Tikhonov regularization with oversmoothing penalties

D. Gerth

Preprint 2016-8

Preprintreihe der Fakultät für Mathematik
ISSN 1614-8835

Tikhonov regularization with oversmoothing penalties

Daniel Gerth *

December 15, 2016

Key words: Tikhonov regularization, convergence rates, ℓ^1-regularization, oversmoothing

MSC: 65J20, 47A52

Abstract

In the last decade ℓ^1-regularization became a powerful and popular tool for the regularization of Inverse Problems. While in the early years sparse solution were in the focus of research, recently also the case that the coefficients of the exact solution decay sufficiently fast was under consideration. In this paper we seek to show that ℓ^1-regularization is applicable and leads to optimal convergence rates even when the exact solution does not belong to ℓ^1 but only to ℓ^2. This is a particular example of oversmoothing regularization, i.e., the penalty implies smoothness properties the exact solution does not fulfill. We will make some statements on convergence also in this general context.

1 Introduction

Our task is to find an $x \in X$ satisfying the ill-posed linear equation

$$Ax = y \tag{1}$$

between Banach spaces X and Y for given $y \in Y$. We assume, however, that only noisy data y^δ is available with $||y - y^\delta|| \leq \delta$ for $\delta > 0$. In order to determine a stable solution of (1) from the noisy data, regularization is required. We employ a Tikhonov-type functional

$$T_\alpha^\delta(x) = \frac{1}{r}||Ax - y^\delta||^r + \alpha\mathcal{R}(x) \tag{2}$$

*Technische Universität Chemnitz, Faculty of Mathematics, D-09107 Chemnitz, Germany, daniel.gerth@mathematik.tu-chemnitz.de Research supported by DFG grant HO 1454/8-2.

where $1 \leq r < \infty$, $\alpha > 0$ is the regularization parameter which has to be chosen appropriately and $\mathcal{R} : \mathcal{D}(\mathcal{R}) \subset X \to \mathbb{R}$ the penalty functional. The minimizer of (2) is the regularized solution, i.e.,

$$x_\alpha^\delta := \min_{x \in \mathcal{D}(A)} T_\alpha^\delta(x). \tag{3}$$

Problems of this form have been discussed widely in the literature, see e.g. [6] and references therein. To the best of the author's knowledge, in all these publications (with the exception of the paper discussed in Section 2) it is assumed that the exact solution, which we donate by $x^\dagger$ throughout this work, attains a finite value of the penalty, $\mathcal{R}(x^\dagger) < \infty$. In this paper we seek to generalize this and investigate the case that $\mathcal{R}(x^\dagger)$ is infinite. Since this is an unusual concept, let us motivate its use first. A practical application may be the approximation of an object emphasizing certain features that the true solution does not possess. For example, in order to save memory one may be interested to find a sparse solution to an Inverse Problem although it is a priori clear that the exact solution is not sparse. Another motivation comes from stochastic regularization methods where x is regarded as a random variable. Defining an infinite dimensional random variable has the effect that the random variable takes finite expectation only in a smoother space. As the main example, an infinite dimensional Gaussian random variable is an element of L_2 with probability zero, but lies in some Sobolev space H^s for some negative value $s < 0$. It may also be that the minimization of the functional (2) is particularly simple for a certain choice of $\mathcal{R}$ or that the functional $\mathcal{R}$ leads to improved convergence behavior of the regularized solutions. For example we will later demonstrate that ℓ^1-regularization does not suffer from the saturation effect well known in classical ℓ^2-Tikhonov regularization. Finally, considering an infinite value of the penalty at the exact solution appears to be a gap in the theory of inverse problems that may spark new ideas also in existing branches of the theory of Inverse Problems.

The paper is organized as follows. In the next Section we discuss Natterer's paper on Tikhonov regularization in Hilbert scales which includes oversmoothing regularization. In Section 3 we discuss existence and stability of the regularized solutions as well as convergence to the exact solution in the general case (2). In order to show convergence rates we pick a particularly structured problem, namely a diagonal operator wich allows us to explicitly calculate convergence rates in Section 4. We discuss the results and compare it with classical ℓ^1 and classical ℓ^2 regularization in Section 5. Finally we numerically verify the theoretical results.

2 Literature survey: Natterer's paper

To the best of the author's knowledge, only one paper exists in the literature that includes a direct statement on oversmoothing regularization, namely the paper [1] by Natterer from 1984. Natterer considers a linear problem (1) between Hilbert spaces X and Y. The regularized solution is obtained by a variational minimization problem

$$x_\alpha^\delta := \min_x ||Ax - y^\delta||^2 + \alpha ||x||_p \tag{4}$$

of type (2) where the penalty is a norm in a Hilbert scale. A family $\{H_s\}_{s\in\mathbb{R}}$ of Hilbert spaces is called Hilbert scale if $H_t \subseteq H_s$ whenever $s < t$ and the inclusion is a continuous embedding, i.e., there exists $c_{s,t} > 0$ such that

$$||x||_s \leq c_{s,t} ||x||_t.$$

Natterer assumes that there is an unbounded, self-adjoint and strictly positive definite operator T in X such that

$$||x||_s = ||T^s x||_X, \qquad s \in \mathbb{R}$$

defines a norm in H_s. this satisfies the condition above.

Assuming $||Ax^\dagger - y^\delta|| \leq \delta$, that $||x^\dagger||_q \leq \rho$ for some $q \geq 0$ and that there exists some $a > 0$ such that with two constants $m, M > 0$

$$m||x||_{-a} \leq ||Ax|| \leq M||x||_{-a} \tag{5}$$

he shows that

$$||x_\alpha^\delta - x^\dagger|| \leq C\delta^{\frac{q}{a+q}} \rho^{\frac{a}{a+q}} \tag{6}$$

provided $p > (q-a)/2)$ if the regularization parameter is chosen via

$$\alpha = C(\rho)\delta^{\frac{2(a+p)}{a+q}}.$$

It is interesting here that p, i.e., the smoothness of the penalty, is bounded from below but not from above. As Natterer states, "there is nothing wrong with high order regularization, even well above the order of the smoothness of the exact solution" (which is here given by q). Even though the exact solution may have infinite p-norm, we not only still obtain regularization but the rate of convergence remains unchanged. The only adjustment to be made is a decrease of the regularization parameter.

Although Natterers paper is quickly summarized, several interesting observations can be made which we will use later. First, note that the convergence rate depends on the smoothness of the exact solution q and the smoothing of the operator a. In particular, the penalty has no influence on the speed of convergence. However, in order to obtain the convergence rate, the regularization parameter has to be adjusted according to the smoothness p of the penalty. In case $q = p$ we have $\alpha \sim \delta^2$. Increasing p above q then yields smaller regularization parameters (or equivalently: larger exponents). On the other hand, in the case $p = (q-a)/2$ we have $\alpha \sim \delta$ which is the largest possible regularization parameter. Next, observe that the smoothness condition (5) relates the norms in x with those in $AX \subset Y$. This means that whenever residuals $||Ax_1 - Ax_2||$ are close, also the distance between the elements x_1 and x_2 is small, although in a weaker norm. Finally we would like to mention that in a Hilbert scale there is a common dense subset among all spaces of the scale, hence any element of H_{s_1} can be approximated arbitrarily well with elements from H_{s_2}, $s_2 > s_1$.

