SpezialForschungsBereich F 32

Karl–Franzens Universität Graz
Technische Universität Graz
Medizinische Universität Graz

A primal-dual semi-smooth Newton method for nonlinear L^1 data fitting problems

Christian Clason Bangti Jin

SFB-Report No. 2010–025 June 2010

A–8010 GRAZ, HEINRICHSTRASSE 36, AUSTRIA

Supported by the
Austrian Science Fund (FWF)

SFB sponsors:

- **Austrian Science Fund** (FWF)
- **University of Graz**
- **Graz University of Technology**
- **Medical University of Graz**
- **Government of Styria**
- **City of Graz**

A primal-dual semi-smooth Newton method for nonlinear L^1 data fitting problems

Christian Clason* Bangti Jin†

June 26, 2010

This work is concerned with L^1 data fitting for nonlinear inverse problems. This formulation is advantageous if the data is corrupted by impulsive noise. However, the problem is not differentiable and lacks local uniqueness, which makes its efficient solution challenging. By considering a regularized primal-dual formulation of this problem, local uniqueness can be shown under a second order sufficient condition and a semi-smooth Newton method becomes applicable. In particular, its super-linear convergence is proved for the discretized optimality system. The convergence of the regularized formulation as the regularization and discretization parameters go to zero is shown. Additionally, approximation properties of the minimizers to nonlinear functionals with L^1 data fitting are analyzed and a strategy for selecting the regularization parameter based on a balancing principle is suggested. The efficiency is illustrated through the model problem of recovering the potential in an elliptic boundary value problem from distributed observational data, for which one- and two-dimensional numerical examples are presented.

1. Introduction

We are interested in the minimization problem

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

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

†Center for Industrial Mathematics, University of Bremen, D-28334 Bremen, Germany, (btjin@math.uni-bremen.de).

arising in the Tikhonov approach for nonlinear inverse problems $S(u) = y^\delta$. Here $y^\delta \in L^\infty(\Omega)$ represents experimental measurements corrupted by (impulsive) noise, $\Omega \subset \mathbb{R}^d$ ($d = 1, 2, 3$) is open and bounded with a Lipschitz boundary $\partial\Omega$ and $S : L^2(\Omega) \to Y$ is the parameter-to-observation mapping. In this work, we assume that Y compactly embeds into $L^2(\Omega)$, which is motivated by parameter identification problems for partial differential equations. In particular, we are interested in various structural properties of the L^1-norm compared with the L^2-norm. Our main goal in this work is to resolve the 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.

The L^1 data fitting in the functional $\mathcal{J}_\alpha$ is motivated by the non-Gaussian nature of the noise under consideration. 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, arise in image processing because of malfunctioning pixels in camera sensors, faulty memory locations in hardware, or transmission in noisy channels [4]. As was demonstrated in, e.g., [11], L^2 data fitting is inadequate for coping with such noise, whereas L^1 data fitting allows accurate reconstructions. Statistically speaking, L^1 data fitting is more robust to outliers than the more conventional L^2 counterpart, in the sense that outliers have less influence on the solution [23]. Consequently, functionals involving L^1 data fitting have recently received considerable interest, e.g., in imaging sciences [27, 11] and parameter identification [5], and are deemed suitable when the data possibly contains outliers.

Because of the practical interest in functionals with L^1 data fitting, there has been a growing interest in analyzing their properties as well as in efficient minimization algorithms. A number of recent works have addressed the analytical properties of L^1 fitting approaches, 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 [6, 2, 34, 13], i.e., when K is the identity operator. In addition, several efficient algorithms [33, 12, 11, 10] have been developed.

However, all these works are only concerned with linear inverse problems, and their analysis and algorithms are not directly applicable to the nonlinear case. In this work, we shall first investigate the properties of the functional $\mathcal{J}_\alpha$ by adopting a general theoretical framework for inverse problems [16, 15]. In particular, the well-posedness of the formulation and convergence rates of the approximation for both a priori and a posteriori parameter choice rules are shown. Additionally, we show an exact recovery result, which is a special feature of L^1 data fitting. We also propose an adaptive rule for choosing regularization parameters based on a balancing principle.

Since the optimality system will in general not have a locally unique solution and thus can not be solved directly with a (semi-smooth) Newton method, we consider a regularized primal-dual version. Under a second order sufficient condition, we prove existence of a locally unique solution and convergence as the regularization parameter tends to zero. The numerical solution of the regularized optimality system is carried out using the finite element method, and we show convergence as the mesh size h tends to zero. The discrete optimality system is then solved by a semi-smooth

Newton method, for which we establish superlinear convergence. To the best of our knowledge, this work represents the first investigation on L^1 data fitting with general nonlinear inverse problems.

To show the applicability of our approach and to illustrate its behavior numerically, we consider as a model problem the mapping $S : u \mapsto y$, where $y \in Y := H^1(\Omega)$ solves the following elliptic boundary value problem

$$\langle \nabla y, \nabla v \rangle_{L^2} + \langle uy, v \rangle_{L^2} = \langle f, v \rangle_{L^2} \tag{1.1}$$

for all $v \in H^1(\Omega)$, where $f \in L^2(\Omega)$ is given.

Remark 1.1. In this work, we assume that the domain of definition of S is $L^2(\Omega)$, in order to focus the presentation on the nonlinear L^1-fitting problem. There is no fundamental difficulty in treating operators defined on a proper subset of $L^2(\Omega)$, 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 problem), the modified optimality system can still be solved using a semi-smooth Newton method after applying a Moreau-Yosida regularization, cf. [21].

This work is organized as follows. In Section 2, we show the well-posedness of the formulation for general nonlinear inverse problems with L^1 data fitting, derive convergence rates and present our parameter choice method. Section 3.1 derives the optimality system for Problem ($\mathcal{P}$), while the regularized saddle-point problem is studied in Section 3.2. Section 4 is devoted to the discretization of the problem and its solution using a semi-smooth Newton method. In Section 5, we present numerical results for our model problem in one and two dimensions. Two appendices verify the assumptions invoked on the operator S and compute the explicit form of the semi-smooth Newton system for the model problem, respectively.

2. Convergence theory

Since convergence rate results for Tikhonov regularization of nonlinear L^1 fitting problems have not yet been derived elsewhere, we first treat the general formulation. Consider the nonlinear operator $S : L^2(\Omega) \to Y \subset L^1(\Omega)$ and the Tikhonov functional $\mathcal{J}_\alpha : L^2(\Omega) \to \mathbb{R}$,

$$\mathcal{J}_\alpha(u) = \|S(u) - y^\delta\|_{L^1} + \frac{\alpha}{2} \|u - u_0\|_{L^2}^2,$$

where u_0 is an initial guess which also plays the role of a selection criterion. We shall denote a minimizer of the functional $\mathcal{J}_\alpha$ by u_α^δ, while u_α will be a minimizer with y^δ replaced by the exact data $y^\dagger$. We assume that $y^\dagger$ is attainable, i.e., that there exists an element $u^\dagger \in L^2(\Omega)$ such that $y^\dagger = S(u^\dagger)$. If this element is not unique, $u^\dagger$ always refers to a u_0-minimum-norm solution, i.e., an element minimizing $\|u - u_0\|_{L^2}$ over the set of solutions to $S(u) = y^\dagger$.

We shall assume:

(A1) The operator S is uniformly bounded in $\mathrm{L}^2(\Omega)$ and completely continuous, i.e., if for $u \in \mathrm{L}^2(\Omega)$, the sequence $\{u_n\} \subset \mathrm{L}^2(\Omega)$ satisfies $u_n \rightharpoonup u$ in $\mathrm{L}^2(\Omega)$, then

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

We first address the well-posedness of the formulation. The proof of the next result is quite standard (c.f., e.g., [16],[15, Chap. 10]), and is thus omitted.

Theorem 2.1. *Under Assumption (A1), problem ($\mathcal{P}$) is well-posed and consistent, i.e.,*

(i) There exists at least one minimizer u_α^δ to problem ($\mathcal{P}$);

(ii) For a sequence of data $\{y_n\}$ such that $y_n \to y^\delta$ in $\mathrm{L}^1(\Omega)$, the sequence of minimizers $\{u_\alpha^n\}$ contains a subsequence converging to u_α^δ;

(iii) If the regularization parameter $\alpha = \alpha(\delta)$ satisfies

$$\lim_{\delta \to 0} \alpha(\delta) = \lim_{\delta \to 0} \tfrac{\delta}{\alpha(\delta)} = 0,$$

then the sequence $\{u^\delta_{\alpha(\delta)}\}$ has a subsequence converging to $u^\dagger$ as $\delta \to 0$.

Then we have the following result on the convergence rate for the a priori parameter choice rule $\alpha = \alpha(\delta) \sim \delta^\varepsilon$ for any $\varepsilon \in (0,1)$. Throughout, C denotes a generic constant, whose value may differ at different occurrences.

Theorem 2.2. *Let $y^\delta \in Y$ satisfy $\|y^\delta - y^\dagger\|_{\mathrm{L}^1} \le \delta$ and $u^\dagger$ be a u_0-minimum norm solution of $S(u) = y^\dagger$. Moreover, let the following conditions hold:*

(i) S is Fréchet differentiable;

(ii) There exists $L > 0$ such that $\|S'(u^\dagger) - S'(z)\|_{\mathrm{L}^2} \le L\|u^\dagger - z\|_{\mathrm{L}^2}$ for all $z \in \mathrm{L}^2(\Omega)$;

(iii) There exists a $w \in \mathrm{L}^\infty(\Omega) \cap \mathrm{L}^2(\Omega)$ with $L\,\|w\|_{\mathrm{L}^2} < 1$ satisfying $u^\dagger - u_0 = S'(u^\dagger)^* w$.

Then for any fixed $\varepsilon \in (0,1)$, the choice $\alpha \sim \delta^\varepsilon$ and δ sufficiently small, we have the estimate

$$\|u_\alpha^\delta - u^\dagger\|_{\mathrm{L}^2} \le C\delta^{\frac{1-\varepsilon}{2}}.$$

Proof. By the minimizing property of u_α^δ and $\|y^\delta - y^\dagger\|_{\mathrm{L}^1} \le \delta$, we obtain

$$\|S(u_\alpha^\delta) - y^\delta\|_{\mathrm{L}^1} + \frac{\alpha}{2}\|u_\alpha^\delta - u_0\|_{\mathrm{L}^2}^2 \le \delta + \frac{\alpha}{2}\|u^\dagger - u_0\|_{\mathrm{L}^2}^2,$$

and hence

$$\|S(u_\alpha^\delta) - y^\delta\|_{\mathrm{L}^1} + \frac{\alpha}{2}\|u_\alpha^\delta - u^\dagger\|_{\mathrm{L}^2}^2 \le \delta + \alpha\langle u^\dagger - u_0, u^\dagger - u_\alpha^\delta\rangle_{\mathrm{L}^2}.$$

Now by the source condition (iii), we obtain

$$\|S(u_\alpha^\delta) - y^\delta\|_{\mathrm{L}^1} + \frac{\alpha}{2}\|u_\alpha^\delta - u^\dagger\|_{\mathrm{L}^2}^2 \le \delta + \alpha\langle w, S'(u^\dagger)(u^\dagger - u_\alpha^\delta)\rangle_{\mathrm{L}^2}.$$

The Fréchet differentiability of S and condition (ii) imply

$$S(u_\alpha^\delta) = S(u^\dagger) + S'(u^\dagger)(u_\alpha^\delta - u_0) + r(u_\alpha^\delta, u^\dagger)$$

with $\|r(u_\alpha^\delta, u^\dagger)\|_{L^2} \le \frac{L}{2}\|u_\alpha^\delta - u^\dagger\|_{L^2}^2$. Combining these estimates leads to

$$\begin{aligned}\|S(u_\alpha^\delta) - y^\delta\|_{L^1} + \frac{\alpha}{2}\|u_\alpha^\delta - u^\dagger\|_{L^2}^2 &\le \delta - \alpha\langle w, (y^\dagger - y^\delta) + (y^\delta - S(u_\alpha^\delta)) + r(u_\alpha^\delta, u^\dagger)\rangle_{L^2} \\ &\le \delta + \alpha\,\|w\|_{L^\infty}\,\delta + \alpha\,\|w\|_{L^\infty}\,\|S(u_\alpha^\delta) - y^\delta\|_{L^1} \\ &\quad + \frac{\alpha}{2}L\,\|w\|_{L^2}\,\|u_\alpha^\delta - u^\dagger\|_{L^2}^2,\end{aligned}$$

and hence

$$(1 - \alpha\,\|w\|_{L^\infty})\|S(u_\alpha^\delta) - y^\delta\|_{L^1} + \frac{\alpha}{2}(1 - L\,\|w\|_{L^2})\|u_\alpha^\delta - u^\dagger\|_{L^2}^2 \le \delta + \alpha\,\|w\|_{L^\infty}\,\delta.$$

Now the desired result follows from the condition $L\,\|w\|_{L^2} < 1$ and the choice of α such that $\alpha\,\|w\|_{L^\infty} < 1$ for δ sufficiently small. □

Remark 2.3. An inspection of the proof shows that a convergence rate of order $\mathcal{O}(\delta^{\frac{1}{2}})$ can be achieved for a choice rule $\alpha(\delta)$ for which the limit $\alpha^* = \lim_{\delta\to 0}\alpha(\delta)$ satisfies $\alpha^* < 1/\,\|w\|_{L^\infty}$ and $\alpha^* > 0$.

