SIAM J. IMAGING SCIENCES
Vol. 3, No. 2, pp. 199–231

A Semismooth Newton Method for L^1 Data Fitting with Automatic Choice of Regularization Parameters and Noise Calibration*

正则化参数

Christian Clason[†], Bangti Jin[‡], and Karl Kunisch[†]

Abstract. This paper considers the numerical solution of inverse problems with an L^1 data fitting term, which is challenging due to the lack of differentiability of the objective functional. Utilizing convex duality, the problem is reformulated as minimizing a smooth functional with pointwise constraints, which can be efficiently solved using a semismooth Newton method. In order to achieve superlinear convergence, the dual problem requires additional regularization. For both the primal and the dual problems, the choice of the regularization parameters is crucial. We propose adaptive strategies for choosing these parameters. The regularization parameter in the primal formulation is chosen according to a balancing principle derived from the model function approach, whereas the one in the dual formulation is determined by a path-following strategy based on the structure of the optimality conditions. Several numerical experiments confirm the efficiency and robustness of the proposed method and adaptive strategy.

Key words. L^1 data fitting, semismooth Newton, Fenchel duality, regularization parameter, balancing principle, model function

AMS subject classifications. 65J22, 65J20, 49M15, 49M29

DOI. 10.1137/090758003

1. Introduction.

This work is concerned with solving the inverse problem

$$Kx = y^\delta,$$

where $K : L^2(\Omega) \to L^2(\Omega)$ is a bounded linear operator, $\Omega \subset \mathbb{R}^n$ is a bounded domain, and $y^\delta \in L^2(\Omega)$ are noisy measurements with noise level $\|y^\dagger - y^\delta\|_{L^1} \leq \delta$ ($y^\dagger$ being the noise-free data). This problem is ill-posed in the sense of Hadamard, and, in particular, the solution often fails to depend continuously on the data. The now standard approach is Tikhonov regularization, which typically incorporates a priori information and amounts to solving a minimization problem of the form

$$\frac{1}{2}\|Kx - y^\delta\|_{L^2}^2 + \alpha R(x),$$

提高稳定性保证唯一性

where R is the regularization term and α is a regularization parameter determining the relative weight of these two terms. The choice of the regularization term R is application dependent,

*Received by the editors May 4, 2009; accepted for publication (in revised form) March 1, 2010; published electronically May 12, 2010. This work was supported by the Austrian Science Fund (FWF) under grant SFB F32 (SFB "Mathematical Optimization and Applications in Biomedical Sciences").
http://www.siam.org/journals/siims/3-2/75800.html

[†]Institute for Mathematics and Scientific Computing, University of Graz, Heinrichstrasse 36, A-8010 Graz, Austria (christian.clason@uni-graz.at, karl.kunisch@uni-graz.at).
[‡]Center for Industrial Mathematics, University of Bremen, D-28334 Bremen, Germany (btjin@informatik.uni-bremen.de). The work of this author was supported by the Alexander von Humboldt Foundation through a postdoctoral research fellowship.

and in what follows, we shall focus on the choice $R(x) = \frac{1}{2}\|x\|_{L^2}^2$, which is suitable for smooth solutions.

The classical Tikhonov regularization uses an L^2 data fitting term, which, statistically 参考凸优化 speaking, is most appropriate for Gaussian noise. The success of this formulation relies crucially on the validity of the Gaussian assumption [24] (no heavy tails, and the noise distribution is symmetric); in some practical applications, however, the noise is non-Gaussian. For instance, the noise may follow a Laplace distribution as in certain inverse problems arising in signal processing [3]. Noise models of impulse 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], and call for the use of L^1 data fitting. The advantage of using the L^1 norm is given by the fact that the solution is L1范数更鲁棒 more robust when compared to the L^2 norm [18]. In particular, a small number of outliers has less influence on the solution, whereas the L^2 formulation needs some extra processing outlier的权值变小了 stage utilizing robust procedures to locate the outliers [30]. These considerations motivate the formulation

$$\mathcal{J}_\alpha(x) = \|Kx - y^\delta\|_{L^1} + \frac{\alpha}{2}\|x\|_{L^2}^2.$$

Recently, minimization of cost functions involving L^1 data fitting has received growing interest in diverse disciplines, e.g., signal processing [3, 2], image processing [27, 28, 8, 24], and distributed parameter identification [6]. Alliney [2] studied the properties of a discrete variational problem and established its relation with recursive median filters. Nikolova [27] showed that in L^1 data fitting for discrete denoising problems, a certain number of data points can be attained exactly, and thus theoretically justified its superior performance over the standard model for certain types of noise. Chan and Esedoḡlu [8] investigated the analytical properties of minimizers and their implication for multiscale image decomposition and parameter selection in the context of total variation image denoising. These results were recently extended and refined by a number of authors [35, 1, 10].

Numerical methods for the solution of L^1 data fitting problems have also received some attention; see, for instance, [24] (an active set algorithm for denoising), [14] (an interior point algorithm for image restoration problems), [29] (a generalization of the classical iteratively reweighted least squares method), [34] (an alternating minimization algorithm for color image restoration), and [7] (a primal-dual algorithm for image restoration). Note that these studies focus on structured matrices, e.g., identity or (block) Toeplitz, instead of general (infinite-dimensional) operators. Except for [24], the above-mentioned works focus on total variation regularization because of their interest in image processing.

This paper focuses on the efficient numerical solution of the L^1 data fitting problem in infinite dimensions using a semismooth Newton method and on the automatic choice of the regularization parameter with a model function approach. Semismooth Newton methods were applied to inverse problems with L^1-type functionals in, e.g., [15] (for sparsity constrained ℓ^2-minimization) and [9] (for ℓ^1-TV image restoration). One particular advantage of semismooth Newton methods in function spaces [17, 31, 21] is their mesh independence (i.e., the necessary number of iterations is independent of the problem size). The model function approach was originally proposed for efficiently solving Morozov's discrepancy equation [20, 25, 33]. However, the discrepancy principle requires an accurate estimate of the noise level δ, which might be unavailable in practice. Therefore, it is useful to estimate the noise level and to develop heuristic parameter choice rules based on this estimate. In [25] it was suggested that one estimate the noise level using an iterative approach involving model functions. Numerically, it was observed that the method gives an excellent approximation of the noise level δ after two or three iterations. However, the method was formulated for least squares data fitting problems, and, also, the mechanism of the iteration remained unexplored.

Our main contributions are as follows. First, we propose and analyze an efficient semismooth Newton method for solving the L^1 data fitting problem. Second, we derive heuristic choice rules for the regularization parameters in the primal and dual problems based on the idea of balancing, which do not require knowledge of the noise level. Third, the convergence property of a fixed point iteration for the automatic parameter choice is investigated.

This paper is organized as follows. In section 2, we treat the primal and a dual formulation of the problem. The regularizing properties, especially the convergence rate results for a priori and a posteriori parameter choice rules, are shown, and optimality conditions are established. Section 3 is devoted to the solution of the dual problem using a semismooth Newton method. The additional regularization guaranteeing superlinear convergence is discussed in section 3.1, while section 3.2 concerns the convergence of the semismooth Newton method. Our adaptive rules for choosing regularization parameters in the primal and dual problems are presented in section 4. We conclude with several numerical experiments involving test problems in one and two dimensions.

2. Properties of minimizers.

2.1. Primal problem. In this section, we consider the properties of the primal problem

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

The functional $\mathcal{J}_\alpha$ is strictly convex and thus has a unique minimizer x_α. The next result, whose proof is rather standard and is thus omitted [12], summarizes the regularizing property of the functional $\mathcal{J}_\alpha$. For the next result, as well as for Theorems 2.3 and 2.4, we shall assume that the noise-free data $y^\dagger$ is attainable; i.e., there exists some $x^\dagger \in L^2$ such that $y^\dagger = Kx^\dagger$.

Theorem 2.1. *For each fixed α, there exists a unique minimizer x_α to the functional $\mathcal{J}_\alpha$ which depends continuously on the data y^δ. Moreover, if the regularization parameter α satisfies $\alpha \to 0$ and if $\lim_{\delta\to 0^+} \delta/\alpha = 0$, then x_α converges to $x^\dagger$, a minimum norm solution of the inverse problem, as $\delta \to 0$.*

We also need the following results on properties of the value function

$$F(\alpha) = \|Kx_\alpha - y^\delta\|_{L^1} + \frac{\alpha}{2}\|x_\alpha\|_{L^2}^2.$$

The proofs can be found in [19, 22].

Theorem 2.2. *The functions $\|Kx_\alpha - y^\delta\|_{L^1}$ and $\|x_\alpha\|_{L^2}^2$ are continuous and, respectively, monotonically increasing and decreasing with respect to α in the sense that*

$$\begin{aligned}(\alpha_1-\alpha_2)(\|Kx_{\alpha_1} - y^\delta\|_{L^1} - \|Kx_{\alpha_2} - y^\delta\|_{L^1}) &\geq 0,\\ (\alpha_1-\alpha_2)(\|x_{\alpha_1}\|_{L^2}^2 - \|x_{\alpha_2}\|_{L^2}^2) &\leq 0.\end{aligned}$$

直观上，权重影响最终的函数值

The value function $F(\alpha)$ is continuous and increasing, and it is differentiable with derivative

$$F'(\alpha) = \frac{1}{2}\|x_\alpha\|_{\mathrm{L}^2}^2.$$

The next result shows a convergence rate result for a priori parameter choice rules under certain source conditions. To explicitly indicate the dependence of the minimizer x_α on the data y^δ, we shall use the notation x_α^δ for the next two results. We will also assume that the following source condition holds: there exists some $w \in \mathrm{L}^\infty$ such that the exact solution $x^\dagger$ satisfies

$$x^\dagger = K^* w. \tag{2.1}$$

Theorem 2.3. *Assume that the source condition (2.1) holds. Then for sufficiently small δ and $\alpha = \mathcal{O}(\delta^\varepsilon)$ with $\varepsilon \in (0,1)$, the minimizer x_α^δ of the functional $\mathcal{J}_\alpha$ satisfies*

$$\|x_\alpha^\delta - x^\dagger\|_{\mathrm{L}^2} \le c\delta^{\frac{1-\varepsilon}{2}}.$$

Proof. By the minimizing property of x_α^δ, we have

$$\|Kx_\alpha^\delta - y^\delta\|_{\mathrm{L}^1} + \frac{\alpha}{2}\|x_\alpha^\delta\|_{\mathrm{L}^2}^2 \le \|Kx^\dagger - y^\delta\|_{\mathrm{L}^1} + \frac{\alpha}{2}\|x^\dagger\|_{\mathrm{L}^2}^2 \le \delta + \frac{\alpha}{2}\|x^\dagger\|_{\mathrm{L}^2}^2. \tag{2.2}$$

Therefore, we have

$$\|Kx_\alpha^\delta - y^\delta\|_{\mathrm{L}^1} + \frac{\alpha}{2}(\|x_\alpha^\delta\|_{\mathrm{L}^2}^2 - \|x^\dagger\|_{\mathrm{L}^2}^2 - 2\langle x^\dagger, x_\alpha^\delta - x^\dagger\rangle) \le \delta - \alpha\langle x^\dagger, x_\alpha^\delta - x^\dagger\rangle,$$

i.e.,

$$\|Kx_\alpha^\delta - y^\delta\|_{\mathrm{L}^1} + \frac{\alpha}{2}\|x_\alpha^\delta - x^\dagger\|_{\mathrm{L}^2}^2 \le \delta - \alpha\langle x^\dagger, x_\alpha^\delta - x^\dagger\rangle.$$

Now by the source condition (2.1) and the triangle inequality we have

$$\begin{aligned}\|Kx_\alpha^\delta - y^\delta\|_{\mathrm{L}^1} + \frac{\alpha}{2}\|x_\alpha^\delta - x^\dagger\|_{\mathrm{L}^2}^2 &\le \delta - \alpha\langle K^*w, x_\alpha^\delta - x^\dagger\rangle \\ &= \delta - \alpha\langle w, Kx_\alpha^\delta - y^\dagger\rangle \\ &\le \delta + \alpha\|w\|_{\mathrm{L}^\infty}(\|Kx_\alpha^\delta - y^\delta\|_{\mathrm{L}^1} + \|y^\delta - y^\dagger\|_{\mathrm{L}^1}).\end{aligned}$$

Rearranging the terms gives

$$(1 - \alpha\|w\|_{\mathrm{L}^\infty})\|Kx_\alpha^\delta - y^\delta\|_{\mathrm{L}^1} + \frac{\alpha}{2}\|x_\alpha^\delta - x^\dagger\|_{\mathrm{L}^2}^2 \le (1 + \alpha\|w\|_{\mathrm{L}^\infty})\delta.$$

The desired convergence rate now follows using $\alpha = \mathcal{O}(\delta^\varepsilon)$. ∎

Next we consider Morozov's classical discrepancy principle [26], which consists in choosing α as a solution of the following nonlinear equation:

$$\|Kx_\alpha^\delta - y^\delta\|_{\mathrm{L}^1} = c\delta \tag{2.3}$$

for some $c \ge 1$. Under the condition that $\lim_{\alpha\to 0^+} \|Kx_\alpha^\delta - y^\delta\|_{\mathrm{L}^1} < c\delta$ and $\lim_{\alpha\to+\infty} \|Kx_\alpha^\delta - y^\delta\|_{\mathrm{L}^1} = \|y^\delta\|_{\mathrm{L}^1} > c\delta$, there exists at least one positive solution to Morozov's equation (2.3), which follows from the continuity and monotonicity results; see Theorem 2.2.

In contrast to a priori choice rules, the discrepancy principle yields a concrete scheme for determining an appropriate regularization parameter, and is mathematically rigorous in that consistency and convergence rates can be established [13, 12]. Its applicability for $\mathcal{J}_\alpha$ follows directly from the results in [22].

The next result shows a convergence rate result for the discrepancy principle. We point out that there have been few convergence rate results for the discrepancy principle for the Tikhonov functional other than the classical L^2-L^2 formulation.

Theorem 2.4. *Assume that the source condition* (2.1) *holds, and that the regularization parameter $\alpha = \alpha(\delta)$ is determined according to the discrepancy principle. Then the minimizer x_α^δ of the functional $\mathcal{J}_\alpha$ satisfies*

$$\|x_\alpha^\delta - x^\dagger\|_{L^2} \leq [2(c+1)\|w\|_{L^\infty}]^{\frac{1}{2}}\delta^{\frac{1}{2}}.$$

Proof. Relation (2.2) together with Morozov's equation (2.3) for $\alpha(\delta)$ gives

$$\|x_\alpha^\delta\|_{L^2}^2 \leq \|x^\dagger\|_{L^2}^2,$$

from which it follows that

$$\begin{aligned}
\|x_\alpha^\delta - x^\dagger\|_{L^2}^2 &\leq 2\langle x^\dagger, x^\dagger - x_\alpha^\delta\rangle = 2\langle K^*w, x^\dagger - x_\alpha^\delta\rangle \\
&\leq 2\|w\|_{L^\infty}\|Kx^\dagger - Kx_\alpha^\delta\|_{L^1} \\
&\leq 2\|w\|_{L^\infty}(\|Kx^\dagger - y^\delta\|_{L^1} + \|Kx_\alpha^\delta - y^\delta\|_{L^1}) \\
&\leq 2\|w\|_{L^\infty}(\delta + c\delta),
\end{aligned}$$

again by (2.3). This yields the desired estimate. ■

2.2. Dual problem. In this section, we consider the problem

$$(\mathcal{P}^*) \qquad \boxed{\begin{cases} \displaystyle\min_{p\in L^2} \frac{1}{2\alpha}\|K^*p\|_{L^2}^2 - \langle p, y^\delta\rangle_{L^2} \\ \text{s.t.} \qquad \|p\|_{L^\infty} \leq 1, \end{cases}}$$

which we will show to be the dual problem of $(\mathcal{P})$.

