Convergence rates in ℓ^1-regularization if the sparsity assumption fails

Martin Burger[1], Jens Flemming[2], Bernd Hofmann[2]

[1]Westfälische Wilhelms-Universität Münster, Institut für Numerische und Angewandte Mathematik, Einsteinstr. 62, 48149 Münster, Germany. [2]Technische Universität Chemnitz, Fakultät für Mathematik, 09107 Chemnitz, Germany

E-mail: `martin.burger@wwu.de`,`jens.flemming@mathematik.tu-chemnitz.de`, `hofmannb@mathematik.tu-chemnitz.de`

Abstract. Variational sparsity regularization based on ℓ^1-norms and other nonlinear functionals has gained enormous attention recently, both with respect to its applications and its mathematical analysis. A focus in regularization theory has been to develop error estimation in terms of regularization parameter and noise strength. For this sake specific error measures such as Bregman distances and specific conditions on the solution such as source conditions or variational inequalities have been developed and used.

In this paper we provide, for a certain class of ill-posed linear operator equations, a convergence analysis that works for solutions that are not completely sparse, but have a fast decaying nonzero part. This case is not covered by standard source conditions, but surprisingly can be treated with an appropriate variational inequality. As a consequence the paper also provides the first examples where the variational inequality approach, which was often believed to be equivalent to appropriate source conditions, can indeed go farther than the latter.

1. Introduction

Variational problems of the form

$$\frac{1}{2}\|Ax - y^\delta\|^2 + \alpha\|x\|_{\ell^1(\mathbb{N})} \to \min_{x \in \ell^1(\mathbb{N})} \tag{1.1}$$

have become an important and standard tool in the regularization of operator equations $Ax = y$. In the field of compressed sensing, with a usually not injective but quite well-conditioned finite-dimensional operator A the sparsity modeled via the ℓ^1-minimization yields the appropriate prior knowledge to uniquely restore the desired solution. In inverse problems, with usually an infinite-dimensional compact operator A,

the sparsity prior allows for stable reconstructions even in presence of noise. There is a comprehensive literature concerning the ℓ^1-regularization of ill-posed problems under sparsity constraints including assertions on convergence rates (cf. e.g. [7] and [2, 4, 6, 13, 15, 16, 22, 23, 24, 25, 26, 27]).

A natural question arising in problems of this type is the asymptotic analysis of such variational problems as $\alpha \to 0$ respectively $y^\delta \to y$, where y are the data that would be produced by an exact solution of the problem. While it is straight-forward to show convergence of subsequences in weak topologies, quantitative error estimates are much more involved. For a long time it was even an open question what are the right conditions and error measures to perform such an analysis. In the last years, following [5], the use of Bregman distances has been evolved as a standard tool. First estimates were based on source conditions of the form (cf. [6] for the ℓ^1-case and [23] for further analysis in a compressed sensing framework)

$$\exists w : \quad A^*w \in \partial\|x^\dagger\|_{\ell^1(\mathbb{N})}, \tag{1.2}$$

where $x^\dagger$ is the exact solution. Extensions to approximate source conditions (cf. [18]) and a different, but seemingly equivalent approach based on variational inequalities (cf. [10]) have been developed subsequently. In the case of ℓ^1-regularization it has been shown that source conditions are directly related to sparsity, hence error estimates have been derived with constants depending on the sparsity level.

However, one can also consider solutions which are only merely sparse, i.e. few of the components being large and the majority being small and decaying fast to zero. Such a model is actually more realistic in many cases, e.g. when applying wavelets to audio signals or natural images. It is the basis of compression algorithms that most of the coefficients are very small and can be ignored, i.e. set to zero. In inversion methods, this type of a-priori information can be analyzed for ℓ^1-regularization with two perspectives. The first is to assume that the relevant solution is indeed sparse, i.e. we are interested in a sparse approximation $\tilde{x}^\dagger$ to $x^\dagger$. In such a case one should clearly analyze a systematic error $A(\tilde{x}^\dagger - x^\dagger)$ in addition to the usual error, which is however not straightforward under general assumptions. The second approach, which we will adopt in this paper, is to really analyze the error in approximating $x^\dagger$ using a natural separation into the (few) large and the remaining small entries.

The further goal of the paper is twofold. On the one hand, we are going to derive novel convergence rates for Tikhonov regularized solutions of linear ill-posed operator equations, where the penalty functional is the ℓ^1-norm. We will prove convergence rates when the exact solution is not sparse but in $\ell^1(\mathbb{N})$. Moreover, we will formulate the specific manifestations of solution smoothness in this case, also essentially based on the decay rate of solution components. On the other hand, we give a first example for an application of the variational inequality approach (see for details [9, 14, 19, 20, 21]) when neither source conditions nor approximate source conditions in the Banach space setting (cf. [28, Section 3.2.2]) are available. The necessity of exploiting source conditions in the course of constructing variational inequalities for obtaining convergence rates in

Tikhonov regularization was up to now considered as a weakness of the approach.

The paper is organized as follows: In Section 2 we will fix the basic problem setup, notations, and assumptions, and then proceed to an overview of smoothness conditions for proving convergence rates of variational regularizations in Section 3. In Section 4 we verify that non-sparse signals indeed fail to satisfy source conditions and even approximate source conditions. In Section 5 we derive a new variational inequality and use it to prove convergence rates for approximately sparse solutions.

2. Problem Setting and Assumptions

Let $\widetilde{A} \in \mathcal{L}(\widetilde{X}, Y)$ be an injective and bounded linear operator mapping between the infinite dimensional Banach spaces $\widetilde{X}$ and Y with norms $\|\cdot\|_{\widetilde{X}}$ and $\|\cdot\|_Y$ as well as with topological duals $\widetilde{X}^*$ and Y^*, respectively, where we assume that the *range* of $\widetilde{A}$ is *not closed*. This means that we have $\mathcal{R}(\widetilde{A}) \neq \overline{\mathcal{R}(\widetilde{A})}^Y$, which is equivalent to the fact that the inverse

$$\widetilde{A}^{-1} : \mathcal{R}(\widetilde{A}) \subset Y \to \widetilde{X}$$

is an unbounded linear operator and that there is no constant $0 < C < \infty$ such that

$$\|\widetilde{x}\|_{\widetilde{X}} \leq C \|\widetilde{A}\widetilde{x}\|_Y \quad \text{for all} \quad \widetilde{x} \in X.$$

Moreover, we denote by $\{u_k\}_{k \in \mathbb{N}} \subset \widetilde{X}$ a *bounded Schauder basis* in $\widetilde{X}$. This means that there is some $K > 0$ such that

$$\|u_k\|_{\widetilde{X}} \leq K \quad \text{for all} \quad k \in \mathbb{N} \tag{2.1}$$

and any element $\widetilde{x} \in \widetilde{X}$ can be represented as an infinite series $\widetilde{x} = \sum_{k=1}^{\infty} x_k u_k$ convergent in $\widetilde{X}$ with uniquely determined coefficients $x_k \in \mathbb{R}$ in the sense of $\lim_{n \to \infty} \|\widetilde{x} - \sum_{k=1}^{n} x_k u_k\|_{\widetilde{X}} = 0$. In the sequel we always consider the coefficients x_k, $k = 1, 2, ...$, as components of an infinite real sequence $x := (x_1, x_2, ...)$ and following the setting in [11] we assume that $L : \ell^1(\mathbb{N}) \to \widetilde{X}$ is the *synthesis operator* defined as

$$Lx := \sum_{k=1}^{\infty} x_k u_k \in \widetilde{X} \quad \text{for} \quad x = (x_1, x_2, ...) \in \ell^1(\mathbb{N}).$$

Evidently, L is a well-defined, injective and, as one simply verifies, also bounded linear operator, i.e. $L \in \mathcal{L}(\ell^1(\mathbb{N}), \widetilde{X})$.

As usual

$$\|x\|_{\ell^q(\mathbb{N})} := \left(\sum_{k=1}^{\infty} |x_k|^q \right)^{1/q}$$

describe the norms in the Banach spaces $\ell^q(\mathbb{N})$, $1 \leq q < \infty$, and $\|x\|_{\ell^\infty(\mathbb{N})} := \sup_{k \in \mathbb{N}} |x_k|$ the norm in $\ell^\infty(\mathbb{N})$. The same norm $\|x\|_{c_0} := \sup_{k \in \mathbb{N}} |x_k|$ is used for the Banach space c_0

of infinite sequences tending to zero. By $\ell^0(\mathbb{N})$ we denote the set of *sparse* sequences, where $x_k \neq 0$ only occurs for a finite number of components.

