Convergence rates for ℓ^1-regularization without the help of a variational inequality

Daniel Gerth *

January 12, 2017

Key words: sparsity promoting regularization, Tikhonov regularization, source condition, variational source condition, convergence rates

MSC: 65J20, 47A52

Abstract

We show that convergence rates for ℓ^1-regularization can be obtained in an elementary way without requiring a classical source condition and without the help of a variational inequality. For the specific case of a diagonal operator we improve the convergence rate compared to the literature and conduct numerical experiments which verify the predicted rate. The idea of the proof is rather generic and might be used in other settings as well. By construction the obtained convergence rates are optimal in a certain sense.

1 Introduction and main theorem

We consider a bounded, injective, linear operator $A : \ell^1 \to Y$ mapping absolutely summable real sequences into some real Banach space Y. We have to deal with the ill-posed equation

$$Ax = y, \quad x \in \ell^1, \tag{1.1}$$

where instead of the true data $y \in Y$ only perturbed data y^δ satisfying $||y - y^\delta||_Y \leq \delta$, $\delta > 0$, is available. In order to find an approximation

*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 and HO1454/10-1.

to the unique exact solution $x^\dagger$ for the exact data $y \in Y$ we employ ℓ^1-regularization to determine an approximate solution to (1.1). Hence we solve

$$T_\alpha^\delta(x) := \frac{1}{p}\|Ax - y^\delta\|_Y^p + \alpha\|x\|_{\ell^1} \to \min_{x \in \ell^1} \tag{1.2}$$

with some $1 < p < \infty$. The regularization parameter $\alpha > 0$ is used to balance between the residual of the approximate solution and its value of the penalty term $||\cdot||_{\ell^1}$. The minimizer of (1.2) is denoted by x_α^δ. In order to guarantee the existence of the minimizers we make the following assumption.

Assumption 1. Let the linear operator $A : \ell^1 \to Y$ be sequentially weak*-to-weak continuous.

In [18] it has been shown that the weak*-to-weak continuity of A is equivalent to both the condition $\mathcal{R}(A^*) \subseteq c_0$ and the condition $Ae_i \rightharpoonup 0$. Here and in the remainder of the paper, $A^* : Y^* \to (\ell^1)^* = \ell^\infty$ denotes the adjoint of A, $\mathcal{R}$ the range of an operator and e_i, $i \in \mathbb{N}$ are the canonical basis in ℓ^1, i.e., the k-th component of e_i equals 1 for $i = k$ and 0 otherwise. To the best of the authors knowledge Assumption 1 or its equivalents correspond to the weakest condition for the existence of minimizers of (1.2) currently available in the literature, cf. [18, 12, 4]. Assumption 1 is for example fulfilled if A has a bounded extension to some ℓ^q-space with $q > 1$. This is often the case in practical applications, in particular the case that A is factored through a Hilbert-space occurs frequently. As a counterexample, the identity $Id : \ell^1 \to \ell^1$ is not weak*-to-weak continuous as it is simple to show that $\mathcal{R}(Id^*) = \ell^\infty$. Another direct consequence of Assumption 1 is the following property as was shown in [12], where

$$P_n : \ell^\infty \to \ell^\infty, \quad P_n x := (x_1, \ldots, x_n, 0, \ldots) \tag{1.3}$$

denotes the cut-off after the n-th entry.

Property 2. There exist a real sequence $(\gamma_n)_{n \in \mathbb{N}}$ and a constant $\mu \in [0, 1)$ such that for each $n \in \mathbb{N}$ and each $\xi \in \ell^\infty$ with

$$\xi_k \begin{cases} \in \{-1, 0, 1\}, & \text{if } k \le n, \\ = 0, & \text{if } k > n \end{cases} \tag{1.4}$$

there exists some $\eta = \eta(n, \xi, \mu)$ in Y^* such that

(i) $P_n A^* \eta = \xi$,

(ii) $||(I - P_n)A^*\eta||_{\ell^\infty} \le \mu$,

(iii) $\|\eta\|_{Y^*} \leq \gamma_n$ for all ξ as in (1.4).

We cite [12, Proposition 12] to clarify the relation between Assumption 1 and Property 2.

