Electronic Transactions on Numerical Analysis.
Volume 46, pp. 233–244, 2017.

ISSN 1068–9613.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

CONVERGENCE RATES FOR ℓ^1-REGULARIZATION WITHOUT THE HELP OF A VARIATIONAL INEQUALITY*

DANIEL GERTH†

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 found in the literature and conduct numerical experiments that illustrate the predicted rate. The idea of the proof is rather generic and might be used in other settings as well.

Key words. ℓ^1-regularization, Tikhonov regularization, variational inequality, convergence rates

AMS subject classifications. 65J20, 47A52

1. Introduction and main theorem.

We seek to solve the linear, ill-posed equation

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

for $x \in \ell^1$ where instead of the true data y in some Banach space Y only perturbed data y^δ satisfying $\|y - y^\delta\|_Y \leq \delta$, $\delta > 0$, are available. Throughout this paper we assume the following properties of the operator.

ASSUMPTION 1.1. *Let the bounded linear operator $A : \ell^1 \to Y$ which maps absolutely summable real sequences into some real Banach space Y be* injective *and* sequentially weak*-to-weak continuous.

The latter property makes sense as ℓ^1 is the dual of c_0, the space of all sequences converging to zero. The weak*-to-weak continuity thus means that for any sequence $\{x_n\}_{n=1}^\infty \in \ell^1$

$$\langle \eta, Ax_n\rangle_{Y^*\times Y} \to \langle \eta, Ax\rangle_{Y^*\times Y} \forall \eta \in Y^* \text{ whenever } \langle x_n, \zeta\rangle_{\ell^1\times c_0} \to \langle x, \zeta\rangle_{\ell^1\times c_0} \forall \zeta \in c_0.$$

Here and further on the brackets $\langle\cdot,\cdot\rangle$ denote the duality pairing between the indicated spaces and Y^* is the dual space of Y. In order to find an approximation to the unique exact solution $x^\dagger$ to the (unknown) exact data $y = Ax^\dagger \in Y$ we employ ℓ^1-regularization, i.e., 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. To the best of the author's knowledge, the weak*-to-weak continuity as in Assumption 1.1 corresponds to the weakest sufficient condition for the existence of minimizers of (1.2) currently available in the literature; see [6, 7, 17]. In [6] it has been shown that the weak*-to-weak continuity of A is equivalent to the condition $\mathcal{R}(A^*) \subseteq c_0$ as well as the condition $Ae_i \rightharpoonup 0$. Here and in the remainder of the paper, $A^* : Y^* \to \ell^\infty = (\ell^1)^*$ denotes the adjoint of A, $\mathcal{R}$ the range of an operator, and e_i, $i \in \mathbb{N}$, is the canonical basis in ℓ^1, i.e., the k-th component of e_i equals 1 for $i = k$ and 0 otherwise. Assumption 1.1 is for example fulfilled if A is injective and 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 the Hilbert space ℓ^2 occurs frequently. As a counterexample, the identity $\mathrm{Id} : \ell^1 \to \ell^1$

*Received February 13, 2017. Accepted June 28, 2017. Published online on July 7, 2017. Recommended by S. Kindermann. Work supported by DFG grants HO 1454/8-2 and HO1454/10-1.
†Technische Universität Chemnitz, Fakultät für Mathematik, D-09107 Chemnitz, Germany (daniel.gerth@mathematik.tu-chemnitz.de).

ETNA
Kent State University and
Johann Radon Institute (RICAM)

is not weak*-to-weak continuous as it is easy to show that $\mathcal{R}(\mathrm{Id}^*) = \ell^\infty$. Another direct consequence of Assumption 1.1 is the following property as shown in [7]. To this end let

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

denote the cut-off operator after the n-th component.

PROPERTY 1.2. *There exist a sequence of real numbers $(\gamma_n)_{n\in\mathbb{N}}$ 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 is $\eta = \eta(n,\xi,\mu)$ in Y^* fulfilling

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

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

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

We cite [7, Prop. 12] to clarify the relation between Assumption 1.1 and Property 1.2.