In this paper, the focus is on elements $\widetilde{x} = Lx \in \widetilde{X}$ which correspond to sequences $x \in \ell^1(\mathbb{N})$ and we choose the Schauder basis $\{u_k\}_{k\in\mathbb{N}}$ such that

$$\overline{\mathcal{R}(L)}^{\widetilde{X}} = \widetilde{X}. \tag{2.2}$$

Precisely, we set

$$X := \ell^1(\mathbb{N}), \qquad A := \widetilde{A} \circ L \in \mathcal{L}(\ell^1(\mathbb{N}), Y),$$

and hence use an operator composition as has been introduced in [29]. Noting that A is also injective since $\widetilde{A}$ and L are, we are searching for *stable approximations* to elements $x \in \ell^1(\mathbb{N})$ satisfying the linear operator equation

$$Ax = y, \qquad x \in X, \quad y \in Y, \tag{2.3}$$

in an approximate manner, where instead of the exact right-hand side $y \in \mathcal{R}(A)$ only *noisy data* $y^\delta \in Y$ satisfying the noise model

$$\|y - y^\delta\|_Y \leq \delta \tag{2.4}$$

with noise level $\delta > 0$ are given.

Proposition 2.1 *Under the given setting including the conditions (2.1) and (2.2) the linear operator equation (2.3) is ill-posed, i.e., we have $\mathcal{R}(A) \neq \overline{\mathcal{R}(A)}^Y$.*

Proof. By the continuity of $\widetilde{A}$ and by (2.2) $\mathcal{R}(A)$ is dense in $R(\widetilde{A})$. If $\mathcal{R}(A)$ would be closed then we had $\mathcal{R}(A) = \mathcal{R}(\widetilde{A})$ and hence $\mathcal{R}(\widetilde{A})$ would be closed, too, which contradicts our assumptions. $\qquad\square$

In the sequel, let $\langle v^*, v \rangle_{B^* \times B}$ denote the dual paring of an element v from the Banach space B with an element v^* from its dual space B^*. Furthermore, we denote by $e_k := (0, 0, ..., 0, 1, 0, ...)$, with 1 at the k-th position for $k = 1, 2, ...$, the elements of the normalized canonical Schauder basis in $\ell^q(\mathbb{N})$, $1 \leq q < \infty$, which is also a Schauder basis in c_0. That means, we find $\lim_{n\to\infty} \|x - \sum_{k=1}^n x_k e_k\|_{c_0} = 0$ for all $x = (x_1, x_2, ...) \in c_0$ and $\lim_{n\to\infty} \|x - \sum_{k=1}^n x_k e_k\|_{\ell^q(\mathbb{N})} = 0$ for all $x = (x_1, x_2, ...) \in \ell^q(\mathbb{N})$, $1 \leq q < \infty$, but not for $q = \infty$. Moreover, we suppose that the following standing assumptions hold:

Assumption 2.2

(a) *The element $x^\dagger \in \ell^1(\mathbb{N})$ solves the operator equation (2.3).*

(b) *We have the limit condition $\lim_{k\to\infty} \|Ae_k\|_Y = 0$.*

(c) *For all $k \in \mathbb{N}$ there exist $f_k \in Y^*$ such that $e_k = A^* f_k$, i.e., we have*
$$x_k = \langle e_k, x \rangle_{\ell^\infty(\mathbb{N}) \times \ell^1(\mathbb{N})} = \langle f_k, Ax \rangle_{Y^* \times Y} \quad \text{for all } x = (x_1, x_2, ...) \in \ell^1(\mathbb{N}).$$

Remark 2.3 One simply sees that Assumption 2.2 is appropriate for the introduced model of the linear operator equation (2.3) with injective forward operator $A = \widetilde{A} \circ L$ and its ill-posedness verified by Proposition 2.1. Firstly, by item (c) the operator

$A : \ell^1(\mathbb{N}) \to Y$ is injective. Namely, we have $x_k = \langle e_k, x \rangle_{\ell^\infty(\mathbb{N}) \times \ell^1(\mathbb{N})} = 0$ for all $k \in \mathbb{N}$ whenever $x = (x_1, x_2, ...) \in \ell^1(\mathbb{N})$ and $Ax = 0$. This yields $x = 0$ and hence the injectivity of A. From the injectivity of A, however, we have that $x^\dagger$ from item (a) of Assumption 2.2 is the uniquely determined solution of (2.3) for given $y \in \mathcal{R}(A)$. Secondly, from item (b) of Assumption 2.2 it follows that there is no constant $0 < C < \infty$ such that $1 = \|e_k\|_{\ell^1(\mathbb{N})} \leq C \|Ae_k\|_Y$ for all $k \in \mathbb{N}$ and hence we have that $\mathcal{R}(A)$ is a non-closed subset of the Banach space Y. Consequently, the linear operator equation (2.3) is ill-posed under Assumption 2.2 and the inverse operator $A^{-1} : \mathcal{R}(A) \subset Y \to \ell^1(\mathbb{N})$, which exists due to the injectivity of A, is an unbounded linear operator. Hence, the stable approximate solution of (2.3) based on noisy data $y^\delta \in Y$ satisfying (2.4) requires some kind of regularization. □

Note that by the closed range theorem the range $\mathcal{R}(A^*)$ is a non-closed subset of $\ell^\infty(\mathbb{N})$, but not dense in $\ell^\infty(\mathbb{N})$, since it is a subset of c_0, as the following proposition indicates, and c_0 is not dense in $\ell^\infty(\mathbb{N})$ with respect to the supremum norm.

Proposition 2.4 *Under Assumption 2.2 we have* $\overline{\mathcal{R}(A^*)}^{\ell^\infty(\mathbb{N})} = c_0$.

Proof. First we show that $\mathcal{R}(A^*) \subseteq c_0$. For $w \in Y^*$ we obvioulsy have $A^*w \in \ell^\infty(\mathbb{N})$ and

$$|[A^*w]_k| = |\langle A^*w, e_k \rangle_{\ell^\infty(\mathbb{N}) \times \ell^1(\mathbb{N})}| = |\langle w, Ae_k \rangle_{Y^* \times Y}| \leq \|w\|_{Y^*} \|Ae_k\|_Y.$$

Thus by Assumption 2.2 (b), $A^*w \in c_0$.

It remains to show that each $z \in c_0$ can be approximated by a sequence $\{A^*w_n\}_{n \in \mathbb{N}}$ with $w_n \in Y^*$. For this purpose we define $w_n = \sum_{k=1}^{n} z_k f_k$ with f_k from Assumption 2.2 (c). Then

$$\|z - A^*w_n\|_{\ell^\infty(\mathbb{N})} = \left\| \sum_{k=n+1}^{\infty} z_k e_k \right\|_{\ell^\infty(\mathbb{N})}$$

and therefore $\|z - A^*w_n\|_{\ell^\infty(\mathbb{N})} \to 0$ as $n \to \infty$, since $z \in c_0$. □

Remark 2.5 An inspection of the proof shows that for Proposition 2.4 the condition (b) of Assumption 2.2 can be weakened to

(b') We have weak convergence $Ae_k \rightharpoonup 0$ in Y as $k \to \infty$.

Further, item (c) of Assumption 2.2 implies that

$$|x_k| \leq \|f_k\|_{Y^*} \|Ax\|_Y \qquad \text{for all} \quad x \in \ell^1(\mathbb{N}) \quad \text{and} \quad k \in \mathbb{N}. \tag{2.5}$$

Applying this inequality to $x = e_k$ we obtain $1 \leq \|f_k\|_{Y^*} \|Ae_k\|_Y$ and therefore on the one hand

$$\|f_k\|_{Y^*} \geq \frac{1}{\|A\|} \tag{2.6}$$

and on the other hand by exploiting item (b)

$$\lim_{k \to \infty} \|f_k\|_{Y^*} \geq \lim_{k \to \infty} \frac{1}{\|Ae_k\|_Y} = \infty. \tag{2.7}$$

If $\ell_g^\infty(\mathbb{N})$ denotes the weighted ℓ^∞-space with the positive weights $g = (1/\|f_1\|_{Y^*}, 1/\|f_2\|_{Y^*}, ...)$ and norm

$$\|x\|_{\ell_g^\infty(\mathbb{N})} := \sup_{k \in \mathbb{N}} \frac{|x_k|}{\|f_k\|_{Y^*}},$$

from (2.6) we have that

$$\|x\|_{\ell_g^\infty(\mathbb{N})} \leq \|A\| \|x\|_{\ell^\infty(\mathbb{N})} \qquad \text{for all} \quad x \in \ell^\infty(\mathbb{N})$$

and from (2.5) that

$$\|x\|_{\ell_g^\infty(\mathbb{N})} \leq \|Ax\|_Y \qquad \text{for all} \quad x \in \ell^1(\mathbb{N})$$