Proposition 3. Let Assumption 1 hold. Then the following statements are equivalent:

i) For every $0 < \mu < 1$ and every ξ as in Property 2, there is $\eta \in Y^*$ such that Property 2 holds.

ii) $e_i \in \overline{\mathcal{R}(A^*)}$ $\forall i \in \mathbb{N}$ where the closure is taken w.r.t. the norm in ℓ^∞

iii) $\overline{\mathcal{R}(A^*)} = c_0$, i.e., the Banach space of sequences converging to zero

iv) A is injective

Thus letting A injective and weak*-to-weak continuous allows to work with Property 2. In the context of the development of regularization theory for ℓ^1-regularization, item ii) is very interesting. For a survey and on ℓ^1-regularization theory we refer to [19]. We will give a brief summary here. The seminal paper [1] sparked the investigation of sparsity-promoting inversion methods, with ℓ^1-regularization being one of the most prominent examples. In the context of inverse and ill-posed problems the question of convergence rates is of highest interest, i.e., one is interested in estimates of the form

$$||x_\alpha^\dagger - x^\dagger|| \leq C\varphi(\delta) \tag{1.5}$$

where C is a positive constant and φ an index function (a continuous, concave and monotonically increasing function with $\varphi(0) = 0$). First results where already given in [1]. Later the focus shifted to sparse solutions, where $x^\dagger$ as solution of the unperturbed equation (1.1) has only finitely many nonzero elements. Convergence rates in this case can be found in, e.g., [20, 9, 8] using different kinds of smoothness properties of A and $x^\dagger$ to derive the rate. Recently the case that $x^\dagger$ is an infinite sequence in ℓ^1 has gathered attention, starting with the paper [2]. There, a smoothness condition on the canonical basis with respect to the operator is crucial. Namely, the authors assume that for each $i \in \mathbb{N}$ there is $f_i \in Y^*$ such that $A^* f_i = e_i$, i.e.,

$$e_i \in \mathcal{R}(A^*) \quad \forall i \in \mathbb{N}. \tag{1.6}$$

Such a condition already appeared in [20] and can be traced back to [13]. While such a condition holds for various types of inverse problems, see [14], it does not hold in general. A counterexample was presented in [15]. This lead to relaxed assumptions similar to (1.6) in [15] and [16]. The assumption made in [16] in principle coincides with Property 2. It is an important contribution of [12] to show such an assumption (Property 2 in this paper) is already a consequence of Assumption 1. In this sense, (1.6) is generalized to item ii) in Proposition 3. Please note that (1.6) implies Property 2 with $\mu = 0$.

All proofs for convergence rates in the literature rely on some condition relating the smoothness of the solution and the smoothing properties of the operator. Classically such a relation is expressed via the assumption $x^\dagger \in \mathcal{R}((A^*A)^\nu)$ for some $\nu > 0$, see for example [5, 6]. In recent years, variational inequalities (sometimes also called variational source conditions) have been used as linking condition. For ℓ^1-regularization it was shown in [2] that the inequality

$$||x - x^\dagger||_{\ell^1} \leq ||x||_{\ell^1} - ||x^\dagger||_{\ell^1} + \varphi(||Ax - Ax^\dagger||_Y) \tag{1.7}$$

holds for all $x \in \ell^1$. There, φ is given via

$$\varphi(t) := 2 \inf_{n \in \mathbb{N}} \left(\gamma_n t + ||(I - P_n)x^\dagger||_{\ell^1} \right) \tag{1.8}$$

with $\gamma_n = \sum_{i=1}^n ||f_i||_{Y^*}$ where due to assuming (1.6) the $f_i \in Y^*$ are such that $A^* f_i = e_i$, $i \in \mathbb{N}$. From (1.7) and (1.8), a convergence rate of the form (1.5) with the same φ as in (1.8) follows by standard arguments. These can be found, e.g., in [3].

The aim of this paper is to show that such a link condition as for example the variational inequality (1.7) is not necessary to obtain convergence rates. In Section 2 we provide a proof of the following Theorem 4 leading to essentially the same function φ as in (1.8) in an elementary way. In Section 3 we investigate a particular operator for which we can calculate a convergence rate as in (1.8) explicitly and show that the theoretical rate is achieved in numerical experiments.

