Convergence rates for ℓ^1-regularization without injectivity-type assumptions

Jens Flemming*

June 3, 2016

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

MSC: 65J20, 47A52

Abstract

Existing convergence rate results for sparsity promoting regularization of Tikhonov-type rely on injectivity of the considered operator or at least on slightly weakened injectivity assumptions (finite basis injectivity or restricted isometry property). We extend such results to non-injective operators by formulating a suitable variational source condition, which then is characterized in the language of range conditions with respect to the range of the adjoint operator.

As a special case we consider operator equations with uniquely determined 1-norm minimizing solutions. Based on the developed characterization of a variational source condition we also provide convergence rates for the case that solutions are not sparse.

1 Setting and notation

1.1 Setting

Let Y be some Banach space over the real numbers and denote by ℓ^1 the set of all real-valued absolutely summable sequences $x = (x_k)_{k\in\mathbb{N}}$ equipped

*Technische Universität Chemnitz, Faculty of Mathematics, D-09107 Chemnitz, Germany, jens.flemming@mathematik.tu-chemnitz.de. Research supported by DFG grant FL 832/1-2.

with the norm

$$\|x\| = \sum_{k\in\mathbb{N}} |x_k|.$$

This is a Banach space, too.

We investigate linear operator equations

$$A\,x = y^\dagger \tag{1.1}$$

with a bounded linear operator $A : \ell^1 \to Y$, where $y^\dagger$ belongs to the range of A and x is the element to be determine or to be approximated.

Instead of $y^\dagger$ there might be only a noisy version y^δ be accessible, which not necessarily belongs to the range $\mathcal{R}(A)$ of A, and we assume that both elements are connected by

$$\|y^\delta - y^\dagger\| \leq \delta.$$

The non-negative number δ can be regarded as the noise level.

As will be made precise in Section 2, equations (1.1) with an operator working on the space ℓ^1 are typically ill-posed and, thus, require regularization. To find stable approximate solutions we search for minimizers of the Tikhonov-type functional

$$T_\alpha^{y^\delta}(x) := \frac{1}{p}\,\|A\,x - y^\delta\|^p + \alpha\,\|x\| \tag{1.2}$$

over all x in ℓ^1. Here, the exponent satisfies $p \geq 1$ and α is a positive regularization parameter controlling the trade-off between data-fitting and stability. Typically, the minimizers are sparse (cf. Proposition 2.2). That is, only finitely many components are not zero. Such sparsity promoting regularization methods are widely used in practice.

Our aim is to establish a variational source condition

$$\beta\,E(x,x^\dagger) \leq \|x\| - \|x^\dagger\| + \varphi(\|A\,x - A\,x^\dagger\|) \quad \text{for all } x \in \ell^1, \tag{1.3}$$

where β is a non-negative constant, $x^\dagger$ is a solution to (1.1), E is some functional expressing the deviation of x from $x^\dagger$ (details will be provided in Section 4), and φ is a concave index function. Here, a non-negative function $\varphi : [0,\infty) \to [0,\infty)$ is called index function if it is continuous and strictly increasing and satisfies $\varphi(0) = 0$. Variational source conditions were introduced in [15] and further developed during the past years in [18, 14, 4, 12, 8].

Denoting by x_α^δ some minimizer of $T_\alpha^{y^\delta}$, such a variational source condition yields estimates

$$E(x_\alpha^\delta, x^\dagger) \leq c\,\varphi(\delta) \quad \text{for all } \delta > 0 \tag{1.4}$$

with some positive constant c, if α is chosen appropriately in dependence of δ.

Such estimates (1.4) already have been obtained with general concave φ for injective A in [6] and with linear φ for almost injective A in [13] (see also references therein). Here, almost injective means, roughly speaking, that injectivity is required on finite-dimensional subspaces (finite basis injectivity, restricted isometry property). For a detailed discussion of such conditions we refer to [13]. In finite-dimensional settings estimates (1.4) were first obtained in [7]. In the present paper we do not require any injectivity-type assumptions for proving convergence rates in an infinite-dimensional setting. Thus, we overcome the limitations of the convergence rates results in [13] and [6]. Even if applied to finite-dimensional spaces our results extend existing ones. Further discussion of sufficient conditions for convergence rates for ℓ^1-regularization can be found in [17, 19, 3, 2, 9, 10]

1.2 ℓ^1 and related spaces

In the sequel we need several facts about the space ℓ^1 and related notation, which we now recall. By c_0 we denote the Banach space of all real-valued sequences converging to zero equipped with the norm

$$\|u\| = \sup_{k \in \mathbb{N}} |u_k|.$$

Its topological dual is ℓ^1, allowing us to introduce the notion of weak* convergence in ℓ^1. We say that a sequence $(x^{(n)})_{n\in\mathbb{N}}$ in ℓ^1 converges weakly* to x if

$$\lim_{n\to\infty} \langle x^{(n)}, u \rangle = \langle x, u \rangle \quad \text{for all } u \in c_0.$$

Here, $\langle \cdot, \cdot \rangle$ denotes the duality pairing between a space and its dual. The first argument is always the element from the dual of the space from which the second argument comes.

With respect to the weak* topology bounded sets in ℓ^1 are sequentially relatively compact. That is, each sequence in a bounded set contains a subsequence which converges weakly* to an element in ℓ^1. In addition the norm of ℓ^1 is weak* sequentially lower semi-continuous, meaning that for each sequence $(x^{(n)})_{n\in\mathbb{N}}$ in ℓ^1 converging weakly* to some x in ℓ^1 we have

$$\|x\| \leq \liminf_{n\to\infty} \|x^{(n)}\|.$$

The topological dual of ℓ^1 is the space ℓ^∞ of all real-valued bounded sequences with the norm

$$\|\xi\| = \sup_{k\in\mathbb{N}} |\xi_k|.$$

But the dual of ℓ^∞ is not ℓ^1.

1.3 Convex analysis

For a proper convex function $f : X \to (-\infty, \infty]$ on some Banach space X we denote by

$$\partial f(x) := \{\xi \in X^* : f(\tilde{x}) \geq f(x) + \langle \xi, \tilde{x} - x\rangle \text{ for all } \tilde{x} \in X\}$$

the subdifferential of f at x. The elements of $\partial f(x)$ are called subgradients.

Given a convex subset A of some Banach space X and an element x of A the set

$$N_A(x) := \{\xi \in X^* : \langle \xi, \tilde{x} - x\rangle \leq 0 \text{ for all } \tilde{x} \in A\} \tag{1.5}$$

is called normal cone of A at x.

2 Existence, stability, convergence

We start with an extension of results obtained in [6, Prop. 2.4, Lemma 2.7].

Lemma 2.1. *Denoting by $e^{(n)}$ the standard unit sequences in ℓ^1 with a one at position n and zeros else, the following assertions are equivalent.*

(i) $(A\,e^{(n)})_{n\in\mathbb{N}}$ *converges weakly to zero.*

(ii) $\mathcal{R}(A^*) \subseteq c_0$.

(iii) *A is sequentially weak*-to-weak continuous.*

Proof. Let (i) be satisfied. Then for each $A^*\,\eta$ from $\mathcal{R}(A^*)$ we have

$$[A^*\,\eta]_k = \langle A^*\,\eta, e^{(k)}\rangle = \langle \eta, A\,e^{(k)}\rangle \to 0 \quad \text{if } k \to \infty,$$

that is, $A^*\,\eta \in c_0$.

Now let (ii) be true. If we take a weakly* convergent sequence $(x^{(n)})_{n\in\mathbb{N}}$ with limit x, then $\langle \eta, A\,x^{(n)}\rangle = \langle A^*\,\eta, x^{(n)}\rangle$ and since $A^*\,\eta$ belongs to c_0 and ℓ^1 is the dual of c_0 we may write $\langle A^*\,\eta, x^{(n)}\rangle = \langle x^{(n)}, A^*\,\eta\rangle$. Thus,

$$\langle \eta, A\,x^{(n)}\rangle = \langle x^{(n)}, A^*\,\eta\rangle \to \langle x, A^*\,\eta\rangle \quad \text{if } n \to \infty,$$

showing

$$\lim_{n\to\infty} \langle \eta, A\,x^{(n)}\rangle = \langle \eta, A\,x\rangle \quad \text{for all } \eta \in Y^*.$$

Finally, from (iii) and from the obvious fact that $(e^{(n)})_{n\in\mathbb{N}}$ converges weakly* to zero we immediately obtain (i). □

Note that the conditions in the lemma are satisfied if A can be extended to a bounded linear operator on ℓ^2, the space of square-summable sequences. In [10] it was shown that equations (1.1) for which $\mathcal{R}(A^*) \subseteq c_0$ holds always are ill-posed of Nashed's ill-posedness type II (i.e., $\mathcal{R}(A)$ contains no closed infinite dimensional subspace).