Hence the norm in $\ell_g^\infty(\mathbb{N})$ is weaker than the standard supremum norm in $\ell^\infty(\mathbb{N})$ and we have $0 < C < \infty$ such that $\|Gx\|_{\ell_g^\infty(\mathbb{N})} \leq C \|Ax\|_Y$ for all $x \in \ell^1(\mathbb{N})$, where G is the embedding operator from $\ell^1(\mathbb{N})$ to $\ell_g^\infty(\mathbb{N})$ the behaviour of which characterizes the nature of ill-posedness of problem (2.3), and we refer to [22, Remark 3.5] for similar considerations in ℓ^q-regularization. □

Example 2.6 (diagonal operator) In order to get some more insight into details, we consider the situation of singular value decomposition (SVD) for $\widetilde{A}$, where $\widetilde{X}$ and Y are infinite dimensional *separable Hilbert spaces* and $\widetilde{A} : \widetilde{X} \to Y$ is a *compact* linear operator with the singular system $\{\sigma_k, u_k, v_k\}_{k \in \mathbb{N}}$. As usual $\{\sigma_k\}_{k \in \mathbb{N}}$ denotes the decreasingly ordered sequence of positive singular values tending to zero as $k \to \infty$ and $\{u_k\}_{k \in \mathbb{N}}$ as well as $\{v_k\}_{k \in \mathbb{N}}$ are complete orthonormal systems in $\widetilde{X}$ and $\overline{\mathcal{R}(A)}^Y$, respectively, such that we have $\widetilde{A}u_k = \sigma_k v_k$ and $\widetilde{A}^* v_k = \sigma_k u_k$ for all $k \in \mathbb{N}$. The adjoint of $\widetilde{A}$ maps as $\widetilde{A}^* : Y \to \widetilde{X}$, since we consider Hilbert spaces and can identify the dual spaces with the original ones. Moreover, we denote by $\langle \cdot, \cdot \rangle_{\widetilde{X}}$ and $\langle \cdot, \cdot \rangle_Y$ the inner products in $\widetilde{X}$ and Y.

Using the specific orthonormal system $\{u_k\}_{k \in \mathbb{N}}$ from the singular system of $\widetilde{A}$ as Schauder basis in $\widetilde{X}$, i.e., for any $\widetilde{x} \in \widetilde{X}$ we have $\widetilde{x} = \sum_{k \in N} x_k u_k$ with Fourier coefficients $x_k = \langle \widetilde{x}, u_k \rangle_{\widetilde{X}}$ as components of the infinite sequence $x = \{x_k\}_{k \in \mathbb{N}} \in \ell^2(\mathbb{N})$, we obtain a compact and injective linear operator $A : X = \ell^1(\mathbb{N}) \to Y$ with *diagonal structure* due to $Ae_k = \widetilde{A}u_k$, $k \in \mathbb{N}$, characterized by $Ax = \sum_{k \in \mathbb{N}} x_k \sigma_k v_k$ for all $x \in \ell^1(\mathbb{N})$. Namely, we have compactness and injectivity of $A = \widetilde{A} \circ L$ since the injective operator $\widetilde{A}$ is compact and the linear synthesis operator $L : \ell^1(\mathbb{N}) \to \widetilde{X}$ is bounded and injective as the composition $L = \mathcal{U} \circ \mathcal{E}$ of the bounded injective embedding operator $\mathcal{E}$ from $\ell^1(\mathbb{N})$ to $\ell^2(\mathbb{N})$ with $\overline{\mathcal{R}(\mathcal{E})}^{\ell^2(\mathbb{N})} = \ell^2(\mathbb{N})$ and the unitary operator $\mathcal{U}$ coming from the isometry between $\ell^2(\mathbb{N})$ and $\widetilde{X}$. We have $\overline{\mathcal{R}(L)}^{\widetilde{X}} = \widetilde{X}$ and note that both conditions (2.1) and (2.2) are satisfied for that example.

It is important to mention that also the items (b) and (c) of Assumption 2.2 are satisfied. Item (b) is evident because of $\|Ae_k\|_Y = \|\widetilde{A}u_k\|_{\widetilde{X}} = \sigma_k \to 0$ as $k \to \infty$. For the operator $A^* : Y \to X^* = \ell^\infty(\mathbb{N})$ we have for all $i, k \in \mathbb{N}$ by definition

$$\langle A^* v_k, e_i \rangle_{\ell^\infty(\mathbb{N}) \times \ell^1(\mathbb{N})} = \langle v_k, Ae_i \rangle_Y = \langle v_k, \widetilde{A}u_i \rangle_Y = \sigma_i \langle v_k, v_i \rangle_Y = \begin{cases} \sigma_k & \text{if} \quad k = i \\ 0 & \text{if} \quad k \neq i \end{cases} .$$

This can be rewritten as $e_k = A^* f_k$ for $f_k = \frac{1}{\sigma_k} v_k$ and yields item (c). We still note that $\|f_k\|_Y = \frac{1}{\sigma_k}$ tends to infinity as $k \to \infty$. $\qquad \square$

Our focus in on situations where we conjecture that the solutions of (2.3) are sparse, i.e. $x^\dagger \in \ell^0(\mathbb{N})$, or at least that the coefficients $x_k^\dagger$ in $x^\dagger \in \ell^1(\mathbb{N})$ are negligible for sufficiently large $k \in \mathbb{N}$. Then it makes sense to use the ℓ^1-*regularization*, and the regularized solutions $x_\alpha^\delta \in \ell^1(\mathbb{N})$ are minimizers of the extremal problem

$$T_\alpha(x) := \frac{1}{p}\|Ax - y^\delta\|_Y^p + \alpha \, \|x\|_{\ell^1(\mathbb{N})} \to \min_{x \in X = \ell^1(\mathbb{N})} \tag{2.8}$$

of Tikhonov type with regularization parameters $\alpha > 0$ and a misfit norm exponent $1 < p < \infty$. Then the sublinear and continuous penalty functional $\Omega(x) := \|x\|_{\ell^1(\mathbb{N})}$ possessing finite values for all $x \in X$ is convex. Since $X = \ell^1(\mathbb{N})$ is a *non-reflexive* Banach space we need some topology τ_X in X which is weaker than the canonical weak topology in X in order to ensure stabilizing properties of the regularized solutions. In other words, Ω must be a *stabilizing functional* in the sense that the sublevel sets

$$\mathcal{M}^\Omega(c) := \{x \in X : \Omega(x) \le c\}$$

are τ_X-sequentially precompact subsets of X for all $c \ge 0$. Since $Z := c_0$ with $Z^* = \ell^1(\mathbb{N})$ is a separable predual Banach space of X, we can use the associated weak*-topology as τ_X (cf. [28, Remark 4.9 and Lemma 4.10]). Note that Ω under consideration here is also sequentially lower semi-continuous with respect to the weak*-topology. If the operator A can be proven to transform weak*-convergent sequences in $\ell^1(\mathbb{N})$ to weakly convergent sequences in the Banach space Y, which will be done by Lemma 2.7 below, then existence and stability and weak*-convergence of regularized solutions can be ensured. We will summarize this in Proposition 2.8.