Theorem 4. *Let $A : \ell^1 \to Y$ be a bounded and injective linear operator that is additionally weak*-to-weak continuous, i.e., fulfills Assumption 1. Then there is an index function $\varphi : [0, \infty) \to [0, \infty)$ of the form*

$$\varphi(t) := \inf_{n \in \mathbb{N}} \left((1 + \mu) \sum_{k=n+1}^{\infty} |x_k^\dagger| + \gamma_n t \right), \tag{1.9}$$

with μ and γ_n from Property 2 such that

$$||x_\alpha^\dagger - x^\dagger|| \leq C\varphi(\delta)$$

holds with a constant $C > 0$ whenever the regularization parameter is chosen a priori such that $c_1 \frac{\delta^p}{\varphi(\delta)} \le \alpha(\delta) \le c_2 \frac{\delta^p}{\varphi(\delta)}$ for some $0 < c_1 \le c_2 < \infty$ or a posteriori via the two-sided discrepancy principle, i.e., x_α^δ satisfies

$$\tau_1 \delta \le ||Ax_\alpha^\delta - y^\delta||_Y \le \tau_2 \delta$$

for some $1 < \tau_1 \le \tau_2$.

In the proof we will explicitly work out the constants C. We avoided noting them down here due to their complicated structure. Without going into further details we would like to remark that the case of a sparse exact solution $x^\dagger$, i.e. $\text{supp}(x^\dagger) \subseteq \{1, \ldots, n_0\}$, $n_0 \in \mathbb{N}$, is covered in our assumptions. We then obtain

$$\varphi(\delta) = \gamma_{n_0+1} \delta$$

which is the known linear rate.

2 Proof of the main theorem

First observe that for arbitrary $n \in \mathbb{N}$ we can split

$$||x_\alpha^\delta - x^\dagger||_{\ell^1} = ||(I - P_n)(x_\alpha^\delta - x^\dagger)||_{\ell^1} + ||P_n(x_\alpha^\delta - x^\dagger)||_{\ell^1} \tag{2.1}$$

with the projectors P_n from (1.3) and using the triangle inequality we get

$$||x_\alpha^\delta - x^\dagger||_{\ell^1} \le ||(I - P_n)x_\alpha^\delta||_{\ell^1} + ||P_n(x_\alpha^\delta - x^\dagger)||_{\ell^1} + ||(I - P_n)x^\dagger||_{\ell^1}. \tag{2.2}$$

We start with the term in the middle of the right-hand side. It is for some $\xi \in \ell^\infty$ as in Property 2 and by Property 2

$$\begin{aligned}
||P_n(x_\alpha^\delta - x^\dagger)||_{\ell^1} &= \langle \xi, x_\alpha^\delta - x^\dagger \rangle_{\ell^\infty \times \ell^1} = \langle P_n A^* \eta, x_\alpha^\delta - x^\dagger \rangle_{\ell^\infty \times \ell^1} \\
&= \langle P_n A^* \eta - A^* \eta, x_\alpha^\delta - x^\dagger \rangle_{\ell^\infty \times \ell^1} + \langle A^* \eta, x_\alpha^\delta - x^\dagger \rangle_{\ell^\infty \times \ell^1} \\
&= -\langle (I - P_n) A^* \eta, (I - P_n)(x_\alpha^\delta - x^\dagger) \rangle_{\ell^\infty \times \ell^1} + \langle A^* \eta, x_\alpha^\delta - x^\dagger \rangle_{\ell^\infty \times \ell^1} \\
&\le \mu ||(I - P_n)(x_\alpha^\delta - x^\dagger)||_{\ell^1} + \gamma_n ||Ax_\alpha^\delta - Ax^\dagger||_Y.
\end{aligned}$$

From here we obtain

$$||P_n(x - x^\dagger)||_{\ell^1} \le \mu ||(I - P_n)x_\alpha^\delta||_{\ell^1} + \mu ||(I - P_n)x^\dagger||_{\ell^1} + \gamma_n ||Ax - Ax^\dagger||_Y \tag{2.3}$$

which plugged into (2.2) yields

$$||x_\alpha^\delta - x^\dagger||_{\ell^1} \le (1 + \mu)||(I - P_n)x_\alpha^\delta||_{\ell^1} + \gamma_n ||Ax_\alpha^\delta - Ax^\dagger||_Y + (1 + \mu)||(I - P_n)x^\dagger||_{\ell^1}. \tag{2.4}$$

If A is factored through a Hilbert space and the basis is smooth enough such that (1.6) holds (and consequently $\mu = 0$), then the step from $||P_n(x_\alpha^\delta - x^\dagger)||_{\ell^1}$ to $\gamma_n||Ax_\alpha^\delta - Ax^\dagger||_Y$ corresponds to the observation that the operator is no longer ill-posed but merely ill-conditioned when operating on a finite dimensional subspace, the condition number being based on the dimension of the subspace.

