Mathematics

Research article

Oversmoothing regularization with ℓ^1-penalty term

Daniel Gerth* and Bernd Hofmann

Faculty for Mathematics, Chemnitz University of Technology, 09107 Chemnitz, Germany

* **Correspondence:** Email: daniel.gerth@mathematik.tu-chemnitz.de; Tel: +49 371 531 30048;
  Fax: +49 371 531 22509.

Abstract: In Tikhonov-type regularization for ill-posed problems with noisy data, the penalty functional is typically interpreted to carry a-priori information about the unknown true solution. We consider in this paper the case that the corresponding a-priori information is too strong such that the penalty functional is oversmoothing, which means that its value is infinite for the true solution. In the case of oversmoothing penalties, convergence and convergence rate assertions for the regularized solutions are difficult to derive, only for the Hilbert scale setting convincing results have been published. We attempt to extend this setting to ℓ^1-regularization when the solutions are only in ℓ^2. Unfortunately, we have to restrict our studies to the case of bounded linear operators with diagonal structure, mapping between ℓ^2 and a separable Hilbert space. But for this subcase, we are able to formulate and to prove a convergence theorem, which we support with numerical examples.

1. Introduction

In this paper we consider the problem of finding a stable approximate solution to the ill-posed linear operator equation

$$Ax = y, \tag{1.1}$$

where $A : X := \ell^2 \to Y$ is a bounded linear operator with non-closed range, i.e., range$(A) \neq \overline{\text{range}(A)}$, in the separable Hilbert space Y with norm $\| \cdot \|$. We recall that for $1 \le p \le \infty$ the spaces ℓ^p are Banach spaces consisting of (infinite) series $x = \{x_i\}_{i=1}^\infty$ with norms $\|x\|_{\ell^p} = \sqrt[p]{\sum_{i=1}^\infty |x_i|^p}$ for $1 \le p < \infty$ and $\| \cdot \|_{\ell^\infty} = \sup_{i \in \mathbb{N}} |x_i|$. For $p = 2$ one obtains a Hilbert space. We assume that in (1.1) only noisy data y^δ is available satisfying the inequality

$$\|y - y^\delta\| \le \delta \tag{1.2}$$

with known noise level $\delta > 0$. In order to determine a stable solution to (1.1) from the noisy data, regularization is required. We are searching for approximations x_α^δ of the exact solution $x^\dagger$ with $Ax^\dagger = y$ based on the noisy data y^δ, which are calculated as ℓ^1-regularized solutions

$$x_\alpha^\delta := \arg\min_{x \in \ell^2} \left\{ \frac{1}{2} \|Ax - y^\delta\|^2 + \alpha\|x\|_{\ell^1} \right\} \tag{1.3}$$

with regularization parameters $\alpha > 0$. This approach was extensively studied in the last decade and we refer to the brief discussion in Section 4. We remark that the Banach space ℓ^1 is non-reflexive, which makes its analysis challenging. In particular, it was shown in [3, Prop 3.3] that (1.1) is always ill-posed when the domain of A is restricted to ℓ^1 and Y is a reflexive Banach space. On the other hand, ℓ^1-regularization in the case $\|x^\dagger\|_{\ell^1} < \infty$ is well understood as a prominent version of sparsity promoting regularization. Hence, it seems to be of great interest whether the ℓ^1-regularization also makes sense in the *oversmoothing* case, which is characterized by the fact that $x^\dagger \in \ell^2 \setminus \ell^1$ such that $\|x^\dagger\|_{\ell^1} = \infty$. The scenario of oversmoothing regularization has, to the best of the authors knowledge, only been treated in the setting of regularization in Hilbert scales, see Section 2. There, the specific structure of the problem yields link conditions as tools to jump between different scale levels and thus to handle the occurring analytical problems arising from the oversmoothing character of the penalty. For other space settings the corresponding tools are missing, and this paper is an attempt to overcome this deficit. We therefore stress that in this paper we are not interested in specific applications, nor in improving any sort of reconstruction quality. Equation (1.1) and procedure (1.3) merely constitute a new trial case for modeling the oversmoothing regularization.

Let us now discuss our motivation for investigating this regularization idea. As we will see in our model problem, oversmoothing ℓ^1-regularization already has one advantage over the classical Tikhonov regularization with penalty $\|\cdot\|_{\ell^2}^2$, i.e.,

$$x_\alpha^\delta := \arg\min_{x \in \ell^2} \left\{ \frac{1}{2} \|Ax - y^\delta\|^2 + \alpha\|x\|_{\ell^2}^2 \right\}. \tag{1.4}$$

Namely, it does not suffer from the saturation effect of the ℓ^2-regularized approach, and the proposed a priori choice of the regularization parameter coincides with the discrepancy principle for all cases under consideration. Hence, one motivation for oversmoothing regularization are possibly improved convergence properties. But there are more potential benefits. An oversmoothing penalty might yield an easier optimization problem. For example, when $x^\dagger \in \ell^p \setminus \ell^2$ with $p > 2$, classical Tikhonov regularization (1.4) would be oversmoothing. However, since it is much easier to find the minimizer of the ℓ^2-penalized Tikhonov functional (which corresponds to solving a linear system) than minimizing a functional with arbitrary ℓ^p-penalty, the former might be preferred. A third aspect for oversmoothing regularization is feature extraction. In terms of our model problem, ℓ^1-regularization is known to yield sparse regularized solutions, i.e., only finitely many components are non-zero. As we will see later, ℓ^1-regularization in some sense selects the best (least) discretization level to achieve a certain residual. Therefore, the approximate solutions need less storage space than ℓ^2-regularized solutions. In general,

other features than sparsity could be of interest. Finally, the analysis of oversmoothing regularization is challenging and might provide new insights for general regularization approaches.

The paper is organized as follows: In Section 2 we sketch the existing results on oversmoothing regularization in Hilbert scales. We proceed with properties of ℓ^1-regularized solutions and discuss their implications in Section 3. After that we move to convergence rates for ℓ^1-regularization including the oversmoothing case in Section 4. In this context, we restrict our studies to diagonal operators and present a specific simplified model problem, which is the basis for the subsequent numerical experiments. In Section 5 we compare our derived convergence rates with the ones known from classical Tikhonov regularization with ℓ^2-penalty. Finally, numerical experiments supporting the theoretical results are presented in Section 6.

2. Literature survey

To the best of the authors knowledge, oversmoothing regularization has so far only been investigated in the setting of regularization in Hilbert scales, starting with the seminal paper by Natterer [6]. There, Natterer considers a linear problem (1.1) between Hilbert spaces X and Y. The regularized solutions are obtained by the minimization problem

$$x_\alpha^\delta := \arg\min_{x \in X} \left\{ \|Ax - y^\delta\|^2 + \alpha\|x\|_p^2 \right\}, \tag{2.1}$$

where the penalty is a norm in a Hilbert scale. In this context, a family $\{X_s\}_{s \in \mathbb{R}}$ of Hilbert spaces with $X_0 := X$ is called Hilbert scale if $X_t \subseteq X_s$ whenever $s < t$ and the inclusion is a continuous embedding, i.e., there exists $c_{s,t} > 0$ such that $\|x\|_s \leq c_{s,t}\|x\|_t$. The Hilbert scale is induced by an unbounded, self-adjoint, and strictly positive definite operator T in X such that $\|x\|_s := \|T^s x\|_X$, $s \in \mathbb{R}$, defines a norm in X_s.

Under the noise model (1.2) in combination with the smoothness assumption $\|x^\dagger\|_q \leq \rho$ for some $q > 0$, and provided that there exists some $a > 0$ such that with two constants $m, M > 0$ and for all $x \in X$ the inequality chain

$$m\|x\|_{-a} \leq \|Ax\| \leq M\|x\|_{-a} \tag{2.2}$$

holds true, Natterer shows in [6] the error estimate

$$\|x_\alpha^\delta - x^\dagger\|_X \leq C\delta^{\frac{q}{a+q}}\rho^{\frac{a}{a+q}} \tag{2.3}$$

whenever the penalty smoothness p in (2.1) is sufficiently high, in the sense that $p \geq (q - a)/2$, and for the a-priori parameter choice

$$\alpha(\delta) = c(\rho)\delta^{\frac{2(a+p)}{a+q}} \tag{2.4}$$

with appropriately chosen constant $c(\rho) > 0$. It is interesting that here the index p, which characterizes the smoothness of the penalty in (2.1), is limited for obtaining the rate result (2.3) by a lower bound, but not by an upper bound. Really, Natterer states that "there is nothing wrong with high order regularization, even well above the order of the smoothness of the exact solution". Even though $x^\dagger$ may have arbitrarily small q-norms with $q > 0$, one still obtains the order optimal rate of convergence. The only adjustment to be made is an appropriate decrease of the regularization parameter.

Recently, Natterer's results have been extended to ill-posed inverse problems with nonlinear forward operators by Mathé and Hofmann [7, 8]. In the first paper, the regularization parameter is

chosen according to the discrepancy principle, while the second paper shows convergence rates under, in principle, the same a-priori parameter choice as occurring in (2.4). In both papers the obtained convergence rate coincides with that in (2.3), i.e., $\|x_\alpha^\delta - x^\dagger\|_X = O\left(\delta^{\frac{q}{a+q}}\right)$ as $\delta \to 0$.

3. Properties of regularized solutions

In this section we discuss some basic properties of oversmoothing ℓ^1-regularization or the lack thereof. The results are easily generalized to arbitrary penalty functionals in Tikhonov-type regularization, as long as certain standard assumptions are fulfilled. We refer to, e.g., [9, Assumption 3.11 and Assumption 3.22] or [10, Assumption 8.1] for details. Basic properties of ℓ^1-regularization are well known. As the following lemma shows, existence and stability of regularized solutions with respect to small perturbations in the data are not influenced by the oversmoothing setting.

Lemma 3.1. *For all fixed $\alpha, \delta > 0$, a minimizer of (1.3) exists and has a finite ℓ^1-norm. Furthermore, if $\{y_k\}$ is a sequence norm-converging in Y to y^δ, i.e. $y_k \to y^\delta$, then every associated sequence $\{x_k\}$ with*

$$x_k \in \arg\min_{x \in \ell^2} \left\{ \frac{1}{2}\|Ax - y_k\|^2 + \alpha\|x\|_{\ell^1} \right\}$$

has a subsequence that converges weakly. The weak limit of every such convergent subsequence $\{x_{\tilde{k}}\}$ of $\{x_k\}$ is a minimizer $\tilde{x}$ of (1.3). Moreover, $\|x_{\tilde{k}}\|$ converges to $\|\tilde{x}\|$.

Proof. Take $\alpha > 0$ arbitrary but fixed. Since $\ell^1 \subset \ell^2$, there exists $\tilde{x} \in \ell^2$ with finite values $\|A\tilde{x} - y^\delta\|$ and $\|\tilde{x}\|_{\ell^1}$. This implies

$$\|x_k\|_{\ell^1} \leq \frac{1}{\alpha}\left(\frac{1}{2}\|Ax_k - y_k\|^2 + \alpha\|x_k\|_{\ell^1}\right) \leq \frac{1}{\alpha}\left(\frac{1}{2}\|A\tilde{x} - y_k\|^2 + \alpha\|\tilde{x}\|_{\ell^1}\right) < \infty$$

by the optimality of the minimizer. The remaining part of the proof follows from standard arguments [9, 10]. $\qquad \square$

We move on to identify some necessary conditions for the convergence of regularized solutions for $\delta \to 0$ in the case of oversmoothing regularization.