Lemma 2.7 *Under the assumptions stated above the operator $A : \ell^1(\mathbb{N}) \to Y$ is always sequentially weak*-to-weak continuous, i.e., for a weak* convergent sequence $x^{(n)} \rightharpoonup^* \overline{x}$ in ℓ^1 we have weak convergence as $Ax^{(n)} \rightharpoonup A\overline{x}$ in Y.*

Proof. Since the separable Banach space c_0, which has the same supremum norm like ℓ^∞, is a predual space of ℓ^1, i.e. any element $x = (x_1, x_2, ...) \in \ell^1$ is a bounded linear functional on c_0, the weak*-convergence $x^{(n)} \rightharpoonup^* \overline{x}$ in ℓ^1 can be written as

$$\lim_{n \to \infty} \langle x^{(n)}, g \rangle_{\ell^1 \times c_0} = \lim_{n \to \infty} \sum_{k \in \mathbb{N}} x_k^{(n)} g_k = \sum_{k \in \mathbb{N}} \overline{x}_k g_k = \langle \overline{x}, g \rangle_{\ell^1 \times c_0} \quad \text{for all} \quad g = (g_1, g_2, ...) \in c_0.$$

With the bounded linear operator $A^* : Y^* \to \ell^\infty$ we can further conclude from $\mathcal{R}(A^*) \subseteq c_0$, which follows from Proposition 2.4 , that $A^* f \in c_0$ for all $f \in Y^*$ and that

$$\langle f, Ax^{(n)} \rangle_{Y^* \times Y} = \langle A^* f, x^{(n)} \rangle_{\ell^\infty, \ell^1} = \langle x^{(n)}, A^* f \rangle_{\ell^1 \times c_0} \quad \text{for all} \quad f \in Y^*.$$

Hence we have

$$\lim_{n \to \infty} \langle f, Ax^{(n)} \rangle_{Y^* \times Y} = \lim_{n \to \infty} \langle x^{(n)}, A^* f \rangle_{\ell^1 \times c_0} = \langle \overline{x}, A^* f \rangle_{\ell^1 \times c_0} = \langle f, A\overline{x} \rangle_{Y^* \times Y} \quad \text{for all} \quad f \in Y^*,$$

which yields the weak convergence in Y and completes the proof. $\qquad \square$

Proposition 2.8 *For all $\alpha > 0$ and all $y^\delta \in Y$ there exist uniquely determined minimizers $x_\alpha^\delta \in \ell^1(\mathbb{N})$ of the Tikhonov functional T_α from (2.8). These regularized solutions satisfy the condition $x_\alpha^\delta \in \ell^0(\mathbb{N})$, and they are stable with respect to the data, i.e., small perturbations in y^δ in the norm topology of Y lead only to small changes in x_α^δ with respect to the weak*-topology in $\ell^1(\mathbb{N})$. If $\delta_n \to 0$ and if the regularization parameters $\alpha_n = \alpha(\delta_n, y^{\delta_n})$ are chosen such that $\alpha_n \to 0$ and $\frac{\delta_n^p}{\alpha_n} \to 0$ as $n \to \infty$, then $x_{\alpha_n}^{\delta_n}$ converges in the weak*-topology of $\ell^1(\mathbb{N})$ to the uniquely determined solution $x^\dagger$ of the operator equation (2.3) and moreover we have $\lim_{n\to\infty} \|x_{\alpha_n}^{\delta_n}\|_{\ell^1(\mathbb{N})} = \|x^\dagger\|_{\ell^1(\mathbb{N})}$.*

Proof. Taking into account Lemma 2.7 the properties of existence, stability and convergence of regularized solution follow directly from the general theory of Tikhonov regularization in Banach spaces (cf., e.g., [19, §3], [28, Section 4.1] and [6, 11]). The regularized solutions x_α^δ are uniquely determined for all $\alpha > 0$ and $y^\delta \in Y$, because the Tikhonov functional T_α is strictly convex due to the injectivity of A. Since also the solution $x^\dagger$ of the operator equation (2.3) is uniquely determined for injective A, all weak*-convergent subsequences of $x_{\alpha_n}^{\delta_n}$ have $x^\dagger$ as common limit element. The last missing fact that $x_\alpha^\delta \in \ell^0(\mathbb{N})$ for fixed $\alpha > 0$ is proven in [22, Lemma 2.1] if Y is a Hilbert space. For Banach spaces Y the proof is analogous if one observes that for each minimizer x_α^δ there is some $\xi \in \ell^\infty(\mathbb{N})$ such that

$$\xi \in A^* \partial \left(\frac{1}{p} \| \cdot - y^\delta \|_Y^p \right) (Ax_\alpha^\delta) \subseteq \mathcal{R}(A^*) \subseteq c_0 \quad \text{and} \quad -\xi \in \partial(\alpha \| \cdot \|_{\ell^1(\mathbb{N})})(x_\alpha^\delta).$$

□

Remark 2.9 Since the Banach space $\ell^1(\mathbb{N})$ fails to satisfy the weak* Kadec-Klee property, we cannot conclude from $x_{\alpha_n}^{\delta_n} \rightharpoonup^* x^\dagger$ and $\lim_{n\to\infty} \|x_{\alpha_n}^{\delta_n}\|_{\ell^1(\mathbb{N})} = \|x^\dagger\|_{\ell^1(\mathbb{N})}$, both claimed and proven in the context of Proposition 2.8, that we have norm convergence

$$\lim_{n\to\infty} \|x_{\alpha_n}^{\delta_n} - x^\dagger\|_{\ell^1(\mathbb{N})} = 0. \tag{2.9}$$

However, we will prove convergence rates in Section 5 which directly imply (2.9). To understand the role of stated assumptions one should recall that convergence of Tikhonov regularized solutions, here in a weak* sense, can be stated if the misfit term and the penalty term possess the corresponding properties, whereas *convergence rates* can only be proven if we have some kind of *smoothness of the solution $x^\dagger$ with respect to the forward operator A*. By inspection of proof details one can see that Proposition 2.8 does not make use of item (c) of Assumption 2.2, whereas all rate results in Section 5 need this item. Indeed, the requirement that e_k belongs to the range of the adjoint operator $A^* : Y^* \to \ell^\infty(\mathbb{N})$ for all $k \in \mathbb{N}$ is the most restrictive one in our setting. An assumption of this nature was already used under sparsity constraints in the papers [4, 11, 22], partly only for all k satisfying the condition $x_k^\dagger \neq 0$. Since our focus is on the case that sparsity fails but $x^\dagger$ is in $\ell^1(\mathbb{N})$, item (c) of Assumption 2.2 together with the decay rate of $|x_k^\dagger| \to 0$ express the solution smoothness with respect to A. This is a somewhat curious form of smoothness expression, because $x^\dagger$ and A are only coupled by

taking the detour through the Schauder basis $\{u_k\}_{k\in\mathbb{N}}$ which prescribes the properties of the unit elements e_k.

One easily shows that item (c) of Assumption 2.2 is fulfilled if there is a multiplication operator $D : \ell^1(\mathbb{N}) \to \ell^1(\mathbb{N})$ defined as $[Dx]_k := d_k x_k$ with $|d_k| > 0$ for all $k \in \mathbb{N}$ such that

$$\|Ax\|_Y \geq \|Dx\|_{\ell^1(\mathbb{N})} \qquad \text{for all} \quad x \in \ell^1(\mathbb{N}). \tag{2.10}$$

Even if the condition (2.10) seems to be not too far from the diagonal case mentioned in Example 2.6 and mimics in some sense a singular value decomposition, this situation is matched by well-studied wavelet or wavelet-vaguelette approaches to inverse problems. To date we have no approach to norm convergence (2.9) and rates in ℓ^1-regularization when both sparsity assumption and item (c) of Assumption 2.2 fail.

3. Manifestations of smoothness for convergence rates

It is well-known that for linear ill-posed operator equations (2.3) with $A \in \mathcal{L}(X, Y)$, formulated in Banach spaces X and Y with some solution $x^\dagger \in X$, *convergence rates of regularized solutions*

$$E(x^\delta_{\alpha(\delta,y^\delta)}, x^\dagger) = \mathcal{O}(\varphi(\delta)) \qquad \text{as} \qquad \delta \to 0 \tag{3.1}$$

evaluated by some *nonnegative error measure* E and some *index function* φ require additional properties of $x^\dagger$ which express some kind of smoothness of the solution with respect to the forward operator $A : X \to Y$ and its adjoint $A^* : Y^* \to X^*$. We call a function $\varphi : (0, \infty) \to (0, \infty)$ index function if it is continuous, strictly increasing and satisfies the limit condition $\lim_{t\to+0} \varphi(t) = 0$. Moreover, we denote by x^δ_α the minimizers of

$$\frac{1}{p}\|Ax - y^\delta\|_Y^p + \alpha\,\Omega(x) \to \min_{x\in X},$$

for $1 < p < \infty$ and some convex stabilizing penalty functional $\Omega : X \to [0, \infty]$. Then the original form of smoothness is given by *source conditions*

$$\xi = A^*w, \qquad w \in Y^*, \tag{3.2}$$

for subgradients $\xi \in \partial\Omega(x^\dagger) \subseteq X^*$, and the error can be measured by the Bregman distance

$$E(x, x^\dagger) := \Omega(x) - \Omega(x^\dagger) - \langle\xi, x - x^\dagger\rangle_{X^*\times X} \tag{3.3}$$

as introduced in [5]. Then convergence rates (3.1) with $\varphi(t) = t$ can be derived under appropriate choices of the regularization parameter $\alpha > 0$ (cf. [5, 19, 27]).