We continue with the right-most term in (2.2) and (2.4) which describes the smoothness of the solution as it measures the behavior of the tail of $x^\dagger$. It is simply

$$||(I - P_n)x^\dagger||_{\ell^1} = \sum_{i=n+1}^{\infty} |x_i^\dagger|. \tag{2.5}$$

For now we neglect the term $||(I - P_n)x_\alpha^\delta||_{\ell^1} = 0$ in (2.4) as obviously in the best possible case it vanishes. In this case we have

$$||x_\alpha^\delta - x^\dagger||_{\ell^1} \le \gamma_n||Ax_\alpha^\delta - Ax^\dagger||_Y + (1 + \mu) \sum_{i=n+1}^{\infty} |x_i^\dagger|$$

and its lowest value is given by taking the infimum over all n. Therefore we define

$$\varphi(t) := \inf_{n\in\mathbb{N}} \gamma_n t + (1 + \mu) \sum_{i=n+1}^{\infty} |x_i^\dagger| \tag{2.6}$$

which is the same rate as (1.8) but with a smaller constant. Therefore the proof that φ constitutes an index function is as in [2, Theorem 5.2].

The splitting of (2.2) is based on properties of the norm we measure the error in. The rate (2.6) depends on the properties of A and the exact solution $x^\dagger$. There are only two properties that have to be shown for a concrete regularization method. First, one has to have an estimate of $||Ax_\alpha^\delta - Ax^\dagger||$ which is usually not hard to obtain. More tricky is to show that the term $||(I - P_n)x_\alpha^\delta||_{\ell^1}$, which we so far neglected, is not too large. It turns out that for ℓ^1-regularization estimates for both terms can be derived directly from the Tikhonov-functional (1.2).

Namely, it is

$$\frac{1}{p}||Ax_\alpha^\delta - y^\delta||_Y^p + \alpha||x_\alpha^\delta||_{\ell^1} \le \frac{1}{p}||Ax^\dagger - y^\delta||_Y^p + \alpha||x^\dagger||_{\ell^1} \tag{2.7}$$

as the x_α^δ are the minimizers of the functional. Now splitting the ℓ^1-terms $||x||_{\ell^1} = ||P_n x||_{\ell^1} + ||(I - P_n)x||_{\ell^1}$ with $x = x_\alpha^\delta$ and $x = x^\dagger$, using the triangle inequality

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

and substracting the term $||P_n x_\alpha^\delta||_{\ell^1}$ from both sides we obtain

$$\frac{1}{p}||Ax_\alpha^\delta - y^\delta||_Y^p + \alpha||(I-P_n)x_\alpha^\delta||_{\ell^1}$$
$$\leq \frac{1}{p}||Ax^\dagger - y^\delta||_Y^p + \alpha(||(I-P_n)x^\dagger||_{\ell^1} + ||P_n(x^\dagger - x_\alpha^\delta)||_{\ell^1}) \qquad (2.8)$$

Substituting (2.3) yields after reordering

$$\frac{1}{p}||Ax_\alpha^\delta - y^\delta||_Y^p + \alpha(1-\mu)||(I-P_n)x_\alpha^\delta||_{\ell^1}$$
$$\leq \frac{1}{p}||Ax^\dagger - y^\delta||_Y^p + \alpha((1+\mu)||(I-P_n)x^\dagger||_{\ell^1} + \gamma_n||Ax_\alpha^\delta - Ax^\dagger||)$$

So far the dimension n was free. Now we fix it to be n_φ, which is where the infimum in (2.6) is attained. For $n = n_\varphi$ it is

$$\gamma_{n_\varphi}||Ax_\alpha^\delta - Ax^\dagger)|| + (1+\mu)||(I-P_{n_\varphi})x^\dagger||_{\ell^1} = \varphi(||Ax_\alpha^\delta - Ax^\dagger||) \leq 2\varphi(||Ax_\alpha^\delta - y^\delta||) \qquad (2.9)$$

where for the last inequality we assumed that $||Ax - y^\delta||_Y \geq \delta$ as in the opposite case we have $\varphi(||Ax_\alpha^\delta - Ax^\dagger||_Y) \leq 2\varphi(\delta)$ trivially. Note to this end that for a concave index function φ it holds $\varphi(C\cdot) \leq C\varphi(\cdot)$, $C \geq 1$, see [3]. With this (2.8) reads

$$\frac{1}{p}||Ax_\alpha^\delta - y^\delta||_Y^p + \alpha(1-\mu)||(I-P_{n_\varphi})x_\alpha^\delta||_{\ell^1} \leq \frac{1}{p}||Ax^\dagger - y^\delta||_Y^p + 2\alpha\varphi(||Ax_\alpha^\delta - y^\delta||). \qquad (2.10)$$

