# Necessary and Sufficient Conditions for Linear Convergence of $\ell^1$-Regularization

MARKUS GRASMAIR
*Computational Science Center, University of Vienna, Austria*

MARKUS HALTMEIER
*Computational Science Center, University of Vienna, Austria*

AND

OTMAR SCHERZER
*Computational Science Center, University of Vienna, Austria*
*RICAM, University of Linz, Austria*

**Abstract**

Motivated from the theoretical and practical results in compressed sensing, efforts have been undertaken by the inverse problems community to derive analogous results, for instance linear convergence rates, for Tikhonov regularization with $\ell^1$-penalty term for the solution of ill-posed equations. Conceptually, the main difference between these two fields is that regularization in general is an unconstrained optimization problem, while in compressed sensing a constrained one is used. Since the two methods have been developed in two different communities, the theoretical approaches to them appear to be rather different: In compressed sensing, the *restricted isometry property* seems to be central for proving linear convergence rates, whereas in regularization theory *range* or *source* conditions are imposed. The paper gives a common meaning to the seemingly different conditions and puts them into perspective with the conditions from the respective other community. A particularly important observation is that the range condition together with an injectivity condition is weaker than the restricted isometry property. Under the weaker conditions, linear convergence rates can be proven for compressed sensing and for Tikhonov regularization. Thus existing results from the literature can be improved based on a unified analysis. In particular the range condition is shown to be the weakest possible condition that permits the derivation of linear convergence rates for Tikhonov regularization with a–priori parameter choice.

**Key words.** $\ell^1$-Regularization, Sparsity, Convergence Rates, Compressed Sensing, Inverse Problems, Ill-posed Equations, Regularization, Residual Method, Tikhonov Regularization.

**AMS subject classifications.** 65J20; 47J06; 49J27; 65J22

## 1 Introduction

The problem of solving ill-posed linear operator equations of the form

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

is constantly encountered in various mathematical applications. If the available data $y$ are a perturbation of some true data $y^\dagger$, the ill-posedness prohibits a direct solution by inverting $\mathbf{A}$. As a remedy, regularization methods are used, which calculate from $y$ an approximation $x^\delta$ of the solution of (1.1), thereby also permitting a certain defect $\mathbf{A}x^\delta - y \neq 0$.

Regularization methods rely on a regularization functional R that is adapted to the particular application and the solution $x^\dagger$ to be recovered. Both in theory and practice, estimates for the data error in $y$ have to be available, i.e., the constant $\delta$ in the inequality $\|y^\dagger - y\| \leq \delta$ has to be known. Using an appropriate regularization functional and appropriately taking into account $\delta$, makes regularization methods stable with respect to data perturbations and convergent to $x^\dagger$ when $\delta$ tends to zero (see for instance [19, 27, 34]).

In this paper, we consider three kinds of regularization methods:

*Residual method*: Fix $\tau \geq 1$ and solve the constrained minimization problem

$$\mathrm{R}(x) \to \min \qquad \text{subject to } \|\mathbf{A}x - y\| \leq \tau\delta\,. \tag{1.2}$$

*Tikhonov regularization with discrepancy principle*: Fix $\tau \geq 1$ and minimize the *Tikhonov functional*

$$\mathrm{T}_{\alpha,y}(x) := \|\mathbf{A}x - y\|^2 + \alpha \mathrm{R}(x)\,, \tag{1.3}$$

where $\alpha > 0$ is chosen in such a way that Morozov's discrepancy principle is satisfied, i.e., $\|\mathbf{A}x_\alpha - y\| = \tau\delta$ with $x_\alpha \in \arg\min_x \mathrm{T}_{\alpha,y}(x)$.

*Tikhonov regularization with a–priori parameter choice*: Fix $C > 0$ and minimize the Tikhonov functional (1.3) with a parameter choice

$$\alpha = C\delta\,. \tag{1.4}$$

Classical regularization methods assume sufficient smoothness of the solution $x^\dagger$ of (1.1) and therefore use a squared Sobolev norm for the regularization term R. In the recent years, however, *sparse regularization* has been established as a powerful alternative to the standard methods [14, 1, 4, 23, 24, 28]. Here the core assumption is that the solution $x^\dagger$ has a sparse expansion with respect to some given basis or frame $(\phi_\lambda)_{\lambda\in\Lambda}$ of the domain $X$ of the operator $\mathbf{A}$. In this context, *sparsity* signifies that only a few coefficients $\langle \phi_\lambda, x^\dagger\rangle$ of the solution $x^\dagger$ are non-zero. This criterion of sparsity can be implemented with all three regularization methods by using the regularization term

$$\mathrm{R}(x) := \|x\|_{\ell^1} := \sum_{\lambda\in\Lambda} |\langle \phi_\lambda, x\rangle|\,. \tag{1.5}$$

Sparsity is also the base of the active field of *compressed sensing* [10, 12, 16]. In this application, the nullspace of $\mathbf{A}$, $\ker(\mathbf{A})$, is assumed to be high dimensional, and as a consequence, solving (1.1) is a vastly under-determined problem. Still, under certain conditions it is possible to recover $x^\dagger$ exactly, by solving the convex minimization problem

$$\mathrm{R}(x) = \|x\|_{\ell^1} \to \min \quad \text{such that } \mathbf{A}x = y^\dagger . \tag{1.6}$$

Indeed, it has been shown that, under certain assumptions, the solution of (1.6) is the sparsest solution of $\mathbf{A}x = y^\dagger$, that is, the solution with the smallest number of non-zero coefficients (see [10, 15]). In [12], this *exact reconstruction property* is shown under the assumptions that the true solution $x^\dagger$ is sufficiently sparse and the operator $\mathbf{A}$ is close to being an isometry on certain low dimensional subspaces of $X$. The latter condition is generally termed the *restricted isometry property* of $\mathbf{A}$.

The same property allows the derivation of stability results for the constrained minimization problem (1.2) with $\ell^1$-regularization term (see [11, 16, 22, 36]), i.e., for

$$\mathrm{R}(x) = \|x\|_{\ell^1} \to \min \quad \text{such that } \|\mathbf{A}x - y\| \le \tau\delta . \tag{1.7}$$

In this context the restricted isometry property implies the existence of a constant $c > 0$, only depending on the number $s$ of non-zero coefficients of the solution $x^\dagger$ of (1.1), such that the solution of the constrained minimization problem $x^\delta$ of (1.7) satisfies (see [11])

$$\|x^\delta - x^\dagger\| \le c\delta . \tag{1.8}$$

In [24], the linear convergence rate (1.8) for the $\ell^1$ residual method has been derived recently under different assumptions. Instead of restricted isometry, the following two assumptions are postulated: first, restricted injectivity of $\mathbf{A}$, and second, that $x^\dagger$ satisfies a *range condition*, which is commonly imposed in regularization theory. In the general context, for arbitrary convex regularization functionals, this condition states that there exists some $\eta$ such that

$$\mathbf{A}^*\eta \in \partial\mathrm{R}(x^\dagger) . \tag{1.9}$$

Here $\mathbf{A}^*$ denotes the adjoint of the linear operator $\mathbf{A}$, and $\partial\mathrm{R}(x^\dagger)$ denotes the subdifferential of the convex functional R at $x^\dagger$.

Also in the case of Tikhonov regularization with a–priori parameter choice $\alpha = C\delta$, a linear convergence rate has been derived under the assumptions of range condition (1.9) and restricted injectivity (see [23, 34]).

These results underline that the convergence analysis for both methods, constrained and unconstrained, can be performed under the same conditions. The relations between constrained and unconstrained regularization go even beyond this connection. Indeed, it can be shown that the constrained minimization problem (1.7) is equivalent to Tikhonov regularization, when the regularization parameter $\alpha$ is chosen according to Morozov's discrepancy principle (see Proposition 2.2

below). As a consequence, the convergence rates results for constrained regularization (as for instance in [6, 35, 36]) carry over directly to Tikhonov regularization with Morozov's parameter selection criterion.

In this paper we carefully investigate the conditions under which linear convergence rates for constrained and unconstrained sparsity regularization have been derived. In particular, we study the relation between the range condition (1.9), restricted injectivity, and the restricted isometry property from compressed sensing. Having shown that restricted isometry can be considered a special case of a range condition, we consider the question whether the latter is necessary for obtaining linear rates (1.8). As a partial answer, we state in Theorem 4.7 that, for Tikhonov regularization with a–priori parameter choice strategy, the range condition combined with restricted injectivity is indeed not only sufficient, but also necessary for a linear convergence rate. Thus, the stated conditions are the weakest possible ones for linear convergence rates.

As a consequence of the previous discussion, most existing linear convergence rates results for compressed sensing can be generalized by replacing the assumption of restricted isometry property by the assumptions of restricted injectivity and the range condition.

## 2 Notational Preliminaries

Let $X$ and $Y$ be separable Hilbert spaces with inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$, and let $\mathbf{A}\colon X \to Y$ be a bounded linear operator. Assume that we are given an orthonormal basis $(\phi_\lambda)_{\lambda\in\Lambda} \subset X$ of the space $X$. Here, $\Lambda$ is some index set; since $X$ is separable, it follows that $\Lambda$ can at most be countable.

By $\mathrm{R}\colon X \to \mathbb{R} \cup \{\infty\}$ we denote the $\ell^1$-regularization term defined in (1.5). For $y^\dagger \in \operatorname{ran}(\mathbf{A})$, a solution of (1.6) is called an R-minimizing solution of the equation $\mathbf{A}x = y^\dagger$. Finally, for $y \in Y$ and $\alpha > 0$, we denote by $\mathrm{T}_{\alpha,y}\colon X \to \mathbb{R}\cup\{\infty\}$ the $\ell^1$-Tikhonov functional defined by (1.3).

