L^{∞} FITTING FOR INVERSE PROBLEMS WITH UNIFORM NOISE

Christian Clason*

August 16, 2012

For inverse problems where the data are corrupted by uniform noise such as arising from quantization errors, the L^{∞} norm is a more robust data fitting term than the standard L^2 norm. Well-posedness and regularization properties for linear inverse problems with L^{∞} data fitting are shown, and the automatic choice of the regularization parameter is discussed. After introducing an equivalent reformulation of the problem and a Moreau–Yosida approximation, a superlinearly convergent semi-smooth Newton method becomes applicable for the numerical solution of L^{∞} fitting problems. Numerical examples illustrate the performance of the proposed approach as well as the qualitative behavior of L^{∞} fitting.

1 INTRODUCTION

This work is concerned with the inverse problem

$$Kx = y^{\delta}$$

for a bounded linear operator K and given data y^{δ} corrupted by uniformly distributed noise. Apart from being often used in numerical tests of reconstruction algorithms, such noise appears as a statistical model of quantization errors and is therefore of relevance in any inverse problem where digital acquisition and processing of measured data plays a significant role [Widrow and Kollár 2008; Shykula and Seleznjev 2006]. Although advances in the resolution of analog-to-digital converters and in the floating-point precision of microprocessors have made these concerns less important for modern measurement equipment, they have become pertinent again in the context of wireless sensor networks. These consist of a large number of small, low-cost, usually battery-powered, densely distributed sensors which transmit gathered

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

data to a central location [Gharavi and Kumar 2003; Arampatzis et al. 2005]. Such networks have attracted increasing attention in recent years due to their wide range of applications, e.g., in environmental monitoring where they can be used for quickly locating sources of a contaminant from distributed measurement of its concentration [Polastre et al. 2004; Doolin and Sitar 2005]. However, their communication is limited by their low power and shared bandwidth, which requires the data to be highly compressed before transmission, leading to significant quantization errors [Niu and Varshney 2006; Schizas et al. 2007]. More robust algorithms for state estimation from quantized data would therefore allow higher compression rates and therefore extended lifetime of the sensors. In this work, we are thus especially (but not exclusively) interested in inverse problems where K is the solution operator for a (linear) partial differential equation.

Since this problem is ill-posed, regularization needs to be applied. For uniform noise, the L^∞ norm is an appropriate term for measuring the data misfit due to its connection with the maximum likelihood estimator for this noise type (see, e.g., [Boyd and Vandenberghe 2004, Chapter 7.1.1]). This leads to minimizing a Tikhonov functional of the type

$$\min_x \left\|Kx - y^\delta\right\|_{L^\infty} + \alpha \|x\|^2 \tag{1.1}$$

or – if the noise level δ is known – a Morozov functional of the type

$$\min_x \|x\|^2 \quad \text{subject to} \quad \left\|Kx - y^\delta\right\|_{L^\infty} \leqslant \delta.$$

(These will be made precise below, see Section 2.) The difficulty arises from the non-differentiability of the L^∞ norm. This may be the reason why inverse problems in L^∞ have received rather little attention in the mathematical literature, even though there has been considerable recent progress in the regularization theory in Banach spaces (see, e.g., [Burger and Osher 2004; Resmerita 2005; Resmerita and Scherzer 2006; Hofmann et al. 2007; Pöschl 2009; Scherzer et al. 2009]). Numerical methods for minimizing L^∞ functionals have been investigated in [Williams and Kalogiratou 1993a; Williams and Kalogiratou 1993b] for curve fitting and parameter estimation for ordinary differential equations and in [Grund and Rösch 2001; Prüfert and Schiela 2009; Clason, Ito, et al. 2010] for optimal control of partial differential equations. There has also be some recent interest in L^∞ functionals in the context of geometric vision [Hartley and Schaffalitzky 2004; Sim and Hartley 2006; Seo and Hartley 2007].

Our main interest thus lies in deriving an efficient method for the numerical solution of inverse problems with L^∞ fitting. Following [Grund and Rösch 2001; Prüfert and Schiela 2009; Clason, Ito, et al. 2010], our approach is based on an equivalent formulation of (1.1):

$$\min_{c,x} c + \alpha \|x\|^2 \quad \text{subject to} \quad \left\|Kx - y^\delta\right\|_{L^\infty(\Omega)} \leqslant c.$$

This can be interpreted as an "augmented Morozov regularization" for the joint estimation of the unknown parameter x and the noise level $\delta = c$. (In fact, if δ is known, the proposed approach can be used for the numerical solution of the Morozov functional by fixing $c = \delta$, see Remark 4.7 below.) For this reformulation, we derive optimality conditions, introduce a Moreau–Yosida approximation and show its convergence, and prove superlinear convergence of a semi-smooth Newton method. We also address the automatic choice of the regularization parameter α using a simple fixed-point iteration.

This paper is organized as follows. In Section 2, we address well-posedness and convergence of a slight generalization of the Tikhonov functional (1.1). Section 3 is concerned with the fixed-point algorithm for the automatic choice of the regularization parameter. The numerical solution of the L^∞ fitting problem is discussed in Section 4. Finally, numerical examples for one- and two-dimensional model problems are presented in Section 5.

2 WELL-POSEDNESS AND REGULARIZATION PROPERTIES

We consider for $1 \leqslant p < \infty$ the problem

$$\min_{x\in X} \frac{1}{p} \left\|Kx - y^\delta\right\|^p_{L^\infty(\Omega)} + \frac{\alpha}{2} \left\|x - x_0\right\|^2_X, \tag{$\mathcal{P}$}$$

where $K : X \to L^\infty(\Omega)$ is a bounded linear operator defined on the Hilbert space X, $\Omega \subset \mathbb{R}^n$ is a bounded domain, $x_0 \in X$ is given, and $y^\delta \in L^\infty(\Omega)$ are noisy measurements with noise level $\left\|y^\dagger - y^\delta\right\|_{L^\infty(\Omega)} \leqslant \delta$ (with $y^\dagger = Kx^\dagger$ being the noise-free data). If the kernel of K is non-trivial, we denote by $x^\dagger$ the x_0-minimum norm solution, i.e., the minimizer of $\|x - x_0\|_X$ over the set $\left\{x \in X : Kx = y^\dagger\right\}$. Our main assumption on K (needed for convergence of the Moreau–Yosida approximation, see Theorem 4.2) is that

$$x_n \rightharpoonup x^\dagger \text{ in } X \quad \text{implies} \quad Kx_n \to Kx^\dagger \text{ in } L^\infty(\Omega). \tag{2.1}$$

This holds if K is a compact operator or maps into a space compactly embedded into $L^\infty(\Omega)$ (as is commonly the case if K is the solution operator for a partial differential equation).

The results of this section are standard (see, e.g., [Engl et al. 1996, Chapters 5, 10], [Scherzer et al. 2009, Chapter 3.2]), and are given here to make the presentation self-contained. The first result concerns the well-posedness of ($\mathcal{P}$).

Theorem 2.1. *For $\alpha > 0$ and given y^δ,*

(i) there exists a unique solution $x^\delta_\alpha \in X$ to problem ($\mathcal{P}$);

(ii) for a sequence of data $\{y_n\}_{n\in\mathbb{N}}$ such that $y_n \to y^\delta$ in $L^\infty(\Omega)$, the sequence $\left\{x^n_\alpha\right\}_{n\in\mathbb{N}}$ of minimizers contains a subsequence converging to x^δ_α;

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

$$\lim_{\delta\to 0} \alpha(\delta) = \lim_{\delta\to 0} \frac{\delta^p}{\alpha(\delta)} = 0,$$

then the family $\left\{x^\delta_{\alpha(\delta)}\right\}_{\delta>0}$ has a subsequence converging to $x^\dagger$ as $\delta \to 0$.

Rates for the convergence in (iii) can be obtained under a source condition. Here, we assume the following condition: There exists a $w \in L^\infty(\Omega)^*$, i.e., a continuous linear functional on $L^\infty(\Omega)$, such that

$$x^\dagger - x_0 = K^* w. \tag{2.2}$$

For a discussion of source conditions and obtainable convergence rates, see [Scherzer et al. 2009, Chapter 3.2]. More general smoothness assumptions and their relation to source conditions are discussed in [Hofmann et al. 2007; Flemming and Hofmann 2011].

Theorem 2.2. *If the source condition (2.2) holds, and $\alpha = \mathcal{O}(\delta^\varepsilon)$ with $\varepsilon \in (0,1)$ in case $p = 1$ and $\alpha = \mathcal{O}(\delta^{p-1})$ in case $p > 1$, then the minimizer x_α^δ of $(\mathcal{P})$ satisfies*

$$\|x_\alpha^\delta - x^\dagger\|_X \leqslant \begin{cases} c\delta^{\frac{1-\varepsilon}{2}} & \text{if } p = 1, \\ c\delta^{\frac{1}{2}} & \text{if } p > 1. \end{cases}$$

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

$$\begin{aligned} \frac{1}{p}\|Kx_\alpha^\delta - y^\delta\|_{L^\infty(\Omega)}^p + \frac{\alpha}{2}\|x_\alpha^\delta - x_0\|_X^2 &\leqslant \frac{1}{p}\|Kx^\dagger - y^\delta\|_{L^\infty(\Omega)}^p + \frac{\alpha}{2}\|x^\dagger - x_0\|_X^2 \\ &\leqslant \frac{\delta^p}{p} + \frac{\alpha}{2}\|x^\dagger - x_0\|_X^2 \end{aligned}$$

and hence

$$\frac{1}{p}\|Kx_\alpha^\delta - y^\delta\|_{L^\infty(\Omega)}^p + \frac{\alpha}{2}\|x_\alpha^\delta - x^\dagger\|_X^2 \leqslant \frac{\delta^p}{p} + \alpha\langle x^\dagger - x_0, x^\dagger - x_\alpha^\delta\rangle_X.$$