The a priori choice gives only an order of magnitude for α and is thus practically inconvenient to use. In contrast, Morozov's discrepancy principle [29, 26] enables constructing a concrete scheme for determining the regularization parameter α. Specifically, we choose $\alpha = \alpha(\delta)$ such that

$$\|S(u_\alpha^\delta) - y^\delta\|_{L^1} = c\delta,$$

where $c \ge 1$ is a constant.

Theorem 2.4 (discrepancy principle)**.** *Assume that conditions (i)-(iii) in Theorem 2.2 are satisfied. Then for the choice α determined by the discrepancy principle, there exists a constant $C > 0$ such that*

$$\|u_\alpha^\delta - u^\dagger\|_{L^2} \le C\delta^{\frac{1}{2}}.$$

Proof. By the definition of the discrepancy principle, we have

$$\|u_\alpha^\delta - u_0\|_{L^2}^2 \le \|u^\dagger - u_0\|_{L^2}^2,$$

and hence

$$\|u_\alpha^\delta - u^\dagger\|_{L^2}^2 \le 2\langle u^\dagger - u_0, u^\dagger - u_\alpha^\delta\rangle_{L^2}.$$

Now by the source condition, we obtain

$$\|u_\alpha^\delta - u^\dagger\|_{L^2}^2 \le 2\langle w, S'(u^\dagger)(u^\dagger - u_\alpha^\delta)\rangle_{L^2}.$$

The Fréchet differentiability of S implies

$$S(u_\alpha^\delta) = S(u^\dagger) + S'(u^\dagger)(u_\alpha^\delta - u^\dagger) + r(u_\alpha^\delta, u^\dagger),$$

with $\|r(u_\alpha^\delta, u^\dagger)\|_{L^2} \leq \frac{L}{2}\|u_\alpha^\delta - u^\dagger\|_{L^2}^2$. Combining these estimates leads to

$$\begin{aligned}\|u_\alpha^\delta - u^\dagger\|_{L^2}^2 &\leq -2\langle w, (y^\dagger - y^\delta) + (y^\delta - S(u_\alpha^\delta)) + r(u_\alpha^\delta, u^\dagger)\rangle_{L^2} \\ &\leq 2\,\|w\|_{L^\infty}\,\delta + 2\,\|w\|_{L^\infty}\,\|S(u_\alpha^\delta) - y^\delta\|_{L^1} + L\,\|w\|_{L^2}\,\|u_\alpha^\delta - u^\dagger\|_{L^2}^2,\end{aligned}$$

and hence, by appealing to the discrepancy principle again,

$$(1 - L\,\|w\|_{L^2})\|u_\alpha^\delta - u^\dagger\|_{L^2}^2 \leq 2(1+c)\,\|w\|_{L^\infty}\,\delta.$$

Now the condition $L\,\|w\|_{L^2} < 1$ implies

$$\|u_\alpha^\delta - u^\dagger\|_{L^2} \leq \sqrt{\tfrac{2(1+c)}{1-L\|w\|_{L^2}}}\,\delta^{\frac{1}{2}}.$$

This shows the desired estimate with $C = \sqrt{\frac{2(1+c)}{1-L\|w\|_{L^2}}}$. □

The efficient solution of Morozov's principle can be carried out by either a two-parameter algorithm based on model functions or the secant method, and we refer to [26] for details.

The next result shows an interesting property of L^1 data fitting in the case of exact data: the regularized reconstruction u_α coincides with the exact solution $u^\dagger$ if the regularization parameter α is sufficiently small. This is in sharp contrast with L^2 data fitting, where the Tikhonov minimizer is different from the true solution for every $\alpha > 0$.

Theorem 2.5 (exact recovery)**.** *Assume that conditions (i)-(iii) in Theorem 2.2 are satisfied. Then, $u_\alpha = u^\dagger$ holds for $\alpha > 0$ sufficiently small.*

Proof. By the minimizing properties of u_α, we obtain

$$\|S(u_\alpha) - y^\dagger\|_{L^1} + \frac{\alpha}{2}\|u_\alpha - u_0\|_{L^2}^2 \leq \frac{\alpha}{2}\|u^\dagger - u_0\|_{L^2}^2,$$

and consequently

$$\|S(u_\alpha) - y^\dagger\|_{L^1} + \frac{\alpha}{2}\|u_\alpha - u^\dagger\|_{L^2}^2 \leq \alpha\langle u^\dagger - u_0, u_\alpha - u^\dagger\rangle_{L^2}.$$

By utilizing the source condition, we arrive at

$$\|S(u_\alpha) - y^\dagger\|_{L^1} + \frac{\alpha}{2}\|u_\alpha - u^\dagger\|_{L^2}^2 \leq \alpha\langle w, S'(u^\dagger)(u_\alpha - u^\dagger)\rangle_{L^2}.$$

As before, we obtain by the Fréchet differentiability of S that

$$[t] \quad \|S(u_\alpha) - y^\dagger\|_{L^1} + \frac{\alpha}{2}\|u_\alpha - u^\dagger\|_{L^2}^2 \le -\alpha\langle w, (y^\dagger - S(u_\alpha)) + r(u_\alpha, u^\dagger)\rangle_{L^2} \le \alpha\,\|w\|_{L^\infty}\,\|S(u_\alpha) - y^\dagger\|_{L^1} + \frac{\alpha}{2}L\,\|w\|_{L^2}\,\|u_\alpha - u^\dagger\|_{L^2}^2.$$

Rearranging the terms gives

$$(1 - \alpha\,\|w\|_{L^\infty})\|S(u_\alpha) - y^\dagger\|_{L^1} + \frac{\alpha}{2}(1 - L\,\|w\|_{L^2})\|u_\alpha - u^\dagger\|_{L^2}^2 \le 0.$$

From this inequality, for $\alpha \le 1/\,\|w\|_{L^\infty}$, we have $\|u_\alpha - u^\dagger\|_{L^2} = 0$, i.e. $u_\alpha = u^\dagger$. This shows the desired assertion. □

The main ingredient for deriving the preceding convergence rate results is the source condition, i.e. condition (iii) of Theorem 2.2. It is well-known that without such a condition the convergence of u_α^δ to $u^\dagger$ can be arbitrarily slow [15]. The source condition represents a certain smoothness assumption on the solution; see [16] for an illustration in the context of parameter identification for partial differential equations. The Lipschitz condition on the derivative together with the smallness condition on the source representer w controls the nonlinear term. Alternative nonlinearity conditions (cf., e.g., [20]) might also be used to derive similar convergence rates results.

In practice, the noise level δ may be unknown, rendering the above rules inapplicable. To circumvent this problem, we propose a heuristic parameter choice rule based on the following balancing principle [11]: Choose α such that

$$(\sigma - 1)\|S(u_\alpha^\delta) - y^\delta\|_{L^1} - \frac{\alpha}{2}\|u_\alpha^\delta\|_{L^2}^2 = 0 \tag{2.1}$$

is satisfied. The underlying idea of the principle is to balance the data fitting term with the regularization term, and the parameter $\sigma > 1$ controls the trade-off between the two terms. We point out that this principle does not require a knowledge of the noise level. It has been successfully applied to linear inverse problems with L^1 data fitting [11, 10].

We compute a solution α^* to the balancing equation (2.1) by the following simple fixed point algorithm proposed in [10]:

$$\alpha_{k+1} = (\sigma - 1)\frac{\|S(u_{\alpha_k}^\delta) - y^\delta\|_{L^1}}{\frac{1}{2}\|u_{\alpha_k}^\delta\|_{L^2}^2}. \tag{2.2}$$

This fixed point algorithm can be derived formally from the model function approach [11]. The convergence can be proven similar to [10], by observing that the proof given there does not depend on the linearity of the forward operator.

Theorem 2.6. *If the initial guess α_0 satisfies $(\sigma - 1)\|S(u_{\alpha_0}^\delta) - y^\delta\|_{L^1} - \frac{\alpha_0}{2}\|u_{\alpha_0}^\delta\|_{L^2}^2 < 0$, then the sequence $\{\alpha_k\}$ generated by the fixed point algorithm is monotonically decreasing and converges to a solution of* (2.1).

3. Optimality system and its regularization

This section derives the optimality system for problem ($\mathcal{P}$) and studies a regularized version which is more suitable for numerical solution. The convergence of solutions of the regularized system to solutions of ($\mathcal{P}$) will be established, and the local uniqueness of a solution to the regularized system will also be studied. To avoid unnecessary notation, we shall suppress the superscript δ and denote minimizers of ($\mathcal{P}$) by u_α from here on.

3.1. Optimality system

In addition to Assumption (A1), we also assume the following properties of the operator S: there exists a neighborhood $U \subset L^2(\Omega)$ of u_α such that

(A2) S is twice Fréchet differentiable.

(A3) There exists a $C > 0$ independent of u, such that

$$\left\|S'(u)h\right\|_{L^2} \leq C \left\|h\right\|_{L^2}$$

holds for all $u \in U$ and $h \in L^2(\Omega)$.

(A4) There exists a $C > 0$ independent of u, such that

$$\left\|S''(u)(h,h)\right\|_{L^2} \leq C \left\|h\right\|_{L^2}^2$$

holds for all $u \in U$ and $h \in L^2(\Omega)$.

The twice-differentiability of S in (A2) is needed for performing the Newton step, see Section 4, as well as ensuring strict differentiability required for the chain rule, see the proof of Theorem 3.1. The a priori estimate in (A3) is required for analyzing the convergence of the regularized solutions, while (A4) will be used to show local uniqueness of the regularized problem and locally superlinear convergence of the semi-smooth Newton method. These assumptions are satisfied for our model problem, see Appendix A.

Theorem 3.1. *For any local minimizer $u_\alpha \in L^2(\Omega)$ of problem ($\mathcal{P}$) there exists a $p_\alpha \in L^\infty(\Omega)$ such that the following relations hold:*

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

Here, $S'(u)^*$ denotes the adjoint of $S'(u)$ with respect to the $L^2(\Omega)$ inner product. Note that both $S(u)$ and y^δ are in $L^2(\Omega)$, and so the duality pairing $\langle S(u) - y^\delta, p\rangle_{L^1,L^\infty}$ reduces to the standard L^2-inner product.

Proof. Setting

$$\mathcal{F} : \mathrm{L}^2(\Omega) \to \mathbb{R}, \quad u \mapsto \frac{\alpha}{2} \|u\|_{\mathrm{L}^2}^2,$$
$$\mathcal{G} : \mathrm{L}^1(\Omega) \to \mathbb{R}, \quad v \mapsto \|v\|_{\mathrm{L}^1},$$

we have that

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

Since by Assumption (A2) the operator S is twice Fréchet differentiable (which implies strict differentiability) and $\mathcal{G}$ is real-valued and convex, the sum and chain rules for the generalized gradient [9, Ths. 2.3.3, 2.3.10] yield that for all $u \in \mathrm{L}^2(\Omega)$, the functional $\mathcal{J}_\alpha$ is Lipschitz continuous near u and

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

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., [9, Prop. 2.3.2]) thus implies the existence of a subgradient $p_\alpha \in \partial \mathcal{G}(S(u_\alpha) - y^\delta) \subset \mathrm{L}^\infty(\Omega)$ such that

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

holds, which is the first relation of (OS). Since $\mathcal{G}$ is convex, the generalized gradient reduces to the convex subdifferential (cf. [9, 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., [14, Chap. I.4]), given by the indicator function of the unit ball $B := \{p \in \mathrm{L}^\infty(\Omega) : \|p\|_{\mathrm{L}^\infty} \leq 1\}$. The subdifferential of $\mathcal{G}^*$ coincides with the normal cone to B, from which 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 $p \in \mathrm{L}^\infty(\Omega)$ 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).

Corollary 3.2. *Let u_α be a minimizer of problem ($\mathcal{P}$), and $p_\alpha \in \mathrm{L}^\infty(\Omega)$ as given by Theorem 3.1. Then the following relations hold for any $p \in \mathrm{L}^\infty(\Omega)$, $p \geq 0$:*

$$\begin{aligned}
\langle S(u_\alpha) - y^\delta, p \rangle_{\mathrm{L}^2} &= 0 \qquad \text{if } \operatorname{supp} p \subset \{x : |p_\alpha(x)| < 1\}, \\
\langle S(u_\alpha) - y^\delta, p \rangle_{\mathrm{L}^2} &\geq 0 \qquad \text{if } \operatorname{supp} p \subset \{x : p_\alpha(x) = 1\}, \\
\langle S(u_\alpha) - y^\delta, p \rangle_{\mathrm{L}^2} &\leq 0 \qquad \text{if } \operatorname{supp} p \subset \{x : p_\alpha(x) = -1\}.
\end{aligned}$$

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

3.2. Regularized optimality system

Since the $L^1(\Omega)$ norm is not strictly convex, we cannot expect even local uniqueness of the solution of problem ($\mathcal{P}$). As this poses a problem for the numerical solution by a semi-smooth Newton method, we introduce a regularization of problem ($\mathcal{P}$). Specifically, we choose $\beta > 0$ and consider

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

Proposition 3.3. *There exists a saddle point* $(u_\beta, p_\beta) \in L^2(\Omega)\times L^\infty(\Omega)$ *of problem* ($\mathcal{P}_\beta$) *satisfying the optimality system*

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

for all $p \in L^\infty(\Omega)$ *with* $\|p\|_{L^\infty} \leq 1$.