We use the following definitions of sparsity:

**Definition 2.1.** For $x \in X$, the set

$$\operatorname{supp}(x) := \{\lambda \in \Lambda : \langle \phi_\lambda, x\rangle \neq 0\}$$

denotes the *support* of $x$ with respect to the basis $(\phi_\lambda)_{\lambda\in\Lambda} \subset X$. If $|\operatorname{supp}(x)| \le s$ for some $s \in \mathbb{N}$, then the element $x$ is called *s-sparse*. It is called *sparse*, if it is $s$-sparse for some $s \in \mathbb{N}$, that is, $|\operatorname{supp}(x)| < \infty$.

For a subset of indices $\Lambda' \subset \Lambda$, we denote by

$$X_{\Lambda'} := \overline{\operatorname{span}\{\phi_\lambda : \lambda \in \Lambda'\}} \subset X$$

the (closed) subspace spanned of all basis elements $\phi_\lambda$ with $\lambda \in \Lambda'$. With the restriction of the inner product on $X$ to $X_{\Lambda'}$, this space is again a Hilbert space.

Moreover, we denote by $\mathrm{i}_{\Lambda'}\colon X_{\Lambda'} \to X$ the embedding defined by $\mathrm{i}_{\Lambda'} x = x$. Its adjoint operator is the projection $\pi_{\Lambda'}\colon X \to X_{\Lambda'}$, which is defined by

$$\pi_{\Lambda'}(x) = \sum_{\lambda \in \Lambda'} \langle \phi_\lambda, x \rangle \phi_\lambda \; .$$

In addition, we denote for an operator $\mathbf{A}\colon X \to Y$ by $\mathbf{A}_{\Lambda'} := \mathbf{A} \circ \mathrm{i}_{\Lambda'}\colon X_{\Lambda'} \to Y$ its restriction to $X_{\Lambda'}$. Then the adjoint of $\mathbf{A}_{\Lambda'}$ is an operator

$$\mathbf{A}^*_{\Lambda'} := (\mathbf{A}_{\Lambda'})^*\colon Y \to X_{\Lambda'} \; .$$

Before we present some linear convergence rates results, we first cite the equivalence of the residual method and Tikhonov regularization with a–posteriori parameter choice strategy. For simplicity, we assume in the following that the parameter $\tau$ required for the residual method and for the application of Morozov's discrepancy principle equals 1.

**Proposition 2.2.** *Assume that the operator $\mathbf{A}\colon X \to Y$ has dense range. Then the residual method and Tikhonov regularization with an a–posteriori parameter choice by means of the discrepancy principle are equivalent in the following sense:*

*Let $y \in Y$ and $\delta > 0$ satisfy $\|y\| > \delta$. Then $x^\delta$ solves the constrained problem (1.2), if and only if $\|\mathbf{A}x^\delta - y\| = \delta$ and there exists $\alpha > 0$ such that $x^\delta$ minimizes (1.3).*

*Proof.* This follows by applying [27, Theorems 3.5.2, 3.5.5] where $D = \ell^1(\Lambda) \subset \ell^2(\Lambda)$ and $L\colon D \to \ell^1(\Lambda)$ is the identity operator on $\ell^1(\Lambda)$. □

The condition $\|y\| > \delta$ appears frequently in regularization theory. It states that the signal to noise ratio $\|y\|/\delta$ must be greater that 1. In other words, more information has to be contained in the data than in the noise.

## 3 Sparse Regularization

In this section we review, and slightly refine, the convergence rates results for $\ell^1$–regularization, which have been derived in [23, 24]. In addition, we compare these results with general convergence rates results with respect to the *Bregman distance* (see [6, 23, 24, 25, 26, 28, 29, 31, 32, 34]).

To that end, we require the notion of the *subdifferential* of the convex function $\mathrm{R}$ at $x \in X$, denoted by $\partial\mathrm{R}(x) \subset X$. It consists of all elements $\xi \in X$, called *subgradients*, satisfying $\langle \phi_\lambda, \xi \rangle = \operatorname{sign}\bigl(\langle \phi_\lambda, x \rangle\bigr)$ for $\lambda \in \operatorname{supp}(x)$ and $\langle \phi_\lambda, \xi \rangle \in [-1, 1]$ for $\lambda \notin \operatorname{supp}(x)$. Every subgradient being an element of the Hilbert space $X$, it follows that the sequence of coefficients $(\langle \phi_\lambda, \xi \rangle)_\lambda$ is square summable. In particular, the value $\langle \phi_\lambda, \xi \rangle = \pm 1$ can be attained only finitely many times. Thus the subdifferential $\partial\mathrm{R}(x)$ is non-empty if and only if $x$ is sparse.

**Lemma 3.1.** *Let $x^\dagger \in X$ satisfy $\mathbf{A}x^\dagger = y^\dagger$. Assume that $\mathbf{A}^*\eta \in \partial\mathrm{R}(x^\dagger)$ for some $\eta \in Y$. Then $x^\dagger$ is an $\mathrm{R}$-minimizing solution of the equation $\mathbf{A}x = y^\dagger$.*

*Proof.* This is a special case of [18, Chap. III, Prop. 4.1], applied to the function $J(x,\mathbf{A}x) = \mathrm{R}(x)$ if $\mathbf{A}x = y^\dagger$ and $J(x,\mathbf{A}x) = +\infty$ else. □

**Definition 3.2.** Let $x^\dagger$, $x \in X$ and assume that $\partial\mathrm{R}(x^\dagger) \neq \emptyset$. For $\xi \in \partial\mathrm{R}(x^\dagger)$ the *Bregman distance* of $x$ and $x^\dagger$ with respect to $\xi$ and R is defined as

$$D_\xi^{\mathrm{R}}(x,x^\dagger) := \mathrm{R}(x) - \mathrm{R}(x^\dagger) - \langle \xi, x - x^\dagger \rangle \,.$$

*Remark* 3.3. Because the functional R is positively homogeneous, it follows that $\mathrm{R}(x^\dagger) = \langle \xi, x^\dagger \rangle$ for every $\xi \in \partial\mathrm{R}(x^\dagger)$. Thus, the Bregman distance of $x$ and $x^\dagger$ with respect to $\xi$ simplifies to

$$D_\xi^{\mathrm{R}}(x,x^\dagger) = \mathrm{R}(x) - \langle \xi, x \rangle = \sum_{\lambda\in\Lambda} \Big( |\langle \phi_\lambda, x\rangle| - \langle \phi_\lambda, \xi \rangle \langle \phi_\lambda, x \rangle \Big) \,.$$

The following two results are concerned with convergence rates with respect to the Bregman distance for the residual method and for Tikhonov regularization with a–priori parameter selection (1.4). The first lemma has been derived in [6, Theorem 3] for arbitrary convex functionals R (see also [24, Section 5]). It can be applied to obtain convergence rates for the residual method and, by means of Proposition 2.2, the same convergence rate holds for Tikhonov regularization with an a–posteriori parameter choice in form of the discrepancy principle.

**Lemma 3.4.** *Let $x^\dagger \in X$ satisfy $\mathbf{A}x^\dagger = y^\dagger$ and assume that $\mathbf{A}^*\eta \in \partial\mathrm{R}(x^\dagger)$ for some $\eta \in Y$. Let $y \in Y$ satisfy $\|y - y^\dagger\| \leq \delta$ for some $\delta \geq 0$. Then*

$$D_{\mathbf{A}^*\eta}^{\mathrm{R}}(x^\delta, x^\dagger) \leq 2\|\eta\|\delta \tag{3.1}$$

*for every*

$$x^\delta \in \arg\min\{\mathrm{R}(x) : \|\mathbf{A}x - y\| \leq \delta\} \,.$$

*Proof.* Because $\|\mathbf{A}x^\dagger - y\| = \|y - y^\dagger\| \leq \delta$, the definition of $x^\delta$ implies that

$$\mathrm{R}(x^\delta) \leq \mathrm{R}(x^\dagger) \,.$$

Thus,

$$\begin{aligned} D_{\mathbf{A}^*\eta}^{\mathrm{R}}(x^\delta, x^\dagger) &= \mathrm{R}(x^\delta) - \mathrm{R}(x^\dagger) - \langle \mathbf{A}^*\eta, x^\delta - x^\dagger \rangle \leq -\langle \eta, \mathbf{A}(x^\delta - x^\dagger) \rangle \\ &\leq \|\eta\| \|\mathbf{A}(x^\delta - x^\dagger)\| \leq \|\eta\| (\|\mathbf{A}x^\delta - y\| + \|y - y^\dagger\|) \leq 2\|\eta\|\delta \,. \end{aligned}$$

Here, the last estimate follows again from the definition of $x^\delta$. □

Next we state the corresponding result for Tikhonov Regularization with a–priori parameter choice. The given estimate is a slight improvement over [34, Proposition 3.41] (see also the analogous estimates in [19, Theorem 10.4]).

