Research Network FWF S105

Photoacoustic Imaging in
Medicine and Biology

http://pai.uibk.ac.at

Sparse Regularization with l^q Penalty Term

M. Grasmair, M. Haltmeier,
and O. Scherzer

first revision, August 2008
(original version: June 2008)

PAI Report No. 1

Der Wissenschaftsfonds.

universität
innsbruck

Sparse Regularization with l^q Penalty Term

Markus Grasmair[1] Markus Haltmeier[1]
Otmar Scherzer[1,2]

[1]Department of Mathematics
University of Innsbruck
Technikerstr. 21a
6020 Innsbruck, Austria

[2]Radon Institute of Computational
and Applied Mathematics
Altenberger Str. 69
4040 Linz, Austria

August 25, 2008

Abstract

We consider the stable approximation of sparse solutions to non-linear operator equations by means of Tikhonov regularization with a subquadratic penalty term. Imposing certain assumptions, which for a linear operator are equivalent to the standard range condition, we derive the usual convergence rate $O(\sqrt{\delta})$ of the regularized solutions in dependence of the noise level δ. Particular emphasis lies on the case, where the true solution is known to have a sparse representation in a given basis. In this case, if the differential of the operator satisfies a certain injectivity condition, we can show that the actual convergence rate improves up to $O(\delta)$.

MSC: 65J20; 65J22, 49N45.

1 Introduction

A widely used technique for the approximate solution of an ill-posed, possibly non-linear operator equation

$$F(u) = v \tag{1}$$

on a Hilbert space U is Tikhonov regularization, which can be formulated as minimization of the functional

$$\mathcal{T}(u) = \|F(u) - v\|^2 + \alpha \mathcal{R}(u) \ .$$

The first term ensures that the minimizer u_α will indeed approximately solve the equation, while the second term stabilizes the process of inverting F and forces u_α to satisfy certain regularity properties incorporated into $\mathcal{R}$. Originally, Tikhonov applied this method to the stable solution of the Fredholm equation. Requiring differentiability of u_α, he used the square of a higher order weighted Sobolev norm as penalty term [26, 27].

Recently, the focus has shifted from the postulation of differentiability properties to sparsity constraints [10, 3, 7, 8, 9, 11, 14, 17, 18, 23, 28]. Here, one requires the expansion of u_α with respect to some given orthonormal basis $(\phi_i)_{i\in\mathbb{N}}$

of U to be sparse in the sense that only finitely many coefficients are different from zero. This can be achieved with regularization functionals

$$\mathcal{R}(u) = \sum_{i \in \mathbb{N}} |\langle \phi_i, u \rangle|^q \,, \qquad 0 \le q \le 2 \,. \tag{2}$$

In fact, sparsity of the solution is not necessarily guaranteed for $q > 1$. The lack of convexity of $\mathcal{R}$, however, makes a choice $q < 1$ inconvenient both for theoretical analysis and the actual computation of a minimizer. On the other hand, the assumption $q \le 2$ is used to obtain coercivity of the regularization functional, which in turn implies the existence of minimizers of $\mathcal{T}$. For these reasons we only consider the case $1 \le q \le 2$.

We concentrate our analysis on the well-posedness of the regularization method and the derivation of convergence rates. For that purpose we assume that only noisy data v^δ is given, which satisfies $\|v^\delta - v\| \le \delta$. We denote by u_α^δ the minimizer of the regularization functional with noisy data v^δ and regularization parameter α, and by $u^\dagger$ an $\mathcal{R}_q$-minimizing solution of $F(u) = v$. Then the question is how the distance $\|u_\alpha^\delta - u^\dagger\|$ depends on the noise level δ and the regularization parameter α.

Dismissing for the moment the assumption of sparsity, we derive for a parameter choice $\alpha \sim \delta$ a convergence rate $\|u_\alpha^\delta - u^\dagger\| = O(\sqrt{\delta})$ provided $1 < q \le 2$ and a source condition is satisfied (see Proposition 12). In the linear case this condition is the usual range condition $\partial\mathcal{R}(u^\dagger) \cap \operatorname{range}(F^*) \neq \emptyset$, where F^* denotes the adjoint of the operator F (see Proposition 11). Similar results have been derived recently [20, 22]. In the non-linear case we impose a different assumption, which for sparsity regularization generalizes common source conditions involving the Bregman distance [19, 24, 25].