Theorem 3.1. *Let $x^\dagger \in \ell^2$ with $\|x^\dagger\|_{\ell^1} = \infty$ denote a solution to (1.1). If the Tikhonov regularized solutions (1.3) under consideration are weakly convergent to $x^\dagger$ for an a priori or a posteriori parameter choice rule $\alpha = \alpha(\delta, y^\delta)$, then the following items must hold for a sequence $x_k := x_{\alpha_k}^{\delta_k} \rightharpoonup x^\dagger$ as $\delta_k \to 0$:*

a) $\lim_{k \to \infty} \|x_k\|_{\ell^1} = \infty$,
b) $\lim_{k \to \infty} \alpha_k = 0$,
c) $\lim_{k \to \infty} \alpha_k \|x_k\|_{\ell^1} \leq C < \infty$.

Proof. By Lemma 3.1, $\|x_k\|_{\ell^1} < \infty$ for all $k \in \mathbb{N}$. If, however, we assume that there is a subsequence $\{x_{k_j}\}$ with $\|x_{k_j}\|_{\ell^1} \leq c < \infty$ uniformly for all $j \in \mathbb{N}$, then the assumed weak convergence $x_{k_j} \rightharpoonup x^\dagger$ in X implies that $\|x^\dagger\|_{\ell^1} \leq c$. This contradicts the assumption $\|x^\dagger\|_{\ell^1} = \infty$ and yields item a) of the theorem.

Now take some fixed $\tilde{x} \in \ell^1$ and keep in mind the definition of the x_k as minimizers of the functional (1.3). It is

$$\frac{1}{2}\|Ax_k - y^{\delta_k}\|^2 + \alpha_k\|x_k\|_{\ell^1} \leq \frac{1}{2}\|A\tilde{x} - y_k^\delta\|^2 + \alpha_k\|\tilde{x}\|_{\ell^1}$$
$$\leq \delta_k^2 + \|A\tilde{x} - y\|^2 + \alpha_k\|\tilde{x}\|_{\ell^1}.$$

Therefore

$$\alpha_k(\|x_k\|_{\ell^1} - \|\tilde{x}\|_{\ell^1}) \leq \delta_k^2 + \|A\tilde{x} - y\|^2 \leq C < \infty.$$

Since $\|\tilde{x}\|_{\ell^1} < \infty$ we need $\lim_{k\to\infty} \alpha_k = 0$ in order to allow $\lim_{k\to\infty} \|x_k\|_{\ell^1} = \infty$ as necessary due to a). Additionally, the product $\alpha_k\|x_k\|_{\ell^1}$ has to stay bounded, yielding together b) and c). □

The next step would be to show (weak) convergence of the regularized solutions to the exact solution as $\delta \to 0$. However, no such result is known. Even for the Hilbert scale setting of Section 2 no general convergence assertion appears to be available. In the standard setting $\|x^\dagger\|_{\ell^1} < \infty$, one has due to the optimality of the x_α^δ

$$\frac{1}{2}\|Ax_\alpha^\delta - y^\delta\|^2 + \alpha\|x_\alpha^\delta\|_{\ell^1} \leq \frac{1}{2}\delta^2 + \alpha\|x^\dagger\|_{\ell^1} < \infty. \tag{3.1}$$

Requiring $\alpha \to 0$ and $\delta^2/\alpha \to 0$ as $\delta \to 0$ then ensures (weak) convergence of the regularized solutions to the exact one. In particular, the ℓ^1-norm of the regularized solutions $\|x_\alpha^\delta\|_{\ell^1}$ remains bounded by a constant for all $\delta > 0$. In the oversmoothing setting, the right-hand side in (3.1) is infinite, hence provides no useful information. It appears natural to replace $x^\dagger$ by suitable auxiliary elements to bound the Tikhonov functional, but that is not enough. Let $\{x_\delta\}_{\delta>0}$ be a sequence with

$$\|x_\delta\|_{\ell^1} = \inf_{x\in\ell^2, \|Ax-y\|\leq c\delta} \|x\|_{\ell^1}$$

for fixed constant $c > 0$. Using the $\{x_\delta\}_{\delta>0}$ to bound the Tikhonov functional, we obtain

$$\frac{1}{2}\|Ax_\alpha^\delta - y^\delta\|^2 + \alpha\|x_\alpha^\delta\|_{\ell^1} \leq \frac{1}{2}\|Ax_\delta - y^\delta\|^2 + \alpha\|x_\delta\|_{\ell^1}$$
$$\leq (c+1)\delta^2 + \alpha\|x_\delta\|_{\ell^1}. \tag{3.2}$$

For any choice of $\alpha = \alpha(\delta)$ one obtains $\|x_\alpha^\delta\|_{\ell^1} \leq c\frac{\delta^2}{\alpha} + \|x_\delta\|_{\ell^1}$. Even if $\delta^2/\alpha \to 0$, this does not yield a bound for $\|x_\alpha^\delta\|_{\ell^1}$ independent of δ, since $\|x_\delta\|_{\ell^1} \to \infty$. Therefore one cannot infer the existence of a (weakly) convergent subsequence among the regularized solutions x_α^δ as is the argument in the standard, non-oversmoothing, convergence results. For the oversmoothing regularization one would need that the x_α^δ are bounded in a norm weaker than the ℓ^1-norm, for example, in the ℓ^2-norm. It is currently not clear how such a connection can be established. At this point we also mention that the oversmoothing approach prevents the use of state-of-the-art regularization theory. In recent years, *variational inequalities* (sometimes also called variational source conditions) have emerged as a powerful tool in the theory of Banach-space regularization, and we only refer, for example, to the papers [4, 5]. For ℓ^1 regularization, in [11] a variational inequality of the form

$$\|x - x^\dagger\|_{\ell^1} \leq \|x\|_{\ell^1} - \|x^\dagger\|_{\ell^1} + \varphi(\|Ax - Ax^\dagger\|) \tag{3.3}$$

for a well-defined concave index function φ and valid for all $x \in \ell^1$ was established and used for deriving convergence rates in the case $x^\dagger \in \ell^1$. Clearly, this concept is inapplicable when $x^\dagger \in \ell^2 \setminus \ell^1$ and hence $\|x^\dagger\|_{\ell^1} = \infty$, but it could be an idea to measure the error in ℓ^2-norms. Therefore, it seems to be interesting that not even for $x^\dagger \in \ell^1$ a variational inequality of the form

$$\|x - x^\dagger\|_{\ell^2}^\kappa \leq \|x\|_{\ell^1} - \|x^\dagger\|_{\ell^1} + \varphi(\|Ax - Ax^\dagger\|)$$

with $\kappa = 1$ or $\kappa = 2$ is known as an alternative to (3.3).

4. Convergence rates for diagonal operators

In this section, we derive convergence rates for the ℓ^1-penalized Tikhonov-type regularization (1.3) to (1.1). This method became a popular and powerful tool in the last decade, sparked by the seminal paper [13]. Since then, many authors have contributed to its theory and application. Here we only mention the papers [14–16]. As is typical in ℓ^1-regularization, we assume that A is *injective*. For the non-injective case we refer to [17]. The vast majority of papers connected to ℓ^1-regularization assumes sparsity in the sense that $x^\dagger \in \ell^0$, i.e., that it has only finitely many non-zero components. However, in [11] for the first time the situation that the exact solution $x^\dagger$ is not sparse, but only $x^\dagger \in \ell^1$, was explored. The results were later refined and extended in [3, 12, 18, 19]. In some sense, this paper is a continuation of this development as we are now interested in the case that $x^\dagger$ is not even in ℓ^1, but $x^\dagger \in \ell^2 \setminus \ell^1$. Due to this we will employ the ℓ^2-norm to measure the speed of convergence, and we seek for an index function φ, i.e., a continuous, monotonically increasing, and concave function with $\varphi(0) = 0$, such that

$$\|x_\alpha^\delta - x^\dagger\|_{\ell^2} \leq C\varphi(\delta)$$

holds with some constant $C > 0$. We will show that an appropriate choice of the regularization parameter yields such a function φ.

It is well-known that ℓ^1-regularization yields sparse solutions. We will see that for showing convergence rates when $x^\dagger \notin \ell^1$ it is essential to estimate the support of the regularized solutions. In order to do this, we rely on the explicit calculation of the minimizers. Therefore, we have to restrict ourselves to diagonal operators for this paper. We denote by $\{e^{(i)}\}_{i=1}^\infty$ the canonical orthonormal basis in ℓ^2 with components $e_k^{(k)} = 1$ and $e_k^{(i)} = 0$ for $k \neq i$. Moreover, we denote by $\{v^{(i)}\}_{i=1}^\infty$ an orthonormal basis in the range closure $\overline{\text{range}(A)}$ of Y. Then we say that a compact operator $A : \ell^2 \to Y$ with decreasingly ordered singular values $\{\sigma_i\}_{i=1}^\infty$ is of *diagonal structure* if

$$Ae^{(i)} = \sigma_i v^{(i)} \quad \text{and} \quad e^{(i)} = \frac{1}{\sigma_i}A^*v^{(i)} \quad (i = 1, 2, ...).$$

This model includes compact linear operators mapping between general Hilbert spaces X and Y, since such operators can be diagonalized with respect to their singular system [1, 2].

We now present a way of constructing functions φ that serve as prototypes for convergence rates. To this end, we define, for all $n \in \mathbb{N}$, the linear projectors

$$P_n : \ell^2 \to \ell^2, \quad P_n x = \{x_i\}_{i=1}^n. \tag{4.1}$$

Lemma 4.1. *Let $A : \ell^2 \to Y$ be a compact linear operator with diagonal structure as introduced above, possessing the singular values*

$$\|A\| = \sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq \sigma_{n+1} \geq \ldots \to 0 \qquad as \quad n \to \infty,$$

and let $x^\dagger \in \ell^2$ denote the uniquely determined solution to the operator equation (1.1)*, and set*

$$\mathcal{T}(x^\dagger, n) := \|(I - P_n)x^\dagger\|_{\ell^2} = \sqrt{\sum_{i=n+1}^{\infty} \left|x_i^\dagger\right|^2}. \tag{4.2}$$

Then for any $x \in \ell^2$ it is, for all $n \in \mathbb{N}$,

$$\|x - x^\dagger\|_{\ell^2} \leq \|(I - P_n)x\|_{\ell^2} + \sigma_n^{-1}\|Ax - Ax^\dagger\| + \mathcal{T}(x^\dagger, n). \tag{4.3}$$

Proof. We have, for any $x \in \ell^2$, with the linear projectors from (4.1) the relation

$$\|x - x^\dagger\|_{\ell^2} \leq \|(I - P_n)x\|_{\ell^2} + \|P_n(x - x^\dagger)\|_{\ell^2} + \|(I - P_n)x^\dagger\|_{\ell^2} \tag{4.4}$$

which holds for all $n \in \mathbb{N}$. We keep n arbitrary but fixed and start estimating the last term which describes the decay of the tail of the solution and thus its smoothness. It is a fixed function of n which goes to zero as $n \to \infty$, and we employ the convention (4.2).