**Lemma 3.5.** *Let $x^\dagger \in X$ satisfy $\mathbf{A}x^\dagger = y^\dagger$. Assume that $\mathbf{A}^*\eta \in \partial \mathrm{R}(x^\dagger)$ for some $\eta \in Y$. Moreover, let $\alpha > 0$ and assume that $y \in Y$ satisfies $\|y - y^\dagger\| \leq \delta$ for some $\delta \geq 0$. Then*

$$\|\mathbf{A}x_\alpha - y\| \leq \delta + \alpha\|\eta\|\,, \tag{3.2}$$

$$D^{\mathrm{R}}_{\mathbf{A}^*\eta}(x_\alpha, x^\dagger) \leq \frac{(\delta + \alpha\|\eta\|/2)^2}{\alpha}\,, \tag{3.3}$$

*for every $x_\alpha \in \arg\min \mathrm{T}_{\alpha,y}$.*

*Proof.* By definition of $x_\alpha$, we have

$$\|\mathbf{A}x_\alpha - y\|^2 + \alpha \mathrm{R}(x_\alpha) \leq \|\mathbf{A}x^\dagger - y\|^2 + \alpha\mathrm{R}(x^\dagger) \leq \delta^2 + \alpha\mathrm{R}(x^\dagger)\,.$$

Moreover

$$\begin{aligned} D^{\mathrm{R}}_{\mathbf{A}^*\eta}(x_\alpha, x^\dagger) &= \mathrm{R}(x_\alpha) - \mathrm{R}(x^\dagger) - \langle \mathbf{A}^*\eta, x_\alpha - x^\dagger\rangle \\ &\leq \mathrm{R}(x_\alpha) - \mathrm{R}(x^\dagger) + \|\eta\|\|\mathbf{A}x_\alpha - y^\dagger\| \\ &\leq \mathrm{R}(x_\alpha) - \mathrm{R}(x^\dagger) + \|\eta\|\delta + \|\eta\|\|\mathbf{A}x_\alpha - y\|\,. \end{aligned}$$

Therefore

$$\begin{aligned} \delta^2 &\geq \|\mathbf{A}x_\alpha - y\|^2 + \alpha\big(\mathrm{R}(x_\alpha) - \mathrm{R}(x^\dagger)\big) \\ &\geq \|\mathbf{A}x_\alpha - y\|^2 + \alpha\big(D^{\mathrm{R}}_{\mathbf{A}^*\eta}(x_\alpha, x^\dagger) - \|\eta\|\delta - \|\eta\|\|\mathbf{A}x_\alpha - y\|\big)\,. \end{aligned} \tag{3.4}$$

Together with the estimate

$$\alpha\|\eta\|\|\mathbf{A}x_\alpha - y\| \leq \|\mathbf{A}x_\alpha - y\|^2 + \frac{\alpha^2\|\eta\|^2}{4}\,,$$

equation (3.4) implies (3.3).

Since the Bregman distance is non-negative, it follows from (3.4) that

$$\begin{aligned} 0 &\geq \|\mathbf{A}x_\alpha - y\|^2 - \alpha\|\eta\|\|\mathbf{A}x_\alpha - y\| - \delta^2 - \delta\alpha\|\eta\| \\ &= \big(\|\mathbf{A}x_\alpha - y\| + \delta\big)\big(\|\mathbf{A}x_\alpha - y\| - \alpha\|\eta\| - \delta\big)\,, \end{aligned}$$

proving (3.2). □

*Remark* 3.6. If R is any convex regularization term, $\xi \in \partial\mathrm{R}(x^\dagger)$ and $D^{\mathrm{R}}_{\xi}(x; x^\dagger)$ denotes the Bregman distance of $x$ and $x^\dagger$ with respect to $\xi$ and R, then the estimates in Lemmas 3.4 and 3.5 remain to hold true. The special structure of the $\ell^1$-regularization term has not been used for their derivation.

*Remark* 3.7. In [6, Equation (3.3)], the even sharper estimate $D^{\mathrm{R}}_{\mathbf{A}^*\eta}(x_\alpha, x^\dagger) \leq \big(\delta^2 + \alpha^2\|\eta\|^2/4\big)/\alpha$ is claimed. Carrying out the proof in [6], however, also only results in (3.3).

Lemmas 3.4 and 3.5 provide bounds for the Bregman distance of regularized solutions and exact solutions of $\mathbf{A}x = y^\dagger$. Because the functional R is not strictly convex, the Bregman distance of two different elements can vanish. Thus, the estimates in terms of the Bregman distance yield no estimates in terms of the norm. In the following, we will clarify what kind of estimates they do entail. To that end we define, for $\eta \in Y$, the set

$$\Gamma[\eta] := \left\{\lambda \in \Lambda : |\langle \phi_\lambda, \mathbf{A}^*\eta\rangle| \geq 1\right\} \tag{3.5}$$

and the number

$$m[\eta] := \max\left\{|\langle \phi_\lambda, \mathbf{A}^*\eta\rangle| : \lambda \notin \Gamma[\eta]\right\} . \tag{3.6}$$

Because $\mathbf{A}^*\eta \in X$, and $(\phi_\lambda)_\lambda$ is an orthonormal basis, the sequence of coefficients $|\langle \phi_\lambda, \mathbf{A}^*\eta\rangle|$ converges to zero, and therefore the number $m[\eta]$ is well defined and satisfies $m[\eta] < 1$.

Now assume that $\eta \in Y$ is such that $\mathbf{A}^*\eta \in \partial \mathrm{R}(x^\dagger)$. Then we can estimate, for $x \in X$, the Bregman distance by (cf. Remark 3.3)

$$D^{\mathrm{R}}_{\mathbf{A}^*\eta}(x, x^\dagger) = \sum_{\lambda\in\Lambda}\Big(|\langle \phi_\lambda, x\rangle| - \langle \phi_\lambda, \mathbf{A}^*\eta\rangle\langle \phi_\lambda, x\rangle\Big) \geq (1 - m[\eta]) \sum_{\lambda\notin\Gamma[\eta]} |\langle \phi_\lambda, x\rangle| . \tag{3.7}$$

Therefore, if the Bregman distance of $x$ and $x^\dagger$ converges linearly to zero, so does the right hand side of (3.7). With the notation of Lemma 3.4 we consequently obtain that the projection $\pi_{\Lambda\setminus\Gamma[\eta]}x^\delta$ of $x^\delta$ converges linearly to $0 = \pi_{\Lambda\setminus\Gamma[\eta]}x^\dagger$. Similarly, with the notation of Lemma 3.5 the projection $\pi_{\Lambda\setminus\Gamma[\eta]}x_\alpha$ converges linearly. No information, however, can be obtained from (3.7) about the projections of $x^\delta$ and $x_\alpha$ to the space $X_{\Gamma[\eta]}$.

In contrast, in [23, 24, 34], we have derived linear convergence rates on the whole space $X$, assuming an additional injectivity condition originally introduced in [5, 28]. We follow the argumentation there, but using a slightly weaker condition.

*Condition* 3.8. The element $x^\dagger \in X$ solves the equation $\mathbf{A}x = y^\dagger$. In addition, the following conditions hold:

(1) *Source condition:* There exists some $\eta \in Y$ such that $\mathbf{A}^*\eta \in \partial \mathrm{R}(x^\dagger)$.
(2) *Restricted injectivity:* The mapping $\mathbf{A}_{\Gamma[\eta]} = \mathbf{A} \circ \mathrm{i}_{\Gamma[\eta]} : X_{\Gamma[\eta]} \to Y$ is injective.

*Remark* 3.9. (1) The first item of condition 3.8 implies that $|\langle \phi_\lambda, \mathbf{A}^*\eta\rangle| \leq 1$. Therefore we can rewrite $\Gamma[\eta]$ as

$$\Gamma[\eta] = \{\lambda \in \Lambda : |\langle \phi_\lambda, \mathbf{A}^*\eta\rangle| = 1\} .$$

(2) If Condition 3.8 is satisfied, then $X_{\Gamma[\eta]}$ is finite dimensional and $\mathbf{A}_{\Gamma[\eta]}$ is injective. Consequently, it has an inverse

$$\mathbf{A}^{-1}_{\Gamma[\eta]} : \operatorname{ran}(\mathbf{A}_{\Gamma[\eta]}) \to X_{\Gamma[\eta]} .$$

In particular, the operator norm of $\mathbf{A}_{\Gamma[\eta]}^{-1}$, denoted by $\|\mathbf{A}_{\Gamma[\eta]}^{-1}\|$, exists and is finite.

**Lemma 3.10.** *Let $x^\dagger$ and $\eta$ satisfy Condition 3.8. Then*

$$\|x - x^\dagger\| \le \|\mathbf{A}_{\Gamma[\eta]}^{-1}\| \|\mathbf{A}x - y^\dagger\| + \frac{1 + \|\mathbf{A}_{\Gamma[\eta]}^{-1}\| \|\mathbf{A}\|}{1 - m[\eta]} D^{\mathrm{R}}_{\mathbf{A}^*\eta}(x, x^\dagger) \tag{3.8}$$

*for every $x \in X$. Here $m[\eta] \in [0,1)$ is defined by (3.6).*

*Proof.* We have the estimate

$$\begin{aligned}
\|x - x^\dagger\| &\le \|\pi_{\Gamma[\eta]}x - x^\dagger\| + \|\pi_{\Lambda\setminus\Gamma[\eta]}x\| \\
&\le \|\mathbf{A}_{\Gamma[\eta]}^{-1}\| \|\mathbf{A}_{\Gamma[\eta]}(\pi_{\Gamma[\eta]}x - x^\dagger)\| + \|\pi_{\Lambda\setminus\Gamma[\eta]}x\| \\
&\le \|\mathbf{A}_{\Gamma[\eta]}^{-1}\| \|\mathbf{A}(x - x^\dagger)\| + \big(1 + \|\mathbf{A}_{\Gamma[\eta]}^{-1}\| \|\mathbf{A}\|\big) \|\pi_{\Lambda\setminus\Gamma[\eta]}x\| \,.
\end{aligned}$$

Moreover, we obtain from (3.7) that

$$\|\pi_{\Lambda\setminus\Gamma[\eta]}x\| \le \mathrm{R}(\pi_{\Lambda\setminus\Gamma[\eta]}x) = \sum_{\lambda \notin \Gamma[\eta]} |\langle \phi_\lambda, x\rangle| \le \frac{1}{1 - m[\eta]} D^{\mathrm{R}}_{\mathbf{A}^*\eta}(x, x^\dagger) \,.$$

This proves the assertion. □

*Remark* 3.11. Lemma 3.10 is a typical example of an *a–posteriori* error estimate. Such estimates are widely used in numerics of partial differential equations (see e.g. [37]), where the error of some numerical approximation to the exact solution of a partial differential equation is estimated. There, typically, a finite dimensional subspace $X_h$, for instance the finite element space, is considered. A–posteriori estimates look very similar as those in Lemma 3.10, after replacing the Bregman distance errors by boundary data errors and the source condition by smoothness assumptions. In the PDE context, $\mathbf{A}$ denotes evaluation of the PDE.