Now by the source condition (2.2), we have

$$\begin{aligned} \frac{1}{p}\|Kx_\alpha^\delta - y^\delta\|_{L^\infty(\Omega)}^p + \frac{\alpha}{2}\|x_\alpha^\delta - x^\dagger\|_X^2 &\leqslant \frac{\delta^p}{p} + \alpha\langle K^*w, x^\dagger - x_\alpha^\delta\rangle_X \\ &= \frac{\delta^p}{p} + \alpha\langle w, y^\dagger - Kx_\alpha^\delta\rangle_{L^\infty(\Omega)^*, L^\infty(\Omega)} \\ &\leqslant \frac{\delta^p}{p} + \alpha\,\|w\|_{L^\infty(\Omega)^*}\,\|Kx_\alpha^\delta - y^\dagger\|_{L^\infty(\Omega)}. \end{aligned}$$

Inserting the productive zero $0 = y^\delta - y^\delta$ on the right hand side and applying the triangle inequality yields

$$\begin{aligned} &\frac{1}{p}\|Kx_\alpha^\delta - y^\delta\|_{L^\infty(\Omega)}^p + \frac{\alpha}{2}\|x_\alpha^\delta - x^\dagger\|_X^2 \\ &\qquad \leqslant \frac{\delta^p}{p} + \alpha\,\|w\|_{L^\infty(\Omega)^*}\,(\|Kx_\alpha^\delta - y^\delta\|_{L^\infty(\Omega)} + \|y^\dagger - y^\delta\|_{L^\infty(\Omega)}). \end{aligned} \tag{2.3}$$

If $p = 1$, we have

$$(1 - \alpha\,\|w\|_{L^\infty(\Omega)^*})\|Kx_\alpha^\delta - y^\delta\|_{L^\infty(\Omega)} + \frac{\alpha}{2}\|x_\alpha^\delta - x^\dagger\|_X^2 \leqslant (1 + \alpha\,\|w\|_{L^\infty(\Omega)^*})\delta,$$

from which the desired convergence rate follows by choosing choosing $\alpha = \mathcal{O}(\delta^{\varepsilon})$. For $p > 1$, we use Young's inequality $ab \leqslant \frac{1}{p}a^p + \frac{1}{p'}b^{p'}$ for $p' = \frac{p}{p-1}$, $a = \|Kx_\alpha^\delta - y^\delta\|_{L^\infty(\Omega)}$ and $b = \alpha \|w\|_{L^\infty(\Omega)^*}$, and rearrange terms to deduce

$$-\frac{1}{p}\|Kx_\alpha^\delta - y^\delta\|_{L^\infty(\Omega)}^p \leqslant -\alpha \|w\|_{L^\infty(\Omega)^*} \|Kx_\alpha^\delta - y^\delta\|_{L^\infty(\Omega)} + \frac{1}{p'}(\alpha \|w\|_{L^\infty(\Omega)^*})^{p'}.$$

Hence, by adding the last term on the right hand side to both sides of (2.3), we obtain

$$\frac{\alpha}{2}\|x_\alpha^\delta - x^\dagger\|_X^2 \leqslant \frac{\delta^p}{p} + \alpha \|w\|_{L^\infty(\Omega)^*} \delta + \frac{1}{p'}(\alpha \|w\|_{L^\infty(\Omega)^*})^{p'}.$$

Taking $\alpha = \mathcal{O}(\delta^{p-1})$ then yields the claimed estimate. □

We remark that for $p = 1$, ($\mathcal{P}$) is an exact penalization, i.e., there exists $\alpha^* > 0$ such that for all $\alpha < \alpha^*$, the minimizer x_α^0 of ($\mathcal{P}$) with exact data $y^\dagger$ satisfies $x_\alpha^0 = x^\dagger$, see [Burger and Osher 2004; Hofmann et al. 2007].

Next we consider Morozov's discrepancy principle [Morozov 1966], which consists in choosing α such that

$$\|Kx_\alpha^\delta - y^\delta\|_{L^\infty(\Omega)} = \tau\delta$$

for some $\tau > 1$.

Theorem 2.3. *Assume that the source condition (2.2) holds, and that the regularization parameter $\alpha = \alpha(\delta)$ is determined according to the discrepancy principle. Then the minimizer x_α^δ of ($\mathcal{P}$) satisfies*

$$\|x_\alpha^\delta - x^\dagger\|_X \leqslant c\delta^{\frac{1}{2}}.$$

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

$$\|x_\alpha^\delta - x_0\|_X^2 \leqslant \|x^\dagger - x_0\|_X^2,$$

from which it follows that

$$\begin{aligned}
\|x_\alpha^\delta - x^\dagger\|_X^2 &\leqslant 2\langle x^\dagger - x_0, x^\dagger - x_\alpha^\delta\rangle_X = 2\langle K^*w, x^\dagger - x_\alpha^\delta\rangle_X \\
&\leqslant 2\|w\|_{L^\infty(\Omega)^*} \|Kx^\dagger - Kx_\alpha^\delta\|_{L^\infty(\Omega)} \\
&\leqslant 2\|w\|_{L^\infty(\Omega)^*} (\|y^\dagger - y^\delta\|_{L^\infty(\Omega)} + \|y^\delta - Kx_\alpha^\delta\|_{L^\infty(\Omega)}) \\
&\leqslant 2\|w\|_{L^\infty(\Omega)^*} (1+\tau)\delta,
\end{aligned}$$

and hence we obtain the claimed estimate. □

Remark 2.4. If K is an adjoint operator, i.e., there exists $K_* : L^1(\Omega) \to X$ such that $(K_*)^* = K$, the source condition can be stated as: There exists $w \in L^1(\Omega)$ such that $x^\dagger - x_0 = K_* w$.

3 PARAMETER CHOICE

Morozov's discrepancy principle requires knowledge of the noise level, which is often not available in practice. Here, we use a heuristic choice rule derived from a balancing principle [Clason et al. 2010b; Clason et al. 2010a; Clason and Jin 2012], which involves auto-calibration of the noise level. Although there is no rigorous justification, we can give a brief motivation of this principle. Recall that the Morozov discrepancy principle chooses α such that the residual in the appropriate norm is on the order of the noise level δ. In an iterative scheme, one would start with a large parameter and reduce it until this condition is satisfied, making use of the fact that the norm of the residual is monotonically increasing as a function of α (see Lemma 3.1 below). On the other hand, the regularization term is monotonically decreasing; one could therefore equally choose α such that the regularization term reaches a certain value, which is proportional to the noise level δ. If δ is not known, the current residual can be used in this approach to give an estimate of the noise level. If the true data and the noise are sufficiently structurally different, it can be expected that

$$\|Kx_\alpha - y^\delta\|_{L^\infty(\Omega)} \approx \delta$$

for a reasonable range of α. (A similar assumption can be used to show convergence of minimization-based noise level-free parameter choice rules [Kindermann 2011].) This motivates considering the following heuristic principle: Choose $\alpha > 0$ such that the balancing equation

$$\frac{\alpha}{2}\|x_\alpha - x_0\|_X^2 = \sigma\|Kx_\alpha - y^\delta\|_{L^\infty(\Omega)} \tag{3.1}$$

is satisfied. (Note that $\|Kx^\dagger - y^\delta\|_{L^\infty(\Omega)}$ rather than $\|Kx^\dagger - y^\delta\|_{L^\infty(\Omega)}^p$ is the true noise level by definition.) Here, σ is a proportionality constant which depends on K and X, but not on δ.

We can compute a solution α^* to (3.1) by the following simple fixed-point algorithm proposed in [Clason et al. 2010a]:

$$\alpha_{k+1} = \sigma\frac{\|Kx_{\alpha_k} - y^\delta\|_{L^\infty(\Omega)}}{\frac{1}{2}\|x_{\alpha_k} - x_0\|_X^2}. \tag{3.2}$$

This fixed-point algorithm can be derived formally from the model function approach [Clason et al. 2010b]. The convergence can be proven similarly as in [Clason et al. 2010a]. We start by arguing monotonicity of the data fitting and of the regularization term.

Lemma 3.1. *The functions $\|Kx_\alpha - y^\delta\|_{L^\infty(\Omega)}$ and $\|x_\alpha - x_0\|_X$ are monotonic in α, in the sense that for $\alpha_1, \alpha_2 > 0$,*

$$\left(\|Kx_{\alpha_1} - y^\delta\|_{L^\infty(\Omega)} - \|Kx_{\alpha_2} - y^\delta\|_{L^\infty(\Omega)}\right)(\alpha_1 - \alpha_2) \geqslant 0$$

and

$$\left(\|x_{\alpha_1} - x_0\|_X^2 - \|x_{\alpha_2} - x_0\|_X^2\right)(\alpha_1 - \alpha_2) \leqslant 0.$$

Proof. The minimizing property of x_{α_1} and x_{α_2} yields

$$\begin{aligned}
\frac{1}{p}\left\|Kx_{\alpha_1}-y^\delta\right\|^p_{L^\infty(\Omega)}+\frac{\alpha_1}{2}\left\|x_{\alpha_1}-x_0\right\|^2_X &\leqslant \frac{1}{p}\left\|Kx_{\alpha_2}-y^\delta\right\|^p_{L^\infty(\Omega)}+\frac{\alpha_1}{2}\left\|x_{\alpha_2}-x_0\right\|^2_X,\\
\frac{1}{p}\left\|Kx_{\alpha_2}-y^\delta\right\|^p_{L^\infty(\Omega)}+\frac{\alpha_2}{2}\left\|x_{\alpha_2}-x_0\right\|^2_X &\leqslant \frac{1}{p}\left\|Kx_{\alpha_1}-y^\delta\right\|^p_{L^\infty(\Omega)}+\frac{\alpha_2}{2}\left\|x_{\alpha_1}-x_0\right\|^2_X.
\end{aligned}$$

Adding these two inequalities together gives the second estimate. The first one can be obtained by dividing the two inequalities by α_1 and α_2, respectively, adding them together, and using the monotonicity of $t \mapsto t^p$ for $p \geqslant 1$ and $t \geqslant 0$. $\square$

We shall denote by

$$r(\alpha) = \sigma\|Kx_\alpha - y^\delta\|_{L^\infty(\Omega)} - \frac{\alpha}{2}\|x_\alpha - x_0\|_X$$

the residual in (3.1). The next lemma shows the monotonicity of the iteration (3.2).