Next we estimate the middle term $\|P_n(x - x^\dagger)\|_{\ell^2}$ in (4.4). In order to do so we recall the notion of the modulus of continuity, given by

$$\omega(M, \theta) := \sup\{\|x\| : x \in M, \|Ax\| \leq \theta\}$$

where $M \subset \ell^2$. This quantity is essentially related to minimal errors of any regularization method for noisy data. Since $P_n(x - x^\dagger) \in \ell_n^2 := \mathrm{span}\{e_1, \ldots, e_n\}$, we can use tools from approximation theory to estimate its norm. In [20, Proposition 3.9] it has been shown that for any finite dimensional space X_n

$$\omega(X_n, \theta) = \frac{\theta}{\Theta(A, n)},$$

where the *modulus of injectivity* $\Theta(A, n)$ is defined as

$$\Theta(A, n) := \inf_{0 \neq x \in X_n} \frac{\|Ax\|}{\|x\|}.$$

For diagonal operators it is $\Theta(A, n) = \sigma_n$ and thus

$$\omega(\ell_n^2, \theta) = \sigma_n^{-1}\theta.$$

Noting that for diagonal operators it also holds that

$$\|AP_n(x - x^\dagger)\| \leq \|A(x - x^\dagger)\|, \tag{4.5}$$

it is therefore

$$\|P_n(x - x^\dagger)\|_{\ell^2} \leq \omega(\ell_n^2, \|AP_n(x - x^\dagger)\|) = \sigma_n^{-1}\|AP_n(x - x^\dagger)\| \leq \sigma_n^{-1}\|A(x - x^\dagger)\|. \tag{4.6}$$

Inserting this and (4.2) into (4.4), we obtain (4.3). $\qquad\square$

In the best case scenario the term $\|(I - P_n)x\|_{\ell^2}$ in (4.3) vanishes (note that this is a crucial point to be shown for the ℓ^1-regularized approximations to $x^\dagger$) and only the two rightmost terms remain. The best possible convergence rate is therefore determined by taking the infimum of those two expressions with respect to n. This yields a prototype rate function

$$\varphi(t) = \inf_{n\in\mathbb{N}} \left\{ \frac{t}{\sigma_n} + \|(I - P_n)x^\dagger\|_{\ell^2} \right\}. \tag{4.7}$$

Note that as infimum over affine functions φ is concave, and it is also monotonically increasing with $\varphi(0) = 0$. Since for fixed $t \geq 0$ the infimum is taken over a countable set and $\sigma_n^{-1}t \to \infty$ monotonically as $n \to \infty$ while $\mathcal{T}(x^\dagger, n) \to 0$ monotonically, the infimum is attained as minimum and the corresponding index

$$n_{\inf}(t) := \arg\min_{n\in\mathbb{N}} \left\{ \frac{t}{\sigma_n} + \|(I - P_n)x^\dagger\|_{\ell^2} \right\} \tag{4.8}$$

is well defined. In order to show $\|x - x^\dagger\|_{\ell^2} \leq \varphi(\|Ax - Ax^\dagger\|)$ for the ℓ^1-regularized solutions $x = x_\alpha^\delta$ from (1.3), we need the following assumptions.

Assumption 4.1. *(a) The singular values of A fulfill, for all $n \in \mathbb{N}$,*

$$1 \leq \frac{\sigma_n}{\sigma_{n+1}} \leq C_\sigma$$

with a constant $1 \leq C_\sigma < \infty$.
(b) The tail is monotone, i.e., $|x_n^\dagger|$ is monotonically decreasing for all sufficiently large n.
(c) The convergence rate $\varphi(t)$ is not dominated by the tail, i.e., there exists a constant $\tilde{C}$ such that

$$\mathcal{T}(x^\dagger, n_{\inf}(t)) \leq \tilde{C}\sigma_{n_{\inf}(t)}^{-1}t.$$

(d) For the true solution $x^\dagger$ it holds

$$|x_{n_{\inf}(t)+1}^\dagger| \leq C\frac{\sigma_{n_{\inf}(t)}^{-1}t}{\sqrt{n_{\inf}(t)}}$$

for a constant $0 \leq C < \infty$ and sufficiently small $t > 0$.

Part (a) limits the ill-posedness of the forward operators. A polynomial decay of the singular values fulfills the condition, even an exponential decay of order $\sigma_i = e^{-i}$ is permitted, but for any $\epsilon > 0$ a decay $\sigma_i = e^{-i^{1+\epsilon}}$ violates the condition. Note that the left-hand inequality is trivial, but will be used in a proof later. Part (b) is required purely for technical reasons. Part (c) and (d) link the smoothness of the true solution with the ill-posedness of the forward operator. They will be discussed in examples below. We remark that similar restrictions are standing assumptions in the Hilbert-scale setting of Section 2: Condition (2.2) ensures that the singular values of A fall asymptotically as i^{-a}, and the condition $p \geq (q - a)/2$ implies that the solution must not be significantly smoother than the penalty.

For later use we mention two consequences of Assumption 4.1.

Lemma 4.2. *Let A be as in Lemma 4.1, and let Assumption 4.1 (a) and (c) hold. Then*

$$\frac{1}{1 + \tilde{C}}\varphi(t) \leq \sigma_{n_{\inf}(t)}^{-1}t \leq \varphi(t). \tag{4.9}$$

Furthermore, it is

$$\|AP_{n_{\inf}(t)}x^\dagger - Ax^\dagger\| \leq \tilde{C}t. \tag{4.10}$$

Proof. The right-hand inequality of (4.9) follows from the definition of φ in (4.7). Similarly, it is

$$\varphi(t) = \sigma_{n_{\inf}(t)}^{-1} t + \mathcal{T}(x^\dagger, n_{\inf}(t)) \le (1 + \tilde{C})\sigma_{n_{\inf}(t)}^{-1} t,$$

which yields the left-hand inequality. To obtain the estimate of the residual, we observe that

$$\|AP_{n_{\inf}(t)}x^\dagger - Ax^\dagger\|^2 = \sum_{i=n_{\inf}(t)+1}^{\infty} |\sigma_i x_i^\dagger|^2 \le \sigma_{n_{\inf}(t)+1}^2 \sum_{i=n_{\inf}(t)+1}^{\infty} |x_i^\dagger|^2 \le \sigma_{n_{\inf}(t)}^2 \mathcal{T}(x^\dagger, n_{\inf}(t))^2.$$

Inserting Assumption 4.1 (c) immediately yields (4.10). $\qquad\square$

4.1. A-priori parameter choice

In order to show convergence rates based on Lemma 4.1, two terms need to be estimated in (4.3) for $x = x_\alpha^\delta$: the residual $\|Ax_\alpha - Ax^\dagger\|$ and the tail $\|(I - P_n)x_\alpha^\delta\|_{\ell^2}$.

Lemma 4.3. *Let A be as in Lemma 4.1, and let Assumption 4.1 (c) hold. Then, with the a-priori choice*

$$\alpha(\delta) = c_\alpha \frac{\delta^2}{\sqrt{n_{\inf}(\delta)}\varphi(\delta)}, \tag{4.11}$$

with constant $0 < c_\alpha < \infty$, the minimizers x_α^δ of (1.3) satisfy

$$\|Ax_\alpha^\delta - y^\delta\| \le ((\tilde{C} + 1)^2 + 4c_\alpha)\delta. \tag{4.12}$$

In particular, it is

$$\|Ax_\alpha^\delta - Ax^\dagger\| \le c\delta \tag{4.13}$$

with positive constant $c = (\tilde{C} + 1)^2 + 4c_\alpha + 1$.

Proof. From the Tikhonov functional (1.3) we have for all $n \in \mathbb{N}$

$$\frac{1}{2}\|Ax_\alpha^\delta - y^\delta\|^2 + \alpha\|x_\alpha^\delta\|_{\ell^1} \le \frac{1}{2}\|AP_n x^\dagger - y^\delta\|^2 + \alpha\|P_n x^\dagger\|_{\ell^1},$$

which with

$$\|x_\alpha^\delta\|_{\ell^1} = \|P_n x_\alpha^\delta\|_{\ell^1} + \|(I - P_n)x_\alpha^\delta\|_{\ell^1}$$

and

$$\|P_n x^\dagger\|_{\ell^1} \le \|P_n(x^\dagger - x_\alpha^\delta)\|_{\ell^1} + \|P_n x_\alpha^\delta\|_{\ell^1}$$

yields

$$\frac{1}{2}\|Ax_\alpha^\delta - y^\delta\|^2 + \alpha\|(I - P_n)x_\alpha^\delta\|_{\ell^1} \le \frac{1}{2}\|AP_n x^\dagger - y^\delta\|^2 + \alpha\|P_n(x^\dagger - x_\alpha^\delta)\|_{\ell^1}$$

$$\le \frac{1}{2}\|AP_n x^\dagger - y^\delta\|^2 + \alpha\sqrt{n}\|P_n(x^\dagger - x_\alpha^\delta)\|_{\ell^2}. \tag{4.14}$$

Now fix $n = n_{\inf}(\delta)$. Using (4.6) and (4.9) on the right-hand side we have

$$\|P_{n_{\inf(\delta)}}(x^\dagger - x_\alpha^\delta)\|_{\ell^2} \le \sigma_{n_{\inf(\delta)}}^{-1}\|Ax_\alpha^\delta - Ax^\dagger\|$$
$$\le \sigma_{n_{\inf(\delta)}}^{-1}(\|Ax_\alpha^\delta - y^\delta\| + \|Ax^\dagger - y^\delta\|)$$
$$\le 2\sigma_{n_{\inf(\delta)}}^{-1}\max\{\|Ax_\alpha^\delta - y^\delta\|, \delta\}$$
$$\le 2\varphi(\|Ax_\alpha^\delta - y^\delta\|), \tag{4.15}$$

where in the last estimate we have assumed $\|Ax_\alpha^\delta - y^\delta\| > \delta$, as otherwise the assertion of the lemma is trivially fulfilled. We combine this estimate with (4.14) and again set $n = n_{\inf}(\delta)$. This yields

$$\frac{1}{2}\|Ax_\alpha^\delta - y^\delta\|^2 \le \frac{1}{2}\|AP_{n_{\inf(\delta)}}x^\dagger - y^\delta\|^2 + 2\alpha\sqrt{n_{\inf}(\delta)}\varphi(\|Ax_\alpha^\delta - y^\delta\|). \tag{4.16}$$

Note that by (4.10) we have $\|AP_{n_{\inf(\delta)}}x^\dagger - y^\delta\| \le (\tilde{C} + 1)\delta$. Inserting the parameter choice (4.11) into (4.16), we continue analogously to [5, Corollary 1]. Namely, it is by concavity of φ

$$\frac{1}{2}\|Ax_\alpha^\delta - y^\delta\|^2 \le \frac{(\tilde{C} + 1)^2}{2}\delta^2 + 2c_\alpha\frac{\delta^2}{\sqrt{n_{inf}(\delta)}\varphi(\delta)}\sqrt{n_{inf}(\delta)}\varphi(\|Ax_\alpha^\delta - y^\delta\|)$$
$$\le \frac{(\tilde{C} + 1)^2}{2}\delta^2 + 2c_\alpha\delta^2\frac{\varphi(\|Ax_\alpha^\delta - y^\delta\|)}{\varphi(\delta)}$$
$$\le \delta^2\left(\frac{(\tilde{C} + 1)^2}{2} + 2c_\alpha\right)\frac{\varphi(\|Ax_\alpha^\delta - y^\delta\|)}{\varphi(\delta)}$$
$$\le \delta^2\left(\frac{(\tilde{C} + 1)^2}{2} + 2c_\alpha\right)\frac{\|Ax_\alpha^\delta - y^\delta\|\varphi(\delta)}{\delta\varphi(\delta)}$$
$$= \delta\left(\frac{(\tilde{C} + 1)^2}{2} + 2c_\alpha\right)\|Ax_\alpha^\delta - y^\delta\|.$$

This yields $\|Ax_\alpha^\delta - y^\delta\| \le ((\tilde{C} + 1)^2 + 4c_\alpha)\delta$ for α from (4.11).