A–posteriori error estimates can be used for local grid refinement as an indicator for refinement strategy, for instance, when the evaluation of $\mathbf{A}$ for the finite element approximation, produces significant errors relative to the given right hand side. A similar motivation has been considered for quadratic regularization in [33] for nonlinear ill–posed problems. The estimates there have also the same structure as here.

**Proposition 3.12.** *Let Condition 3.8 be satisfied, and let $y \in Y$ satisfy $\|y^\dagger - y\| \le \delta$. Then we have for all $x^\delta \in \arg\min\{\mathrm{R}(x) : \|\mathbf{A}x - y\| \le \delta\}$ that*

$$\|x^\delta - x^\dagger\| \le c_\eta \delta \tag{3.9}$$

*with*

$$c_\eta = 2\left( \|\mathbf{A}_{\Gamma[\eta]}^{-1}\| + \frac{1 + \|\mathbf{A}_{\Gamma[\eta]}^{-1}\| \|\mathbf{A}\|}{1 - m[\eta]} \|\eta\| \right) .$$

*Proof.* This follows from Lemmas 3.4 and 3.10. □

**Proposition 3.13.** *Let Condition 3.8 be satisfied and consider the parameter choice $\alpha = C\delta$ for some $C > 0$. Then we have for all $y$ satisfying $\|y - y^\dagger\| \le \delta$ and all $x_\alpha \in \arg\min \mathrm{T}_{\alpha,y}$ that*

$$\|\mathbf{A}x_\alpha - y\| \le c_\eta^{(1)}\delta\,, \qquad \|x_\alpha - x^\dagger\| \le c_\eta^{(2)}\delta\,, \tag{3.10}$$

*where*

$$c_\eta^{(1)} := 1 + C\|\eta\|\,,$$

$$c_\eta^{(2)} := \|\mathbf{A}_{\Gamma[\eta]}^{-1}\| c_\eta^{(1)} + \frac{1 + \|\mathbf{A}_{\Gamma[\eta]}^{-1}\|\|\mathbf{A}\|}{1 - m[\eta]} \frac{\left(1 + C\|\eta\|/2\right)^2}{C}\,.$$

*Proof.* This follows from Lemmas 3.5 and 3.10. □

*Remark* 3.14. Notice that the linear convergence of the residuum $\|\mathbf{A}x_\alpha - y\|$ only requires $x^\dagger$ to satisfy the source condition (see Lemma 3.5). Conversely, we shall see in the next section that the source condition is in fact equivalent to the linear convergence of the residual.

## 4 Necessary Conditions

In the previous section we have verified that Condition 3.8 implies linear convergence of $\|x^\dagger - x_\alpha\|$ as $\alpha \to 0$. In the following we shall see that Condition 3.8 is in some sense the weakest possible condition that guarantees this linear rate of convergence.

First we prove the necessity of the source conditions for obtaining convergence rates for Tikhonov regularization with *a–priori* parameter choice:

**Lemma 4.1.** *Let $(\delta_k)_{k\in\mathbb{N}}$ be a sequence of positive numbers converging to zero as $k \to \infty$. Moreover, let $(y_k)_{k\in\mathbb{N}} \subset Y$ satisfy $\|y_k - y\| \le \delta_k$ and let $x_k \in \arg\min \mathrm{T}_{\alpha_k,y_k}$, with $\alpha_k \ge C\delta_k$ for some fixed $C > 0$.*

*If $\|x_k - x^\dagger\| \to 0$ as $k \to \infty$ and if there exists a constant $c > 0$ such that $\|\mathbf{A}x_k - y_k\| \le c\delta_k$, then $\mathrm{ran}(\mathbf{A}^*) \cap \partial\mathrm{R}(x^\dagger) \ne \emptyset$.*

*Proof.* Denote $\eta_k := (y_k - \mathbf{A}x_k)/\alpha_k$. By assumption the sequence $(\eta_k)_{k\in\mathbb{N}}$ is bounded in $Y$ and thus admits a subsequence $(\eta_{k_l})_{l\in\mathbb{N}}$ weakly converging to some $\eta \in Y$. Since $\mathbf{A}$ is a bounded operator, and therefore also the adjoint $\mathbf{A}^*$, this implies that $\mathbf{A}^*\eta_{k_l}$ weakly converges to $\mathbf{A}^*\eta$ in $X$, in signs $\mathbf{A}^*\eta_{k_l} \rightharpoonup \mathbf{A}^*\eta$.

The assumption $x_{k_l} \in \arg\min \mathrm{T}_{\alpha_{k_l},y_{k_l}}$ implies that

$$0 \in \partial\mathrm{T}_{\alpha_{k_l},y_{k_l}}(x_{k_l}) = \mathbf{A}^*(\mathbf{A}x_{k_l} - y_{k_l}) + \alpha_{k_l}\partial\mathrm{R}(x_{k_l})\,,$$

and therefore $\mathbf{A}^*\eta_{k_l} \in \partial\mathrm{R}(x_{k_l})$. The multifunction $\partial\mathrm{R}$ is maximal monotone and therefore its graph is sequentially strongly-weakly closed (see [13, Chapter V, p. 175]). Because $x_{k_l} \to x^\dagger$ and $\mathbf{A}^*\eta_{k_l} \rightharpoonup \mathbf{A}^*\eta$ as $l \to \infty$, this implies that $\mathbf{A}^*\eta \in \partial\mathrm{R}(x^\dagger)$ and proves the assertion. □

*Remark* 4.2. The proof of Lemma 4.1 is similar to the proofs given in [17] for entropy regularization and in [30] for $\ell^p$ regularization with $p > 1$. In fact, the above result is equally true for every proper, convex, and lower semi-continuous regularization functional. Under this assumption the same proof applies, because the subdifferential of every such functional is maximal monotone (see [3, Chapter II, Theorem 2.1]).

As a next step, we show the necessity of the restricted injectivity in Condition 3.8, Item 2. To this end, we show that Condition 3.8 is equivalent to the following:

*Condition* 4.3. The element $x^\dagger \in X$ solves the equation $\mathbf{A}x = y^\dagger$. In addition, the following conditions hold:

(1) *Strong source condition:* There exists some $\hat{\eta} \in Y$ with

$$\mathbf{A}^*\hat{\eta} \in \partial \mathrm{R}(x^\dagger) \quad \text{and} \quad |\langle \phi_\lambda, \mathbf{A}^*\hat{\eta}\rangle| < 1 \text{ for } \lambda \notin \operatorname{supp}(x^\dagger)\,. \tag{4.1}$$

(2) The restricted mapping $\mathbf{A}_{\operatorname{supp}(x^\dagger)}$ is injective.

*Remark* 4.4. In [5], a similar condition has been applied in order to derive a linear rate of convergence for an iterative soft-thresholding algorithm for minimizing the $\ell^1$-Tikhonov functional. There, a sequence $x^\dagger$ satisfying the first condition in 4.3 has been called to possess a *strict sparsity pattern*.

The proof of the equivalence of Conditions 3.8 and 4.3 mainly relies on the following Lemma.

**Lemma 4.5.** *Assume that $x^\dagger \in X$ is the* unique R*-minimizing solution of the equation $\mathbf{A}x = y^\dagger$. Moreover, assume that $x^\dagger$ is sparse.*

*Denote the support of $x^\dagger$ by $\Omega := \operatorname{supp}(x^\dagger)$. Then the following hold:*

- *The restricted mapping $\mathbf{A}_\Omega$ is injective.*
- *For every finite set $\Omega' \subset \Lambda$ with $\Omega \cap \Omega' = \emptyset$ there exists $\theta \in Y$ such that*

$$\langle \phi_\lambda, \mathbf{A}^*\theta\rangle = \operatorname{sign}(\langle \phi_\lambda, x^\dagger\rangle) \text{ for } \lambda \in \Omega \;\text{ and }\; |\langle \phi_\lambda, \mathbf{A}^*\theta\rangle| < 1 \text{ for } \lambda \in \Omega'\,. \tag{4.2}$$

*Proof.* The proof is divided in two parts. First we show the injectivity of $\mathbf{A}_\Omega$, then we prove (4.2).

- In order to verify that $\mathbf{A}_\Omega$ is injective, we have to show that $\ker(\mathbf{A}) \cap X_\Omega = \{0\}$. After possibly replacing some of the basis vectors $\phi_\lambda$ by $-\phi_\lambda$ we may assume without loss of generality that $\operatorname{sign}(\langle \phi_\lambda, x^\dagger\rangle) = 1$ for every $\lambda \in \Omega$.

  Since $x^\dagger$ is the unique R-minimizing solution of $\mathbf{A}x = y^\dagger$, it follows that

$$\mathrm{R}(x^\dagger + tw) > \mathrm{R}(x^\dagger) \qquad \text{for all } w \in \ker(\mathbf{A}) \setminus \{0\} \text{ and all } t \neq 0\,. \tag{4.3}$$

  Because $\Omega$ is a finite set, the mapping

$$t \mapsto \mathrm{R}(x^\dagger + tw) = \sum_{\lambda \in \Omega} |\langle \phi_\lambda, x^\dagger\rangle + t\langle \phi_\lambda, w\rangle| + |t| \sum_{\lambda \notin \Omega} |\langle \phi_\lambda, w\rangle|$$

is piecewise linear. Taking the one sided directional derivative with respect to $t$, it follows from (4.3) that for every $w \in \ker(\mathbf{A}) \setminus \{0\}$

$$\sum_{\lambda\in\Omega} \langle \phi_\lambda, w\rangle + \sum_{\lambda\notin\Omega} |\langle \phi_\lambda, w\rangle| > 0 \,.$$

For deriving the last inequality, we have used that all basis elements have been normalized such that $\operatorname{sign}(\langle \phi_\lambda, x^\dagger\rangle) = 1$. Applying the last inequality to $-w$ instead of $w$, we deduce that, in fact,

$$\sum_{\lambda\notin\Omega} |\langle \phi_\lambda, w\rangle| > \Big|\sum_{\lambda\in\Omega} \langle \phi_\lambda, w\rangle\Big| \qquad \text{for every } w \in \ker(\mathbf{A}) \setminus \{0\} \,. \tag{4.4}$$

In particular, it follows that $\sum_{\lambda\notin\Omega}|\langle \phi_\lambda, w\rangle| > 0$ for every $w \in \ker(\mathbf{A}) \setminus \{0\}$.