If, furthermore, the solution $u^\dagger$ of the operator equation is known to be sparse, then the convergence rates of the regularized solutions to $u^\dagger$ can be shown to be $O(\delta^{1/q})$ where $1 \le q \le 2$ is the exponent in the regularization term (2) (see Theorems 14 and 15). To that end we require the derivative of F at $u^\dagger$ to be invertible on certain finite dimensional subspaces, a condition introduced in [4] for linear operators as 'finite basis injectivity property'. This improved convergence rate provides a theoretical justification for the usage of subquadratic penalty terms for regularization with sparsity constraints.

Our results reveal a fundamental difference between quadratic and non-quadratic Tikhonov regularization. Neubauer [21] has derived a saturation result for quadratic regularization in a Hilbert space setting with a linear operator F. He has shown that, apart from the trivial case $u^\dagger = 0$, the convergence rates cannot be better than $O(\delta^{2/3})$. The present article shows that this rate can be beaten by sparse regularization when applied to the recovery of sparse data.

2 Notational Preliminaries

All along this paper we assume that V is a reflexive Banach space and U is a Hilbert space in which a frame $(\phi_i)_{i \in \mathbb{N}} \subset U$ is given. That is, there exist $0 < C_1 \le C_2 < \infty$ such that

$$C_1 \|u\|^2 \le \sum_{i \in \mathbb{N}} |\langle \phi_i, u \rangle|^2 \le C_2 \|u\|^2 \qquad \text{for every } u \in U \,.$$

The operator $F\colon \operatorname{dom}(F) \subseteq U \to V$ is assumed to be weakly sequentially closed and $\operatorname{dom}(F) \cap \operatorname{dom}(\mathcal{R}_q) \neq \emptyset$. Examples for weakly sequentially closed operators are linear bounded operators restricted to convex domains, which naturally arise for instance in image restoration problems or tomographic applications [15, 25]. Truly nonlinear operators arise in schlieren imaging [25] or simultaneous activity and attenuation reconstruction in emission tomography [12]. See also [1, 23] for the application of sparsity constraints to inverse problems.

We define the regularization functional $\mathcal{R}_q \colon U \to \mathbb{R} \cup \{\infty\}$ by

$$\mathcal{R}_q(u) := \sum_{i \in \mathbb{N}} w_i |\langle \phi_i, u \rangle|^q \,,$$

where $1 \leq q \leq 2$ and there exists $w_{\min} > 0$ such that $w_i \geq w_{\min}$ for all $i \in \mathbb{N}$. Note that $\mathcal{R}_q$ is convex and weakly lower semi-continuous as the sum of non-negative convex and weakly continuous functionals.

The subdifferential of $\mathcal{R}_q$ at u is denoted by $\partial \mathcal{R}_q(u) \subset U$. If $q > 1$, then $\partial \mathcal{R}_q(u)$ is at most single valued and is identified with its single element.

For the approximate solution of the operator equation $F(u) = v$ we consider the minimization of the regularization functional

$$\mathcal{T}_{\alpha,v}^{p,q}(u) := \begin{cases} \left\| F(u) - v \right\|^p + \alpha \mathcal{R}_q(u) \,, & \text{if } u \in \operatorname{dom}(F) \cap \operatorname{dom}(\mathcal{R}_q) \,, \\ +\infty \,, & \text{if } u \notin \operatorname{dom}(F) \cap \operatorname{dom}(\mathcal{R}_q) \,, \end{cases}$$

with some $\alpha > 0$ and $p \geq 1$.

In order to prove convergence rates results we impose an additional assumption concerning the interaction of F and $\mathcal{R}_q$ in a neighborhood of an $\mathcal{R}_q$-minimizing solution of $F(u) = v$. Here $u^\dagger \in U$ is called $\mathcal{R}_q$-minimizing solution, if $F(u^\dagger) = v$ and

$$\mathcal{R}_q(u^\dagger) = \min\bigl\{ \mathcal{R}_q(u) : F(u) = v \bigr\} \,.$$

Assumption 1. *The equation $F(u) = v$ has an $\mathcal{R}_q$-minimizing solution $u^\dagger$ and there exist β_1, $\beta_2 > 0$, $r > 0$, $\sigma > 0$, and $\rho > \mathcal{R}_q(u^\dagger)$ such that*

$$\mathcal{R}_q(u) - \mathcal{R}_q(u^\dagger) \geq \beta_1 \|u - u^\dagger\|^r - \beta_2 \|F(u) - F(u^\dagger)\| \tag{3}$$

for all $u \in \operatorname{dom}(F)$ satisfying $\mathcal{R}_q(u) < \rho$ and $\|F(u) - F(u^\dagger)\| < \sigma$.