Ignoring the second term on the left hand side and inserting the estimate for the data error we have that

$$||Ax_\alpha^\delta - y^\delta||_Y^p \leq \delta^p + 2p\alpha\varphi(||Ax_\alpha^\delta - y^\delta||). \qquad (2.11)$$

From here we follow the lines of the proof of [3], Corollary 1, to deduce that with the parameter choice

$$c_1\frac{\delta^p}{\varphi(\delta)} \leq \alpha \leq c_2\frac{\delta^p}{\varphi(\delta)}, \qquad 0 < c_1 \leq c_2 < \infty \qquad (2.12)$$

it holds

$$||Ax_\alpha^\delta - y^\delta||_Y \leq \tilde{c}_p\delta. \qquad (2.13)$$

with

$$\tilde{c}_p = (1+2pc_2)^{1/(p-1)}. \qquad (2.14)$$

We now move to the term $||(I-P_n)x_\alpha^\delta||_{\ell^1}$ by going back to (2.10) and this time ignoring the first term on the left hand side. This yields, recalling $0 \le \mu < 1$,

$$\alpha(1-\mu)||(I-P_n)x_\alpha^\delta||_{\ell^1} \le ||Ax^\dagger - y^\delta||_Y^p + 2p\alpha\varphi(||Ax_\alpha^\delta - y^\delta||). \tag{2.15}$$

Inserting the parameter choice (2.12) gives

$$||(I-P_{n_\varphi})x_\alpha^\delta||_{\ell^1} \le \frac{\delta^2\varphi(\delta)}{2c_1(1-\mu)\delta^2} + \frac{2p}{1-\mu}\varphi(||Ax_\alpha^\delta - y^\delta||) \tag{2.16}$$

so that using (2.13)

$$||(I-P_{n_\varphi})x_\alpha^\delta||_{\ell^1} \le \bar{c}_p\varphi(\delta) \tag{2.17}$$

holds with

$$\bar{c}_p = \frac{1}{1-\mu}\left(\frac{1}{2c_1} + 2p\tilde{c}_p\right).$$

Going back to (2.4) we have

$$||x_\alpha^\delta - x^\dagger||_{\ell^1} \le (1+\mu)\bar{c}_p\varphi(\delta) + \varphi(2\tilde{c}_p\delta) \le c_p\varphi(\delta) \tag{2.18}$$

with the constant

$$c_p = \frac{1}{1-\mu}\left(\frac{1+\mu}{2c_1} + 4(1+2pc_2)^{1/(p-1)}\right).$$

In practice, an explicit expression for $\varphi(\delta)$ is only available in special cases, rendering the a-priori choice (2.12) useless otherwise. One way out of this dilemma is to use a discrepancy principle for the choice of the parameter. We shall employ here the strong discrepancy principle, i.e., given $1 < \tau_1 \le \tau_2$ we select α such that

$$\tau_1\delta \le ||Ax_\alpha^\delta - y^\delta|| \le \tau_2\delta. \tag{2.19}$$

At first let us show that under the assumption $||Ax_\alpha^\delta - y^\delta|| > \tau_1\delta$ for $\tau_1 > 1$ the regularization parameter can not become too small. We have from (2.10)

$$0 \le \delta^p - ||Ax_\alpha^\delta - y^\delta||_Y^p + 2p\alpha\varphi(||Ax_\alpha^\delta - Ax^\dagger||) \tag{2.20}$$

(Note that we changed the argument of $\varphi(\cdot)$ back to the basic estimate which is a tighter upper bound, see (2.11)). It follows with the same argumentation as in [3] that

$$\alpha \ge 2^{1-p}\frac{\tau_1^p - 1}{\tau_1^p + 1}\Phi((\tau_1 - 1)\delta). \tag{2.21}$$

with $\Phi(t) := \frac{t^2}{\varphi(t)}$. This also explains why we need to require $\tau_1 > 1$ in (2.19).