Thus, for every $w \in \ker(\mathbf{A}) \setminus \{0\}$ there exists some $\lambda \notin \Omega$ such that $\langle \phi_\lambda, w\rangle \neq 0$, and therefore $\ker(\mathbf{A}) \cap X_\Omega = \{0\}$. This, however, amounts to saying that the restricted mapping $\mathbf{A}_\Omega$ is injective.

- Now let $\Omega' \subset \Lambda$ be any finite subset of $\Lambda$ with $\Omega \cap \Omega' = \emptyset$. Inequality (4.4) and the finiteness of $\Omega \cup \Omega'$ imply the existence of a constant $\mu$ with $0 < \mu < 1$ such that

$$\mu \sum_{\lambda\in\Omega'} |\langle \phi_\lambda, w\rangle| \geq \Big|\sum_{\lambda\in\Omega} \langle \phi_\lambda, w\rangle\Big| \qquad \text{for every } w \in \ker(\mathbf{A}) \cap X_{\Omega\cup\Omega'} \,. \tag{4.5}$$

Assume now for the moment that $\xi$ is an element in $\operatorname{ran}(\mathbf{A}^*_{\Omega\cup\Omega'})$. Then there exists $\theta \in Y$ such that $\xi = \mathbf{A}^*_{\Omega\cup\Omega'}\theta$. The identity

$$\pi_{\Omega\cup\Omega'} \circ \mathbf{A}^* = (\mathbf{A} \circ \mathrm{i}_{\Omega\cup\Omega'})^* = \mathbf{A}^*_{\Omega\cup\Omega'}$$

therefore implies that

$$\langle \phi_\lambda, \xi\rangle = \langle \phi_\lambda, \mathbf{A}^*_{\Omega\cup\Omega'}\theta\rangle = \langle \phi_\lambda, \mathbf{A}^*\theta\rangle \qquad \text{for every } \lambda \in \Omega \cup \Omega' \,.$$

By assumption, $X_{\Omega\cup\Omega'}$ is finite dimensional and therefore the identity $\operatorname{ran}(\mathbf{A}^*_{\Omega\cup\Omega'}) = \big(\ker(\mathbf{A}_{\Omega\cup\Omega'})\big)^\perp$ holds true, where $^\perp$ denotes the orthogonal complement in $X_{\Omega\cup\Omega'}$. This shows that (4.2) is equivalent to the existence of some

$$\xi \in \operatorname{ran}(\mathbf{A}^*_{\Omega\cup\Omega'}) = \big(\ker(\mathbf{A}_{\Omega\cup\Omega'})\big)^\perp \subset X_{\Omega\cup\Omega'}$$

with

$$\langle \phi_\lambda, \xi\rangle = \operatorname{sign}(\langle \phi_\lambda, x^\dagger\rangle) = 1 \text{ for } \lambda \in \Omega \quad \text{ and } \quad |\langle \phi_\lambda, \xi\rangle| < 1 \text{ for } \lambda \in \Omega' \,. \tag{4.6}$$

In the following we make use of the element $\zeta \in X_{\Omega\cup\Omega'}$ defined by $\langle \phi_\lambda, \zeta\rangle = 1$ for $\lambda \in \Omega$ and $\langle \phi_\lambda, \zeta\rangle = 0$ for $\lambda \in \Omega'$.

If $\zeta \in (\ker(\mathbf{A}_{\Omega\cup\Omega'}))^\perp$, then we choose $\xi := \zeta$ and (4.6), and in consequence (4.2), follows.

If, on the other hand, $\zeta \notin \big(\ker(\mathbf{A}_{\Omega\cup\Omega'})\big)^{\perp}$, then there exists a basis $\big(w^{(1)},\dots,w^{(s)}\big)$ of $\ker(\mathbf{A}_{\Omega\cup\Omega'})$ such that

$$1 = \langle \zeta, w^{(i)} \rangle = \sum_{\lambda\in\Omega} \langle \phi_\lambda, \zeta\rangle \langle \phi_\lambda, w^{(i)}\rangle = \sum_{\lambda\in\Omega} \langle \phi_\lambda, w^{(i)}\rangle \quad \text{for every } 1 \le i \le s\,. \tag{4.7}$$

Consider now the constrained minimization problem on $X_{\Omega'}$,

$$\max_{\lambda\in\Omega'} |\langle \phi_\lambda, \hat{\zeta}\rangle| \to \min \qquad \text{subject to } \langle \hat{\zeta}, w^{(i)}\rangle = -1 \text{ for } i \in \{1,\dots,s\}\,. \tag{4.8}$$

Because of the equality $\langle \zeta, w^{(i)}\rangle = 1$, the admissible vectors $\hat{\zeta}$ in (4.8) are precisely those for which $\xi := \zeta + \hat{\zeta} \in \big(\ker(\mathbf{A}_{\Omega\cup\Omega'})\big)^{\perp}$. Thus, the task of finding $\xi$ satisfying (4.6) reduces to showing that the value of (4.8) is strictly smaller than 1.

Now note that the dual of the convex function $\hat{\zeta} \mapsto \max_{\lambda\in\Omega'} |\langle \phi_\lambda, \hat{\zeta}\rangle|$ is the function

$$X_{\Omega'} \ni \hat{\zeta} \mapsto \begin{cases} 0, & \text{if } \sum_{\lambda\in\Omega'} \big|\langle \phi_\lambda, \hat{\zeta}\rangle\big| \le 1\,, \\ +\infty, & \text{if } \sum_{\lambda\in\Omega'} \big|\langle \phi_\lambda, \hat{\zeta}\rangle\big| > 1\,. \end{cases}$$

Recalling that $\langle \hat{\zeta}, w^{(i)}\rangle = \sum_{\lambda\in\Omega'} \langle \phi_\lambda, \hat{\zeta}\rangle \langle \phi_\lambda, w^{(i)}\rangle$, it follows that the dual problem to (4.8) is the following constrained problem on $\mathbb{R}^s$ (see for instance [2, Sec. 5.3]):

$$\mathrm{S}(p) := -\sum_{i=1}^{s} p_i \to \min \qquad \text{subject to } \sum_{\lambda\in\Omega'} \Big| \sum_{i=1}^{s} p_i \langle \phi_\lambda, w^{(i)}\rangle \Big| \le 1\,. \tag{4.9}$$

From (4.7) we obtain that

$$\sum_{\lambda\in\Omega} \sum_{i=1}^{s} p_i \langle \phi_\lambda, w^{(i)}\rangle = \sum_{i=1}^{s} p_i = -\mathrm{S}(p)$$

for every $p \in \mathbb{R}^s$. Taking $w = \sum_{i=1}^{s} p_i w^{(i)}$, inequality (4.5) therefore implies that

$$\mu \sum_{\lambda\in\Omega'} \Big| \sum_{i=1}^{s} p_i \langle \phi_\lambda, w^{(i)}\rangle \Big| \ge \Big| \sum_{\lambda\in\Omega} \sum_{i=1}^{s} p_i \langle \phi_\lambda, w^{(i)}\rangle \Big| = \Big| \sum_{i=1}^{s} p_i \Big| = |\mathrm{S}(p)|$$

for every $p \in \mathbb{R}^s$. Then from (4.9) it follows that $|\mathrm{S}(p)| \le \mu$ for every admissible vector $p \in \mathbb{R}^s$ for the problem (4.9). Thus the value of $\mathrm{S}(p)$ in (4.9) is greater or equal than $-\mu$. Since the value of the primal problem (4.8) is the negative of the value of the dual problem (4.9), this shows that the value of (4.9) is at most $\mu$.

This proves that the value $\mu$ of (4.8) is strictly smaller than 1 and, as we have shown above, this proves the assertion (4.6).

□

**Proposition 4.6.** *Assume that $x^\dagger$ is the* unique R*-minimizing solution of the equation $\mathbf{A}x = y^\dagger$. Moreover, let there exist some $\eta \in Y$ satisfying $\mathbf{A}^*\eta \in \partial\mathrm{R}(x^\dagger)$. Then Condition 4.3 is satisfied.*

*Proof.* The injectivity of $\mathbf{A}_{\mathrm{supp}(x^\dagger)}$ follows from Lemma 4.5.