The second assertion follows from this, the noise assumption (1.2), and the triangle inequality. $\quad\square$

Lemma 4.4. *Let A be as in Lemma 4.1, and let Assumption 4.1 (a)–(d) hold. Let α be chosen according to (4.11) such that $c_\alpha \ge C(\tilde{C} + 1)$. Then the minimizers x_α^δ of (1.3) satisfy*

$$\|(I - P_{n_{\inf(\delta)}})x_\alpha^\delta\|_{\ell^2} \le C_\sigma\varphi(\delta). \tag{4.17}$$

Proof. The diagonal structure of the operator allows to calculate the minimizer of (1.3) explicitly, and (1.3) reads

$$\frac{1}{2}\|Ax - y^\delta\|^2 + \alpha\|x\|_{\ell^1} = \frac{1}{2}\sum_{i\in\mathbb{N}}(\sigma_i x_i - y_i^\delta)^2 + \alpha\sum_{i\in\mathbb{N}}|x_i|. \tag{4.18}$$

Since the components are decoupled, the first order optimality condition for the above functional is

$$\frac{\partial}{\partial x_i}\left(\frac{1}{2}\sum_{i\in\mathbb{N}}(\sigma_i x_i - y_i^\delta)^2 + \alpha\sum_{i\in\mathbb{N}}|x_i|\right) = 0 \quad \forall i \in \mathbb{N},$$

i.e., for each $i \in \mathbb{N}$,

$$\sigma_i(\sigma_i x_i - y_i^\delta) + \alpha \mathrm{sgn}(x_i) = 0$$

where

$$\mathrm{sgn}(x) := \begin{cases} 1 & x > 0 \\ -1 & x < 0 \\ \in [-1, 1] & x = 0 \end{cases}.$$

Consider the case $[x_\alpha^\delta]_i > 0$. Then

$$[x_\alpha^\delta]_i = \frac{y_i^\delta}{\sigma_i} - \frac{\alpha}{\sigma_i^2} > 0.$$

On the other hand, for $[x_\alpha^\delta]_i < 0$ we have

$$[x_\alpha^\delta]_i = \frac{y_i^\delta}{\sigma_i} + \frac{\alpha}{\sigma_i^2} < 0$$

and consequently

$$[x_\alpha^\delta]_i = 0 \Leftrightarrow \left| \frac{y_i^\delta}{\sigma_i} \right| \le \frac{\alpha}{\sigma_i^2}. \tag{4.19}$$

We will only consider the case $[x_\alpha^\delta]_i > 0$ further, the results for $[x_\alpha^\delta]_i < 0$ are analogous with inverted sign. First let y be exact, noise-free data, i.e., $y_i = \sigma_i x_i^\dagger$ for all $i \in \mathbb{N}$, where the minimizer of (4.18) (with y^δ temporarily replaced by y) is denoted by x_α. Then (4.19) yields

$$[x_\alpha]_i = 0 \Leftrightarrow \left| x_i^\dagger \right| \le \frac{\alpha}{\sigma_i^2}. \tag{4.20}$$

Inserting the parameter choice (4.11) and considering $i = n_{\mathrm{inf}}(\delta) + 1$, we find with Assumption 4.1 (a) and (d), and with (4.9) that

$$\begin{aligned}
\frac{\alpha}{\sigma_{n_{\mathrm{inf}}(\delta)+1}^2} &= \frac{\alpha}{\sigma_{n_{\mathrm{inf}}(\delta)}^2} \frac{\sigma_{n_{\mathrm{inf}}(\delta)}^2}{\sigma_{n_{\mathrm{inf}}(\delta)+1}^2} \\
&\ge c_\alpha \frac{(\sigma_{n_{\mathrm{inf}}(\delta)}^{-1} \delta)^2}{\sqrt{n_{\mathrm{inf}}(\delta)} \varphi(\delta)} \\
&\ge \frac{c_\alpha}{\tilde{C}+1} \frac{\sigma_{n_{\mathrm{inf}}(\delta)}^{-1} \delta}{\sqrt{n_{\mathrm{inf}}(\delta)}} \\
&\ge \frac{c_\alpha}{C(\tilde{C}+1)} |x_{n_{\mathrm{inf}}(\delta)+1}^\dagger|.
\end{aligned}$$

Now, as long as $c_\alpha \ge C(\tilde{C}+1)$, this implies $[x_\alpha]_{n_{\mathrm{inf}}(\delta)+1} = 0$. Because $\frac{\alpha}{\sigma_i^2}$ is increasing in i while $|x_i^\dagger|$ decreases, we conclude that for $n_{\mathrm{inf}}(\delta)$ sufficiently large (i.e., δ sufficiently small), $[x_\alpha]_{n_{\mathrm{inf}}+1} = 0$. Under Assumption 4.1 (b) all entries $[x_\alpha]_m$, $m > n_{\mathrm{inf}}(\delta)$, must then also be zero. Hence, in the noise-free case, we have $\|(I - P_{n_{\mathrm{inf}}(\delta)})x_\alpha\|_{\ell^2} = 0$. Therefore, for noisy data and under the same parameter choice, any

contribution to $\|(I - P_{n_{\inf}(\delta)})x_\alpha^\delta\|_{\ell^2}$ must be due to the noise in the data. Consider the model $y - y^\delta = \delta\zeta$ with some $\zeta \in \ell^2$, $\|\zeta\|_{\ell^2} = 1$. Then $\|y - y^\delta\| = \delta$ and

$$[x_\alpha^\delta]_i > 0 \quad \Leftrightarrow \quad x_i^\dagger + \frac{\delta\zeta_i}{\sigma_i} - \frac{\alpha}{\sigma_i^2} > 0. \tag{4.21}$$

The components $x_i^\dagger$ are fixed and decreasing. The regularization parameter α is fixed for all i, hence α/σ_i^2 grows with increasing index. Therefore, the smaller i, the (potentially) larger the components $[x_\alpha^\delta]_i$ could become. To estimate the tail $\|(I - P_{n_{\inf}(\delta)})x_\alpha^\delta\|_{\ell^2}$ we are only interested in the higher indices $i > n_{\inf}(\delta)$. Due to the asymptotics of the individual terms, $\|(I - P_{n_{\inf}(\delta)})x_\alpha^\delta\|_{\ell^2}$ therefore becomes largest when all its mass is concentrated in the lowest possible component $n_{\inf}(\delta) + 1$, i.e., $\zeta_{n_{\inf}(\delta)+1} = 1$ and $(I - P_{n_{\inf}(\delta)})x_\alpha^\delta = ([x_\alpha^\delta]_{n_{\inf}(\delta)+1}, 0, \dots)$. From the noise-free considerations above we already obtained $x_i^\dagger \le \frac{\alpha}{\sigma_i^2}$. Hence, any positive contribution in (4.21) comes from the noise. With an analogous argument for the case $[x_\alpha^\delta]_i < 0$ we therefore get

$$|[x_\alpha^\delta]_{n_{\inf}(\delta)+1}| \le \frac{\delta}{\sigma_{n_{\inf}(\delta)+1}}.$$

Using Assumption 4.1 (c) we obtain

$$\frac{\delta}{\sigma_{n_{\inf}(\delta)+1}} = \frac{\delta}{\sigma_{n_{\inf}(\delta)}} \frac{\sigma_{n_{\inf}(\delta)}}{\sigma_{n_{\inf}(\delta)+1}} \le C_\sigma \sigma_{n_{\inf}(\delta)}^{-1} \delta \le C_\sigma \varphi(\delta).$$

$\square$

Putting all the pieces together yields the main theorem.

Theorem 4.1. *Let $A : \ell^2 \to Y$ be a compact linear operator with diagonal structure, possessing the singular values*

$$\|A\| = \sigma_1 \ge \sigma_2 \ge \dots \ge \sigma_n \ge \sigma_{n+1} \ge \dots \to 0 \qquad as \quad n \to \infty.$$

Let $x^\dagger \in \ell^2$ denote the uniquely determined solution to the operator equation (1.1), and let Assumption 4.1 hold. Then the ℓ^1-regularized solutions x_α^δ from (1.3) with noisy data y^δ obeying the noise model $\|y - y^\delta\| \le \delta$ satisfy for sufficiently small $\delta > 0$ the estimate

$$\|x_\alpha^\delta - x^\dagger\|_{\ell^2} \le (C_\sigma + c)\varphi(\delta)$$

with c as in (4.13) and with the concave index function

$$\varphi(t) = \inf_{n \in \mathbb{N}} \left\{ \frac{t}{\sigma_n} + \|(I - P_n)x^\dagger\|_{\ell^2} \right\},$$

provided the regularization parameter is chosen a priori as

$$\alpha(\delta) = c_\alpha \frac{\delta^2}{\sqrt{n_{\inf}(\delta)}\varphi(\delta)},$$

with constant $c_\alpha \ge C(\tilde{C} + 1)$. The integers

$$n_{\inf}(\delta) := \arg\min_{n \in \mathbb{N}} \left\{ \frac{\delta}{\sigma_n} + \|(I - P_n)x^\dagger\|_{\ell^2} \right\}$$

can be found for all $\delta > 0$.

Proof. Lemma 4.1 and the discussion thereafter provide us with the prototype of the convergence rate $\varphi(t)$ (4.7) and with $n_{\inf}(t)$ (4.8). Consider (4.3) with x replaced by x_α^δ, i.e.,

$$\|x_\alpha^\delta - x^\dagger\|_{\ell^2} \leq \|(I - P_n)x_\alpha^\delta\|_{\ell^2} + \sigma_n^{-1}\|Ax_\alpha^\delta - Ax^\dagger\| + \mathcal{T}(x^\dagger, n). \tag{4.22}$$

Fix $n = n_{\inf}(\delta)$. From Lemma 4.3 we then have $\|Ax_\alpha^\delta - Ax^\dagger\| \leq c\delta$ with $c > 1$, hence

$$\|x_\alpha^\delta - x^\dagger\|_{\ell^2} \leq \|(I - P_{n_{\inf}(\delta)})x_\alpha^\delta\|_{\ell^2} + c(\sigma_{n_{\inf}(\delta)}^{-1}\delta + \mathcal{T}(x^\dagger, n_{\inf}(\delta))).$$

The term in brackets, by definition, attains the value $\varphi(\delta)$. The remaining term $\|(I - P_{n_{\inf}(\delta)})x_\alpha^\delta\|_{\ell^2}$ was estimated in Lemma 4.4. Therefore

$$\|x_\alpha^\delta - x^\dagger\|_{\ell^2} \leq (C_\sigma + c)\varphi(\delta).$$

$\square$

4.2. The discrepancy principle

The a-priori choice requires, in principle, the knowledge of the exact solution and is thus unfeasible in practice. In the following we comment on the discrepancy principle as an a-posteriori parameter choice. We begin with a helpful lemma.