In order to show that the discrepancy principle (2.19) yields the same convergence rate as the a-priori choice, we need to show again that $||Ax_\alpha^\delta - y^\delta|| \leq C\delta$ and $||(I - P_{n_\varphi})x_\alpha^\delta||_{\ell^1} \leq C\varphi(\delta)$ hold with appropriate constants. The first property follows trivially from (2.19) with $C = \tau_2$. The second one follows since (2.15) yields, together with (2.19),

$$||(I - P_n)x_\alpha^\delta||_{\ell^1} \leq \frac{1}{2(1-\mu)} \frac{\delta^2}{\alpha} + 2\varphi(\tau_2\delta). \tag{2.22}$$

Namely, inserting (2.21) with $\tau = \tau_1$ results in the inequality

$$||(I - P_n)x_\alpha^\delta||_{\ell^1} \leq \tilde{c}\varphi(\delta)$$

where the constant $\tilde{c}$ is given by the expression

$$\tilde{c} = \frac{1}{2^p(1-\mu)} \frac{(\tau_1 - 1)^{(1-p)}(\tau_1^p + 1)}{\tau_1^p - 1} + 2\tau_2$$

This proves

$$||x_\alpha^\delta - x^\dagger||_{\ell^1} \leq c\varphi(\delta).$$

with $c = \tilde{c} + 2\tau_2$. This completes the proof.

Remark 5. By construction, the rate φ in (2.6) is not only the optimal convergence rate for ℓ^1-regularization but for *any* regularization method that measures the regularization error in the ℓ^1-norm. We speak of optimal here in the sense that the rate depends on the parameters γ_n which are usually difficult to obtain, even more so with a sharp estimate. As seen above, however, in particular the discrepancy principle does not need any information on the γ_n. We assume that the optimal combination of the parameters μ and γ_n in Property 2 is selected automatically but further investigation is necessary in this direction.

3 Case study: a diagonal operator

Examples of convergence rates of type (1.8),(2.6) can be found in [2, 14, 16, 15]. We will not recall them but focus on a particular problem for which we improve the known convergence rate and show that numerically our rate is achieved.

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. Hence $\widetilde{A}$ has the singular system $\{\sigma_i, u_i, v_i\}_{i\in\mathbb{N}}$ with decreasingly ordered singular values σ_i tending to zero and $\{u_i\}$, $\{v_i\}$ are complete orthonormal systems in $\widetilde{X}$ and $\overline{\mathcal{R}(A)}$, 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 may 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 write any $\tilde{x} \in \tilde{X}$ via $\tilde{x} = \sum_{i\in\mathbb{N}} x_i u_i$ where $x_i = \langle \tilde{x}, u_i\rangle$ with the scalar product in $\widetilde{X}$. The synthesis operator $L : \ell^2 \to \tilde{X}$ maps the variables $x = \{x_i\}_{i\in\mathbb{N}} \in \ell^2(\mathbb{N})$ to an element $\tilde{x}$ as above. Using the embedding $\mathcal{E} : \ell^1 \to \ell^2$ we finally obtain our linear bounded operator $A : \ell^1 \to Y$ via the composition $A = \tilde{A} \circ L \circ \mathcal{E}$. This is a diagonal operator since $Ae_i = \widetilde{A}u_i = \sigma_i v_i$ for all $i \in \mathbb{N}$. It fulfills (1.6) since $A^* \frac{v_i}{\sigma_i} = e_i$. In particular we may choose $\mu = 0$ in the following. In general it will still be difficult to calculate the convergence rate in (2.6). In order to keep the computations simple and to be able to track the constants, we will therefore assume that $\sigma_i = i^{-\beta}$ and $x_i = \langle x, u_i\rangle = i^{-\eta}$ for positive values β and $\eta > 1$ (such that $x^\dagger \in \ell^1$). We then have

$$\sum_{i=n+1}^{\infty} |x_i^\dagger| = \sum_{i=n+1}^{\infty} i^{-\eta} \leq \frac{1}{\eta - 1} n^{1-\eta}.$$

In order to estimate γ_n in (2.6) we follow [17, Example 3.8] where it was shown that

$$\gamma_n \leq \sup_{\substack{a_i \in \{-1,0,1\} \\ i=1,\ldots,n}} || \sum_{i=1}^{n} a_i f_i ||_Y = \sqrt{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}.$$

Instead of proceeding by estimating $\sqrt{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}} \leq \sum_{i=1}^{n} \frac{1}{\sigma_i}$, however, we evaluate the sum directly and take the square root afterwards. This leads to

$$\gamma_n \leq \sqrt{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}} = \sqrt{\sum_{i=1}^{n} i^{2\beta}} \leq \sqrt{\frac{1}{2\beta + 1}} n^{\beta + \frac{1}{2}}$$

in comparison to the original $\gamma_n \leq C n^{\beta+1}$, $C > 0$.