Theorem 2.5. *The dual problem of $(\mathcal{P})$ is $(\mathcal{P}^*)$, which has at least one solution, and the solutions $x_\alpha \in L^2$ of $(\mathcal{P})$ and $p_\alpha \in L^2$ of $(\mathcal{P}^*)$ are related by*

$$(2.4) \qquad \begin{cases} K^*p_\alpha = \alpha x_\alpha, \\ \quad 0 \leq \langle Kx_\alpha - y^\delta, p - p_\alpha\rangle_{L^2} \end{cases}$$

for all $p \in L^2$ with $\|p\|_{L^\infty} \leq 1$.

Proof. We apply Fenchel duality [11], setting

$$\begin{aligned}
&\mathcal{F} : L^2 \to \mathbb{R}, && \mathcal{F}(x) = \frac{\alpha}{2}\|x\|_{L^2}^2, \\
&\mathcal{G} : L^2 \to \mathbb{R}, && \mathcal{G}(x) = \|x - y^\delta\|_{L^1}, \\
&\Lambda : L^2 \to L^2, && \Lambda x = Kx.
\end{aligned}$$

The Fenchel conjugates of $\mathcal{F}$ and $\mathcal{G}$ are given by

$$\mathcal{F}^* : \mathrm{L}^2 \to \mathbb{R}, \qquad \mathcal{F}^*(q) = \frac{1}{2\alpha} \|q\|_{\mathrm{L}^2}^2 ,$$

$$\mathcal{G}^* : \mathrm{L}^2 \to \mathbb{R} \cup \{\infty\}, \qquad \mathcal{G}^*(q) = \begin{cases} \langle q, y^\delta \rangle_{\mathrm{L}^2} & \text{if } \|q\|_{\mathrm{L}^\infty} \le 1, \\ \infty & \text{if } \|q\|_{\mathrm{L}^\infty} > 1. \end{cases}$$

Since the functionals $\mathcal{F}$ and $\mathcal{G}$ are convex lower semicontinuous, proper, and continuous at $x_0 = 0 = Kx_0$, and K is a continuous linear operator, the Fenchel duality theorem states that

$$\inf_{x \in \mathrm{L}^2} \mathcal{F}(x) + \mathcal{G}(\Lambda x) = \sup_{p \in \mathrm{L}^2} -\mathcal{F}^*(\Lambda^* p) - \mathcal{G}^*(-p) \tag{2.5}$$

holds, and that the right-hand side of (2.5) has at least one solution.

Furthermore, the equality in (2.5) is attained at (x_α, p_α) if and only if

$$\begin{cases} \Lambda^* p_\alpha \in \partial \mathcal{F}(x_\alpha), \\ -p_\alpha \in \partial \mathcal{G}(\Lambda x_\alpha). \end{cases}$$

Since $\mathcal{F}$ is Fréchet-differentiable, the first relation of (2.4) follows by direct calculation. Recall that by the definition of the subgradient

$$-p_\alpha \in \partial \mathcal{G}(\Lambda x_\alpha) \Leftrightarrow \Lambda x_\alpha \in \partial \mathcal{G}^*(-p_\alpha)$$

holds. Subdifferential calculus then yields

$$\Lambda x_\alpha - y^\delta \in \partial I_{\{\|-p_\alpha\|_{\mathrm{L}^\infty} \le 1\}},$$

where I_S denotes the indicator function of the set S, whose subdifferential coincides with the normal cone at S (cf., e.g., [21, Ex. 4.21]). We thus obtain that

$$0 \ge \langle Kx_\alpha - y^\delta, p + p_\alpha \rangle_{\mathrm{L}^2}$$

for all $p \in \mathrm{L}^2$ with $\|p\|_{\mathrm{L}^\infty} \le 1$, from which the second relation follows. ■

Remark 2.6. The solution of problem $(\mathcal{P}^*)$ is no longer unique; rather, any solution p_α can be written as $p_\alpha = \tilde{p}_\alpha + \hat{p}_\alpha$ with $\tilde{p}_\alpha \in \ker K^*$ and a unique $\hat{p}_\alpha \in (\ker K^*)^\perp$. Nevertheless, the corresponding primal solution x_α calculated using the first extremality relation (2.4) will be unique. The treatment of the nonuniqueness in the numerical solution of problem $(\mathcal{P}^*)$ will be discussed in section 3.1.

Assisted with Theorem 2.5, we can now derive the first order optimality conditions for problem $(\mathcal{P}^*)$.

Corollary 2.7. *Let $p_\alpha \in \mathrm{L}^2$ be a solution of $(\mathcal{P}^*)$. Then there exists $\lambda_\alpha \in \mathrm{L}^2$ such that*

$$\begin{cases} \frac{1}{\alpha} KK^* p_\alpha - y^\delta + \lambda_\alpha = 0, \\ \langle \lambda_\alpha, p - p_\alpha \rangle_{\mathrm{L}^2} \le 0 \end{cases} \tag{2.6}$$

holds for all $p \in \mathrm{L}^2$ with $\|p\|_{\mathrm{L}^\infty} \le 1$.

Proof. By applying $\frac{1}{\alpha}K$ to the first relation in (2.4) and setting $\lambda_\alpha = -(Kx_\alpha - y^\delta)$, we immediately obtain the existence of a Lagrange multiplier satisfying (2.6). ■

The following structural information for the solution of problem ($\mathcal{P}$) is a direct consequence of (2.4).

Corollary 2.8. *Let x_α be the minimizer of ($\mathcal{P}$). Then the following holds for any $p \in \mathrm{L}^2$, $p \geq 0$:*

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

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 x_α.

3. Solution by semismooth Newton method.

3.1. Regularization. If the inversion of K is ill-posed, problem ($\mathcal{P}^*$) remains ill-posed in spite of the pointwise bounds on p. To counter this and to ensure superlinear convergence of the semismooth Newton method for solving the constrained optimization problem, we introduce the regularized problem

$$(\mathcal{P}^*_\beta) \qquad \boxed{\begin{cases} \min\limits_{p\in \mathrm{H}^1} \dfrac{1}{2\alpha}\|K^*p\|^2_{\mathrm{L}^2} + \dfrac{\beta}{2}\|\nabla p\|^2_{\mathrm{L}^2} - \langle p, y^\delta\rangle_{\mathrm{L}^2} \\ \text{s.t.} \qquad \|p\|_{\mathrm{L}^\infty} \leq 1 \end{cases}}$$

for $\beta > 0$. The interplay between the pointwise bound on p and the seminorm regularization term will enable an easy choice for the regularization parameter β, which will be explained in section 4.2. We assume that $\ker K^* \cap \ker \nabla = \{0\}$; i.e., constant functions do not belong to the kernel of K^*. Under this assumption the inner product $\frac{1}{\alpha}\langle K^*\cdot, K^*\cdot\rangle + \beta\langle\nabla\cdot,\nabla\cdot\rangle$ induces an equivalent norm on H^1, and problem ($\mathcal{P}^*_\beta$) admits a unique solution p_β. This assumption can be removed if the seminorm regularization is replaced by the full H^1 norm.

To solve ($\mathcal{P}^*_\beta$) numerically, we introduce a Moreau–Yosida regularization of the box constraints and consider

对应p的约束

$$(\mathcal{P}^*_{\beta,c}) \quad \min_{p\in \mathrm{H}^1} \frac{1}{2\alpha}\|K^*p\|^2_{\mathrm{L}^2} + \frac{\beta}{2}\|\nabla p\|^2_{\mathrm{L}^2} - \langle p, y^\delta\rangle_{\mathrm{L}^2} + \frac{1}{2c}\|\max(0, c(p-1))\|^2_{\mathrm{L}^2} + \frac{1}{2c}\|\min(0, c(p+1))\|^2_{\mathrm{L}^2}$$

惩罚项

for $c > 0$, where the max and min are taken pointwise. For fixed β and c, under the above assumption, $\frac{1}{2\alpha}\|K^*p\|^2_{\mathrm{L}^2} + \frac{\beta}{2}\|\nabla p\|^2_{\mathrm{L}^2}$ is strictly convex and hence problem ($\mathcal{P}^*_{\beta,c}$) admits a unique minimizer p_c. The optimality system for ($\mathcal{P}^*_{\beta,c}$) is given by

$$\begin{cases} \dfrac{1}{\alpha}KK^*p_c - \beta\Delta p_c - y^\delta + \lambda_c = 0, \\ \lambda_c = \max(0, c(p_c - 1)) + \min(0, c(p_c + 1)). \end{cases} \tag{3.1}$$

分段求导

This yields higher regularity for the Lagrange multiplier λ_c: since the max (and min) operator is continuous from $W^{1,\infty}$ to $W^{1,\infty}$ (cf., e.g., [5]) and $p_c \in H^1$ by construction, the second equation of (3.1) ensures that $\lambda_c \in H^1$ as well.

Now we address the convergence, as $c \to \infty$, of the solutions of (3.1) to that of problem $(\mathcal{P}_\beta^*)$. To this end, we introduce the optimality system for problem $(\mathcal{P}_\beta^*)$:

$$\begin{cases} \frac{1}{\alpha} K K^* p_\beta - \beta \Delta p_\beta - y^\delta + \lambda_\beta = 0, \\ \langle \lambda_\beta, p - p_\beta \rangle_{L^2} \leq 0 \end{cases} \tag{3.2}$$

for all $p \in H^1$ with $\|p\|_{L^\infty} \leq 1$ and a $\lambda_\beta \in (H^1)^*$, the dual space of H^1.

Theorem 3.1. *For $\beta > 0$ fixed, let $(p_c, \lambda_c) \in H^1 \times (H^1)^*$ be the solution of (3.1) for $c > 0$, and let $(p_\beta, \lambda_\beta) \in H^1 \times (H^1)^*$ be the solution of (3.2). Then we have, as $c \to \infty$,*

$$\begin{aligned} p_c &\to p_\beta \quad \text{in } H^1, \\ \lambda_c &\rightharpoonup \lambda_\beta \quad \text{in } (H^1)^*. \end{aligned}$$

Proof. From the optimality system (3.1), we have that pointwise in $x \in \Omega$

$$\lambda_c p_c = \max(0, c(p_c - 1))p_c + \min(0, c(p_c + 1))p_c = \begin{cases} c(p_c - 1)p_c, & p_c \geq 1, \\ 0, & |p_c| < 1, \\ c(p_c + 1)p_c, & p_c \leq -1 \end{cases}$$

holds and thus, since $\lambda_c \in W^{1,\infty} \subset L^2$, that

$$\langle \lambda_c, p_c \rangle_{L^2} \geq \frac{1}{c} \|\lambda_c\|_{L^2}^2 . \tag{3.3}$$

Inserting p_c as a test function in the variational form of (3.1),

$$\frac{1}{\alpha} \langle K^* p_c, K^* v \rangle_{L^2} + \beta \langle \nabla p_c, \nabla v \rangle_{L^2} - \langle y^\delta, v \rangle_{L^2} + \langle \lambda_c, v \rangle_{(H^1)^*, H^1} = 0, \tag{3.4}$$

for all $v \in H^1$, yields

$$\frac{1}{\alpha} \|K^* p_c\|_{L^2}^2 + \beta \|\nabla p_c\|_{L^2}^2 + \frac{1}{c} \|\lambda_c\|_{L^2}^2 \leq \|p_c\|_{L^2} \|y^\delta\|_{L^2}, \tag{3.5}$$

and by recalling that by assumption the first two terms define an equivalent norm on H^1, we deduce that $\|p_c\|_{H^1} \leq C \left\|y^\delta\right\|_{L^2}$ for some constant C. Moreover,

$$\begin{aligned} \|\lambda_c\|_{(H^1)^*} &= \sup_{\substack{v \in H^1, \\ \|v\|_{H^1} \leq 1}} \langle \lambda_c, v \rangle_{(H^1)^*, H^1} \\ &= \sup_{\substack{v \in H^1, \\ \|v\|_{H^1} \leq 1}} \left[-\frac{1}{\alpha} \langle K^* p_c, K^* v \rangle_{L^2} - \beta \langle \nabla p_c, \nabla v \rangle_{L^2} + \langle y^\delta, v \rangle_{L^2} \right] \\ &\leq \sup_{\substack{v \in H^1, \\ \|v\|_{H^1} \leq 1}} \left[C_1 \|p_c\|_{H^1} \|v\|_{H^1} + \|v\|_{L^2} \|y^\delta\|_{L^2} \right] \\ &\leq (CC_1 + 1) \sup_{\substack{v \in H^1, \\ \|v\|_{H^1} \leq 1}} \|v\|_{H^1} \|y^\delta\|_{L^2} =: K < \infty, \end{aligned}$$

where C_1 is another norm equivalence constant. Thus, (p_c, λ_c) is uniformly bounded in $\mathrm{H}^1 \times (\mathrm{H}^1)^*$, and there exists some $(\tilde{p}, \tilde{\lambda}) \in \mathrm{H}^1 \times (\mathrm{H}^1)^*$ such that

$$(p_c, \lambda_c) \rightharpoonup (\tilde{p}, \tilde{\lambda}) \quad \text{in } \mathrm{H}^1 \times (\mathrm{H}^1)^*.$$

Passing to the limit in (3.4), we obtain

$$\frac{1}{\alpha} \langle K^*\tilde{p}, K^*v\rangle_{\mathrm{L}^2} + \beta \langle \nabla\tilde{p}, \nabla v\rangle_{\mathrm{L}^2} - \langle y^\delta, v\rangle_{\mathrm{L}^2} + \left\langle \tilde{\lambda}, v\right\rangle_{(\mathrm{H}^1)^*,\mathrm{H}^1} = 0$$

for all $v \in \mathrm{H}^1$.

