The impact of a curious type of smoothness conditions on convergence rates in ℓ^1-regularization

RADU IOAN BOŢ and BERND HOFMANN *

March 1, 2013

Abstract

Tikhonov-type regularization of linear and nonlinear ill-posed problems in abstract spaces under sparsity constraints gained relevant attention in the past years. Since under some weak assumptions all regularized solutions are sparse if the ℓ^1-norm is used as penalty term, the ℓ^1-regularization was studied by numerous authors although the non-reflexivity of the Banach space ℓ^1 and the fact that such penalty functional is not strictly convex lead to serious difficulties. We consider the case that the sparsity assumption is narrowly missed. This means that the solutions may have an infinite number of nonzero but fast decaying components. For that case we formulate and prove convergence rates results for the ℓ^1-regularization of nonlinear operator equations. In this context, we outline the situations of Hölder rates and of an exponential decay of the solution components.

MSC2010 subject classification: 47J06, 65J20, 47A52, 49J40

Keywords: Nonlinear ill-posed problems, Tikhonov-type regularization, ℓ^1-regularization, sparsity constraints, convergence rates, solution decay, variational inequalities, source conditions, discrepancy principle.

1 Introduction

In the last ten years there was a substantial progress with respect to the error analysis including convergence rates results for regularized solutions to inverse problems in Banach spaces. Such problems can be formulated as operator equations

$$G(z) = y, \qquad z \in \mathcal{D}(G) \subseteq Z,\ y \in Y, \tag{1.1}$$

with a nonlinear and *smoothing* (e.g. compact) forward operator $G : \mathcal{D}(G) \subseteq Z \to Y$ mapping between the Banach spaces Z and Y with norms $\|\cdot\|_Z$ and $\|\cdot\|_Y$, respectively.

*Department of Mathematics, Chemnitz University of Technology, 09107 Chemnitz, GERMANY. Email: bot@mathematik.tu-chemnitz.de, hofmannb@mathematik.tu-chemnitz.de.

The impact of the smoothing character of G consists in the *ill-posedness* of the problem, which means that small perturbations in the right-hand side y of equation (1.1) may lead to significant errors in the solution. Moreover, solutions $z \in \mathcal{D}(G)$ need not exist for all $y \in Y$ and if they exist they need not be uniquely determined. We assume attainability, i.e. the element y belongs to the range $\mathcal{R}(G) = G(\mathcal{D}(G))$ of G, but only noisy data $y^\delta \in Y$ are available, which satisfy the deterministic noise model

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

with given noise level $\delta > 0$. Based on such data the stable approximate solution of equations of this type is required for numerous identification problems in physics, geosciences, imaging, and finance (see e.g. [26, Chapter 1] and [4, 18, 20, 23, 24, 28]). Consequently, the equation (1.1) must be regularized and the most prominent approach for Banach space regularization is the *Tikhonov-type* or *variational regularization*, where regularized solutions are minimizers of the extremal problem

$$\frac{1}{p}\|G(z) - y^\delta\|_Y^p + \alpha\,\Omega(z) \to \min, \quad \text{subject to} \quad z \in \mathcal{D}(G) \subseteq Z\,, \tag{1.3}$$

with a *regularization parameter* $\alpha > 0$, a *convex* penalty functional $\Omega : Z \to [0,\infty]$ and some exponent $1 \le p < \infty$ of the data misfit term, where we refer to [15] and [26, Chapt. 3] for standard assumptions to be made on $G, \mathcal{D}(G)$ and Ω. In this context, we also refer to [16] for a discussion of appropriate choices of the regularization parameter $\alpha > 0$ depending on the noise level δ (a priori choice) and alternatively on δ and y^δ (a posteriori choice).

For the following studies we assume to have a *bounded Schauder basis* $\{u_k\}_{k\in\mathbb{N}}$ in the Banach space Z such that the element z to be identified can be written as $z = \sum_{k=1}^{\infty} x_k u_k$ with uniquely determined coefficients $x_k \in \mathbb{R}$ in the sense that $\lim_{n\to\infty} \|z - \sum_{k=1}^{n} x_k u_k\|_Z = 0$. Our focus is on a situation where the solution z of equation (1.1) tends to be *sparse*. The treatment of sparsity in ill-posed problems has gained enormous attention recently and we refer, e.g., to [7, Section 1] for literature. In this paper, we conjecture that only a small number of coefficients x_k is relevant. Either only a finite number of coefficients is nonzero or at least the nonzero coefficients for larger k are negligibly small. In any case we assume that $\sum_{k=1}^{\infty} |x_k| < \infty$, or for short $x := (x_1, x_2, ...) \in \ell^1$. As usual we consider in the sequel for $1 \le q < \infty$ the Banach spaces ℓ^q of infinite sequences of real numbers equipped with the norms $\|x\|_{\ell^q} := \left(\sum_{k=1}^{\infty} |x_k|^q\right)^{1/q}$ and for $q = \infty$ with the norm $\|x\|_{\ell^\infty} := \sup_{k\in\mathbb{N}} |x_k|$. The latter attains the same form as the norm $\|x\|_{c_0} := \sup_{k\in\mathbb{N}} |x_k|$ of the space c_0 of infinite sequences tending to zero. By ℓ^0 we denote the set of sparse sequences, where $x_k \ne 0$ only occurs for a finite number of components.