Proof. For any fixed $u \in L^2(\Omega)$, consider the problem

$$\max_{\|p\|_{L^\infty}\leq 1} \langle S(u) - y^\delta, p\rangle_{L^2} - \frac{\beta}{2}\|p\|_{L^2}^2 . \tag{3.4}$$

The set $B := \{p \in L^\infty(\Omega) : \|p\|_{L^\infty} \leq 1\}$ is convex and weak-$\star$ sequentially compact in $L^\infty(\Omega)$. Furthermore, $j_u(p) := -\langle p, S(u) - y^\delta\rangle_{L^2} + \frac{\beta}{2}\|p\|_{L^2}^2$ is strictly convex and weak-$\star$ lower semicontinuous. By standard arguments (e.g., [35, Thms. 38.12, 38.C]), problem (3.4) has a unique solution in B, denoted by $p(u)$, which satisfies the necessary optimality condition

$$\langle S(u) - y^\delta - \beta p(u), p - p(u)\rangle_{L^2} \leq 0$$

for all $p \in B$. We next show the complete continuity of $u \mapsto p(u)$ in $L^2(\Omega)$. For any $u, v \in L^2(\Omega)$, consider the two solutions $p(u), p(v) \in B$ of (3.4). Since both are feasible, we can test the optimality condition for $p(u)$ with $p(v)$ and vice versa, and after adding them we obtain

$$\langle S(u) - S(v) - \beta(p(u) - p(v)), p(v) - p(u)\rangle_{L^2} \leq 0.$$

We then deduce by the Cauchy-Schwarz inequality that

$$\beta\|p(u) - p(v)\|_{L^2} \leq \|S(u) - S(v)\|_{L^2} .$$

By Assumption (A1), $u \rightharpoonup v$ in $L^2(\Omega)$ implies $S(u) \to S(v)$ in $L^2(\Omega)$, from which we deduce that $p(u) \to p(v)$ in $L^2(\Omega)$.

Now we return to problem ($\mathcal{P}_\beta$), which by the above considerations can be written as

$$\min_{u\in L^2(\Omega)} \langle S(u) - y^\delta, p(u)\rangle_{L^2} + \frac{\alpha}{2}\|u\|_{L^2}^2 - \frac{\beta}{2}\|p(u)\|_{L^2}^2 .$$

Since Assumption (A1) implies that $S(u)$ is bounded uniformly for $u \in L^2(\Omega)$ and $\|p(u)\|_{L^\infty} \leq 1$ holds, the functional is coercive. Furthermore, complete continuity of $S(u)$ and $p(u)$ and weak lower semicontinuity of norms yields again existence of a minimizer $u_\beta \in L^2(\Omega)$. Setting $p_\beta := p(u_\beta) \in L^\infty(\Omega)$ with $\|p_\beta\|_{L^\infty} \leq 1$, we obtain the existence of a saddle point to ($\mathcal{P}_\beta$).

Setting $J_\beta : L^2(\Omega) \times B \to \mathbb{R}$,

$$J_\beta(u,p) := \langle S(u) - y^\delta, p\rangle_{L^2} + \frac{\alpha}{2}\|u\|_{L^2}^2 - \frac{\beta}{2}\|p\|_{L^2}^2,$$

the first order necessary optimality conditions for a saddle point of J_β are

$$(0,0) \in (\partial_u J_\beta(u_\beta,p_\beta), -\partial_p J_\beta(u_\beta,p_\beta) + N_B(p_\beta)),$$

where $N_B(p)$ denotes the normal cone to B at p. Since J_β is Fréchet differentiable, we arrive at the desired relation (OS_β) by direct calculation. □

Next, we show the convergence of solutions of the regularized problem ($\mathcal{P}_\beta$) to a solution of the L^1 fitting problem ($\mathcal{P}$).

Theorem 3.4. *Let $\beta_n \to 0$. Then the sequence of saddle points $\{(u_{\beta_n}, p_{\beta_n})\} \subset L^2(\Omega) \times L^\infty(\Omega)$ of ($\mathcal{P}_\beta$) has a subsequence converging strongly in $L^2(\Omega)$ and weakly-$\star$ in $L^\infty(\Omega)$, respectively, to a solution $(u_\alpha, p_\alpha) \in L^2(\Omega) \times L^\infty(\Omega)$ of (OS).*

Proof. Let $(u_n, p_n) \equiv (u_{\beta_n}, p_{\beta_n})$ be a saddle point of ($\mathcal{P}_\beta$). Since all p_n satisfy $\|p_n\|_{L^\infty} \leq 1$, the sequence $\{p_n\}$ is uniformly bounded in $L^\infty(\Omega)$. Taking the inner product of the first equation of (OS_β) with u_n, we obtain

$$\begin{aligned}\alpha\|u_n\|_{L^2}^2 &= -\left\langle S'(u_n)^* p_n, u_n\right\rangle_{L^2} = -\left\langle S'(u_n)u_n, p_n\right\rangle_{L^2} \\ &\leq \left\|S'(u_n)u_n\right\|_{L^2}\|p_n\|_{L^2} \leq C\|u_n\|_{L^2},\end{aligned}$$

by the a priori estimate for $S'(u_n)$ from Assumption (A3) and feasibility of the p_n. Consequently, the sequence $\{u_n\}$ is uniformly bounded in $L^2(\Omega)$. Thus there exists a subsequence of $\{(u_n, p_n)\}$, also denoted by $\{(u_n, p_n)\}$, and some $(u_\alpha, p_\alpha) \in L^2(\Omega) \times L^\infty(\Omega)$ such that $u_n \rightharpoonup u_\alpha$ in $L^2(\Omega)$ and $p_n \overset{\star}{\rightharpoonup} p_\alpha$ in $L^\infty(\Omega)$. Furthermore, we have that $\|p_\alpha\|_{L^\infty} \leq \liminf_{n\to\infty}\|p_n\|_{L^\infty} \leq 1$ (cf., e.g., [18, Thms. 1.A.1,2]).

Next, by the saddle point property of (u_n, p_n), we have that for any (u,p) there holds

$$\begin{aligned}(3.5)\quad &\langle S(u_n) - y^\delta, p\rangle_{L^2} + \frac{\alpha}{2}\|u_n\|_{L^2}^2 - \frac{\beta_n}{2}\|p\|_{L^2}^2 \\ &\leq \langle S(u_n) - y^\delta, p_n\rangle_{L^2} + \frac{\alpha}{2}\|u_n\|_{L^2}^2 - \frac{\beta_n}{2}\|p_n\|_{L^2}^2 \\ &\leq \langle S(u) - y^\delta, p_n\rangle_{L^2} + \frac{\alpha}{2}\|u\|_{L^2}^2 - \frac{\beta_n}{2}\|p_n\|_{L^2}^2.\end{aligned}$$

Now by the complete continuity of S from Assumption (A1), we have $S(u_n) \to S(u_\alpha)$ in $L^2(\Omega)$, and consequently

$$\langle S(u_n) - y^\delta, p_n\rangle_{L^2} \to \langle S(u_\alpha) - y^\delta, p_\alpha\rangle_{L^2}$$

holds as $n \to \infty$. Together with the weak lower-semicontinuity of norms, this indicates that the limit (u_α, p_α) satisfies

$$\begin{aligned}\langle S(u_\alpha) - y^\delta, p\rangle_{L^2} + \frac{\alpha}{2}\|u_\alpha\|_{L^2}^2 \\ \leq \langle S(u_\alpha) - y^\delta, p_\alpha\rangle_{L^2} + \frac{\alpha}{2}\|u_\alpha\|_{L^2}^2 \\ \leq \langle S(u) - y^\delta, p_\alpha\rangle_{L^2} + \frac{\alpha}{2}\|u\|_{L^2}^2,\end{aligned}$$

for any $(u,p) \in L^2(\Omega) \times L^\infty(\Omega)$. Thus, the limit (u_α, p_α) is a saddle point of

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

Noting that $\max_{\|p\|_{L^\infty}\leq 1}\langle S(u) - y^\delta, p\rangle_{L^2} = \|S(u) - y^\delta\|_{L^1}$ for $S(u) - y^\delta \in L^2(\Omega)$ concludes the proof of the first assertion.

To show the strong convergence of the subsequence $\{u_n\}$ in $L^2(\Omega)$, it suffices to show that $\limsup_{n\to\infty}\|u_n\|_{L^2} \leq \|u_\alpha\|_{L^2}$. To this end, we assume the contrary, i.e. that there exists a subsequence of $\{u_n\}$, also denoted by $\{u_n\}$, satisfying $u_n \rightharpoonup u_\alpha$ in $L^2(\Omega)$ and $\lim_{n\to\infty}\|u_n\|_{L^2} =: c > \|u_\alpha\|_{L^2}$. Setting $u = u_\alpha$ in the second inequality of (3.5), and taking the limit as $n \to \infty$, we arrive at

$$\langle S(u_\alpha) - y^\delta, p_\alpha\rangle_{L^2} + \frac{\alpha}{2}c^2 \leq \langle S(u_\alpha) - y^\delta, p_\alpha\rangle_{L^2} + \frac{\alpha}{2}\|u_\alpha\|_{L^2}^2,$$

i.e. $c \leq \|u_\alpha\|_{L^2}$, which is in contradiction with the assumption. Therefore the subsequence converges strongly in $L^2(\Omega)$. □

Theorem 3.4 suggests naturally a numerical continuation procedure: Given a sequence of parameters $\beta_n = \beta_0 q^n$ for $q < 1$ and $n = 0, \ldots, n_{\max}$, we compute the saddle point $(u_{\beta_n}, p_{\beta_n})$ using the previous saddle point $(u_{\beta_{n-1}}, p_{\beta_{n-1}})$ as initial guess.

We can now give a sufficient condition for the local uniqueness of the saddle point of the regularized system (OS_β). We assume that (u_β, p_β) is a solution of (OS_β) and that there exists a constant $\gamma > 0$ such that

$$\langle S''(u_\beta)(h,h), p_\beta\rangle_{L^2} + \alpha\|h\|_{L^2}^2 \geq \gamma\|h\|_{L^2}^2 \tag{3.6}$$

holds for all $h \in L^2(\Omega)$. This is related to standard second-order sufficient optimality conditions in PDE-constrained optimization (cf., e.g., [32, Chap. 4.10]). In general, it is not possible to check such second order sufficient conditions a priori (but see [31] for their numerical verification in a semilinear problem). Here, the condition is satisfied for either large α or small p_β, since

$$\langle S''(u_\beta)(h,h), p_\beta\rangle_{L^2} + \alpha \|h\|_{L^2}^2 \geq (\alpha - C\|p_\beta\|_{L^2}) \|h\|_{L^2}^2 \tag{3.7}$$

holds by the uniform a priori estimate on S'' from Assumption (A4). For large β, $\|p_\beta\|_{L^2}$ is small by virtue of the L^2-penalty term on p, while for small β, the convergence result of Theorem 3.4 indicates that we might expect p_β to be close to p_α, which acts as a noise indicator for $S(u_\alpha) - y^\delta$. 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 taken only for small noise levels.

The next result shows the local uniqueness of the saddle point to (OS_β).

Proposition 3.5. *If $(u_\beta, p_\beta) \in L^2(\Omega) \times L^\infty(\Omega)$ satisfies condition (3.6), then there exists an $\varepsilon > 0$ such that for all $(u,p) \in L^2(\Omega) \times L^\infty(\Omega)$ with $\|(u,p) - (u_\beta, p_\beta)\|_{L^2} \leq \varepsilon$ and $(u,p) \neq (u_\beta, p_\beta)$, the following strict inequality is satisfied:*

$$J_\beta(u_\beta, p) < J_\beta(u_\beta, p_\beta) < J_\beta(u, p_\beta).$$

Furthermore, the solution of the optimality system (OS_β) is unique in this neighborhood of (u_β, p_β).

Proof. First, we observe that the first strict inequality follows directly from the strict concavity of the functional with respect p for fixed u_β. Next, by condition (3.6) and Assumption (A4), there exists an $\varepsilon > 0$ such that

$$\begin{aligned}\langle S''(u)(h,h), p\rangle_{L^2} + \alpha \|h\|_{L^2}^2 &= \langle S''(u)(h,h), p_\beta\rangle_{L^2} + \alpha \|h\|_{L^2}^2 + \langle S''(u)(h,h), p - p_\beta\rangle_{L^2} \\ &\geq \gamma \|h\|_{L^2}^2 - C\varepsilon \|h\|_{L^2}^2 \geq \frac{\gamma}{2} \|h\|_{L^2}^2\end{aligned}$$

holds for all (u,p) with $\|(u,p) - (u_\beta, p_\beta)\|_{L^2} \leq \varepsilon$ if ε is sufficiently small. Hence, $J_\beta(\cdot, p)$ is strictly convex in this neighborhood of (u_β, p_β). Consequently we obtain the second part of the assertion of strict inequality.