3 Convergence for general penalty functionals

In this section we will show that all basic properties we seek to find in a regularization method also can be expected in the oversmoothing case. We will discuss the existence of minimizers, their stability w.r.t noise in the right-hand side and convergence of the approximate solutions when the noise goes to zero. The following assumption is standard in regularization theory, see for example [6, 12]

Assumption 3.1.

a) *X and Y are infinite dimensional Banach spaces where in addition to the norm topologies $||\cdot||_X$ and $||\cdot||_Y$ weaker topologies τ_X and τ_Y are under consideration for X and Y, respectively, such that the norm $||\cdot||_Y$ is τ_Y-sequentially lower semicontinuous*

b) *The domain $\mathcal{D}(A)$ is a convex and τ_X-sequentially closed subset of X*

c) *The operator $A : \mathcal{D}(A) \subset X \to Y$ is $\tau_X - \tau_Y$-continuous*

d) *$\mathcal{R} : X \to [0, \infty]$ is a τ_X-sequentially lower semicontinuous, convex and proper functional, the latter meaning*

$$\mathcal{D}(\mathcal{R}) := \{x \in X : \mathcal{R}(x) < \infty\} \neq \emptyset.$$

Moreover, we assume $\mathcal{D} := \mathcal{D}(A) \cap \mathcal{D}(\mathcal{R}) \neq \emptyset$.

e) *The penalty functional $\mathcal{R}$ is assumed to be stabilizing in the sense that the sublevel sets*

$$\mathcal{M}_{\mathcal{R}}(c) := \{x \in X : \mathcal{R}(x) \leq c\}$$

are sequentially precompact w.r.t. the topology τ_X in X.

We start with the existence of a minimizer of (2).

Lemma 3.1 (existence). *For all fixed $\alpha, \delta > 0$ a minimizer of* (2) *exists.*

Proof. Take $\alpha > 0$ arbitrary but fixed. Since by Assumption 3.1 d) $\mathcal{D} := \mathcal{D}(A) \cap \mathcal{D}(\mathcal{R}) \neq \emptyset$ there exists $\tilde{x} \in X$ with $||Ax - y^\delta|| < \infty$, $\mathcal{R}(\tilde{x}) < \infty$ and hence $T_\alpha^\delta(\tilde{x}) < \infty$. The remainder of the proof follows from standard arguments. We include it for the convenience of the reader. There exists a sequence $\{x_k\}_{k\in\mathbb{N}} \in \mathcal{D}$ such that

$$\lim_{k\to\infty} T_\alpha^\delta(x_k) = c := \inf\{T_\alpha^\delta(x) : x \in \mathcal{D}\} \leq T_\alpha^\delta(\tilde{x}) < \infty.$$

Thus, $\{T_\alpha^\delta(x_k)\}$ is a bounded sequence, in particular $\alpha\mathcal{R}(x_k) < \infty$ and since $\alpha > 0$ is fixed $\mathcal{R}(x_k) < \infty$. By Assumption 3.1 $\{x_k\}$ has a subsequence $\{x_{k_j}\}$ weakly converging to some $\hat{x} \in \mathcal{D}$ since $\mathcal{D}$ is convex and closed and therefore weakly closed. Since A is assumed to be weak-to-weak continuous, $A(x_{k_j})-y^\delta \rightharpoonup A(\hat{x})-y^\delta$ in Y and, since the norm is weakly lower semi-continuous

$$||A\hat{x}-y^\delta|| \le \liminf_{j\to\infty} ||Ax_{k_j}-y^\delta||.$$

On the other hand, for the lower-semicontinuous functional $\mathcal{R}$, which is weakly lower-semicontinuous, we have

$$\mathcal{R}(\hat{x}) \le \liminf_{j\to\infty} \mathcal{R}(x_{k_j}).$$

Thus $\hat{x}$ minimizes T_α^δ. □

Since existence of a minimizer to (2) depends on the properties of the penalty $\mathcal{R}$ and not on the exact solution the Lemma is no surprise. Similarly one sees that stability of the minimizers with respect to disturbances in the right hand side carries over from the standard theory.

Lemma 3.2 (**stability**). *Let Assumption 3.1 hold and let $\alpha > 0$. If $\{y_k\}$ is a sequence converging to y^δ in Y with respect to the norm topology, then every sequence $\{x_k\}$ with*

$$x_k \in \arg\min\{T_\alpha^{y_k}(x) : x \in X\}$$

has a subsequence that converges with respect to τ_X. The limit of every τ_X-convergent subsequence $\{x_{\tilde{k}}\}$ of $\{x_k\}$ is a minimizer $\tilde{x}$ of $T_\alpha^{y^\delta}$ and $\mathcal{R}(x_{\tilde{k}})$ converges to $\mathcal{R}(\tilde{x})$.

Proof. Since $\alpha > 0$ every minimizer $\tilde{x}$ of (2) fulfills $\mathcal{R}(\tilde{x}) < \infty$. The proof then follows by standard arguments. □

The next step is to investigate convergence of the minimizers to the exact solution when the noise tends to zero. In standard theory, the conditions $\alpha \to 0$ and $\frac{\delta^2}{\alpha} \to 0$ as $\delta \to 0$ ensure the weak convergence of subsequences to the exact solution. We start our considerations by identifying some necessary conditions for this statement in the case of oversmoothing regularization.

Theorem 3.1. *Let $x^\dagger \in \mathcal{D}$ with $\mathcal{R}(x^\dagger) = \infty$ denote a solution to (1). If the Tikhonov regularization under consideration is weakly convergent to $x^\dagger$ for a parameter choice rule $\alpha = \alpha(\delta, y^\delta)$, then the following items must hold for a sequence $x_k := x_{\alpha_k}^{\delta_k}$ with $\delta_k \to 0$.*

a) $\lim_{k\to\infty} \mathcal{R}(x_k) = \infty$

b) $\lim_{k\to\infty} \alpha_k = 0$.

c) $\lim_{k\to\infty} \alpha_k \mathcal{R}(x_k) \le C < \infty$

Proof. By Lemma 3.1, $\mathcal{R}(x_k) < \infty$ for all $k \in \mathbb{N}$. If, however, we assume that there is a subsequence $\{x_{k_j}\}$ with $\mathcal{R}(x_{k_j}) \le c < \infty$ uniformly for all $j \in \mathbb{N}$, then the assumed weak convergence $x_{k_j} \rightharpoonup x^\dagger$ in X implies that $\mathcal{R}(x^\dagger) \le c$. This

contradicts the assumption $\mathcal{R}(x^\dagger) = \infty$ and yields a).
Now take some fixed $\hat{x} \in \mathcal{D}$ and keep in mind the definition of the x_k as minimizers of the functional (2). It is

$$\begin{aligned}\frac{1}{r}||Ax_k - y^{\delta_k}||^r + \alpha_k \mathcal{R}(x_k) &\le \frac{1}{r}||A\hat{x} - y_k^\delta||^r + \alpha_k \mathcal{R}(\hat{x}) \\ &\le \frac{2^{r-1}}{r}\delta_k^r + \frac{2^{r-1}}{r}||A\hat{x} - y||^r + \alpha_k \mathcal{R}(\hat{x}).\end{aligned}$$

Therefore

$$\alpha_k(\mathcal{R}(x_k) - \mathcal{R}(\hat{x})) \le c(r)\,(\delta_k^r + ||A\hat{x} - y||^r) \le C < \infty.$$

Since $\mathcal{R}(\hat{x}) < \infty$ we need $\lim_{k\to\infty} \alpha_k = 0$ in order to allow $\lim_{k\to\infty} \mathcal{R}(x_k) = \infty$ as necessary due to a). Additionally, the product $\alpha_k \mathcal{R}(x_k)$ has to stay bounded, yielding c). □

The regularization parameter is a free parameter that has to be chosen appropriately. Theorem 3.1 states that in order to hope to obtain a sequence of regularized solutions weakly convergent to $x^\dagger$, then the regularization parameter has to go to zero, in particular, be bounded from above. An immediate consequence of this is that indeed condition c) of Theorem 3.1 is fulfilled.