PROPOSITION 1.3. *Let A be sequentially weak*-to-weak continuous. Then the following statements are equivalent.*

i) For every $0 < \mu < 1$ and every ξ as in Property 1.2, there is $\eta \in Y^*$ such that Property 1.2 holds,

ii) $e_i \in \overline{\mathcal{R}(A^*)}$ $\forall i \in \mathbb{N}$, where the closure is taken with respect to the norm in ℓ^∞,

iii) $\overline{\mathcal{R}(A^*)} = c_0$,

iv) A is injective.

One can show that $\gamma_n \to \infty$ as $n \to \infty$ whenever $\mathcal{R}(A) \ne \overline{\mathcal{R}(A^*)}$, i.e., A is ill-posed. Thus, letting A be as in Assumption 1.1, allows us to work with Property 1.2. For a survey on ℓ^1-regularization theory we refer to [14]. We will only give a brief summary here. The seminal paper [4] 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

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

where C is a positive constant and φ a concave index function; i.e., a continuous and monotonically increasing function with $\varphi(0) = 0$. First results were already given in [4]. Later the focus shifted to sparse solutions where $x^\dagger$ as solution of the unperturbed equation (1.1) has only finitely many non-zero elements. Convergence rates in this case can be found in, e.g., [2, 12, 15], where different kinds of smoothness properties of A and $x^\dagger$ are used to derive the rates. Starting with the paper [3], the case that $x^\dagger$ is an infinite sequence in ℓ^1 has gathered attention recently. There, a smoothness condition on the canonical basis with respect to the operator is crucial. Namely, the authors assume that

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

Such a condition already appeared in [2] and can be traced back to [11]. While (1.6) holds for various types of inverse problems (see [1]), it does not hold in general. A counterexample was presented in [8]. The authors of [8] thus formulated a relaxed assumption which was again slightly generalized in [10]. The assumption made in [10] in principle coincides with Property 1.2. It is an important contribution of [7] to show that Property 1.2 is already a consequence of Assumption 1.1. Using the weak*-to-weak continuity of A, (1.6) is generalized to item ii) in Proposition 1.3. Please note that (1.6) implies Property 1.2 with $\mu = 0$.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

All proofs of convergence rates in the literature rely on some condition relating the smoothness of the solution and the smoothing properties of the operator. Such a relation is typically expressed via the assumption $x^\dagger \in \mathcal{R}((A^*A)^\nu)$ for some $\nu > 0$; see for example [5, 16]. In recent years, variational inequalities (sometimes also called variational source conditions) have been used as link condition. For ℓ^1-regularization it was shown in [3] that

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

holds for all $x \in \ell^1$ assuming (1.6). Here, φ is a concave index function given by

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

with $\gamma_n = \sum_{i=1}^n \|f_i\|_{Y^*}$, where due to (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 [13].

The aim of this paper is to show that a link condition as for example the variational inequality (1.7) is not necessary to obtain convergence rates. While the use of a variational inequality may be more elegant, it is interesting that all the required information is already contained in the Tikhonov functional (1.2). There is no particular technical difficulty in our approach in addition to the techniques used in previous works to derive a variational inequality. In fact, we use several results from this framework. In Section 2 we provide a proof of the following Theorem 1.4 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 in Section 4. Our main result is the following.

THEOREM 1.4. *Let $A : \ell^1 \to Y$ be as in Assumption 1.1. Then there is a concave index function $\varphi : [0, \infty) \to [0, \infty)$ of the form*

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

with μ and γ_n from Property 1.2 such that

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

holds with a constant $C > 0$ whenever the regularization parameter is chosen a priori such that $c_1 \frac{\delta^p}{\varphi(\delta)} \leq \alpha(\delta) \leq c_2 \frac{\delta^p}{\varphi(\delta)}$ for some $0 < c_1 \leq c_2 < \infty$ or a posteriori via the two-sided discrepancy principle, i.e., x_α^δ satisfies $\tau_1\delta \leq \|Ax_\alpha^\delta - y^\delta\|_Y \leq \tau_2\delta$ for some $1 < \tau_1 \leq \tau_2$.