Furthermore, the strict convexity of $J_\beta(\cdot, p)$ and $-J_\beta(u, \cdot)$ in the neighborhood implies that the operator $T : L^2(\Omega) \times L^\infty(\Omega) \to L^2(\Omega) \times (L^\infty(\Omega))^*$,

$$T(u,p) = \begin{pmatrix} \partial_u J_\beta(u,p) \\ -\partial_p J_\beta(u,p) + N_B(p) \end{pmatrix},$$

where $N_B(p)$ again denotes the normal cone to B at p, is strictly monotone. Since T is also continuous and coercive, the main theorem on monotone operators (e.g., [36, Th. 26.A]) asserts that the generalized equation $(0,0) \in T(u,p)$ has a unique solution in the neighborhood of (u_β, p_β). $\square$

4. Numerical solution of regularized system

In this section, we discuss the numerical solution of the regularized optimality system (OS_β). We introduce a finite element discretization of the saddle point problem, analyze its convergence, and apply a semi-smooth Newton method to the discrete optimality system, for which we show superlinear convergence.

4.1. Discretization

In order to numerically compute a saddle point of the regularized problem ($\mathcal{P}_\beta$), we need to discretize the problem suitably. Here we employ the finite element method. Let $\{\mathcal{T}_h\}_{h>0}$ be a quasi-uniform family of shape regular triangulations of the (polyhedral) domain Ω with $h = \max\{\operatorname{diam}(K) : K \in \mathcal{T}_h\}$. We define the discrete spaces

$$\begin{aligned} V_h &= \left\{v_h \in C(\overline{\Omega}) : v_h|_K \in \mathcal{Q}_1(K) \text{ for all } K \in \mathcal{T}_h\right\}, \\ U_h &= \left\{u_h \in L^2(\Omega) : u_h|_K \in \mathcal{Q}_0(K) \text{ for all } K \in \mathcal{T}_h\right\}, \end{aligned}$$

where the space $\mathcal{Q}_1(K)$ consists of all (bi-/tri-)linear functions defined on the element K and the space $\mathcal{Q}_0(K)$ consists of all constant functions on the element K. We also recall the L^2-orthogonal projection Π_h^0 onto U_h, defined by

$$\Pi_h^0 v|_K = \frac{1}{|K|}\int_K v(x)dx \qquad \text{for all } K \in \mathcal{T}_h,$$

which satisfies the estimates (cf., e.g., [17, Prop. 1.135])

$$\begin{aligned} \left\|\Pi_h^0 v\right\|_{L^2} &\le \|v\|_{L^2} && \text{for all } v \in L^2(\Omega), \\ \left\|\Pi_h^0 v - v\right\|_{L^2} &\le Ch\,\|v\|_{H^1} && \text{for all } v \in H^1(\Omega). \end{aligned}$$

We will take $U_h \subset L^2(\Omega) \cap L^\infty(\Omega)$ as the space for the discrete parameter u_h and dual variable p_h, and $V_h \subset H^1(\Omega)$ as the discrete state space, i.e., we define the discretized operator $S_h(u_h) : U_h \to V_h$ by the finite element approximation of the operator $S(u)$. The corresponding discretization of problem ($\mathcal{P}_\beta$) is then given by

$$(\mathcal{P}_h) \qquad \min_{u_h \in U_h} \max_{\|p_h\|_{L^\infty} \le 1} \langle S_h(u_h) - y^\delta, p_h\rangle_{L^2} + \frac{\alpha}{2}\|u_h\|_{L^2}^2 - \frac{\beta}{2}\|p_h\|_{L^2}^2.$$

By virtue of the finite element approximation, it is reasonable to assume that the discrete operator S_h inherits the required properties of S and has the necessary approximation properties, i.e., that there exists a neighborhood $U \subset U_h$ of $\overline{u}_h$, where $(\overline{u}_h, \overline{p}_h)$ is the saddle point of ($\mathcal{P}_h$), such that

(A1$_h$) S_h is uniformly bounded, and if the sequence $\{u_h\} \subset U_h$ satisfies $u_h \rightharpoonup u$ in $L^2(\Omega)$ as $h \to 0$ for $u \in L^2(\Omega)$, then

$$S_h(u_h) \to S(u) \quad \text{in } L^2(\Omega).$$

(A2$_h$) S_h is twice Fréchet differentiable.

(A3$_h$) There exists a $C > 0$ independent of u_h and h, such that

$$\left\|S_h'(u_h)v_h\right\|_{L^2} \le C\left\|v_h\right\|_{L^2}$$

holds for all $u_h \in U$ and $v_h \in U_h$.

(A4$_h$) There exists a $C > 0$ independent of u_h and h, such that

$$\left\|S''(u_h)(v_h, v_h)\right\|_{L^2} \le C\left\|v_h\right\|_{L^2}^2$$

holds for all $u_h \in U$ and $v_h \in U_h$.

These assumptions hold for our model problem, as is verified (where not immediate) in Appendix A.

The existence of a saddle point $(\overline{u}_h, \overline{p}_h) \in U_h \times U_h$ satisfying

$$\text{(OS}_h\text{)}\qquad \begin{cases} S_h'(\overline{u}_h)^*\overline{p}_h + \alpha\overline{u}_h = 0, \\ \langle S_h(\overline{u}_h) - y^\delta - \beta\overline{p}_h, p - \overline{p}_h\rangle_{L^2} \le 0, \end{cases}$$

for all $p \in U_h$ with $\|p\|_{L^\infty} \le 1$ now follows as before. The first equation in (OS$_h$) should be understood in the weak sense.

The next result shows the convergence of the saddle points $(\overline{u}_h, \overline{p}_h)$ as the mesh size h decreases to zero.

Theorem 4.1. *The sequence of saddle points $\{(\overline{u}_h, \overline{p}_h)\} \subset U_h \times U_h$ of ($\mathcal{P}_h$) has a subsequence converging weakly in $L^2(\Omega)$ and weakly-$\star$ in $L^\infty(\Omega)$, respectively, to a solution $(u_\beta, p_\beta) \in L^2(\Omega) \times L^\infty(\Omega)$ of ($\mathcal{P}_\beta$) as $h \to 0$. Furthermore, the convergence of the subsequence $(\overline{u}_h, \overline{p}_h)$ is strong in $L^2(\Omega) \times L^2(\Omega)$.*

Proof. First, we observe that $\{\overline{p}_h\}$ satisfies $\|\overline{p}_h\|_{L^\infty} \le 1$. Next, due to the optimality of $(\overline{u}_h, \overline{p}_h)$, we obtain

$$\begin{aligned} \alpha\|\overline{u}_h\|_{L^2}^2 &= -\left\langle S_h'(\overline{u}_h)^*\overline{p}_h, \overline{u}_h\right\rangle_{L^2} = -\left\langle S_h'(\overline{u}_h)\overline{u}_h, \overline{p}_h\right\rangle_{L^2} \\ &\le \|S_h'(\overline{u}_h)\overline{u}_h\|_{L^2}\|\overline{p}_h\|_{L^2} \le C\|\overline{u}_h\|_{L^2} \end{aligned}$$

by the boundedness of $\{\overline{p}_h\}$ and Assumption (A3$_h$). Consequently, $\{\overline{u}_h\}$ is uniformly bounded in $L^2(\Omega)$. Thus there exists a subsequence of $\{(\overline{u}_h, \overline{p}_h)\}$, also denoted by $\{(\overline{u}_h, \overline{p}_h)\}$, and some $(u_\beta, p_\beta) \in L^2(\Omega) \times L^\infty(\Omega)$ such that

$$\overline{u}_h \rightharpoonup u_\beta \text{ in } L^2(\Omega), \qquad \overline{p}_h \overset{\star}{\rightharpoonup} p_\beta \text{ in } L^\infty(\Omega).$$

Furthermore, we have that $\|p_\beta\|_{L^\infty} \le \liminf_{h\to 0}\|\overline{p}_h\|_{L^\infty} \le 1$ (cf., e.g., [18, Thms. 1.A.1,2]).

Next we recall the density of $C^\infty(\overline{\Omega})$ in $L^2(\Omega)$ and in $L^\infty(\Omega)$ with respect to the weak-$\star$ topology [19, 25]. Using the standard mollifier construction, we can thus find for any $(u,p) \in L^2(\Omega) \times L^\infty(\Omega)$ a sequence $(u^\varepsilon, p^\varepsilon) \in C^\infty(\overline{\Omega}) \times C^\infty(\overline{\Omega})$ such that as $\varepsilon \to 0$,

$$u^\varepsilon \to u \text{ in } L^2(\Omega), \qquad p^\varepsilon \to p \text{ in } L^2(\Omega), \quad p^\varepsilon \overset{\star}{\rightharpoonup} p \text{ in } L^\infty(\Omega).$$

Also, by this construction, $\|p^\varepsilon\|_{L^\infty} \leq 1$ holds.

For any admissible $(u,p) \in L^2(\Omega) \times L^\infty(\Omega)$, the pair $(\Pi_h^0 u^\varepsilon, \Pi_h^0 p^\varepsilon)$ is in $U_h \times U_h$ and satisfies $\|\Pi_h^0 p^\varepsilon\|_{L^\infty} \leq 1$, and thus it is a valid test function. By the saddle point property of $(\overline{u}_h, \overline{p}_h)$, we have

$$\begin{aligned}(4.1)\quad \langle S_h(\overline{u}_h) - y^\delta, \Pi_h^0 p^\varepsilon\rangle_{L^2} &+ \frac{\alpha}{2}\|\overline{u}_h\|_{L^2}^2 - \frac{\beta}{2}\|\Pi_h^0 p^\varepsilon\|_{L^2}^2 \\ &\leq \langle S_h(\overline{u}_h) - y^\delta, \overline{p}_h\rangle_{L^2} + \frac{\alpha}{2}\|\overline{u}_h\|_{L^2}^2 - \frac{\beta}{2}\|\overline{p}_h\|_{L^2}^2 \\ &\leq \langle S_h(\Pi_h^0 u^\varepsilon) - y^\delta, \overline{p}_h\rangle_{L^2} + \frac{\alpha}{2}\|\Pi_h^0 u^\varepsilon\|_{L^2}^2 - \frac{\beta}{2}\|\overline{p}_h\|_{L^2}^2 .\end{aligned}$$

Now, let h go to zero in (4.1). By the weak lower-semicontinuity of norms and the complete continuity of S_h and the convergence property of Π_h^0, we thus deduce that the limit $(u_\beta, p_\beta) \in L^2(\Omega) \times L^\infty(\Omega)$ satisfies

$$\begin{aligned}\langle S(u_\beta) - y^\delta, p^\varepsilon\rangle_{L^2} &+ \frac{\alpha}{2}\|u_\beta\|_{L^2}^2 - \frac{\beta}{2}\|p^\varepsilon\|_{L^2}^2 \\ &\leq \langle S(u_\beta) - y^\delta, p_\beta\rangle_{L^2} + \frac{\alpha}{2}\|u_\beta\|_{L^2}^2 - \frac{\beta}{2}\|p_\beta\|_{L^2}^2 \\ &\leq \langle S(u^\varepsilon) - y^\delta, p_\beta\rangle_{L^2} + \frac{\alpha}{2}\|u^\varepsilon\|_{L^2}^2 - \frac{\beta}{2}\|p_\beta\|_{L^2}^2 .\end{aligned}$$

Letting ε tend to zero shows that the limit (u_β, p_β) is a saddle point of

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

This shows the first assertion.

To show strong convergence of the subsequence $\{\overline{p}_h\}$ to p_β in $L^2(\Omega)$, it suffices to show that $\limsup_{h\to 0} \|\overline{p}_h\|_{L^2} \leq \|p_\beta\|_{L^2}$. To this end, we assume the contrary, i.e., the existence of a subsequence of $\{\overline{p}_h\}$, also denoted by $\{\overline{p}_h\}$, such that $\overline{p}_h \overset{\star}{\rightharpoonup} p_\beta$ in $L^\infty(\Omega)$ and $\lim_{h\to 0}\|\overline{p}_h\|_{L^2} =: c > \|p_\beta\|_{L^2}$. On the other hand, the limit p_β is feasible, so we can insert it (after mollification and projection) in the first inequality of (4.1) to obtain

$$\langle S_h(\overline{u}_h) - y^\delta, \Pi_h^0 p_\beta^\varepsilon - \overline{p}_h\rangle_{L^2} + \frac{\beta}{2}\|\overline{p}_h\|_{L^2}^2 \leq \frac{\beta}{2}\|\Pi_h^0 p_\beta^\varepsilon\|_{L^2}^2.$$

Letting $h \to 0$, the strong convergence of $S_h(\overline{u}_h)$ to $S(u_\beta)$, weak convergence of $\overline{p}_h$ to p_β, and the assumption implies that

$$\langle S(u_\beta) - y^\delta, p_\beta^\varepsilon - p_\beta\rangle_{L^2} + \frac{\beta}{2}c^2 \leq \frac{\beta}{2}\left\|p_\beta^\varepsilon\right\|_{L^2}^2 .$$

By taking $\varepsilon \to 0$, we obtain the contradiction $\|p_\beta\|_{L^2} < c \leq \|p_\beta\|_{L^2}$ and thus convergence of $L^2(\Omega)$-norms for the subsequence $\{\overline{p}_h\}$. Since weak-$\star$ convergence in $L^\infty(\Omega)$ implies weak convergence in $L^2(\Omega)$, we deduce that $\overline{p}_h \to p_\beta$ in $L^2(\Omega)$. The remaining assertion for $\{\overline{u}_h\}$ can be shown analogously. □