Lemma 3.2. *The sequence of regularization parameters* $\{\alpha_k\}_{k\in\mathbb{N}}$ *generated by the fixed-point algorithm is monotonically increasing if* $r(\alpha_0) > 0$ *and monotonically decreasing if* $r(\alpha_0) < 0$.

Proof. We argue by induction. If $r(\alpha_0) > 0$, then by the definition of the iteration, we have

$$\alpha_1 = \sigma\frac{\|Kx_{\alpha_0} - y^\delta\|_{L^\infty(\Omega)}}{\frac{1}{2}\|x_{\alpha_0} - x_0\|_X^2} > \alpha_0,$$

and similarly (with the opposite inequality) if $r(\alpha_0) < 0$. Now for any $k > 1$,

$$\begin{aligned}
\alpha_{k+1} - \alpha_k &= \sigma\left(\frac{\|Kx_{\alpha_k} - y^\delta\|_{L^\infty(\Omega)}}{\frac{1}{2}\|x_{\alpha_k} - x_0\|_X^2} - \frac{\|Kx_{\alpha_{k-1}} - y^\delta\|_{L^\infty(\Omega)}}{\frac{1}{2}\|x_{\alpha_{k-1}} - x_0\|_X^2}\right)\\
&= \frac{2\sigma}{\|x_{\alpha_k} - x_0\|_X^2\|x_{\alpha_{k-1}} - x_0\|_X^2}\\
&\qquad\Big[\|Kx_{\alpha_k} - y^\delta\|_{L^\infty(\Omega)}(\|x_{\alpha_{k-1}} - x_0\|_X^2 - \|x_{\alpha_k} - x_0\|_X^2)\\
&\qquad\quad + \|x_{\alpha_k} - x_0\|_X^2(\|Kx_{\alpha_k} - y^\delta\|_{L^\infty(\Omega)} - \|Kx_{\alpha_{k-1}} - y^\delta\|_{L^\infty(\Omega)})\Big].
\end{aligned}$$

By Lemma 3.1, the two terms in parentheses both have the sign of $(\alpha_k - \alpha_{k-1})$, and thus the whole sequence is monotonic. $\square$

Theorem 3.3. *If the initial guess* α_0 *satisfies* $r(\alpha_0) < 0$, *then the sequence* $\{\alpha_k\}$ *generated by the fixed-point algorithm converges to a solution to* (3.1).

Proof. By Lemma 3.2 and $r(\alpha_0) < 0$, the sequence $\{\alpha_k\}$ is monotonically decreasing. Since by definition (3.2) it is clearly bounded from below by zero, convergence follows. $\square$

Note that Theorem 3.3 gives a constructive method of choosing a suitable parameter σ: Set α_0 sufficiently large (e.g., $\alpha_0 = 1$) and select σ small enough such that $r(\alpha_0) < 0$ is satisfied.

Remark 3.4. The convergence of the fixed-point iteration solely depends on the monotonicity properties of the fitting and of the regularization term. It can therefore be applied for finding solutions to the balancing equation

$$\alpha\mathcal{R}(x_\alpha) = \sigma\mathcal{F}(x_\alpha; y^\delta)$$

for minimizers x_α of the Tikhonov functional

$$\varphi\left(\mathcal{F}(x; y^\delta)\right) + \alpha\mathcal{R}(x),$$

where φ is any monotone, real-valued function and $\mathcal{R}$, $\mathcal{F}$ are arbitrary functionals for which the minimization problem is well-posed.

4 NUMERICAL SOLUTION

The numerical solution of problem ($\mathcal{P}$) is based on a sequence of Moreau–Yosida approximations of an equivalent formulation of ($\mathcal{P}$), which can be solved using a superlinearly convergent semi-smooth Newton method.

4.1 REFORMULATION

We begin by introducing an equivalent formulation of ($\mathcal{P}$) that allows making use of techniques developed for optimization problems for partial differential equations with state constraints (see [Grund and Rösch 2001; Prüfert and Schiela 2009; Clason, Ito, et al. 2010]). Since we wish to apply a Newton-type method, we fix $p = 2$ from here on (guaranteeing positive definiteness of the Hessian; see Theorem 4.4). Note that the value of p only influences the trade-off between minimizing the L^∞ norm of the residual and minimizing the norm of x, but not the relevant structural properties of the functional (in particular the geometry of the unit ball with respect to $\|\cdot\|_{L^\infty}^p$). Without loss of generality, we also set $x_0 = 0$ and consider

$$(\mathcal{P}_c) \qquad \min_{(x,c)\in X\times\mathbb{R}} \frac{c^2}{2} + \frac{\alpha}{2}\|x\|_X^2 \quad \text{subject to} \quad \left\|Kx - y^\delta\right\|_{L^\infty(\Omega)} \leqslant c.$$

The strict convexity of the equivalent problem ($\mathcal{P}$) directly yields the existence of a unique minimizer (x^*, c^*) with $x^* = x_\alpha^\delta$ from Theorem 2.1 and $c^* = \left\|Kx_\alpha^\delta - y^\delta\right\|_{L^\infty(\Omega)}$.

For a *differentiable* strictly convex functional J, the minimizer x^* can be found by computing a stationary point, which satisfies the optimality condition $J'(x^*) = 0$. For nondifferentiable problems such as ($\mathcal{P}_c$), a similar equivalence holds, although the optimality conditions become more complicated and a so-called regular point condition needs to be satisfied (see, e.g., [Maurer and Zowe 1979; Ito and Kunisch 2008]).

In the following, $j : X \to X^*$ denotes the (linear) duality mapping of the Hilbert space X, i.e., $j(u) = \partial\left(\frac{1}{2}\|\cdot\|_X^2\right)(u)$, and $\langle\cdot,\cdot\rangle_{L^\infty(\Omega)^*, L^\infty(\Omega)}$ the duality pairing between $L^\infty(\Omega)$ and its topological dual.

Theorem 4.1. *Let* $(x^*, c^*) \in X \times \mathbb{R}$ *be the solution to* ($\mathcal{P}_c$)*. Then there exist* $\lambda_1, \lambda_2 \in L^\infty(\Omega)^*$ *with*

$$\langle \lambda_1, \varphi \rangle_{L^\infty(\Omega)^*, L^\infty(\Omega)} \leqslant 0, \qquad \langle \lambda_2, \varphi \rangle_{L^\infty(\Omega)^*, L^\infty(\Omega)} \geqslant 0 \tag{4.1}$$

for all $\varphi \in L^\infty(\Omega)$ *with* $\varphi \geqslant 0$ *such that*

$$\text{(OS)} \qquad \begin{cases} \alpha j(x^*) = \langle \lambda_1 + \lambda_2, Kx^* \rangle_{L^\infty(\Omega)^*, L^\infty(\Omega)}, \\ \quad c^* = \langle \lambda_1 - \lambda_2, -1 \rangle_{L^\infty(\Omega)^*, L^\infty(\Omega)}, \\ \quad 0 = \left\langle \lambda_1, Kx^* - y^\delta - c^* \right\rangle_{L^\infty(\Omega)^*, L^\infty(\Omega)}, \\ \quad 0 = \left\langle \lambda_2, Kx^* - y^\delta + c^* \right\rangle_{L^\infty(\Omega)^*, L^\infty(\Omega)}. \end{cases}$$

Proof. Let $G : X \times \mathbb{R} \to L^\infty(\Omega) \times L^\infty(\Omega)$ be defined by

$$G(x, c) = \begin{pmatrix} Kx - y^\delta - c \\ -Kx + y^\delta - c \end{pmatrix},$$

and let K denote the non-positive cone in $L^\infty(\Omega)$, i.e.,

$$K = \{y \in L^\infty(\Omega) : y \leqslant 0\}.$$

With $J : X \times \mathbb{R} \to \mathbb{R}$,

$$J(x, c) = \frac{c^2}{2} + \frac{\alpha}{2} \|x\|_X^2,$$

we can express ($\mathcal{P}_c$) as

$$\min_{(x,c) \in X \times \mathbb{R}} J(x, c) \qquad \text{subject to} \qquad G(x, c) \in K \times K. \tag{4.2}$$

The regular point condition [Maurer and Zowe 1979; Ito and Kunisch 2008] for (4.2) is

$$0 \in \operatorname{int}\left(G(x^*, c^*) + G'(x^*, c^*)(X \times \mathbb{R}) - K \times K\right), \tag{4.3}$$

where int denotes the topological interior and G' is the Fréchet derivative of G. To verify (4.3), we need to find $\bar{x} \in X$ and $\bar{c} \in \mathbb{R}$ such that

$$\begin{aligned} (Kx^* - y^\delta - c^*) + K\bar{x} - \bar{c} &< 0, \\ (-Kx^* + y^\delta - c^*) - K\bar{x} - \bar{c} &< 0. \end{aligned}$$

Since the minimizer (x^*, c^*) satisfies the L^∞ bound, the terms in parentheses are non-positive almost everywhere, and thus these conditions are satisfied for $\bar{x} = 0$ and arbitrary $\bar{c} > 0$. From [Maurer and Zowe 1979, Theorem 3.2], we then obtain the existence of $(\lambda_1, -\lambda_2)$ in the dual cone of $K \times K$ (i.e., satisfying (4.1)), such that with $Y := L^\infty(\Omega) \times L^\infty(\Omega)$,

$$J'(x^*, c^*) = \langle (\lambda_1, -\lambda_2), G'(x^*, c^*) \rangle_{Y^*, Y}$$

and

$$\langle (\lambda_1, -\lambda_2), G(x^*, c^*) \rangle_{Y^*, Y} = 0$$

hold. Inserting the explicit form of J', G, and G' yields (OS). $\square$

4.2 Moreau–Yosida approximation

To avoid dealing with the dual space of $L^\infty(\Omega)$, we consider for $\gamma > 0$ the Moreau–Yosida approximation

$$(\mathcal{P}_\gamma) \qquad \min_{(x,c)\in X\times\mathbb{R}} \frac{c^2}{2} + \frac{\alpha}{2}\|x\|_X^2 + \frac{\gamma}{2}\left\|\max(0, Kx - y^\delta - c)\right\|_{L^2(\Omega)}^2 + \frac{\gamma}{2}\left\|\min(0, Kx - y^\delta + c)\right\|_{L^2(\Omega)}^2,$$

where the max and min are to be understood pointwise in Ω. Since this is a strictly convex and weakly lower semi-continuous problem in c and x, there exists a unique solution $(x_\gamma, c_\gamma) \in X \times \mathbb{R}$.