In our setting the *synthesis operator* $L : \ell^1 \to Z$ defined as $Lx := \sum_{k=1}^{\infty} x_k u_k$ is a well-defined, injective and bounded linear operator. Even if our focus is on a *nearly sparse* situation we follow a standard approach in regularization under sparsity constraints and consider with $X := \ell^1$ the composition $F = G \circ L : \mathcal{D}(F) \subset X \to Y$ as forward operator with a domain $\mathcal{D}(F) = \{x \in \ell^1 : Lx \in \mathcal{D}(G)\}$ and as a consequence the nonlinear operator equation

$$F(x) = y, \qquad x \in \mathcal{D}(F) \subseteq X = \ell^1, \ y \in Y, \tag{1.4}$$

as an implementation of (1.1) for the specified situation under consideration here. Then in order to induce sparsity we consider as convex penalty functional $\Omega(x) := \|x\|_{\ell^1}$ and hence as regularized solutions the minimizers $x_\alpha^\delta \in \mathcal{D}(F)$ of the extremal problem

$$\frac{1}{p}\|F(x) - y^\delta\|_Y^p + \alpha\, \|x\|_{\ell^1} \to \min, \quad \text{subject to} \quad x \in \mathcal{D}(F) \subseteq \ell^1, \tag{1.5}$$

again with exponents $1 \le p < \infty$. Such regularized solutions are sparse, i.e. $x_\alpha^\delta \in \ell^0$, if F is locally Lipschitz at x_α^δ (see [11, Theorem 1.2]). We say that a solution $x^\dagger \in \mathcal{D}(F) \subseteq \ell^1$ to equation (1.4) is an ℓ^1-*norm minimizing solution* if $\|x^\dagger\|_{\ell^1} = \min_{\tilde{x} \in \mathcal{D}(F):\, F(\tilde{x}) = y} \|\tilde{x}\|_{\ell^1}$. Since ℓ^1 is not a strictly convex Banach space, ℓ^1-norm minimizing solutions need not be uniquely determined.

Since the ill-posedness of the original problem (1.1) in general carries over to the problem (1.4) (see a detailed proof for linear G in [7, Prop. 2.1]) it is well-known that, for appropriate choices of the regularization parameters $\alpha = \alpha(\delta, y^\delta)$, *convergence rates*

$$E(x^\delta_{\alpha(\delta,y^\delta)}, x^\dagger) = O(\varphi(\delta)) \qquad \text{as} \qquad \delta \to 0 \tag{1.6}$$

of regularized solutions x_α^δ to exact solutions $x^\dagger$ for some nonnegative error measure E and some index function $\varphi : (0,\infty) \to (0,\infty)$ (continuous and strictly increasing function with $\lim_{t \to +0} \varphi(t) = 0$) can only occur if $x^\dagger$ satisfies some *smoothness condition* with respect to the forward operator and if, additionally, $x^\dagger$ matches the *nonlinear structure* of the forward operator F. In Banach space variational regularization with strictly convex penalty functionals Ω it is common to use as error measure the *Bregman distance* $E(x, x^\dagger) = \Omega(x) - \Omega(x^\dagger) - \langle \xi^\dagger, x - x^\dagger \rangle_{X^* \times X}$ with a subgradient $\xi^\dagger \in \partial\Omega(x^\dagger) \subset X^*$ (see [8, 21, 22]), where we denote by X^* the dual space of X and by $\langle \cdot, \cdot \rangle_{X^* \times X}$ the dual pairing between X and X^*). For such penalties, appropriate smoothness conditions attain the form of *source conditions* (sourcewise representations), as we know them from Hilbert space regularization, but here in the form

$$\xi^\dagger = (F'(x^\dagger))^* v, \quad v \in Y^*, \tag{1.7}$$

where $(F'(x^\dagger))^* : Y^* \to X^*$ denotes the adjoint operator of the Gâteaux derivative $F'(x^\dagger) : X \to Y$ of F at the point $x^\dagger$. For ℓ^1-regularization (1.5) the Bregman distance is not preferred as error measure E, because then $E(x, x^\dagger)$ can be zero although x and $x^\dagger$ are different elements (see [19]). Therefore, in [11] a rather curious form of smoothness conditions for obtaining convergence rates in the ℓ^1-setting outlined above was suggested, which attains in our terms the form

$$e_k = (F'(x^\dagger))^* f_k, \quad f_k \in Y^*, \quad k = 1, 2, ... \tag{1.8}$$

That means, for every $k \in \mathbb{N}$ there exist source elements $f_k \in Y^*$ with respect to the Gâteaux derivative of F at $x^\dagger$ for a sourcewise representation of the k-th unit sequence $e_k := (0, 0, ..., 0, 1, 0, ...)$ with 1 in the k-th component. Here we have $X^* = \ell^\infty$ and hence