We next verify the feasibility of $\tilde{p}$. By pointwise inspection similar to (3.3), we obtain that

$$\frac{1}{c} \|\lambda_c\|^2_{\mathrm{L}^2} = c \|\max(0, p_c - 1)\|^2_{\mathrm{L}^2} + c \|\min(0, p_c + 1)\|^2_{\mathrm{L}^2}.$$

From (3.5), we have that $\frac{1}{c} \|\lambda_c\|^2_{\mathrm{L}^2} \leq C \left\|y^\delta\right\|^2_{\mathrm{L}^2}$, so that

$$\begin{aligned} \|\max(0, p_c - 1)\|^2_{\mathrm{L}^2} &\leq \frac{1}{c} C \|y^\delta\|^2_{\mathrm{L}^2} \to 0, \\ \|\min(0, p_c + 1)\|^2_{\mathrm{L}^2} &\leq \frac{1}{c} C \|y^\delta\|^2_{\mathrm{L}^2} \to 0 \end{aligned}$$

as $c \to \infty$. Since $p_c \to \tilde{p}$ strongly in L^2, this implies that

$$-1 \leq \tilde{p}(x) \leq 1 \quad \text{for all } x \in \Omega.$$

It remains to show that the second equation of (3.2) holds. First, the minimizing property of p_c yields that

$$\frac{1}{2\alpha} \|K^* p_c\|^2_{\mathrm{L}^2} + \frac{\beta}{2} \|\nabla p_c\|^2_{\mathrm{L}^2} - \langle p_c, y^\delta\rangle_{\mathrm{L}^2} \leq \frac{1}{2\alpha} \|K^* p\|^2_{\mathrm{L}^2} + \frac{\beta}{2} \|\nabla p\|^2_{\mathrm{L}^2} - \langle p, y^\delta\rangle_{\mathrm{L}^2}$$

holds for all feasible $p \in \mathrm{H}^1$. Therefore, we have that

$$\limsup_{c\to\infty} \left[\frac{1}{2\alpha} \|K^* p_c\|^2_{\mathrm{L}^2} + \frac{\beta}{2} \|\nabla p_c\|^2_{\mathrm{L}^2} - \langle p_c, y^\delta\rangle_{\mathrm{L}^2}\right] \leq \frac{1}{2\alpha} \|K^*\tilde{p}\|^2_{\mathrm{L}^2} + \frac{\beta}{2} \|\nabla \tilde{p}\|^2_{\mathrm{L}^2} - \langle \tilde{p}, y^\delta\rangle_{\mathrm{L}^2}$$

and consequently $p_c \to \tilde{p}$ strongly in H^1. Now observe that

$$\langle \lambda_c, p - p_c\rangle_{(\mathrm{H}^1)^*,\mathrm{H}^1} = \langle \max(0, c(p_c - 1)), p - p_c\rangle_{\mathrm{L}^2} + \langle \min(0, c(p_c + 1)), p - p_c\rangle_{\mathrm{L}^2} \leq 0$$

holds for all $p \in \mathrm{H}^1$ with $\|p\|_{\mathrm{L}^\infty} \leq 1$. Thus

$$\left\langle \tilde{\lambda}, p - \tilde{p}\right\rangle_{(\mathrm{H}^1)^*,\mathrm{H}^1} \leq 0$$

is satisfied for all $p \in \mathrm{H}^1$ with $\|p\|_{\mathrm{L}^\infty} \leq 1$. Therefore, $(\tilde{p}, \tilde{\lambda}) \in \mathrm{H}^1 \times (\mathrm{H}^1)^*$ satisfies (3.2), and since the solution of (3.2) is unique, $\tilde{p} = p_\beta$ and $\tilde{\lambda} = \lambda_\beta$ follows. ■

Next we address the convergence of the solution of ($\mathcal{P}_\beta^*$) as $\beta \to 0$ to a solution of ($\mathcal{P}^*$), which might be nonunique if the operator K is not injective. The functional in ($\mathcal{P}^*$) is convex, as is the set of all minimizers, and thus, if the problem has a solution in H^1, there exists an element with minimal H^1 seminorm, denoted by $p^\dagger$.

Theorem 3.2. *Let $\{\beta_n\} \to 0$. Then the sequence of minimizers $\{p_{\beta_n}\}$ of ($\mathcal{P}_\beta^*$) has a subsequence converging weakly to a minimizer of problem ($\mathcal{P}^*$). If the operator K is injective or there exists a unique $p^\dagger$ as defined above, then the whole sequence converges weakly to $p^\dagger$.*

Proof. Since the minimizers $p_n := p_{\beta_n}$ of ($\mathcal{P}_\beta^*$) satisfy $\|p_n\|_{\mathrm{L}^\infty} \le 1$, the sequence $\{p_n\}$ is uniformly bounded in L^2 independently of n. Therefore, there exists a subsequence, also denoted by $\{p_n\}$, converging weakly in L^2 to some $p^* \in \mathrm{L}^2$. By the weak lower semicontinuity of norms, we have that

$$\|K^*p^*\|_{\mathrm{L}^2}^2 \le \liminf_{n\to\infty} \|K^*p_n\|_{\mathrm{L}^2}^2, \quad \langle p^*, y^\delta\rangle_{\mathrm{L}^2} = \lim_{n\to\infty} \langle p_n, y^\delta\rangle_{\mathrm{L}^2},$$

and

$$\|p^*\|_{\mathrm{L}^\infty} \le \liminf_{n\to\infty} \|p_n\|_{\mathrm{L}^\infty} \le 1.$$

Therefore, by the minimizing property of p_n, for any fixed $p \in \mathrm{H}^1$ we have that

$$\begin{aligned}\frac{1}{2\alpha}\|K^*p^*\|_{\mathrm{L}^2}^2 - \langle p^*, y^\delta\rangle_{\mathrm{L}^2} &\le \liminf_{n\to\infty}\left(\frac{1}{2\alpha}\|K^*p_n\|_{\mathrm{L}^2}^2 - \langle p_n, y^\delta\rangle_{\mathrm{L}^2} + \frac{\beta_n}{2}\|\nabla p_n\|_{\mathrm{L}^2}^2\right)\\ &\le \liminf_{n\to\infty}\left(\frac{1}{2\alpha}\|K^*p\|_{\mathrm{L}^2}^2 - \langle p, y^\delta\rangle_{\mathrm{L}^2} + \frac{\beta_n}{2}\|\nabla p\|_{\mathrm{L}^2}^2\right)\\ &= \frac{1}{2\alpha}\|K^*p\|_{\mathrm{L}^2}^2 - \langle p, y^\delta\rangle_{\mathrm{L}^2}.\end{aligned}$$

Therefore, p^* is a minimizer of problem ($\mathcal{P}^*$) over H^1. Now the density of H^1 in L^2 shows that p^* is also a minimizer of problem ($\mathcal{P}^*$) over L^2.

Now by the minimizing properties of $p^\dagger$ and p_n, we have that

$$\begin{aligned}\frac{1}{2\alpha}\|K^*p_n\|_{\mathrm{L}^2}^2 - \langle p_n, y^\delta\rangle_{\mathrm{L}^2} + \frac{\beta_n}{2}\|\nabla p_n\|_{\mathrm{L}^2}^2 &\le \frac{1}{2\alpha}\|K^*p^\dagger\|_{\mathrm{L}^2}^2 - \langle p^\dagger, y^\delta\rangle_{\mathrm{L}^2} + \frac{\beta_n}{2}\|\nabla p^\dagger\|_{\mathrm{L}^2}^2,\\ \frac{1}{2\alpha}\|K^*p^\dagger\|_{\mathrm{L}^2}^2 - \langle p^\dagger, y^\delta\rangle_{\mathrm{L}^2} &\le \frac{1}{2\alpha}\|K^*p_n\|_{\mathrm{L}^2}^2 - \langle p_n, y^\delta\rangle_{\mathrm{L}^2}.\end{aligned}$$

Adding these two inequalities together, we deduce that

$$\|\nabla p_n\|_{\mathrm{L}^2}^2 \le \|\nabla p^\dagger\|_{\mathrm{L}^2}^2,$$

which together with the weak lower semicontinuity of the H^1 seminorm yields

$$\|\nabla p^*\|_{\mathrm{L}^2}^2 \le \|\nabla p^\dagger\|_{\mathrm{L}^2}^2;$$

i.e., p^* is a minimizer with minimal H^1 seminorm. If K is injective or $p^\dagger$ is unique, then it follows that $p^* = p^\dagger$. Consequently, each subsequence has a subsequence converging weakly to $p^\dagger$, and the whole sequence converges weakly. ∎

Remark 3.3. For the numerical solution of the dual problem, we will let $\beta \to 0$ for some fixed $c > 0$ (cf. section 4.2). Thus it is useful to have the convergence of the solution $p_c = p_{\beta,c}$ of $(\mathcal{P}^*_{\beta,c})$ as $\beta \to 0$. The proof of this result is similar to that of Theorem 3.2 and is given in Appendix A.

Remark 3.4. For completeness, we also state how the regularization introduced in this section affects the primal problem. Setting

$$
\begin{aligned}
&\mathcal{F} : \mathrm{L}^2 \to \mathbb{R}, && \mathcal{F}(p) = -\langle p, y^\delta \rangle_{\mathrm{L}^2} + \frac{1}{2c} \|\max(0, c(p-1))\|^2_{\mathrm{L}^2} \\
& && \qquad + \frac{1}{2c} \|\min(0, c(p+1))\|^2_{\mathrm{L}^2}, \\
&\mathcal{G} : \mathrm{L}^2 \times (\mathrm{L}^2)^n \to \mathbb{R}, && \mathcal{G}(p,q) = \frac{1}{2\alpha} \|p\|^2_{\mathrm{L}^2} + \frac{\beta}{2} \|q\|^2_{\mathrm{L}^2}, \\
&\Lambda : \mathrm{H}^1 \to \mathrm{L}^2 \times (\mathrm{L}^2)^n, && \Lambda p = (K^* p, \nabla p),
\end{aligned}
$$

and calculating the corresponding duals, we find that the dual of problem $(\mathcal{P}^*_{\beta,c})$ is

$$
\min_{\substack{x \in \mathrm{L}^2, \\ z \in \mathrm{H}(\mathrm{div})}} \|Kx + \operatorname{div} z - y^\delta\|_{\mathrm{L}^1} + \frac{1}{2c} \|Kx + \operatorname{div} z - y^\delta\|^2_{\mathrm{L}^2} + \frac{\alpha}{2} \|x\|^2_{\mathrm{L}^2} + \frac{1}{2\beta} \|z\|^2_{(\mathrm{L}^2)^n}.
$$

3.2. Semismooth Newton method. The regularized optimality system (3.1) can be solved efficiently using a semismooth Newton method (cf. [17, 31]), which is superlinearly convergent. For this purpose, we consider (3.1) as a nonlinear equation $F(p) = 0$ with $F : \mathrm{H}^1 \to (\mathrm{H}^1)^*$,

$$
F(p) := \frac{1}{\alpha} KK^* p - \beta \Delta p + \max(0, c(p-1)) + \min(0, c(p+1)) - y^\delta.
$$

It is known (cf., e.g., [21, Ex. 8.14]) that the projection operator

$$
P(p) := \max(0, (p-1)) + \min(0, (p+1))
$$

is semismooth from L^q to L^p if and only if $q > p$ and that in this case it has as Newton derivative

$$
D_N P(p) h = h \chi_{\{|p|>1\}} := \begin{cases} h(x) & \text{if } |p(x)| > 1, \\ 0 & \text{if } |p(x)| \leq 1. \end{cases}
$$

微分，h为增量

Since sums of Frechét-differentiable functions and semismooth functions are semismooth (with canonical Newton derivatives), we find that F is semismooth and that its Newton derivative is

$$
D_N F(p) h = \frac{1}{\alpha} KK^* h - \beta \Delta h + c h \chi_{\{|p|>1\}}.
$$

A semismooth Newton step consists in solving for $p^{k+1} \in \mathrm{H}^1$ the equation

$$
D_N F(p^k)(p^{k+1} - p^k) = -F(p^k). \tag{3.6}
$$

牛顿法

Algorithm 1. Semismooth Newton method for (3.1).

1: Set $k = 0$, choose $p^0 \in \mathrm{H}^1$.
2: **repeat**
3: Set

$$\mathcal{A}_k^+ = \left\{x \in \Omega : \, p^k(x) > 1\right\},$$
$$\mathcal{A}_k^- = \left\{x \in \Omega : \, p^k(x) < -1\right\},$$
$$\mathcal{A}_k = \mathcal{A}_k^+ \cup \mathcal{A}_k^-.$$

4: Solve for $p^{k+1} \in \mathrm{H}^1$:

$$\frac{1}{\alpha} K K^* p^{k+1} - \beta \Delta p^{k+1} + c \chi_{\mathcal{A}_k} p^{k+1} = y^\delta + c(\chi_{\mathcal{A}_k^+} - \chi_{\mathcal{A}_k^-}).$$

5: Set $k = k + 1$.
6: **until** $(\mathcal{A}_k^+ = \mathcal{A}_{k-1}^+)$ and $(\mathcal{A}_k^- = \mathcal{A}_{k-1}^-)$, or $k = k_{max}$

Upon defining the active and inactive sets

$$\mathcal{A}_k^+ := \left\{x : \, p^k(x) > 1\right\}, \quad \mathcal{A}_k^- := \left\{x : \, p^k(x) < -1\right\}, \quad \mathcal{A}_k := \mathcal{A}_k^+ \cup \mathcal{A}_k^-,$$

the step (3.6) can be written explicitly as finding $p^{k+1} \in \mathrm{H}^1$ such that

$$\frac{1}{\alpha} K K^* p^{k+1} - \beta \Delta p^{k+1} + c \chi_{\mathcal{A}_k} p^{k+1} = y^\delta + c(\chi_{\mathcal{A}_k^+} - \chi_{\mathcal{A}_k^-}). \tag{3.7}$$

The resulting semismooth Newton method is given as Algorithm 1.

Theorem 3.5. *If $\|p_c - p^0\|_{\mathrm{H}^1}$ is sufficiently small, the sequence of iterates $\{p^k\}$ of Algorithm 1 converges superlinearly in H^1 to the solution p_c of (3.1) as $k \to \infty$.*

Proof. Since F is semismooth, it suffices to show that $(D_N F)^{-1}$ is uniformly bounded. Let $g \in (\mathrm{H}^1)^*$ be given. By assumption, the inner product $\beta \langle \nabla \cdot, \nabla \cdot \rangle_{\mathrm{L}^2} + \frac{1}{\alpha} \langle K^* \cdot, K^* \cdot \rangle_{\mathrm{L}^2}$ induces an equivalent norm on H^1, and thus the Lax–Milgram theorem ensures the existence of a unique $\varphi \in \mathrm{H}^1$ such that

$$\beta \langle \nabla \varphi, \nabla v \rangle_{\mathrm{L}^2} + \frac{1}{\alpha} \langle K^* \varphi, K^* v \rangle_{\mathrm{L}^2} + c \langle \chi_{\mathcal{A}} \varphi, v \rangle_{\mathrm{L}^2} = \langle g, v \rangle_{(\mathrm{H}^1)^*, \mathrm{H}^1}$$

holds for all $v \in \mathrm{H}^1$, independently of $\mathcal{A}$. Furthermore, φ satisfies

$$\|\varphi\|_{\mathrm{H}^1}^2 \leq C \|g\|_{\mathrm{H}^{1*}}^2,$$

with a constant C depending only on α, β, K, and Ω. This yields the desired uniform bound. The superlinear convergence now follows from standard results (e.g., [21, Thm. 8.16]). ■

The termination criterion in Algorithm 1, step 6, can be justified as follows.

Proposition 3.6. *If $\mathcal{A}_{k+1}^+ = \mathcal{A}_k^+$ and $\mathcal{A}_{k+1}^- = \mathcal{A}_k^-$ hold, then p^{k+1} satisfies $F(p^{k+1}) = 0$.*

This can be verified by simple inspection and is shown in [21, Rem. 7.1.1].

4. Adaptive choice of regularization parameters.

4.1. Choice of α by a model function approach. In this section, we propose a fixed point algorithm for adaptively determining the regularization parameter α based on a balancing principle and the model function approach. Specifically, we first state a balancing principle based on the value function introduced in section 2.1 and prove the existence of a regularization parameter α^* satisfying it. Then we propose a fixed point iteration which computes α^* by making use of a model function, which is a rational interpolant of the value function. The last part is devoted to the proof of convergence of the fixed point iteration.

4.1.1. Balancing principle. The idea of our regularization parameter choice method is to balance the data fitting term $\varphi(\alpha) = \|Kx_\alpha - y^\delta\|_{L^1}$ and the penalty term $\alpha F'(\alpha) = \frac{\alpha}{2}\|x_\alpha\|_{L^2}^2$. The balancing principle thus consists in choosing the regularization parameter α^* as the solution of the equation

$$(\sigma - 1)\varphi(\alpha^*) = \alpha^* F'(\alpha^*), \tag{4.1}$$

where $\sigma > 1$ controls the relative weight between these two terms. Similar balancing ideas underlie a number of heuristic parameter choice rules, e.g., the local minimum criterion [12], the zero-crossing method [23], and the L-curve criterion [16].