Lemma 3.3. *Let $\{y_\delta\}$ be a sequence of noisy data with $||y^{\delta_k} - y|| \le \delta_k \to 0$ for $k \to \infty$. Let $0 \le \alpha_k = \alpha_k(\delta_k, y_k^\delta) \le \bar{\alpha} < \infty$ for all $k \in \mathbb{N}$. Then the minimizers $x_{\alpha_k}^{\delta_k}$ fulfill*

$$T_{\alpha_k}^{\delta_k}(x_{\alpha_k}^{\delta_k}) \le C < \infty$$

and consequently

$$\alpha_k \mathcal{R}(x_{\alpha_k}^{\delta_k}) \le C < \infty$$

for all $k \in \mathbb{N}$ with some constant $C > 0$.

Proof. Since the α_k are bounded the claim follows immediately from the proof of Theorem 3.1. □

In order to show convergence of the approximate solutions when the noise tends to zero, we require that $x^\dagger$ can be approximated arbitrarily well with objects in $\mathcal{D}(\mathcal{R})$.

Assumption 3.2. *There exists at least one sequence $\{\tilde{x}_j\} \subset \mathcal{D}(\mathcal{R}) \cap X$ such that $\tilde{x}_j \rightharpoonup x^\dagger$ and $\mathcal{R}(\tilde{x}_j) < \infty$ for all $j \in \mathbb{N}$.*

This is for example the case if $\mathcal{D}(\mathcal{R})$ is dense in X (or $\mathcal{D}(A)$, respectively). Later we will more specifically assume that there exists a linear projector $P_n : X \to \mathcal{D}(\mathcal{R})$ such that $\mathcal{R}(P_n x^\dagger) < \infty$ for all $n \in \mathbb{N}$ and $\lim_{n\to\infty} P_n x^\dagger = x^\dagger$.

Assumption 3.2 allows to show that the regularized solutions converge to the solution $x^\dagger$ when the noise goes to zero. Indeed they fulfill $\mathcal{R}(x_\alpha^\delta) \to \infty$. The only requirement is that the regularization parameter goes to zero, no additional restriction is necessary.

Lemma 3.4. *Let $\{y^{\delta_k}\}$ be a sequence of noisy data with $||y^{\delta_k} - y|| \leq \delta_k \to 0$ for $k \to \infty$. Let $0 \leq \alpha_k = \alpha_k(\delta_k, y^{\delta_k}) \leq \bar{\alpha} < \infty$ be a parameter choice fulfilling $\alpha_k \to 0$ as $k \to \infty$. Then among the minimizers $x_{\alpha_k}^{\delta_k}$ there is a subsequence weakly converging to $x^\dagger$. In particular, it is*

$$\mathcal{R}(x_{\alpha_k}^{\delta_k}) \to \infty \quad \text{as} \quad k \to \infty.$$

Proof. Assumption 3.2 yields the existence of a sequence $\tilde{x}_j \rightharpoonup x^\dagger$ in $\mathcal{D}$ with $\mathcal{R}(\tilde{x}_j) < \infty \; \forall j$. Due to Assumption 3.1 c) this sequence also fulfills $A(\tilde{x}_j) \rightharpoonup A(x^\dagger)$. Since $\alpha_k \to 0$ as $k \to \infty$ we can pick a subsequence α_{k_j} which we simply denote by α_j such that

$$\lim_{j\to\infty} \alpha_j \mathcal{R}(\tilde{x}_j) = 0. \tag{7}$$

Since the $x_{\alpha_k}^{\delta_k}$ are minimizers of (2), the functional value has to be smaller than the one of the $\tilde{x}_j$ at the respective α_j, i.e.

$$0 \leq T_{\alpha_j}^{\delta_j}(x_{\alpha_j}^{\delta_j}) \leq T_{\alpha_j}^{\delta_j}(\tilde{x}_j) \quad \forall j. \tag{8}$$

By assumption,

$$0 \leq \liminf_{j\to\infty} T_{\alpha_j}^{\delta_j}(\tilde{x}_j) \leq \liminf_{j\to\infty} \left(\frac{2^{r-1}}{r} ||A\tilde{x}_j - Ax^\dagger||^r + \frac{2^{r-1}}{r} \delta_j^r + \alpha_j \mathcal{R}(\tilde{x}_j) \right) = 0.$$

Note that the $\tilde{x}_j$ are not necessary minimizers of (2). These minimizers however cannot perform worse, see (8), and hence there exists a subsequence of $\{x_{\alpha_k}^{\delta_k}\}$, which we denote by $\{x_j\}$ with

$$0 \leq \liminf_{j\to\infty} T_{\alpha_j}^{\delta_j}(x_j) = 0$$

and consequently

$$\liminf_{j\to\infty} ||Ax_j - y^{\delta_j}|| = 0.$$

Denote with $\hat{x}$ the weak limit of the sequence $\{x_l\}$ (or move to a subsubsequence, since the functional is bounded). It holds in particular

$$||A\hat{x} - y||^r \leq \liminf_{j\to\infty} ||Ax_j - y||^r \leq \liminf_{j\to\infty} \frac{2^{r-1}}{r} ||Ax_j - y^{\delta_j}||^r + \lim_{j\to\infty} \frac{2^{r-1}}{r} \delta_j^p = 0,$$

i.e., $\hat{x}$ is a solution to (1). As in Theorem (3.1) a) it is $\lim_{j\to\infty} \mathcal{R}(x_{\alpha_k}^{\delta_k}) = \infty$. Otherwise, there would need to exist a solution to (1) with finite value of $\mathcal{R}(\cdot)$. □

The proof also holds in the case that $\mathcal{R}(x^\dagger) < \infty$. If (1) admits more than one solution, however, weak convergence to the minimum-norm solution, i.e. a solution with minimal $\mathcal{R}$-value among all possible solutions, in the classical sense is not guaranteed anymore. Namely, if the exact solution is not unique one will get convergence to some solution, but it is not clear which. In the standard theory convergence to a minimum-norm solution is obtained via the condition $\delta^p/\alpha \to 0$.

In the oversmoothing scenario we strongly believe that a similar statement can be made. However, at this point there is still a gap in the proof of the corresponding theorem. We shall give the Theorem and its proof in the following and indicate the so far missing link. Since a minimum-norm solution with respect to $\mathcal{R}$ does not exist in our context, the idea is to find a functional $\mathcal{S}$ related to $\mathcal{R}$ for which $\mathcal{S}(x^\dagger) < \infty$ and use this to define $\mathcal{S}$-minimum norm solutions. The following Assumption links $\mathcal{R}$ and $\mathcal{S}$. We suppose that a third condition is needed to complete the proof of the convergence theorem below. Note that later when we discuss convergence rates we get stuck at precisely the same point.

