_{\mathrm{L}^2(\Gamma_i)} \|y(u+h_1) - y(u)\|_{\mathrm{H}^1},$$

from which it follows directly that $y'(u)h_1$ defined above is indeed the Fréchet derivative of $y(u)$ at u. By arguing similarly and using the first assertion, the second assertion follows. □

Together with the linearity of the trace operator, we obtain $S'(u)h_1 = y'(u)h_1|_{\Gamma_c} \in \mathrm{H}^{\frac{1}{2}}(\Gamma_c)$ and $S''(u)(h_1,h_2) = y''(u)(h_1,h_2)|_{\Gamma_c} \in \mathrm{H}^{\frac{1}{2}}(\Gamma_c)$, and thus property (A2). Finally, properties (A3) and (A4) follow directly from the estimates in Lemma A.3 and the trace theorem [1].

We again give the necessary steps in a Krylov subspace method for the solution to (3.4). For given $u \in \mathrm{L}^2(\Gamma_i)$, $F(u)$ is computed by the following steps:

1: Solve for $y \in \mathrm{H}^1(\Omega)$ in

$$\langle \nabla y, \nabla v\rangle_{\mathrm{L}^2} + \langle uy, v\rangle_{\mathrm{L}^2(\Gamma_i)} = \langle f, v\rangle_{\mathrm{L}^2(\Gamma_c)} \qquad \text{for all } v \in \mathrm{H}^1(\Omega).$$

2: Solve for $p \in \mathrm{H}^1(\Omega)$ in

$$\langle \nabla p, \nabla v\rangle_{\mathrm{L}^2} + \langle up, v\rangle_{\mathrm{L}^2(\Gamma_i)} = -\langle \mathrm{sign}_\beta(y|_{\Gamma_c} - y^\delta), v\rangle_{\mathrm{L}^2(\Gamma_c)} \qquad \text{for all } v \in \mathrm{H}^1(\Omega).$$

3: Set $F(u) = \alpha u + y|_{\Gamma_i} p|_{\Gamma_i}$.

For given $\delta u \in \mathrm{L}^2(\Gamma_i)$, the application of $D_N F(u)$ on δu is computed by:

1: Solve for $\delta y \in \mathrm{H}^1(\Omega)$ in

$$\langle \nabla \delta y, \nabla v\rangle_{\mathrm{L}^2} + \langle u\delta y, v\rangle_{\mathrm{L}^2(\Gamma_i)} = -\langle y\delta u, v\rangle_{\mathrm{L}^2(\Gamma_i)} \qquad \text{for all } v \in \mathrm{H}^1(\Omega).$$

2: Solve for $\delta p \in \mathrm{H}^1(\Omega)$ in

$$\langle \nabla \delta p, \nabla v\rangle_{\mathrm{L}^2} + \langle up, v\rangle_{\mathrm{L}^2(\Gamma_i)} = -\langle \tfrac{1}{\beta}\chi_{\mathcal{I}}(\delta y|_{\Gamma_c}), v\rangle_{\mathrm{L}^2(\Gamma_c)} - \langle p\delta u, v\rangle_{\mathrm{L}^2(\Gamma_i)} \qquad \text{for all } v \in \mathrm{H}^1(\Omega).$$

3: Set $D_N F(u) = \alpha\delta u + p|_{\Gamma_i}(\delta y)|_{\Gamma_i} + y|_{\Gamma_i}(\delta p)|_{\Gamma_i}$.

A.3. Diffusion coefficient problem

In this model problem, the operator S maps $u \in \mathcal{X} = \mathrm{H}^1(\Omega)$ to the solution $y \in \mathcal{Y} = \mathrm{W}_0^{1,q}(\Omega)$, for some $q > 2$, of (1.3), and the admissible set is $U = \{u \in \mathrm{H}^1(\Omega) : \lambda \leq u \leq \lambda^{-1}\}$ for some fixed $\lambda \in (0,1)$. The following estimate is an immediate consequence of Theorem 1 in [24], where $Q > 2$ is a constant depending only on λ and Ω.

Lemma A.4. *There exists a number $Q > 2$ depending only on λ and Ω, such that for any $u \in U$ and $q \in (2, Q)$, problem (1.3) has a unique solution $y \in \mathrm{W}_0^{1,q}(\Omega)$ which satisfies*

$$\|y\|_{\mathrm{W}^{1,q}} \leq C\|f\|_{\mathrm{L}^q}.$$

From this, the uniform boundedness of S follows since $f \in \mathrm{L}^q(\Omega)$ is fixed. We next address the complete continuity of S.