Theorem 4.1. *For σ sufficiently close to 1 and $y^\delta \neq 0$, there exists at least one positive solution to the balancing equation* (4.1).

For the proof of this result, we need the next lemma, which is proved in Appendix B.

Lemma 4.2. *The following limits hold true:*

$$\lim_{\alpha\to 0^+} \frac{\alpha}{2}\|x_\alpha\|_{L^2}^2 = \lim_{\alpha\to+\infty} \frac{\alpha}{2}\|x_\alpha\|_{L^2}^2 = 0.$$

In the following, we will repeatedly make use of the residual in (4.1):

$$r(\alpha) = \alpha F'(\alpha) - (\sigma - 1)\varphi(\alpha). \tag{4.2}$$

By Theorem 2.2, the functions $\varphi(\alpha)$ and $F'(\alpha)$ are continuous, and thus the function $r(\alpha)$ is continuous. We can now prove existence of a positive solution to the balancing equation (4.1).

Proof of Theorem 4.1. Lemma 4.2 shows that the following limits hold for $\sigma > 1$:

$$\begin{aligned}\lim_{\alpha\to 0^+} r(\alpha) &= -(\sigma-1)\lim_{\alpha\to 0^+}\|Kx_\alpha - y^\delta\|_{L^1} \le 0,\\ \lim_{\alpha\to+\infty} r(\alpha) &= -(\sigma-1)\|y^\delta\|_{L^1} < 0.\end{aligned}$$

However, $\|Kx_\alpha - y^\delta\|_{L^1} \le \|y^\delta\|_{L^1}$, and $\sup_{\alpha\in(0,+\infty)} \frac{\alpha}{2}\|x_\alpha\|_{L^2}^2 > 0$. Consequently, we have

$$r(\alpha) = \frac{\alpha}{2}\|x_\alpha\|_{L^2}^2 - (\sigma-1)\|Kx_\alpha - y^\delta\|_{L^1} \ge \frac{\alpha}{2}\|x_\alpha\|_{L^2}^2 - (\sigma-1)\|y^\delta\|_{L^1}.$$

Therefore, there exists a $\sigma_0 > 1$ such that $\sup_{\alpha\in(0,+\infty)} r(\alpha) > 0$ for all $\sigma \in (1, \sigma_0)$, and the existence of a positive solution follows. ∎

4.1.2. Model function and fixed point iteration. To find a solution of the balancing equation, we write (4.1) as

$$F(\alpha^*) = \sigma(F(\alpha^*) - \alpha^* F'(\alpha^*))$$

and consider a fixed point iteration, where α^{k+1} is chosen as the solution of

$$F(\alpha^{k+1}) = \sigma(F(\alpha^k) - \alpha^k F'(\alpha^k)). \tag{4.3}$$

To compute this solution, we make use of the model function approach, proposed in [20] for determining regularization parameters, which locally approximates the value function $F(\alpha)$ by rational polynomials. In this paper, we consider a model function of the form

$$m(\alpha) = b + \frac{s}{t+\alpha}.$$

Noting that $x_\alpha \to 0$ for $\alpha \to \infty$, we fix $b = \|y^\delta\|_{L^1}$ to match the asymptotic behavior of $F(\alpha)$ (although larger values of b work as well for our purposes). The parameters s and t are determined by the interpolation conditions

$$m(\alpha) = F(\alpha), \quad m'(\alpha) = F'(\alpha),$$

which together with the definition of $m(\alpha)$ give

$$b + \frac{s}{t+\alpha} = F(\alpha), \quad -\frac{s}{(t+\alpha)^2} = F'(\alpha).$$

The parameters s and t can be derived explicitly. We recall that by Theorem 2.2, we have $F'(\alpha) = \frac{1}{2}\|x_\alpha\|^2_{L^2}$, and this value can be calculated without any extra computational effort. If we replace the left-hand side of (4.3) by the value $m_k(\alpha^{k+1})$ of the model function $m_k(\alpha)$ derived from the interpolation conditions at α^k, we can thus explicitly compute a new iterate α^{k+1}. The resulting iteration is given as Algorithm 2. One of its salient features lies in not requiring knowledge of the noise level. Indeed, since $F(0)$ represents a lower bound for the noise level, the quantity $m(0)$ may be taken as a valid estimate of the noise level if the model function $m(\alpha)$ reasonably approximates the value function $F(\alpha)$ in the neighborhood of $\alpha = 0$.

Having found a fixed point $\bar\alpha$ of Algorithm 2, we find—using the fact that $\hat m = m(\bar\alpha)$ satisfies the interpolation condition at $\bar\alpha$—that $\bar\alpha$ is a solution of the balancing equation (4.1). We can show that under some very general assumptions, Algorithm 2 converges locally to such a fixed point. To this end, we call a solution α^* to (4.1) a regular attractor if there exists an $\varepsilon > 0$ such that $r(\alpha) < 0$ for $\alpha \in (\alpha^* - \varepsilon, \alpha^*)$ and $r(\alpha) > 0$ for $\alpha \in (\alpha^*, \alpha^* + \varepsilon)$.

Theorem 4.3. *Assume that α_0 satisfies $r(\alpha_0) < 0$ and that it is close to a regular attractor α^*. Then the sequence $\{\alpha_k\}$ generated by Algorithm 2 converges to α^*.*

The proof is given in the next subsection.

4.1.3. Convergence of fixed point iteration. To analyze the convergence of the fixed point algorithm, we first observe that

$$\begin{aligned}\alpha_{k+1} &= \frac{s_k}{\sigma\hat m - b} - t_k \\ &= \frac{(F(\alpha_k) - b)^2 - (b - F(\alpha_k) - \alpha_k F'(\alpha_k))(b - \sigma\varphi(\alpha_k))}{F'(\alpha_k)(b - \sigma\varphi(\alpha_k))},\end{aligned}$$

Algorithm 2. Fixed point algorithm for balancing equation.

```
Set k = 0, choose α_0 > 0, b ≥ ‖y‖_{L^1}, and σ > 1.
repeat
    Compute x_{α_k} by path-following semismooth Newton method (Algorithm 3).
    Compute F(α_k) and F'(α_k).
    Construct the model function m_k(α) = b + s_k/(t_k+α) from the interpolation condition at α_k
    by setting
        s_k = -(b - F(α_k))^2 / F'(α_k),
        t_k = (b - F(α_k)) / F'(α_k) - α_k.
    Solve for α_{k+1} in m_k(α_{k+1}) = σ(F(α_k) - α_k F'(α_k)), i.e.,
        m̂_k = F(α_k) - α_k F'(α_k),
        α_{k+1} = s_k / (σ m̂_k - b) - t_k.
    Set k = k + 1.
until k = k_max
```

where $\varphi(\alpha) = \|Kx_\alpha - y^\delta\|_{\mathrm{L}^1}$ denotes again the norm of the residual. The numerator can be simplified as follows:

$$
\begin{aligned}
&(F(\alpha_k) - b)^2 - (b - F(\alpha_k) - \alpha_k F'(\alpha_k))(b - \sigma\varphi(\alpha_k)) \\
&\qquad\qquad = (\alpha_k F'(\alpha_k))^2 + (\sigma - 1)\varphi(\alpha_k)[b - F(\alpha_k) - \alpha_k F'(\alpha_k)].
\end{aligned}
$$

Therefore, the fixed point iteration reads as follows:

$$
\alpha_{k+1} = \frac{(\alpha_k F'(\alpha_k))^2 + (\sigma - 1)\varphi(\alpha_k)[b - F(\alpha_k) - \alpha_k F'(\alpha_k)]}{F'(\alpha_k)(b - \sigma\varphi(\alpha_k))} =: \alpha_k \frac{N_k}{D_k}. \tag{4.4}
$$

Under the assumption $b > \sigma\|y^\delta\|_{\mathrm{L}^1}$, the denominator D_k is positive, and thus the iteration is well defined. Moreover, the following identity holds:

$$
N_k - D_k = [(\sigma - 1)\varphi(\alpha_k) - \alpha_k F'(\alpha_k)][b - F(\alpha_k)].
$$

Therefore, it follows from (4.4) that if $r(\alpha_k) = \alpha_k F'(\alpha_k) - (\sigma - 1)\varphi(\alpha_k) > 0$, then $N_k < D_k$

and consequently $\alpha_{k+1} < \alpha_k$; otherwise $\alpha_{k+1} > \alpha_k$ holds. Next consider

$$\begin{aligned}\alpha_{k+1} - \alpha_k = \alpha_k \frac{N_k}{D_k} - \alpha_k &= \frac{N_k - D_k}{F'(\alpha_k)(b - \sigma\varphi(\alpha_k))} \\ &= \frac{[(\sigma-1)\varphi(\alpha_k) - \alpha_k F'(\alpha_k)][b - F(\alpha_k)]}{F'(\alpha_k)(b - \sigma\varphi(\alpha_k))} \\ &= \left[(\sigma-1)\frac{\varphi(\alpha_k)}{F'(\alpha_k)} - \alpha_k\right] \frac{b - F(\alpha_k)}{b - \sigma\varphi(\alpha_k)} \\ &= [T(\alpha_k) - \alpha_k] \frac{b - F(\alpha_k)}{b - \sigma\varphi(\alpha_k)},\end{aligned}$$

where the operator $T(\alpha)$ is defined by

$$T(\alpha) = (\sigma - 1)\frac{\varphi(\alpha)}{F'(\alpha)}. \tag{4.5}$$

The auxiliary operator T can be regarded as the asymptotic of the operator $\alpha_k \frac{N_k}{D_k}$ in (4.4) as the scalar b tends to $+\infty$. For $b > \sigma\|y^\delta\|_{L^1}$, the inequality

$$\omega_k := \frac{b - F(\alpha_k)}{b - \sigma\varphi(\alpha_k)} > 0$$

holds true, and the fixed point iteration (4.4) can be rewritten as

$$\alpha_{k+1} = \omega_k T(\alpha_k) + (1 - \omega_k)\alpha_k.$$

Therefore, the fixed point iteration (4.4) can be regarded as a relaxation of the iteration $\alpha_{k+1} = T(\alpha_k)$ with a dynamically updated relaxation parameter ω_k. Note that both iterations have the same fixed point. Moreover, we have

$$\omega_k < 1 \text{ if and only if } \alpha_k F'(\alpha_k) > (\sigma - 1)\varphi(\alpha_k).$$

The next result shows the monotonicity of the operator T.

Lemma 4.4. *The operator T is monotone in the sense that if $0 < \alpha_0 < \alpha_1$, then*

$$T(\alpha_0) \le T(\alpha_1).$$

Proof. By Theorem 2.2, we have

$$\varphi(\alpha_0) \le \varphi(\alpha_1), \quad F'(\alpha_0) \ge F'(\alpha_1).$$

The result now follows directly from the definition of the operator T; see (4.5). ■

The next lemma shows the monotonic convergence of the sequence $\{T^k(\alpha_0)\}$. This iteration itself is of independent interest because of its simplicity and monotonic convergence, which is useful in practice.

Lemma 4.5. *For any initial guess α_0, the sequence $\{T^k(\alpha_0)\}$ is monotonic. Furthermore, it is monotonically decreasing (respectively, increasing) if $r(\alpha_0) > 0$ (respectively, $r(\alpha_0) < 0$).*

Proof. Let $\alpha_k = T^k(\alpha_0)$. Then we have

$$\begin{aligned}\alpha_{k+1} - \alpha_k &= T^{k+1}(\alpha_0) - T^k(\alpha_0)\\ &= (\sigma-1)\frac{\varphi(\alpha_k)}{F'(\alpha_k)} - (\sigma-1)\frac{\varphi(\alpha_{k-1})}{F'(\alpha_{k-1})}\\ &= (\sigma-1)\frac{\varphi(\alpha_k)F'(\alpha_{k-1}) - \varphi(\alpha_{k-1})F'(\alpha_k)}{F'(\alpha_{k-1})F'(\alpha_k)}\\ &= (\sigma-1)\frac{\varphi(\alpha_k)[F'(\alpha_{k-1}) - F'(\alpha_k)] + F'(\alpha_k)[\varphi(\alpha_k) - \varphi(\alpha_{k-1})]}{F'(\alpha_{k-1})F'(\alpha_k)}.\end{aligned}$$

By Theorem 2.2 both terms in square brackets have the same sign as $\alpha_k - \alpha_{k-1}$, which shows the desired monotonicity.

Let $r(\alpha)$ again denote the residual in the balancing equation as defined by (4.2). Now if $r(\alpha_0) > 0$ holds, by the definition of the function r, we have

$$\alpha_0 F'(\alpha_0) - (\sigma-1)\varphi(\alpha_0) > 0,$$

which after rearranging the terms gives

$$\alpha_0 > (\sigma-1)\frac{\varphi(\alpha_0)}{F'(\alpha_0)} = T(\alpha_0).$$

The second assertion follows directly from this inequality and the first statement. ■

Remark 4.6. The iteration produces a strictly monotonic sequence before reaching a solution to (4.1). If two consecutive steps coincide, then a solution has been found and we can stop the iteration. Upon reaching a solution α^*, there holds $r(\alpha^*) = 0$. In our subsequent analysis, this trivial case will be excluded.

Remark 4.7. The sequence $\{T^k(\alpha_0)\}$ can diverge to $+\infty$. This can be remedied by further regularizing the operator T by setting

$$T_r(\alpha) = (\sigma-1)\frac{\varphi(\alpha)}{F'(\alpha)+\gamma}$$

for some small number $\gamma > 0$. This preserves the monotonicity of the iterates and ensures the upper bound $(\sigma-1)\|y^\delta\|_{L^1}/\gamma$, which together with the trivial lower bound 0 and the monotonicity guarantees convergence of $T^k(\alpha_0)$. Moreover, the second part of Lemma 4.5 classifies the positive solutions of (4.1), and the sign of the function $r(\alpha)$ provides an explicit characterization for that classification. To illustrate this point, let α^* be a solution to (4.1). The iterate $T^k(\alpha_0)$ converges to α^* for α_0 in the neighborhood of α^* if and only if

$$r(\alpha)\begin{cases} <0, & \alpha \in (\alpha^*-\varepsilon, \alpha^*),\\ >0, & \alpha\in(\alpha^*, \alpha^*+\varepsilon),\end{cases}$$

for sufficiently small ε, and it diverges from α^* for α_0 in the neighborhood of α^* if and only if

$$r(\alpha)\begin{cases} >0, & \alpha \in (\alpha^*-\varepsilon, \alpha^*),\\ <0, & \alpha\in(\alpha^*, \alpha^*+\varepsilon).\end{cases}$$

In the case that $r(\alpha)$ has the same sign on $(\alpha^*-\varepsilon,\alpha^*)$ and $(\alpha^*,\alpha^*+\varepsilon)$, the iterate can converge to α^* only for α_0 in its one-sided neighborhood. If $r(\alpha)>0$ on $(\alpha^*-\varepsilon,\alpha^*)\cup(\alpha^*,\alpha^*+\varepsilon)$, then the iterates $T^k(\alpha_0)$ converge if $r(\alpha_0)>0$, and vice versa for $r<0$.

We shall also need a "sign-preserving" property of the operator T: the function $r(\alpha)$ cannot vanish on the open interval between α_0 and the limit α^* of the sequence $\{T^k(\alpha_0)\}$.