4.2. Semi-smooth Newton method

To solve the discrete optimality system (OS_h) using a semi-smooth Newton method, we rewrite the variational inequality using a complementarity function. First, by the fact that $U_h, V_h \subset L^2(\Omega)$ and the definition of the L^2-orthogonal projection Π_h^0, we have that

$$\langle S_h(u_h) - y^\delta - \beta p_h, p - p_h\rangle_{L^2} = \langle \Pi_h^0(S_h(u_h) - y^\delta) - \beta p_h, p - p_h\rangle_{L^2}$$

holds for all $p_h, p \in U_h$. Since all occurring functions on the right hand side are now piecewise constant, we can interpret the variational inequality element-wise on each element $K \in \mathcal{T}_h$ and express it using a complementarity function for any $c > 0$ as [7, 24]

$$\begin{aligned}(4.2)\quad \Pi_h^0(S_h(u_h) - y^\delta) - \beta p_h - \max(0, \Pi_h^0(S_h(u_h) - y^\delta) - \beta p_h + c(p_h - 1))\\ - \min(0, \Pi_h^0(S_h(u_h) - y^\delta) - \beta p_h + c(p_h + 1)) = 0.\end{aligned}$$

The discrete optimality system (OS_h) can thus be written as an operator equation $T_h(u_h, p_h) = (0, 0)$ for $T_h : U_h \times U_h \to U_h \times U_h$,

$$T_h(u_h, p_h) = \begin{pmatrix} S_h'(u_h)^* p_h + \alpha u_h \\ \Pi_h^0(S_h(u_h) - y^\delta) - \beta p_h - P_h(u_h, p_h)\end{pmatrix},$$

where, for the sake of brevity, we have introduced $P_h : U_h \times U_h \to U_h$,

$$\begin{aligned}P_h(u_h, p_h) = \max(0, \Pi_h^0(S_h(u_h) - y^\delta) - \beta p_h + c(p_h - 1))\\ + \min(0, \Pi_h^0(S_h(u_h) - y^\delta) - \beta p_h + c(p_h + 1)).\end{aligned}$$

We next argue Newton differentiability of T_h. Since U_h is finite dimensional, we know (e.g., from [30]) that the function $z_h \mapsto \max(0, z_h)$ is Newton differentiable from U_h to U_h with its Newton derivative given element-wise by

$$D\max(0, z_h)|_K = \begin{cases} 1, & \text{if } z_h|_K > 0, \\ 0, & \text{if } z_h|_K \leq 0.\end{cases}$$

An analogous statement holds for the min function. It follows that P_h is Newton differentiable with respect to p_h from U_h to U_h for fixed u_h, with its Newton derivative with respect to p_h given by

$$D_p P_h(u_h, p_h) = (c - \beta)\chi_{\mathcal{A}}.$$

Here, $\chi_{\mathcal{A}}$ is the indicator function of the active set defined by

$$\begin{aligned}
\mathcal{A}_+ &:= \bigcup \left\{ K \in \mathcal{T}_h : \left(\Pi_h^0(S_h(u_h) - y^\delta) - \beta p_h + c(p_h - 1) \right) |_K > 0 \right\}, \\
\mathcal{A}_- &:= \bigcup \left\{ K \in \mathcal{T}_h : \left(\Pi_h^0(S_h(u_h) - y^\delta) - \beta p_h + c(p_h + 1) \right) |_K < 0 \right\}, \\
\mathcal{A} &:= \mathcal{A}_+ \cup \mathcal{A}_-,
\end{aligned}$$

and we similarly define the indicator function $\chi_{\mathcal{I}}$ of the inactive set $\mathcal{I} := \Omega \setminus \mathcal{A}$.

By the same arguments, we have that for fixed p_h, P_h is Newton differentiable with respect to u_h from U_h to U_h. By the chain rule and the Fréchet differentiability of S_h, we can calculate its Newton derivative with respect to u_h acting on $v_h \in U_h$ as

$$D_u P_h(u_h, p_h) v_h = \chi_{\mathcal{A}} \Pi_0^h (S_h'(u_h) v_h).$$

Given $(u_h^k, p_h^k) \in U_h \times U_h$, one Newton step towards the solution of the optimality system consists in solving for the increments $(\delta u, \delta p) \in U_h \times U_h$ in

$$\text{(4.3)} \quad \begin{cases} S_h''(u_h^k)(\delta u)^* p_h^k + \alpha \delta u + S_h'(u_h^k)^* \delta p = -S_h'(u_h^k)^* p_h^k - \alpha u_h^k, \\ \chi_{\mathcal{I}^k} \Pi_h^0 (S_h'(u_h^k) \delta u) - (\beta \chi_{\mathcal{I}^k} + c \chi_{\mathcal{A}^k}) \delta p = -\chi_{\mathcal{I}^k} \left(\Pi_0^h (S_h(u_h^k) - y^\delta) - \beta p_h^k \right) \\ \qquad\qquad + c(\chi_{\mathcal{A}_+^k}(p_h^k - 1) + \chi_{\mathcal{A}_-^k}(p_h^k + 1)), \end{cases}$$

and setting $u_h^{k+1} = u_h^k + \delta u$ and $p_h^{k+1} = p_h^k + \delta p$. Both equations should now be understood to hold element-wise. Given a way to compute the action of the derivatives $S_h'(u_h) v_h$, $S_h'(u_h)^* v_h$ and $[S_h''(u_h) v_h]^* p_h$ for given $u_h, p_h, v_h \in U_h$, this discrete system can be solved using an iterative solver such as BiCGstab or GMRES (since the system is not symmetric). For our model problem, the explicit form of the Newton system and the full algorithm are given in Appendix B.

It remains to show the uniform well-posedness of the discrete Newton system, from which superlinear convergence of the semi-smooth Newton method follows by standard arguments. For this, we use a second-order condition for the discrete problem similar to (3.6), which again holds (for h sufficiently small) if (3.7) is satisfied.

Lemma 4.2. *If condition (3.7) holds, then there exists an $\varepsilon > 0$ such that for all $(u_h, p_h) \in U_h \times U_h$ with $\|(u_h, p_h) - (\overline{u}_h, \overline{p}_h)\|_{L^2} \leq \varepsilon$ and h sufficiently small,*

$$\langle S_h''(u_h)(\delta u, \delta u), p_h \rangle_{L^2} + \alpha \|\delta u\|_{L^2}^2 \geq \frac{\gamma}{4} \|\delta u\|_{L^2}^2$$

holds for all $\delta u \in U_h$.

Proof. Using Assumption (A4$_h$), we have that for any $p_h, u_h, \delta u \in U_h$,

$$\begin{aligned}
\langle S_h''(u_h)(\delta u, \delta u), p_h \rangle_{L^2} &= \langle S_h''(u_h)(\delta u, \delta u), p_\beta + (\overline{p}_h - p_\beta) + (p_h - \overline{p}_h) \rangle_{L^2} \\
&\geq -C(\|p_\beta\|_{L^2} + \|\overline{p}_h - p_\beta\|_{L^2} + \|p_h - \overline{p}_h\|_{L^2}) \|\delta u\|_{L^2}^2
\end{aligned}$$

holds. Now, condition (3.7) and the strong convergence of $\overline{p}_h$ to the saddle point p_β (Theorem 4.1) implies that for h and ε sufficiently small,

$$\langle S_h''(u_h)(\delta u, \delta u), p_h\rangle_{L^2} + \alpha \|\delta u\|_{L^2}^2 \geq \frac{\gamma}{4} \|\delta u\|_{L^2}^2$$

holds for all p_h with $\|p_h - \overline{p}_h\|_{L^2} \leq \varepsilon$. □

The following result shows the uniform invertibility of the Newton system.

Proposition 4.3. *If condition (3.7) holds, then for each $(u_h, p_h) \in U_h \times U_h$ sufficiently close to a saddle point $(\overline{u}_h, \overline{p}_h)$ of $(\mathcal{P}_h)$, the mapping $M : U_h \times U_h \to U_h \times U_h$,*

$$M := \begin{pmatrix} S_h''(u_h)(\cdot)^* p_h + \alpha I & S_h'(u_h)^* \\ \chi_{\mathcal{I}} \Pi_h^0 S_h'(u_h) & -(\beta\chi_{\mathcal{I}} + c\chi_{\mathcal{A}}) \end{pmatrix}$$

is invertible, and there exists a constant $C > 0$ independent of (u_h, p_h) such that

$$\|M^{-1}\| \leq C$$

holds.

Proof. For given $(w_1, w_2) \in U_h \times U_h$, we need to find $(\delta u, \delta p) \in U_h \times U_h$ satisfying

$$\begin{cases} S_h''(u_h)(\delta u)^* p_h + \alpha\delta u + S_h'(u_h)^*\delta p = w_1, \\ \chi_{\mathcal{I}} \Pi_h^0 S_h'(u_h)\delta u - (\beta\chi_{\mathcal{I}} + c\chi_{\mathcal{A}})\delta p = w_2. \end{cases} \tag{4.4}$$

Since $\beta\chi_{\mathcal{I}} + c\chi_{\mathcal{A}} > 0$ everywhere, we can rewrite the second equation of (4.4) as

$$\delta p = \beta^{-1}\chi_{\mathcal{I}} \Pi_h^0 S_h'(u_h)\delta u - (\beta\chi_{\mathcal{I}} + c\chi_{\mathcal{A}})^{-1} w_2. \tag{4.5}$$

Inserting this into the first equation of (4.4) yields

$$(S_h''(u_h)\delta u)^* p_h + \alpha\delta u + S_h'(u_h)^* \left(\beta^{-1}\chi_{\mathcal{I}} \Pi_h^0 S_h'(u_h)\delta u\right) \\ = w_1 + S_h'(u_h)^*(\beta\chi_{\mathcal{I}} + c\chi_{\mathcal{A}})^{-1} w_2.$$

Taking the inner product of this equation with δu and noting the definition of the projection Π_h^0, we obtain

$$\left\langle S_h''(u_h)(\delta u, \delta u), p_h\right\rangle_{L^2} + \alpha \|\delta u\|_{L^2}^2 + \beta^{-1} \left\|\Pi_h^0 S_h'(u_h)\delta u\right\|_{L^2(\mathcal{I})}^2 \\ = \langle w_1, \delta u\rangle_{L^2} + \left\langle (\beta\chi_{\mathcal{I}} + c\chi_{\mathcal{A}})^{-1} w_2, S_h'(u_h)\delta u\right\rangle_{L^2}.$$

By Lemma 4.2, the Cauchy-Schwarz inequality and Assumption (A3_h) we deduce that

$$\frac{\gamma}{4} \|\delta u\|_{L^2}^2 \leq \|w_1\|_{L^2} \|\delta u\|_{L^2} + \beta^{-1} \|w_2\|_{L^2} \|\delta u\|_{L^2}. \tag{4.6}$$

Going back to (4.5) and using again the Cauchy-Schwarz inequality, we find that

$$\|\delta p\|_{L^2} \leq (\beta\chi_{\mathcal{I}} + c\chi_{\mathcal{A}})^{-1} \left\|S_h'(u_h)\delta u\right\|_{L^2} + \beta^{-1} \|w_2\|_{L^2}$$

holds. Using again Assumption (A3$_h$) and (4.6), we obtain the existence of a constant $C \equiv C(\beta, c, \gamma)$ such that

$$\|\delta p\|_{L^2} \leq C(\|w_2\|_{L^2} + \|w_1\|_{L^2}).$$

This implies the claim. □

Thus, system (4.3) is semi-smooth, and from standard results (e.g., [30],[24, Th. 8.5]) we deduce the following convergence result for the semi-smooth Newton method.

Theorem 4.4. *If condition (3.7) holds, the Newton iteration (4.3) converges superlinearly to the solution $(\overline{u}_h, \overline{p}_h)$ of (OS$_h$) for every $c, \beta > 0$, provided that (u_h^0, p_h^0) is sufficiently close to $(\overline{u}_h, \overline{p}_h)$.*

Remark 4.5. A semi-smooth Newton method in function space would directly imply mesh independence of the superlinear convergence (cf. [22]), which is desirable in practice. Although the optimality system (OS$_\beta$) possesses the necessary norm gap to have Newton differentiability of the pointwise max and min in infinite dimensions, the system seems not to be semi-smooth. This would require uniform invertibility of the Newton derivative with respect to p, which is hindered by the fact that the regularization term $\|p\|_{L^2}^2$ is not sufficient to ensure surjectivity from $L^\infty(\Omega)$ to $L^2(\Omega)$. On the other hand, a stronger penalty (leading to corresponding duality mapping in the Newton system) is computationally inconvenient. Despite of this, mesh-independence of the semi-smooth method is still observed for the model problem, see Section 5.

5. Numerical examples

We now present some numerical results for the model problem (1.1) for $d = 1, 2$ to illustrate salient features of the proposed approach. Since the model problem is only well-posed for $u(x) \geq c > 0$ for almost all $x \in \Omega$, we have selected the true solution and the initial guess to be in the interior of the domain of definition $D(S)$. In all computations, we observed that the iterates stay well within $D(S)$.

We have implemented the semi-smooth Newton method as given in Appendix B (Algorithm 1). The iteration was terminated if the active sets did not change and the norm of the gradient was smaller than 10^{-6}, or if 20 iterations were reached. The free parameter c in the complementarity function (4.2) was set to $c = 10^3$; in all our tests, the results depended only very weakly on its value.