Lemma A.5. *Let $\{u_n\} \subset U$ be a sequence converging weakly in $\mathrm{H}^1(\Omega)$ to $u^* \in U$, then*

$$S(u_n) \to S(u^*) \quad \text{in } \mathrm{L}^2(\Omega).$$

Proof. For $u_n \in U$, set $y_n = S(u_n) \in \mathrm{W}_0^{1,q}(\Omega)$. By the a priori estimate from Lemma A.4, the sequence $\{y_n\}$ is uniformly bounded in $\mathrm{W}^{1,q}$ and has a convergent subsequence also denoted by $\{y_n\}$, such that there exists $y^* \in \mathrm{W}_0^{1,q}(\Omega)$ with

$$y_n \rightharpoonup y^* \text{ in } \mathrm{W}^{1,q}(\Omega).$$

The Rellich-Kondrachov embedding theorem [1, Th. 6.3] implies

$$u_n \to u^* \text{ in } \mathrm{L}^p(\Omega)$$

for any $p < +\infty$. In particular, we will take p such that $\frac{1}{2} + \frac{1}{p} + \frac{1}{q} = 1$. Then we have

$$|\langle (u_n - u^*)\nabla y_n, \nabla v\rangle_{\mathrm{L}^2}| \leq \|u_n - u^*\|_{\mathrm{L}^p}\|\nabla y_n\|_{\mathrm{L}^q}\|\nabla v\|_{\mathrm{L}^2} \to 0$$

by the weak convergence of $\{y^n\}$ in $\mathrm{W}^{1,q}(\Omega)$ and the strong convergence of $\{u_n\}$ in $\mathrm{L}^p(\Omega)$. Therefore, we have

$$\lim_{n\to\infty} \langle u_n \nabla y_n, \nabla v\rangle_{\mathrm{L}^2} = \lim_{n\to\infty} \left(\langle (u_n - u^*)\nabla y_n, \nabla v\rangle_{\mathrm{L}^2} + \langle u^* \nabla y_n, \nabla v\rangle_{\mathrm{L}^2} \right) = \langle u^* \nabla y^*, \nabla v\rangle_{\mathrm{L}^2}.$$

Now passing to the limit in the weak formulation indicates that y^* satisfies

$$\langle u^* \nabla y^*, \nabla v\rangle_{\mathrm{L}^2} = \langle f, v\rangle_{\mathrm{L}^2} \quad \text{for all } v \in \mathrm{H}_0^1(\Omega),$$

i.e., $y^* = S(u^*)$. Since every subsequence has itself a subsequence converging weakly in $\mathrm{W}^{1,q}$ to $S(u^*)$, the whole sequence converges weakly. Applying again the Rellich-Kondrachov embedding theorem [1] for $p = 2$ completes the proof of the lemma. □

The above two statements imply that property (A1) holds. The next statement yields the remaining properties (A2), (A3) and (A4).

Lemma A.6. *The operator $S : U \to \mathrm{W}_0^{1,q}(\Omega)$ is twice Fréchet differentiable, and for every $u \in U$ and all admissible directions $h_1, h_2 \in \mathrm{H}^1(\Omega)$, the derivatives are given by*

(i) $S'(u)h_1 \in \mathrm{W}_0^{1,q}(\Omega)$ *is the solution* z *of*

$$\langle u\nabla z, \nabla v\rangle_{\mathrm{L}^2} = -\langle h_1 \nabla S(u), \nabla v\rangle_{\mathrm{L}^2} \quad \text{for all } v \in \mathrm{H}_0^1(\Omega),$$

and the following estimate holds

$$\|S'(u)h_1\|_{\mathrm{W}^{1,q}} \le C\|h_1\|_{\mathrm{H}^1}.$$

(ii) $S''(u)(h_1, h_2) \in \mathrm{W}_0^{1,q}(\Omega)$ *is the solution* z *of*

$$\langle u\nabla z, \nabla v\rangle_{\mathrm{L}^2} = -\langle h_1 \nabla S'(u)h_2 + h_2\nabla S'(u)h_1, \nabla v\rangle_{\mathrm{L}^2} \quad \text{for all } v \in \mathrm{H}_0^1(\Omega),$$

and the following estimate holds

$$\|S''(u)(h_1, h_2)\|_{\mathrm{W}^{1,q}} \le C\|h_1\|_{\mathrm{H}^1}\|h_2\|_{\mathrm{H}^1}.$$