If the subgradient $\xi \in X^*$ fails to satisfy (3.2), then one can use *approximate source conditions* and ask for the degree of violation of ξ with respect to the benchmark source condition (3.2). This violation is, for example, expressed by properties of the strictly positive, convex and nonincreasing distance function

$$d_\xi(R) := \inf_{w\in Y^*:\,\|w\|_{Y^*}\leq R} \|\xi - A^*w\|_{X^*}. \tag{3.4}$$

If the limit condition

$$\lim_{R \to \infty} d_\xi(R) = 0 \tag{3.5}$$

holds, then one can prove convergence rates (3.1) with E from (3.3) and φ depending on d_ξ (cf. [3, 17], [28, Chapter 3], and [20, Appendix A]). If, for example, the Bregman distance is q-coercive with $q \geq 2$ and $1/q + 1/q^* = 1$, then we have

$$\varphi(t) = \left[d_\xi\left(\Phi^{-1}(t)\right)\right]^{q^*}, \qquad \text{where} \quad \Phi(R) := \frac{[d_\xi(R)]^{q^*}}{R}.$$

If, however, the distance function d_ξ does not satisfy (3.5), this approach fails. As mentioned in [3] such situation is only possible if the biadjoint operator $A^{**} : X^{**} \to Y^{**}$ mapping between the bidual spaces of X and Y is not injective.

An alternative manifestation of solution smoothness is given by *variational inequalities* (cf., e.g., [19, 21]), where in the sequel our focus will be on the variant formulated in Assumption 3.1, originally suggested in [9, 12].

Assumption 3.1 *For given nonnegative error measure E, convex stabilizing functional Ω and $x^\dagger \in X$, let there exist a concave index function φ, a set $\mathcal{M} \subseteq X$ with $x^\dagger \in \mathcal{M}$ and constants $\beta > 0$ as well as $C > 0$ such that*

$$\beta \, E(x, x^\dagger) \leq \Omega(x) - \Omega(x^\dagger) + C \, \varphi\left(\|A(x - x^\dagger)\|_Y\right) \quad \text{for all} \quad x \in \mathcal{M}. \tag{3.6}$$

In the case of an *appropriate choice of the regularization parameter* $\alpha > 0$ the index function φ in the variational inequality (3.6) immediately determines the convergence rate of Tikhonov-regularized solutions x_α^δ to $x^\dagger$. Proofs of the assertions of the following proposition can be found in [20, Theorem 1] and [9, Chapter 4]. We also refer to [1].

Proposition 3.2 *Let the regularization parameter be chosen a priori as $\alpha = \alpha(\delta) := \frac{\delta^p}{\varphi(\delta)}$ or a posteriori as $\alpha = \alpha(\delta, y^\delta)$ according to the strong discrepancy principle*

$$\tau_1 \, \delta \leq \|A x_\alpha^\delta - y^\delta\|_Y \leq \tau_2 \, \delta \tag{3.7}$$

for prescribed constants $1 \leq \tau_1 \leq \tau_2 < \infty$. Then under Assumption 3.1 we have the convergence rate (3.1) whenever in both cases all regularized solutions x_α^δ for sufficiently small $\delta > 0$ belong to $\mathcal{M}$.

Remark 3.3 As was shown in [20] the strong discrepancy principle (3.7) in Proposition 3.2 can be replaced with the more traditional (sequential) discrepancy principle, where with sufficiently large $\alpha_0 > 0$, $0 < \zeta < 1$, and

$$\Delta_\zeta := \{\alpha_j : \alpha_j := \zeta^j \alpha_0, \; j = 1, 2, ...\}$$

the regularization parameter $\alpha = \alpha(\delta, y^\delta)$ is chosen as the largest parameter within Δ_ζ such that $\|A x_\alpha - y^\delta\|_Y \leq \tau \delta$ for prescribed $\tau > 1$. This, however, is more of interest if the forward operator is nonlinear and duality gaps can hinder regularization parameters $\alpha = \alpha(\delta, y^\delta)$ to fulfil (3.7) simultaneously with lower and upper bounds. □

4. Failure of approximate source conditions

Now we return to the situation $X = \ell^1(\mathbb{N})$ of ℓ^1-regularization introduced in Section 2. Precisely, in the notation of Section 3 we consider $\Omega(x) = \|x\|_{\ell^1(\mathbb{N})}$ and pose the following additional assumption.

Assumption 4.1 *Let the solution* $x^\dagger = (x_1^\dagger, x_2^\dagger, ...) \in \ell^1(\mathbb{N})$ *of (2.3) be non-sparse, i.e.,* $x^\dagger \notin \ell^0(\mathbb{N})$ *and hence there is an infinite subsequence* $\{x_{k_n}^\dagger \neq 0\}_{n=1}^\infty$ *of nonzero components of* $x^\dagger$.

Lemma 4.2 *Let Assumptions 2.2 and 4.1 hold and let* $\xi \in \ell^\infty(\mathbb{N})$ *be an arbitrary element of* $X^* = \ell^\infty(\mathbb{N})$. *Then on the one hand,* $d_\xi(R) \to 0$ *for* $R \to \infty$ *if and only if* $\xi \in c_0$. *On the other hand, if* $s := \limsup_{k \to \infty} |\xi_k| > 0$ *we have* $d_\xi(R) \geq s$ *for all* $R \geq 0$.

Proof. By definition of the distance function, $d_\xi(R) \to 0$ for $R \to \infty$ if and only if $\xi \in \overline{\mathcal{R}(A^*)}^{\ell^\infty(\mathbb{N})}$. Hence the first part of the assertion is a consequence of Proposition 2.4 where $\overline{\mathcal{R}(A^*)}^{\ell^\infty(\mathbb{N})} = c_0$ was proven.

For proving the second part of the assertion we take a subsequence $\{\xi_{l_n}\}_{n \in \mathbb{N}}$ with $|\xi_{l_n}|_{\ell^\infty(\mathbb{N})} \to s$ as $n \to \infty$ and assume $w \in Y^*$ with $\|w\|_{Y^*} \leq R$. Because of

$$|[A^*w]_{l_n}| \leq \|w\|_{Y^*}\|Ae_{l_n}\|_Y \leq R\|Ae_{l_n}\|_Y \to 0 \quad \text{as} \quad n \to \infty,$$

we see that $\|\xi - A^*w\|_{\ell^\infty(\mathbb{N})} \geq \sup_{n \in \mathbb{N}} |\xi_{l_n} - [A^*w]_{l_n}| \geq s$. Thus, $d_\xi(R) \geq s$ for all $R > 0$. This completes the proof. □

The failure of benchmark source conditions (3.2) for $\xi \in \partial\|x^\dagger\|_{\ell^1(\mathbb{N})}$ under Assumption 4.1 has been mentioned previously by different authors (cf., e.g., [6, 16, 22, 25]), but Proposition 4.3 will show that even approximate source conditions fail.

Proposition 4.3 *Under the Assumptions 2.2 and 4.1 the benchmark source condition (3.2) always fails. Also the method of approximate source conditions is not applicable, because we have for the corresponding distance function*

$$d_\xi(R) = \inf_{w \in Y^*: \|w\|_{Y^*} \leq R} \|\xi - A^*w\|_{\ell^\infty(\mathbb{N})} \geq 1 \qquad \text{for all} \qquad R > 0. \quad (4.1)$$

Proof. As is well-known the subgradients $\xi = (\xi_1, \xi_2, ...) \in \partial\|x\|_{\ell^1(\mathbb{N})}$ can be made explicit as

$$\xi_k \begin{cases} = 1 & \text{if} \quad x_k > 0, \\ \in [-1, 1] & \text{if} \quad x_k = 0, \qquad k \in \mathbb{N}. \\ = -1 & \text{if} \quad x_k < 0, \end{cases}$$

So we have $\xi \notin c_0$ by Assumption 4.1. Moreover, by this assumption we also have $|\xi_{k_n}| = 1$ for all $n \in \mathbb{N}$, that is $\limsup_{k \to \infty} |\xi_k| = 1$. Lemma 4.2 thus shows $d_\xi(R) \geq 1$ for all $R > 0$. This means that $d_\xi(R)$ does not tend to zero as $R \to \infty$ and, in particular, that $\xi \notin \mathcal{R}(A^*)$ as a direct consequence of $\xi \notin c_0$, since this would imply $d_\xi(R) = 0$ for sufficiently large $R > 0$. □

Remark 4.4 Since (4.1) implies that $d_\xi(R) \not\to 0$ as $R \to \infty$ and (3.5) fails, this is an example for the case that the biadjoint operator A^{**} is not injective although A is injective, where we have $A^{**} : (\ell^\infty(\mathbb{N}))^* \to Y^{**}$ here. □

5. Derivation of a variational inequality

As outlined in the previous section source conditions and even approximate source conditions do not hold if the searched for solution $x^\dagger \in \ell^1(\mathbb{N})$ to equation (2.3) is not sparse. In the following we derive a variational inequality as introduced in Assumption 3.1 with $\mathcal{M} = X = \ell^1(\mathbb{N})$ and $\beta = 1$ in the case of a non-sparse solution. By Proposition 3.2 we then directly obtain a convergence rate. Since the index function φ constructed below is not explicitly available for choosing $\alpha > 0$, a posteriori choices of the regularization parameter are to be preferred, and in particular the strong discrepancy principle (3.7) ensures the convergence rate

$$\|x_\alpha^\delta - x^\dagger\|_{\ell^1(\mathbb{N})} = \mathcal{O}(\varphi(\delta)) \qquad \text{as} \qquad \delta \to 0.$$

We need the following lemma.