$(F'(x^\dagger))^* : Y^* \to \ell^\infty$. Note that (1.8) really characterizes the smoothness of a solution $x^\dagger \in \ell^1$ with respect to F, but in an implicit manner via the unit sequences $\{e_k\}_{k\in\mathbb{N}}$ which form a Schauder basis in all spaces $\ell^q,\ 1 \le q < \infty$, and also in c_0. This becomes clear if one rewrites $x^\dagger \in \ell^1$ as $\sum_{k=1}^{\infty} |\langle e_k, x^\dagger\rangle_{\ell^\infty\times\ell^1}| < \infty$. Namely, this condition and the decay rate of the values $|\langle e_k, x^\dagger\rangle_{\ell^\infty\times\ell^1}| \to 0$ for large k connect via (1.8) the components of $x^\dagger$ with the forward operator F by favour of its derivative $F'(x^\dagger)$. For linear operators $F : \ell^1 \to Y$ a condition of type (1.8), simplified as $e_k = F^* f_k$ for all $k \in \mathbb{N}$, was also successfully employed in [19] and [6] for obtaining convergence rates under sparsity constraints and in [7] when the sparsity assumption fails. In this context, Example 2.6 and Remark 2.9 in [7] indicate that a condition of the form (1.8) is fulfilled in a natural way for a Hilbert space Y and when the forward operator is assumed to be continuous from ℓ^2 to Y and not too far from a diagonal structure. On the other hand, [2] shows that a condition of type (1.8) is not so rare and occurs in relevant practical applications.

In the following study we will formulate and prove assertions on convergence rates (1.6) for the ℓ^1-regularization (1.5) with the error measure $E(x, x^\dagger) = \|x - x^\dagger\|_{\ell^1}$ when the sparsity assumption fails, but under the smoothness condition (1.8) and under the condition

$$\|F'(x^\dagger)(x - x^\dagger)\|_Y \le \sigma(\|F(x) - F(x^\dagger)\|_Y) \quad \text{for all} \quad x \in \mathcal{M} \subseteq X = \ell^1 \tag{1.9}$$

on the structure of nonlinearity which was introduced in our paper [5]. Here, σ denotes an in general concave index function and $\mathcal{M}$ an appropriate subset of X containing all regularized solutions x_α^δ for sufficiently small $\delta > 0$. With the assertions presented below we extend the results from [11] to the non-sparse case and the results from [7] to the case of a nonlinear forward operator F. Note that the otherwise common source condition (1.7) fails if $x^\dagger \notin \ell^0$ (see [7, Section 4] and Remark 2.6 below).

The paper is organized as follows: In Section 2 we will collect the standing assumptions used for the mathematical model under consideration and formulate two technical lemmas. The main result will be formulated as a convergence rate theorem and proven in Section 3. Conclusions in Section 4 concerning open questions and future work complete the paper.

2 Model assumptions and two technical lemmas

In this section, we collect the necessary assumptions for the model under consideration in order to formulate convergence rate results in the subsequent section.

Assumption 2.1

(a) Let Y be a Banach space, for which we consider in addition to the norm-topology $\|.\|_Y$ the weak topology '$\rightharpoonup$'. That means,

$$w_n \rightharpoonup w_0 \quad \text{in} \quad Y \quad \Longleftrightarrow \quad \langle v, w_n\rangle_{Y^*\times Y} \to \langle v, w_0\rangle_{Y^*\times Y} \ \ \forall v \in Y^*.$$

(b) In $X = \ell^1$ with predual space c_0, i.e. $c_0^* = X$, we consider the weak*-topology '$\rightharpoonup^*$', where $g_n \rightharpoonup^* g_0$ in X means that

$$\langle g_n, f\rangle_{\ell^1\times c_0} = \langle f, g_n\rangle_{\ell^\infty\times\ell^1} \to \langle g_0, f\rangle_{\ell^1\times c_0} = \langle f, g_0\rangle_{\ell^\infty\times\ell^1} \ \ \forall f \in c_0 \subset \ell^\infty.$$

(c) Let $\mathcal{D}(F)$ be a nonempty and weak* closed subset of X.

(d) Let $F : \mathcal{D}(F) \subseteq X \to Y$ be weak*-to-weak sequentially continuous, i.e.

$$g_n \rightharpoonup^* g_0 \quad \text{in} \quad X \ \text{with} \quad g_n \in \mathcal{D}(F) \quad \Longrightarrow \quad F(g_n) \rightharpoonup F(g_0) \quad \text{in} \quad Y.$$

(e) Let $y \in \mathcal{R}(F) := F(\mathcal{D}(F))$, i.e. the nonlinear operator equation (1.4) has a solution.

Because of its importance for ℓ^1-regularization we formulate and prove with the following Lemma 2.2 the weak* Kadec-Klee property of the ℓ^1-norm. As mentioned in [11], the proof is analogously to the proof of the weak Kadec-Klee property of the ℓ^1-norm given in Lemma 2 in [14].