In Section 4 below we show that Assumption 1 with $r = 2$ follows from the standard conditions stated in general convergence rates results in a Banach space setting [5, 19, 24], which in turn generalize the standard conditions in a Hilbert space setting [15, 16]. Moreover, the assumption is equivalent to the standard source condition $\partial \mathcal{R}(u^\dagger) \cap \operatorname{range}(F^*) \neq \emptyset$ in the particular case of a linear and bounded operator F (see Proposition 11).

3 Well-Posedness and Convergence Rates

In this section we prove the well-posedness of the regularization method. By this we mean that minimizers u_α^δ of the regularization functional $\mathcal{T}_{\alpha,v^\delta}^{p,q}$ exist for every $\alpha > 0$, continuously depend on the data v^δ, and converge to a solution of $F(u) = v$ as the noise level approaches zero, provided the regularization parameter α is chosen appropriately.

These results are analogous to results obtained for standard quadratic Tikhonov regularization in Hilbert spaces (see e.g. [15]). Also the mathematical techniques employed in the proofs of existence, weak stability, and convergence are similar. Some extra work is needed, however, for the passage from weak stability and convergence to stability and convergence with respect to $\mathcal{R}_q$.

Lemma 2. *Let $1 \le q \le 2$. Assume that $(u_k)_{k\in\mathbb{N}} \subset U$ weakly converges to $u \in U$ and that $\mathcal{R}_q(u_k)$ converges to $\mathcal{R}_q(u)$. Then $\mathcal{R}_q(u_k - u) \to 0$.*

Proof. The assumption $\mathcal{R}_q(u_k) \to \mathcal{R}(u)$ implies that

$$\begin{aligned}
&\limsup_k \mathcal{R}_q(u_k - u)\\
&\qquad = \limsup_k \Big[2\big(\mathcal{R}_q(u_k) + \mathcal{R}_q(u)\big) - 2\big(\mathcal{R}_q(u_k) + \mathcal{R}_q(u)\big) + \mathcal{R}_q(u_k - u)\Big]\\
&\qquad = 4\,\mathcal{R}_q(u) - \liminf_k \sum_{i\in\mathbb{N}} w_i \Big[2\,|\langle\phi_i, u_k\rangle|^q + 2\,|\langle\phi_i, u\rangle|^q - |\langle\phi_i, u_k - u\rangle|^q\Big]\,.
\end{aligned}$$

Using Fatou's Lemma we obtain that

$$\begin{aligned}
&-\liminf_k \sum_{i\in\mathbb{N}} w_i \Big[2\,|\langle\phi_i, u_k\rangle|^q + 2\,|\langle\phi_i, u\rangle|^q - |\langle\phi_i, u_k - u\rangle|^q\Big]\\
&\qquad\qquad \le -\sum_{i\in\mathbb{N}} \liminf_k w_i \Big[2\,|\langle\phi_i, u_k\rangle|^q + 2\,|\langle\phi_i, u\rangle|^q - |\langle\phi_i, u_k - u\rangle|^q\Big]\,.
\end{aligned}$$

Now, the weak convergence of $(u_k)_{k\in\mathbb{N}}$ shows that $\langle\phi_i, u_k\rangle \to \langle\phi_i, u\rangle$ for all $i \in \mathbb{N}$. Therefore it follows that

$$-\sum_{i\in\mathbb{N}} \liminf_k w_i \Big[2\,|\langle\phi_i, u_k\rangle|^q + 2\,|\langle\phi_i, u\rangle|^q - |\langle\phi_i, u_k - u\rangle|^q\Big] = -4\sum_{i\in\mathbb{N}} w_i |\langle\phi_i, u\rangle|^q\,.$$

Combining the above inequality and equalities we see that

$$\limsup_k \mathcal{R}_q(u_k - u) \le 4\,\mathcal{R}_q(u) - 4\sum_{i\in\mathbb{N}} w_i |\langle\phi_i, u\rangle|^q = 0$$

or, equivalently, that $\mathcal{R}_q(u_k - u) \to 0$. $\square$

Remark 3. Convergence with respect to $\mathcal{R}_q$ implies convergence with respect to the norm, which is an easy consequence of the inequality

$$\Big(\sum_{i\in\mathbb{N}} |c_i|^t\Big)^{1/t} \le \Big(\sum_{i\in\mathbb{N}} |c_i|^s\Big)^{1/s} =: |c|_s \tag{4}$$

for $c = (c_i)_{i\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ and $0 < s \le t < \infty$. The inequality (4) easily follows for $0 < |c|_s < \infty$ from the inequality

$$\sum_{i\in\mathbb{N}} \Big(\frac{|c_i|}{|c|_s}\Big)^t \le \sum_{i\in\mathbb{N}} \Big(\frac{|c_i|}{|c|_s}\Big)^s = 1\,.$$

In particular, this shows that

$$\mathcal{R}_q(u) \geq w_{\min} \sum_{i\in\mathbb{N}} |\langle \phi_i, u\rangle|^q \geq w_{\min} \Big(\sum_{i\in\mathbb{N}} |\langle \phi_i, u\rangle|^2 \Big)^{q/2} \geq w_{\min} C_1^{q/2} \|u\|^q \tag{5}$$

for every $u \in U$. Therefore, Lemma 2 implies [10, Lemma 4.3], where the authors show convergence of the sequence $(u_k)_{k\in\mathbb{N}}$ with respect to the norm.