Proof. Again, the characterization of the derivatives are obtained by direct calculation. Set $y = S(u) \in \mathrm{W}_0^{1,q}(\Omega)$. By Lemma A.4 and Hölder's inequality, we get

$$\begin{aligned}\|S'(u)h_1\|_{\mathrm{W}^{1,q}} &\le C\|h_1\nabla y\|_{\mathrm{L}^q} \le \|h_1\|_{\mathrm{L}^p}\|\nabla y\|_{\mathrm{L}^{q'}} \\ &\le C\|h_1\|_{\mathrm{H}^1}\|\nabla y\|_{\mathrm{L}^{q'}} \le C\|h_1\|_{\mathrm{H}^1},\end{aligned}$$

with $q' \in (q, Q)$ and $\frac{1}{q} = \frac{1}{p} + \frac{1}{q'}$, where we have used the Sobolev embedding theorem and the estimate in Lemma A.4. Analogously, we deduce that

$$\|S(u + h_1) - S(u)\|_{\mathrm{W}^{1,\tilde{q}}} \le C\|h_1\|_{\mathrm{H}^1}.$$

where the exponent $\tilde{q}$ satisfies $\tilde{q} \in (q, Q)$. Next let $w = S(u + h_1) - S(u) - S'(u)h_1$, which satisfies

$$\langle u\nabla w, \nabla v\rangle_{\mathrm{L}^2} = -\langle h_1 \nabla(S(u + h_1) - S(u)), \nabla v\rangle_{\mathrm{L}^2} \quad \text{for all } v \in \mathrm{H}_0^1(\Omega),$$

Repeating the proof of the preceding estimate, we derive

$$\|w\|_{W^{1,p}} \le C\|h_1\|_{H^1}\|S(u + h_1) - S(u)\|_{W^{1,\tilde{q}}}.$$

Combining these estimates yields the first assertion, i.e. $S'(u)h_1$ defined above is indeed the Fréchet derivative of the forward operator $S(u) : H^1 \to W_0^{1,p}(\Omega)$, and it satisfies the desired estimate. Similarly, the second assertion follows from Lemma A.4 and the first assertion. □

We finally address the steps required in a Krylov subspace method for the solution to (3.4). For given $u \in \mathrm{H}^1(\Omega)$, $F(u)$ is computed by the following steps:

1: Solve for $y \in \mathrm{H}_0^1(\Omega)$ in

$$\langle u\nabla y, \nabla v\rangle_{\mathrm{L}^2} = \langle f, v\rangle_{\mathrm{L}^2} \quad \text{for all } v \in \mathrm{H}_0^1(\Omega).$$

2: Solve for $p \in \mathrm{H}_0^1(\Omega)$ in

$$\langle u\nabla p, \nabla v\rangle_{\mathrm{L}^2} = \langle \mathrm{sign}_\beta(y - y^\delta), v\rangle_{\mathrm{L}^2} \quad \text{for all } v \in \mathrm{H}_0^1(\Omega).$$

3: Set $F(u) = \alpha(-\Delta u + u) - \nabla y \cdot \nabla p$.

For given $\delta u \in \mathrm{H}^1(\Omega)$, the application of $D_N F(u)$ on δu is computed by:

1: Solve for $\delta y \in \mathrm{H}_0^1(\Omega)$ in

$$\langle u \nabla \delta y, \nabla v \rangle_{\mathrm{L}^2} = \langle \delta u \nabla y, \nabla v \rangle_{\mathrm{L}^2} \qquad \text{for all } v \in \mathrm{H}_0^1(\Omega).$$

2: Solve for $\delta p \in \mathrm{H}_0^1(\Omega)$ in

$$\langle u \nabla \delta p, \nabla v \rangle_{\mathrm{L}^2} = -\langle \tfrac{1}{\beta} \chi_{\mathcal{I}} \delta y, v \rangle_{\mathrm{L}^2} + \langle \delta u \nabla p, \nabla v \rangle_{\mathrm{L}^2} \qquad \text{for all } v \in \mathrm{H}_0^1(\Omega).$$

3: Set $D_N F(u) = \alpha(-\Delta \delta u + \delta u) - \nabla \delta y \cdot \nabla p - \nabla y \cdot \nabla \delta p$.