It is now simple calculus to find the convergence rate from (2.6). We obtain

$$\varphi(\delta) = c_\beta \delta^{\frac{\eta - 1}{\eta + \beta - \frac{1}{2}}} \tag{3.1}$$

with the constant

$$c_\beta = \left(\frac{1}{2}\sqrt{2\beta+1}\right)^{-\frac{1}{\eta+\beta-\frac{1}{2}}}.$$

In comparison to the rates presented in [2, 17] this improves the exponent from $\frac{\eta-1}{\eta+\beta}$ to $\frac{\eta-1}{\eta+\beta-\frac{1}{2}}$ and gives an explicit constant.

Finally we want to verify these rates numerically. In order to arrive at the same setting as above we start with the Voltera operator

$$[\tilde{A}x](s) = \int_0^s x(t)\,dt. \tag{3.2}$$

We then discretize $\tilde{A}$ with the rectangular rule at $N = 400$ points. In order to ensure our desired properties, we compute the SVD of the resulting matrix and manually set its singular values σ_i to $\sigma_i = i^{-\beta}$. 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 $\langle x^\dagger, v_i\rangle = i^{-\eta}$ holds for various values of $\eta > 0$. 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

$$x_\alpha^\delta = \text{argmin}\frac{1}{2}||Ax - y^\delta||^2 + \alpha||x||_{\ell_1},$$

where the ℓ^1-norm was taken of the coefficients with respect to the basis originating from the SVD. The minimizer was obtained via iterative soft shrinking [1]. The regularization parameter was chosen a priori according to (2.12) with $c_1 = c_2 = 1$. The constant c_p in (2.18) takes in our case $p = 2$ the value $c_p = 20.5$. We computed the reconstruction error in the ℓ_1 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

$$||P_{400}(x_\alpha^\delta - x^\dagger)||_{\ell^1}$$

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

$$||P_{400}(x_\alpha^\delta - x^\dagger)||_{\ell^1} + ||(I - P_{400})x^\dagger||_{\ell^1}. \tag{3.3}$$

η	α, a	measured rate, e	predicted rate, e	residual, d
1.01	1.99	0.0065	0.0066	1.01
1.05	1.97	0.0332	0.0322	1.008
1.1	1.94	0.0625	0.0625	1.009
1.3	1.83	0.1691	0.1667	1.006
1.5	1.75	0.2588	0.25	1.005
2	1.6	0.3961	0.4	1.003
2.5	1.5	0.498	0.5	0.996

Table 1: Convergence rates for $\beta = 1$ and various values η. α in the form $\alpha = \delta^a$. Measured and predicted regularization error in the form $||x_\alpha^\delta - x^\dagger||_{\ell^1} = c\delta^e$, cf. (3.1). Residual in the form $||Ax_\alpha^\delta - y^\delta|| = c\delta^d$.

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. For each fixed η, β we calculated the regression for the conjecture $||x_\alpha^\delta - x^\dagger|| = c\delta^e$. From the theoretical part the constant c is given by $c = c_p c_\beta$ with c_p from (2.18) and c_β from (3.1). In the following tables we present observed convergence rates for $\beta = 1$ and $\beta = 2$, respectively, and various values of $\eta > 1$. We provide the exponent e of the rate according to the regression and compare it with the theoretical exponent. In all tests we observed a nice fit. We also monitored the constants. The theoretical upper bounds $c \approx 22.5.6$ and $c \approx 19.6$ for $\beta = 1$ and $\beta = 2$, respectively, have not been exceeded, the highest observed constant being $c = 15.66$. However as η increases the measured constant decreases. For example for $\eta = 2.5$ and $\beta = 1$ we obtained $c = 0.74$. It is an open topic to understand this behavior and tighten the constants in the convergence rate. In the tables we additionally give the residual obtained from another regression for the conjecture $||Ax_\alpha^\delta - y^\delta||^2 = C\delta^d$ with d given. This hovers nicely around the theoretical value $d = 1$. We observed hat the theoretical constant $C = c_p = 5$ from (2.14) is not hit as the observed constant C obtained values between 1.1 and 1.4.