Lemma 4.5. *Let Assumption 4.1 hold and $\alpha_*(\delta)$ be any choice of the regularization parameter in* (1.3) *such that*

$$\tau_1\delta \leq \|Ax_{\alpha_*}^\delta - y^\delta\| \leq \tau_2\delta$$

for $(1 + \tilde{C}) < \tau_1 \leq \tau_2 < \infty$. Then $\bar{c}\alpha_* \geq \alpha$ with the a-priori choice of α from (4.11) *and $\bar{c} = \frac{2c_\alpha(\tau_2+1)}{\tau_1^2 - (\tilde{C}+1)^2}$.*

Proof. From the minimizing property of the Tikhonov-functional we have

$$\frac{1}{2}\|Ax_{\alpha_*}^\delta - y^\delta\|^2 + \alpha_*\|x_{\alpha_*}^\delta\|_{\ell^1} \leq \frac{1}{2}\|AP_{n_{\inf}(\delta)}x^\dagger - y^\delta\|^2 + \alpha_*\|P_{n_{\inf}(\delta)}x^\dagger\|_{\ell^1}.$$

Similar to the proof of Lemma 4.3 this implies

$$\frac{1}{2}\|Ax_{\alpha_*}^\delta - y^\delta\|^2 \leq \frac{1}{2}\|AP_{n_{\inf}(\delta)}x^\dagger - y^\delta\|^2 + \alpha_* \sqrt{n_{\inf}(\delta)}\varphi(\|Ax_\alpha^\delta - Ax^\dagger\|). \tag{4.23}$$

Note that for this results we have replaced (4.15) by

$$\|P_{n_{\inf}(\delta)}(x^\dagger - x_\alpha^\delta)\|_{\ell^2} \leq \sigma_{n_{\inf}(\delta)}^{-1}\|Ax_\alpha^\delta - Ax^\dagger\| \leq \varphi(\|Ax_\alpha^\delta - Ax^\dagger\|).$$

Using $\|AP_{n_{\inf}(\delta)}x^\dagger - y^\delta\| \leq (\tilde{C} + 1)\delta$ (cf. Lemma 4.2) and $\tau_1\delta \leq \|Ax_{\alpha_*}^\delta - y^\delta\|$ in (4.23) we have

$$\frac{\tau_1^2\delta^2}{2} \leq \frac{(\tilde{C} + 1)^2}{2}\delta^2 + \alpha_* \sqrt{n_{\inf}(\delta)}\varphi(\|Ax_\alpha^\delta - Ax^\dagger\|),$$

and since $\tau_1 > (1 + \tilde{C})$

$$\delta^2 \leq \frac{2}{\tau_1^2 - (\tilde{C} + 1)^2}\alpha_* \sqrt{n_{\inf}(\delta)}\varphi(\|Ax_\alpha^\delta - Ax^\dagger\|).$$

Using the upper bound $\|Ax_{\alpha_*}^\delta - y^\delta\| \leq \tau_2\delta$ and the concavity of φ, it is

$$\varphi(\|Ax_\alpha^\delta - Ax^\dagger\|) \leq \varphi(\|Ax_\alpha^\delta - y^\delta\| + \|y^\delta - Ax^\dagger\|) \leq (\tau_2 + 1)\varphi(\delta)$$

and we obtain, with α from (4.11),

$$\alpha = c_\alpha \frac{\delta^2}{\sqrt{n_{\inf}(\delta)}\varphi(\delta)} \leq c_\alpha \frac{2(\tau_2 + 1)}{\tau_1^2 - (\tilde{C} + 1)^2}\alpha_*.$$

$\square$

Theorem 4.2. *Let $A : \ell^2 \to Y$ be a compact linear operator with diagonal structure, possessing the singular values*

$$\|A\| = \sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq \sigma_{n+1} \geq \ldots \to 0 \qquad as \quad n \to \infty.$$

Let $x^\dagger \in \ell^2$ denote the uniquely determined solution to the operator equation (1.1), and let Assumption 4.1 hold. Then the ℓ^1-regularized solutions x_α^δ from (1.3) with noisy data y^δ obeying the noise model $\|y - y^\delta\| \leq \delta$ satisfy for sufficiently small $\delta > 0$ the estimate

$$\|x_\alpha^\delta - x^\dagger\|_{\ell^2} \leq (\tau_2 + C_\sigma + 1)\varphi(\delta)$$

with a constant $0 < c < \infty$ and with the concave index function

$$\varphi(t) = \inf_{n \in \mathbb{N}}\left\{\frac{t}{\sigma_n} + \|(I - P_n)x^\dagger\|_{\ell^2}\right\},$$

provided the regularization parameter is chosen a posteriori such that

$$\tau_1\delta \leq \|Ax_\alpha^\delta - y^\delta\| \leq \tau_2\delta, \tag{4.24}$$

with parameters $1 + \tilde{C} < \tau_1 \leq \tau_2 < \infty$ such that $\frac{2C(\tilde{C}+1)(\tau_2+1)}{\tau_1^2-(\tilde{C}+1)^2} \leq 1$.

Proof. We start as in the proof of Theorem 4.1 and consider (4.22). Due to the parameter choice (4.24) and the triangle inequality we have

$$\|Ax_\alpha^\delta - Ax^\dagger\| \leq (\tau_2 + 1)\delta,$$

hence it is, with $n = n_{\inf}(\delta)$,

$$\|x_\alpha^\delta - x^\dagger\|_{\ell^2} \leq \|(I - P_{n_{\inf}(\delta)})x_\alpha^\delta\|_{\ell^2} + (\tau_2 + 1)(\sigma_{n_{\inf}(\delta)}^{-1}\delta + \mathcal{T}(x^\dagger, n_{\inf}(\delta))) = \|(I - P_{n_{\inf}(\delta)})x_\alpha^\delta\|_{\ell^2} + (\tau_2 + 1)\varphi(\delta).$$

According to Lemma 4.3, the regularization parameter α_* obtained from the discrepancy principle is larger than the a-priori parameter a from (4.11), $\alpha \leq \alpha_*$, if $\bar{c} \leq 1$. The minimal admissible constant c_α in Lemma 4.4 is $c_\alpha = C(\tilde{C} + 1)$. Inserting this into $\bar{c}$ yields $\alpha \leq \alpha_*$ if $\frac{2C(\tilde{C}+1)(\tau_2+1)}{\tau_1^2-(\tilde{C}+1)^2} \leq 1$. It now follows analogously as in the proof of Lemma 4.4 that for noise-free data $\|(I - P_{n_{\inf}(\delta)})x_\alpha^\delta\|_{\ell^2} = 0$ and in the worst-case noise scenario $\|(I - P_{n_{\inf}(\delta)})x_\alpha^\delta\|_{\ell^2} \leq C_\sigma\varphi(\delta)$. Hence

$$\|x_\alpha^\delta - x^\dagger\|_{\ell^2} \leq (\tau_2 + C_\sigma + 1)\varphi(\delta).$$

$\square$

4.3. Model cases

In order to exemplify and illustrate the general theory in more detail, we will now consider some simple model scenarios. As before, we assume that $A : \ell^2 \to \ell^2$ is diagonal, $[Ax]_i = \sigma_i x_i$, $i = 1, 2, \ldots$. For simplicity, we will denote all constants by a generic constant $0 < c < \infty$.

4.3.1. Polynomial singular values, polynomially decaying solution

Let $\sigma_i = i^{-\beta}$ and $x_i^\dagger = i^{-\eta}$ for positive values β and η and all $i \in \mathbb{N}$. In particular, for $\frac{1}{2} < \eta \le 1$ this yields a case of oversmoothing regularization $\|x^\dagger\|_{\ell^1} = \infty$, whereas for $\eta > 1$ we have the classical model with $\|x^\dagger\|_{\ell^1} < \infty$.

Theorem 4.3. *Let A be diagonal with singular values $\sigma_i = i^{-\beta}$, $\beta > 0$. Let $x^\dagger \in \ell^2$ be such that $[x^\dagger]_i = i^{-\eta}$ for $\eta > \frac{1}{2}$. Then the ℓ^1-regularized solutions x_α^δ from (1.3) to (1.1), with noisy data y^δ obeying $\|y - y^\delta\| \le \delta$, satisfy*

$$\|x_\alpha^\delta - x^\dagger\|_{\ell^2} \le c\delta^{\frac{2\eta-1}{2\eta+2\beta-1}}, \tag{4.25}$$

provided the regularization parameter is chosen a priori as

$$\alpha(\delta) = c\delta^{\frac{4\beta+2\eta}{2\eta+2\beta-1}} \tag{4.26}$$

or according to the discrepancy principle (4.24).

Proof. It remains to calculate the quantities occurring in the proof of Theorem 4.1 explicitly.

The tail of the exact solution satisfies

$$\|(I - P_n)x^\dagger\|_{\ell^2} = \sqrt{\sum_{i=n+1}^{\infty} |i^{-\eta}|^2} \le \sqrt{\frac{1}{2\eta - 1}(n + 1)^{1-2\eta}} \le (2\eta - 1)^{-\frac{1}{2}} n^{\frac{1}{2}-\eta}. \tag{4.27}$$

Inserting the structure of A, the rate prototype (4.7) becomes

$$\varphi(t) = \inf_{n \in \mathbb{N}} \left\{ n^\beta t + c_\eta n^{\frac{1}{2}-\eta} \right\}. \tag{4.28}$$

It is simple calculus to show

$$\varphi(t) = ct^{\frac{2\eta-1}{2\eta+2\beta-1}}$$

and

$$n_{inf}(t) = \left\lceil ct^{-\frac{2}{2\eta+2\beta-1}} \right\rceil \tag{4.29}$$

where $\lceil \cdot \rceil$ denotes the closest integer. Inserting the previous results into the parameter choice (4.11) yields

$$\alpha(\delta) = \frac{\delta^2}{\sqrt{n_{inf}(\delta)}\varphi(\delta)} = c\delta^{\frac{4\beta+2\eta}{2\eta+2\beta-1}}. \tag{4.30}$$

$\square$

We mention that this model problem fulfills Assumption 4.1. Namely, we have $x_{n+1} < x_n = n^{-\eta}$ for all $n \in \mathbb{N}$, and from (4.27) we obtain

$$\frac{\|(I - P_n)x^\dagger\|_{\ell^2}}{\sqrt{n}} \le cn^{-\eta}.$$

Also Assumption 4.1 (c) is fulfilled as

$$1 \le \frac{\sigma_n}{\sigma_{n+1}} \le \frac{n^{-\beta}}{(n+1)^{-\beta}} = \left(\frac{n+1}{n}\right)^{\beta} \le 2^{\beta} =: C_{\sigma}.$$

Finally, we observe that

$$\frac{\delta^2}{\alpha} = \frac{\delta^2}{c\delta^{\frac{4\beta+2\eta}{2\eta+2\beta-1}}} = c\delta^{\frac{2\eta-2}{2\eta+2\beta-1}}$$

goes to infinity when $\frac{1}{2} < \eta < 1$, stays constant for $\eta = 1$, and goes to zero for $\eta > 1$. That means there is a seamless transition between the regime of oversmoothing regularization and that of classical regularization. Numerical experiments are presented in Section 6.