B. Tables

r	δ	α_o	α_b	e_o	e_b
0.1	6.46e-2	4.65e-3	1.66e-3	2.64e-4	5.90e-4
0.2	1.13e-1	5.53e-3	2.91e-3	3.49e-4	4.11e-4
0.3	1.68e-1	4.17e-3	4.32e-3	4.60e-4	5.06e-4
0.4	2.26e-1	3.82e-3	5.81e-3	6.26e-4	8.97e-4
0.5	2.94e-1	3.34e-3	7.56e-3	3.54e-3	5.40e-3
0.6	3.19e-1	6.72e-3	8.20e-3	1.31e-2	2.05e-2
0.7	3.86e-1	1.35e-2	9.93e-3	8.47e-3	8.95e-3
0.8	4.38e-1	6.40e-3	1.13e-2	9.27e-3	1.89e-2
0.9	4.85e-1	1.51e-2	1.28e-2	2.04e-1	2.04e-1

Table 1: Comparison of the balancing principle (α_b, e_b) with the sampling-based optimal choice (α_o, e_o) for the 1d inverse potential problem.

Table 2: Convergence of path-following method. For each step k, the parameter $\beta(k)$, number $it(k)$ of SSN iterations and L^2-error $e(k)$ are shown.

$\beta(k)$	1.00e0	5.00e-1	2.50e-1	1.25e-1	6.25e-2	3.12e-2	1.56e-2	7.81e-3	3.91e-3	1.95e-3
$it(k)$	6	4	4	3	3	3	3	3	3	3
$e(k)$	1.83e-1	1.33e-1	1.08e-1	8.77e-2	7.17e-2	5.99e-2	4.63e-2	3.43e-2	2.63e-2	2.11e-2
$\beta(k)$	9.77e-4	4.88e-4	2.44e-4	1.22e-4	6.10e-5	3.05e-5	1.53e-5	7.63e-6	3.81e-6	1.91e-6
$it(k)$	3	3	3	3	3	3	3	3	3	4
$e(k)$	1.70e-2	1.33e-2	1.03e-2	8.06e-3	6.29e-3	4.90e-3	3.80e-3	2.94e-3	2.31e-3	1.85e-3
$\beta(k)$	9.54e-7	4.77e-7	2.38e-7	1.19e-7	5.96e-8	2.98e-8	1.49e-8	7.45e-9	3.73e-9	1.86e-9
$it(k)$	4	5	5	9	14	20	20	20	20	20
$e(k)$	1.51e-3	1.28e-3	1.12e-3	1.00e-3	9.32e-4	9.08e-4	8.90e-4	8.66e-4	8.65e-4	8.66e-4

Table 3: Convergence behavior of the SSN method (for fixed α, β) for the 1d inverse potential problem. Shown are the problem size N, the number $n(k)$ of elements that changed between active and inactive sets and residual norm $r(k) \equiv \|F(u)\|_{L^2}$ after each iteration k.

N	k	1	2	3	4	5
101	$n(k)$	88	0	0	0	0
	$r(k)$	4.40e-2	1.51e-2	8.62e-4	6.58e-6	2.86e-10
1001	$n(k)$	791	6	5	0	0
	$r(k)$	1.23e-1	1.78e-2	2.30e-3	3.53e-5	9.99e-9
10001	$n(k)$	7803	91	16	1	0
	$r(k)$	1.21e-1	1.67e-2	1.68e-3	1.90e-5	2.47e-9

Table 4: Computing times (in seconds) for SSN method (t_s) and fixed-point iteration (t_b) and L^2-error e for the 1d inverse potential problem. Shown are the problem size N, the mean ($\{t_s, t_b, e\}_m$) and standard deviation ($\{t_s, t_b, e\}_s$) over ten noise realizations.

N	100	200	400	800	1600	3200	6400	12800
$t_{s,m}$	1.25	1.75	5.28	12.09	19.40	29.66	55.33	107.87
$t_{s,s}$	0.48	0.45	3.31	4.57	7.22	4.45	11.44	25.92
$t_{b,m}$	7.12	9.63	14.42	39.04	54.19	80.30	131.72	234.00
$t_{b,s}$	3.42	6.21	7.63	17.14	16.79	17.25	38.08	75.21
e_m	8.98e-1	1.51e+0	2.88e-3	9.17e-4	6.22e-4	3.52e-4	2.76e-4	2.78e-4
e_s	2.46e+0	3.16e+0	2.05e-3	5.76e-4	6.83e-4	9.36e-5	4.29e-5	6.93e-5