Proposition 2.2. *Let $\mathcal{R}(A^*) \subseteq c_0$. Then the following assertions are true.*

- *(i) Existence: There exist solutions to* (1.1) *with minimal norm (referred to as norm minimizing solutions) and there exist minimizers of the Tikhonov-type functional* (1.2)*. Further, all minimizer of $T_\alpha^{y^\delta}$ have only finitely many non-zero components.*
- *(ii) Stability: If $(y_n)_{n\in\mathbb{N}}$ converges to y^δ and if $(x^{(n)})_{n\in\mathbb{N}}$ is a corresponding sequence of minimizers of $T_\alpha^{y_n}$, then this second sequence has a weakly* convergent subsequence and each weak* convergent subsequence converges weakly* to a minimizer of $T_\alpha^{y^\delta}$.*
- *(iii) Convergence: If $(\delta_n)_{n\in\mathbb{N}}$ converges to zero and if $(y_n)_{n\in\mathbb{N}}$ satisfies $\|y_n - y^\dagger\| \le \delta_n$, then there is a sequence $(\alpha_n)_{n\in\mathbb{N}}$ such that each corresponding sequence of minimizers of $T_{\alpha_n}^{y_n}$ contains a weakly* convergent subsequence. Each such subsequence converges in norm to some norm minimizing solution of* (1.1)*.*

Proof. By the weak* sequential lower semi-continuity of the ℓ^1-norm and by the weak* sequential relative compactness of bounded sets in ℓ^1 we have weak* sequential compactness of closed balls in ℓ^1 centered at zero. Taking also the sequential weak*-to-weak continuity of A into account (see Lemma 2.1) we may apply standard results on Tikhonov-type regularization methods in Banach spaces [15, 8, 20]. Note that the ℓ^1-norm satisfies the so called weak* Kadec-Klee property, which yields convergence in norm in item (iii) of the proposition.

It only remains to show that each minimizer of (1.2) has only finitely many non-zero components. This is a consequence of $\mathcal{R}(A^*) \subseteq c_0$. By standard arguments from convex analysis we see that some x is a minimizer of $T_\alpha^{y^\delta}$ if and only if there is some ξ in ℓ^∞ such that

$$-\xi \in \alpha\,\partial\|\cdot\|(x) \quad \text{and} \quad \xi \in A^*\,\partial(\|A\,\cdot\,-y^\delta\|^p)(x).$$

Thus, $|\xi_k| = \alpha$ whenever $x_k \neq 0$ and ξ belongs to c_0. This is only possible if x has at most finitely many non-zero components. □

3 Distance to norm minimizing solutions

We do not assume injectivity of A. Thus, there might by many solutions to (1.1). We denote the set of all solutions by

$$L := \{x \in \ell^1 : A\,x = y^\dagger\}.$$

Even restricting our attention to norm minimizing solutions does not guarantee uniqueness, because the norm of ℓ^1 is not strictly convex. The set of all norm minimizing solutions will be denoted by

$$S := \{x \in L : \|x\| \leq \|\tilde{x}\| \text{ for all } \tilde{x} \in L\}.$$

Obviously, all elements in S have the same norm and we denote this value by $\|S\|$. In addition we immediately see that S is closed and convex.

For x in ℓ^1 we denote by

$$\operatorname{dist}(x, S) := \inf_{x^\dagger \in S} \|x - x^\dagger\|$$

the distance of x to the set S of norm minimizing solutions.

Proposition 3.1. *Let $\mathcal{R}(A^*) \subseteq c_0$. Then for each x in ℓ^1 there is some $x^\dagger$ in S such that* $\operatorname{dist}(x, S) = \|x - x^\dagger\|$.

Proof. This can be shown by standard arguments. □

The next proposition states that all norm minimizing solutions lie in the same orthant.

Proposition 3.2. *For each k in $\mathbb{N}$ we have either $x_k^\dagger \geq 0$ for all $x^\dagger$ in S or $x_k^\dagger \leq 0$ for all $x^\dagger$ in S.*

Proof. Assume that there are $x^\dagger$ and $\tilde{x}^\dagger$ in S with $x^\dagger_k < 0$ and $\tilde{x}^\dagger_k > 0$ for some k. Set

$$t := \frac{\tilde{x}^\dagger_k}{\tilde{x}^\dagger_k - x^\dagger_k}.$$

Then $t \in (0,1)$ and the convex combination $t\,x^\dagger + (1-t)\,\tilde{x}^\dagger$ belongs to S. We now have

$$\begin{aligned}\|t\,x^\dagger + (1-t)\,\tilde{x}^\dagger\| &= \sum_{l\neq k} |t\,x^\dagger_l + (1-t)\,\tilde{x}^\dagger_l| \leq t \sum_{l\neq k} |x^\dagger_l| + (1-t) \sum_{l\neq k} |\tilde{x}^\dagger_l| \\ &= \|S\| - \bigl(t\,|x^\dagger_k| + (1-t)\,|\tilde{x}^\dagger_k|\bigr) < \|S\|,\end{aligned}$$

which is not possible for an element in S. □

Justified by the proposition we define a sequence $\sigma^S = (\sigma^S_k)_{k\in\mathbb{N}}$ by

$$\sigma^S_k := \begin{cases} 1, & \text{if there are } x^\dagger \text{ in } S \text{ with } x^\dagger_k > 0, \\ -1, & \text{if there are } x^\dagger \text{ in } S \text{ with } x^\dagger_k < 0, \\ 0, & \text{if } x^\dagger_k = 0 \text{ for all } x^\dagger \text{ in } S. \end{cases}$$

Further we introduce the set

$$\mathbb{1}^S := \bigl\{(\sigma_k)_{k\in\mathbb{N}} : \sigma_k \in \{-1,0,1\} \text{ for all } k \text{ and } \sigma_k = 0 \text{ if } \sigma^S_k = 0\bigr\}$$

and for each $\sigma \in \mathbb{1}^S$ subsets $S(\sigma)$ of S by

$$\begin{aligned} S(\sigma) := \{x^\dagger \in S :\ & \text{there is some } \xi \in N_S(x^\dagger) \text{ with} \\ & \xi_k = \sigma_k \text{ if } \sigma_k \neq 0, \quad \xi_k \in (-1,1) \text{ if } \sigma_k = 0,\ \sigma^S_k \neq 0\}. \end{aligned}$$

Here, $N_{x^\dagger}(S)$ denotes the normal cone of S at $x^\dagger$, see (1.5). We can regard $S(\sigma)$ as the face of S visible from direction σ.

Lemma 3.3. *Let $\mathcal{R}(A^*) \subseteq c_0$. If $\sigma \in \mathbb{1}^S$ has only finitely many non-zero components, we have $S(\sigma) \neq \emptyset$.*

Proof. Setting $\xi := \sigma$ we show that there is some $x^\dagger$ in S with $\xi \in N_S(x^\dagger)$, that is, $x^\dagger$ maximizes $\langle \xi, x\rangle$ over all x in S. Let $(x^{(n)})_{n\in\mathbb{N}}$ be a sequence in S with

$$\langle \xi, x^{(n)}\rangle \to c := \sup_{x\in S} \langle \xi, x\rangle.$$

This sequence is bounded and thus contains a subsequence converging weakly* to some $x^\dagger$ in ℓ^1. The sequential weak*-to-weak continuity of A (see Proposition 2.1) guarantees $x^\dagger \in L$ and the sequential weak* lower semi-continuity of the ℓ^1-norm yields $x^\dagger \in S$. Denoting the subsequence again by $(x^{(n)})_{n\in\mathbb{N}}$ and noting that $\xi \in c_0$ we further obtain

$$c = \lim_{n\to\infty} \langle \xi, x^{(n)} \rangle = \lim_{n\to\infty} \langle x^{(n)}, \xi \rangle = \langle x^\dagger, \xi \rangle = \langle \xi, x^\dagger \rangle. \tag{3.1}$$

Thus, $x^\dagger$ indeed maximizes $\langle \xi, \cdot \rangle$ over S. □

Now we restrict our attention to subsets of ℓ^1 on which $\operatorname{dist}(x, S)$ is almost affine. For $\sigma \in \mathbb{1}^S$ and $x^\dagger \in S$ we define

$$\begin{aligned} M_{x^\dagger}(\sigma) := \{x \in \ell^1 : \; & x_k \geq x_k^\dagger \text{ if } \sigma_k = 1, \\ & x_k \leq x_k^\dagger \text{ if } \sigma_k = -1, \\ & x_k = x_k^\dagger \text{ if } \sigma_k = 0,\, \sigma_k^S \neq 0\}. \end{aligned}$$

The sets $M_{x^\dagger}(\sigma)$ are obviously closed and convex and we always have $x^\dagger \in M_{x^\dagger}(\sigma)$.