Lemma 4.8. *For any α_0 such that $\{T^k(\alpha_0)\}$ converges to α^*, the function $r(\alpha)$ does not vanish on the open interval* $(\min(\alpha_0,\alpha^*),\max(\alpha_0,\alpha^*))$.

Proof. Without loss of generality we assume that $\alpha_0<T(\alpha_0)$, as the other case can be treated similarly. Assume that the assertion is false. Then there exists an $\alpha\in(\alpha_0,\alpha^*)$ such that $r(\alpha)=0$; i.e., $T(\alpha)=\alpha$. By Lemma 4.5, there exists some $k\in\mathbb{N}$ such that

$$T^k(\alpha_0)\le\alpha<T^{k+1}(\alpha_0).$$

However, by Lemma 4.4, we have

$$\alpha<T^{k+1}(\alpha_0)\le T(\alpha)\le T^{k+2}(\alpha_0),$$

which is a contradiction to $T(\alpha)=\alpha$. ∎

We note that in Lemma 4.8, α^* can take the value $+\infty$; i.e., the convergence can be understood in a generalized sense. Using Lemmas 4.5 and 4.8, we can now state a monotone convergence result for the fixed point algorithm.

Theorem 4.9. *Assume that α_0 satisfies $r(\alpha_0)>0$. Then the sequence $\{\alpha_k\}$ generated by the fixed point iteration (4.4) is monotonically decreasing and converges to a solution of (4.1).*

Proof. Since $r(\alpha_0)>0$, we have

$$0<\omega_0=\frac{b-F(\alpha_0)}{b-F(\alpha_0)+r(\alpha_0)}<\frac{b-F(\alpha_0)}{b-F(\alpha_0)}=1. \tag{4.6}$$

Due to Lemma 4.5, the auxiliary sequence $\{T^k(\alpha_0)\}$ is monotonically decreasing and bounded below by zero and thus converges to some α^*. In particular, $T(\alpha_0)<\alpha_0$, which together with (4.6) implies that

$$\alpha_1=\omega_0T(\alpha_0)+(1-\omega_0)\alpha_0\in(T(\alpha_0),\alpha_0). \tag{4.7}$$

Consequently we have $\alpha^*\le T(\alpha_0)<\alpha_1<\alpha_0$. Lemmas 4.8 and (4.7) imply that $r(\alpha_1)>0$. Now assume that α_k generated by the algorithm satisfies $r(\alpha_k)>0$. Then by the definition of the operator T, we have $T(\alpha_k)<\alpha_k$. Appealing to the preceding arguments, we have $\omega_k\in(0,1)$ and

$$\alpha_{k+1}=\omega_kT(\alpha_k)+(1-\omega_k)\alpha_k\in(T(\alpha_k),\alpha_k),$$

and thus $\alpha^*\le T(\alpha_k)<\alpha_{k+1}<\alpha_k$. This shows that the sequence $\{\alpha_k\}_{k=0}^\infty$ is monotonically decreasing and bounded from below by α^*, and thus converges to some $\alpha^\dagger$. Upon convergence, the limit $\alpha^\dagger$ satisfies

$$\alpha^\dagger=\alpha^\dagger\frac{(\alpha^\dagger F'(\alpha^\dagger))^2+(\sigma-1)\varphi(\alpha^\dagger)[b-F(\alpha^\dagger)-\alpha^\dagger F'(\alpha^\dagger)]}{\alpha^\dagger F'(\alpha^\dagger)(b-\sigma\varphi(\alpha^\dagger))}$$

due to the continuous dependence of $F(\alpha)$, $F'(\alpha)$, and $\varphi(\alpha)$ on α (Theorem 2.2). Simplifying the equation shows that $\alpha^\dagger$ is a solution to (4.1). Moreover, from Lemma 4.8 we deduce that there is no other solution to (4.1) in the open interval (α^*, α_0), and thus that $\alpha^\dagger = \alpha^*$. ■

Next we address the convergence behavior of the algorithm for the case $r(\alpha_0) < 0$.

Theorem 4.10. *Assume that the initial guess α_0 satisfies $r(\alpha_0) < 0$. Then either the sequence $\{\alpha_k\}$ generated by the fixed point iteration (4.4) is monotonically increasing or there exists some $k_0 \in \mathbb{N}$ such that $r(\alpha_k) \geq 0$ for all $k \geq k_0$.*

Proof. Since $r(\alpha_0) < 0$, we have

$$\omega_0 = \frac{b - F(\alpha_0)}{b - F(\alpha_0) + r(\alpha_0)} > \frac{b - F(\alpha_0)}{b - F(\alpha_0)} = 1.$$

From Lemma 4.5, we deduce that $\alpha_0 < T(\alpha_0)$ and, moreover, the auxiliary sequence $\{T^k(\alpha_0)\}$ is monotonically increasing. Consequently, we have

$$\alpha_1 = \omega_0 T(\alpha_0) + (1 - \omega_0)\alpha_0 = T(\alpha_0) + (\omega_0 - 1)(T(\alpha_0) - \alpha_0) > T(\alpha_0).$$

In particular, $\alpha_0 < \alpha_1$. Now α_1 can satisfy either $r(\alpha_1) < 0$ or $r(\alpha_1) > 0$. For the latter case, we appeal to Theorem 4.9, and we have $r(\alpha_k) \geq 0$ for $k \geq 1$. Otherwise $r(\alpha_1) < 0$ and hence as above $\alpha_1 < T(\alpha_1) < \alpha_2$. The claim now follows by induction. ■

We can now show the convergence of the fixed point iteration.

Proof of Theorem 4.3. By Theorem 4.10, we have that the sequence $\{\alpha_k\}$ is monotonically increasing or that there exists some $k_0 \in \mathbb{N}$ such that $r(\alpha_k) \geq 0$ for all $k \geq k_0$. Moreover, by the definition of a regular attractor, $r(\alpha) > 0$ holds for all $\alpha \in (\alpha^*, \alpha^* + \varepsilon)$ for some $\varepsilon > 0$, and by Lemma 4.2, we have

$$\lim_{\alpha \to +\infty} r(\alpha) = -(\sigma - 1) \lim_{\alpha \to +\infty} F(\alpha) = -(\sigma - 1)\|y^\delta\|_{L^1} < 0.$$

Therefore, by the continuity of the function $r(\alpha)$ (cf. Theorem 2.2), there exists at least one solution to (4.1) on the interval $(\alpha^*, +\infty)$. Denote the smallest solution of (4.1) larger than α^* by α^{**}, and set $c = \frac{2}{b - \|y^\delta\|_{\mathrm{L}^1}}$. Since the function $r(\alpha)$ is continuous and $r(\alpha^*) = 0$, for any $\delta > 0$ we can choose ε such that

$$|r(\alpha)| < \delta \quad \text{for all } \alpha \in (\alpha^* - \varepsilon, \alpha^* + \varepsilon).$$

We now choose δ such that $\delta < \min\{\frac{\alpha^{**} - \alpha^*}{c\alpha^*}, \frac{b - \|y^\delta\|_{\mathrm{L}^1}}{2}\}$ and pick ε accordingly. Consequently, we have for $\alpha \in (\alpha^* - \varepsilon, \alpha^* + \varepsilon)$

$$\begin{aligned} \omega(\alpha) - 1 &= \frac{b - F(\alpha)}{b - F(\alpha) + r(\alpha)} - 1 = \frac{-r(\alpha)}{b - F(\alpha) + r(\alpha)} \\ &\leq \frac{-r(\alpha)}{b - \|y^\delta\|_{\mathrm{L}^1} - \delta} \leq \frac{\delta}{b - \|y^\delta\|_{\mathrm{L}^1} - \delta} < c\delta \end{aligned}$$

and, in particular,

$$\alpha_0 < \alpha_1 = T(\alpha_0) + (\omega_0 - 1)(T(\alpha_0) - \alpha_0) < T(\alpha_0) + c\delta T(\alpha_0) < T(\alpha_0) + c\delta\alpha^*,$$

where we have used that $\omega_0 > 1$ and $\alpha^* > T(\alpha_0) > \alpha_0$.

This implies $\alpha_1 < T(\alpha_0) + c\delta\alpha^* < \alpha^{**}$. Therefore, we have either $r(\alpha_1) < 0$ with $\alpha_0 < \alpha_1 < \alpha^*$ or $r(\alpha_1) \geq 0$ with $\alpha^* \leq \alpha_1 < \alpha^{**}$. In the latter case, the convergence of α_k to α^* follows directly from Theorem 4.9, and thus it suffices to consider the former case. We proceed by induction and assume that α_k satisfies $r(\alpha_k) < 0$. By repeating the preceding argument, we deduce that $\alpha_k < \alpha_{k+1}$. Again, either $r(\alpha_{k+1}) < 0$ and convergence to α^* follows, or $r(\alpha_{k+1}) \geq 0$ and we can proceed as before. If $r(\alpha_k) < 0$ for all k, then the sequence $\{\alpha_k\}$ is monotonically increasing and bounded from above by α^*. It thus converges to some $\alpha^\dagger$. Analogously to Theorem 4.9, we can show that $\alpha^\dagger$ is a solution to (4.1). The conclusion $\alpha^\dagger = \alpha^*$ now follows from Lemma 4.8. ∎

Remark 4.11. Note that our derivations are valid as well for other Tikhonov functionals, e.g., L^2 data fitting with total variation regularization. A differentiability result of the cost functional with respect to the regularization parameter as in Theorem 2.2 is an essential ingredient of this approach.

4.2. Choice of β by a path-following method. Since the introduction of the H^1 smoothing alters the structure of the primal problem (cf. Remark 3.4), the value of β should be as small as possible. However, the regularized dual problem ($\mathcal{P}^*_{\beta,c}$) becomes increasingly ill-conditioned as $\beta \to 0$ due to the ill-conditioning of the discretized operator KK^* and the rank-deficiency of the diagonal matrix corresponding to the (discrete) active set. Therefore, the respective system matrix will eventually become numerically singular for some small $\beta > 0$.

One possible remedy is a path-following strategy: starting with a large β_0, we reduce its value as long as the system is still solvable and take the solution corresponding to the smallest such value. The question remains how to automatically select the stopping index without a priori knowledge or expensive computations for estimating the condition number by, e.g., singular value decomposition. To select an appropriate stopping index, we exploit the structure of the (infinite-dimensional) box constraint problem: the correct solution should be nearly feasible for c sufficiently large, and therefore the discretized solution should satisfy $\|p_\beta\|_\infty \leq \tau$ for some $\tau \approx 1$. Recall that for the linear system corresponding to (3.7), the right-hand side f satisfies $\|f\|_\infty \approx c \gg 1$, which should be balanced by the diagonal matrix $c\chi_{\mathcal{A}}$ in order for the solution to be feasible. If the matrix is nearly singular, this will no longer be the case, and the solution p blows up and consequently violates the feasibility condition, i.e., $\|p_\beta\|_\infty \gg 1$. Once this happens, we take the last iterate which is still (close to) feasible and return it as the solution.

This whole procedure is summarized in Algorithm 3. Here, β_{min} can be set to machine precision, and β_0 may be initialized with 1.

5. Numerical examples. We now present some numerical results to illustrate salient features of the semismooth Newton method as well as the adaptive regularization parameter choice rules. The first two benchmark examples, `deriv2` and `heat`, are taken from [16] and are available in the companion MATLAB package Regularization Tools (http://www2.imm.dtu.dk/~pch/Regutools/). The third example is an inverse source problem for the two-dimensional Laplace operator.

Unless otherwise stated the first two examples are discretized into linear systems of size $n = 100$, and the parameters are set as follows: in the fixed point algorithm, Algorithm 2,

Algorithm 3. Path-following method to solve L^1 data fitting problem for fixed α.

1: Set $k = 0$, choose $\beta_0 > 0$, $q < 1$, $\beta_{min} > 0$, $\tau \gg 1$.
2: **repeat**
3: Compute $p_{\beta_{k+1}}$ using semismooth Newton method with $p^0 = p_{\beta_k}$ (Algorithm 1).
4: Set $\beta_{k+1} = q \cdot \beta_k$.
5: Set $k = k + 1$.
6: **until** $\|p_{\beta_k}\|_{L^\infty} > \tau$ or $\beta_k < \beta_{min}$
7: **return** $x = \frac{1}{\alpha} K^* p_{\beta_{k-1}}$

$\sigma = 1.05$, $\alpha = 0.01$, and $b = \|y^\delta\|_{L^1}$; in the path-following algorithm, Algorithm 3, $\beta_0 = 1$, $q = \frac{1}{5}$, $\beta_{min} = 10^{-16}$ (floating point machine precision), and $\tau = 10$; in the semismooth Newton algorithm, Algorithm 1, $k_{max} = 10$. The penalty parameter was chosen as $c = 10^9$.

We compare the performance of the proposed method with two other algorithms: the iteratively regularized least squares (IRLS) method [32, 29] and a splitting approach using an alternating direction minimization (ADM) method [34]. Since these algorithms were not originally designed for the L^1 model under consideration and numerical implementations are not freely available, we have adapted the algorithms. The details and their respective parameters are described in Appendix C. All parameters in the benchmark algorithms were chosen for optimal performance with the reconstruction error being the same as that from the path-following semismooth Newton algorithm.

The noisy data y^δ is generated pointwise by setting

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

where ξ follows a normal distribution with mean 0 and standard deviation 1, and $\varepsilon = \epsilon \max |y^\delta|$ with ϵ being the relative noise level. Unless otherwise stated we set $r = 0.3$ and $\epsilon = 1$. All computations were performed with MATLAB version 2009b on a single core of a 2.4 GHz workstation with 4 GByte RAM. MATLAB codes implementing the algorithm presented in this paper can be downloaded from http://www.uni-graz.at/~clason/codes/l1fitting.zip.