Define now

$$\Omega' := \Gamma(\eta) \setminus \mathrm{supp}(x^\dagger) = \left\{ \lambda \in \Lambda \setminus \mathrm{supp}(x^\dagger) : |\langle \phi_\lambda, \mathbf{A}^*\eta \rangle| = 1 \right\} .$$

Because $\mathbf{A}^*\eta \in \partial\mathrm{R}(x^\dagger) \subset \ell^2(\Lambda)$, the set $\Omega'$ is finite. Let now $\theta \in Y$ satisfy (4.2), and define

$$\hat{\eta} = \Big(1 - \frac{1 - m[\eta]}{2\|\theta\|_\infty}\Big)\eta + \frac{1 - m[\eta]}{2\|\theta\|_\infty}\,\theta .$$

Then one easily verifies that $\hat{\eta}$ satisfies the required condition (4.1). □

**Theorem 4.7.** *Let $x^\dagger \in X$ satisfy $\mathbf{A}x^\dagger = y^\dagger$.*

*Then the following statements are equivalent:*

(1) *$x^\dagger$ satisfies Condition 4.3.*
(2) *$x^\dagger$ satisfies Condition 3.8.*
(3) *For every $C > 0$ there exists $c^{(1)} > 0$ such that*

$$\|x_\alpha - x^\dagger\| \le c^{(1)}\delta$$

*whenever $x_\alpha \in \arg\min_x \mathrm{T}_{\alpha,y}(x)$ with $\|y^\dagger - y\| \le \delta$ and $\alpha = C\delta$.*

(4) *$x^\dagger$ is the unique* R*-minimizing solution of the equation $\mathbf{A}x = y^\dagger$. Moreover, for every $C > 0$ there exists $c^{(2)} > 0$ such that*

$$\|\mathbf{A}x_\alpha - y^\dagger\| \le c^{(2)}\delta$$

*whenever $x_\alpha \in \arg\min_x \mathrm{T}_{\alpha,y^\dagger}(x)$ with $\|y^\dagger - y\| \le \delta$ and $\alpha = C\delta$.*

*Proof.* Item 1 obviously implies Item 2. The implication 2 $\implies$ 3 has been shown in Proposition 3.13. Moreover, the implication 3 $\implies$ 4 is trivial, the operator $\mathbf{A}$ being linear and bounded.

Now assume that Item 4 holds. Consider sequences $\delta_k \to 0$ and $(y_k)_{k\in\mathbb{N}} \subset Y$ with $\|y_k - y\| \le \delta_k$. Choose any $x_k \in \arg\min_x \mathrm{T}_{\alpha_k,y_k}(x)$, where $\alpha_k = C\delta_k$. Then the uniqueness of $x^\dagger$ implies that $x_k \to x^\dagger$ (see [23, Prop. 7]). As a consequence, Lemma 4.1 applies, which shows the existence of $\eta \in Y$ with $\mathbf{A}^*\eta \in \partial\mathrm{R}(x^\dagger)$. Again using the uniqueness of $x^\dagger$, Proposition 4.6 now implies that Condition 4.3 holds, which concludes the proof. □

## 5 Application to Compressed Sensing

The *restricted isometry property* (also known as *uniform uncertainty principle*) is the key ingredient in compressed sensing for proving linear error estimates for the $\ell^1$-residual method in finite dimensional spaces (see [10]). Below we introduce and exploit this notation on arbitrary separable Hilbert spaces.

**Definition 5.1.** The *s-restricted isometry constant* $\vartheta_s$ of $\mathbf{A}$ is defined as the smallest number $\vartheta \geq 0$ that satisfies

$$(1-\vartheta)\|x\|_2^2 \leq \|\mathbf{A}x\|^2 \leq (1+\vartheta)\|x\|_2^2 \tag{5.1}$$

for all $s$-sparse $x \in X$. The $(s,s')$*-restricted orthogonality constant* $\vartheta_{s,s'}$ of $\mathbf{A}$ is defined as the smallest number $\vartheta \geq 0$ such that

$$|\langle \mathbf{A}x, \mathbf{A}x' \rangle| \leq \vartheta \|x\| \|x'\| \tag{5.2}$$

for all $s$-sparse $x \in X$ and $s'$-sparse $x' \in X$ with $\operatorname{supp} x \cap \operatorname{supp} x' = \emptyset$.

Because $\mathbf{A}$ is a bounded operator, the estimate (5.2) and the upper bound in (5.1) are satisfied for $\vartheta \geq \|\mathbf{A}\|^2$. Moreover, the lower bound in (5.1) is trivially satisfied for $\vartheta \geq 1$. Therefore, $\vartheta_s$ and $\vartheta_{s,s'}$ are well defined. Moreover, for every $\Omega \subset \Lambda$ with $|\Omega| \leq s$, the estimates $\|\mathbf{A}_\Omega^{-1}\| \leq 1/\sqrt{1-\vartheta_s}$ and $\|\mathbf{A}_\Omega\| \leq \sqrt{1+\vartheta_s}$ hold.

**Definition 5.2** (See [12])**.** The mapping $\mathbf{A}$ satisfies the *s-restricted isometry property*, if $\vartheta_s + \vartheta_{s,s} + \vartheta_{s,2s} < 1$.

The following Proposition is an important auxiliary result, which states that the $s$-restricted isometry property implies Condition 3.8 to hold for every $s$-sparse element.

**Proposition 5.3.** *Assume that* $\mathbf{A}$ *satisfies the s-restricted isometry property and let* $x^\dagger$ *be an s-sparse solution of the equation* $\mathbf{A}x = y^\dagger$*. Then* $x^\dagger$ *satisfies Condition 4.3. Moreover, the required source element* $\hat{\eta} \in \partial \mathrm{R}(x^\dagger)$ *can be chosen in such a way, that*

$$\|\hat{\eta}\| \leq W_s := \frac{\sqrt{s}}{\sqrt{1-\vartheta_s}} \frac{\vartheta_{s,s}}{1-\vartheta_s-\vartheta_{s,2s}}\,, \tag{5.3}$$

$$|\langle \phi_\lambda, \mathbf{A}^* \hat{\eta} \rangle| \leq M_s := \frac{\vartheta_{s,s}}{1-\vartheta_s-\vartheta_{s,2s}} < 1 \quad \text{for } \lambda \notin \operatorname{supp}(x^\dagger)\,. \tag{5.4}$$

*Notice that the bounds* $M_s$ *and* $W_s$ *are independent of* $x^\dagger$ *and* $\hat{\eta}$*, and that (5.4) implies the inequality* $m[\hat{\eta}] \leq M_s$ *(see (3.6) for the definition of* $m[\hat{\eta}]$*).*

*Proof.* In a finite dimensional setting, Proposition 5.3 follows from [12, Proof of Lemma 2.2]. The arguments given there, however, also apply to an infinite dimensional setting. Nevertheless, a proof of this important result for the general setting is given in Appendix A, as the notation of [12] significantly differs from ours. □

**Theorem 5.4.** *Assume that* $\mathbf{A}$ *satisfies the s-restricted isometry property and that* $x^\dagger$ *is an s-sparse solution of the equation* $\mathbf{A}x = y^\dagger$*. Then we have for all y satisfying* $\|y^\dagger - y\| \leq \delta$*, and all* $x^\delta \in \arg\min\{\mathrm{R}(x) : \|\mathbf{A}x - y\| \leq \delta\}$ *that*

$$\|x^\delta - x^\dagger\| \leq c_s \delta\,, \tag{5.5}$$

*where*

$$c_s = \frac{2}{\sqrt{1-\vartheta_s}} \left( 1 + \sqrt{s}\, \vartheta_{s,s} \frac{1+(1-\vartheta_s)^{-1/2}\|\mathbf{A}\|}{1-\vartheta_s-\vartheta_{s,s}-\vartheta_{s,2s}} \right)\,. \tag{5.6}$$

*Note that the constant $c_s$ is independent of the particular element $x^\dagger$.*

*Proof.* Lemma 5.3 implies that every $s$-sparse element $x^\dagger$ fulfills Condition 4.3 with a source element $\hat{\eta} \in Y$ that satisfies the estimates $\|\hat{\eta}\| \le W_s$ and $m[\hat{\eta}] \le M_s$. Together with the inequality $\|\mathbf{A}^{-1}_{\operatorname{supp}(x^\dagger)}\| \le (1-\vartheta_s)^{-1/2}$ and Proposition 3.12 it follows that (5.5) holds with

$$c_s = 2\left((1-\vartheta_s)^{-1/2} + \frac{1+(1-\vartheta_s)^{-1/2}\|\mathbf{A}\|}{1-M_s}\,W_s\right).$$

Insertion of the definitions of $M_s$ and $W_s$ (see (5.3) and (5.4)) shows the equality (5.6) and concludes the proof. □

*Remark* 5.5. Notice the qualitative difference between Proposition 3.12 and Theorem 5.4: In Proposition 3.12, the constant $c_\eta$ depends on $x^\dagger$, whereas the estimate of Theorem 5.4 holds uniformly for all $s$-sparse $x^\dagger$. On the other hand, Proposition 3.12 may provide convergence rates for certain sparse solutions, even if $\mathbf{A}$ does not satisfy any restricted isometry property.

Uniform linear estimates have first been obtained in [10] for the $\ell^1$ residual method. Applying Lemma 5.3 and Proposition 3.13, we now show that the same type of result also holds for $\ell^1$ Tikhonov regularization, where such uniform rates have not been given so far.

**Theorem 5.6.** *Assume that $\mathbf{A}$ satisfies the s-restricted isometry property and that $x^\dagger$ is s-sparse, and consider the parameter choice $\alpha = C\delta$ for some $C > 0$. Then, for all y satisfying $\|\mathbf{A}x^\dagger - y\| \le \delta$ and all $x_\alpha \in \arg\min \mathrm{T}_{\alpha,y}$, we have*

$$\|\mathbf{A}x_\alpha - y\| \le c_s^{(1)}\delta\,, \qquad \|x_\alpha - x^\dagger\| \le c_s^{(2)}\delta\,, \tag{5.7}$$

*where*

$$\begin{aligned} c_s^{(1)} &:= 1 + CW_s\,,\\ c_s^{(2)} &:= (1-\vartheta_s)^{-1}c_s^{(1)} + \frac{1+(1-\vartheta_s)^{-1}\|\mathbf{A}\|}{1-M_s}\,\frac{(1+CW_s/2)^2}{C}\,. \end{aligned} \tag{5.8}$$

*Here $M_s$ and $W_s$ are defined in (5.3) and (5.4).*

*Proof.* Lemma 5.3 states that every sparse element satisfies Condition 4.3 and the required source element $\hat{\eta} \in Y$ satisfies the estimates $\|\hat{\eta}\| \le W_s$ and $m[\hat{\eta}] \le M_s$. Together with the inequality $\|\mathbf{A}^{-1}_{\operatorname{supp}(x^\dagger)}\| \le (1-\vartheta_s)^{-1/2}$ and Proposition 3.13, this shows the assertions. □

## Finite Dimensional Results

We now assume that $\Lambda$ is finite, which is usually assumed in the compressed sensing literature. In this case, linear convergence already follows from the uniqueness of a R minimizing solution (1.6).