In the proof we will explicitly determine the constants C. We avoid their exact expressions in the theorem 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., for some $n_0 \in \mathbb{N}$, $\operatorname{supp}(x^\dagger) \subseteq \{1, \ldots, n_0\}$, is covered in our assumptions. We then obtain

$$\varphi(\delta) = \gamma_{n_0}\delta,$$

which is the known linear rate.

2. Proof of the main theorem. First observe that for arbitrary $n \in \mathbb{N}$ we can split

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

with the projectors P_n from (1.3). Using the triangle inequality we get

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

We will estimate the terms on the right-hand side of (2.1). Consider first the term in the middle. With some $\xi \in \ell^\infty$ as in Property 1.2 and by Property 1.2, we have

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

From this we obtain

$$\|P_n(x_\alpha^\delta - x^\dagger)\|_{\ell^1} \leq \mu\|(I-P_n)x_\alpha^\delta\|_{\ell^1} + \mu\|(I-P_n)x^\dagger\|_{\ell^1} + \gamma_n\|Ax - Ax^\dagger\|_Y, \tag{2.2}$$

which substituted into (2.1) yields

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

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 $\gamma_n\|Ax_\alpha^\delta - Ax^\dagger\|_Y$ in (2.2) corresponds to the observation that the operator is no longer ill-posed but merely ill-conditioned when operating on a finite-dimensional subspace with the condition number being based on the dimension of the subspace.

We continue with the right-most terms in (2.1) and (2.3). It describes the smoothness of the solution as it measures the behavior of the tail of $x^\dagger$. We simply have

$$\|(I-P_n)x^\dagger\|_{\ell^1} = \sum_{i=n+1}^{\infty} |x_i^\dagger|. \tag{2.4}$$

Let us for now neglect the term $\|(I-P_n)x_\alpha^\delta\|_{\ell^1}$ in (2.3) as in the best possible case it vanishes. Then we have from (2.3) and (2.4)

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

Its lowest value is given by taking the infimum over all n. Therefore, we define

$$\varphi(t) := \inf_{n\in\mathbb{N}} \gamma_n t + (1+\mu)\sum_{i=n+1}^{\infty}|x_i^\dagger|, \tag{2.5}$$

which is the same rate as (1.8) but with a smaller constant. Hence, the proof that φ constitutes a concave index function follows the lines of [3, Theorem 5.2].

So far we only used properties of X and Y without any reference to the regularization method. Indeed, the splitting of (2.1) is based on properties of the norm we measure the error in. The rate (2.5) 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

ETNA
Kent State University and
Johann Radon Institute (RICAM)

that for ℓ^1-regularization, estimates for both terms can be derived directly from the Tikhonov functional (1.2). Namely, it holds that

$$\frac{1}{p}\|Ax_\alpha^\delta - y^\delta\|_Y^p + \alpha\|x_\alpha^\delta\|_{\ell^1} \le \frac{1}{p}\|Ax^\dagger - y^\delta\|_Y^p + \alpha\|x^\dagger\|_{\ell^1} \tag{2.6}$$

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 $x = x^\dagger$, respectively, using the triangle inequality

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

and subtracting the term $\|P_n x_\alpha^\delta\|_{\ell^1}$ from both sides, we obtain from (2.6)

$$\begin{aligned}&\frac{1}{p}\|Ax_\alpha^\delta - y^\delta\|_Y^p + \alpha\|(I-P_n)x_\alpha^\delta\|_{\ell^1}\\ &\le \frac{1}{p}\|Ax^\dagger - y^\delta\|_Y^p + \alpha(\|(I-P_n)x^\dagger\|_{\ell^1} + \|P_n(x_\alpha^\delta - x^\dagger)\|_{\ell^1}).\end{aligned}$$