Assumption 3.3. *a) There exists a family of linear projectors $P_n : X \to X$ depending on a real valued variable n such that $\lim_{n\to\infty} P_n x = x$ for all $x \in X$ and $\mathcal{R}(P_n x^\dagger) < \infty$ for all $n \in \mathbb{N}$*

b) There exists a linear functional $\mathcal{S} : X \to \mathbb{R}$ with $\mathcal{D}(\mathcal{R}) \subset \mathcal{D}(\mathcal{S}) \subseteq X$ with $\mathcal{S}(x^\dagger) < \infty$, where the values $\mathcal{S}(P_n x^\dagger)$ increase monotonically and a monotonically increasing function $f : \mathbb{R} \to \mathbb{R}$ with $\lim_{n\to\infty} f(n) = \infty$ such that $\mathcal{R}(P_n x) \le f(n)\mathcal{S}(P_n x)$ for all $x \in \mathcal{D}(\mathcal{R})$

Assumption 3.3 is not finalized yet so let us quickly discuss it. Point a) is a special case of the requirement that $x^\dagger$ can be approximated arbitrarily well with elements in $\mathcal{D}(\mathcal{R})$, i.e., Assumption 3.2. Item b) requires that these projectors allow to measure the growth of $\mathcal{R}(P_n x^\dagger)$ with respect to n and a weaker functional $\mathcal{S}$. This condition is required to control the residual of the regularized solution. Unfortunately those two conditions are not enough. The difficulty in the following proof as well as in the proof for the convergence rates is to control the $\mathcal{R}$-value of the regularized solution or its tail, respectively. We need some condition that allows to relate $\mathcal{R}$-based inequalities with $\mathcal{S}$-based inequalities.

For the following proof note that since by construction $P_n x^\dagger \to x^\dagger$ it follows that for each $0 < \delta \le \delta_0$ there is $n_0(\delta)$ such that $||AP_n x^\dagger - Ax^\dagger|| \le c\delta$ for all $n > n_0(\delta)$ where the constant $c > 1$ is independent of n and δ.

Lemma 3.5. *Let $\{y^{\delta_k}\}$ be a sequence of noisy data with $||y^{\delta_k} - y|| \le \delta_k \to 0$ for $k \to \infty$ and assume the requirements of Assumption* (3.3) *hold true. Let $n_0(\delta) = \inf_{n\in\mathbb{N}}\{||AP_n x^\dagger - Ax^\dagger|| \le c\delta\}$ for fixed $c \ge 0$. Let $\alpha_k = \alpha_k(\delta_k, y^{\delta_k})$ be a parameter choice fulfilling $\alpha_k f(n_0(\delta_k)) \to 0$ and $\frac{\delta_k^p}{\alpha} \le Cf(n_0(\delta_k))$ for all $k \to \infty$. Then the sequence of minimizers $x_{\alpha_k}^{\delta_k}$ converges to a solution of* (1) *and*

$$\mathcal{R}(x_{\alpha_k}^{\delta_k}) \to \infty \quad \text{as} \quad k \to \infty.$$

Proof. From the minimizing property we have

$$\frac{1}{r}||A(x_{\alpha_k}^{\delta_k}) - y^{\delta_k}||^r + \alpha_k \mathcal{R}(x_{\alpha_k}^{\delta_k}) \le \frac{1}{r}||A(P_n x^\dagger) - y^{\delta_k}||^r + \alpha_k \mathcal{R}(P_n x^\dagger)$$

for all $n \in \mathbb{N}$. Now pick any n with $||AP_n x^\dagger - Ax^\dagger|| \le c\delta_k$. Then

$$\begin{aligned}\alpha_k \mathcal{R}(x_{\alpha_k}^{\delta_k}) &\le \frac{1}{r}||A(P_n x^\dagger) - y^{\delta_k}||^r + \alpha_k \mathcal{R}(P_n x^\dagger)\\ &\le \frac{c}{r}\delta_k^r + \alpha_k f(n_0(\delta_k))\mathcal{S}(P_n x^\dagger)\\ &\le \frac{c}{r}\delta_k^r + \alpha_k f(n_0(\delta_k))\mathcal{S}(x^\dagger) \longrightarrow 0 \text{ as } k \to \infty\end{aligned}$$

Note that also

$$\mathcal{R}(x_{\alpha_k}^{\delta_k}) \le \frac{c\delta_k^r}{r\alpha_k} + f(n_0(\delta_k))\mathcal{S}(P_n x^\dagger) \le Cf(n_0(\delta_k)),$$

with some constant $C > 1$. This means that for sufficiently large δ_k, $\mathcal{R}(x_{\alpha_k}^{\delta_k})$ grows approximately as fast as $f(n_0(\delta_k))$. On the other hand, the minimizing property also implies

$$\frac{1}{r}||A(x_{\alpha_k}^{\delta_k}) - y^{\delta_k}||^r \le \frac{1}{r}||A(P_n x^\dagger) - y^{\delta_k}||^r + \alpha_k \left(\mathcal{R}(P_n x^\dagger) - \mathcal{R}(x_{\alpha_k}^{\delta_k})\right)$$

for all $n \in \mathbb{N}$. If $\mathcal{R}(P_n x^\dagger) \le \mathcal{R}(x_{\alpha_k}^{\delta_k})$ this immediately yields

$$\frac{1}{r}||A(x_{\alpha_k}^{\delta_k}) - y^{\delta_k}||^r \le \frac{1}{r}\delta_k^r. \longrightarrow 0 \text{ as } k \to \infty$$

If $\mathcal{R}(P_n x^\dagger) > \mathcal{R}(x_{\alpha_k}^{\delta_k})$, then

$$\begin{aligned}\frac{1}{r}||A(x_{\alpha_k}^{\delta_k}) - y^{\delta_k}||^r &\le \frac{1}{r}||A(P_n x^\dagger) - y^{\delta_k}||^r + \alpha_k f(n)\mathcal{S}(P_n x^\dagger)\\ &\le \frac{c}{r}\delta_k^r + \alpha_k f(n_0(\delta_k))S(x^\dagger) \longrightarrow 0 \text{ as } k \to \infty\end{aligned}$$

In any case, we have shown that

$$\mathcal{T}_{\alpha_k}^{\delta_k}(x_{\alpha_k}^{\delta_k}) \to 0 \text{ as } k \to 0$$

from which the convergence follows by standard arguments.

If from the relations

$$\alpha_k \mathcal{R}(x_{\alpha_k}^{\delta_k}) \le \frac{c}{r}\delta_k^r + \alpha_k \mathcal{R}(P_n x^\dagger) \tag{9}$$

and $\mathcal{R}(P_n x^\dagger) \le f(n_0(\delta))\mathcal{S}(x^\dagger)$ we could conclude

$$\alpha_k f(n_0(\delta_k))\mathcal{S}(x_{\alpha_k}^{\delta_k}) \le \frac{c}{r}\delta_k^r + \alpha_k f(n_0(\delta_k))\mathcal{S}(P_n x^\dagger) \tag{10}$$

it would follow

$$\mathcal{S}(x_{\alpha_k}^{\delta_k}) \le \frac{c\delta_k^p}{\alpha_k f(n_0(\delta_k))} + \mathcal{S}(P_n x^\dagger).$$

Since the term in the middle goes to zero as $k \to \infty$ and $x_n^\dagger \to x^\dagger$ as $n \to \infty$ this would mean the regularized solutions converge to the solution of (1) with minimal $\mathcal{S}$-value. □

We quickly want to relate this to the diagonal case for ℓ^1-regularization with an ℓ^2-solution as presented in Section 4. Let P_n denote the cut-off projectors . Then $||P_n x||_{\ell^1} \leq \sqrt{n}||P_n x||_{\ell^2}$ for all $x \in \ell^2$. Later we calculate that there is $n_{inf}(\delta) = c\delta^{-\frac{2}{2\beta+2\eta+1}}$ such that $||Ax^\dagger_{n_{inf}} - Ax^\dagger|| \sim \delta$. Therefore,

$$||P_{n_{inf}}x^\dagger||_{\ell^1} \leq \sqrt{n_{inf}(\delta)}||P_{n_{inf}}x^\dagger||_{\ell^2} = c\delta^{-\frac{1}{2\beta+2\eta+1}}||P_{n_{inf}}x^\dagger||_{\ell^2}.$$

The resulting correction of the regularization parameter with the factor $\delta^{\frac{1}{2\beta+2\eta+1}}$ i.e.,

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

coincides precisely with the adjustment needed for the regularization parameter as observed in the experiments of Section (6).