Proposition 3.4. *Let $\sigma \in \mathbb{1}^S$ and let $x^\dagger \in S(\sigma)$. Then*

$$\operatorname{dist}(x, S) = \langle \sigma, x - x^\dagger \rangle + \sum_{k:\, \sigma_k^S = 0} |x_k| \qquad \textit{for all } x \in M_{x^\dagger}(\sigma).$$

Proof. As a standard result of convex analysis we have $\operatorname{dist}(x, S) = \|x - x^\dagger\|$ if and only if there is some ξ in the normal cone $N_S(x^\dagger)$ such that $\xi \in -\partial\|x - \cdot\,\|(x^\dagger)$. On the one hand we have

$$\begin{aligned} -\partial\|x - \cdot\,\|(x^\dagger) = \partial\|\cdot\|(x - x^\dagger) = \{\tilde{\xi} \in \ell^\infty : \; & \tilde{\xi}_k = 1 \text{ if } x_k > x_k^\dagger, \\ & \tilde{\xi}_k = -1 \text{ if } x_k < x_k^\dagger, \\ & \tilde{\xi}_k \in [-1, 1] \text{ if } x_k = x_k^\dagger\}. \end{aligned}$$

On the other hand, $x^\dagger \in S(\sigma)$ and $x \in M_{x^\dagger}(\sigma)$ imply that there is some ξ in $N_S(x^\dagger)$ such that

$$\xi_k \begin{cases} = 1, & \text{if } x_k > x_k^\dagger, \\ = -1, & \text{if } x_k < x_k^\dagger, \\ \in [-1, 1], & \text{if } x_k = x_k^\dagger \end{cases} \quad \text{for all } k \text{ with } \sigma_k^S \neq 0.$$

If we now define $\tilde{\xi}$ by

$$\tilde{\xi}_k := \begin{cases} \xi_k, & \text{if } \sigma_k^S \neq 0, \\ 1, & \text{if } \sigma_k^S = 0,\, x_k \geq 0, \\ -1, & \text{if } \sigma_k^S = 0,\, x_k < 0, \end{cases}$$

we immediately see, that $\tilde{\xi} \in -\partial\|x - \cdot\,\|(x^\dagger)$ (remember $x^\dagger_k = 0$ if $\sigma^S_k = 0$). From $\xi \in N_S(x^\dagger)$ we have

$$\langle \xi, \tilde{x}^\dagger - x^\dagger \rangle \leq 0 \quad \text{for all } \tilde{x}^\dagger \in S,$$

which together with

$$\langle \tilde{\xi}, \tilde{x}^\dagger - x^\dagger \rangle = \sum_{k\colon \sigma^S_k \neq 0} \tilde{\xi}_k \, (\tilde{x}^\dagger_k - x^\dagger_k) = \sum_{k\colon \sigma^S_k \neq 0} \xi_k \, (\tilde{x}^\dagger_k - x^\dagger_k) = \langle \xi, \tilde{x}^\dagger - x^\dagger \rangle$$

yields that $\tilde{\xi}$ is in $N_S(x^\dagger)$, too. This proves $\operatorname{dist}(x, S) = \|x - x^\dagger\|$.

As the second step we observe that $x \in M_{x^\dagger}(\sigma)$ yields

$$|x_k - x^\dagger_k| = \sigma_k \, (x_k - x^\dagger_k) \quad \text{if } \sigma^S_k \neq 0.$$

Thus,

$$\|x - x^\dagger\| = \sum_{k\colon \sigma^S_k \neq 0} |x_k - x^\dagger_k| + \sum_{k\colon \sigma^S_k = 0} |x_k| = \langle \sigma, x - x^\dagger \rangle + \sum_{k\colon \sigma^S_k = 0} |x_k|.$$

□

Corollary 3.5. *For each $\sigma \in \mathbb{1}^S$ and each $x^\dagger \in S(\sigma)$ we have*

$$S \cap M_{x^\dagger}(\sigma) = \{x^\dagger\}.$$

Proof. Assume that there is a second solution $\tilde{x}^\dagger$ in $S \cap M_{x^\dagger}(\sigma)$. Then from Proposition 3.4 (and even more easily from its proof) we obtain

$$0 = \operatorname{dist}(\tilde{x}^\dagger, S) = \|\tilde{x}^\dagger - x^\dagger\|.$$

Thus, $\tilde{x}^\dagger = x^\dagger$. □

We close this section with the following important observation.

Proposition 3.6. *The sets $M_{x^\dagger}(\sigma)$ cover the whole space ℓ^1, that is,*

$$\ell^1 = \bigcup_{\sigma \in \mathbb{1}^S} \bigcup_{x^\dagger \in S(\sigma)} M_{x^\dagger}(\sigma). \tag{3.2}$$

Proof. For fixed x in ℓ^1 let $x^\dagger$ be a minimizer of $\|x - \cdot\,\|$ over S. Then there is some ξ in the normal cone $N_S(x^\dagger)$ such that $\xi \in -\partial\|x - \cdot\,\|(x^\dagger)$. Thus, we know

$$\xi_k = 1 \text{ if } x_k > x^\dagger_k, \quad \xi_k = -1 \text{ if } x_k < x^\dagger_k, \quad \xi_k \in [-1, 1] \text{ if } x_k = x^\dagger_k.$$

If we now define σ by

$$\sigma_k := \begin{cases} 1, & \text{if } \xi_k = 1,\ \sigma_k^S \neq 0, \\ -1, & \text{if } \xi_k = -1,\ \sigma_k^S \neq 0, \\ 0, & \text{if } \xi_k \in (-1,1) \text{ or } \sigma_k^S = 0, \end{cases}$$

then $\sigma \in \mathbb{1}^S$, $x^\dagger \in S(\sigma)$ and $x \in M_{x^\dagger}(\sigma)$. □

4 A variational source condition

Having finished the study of the distance $\operatorname{dist}(x, S)$ between an element x in ℓ^1 and the set S of norm minimizing solutions we now want to establish a variational source condition (1.3) with error functional

$$E(x, x^\dagger) = \operatorname{dist}(x, S),$$

where $x^\dagger$ is some element of S, and with a linear index function $\varphi(t) = \gamma\, t$, $\gamma > 0$. The desired variational source condition reads

$$\beta \operatorname{dist}(x, S) \leq \|x\| - \|x^\dagger\| + \gamma \, \|A\, x - A\, x^\dagger\| \quad \text{for all } x \in \ell^1$$

or, taking into account that $\|x^\dagger\|$ and $A\, x^\dagger$ do not depend on the concrete choice of $x^\dagger$ from S,

$$\beta \operatorname{dist}(x, S) \leq \|x\| - \|S\| + \gamma \, \|A\, x - A\, S\| \quad \text{for all } x \in \ell^1. \tag{4.1}$$

It suffices to consider $\beta \in (0, 1]$, because a variational source condition with $\beta > 1$ always implies a variational source condition with $\beta \leq 1$.

At first we split the variational source condition into 'smaller' ones. Here and in the sequel we use the notation introduced in Section 3.