We next address convergence of (x_γ, c_γ) to the solution (x^*, c^*) to ($\mathcal{P}_c$).

Theorem 4.2. *As $\gamma \to \infty$, (x_γ, c_γ) converges strongly to (x^*, c^*) in $X \times \mathbb{R}$.*

Proof. Let

$$\lambda_{\gamma,1} = \gamma \max(0, Kx_\gamma - y^\delta - c_\gamma), \qquad \lambda_{\gamma,2} = \gamma \min(0, Kx_\gamma - y^\delta + c_\gamma).$$

Due to the optimality of (x_γ, c_γ) and the feasibility of (x^*, c^*), we have for all $\gamma > 0$ that

$$\frac{(c_\gamma)^2}{2} + \frac{\alpha}{2}\|x_\gamma\|_X^2 + \frac{1}{2\gamma}\|\lambda_{\gamma,1}\|_{L^2(\Omega)}^2 + \frac{1}{2\gamma}\|\lambda_{\gamma,2}\|_{L^2(\Omega)}^2 \leqslant \frac{(c^*)^2}{2} + \frac{\alpha}{2}\|x^*\|_X^2 \tag{4.5}$$

and hence that the families

$$\{x_\gamma\}_{\gamma>0}, \quad \{c_\gamma\}_{\gamma>0}, \quad \left\{\gamma^{-1}\|\lambda_{\gamma,1}\|_{L^2(\Omega)}^2\right\}_{\gamma>0}, \quad \left\{\gamma^{-1}\|\lambda_{\gamma,2}\|_{L^2(\Omega)}^2\right\}_{\gamma>0}$$

are bounded. Consequently, there exists a sequence $\{\gamma_k\}_{k\in\mathbb{N}}$ and $(\hat{x}, \hat{c}) \in X \times \mathbb{R}$ such that $(x_{\gamma_k}, c_{\gamma_k})$ converges to $(\hat{x}, \hat{c})$ and

$$\left\|\max(0, Kx_{\gamma_k} - y^\delta - c_{\gamma_k})\right\|_{L^2(\Omega)}^2 \leqslant \frac{C}{\gamma_k} \to 0$$

for $k \to \infty$, and similarly for $\min(0, Kx_{\gamma_k} - y^\delta + c_{\gamma_k})$. Since $Kx_{\gamma_k} \to K\hat{x}$ strongly in $L^\infty(\Omega)$ by assumption (2.1), this implies that

$$\left\|K\hat{x} - y^\delta\right\|_{L^\infty(\Omega)} \leqslant \hat{c}.$$

Taking the limit in (4.5), we thus find that $(\hat{x}, \hat{c})$ coincides with the unique solution (x^*, c^*) to ($\mathcal{P}_c$). Due to uniqueness of (x^*, c^*), the whole family $\{(x_\gamma, c_\gamma)\}_{\gamma>0}$ converges in $X \times \mathbb{R}$ to (x^*, c^*). □

Since ($\mathcal{P}_\gamma$) is differentiable and strictly convex, straightforward computation yields the (necessary and sufficient) optimality conditions

$$(\mathrm{OS}_\gamma) \qquad \begin{cases} \alpha j(x_\gamma) + \gamma K^*\left(\max(0, Kx_\gamma - y^\delta - c_\gamma) + \min(0, Kx_\gamma - y^\delta + c_\gamma)\right) = 0, \\ c_\gamma + \gamma\left\langle -\max(0, Kx_\gamma - y^\delta - c_\gamma) + \min(0, Kx_\gamma - y^\delta + c_\gamma), 1\right\rangle_{L^2(\Omega)} = 0. \end{cases}$$

Remark 4.3. Similarly to [Clason, Ito, et al. 2010, Theorem 3.1], one can show that as $\gamma \to \infty$,

$$\begin{aligned} -\gamma \max(0, Kx_\gamma - y^\delta - c_\gamma) &\rightharpoonup \lambda_1, \\ \gamma \min(0, Kx_\gamma - y^\delta + c_\gamma) &\rightharpoonup \lambda_2, \end{aligned}$$

weakly-$\star$ in $L^\infty(\Omega)^*$, with λ_1 and λ_2 as given by Theorem 4.1.

4.3 SEMI-SMOOTH NEWTON METHOD

To solve the optimality system (OS_γ) with a semi-smooth Newton method [Kummer 1992; Chen et al. 2000; Hintermüller et al. 2002; Ulbrich 2002], we consider it as an operator equation $F(x, c) = 0$ for $F : X \times \mathbb{R} \to X^* \times \mathbb{R}$,

$$F(x, c) = \begin{pmatrix} \alpha j(x) + \gamma K^* \left(\max(0, Kx - y^\delta - c) + \min(0, Kx - y^\delta + c)\right) \\ c + \gamma \left\langle -\max(0, Kx - y^\delta - c) + \min(0, Kx - y^\delta + c), 1\right\rangle_{L^2(\Omega)} \end{pmatrix}.$$

We now argue the Newton differentiability of F. Recall that a mapping $F : X \to Y$ between Banach spaces X and Y is Newton differentiable at $x \in X$ if there exists a neighborhood $N(x)$ and a mapping $G : N(x) \to L(X, Y)$ with

$$\lim_{\|h\| \to 0} \frac{\|F(x + h) - F(x) - G(x + h)h\|_Y}{\|h\|_X} \to 0.$$

(Note that in contrast with Fréchet differentiability, the linearization is taken in a neighborhood $N(x)$ of x.) Any mapping $D_N F \in \{G(s) : s \in N(x)\}$ is then a Newton derivative of F at x.

Now we have (e.g., from [Ito and Kunisch 2008, Example 8.14]; see also [Schiela 2008]) that the function $z \mapsto \max(0, z)$ is Newton differentiable from $L^p(\Omega)$ to $L^q(\Omega)$ for any $p > q \geqslant 1$. Furthermore, the chain rule for Newton derivatives (e.g., [Ito and Kunisch 2008, Lemma 8.15]) yields that for a linear operator B with range contained in $L^p(\Omega)$, the Newton derivative of $\max(0, Bv)$ at v in direction δv is given pointwise almost everywhere by

$$\left(D_N \max(0, Bv)\delta v\right)(t) = \begin{cases} (B\delta v)(t), & \text{if } v(t) > 0, \\ 0, & \text{if } v(t) \leqslant 0. \end{cases}$$

Since for any $x \in X$ we have $Kx \in L^\infty(\Omega)$, the mapping

$$x \mapsto \max(0, Kx - y^\delta - c) \tag{4.6}$$

for fixed c is Newton-differentiable from X to $L^1(\Omega) \subset L^\infty(\Omega)^*$ with Newton derivative in direction $\delta x \in X$ given by $(K\delta x)\chi_1$, where

$$\chi_1(t) = \begin{cases} 1 & \text{if } (Kx - y^\delta - c)(t) > 0, \\ 0 & \text{if } (Kx - y^\delta - c)(t) \leqslant 0. \end{cases} \tag{4.7}$$

Similarly, the embedding that maps $c \in \mathbb{R}$ to the constant function $t \mapsto c \in L^\infty(\Omega)$ yields Newton-differentiability of (4.6) with respect to c for fixed $x \in X$ from $\mathbb{R}$ to $L^1(\Omega) \subset L^\infty(\Omega)^*$, with Newton derivative in direction $\delta c \in \mathbb{R}$ given by $(-\delta c)\chi_1$. One proceeds analogously for the min terms by defining

$$\chi_2(t) = \begin{cases} 1 & \text{if } (Kx - y^\delta + c)(t) < 0, \\ 0 & \text{if } (Kx - y^\delta + c)(t) \geqslant 0. \end{cases} \tag{4.8}$$

Altogether, F is Newton-differentiable from $X \times \mathbb{R} \to X^* \times \mathbb{R}$ with Newton derivative at $(x, c) \in X \times \mathbb{R}$ given by

$$D_N F(x,c)(\delta x, \delta c) = \begin{pmatrix} \alpha j'(x)\delta x + \gamma K^*((\chi_1 + \chi_2)K\delta x) + \gamma \delta c K^*(-\chi_1 + \chi_2) \\ \gamma \langle -\chi_1 + \chi_2, K\delta x\rangle_{L^2(\Omega)} + \left(1 + \gamma \langle \chi_1 + \chi_2, 1\rangle_{L^2(\Omega)}\right) \delta c \end{pmatrix},$$

For given (x^k, c^k), a semi-smooth Newton step consists in solving for $(\delta x, \delta c) \in X \times \mathbb{R}$ in

$$D_N F(x^k, c^k)(\delta x, \delta c) = -F(x^k, c^k) \tag{4.9}$$

and setting $x^{k+1} = x^k + \delta x$, $c^{k+1} = c^k + \delta c$. It remains to show uniform invertibility of the Newton step, which will imply local superlinear convergence of the sequence of iterates (x^k, c^k).