4 Convergence rates for ℓ^1-regularization with a diagonal operator

The derivation of convergence rates is typically much more difficult than the convergence statements of the previous section. We will therefore step away from the general penalty term as introduced in Section 1 and focus on ℓ^1-regularization. This strategy defines the approximate solution to (1) via

$$x_\alpha^\delta := \operatorname*{argmin}_{x\in D(A)} \frac{1}{2}||Ax - y^\delta||^2 + \alpha||x||_{\ell^1}. \tag{11}$$

For the sake of simplicity we also set the exponent of the residual to the standard value $r = 2$. ℓ^1-regularization became a popular and powerful tool in the last decade, sparked by the seminal paper [2]. Since then, many authors have contributed to its theory and application. Here we only mention the papers [11, 10, 9, 3].The vast majority of papers connected to ℓ^1-regularization assumes that $x^\dagger \in \ell^0$, i.e. it has only finitely many non-zero components. However, in [3] for the first time the situation that the exact solution $x^\dagger$ is not sparse, but only $x^\dagger \in \ell^1$ was explored. In some sense this paper is a continuation of this trend as we now assume that $x^\dagger$ is not even in ℓ^1, but $x^\dagger \in \ell^2\backslash\ell^1$. Due to this we will employ the ℓ^2-norm to measure the speed of convergence, i.e. we seek a positive function $\varphi(\delta)$ such that

$$||x_\alpha^\delta - x^\dagger||_{\ell^2} \leq C\varphi(\delta), \qquad C > 0.$$

We will see that even in this case convergence rates can be derived. Often it is only possible to state that a function describing the convergence rate exists and that it indeed bounds the regularization error from above. In order to be able to explicitly calculate this function we will restrict ourselves to a particular scenario. Namely, we consider the case of a compact operator between Hilbert spaces. This allows us to use its singular system for the calculus. Assume $\widetilde{A} : \widetilde{X} \to Y$ to be a compact linear operator between infinite dimensional separable Hilbert spaces $\widetilde{X}$ and Y. Then $\widetilde{A}$ has a singular system $\{\sigma_i, u_i, v_i\}_{i\in\mathbb{N}}$ with nonnegative, decreasingly ordered singular values σ_i tending to zero and $\{u_i\}$, $\{v_i\}$ are complete orthonormal systems in $\widetilde{X}$ and $\overline{\mathrm{ran}(A)}$, respectively. We have $\widetilde{A}u_i = \sigma_i v_i$ and $\widetilde{A}^* v_i = \sigma_i u_i$. Since we consider Hilbert-spaces, we can identify the dual spaces with the original ones. Hence $\widetilde{A}^* : Y \to \widetilde{X}$.

Using the $\{u_i\}$ as Schauder basis in $\widetilde{X}$ we have $\tilde{x} = \sum_{i\in\mathbb{N}} x_i u_i$ where $x_i = \langle \tilde{x}, u_i\rangle$ with the scalar product in $\widetilde{X}$. Interpreting the coefficients as an infinite series we introduce the synthesis operator $L : \ell^2(\mathbb{N}) \to X$, $x = (x_i)_{i=}^{\infty} \mapsto \sum_{i\in\mathbb{N}} x_i u_i$ to obtain via the composition $A = \tilde{A} \circ L$ a compact linear operator $A : X = \ell^2(\mathbb{N}) \to Y$. A still has diagonal structure as $Ae_i = \widetilde{A}u_i$ for all $i \in \mathbb{N}$, characterized by $Ax = \sum_{i\in\mathbb{N}} x_i \sigma_i v_i$. We use this special structure for the characterization of the smoothness properties of the operator and the exact solution. Namely, we will assume that $\sigma_i = i^{-\beta}$ and $|\langle x, u_i\rangle| = i^{-\eta}$ for positive values β and η.

In the remainder of the section we will proof the following Theorem up to a final gap that has not been closed yet.

Theorem 4.1. *Let A be as above with singular values $\sigma_i = i^{-\beta}$, $\beta > 0$ and let y^{δ_k} be a sequence fulfilling $||y - y^{\delta_k}|| \le \delta_k$ for a sequence $\delta_k \to 0$. Let $x^\dagger$ be such that $|\langle x, u_i\rangle| = i^{-\eta}$ for $\eta > \frac{1}{2}$. Then, with the a priori parameter choice*

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

the ℓ^1-regularized solution of (11) *fulfill*

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

even if $\frac{1}{2} < \eta \le 1$, i.e., $x^\dagger \in \ell^2 \backslash \ell^1$.

We chose the ℓ^1-ℓ^2-situation for several reasons. As mentioned above, ℓ^1-regularization is a very popular method. In practice however it is not always clear whether or not the exact solution is sparse, or what its precise smoothness characteristics are. In contrast to Natterer's Hilbert space theory, ℓ^1 is a Banach space. Even more, ℓ^1 is not even reflexive, in contrast to the ℓ^p-spaces for $1 < p < \infty$. Thus ℓ^1 is far away from the Hilbert space structure of ℓ^2 and we may hope to find results that can be generalized to a large family of Banach spaces.

We start with the derivation of a governing inequality. We have with the linear projector

$$P_n : \ell^2 \to \ell^2, \quad x_n := P_n(x) = \{x_i\}_{i=1,\ldots,n},$$

the relation

$$\begin{aligned} ||x - x^\dagger||_{\ell^2} &\le ||x - x_n^\dagger||_{\ell^2} + ||x_n^\dagger - x^\dagger||_{\ell^2} \\ &\le ||(I - P_n)x||_{\ell^2} + ||P_n(x - x^\dagger)||_{\ell^2} + ||(I - P_n)x^\dagger||_{\ell^2} \qquad (12) \end{aligned}$$

Until further notice we keep the parameter $n \in \mathbb{N}$ arbitrary but fixed and start estimating the last term which describes the decay of the tail of the solution and thus its smoothness. One easily sees that in in our special situation with $\langle x^\dagger, u_i \rangle = \sigma_i^{-\eta}$ it is

$$\begin{aligned} ||x_n^\dagger - x^\dagger||_{\ell^2} &= \sqrt{\sum_{i=n+1}^{\infty} |\langle x^\dagger, u_i \rangle|^2} = \sqrt{\sum_{i=n+1}^{\infty} |i^{-\eta}|^2} \\ &= \sqrt{\sum_{i=n+1}^{\infty} i^{-2\eta}} \leq \sqrt{\frac{1}{2\eta - 1}(n+1)^{1-2\eta}} \\ &\leq (2\eta - 1)^{-\frac{1}{2}} n^{\frac{1}{2}-\eta}. \end{aligned}$$

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

$$\omega(M, \delta) := \sup\{||x|| : x \in M, ||Ax|| \leq \delta\}. \tag{13}$$

This quantity is essentially related to minimal errors of any regularization method for noisy data. Since $P_n(x - x^\dagger) \in \operatorname{span}\{u_1, \ldots, u_n\} =: X_n$, we can use tools from approximation theory to estimate its norm. In [13], Proposition 3.9, it has been shown that

$$\omega(X_n, \delta) = \frac{\delta}{\sigma_n}$$

where $\frac{1}{j(A, X_n)}$ is the inverse of the *modulus of injectivity* $j(A, X_n)$, defined as