5.1. Example 1: `deriv2`. This example involves computing the second derivative of a function; i.e., the operator K is a Fredholm integral operator of the first kind:

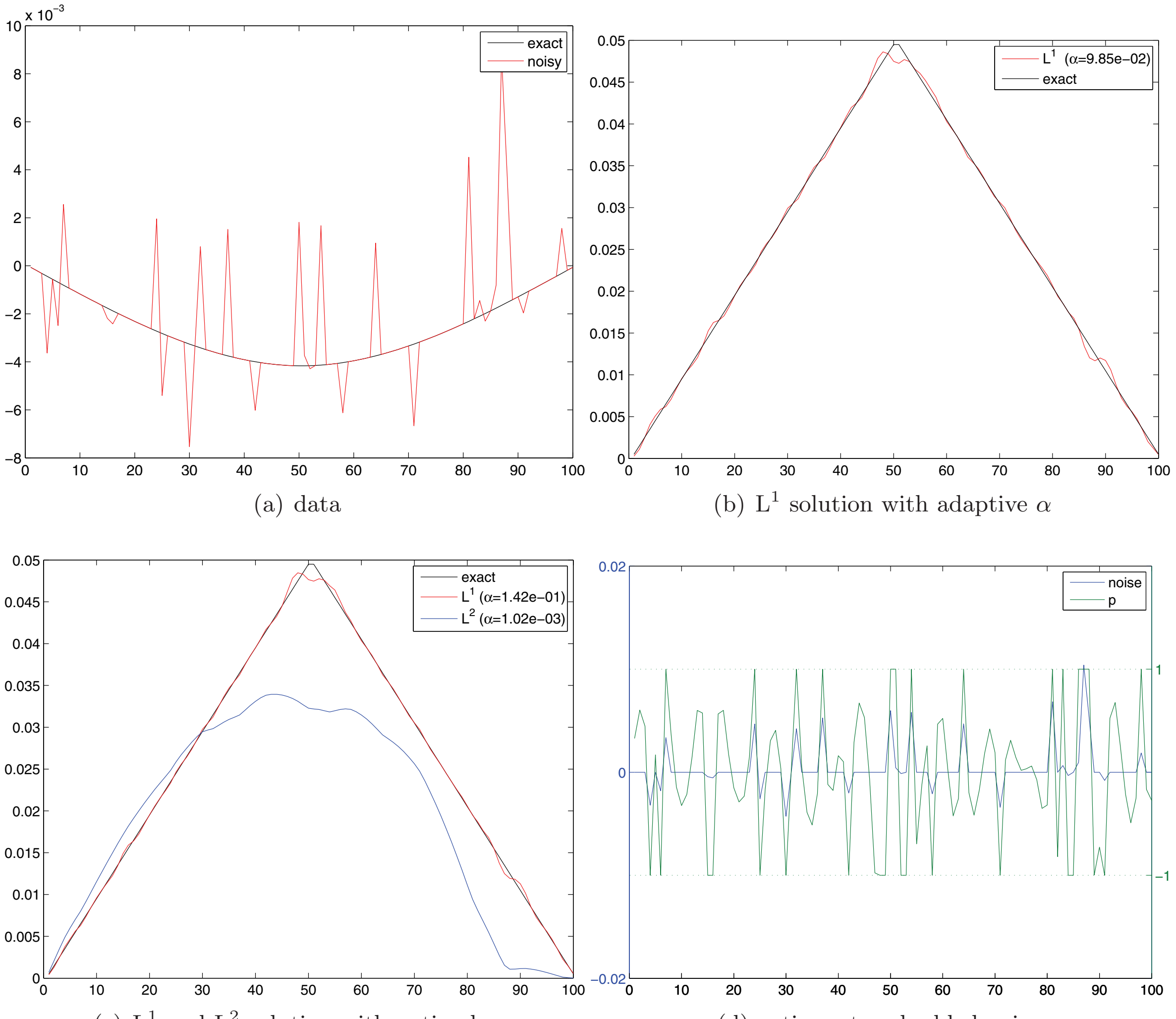
$$(Kx)(t) = \int_0^1 k(s,t)x(s)\,ds.$$

Here, the kernel $k(s,t)$ and the exact solution $x(t)$ are given by

$$k(s,t) = \begin{cases} s(t-1), & s < t, \\ t(s-1), & s \geq t, \end{cases} \qquad x(t) = \begin{cases} t, & t < \frac{1}{2}, \\ 1-t & \text{otherwise,} \end{cases}$$

respectively. The problem is discretized using a Galerkin method. This problem is mildly ill-posed, and the condition number of the matrix is 1.216×10^4.

A typical realization of noisy data is displayed in Figure 1(a). The corresponding reconstruction with the adaptively chosen parameter $\alpha_b = 9.854 \times 10^{-2}$ is shown in Figure 1(b)



(a) data (b) L^1 solution with adaptive α

(c) L^1 and L^2 solution with optimal α (d) active set and added noise

Figure 1. *Results for test problem* `deriv2`.

and agrees very well with the exact solution almost everywhere. The convergence of the fixed point algorithm is fairly fast, usually within two iterations. For comparison, we also computed the optimal value α_{opt} of the regularization parameter by sampling α at 100 points uniformly distributed over the range $[10^{-5}, 1]$ in a logarithmic scale. This yields $\alpha_{\text{opt}} = 1.418 \times 10^{-1}$, which is very close to α_{b}. The result, shown in Figure 1(c), is practically identical to that of the adaptive strategy. The optimal reconstruction using L^2 data fitting, also shown in Figure 1(c), is far inferior.

In Figure 1(d), we show the dual solution p and the data noise. Observe that p serves as a good indicator of noise, as both the location and the sign of nonzero noise components are accurately detected. This numerically corroborates Corollary 2.8.

To illustrate the performance of the semismooth Newton (SSN) method, we compare the computing time and reconstruction error e, defined as $e = \|x_\alpha - x^\dagger\|_{\mathrm{L}^2}$, for different problem

Table 1

Computing time (in seconds) and reconstruction errors for the SSN vs. IRLS and ADM methods (`deriv2`*).*

n	50	100	200	400	800	1600
t_{ssn}	0.0107	0.0223	0.0846	0.4430	3.3032	24.8826
t_{irls}	0.0131	0.0497	0.3079	2.3132	17.3116	132.2214
t_{adm}	0.0720	0.1801	0.6400	3.5907	21.5150	152.5250
e_{ssn}	1.0847e-2	3.4164e-3	7.7573e-4	5.9145e-4	2.9626e-4	3.1714e-4
e_{irls}	1.0753e-2	3.5412e-3	8.0371e-4	5.8767e-4	2.6148e-4	2.0400e-4
e_{adm}	1.0962e-2	3.8802e-3	1.2056e-3	8.3158e-4	4.7506e-4	3.9425e-4

Table 2

Iterates in the path-following method for β (`deriv2`*).*

β	Iterations	e	$F(x)$	$\|\nabla p\|_{L^2}$
1.000e+0	2	2.860e-2	2.794e-3	2.589e-3
4.000e-2	2	2.362e-2	2.438e-3	5.155e-2
1.600e-3	2	7.729e-3	1.302e-3	2.148e-1
6.400e-5	2	7.926e-3	1.096e-3	5.760e-1
2.560e-6	7	2.096e-2	1.074e-3	4.240e+0
1.024e-7	6	9.300e-3	8.986e-4	7.423e+0
4.096e-9	4	3.681e-3	8.646e-4	6.999e+0
1.638e-10	2	1.448e-3	8.625e-4	5.994e+0
6.554e-12	3	5.334e-4	8.622e-4	6.389e+0
1.311e-12	10	3.792e-4	8.622e-4	6.884e+0

sizes and $r = 0.7$ (averaged over ten runs with different noise realizations) with that of the IRLS and ADM methods in Table 1. For all problem sizes under consideration, the SSN method is significantly faster than the IRLS and ADM methods. The results of all three methods are very close to each other, with the ADM method showing less accuracy.

The convergence behavior of the path-following method is shown in Table 2. For moderate values of β, the SSN method exhibits superlinear convergence as indicated by the convergence after two iterations. This property is lost when β becomes too small, but the method still converges after very few iterations due to our path-following strategy. Interestingly, while the functional value F keeps on decreasing as β decreases, the error e experiences some transition at $\beta = 2.560 \times 10^{-6}$. This might be attributed to the change from the dominance of the H^1 term (β) to that of the L^2 term (α) in the regularized dual formulation ($\mathcal{P}^*_{\beta,c}$).

Finally, we compare the parameters chosen by the balancing principle (BP) with those obtained from the discrepancy principle (DP) and the optimal choice. The chosen parameters α and corresponding errors e for different noise parameters (r, ϵ) are shown in Table 3, where the subscripts b, d, and opt refer to the BP, the DP, and the optimal choice, respectively. For the DP, we utilize the exact noise level δ. We observe that the results by the BP and DP are largely comparable in terms of the error e despite the discrepancies in the regularization parameter. Also, the regularization parameter determined by the BP increases at the same rate as the noise level δ, whereas the one determined by the DP seems largely independent of δ, especially for fixed r. This causes slight underregularization in the BP for low noise levels. Nonetheless, the noise level δ is estimated very accurately by δ_{b}. Interestingly, we observe

Table 3
Comparison of BP with DP (`deriv2`).

(r,ϵ)	δ	$\delta_{\rm b}$	$\alpha_{\rm b}$	$\alpha_{\rm d}$	$\alpha_{\rm opt}$	$e_{\rm b}$	$e_{\rm d}$	$e_{\rm opt}$
(0.3, 0.1)	7.291e-5	7.284e-5	8.656e-3	1.612e-1	2.056e-1	2.849e-3	2.240e-4	1.949e-4
(0.3, 0.3)	2.187e-4	2.186e-4	2.621e-2	1.125e-1	2.056e-1	9.368e-4	2.392e-4	1.949e-4
(0.3, 0.5)	3.645e-4	3.644e-4	4.372e-2	1.266e-1	2.056e-1	5.971e-4	2.358e-4	1.949e-4
(0.3, 0.7)	5.104e-4	5.101e-4	6.123e-2	1.458e-1	2.056e-1	3.695e-4	2.304e-4	1.949e-4
(0.3, 0.9)	6.562e-4	6.557e-4	7.874e-2	1.620e-1	2.056e-1	3.851e-4	2.035e-4	1.949e-4
(0.1, 0.3)	3.901e-5	3.901e-5	4.677e-3	9.543e-2	1.707e-1	7.963e-4	4.274e-5	3.904e-5
(0.3, 0.3)	2.187e-4	2.186e-4	2.621e-2	1.125e-1	2.056e-1	9.368e-4	2.392e-4	1.949e-4
(0.5, 0.3)	4.128e-4	4.125e-4	4.930e-2	3.048e-1	3.593e-1	1.909e-3	3.811e-4	3.711e-4
(0.7, 0.3)	5.822e-4	5.798e-4	6.730e-2	4.065e-1	3.944e-1	5.438e-3	1.791e-3	1.794e-3
(0.9, 0.3)	8.238e-4	7.991e-4	6.602e-2	5.147e-1	4.750e-1	1.797e-2	2.927e-3	2.859e-3

that the two factors of the noise, i.e., r and ϵ, have drastically different effects on the inverse solution: the results seem relatively independent of the ϵ for fixed r, whereas for fixed ϵ, the error e deteriorates rapidly as the noise percentage r increases. In particular, $\alpha_{\rm opt}$ seems relatively independent of ϵ for fixed r, and increases at the rate of r for fixed ϵ. Finally, with the knowledge of the exact noise level δ, the DP achieves optimal convergence rate in that its error is roughly the same as that with the optimal parameter.

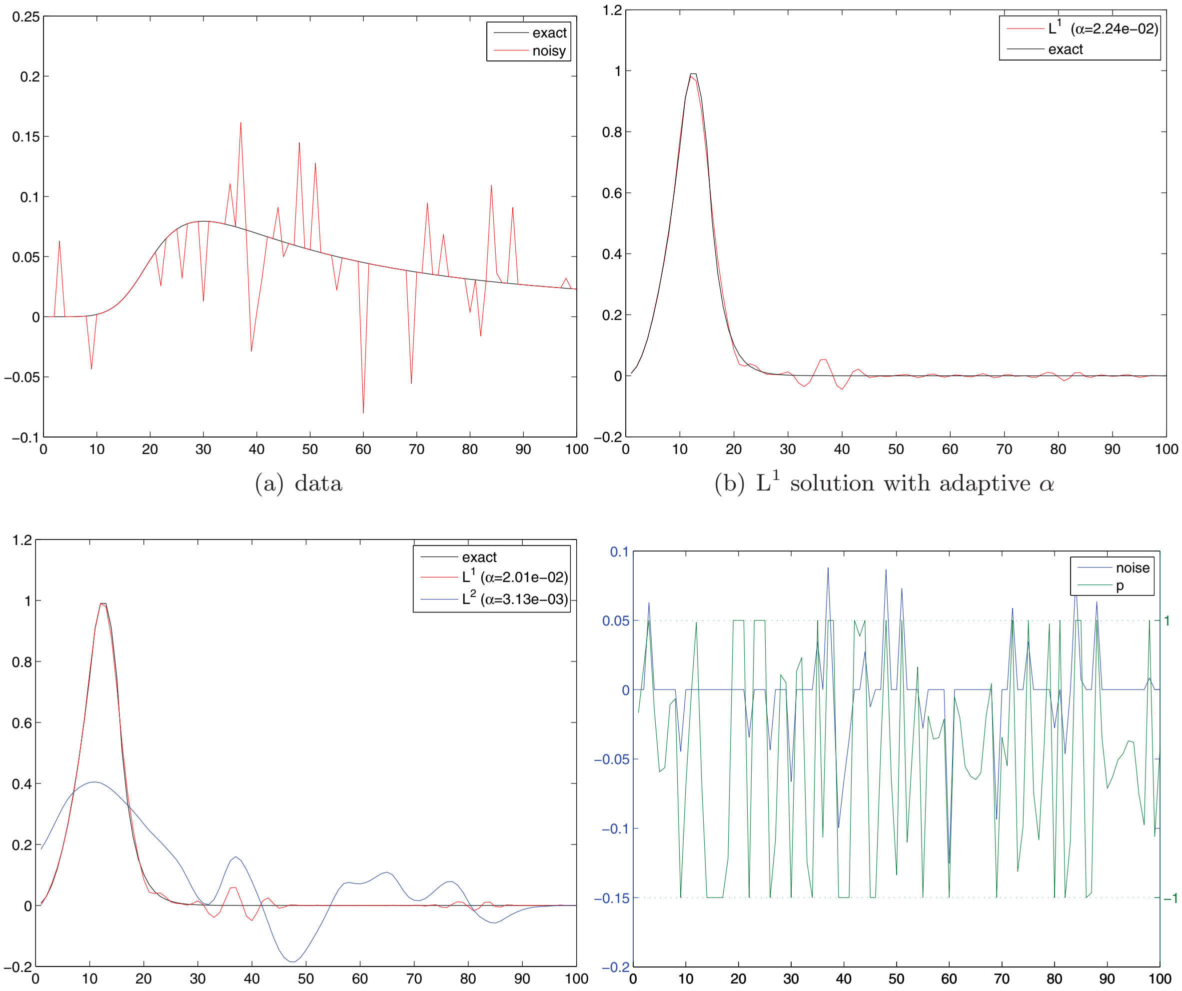
5.2. Example 2: `heat`. This example is an inverse heat conduction problem, posed as a Volterra integral equation of the first kind. The kernel $k(s,t)$ and the exact solution $x(t)$ are given by

$$k(s,t) = \frac{(s-t)^{-\frac{3}{2}}}{2\sqrt{\pi}} e^{-\frac{1}{4(s-t)}}, \qquad x(t) = \begin{cases} 75t^2, & u \leq 2, \\ \frac{3}{4} + (u-2)(3-u), & 2 < u \leq 3, \\ \frac{3}{4} e^{-2(u-3)}, & 3 < u \leq 10, \\ 0 & \text{otherwise}, \end{cases}$$

with $u = 20t$ and the integration interval $[0,1]$. The integral equation is discretized using collocation and the midpoint rule. This problem is exponentially ill-posed, and the condition number is 8.217×10^{36}.

The results are given in Figure 2. Again, the reconstruction with automatically chosen parameter $\alpha_{\rm b} = 2.239 \times 10^{-2}$ is very close to the exact solution and to the optimal reconstruction with $\alpha_{\rm opt} = 2.009 \times 10^{-2}$, while the L^2 reconstruction is vastly inferior. The performance and convergence of the path-following SSN method are similar to those in Example 1 (cf. Tables 4 and 5). Also, the adaptive strategy yields results comparable with those of the discrepancy principle and optimal choice; see Table 6.

The convergence of the fixed point algorithm is now even faster: the convergence is achieved in one iteration. This may be attributed to the fact that the spectrum of the matrix spans a much broader range because of its exponential ill-posedness, and thus the residual is less sensitive to the variation of the regularization parameter. This consequently accelerates the convergence of the fixed point algorithm. Again the noise level is estimated very accurately, while the chosen regularization parameter is now closer to the optimal one compared to Example 1 and sometimes even outperforms the DP with exact noise level (cf. Table 6).



(a) data

(b) L^1 solution with adaptive α

(c) L^1 and L^2 solution with optimal α

(d) active set and added noise

Figure 2. *Results for test problem* `heat`.