Substituting (2.2) yields after reordering

$$\begin{aligned}&\frac{1}{p}\|Ax_\alpha^\delta - y^\delta\|_Y^p + \alpha(1-\mu)\|(I-P_n)x_\alpha^\delta\|_{\ell^1}\\ &\le \frac{1}{p}\|Ax^\dagger - y^\delta\|_Y^p + \alpha((1+\mu)\|(I-P_n)x^\dagger\|_{\ell^1} + \gamma_n\|Ax_\alpha^\delta - Ax^\dagger\|).\end{aligned} \tag{2.7}$$

So far the dimension n was free. Now we fix it to n_φ which is where the infimum in (2.5) is attained. Its existence follows since the infimum is taken over a countable set and since $\gamma_n \to \infty$ as $n \to \infty$. For $n = n_\varphi$ it is

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

where for the last inequality we assumed that $\|Ax - y^\delta\|_Y \ge \delta$ as in the opposite case we have $\varphi(\|Ax_\alpha^\delta - Ax^\dagger\|_Y) \le 2\varphi(\delta)$ trivially. Note to this end that for a concave index function φ it holds that $\varphi(C\cdot) \le C\varphi(\cdot)$, $C \ge 1$; see [13]. With this, (2.7) reads

$$\frac{1}{p}\|Ax_\alpha^\delta - y^\delta\|_Y^p + \alpha(1-\mu)\|(I-P_{n_\varphi})x_\alpha^\delta\|_{\ell^1} \le \frac{1}{p}\|Ax^\dagger - y^\delta\|_Y^p + 2\alpha\varphi(\|Ax_\alpha^\delta - y^\delta\|). \tag{2.8}$$

Ignoring the second term on the left hand side and using $\|y - y^\delta\|_Y \le \delta$ we have that

$$\|Ax_\alpha^\delta - y^\delta\|_Y^p \le \delta^p + 2p\alpha\varphi(\|Ax_\alpha^\delta - y^\delta\|). \tag{2.9}$$

From here we follow the proof of [13, Corollary 1] to deduce that with the parameter choice

$$c_1\frac{\delta^p}{\varphi(\delta)} \le \alpha \le c_2\frac{\delta^p}{\varphi(\delta)}, \qquad 0 < c_1 \le c_2 < \infty, \tag{2.10}$$

it holds that

$$\|Ax_\alpha^\delta - y^\delta\|_Y \le \tilde{c}_p\delta \tag{2.11}$$

with

$$\tilde{c}_p = (1 + 2pc_2)^{1/(p-1)}. \tag{2.12}$$

We now move to the term $\|(I - P_n)x_\alpha^\delta\|_{\ell^1}$ by going back to (2.8) and this time ignoring the first term on the left-hand side. This yields, recalling that $0 \le \mu < 1$,

$$\alpha(1-\mu)\|(I - P_n)x_\alpha^\delta\|_{\ell^1} \le \|Ax^\dagger - y^\delta\|_Y^p + 2p\alpha\varphi(\|Ax_\alpha^\delta - y^\delta\|). \tag{2.13}$$

Inserting the parameter choice (2.10) gives

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

so that with (2.11)

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

holds with the constant

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

Going back to (2.3) we have

$$\|x_\alpha^\delta - x^\dagger\|_{\ell^1} \le (1+\mu)\bar{c}_p\varphi(\delta) + \varphi(2\tilde{c}_p\delta) \le c_p\varphi(\delta) \tag{2.14}$$

with the constant

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

Thus Theorem 1.4 is proven for the a-priori parameter choice. In practice an explicit expression for $\varphi(\delta)$ is only available in special cases, rendering the a-priori choice (2.10) useless otherwise. One way out of this dilemma is to use an a posteriori choice of the regularization parameter. We shall employ here the two-sided discrepancy principle, i.e., given $1 < \tau_1 \le \tau_2$ we select the regularization parameter α such that

$$\tau_1\delta \le \|Ax_\alpha^\delta - y^\delta\| \le \tau_2\delta. \tag{2.15}$$