Following [11], the continuation strategy for β was implemented as follows. We set $\beta_0 = 1$ and $\beta_{n+1} = \beta_n/2$ until $\beta_n < 10^{-12}$ holds or the semi-smooth Newton method did not converge in the given number of iterations. In the latter case, we return the last feasible saddle point $(u_{\beta_{n-1}}, p_{\beta_{n-1}})$. The whole procedure will hereafter be denoted by *path-following semi-smooth Newton method*. The reduced semi-smooth Newton system (B.7) is solved using BiCGstab (with tolerance 10^{-6} and maximum number of iterations 1000). For the automatic parameter choice by the balancing principle, we have set $\sigma = 1.03$.

All computations were performed with MATLAB (R2010a) on a four-core 2.4 GHz workstation with 4 GByte of RAM. MATLAB codes with our implementation can be downloaded from `http://www.uni-graz.at/~clason/codes/l1fitnonlin.zip`.

5.1. One-dimensional example

In our first example, we take $\Omega = [-1, 1]$, $f(x) = 1$ and

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

Unless otherwise stated, we discretize Ω into $N = 1000$ elements of uniform size, and set $u_h^\dagger = \Pi_h^0 u^\dagger$. We generate the noisy data y_h^δ by corrupting the projection of the exact data onto U_h. Specifically, we set element-wise

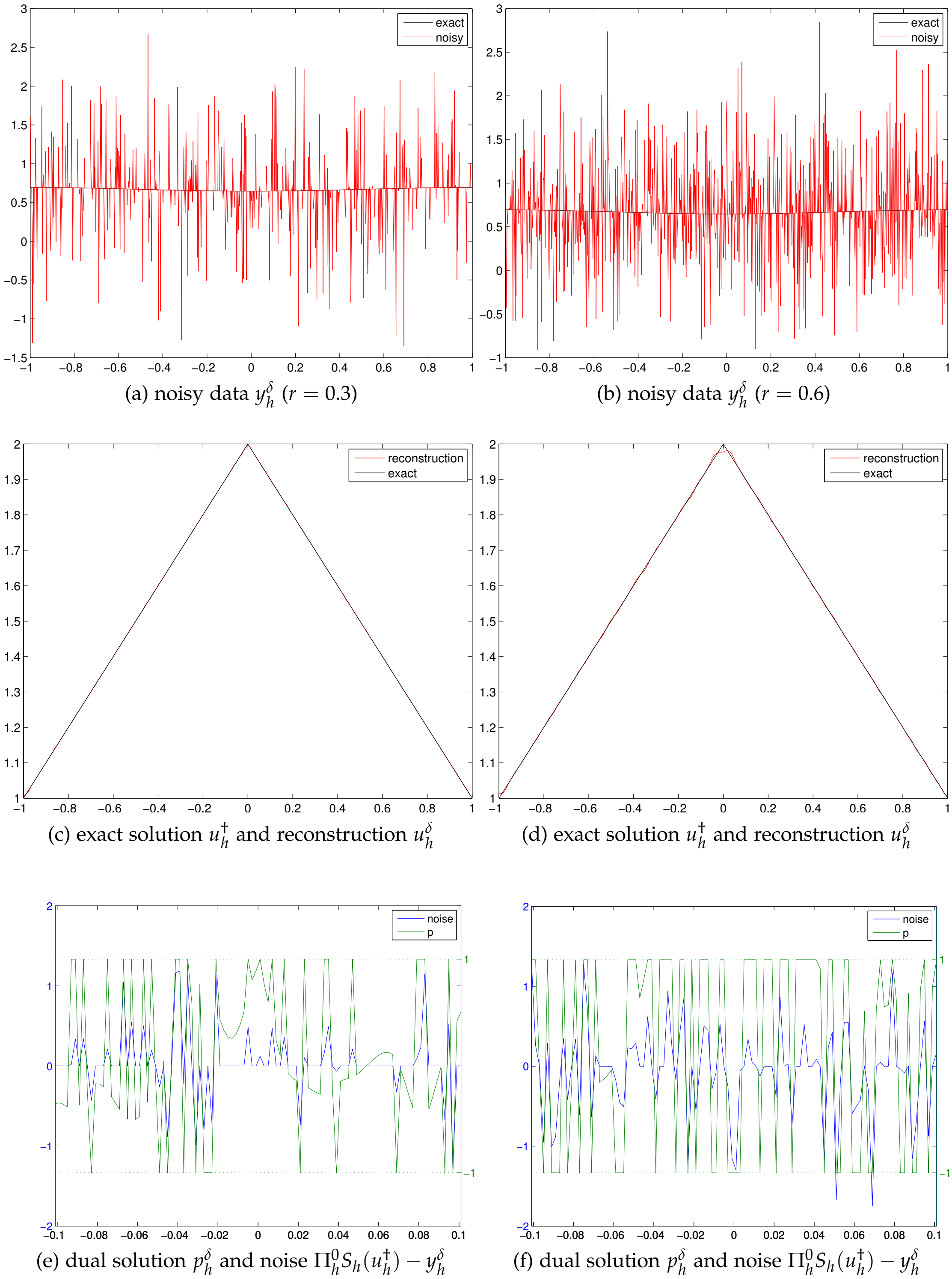
$$y_h^\delta = \begin{cases} y_h^\dagger + \|y_h^\dagger\|_{L^\infty}\xi, & \text{with probability } r, \\ y_h^\dagger, & \text{otherwise,} \end{cases}$$

where ξ follows the standard normal distribution and $y_h^\dagger = \Pi_h^0 S_h(u_h^\dagger)$. The noise level δ is defined by $\delta = \|y_h^\delta - y_h^\dagger\|_{L^1}$. A typical realization of noisy data is displayed in Figure 1a for $r = 0.3$ ($\delta = 1.859 \times 10^{-1}$) and Fig. 1b for $r = 0.6$ ($\delta = 3.203 \times 10^{-1}$). The corresponding reconstructions u_h^δ with α chosen according to the balancing principle are shown in Figures 1c and 1d, respectively. In both cases, the fixed point iteration (2.2) converged after 4 iterations, yielding $\alpha = 4.781 \times 10^{-3}$ for $r = 0.3$ and $\alpha = 8.236 \times 10^{-3}$ for $r = 0.6$. The L^2-reconstruction error $e = \|\bar{u}^h - u_h^\dagger\|_{L^2}$ is 6.166×10^{-4} for $r = 0.3$ and 2.799×10^{-3} for $r = 0.6$.

In Figures 1e and 1f, we show an enlarged sample of the dual solution p_h^δ and the data noise $\Pi_h^0 S_h(u_h^\dagger) - y_h^\delta$ for both noise cases. Observe that p_h^δ indeed serves as a good indicator of noise, as both locations and signs of nonzero noise components are accurately detected. This numerically corroborates Corollary 3.2.

The behavior of the balancing principle is further illustrated in Table 1, where we compare the balancing parameter α_b for different noise levels with the "optimal" parameter, which is obtained by sampling the interval $[0.1\alpha_b, 10\alpha_b]$ uniformly with 20 parameters and taking as α_{opt} the one with smallest reconstruction error e_{opt}. We observe that both the regularization parameters and corresponding reconstruction errors by the two approaches are of comparable size. For completeness, we also show the mean and standard deviation of balancing parameters and reconstruction errors obtained for ten realizations of the same noise level. The standard deviation shows the stability of the balancing principle with respect to noise realization.

To illustrate the convergence of the semi-smooth Newton method, we fix $r = 0.3$, $\alpha = 4 \times 10^{-3}$ and $\beta = 10^{-2}$, and solve the optimality system (OS_h) using Algorithm 1. Table 2 shows the number of elements that changed between active and inactive



(a) noisy data y_h^δ ($r = 0.3$)
(b) noisy data y_h^δ ($r = 0.6$)
(c) exact solution $u_h^\dagger$ and reconstruction u_h^δ
(d) exact solution $u_h^\dagger$ and reconstruction u_h^δ
(e) dual solution p_h^δ and noise $\Pi_h^0 S_h(u_h^\dagger) - y_h^\delta$
(f) dual solution p_h^δ and noise $\Pi_h^0 S_h(u_h^\dagger) - y_h^\delta$

Figure 1: Results for different noise levels (left: $r = 0.3$, right: $r = 0.6$).

Table 1: Numerical results for different noise percentages r. Shown are noise level δ, optimal regularization parameter α_{opt}, balancing parameter α_{b} and its average $\overline{\alpha}_{\text{b}}$, and the corresponding reconstruction errors e_{opt}, e_{b} and $\overline{e}_{\text{b}}$.

r	δ	α_{opt}	α_{b}	$\overline{\alpha}_{\text{b}}$	e_{opt}	e_{b}	$\overline{e}_{\text{b}}$
0.1	4.678e-2	3.254e-3	1.203e-3	1.379e-3±2.28e-4	1.531e-4	2.585e-4	3.503e-4±7.94e-5
0.2	1.117e-1	3.281e-3	2.872e-3	2.863e-3±3.07e-4	5.479e-4	5.351e-4	4.069e-4±1.27e-4
0.3	1.860e-1	2.970e-3	4.783e-3	4.186e-3±1.85e-4	5.775e-4	6.716e-4	7.933e-4±3.84e-4
0.4	2.046e-1	6.008e-3	5.260e-3	5.752e-3±2.97e-4	7.248e-4	1.067e-3	1.599e-3±4.43e-4
0.5	2.889e-1	4.613e-3	7.428e-3	7.137e-3±2.34e-4	2.135e-3	2.635e-3	1.842e-3±7.43e-4
0.6	3.361e-1	5.368e-3	8.644e-3	8.541e-3±3.70e-4	2.457e-3	2.926e-3	3.607e-3±1.24e-3
0.7	3.887e-1	6.209e-3	9.997e-3	1.003e-2±3.83e-4	6.537e-3	8.930e-3	9.477e-3±6.31e-3
0.8	4.446e-1	7.106e-3	1.144e-2	1.132e-2±2.92e-4	1.842e-2	2.391e-2	1.959e-2±1.53e-2
0.9	5.034e-1	8.060e-3	1.298e-2	1.287e-2±7.51e-4	2.414e-2	3.766e-2	8.273e-2±4.84e-2

Table 2: Convergence behavior of the semi-smooth Newton method (for fixed α, β). Shown are the number $n(k)$ of elements that changed between active and inactive sets and norm of the residual $r(k) := \|T_h(u_h^k, p_h^k)\|$ after each iteration k.

k	1	2	3	4	5	6	7
$n(k)$	808	245	262	217	46	0	0
$r(k)$	3.503e+4	3.426e+1	3.988e+2	2.176e+2	3.508e-3	5.618e-5	8.465e-9

sets as well as the norm of the residual $\|T_h(u_h^k, p_h^k)\|_{L^2}$ after iteration k. The locally superlinear convergence can be observed very clearly.

We next address the overall performance of the proposed approach. Table 3 shows the time for one path-following semi-smooth Newton method (with the final α, since it depends on this value) and for the full balancing principle iteration for different numbers of elements N. For completeness, we also show the final reconstruction error. Again, we give mean and standard deviation over ten different noise realizations (with $r = 0.3$). As can be seen, both the balancing principle iteration and the path-following semi-smooth Newton method scale very well with the number of elements, indicating that we have the desired mesh independence of the Newton method. This is further supported by the fact that at least for all $\beta > 10^{-2}$, the required number of semi-smooth Newton iterations was either 4 or 5 for all values of N. We also point out that the computational cost of calculating the balancing parameter is only between two and three times the cost of solving the L^1 fitting problem for a fixed (reasonable) parameter.

5.2. Two-dimensional example

To illustrate the applicability of the proposed approach to higher-dimensional parameter identification problems, we also consider the model problem in two dimensions.

Table 3: Computing times (in seconds) for semi-smooth Newton method (t_s) and balancing principle iteration (t_b) and reconstruction error e for different problem sizes N. Shown are the mean and standard deviation over ten noise realizations.

N	100	200	400	800	1600	3200	6400	12800
$t_{s,m}$	10.06	13.60	18.17	31.41	52.83	118.71	190.52	397.41
$t_{s,s}$	4.64	5.25	8.54	9.27	19.65	27.71	119.26	169.41
$t_{b,m}$	25.88	35.07	38.34	70.24	113.37	247.45	450.89	969.85
$t_{b,s}$	5.09	5.58	9.96	13.50	32.64	42.53	87.83	207.85
e_m	2.83e-2	5.91e-3	2.18e-3	1.07e-3	5.53e-4	3.41e-4	3.24e-4	3.23e-4
e_s	1.65e-2	1.55e-3	7.01e-4	4.43e-4	2.12e-4	1.20e-4	1.20e-4	8.17e-5

Specifically, we discretize the domain $[-1,1]^2$ by the standard uniform triangulation, i.e, we create a tensor product mesh of m equally spaced grid points in each direction and split each grid square along the diagonal into two congruent triangles. The choice $m = 128$ yields $N = 2(128-1)^2 = 32258$ elements. Here, the exact solution, shown in Figure 2c, is given by

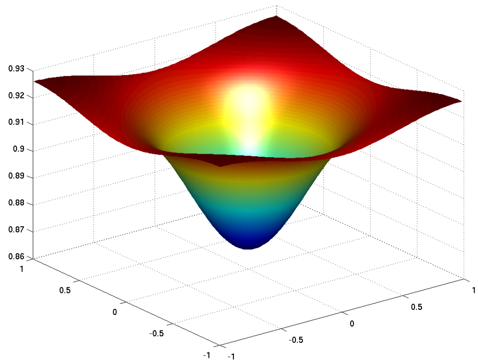
$$u^\dagger(x_1,x_2) = \begin{cases} 1+\cos(x_1\pi)\cos(x_2\pi), & \text{if } |x_1| \leq \frac{1}{2}, |x_2| \leq \frac{1}{2}, \\ 1, & \text{otherwise.} \end{cases}$$

The right hand side is again $f(x_1,x_2) = 1$. The exact and noisy data with $r = 0.3$ ($\delta = 2.214 \times 10^{-1}$) are given in Figure 2a and Figure 2b, respectively. The corresponding reconstruction, shown in Figure 2d, is an excellent approximation of the true solution, with the reconstruction error e being 3.502×10^{-3}. In particular, the magnitude of the peak is correctly recovered, and the homogeneous background is relatively artifact-free. Again, the fixed point algorithm converged in four iterations to a value $\alpha_b = 1.050 \times 10^{-2}$.