5.3. Example 3: Inverse source problem in two dimensions. As a two-dimensional test problem, we consider the inverse source problem for the Laplacian on the unit square $[0,1]^2$ with a homogeneous Dirichlet boundary condition, i.e., $K = (-\Delta)^{-1}$. The exact solution $x(s,t)$ is given by (cf. Figure 3(a))

$$x(s,t) = \begin{cases} \sin 2\pi(s-\frac{1}{4}) \sin 2\pi(t-\frac{1}{4}), & |s-\frac{1}{2}| \leq \frac{1}{4},\ |t-\frac{1}{2}| \leq \frac{1}{4}, \\ 0 & \text{otherwise}. \end{cases}$$

The problem is discretized on a 64×64 mesh using the standard five-point stencil, resulting in a linear system of size $n = 4096$. The problem is mildly ill-posed, and the estimated condition number is 2.689×10^3. The CPU time for one reconstruction using Algorithm 3 is 6.5 seconds.

The noisy data for this problem is given in Figure 3(b). The corresponding numerical solution of the inverse source problem, shown in Figure 3(c), is a good approximation of the

Table 4

Computing time (in seconds) and reconstruction errors for the SSN vs. IRLS and ADM methods (`heat`).

n	50	100	200	400	800	1600
t_{ssn}	0.0089	0.0197	0.0808	0.4365	4.1769	27.5007
t_{irls}	0.0131	0.0515	0.3254	2.3720	18.6916	132.5247
t_{adm}	0.0729	0.1826	0.7952	4.3687	25.0411	154.9379
e_{ssn}	1.4743	2.0263e-1	1.3621e-1	1.4336e-1	1.3319e-1	1.3615e-1
e_{irls}	1.4610	1.7857e-1	1.3587e-1	1.4334e-1	1.3322e-1	1.3618e-1
e_{adm}	1.4657	1.7913e-1	1.3595e-1	1.4333e-1	1.3335e-1	1.3638e-1

Table 5

Iterates in the path-following method for β (`heat`).

β	Iterations	e	$F(x)$	$\|\nabla p\|_{\mathrm{L}^2}$
1.000e+0	2	2.248e-1	3.274e-2	2.951e-2
4.000e-2	2	1.855e-1	1.963e-2	1.090e-1
1.600e-3	2	1.610e-1	1.713e-2	2.191e-1
6.400e-5	2	1.556e-1	1.737e-2	1.408e+0
2.560e-6	6	2.250e-1	1.644e-2	5.852e+0
1.024e-7	4	7.042e-2	1.435e-2	6.436e+0
4.096e-9	3	2.043e-2	1.415e-2	6.902e+0
1.638e-10	10	1.563e-2	1.414e-2	7.361e+0
3.277e-11	10	1.546e-2	1.414e-2	8.960e+0

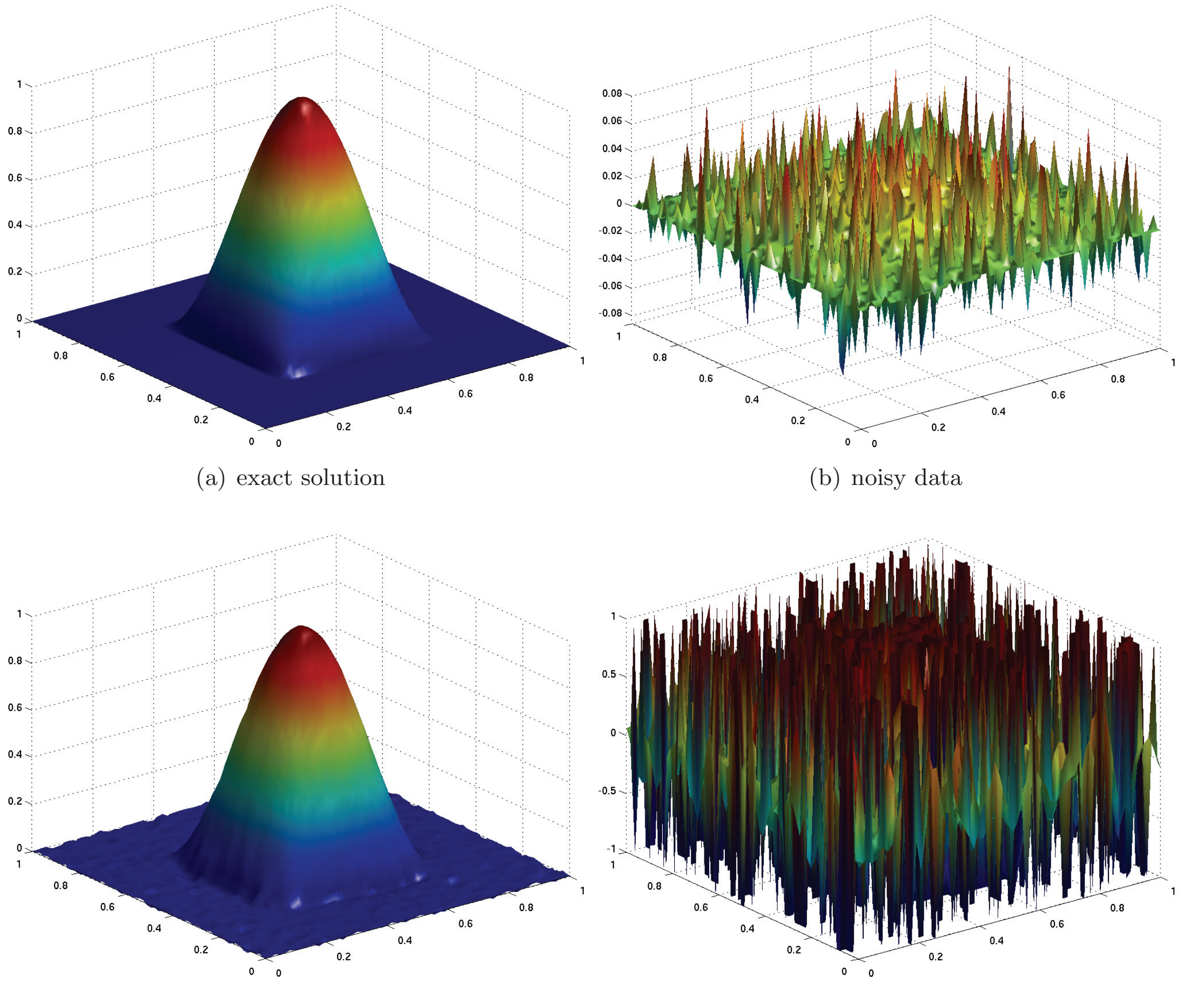
Table 6

Comparison of BP with DP (`heat`).

(r, ϵ)	δ	δ_{b}	α_{b}	α_{d}	α_{opt}	e_{b}	e_{d}	e_{opt}
(0.3, 0.1)	1.390e-3	1.335e-3	1.402e-3	1.910e-2	1.830e-2	1.860e-1	2.021e-2	2.026e-2
(0.3, 0.3)	4.170e-3	4.155e-3	6.638e-3	1.906e-2	1.830e-2	4.515e-2	2.021e-2	2.026e-2
(0.3, 0.5)	6.950e-3	6.939e-3	1.135e-2	1.908e-2	1.830e-2	2.706e-2	2.022e-2	2.026e-2
(0.3, 0.7)	9.731e-3	9.719e-3	1.604e-2	2.045e-2	1.830e-2	2.103e-2	2.032e-2	2.026e-2
(0.3, 0.9)	1.251e-2	1.249e-2	2.083e-2	1.909e-2	1.830e-2	2.037e-2	2.022e-2	2.026e-2
(0.1, 0.3)	7.438e-4	7.439e-4	1.227e-3	1.374e-2	7.742e-4	1.727e-3	2.736e-3	5.980e-4
(0.3, 0.3)	4.170e-3	4.155e-3	6.638e-3	1.906e-2	1.830e-2	4.515e-2	2.021e-2	2.026e-2
(0.5, 0.3)	7.871e-3	7.799e-3	1.225e-2	4.718e-2	3.199e-2	5.635e-2	4.026e-2	3.772e-2
(0.7, 0.3)	1.110e-2	1.074e-2	2.254e-2	3.995e-2	7.390e-2	1.118e-1	1.130e-1	1.034e-1
(0.9, 0.3)	1.570e-2	1.470e-2	2.247e-2	1.553e-1	5.094e-2	1.662e-1	1.487e-1	1.388e-1

exact one. Note in particular that the magnitude of the peak is correctly recovered. The L^2 norm of the reconstruction error is $e = 7.526 \times 10^{-3}$. The fixed point algorithm converges in three iterations to the value $\alpha_{\mathrm{b}} = 8.797 \times 10^{-3}$. The estimated noise level is $\delta_{\mathrm{b}} = 5.475 \times 10^{-3}$, which is very close to the exact one $\delta = 5.490 \times 10^{-3}$. For completeness, we show also the dual solution in Figure 3(d).

6. Conclusion. We have presented a semismooth Newton method for the numerical solution of inverse problems with L^1 data fitting together with an adaptive method for the choice of regularization parameters. The main advantage of the adaptive strategy is that no knowledge of the noise level is necessary, and it can, in fact, provide an excellent estimate of the



(a) exact solution (b) noisy data

(c) L^1 reconstruction (d) dual solution p

Figure 3. *Results for two-dimensional inverse source problem.*

noise level. This is important for some practical applications. The convergence of the fixed point iteration was analyzed. In practice it is usually achieved within two or three iterations. The value for the regularization parameter obtained by the proposed technique based on the balancing principle derived from the model function approach was always fairly close to the optimal one.

Similarly, the semismooth Newton method allows an efficient numerical solution of L^1 data fitting problems. In our examples, the proposed method was significantly faster than the iteratively reweighted least squares method and the ADM method. In practice, implementations of the latter two methods would include early termination criteria based, e.g., on the norm of the difference of consecutive iterates, which would accelerate the methods, although at the risk of loss of accuracy. Similar strategies could, of course, also be applied in our method.

On the other hand, we consider the fact that the semismooth Newton method performs very well without the introduction of heuristic tolerance-based termination criteria to be one of its main advantages.

The path-following strategy proved to be an efficient and simple strategy for achieving the conflicting goals of minimizing the effect of the additional smoothing term on the primal problem and maintaining the numerical stability of the dual problem.

The proposed approach can also be extended to more general functionals (e.g., those involving total variation terms), which will be the focus of a subsequent work.

Appendix A. Convergence of smoothing for penalized box constraints. Here we show the convergence of the solutions of ($\mathcal{P}^*_{\beta,c}$) as β tends to zero to a solution of

$$(\mathcal{P}^*_c) \quad \min_{p\in \mathrm{L}^2} \frac{1}{2\alpha}\, \|K^*p\|^2_{\mathrm{L}^2} \;-\; \langle p, y^\delta\rangle_{\mathrm{L}^2} \;+\; \frac{1}{2c}\, \|\max(0, c(p-1))\|^2_{\mathrm{L}^2} \;+\; \frac{1}{2c}\, \|\min(0, c(p+1))\|^2_{\mathrm{L}^2}\,.$$

For this problem, the solution might be nonunique if the operator K is not injective. Again, the functional in ($\mathcal{P}^*_c$) is convex, and so is the set of all minimizers, and thus, if the problem has a solution in H^1, there exists an element with minimal H^1 seminorm, denoted by $p^\dagger$.

Theorem A.1. *Let $\{\beta_n\}$ be a vanishing sequence. Then the sequence of minimizers $\{p_{\beta_n,c}\}$ of ($\mathcal{P}^*_{\beta,c}$) has a subsequence converging weakly to a minimizer of problem ($\mathcal{P}^*_c$). If the operator K is injective or there exists a unique $p^\dagger$ as defined above, then the whole sequence converges weakly to $p^\dagger$.*

Proof. Let $\mathcal{A}^+ = \{x \in \Omega : p(x) > 1\}$ and $\mathcal{A}^- = \{x \in \Omega : p(x) < -1\}$. We denote the positive and negative parts of a function p by $(p)^+$ and $(p)^-$, respectively. The functional in ($\mathcal{P}^*_{\beta,c}$) can then be written as

$$\frac{1}{2\alpha}\|K^*p\|^2_{\mathrm{L}^2} - \langle p, y^\delta\rangle_{\mathrm{L}^2} + \frac{\beta}{2}\|\nabla p\|^2_{\mathrm{L}^2} + \frac{1}{2c}(\|c(p-1)^+\|^2_{\mathrm{L}^2} + \|c(p+1)^-\|^2_{\mathrm{L}^2}).$$

Now observe that

$$\begin{aligned}\|(p-1)^+\|^2_{\mathrm{L}^2} = \int_\Omega ((p-1)^+)^2\,dx &= \int_{\mathcal{A}^+} p^2 - 2p + 1\,dx \\ &= \|p\|^2_{\mathrm{L}^2(\mathcal{A}^+)} + |\mathcal{A}^+| - 2\int_{\mathcal{A}^+} p\,dx.\end{aligned}$$

Note also that

$$\int_{\mathcal{A}^+} p\,dx \le \|p\|_{\mathrm{L}^2(\mathcal{A}^+)}|\mathcal{A}^+|^{1/2} \le \frac{1}{4}\|p\|^2_{\mathrm{L}^2(\mathcal{A}^+)} + |\mathcal{A}^+|.$$

Combining these two inequalities gives

$$\|(p-1)^+\|^2_{\mathrm{L}^2} \ge \frac{1}{2}\|p\|^2_{\mathrm{L}^2(\mathcal{A}^+)} - |\mathcal{A}^+|.$$

Similarly, we have that

$$\|(p+1)^-\|^2_{\mathrm{L}^2} \ge \frac{1}{2}\|p\|^2_{\mathrm{L}^2(\mathcal{A}^-)} - |\mathcal{A}^-|.$$

Without loss of generality, we may assume that $c \geq 1$. Then by the minimizing property of $p_n \equiv p_{\beta_n,c}$, we have that

$$\begin{aligned}\frac{1}{2\alpha}\|K^*p_n\|_{L^2}^2 - \langle p_n, y^\delta\rangle_{L^2} + \frac{\beta_n}{2}\|\nabla p_n\|_{L^2}^2 + \frac{1}{2}(\|(p_n-1)^+\|_{L^2}^2 + \|(p_n+1)^-\|_{L^2}^2)\\ \leq \frac{1}{2\alpha}\|K^*p_n\|_{L^2}^2 - \langle p_n, y^\delta\rangle_{L^2} + \frac{\beta_n}{2}\|\nabla p_n\|_{L^2}^2 + P_c(p_n) \leq 0,\end{aligned}$$

where for the sake of brevity we have set

$$P_c(p) := \frac{1}{2c}\left(\|c(p-1)^+\|_{L^2}^2 + \|c(p+1)^-\|_{L^2}^2\right).$$

This together with the inequalities above implies that

$$\begin{aligned}\frac{1}{2\alpha}\|K^*p_n\|_{L^2}^2 + \frac{\beta_n}{2}\|\nabla p_n\|_{L^2}^2 + \frac{1}{4}\|p_n\|_{L^2}^2 &\leq |\Omega| + \langle p_n, y^\delta\rangle_{L^2}\\ &\leq |\Omega| + \|p_n\|_{L^2}\|y^\delta\|_{L^2}\\ &\leq |\Omega| + \frac{1}{8}\|p_n\|_{L^2}^2 + 2\|y^\delta\|_{L^2}^2.\end{aligned}$$

This in particular implies that the sequence $\{p_n\}$ is uniformly bounded in L^2 independently of n. Therefore, there exists a subsequence, also denoted by $\{p_n\}$, converging weakly in L^2 to some $p^* \in L^2$. By the weak lower semicontinuity of norms, we have

$$\|K^*p^*\|_{L^2}^2 \leq \liminf_{n\to\infty}\|K^*p_n\|_{L^2}^2, \qquad \langle p^*, y^\delta\rangle_{L^2} = \lim_{n\to\infty}\langle p_n, y^\delta\rangle_{L^2},$$

and, moreover, by the convexity of the operators max and min, we have weak lower semicontinuity of the corresponding terms

$$\begin{aligned}\|(p^*-1)^+\|_{L^2}^2 &\leq \liminf_{n\to\infty}\|(p_n-1)^+\|_{L^2}^2,\\ \|(p^*+1)^-\|_{L^2}^2 &\leq \liminf_{n\to\infty}\|(p_n+1)^-\|_{L^2}^2.\end{aligned}$$

By the minimizing property of p_n, we thus have for any fixed $p \in H^1$ that

$$\begin{aligned}&\frac{1}{2\alpha}\|K^*p^*\|_{L^2}^2 - \langle p^*, y^\delta\rangle_{L^2} + P_c(p^*)\\ &\quad\leq \liminf_{n\to\infty}\left(\frac{1}{2\alpha}\|K^*p_n\|_{L^2}^2 - \langle p_n, y^\delta\rangle_{L^2} + \frac{\beta_n}{2}\|\nabla p_n\|_{L^2}^2 + P_c(p_n)\right)\\ &\quad\leq \liminf_{n\to\infty}\left(\frac{1}{2\alpha}\|K^*p\|_{L^2}^2 - \langle p, y^\delta\rangle_{L^2} + \frac{\beta_n}{2}\|\nabla p\|_{L^2}^2 + P_c(p)\right)\\ &\quad= \frac{1}{2\alpha}\|K^*p\|_{L^2}^2 - \langle p, y^\delta\rangle_{L^2} + \frac{1}{2c}\left(\|c(p-1)^+\|_{L^2}^2 + \|c(p+1)^-\|_{L^2}^2\right).\end{aligned}$$

Therefore, p^* is a minimizer of problem ($\mathcal{P}_c^*$) over H^1. Now the density of H^1 in L^2 shows that p^* is also a minimizer of problem ($\mathcal{P}_c^*$) over L^2.