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 (2.8)

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

Note that we changed the argument of $\varphi(\cdot)$ back to the basic estimate which is a tighter upper bound; see (2.9). It follows with the same argumentation as in [13] that

$$\alpha \ge 2^{1-p}\frac{\tau_1^p - 1}{\tau_1^p + 1}\Phi((\tau_1 - 1)\delta) \tag{2.16}$$

with $\Phi(t) := \frac{t^2}{\varphi(t)}$. In order for (2.16) to be meaningful we need to require $\tau_1 > 1$ in (2.15). To prove that the discrepancy principle (2.15) yields the same convergence rate as the a-priori choice, we need to show that also now $\|Ax_\alpha^\delta - y^\delta\| \le C\delta$ and $\|(I - P_{n_\varphi})x_\alpha^\delta\|_{\ell^1} \le C\varphi(\delta)$

hold with appropriate constants. The first property with $C = \tau_2$ is trivial by (2.15). The second one follows since (2.13) and (2.15) together yield

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

Inserting (2.16) 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 $c = \tilde{c} + 2\tau_2$ under the discrepancy principle. The proof of Theorem 1.4 is complete.

REMARK 2.1. By construction, the rate φ in (2.5) is optimal up to the behavior of the sequence $\{\gamma_n\}_{n\in\mathbb{N}}$ in Property 1.2 and to a lesser extent the parameter μ. In this sense our result not only yields the optimal convergence rate for ℓ^1-regularization but for *any* regularization method of the operator A that measures the regularization error in the ℓ^1-norm. It is an open problem to investigate the relation between μ and γ_n in Property 1.2. Results in this direction might lead to a clear definition of optimal convergence rates in ℓ^1-regularization. However, in particular the discrepancy principle does not need any information on the γ_n. We believe that the optimal combination of the parameters μ and γ_n in Property 1.2 is selected automatically.

3. Case study: a diagonal operator. Examples of convergence rates of type (1.8), (2.5) can be found in [1, 3, 8, 10]. 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 an injective, compact operator between Hilbert spaces. This allows us to use its singular system. 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 $\{u_i\}$ 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 $L : \ell^2 \to \widetilde{X}$ 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 $A = \tilde{A} \circ L \circ \mathcal{E}$. Due to the factorization through ℓ^2, the operator A is weak*-to-weak continuous. Note that A still 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) due to $A^* \frac{v_i}{\sigma_i} = e_i \ \forall i \in \mathbb{N}$. This means we have $\mu = 0$ in Property 1.2. In general it will still be difficult to calculate the convergence rate in (2.5). In order to keep the computations simple and to be able to track the constants, we assume that $\sigma_i = i^{-\beta}$ and $x_i^\dagger = \langle x^\dagger, u_i\rangle = i^{-\eta}$ for some positive β and $\eta > 1$ such that $x^\dagger \in \ell^1$. Then,

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

ETNA
Kent State University and
Johann Radon Institute (RICAM)

In order to estimate γ_n in (2.5), we follow [9, Example 3.8], where it was shown that

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

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

$$\gamma_n \leq \sqrt{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}} = \sqrt{\sum_{i=1}^{n} i^{2\beta}} \simeq \sqrt{\frac{1}{2\beta+1}} n^{\beta+\frac{1}{2}}, \tag{3.1}$$

which is an improvement in comparison to the original bound $\gamma_n \leq Cn^{\beta+1}$, $C > 0$. In (3.1) the constant $\sqrt{\frac{1}{2\beta+1}}$ is meant asymptotically as $n \to \infty$ indicated by the symbol "$\simeq$". We will continue with this constant in order to establish estimates for further constants that are as sharp as possible.