Lemma 2.2 (weak* Kadec-Klee property of ℓ^1) *If*

$$x^{(n)} \rightharpoonup^* \bar{x} \quad \text{in} \quad \ell^1 \quad \text{and} \quad \lim_{n\to\infty} \|x^{(n)}\|_{\ell^1} = \|\bar{x}\|_{\ell^1} \quad \text{for} \quad n \to \infty,$$

then we have

$$\lim_{n\to\infty} \|x^{(n)} - \bar{x}\|_{\ell^1} = 0.$$

Proof (analogously to [14, Lemma 2]). By using the assumption $\lim_{n\to\infty} \|x^{(n)}\|_{\ell^1} = \|\bar{x}\|_{\ell^1}$ we have

$$\limsup_{n\to\infty} \|x^{(n)} - \bar{x}\|_{\ell^1} = \limsup_{n\to\infty} \{2(\|x^{(n)}\|_{\ell^1} + \|\bar{x}\|_{\ell^1}) - 2(\|x^{(n)}\|_{\ell^1} + \|\bar{x}\|_{\ell^1}) + \|x^{(n)} - \bar{x}\|_{\ell^1}\}$$

$$= 4\|\bar{x}\|_{\ell^1} - \liminf_{n\to\infty} \sum_{k=1}^{\infty} \left(2|\langle e_k, x^{(n)}\rangle_{\ell^\infty\times\ell^1}| + 2|\langle e_k, \bar{x}\rangle_{\ell^\infty\times\ell^1}| - |\langle e_k, x^{(n)} - \bar{x}\rangle_{\ell^\infty\times\ell^1}|\right).$$

Then due to $2(\|x^{(n)}\|_{\ell^1} + \|\bar{x}\|_{\ell^1}) - \|x^{(n)} - \bar{x}\|_{\ell^1} \geq 0$ the Lemma of Fatou yields

$$-\liminf_{n\to\infty} \sum_{k=1}^{\infty} \left(2|\langle e_k, x^{(n)}\rangle_{\ell^\infty\times\ell^1}| + 2|\langle e_k, \bar{x}\rangle_{\ell^\infty\times\ell^1}| - |\langle e_k, x^{(n)} - \bar{x}\rangle_{\ell^\infty\times\ell^1}|\right)$$

$$\leq -\sum_{k=1}^{\infty} \liminf_{n\to\infty} \left(2|\langle e_k, x^{(n)}\rangle_{\ell^\infty\times\ell^1}| + 2|\langle e_k, \bar{x}\rangle_{\ell^\infty\times\ell^1}| - |\langle e_k, x^{(n)} - \bar{x}\rangle_{\ell^\infty\times\ell^1}|\right).$$

and since $e_k \in c_0$ for all $k \in \mathbb{N}$ the weak*-convergence $x^{(n)} \rightharpoonup^* \bar{x}$ in ℓ^1 implies that

$$-\sum_{k=1}^{\infty} \liminf_{n\to\infty} \left(2|\langle e_k, x^{(n)}\rangle_{\ell^\infty\times\ell^1}| + 2|\langle e_k, \bar{x}\rangle_{\ell^\infty\times\ell^1}| - |\langle e_k, x^{(n)} - \bar{x}\rangle_{\ell^\infty\times\ell^1}|\right) = -4\|\bar{x}\|_{\ell^1}.$$

The formulae derived above can be combined to

$$\limsup_{n\to\infty} \|x^{(n)} - \bar{x}\|_{\ell^1} \leq 4\|\bar{x}\|_{\ell^1} - 4\|\bar{x}\|_{\ell^1} = 0,$$

which can be rewritten as $\lim_{n\to\infty} \|x^{(n)} - \bar{x}\|_{\ell^1} = 0$. □

Under Assumption 2.1 we prove in the following proposition the *existence* of ℓ^1-norm minimizing solutions for (1.4).

Proposition 2.3 *The nonlinear operator equation (1.4) admits at least one ℓ^1-norm minimizing solution.*

Proof. Item (e) of Assumption 2.1 guarantees that

$$\zeta := \inf\{\|x\|_{\ell^1} : x \in \mathcal{D}(F), F(x) = y\} \in \mathbb{R}.$$

For every $n \in \mathbb{N}$ there exists $x_n \in \mathcal{D}(F), F(x_n) = y$, such that $\zeta \le \|x_n\|_{\ell^1} \le \zeta + 1/n$. Since $\{x_n\}_{n\in\mathbb{N}}$ is a subset of $\{x \in \ell^1 : \|x\|_{\ell^1} \le \zeta + 1\}$, which is a weak* sequentially compact set, there exist a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ and an element $x^\dagger \in X$ such that $x_{n_k} \rightharpoonup^* x^\dagger$ as $k \to \infty$. The Items (c) and (d) in Assumption 2.1 ensure that $x^\dagger \in \mathcal{D}(F)$ and $F(x^\dagger) = y$, respectively. On the other hand, the weak* lower semicontinuity of the ℓ^1-norm in X implies that $\|x^\dagger\|_{\ell^1} \le \liminf\limits_{k\to\infty} \|x_{n_k}\|_{\ell^1} = \zeta$, which proves that $x^\dagger$ is an ℓ^1-norm minimizing solution of (1.4). □

Under Assumption 2.1 we also have the following Proposition 2.4 on the *existence* of ℓ^1-regularized solutions, their *stability* with respect to perturbations in the data y^δ and their *convergence* to some ℓ^1-norm minimizing solution $x^\dagger$. Note that existence results like formulated in Proposition 2.3 above and in Item (i) of Proposition 2.4 below follow under Assumption 2.1 directly from the general theory of Tikhonov regularization in topological spaces (see [27, Section 2.4]).