Another immediate consequence of (5) is the weak coercivity of the functional $\mathcal{R}_q$. ∎

Lemma 4. *Let $(u_k)_{k\in\mathbb{N}} \subset \mathrm{dom}(F)$ and $(v_k)_{k\in\mathbb{N}} \subset V$. Assume that the sequence $(v_k)_{k\in\mathbb{N}}$ is bounded in V and that there exist $\alpha > 0$ and $M > 0$ such that $\mathcal{T}^{p,q}_{\alpha,v_k}(u_k) < M$ for all $k \in \mathbb{N}$. Then there exist $u \in \mathrm{dom}(F)$ and a subsequence $(u_{k_j})_{j\in\mathbb{N}}$ such that $u_{k_j} \rightharpoonup u$ and $F(u_{k_j}) \rightharpoonup F(u)$.*

Proof. The coercivity of $\mathcal{R}_q$ and the estimate $\mathcal{T}^{p,q}_{\alpha,v_k}(u_k) \geq \alpha \mathcal{R}_q(u_k)$ imply that the sequence $(u_k)_{k\in\mathbb{N}}$ is bounded in U. Similarly, since $(v_k)_{k\in\mathbb{N}}$ is bounded, also the sequence $\big(F(u_k)\big)_{k\in\mathbb{N}}$ is bounded in V. Therefore there exist a subsequence $(u_{k_j})_{j\in\mathbb{N}}$ and $u \in U$, $y \in V$, such that $(u_{k_j})_{j\in\mathbb{N}}$ weakly converges to u and $\big(F(u_{k_j})\big)_{j\in\mathbb{N}}$ weakly converges to y. Since F is weakly sequentially closed, it follows that $u \in \mathrm{dom}(F)$ and $F(u) = y$. □

The ideas of the following proofs are based on [19, Section 3]. Still, we provide short proofs, since our assumptions are slightly different from [19], where weak continuity of the operator F is assumed.

Proposition 5 (Existence). *For every $v^\delta \in V$ the functional $\mathcal{T}^{p,q}_{\alpha,v^\delta}$ has a minimizer in U.*

Proof. Let $(u_k)_{k\in\mathbb{N}}$ satisfy

$$\lim_{k\to\infty} \mathcal{T}^{p,q}_{\alpha,v^\delta}(u_k) = \inf\{\mathcal{T}^{p,q}_{\alpha,v^\delta}(u) : u \in U\} .$$

Lemma 4 shows that there exists a subsequence $(u_{k_j})_{j\in\mathbb{N}}$ weakly converging to some $u \in U$ such that $F(u_{k_j}) \rightharpoonup F(u)$. Therefore the weak sequential lower semi-continuity of $\mathcal{T}^{p,q}_{\alpha,v^\delta}$ implies that u is a minimizer of $\mathcal{T}^{p,q}_{\alpha,v^\delta}$. □

Proposition 6 (Stability). *Let $(v_k)_{k\in\mathbb{N}}$ converge to $v^\delta \in V$ and let*

$$u_k \in \arg\min\big\{\mathcal{T}^{p,q}_{\alpha,v_k}(u) : u \in U\big\} .$$

Then there exists a subsequence $(u_{k_j})_{j\in\mathbb{N}}$ and a minimizer u_α^δ of $\mathcal{T}^{p,q}_{\alpha,v^\delta}$ such that $\mathcal{R}_q(u_\alpha^\delta - u_{k_j}) \to 0$. If the minimizer u_α^δ is unique, then $(u_k)_{k\in\mathbb{N}}$ converges to u_α^δ with respect to $\mathcal{R}_q$.

Proof. From Lemma 4 we obtain the existence of a subsequence $(u_{k_j})_{j\in\mathbb{N}}$ weakly converging to some $u \in \mathrm{dom}(F)$ such that $F(u_{k_j}) \rightharpoonup F(u)$. Since $v_k \to v^\delta$, it follows that $\mathcal{T}^{p,q}_{\alpha,v^\delta}(u) \leq \liminf_j \mathcal{T}^{p,q}_{\alpha,v_{k_j}}(u_{k_j})$.