We only give examples for the a-priori parameter choice here. In our experiments the discrepancy principle (2.19) yields essentially identical results.

η	$\alpha = \delta^a$, a	measured rate, e	predicted rate, e	residual, d
1.01	1.996	0.0039	0.0038	0.999
1.05	1.97	0.02	0.0196	1.001
1.1	1.96	0.0388	0.0385	0.998
1.3	1.89	0.1063	0.1071	1.002
1.5	1.83	0.1681	0.1667	0.999
2	1.71	0.2864	0.2857	1.000
2.5	1.625	0.3617	0.375	1.002

Table 2: Convergence rates for $\beta = 2$ and various values η. α in the form $\alpha = \delta^a$. Measured and predicted regularization error in the form $||x_\alpha^\delta - x^\dagger||_{\ell^1} = c\delta^e$, cf. (3.1). Residual in the form $||Ax_\alpha^\delta - y^\delta|| = c\delta^d$.

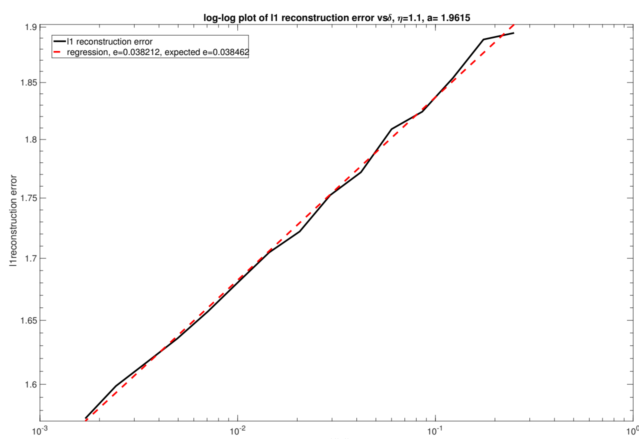

Figure 1: Exemplary observed convergence rate for $\eta = 1.1$, $\beta = 2$ and $\alpha = 1.96$ according to (2.12). The regression for the conjecture $||x_\alpha^\delta - x^\dagger|| = c\delta^e$ is given dashed.

4 Conclusion

We have shown that convergence rates for ℓ^1-regularization can be derived based on the tail of the true solution $x^\dagger$ in ℓ^1 and the smoothness of the operator expressed by Property 2. These define the rate. The strategy of the proof is based on a splitting of the regularization error in a finite dimensional part and two infinite dimensional tail terms. Using this, for any regularization method one only has to show two properties. First it is required that $||Ax_\alpha^\delta - y^\delta||_Y \sim \delta$ and second one needs to control the tail of the regularized solution in the ℓ^1-norm. For ℓ^1-regularization both estimates can be derived directly from the Tikhonov-functional. For a specific problem involving a diagonal operator we could verify that the predicted convergence rates are achieved numerically.

References

[1] 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

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

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

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

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

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

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

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

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

[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), John Wiley, New Jersey 2015

[11] B. Hofmann, P.Mathé 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

[12] D. Gerth, J. Flemming *Convergence rates for ℓ^1-regularization without source-type conditions* preprint, 2017

[13] M. Grasmair, *Well-posedness and convergence rates for sparse regularization with sublinear ℓ^q penalty term.* Inverse Problems and Imaging **3** (3), 2009

[14] S. W. Anzengruber, B. Hofmann, R. Ramlau *On the interplay of basis smoothness and specific range conditions occurring in sparsity regularization* Inverse problems **29** 125002, 2013

[15] J. Flemming and M. Hegland *Convergence rates in ell^1-regularization when the basis is not smooth enough* Applicable Analysis **94** (3), 2015

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

[17] J. Flemming, B. Hofmann and I. Veselic *On ℓ^1-regularization in Light of Nashed's Ill-posedness Concept* Comput. Methods Appl. Math. **15**, 2015

[18] J. Flemming *Convergence rates for ℓ^1-regularization without injectivity-type assumptions* Inverse Problems **32** 095001, 2016

[19] B. Jin and P. Maaß *Sparsity regularization for parameter identification* Inverse Problems **18**, 2012

[20] Bredies K and Lorenz D A 2009 Linear convergence of iterative soft-thresholding *Journal of Fourier Analysis and Applications* **14** 5-6 813–837