Lemma 5.1 *For all $x \in \ell^1(\mathbb{N})$ and all $n \in \mathbb{N}$ the inequality*

$$\|x - x^\dagger\|_{\ell^1(\mathbb{N})} - \|x\|_{\ell^1(\mathbb{N})} + \|x^\dagger\|_{\ell^1(\mathbb{N})} \le 2 \left(\sum_{k=n+1}^\infty |x_k^\dagger| + \sum_{k=1}^n |x_k - x_k^\dagger| \right)$$

is true.

Proof. For $n \in \mathbb{N}$ define the projection $P_n : \ell^1(\mathbb{N}) \to \ell^1(\mathbb{N})$ by $P_n x := (x_1, \ldots, x_n, 0, \ldots)$. Then obviously

$$\|x\|_{\ell^1(\mathbb{N})} = \|P_n x\|_{\ell^1(\mathbb{N})} + \|(I - P_n)x\|_{\ell^1(\mathbb{N})}$$

for all $x \in \ell^1(\mathbb{N})$. Based on this equality we see that

$$\begin{aligned}
\|x - &x^\dagger\|_{\ell^1(\mathbb{N})} - \|x\|_{\ell^1(\mathbb{N})} + \|x^\dagger\|_{\ell^1(\mathbb{N})} \\
&= \|P_n(x - x^\dagger)\|_{\ell^1(\mathbb{N})} + \|(I - P_n)x^\dagger\|_{\ell^1(\mathbb{N})} + \|(I - P_n)(x - x^\dagger)\|_{\ell^1(\mathbb{N})} \\
&\quad - \|(I - P_n)x\|_{\ell^1(\mathbb{N})} + \|P_n x^\dagger\|_{\ell^1(\mathbb{N})} - \|P_n x\|_{\ell^1(\mathbb{N})}.
\end{aligned}$$

Consequently, together with the triangle inequalities

$$\|(I - P_n)(x - x^\dagger)\|_{\ell^1(\mathbb{N})} \le \|(I - P_n)x\|_{\ell^1(\mathbb{N})} + \|(I - P_n)x^\dagger\|_{\ell^1(\mathbb{N})}$$

and

$$\|P_n x^\dagger\|_{\ell^1(\mathbb{N})} \le \|P_n(x - x^\dagger)\|_{\ell^1(\mathbb{N})} + \|P_n x\|_{\ell^1(\mathbb{N})}$$

we obtain

$$\|x - x^\dagger\|_{\ell^1(\mathbb{N})} - \|x\|_{\ell^1(\mathbb{N})} + \|x^\dagger\|_{\ell^1(\mathbb{N})} \le 2 \left(\|P_n(x - x^\dagger)\|_{\ell^1(\mathbb{N})} + \|(I - P_n)x^\dagger\|_{\ell^1(\mathbb{N})} \right).$$

□

Theorem 5.2 *The variational inequality (3.6) holds true with $\beta = 1$, $\mathcal{M} = X = \ell^1(\mathbb{N})$, $E(x, x^\dagger) = \|x - x^\dagger\|_{\ell^1(\mathbb{N})}$, $\Omega(x) = \|x\|_{\ell^1(\mathbb{N})}$, and*

$$\varphi(t) = 2 \inf_{n \in \mathbb{N}} \left(\sum_{k=n+1}^{\infty} |x_k^\dagger| + t \sum_{k=1}^{n} \|f_k\|_{Y^*} \right). \tag{5.1}$$

The function φ is a concave index function.

Proof. By Assumption 2.2(c) we have

$$\sum_{k=1}^{n} |x_k - x_k^\dagger| = \sum_{k=1}^{n} |\langle e_k, x - x^\dagger \rangle_{\ell^\infty(\mathbb{N}) \times \ell^1(\mathbb{N})}| \leq \sum_{k=1}^{n} \|f_k\|_{Y^*} \|A(x - x^\dagger)\|_Y$$

for all $n \in \mathbb{N}$ and with the help of Lemma 5.1 we obtain

$$\|x - x^\dagger\|_{\ell^1(\mathbb{N})} - \|x\|_{\ell^1(\mathbb{N})} + \|x^\dagger\|_{\ell^1(\mathbb{N})} \leq 2 \left(\sum_{k=n+1}^{\infty} |x_k^\dagger| + \|A(x - x^\dagger)\|_Y \sum_{k=1}^{n} \|f_k\|_{Y^*} \right)$$

for all $n \in \mathbb{N}$. Taking the infimum over n on the right-hand side yields the desired variational inequality.

It remains to show that φ is a *concave index function*. As an infimum of affine functions φ is concave and upper semi-continuous. Since $\varphi(t) < \infty$ for $t \in [0, \infty)$, φ is even continuous on $(0, \infty)$. Together with $\varphi(0) = 0$ the upper semi-continuity in $t = 0$ yields $\varphi(t) \to 0$ if $t \to +0$, that is, continuity in $t = 0$. To show that φ is strictly increasing take $t_1, t_2 \in [0, \infty)$ with $t_1 < t_2$. Since by (2.7) the infimum in the definition of $\varphi(t_2)$ is attained at some $n_2 \in \mathbb{N}$, we obtain

$$\varphi(t_1) \leq 2 \left(\sum_{k=n_2+1}^{\infty} |x_k^\dagger| + t_1 \sum_{k=1}^{n_2} \|f_k\|_{Y^*} \right) < \varphi(t_2).$$

$\square$

Example 5.3 (Hölder rates) As one sees the rate function φ in Theorem 5.2 depends on decay properties of the solution components $|x_k^\dagger|$ for $k \to \infty$ and on the growth properties of the finite sums $\sum_{k=1}^{n} \|f_k\|_{Y^*}$ for $n \to \infty$. If decay and growth are of monomial type as

$$\sum_{k=n+1}^{\infty} |x_k^\dagger| \leq K_1 \, n^{-\mu}, \qquad \sum_{k=1}^{n} \|f_k\|_{Y^*} \leq K_2 \, n^{\nu}, \qquad \mu, \nu > 0, \tag{5.2}$$

with some constants $K_1, K_2 > 0$, then we have from Theorem 5.2 (see formula (5.1)) and Proposition 3.2 that the strong discrepancy principle for choosing the regularization parameter $\alpha > 0$ yields the *Hölder convergence rates*

$$\|x_\alpha^\delta - x^\dagger\|_{\ell^1(\mathbb{N})} = \mathcal{O}\left(\delta^{\frac{\mu}{\mu+\nu}} \right) \qquad \text{as} \qquad \delta \to 0. \tag{5.3}$$

Whenever exponents $\hat\mu > 1$, $\hat\nu > 0$, and constants $\hat{K}_1, \hat{K}_2 > 0$ exist such that

$$|x_k^\dagger| \leq \hat{K}_1 \, n^{-\hat\mu}, \qquad \|f_k\|_{Y^*} \leq \hat{K}_2 \, n^{\hat\nu}, \qquad \text{for all} \quad k \in \mathbb{N}, \tag{5.4}$$

the rate result (5.3) can be rewritten as

$$\|x_\alpha^\delta - x^\dagger\|_{\ell^1(\mathbb{N})} = \mathcal{O}\left(\delta^{\frac{\hat{\mu}-1}{\hat{\mu}+\hat{\nu}}}\right) \qquad \text{as} \qquad \delta \to 0. \tag{5.5}$$