Lemma 4.1. *The variational source condition* (4.1) *on* ℓ^1 *is satisfied if and only if for each* $\sigma \in \mathbb{1}^S$ *and each* $x^\dagger \in S(\sigma)$ *we have*

$$\beta \, \langle \sigma, x - x^\dagger \rangle + \beta \sum_{k:\, \sigma_k^S = 0} |x_k| \leq \|x\| - \|x^\dagger\| + \gamma \|A\, x - A\, x^\dagger\| \tag{4.2}$$

for all $x \in M_{x^\dagger}(\sigma)$.

Proof. This is a direct consequence of Propositions 3.4 and 3.6. □

Lemma 4.2. *For $\sigma \in \mathbb{1}^S$ and $x^\dagger \in S(\sigma)$ the variational source condition (4.2) on $M_{x^\dagger}(\sigma)$ is satisfied if and only if there is some η in Y^* with $\|\eta\| \leq \frac{\gamma}{1+\beta}$ such that*

$$\begin{cases} [A^* \eta]_k \in [-\mu, \mu], & \text{if } \sigma_k^S = 0, \\ \sigma_k [A^* \eta]_k \leq \mu, & \text{if } \sigma_k^S \neq 0,\, x_k^\dagger = 0,\, \sigma_k \neq 0, \\ \sigma_k^S [A^* \eta]_k \leq \mu, & \text{if } x_k^\dagger \neq 0,\, \sigma_k = \sigma_k^S, \\ \sigma_k^S [A^* \eta]_k \geq 1, & \text{if } x_k^\dagger \neq 0,\, \sigma_k = -\sigma_k^S \end{cases}$$

for all k, where $\mu := \frac{1-\beta}{1+\beta} \in [0, 1)$.

Proof. We rewrite (4.2) as

$$\|x^\dagger\| \leq -\beta \langle \sigma, x - x^\dagger \rangle + (1-\beta) \sum_{k\colon \sigma_k^S = 0} |x_k| + \sum_{k\colon \sigma_k^S \neq 0} |x_k| + \gamma \|A\,x - A\,x^\dagger\|$$

and, taking into account that $x_k^\dagger = 0$ if $\sigma_k^S = 0$, see that $x^\dagger$ is a minimizer of the convex functional on the right-hand side with respect to x in $M_{x^\dagger}(\sigma)$. Thus, there is some ξ in the normal cone $N_{M_{x^\dagger}(\sigma)}(x^\dagger)$ such that $-\xi$ belongs to the subdifferential of the functional at $x^\dagger$. This subdifferential is the sum of the subdifferential for each summand. We have

$$\begin{aligned} N_{M_{x^\dagger}(\sigma)}(x^\dagger) = \{\xi \in \ell^\infty : {} & \xi_k = 0 \text{ if } \sigma_k^S = 0, \\ & \xi_k \leq 0 \text{ if } \sigma_k = 1, \\ & \xi_k \geq 0 \text{ if } \sigma_k = -1\}, \end{aligned}$$

$$\partial(-\beta \langle \sigma, \cdot - x^\dagger \rangle)(x^\dagger) = -\beta\, \sigma,$$

$$\begin{aligned} \partial\left(x \mapsto (1-\beta) \sum_{k\colon \sigma_k^S = 0} |x_k|\right)(x^\dagger) = \{\tilde{\xi} \in \ell^\infty : {} & \tilde{\xi}_k \in [-(1-\beta), 1-\beta] \text{ if } \sigma_k^S = 0, \\ & \tilde{\xi}_k = 0 \text{ if } \sigma_k^S \neq 0\}, \end{aligned}$$

$$\begin{aligned} \partial\left(x \mapsto \sum_{k\colon \sigma_k^S \neq 0} |x_k|\right)(x^\dagger) = \{\tilde{\xi} \in \ell^\infty : {} & \tilde{\xi}_k = 0 \text{ if } \sigma_k^S = 0, \\ & \tilde{\xi}_k = 1 \text{ if } x_k^\dagger > 0 \\ & \tilde{\xi}_k = -1 \text{ if } x_k^\dagger < 0 \\ & \tilde{\xi}_k \in [-1, 1] \text{ if } \sigma_k^S \neq 0,\, x_k^\dagger = 0\}, \end{aligned}$$

and

$$\partial(\gamma\|A\,\cdot\,-A\,x^\dagger\|)(x^\dagger)=\{A^*\,\eta:\eta\in Y^*,\,\|\eta\|\leq\gamma\}.$$

From these equations we see that there is some η in Y^* with $\|\eta\|\leq\gamma$ such that

$$\begin{cases}-[A^*\,\eta]_k\in[-(1-\beta),1-\beta], & \text{if } \sigma_k^S=0,\\ -\sigma_k\,[A^*\,\eta]_k\leq 1-\beta, & \text{if } \sigma_k^S\neq 0,\ x_k^\dagger=0,\ \sigma_k\neq 0,\\ -\sigma_k^S\,[A^*\,\eta]_k\leq 1-\beta, & \text{if } x_k^\dagger\neq 0,\ \sigma_k=\sigma_k^S,\\ -\sigma_k^S\,[A^*\,\eta]_k\geq 1+\beta, & \text{if } x_k^\dagger\neq 0,\ \sigma_k=-\sigma_k^S.\end{cases}$$

Replacing η by $-(1+\beta)\,\eta$ completes the proof. □

Theorem 4.3. *The variational source condition* (4.1) *on ℓ^1 is satisfied if and only if for each $\sigma\in\mathbb{1}^S$ and each $x^\dagger\in S(\sigma)$ there is some η in Y^* with $\|\eta\|\leq\frac{\gamma}{1+\beta}$ such that*

$$\begin{cases}[A^*\,\eta]_k\in[-\mu,\mu], & \text{if } \sigma_k^S=0,\\ \sigma_k\,[A^*\,\eta]_k\leq \mu, & \text{if } \sigma_k^S\neq 0,\ x_k^\dagger=0,\ \sigma_k\neq 0,\\ \sigma_k^S\,[A^*\,\eta]_k\leq \mu, & \text{if } x_k^\dagger\neq 0,\ \sigma_k=\sigma_k^S,\\ \sigma_k^S\,[A^*\,\eta]_k\geq 1, & \text{if } x_k^\dagger\neq 0,\ \sigma_k=-\sigma_k^S\end{cases}\tag{4.3}$$

for all k, where $\mu:=\frac{1-\beta}{1+\beta}\in[0,1)$.

Proof. This is a direct consequence of Lemma 4.1 and Lemma 4.2. □

From a variational source condition (4.1) we obtain the error estimate

$$\operatorname{dist}(x_\alpha^\delta,S)\leq c\,\delta$$

for all sufficiently small $\delta>0$ with some constant $c>0$, if the regularization parameter α is chosen appropriately, for example proportional to δ or by the two-sided discrepancy principle (see [8] for both variants) or by the sequential discrepancy principle [16, 1]. Here, x_α^δ again denotes a minimizer of the Tikhonov-type functional (1.2).

Remark 4.4. Let $\mathcal{R}(A^*)\subseteq c_0$. Then Theorem 4.3 implies that a variational source condition (4.1) can only be satisfied if all solutions in S are sparse. To see this choose $\sigma=-\sigma^S$. Then $|[A^*\,\eta]_k|\geq 1$ on the support of $x^\dagger$, which is only possible if the support is finite.

Remark 4.5. Denoting by $e^{(k)}$ the standard unit sequence (one at position k, zero else) the authors of [6] used the assumption

$$e^{(k)} \in \mathcal{R}(A^*) \quad \text{for all} \quad k \in \mathbb{N} \tag{4.4}$$

to obtain a variational source condition. Such an assumption can also be found in [11]. Obviously, condition (4.3) is weaker than (4.4) because each finitely supported element in $\mathcal{R}(A^*)$ is a linear combination of the $e^{(k)}$. In [6] it was shown that (4.4) implies injectivity of A whereas our characterization (4.3) of a variational source condition does not imply injectivity (cf. Section 7).

We close this section with three remarks which reduce the set of elements σ and $x^\dagger$ for which condition (4.3) has to be verified in order to obtain convergence rates.