Finally, by the minimizing property of $p^\dagger$ and p_n, we have

$$\begin{aligned}
&\frac{1}{2\alpha}\|K^*p_n\|_{\mathrm{L}^2}^2 - \langle p_n, y^\delta\rangle_{\mathrm{L}^2} + \frac{\beta_n}{2}\|\nabla p_n\|_{\mathrm{L}^2}^2 + P_c(p_n)\\
&\quad\le \frac{1}{2\alpha}\|K^*p^\dagger\|_{\mathrm{L}^2}^2 - \langle p^\dagger, y^\delta\rangle_{\mathrm{L}^2} + \frac{\beta_n}{2}\|\nabla p^\dagger\|_{\mathrm{L}^2}^2 + P_c(p^\dagger),\\
&\frac{1}{2\alpha}\|K^*p^\dagger\|_{\mathrm{L}^2}^2 - \langle p^\dagger, y^\delta\rangle_{\mathrm{L}^2} + P_c(p^\dagger) \le \frac{1}{2\alpha}\|K^*p_n\|_{\mathrm{L}^2}^2 - \langle p_n, y^\delta\rangle_{\mathrm{L}^2} + P_c(p_n).
\end{aligned}$$

Adding these two inequalities together, we deduce that

$$\|\nabla p_n\|_{\mathrm{L}^2}^2 \le \|\nabla p^\dagger\|_{\mathrm{L}^2}^2,$$

which together with the weak lower semicontinuity of the seminorm yields

$$\|\nabla p^*\|_{\mathrm{L}^2}^2 \le \|\nabla p^\dagger\|_{\mathrm{L}^2}^2,$$

i.e., that p^* is a minimizer with minimal H^1 seminorm. If K is injective or $p^\dagger$ is unique, then it follows that $p^* = p^\dagger$. Consequently, each subsequence has a subsequence converging weakly to $p^\dagger$, and the whole sequence converges weakly. ■

Appendix B. Proof of Lemma 4.2. By Theorem 2.2, the function $\|Kx_\alpha - y^\delta\|_{\mathrm{L}^1}$ is continuous and increasing as a function of α. Therefore the following limits exist:

$$\lim_{\alpha\to 0^+}\|Kx_\alpha - y^\delta\|_{\mathrm{L}^1}, \qquad \lim_{\alpha\to+\infty}\|Kx_\alpha - y^\delta\|_{\mathrm{L}^1}.$$

By the minimizing property of x_α, we have

$$\|Kx_\alpha - y^\delta\|_{\mathrm{L}^1} + \frac{\alpha}{2}\|x_\alpha\|_{\mathrm{L}^2}^2 \le \|Kx - y^\delta\|_{\mathrm{L}^1} + \frac{\alpha}{2}\|x\|_{\mathrm{L}^2}^2 \quad \text{for all } x \in \mathrm{L}^2.$$

Taking $x = 0$, this gives

$$\|Kx_\alpha - y^\delta\|_{\mathrm{L}^1} + \frac{\alpha}{2}\|x_\alpha\|_{\mathrm{L}^2}^2 \le \|y^\delta\|_{\mathrm{L}^1}.$$

Letting α tend to ∞, we deduce that

$$0 \le \lim_{\alpha\to+\infty}\|x_\alpha\|_{\mathrm{L}^2}^2 \le \lim_{\alpha\to+\infty}\frac{2}{\alpha}\|y^\delta\|_{\mathrm{L}^1} = 0,$$

i.e., $\lim_{\alpha\to+\infty} x_\alpha = 0$. From this we derive that

$$\lim_{\alpha\to+\infty}\|Kx_\alpha - y^\delta\|_{\mathrm{L}^1} = \|y^\delta\|_{\mathrm{L}^1}.$$

Appealing again to the minimizing property, we obtain

$$\lim_{\alpha\to+\infty}\frac{\alpha}{2}\|x_\alpha\|_{\mathrm{L}^2}^2 = 0.$$

Let $\theta = \inf_{x\in\mathrm{L}^2}\|Kx - y^\delta\|_{\mathrm{L}^1}$. By monotonicity and continuity of $\left\|Kx_\alpha - y^\delta\right\|_{\mathrm{L}^1}$, we have that

$$\theta = \lim_{\alpha\to 0^+}\|Kx_\alpha - y^\delta\|_{\mathrm{L}^1}. \tag{B.1}$$

By the definition of the infimum, there exists an x^ε such that

$$\theta \leq \|Kx^\varepsilon - y^\delta\|_{\mathrm{L}^1} \leq \theta + \varepsilon.$$

Now the minimizing property of x_α yields

$$\theta \leq \|Kx_\alpha - y^\delta\|_{\mathrm{L}^1} + \frac{\alpha}{2}\|x_\alpha\|_{\mathrm{L}^2}^2 \leq \|Kx^\varepsilon - y^\delta\|_{\mathrm{L}^1} + \frac{\alpha}{2}\|x^\varepsilon\|_{\mathrm{L}^2}^2 \leq \theta + \varepsilon + \frac{\alpha}{2}\|x^\varepsilon\|_{\mathrm{L}^2}^2.$$

Letting α tend to zero, we conclude that

$$\theta \leq \lim_{\alpha\to 0^+} \left\{\|Kx_\alpha - y^\delta\|_{\mathrm{L}^1} + \frac{\alpha}{2}\|x_\alpha\|_{\mathrm{L}^2}^2\right\} \leq \theta + \varepsilon,$$

and, since ε is arbitrary, we have

$$\theta \leq \lim_{\alpha\to 0^+} \left\{\|Kx_\alpha - y^\delta\|_{\mathrm{L}^1} + \frac{\alpha}{2}\|x_\alpha\|_{\mathrm{L}^2}^2\right\} \leq \theta,$$

which together with (B.1) implies that $\lim_{\alpha\to 0^+} \frac{\alpha}{2}\|x_\alpha\|_{\mathrm{L}^2}^2 = 0$. ∎

Appendix C. Benchmark algorithms. For the reader's convenience, we briefly sketch the implemented version and the values of all occurring parameters of the benchmark methods. For a detailed description, we refer to references [32, 29, 34]. Since the proposed semismooth Newton method solves the discrete optimality system exactly, we avoided introducing early termination criteria in the benchmark algorithms to allow for a fair comparison. Instead, we fixed the number of iterations such that their performance was optimal while still giving the same reconstruction errors as the semismooth Newton method. In practice, one would add termination criteria based, e.g., on the norm of the difference of iterates, which would accelerate the benchmark methods as well as the semismooth Newton method.

C.1. Iteratively reweighted least squares. The basic idea of the approach is to approximate the (discrete) L^1 norm from above by a quadratic function $Q(x,z)$ such that $Q(x,z) \geq \|z\|_{\mathrm{L}^1}$ and $Q(x,x) = \|x\|_{\mathrm{L}^1}$. One such choice is given by

$$Q(x,z) = \|x\|_{\mathrm{L}^1} + \tfrac{1}{2}\left(z^T W(x) z - x^T W(x) x\right),$$

where $W(x)$ is a diagonal matrix with entries $|x_i|^{-1}$. The L^1 norm is then minimized by iteratively solving for given x^k the smooth minimization problem with $Q(x^k, x)$ in place of $\|x\|_{\mathrm{L}^1}$. To avoid division by zero in the definition of $W(x)$, an additional regularization parameter ε is introduced, which is set to $\varepsilon = 10^{-6}$, again for best performance, with the reconstruction error being the same as that from the proposed algorithm. The maximum number of iterations k_{max} is set to 40.

C.2. Alternating direction minimization. This approach consists in introducing the splitting $z = Kx - y^\delta$ and minimizing for $\beta > 0$ the functional

$$\|z\|_{\mathrm{L}^1} + \frac{\beta}{2}\left\|Kx - y^\delta - z\right\|_{\mathrm{L}^2}^2 + \frac{\alpha}{2}\|x\|_{\mathrm{L}^2}^2.$$

The minimization is carried out by alternately minimizing with respect to z and with respect to x, which are both explicitly solvable. If the penalty parameter β goes to infinity, the solution converges to the solution x_α of problem ($\mathcal{P}$). We therefore employ a continuation strategy for β, as was adopted in [34]. The full procedure is given in Algorithm 5. The parameters were chosen as $\beta_0 = 1$, $\beta_{max} = 2^{16}$, $q = 2$, and $k_{max} = 30$.

Algorithm 4. IRLS.

Choose tol, k_{max}, set $x_0 = 0$.
for $k = 0, \dots, k_{max}$ **do**
  Set $W = \operatorname{diag}\left(\max(|Kx_k - y^\delta|, \varepsilon)^{-1}\right)$.
  Set x_{k+1} as solution of $(K^*WK + \alpha I)x = (K^*W)y^\delta$.
end for

Algorithm 5. ADM.

Choose k_{max}, β_0, β_{max}, q.
Set $x_0 = z_0 = 0$, $\beta = \beta_0$.
repeat
  for $k = 0, \dots, k_{max}$ **do**
    Set $z_{k+1} = \operatorname{sgn}(Kx_k - y^\delta) \cdot \max(|Kx_k - y^\delta| - \beta^{-1}, 0)$.
    Set x_{k+1} as solution of $(K^*K + \frac{\alpha}{\beta} I)x = K^*(y^\delta + z_{k+1})$.
  end for
  Set $x_0 = x_k$, $\beta = q\beta$.
until $\beta > \beta_{max}$
return x_k

Acknowledgments. This work was carried out during the second author's visit to the Institute of Mathematics and Scientific Computing, Karl-Franzens-Universität Graz. He would like to thank Professor Karl Kunisch and the Institute for their hospitality.

REFERENCES

[1] W. K. Allard, *Total variation regularization for image denoising,* I. *Geometric theory*, SIAM J. Math. Anal., 39 (2007), pp. 1150–1190.
[2] S. Alliney, *A property of the minimum vectors of a regularizing functional defined by means of the absolute norm*, IEEE Trans. Signal Process., 45 (1997), pp. 913–917.
[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, ed., *Handbook of Image and Video Processing (Communications, Networking and Multimedia)*, Academic Press, Orlando, FL, 2005.
[5] H. Brezis, *Analyse fonctionnelle*, Masson, Paris, 1983.
[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] R. H. Chan, Y. Dong, and M. Hintermüller, *An efficient two-phase $L1$-TV method for restoring blurred images with impulse noise*, IEEE Trans. Image Process., to appear.
[8] T. F. Chan and S. Esedoḡlu, *Aspects of total variation regularized L^1 function approximation*, SIAM J. Appl. Math., 65 (2005), pp. 1817–1837.
[9] 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.
[10] V. Duval, J.-F. Aujol, and Y. Gousseau, *The $TVL1$ model: A geometric point of view*, Multiscale Model. Simul., 8 (2009), pp. 154–189.
[11] I. Ekeland and R. Témam, *Convex Analysis and Variational Problems*, Classics Appl. Math. 28, SIAM, Philadelphia, 1999.
[12] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.

[13] 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.

[14] H. Fu, M. K. Ng, M. Nikolova, and J. L. Barlow, *Efficient minimization methods of mixed $\ell2$-$\ell1$ and $\ell1$-$\ell1$ norms for image restoration*, SIAM J. Sci. Comput., 27 (2006), pp. 1881–1902.

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

[16] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*, SIAM Monogr. Math. Model. Comput. 4, SIAM, Philadelphia, 1998.

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

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

[19] K. Ito, B. Jin, and J. Zou, *A New Choice Rule for Regularization Parameters in Tikhonov Regularization*, Technical report 2008-07(362), Department of Mathematics, Chinese University of Hong Kong, Hong Kong, 2008.

[20] K. Ito and K. Kunisch, *On the choice of the regularization parameter in nonlinear inverse problems*, SIAM J. Optim., 2 (1992), pp. 376–404.

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

[22] B. Jin and J. Zou, *Iterative Parameter Choice by Discrepancy Principle*, Technical report 2010-02(369), Department of Mathematics, Chinese University of Hong Kong, Hong Kong, 2010.

[23] P. R. Johnston and R. M. Gulrajani, *An analysis of the zero-crossing method for choosing regularization parameters*, SIAM J. Sci. Comput., 24 (2002), pp. 428–442.

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

[25] K. Kunisch and J. Zou, *Iterative choices of regularization parameters in linear inverse problems*, Inverse Problems, 14 (1998), pp. 1247–1264.

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

[27] M. Nikolova, *Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers*, SIAM J. Numer. Anal., 40 (2002), pp. 965–994.

[28] M. Nikolova, *A variational approach to remove outliers and impulse noise*, J. Math. Imaging Vision, 20 (2004), pp. 99–120.

[29] P. Rodríguez and B. Wohlberg, *Efficient minimization method for a generalized total variation functional*, IEEE Trans. Image Process., 18 (2009), pp. 322–332.

[30] P. J. Rousseeuw and A. M. Leroy, *Robust regression and outlier detection*, John Wiley & Sons, New York, 1987.

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

[32] R. Wolke and H. Schwetlick, *Iteratively reweighted least squares: Algorithms, convergence analysis, and numerical comparisons*, SIAM J. Sci. Statist. Comput., 9 (1988), pp. 907–921.

[33] J. Xie and J. Zou, *An improved model function method for choosing regularization parameters in linear inverse problems*, Inverse Problems, 18 (2002), pp. 631–643.

[34] 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.

[35] 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.