For the diagonal situation of Example 2.6 in Hilbert spaces (see also Remark 5.5 below) the exponent $\hat{\nu} > 0$ can be interpreted as the corresponding *degree of ill-posedness* expressed by the decay rate $\sigma_n \sim n^{-\hat{\nu}}$ of the singular values of $\widetilde{A}$. $\qquad\square$

Remark 5.4 If $x^\dagger$ is sparse and if $\bar{n} \in \mathbb{N}$ is the largest n for which $x_n^\dagger \neq 0$, then Theorem 5.2 yields a variational inequality with

$$\varphi(t) \leq 2 \left(\sum_{k=1}^{\bar{n}} \|f_k\|_{Y^*}\right) t.$$

Consequently the $\mathcal{O}$-constant in the corresponding convergence rate $\|x_\alpha^\delta - x^\dagger\|_{\ell^1(\mathbb{N})} = \mathcal{O}(\delta)$ depends on the size of the support of $x^\dagger$. $\qquad\square$

Remark 5.5 In this remark, we revisit the situation of a diagonal operator A based on the singular value decomposition of an injective compact linear operator $\widetilde{A}$ mapping between infinite dimensional separable Hilbert spaces $\widetilde{X}$ and Y, which has been introduced in Example 2.6 and for which the proof of Theorem 5.2 yields for all $n \in \mathbb{N}$ the error estimate

$$\|x_\alpha^\delta - x^\dagger\|_{\ell^1(\mathbb{N})} \leq \|x_\alpha^\delta\|_{\ell^1(\mathbb{N})} - \|x^\dagger\|_{\ell^1(\mathbb{N})} + 2\left(\sum_{k=n+1}^\infty |x_k^\dagger| + \|A(x_\alpha^\delta - x^\dagger)\|_Y \sum_{k=1}^n \|f_k\|_{Y^*}\right)$$

for regularized solutions x_α^δ minimizing the Tikhonov functional (2.8). If the regularization parameter $\alpha = \alpha(\delta, y^\delta)$ is chosen according to the strong discrepancy principle (3.7), we have $\|A(x_\alpha^\delta - x^\dagger)\|_Y \leq \|Ax_\alpha^\delta - y^\delta\|_Y + \|Ax^\dagger - y^\delta\|_Y \leq (\tau_2 + 1)\delta$, $\|x_\alpha^\delta\|_{\ell^1(\mathbb{N})} - \|x^\dagger\|_{\ell^1(\mathbb{N})} \leq 0$ and hence for all $n \in \mathbb{N}$

$$\|x_\alpha^\delta - x^\dagger\|_{\ell^1(\mathbb{N})} \leq 2 \sum_{k=n+1}^\infty |x_k^\dagger| + 2(\tau_2+1) \sum_{k=1}^n \|f_k\|_Y \,\delta. \tag{5.6}$$

Under the power-rate conditions (5.4) an optimal choice of $n_{opt} = n_{opt}(\delta)$ yields the Hölder convergence rate (5.5) of order $\delta^{\frac{\hat{\mu}-1}{\hat{\mu}+\hat{\nu}}}$.

For ℓ^2-regularization Hölder convergence rates are usually consequences of Hölder source conditions

$$\widetilde{x}^\dagger = \left(\widetilde{A}^*\widetilde{A}\right)^\kappa \widetilde{w} \tag{5.7}$$

with some exponent $\kappa > 0$ and some source element $\widetilde{w} \in \widetilde{X}$. The best possible ℓ^2-error rate is then of order $\delta^{\frac{2\kappa}{2\kappa+1}}$ (cf., e.g., [8, Chapter 4]). In our SVD setting we have $\widetilde{x}^\dagger = \sum_{k\in\mathbb{N}} x_k^\dagger u_k$ and with (5.7) the equations $x_n^\dagger = \sigma_n^{2\kappa}\langle \widetilde{w}, u_k\rangle_{\widetilde{X}}$. This implies with $\sigma_n \sim n^{-\hat{\nu}}$ the inequalities $|x_n^\dagger| \leq \hat{K}_1 \, n^{-2\kappa\hat{\nu}}$ for all $n \in \mathbb{N}$, and hence along the lines of Example 5.3 the convergence rate

$$\|x_\alpha^\delta - x^\dagger\|_{\ell^1(\mathbb{N})} = \mathcal{O}\left(\delta^{\frac{2\kappa - \frac{1}{\hat{\nu}}}{2\kappa+1}}\right) = \mathcal{O}\left(\delta^{\frac{1-\frac{1}{2\kappa\hat{\nu}}}{1+\frac{1}{2\kappa}}}\right) \qquad \text{as} \qquad \delta \to 0.$$

Evidently, this makes only sense for $2\kappa\hat{\nu} > 1$ which is a smoothness condition ensuring that $x^\dagger \in \ell^1(\mathbb{N})$.

An error estimate similar to (5.6) can also be derived by *spectral cut-off* using the minimizer $x^{(n)} := \sum_{k=1}^{n} \frac{\langle y^\delta, v_k \rangle_Y}{\sigma_k} e_k$ of the extremal problem

$$\|Ax - y^\delta\|_Y \to \min_{x \in \operatorname{span}(e_1, e_2, \ldots, e_n)}.$$

Precisely, under Assumption 2.2 (c) and taking into account that $x^{(n)}$ is in $\operatorname{span}(e_1, e_2, \ldots, e_n)$ we can proceed similar to the proof of Theorem 5.2 and estimate as

$$\|x^{(n)} - x^\dagger\|_{\ell^1(\mathbb{N})} \leq \sum_{k=n+1}^{\infty} |x_k^\dagger| + \sum_{k=1}^{n} \|f_k\|_Y (\|Ax^{(n)} - y^\delta\|_Y + \delta).$$

Since $\|Ax^{(n)} - y^\delta\|_Y^2 = \sum_{k=n+1}^{\infty} |\langle y^\delta, v_k \rangle_Y|^2 \to 0$ as $n \to \infty$ we can always choose $n_{dis} = n_{dis}(\delta, y^\delta)$ such that $\|Ax^{(n_{dis})} - y^\delta\|_Y \leq \delta$ and we have under the conditions (5.2) the error estimate

$$\|x^{(n_{dis})} - x^\dagger\|_{\ell^1(\mathbb{N})} \leq K_1 (n_{dis})^{-\mu} + 2 K_2 (n_{dis})^\nu \delta, \tag{5.8}$$

but in contrast to the approach of ℓ^1-regularization not a general δ-rate which is independent of the specific data element y^δ. $\qquad\square$

Theorem 5.2 yields a function φ and a corresponding variational inequality (3.6) for arbitrary $x^\dagger \in X = \ell^1(\mathbb{N})$. But as a consequence of Assumption 2.2 (c) the proof of Theorem 5.2 only works for injective operators A. The converse assertion would be that if for each $\bar{x} \in X$ there is some index function φ such that the variational inequality

$$\|x - \bar{x}\|_{\ell^1(\mathbb{N})} \leq \|x\|_{\ell^1(\mathbb{N})} - \|\bar{x}\|_{\ell^1(\mathbb{N})} + \varphi(\|A(x - \bar{x})\|_Y) \quad \text{for all} \quad x \in X$$

is true, then A is injective. The following proposition states that even the existence of only one such $\bar{x}$ is sufficient for injectivity of A.

Proposition 5.6 *If there is some $\bar{x} \in \ell^1(\mathbb{N})$ with $\bar{x}_k \neq 0$ for all $k \in \mathbb{N}$ such that the variational inequality*

$$\beta\|x - \bar{x}\|_{\ell^1(\mathbb{N})} \leq \|x\|_{\ell^1(\mathbb{N})} - \|\bar{x}\|_{\ell^1(\mathbb{N})} + \varphi(\|A(x - \bar{x})\|_Y) \quad \text{for all} \quad x \in \ell^1(\mathbb{N}) \tag{5.9}$$

is satisfied with $\beta > 0$ and an arbitrary index function φ, then A is injective.

Proof. For $x \in \mathcal{N}(A) \setminus \{0\}$, where $\mathcal{N}(A)$ denotes the null-space of A, the variational inequality (5.9) applied to $\bar{x} + x$ yields

$$h(x) := \beta\|x\|_{\ell^1(\mathbb{N})} - \|\bar{x} + x\|_{\ell^1(\mathbb{N})} + \|\bar{x}\|_{\ell^1(\mathbb{N})} \leq \varphi(\|Ax\|_Y) = 0.$$

If there is some $t > 0$ such that $h(tx) > 0$ we have a contradiction to $tx \in \mathcal{N}(A)$. Thus, we can assume $h(tx) \leq 0$ for all $t > 0$.