Remark 4.6. For fixed $\sigma \in \mathbb{1}^S$ condition (4.3) is satisfied for all $x^\dagger \in S(\sigma)$ if and only if it is satisfied for all $x^\dagger \in S(\sigma)$ having maximal support. Here we say that some $x^\dagger$ from $S(\sigma)$ has maximal support if there is no $\tilde{x}^\dagger$ in $S(\sigma)$ with $\{k \in \mathbb{N} : \tilde{x}^\dagger_k \neq 0\} \supsetneq \{k \in \mathbb{N} : x^\dagger_k \neq 0\}$.

Remark 4.7. Let $\sigma \in \mathbb{1}^S$. If $\sigma_k = \sigma^S_k$ for all k with $\sigma_k \neq 0$ and with $x^\dagger_k \neq 0$ for at least one $x^\dagger \in S(\sigma)$, then condition (4.3) is satisfied with $\eta = 0$.

Remark 4.8. Let $\sigma \in \mathbb{1}^S$ and $\tilde{\sigma} \in \mathbb{1}^S$ such that $\tilde{\sigma}$ has smaller support than σ, that is, $\sigma_k \neq 0$ whenever $\tilde{\sigma}_k \neq 0$. Further, let $x^\dagger$ be in $S(\sigma)$ and also in $S(\tilde{\sigma})$. Then condition (4.3) is satisfied for $\tilde{\sigma}$ if it is satisfied for σ.

5 Unique norm minimizing solution

We consider the case that the set of norm minimizing solutions contains only one element, that is,

$$S = \{x^\dagger\}.$$

Note that this does not necessarily imply injectivity of A. The variational source condition (4.1) now reads

$$\beta \, \|x - x^\dagger\| \leq \|x\| - \|x^\dagger\| + \gamma \, \|A\,x - A\,x^\dagger\| \quad \text{for all } x \in \ell^1 \tag{5.1}$$

and Theorem 4.3 can be refined as follows.

Theorem 5.1. *Let $S = \{x^\dagger\}$. Then the variational source condition* (5.1) *on ℓ^1 is satisfied if and only if for each $\sigma \in \mathbb{1}^S$ there is some η in Y^* with $\|\eta\| \leq \frac{\gamma}{1+\beta}$ such that*

$$\begin{cases} [A^*\,\eta]_k \in [-\mu, \mu], & \text{if } x^\dagger_k = 0, \\ \sigma^S_k\,[A^*\,\eta]_k = \mu, & \text{if } x^\dagger_k \neq 0,\ \sigma_k = \sigma^S_k, \\ \sigma^S_k\,[A^*\,\eta]_k = 1, & \text{if } x^\dagger_k \neq 0,\ \sigma_k = -\sigma^S_k \end{cases}$$

for all k, where $\mu := \frac{1-\beta}{1+\beta} \in [0,1)$.

Proof. We apply Theorem 4.3 to the case $S = \{x^\dagger\}$. Note that $\sigma \in \mathbb{1}^S$ if and only if its support coincides with the support of $x^\dagger$, and that σ^S_k provides the sign of $x^\dagger_k$ for each k. Further, the normal cone in the definition of $S(\sigma)$ is $N_S(x^\dagger) = l^\infty$, which allows to choose $\xi = \sigma$ in that definition. We immediately obtain $S(\sigma) = \{x^\dagger\}$ for each $\sigma \in \mathbb{1}^S$ and therefore Theorem 4.3 states the the variational source condition (5.1) holds if and only if for each $\sigma \in \mathbb{1}^S$ there is some η with $\|\eta\| \leq \frac{\gamma}{1-\beta}$ such that

$$\begin{cases} [A^*\,\eta]_k \in [-\mu, \mu], & \text{if } x^\dagger_k = 0, \\ \sigma^S_k\,[A^*\,\eta]_k \leq \mu, & \text{if } x^\dagger_k \neq 0,\ \sigma_k = \sigma^S_k, \\ \sigma^S_k\,[A^*\,\eta]_k \geq 1, & \text{if } x^\dagger_k \neq 0,\ \sigma_k = -\sigma^S_k \end{cases} \tag{5.2}$$

for each k.

Now fix $\sigma \in \mathbb{1}^S$ and let $k_1, k_2, \ldots$ be an enumeration (finite or infinite) of all indices k satisfying $\sigma_k \neq 0$. Note that $x^\dagger_{k_n} \neq 0$ for all n. We prove the theorem by induction over n.

Let $\bar{\sigma} \in \mathbb{1}^S$ satisfy $\bar{\sigma}_{k_1} = \sigma^S_{k_1}$ and let $\tilde{\sigma}$ be the same except for $\tilde{\sigma}_{k_1} = -\sigma^S_{k_1}$. Then there are $\bar{\eta}$ and $\tilde{\eta}$ such that (5.2) holds with σ replaced by $\bar{\sigma}$ and $\tilde{\sigma}$, respectively. At index k_1 we have $\sigma^S_{k_1}[A^*\,\bar{\eta}]_{k_1} \leq \mu$ and $\sigma^S_{k_1}[A^*\,\tilde{\eta}]_{k_1} \geq 1$. Thus, there exists a convex combination $\eta^{(1)}(\bar{\sigma})$ of $\bar{\eta}$ and $\tilde{\eta}$ which satisfies $\sigma^S_{k_1}[A^*\,\eta^{(1)}(\bar{\sigma})]_{k_1} = \mu$ (if $\sigma_{k_1} = \sigma^S_{k_1}$) or $\sigma^S_{k_1}[A^*\,\eta^{(1)(\bar{\sigma})}]_{k_1} = 1$ (if $\sigma_{k_1} = -\sigma^S_{k_1}$). In addition, such an element $\eta^{(1)}(\bar{\sigma})$ satisfies (5.2) with σ replaced by $\bar{\sigma}$ for all other indices $k \neq k_1$.

Now let $\bar{\sigma} \in \mathbb{1}^S$ satisfy $\bar{\sigma}_{k_l} = \sigma_{k_l}$ for $l = 1, \ldots, n-1$ and $\bar{\sigma}_{k_n} = \sigma^S_{k_n}$. Further, let $\tilde{\sigma}$ be the same except for $\tilde{\sigma}_{k_n} = -\sigma^S_{k_n}$. Assume that there is $\eta^{(n-1)}(\bar{\sigma})$ such that (5.2) holds for all k and such that for $k_1, \ldots, k_{n-1}$ it holds with equality signs. The existence of such an $\eta^{(n-1)}(\bar{\sigma})$ has been shown above for $n = 2$. Again there are $\bar{\eta}$ and $\tilde{\eta}$ such that (5.2) holds with σ replaced by $\bar{\sigma}$ and $\tilde{\sigma}$, respectively. At index k_n we have $\sigma^S_{k_n}[A^*\,\bar{\eta}]_{k_n} \leq \mu$ and

$\sigma^S_{k_n}[A^*\,\tilde\eta]_{k_n} \geq 1$. Thus, there exists a convex combination $\eta^{(n)}(\bar\sigma)$ of $\bar\eta$ and $\tilde\eta$ which satisfies $\sigma^S_{k_n}[A^*\,\eta^{(n)}(\bar\sigma)]_{k_n} = \mu$ (if $\sigma_{k_n} = \sigma^S_{k_n}$) or $\sigma^S_{k_n}[A^*\,\eta^{(n)(\bar\sigma)}]_{k_n} = 1$ (if $\sigma_{k_n} = -\sigma^S_{k_n}$). In addition, such an element $\eta^{(n)}(\bar\sigma)$ satisfies (5.2) with σ replaced by $\bar\sigma$ for all other indices $k \neq k_n$.

So far we have shown that for each $\sigma \in \mathbb{1}^S$ and each n we can construct $\eta^{(n)}(\sigma)$ which satisfies (5.2), where we can replace inequality by equality signs at indices $k_1, \ldots, k_n$. Consequently we find η such that equality holds at all indices k at which $\sigma_k \neq 0$. □

Remark 5.2. Analogously to Remark 4.8 we can replace $\mathbb{1}^S$ in Theorem 5.1 by the set of all σ which satisfy $\sigma_k = \pm 1$ if $\sigma^S_k \neq 0$ and $\sigma_k = 0$ else.