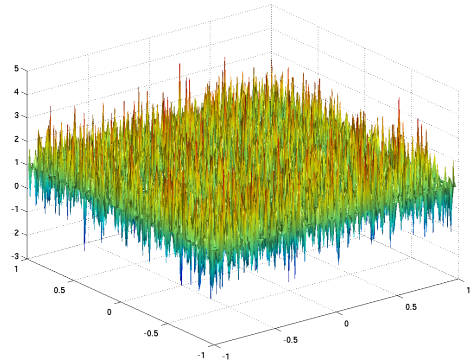
6. Conclusion

In this paper we have studied nonlinear L^1 data fitting problems and presented a path-following semi-smooth Newton method for their numerical solution based on a finite element approach. In particular, we have shown regularizing properties, convergence rates for a priori and a posteriori parameter choice rules and the exact recovery property for nonzero α of the L^1 data fitting model. The latter is in direct contrast to the L^2 data fitting, where the Tikhonov minimizer is different from the true solution for every $\alpha > 0$.

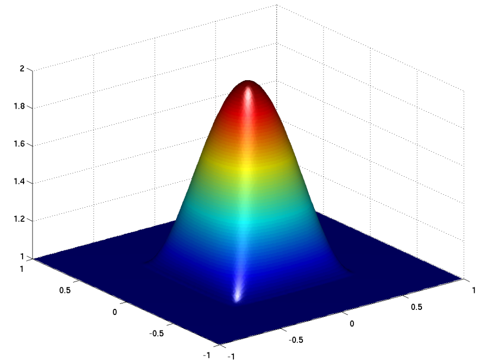
The semi-smooth Newton is based on a regularized primal-dual saddle-point formulation of the L^1 fitting functional, for which superlinear convergence in the discrete setting was shown and numerically observed. Furthermore, we have also observed



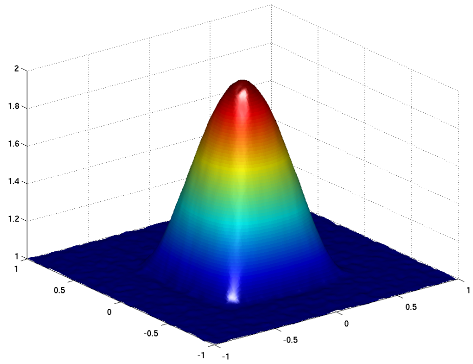
(a) exact data $y_h^\dagger$



(b) noisy data y_h^δ



(c) exact solution $u_h^\dagger$



(d) reconstruction u_h^δ

Figure 2: Results for the two-dimensional example

mesh independence of the method, which suggests that it might be possible to show semi-smoothness also for the infinite-dimensional optimality system. As the model examples showed, this allows the efficient numerical solution of the L^1 parameter identification problem.

The balancing principle proved to be an effective parameter choice method, which required little a priori information (such as the noise level or a reasonable range for sampling), while adding only a small computational overhead over the solution of a single minimization problem.

The presented approach can be extended in several directions. As noted in Remark 1.1, 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. Also of interest would be considering parameter identification problems where the parameter u is not in $L^2(\Omega)$, such as identifying the principal coefficient in an elliptic equation (which would require $u \in \mathrm{H}^1(\Omega)$ or BV) or coefficients in boundary conditions.

A. Verification of assumptions for model problem

For completeness, we collect in this section some results which verify the continuity and differentiability assumptions (A1)–(A4) for our model problem (1.1), i.e., where the solution operator $S : \mathrm{L}^2(\Omega) \to \mathrm{H}^1(\Omega)$ maps

$$u \in D(S) = \left\{u \in \mathrm{L}^2(\Omega) : u(x) \geq c > 0 \text{ a.e. in } \Omega\right\}$$

to the solution y of

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

for a given $f \in \mathrm{L}^2(\Omega)$. As stated in Remark 1.1, for the sake of presentation we do not explicitly enforce the constraint $u \in D(S)$ in the derivations above, assuming that all sequences are contained in a neighborhood of a solution in the interior of $D(S)$.

Since for all $u \in D(S)$, the bilinear form in (A.1) defines an inner product on $\mathrm{H}^1(\Omega)$, we obtain the well-posedness of the state equation from standard arguments.

Lemma A.1. *For every $u \in D(S)$, the problem (A.1) has a unique solution $y \in \mathrm{H}^1(\Omega)$ which satisfies the a priori estimate*

$$\|y\|_{\mathrm{H}^1} \leq C. \tag{A.2}$$

From this, the uniform boundedness of S follows since $f \in \mathrm{L}^2(\Omega)$ is fixed. We next address complete continuity of the solution operator S.

Lemma A.2. *Let $\{u_n\} \subset D(S)$ be a sequence converging weakly in $\mathrm{L}^2(\Omega)$ to $u^* \in D(S)$, then*

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

Proof. For $u_n \in D(S)$, set $y_n = S(u_n) \in \mathrm{H}^1(\Omega)$. By the a priori estimate (A.2), the sequence $\{y_n\}$ is uniformly bounded in H^1 and has a convergent subsequence also denoted by $\{y_n\}$, such that there exists $y^* \in \mathrm{H}^1(\Omega)$ with

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

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

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

for any $p < +\infty$ if $d = 2$ and for $p < 6$ if $d = 3$. In particular, we will take $p = 4$. Then we have

$$\lim_{n\to\infty} \langle \nabla y_n, \nabla v \rangle_{\mathrm{L}^2} = \langle \nabla y^*, \nabla v \rangle_{\mathrm{L}^2}$$

and

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

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

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

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

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

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

The above two statements imply that Assumption (A1) holds. The next statement follows from direct calculation and yields Assumption (A2):

Lemma A.3. *The operator $S : D(S) \to H^1(\Omega)$ is twice continuously Fréchet differentiable. Furthermore, for every $u \in D(S)$, we have the following characterizations of the derivatives acting on $h_1, h_2 \in L^2(\Omega)$:*

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

$$\langle \nabla z, \nabla v\rangle_{L^2} + \langle uz, v\rangle_{L^2} = -\langle h_1 S(u), v\rangle_{L^2} \qquad \textit{for all } v \in H^1(\Omega).$$

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

$$\langle \nabla w, \nabla v\rangle_{L^2} + \langle uw, v\rangle_{L^2} = -\left\langle h_1 S'(u)h_2 + h_2 S'(u)h_1, v\right\rangle_{L^2} \qquad \textit{for all } v \in H^1(\Omega). \tag{A.3}$$

Similarly to the case of the state equation (A.1), we can derive the following a priori estimates, which verify Assumptions (A3) and (A4).

Lemma A.4. *There exists a constant $C > 0$, independent of u, such that for all $u \in D(S)$ and $h_1, h_2 \in L^2(\Omega)$, the following estimates hold:*

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

Proof. The first estimate follows by direct calculation. Testing now (A.3) with $w = S''(u)(h_1, h_2)$, we deduce by Hölder's inequality and the Sobolev embedding theorem [1, Th. 4.12]

$$\begin{aligned}
\|S''(u)(h_1,h_2)\|_{H^1}^2 &\le C\Big(\|h_1\|_{L^2}\|S'(u)h_2\|_{L^4}\|S''(u)(h_1,h_2)\|_{L^4} \\
&\qquad + \|h_2\|_{L^2}\|S'(u)h_1\|_{L^4}\|S''(u)(h_1,h_2)\|_{L^4}\Big) \\
&\le C\Big(\|h_1\|_{L^2}\|S'(u)h_2\|_{H^1}\|S''(u)(h_1,h_2)\|_{H^1} \\
&\qquad + \|h_2\|_{L^2}\|S'(u)h_1\|_{H^1}\|S''(u)(h_1,h_2)\|_{H^1}\Big) \\
&\le C\|h_1\|_{L^2}\|h_2\|_{L^2}\|S''(u)(h_1,h_2)\|_{H^1},
\end{aligned}$$

where the last inequality follows from the first estimate. The second estimate now follows directly. □

The last lemma concerns the finite element approximation $S_h(u_h) : U_h \to V_h$, which for $u_h \in D(S_h) \subset U_h$ is defined as the solution $y_h \in V_h$ of

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

The discrete analogues of Lemmas A.1, A.3 and A.4 can be proved analogously to their continuous counterparts, cf. [28]. In particular, we note that the constant C in Lemma A.4 depends only on the Sobolev embedding constant and are thus independent of h. We only show that Assumption (A1$_h$) holds, and refer the explicit form of the discrete derivatives to Appendix B.

Lemma A.5. *Assume that the sequence* $\{u_h\} \subset D(S_h)$ *satisfies* $u_h \rightharpoonup u$ *in* $L^2(\Omega)$ *as* $h \to 0$. *Then* $S_h(u_h) \to S(u)$ *in* $L^2(\Omega)$ *as* $h \to 0$.

Proof. First we recall (e.g., from [8, Th. 3.2.3]) the $H^1(\Omega)$ projection operator Π_h^1 onto V_h which satisfies for any $v \in H^1(\Omega)$,

$$\lim_{h \to 0} \|\Pi_h^1 v - v\|_{H^1} = 0.$$

By the discrete analogue of Lemma A.1, we have that the sequence $\{S_h(u_h)\}$ is uniformly bounded in $H^1(\Omega)$. Therefore, there exists a subsequence, also denoted by $S_h(u_h)$, which converges weakly to some y. We next show that $y = S(u)$. To this end, for any $v \in H^1(\Omega)$, we take $\Pi_h^1 v \in V_h$ as the test function in the weak form to get

$$\langle \nabla y_h, \nabla \Pi_h^1 v \rangle_{L^2} + \langle u_h y_h, \Pi_h^1 v \rangle_{L^2} = \langle f, \Pi_h^1 v \rangle_{L^2}.$$

By the property of Π_h^1, we have

$$\langle \nabla y_h, \nabla \Pi_h^1 v \rangle_{L^2} \to \langle \nabla y, \nabla v \rangle_{L^2}, \quad \langle f, \Pi_h^1 v \rangle_{L^2} \to \langle f, v \rangle_{L^2} \quad \text{as } h \to 0.$$

Next noting the compact embedding of $H^1(\Omega)$ into $L^4(\Omega)$ in $\mathbb{R}^d$ ($d = 1, 2, 3$), we have $y_h \to y$ in L^4, and $\Pi_h^1 v \to v$ in $L^4(\Omega)$. This together with the property of the projection operator Π_h^1 gives

$$\langle u_h y_h, \Pi_h^1 v \rangle_{L^2} \to \langle u y, v \rangle_{L^2} \quad \text{as } h \to 0.$$

Consequently, the limit y satisfies

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

i.e. $y = S(u)$. Since every subsequence has a subsequence converging to y weakly in $H^1(\Omega)$, the whole sequence $\{S_h(u_h)\}$ converges weakly. The claim then follows again from the compact embedding of $H^1(\Omega)$ in $L^2(\Omega)$. □

B. Computation of Newton system for model problem

Here we compute the explicit form of the Newton system for our model problem using the Lagrangian formalism. We again consider the discrete saddle point problem

$$(\mathcal{P}_h) \qquad \min_{u_h\in U_h} \max_{\|p_h\|_{L^\infty}\le 1} \langle S_h(u_h) - y^\delta, p_h\rangle_{L^2} + \frac{\alpha}{2}\|u_h\|_{L^2}^2 - \frac{\beta}{2}\|p_h\|_{L^2}^2 .$$

with $(u_h, p_h) \in U_h \times U_h$ and $S_h : U_h \to V_h$. For the sake of readability, we omit the subscript h on the variables from now on and denote the inner product on $L^2(\Omega)$ by $\langle u, v\rangle$. We first introduce the state variable $y := S_h(u) \in V_h$ and the Lagrange multiplier $\lambda \in V_h$ corresponding to the equality constraint $S_h(u) = y$ in weak form, and consider the saddle point problem

$$\min_{u\in U_h, y\in V_h} \max_{\substack{p\in U_h,\lambda\in V_h\\ \|p\|_{L^\infty}\le 1}} L(u, y, p, \lambda)$$

for the Lagrangian

$$L(u, y, p, \lambda) = \langle y - y^\delta, p\rangle + \frac{\alpha}{2}\|u\|_{L^2}^2 - \frac{\beta}{2}\|p\|_{L^2}^2 + \langle \nabla\lambda, \nabla y\rangle + \langle \lambda, uy - f\rangle .$$

To keep the notation concise, we introduce $X_h := U_h \times V_h \times U_h \times V_h$, $x := (u, y, p, \lambda)^T$ and similarly $\delta x_1, \delta x_2$. For a given $x \in X_h$, a Newton step then consists in computing $\delta x_2 \in X_h$ satisfying

$$\left\langle \delta x_1, \nabla^2 L(x)\delta x_2\right\rangle = -\langle \nabla L(x), \delta x_1\rangle \qquad \text{for all } \delta x_1 \in X_h.$$

Here, $\nabla L = (L_u, L_y, L_p, L_\lambda)^T$ denotes the gradient of L, and

$$\nabla^2 L = \begin{pmatrix} L_{uu} & L_{uy} & L_{up} & L_{u\lambda}\\ L_{yu} & L_{yy} & L_{yp} & L_{y\lambda}\\ L_{pu} & L_{py} & L_{pp} & L_{p\lambda}\\ L_{\lambda u} & L_{\lambda y} & L_{\lambda p} & L_{\lambda\lambda}\end{pmatrix}$$

denotes the Hessian of L (with the second partial (Newton) derivatives L_{uu} etc.).