Theorem 4.4. *For every $\alpha, \gamma > 0$, the sequence (x^k, c^k) of iterates in (4.9) converges superlinearly to the solution (x_γ, c_γ) to (OS_γ), provided that (x^0, c^0) is sufficiently close to (x_γ, c_γ).*

Proof. For arbitrary $(x, c) \in X \times \mathbb{R}$ and $(\delta x, \delta c) \in X \times \mathbb{R}$, we have

$$\begin{aligned} \langle (\delta x, \delta c), D_N F(x,c)(\delta x, \delta c)\rangle_{(X\times\mathbb{R})^*,(X\times\mathbb{R})} &= \alpha \|\delta x\|_X^2 + \delta c^2 \\ &+ \gamma \left(\|(\chi_1 + \chi_2)K\delta x\|_{L^2(\Omega)}^2 + 2\delta c \langle -\chi_1 + \chi_2, K\delta x\rangle_{L^2(\Omega)} + \delta c^2 \|\chi_1 + \chi_2\|_{L^2(\Omega)}^2 \right). \end{aligned}$$

Since χ_1 and χ_2 are characteristic functions of disjoint sets, we can estimate separately

$$\begin{aligned} \|\chi_1 K\delta x\|_{L^2(\Omega)}^2 - 2\langle \delta c \chi_1, K\delta x\rangle_{L^2(\Omega)} + \|\delta c \chi_1\|_{L^2(\Omega)}^2 &= \|\chi_1(K\delta x - \delta c)\|_{L^2(\Omega)}^2 \geqslant 0, \\ \|\chi_2 K\delta x\|_{L^2(\Omega)}^2 + 2\langle \delta c \chi_2, K\delta x\rangle_{L^2(\Omega)} + \|\delta c \chi_2\|_{L^2(\Omega)}^2 &= \|\chi_2(K\delta x + \delta c)\|_{L^2(\Omega)}^2 \geqslant 0. \end{aligned}$$

This implies

$$\langle (\delta x, \delta c), D_N F(x,c)(\delta x, \delta c)\rangle_{(X\times\mathbb{R})^*,(X\times\mathbb{R})} \geqslant \alpha \|\delta x\|_X^2 + \delta c^2$$

and thus that $D_N F(x, c)$ is an isomorphism independent of (x, c). The local superlinear convergence now follows from standard results (e.g., [Ito and Kunisch 2008, Theorem 8.16]). □

Algorithm 1 Semi-smooth Newton method with continuation

```
Choose (x⁰, c⁰), γ⁰, τ > 1, ε > 0, k*, γ*; set j = 0
repeat
    Increment j ← j + 1
    Set x₀ = x^{j−1}, k = 0
    repeat
        Increment k ← k + 1
        Compute indicator function of active sets : χ₁^k, χ₂^k from (4.7) and (4.8)
        Solve for δx, δc in (4.9)
        Update x^k = x^{k−1} + δx, c^k = c^{k−1} + δc
    until χ₁^{k+1} = χ₁^k and χ₂^{k+1} = χ₂^k, or k = k*
    Set x^j = x_k, c^j = c_k
    Set γ^j = τγ^{j−1}
until ‖Kx^j − y^δ‖_{L∞(Ω)} < c + ε or γ = γ*
```

The following property (e.g., [Ito and Kunisch 2008, Remark 7.1.1]) yields an objective stopping criterion for the semismooth Newton method. Let

$$\mathcal{A}_1^k = \left\{t \in \Omega : (Kx^k - y^\delta - c^k)(t) > 0\right\},$$
$$\mathcal{A}_2^k = \left\{t \in \Omega : (Kx^k - y^\delta + c^k)(t) < 0\right\}.$$

denote the sets of points where the L^∞ norm bound is violated in iteration k.

Proposition 4.5. *If* $\mathcal{A}_1^{k+1} = \mathcal{A}_1^k$ *and* $\mathcal{A}_2^{k+1} = \mathcal{A}_2^k$, *then* $F(x^{k+1}, c^{k+1}) = 0$.

To deal with the local convergence of Newton's method, we make use of a continuation strategy in the numerical computation: Solve (OS_γ) for fixed $\gamma_k > 0$, choose $\gamma_{k+1} > \gamma_k$, and compute the next solution $(x_{\gamma_{k+1}}, c_{\gamma_{k+1}})$ using $(x_{\gamma_k}, c_{\gamma_k})$ as starting point. If γ_0 is sufficiently small (e.g., $\gamma_0 = 1$), one can expect convergence of the continuation scheme for any reasonable choice of (x_0, c_0) (e.g., $(x_0, c_0) = (0, 0)$ in the absence of a priori information). The full procedure for computing a numerical approximation of the solution to problem ($\mathcal{P}_c$) is given as Algorithm 1.

Finally, we remark on how the presented approach can be simplified in special cases.

Remark 4.6. In the case where K is the solution operator for a linear partial differential equation, i.e., $K = A^{-1}$ for a partial differential operator $A : Y \to Y^* \supset X$ on the reflexive Banach space Y, (OS_γ) can be reformulated in a more convenient way by introducing $y = A^{-1}x$ as an independent variable and using a Lagrange multiplier approach to enforce the constraint $Ay = x$. This leads to a (semi-smooth) block optimality system, which in many cases can again be reduced to a pair of equations for (y, c) only. Take $X = L^2(\Omega)$, i.e., $j(x) = x$, and assume that A is an isomorphism from $\mathcal{W} = H_0^1(\Omega) \cap H^2(\Omega)$ to $\mathcal{W}^*$. Due to the embedding $\mathcal{W} \hookrightarrow C_0(\Omega)$, we have that the range $A^{-1}(X)$ embeds compactly into $L^\infty(\Omega)$. Inserting $y_\gamma = A^{-1}x_\gamma \in \mathcal{W}$ into the first equation of (OS_γ) yields

$$\alpha A y_\gamma + \gamma A^{-*}\left(\max(0, y_\gamma - y^\delta - c_\gamma) + \min(0, y_\gamma - y^\delta + c_\gamma)\right) = 0.$$

Since the term in parentheses is in $L^\infty(\Omega)$, the mapping properties of A^{-*} yield that $Ay_\gamma \in \mathcal{W}$. We can thus apply A^* to the whole equation to obtain that (y_γ, c_γ) satisfies $F(y_\gamma, c_\gamma) = 0$ for $F : \mathcal{W} \times \mathbb{R} \to \mathcal{W}^* \times \mathbb{R}$,

$$F(y,c) = \begin{pmatrix} \alpha A^*Ay + \gamma\left(\max(0, y - y^\delta - c) + \min(0, y - y^\delta + c)\right) \\ c + \gamma\left\langle -\max(0, y - y^\delta - c) + \min(0, y - y^\delta + c), 1\right\rangle_{L^2(\Omega)} \end{pmatrix}.$$

Since $y \in \mathcal{W}$, the function F is semi-smooth with Newton-derivative

$$D_N F(y,c)(\delta y, \delta c) = \begin{pmatrix} \alpha A^*A\delta y + \gamma(\chi_1 + \chi_2)\delta y + \gamma\delta c(-\chi_1 + \chi_2) \\ \gamma\left\langle -\chi_1 + \chi_2, \delta y\right\rangle_{L^2(\Omega)} + \left(1 + \gamma\left\langle \chi_1 + \chi_2, 1\right\rangle_{L^2(\Omega)}\right)\delta c \end{pmatrix}.$$

Superlinear convergence of the semi-smooth Newton method can then be proven analogously to Theorem 4.4, using the fact that A is an isomorphism from $\mathcal{W}$ to $\mathcal{W}^*$. Given y_γ, we can then compute $x_\gamma = Ay_\gamma$. Note that due to the linearity of the operators, we can further reformulate the Newton step in terms of the new iterate (y^{k+1}, c^{k+1}) only:

$$\begin{pmatrix} \alpha A^*A + \gamma(\chi_2 + \chi_1) & \gamma(\chi_2 - \chi_1) \\ \gamma\left\langle \chi_2 - \chi_1, \cdot\right\rangle_{L^2(\Omega)} & 1 + \gamma\left\langle \chi_1 + \chi_2, 1\right\rangle_{L^2(\Omega)} \end{pmatrix} \begin{pmatrix} y^{k+1} \\ c^{k+1} \end{pmatrix} = \begin{pmatrix} \gamma(\chi_2 + \chi_1)y^\delta \\ \gamma(\chi_2 - \chi_1)y^\delta \end{pmatrix}.$$

Remark 4.7. The presented approach can also be applied to the Morozov regularization

$$\min_{x \in X} \frac{1}{2}\|x\|_X^2 \quad \text{subject to} \quad \|Kx - y^\delta\|_{L^\infty(\Omega)} \leqslant \delta$$

by fixing $c = \delta$ in the above derivations. Applying the same Moreau–Yosida regularization as above yields the optimality conditions $F(x_\gamma) = 0$ for $F : X \to X^*$,

$$F(x) = \alpha j(x) + \gamma K^*\left(\max(0, Kx - y^\delta - \delta) + \min(0, Kx - y^\delta + \delta)\right),$$

with Newton derivative

$$D_N F(x)\delta x = (\alpha j'(x) + \gamma K^*(\chi_1 + \chi_2)K)\,\delta x.$$

Well-posedness, convergence as $\gamma \to \infty$ and superlinear convergence of the semi-smooth Newton method can be shown as for the Tikhonov regularization (with obvious simplifications).

5 NUMERICAL EXAMPLES

In this section, we illustrate the effectiveness of the L^∞ fitting approach as well as some of its qualitative features by way of one- and two-dimensional model problems. The Matlab implementation for both examples can be downloaded as http://www.uni-graz.at/~clason/codes/linffitting.zip. All numerical tests were performed with Matlab (R2011b) on a single core of a 3.4 GHz workstation with 16 GByte of RAM.

5.1 INVERSE HEAT CONDUCTION PROBLEM

We first consider as a standard benchmark example an inverse heat conduction problem, posed as a Volterra integral equation of the first kind (problem `heat` in [Hansen 2007]).[1] Here, $\Omega = (0,1)$, $X = L^2(\Omega)$, and $(Kx)(t) = \int_0^t k(s,t)x(s)\,ds$. Hence, K is a compact linear operator from $L^2(\Omega)$ to $L^\infty(\Omega)$. The kernel $k(s,t)$ and the exact solution $x^\dagger(t)$ are given by

$$k(s,t) = \frac{(s-t)^{-\frac{3}{2}}}{2\sqrt{\pi}} e^{-\frac{1}{4(s-t)}}, \qquad x^\dagger(t) = \begin{cases} 75t^2 & 0 \leqslant t \leqslant \frac{1}{10}, \\ \frac{3}{4} + (20t-2)(3-20t) & \frac{1}{10} < t \leqslant \frac{3}{20}, \\ \frac{3}{4}e^{-2(20t-3)} & \frac{3}{20} < t \leqslant \frac{1}{2}, \\ 0 & \text{otherwise.} \end{cases}$$

The noisy data are generated by setting

$$y^\delta(t) = Kx^\dagger(t) + \xi(t), \qquad t \in (0,1),$$

where $\xi(t)$ is a uniformly distributed random value in the range $[-d\,y_{\max}, d\,y_{\max}]$ for a noise parameter $d > 0$ and $y_{\max} = \|Kx^\dagger\|_\infty$.