Corollary 5.3. *Let $S = \{x^\dagger\}$. Then the variational source condition* (5.1) *on ℓ^1 is satisfied if and only if for each $\sigma \in \mathbb{1}^S$ with $\sigma_k \neq 0$ if $\sigma^S_k \neq 0$ there is some η in Y^* with $\|\eta\| \leq \frac{\gamma}{1+\beta}$ such that*

$$\begin{cases} [A^*\,\eta]_k = \sigma^S_k, & \text{if } x^\dagger_k \neq 0,\ \sigma_k = -\sigma^S_k, \\ [A^*\,\eta]_k \in [-\mu, \mu], & \text{if } x^\dagger_k = 0 \ \text{ or } \ x^\dagger_k \neq 0,\ \sigma_k = \sigma^S_k \end{cases} \tag{5.3}$$

for all k, where $\mu := \frac{1-\beta}{1+\beta} \in [0, 1)$.

Proof. This is a direct consequence of Theorem 5.1 (necessity) and Theorem 4.3 (sufficiency). □

Note that the condition $[A^*\,\eta]_k \in [-\mu, \mu]$ if $x^\dagger_k = 0$ in (5.3) and corresponding conditions in Theorems 4.3 and 5.1 are closely related to a property called *strict sparsity pattern* in [5, Definition 2] and *strong source condition* in [13, Condition 4.3].

6 Non-sparse solutions

We now extend Theorem 4.3 to solution sets S which may contain non-sparse solutions (cf. Remark 4.4), see [6] for a similar result in case of injective A. The aim is to obtain a variational source condition

$$\beta\,\mathrm{dist}(x, S) \leq \|x\| - \|S\| + \varphi(\|A\,x - A\,S\|) \quad \text{for all } x \in \ell^1 \tag{6.1}$$

with some concave index function φ, which depends on the decay of the solutions' components. Here, again, $\|S\|$ denotes the norm of the norm minimizing solutions.

A sufficient condition for such a variational source condition can be deduced from the characterization (4.3) in Theorem 1.1.

Theorem 6.1. *Assume that*

$$\lim_{n\to\infty} \sup_{x^\dagger\in S} \sum_{k>n} |x_k^\dagger| = 0 \tag{6.2}$$

and let $(\gamma_n)_{n\in\mathbb{N}}$ *be a sequence of positive numbers with*

$$\lim_{n\to\infty} \gamma_n = \infty.$$

Then the variational source condition (6.1) *on* ℓ^1 *is satisfied with*

$$\varphi(t) = \inf_{n\in\mathbb{N}} \left(2 \sup_{x^\dagger\in S} \sum_{k>n} |x_k^\dagger| + \gamma_n\, t \right)$$

if for each $n \in \mathbb{N}$, *each* $\sigma \in \mathbb{1}^S$ *and each* $x^\dagger \in S(\sigma)$ *there are a non-negative constant* γ_n *and some* η *in* Y^* *with* $\|\eta\| \leq \frac{\gamma_n}{1+\beta}$ *such that*

$$\begin{cases} [A^*\,\eta]_k \in [-\mu,\mu], & \text{if } \sigma_k^S = 0 \text{ or } k > n, \\ \sigma_k\,[A^*\,\eta]_k \leq \mu, & \text{if } \sigma_k^S \neq 0,\ x_k^\dagger = 0,\ \sigma_k \neq 0,\ k \leq n, \\ \sigma_k^S\,[A^*\,\eta]_k \leq \mu, & \text{if } x_k^\dagger \neq 0,\ \sigma_k = \sigma_k^S,\ k \leq n, \\ \sigma_k^S\,[A^*\,\eta]_k \geq 1, & \text{if } x_k^\dagger \neq 0,\ \sigma_k = -\sigma_k^S,\ k \leq n \end{cases}$$

for all k, *where* $\mu := \frac{1-\beta}{1+\beta} \in [0,1)$.

Proof. Fix x in ℓ^1. By Proposition 3.6 there are σ in $\mathbb{1}^S$ and $x^\dagger$ in $S(\sigma)$ such that x is in $M_{x^\dagger}(\sigma)$. Proposition 3.4 yields

$$\beta\,\mathrm{dist}(x,S) - \|x\| + \|x^\dagger\| = \beta\,\langle\sigma, x - x^\dagger\rangle + \beta \sum_{k:\,\sigma_k^S=0} |x_k| - \|x\| + \|x^\dagger\|.$$

This can be written as a sum

$$\beta\,\mathrm{dist}(x,S) - \|x\| + \|x^\dagger\| = \sum_{k\in\mathbb{N}} a_k$$

with a_k depending only on x_k and $x_k^\dagger$ and we have

$$a_k = \begin{cases} -(1-\beta)\,|x_k|, & \text{if } \sigma_k^S = 0, \\ -(1-\beta)\,|x_k - x_k^\dagger|, & \text{if } \sigma_k^S \neq 0,\ x_k^\dagger = 0,\ \sigma_k \neq 0 \\ & \text{or if } x_k^\dagger \neq 0,\ \sigma_k = \sigma_k^S, \\ -\beta\,\sigma_k^S\,x_k - |x_k| + (1+\beta)\,|x_k^\dagger|, & \text{if } x_k^\dagger \neq 0,\ \sigma_k = -\sigma_k^S, \\ 0 & \text{if } \sigma_k^S \neq 0,\ \sigma_k = 0. \end{cases}$$

Now let η be as in the theorem. Then

$$\gamma_n \,\|A\,x - A\,x^\dagger\| \geq -(1+\beta)\langle \eta, A\,x - A\,x^\dagger\rangle = -(1+\beta)\langle A^*\,\eta, x - x^\dagger\rangle$$

and, because $x \in M_{x^\dagger}(\sigma)$, we see

$$\gamma_n \,\|A\,x - A\,x^\dagger\| \geq -(1+\beta) \sum_{k:\,\sigma_k^S \neq 0} [A^*\,\eta]_k\,\sigma_k\,|x_k - x_k^\dagger| - (1+\beta) \sum_{k:\,\sigma_k^S = 0} [A^*\,\eta]_k\,x_k.$$

Using the properties of $A^*\,\eta$ we obtain

$$2 \sum_{k>n} |x_k^\dagger| - \gamma_n \,\|A\,x - A\,x^\dagger\| \geq \sum_{n \in \mathbb{N}} b_n$$

with

$$b_k \geq \begin{cases} -(1-\beta)\,|x_k|, & \text{if } \sigma_k^S = 0, \\ -(1-\beta)\,|x_k - x_k^\dagger|, & \text{if } \sigma_k^S \neq 0,\ x_k^\dagger = 0,\ \sigma_k \neq 0,\ k \leq n, \\ & \text{or if } x_k^\dagger \neq 0,\ \sigma_k = \sigma_k^S,\ k \leq n, \\ (1+\beta)\,|x_k - x_k^\dagger|, & \text{if } x_k^\dagger \neq 0,\ \sigma_k = -\sigma_k^S,\ k \leq n, \\ 2\,|x_k^\dagger| - (1-\beta)\,|x_k - x_k^\dagger|, & \text{if } \sigma_k^S \neq 0,\ \sigma_k \neq 0,\ k > n, \\ 2\,|x_k^\dagger| & \text{if } \sigma_k^S \neq 0,\ \sigma_k = 0. \end{cases}$$

It is not hard to show that $a_k \leq b_k$ for all k. Thus,

$$\beta \operatorname{dist}(x, S) - \|x\| + \|x^\dagger\| \leq 2 \sum_{k>n} |x_k^\dagger| - \gamma_n \,\|A\,x - A\,x^\dagger\|.$$

Taking the supremum over all $x^\dagger$ and the infimum over all n the variational source condition (6.1) is proven and it remains to show that the function φ is a concave index function.

Obviously, φ is non-negative. As an infimum of affine functions it further is concave and upper semi-continuous. Thus, φ is continuous on the interior $(0,\infty)$ of its domain. Together with

$$\varphi(0) = \inf_{n \in \mathbb{N}} \left(2 \sup_{x^\dagger \in S} \sum_{k>n} |x_k^\dagger| \right) = 0$$

we obtain continuity on $[0,\infty)$. To prove that φ is strictly increasing we take $t_1, t_2 \in [0,\infty)$ with $t_1 < t_2$. The infimum in the definition of $\varphi(t_2)$ is attained at some n_2. Thus,

$$\varphi(t_1) \leq 2 \sup_{x^\dagger \in S} \sum_{k>n_2} |x_k^\dagger| + \gamma_{n_2}\,t_1 < \varphi(t_2).$$

□

Note that condition (6.2) may be violated in some cases. For example if S is the convex hull of the standard unit sequences $\{e^{(1)}, e^{(2)}, \ldots\}$ in ℓ^1, then

$$\sup_{x^\dagger \in S} \sum_{k>n} |x_k^\dagger| \geq \sum_{k>n} \left|e_k^{(n+1)}\right| = 1 \tag{6.3}$$

for all n.