Proposition 2.4

(i) For all $\alpha > 0$ and $y^\delta \in Y$ there exist regularized solutions

$$x_\alpha^\delta \in \operatorname*{arg\,min}_{x\in\mathcal{D}(F)} \left\{ \frac{1}{p}\|F(x) - y^\delta\|_Y^p + \alpha \|x\|_{\ell^1} \right\}.$$

(ii) Let $\{y^{(n)}\}_{n\in\mathbb{N}} \subset Y$ be a data sequence with $\lim\limits_{n\to\infty} \|y^{(n)} - y^\delta\|_Y = 0$. Then, for a fixed regularization parameter $\alpha > 0$, every corresponding sequence $\{x^{(n)}\}_{n\in\mathbb{N}} \subset \mathcal{D}(F)$ of minimizers $x^{(n)} \in \operatorname*{arg\,min}\limits_{x\in\mathcal{D}(F)}\{\frac{1}{p}\|F(x) - y^{(n)}\|_Y^p + \alpha \|x\|_{\ell^1}\}$ has a subsequence which is norm convergent in ℓ^1 to a minimizer of the functional $\frac{1}{p}\|F(x) - y^\delta\|_Y^p + \alpha \|x\|_{\ell^1}$ over $\mathcal{D}(F)$.

(iii) Let $\{y^{\delta_n}\}_{n\in\mathbb{N}} \subset Y$ be a data sequence with $\|y^{\delta_n} - y\|_Y \le \delta_n$ with $\lim\limits_{n\to\infty} \delta_n = 0$. If

$$\alpha_n \to 0 \qquad \text{and} \qquad \frac{\delta_n^p}{\alpha_n} \to 0 \qquad \text{as} \qquad n \to \infty \tag{2.1}$$

then every sequence $\{x_{\alpha_n}^{\delta_n}\}_{n\in\mathbb{N}} \subset \mathcal{D}(F)$ of corresponding regularized solutions has a subsequence which is norm convergent in ℓ^1 to an ℓ^1-norm minimizing solution $x^\dagger$ of equation (1.4). Furthermore, it holds

$$\lim_{n\to\infty} \|x_{\alpha_n}^{\delta_n}\|_{\ell^1} = \|x^\dagger\|_{\ell^1}. \tag{2.2}$$

(iv) We have sparsity $x_\alpha^\delta \in \ell^0$ of the regularized solutions for all $\alpha > 0$ and $y^\delta \in Y$ whenever F is locally Lipschitz at x_α^δ.

Remarks on proof of Proposition 2.4. For the proof we refer to [11]. In some points the proof in [11] is only a sketch and in this context we refer also to [14] and [26, Section 4.1] for more details. An essential point for the proofs of Proposition 2.4 is the fact that the convex functional $\Omega(x) = \|x\|_{\ell^1}$ is *stabilizing* with respect to the weak*-topology in ℓ^1, i.e., the sublevel sets

$$\mathcal{M}_c := \{x \in \ell^1 : \|x\|_{\ell^1} \le c\} \tag{2.3}$$

are weak* sequentially compact for all $c \ge 0$ (cf. [26, Remark 4.9]). Then concerning (iii) under (2.1) the general theory directly provides us (considering subsequences) with

$$x_{\alpha_n}^{\delta_n} \rightharpoonup^* x^\dagger \quad \text{in} \quad \ell^1 \quad \text{and} \quad \lim_{n\to\infty} \|x_{\alpha_n}^{\delta_n}\|_{\ell^1} = \|x^\dagger\|_{\ell^1} \quad \text{as} \quad n \to \infty.$$

Since ℓ^1 satisfies the weak* Kadec-Klee property, this yields norm convergence $\lim\limits_{n\to\infty} \|x_{\alpha_n}^{\delta_n} - x^\dagger\|_{\ell^1} = 0$. The sparsity $x_\alpha^\delta \in \ell^0$ holds when F is locally Lipschitz at x_α^δ (cf. [11, Theorem 1.2]). □

The following assumptions are essential for obtaining convergence rates in the subsequent section.