Define two index sets

$$I_1(t) := \{k \in \mathbb{N} : t|x_k| \le |\bar{x}_k|\},$$
$$I_2(t) := \{k \in \mathbb{N} : t|x_k| > |\bar{x}_k|\},$$

where $t > 0$. Simple calculations then show that

$$|\bar{x}_k - tx_k| + |\bar{x}_k + tx_k| = \begin{cases} 2|\bar{x}_k|, & k \in I_1(t), \\ 2t|x_k|, & k \in I_2(t), \end{cases}$$

for all $k \in \mathbb{N}$. Thus we have for all $t > 0$

$$h(-tx) = 2\beta\|tx\|_{\ell^1(\mathbb{N})} - \|\bar{x} - tx\|_{\ell^1(\mathbb{N})} - \|\bar{x} + tx\|_{\ell^1(\mathbb{N})} + 2\|\bar{x}\|_{\ell^1(\mathbb{N})} - h(tx)$$
$$\ge 2\beta\|tx\|_{\ell^1(\mathbb{N})} - \|\bar{x} - tx\|_{\ell^1(\mathbb{N})} - \|\bar{x} + tx\|_{\ell^1(\mathbb{N})} + 2\|\bar{x}\|_{\ell^1(\mathbb{N})}$$
$$= 2\left(\|\bar{x}\|_{\ell^1(\mathbb{N})} - \sum_{k \in I_1(t)} |\bar{x}_k|\right) + 2t\left(\beta\|x\|_{\ell^1(\mathbb{N})} - \sum_{k \in I_2(t)} |x_k|\right).$$

Now choose $n \in \mathbb{N}$ such that there is some $k_0 \in \{1, \dots, n\}$ with $x_{k_0} \ne 0$ and such that

$$\sum_{k=n+1}^{\infty} |x_k| < \beta\|x\|_{\ell^1(\mathbb{N})}.$$

Set

$$\bar{t} := \min\left\{\frac{|\bar{x}_k|}{|x_k|} : k \in \{1, \dots, n\}, \ x_k \ne 0\right\}.$$

Then

$$\sum_{k \in I_2(\bar{t})} |x_k| \le \sum_{k=n+1}^{\infty} |x_k| < \beta\|x\|_{\ell^1(\mathbb{N})}$$

and therefore $h(-\bar{t}x) > 0$, which contradicts $-\bar{t}x \in \mathcal{N}(A)$.

In all cases the assumption $x \in \mathcal{N}(A) \setminus \{0\}$ led to a contradiction. Thus, $\mathcal{N}(A) = \{0\}$ is proven. □

6. Conclusion

We have shown that, in the case of ℓ^1-regularization, variational inqualities can significantly increase the range of solutions for which convergence rates can be shown compared to source conditions. Of course, the results are still preliminary since they rely on the injectivity of the operator A, which is an unwanted feature in many setups, in particular if motivated by compressed sensing, and since they moreover assume that the unit elements e_k belong to the range of the adjoint operator A^* for all $k \in \mathbb{N}$. However, it provides an interesting insight into the current borders of regularization theory and a strong motivation to study variational inequality approaches in particular for singular regularization functionals.

Thinking about potential extensions of the approach and weakening of the assumptions one observes that currently several steps are based on the choice of

"minimal" subgradients, i.e. the entries of the subgradient are set to zero outside the support of the solution. From a source condition perspective, it can be seen as the assumption that for all one-sparse solutions a very strong source condition is satisfied by the minimal subgradient. A feasible approach to extend the results of this paper might be to consider larger classes of subgradients, whose absolute value should however be bounded away from one or decay to zero outside the support. The exact type of needed condition remains to be determined in the future.

Acknowledgments

References

[1] S. W. Anzengruber and R. Ramlau. Convergence rates for Morozov's discrepancy principle using variational inequalities. *Inverse Problems*, 27:105007, 18pp., 2011.

[2] M. Benning and M. Burger. Error estimates for general fidelities. *Electron. Trans. Numer. Anal.*, 38:44–68, 2011.

[3] R. I. Boţ and B. Hofmann. An extension of the variational inequality approach for obtaining convergence rates in regularization of nonlinear ill-posed problems. *Journal of Integral Equations and Applications*, 22(3):369–392, 2010.

[4] K. Bredies and D. A. Lorenz. Regularization with non-convex separable constraints. *Inverse Problems*, 25(8):085011, 14pp., 2009.

[5] M. Burger and S. Osher. Convergence rates of convex variational regularization. *Inverse Problems*, 20(5):1411–1421, 2004.

[6] M. Burger, E. Resmerita, and L. He. Error estimation for Bregman iterations and inverse scale space methods in image restoration. *Computing*, 81(2-3):109–135, 2007.

[7] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications in Pure and Applied Mathematics*, 57(11):1413–1457, 2004.

[8] H.W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1996.

[9] J. Flemming. *Generalized Tikhonov Regularization and Modern Convergence Rate Theory in Banach Spaces*. Shaker Verlag, Aachen, 2012.

[10] J. Flemming. Solution smoothness of ill-posed equations in Hilbert spaces: four concepts and their cross connections. *Appl. Anal.*, 91(5):1029–1044, 2012.

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

[12] M. Grasmair. Generalized Bregman distances and convergence rates for non-convex regularization methods. *Inverse Problems*, 26(11):115014, 16pp., 2010.

[13] M. Grasmair. Linear convergence rates for Tikhonov regularization with positively homogeneous functionals. *Inverse Problems*, 27(7):075014, 16, 2011.

[14] M. Grasmair. Variational inequalities and improved convergence rates for Tikhonov regularization

on Banach spaces. Technical report, University of Vienna, Preprint, 2011. `http://arxiv.org/pdf/1107.2471v1.pdf`.

[15] M. Grasmair, M. Haltmeier, and O. Scherzer. Sparse regularization with l^q penalty term. *Inverse Problems*, 24(5):055020, 13pp., 2008.

[16] M. Grasmair, M. Haltmeier, and O. Scherzer. Necessary and sufficient conditions for linear convergence of ℓ^1-regularization. *Comm. Pure Appl. Math.*, 64(2):161–182, 2011.

[17] T. Hein and B. Hofmann. Approximate source conditions for nonlinear ill-posed problems—chances and limitations. *Inverse Problems*, 25(3):035003, 16pp., 2009.

[18] B. Hofmann. Approximate source conditions in Tikhonov-Phillips regularization and consequences for inverse problems with multiplication operators. *Math. Methods Appl. Sci.*, 29(3):351–371, 2006.

[19] B. Hofmann, B. Kaltenbacher, C. Pöschl, and O. Scherzer. A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators. *Inverse Problems*, 23(3):987–1010, 2007.

[20] B. Hofmann and P. Mathé. Parameter choice in Banach space regularization under variational inequalities. *Inverse Problems*, 28(10):104006, 17pp, 2012.

[21] B. Hofmann and M. Yamamoto. On the interplay of source conditions and variational inequalities for nonlinear ill-posed problems. *Appl. Anal.*, 89(11):1705–1727, 2010.

[22] D. A. Lorenz. Convergence rates and source conditions for Tikhonov regularization with sparsity constraints. *J. Inverse Ill-Posed Probl.*, 16(5):463–478, 2008.

[23] M. Möller. *Multiscale Methods for (Generalized) Sparse Recovery and Applications in High Dimensional Imaging*. University Münster, Germany, 2012. PhD thesis, `http://imaging.uni-muenster.de/teaching/Doktoranden/thesisMoeller.pdf`.

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

[25] R. Ramlau and E. Resmerita. Convergence rates for regularization with sparsity constraints. *Electron. Trans. Numer. Anal.*, 37:87–104, 2010.

[26] R. Ramlau and G. Teschke. Sparse recovery in inverse problems. In *Theoretical foundations and numerical methods for sparse recovery*, volume 9 of *Radon Ser. Comput. Appl. Math.*, pages 201–262. Walter de Gruyter, Berlin, 2010.

[27] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational Methods in Imaging*, volume 167 of *Applied Mathematical Sciences*. Springer, New York, 2009.

[28] T. Schuster, B. Kaltenbacher, B. Hofmann, and K.S. Kazimierski. *Regularization Methods in Banach Spaces*, volume 10 of *Radon Ser. Comput. Appl. Math.* Walter de Gruyter, Berlin/Boston, 2012. ISBN: 978-3-11-025524-9.

[29] G. Teschke. Multi-frame representations in linear inverse problems with mixed multi-constraints. *Appl. Comput. Harmon. Anal.*, 22:43–60, 2007.