For the numerical solution of the inverse problem $Kx = y^\delta$, we apply Algorithm 1 (with $\tau = 10$, $k^* = 10$, $(x^0, c^0) = (0,0)$, $\gamma^0 = 1$ and $\gamma^* = 10^{12}$) and discretize the integral equation using collocation and the mid-point rule at $n = 300$ points (unless stated otherwise). The parameters in the fixed-point iteration for the automatic parameter choice are set to $\alpha_0 = 0.1$ and $\sigma = 0.008$. The fixed-point iteration is terminated if the relative change in α is less than 10^{-3} or after 20 iterations.

A typical realization of noisy data is displayed in Figure 1a for $d = 0.3$ and Figure 1c for $d = 0.6$. The fixed-point iteration (3.2) converged after 6 (4) iterations for $d = 0.3$ ($d = 0.6$), and yielded the values 6.50×10^{-3} (1.65×10^{-2}) for the regularization parameter α. The respective reconstructions x_α are shown in Figures 1b and 1d. To measure the accuracy of the solution x_α quantitatively, we compute the L^2-error $e = \|x_\alpha - x^\dagger\|_{L^2}$, which is 2.48×10^{-2} for $d = 0.3$ and 7.56×10^{-2} for $d = 0.6$. For comparison, we also show the solution to the L^2 data fitting problem, where the parameter α has been chosen to give the smallest L^2 error. Clearly, the L^2 reconstructions are significantly less accurate than their L^∞ counterparts, especially at the "tail".

The performance of the automatic parameter choice is further illustrated in Table 1, which compares the balancing parameter α_b with the "optimal", sampling-based parameter α_o for different noise levels. This parameter is obtained by sampling each interval $[0.1\alpha_b, \alpha_b]$ and $[\alpha_b, 10\alpha_b]$ uniformly with 51 parameters and taking as α_o the one with smallest L^2-error $e_o = \|x_{\alpha_o} - x^\dagger\|_{L^2}$. Presented in each case are the mean and standard deviation over ten different noise realizations for a given noise level. Both the regularization parameters and the reconstruction errors agree closely, and the noise level is well estimated by the optimal L^∞ bound c_b. Table 1 also illustrates the robustness of the L^∞ data fitting, since the reconstruction

[1] MATLAB code (version 4.1) and documentation is available from http://www2.imm.dtu.dk/~pch/Regutools.

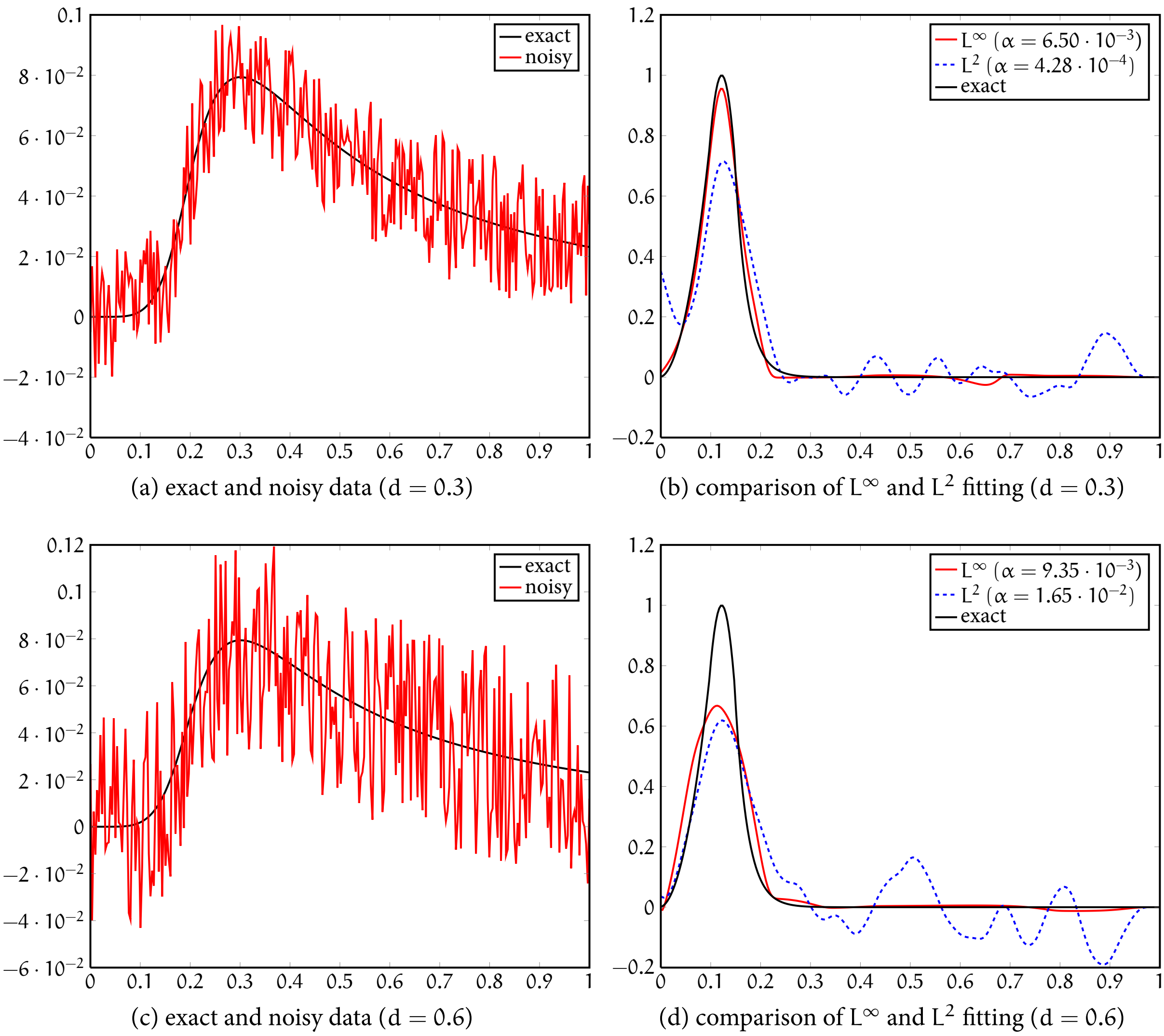


Figure 1: Results for inverse heat conduction problem

Table 1: Comparison of automatic parameter choice (estimated noise level c_b, parameter α_b, reconstruction error e_b) with sampling based optimal choice (α_o, e_o) for different noise parameters d and noise levels δ (shown are the mean and standard deviation over ten different noise realizations)

d	δ		c_b		α_b		e_b		α_o		e_o	
	mean	std	mean	std	mean	std	mean	std	mean	std	mean	std
0.1	7.90e-3	3.7e-5	7.79e-3	1.6e-4	2.26e-3	1.3e-4	4.06e-2	1.7e-2	2.38e-3	7.9e-4	3.76e-2	1.9e-2
0.2	1.58e-2	1.0e-4	1.57e-2	5.1e-4	4.77e-3	4.7e-4	5.06e-2	8.3e-3	4.58e-3	1.8e-3	4.47e-2	1.2e-2
0.3	2.37e-2	5.4e-5	2.35e-2	2.8e-4	7.56e-3	7.0e-4	6.81e-2	1.6e-2	5.62e-3	3.9e-3	6.48e-2	1.9e-2
0.4	3.16e-2	8.8e-5	3.10e-2	1.9e-4	9.66e-3	8.2e-4	6.06e-2	2.1e-2	8.49e-3	5.6e-3	5.58e-2	2.4e-2
0.5	3.96e-2	7.1e-5	3.90e-2	5.7e-4	1.25e-2	1.2e-3	6.29e-2	2.0e-2	1.15e-2	7.0e-3	6.21e-2	2.0e-2
0.6	4.75e-2	1.2e-4	4.67e-2	4.5e-4	1.43e-2	1.8e-3	6.43e-2	3.5e-2	1.39e-2	9.4e-3	5.80e-2	2.9e-2
0.7	5.53e-2	2.1e-4	5.47e-2	4.9e-4	1.85e-2	3.4e-3	7.77e-2	3.2e-2	2.85e-2	2.4e-2	7.27e-2	3.0e-2
0.8	6.34e-2	1.8e-4	6.27e-2	8.1e-4	2.30e-2	4.6e-3	8.81e-2	3.8e-2	2.22e-2	1.2e-2	8.72e-2	3.8e-2
0.9	7.13e-2	1.3e-4	7.03e-2	6.4e-4	2.60e-2	4.2e-3	1.15e-1	3.0e-2	3.07e-2	2.2e-2	1.13e-1	3.1e-2

Table 2: Convergence behavior of the semi-smooth Newton method for fixed $\gamma = 10^2$ (shown are the number of points $n(k)$ that changed in the active sets after iteration k)

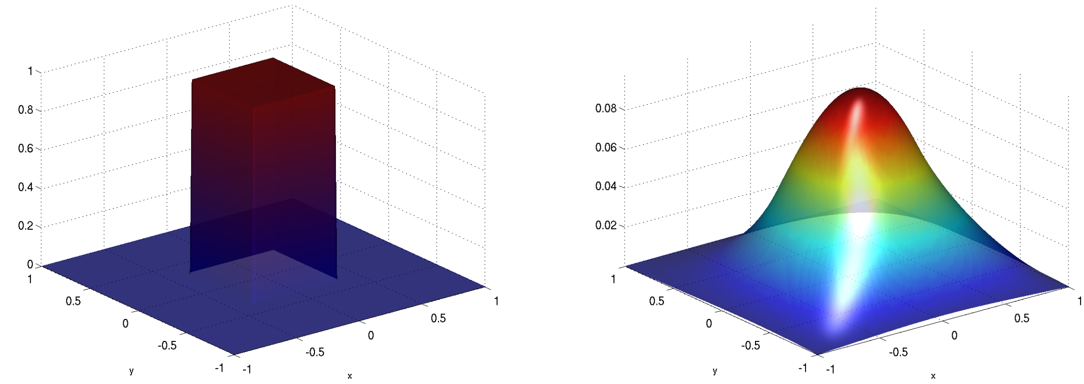
k	1	2	3	4	5	6	7	8
$n(k)$	144	83	39	19	8	1	1	0

error does not significantly increase with increasing noise level. This can be attributed to the fact that the structural properties of the noise (e.g., sign changes of the noise, which is neither more nor less likely for increasing d) is more important than the magnitude.