Assumption 2.5

(a) *For an ℓ^1-norm minimizing solution $x^\dagger \in \mathcal{D}(F)$ of equation (1.4) let $F'(x^\dagger) : X \to Y$ be a linear bounded operator with properties like a Gâteaux derivative of F at $x^\dagger$. Precisely, we suppose for every $x \in \mathcal{D}(F)$ that*

$$\lim_{t\to+0} \frac{1}{t}\left(F(x^\dagger + t(x - x^\dagger)) - F(x^\dagger)\right) = F'(x^\dagger)(x - x^\dagger).$$

(b) *The operator $F'(x^\dagger) : \ell^1 \to Y$ satisfies the weak limit condition $F'(x^\dagger)\, e_k \rightharpoonup 0$ in Y as $k \to \infty$.*

Remark 2.6 From Items (a) and (b) of Assumption 2.5 and (1.8) we have for the adjoint operator $(F'(x^\dagger))^* : Y^* \to \ell^\infty$ the range inclusion $\mathcal{R}((F'(x^\dagger))^*) \subseteq c_0$ (see proof of Proposition 2.4 in [7]). Hence a source condition (1.7) cannot hold if $x^\dagger \notin \ell^0$, because then the subgradient $\xi^\dagger$ is not in c_0 since it contains an infinite number of components with values 1 or -1. If, however, the condition (1.8) is satisfied, then we have that $F'(x^\dagger)$ is *injective*. Namely, (1.8) implies that $|x_k| \le \|f_k\|_{Y^*}\|F'(x^\dagger)\, x\|_Y$ for all $k \in \mathbb{N}$ and all $x = (x_1, x_2, ...) \in \ell^1$, consequently $x_k = 0$ if $\|F'(x^\dagger)\, x\|_Y = 0$. Moreover, for linear ill-posed problems, i.e. if $F : \ell^1 \to Y$ is a bounded linear operator with non-closed range and we have $F'(x^\dagger) = F$ for all $x^\dagger \in \ell^1$, Item (b) of Assumption 2.5 implies that this operator F is weak*-to-weak sequentially continuous (see [7, Lemma 2.7]) as required for the nonlinear forward operator in Item (d) of Assumption 2.1.

Remark 2.7 A typical situation for nonlinear ill-posed problems occurs if F can be extended such that $F : \ell^2 \to Y$ is a *compact* and Gâteaux differentiable operator. Then the linear operator $F'(x^\dagger)$ is also compact (cf., e.g., [9, Theorem 4.19]) and $e_k \rightharpoonup 0$ in ℓ^2 implies that we even have $F'(x^\dagger)\, e_k \to 0$ in the norm topology of Y, which is a stronger condition than Item (d) in Assumption 2.1.

We close the section with a technical lemma that will be used in the subsequent section, for the proof of which we refer to [7, Lemma 5.1].

Lemma 2.8 (norm estimate in ℓ^1) *For all $x = (x_1, x_2, ...) \in \ell^1$, $\bar{x} = (\bar{x}_1, \bar{x}_2, ...) \in \ell^1$, and $n \in \mathbb{N}$ we have the estimate*

$$\|x - \bar{x}\|_{\ell^1} \le \|x\|_{\ell^1} - \|\bar{x}\|_{\ell^1} + 2 \left(\sum_{k=n+1}^{\infty} |\bar{x}_k| + \sum_{k=1}^{n} |x_k - \bar{x}_k| \right). \tag{2.4}$$

3 Convergence rates and examples

For choosing the regularization parameter α there are various modifications of the discrepancy principle available. We use a variant (see also [27, p. 137-139]), which we call *sequential discrepancy principle*, and which was recently analyzed for Banach space regularization in detail in [1, 16]. Given $0 < q < 1$ and sufficiently large $\alpha_0 > 0$, we let

$$\Delta_q := \{\alpha_j > 0 : \ \alpha_j = q^j \alpha_0, \quad j \in \mathbb{N}\}.$$

Definition 3.1 (sequential discrepancy principle) *We say that an element $\alpha \in \Delta_q$ is chosen according to the sequential discrepancy principle (SDP), if for prescribed $\tau > 1$*

$$\|F(x_\alpha^\delta) - y^\delta\|_Y \le \tau\delta < \|F(x_{\alpha/q}^\delta) - y^\delta\|_Y. \tag{3.1}$$

It was shown in [1, Theorem 1] that under weak assumptions $\alpha(\delta, y^\delta)$ chosen according to (SDP) exist and satisfy the limit conditions (2.1). The most relevant assumption for obtaining $\alpha(\delta, y^\delta) \to 0$ as $\delta \to 0$ in this context is that *exact penalization*, in particular occurring for $p = 1$ in the misfit term of (1.5) (cf. [8]), can be avoided. Under (2.1) we have for α from (SDP) with Proposition 2.4 (iii) that $\lim_{\delta \to 0} \|x^\delta_{\alpha(\delta,y^\delta)}\|_{\ell^1} = \|x^\dagger\|_{\ell^1}$. Hence for sufficiently small $\delta > 0$ all regularized solutions $x^\delta_{\alpha(\delta,y^\delta)}$ belong to the sublevel set $\mathcal{M}_c$ (cf. (2.3)) whenever $c > \|x^\dagger\|_{\ell^1}$.