It is now simple calculus to find the convergence rate from (2.5). We obtain

$$\varphi(\delta) = c_\beta \delta^{\frac{\eta-1}{\eta+\beta-\frac{1}{2}}} \tag{3.2}$$

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 [3, 9] this improves the exponent from $\frac{\eta-1}{\eta+\beta}$ to $\frac{\eta-1}{\eta+\beta-\frac{1}{2}}$ and provides an explicit constant.

REMARK 3.1. The polynomial rate as obtained above does not rely on a polynomial decay of the singular values $\sigma_i \sim i^{-\beta}$ of the operator and the values $\langle x^\dagger, u_i \rangle \sim i^{-\eta}$. In fact, assume that σ_i and $\langle x^\dagger, u_i \rangle$ are such that

$$\gamma_n \leq ce^n \quad \text{and} \quad \sum_{i=n+1}^{\infty} |x_i^\dagger| \leq ce^{-n},$$

where for the sake of simplicity we use generic constants $c > 0$. Then one obtains, following the steps above, the convergence rate

$$\varphi(\delta) = c\sqrt{\delta}.$$

The exponentially ill-posed operator yields under the assumption of a particularly smooth exact solution a moderately ill-posed problem. Therefore, it appears to be impossible to infer the smoothness of both the operator and solution from a given convergence rate without a priori knowledge of the type of smoothness of these objects.

4. Numerical examples. Finally we want to illustrate the rates numerically. In order to arrive at the same setting as above we start with the Volterra operator

$$[\tilde{A}x](s) = \int_0^s x(t)\, dt. \tag{4.1}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

Then we 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 $\tilde{A}$ in (4.1) but has 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

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

where the ℓ^1-norm of the coefficients is taken with respect to the basis originating from the SVD. The minimizer is obtained via iterative soft shrinking [4]. The regularization parameter is chosen a priori according to (2.10) with $c_1 = c_2 = 1$. The constant c_p in (2.14) takes—in our case with $p = 2$—the value $c_p = 20.5$. We compute the reconstruction error in the ℓ_1-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 $\|P_{400}(x_\alpha^\delta - x^\dagger)\|_{\ell^1}$ with the projectors P_{400} as in (1.3). In the plots of the convergence rates we display the quantity $\|P_{400}(x_\alpha^\delta - x^\dagger) + (I - P_{400})x^\dagger\|_{\ell^1}$. The second term in the norm 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 due to discretization. For each fixed η, β we calculate the regression for the hypothesis $\|x_\alpha^\delta - x^\dagger\| = c\delta^e$.

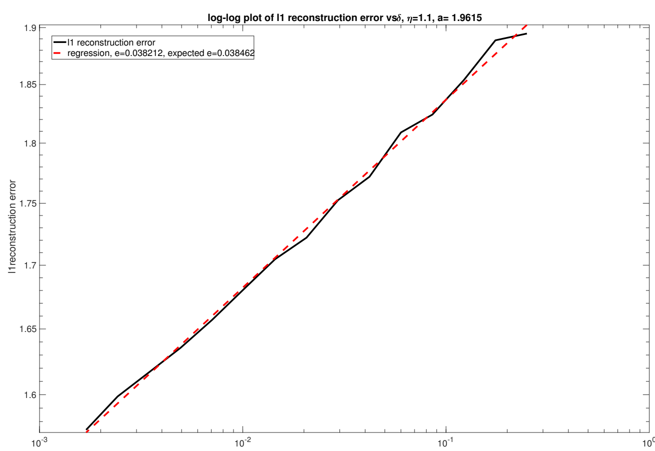

FIG. 4.1. *Observed convergence rate (solid line) for the diagonal operator, $\eta = 1.1$, $\beta = 2$ and $\alpha = \delta^{1.96}$ according to (2.10). The regression for the hypothesis $\|x_\alpha^\delta - x^\dagger\| = c\delta^e$ is given dashed. We obtained $e \approx 0.0382$; from theory we expected $e \approx 0.0385$.*