On the other hand, if $\tilde{u} \in \mathrm{dom}(F)$, then

$$\mathcal{T}^{p,q}_{\alpha,v^\delta}(\tilde{u}) = \lim_k \mathcal{T}^{p,q}_{\alpha,v_k}(\tilde{u}) \geq \liminf_k \mathcal{T}^{p,q}_{\alpha,v_k}(u_k) .$$

Thus $u = u_\alpha^\delta$ is a minimizer of $\mathcal{T}^{p,q}_{\alpha,v^\delta}$.

Now note that also $\mathcal{T}^{p,q}_{\alpha,v^\delta}(u_{k_j}) \to \mathcal{T}^{p,q}_{\alpha,v^\delta}(u)$. Since both $\|\cdot\|^p$ and $\mathcal{R}_q$ are weakly sequentially lower semi-continuous, this implies that $\mathcal{R}_q(u_{k_j}) \to \mathcal{R}_q(u)$. Using Lemma 2, we therefore obtain the convergence of the sequence $(u_{k_j})_{j\in\mathbb{N}}$ with respect to $\mathcal{R}_q$.

In case the minimizer u_α^δ is unique, the convergence of the original sequence $(u_k)_{k\in\mathbb{N}}$ to u_α^δ follows from a subsequence argument. □

Proposition 7 (Convergence). *Assume that the operator equation $F(u) = v$ attains a solution in $\operatorname{dom}(\mathcal{R}_q)$ and that $\alpha\colon \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ satisfies*

$$\alpha(\delta) \to 0 \qquad \text{and} \qquad \frac{\delta^p}{\alpha(\delta)} \to 0 \qquad \text{as } \delta \to 0 \,.$$

Let $\delta_k \to 0$ and let $v_k \in V$ satisfy $\|v_k - v\| \le \delta_k$. Moreover, let $\alpha_k = \alpha(\delta_k)$ and

$$u_k \in \arg\min\left\{\mathcal{T}^{p,q}_{\alpha_k,v_k}(u) : u \in U\right\} .$$

Then there exist an $\mathcal{R}_q$-minimizing solution $u^\dagger$ of $F(u) = v$ and a subsequence $(u_{k_j})_{j\in\mathbb{N}}$ with $\mathcal{R}_q(u^\dagger - u_{k_j}) \to 0$. If the $\mathcal{R}_q$-minimizing solution is unique, then $(u_k)_{k\in\mathbb{N}}$ converges to $u^\dagger$ with respect to $\mathcal{R}_q$.

Proof. Let $\tilde u \in \operatorname{dom}(\mathcal{R}_q)$ be any solution of $F(u) = v$. The definition of u_k implies that

$$\|F(u_k) - v_k\|^p + \alpha_k \mathcal{R}_q(u_k) \le \|F(\tilde u) - v_k\|^p + \alpha_k \mathcal{R}_q(\tilde u) \le \delta_k^p + \alpha_k \mathcal{R}_q(\tilde u) \,.$$

In particular $\|F(u_k) - v_k\| \to 0$ and

$$\limsup_k \mathcal{R}_q(u_k) \le \mathcal{R}_q(\tilde u) + \limsup_k \frac{\delta_k^p}{\alpha_k} = \mathcal{R}_q(\tilde u) \,. \tag{6}$$

This shows that there exists $M > 0$ such that $\mathcal{T}^{p,q}_{\alpha_1,v_k}(u_k) \le M$ for all $k \in \mathbb{N}$. Thus Lemma 4 yields a subsequence $(u_{k_j})_{j\in\mathbb{N}}$ weakly converging to some $u^\dagger \in \operatorname{dom}(F)$ such that $F(u_{k_j}) \rightharpoonup F(u^\dagger)$. Since $\|F(u_{k_j}) - v\| \le \|F(u_{k_j}) - v_{k_j}\| + \|v_{k_j} - v\| \to 0$, it follows that $F(u^\dagger) = v$.

The weak sequential lower semi-continuity of $\mathcal{R}_q$ implies that $\mathcal{R}_q(u^\dagger) \le \liminf_j \mathcal{R}_q(u_{k_j})$. Since (6) holds for every $\tilde u \in \operatorname{dom}(\mathcal{R}_q)$ satisfying $F(\tilde u) = v$, it follows that $u^\dagger$ is an $\mathcal{R}_q$-minimizing solution of $F(u) = v$ and that $\mathcal{R}_q(u_{k_j}) \to \mathcal{R}_q(u^\dagger)$. Lemma 2 now shows that $(u_{k_j})_{j\in\mathbb{N}}$ converges to $u^\dagger$ with respect to $\mathcal{R}_q$.