Finally, we address the performance of the semi-smooth Newton method. Table 2 shows the convergence history of the Newton iteration (in terms of the number of changed points in the active sets $\mathcal{A}_k^+, \mathcal{A}_k^-$ after each iteration) for $d = 0.3$, fixed α computed by the balancing principle and fixed $\gamma = 10^2$, corroborating both the local superlinear convergence (Theorem 4.4) and the finite termination property (Proposition 4.5). The behavior of the full continuation strategy is similarly illustrated in Table 3, demonstrating that a feasible solution (i.e., one attaining the L^∞ bound) is reached at $\gamma = 10^6$ with comparative computational effort.

Table 3: Convergence behavior of the semi-smooth Newton method with continuation (shown are the number of points $n(k)$ that changed in the active sets after iteration k)

γ	1e0				1e1				1e2				1e3					1e4				1e5		1e6
k	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	5	1	2	3	4	1	2	1
$n(k)$	115	25	7	0	91	35	5	0	38	15	5	0	15	8	4	2	0	5	3	1	0	1	0	0



(a) exact solution $x^\dagger$ (b) exact data $y^\dagger$

Figure 2: Two-dimensional test problem: exact solution $x^\dagger$ and data $y^\dagger$

5.2 INVERSE SOURCE PROBLEM IN 2D

Motivated by the problem of detecting the source of a contaminant using distributed and quantized measurements from a sensor network, we consider the inverse source problem for an elliptic partial differential operator on the domain $\Omega \subset \mathbb{R}^2$ with homogeneous Dirichlet boundary conditions, i.e., $K = A^{-1}$, $X = L^2(\Omega)$, where

$$Ay = -a\Delta y + \langle b, \nabla y\rangle_{L^2(\Omega)} + fy$$

for $a \in C^{0,r}(\Omega)$ with $r > 0$ and $a \geqslant a_0 > 0$ pointwise, $b \in C^{0,r}(\Omega)^2$, $f \in L^\infty(\Omega)$ with $f - \nabla \cdot b \geqslant 0$ pointwise. This guarantees (for Ω smooth or a parallelepiped) that A is an isomorphism from $\mathcal{W} = H_0^1(\Omega) \cap H^2(\Omega)$ to $\mathcal{W}^*$ and hence that $y \in C^0(\overline{\Omega})$. Here, we choose $a = 1$, $b = (-2, 0)^T$, $f = 0$, and $\Omega = [0, 1]^2$. The exact solution is given by

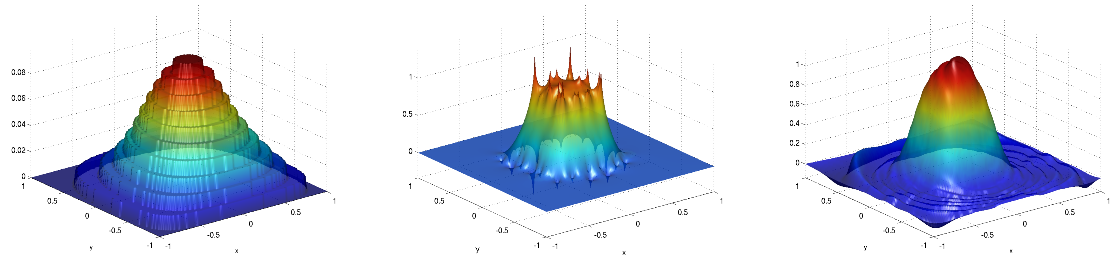
$$x^\dagger(t_1, t_2) = \begin{cases} 1 & \text{if } |t_1| \leqslant \frac{1}{3} \text{ and } |t_2| \leqslant \frac{1}{3}, \\ 0 & \text{otherwise,} \end{cases}$$

see Figure 2a. The exact data $y^\dagger = A^{-1}x^\dagger$ are shown in Figure 2b.

For the numerical solution of the inverse problem $u = A^{-1}y^\delta$, we apply the reformulated algorithm according to Remark 4.6, and discretize the differential operators using standard finite differences on a uniform mesh of size 128×128. The parameters in the semi-smooth Newton method with continuation and the fixed-point iteration are identical to the one-dimensional case.

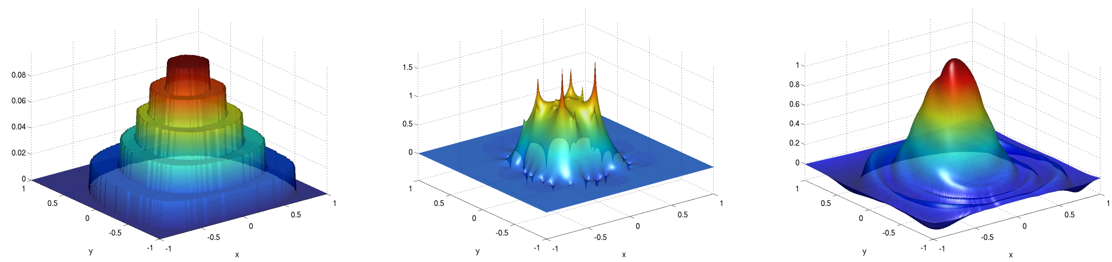
The first example considers data subject to (deterministic) quantization errors, where we set

$$y^\delta(t) = y_s \left[\frac{y^\dagger(t)}{y_s}\right], \qquad y_s = n_b^{-1}\left(\max_{t\in\Omega}\left(y^\dagger(t)\right) - \min_{t\in\Omega}\left(y^\dagger(t)\right)\right),$$



(a) data y^δ (b) L^∞ reconstruction x_α (c) L^2 reconstruction x_α

Figure 3: Reconstructions from quantized data ($n_b = 10$), comparing L^∞ and L^2 fitting



(a) data y^δ (b) L^∞ reconstruction x_α (c) L^2 reconstruction x_α

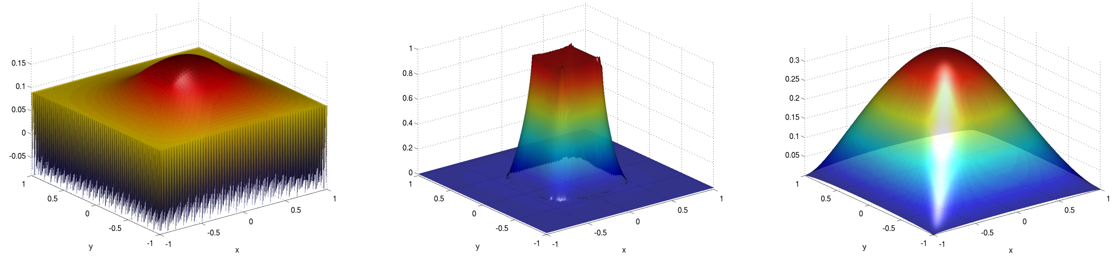
Figure 4: Reconstructions from quantized data ($n_b = 5$), comparing L^∞ and L^2 fitting

with n_b denoting the number of bins and $[s]$ denoting the nearest integer to $s \in \mathbb{R}$ (i.e., the data are rounded to n_b discrete equispaced values, see Figure 3a for $n_b = 10$ and Figure 4a for $n_b = 5$). Again we compare the solutions to the L^∞ fitting problem (where the regularization parameter is chosen using the fixed-point iteration) with reconstructions obtained from standard L^2 fitting (where the parameter is exhaustively selected to yield the lowest L^2 error) in Figures 3b, 3c and 4b, 4c. The difference in reconstruction artifacts can be observed clearly: The L^∞ artifacts are strongly localized and impulse-like, whereas the L^2 reconstruction shows typical ringing. In particular, the support of the exact solution $x^\dagger$ is accurately captured by the L^∞ reconstruction, whereas the L^2 reconstruction is non-zero everywhere.

The second example serves as a "best-case" noise for L^∞ fitting. Based on our observation in the one-dimensional case, we conjecture that the reconstruction error is largest in regions where the sign of the noise does not change. We therefore choose as additive "noise" a checkerboard pattern on the discrete mesh of constant magnitude and alternating sign. Specifically, let $t_{ij} = (t_{1,i}, t_{2,j})$, $1 \leqslant i, j \leqslant 128$, be the grid points of the uniform mesh and set

$$y^\delta(t_{ij}) = y^\dagger(t_{ij}) + (-1)^{i+j} d \|y^\dagger\|_\infty$$

for a noise parameter $d > 0$. For data with $d = 0.9$ (Figure 5a), the L^∞ reconstruction is able to accurately capture support and shape of the true solution, whereas the L^2 reconstruction is



(a) data y^δ (b) L^∞ reconstruction x_α (c) L^2 reconstruction x_α

Figure 5: Reconstructions from checkerboard noise ($d = 0.9$), comparing L^∞ and L^2 fitting

Table 4: Comparison of reconstruction errors for L^∞ fitting (e_∞) and L^2 fitting (e_2) for checkerboard noise of different magnitude d

d	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
e_∞	0.10961	0.1096	0.10961	0.10959	0.10963	0.10962	0.10961	0.1096	0.1096
e_2	0.15625	0.19176	0.22549	0.25568	0.26619	0.26829	0.26978	0.27086	0.27173

far from the target. The robustness of L^∞ fitting in this case is further illustrated by Table 4, where it can be seen that the reconstruction error is virtually independent of the magnitude of the checkerboard noise, in contrast to L^2 fitting.