From a variational source condition (6.1) we obtain the error estimate

$$\operatorname{dist}(x_\alpha^\delta, S) \leq c\, \varphi(\delta)$$

for all sufficiently small $\delta > 0$ with some constant $c > 0$, if the regularization parameter α is chosen appropriately, for example proportional to $\frac{1}{\varphi'(\delta)}$ (if φ is differentiable) or by the two-sided discrepancy principle (see [8] for both variants) or by the sequential discrepancy principle [16, 1]. Here, x_α^δ again denotes a minimizer of the Tikhonov-type functional (1.2).

7 Examples

We provide two very simple examples and a more realistic one to show how the developed results can be applied to non-injective operators. The first example considers multiple norm minimizing solutions. The second and the third one have only one norm minimizing solution and they show, by the way, that the constant β in a variational source condition cannot be chosen arbitrarily close to one.

Example 7.1. For the first example take $Y := \mathbb{R}$, $y^\dagger := 1$ and

$$A\,x := x_1 + x_2.$$

Then the set of solutions is $L = \{x \in \ell^1 : x_2 = 1 - x_1\}$ and the set of norm minimizing solutions is

$$S = \{x \in \ell^1 : x_2 = 1 - x_1,\ x_1 \in [0,1],\ x_k = 0 \text{ for } k > 2\}.$$

Further,

$$A^*\,\eta := (\eta, \eta, 0, \ldots).$$

Figure 1 provides a sketch of the geometric situation.

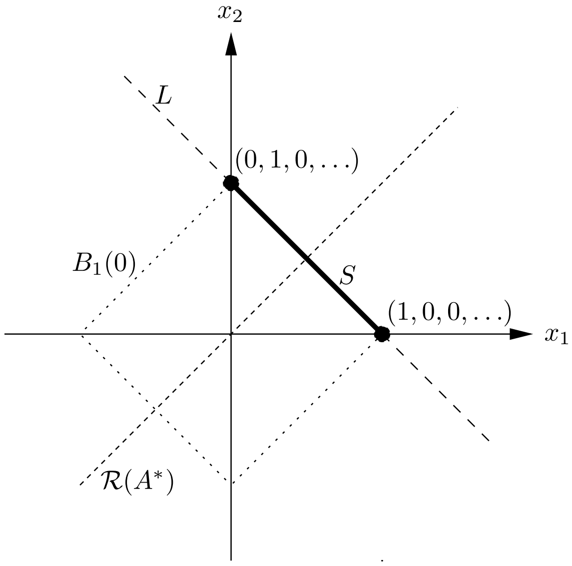

Figure 1: Sketch for the first example of the x_1-x_2-plane with set S of norm minimizing solutions, set L of all solutions, unit ball $B_1(0)$ and 'subspace' $\mathcal{R}(A^*)$.

We now verify condition (4.3) in Theorem 4.3 with $\beta = 1$. First note that $\sigma^S = (1, 1, 0, \ldots)$ and that by Remark 4.8 we only have to consider

$$\begin{aligned}
\sigma^{(1)} &= (1, 1, 0, \ldots), \\
\sigma^{(2)} &= (1, -1, 0, \ldots), \\
\sigma^{(3)} &= (-1, 1, 0, \ldots), \\
\sigma^{(4)} &= (-1, -1, 0, \ldots).
\end{aligned}$$

The corresponding subsets $S(\sigma^{(i)})$ of S are the faces of S looking in direction $\sigma^{(i)}$, that is,

$$\begin{aligned}
S(\sigma^{(1)}) &= S, \\
S(\sigma^{(2)}) &= \{(1, 0, 0, \ldots)\}, \\
S(\sigma^{(3)}) &= \{(0, 1, 0, \ldots)\}, \\
S(\sigma^{(4)}) &= S.
\end{aligned}$$

Taking into account Remark 4.7, only $\sigma^{(4)}$ remains to be considered. Here condition (4.3) is equivalent to $\eta \geq 1$, which is obviously satisfied when

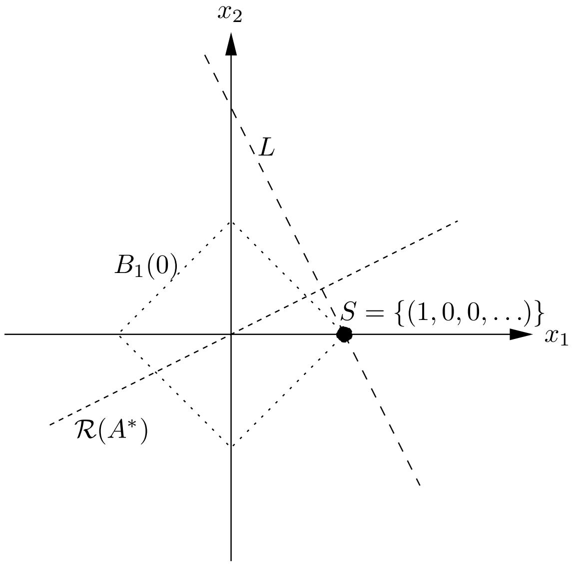

Figure 2: Sketch for the second example of the x_1-x_2-plane with set S of norm minimizing solutions, set L of all solutions, unit ball $B_1(0)$ and 'subspace' $\mathcal{R}(A^*)$.

choosing $\eta = 1$ (by Remark 4.6 we only have to check the condition for $x^\dagger = (\frac{1}{2}, \frac{1}{2}, 0, \ldots)$ for example). Consequently, Theorem (4.3) applies to our first example and yields convergence rates although the operator A is not injective.

Example 7.2. For the second example take $Y := \mathbb{R}$, $y^\dagger := 1$ and

$$A\,x := x_1 + \frac{x_2}{2}.$$

Then the set of solutions is $L = \{x \in \ell^1 : x_2 = 2 - 2\,x_1\}$ and there is only one norm minimizing solution

$$S = \{(1, 0, \ldots)\}.$$

Further,

$$A^*\,\eta := (\eta, \frac{\eta}{2}, 0, \ldots).$$

Figure 2 provides a sketch of the geometric situation.

We now verify condition (5.3) in Corollary 5.3. First note that $\sigma^S = (1, 0, 0, \ldots)$ and so we only have to consider

$$\sigma^{(1)} = (1, 0, \ldots) \quad \text{and} \quad \sigma^{(2)} = (-1, 0, \ldots).$$

For $\sigma^{(1)}$ condition (5.3) is satisfied by $\eta = 0$. For $\sigma^{(2)}$ the condition is equivalent to

$$\eta = 1 \quad \text{and} \quad -\frac{1-\beta}{1+\beta} \leq \frac{\eta}{2} \leq \frac{1-\beta}{1+\beta},$$

which is only possible if $\beta \leq \frac{1}{3}$. Consequently, Corollary 5.3 yields a variational source condition with $\beta \leq \frac{1}{3}$ and corresponding convergence rates for our second non-injective example.

If we had chosen the solution set L to be parallel to the x_2-axis, then $\beta = 1$ would be possible. On the other hand, the more slanting the set L in Figure 2 is, the closer β has to be to zero. The limit case where only $\beta = 0$ would be possible then coincides with the situation discussed in Example 7.1. Generalizing this observation we may say that the constant β in a variational source condition is a 'measure' for paraxiality of the nullspace of A or the range of A^*.