r	δ	α_o	α_b	e_o	e_b
0.1	1.31e+0	1.40e-1	3.38e-2	1.15e-5	4.10e-5
0.2	2.19e+0	1.07e-1	5.62e-2	4.03e-6	1.46e-5
0.3	3.41e+0	2.45e-1	8.76e-2	8.63e-4	1.22e-3
0.4	5.30e+0	6.27e-1	1.36e-1	2.64e-3	5.40e-3
0.5	6.00e+0	5.01e-1	1.54e-1	4.10e-4	1.53e-3
0.6	7.31e+0	4.41e-1	1.88e-1	3.72e-2	6.30e-2
0.7	9.13e+0	3.40e-1	2.35e-1	6.07e-3	6.36e-3
0.8	9.79e+0	2.53e-1	2.53e-1	6.59e-2	6.59e-2
0.9	1.16e+1	6.00e-1	3.16e-1	3.02e-1	3.37e-1

Table 5: Comparison of the balancing principle (α_b, e_b) with the sampling-based optimal choice (α_o, e_o) for the inverse Robin coefficient problem.

r	δ	α_o	α_b	e_o	e_b
0.1	2.08e-2	9.20e-5	1.52e-4	2.34e-5	2.71e-5
0.2	3.81e-2	1.43e-4	2.77e-4	2.17e-5	2.28e-5
0.3	5.82e-2	3.41e-4	3.99e-4	2.68e-5	3.98e-5
0.4	8.54e-2	2.11e-4	5.71e-4	2.59e-5	3.84e-5
0.5	9.45e-2	3.83e-4	5.98e-4	3.56e-5	4.28e-5
0.6	1.22e-1	1.28e-3	7.45e-4	2.82e-4	3.60e-4
0.7	1.42e-1	2.04e-3	8.35e-4	8.01e-4	1.31e-3
0.8	1.55e-1	1.66e-3	8.71e-4	4.82e-4	6.52e-4
0.9	1.80e-1	4.33e-3	9.42e-4	2.11e-3	6.49e-3

Table 6: Comparison of the balancing principle (α_b, e_b) with the sampling-based optimal choice (α_o, e_o) for the 1d inverse diffusion coefficient problem.

Table 7: Computing times (in seconds) for the SSN method (t_s) and fixed point iteration (t_b) and L^2-error e for the 1d inverse diffusion coefficient problem. Shown are the problem size N, the mean ($\{t_s, t_b, e\}_m$) and standard deviation ($\{t_s, t_b, e\}_s$) over ten noise realizations.

N	100	200	400	800	1600	3200	6400	12800
$t_{s,m}$	0.70	2.19	6.49	11.46	25.48	55.34	82.38	167.71
$t_{s,s}$	0.46	1.65	0.91	2.80	5.34	17.19	23.07	31.01
$t_{b,m}$	4.12	8.68	19.27	27.59	52.45	97.32	154.46	332.96
$t_{b,s}$	2.51	3.71	2.20	6.28	9.14	21.96	22.59	39.17
e_m	2.17e-1	8.03e-2	4.94e-5	3.20e-5	2.89e-5	3.33e-5	3.39e-5	3.08e-5
e_s	2.05e-1	1.53e-1	2.66e-5	5.86e-6	4.89e-6	6.74e-6	5.42e-6	3.25e-6

References

[1] R. A. Adams and J. J. F. Fournier, *Sobolev Spaces*, Elsevier/Academic Press, Amsterdam, 2nd ed., 2003.

[2] W. K. Allard, *Total variation regularization for image denoising. I. Geometric theory*, SIAM J. Math. Anal., 39 (2007/08), pp. 1150–1190.

[3] S. Alliney and S. A. Ruzinsky, *An algorithm for the minimization of mixed l_1 and l_2 norms with application to Bayesian estimation*, IEEE Trans. Signal Process., 42 (1994), pp. 618–627.

[4] H. T. Banks and K. Kunisch, *Estimation Techniques for Distributed Parameter Systems*, Birkhäuser, Boston, 1989.

[5] A. C. Bovik, *Handbook of Image and Video Processing (Communications, Networking and Multimedia)*, Academic Press, Inc., Orlando, 2005.

[6] S. Chaabane, J. Ferchichi, and K. Kunisch, *Differentiability properties of the L^1-tracking functional and application to the Robin inverse problem*, Inverse Problems, 20 (2004), pp. 1083–1097.