**Proposition 5.7.** *Assume that $\Lambda$ is finite and the solution $x^\dagger$ of the equation $\mathbf{A}x = y^\dagger$ is unique. Then Condition 4.3 is satisfied for $x^\dagger$. In particular the $\ell^1$-residual method and $\ell^1$-Tikhonov regularization with $\alpha = C\delta$ for some fixed $C > 0$ converge in $X$ linearly towards $x^\dagger$.*

*Proof.* The proof follows from Proposition 4.5 with $\Omega' = \Lambda \setminus \operatorname{supp}(x^\dagger)$. $\square$

In a finite dimensional setting, several modifications of the $s$-restricted isometry property (see Definition 5.2) have been introduced and used to derive unique recovery and linear convergence of the $\ell^1$-residual method (see [7, 8, 9, 10, 20, 21]). As the linear convergence rate of the $\ell^1$-residual method in particular implies unique recovery of $x^\dagger \in X$, each of these modifications implies Condition 4.3, and thus one also obtains linear convergence of $\ell^1$-Tikhonov regularization with a–priori parameter choice $\alpha = C\delta$.

Even more, it follows that the linear convergence of the $\ell^1$ residual method is in fact equivalent to any of the conditions given in Theorem 4.7.

## Appendix: Proof of Proposition 5.3

Following [12] we will construct the required source element $\hat{\eta} = \sum_{k=1}^{\infty} (-1)^k \eta_k$ by recursively defining $\eta_k \in Y$ by means of the following lemma. To that end, we recall the restricted isometry constants $\vartheta_n$ and $\vartheta_{n,n'}$ introduced in Definition 5.1.

**Lemma A.1.** *Let $n \in \mathbb{N}$ be such that the constant $\vartheta_n$ from the restricted isometry property satisfies $\vartheta_n < 1$. Moreover, let $\Omega \subset \Lambda$ satisfy $|\Omega| \leq n$.*

*Then, for every $\xi \in X_\Omega$ and every $n' \in \mathbb{N}$ there exist an element $\eta \in Y$ and a set $\Omega' \subset \Lambda$ with $|\Omega'| \leq n'$ and $\Omega \cap \Omega' = \emptyset$, such that*

$$\pi_\Omega \mathbf{A}^* \eta = \xi \,, \tag{A.1}$$

$$\|\eta\| \leq \frac{1}{\sqrt{1 - \vartheta_n}} \|\xi\| \,, \tag{A.2}$$

$$\|\pi_{\Omega'} \mathbf{A}^* \eta\| \leq \frac{\vartheta_{n,n'}}{1 - \vartheta_n} \|\xi\| \,, \tag{A.3}$$

$$|\langle \phi_\lambda, \mathbf{A}^* \eta \rangle| \leq \frac{\vartheta_{n,n'}}{\sqrt{n'}(1 - \vartheta_n)} \|\xi\| \qquad \textit{for } \lambda \notin \Omega \cup \Omega' \,. \tag{A.4}$$

*Proof.* Define $\eta := \mathbf{A}_\Omega (\mathbf{A}_\Omega^* \mathbf{A}_\Omega)^{-1} \xi$. The identity $\mathbf{A}_\Omega^* = (\mathbf{A} \circ \mathrm{i}_\Omega)^* = \pi_\Omega \circ \mathbf{A}^*$ implies that

$$\pi_\Omega(\mathbf{A}^* \eta) = \mathbf{A}_\Omega^* \mathbf{A}_\Omega (\mathbf{A}_\Omega^* \mathbf{A}_\Omega)^{-1} \xi = \xi \,,$$

which shows (A.1). Because $(1 - \vartheta_n)\|\zeta\|^2 \leq \|\mathbf{A}\zeta\|^2$ for all $\zeta \in X_\Omega$, it follows that

$$\|(\mathbf{A}_\Omega^* \mathbf{A}_\Omega)^{-1}\| \leq \frac{1}{1 - \vartheta_n} \,.$$

Thus

$$\begin{aligned}\|\eta\|^2 &= \left\langle \mathbf{A}_\Omega(\mathbf{A}_\Omega^*\mathbf{A}_\Omega)^{-1}\xi, \mathbf{A}_\Omega(\mathbf{A}_\Omega^*\mathbf{A}_\Omega)^{-1}\xi\right\rangle \\ &= \left\langle (\mathbf{A}_\Omega^*\mathbf{A}_\Omega)^{-1}\xi, (\mathbf{A}_\Omega^*\mathbf{A}_\Omega)(\mathbf{A}_\Omega^*\mathbf{A}_\Omega)^{-1}\xi\right\rangle \\ &\le \|(\mathbf{A}_\Omega^*\mathbf{A}_\Omega)^{-1}\|\|\xi\|^2 \\ &\le \frac{1}{1-\vartheta_n}\|\xi\|^2,\end{aligned}$$

which is (A.2).

Let now $J \subset \Lambda$ be any subset with $|J| \le n'$ and $J \cap \Omega = \emptyset$. Then we have for every $z \in X_J$ that

$$\begin{aligned}|\langle \mathbf{A}_J^*\eta, z\rangle| &= |\langle \mathbf{A}_J^*\mathbf{A}_\Omega(\mathbf{A}_\Omega^*\mathbf{A}_\Omega)^{-1}\xi, z\rangle| \\ &= |\langle \mathbf{A}_\Omega(\mathbf{A}_\Omega^*\mathbf{A}_\Omega)^{-1}\xi, \mathbf{A}_J z\rangle| \\ &\le \vartheta_{n,n'}|\langle (\mathbf{A}_\Omega^*\mathbf{A}_\Omega)^{-1}\xi, z\rangle| \le \frac{\vartheta_{n,n'}}{1-\vartheta_n}\|\xi\|\|z\| .\end{aligned}$$

This implies that

$$\|\mathbf{A}_J^*\eta\| = \max_{z\in X_J}\frac{|\langle \mathbf{A}_J^*\eta, z\rangle|}{\|z\|} \le \frac{\vartheta_{n,n'}}{1-\vartheta_n}\|\xi\| . \tag{A.5}$$

Now define

$$\Omega' := \left\{ \lambda \in \Lambda\setminus\Omega : |\langle \phi_\lambda, \mathbf{A}^*\eta\rangle| > \frac{\vartheta_{n,n'}}{\sqrt{n'}(1-\vartheta_n)}\|\xi\| \right\} .$$

Then (A.4) is satisfied.

Now assume that there exists some $J \subset \Omega'$ with $|J| > n'$. Then the definition of $\Omega'$ would imply that $\|\pi_J \mathbf{A}^*\eta\| = \|\mathbf{A}_J^*\eta\| > \frac{\vartheta_{n,n'}}{1-\vartheta_n}\|\xi\|$, which contradicts (A.5). This shows that $|\Omega'| \le n'$ and, using (A.5), that (A.3) holds. □

*Proof of Proposition 5.3.* Set $\Omega_0 := \operatorname{supp}(x^\dagger)$ and inductively define $\eta_k \in Y$ and $\Omega_k \subset \Lambda$ as follows:

In the case $k = 1$, one applies Lemma A.1 with $n = n' = s$ and

$$\xi = \sum_{\lambda\in\Omega_0} \operatorname{sign}(\langle \phi_\lambda, x\rangle)\,\phi_\lambda .$$

This shows that there exist $\eta_1 \in Y$ and $\Omega_1 \subset \Lambda$ with $|\Omega_1| \le s$ and $\Omega_0 \cap \Omega_1 = \emptyset$, such that

$$
\begin{aligned}
\pi_{\Omega_0}(\mathbf{A}^*\eta_1) &= \sum_{\lambda\in\Omega_0} \operatorname{sign}(\langle \phi_\lambda, x\rangle)\ \phi_\lambda\,, \\
\|\eta_1\| &\le \frac{\sqrt{s}}{\sqrt{1-\vartheta_s}}\,, \\
\|\pi_{\Omega_1}(\mathbf{A}^*\eta_1)\| &\le \frac{\sqrt{s}\ \vartheta_{s,s}}{1-\vartheta_s}\,, \qquad \text{(A.6)} \\
|\langle \phi_\lambda, \mathbf{A}^*\eta_1\rangle| &\le \frac{\vartheta_{s,s}}{(1-\vartheta_s)} \qquad \text{for } \lambda \notin \Omega_0 \cup \Omega_1\,.
\end{aligned}
$$

In the case $k>1$ one applies Lemma A.1 with $n=2s$, $n'=s$, $\Omega = \Omega_0 \cup \Omega_{k-1}$ and $\xi = \pi_{\Omega_{k-1}}(\mathbf{A}^*\eta_{k-1})$. One obtains an element $\eta_k \in Y$ and a set $\Omega_k \subset \Lambda$ with $|\Omega_k| \le s$ and $\Omega_k \cap (\Omega_{k-1} \cup \Omega_0) = \emptyset$, such that