Again, the convergence of the original sequence $(u_k)_{k\in\mathbb{N}}$ to $u^\dagger$ follows from a subsequence argument, if the $\mathcal{R}_q$-minimizing solution $u^\dagger$ is unique. □

In the following we write $\alpha \sim \delta^s$ for $\alpha\colon (0,\infty) \to (0,\infty)$ and $s > 0$, if there exist constants $C \ge c > 0$ and $\delta_0 > 0$, such that $c\delta^s \le \alpha(\delta) \le C\delta^s$ for every $0 < \delta < \delta_0$.

For the next result on convergence rates recall the definition of the exponent r in Assumption 1.

Proposition 8 (Convergence Rates). *Let Assumption 1 hold. Assume that $v^\delta \in V$ satisfies $\|v^\delta - v\| \le \delta$ and $u_\alpha^\delta \in \arg\min\left\{\mathcal{T}^{p,q}_{\alpha,v^\delta}(u) : u \in U\right\}$. For α and δ sufficiently small we obtain the following estimates:*

If $p = 1$ and $\alpha\beta_2 < 1$, then

$$\|u_\alpha^\delta - u^\dagger\|^r \le \frac{(1+\alpha\beta_2)\,\delta}{\alpha\beta_1}\,, \qquad \|F(u_\alpha^\delta) - v^\delta\| \le \frac{(1+\alpha\beta_2)\,\delta}{1-\alpha\beta_2}\,.$$

If $p > 1$, then

$$\begin{aligned}
\|u_\alpha^\delta - u^\dagger\|^r &\le \frac{\delta^p + \alpha\beta_2\delta + (\alpha\beta_2)^{p_*}/p_*}{\alpha\beta_1}\,,\\
\|F(u_\alpha^\delta) - v^\delta\|^p &\le p_*\delta^p + p_*\alpha\beta_2\delta + (\alpha\beta_2)^{p_*}\,.
\end{aligned}$$

Here, p_* is the conjugate of p defined by $1/p_* + 1/p = 1$.

In particular, if $\alpha \sim \delta^{p-1}$, then $\|u_\alpha^\delta - u^\dagger\| = O(\delta^{1/r})$.

Proof. Since u_α^δ minimizes $\mathcal{T}_{\alpha,v^\delta}^{p,q}$, the inequality

$$\|F(u_\alpha^\delta) - v^\delta\|^p + \alpha\mathcal{R}_q(u_\alpha^\delta) \le \|F(u^\dagger) - v^\delta\|^p + \alpha\mathcal{R}_q(u^\dagger)$$

holds. Assumption 1 and the fact that $F(u^\dagger) = v$ therefore imply that

$$\begin{aligned}
\delta^p &\ge \|F(u_\alpha^\delta) - v^\delta\|^p + \alpha\big(\mathcal{R}_q(u_\alpha^\delta) - \mathcal{R}_q(u^\dagger)\big)\\
&\ge \|F(u_\alpha^\delta) - v^\delta\|^p + \alpha\,\beta_1\,\|u_\alpha^\delta - u^\dagger\|^r - \alpha\,\beta_2\,\|F(u_\alpha^\delta) - F(u^\dagger)\|\\
&\ge \|F(u_\alpha^\delta) - v^\delta\|^p + \alpha\,\beta_1\,\|u_\alpha^\delta - u^\dagger\|^r - \alpha\,\beta_2\,\|F(u_\alpha^\delta) - v^\delta\| - \alpha\beta_2\delta\,.
\end{aligned}$$

This shows the assertion in the case $p = 1$.

If $p > 1$, we apply Young's inequality $ab \le a^p/p + b^{p_*}/p_*$ with $a = \|F(u_\alpha^\delta) - v^\delta\|$ and $b = \alpha\beta_2$. Then again the assertion follows. □

Remark 9. Proposition 8 shows that sparsity regularization is an *exact method* for $p = 1$, that is, it yields exact solutions $u^\dagger$ for noise free data and $\alpha < 1/\beta_2$. ■

4 Relations to Source Conditions

We now investigate Assumption 1 more closely and show that it is indeed a generalization of commonly imposed source conditions involving the Bregman distance defined by the functional $\mathcal{R}_q$ (see e.g. [5, 19]). The basis of these results is the following lemma, which relates the Bregman distance to the squared norm on U in case $q > 1$. This result is a consequence of a special case of [2, Lemma 2.7] (see also [6, Corollary 3.7]).

From now on we assume that $(\phi_i)_{i\in\mathbb{N}}$ is an orthonormal basis.