Now we are ready to formulate and prove our main result:

Theorem 3.2 *Under Assumption 2.1 and condition (1.8) let the ℓ^1-norm minimizing solution $x^\dagger$ of equation (1.4) satisfy the nonlinearity condition (1.9) with some concave index function σ and $\mathcal{M} = \mathcal{M}_c$ with $c > \|x^\dagger\|_{\ell^1}$. Moreover, let (SDP) be always applicable, i.e., for sufficiently small $\delta > 0$ there is a well-defined $\alpha = \alpha(\delta, y^\delta)$ in the sense of Definition 3.1. Then we have a convergence rate*

$$\|x^\delta_{\alpha(\delta,y^\delta)} - x^\dagger\|_{\ell^1} = O(\varphi(\delta)) \qquad as \qquad \delta \to 0 \tag{3.2}$$

with the concave index function

$$\varphi(t) = 2 \inf_{n \in N} \left(\sum_{k=n+1}^{\infty} |x_k^\dagger| + \left(\sum_{k=1}^{n} \|f_k\|_{Y^*} \right) \sigma(t) \right). \tag{3.3}$$

Proof. From Lemma 2.8 we obtain for all $x \in \ell^1$ and $n \in \mathbb{N}$

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

and from (1.8) and (1.9) for $x \in \mathcal{M}$

$$\sum_{k=1}^{n} |x_k - x_k^\dagger| = \sum_{k=1}^{n} |\langle e_k, x - x^\dagger \rangle_{\ell^\infty \times \ell^1}| = \sum_{k=1}^{n} |\langle f_k, F'(x^\dagger)(x - x^\dagger) \rangle_{Y^* \times Y}|$$

$$\le \left(\sum_{k=1}^{n} \|f_k\|_{Y^*} \right) \|F'(x^\dagger)(x - x^\dagger)\|_Y \le \left(\sum_{k=1}^{n} \|f_k\|_{Y^*} \right) \sigma\left(\|F(x) - F(x^\dagger)\|_Y\right).$$

Combining this we have for sufficiently small $\delta > 0$ and $x^\delta_{\alpha(\delta,y^\delta)} \in \mathcal{M}$ the inequality

$$\|x^\delta_{\alpha(\delta,y^\delta)} - x^\dagger\|_{\ell^1} \le \|x^\delta_{\alpha(\delta,y^\delta)}\|_{\ell^1} - \|x^\dagger\|_{\ell^1} + \varphi\left(\|F(x^\delta_{\alpha(\delta,y^\delta)}) - F(x^\dagger)\|_Y\right) \tag{3.4}$$

with φ from (3.3). In analogy to [7, Proof of Theorem 5.2] one simply verifies that φ is a concave index function. The inequality (3.4) can be considered as a variational inequality along the lines of Assumption VI in [16] with $\beta = 1$ and $E(x, x^\dagger) = \|x - x^\dagger\|_{\ell^1}$. Then from [16, Theorem 2] we directly have the convergence rate (3.2). □

Remark 3.3 We note that variational inequalities for obtaining convergence rates were introduced for nonlinear ill-posed operator equations in [15] (see also [3, 5, 13, 17, 25, 26]). The specific type used in formula (3.4) was independently developed by Grasmair (see, e.g., [12]) and Flemming (see, e.g., [10]).

One easily sees that the rate function φ in Theorem 3.2 depends on decay properties of the solution components $|x_k^\dagger|$ for $k \to \infty$. The following two examples are presented to illustrate the assertion of Theorem 3.2 for important cases of decay rates of the residuals $\sum\limits_{k=n+1}^{\infty} |x_k^\dagger|$ as $n \to \infty$.

Example 3.4 (Hölder rates) In this example we assume polynomial decay and growth as

$$\sum_{k=n+1}^{\infty} |x_k^\dagger| \le K_1\, n^{-\mu}, \qquad \sum_{k=1}^{n} \|f_k\|_{Y^*} \le K_2\, n^{\nu}, \tag{3.5}$$

with exponents $\mu, \nu > 0$ and corresponding constants $K_1, K_2 > 0$. Moreover we assume that the index function in (1.9) is of the form $\sigma(t) \le K_3\, t^\kappa,\ t > 0$, for exponents $0 < \kappa \le 1$. Then we find from Theorem 3.2 by setting $n^{-\mu} \sim n^\nu t^\kappa$ and hence $n \sim t^{\frac{-\kappa}{\nu+\mu}}$ the *Hölder convergence rates*

$$\|x^\delta_{\alpha(\delta,y^\delta)} - x^\dagger\|_{\ell^1} = \mathcal{O}\left(\delta^{\frac{\mu\kappa}{\mu+\nu}}\right) \qquad \text{as} \qquad \delta \to 0 \tag{3.6}$$

whenever the regularization parameter $\alpha = \alpha(\delta, y^\delta)$ is chosen according to the sequential discrepancy principle. As expected the best possible rate arises from the limit case $\sigma(t) \le K_3 t,\ t > 0$, which is characteristic for the *tangential cone condition* on F. □