(A.7)

$$\pi_{\Omega_0\cup\Omega_{k-1}}(\mathbf{A}^*\eta_k) = \pi_{\Omega_{k-1}}(\mathbf{A}^*\eta_{k-1})\,,$$

$$\|\eta_k\| \le \frac{1}{\sqrt{1-\vartheta_s}}\ \|\pi_{\Omega_k}(\mathbf{A}^*\eta_{k-1})\|\,, \tag{A.8}$$

$$\|\pi_{\Omega_k}(\mathbf{A}^*\eta_k)\| \le \frac{\vartheta_{s,2s}}{1-\vartheta_s}\|\pi_{\Omega_{k-1}}(\mathbf{A}^*\eta_{k-1})\|\,, \tag{A.9}$$

$$|\langle \phi_\lambda, \mathbf{A}^*\eta_k\rangle| \le \frac{\vartheta_{s,2s}}{\sqrt{s}\,(1-\vartheta_s)}\|\pi_{\Omega_{k-1}}(\mathbf{A}^*\eta_{k-1})\| \quad \text{for } \lambda \notin \Omega_0 \cup \Omega_{k-1} \cup \Omega_k\,. \tag{A.10}$$

Using (A.6) and (A.9), it follows from induction that

$$\|\pi_{\Omega_k}(\mathbf{A}^*\eta_k)\| \le \frac{\sqrt{s}\ \vartheta_{s,s}}{1-\vartheta_s}\left(\frac{\vartheta_{s,2s}}{1-\vartheta_s}\right)^{k-1}\,. \tag{A.11}$$

Let

$$\eta := \sum_{k=1}^{\infty} (-1)^k \eta_k\,. \tag{A.12}$$

Then it follows from (A.8) and (A.11), that

$$
\begin{aligned}
\|\eta\| \le \sum_{k=1}^{\infty}\|\eta_k\| &\le \frac{1}{\sqrt{1-\vartheta_s}}\sum_{k=1}^{\infty}\|\pi_{\Omega_k}(\mathbf{A}^*\eta_k)\| \\
&\le \frac{1}{\sqrt{1-\vartheta_s}}\ \frac{\sqrt{s}\ \vartheta_{s,s}}{1-\vartheta_s}\sum_{k=0}^{\infty}\left(\frac{\vartheta_{s,2s}}{1-\vartheta_s}\right)^k = \frac{1}{1-\vartheta_s-\vartheta_{s,2s}}\ \frac{\sqrt{s}\ \vartheta_{s,s}}{\sqrt{1-\vartheta_s}}\,.
\end{aligned}
$$

This shows that the sum in (A.12) converges absolutely and that $\eta$ satisfies estimate (5.3). Moreover, the equalities $\pi_{\Omega_0}(\mathbf{A}^*\eta_1) = \sum_{\lambda\in\Omega_0}\operatorname{sign}(\langle\phi_\lambda, x\rangle)\phi_\lambda$ and

$\pi_{\Omega_0}(\mathbf{A}^*\eta_k) = \pi_{\Omega_0}\pi_{\Omega_k}(\mathbf{A}^*\eta_{k-1}) = 0$ for $k > 1$ (see (A.7)) imply that

$$\pi_{\Omega_0}(\mathbf{A}^*\eta) = \sum_{\lambda \in \Omega_0} \operatorname{sign}(\langle \phi_\lambda, x \rangle)\, \phi_\lambda\,.$$

It remains to show that $\langle \phi_\lambda, \mathbf{A}^*\eta \rangle \leq M_s$ for $\lambda \notin \operatorname{supp}(x^\dagger)$. To that end, note that $\langle \phi_\lambda, \mathbf{A}^*\eta_k \rangle = \langle \phi_\lambda, \mathbf{A}^*\eta_{k+1} \rangle$ if $\lambda \in \Omega_k$. Therefore,

$$\langle \phi_\lambda, \mathbf{A}^*\eta \rangle = \sum_{k=1}^{\infty} (-1)^k \langle \phi_\lambda, \mathbf{A}^*\eta_k \rangle = \sum_{\{k : \lambda \notin \Omega_k \cup \Omega_{k-1}\}} (-1)^k \langle \phi_\lambda, \mathbf{A}^*\eta_k \rangle\,.$$

Consequently, (A.10) and (A.11) imply that

$$\begin{aligned} |\langle \phi_\lambda, \mathbf{A}^*\eta \rangle| &\leq \sum_{\{k : \lambda \notin \Omega_k \cup \Omega_{k-1}\}} \big|\langle \phi_\lambda, \mathbf{A}^*\eta_\lambda \rangle\big| \\ &\leq \frac{\vartheta_{s,s}}{1-\vartheta_s} \sum_{k=0}^{\infty} \left(\frac{\vartheta_{s,2s}}{1-\vartheta_s}\right)^k = \frac{\vartheta_s}{1-\vartheta_s-\vartheta_{s,2s}} = M_s\,, \end{aligned}$$

which concludes the proof. □

**Acknowledgment.** This work has been supported by the Austrian Science Fund (FWF) within the national research networks Industrial Geometry, project 9203-N12, and Photoacoustic Imaging in Biology and Medicine, project S10505-N20.

## Bibliography

[1] S. W. Anzengruber and R. Ramlau. Morozov's discrepancy principle for Tikhonov-type functionals with nonlinear operators. *Inverse Probl.*, 26(2):025001, 17, 2010.
[2] J.-P. Aubin. *Optima and equilibria*, volume 140 of *Graduate Texts in Mathematics*. Springer-Verlag, Berlin, second edition, 1998. An introduction to nonlinear analysis, Translated from the French by Stephen Wilson.
[3] V. Barbu. *Nonlinear Semigroups and Differential Equations in Banach Spaces*. Editura Academiei Republicii Socialiste România, Bucharest, 1976.
[4] T. Bonesky. Morozov's discrepancy principle and Tikhonov-type functionals. *Inverse Probl.*, 25(1):015015, 11, 2009.
[5] K. Bredies and D. Lorenz. Linear convergence of iterative soft-thresholding. *J. Fourier Anal. Appl.*, 14(5-6):813–837, 2008.
[6] M. Burger and S. Osher. Convergence rates of convex variational regularization. *Inverse Probl.*, 20(5):1411–1421, 2004.
[7] T. Cai, L. Wang, and G. Xu. Shifting inequality and recovery of sparse signals. *IEEE Trans. Signal Process.*, 58(3):1300–1308, 2010.
[8] T. Cai, G. Xu, and J. Zhang. On recovery of sparse signals via $\ell^1$ minimization. *IEEE Trans. Inf. Theory*, 55(7):3388–3397, 2009.
[9] E. J. Candès. The restricted isometry property and its implications for compressed sensing. *C. R. Math. Acad. Sci. Paris*, 346(9-10):589–592, 2008.
[10] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory*, 52(2):489–509, 2006.
[11] E. J. Candès, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, 2006.

[12] E. J. Candès and T. Tao. Decoding by linear programming. *IEEE Trans. Inf. Theory*, 51(12):4203–4215, 2005.

[13] I. Cioranescu. *Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems*, volume 62 of *Mathematics and its Applications*. Kluwer, Dordrecht, 1990.

[14] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math.*, 57(11):1413–1457, 2004.

[15] D. L. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via $\ell^1$ minimization. *Proc. Natl. Acad. Sci. USA*, 100(5):2197–2202 (electronic), 2003.

[16] D. L. Donoho, M. Elad, and V. N. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Inf. Theory*, 52(1):6–18, 2006.

[17] P. P. B. Eggermont. Maximum entropy regularization for Fredholm integral equations of the first kind. *SIAM J. Math. Anal.*, 24(6):1557–1576, 1993.

[18] I. Ekeland and R. Temam. *Convex Analysis and Variational Problems*. North-Holland, Amsterdam, 1976.

[19] 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.

[20] S. Foucart. A note on guaranteed sparse recovery via $\ell^1$-minimization. *Appl. Comput. Harmon. Anal.*, 29(1):97–103, 2010.

[21] S. Foucart and M. Lai. Sparsest solutions of underdetermined linear systems via $\ell^q$-minimization for $0 < q \leq 1$. *Appl. Comput. Harmon. Anal.*, 26(3):395–407, 2009.

[22] J. J. Fuchs. Recovery of exact sparse representations in the presence of bounded noise. *IEEE Trans. Inf. Theory*, 51(10):3601–3608, 2005.

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

[24] M. Grasmair, M. Haltmeier, and O. Scherzer. The residual method for regularizing ill-posed problems. Reports of FSP S105 - "Photoacoustic Imaging" 14, University of Innsbruck, Austria, 2009.

[25] T. Hein and B. Hofmann. Approximate source conditions for nonlinear ill-posed problems – chances and limitations. *Inverse Probl.*, 25:035003, 2009.

[26] 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.

[27] V. K. Ivanov, V. V. Vasin, and V. P. Tanana. *Theory of linear ill-posed problems and its applications*. Inverse and Ill-posed Problems Series. VSP, Utrecht, second edition, 2002. Translated and revised from the 1978 Russian original.

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

[29] A. Neubauer. On enhanced convergence rates for tikhonov regularization of nonlinear ill-posed problems in banach spaces. *Inverse Probl.*, 25(6):065009 (10pp), 2009.

[30] E. Ramlau, R. Resmerita. Convergence rates for regularization with sparsity constraints. *Electron. Trans. Numer. Anal.*, 2010. To appear.

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

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

[33] O. Scherzer. A posteriori error estimates for the solution of nonlinear ill-posed operator equations. *Nonlinear Anal.*, 45(4, Ser. A: Theory Methods):459–481, 2001.

[34] 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.

[35] V. P. Tanana. *Methods for solution of nonlinear operator equations*. Inverse and Ill-posed Problems Series. VSP, Utrecht, 1997.

[36] J. A. Tropp. Just relax: convex programming methods for identifying sparse signals in noise. *IEEE Trans. Inf. Theory*, 52(3):1030–1051, 2006.

[37] R. Verfürth. *A Review of A Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques*. John Wiley & Son, Chichester, 1996.

Received Month 200X.