We first compute the components of the gradient $\langle \nabla L(x), \delta x_1\rangle$.

$$\begin{aligned}
\langle L_u(x), \delta u_1\rangle &= \alpha\langle u, \delta u_1\rangle + \langle \lambda y, \delta u_1\rangle ,\\
\left\langle L_y(x), \delta y_1\right\rangle &= \langle p, \delta y_1\rangle + \langle \nabla\lambda, \nabla\delta y_1\rangle + \langle u\lambda, \delta y_1\rangle ,\\
\langle L_\lambda(x), \delta\lambda_1\rangle &= \langle \nabla y, \nabla\delta\lambda_1\rangle + \langle uy, \delta\lambda_1\rangle - \langle f, \delta\lambda_1\rangle .
\end{aligned}$$

The component $\left\langle L_p(x), \delta p_1\right\rangle$ for $\delta p_1 \in U_h$ can again be interpreted element-wise after introducing the orthogonal projection Π_h^0 onto U_h and then expressed for any $c > 0$ as

$$\left\langle L_p(x), \delta p_1\right\rangle = \langle \chi_{\mathcal{I}}(\Pi_h^0(y - y^\delta) - \beta p) - c(\chi_{\mathcal{A}_+}(p-1) + \chi_{\mathcal{A}_-}(p+1)), \delta p_1\rangle,$$

where $\chi_{\mathcal{A}_+}$, $\chi_{\mathcal{A}_-}$ and $\chi_{\mathcal{I}}$ again denote indicator functions of the active and inactive sets.

Similarly, the non-zero components of $\langle \delta x_1, \nabla^2 L(x)\delta x_2\rangle$ are given by

$$\begin{aligned}
\langle \delta u_1, L_{uu}(x)\delta u_2\rangle &= \alpha \langle \delta u_1, \delta u_2\rangle ,\\
\langle \delta u_1, L_{uy}(x)\delta y_2\rangle &= \langle \delta u_1, \lambda\delta y_2\rangle ,\\
\langle \delta u_1, L_{u\lambda}(x)\delta \lambda_2\rangle &= \langle \delta u_1, y\delta \lambda_2\rangle ,\\
\langle \delta y_1, L_{yu}(x)\delta u_2\rangle &= \langle \delta y_1, \lambda\delta u_2\rangle ,\\
\langle \delta y_1, L_{yp}(x)\delta p_2\rangle &= \langle \delta y_1, \delta p_2\rangle ,\\
\langle \delta y_1, L_{y\lambda}(x)\delta \lambda_2\rangle &= \langle \nabla\delta y_1, \nabla\delta \lambda_2\rangle + \langle \delta y_1, u\delta\lambda_2\rangle ,\\
\langle \delta p_1, L_{py}(x)\delta y_2\rangle &= \langle \delta p_1, \chi_{\mathcal{I}}\Pi_h^0\delta y_2\rangle ,\\
\langle \delta p_1, L_{pp}(x)\delta p_2\rangle &= \langle \delta p_1, (\beta\chi_{\mathcal{I}} + c\chi_{\mathcal{A}})\delta p_2\rangle ,\\
\langle \delta \lambda_1, L_{\lambda u}(x)\delta u_2\rangle &= \langle \delta \lambda_1, y\delta u_2\rangle ,\\
\langle \delta \lambda_1, L_{\lambda y}(x)\delta y_2\rangle &= \langle \nabla\delta \lambda_1, \nabla\delta y_2\rangle + \langle \delta\lambda_1, u\delta y_2\rangle .
\end{aligned}$$

Writing now $\delta x \equiv \delta x_2$ for the update and $\tilde{x} \equiv \delta x_1$ for the first variation, we can write the Newton system for $\delta x \in X_h$ as

$$\text{(B.1)}\qquad \begin{cases}
\langle \alpha\delta u + \lambda\delta y + y\delta\lambda, \tilde{u}\rangle = -\langle L_u(x), \tilde{u}\rangle ,\\
\langle \lambda\delta u + \delta p, \tilde{y}\rangle + \langle \nabla\delta\lambda, \nabla\tilde{y}\rangle + \langle u\delta\lambda, \tilde{y}\rangle = -\langle L_y(x), \tilde{y}\rangle ,\\
\langle \chi_{\mathcal{I}}\Pi_h^0\delta y - (\beta\chi_{\mathcal{I}} + c\chi_{\mathcal{A}})\delta p, \tilde{p}\rangle = -\langle L_p(x), \tilde{p}\rangle ,\\
\langle y\delta u, \tilde{\lambda}\rangle + \langle \nabla\delta y, \nabla\tilde{\lambda}\rangle + \langle u\delta y, \tilde{\lambda}\rangle = -\langle L_\lambda(x), \tilde{\lambda}\rangle .
\end{cases}$$

for all $\tilde{x} \in X_h$.

We next compute the reduced system for $(\delta u, \delta p)$. Solving the state equation for given u, the equality constraint is satisfied for $(y(u), u)$ for every λ, and we can choose the Lagrange multiplier such that $L_y(x)$ vanishes. Hence, for given u and p, we compute $y, \lambda \in V_h$ satisfying

$$\text{(B.2)}\qquad \langle \nabla y, \nabla v\rangle + \langle uy, v\rangle = \langle f, v\rangle \qquad \text{for all } v \in V_h,$$

$$\text{(B.3)}\qquad \langle \nabla \lambda, \nabla v\rangle + \langle u\lambda, v\rangle = -\langle p, v\rangle \qquad \text{for all } v \in V_h.$$

Similarly, given $\delta u, \delta p$, we calculate $\delta y, \delta\lambda \in V_h$ from the fourth and second equation of (B.1), respectively, such that

$$\text{(B.4)}\qquad \langle \nabla \delta y, \nabla v\rangle + \langle u\delta y, v\rangle = -\langle y\delta u, v\rangle \qquad \text{for all } v \in V_h,$$

$$\text{(B.5)}\qquad \langle \nabla \delta\lambda, \nabla v\rangle + \langle u\delta\lambda, v\rangle = -\langle \lambda\delta u + \delta p, v\rangle \qquad \text{for all } v \in V_h.$$

Since the test functions $\tilde{u}, \tilde{p}$ are piecewise constant, we can use the projection operator Π_h^0 to write the reduced Newton system element-wise as a set of algebraic equations for the values of δu and δp on each element:

$$\text{(B.6)}\qquad \nabla^2\hat{L}(u,p)\begin{pmatrix}\delta u\\ \delta p\end{pmatrix} = -\begin{pmatrix}L_u(u,p)\\ L_p(u,p)\end{pmatrix},$$

Algorithm 1 Semi-smooth Newton method for the model problem.

Choose $(u^0, p^0) \in U_h$, $\beta > 0$, $c > 0$, set $k = 0$.
repeat
  Compute $y^k, \lambda^k \in V_h$ solving (B.2) and (B.3).
  Compute the active and inactive sets

$$\begin{aligned}
\mathcal{A}_+^k &:= \left\{ K \in \mathcal{T}_h : \left(\Pi_h^0 (y^k - y^\delta) - \beta p^k + c(p^k - 1) \right) |_K > 0 \right\}, \\
\mathcal{A}_-^k &:= \left\{ K \in \mathcal{T}_h : \left(\Pi_h^0 (y^k - y^\delta) - \beta p^k + c(p^k + 1) \right) |_K < 0 \right\}, \\
\mathcal{I}^k &:= \mathcal{T}_h \setminus (\mathcal{A}_-^k \cup \mathcal{A}_+^k).
\end{aligned}$$

  Compute reduced gradient

$$\nabla \hat{L}(u^k, p^k) = \begin{pmatrix} \alpha u^k + \Pi_h^0 (y^k \lambda^k), \\ \chi_{\mathcal{I}^k} (\Pi_h^0 (y^k - y^\delta) - \beta p^k) + c(\chi_{\mathcal{A}_+^k} (p^k - 1) + \chi_{\mathcal{A}_-^k} (p^k + 1)) \end{pmatrix}.$$

  Compute update $\delta u, \delta p \in U_h$ by solving (B.6) using a Krylov method: For given $\delta u, \delta p$, compute δy and $\delta \lambda$ solving (B.4) and (B.5) and set

$$\nabla^2 \hat{L}(u, p)(\delta u, \delta p) = \begin{pmatrix} \alpha \delta u + \Pi_h^0 (\lambda^k \delta y + y^k \delta \lambda) \\ \chi_{\mathcal{I}} \Pi_h^0 \delta y - (\beta \chi_{\mathcal{I}^k} + c \chi_{\mathcal{A}^k}) \delta p \end{pmatrix}.$$

  Set $u^{k+1} = u^k + \delta u$, $p^{k+1} = p^k + \delta p$, $k \leftarrow k + 1$.
until $\mathcal{A}^k = \mathcal{A}^{k-1}$ and $\|\nabla \hat{L}(u^k, p^k)\| \leq \text{tol}$.

where the action of the reduced Hessian $\nabla^2 \hat{L}(u, p)$ on $\delta u, \delta p \in V_h$ is given by

$$\nabla^2 \hat{L}(u, p) \begin{pmatrix} \delta u \\ \delta p \end{pmatrix} := \begin{pmatrix} \alpha \delta u + \Pi_h^0 (\lambda \delta y + y \delta \lambda) \\ \chi_{\mathcal{I}} \Pi_h^0 \delta y - (\beta \chi_{\mathcal{I}} + c \chi_{\mathcal{A}}) \delta p \end{pmatrix}, \tag{B.7}$$

with $\delta y, \delta \lambda$ computed from δu and δp as described above. This can be applied for the solution of the reduced Hessian using a Krylov method. The complete semi-smooth Newton method for the model problem is given in Algorithm 1.

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 the work of the second author was supported by the Alexander von Humboldt Foundation through a postdoctoral research fellowship.

References

[1] R. A. Adams and J. J. F. Fournier, *Sobolev spaces*, vol. 140 of Pure and Applied Mathematics (Amsterdam), Elsevier/Academic Press, Amsterdam, second 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] A. C. Bovik, *Handbook of Image and Video Processing (Communications, Networking and Multimedia)*, Academic Press, Inc., Orlando, 2005.

[5] 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.

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

[7] 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.

[8] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*, North-Holland, Amsterdam, 1978.

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

[10] 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. Scientific Computing, 32 (2010), pp. 1484–1505.

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

[12] 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.

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

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

[15] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, vol. 375 of Mathematics and its Applications, Kluwer Academic Publishers Group, Dordrecht, 1996.

[16] H. W. Engl, K. Kunisch, and A. Neubauer, *Convergence rates for Tikhonov regularisation of nonlinear ill-posed problems*, Inverse Problems, 5 (1989), pp. 523–540.

[17] A. Ern and J.-L. Guermond, *Theory and Practice of Finite Elements*, vol. 159 of Applied Mathematical Sciences, Springer-Verlag, New York, 2004.

[18] L. C. Evans, *Weak Convergence Methods for Nonlinear Partial Differential Equations*, American Mathematical Society, Providence, 1990.

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

[20] T. Hein and B. Hofmann, *Approximate source conditions for nonlinear ill-posed problems – chances and limitations*, Inverse Problems, 25 (2009), p. 035003.

[21] 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).

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

[23] P. J. Huber, *Robust Statistics*, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1981.

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

[25] B. Jin and J. Zou, *Numerical estimation of the Robin coefficient in a stationary diffusion equation*, IMA J. Numer. Anal., (2009), p. drn066.

[26] ——, *Iterative parameter choice by discrepancy principle*, Tech. Rep. 2010-02(369), Department of Mathematics, Chinese University of Hong Kong, 2010.

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

[28] 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.

[29] V. A. Morozov, *On the solution of functional equations by the method of regularization*, Soviet Math. Dokl., 7 (1966), pp. 414–417.

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

[31] A. Rösch and D. Wachsmuth, *Numerical verification of optimality conditions*, SIAM J. Control Optim., 47 (2008), pp. 2557–2581.

[32] F. Tröltzsch, *Optimal Control of Partial Differential Equations: Theory, Methods and Applications*, American Mathematical Society, Providence, 2010. Translated from the German by Jürgen Sprekels.

[33] 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.

[34] 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.

[35] E. Zeidler, *Nonlinear Functional Analysis and its Applications. III. Variational Methods and Optimization*, Springer-Verlag, New York, 1985. Translated from the German by Leo F. Boron.

[36] ——, *Nonlinear Functional Analysis and its Applications. II/B. Nonlinear Monotone Operators*, Springer-Verlag, New York, 1990. Translated from the German by the author and Leo F. Boron.