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

For diagonal operators it is $j(A, X_n) = \sigma_n$ and thus, with $\sigma_n = n^{-\beta}$, we have

$$\omega(X_n, \delta) = n^\beta \delta \tag{14}$$

and therefore

$$||P_n(x - x^\dagger)||_{\ell^2} \leq \omega(X_n, ||AP_n(x - x_n^\dagger)||) = n^\beta(||AP_n(x - x_n^\dagger)|| + \delta),$$

see [13], Lemma 2.2. We need to show later that $||AP_n(x - x_n^\dagger)|| + \delta \leq c\delta$ for some constant $c > 1$. Therefore we neglect the summand δ as in the case $||AP_n(x - x_n^\dagger)|| < \delta$ this criterion is fulfilled anyway. One also easily sees that for diagonal operators it is

$$||AP_n(x - x_n^\dagger)|| \leq ||Ax - x_n^\dagger||.$$

Summing up the results so far, (12) now reads

$$||x - x^\dagger||_{\ell^2} \leq ||(I - P_n)x||_{\ell^2} + n^\beta ||Ax - Ax_n^\dagger||_{\ell^2} + cn^{\frac{1}{2}-\eta} \tag{15}$$

and holds for all $n \in \mathbb{N}$. We are of course only interested in this relation when x is the regularized solution, $x = x_\alpha^\delta$. In the best case scenario the term $||(I - P_n)x||_{\ell^2}$ vanishes and only the two rightmost terms remain. The best possible convergence rate obtainable with our approach is therefore determined by those two expressions. Define

$$\varphi_n(t) := n^\beta t + cn^{\frac{1}{2}-\eta} \tag{16}$$

and

$$\varphi(t) = \inf_{n\in\mathbb{N}} \varphi_n(t) \tag{17}$$

The index for which the infimum is attained is denoted by n_{inf}. Since the infimum is taken over a countable set and $\varphi_n(t) \to \infty$ as $n \to \infty$ such an n_{inf} exists. It is simple calculus to show

$$\varphi(t) = \inf_{n\in\mathbb{N}} \varphi_n(t) = c_{\beta,\eta} t^{\frac{2\eta-1}{2\eta+2\beta-1}} \tag{18}$$

and

$$n_{inf}(t) = c_\eta t^{-\frac{2}{2\eta+2\beta-1}} \tag{19}$$

with $c_\eta > 0$. Before we continue, we shall make a quick credibility check for the convergence rate. Using the singular system of $\widetilde{A}$ we can also calculate the residual approximation $||\widetilde{A}(x_n^\dagger - x^\dagger)||_Y^2$. It is

$$\begin{aligned}
||\widetilde{A}(x_n^\dagger - x^\dagger)||_Y^2 &= \left|\left| \sum_{i=n+1}^{\infty} \sigma_i x_i^\dagger u_i \right|\right|_Y^2 = \sum_{i=n+1}^{\infty} |\sigma_i x_i^\dagger|^2 \\
&= \sum_{i=n+1}^{\infty} |i^{-\beta} i^{-\eta}|^2 = \sum_{i=n+1}^{\infty} i^{-2(\eta+\beta)} \\
&\leq \left| \frac{1}{1-2\eta-2\beta} \right| n^{1-2(\eta+\beta)}
\end{aligned}$$

We insert $n_{inf}(t)$ in this expression and obtain

$$||\widetilde{A}(x_{n_{inf}(t)}^\dagger - x^\dagger)||_Y^2 = c_{\eta,\beta} (t^2)^{\frac{2\eta+2\beta-1}{2\eta+2\beta-1}} = c_{\eta,\beta} t^2, \tag{20}$$

i.e., $||Ax_{n_{inf}(\delta)}^\dagger - Ax^\dagger|| \leq c\delta$. Note that

$$n^\beta ||\widetilde{A}(x_n^\dagger - x^\dagger)||_Y = cn^{\frac{1}{2}-\eta} \to 0 \quad \text{for } \eta > \frac{1}{2},$$

i.e., the the image of A under the approximation of $x^\dagger$ via $x_n^\dagger$ is compatible with the modulus of continuity for all $x \in \ell^2$. In section 6 we observe that the parameter choice (21) indeed yields the rate (18) in numerical experiments.

In order to theoretically verify the convergence rate it remains to show $||Ax_\alpha^\delta - Ax^\dagger|| \leq c\delta^2$ and to estimate $||(I - P_n)x_\alpha^\delta||_{\ell^2}$. For both terms the choice of the regularization parameter is crucial. We will investigate here an a priori parameter choice which takes the form

$$\alpha \sim \frac{\delta^2}{\delta^{\frac{-1}{2\eta+2\beta-1}}\varphi(\delta)}. \tag{21}$$

While we are interested in the discrepancy principle as an a posteriori parameter choice, we while not discuss it for now as we will run into the same problem as with the a priori choice. It is apparent that one this has been overcome the discrepancy principle will lead to the same convergence rate.

Lemma 4.1. *It is with the a-priori parameter choice* (21) $||Ax_\alpha^\delta - y^\delta||^2 \leq C\delta^2$.

Proof. Fix $n = n_{inf}(\delta)$. From the Tikhonov functional we have

$$||Ax_\alpha^\delta - y^\delta||^2 + \alpha||x_\alpha^\delta||_{\ell^1} \leq ||Ax_n^\dagger - y^\delta||^2 + \alpha||x_n^\dagger||_{\ell^1}$$

which with

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

and

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

yields

$$||Ax_\alpha^\delta - y^\delta||^2 + \alpha||(I-P_n)x_\alpha^\delta||_{\ell^1} \leq ||Ax_n^\dagger - y^\delta||^2 + \alpha||P_n(x^\dagger - x_\alpha^\delta)||_{\ell^1}. \tag{22}$$

Therefore, using $||P_n \cdot||_{\ell^1} \leq \sqrt{n}||P_n \cdot||_{\ell^2}$ and adding $\alpha\sqrt{n}||x_n^\dagger - x^\dagger||_{\ell^2}$ on the right hand side one obtains by neglecting the second term on the left hand side

$$||Ax_\alpha^\delta - y^\delta||^2 \leq ||Ax_n^\dagger - y^\delta||^2 + \alpha\sqrt{n}(||P_n(x^\dagger - x_\alpha^\delta)||_{\ell^2} + ||x_n^\dagger - x^\dagger||_{\ell^2}) \tag{23}$$

Observe that the term in brackets is $\varphi(||AP_n(x_\alpha^\delta - x_n^\dagger)||)$ and that

$$\begin{aligned} \varphi(||AP_n(x_\alpha^\delta - x_n^\dagger)||) &\leq \varphi(||Ax_\alpha^\delta - y^\delta|| + ||Ax_n^\dagger - Ax^\dagger||) && (24)\\ &\leq \varphi(2\max\{||Ax_\alpha^\delta - y^\delta||, C\delta\}). && (25) \end{aligned}$$

Therefore we have from (23), with $\sqrt{n} = \sqrt{n_{inf}(\delta)} = C\delta^{\frac{-1}{2\eta+2\beta-1}}$ and $||Ax_n^\dagger - y^\delta||^2 \leq 4\delta^2$

$$||Ax_\alpha^\delta - y^\delta||^2 \leq 4\delta^2 + C\alpha\delta^{\frac{-1}{2\eta+2\beta-1}}\varphi(2\max\{||Ax_\alpha^\delta - y^\delta||, C\delta\}). \tag{26}$$

Using the parameter choice (21) one continues analogously to Corollary 1 in [5]. □

The final step is to show that $||(I-P_n)x_\alpha^\delta|| \leq c\varphi(\delta)$, and this is precisely the open problem. Starting from (22) we have

$$\alpha||(I-P_n)x_\alpha^\delta||_{\ell^1} \leq ||Ax_n^\dagger - y^\delta||^2 + \alpha||P_n(x^\dagger - x_\alpha^\delta)||_{\ell^1}. \tag{27}$$