An example is given in Figure 4.1. From the analysis above, the constant c is found as $c = c_p c_\beta$ with c_p from (2.14) and c_β from (3.2). In Tables 4.1 and 4.2 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 $c \approx 22.56$ and $c \approx 19.6$ for $\beta = 1$ and $\beta = 2$, respectively, have not been exceeded by 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 hypothesis $\|Ax_\alpha^\delta - y^\delta\| = C\delta^d$ with d given. This hovers nicely around the theoretical value $d = 1$. The theoretical constant $C = c_p = 5$ from (2.12) is not achieved as the observed constant C takes values between 1.1 and 1.4. We only give examples for the a-priori parameter choice here. In our experiments the discrepancy principle (2.15) yields essentially identical results.

TABLE 4.1

Diagonal operator: 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. (3.2). Residual in the form $\|Ax_\alpha^\delta - y^\delta\| = c\delta^d$.

η	a	measured e	predicted e	measured 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 4.2

Diagonal operator: 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. (3.2). Residual in the form $\|Ax_\alpha^\delta - y^\delta\| = c\delta^d$.

η	a	measured e	predicted e	measured 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

In a second example we consider a diagonal-dominant operator. The reason that we can not move too far away from the diagonal case is that otherwise we have no analytic reference to compare those rates to, although, of course, we can still observe the convergence rates numerically. The construction of our operator is similar to the one in the first part. We start with (4.1), discretize with the trapezoidal rule on $N = 400$ points, and use the SVD of the resulting matrix. The matrix S containing the singular values is then altered as follows. First we set up the diagonal matrix from the previous setting, now named D, i.e., $d_{i,i} = i^{-\beta}$, $i = 1, \dots, N$, and D is zero otherwise. Then we take a full random matrix $\tilde{S}$ such that each element $\tilde{s}_{ij}$ is uniformly distributed in $[-1, 1]$ and multiply each row $\tilde{s}_i$ by $0.075 i^{-\beta}$. Finally, we combine $S = D + \tilde{S}$. Now S is a full matrix whose largest entry in each row is on the diagonal. All other entries in a row are at most 7.5% of that in magnitude. We also randomized the solution. For p_i, $i = 1, \dots, N$, being uniformly distributed random variables in $[-1, 1]$, we set $\langle x^\dagger, v_i \rangle = i^{-\eta}(1 + 0.25 p_i)$. In contrast to the previous section we now choose the

ETNA
Kent State University and Johann Radon Institute (RICAM)

regularization parameter according to the discrepancy principle. This means we start with a large $\alpha = \alpha_0 > 0$ and reduce it until (2.15) is fulfilled with $\tau_1 = 1.0001$ and $\tau_2 = 1.3$. To further move away from the previous setting we now choose $\beta = 0.5$ and $\beta = 3$. We compare the observed rates to the ones predicted for the fully diagonal operator. The results are shown in Table 4.3 for $\beta = 0.5$ and Table 4.4 for $\beta = 3$. As expected, theoretical and observed convergence rates do not match as well as in the fully diagonal case. However, they are still reasonably close.

TABLE 4.3
Diagonally dominant operator: Convergence rates for $\beta = \frac{1}{2}$ and various values η. α in the form $\alpha = c\delta^a$, chosen via discrepancy principle (2.15). Measured and estimated regularization error in the form $\|x_\alpha^\delta - x^\dagger\|_{\ell^1} = c\delta^e$; cf. (3.2). Residual in the form $\|Ax_\alpha^\delta - y^\delta\| = c\delta^d$.