4.3.2. Polynomial singular values, exponentially decaying solution

We now consider for simplicity the case $\sigma_i = i^{-1}$ and $x_i^{\dagger} = e^{-i}$ for all $i \in \mathbb{N}$. This is no oversmoothing regularization since $x^{\dagger} \in \ell^1$. Using MATHEMATICA®we calculate the tail in (4.7), $\mathcal{T}(x^{\dagger}, n) \le ce^{-n}$. It is then simple calculus to show

$$n_{inf}(t) = c \ln \frac{1}{t}.$$

This yields

$$\varphi(t) = \sigma_{n_{inf}(t)}^{-1} t + \mathcal{T}(x^{\dagger}, n_{inf}(t)) = c(t \ln \frac{1}{t} + t).$$

In contrast to the previous example, the two terms $\sigma_{n_{inf}(\delta)}^{-1} \delta$ and $\mathcal{T}(x^{\dagger}, n_{inf}(\delta))$ are no longer of the same order, but, for sufficiently small δ, $\sigma_{n_{inf}(\delta)}^{-1} \delta$ is the dominating part, i.e., Assumption 4.1 (c) is fulfilled. Also part (d) of Assumption 4.1 holds, since $x_{n_{inf}(\delta)+1}^{\dagger} < x_{n_{inf}(\delta)}^{\dagger} = c\delta$, whereas $\frac{\sigma_{n_{inf}(\delta)}^{-1} \delta}{\sqrt{n_{inf}(\delta)}} = c\delta \sqrt{\ln(\frac{1}{\delta})}$. The predicted rate $\|x_\alpha^\delta - x^{\dagger}\|_{\ell^2} \le c\varphi(\delta) = c\delta \ln \frac{1}{\delta}$ can be verified numerically, see the left plot of Figure 5.

4.3.3. Exponential singular values, polynomially decaying solution

Let now $\sigma_i = e^{-i}$ and, for simplicity, $x_i^{\dagger} = i^{-1}$ for all $i \in \mathbb{N}$. We are again in the oversmoothing regime. To formulate the convergence rate, we recall the definition of the Lambert-W-function, defined as $z = W(ze^z)$. With this, the minimizing argument of

$$\varphi(t) = \inf_{n \in \mathbb{N}} e^n t + cn^{-\frac{1}{2}}$$

can be found, using MATHEMATICA®, to be

$$n_{inf}(t) = \left\lceil \frac{3}{2} W\left(\frac{2}{3} t^{-\frac{3}{2}}\right) \right\rceil,$$

such that

$$\varphi(t) = te^{n_{inf}(t)} + cn_{inf}(t)^{-\frac{1}{2}}.$$

The first term decays faster than the second one, hence the rate is dominated by the tail of the exact solution, and Assumption 4.1 (c) is violated, see Figure 1. Consequently, Theorem 4.1 is not applicable. Indeed, in a numerical experiment, shown in the left part of Figure 6, the measured convergence rate is different from the one predicted here.
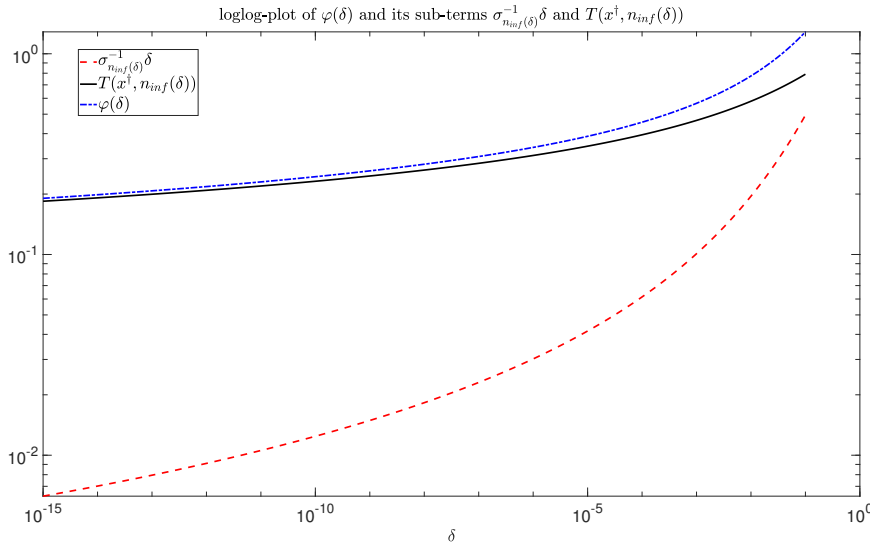
Figure 1. $\delta e^{n_{\inf}(\delta)}$ (red, dashed), $n_{\inf}(\delta)^{-\frac{1}{2}}$ (black, solid), and $\varphi(\delta)$ (blue, dash-dotted), for the scenario of Section 4.3.3, demonstrating the violation of Assumption 4.1 (c).

4.3.4. Exponential singular values, exponentially decaying solution

Let for this example $\sigma_i = e^{-i}$ and $x_i^{\dagger} = e^{-i}$ for all $i \in \mathbb{N}$. Then

$$\varphi(t) = \inf_{n \in \mathbb{N}} e^n t + c e^{-n}$$

and

$$n_{\inf}(t) = \frac{1}{2} \ln\left(\frac{1}{t}\right),$$

hence

$$\varphi(t) = c(e^{\frac{1}{2}\ln(\frac{1}{t})}t + e^{-\frac{1}{2}\ln(\frac{1}{t})}) = c\sqrt{t}.$$

Because $x_{n_{\inf}(\delta)+1}^{\dagger} < x_{n_{\inf}(\delta)}^{\dagger} = c\sqrt{\delta}$, but $\frac{\sigma_{n_{\inf}(\delta)}^{-1}\delta}{\sqrt{n_{\inf}(\delta)}} = c\sqrt{\frac{\delta}{\ln(\frac{1}{\delta})}}$, Assumption 4.1 (d) is (formally) violated. Numerically, however, even for values of δ significantly smaller then the machine-ϵ, a constant can be found such that Assumption 4.1 (d) holds, i.e., $\sqrt{\delta} \leq C(\delta_0)\sqrt{\frac{\delta}{\ln(\frac{1}{\delta})}}$ for $\delta > \delta_0$. A plot exemplifying this can be found in Figure 2. Consequently, our numerically retrieved convergence rate is close to the predicted $\sqrt{\delta}$-rate, see the right part of Figure 6.
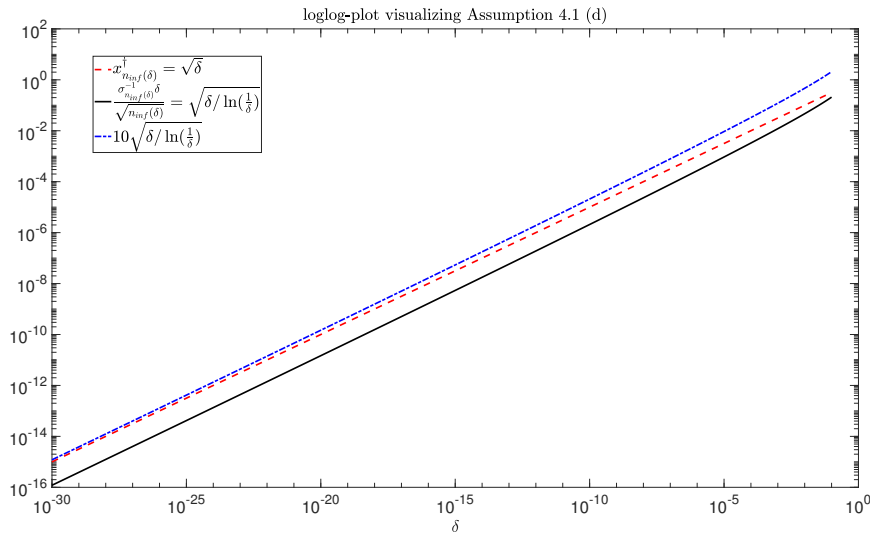
Figure 2. Scenario of Section 4.3.4: In general, this example violates Assumption 4.1 (d), but for $\delta > 10^{-30}$, the assumption is valid with $C = 10$.

4.3.5. Sparse solutions

Our last examples goes back to the setting that sparked the developments around ℓ^1-regularization, namely the case of sparse solutions. Let $x^\dagger = (x_1^\dagger, x_2^\dagger, \ldots, x_{n_0}^\dagger, 0, 0, \ldots)$. Let A be as in Theorem 4.1 with no further restrictions. Then

$$\varphi(t) = \inf_{n \in \mathbb{N}} \sigma_n^{-1} t + \sum_{i=n+1}^\infty |x_i^\dagger| = \min \left\{ \inf_{n \in \mathbb{N}, n \leq n_0} \sigma_n^{-1} t + \sum_{i=n+1}^{n_0} |x_i^\dagger|, \sigma_{n_0+1}^{-1} t \right\},$$

i.e., (only) for sufficiently small values of δ can a linear convergence rate $\varphi(\delta) = c\delta$ be reached. Note that in the literature the linear convergence rate was derived in the ℓ^1-norm, $\|x_\alpha^\delta - x^\dagger\|_{\ell^1} \leq c\delta$, whereas we obtain the same rate in the ℓ^2-norm $\|x_\alpha^\delta - x^\dagger\|_{\ell^2} \leq c\delta$. Since for sufficiently small $\delta > 0$ $n_{\inf}(\delta) = n_0 + 1 = \text{const}$, we also recover from (4.11) the well-known parameter choice $\alpha = c\delta$.

5. Comparison of ℓ^1- and ℓ^2-regularization

In order to get a feeling for the derived convergence rate we compare the result from the previous section with classical, non-oversmoothing, ℓ^1-regularization and ℓ^2-regularization. In order to be able to use explicit expressions for the convergence rates and parameter choice rules, we will only consider the specific model problem of Section 4.3.1. Let us start with classical ℓ^1-regularization, i.e., the approximate solution to (1.1) is obtained via (1.3) but under the assumption that $\|x^\dagger\|_{\ell^1} < \infty$. In [19] a convergence rate

$$\|x_\alpha^\delta - x^\dagger\|_{\ell^1} \leq c\delta^{\frac{\eta-1}{\eta+\beta-\frac{1}{2}}} =: \varphi_{\ell^1}(\delta)$$

was derived using the parameter choice

$$\alpha(\delta) \sim \frac{\delta^2}{\varphi_{\ell^1}(\delta)} = c\delta^{\frac{4\beta+2\eta}{2\eta+2\beta-1}}.$$

Now let us move to ℓ^2-regularization. This corresponds to the classic Tikhonov regularization, i.e., the approximate solution to (1.1) is given by (1.4). It is then well known, see e.g. [1, 2], that under the assumption

$$x^\dagger \in \text{range}((A^*A)^\nu) \tag{5.1}$$

for some $0 \le \nu \le 2$ the best possible convergence rate

$$\|x_\alpha^\delta - x^\dagger\|_{\ell^2} \le c\delta^{\frac{\nu}{\nu+1}} \tag{5.2}$$

can be shown under the a-priori parameter choice

$$\alpha(\delta) \sim \delta^{\frac{2}{\nu+1}}.$$

In the diagonal setting of Section 4.3 the source condition (5.1) can easily be related to the parameters η and β. Namely, (5.1) holds for all ν with $\frac{2\eta-1}{2\beta} > \nu$ [21]. Since we are interested in the largest ν we set $\nu = \frac{2\eta-1}{2\beta}$ for simplicity, acknowledging that actually we should write $\nu = \frac{2\eta-1}{2\beta} + \epsilon$ for arbitrary but small $\epsilon > 0$. With this, the convergence rate becomes

$$\|x_\alpha^\delta - x^\dagger\|_{\ell^2} \le c\delta^{\frac{\nu}{\nu+1}} = c\delta^{\frac{2\eta-1}{2\eta+2\beta-1}} = \varphi(\delta)$$

with $\varphi(\delta)$ from (4.25), and the parameter choice is

$$\alpha(\delta) \sim \frac{\delta^2}{\varphi(\delta)^2} = \frac{\delta^2}{\delta^{\frac{4\eta-2}{2\eta+2\beta-1}}} = \delta^{\frac{4\beta}{2\eta+2\beta-1}}.$$