Alternatively, (23) yields

$$\alpha||(I-P_n)x_\alpha^\delta||_{\ell^1} \le ||Ax_n^\dagger - y^\delta||^2 + \alpha\sqrt{n}\varphi(\delta).$$

Inserting the parameter choice (21) yields

$$||(I-P_n)x_\alpha^\delta||_{\ell^1} \le C\sqrt{n}\varphi(\delta).$$

Without the term $\sqrt{n}$ on the right-hand side this would proof the conjecture for the convergence rate since trivially

$$||(I-P_n)x_\alpha^\delta||_{\ell^2} \le ||(I-P_n)x_\alpha^\delta||_{\ell^1}.$$

However, we would need a relation of the type

$$\sqrt{n}||(I-P_n)x_\alpha^\delta||_{\ell^2} \le ||(I-P_n)x_\alpha^\delta||_{\ell^1}. \tag{28}$$

While for general elements in ℓ^1 this certainly won't hold, we have a particular structure of in $(I-P_n)x_\alpha^\delta$. Namely, by construction its first n coefficients are zero and there is only a finite number of nonzero-elements since $x_\alpha^\delta \in \ell^0$ (see [10] for the last statement). Alternatively, we would immediately obtain the proposed convergence rates if simply

$$||(I-P_n)x_\alpha^\delta||_{\ell^1} = 0$$

would hold, implying $||(I-P_n)x_\alpha^\delta||_{\ell^1} = 0$.

The problem here is precisely the same as in the proof of convergence to a minimum-norm solution, compare (27) and (9) as well as (28) and (10).

Remark 4.1. *Although we restricted ourselves to the case of a diagonal operator with special solution, a generalization to general operators and solutions is possible. Then, however, we might not be able to explicitly state the obtained convergence rate. We could merely show that there is a function φ governing the regularization error which is constructed as in* (16)*, with different factor working on t and different term describing the solution smoothness. Lemma 4.1 would still apply and we would expect that the term $||(I-P_n)x_\alpha^\delta||$ can be handled as well (assuming we would know how to do that in the diagonal case).*

5 Comparison with ℓ^1 and ℓ^2-regularization

In order to get a feeling for the convergence properties we compare the result from the previous section with classical ℓ^1 and classical ℓ^2 regularization. Let us start with ℓ^1-regularization, i.e. the approximate solution to (1) is obtained via (11) but under the assumption that $||x^\dagger||_{\ell^1} < \infty$. Since we are interested in non-sparse solutions, we refer to [3, 4] for this situation. For the diagonal case as in Section 4 they showed a convergence rate

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

see [3, Example 5.3]. Recently it was shown in [14] that actually

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

can be achieved using the parameter choice

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

Now let us move to ℓ^2-regularization. This corresponds to the classic Tikhonov regularization, i.e. the approximate solution to (1) is given by

$$x_\alpha^\delta := \operatorname*{argmin}_{x\in D(A)} \frac{1}{2}||Ax - y^\delta||^2 + \alpha||x||_{\ell^2}. \tag{29}$$

It is then well known, see e.g. [7, 8] that under the assumption

$$x^\dagger \in \operatorname{ran}((A^*A)^\nu) \tag{30}$$

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

$$||x_\alpha^\delta - x^\dagger||_{\ell^1} \leq c\delta^{\frac{\nu}{\nu+1}} \tag{31}$$

can be shown under the a-priori parameter choice

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

In the diagonal setting the source condition (30) can easily be related to the parameters η and β. Namely, (30) holds for all ν with $\frac{2\eta-1}{2\beta} < \nu$. 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

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

and the parameter choice is

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

We summarize the convergence rates and parameter choices in Table 1. One sees that the $\ell^1 - \ell^2$-regularization inherits the parameter choice from the ℓ^1-regularization and the convergence rate from the ℓ^2-regularization. With respect to the ideas of Section 4 this appears to be reasonable. The parameter choice influences the residual $||Ax_\alpha^\delta - y^\delta||$ and the penalty value $||x_\alpha^\delta||_{\ell^1}$. Since it it most important to keep the residual on a level of about δ the ℓ^1-parameter choice is used. Namely, in [14] it has been observed that for ℓ^1-regularization the discrepancy principle and a priori parameter choice always coincide. It is

type	ℓ^1	$\ell^1 - \ell^2$	ℓ^2
rate	$\varphi_{\ell^1}(\delta) = \delta^{\frac{\eta-1}{\eta+\beta-\frac{1}{2}}}$	$\varphi_{\ell^2}(\delta) = c\delta^{\frac{2\eta-1}{2\eta+2\beta-1}}$	$\varphi_{\ell^2}(\delta) = c\delta^{\frac{2\eta-1}{2\eta+2\beta-1}}$
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	$\alpha = \frac{\delta^2}{\varphi_{\ell^1}}$	$\alpha = \frac{\delta^2}{\varphi_{\ell^2}(\delta)\delta^{\frac{-1}{2\eta+2\beta-1}}}$	$\alpha = \frac{\delta^2}{\varphi_{\ell^2}(\delta)}$

Table 1: Comparison of ℓ^1, ℓ^2 and $\ell^1 - \ell^2$ regularization. The parameter choice depends on the penalty functional whereas the convergence rate depends on the norm the regularization error is measured in.

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 (31). In short this means that ℓ^1-regularization has qualification ∞ in comparison to the classical ℓ^2-based Tikhonov regularization which possesses the qualification ν.

The above observations give rise to the assumption that a closed convergence theory for ℓ^q-regularization with convergence rate measured in ℓ^p is possible. This would lead to a theory alike the one in Hilbert scales, see Section 2. We could already find an a priori parameter choice rule as well as a convergence rate for this situation. Numerically observed convergence rates coincide nicely with our conjecture. Since the sketch of the proof is similar to the one in Section 4, it is clear that we have to fix that one first before going to a more general setting.

6 Numerical examples

In this section we consider an operator specifically tailored to the setting of section (4). Our base is the Voltera operator

$$[\tilde{A}x](s) = \int_0^s x(t)\,dt. \tag{32}$$

We discretized $\tilde{A}$ with the rectangular rule at $N = 400$ points. In order to ensure our desired properties $\sigma_n \sim n^{-\beta}$, $\langle x^\dagger, v_n\rangle| \sim n^{-\eta}$, we computed the SVD of the resulting matrix and manually set its singular values σ_n to precisely $n^{-\beta}$. This means that the actual operator A in (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 constructed our solution such that $\langle x^\dagger, v_n\rangle = n^{-\eta}$ holds for various values of $\eta > 0$. We added 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 were computed via

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

where the ℓ^1-norm was taken of the coefficients with respect to the basis originating from the SVD. The regularization parameter was chosen a priori according to (21). We computed the reconstruction error in the ℓ_2 norm as well as the residuals. For larger values of η we could observe the convergence rate directly. For smaller values of η, we had to compensate for the error introduced by the discretization level. Namely, since we used a discretization level $N = 400$, numerically we actually measured

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

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

$$||P_{400}(x_\alpha^\delta - x^\dagger) + (I - P_{400})x^\dagger||_{\ell^2}. \tag{33}$$

The second term can be calculated analytically and is supposed to correct for the fact that we cannot measure the regularization error for larger coefficients, i.e., we add the tail of $x^\dagger$ that can not be observed, keeping in mind that x_α^δ has only finitely many non-zero components.

We show selected plots of the convergence rates in Figure 6 for $\beta = 1$ and Figure 6 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. For the residuals we show the regression for $||Ax_\alpha^\delta - y^\delta||_{\ell^2} = c\delta^d$, d being 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 Table 6 and 6 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 (33). 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.