η	measured a	estimated a	measured e	estimated e	measured d
1.01	1.7	1.99	0.0073	0.0099	1.03
1.05	1.67	1.95	0.038	0.048	1.018
1.1	1.77	1.91	0.081	0.091	1.028
1.3	1.69	1.76	0.21	0.231	1.016
1.5	1.63	1.66	0.316	0.33	1.005
2	1.48	1.5	0.45	0.5	1.003
2.5	1.39	1.4	0.564	0.6	1.005

TABLE 4.4
Diagonally dominant operator: Convergence rates for $\beta = 3$ and various values η. α in the form $\alpha = c\delta^a$, chosen via discrepancy principle (2.15). Measured and estimated regularization error in the form $\|x_\alpha^\delta - x^\dagger\|_{\ell^1} = c\delta^e$; cf. (3.2). Residual in the form $\|Ax_\alpha^\delta - y^\delta\| = c\delta^d$.

η	measured a	estimated a	measured e	estimated e	measured d
1.01	1.88	1.997	0.0021	0.0028	0.98
1.05	1.93	1.99	0.0135	0.014	0.983
1.1	1.98	1.97	0.0238	0.0278	0.986
1.3	1.93	1.92	0.083	0.079	0.976
1.5	1.89	1.88	0.099	0.125	0.98
2	1.82	1.78	0.198	0.222	0.99
2.5	1.72	1.7	0.269	0.3	0.99

5. 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 1.2. These objects define the rate. The strategy of the proof is based on a splitting of the regularization error into 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.

Acknowledgments. The author thanks the anonymous reviewer for his suggestions and Bernd Hofmann (TU Chemnitz) for the fruitful discussions and helpful comments.

REFERENCES

[1] S. W. Anzengruber, B. Hofmann, and R. Ramlau, *On the interplay of basis smoothness and specific range conditions occurring in sparsity regularization*, Inverse Problems, 29 (2013), Art. Id. 125002 (21 pages).

[2] K. Bredies and D. A. Lorenz, *Linear convergence of iterative soft-thresholding*, J. Fourier Anal. Appl., 14 (2008), pp. 813–837.

[3] M. Burger, J. Flemming, and B. Hofmann, *Convergence rates in ℓ^1-regularization if the sparsity assumption fails*, Inverse Problems, 29 (2013), Art. Id. 025013 (16 pages).

[4] 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 (2004), pp. 1413–1457.

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

[6] J. Flemming, *Convergence rates for ℓ^1-regularization without injectivity-type assumptions*, Inverse Problems, 32 (2016), Art. Id. 095001 (19 pages).

[7] J. Flemming and D. Gerth, *Injectivity and weak*-to-weak continuity suffice for convergence rates for ℓ^1-regularization*, J. Inverse Ill-posed Probl., published online 2017 ahead of print.

[8] J. Flemming and M. Hegland, *Convergence rates in ℓ^1-regularization when the basis is not smooth enough*, Appl. Anal., 94 (2015), pp. 464–476.

[9] 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), pp. 279–289.

[10] ———, *A unified approach to convergence rates for ℓ^1-regularization and lacking sparsity*, J. Inverse Ill-posed Probl., 24 (2016), pp. 139–148.

[11] M. Grasmair, *Well-posedness and convergence rates for sparse regularization with sublinear ℓ^q penalty term*, Inverse Probl. Imaging, 3 (2009), pp. 383–387.

[12] M. Grasmair, M. Haltmeier, and O. Scherzer, *Sparse regularization with l^q penalty term*, Inverse Problems, 24 (2008), Art. Id. 055020 (13 pages).

[13] B. Hofmann and P. Mathé, *Parameter choice in Banach space regularization under variational inequalities*, Inverse Problems, 28 (2012), Art. Id. 104006 (17 pages).

[14] B. Jin and P. Maass, *Sparsity regularization for parameter identification problems*, Inverse Problems, 28 (2012), Art. Id. 123001 (70 pages).

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

[16] A. K. Louis, *Inverse und Schlecht Gestellte Probleme*, Teubner, Stuttgart, 1989.

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