We summarize the convergence rates and parameter choices in Table 1. One sees that the $\ell^1 - \ell^2$-regularization, corresponding to ℓ^1-regularization with the ℓ^2-norm as error measure which includes the oversmoothing regularization, inherits the parameter choice from the classical ℓ^1-regularization and the convergence rate from the classical ℓ^2-regularization. The parameter choice influences the residual $\|Ax_\alpha^\delta - y^\delta\|$ and the penalty value $\|x_\alpha^\delta\|_{\ell^1}$. Since it is most important to keep the residual on a level of about δ, the ℓ^1-parameter choice is used. This is in line with [18] where it has been observed that for ℓ^1-regularization the regularization parameter obtained via discrepancy principle and a-priori parameter choice always coincide up to constant factors. It is however somewhat surprising that this property appears to remain even when $x^\dagger \notin \ell^1$. The less smooth the solution is, the smaller α has to be chosen. On the other hand, the optimal convergence rate in ℓ^2 is well known to be given by (5.2). Therefore ℓ^1-regularization yields the optimal (ℓ^2-)convergence rate, even when $x^\dagger \notin \ell^1$. Even more, it does not saturate like classical ℓ^2 regularization which possesses the qualification $\nu = \frac{2\eta-1}{2\beta} \le 2$. Take any $\frac{1}{2} < \eta \le 1$. Then $x^\dagger \in \ell^2 \setminus \ell^1$. If $\beta < \frac{\eta}{2} - \frac{1}{4}$, then $\nu > 2$ and ℓ^2-regularization (1.4) has saturated with the limiting convergence rate $\varphi(\delta) \sim \delta^{2/3}$. Oversmoothing ℓ^1-regularization, on the other hand, would yield a higher rate of convergence since it does not saturate.

Table 1. Comparison of (classical) ℓ^1, (classical) ℓ^2 and (oversmoothing) $\ell^1 - \ell^2$ regularization for the diagonal setting of Section 4.1. The parameter choice depends on the penalty functional whereas the convergence rate depends on the norm the regularization error is measured in.

Type	ℓ^1	$\ell^1 - \ell^2$	ℓ^2
rate	$\varphi_{\ell^1}(\delta) = \delta^{\frac{2\eta-2}{2\eta+2\beta-2}}$	$\varphi_{\ell^2}(\delta) = c\delta^{\frac{2\eta-1}{2\eta+2\beta-1}}$	$\varphi_{\ell^2}(\delta) = c\delta^{\frac{2\eta-1}{2\eta+2\beta-1}}$
α	$\delta^{\frac{4\beta+2\eta}{2\eta+2\beta-1}}$	$\delta^{\frac{4\beta+2\eta}{2\eta+2\beta-1}}$	$\delta^{\frac{4\beta}{2\eta+2\beta-1}}$
α recipe	$\alpha = \dfrac{\delta^2}{\varphi_{\ell^1}}$	$\alpha = \dfrac{\delta^2}{\varphi_{\ell^2}(\delta)\delta^{\frac{-1}{2\eta+2\beta-1}}}$	$\alpha = \dfrac{\delta^2}{\varphi_{\ell^2}(\delta)}$

6. Numerical examples

In this section we consider an operator specifically tailored to the setting of Section 4.1. We start with the Voltera operator

$$[\tilde{A}x](s) = \int_0^s x(t)\,dt$$

and discretize $\tilde{A}$ with the rectangular rule at $N = 400$ points. In order to ensure our desired properties of the model cases of Section 4.3 we compute the SVD of the resulting matrix and manually set its singular values σ_i accordingly. This means that the actual operator A in (1.1) is an operator possessing the same singular vectors $\{u_i\}$ and $\{v_i\}$ as $\tilde{A}$, but different singular values $\{\sigma_i\}$. Using the SVD, we construct our solution such that $x_i^\dagger = \langle x^\dagger, v_i \rangle$ holds according to the scenario. We add random noise to the data $y = Ax^\dagger$ such that $\|y - y^\delta\| = \delta$. The range of δ is such that the relative error is between 25% and 0.2%. The solutions are computed via

$$x_\alpha^\delta = \text{argmin}\left\{\|Ax - y^\delta\|^2 + \alpha\|x\|_{\ell_1}\right\},$$

where the ℓ^1-norm is taken using the coefficients with respect to the basis originating from the SVD. We compute the reconstruction error in the ℓ_2 norm as well as the residuals. For larger values of η we can observe the convergence rate directly. For smaller values of η, we have to compensate for the error introduced by the discretization level. Namely, since we use a discretization level $N = 400$, numerically we actually measure

$$\|P_{400}(x_\alpha^\delta - x^\dagger)\|_{\ell^2}$$

with the projectors P as before being the cut-off after $N = 400$ elements. In the plots of the convergence rates we show

$$\|P_{400}(x_\alpha^\delta - x^\dagger) + (I - P_{400})x^\dagger\|_{\ell^2}. \tag{6.1}$$

The second term can be calculated analytically and is supposed to correct for the fact that we cannot measure the regularization error for coefficients corresponding to larger n, i.e., we add the tail of $x^\dagger$ that can not be observed. Because x_α^δ has only finitely many non-zero components and we use a sufficiently large discretization, we do not make any error with respect to the tail of the solutions.

Our main focus is on the fully diagonal case of Section 4.3.1, i.e, $\sigma_i = i^{-\beta}$ for $\beta > 0$ and $x_i^\dagger = i^{-\eta}$ for $\eta > \frac{1}{2}$. Here, the regularization parameter is chosen a-priori according to (4.26) with $c_\alpha = 1$.

We show selected plots of the convergence rates in Figure 3 for $\beta = 1$ and Figure 4 for $\beta = 2$. The plots are given with logarithmic scales for both axis. In each plot of the convergence rates we added the regression line for the assumption $\|x_\alpha^\delta - x^\dagger\|_{\ell^2} = c\delta^e$. The value of e is given in the legend. The regularization parameter is shown in the title of the figures in the form $\alpha = \delta^a$, a given. The result of our simulations for a larger number of parameters η are shown in Tables 2 and 3 for $\beta = 1$ and $\beta = 2$, respectively. We see that for all values of η the predicted and measured convergence rate coincides nicely. Additionally, the residual remains stable around $\|Ax_\alpha^\delta - y^\delta\| \sim \delta$. For small values of η and $\beta = 1$ the residual is a bit smaller than expected. We suppose this is due to the cut-off of $x^\dagger$ due to the discretization. For correct results we would have to include a tail of the residual similar to (6.1). If η is very large, i.e. the components of the solution decay rapidly, the observed convergence rate is basically linear. We suppose this is due to numerical effects as numerically those solutions are de facto sparse.

Finally, we compare the theoretically predicted convergence rates from the remaining model cases of Section 4.3 to measured rates from numerical experiments. In all experiments, the regularization parameter was chosen according to the discrepancy principle (4.24) with $\tau_1 = 1.1$ and $\tau_2 = 1.3$. We obtained a good match between the theoretical and observed convergence rates in the cases of sections 4.3.2, 4.3.4, and 4.3.4. Only case 4.3.3 showed a significant difference. This is no surprise, since Assumption 4.1 does not hold in this case.

Table 2. Convergence rates for $\beta = 1$ and various values η. α in the form $\alpha = \delta^a$. Measured and predicted regularization error in the form $\|x_\alpha^\delta - x^\dagger\|_{\ell^2} = c\delta^e$. Residual given in the form $\|Ax_\alpha^\delta - y^\delta\| = c\delta^d$.

η	α, a	measured rate, e	predicted rate, e	residual, d
0.55	2.42	0.047	0.048	1.1
0.6	2.36	0.09	0.091	1.08
0.7	2.25	0.163	0.166	1.06
0.8	2.15	0.229	0.23	1.028
0.9	2.07	0.284	0.286	1.007
1	2	0.33	0.333	1.01
1.05	1.97	0.359	0.355	1.006
1.1	1.94	0.372	0.375	1.006
1.3	1.83	0.458	0.444	1.01
1.5	1.75	0.515	0.5	1.01
2	1.6	0.595	0.6	1.01
2.5	1.5	0.659	0.667	0.996
3	1.42	0.698	0.714	0.996
6	1.23	0.81	0.85	0.997

Table 3. Convergence rates for $\beta = 2$ and various values η. α in the form $\alpha = \delta^a$. Measured and predicted regularization error in the form $\|x_\alpha^\delta - x^\dagger\|_{\ell^2} = c\delta^e$. Residual given in the form $\|Ax_\alpha^\delta - y^\delta\| = c\delta^d$.