Lemma 10. *Let $1 < q \le 2$. There exists a constant $c_q > 0$ only depending on q such that*

$$\mathcal{D}_B(\tilde{u}, u) := \mathcal{R}_q(\tilde{u}) - \mathcal{R}_q(u) - \langle\partial\mathcal{R}_q(u), \tilde{u} - u\rangle \ge \frac{c_q\,\|\tilde{u} - u\|^2}{3w_{\min} + 2\mathcal{R}_q(u) + \mathcal{R}_q(\tilde{u})}$$

for all $\tilde{u}$, $u \in \operatorname{dom}(\mathcal{R}_q)$ for which $\partial\mathcal{R}_q(u) \ne \emptyset$, which is equivalent to the assumption that $\sum_{i\in\mathbb{N}} w_i^2\,|\langle\phi_i, u\rangle|^{2(q-1)} < \infty$.

Proof. There exists $d_q > 0$ such that

$$d_q|a-b|^2 \le (|a|^{2-q} + |a-b|^{2-q})\Big[|b|^q - |a|^q - q|a|^{q-1}\operatorname{sgn}(a)\,(b-a)\Big] \tag{7}$$

for all a, $b \in \mathbb{R}$ [13, §5, Eq. 1].

Let $\tilde{u} \neq u \in \operatorname{dom}(\mathcal{R}_q)$. Then

$$\partial\mathcal{R}_q(u) = \sum_{i\in\mathbb{N}} q\, w_i\, |\langle\phi_i, u\rangle|^{q-1} \operatorname{sgn}(\langle\phi_i, u\rangle)\, \phi_i$$

provided that $\partial\mathcal{R}_q(u) \neq \emptyset$. Applying (7), we see that

$$\begin{aligned}
&\mathcal{R}_q(\tilde{u}) - \mathcal{R}_q(u) - \langle\partial\mathcal{R}_q, \tilde{u} - u\rangle \\
&\quad= \sum_{i\in\mathbb{N}} w_i \Big[|\langle\phi_i, \tilde{u}\rangle|^q - |\langle\phi_i, u\rangle|^q - q\,|\langle\phi_i, u\rangle|^{q-1} \operatorname{sgn}(\langle\phi_i, u\rangle)\,\langle\phi_i, \tilde{u} - u\rangle\Big] \\
&\quad\ge d_q \sum_{i\in\mathbb{N}} \frac{w_i\,|\langle\phi_i, \tilde{u} - u\rangle|^2}{|\langle\phi_i, u\rangle|^{2-q} + |\langle\phi_i, \tilde{u} - u\rangle|^{2-q}} \\
&\quad\ge \frac{d_q\, w_{\min}}{\max\{|\langle\phi_i, u\rangle|^{2-q} + |\langle\phi_i, \tilde{u} - u\rangle|^{2-q} : i \in \mathbb{N}\}} \sum_{i\in\mathbb{N}} |\langle\phi_i, \tilde{u} - u\rangle|^2 \\
&\quad\ge \frac{d_q\, w_{\min}}{\max\{2|\langle\phi_i, u\rangle|^{2-q} + |\langle\phi_i, \tilde{u}\rangle|^{2-q} : i \in \mathbb{N}\}} \sum_{i\in\mathbb{N}} |\langle\phi_i, \tilde{u} - u\rangle|^2 \\
&\quad\ge \frac{d_q\, w_{\min}}{3 + \max\{2|\langle\phi_i, u\rangle|^{q} + |\langle\phi_i, \tilde{u}\rangle|^{q} : i \in \mathbb{N}\}} \sum_{i\in\mathbb{N}} |\langle\phi_i, \tilde{u} - u\rangle|^2 \\
&\quad\ge \frac{d_q\, w_{\min}^2}{3w_{\min} + 2\mathcal{R}_q(u) + \mathcal{R}_q(\tilde{u})} \sum_{i\in\mathbb{N}} |\langle\phi_i, \tilde{u} - u\rangle|^2 \,.
\end{aligned} \tag{8}$$

Here, the third and second to last estimates follow from the inequalities $(a+b)^{2-q} \le a^{2-q} + b^{2-q}$ and $a^{q-2} \le 1 + a^q$ for a, $b \ge 0$. Thus the assertion follows by setting $c_q := d_q\, w_{\min}^2$. □

Proposition 11. *Let F be a bounded linear operator on U, $1 < q \le 2$, and $u^\dagger$ an $\mathcal{R}_q$-minimizing solution of $F(u) = v$. Then Assumption 1 with $r = 2$ is equivalent to the source condition*

$$\partial\mathcal{R}_q(u^\dagger) \in \operatorname{range}(F^*) \,. \tag{9}$$

In particular, if $\alpha \sim \delta^{p-1}$, then $\|u_\alpha^\delta - u^\dagger\| = O(\sqrt{\delta})$.