Example 3.5 (exponentially decaying solution components) In contrast to Example 3.4 we assume now that the decay rate of the nonzero solution components is of exponential type. This seems to be a realistic situation if the sparsity assumption is narrowly missed. We take

$$\sum_{k=n+1}^{\infty} |x_k^\dagger| \le K_1 \, \exp\left(-n^\gamma\right), \qquad \sum_{k=1}^{n} \|f_k\|_{Y^*} \le K_2 \, n^\nu, \tag{3.7}$$

with exponents $\gamma, \nu > 0$ and corresponding constants $K_1, K_2 > 0$. For simplicity we consider the limit case $\sigma(t) \le K_3 \, t, \; t > 0$, only. Again by Theorem 3.2 we find an associated convergence rate whenever the regularization parameter $\alpha = \alpha(\delta, y^\delta)$ is chosen according to the sequential discrepancy principle. Precisely, by setting $n^\gamma \sim \log(1/t)$ and hence $\exp\left(-n^\gamma\right) \sim t$ the rate

$$\|x^\delta_{\alpha(\delta,y^\delta)} - x^\dagger\|_{\ell^1} = \mathcal{O}\left(\delta \left(\log\left(\frac{1}{\delta}\right)\right)^{\frac{\nu}{\gamma}}\right) \qquad \text{as} \qquad \delta \to 0. \tag{3.8}$$

As the function $\log\left(\frac{1}{\delta}\right)$ tends to infinity as $\delta \to 0$ the factor $\left(\log\left(\frac{1}{\delta}\right)\right)^{\frac{\nu}{\gamma}}$ lowers the speed of convergence compared to the rate

$$\|x^\delta_{\alpha(\delta,y^\delta)} - x^\dagger\|_{\ell^1} = \mathcal{O}\left(\delta\right) \qquad \text{as} \qquad \delta \to 0, \tag{3.9}$$

which occurs for sparse solutions $x^\dagger \in \ell^0$, but the rate reduction is negligible if the exponent γ is large. □

4 Conclusions

We have shown convergence rates for the ℓ^1-regularization of nonlinear ill-posed operator equations if the smoothness condition (1.8) and the nonlinearity condition (1.9) are satisfied. If at least one of both conditions is not available, one has by Lemma 2.2 at least norm convergence of regularized solutions in the sense of Proposition 2.4 (iii) if the choice of the regularization parameter α satisfies the limit conditions (2.1). It is an interesting open problem and future work even for linear forward operators to formulate alternative smoothness conditions yielding convergence rates if (1.8) is violated.

Acknowledgments

This work was supported by the German Science Foundation (DFG) under the grants BO 2516/4-1 (R.I. Boţ) and HO 1454/8-1 (B. Hofmann).

References

[1] S.W. Anzengruber, B. Hofmann, and P. Mathé. Regularization properties of the discrepancy principle for Tikhonov regularization in Banach spaces. Paper submitted. Preprint 2012-12, Preprint Series of Faculty of Mathematics, Chemnitz University of Technology. `http://nbn-resolving.de/urn:nbn:de:bsz:ch1-qucosa-99353`.

[2] S.W. Anzengruber, B. Hofmann and R. Ramlau. On the interplay of basis smoothness and specific range conditions occurring in ℓ^1-regularization. Paper in preparation, 2013.

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

[4] A. B. Bakushinsky, M. Yu. Kokurin, and A. Smirnova. *Iterative Methods for Ill-Posed Problems – An Introduction*, volume 54 of *Inverse and Ill-Posed Problems Series*. Walter de Gruyter, Berlin, 2011.

[5] 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.

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

[7] M. Burger, J. Flemming and B. Hofmann. Convergence rates in $\ell^{\mathbf{1}}$-regularization if the sparsity assumption fails. *Inverse Problems*, 29(2):025013. 16pp, 2013.

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

[9] D. Colton and R. Kress. *Inverse Acoustic and Electromagnetic Scattering Theory*. Springer, Berlin, 1992.

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

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

[15] 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.

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

[17] 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.

[18] S. I. Kabanikhin. *Inverse and ill-posed problems – Theory and Applications*, volume 55 of *Inverse and Ill-posed Problems Series*. Walter de Gruyter GmbH, Berlin, 2012.

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

[20] J. Mueller and S. Siltanen. *Linear and Nonlinear Inverse Problems with Practical Applications*. SIAM, Philadelphia, 2012.

[21] E. Resmerita. Regularization of ill-posed problems in Banach spaces: convergence rates. *Inverse Problems*, 21(4):1303–1314, 2005.

[22] E. Resmerita and O. Scherzer. Error estimates for non-quadratic regularization and the relation to enhancement. *Inverse Problems*, 22(3):801–814, 2006.

[23] V. G. Romanov. *Investigation Methods for Inverse Problems*, in *Inverse and Ill-posed Problems Series*. VSP, Utrecht, 2002.

[24] O. Scherzer (ed.). *Handbook of Mathematical Methods in Imaging*, in 3 volumes. Springer Science+BusinessMedia, New York, 2011.

[25] 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.

[26] 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.

[27] A. N. Tikhonov, A. S. Leonov, and A. G. Yagola. *Nonlinear Ill-Posed Problems*, volume 1 and 2. Chapman & Hall, London, 1998. Translated from Russian original, Nauka, Moscow, 1995.

[28] A. G. Yagola, I. V. Kochikov, G. M. Kuramshina, and Yu. A. Pentin. *Inverse Problems of Vibrational Spectroscopy*, in *Inverse and Ill-posed Problems Series*. VSP, Utrecht, 1999.