6 CONCLUSION

For measurements subject to uniformly distributed noise, such as arising from statistical models of quantization errors, L^∞ fitting is more robust than standard L^2 fitting. The non-differentiability can be addressed by introducing a Moreau–Yosida regularization together with a continuation scheme, which allows application of a superlinearly convergent semi-smooth Newton method. The regularization parameter can be chosen automatically using a heuristic choice rule that does not require knowledge of the noise level. This approach is useful for a wide variety of linear inverse problems.

For nonlinear problems, the extension would be straightforward (subject to a usual nonlinearity and second order condition, see [Clason and Jin 2012]). By combining the methods of the current work with those of [Clason and Jin 2012], Tikhonov functionals of L^∞-L^1 type (i.e., L^∞ fitting with regularization terms of L^1 type, also known as "Dantzig selector" [Candes and Tao 2007]) could be treated. Finally, a stochastic analogue of the considered uniform noise models in function spaces would be of great interest.

ACKNOWLEDGMENTS

Part of this work was completed while visiting the Isaac Newton Institute for Mathematical Sciences, and the author thanks the institute for the hospitality. Support by the Austrian Science Fund (FWF) under grant SFB F32 (SFB "Mathematical Optimization and Applications in Biomedical Sciences") is gratefully acknowledged.

REFERENCES

Arampatzis, T., Lygeros, J., and Manesis, S. (2005). *A survey of applications of wireless sensors and wireless sensor networks*. In: Proceedings of the 13th IEEE International Conference on Control and Automation. IEEE, pp. 719–724. DOI: 10.1109/.2005.1467103.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press, Cambridge.

Burger, M. and Osher, S. (2004). *Convergence rates of convex variational regularization*. Inverse Problems 20.5, pp. 1411–1421. DOI: 10.1088/0266-5611/20/5/005.

Candes, E. and Tao, T. (2007). *The Dantzig selector: statistical estimation when* p *is much larger than* n. Ann. Statist. 35.6, pp. 2313–2351. DOI: 10.1214/009053606000001523.

Chen, X., Nashed, Z., and Qi, L. (2000). *Smoothing methods and semismooth methods for nondifferentiable operator equations*. SIAM J. Numer. Anal. 38.4, pp. 1200–1216. DOI: 10.1137/S0036142999356719.

Clason, C., Ito, K., and Kunisch, K. (2010). *Minimal invasion: An optimal* L^∞ *state constraint problem*. ESAIM: Math. Model. Numer. Anal. 45.3, pp. 505–522. DOI: 10.1051/m2an/2010064.

Clason, C. and Jin, B. (2012). *A semi-smooth Newton method for nonlinear parameter identification problems with impulsive noise*. SIAM J. Imaging Sci. 5, pp. 505–536. DOI: 10.1137/110826187.

Clason, C., Jin, B., and Kunisch, K. (2010a). *A duality-based splitting method for* ℓ^1*-TV image restoration with automatic regularization parameter choice*. SIAM J. Sci. Comput. 32.3, pp. 1484–1505. DOI: 10.1137/090768217.

Clason, C., Jin, B., and Kunisch, K. (2010b). *A semismooth Newton method for* L^1 *data fitting with automatic choice of regularization parameters and noise calibration*. SIAM J. Imaging Sci. 3.2, pp. 199–231. DOI: 10.1137/090758003.

Doolin, D. and Sitar, N. (2005). *Wireless sensors for wildfire monitoring*. In: Proc. SPIE. Vol. 5765, pp. 477–484. DOI: 10.1117/12.605655.

Engl, H. W., Hanke, M., and Neubauer, A. (1996). *Regularization of Inverse Problems*. Kluwer, Dordrecht, pp. viii+321.

Flemming, J. and Hofmann, B. (2011). *Convergence rates in constrained Tikhonov regularization: equivalence of projected source conditions and variational inequalities*. Inverse Problems 27.8, p. 085001. DOI: 10.1088/0266-5611/27/8/085001.

Gharavi, H. and Kumar, S., eds. (2003). *Proceedings of the IEEE: Special issue on sensor networks and applications*. Vol. 91. 8. IEEE, pp. 1151–1163. DOI: 10.1109/JPROC.2003.814925.

Grund, T. and Rösch, A. (2001). *Optimal control of a linear elliptic equation with a supremum norm functional*. Optim. Methods Softw. 15.3-4, pp. 299–329. DOI: 10.1080/10556780108805823.

Hansen, P. C. (2007). *Regularization Tools version 4.0 for Matlab 7.3*. Numer. Algorithms 46.2, pp. 189–194. DOI: 10.1007/s11075-007-9136-9.

Hartley, R. I. and Schaffalitzky, F. (2004). *L-infinity Minimization in Geometric Reconstruction Problems*. In: CVPR, pp. 504–509. DOI: 10.1109/CVPR.2004.140.

Hintermüller, M., Ito, K., and Kunisch, K. (2002). *The primal-dual active set strategy as a semismooth Newton method*. SIAM J. Optim. 13.3, 865–888 (2003). DOI: 10.1137/S1052623401383558.

Hofmann, B., Kaltenbacher, B., Pöschl, C., and Scherzer, O. (2007). *A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators*. Inverse Problems 23.3, pp. 987–1010. DOI: 10.1088/0266-5611/23/3/009.

Ito, K. and Kunisch, K. (2008). *Lagrange Multiplier Approach to Variational Problems and Applications*. Vol. 15. Advances in Design and Control. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

Kindermann, S. (2011). *Convergence analysis of minimization-based noise level-free parameter choice rules for linear ill-posed problems*. Electron. Trans. Numer. Anal. 38, pp. 233–257. URL: http://etna.math.kent.edu/vol.38.2011/pp233-257.dir/.

Kummer, B. (1992). *Newton's method based on generalized derivatives for nonsmooth functions: convergence analysis*. In: Advances in optimization (Lambrecht, 1991). Vol. 382. Lecture Notes in Econom. and Math. Systems. Springer, Berlin, pp. 171–194.

Maurer, H. and Zowe, J. (1979). *First and second order necessary and sufficient optimality conditions for infinite-dimensional programming problems*. Math. Programming 16.1, pp. 98–110. DOI: 10.1007/BF01582096.

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

Niu, R. and Varshney, P. (2006). *Target location estimation in sensor networks with quantized data*. IEEE Transactions on Signal Processing 54.12, pp. 4519–4528. DOI: 10.1109/TSP.2006.882082.

Polastre, J., Szewczyk, R., Mainwaring, A., Culler, D., and Anderson, J. (2004). *Analysis of Wireless Sensor Networks for Habitat Monitoring*. In: Wireless Sensor Networks. Ed. by C. S. Raghavendra, K. M. Sivalingam, and T. Znati. Springer US, pp. 399–423. DOI: 10.1007/978-1-4020-7884-2_18.

Pöschl, C. (2009). *An overview on convergence rates for Tikhonov regularization methods for non-linear operators*. J. Inverse Ill-Posed Probl. 17.1, pp. 77–83. DOI: 10.1515/JIIP.2009.009.

Prüfert, U. and Schiela, A. (2009). *The minimization of a maximum-norm functional subject to an elliptic PDE and state constraints*. ZAMM 89.7, pp. 536–551. DOI: 10.1002/zamm.200800097.

Resmerita, E. (2005). *Regularization of ill-posed problems in Banach spaces: convergence rates*. Inverse Problems 21.4, pp. 1303–1314. DOI: 10.1088/0266-5611/21/4/007.

Resmerita, E. and Scherzer, O. (2006). *Error estimates for non-quadratic regularization and the relation to enhancement*. Inverse Problems 22.3, pp. 801–814. DOI: 10.1088/0266-5611/22/3/004.

Scherzer, O., Grasmair, M., Grossauer, H., Haltmeier, M., and Lenzen, F. (2009). *Variational Methods in Imaging*. Vol. 167. Applied Mathematical Sciences. Springer, New York.

Schiela, A. (2008). *A simplified approach to semismooth Newton methods in function space*. SIAM J. Optim. 19.3, pp. 1417–1432. DOI: 10.1137/060674375.

Schizas, I., Giannakis, G., and Luo, Z. (2007). *Distributed estimation using reduced-dimensionality sensor observations*. IEEE Transactions on Signal Processing 55.8, pp. 4284–4299. DOI: 10.1109/TSP.2007.895987.

Seo, Y. and Hartley, R. (2007). *A Fast Method to Minimize L^∞ Error Norm for Geometric Vision Problems*. In: ICCV, pp. 1–8. DOI: 10.1109/ICCV.2007.4408913.

Shykula, M. and Seleznjev, O. (2006). *Stochastic structure of asymptotic quantization errors*. Statist. Probab. Lett. 76.5, pp. 453–464. DOI: 10.1016/j.spl.2005.08.022.

Sim, K. and Hartley, R. (2006). *Removing Outliers Using The L^∞ Norm*. In: CVPR. IEEE Computer Society, Washington, DC, USA, pp. 485–494. DOI: 10.1109/CVPR.2006.253.

Ulbrich, M. (2002). *Semismooth Newton methods for operator equations in function spaces*. SIAM J. Optim. 13.3, 805–842 (2003). DOI: 10.1137/S1052623400371569.

Widrow, B. and Kollár, I. (2008). *Quantization Noise: Roundoff Error in Digital Computation, Signal Processing, Control, and Communications*. Cambridge University Press, Cambridge, UK.

Williams, J. and Kalogiratou, Z. (1993a). *Least squares and Chebyshev fitting for parameter estimation in ODEs*. Adv. Comput. Math. 1.3-4, pp. 357–366. DOI: 10.1007/BF02072016.

Williams, J. and Kalogiratou, Z. (1993b). *Nonlinear Chebyshev fitting from the solution of ordinary differential equations*. Numer. Algorithms 5.1-4. Algorithms for approximation, III (Oxford, 1992), pp. 325–337. DOI: 10.1007/BF02108466.