[7] T. F. Chan and S. Esedoḡlu, *Aspects of total variation regularized L^1 function approximation*, SIAM J. Appl. Math., 65 (2005), pp. 1817–1837.

[8] X. Chen, Z. Nashed, and L. Qi, *Smoothing methods and semismooth methods for nondifferentiable operator equations*, SIAM J. Numer. Anal., 38 (2000), pp. 1200–1216.

[9] Z. Chen and J. Zou, *An augmented Lagrangian method for identifying discontinuous parameters in elliptic systems*, SIAM J. Control Optim., 37 (1999), pp. 892–910.

[10] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Classics Appl. Math. 5, SIAM, Philadelphia, second ed., 1990.

[11] C. Clason, B. Jin, and K. Kunisch, *A duality-based splitting method for ℓ^1-TV image restoration with automatic regularization parameter choice*, SIAM J. Sci. Comput., 32 (2010), pp. 1484–1505.

[12] ——, *A semismooth Newton method for L^1 data fitting with automatic choice of regularization parameters and noise calibration*, SIAM J. Imaging Sci., 3 (2010), pp. 199–231.

[13] Y. Dong, M. Hintermüller, and M. Neri, *An efficient primal-dual method for ℓ^1tv image restoration*, SIAM J. Imaging Sci., 2 (2009), pp. 1168–1189.

[14] V. Duval, J.-F. Aujol, and Y. Gousseau, *The TVL1 model: a geometric point of view*, Multiscale Model. Simul., 8 (2009), pp. 154–189.

[15] I. Ekeland and R. Témam, *Convex Analysis and Variational Problems*, Classics Appl. Math. 28, SIAM, Philadelphia, 1999.

[16] L. C. Evans and R. F. Gariepy, *Measure Theory and Fine Properties of Functions*, Studies in Advanced Mathematics, CRC Press, Boca Raton, 1992.

[17] M. Hintermüller, K. Ito, and K. Kunisch, *The primal-dual active set strategy as a semismooth Newton method*, SIAM J. Optim., 13 (2002), pp. 865–888 (2003).

[18] M. Hintermüller and M. Ulbrich, *A mesh-independence result for semismooth Newton methods*, Math. Program., 101 (2004), pp. 151–184.

[19] P. J. Huber, *Robust Statistics*, John Wiley & Sons Inc., New York, 1981.

[20] K. Ito and K. Kunisch, *Lagrange Multiplier Approach to Variational Problems and Applications*, Adv. Des. Control 15, SIAM, Philadelphia, 2008.

[21] B. Jin and J. Zou, *Numerical estimation of the Robin coefficient in a stationary diffusion equation*, IMA J. Numer. Anal., 30 (2010), pp. 677–701.

[22] T. Kärkkäinen, K. Kunisch, and K. Majava, *Denoising of smooth images using L^1-fitting*, Computing, 74 (2005), pp. 353–376.

[23] A. Kröner and B. Vexler, *A priori error estimates for elliptic optimal control problems with a bilinear state equation*, J. Comput. Appl. Math., 230 (2009), pp. 781–802.

[24] N. G. Meyers, *An L^p-estimate for the gradient of solutions of second order elliptic divergence equations*, Ann. Scuola Norm. Sup. Pisa (3), 17 (1963), pp. 189–206.

[25] L. Q. Qi and J. Sun, *A nonsmooth version of Newton's method*, Math. Programming, 58 (1993), pp. 353–367.

[26] S. Stojanovic, *Optimal damping control and nonlinear elliptic systems*, SIAM J. Control Optim., 29 (1991), pp. 594–608.

[27] F. Tröltzsch, *Optimal Control of Partial Differential Equations: Theory, Methods and Applications*, American Mathematical Society, Providence, 2010.

[28] M. Yamamoto and J. Zou, *Simultaneous reconstruction of the initial temperature and heat radiative coefficient*, Inverse Problems, 17 (2001), pp. 1181–1202.

[29] J. Yang, Y. Zhang, and W. Yin, *An efficient tvl1 algorithm for deblurring multichannel images corrupted by impulsive noise*, SIAM J. Sci. Comput., 31 (2009), pp. 2842–2865.

[30] W. W.-G. Yeh, *Review of parameter identification procedures in groundwater hydrology: the inverse problem*, Water Resource Research, 22 (1986), pp. 95–108.

[31] W. Yin, D. Goldfarb, and S. Osher, *The total variation regularized L^1 model for multiscale decomposition*, Multiscale Model. Simul., 6 (2007), pp. 190–211.