Example 7.3. Let $Y := \ell^2$ and let $A := P\,V\,\tilde{A}\,U$ be the composition of the Fourier synthesis operator $U : \ell^1 \to L^2(0,1)$ defined by

$$(U\,x)(t) := x_1 + \sqrt{2}\sum_{l\in\mathbb{N}} x_{2\,l}\,\cos(2\,\pi\,l\,t) + \sqrt{2}\sum_{l\in\mathbb{N}} x_{2\,l+1}\,\sin(2\,\pi\,l\,t)$$

for $t \in (0,1)$, the integration operator $\tilde{A} : L^2(0,1) \to L^2(0,1)$ defined by

$$(\tilde{A}\,\tilde{x})(s) := \int_0^s \tilde{x}(t)\,\mathrm{d}t$$

for $s \in (0,1)$, the Fourier transform $V : L^2(0,1) \to \ell^2$ defined by

$$[V\,\tilde{y}]_1 := \int_0^1 \tilde{y}(s)\,\mathrm{d}s,$$

$$[V\,\tilde{y}]_{2\,l} := \int_0^1 \tilde{y}(s)\,\sqrt{2}\,\cos(2\,\pi\,l\,s)\,\mathrm{d}s,$$

$$[V\,\tilde{y}]_{2\,l+1} := \int_0^1 \tilde{y}(s)\,\sqrt{2}\,\sin(2\,\pi\,l\,s)\,\mathrm{d}s$$

for $l \in \mathbb{N}$, and the projection $P : \ell^2 \to \ell^2$ defined by

$$
\begin{aligned}
[P\,\bar{y}]_1 &= \bar{y}_1, \\
[P\,\bar{y}]_{2\,l} &= 0, \\
[P\,\bar{y}]_{2\,l+1} &= \bar{y}_{2\,l} + \bar{y}_{2\,l+1}
\end{aligned}
$$

for $l \in \mathbb{N}$. In other words, we aim to reconstruct derivatives of functions from incomplete Fourier data under the a priori information that the derivatives are sparse or almost sparse with respect to the Fourier basis. Only sums of the data's cosine and sine coefficients are available, making the operator highly non-injective.

The operator $A : \ell^1 \to \ell^2$ turns out to map a sequence x to a sequence $A\,x$ defined by

$$
\begin{aligned}
[A\,x]_1 &= \frac{1}{2}\,x_1 + \sum_{l \in \mathbb{N}} \frac{1}{\sqrt{2}\,\pi\,l}\,x_{2\,l+1}, \\
[A\,x]_{2\,l} &= 0, \\
[A\,x]_{2\,l+1} &= \frac{1}{2\,\pi\,l}\left(-\sqrt{2}\,x_1 + x_{2\,l} - x_{2\,l+1}\right)
\end{aligned}
$$

for $l \in \mathbb{N}$. The adjoint $A^* = P^*\,V^*\,\tilde{A}^*\,U^* : \ell^2 \to \ell^\infty$ thus is given by

$$
\begin{aligned}
[A^*\,\eta]_1 &= \frac{1}{2}\,\eta_1 - \sum_{l \in \mathbb{N}} \frac{1}{\sqrt{2}\,\pi\,l}\,\eta_{2\,l+1}, \\
[A^*\,\eta]_{2\,l} &= \frac{1}{2\,\pi\,l}\,\eta_{2\,l+1}, \\
[A^*\,\eta]_{2\,l+1} &= \frac{1}{2\,\pi\,l}\left(\sqrt{2}\,\eta_1 - \eta_{2\,l+1}\right)
\end{aligned}
$$

for $l \in \mathbb{N}$. The null space of A is

$$
\mathcal{N}(A) = \left\{(0, w_1, w_1, w_2, w_2, \ldots) \in \ell^1 : \sum_{l \in \mathbb{N}} \frac{1}{l}\,w_l = 0\right\}.
$$

We look at the exact right-hand side $y^\dagger := (0, 0, -\frac{1}{2\,\pi}, 0, -\frac{1}{4\,\pi}, 0, 0, \ldots)$. One easily sees that $x^\dagger = (0, 1, 0, 1, 0, 0, \ldots)$ is a corresponding solution and it turns out that this is the only 1-norm minimizing solution, that is, $S = \{x^\dagger\}$ (here some very basic but longish calculations are necessary).

To verify the assumptions of Corollary 5.3 we have to show that the elements

$$\begin{aligned}
\sigma^{(1)} &= (0,1,0,1,0,0,\ldots),\\
\sigma^{(2)} &= (0,1,0,-1,0,0,\ldots),\\
\sigma^{(3)} &= (0,-1,0,1,0,0,\ldots),\\
\sigma^{(4)} &= (0,-1,0,-1,0,0,\ldots)
\end{aligned}$$

satisfy condition (5.3) for some η. We only mention how to choose η in each case and do not provide all details of the (basic but longish) calculations. Since $\sigma^S = (0,1,0,1,0,0,\ldots)$ we may choose $\eta = 0$ in case of $\sigma^{(1)}$. For $\sigma^{(2)}$ one possible choice is

$$\eta = \left(2\sqrt{2} + \frac{4\pi - 4}{2\pi + \sqrt{2}}, 0, 0, 0, 4\pi, 0, 0, \ldots\right) \qquad \text{if} \quad \mu \geq \frac{2\pi - 2}{2\pi + \sqrt{2}}.$$

Note that for smaller μ there is no η satisfying (5.3) if $\sigma = \sigma^{(2)}$. For $\sigma^{(3)}$ one possible choice is

$$\eta = \left(2\sqrt{2} + \frac{2\pi - 4}{\pi + \sqrt{2}}, 0, 2\pi, 0, 0, \ldots\right) \qquad \text{if} \quad \mu \geq \frac{\pi - 2}{\pi + \sqrt{2}}.$$

Again for smaller μ there is no η satisfying (5.3) if $\sigma = \sigma^{(3)}$. Finally, for $\sigma^{(4)}$ we may choose

$$\eta = \left(\frac{12}{5}\pi, 0, 3\pi, 0, 4\pi, 0, 0, \ldots\right) \qquad \text{if} \quad \mu \geq \frac{2}{5}$$

and for smaller μ there is no η.

Thus, if

$$\mu \geq \frac{2\pi - 2}{2\pi + \sqrt{2}} \approx 0.5564$$

we obtain a variational source condition with

$$\beta \leq \frac{2 + \sqrt{2}}{4\pi - 2 + \sqrt{2}} \approx 0.2850$$

and corresponding convergence rates.

Playing around with this example one also sees that the more non-zero components $x^\dagger$ has the smaller is the best possible β in the variational source condition. Since β enters the $\mathcal{O}$-constant c in the convergence rate result

(1.4) as a factor $\frac{1}{\beta}$ (cf. [8, Theorem 4.11]), the $\mathcal{O}$-constant becomes greater if $x^\dagger$ is 'less sparse'. If the number of non-zero components in $x^\dagger$ goes to infinity, then β goes to zero and consequently the $\mathcal{O}$-constant blows up to infinity. Such situations then can be handled by Theorem 6.1, resulting in slower convergence rates.

References

[1] S. W. Anzengruber, B. Hofmann, and P. Mathé. Regularization properties of the sequential discrepancy principle for Tikhonov regularization in Banach spaces. *Appl. Anal.*, 93:1382–1400, 2014.

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

[3] R. I. Boţ and B. Hofmann. The impact of a curious type of smoothness conditions on convergence rates in ℓ^1-regularization. *Eurasian Journal of Mathematical and Computer Applications*, 1:29–40, 2013.

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

[5] K. Bredies and D. A. Lorenz. Linear Convergence of Iterative Soft-Thresholding. *Journal of Fourier Analysis and Applications*, 14(5–6):813–837, 2008.

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

[7] E. J. Candès, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59:1207–1223, 2006.

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

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

[10] J. Flemming, B. Hofmann, and I. Veselić. On ℓ^1-regularization in light of Nashed's ill-posedness concept. *Comput. Methods Appl. Math.*, 15:279–289, 2015.

[11] M. Grasmair. Well-posedness and convergence rates for sparse regularization with sublinear ℓ^q penalty term. *Inverse Probl. Imaging*, 33:383–387, 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, M. Haltmeier, and O. Scherzer. Necessary and sufficient conditions for linear convergence of ℓ^1-regularization. *Comm. Pure Appl. Math.*, 64:161–182, 2011.

[14] T. Hein and B. Hofmann. Approximate source conditions for nonlinear ill-posed problems—chances and limitations. *Inverse Problems*, 25(3):035033 (16pp), 2009.

[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*, 23:987–1010, 2007.

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

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

[18] C. Pöschl. *Tikhonov Regularization with General Residual Term.* PhD thesis, University of Innsbruck, Innsbruck, Austria, October 2008. Corrected version.

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

[20] 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.