| η	$\alpha = \delta^a$, a	measured rate, e	predicted rate, e	residual, d
0.55	2.22	0.024	0.024	1.0005
0.6	2.19	0.0473	0.0476	1.002
0.7	2.13	0.089	0.091	1.01
0.8	2.09	0.128	0.13	1.006
0.9	2.04	0.166	0.167	1.006
1.01	1.996	0.209	0.203	0.999
1.1	1.96	0.236	0.23	0.994
1.3	1.89	0.284	0.286	1.002
1.5	1.83	0.329	0.333	0.999
1.75	1.76	0.384	0.385	0.996
2	1.71	0.421	0.428	0.999
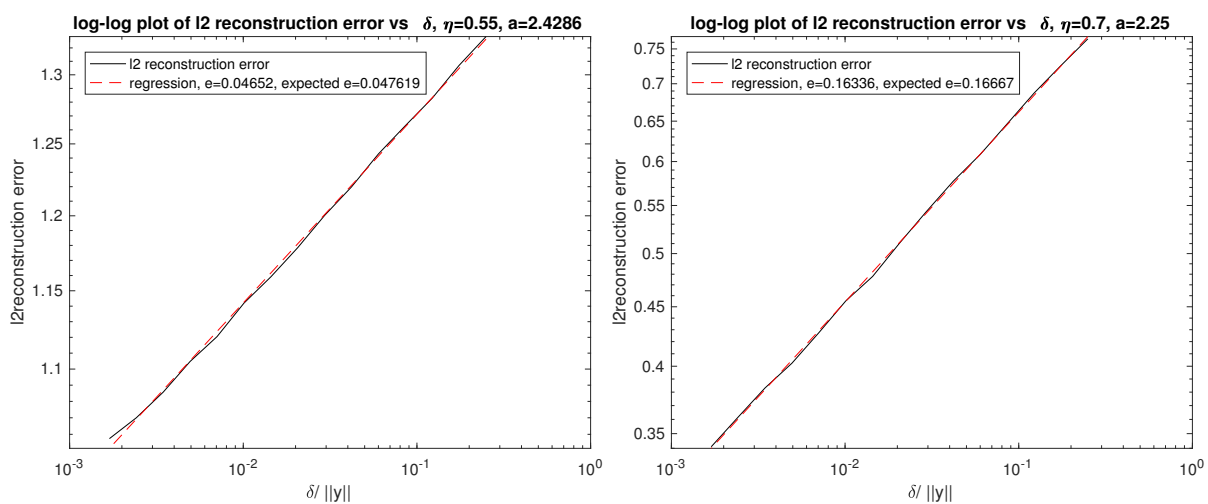
Figure 3. Numerically observed convergence rates for $\beta = 1$ and $\eta \in \{0.55, 0.7\}$. From the measured reconstruction error (solid line) we calculated the regression for the assumption $\|x_\alpha^\delta - x^\dagger\|_{\ell^2} = c\delta^e$, shown in the broken line. The theoretical convergence rate (4.25) is matched almost perfectly.
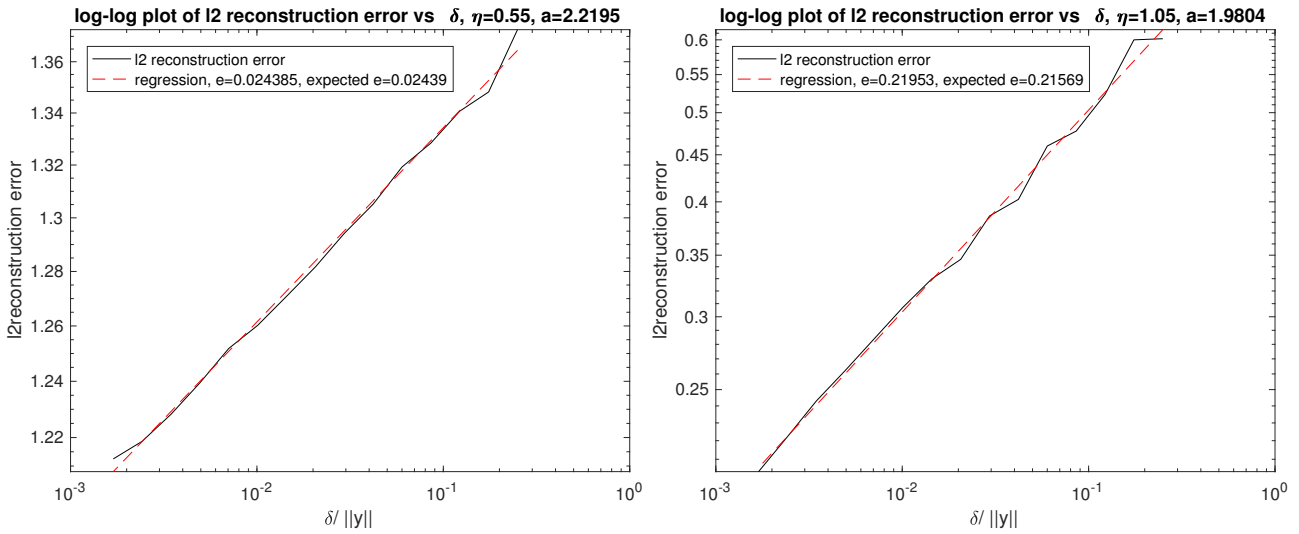
Figure 4. Numerically observed convergence rates for $\beta = 2$ and $\eta \in \{0.55, 1.05\}$. From the measured reconstruction error (solid line) we calculated the regression for the assumption $\|x_\alpha^\delta - x^\dagger\|_{\ell^2} = c\delta^e$, shown in the broken line. The theoretical convergence rate (4.25) is matched almost perfectly.
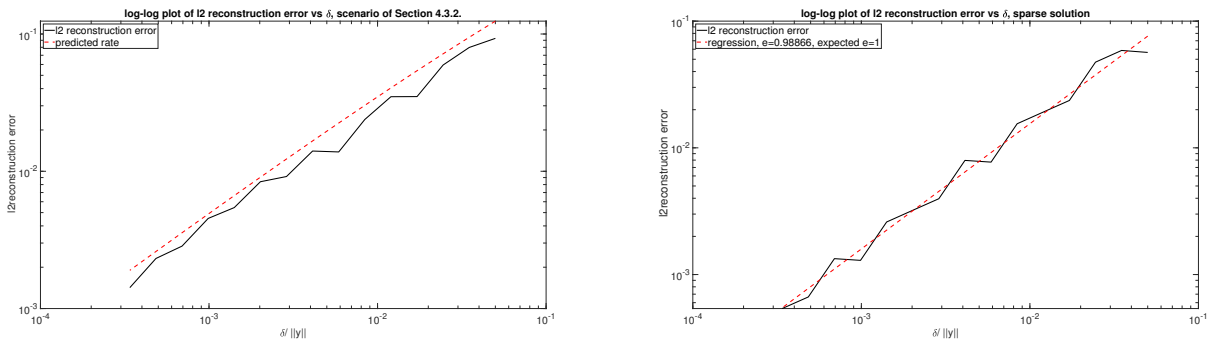
Figure 5. Numerically observed convergence rate for model cases 4.3.2 (left) and the sparsity case 4.3.5 (right) where Assumption 4.1 holds. In both cases we have a good fit between measured and theoretical convergence rate. Note that in the sparsity case, only the first 4 components of $x^\dagger$ were non-zero. The regularized approximation x_α^δ had at most 5 non-zero entries.
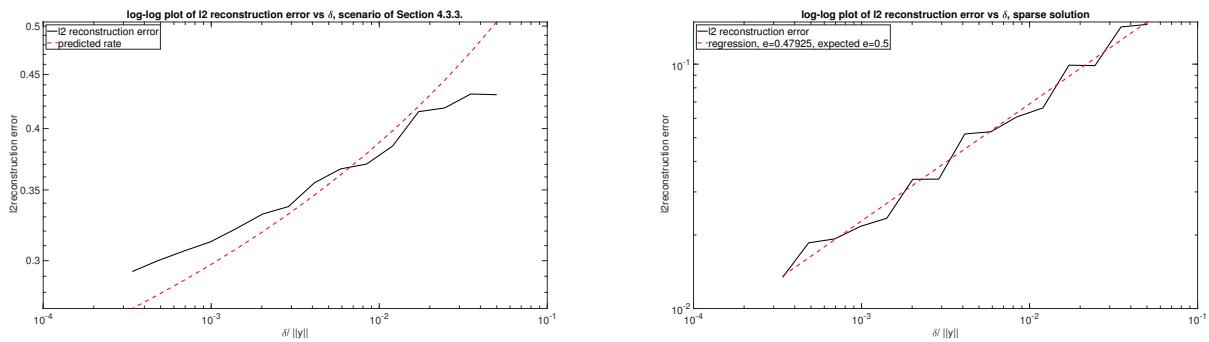
Figure 6. Numerically observed convergence rate for the model cases 4.3.3 (left) and 4.3.4 (right) where Assumption 4.1 is violated. In the left-hand plot, predicted and measured convergence rate show a clear mismatch, indicating that Assumption 4.1 is indeed needed for our analysis. On the right-hand side we see a good match between theoretical and measured rates. We attribute this to the phenomenon discussed in Section 4.3.4, namely that Assumption 4.1 (d) does not hold for all $\delta > 0$, but for all numerically reasonable $\delta > \delta_0 > 0$.

7. Conclusions

We have shown that oversmoothing regularization, i.e., Tikhonov regularization with a penalty that takes the value infinity in the exact solution, yields existence and stability of regularized solutions. We have discussed why it is difficult to show convergence of regularized solutions to the true solution in the limit of vanishing data noise. For the specific case of Tikhonov-regularization with ℓ^1-penalty term under a diagonal operator we have derived convergence rates of regularized solutions to the true solution even if those does not belong to ℓ^1 anymore. The theoretical convergence rates have been verified numerically for a model problem.

Acknowledgments

Conflict of interest

The authors declare there is no conflicts of interest in this paper.

References

1. H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Dordrecht: Kluwer Academic Publishers, 1996.

2. A. K. Louis, *Inverse und schlecht gestellt Probleme*, Stuttgart: Teubner, 1989.

3. D. Gerth and B. Hofmann, *On ℓ^1-regularization under continuity of the forward operator in weaker topologies*. In: New Trends in Parameter Identification for Mathematical Models (Ed.: B. Hofmann, A. Leitao, J.P. Zubelli), Cham: Birkhäuser, (2018), 67–88.

4. B. Hofmann, B. Kaltenbacher, C. Pöschl, et al. *A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators*, Inverse Probl., **23** (2007), 987–1010.

5. B. Hofmann and P. Mathé, *Parameter choice in Banach space regularization under variational inequalities*, Inverse Probl., **28** (2012), 104006.

6. F. Natterer, *Error bounds for Tikhonov regularization in Hilbert scales*, Appl. Anal., **18** (1984), 29–37.

7. B. Hofmann and P. Mathé, *Tikhonov regularization with oversmoothing penalty for non-linear ill-posed problems in Hilbert scales*, Inverse Probl., **34** (2018), 015007.

8. B. Hofmann and P. Mathé, *A priori parameter choice in Tikhonov regularization with oversmoothing penalty for non-linear ill-posed problems*, 2019. Available from: `https://arxiv.org/abs/1904.02014`.

9. T. Schuster, B. Kaltenbacher, B. Hofmann, et al. *Regularization Methods in Banach spaces*, Berlin/Boston: De Gruyter, 2012.

10. B. Hofmann, *On smoothness concepts in regularization for nonlinear inverse problems in Banach spaces*. In: Mathematical and Computational Modeling: With Applications in Natural and Social Sciences, Engineering, and the Arts (Ed.: R. Melnik), New Jersey: John Wiley, (2015), 192–221.

11. M. Burger, J. Flemming and B. Hofmann, *Convergence rates in ℓ^1-regularization if the sparsity assumption fails*, Inverse Probl., **29** (2013), 025013.

12. J. Flemming, B. Hofmann and I. Veselic, *A unified approach to convergence rates for ℓ^1-regularization and lacking sparsity*, J. Inverse Ill-posed Probl., **24** (2016), 139–148.

13. I. Daubechies, M. Defrise and C. De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Commun. Pur. Appl. Math., **57** (2004), 1413–1457.

14. M. Grasmair, M. Haltmeier and O. Scherzer, *Sparse regularization with lq penalty term*, Inverse Probl., **24** (2008), 055020.

15. D. Lorenz, *Convergence rates and source conditions for Tikhonov regularization with sparsity constraints*, J. Inverse ill-posed Probl, **16** (2008), 463–478.

16. R. Ramlau, *Regularization properties of Tikhonov regularization with sparsity constraints*, Electron. T. Numer. Anal., **30** (2008), 54–74.

17. J. Flemming, *Convergence rates for l1-regularization without injectivity-type assumptions*, Inverse Probl., **32** (2016), 095001.

18. D. Gerth, J. Flemming, *Injectivity and weak*-to-weak continuity suffice for convergence rates in l1-regularization*, J. Inverse Ill-posed Probl., **26** (2018), 85–94.

19. D. Gerth, *Convergence rates for ℓ^1-regularization without the help of a variational inequality*, Electron. T. Numer. Anal., **46** (2017), 233–244.

20. B. Hofmann, P. Mathé and M. Schieck, *Modulus of continuity for conditionally stable ill-posed problems in Hilbert space*, J. Inverse Ill-posed Probl., **16** (2008), 567–585.

21. D. Gerth, *Using Landweber iteration to quantify source conditions - a numerical study*, J. Inverse Ill-posed Probl., **27** (2019), 367–383.