References

[1] F. Natterer *Error bounds for Tikhonov regularization in Hilbertscales* Applicable Anal.**18**, p.29–37, 1984

[2] I. Daubechies, M. Defrise and C. De Mol *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint* Comm. Pure Appl. Math. **57** 11 1413–57, 2004

[3] M. Burger, J. Flemming and B. Hofmann *Convergence rates in ℓ^1-regularization if the sparsity assumption fails*, Inverse Problems **29** 025013, 2013

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 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 in the form $||Ax_\alpha^\delta|| = c\delta^d$.

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

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 in the form $||Ax_\alpha^\delta|| = c\delta^d$.

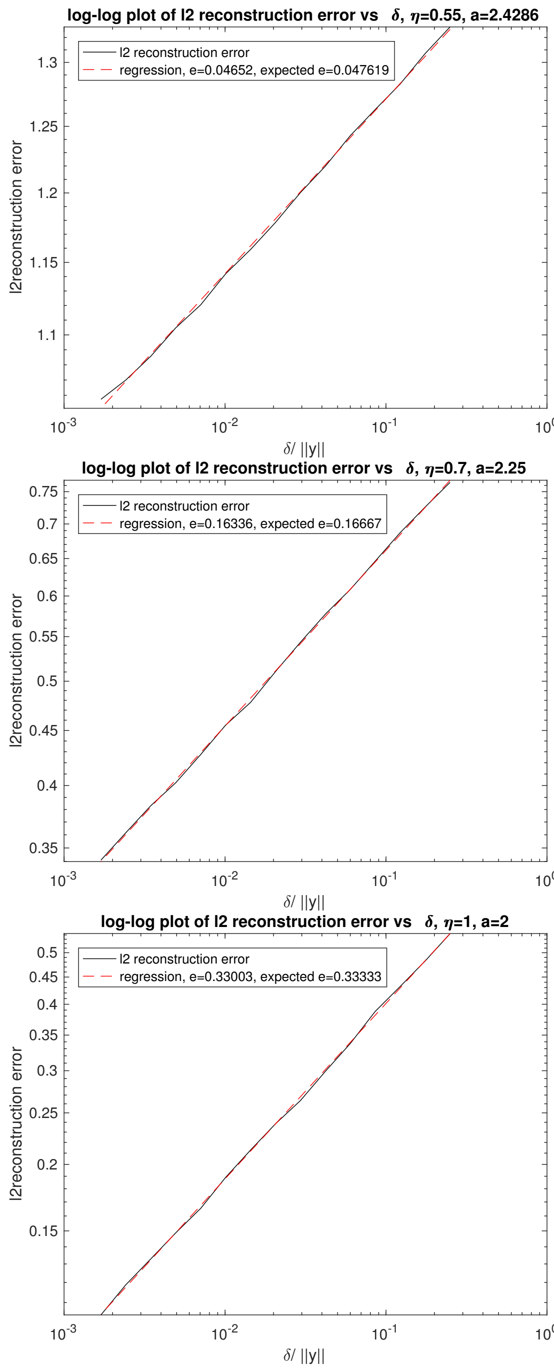

Figure 1: Convergence rates for $\beta = 1$ and $\eta = 0.55, 0.7, 1$. From the measured reconstruction error (solid line) we calculated the regression for the assumption $||x_\alpha^\delta - x^\dagger||_{\ell^2} = c\delta^e$, shown in the broken line.

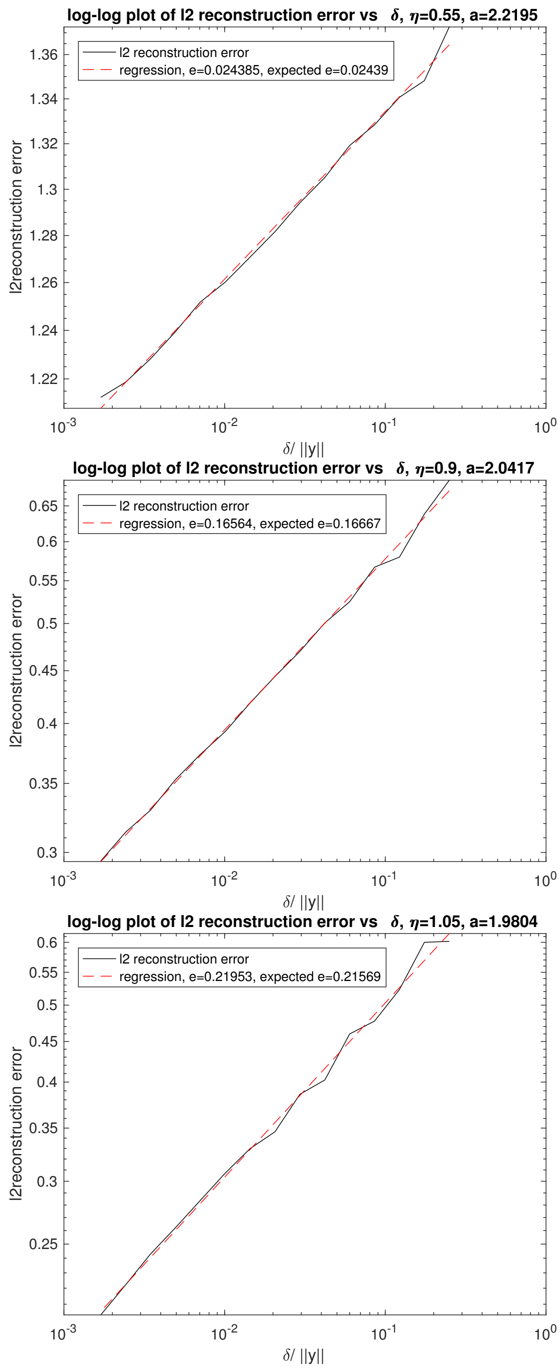

Figure 2: Convergence rates for $\beta = 2$ and $\eta = 0.55, 0.9, 1.05$. From the measured reconstruction error (solid line) we calculated the regression for the assumption $||x_\alpha^\delta - x^\dagger||_{\ell^2} = c\delta^e$, shown in the broken line.

[4] J. Flemming, B. Hofmann and I. Veselic *A unified approach to convergence rates for ℓ^1-regularization and lacking sparsity* Journal of Inverse and Ill-posed Problems **24**, 2016

[5] B. Hofmann and P. Mathè *Parameter choice in Banach space regularization under variational inequalities* Inverse Problems **28** 104006, 2012

[6] T. Schuster, B. Kaltenbacher, B. Hofmann and K. Kazimierski *Regularization Methods in Banach spaces* De Gruyter, Berlin, 2012

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

[8] A. K. Louis *Inverse und schlecht gestellt Probleme* Teubner, Stuttgart, 1989

[9] Ramlau R 2008 Regularization properties of Tikhonov regularization with sparsity constraints *Electron. Trans. Numer. Anal.* **30** 54–74

[10] Lorenz D A 2008 Convergence rates and source conditions for Tikhonov regularization with sparsity constraints *J. Inverse ill-posed Probl* **16** 5 463–478

[11] Grasmair M, Haltmeier M and Scherzer O 2008 Sparse regularization with lq penalty term *Inverse Problems* **24** 5 055020

[12] 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), John Wiley, New Jersey 2015

[13] B. Hofmann, P.Math and M.Schieck *Modulus of continuity for conditionally stable ill-posed problems in Hilbert space* J. Inv. Ill-posed Problems **16** p 567-585, 2008

[14] D. Gerth, J. Flemming *Convergence rates for ℓ^1-regularization without source-type conditions* preprint, 2016