Proof. First assume that (9) holds. The condition $\partial\mathcal{R}_q(u^\dagger) \in \operatorname{range}(F^*)$ implies the existence of a constant $\hat{C} > 0$ such that

$$|\langle\partial\mathcal{R}_q(u^\dagger), u - u^\dagger\rangle| \le \hat{C}\|F(u - u^\dagger)\| \tag{10}$$

for all $u \in U$. Together with Lemma 10 this yields the inequality

$$\begin{aligned}
\mathcal{R}_q(u) - \mathcal{R}_q(u^\dagger) &\ge \frac{c_q}{3w_{\min} + 2\mathcal{R}_q(u^\dagger) + \mathcal{R}_q(u)}\,\|u - u^\dagger\|^2 + \langle\partial\mathcal{R}_q(u^\dagger), u - u^\dagger\rangle \\
&\ge \frac{c_q}{3w_{\min} + 2\mathcal{R}_q(u^\dagger) + \mathcal{R}_q(u)}\,\|u - u^\dagger\|^2 - \hat{C}\|F(u - u^\dagger)\| \,.
\end{aligned}$$

Thus, Assumption 1 is satisfied if we choose $r = 2$, $\rho = \mathcal{R}_q(u^\dagger) + w_{\min}$, $\beta_1 = c_q/(4w_{\min} + 3\mathcal{R}_q(u^\dagger))$, and $\beta_2 = \hat{C}$.

In order to show the converse implication, let Assumption 1 be satisfied for $r = 2$, that is, there exist β_1, $\beta_2 > 0$ such that

$$\beta_1 \|u - u^\dagger\|^2 \le \mathcal{R}_q(u) - \mathcal{R}_q(u^\dagger) + \beta_2 \|F(u - u^\dagger)\|$$

in a neighborhood of $u^\dagger$. Both sides of this inequality are convex functions in the variable u that agree for $u = u^\dagger$. This implies that the subgradient at $u^\dagger$ of the left hand side, which equals zero, is contained in the subgradient at $u^\dagger$ of the right hand side. In other words,

$$0 \in \partial \mathcal{R}_q(u^\dagger) + \beta_2 F^* \partial\big(\|F(u - u^\dagger)\|\big) .$$

Consequently the source condition (9) holds. □

The following result states that the condition proposed in [19] for obtaining convergence rates in the non-linear, non-smooth case also follows from Assumption 1 with exponent $r = 2$.

Proposition 12. *Let $1 < q < 2$ and $u^\dagger$ an $\mathcal{R}_q$-minimizing solution of $F(u) = v$. Assume that there exist $0 \le \gamma_1 < 1$, $\gamma_2 > 0$, and $\rho > \mathcal{R}_q(u^\dagger)$ such that*

$$\big\langle \partial \mathcal{R}_q(u^\dagger), u^\dagger - u \big\rangle \le \gamma_1 \, \mathcal{D}_B(u, u^\dagger) + \gamma_2 \, \|F(u) - F(u^\dagger)\| \qquad (11)$$

for all $u \in \mathrm{dom}(F)$ with $\mathcal{R}_q(u) < \rho$. Then Assumption 1 holds with $r = 2$. In particular, if $\alpha \sim \delta^{p-1}$, then $\|u_\alpha^\delta - u^\dagger\| = O(\sqrt{\delta})$.

Proof. Using (11) and Lemma 10 we obtain that

$$\begin{aligned} \gamma_1\big(\mathcal{R}_q(u) - \mathcal{R}_q(u^\dagger)\big) &\ge -(1-\gamma_1)\big\langle \partial \mathcal{R}_q(u^\dagger), u - u^\dagger \big\rangle - \gamma_2 \, \|F(u) - F(u^\dagger)\| \\ &\ge \tilde{\beta} \|u - u^\dagger\|^2 - \gamma_2 \, \|F(u) - F(u^\dagger)\| - (1-\gamma_1)\big(\mathcal{R}_q(u) - \mathcal{R}_q(u^\dagger)\big) , \end{aligned}$$

where $\tilde{\beta} := (1-\gamma_1)\, c_q/(3w_{\min} + 2\mathcal{R}_q(u^\dagger) + \rho)$. Thus Assumption 1 follows with $\beta_1 = \tilde{\beta}/(1 + 2\gamma_1)$ and $\beta_2 = \gamma_2/(1 + 2\gamma_1)$. □

5 Convergence Rates for Sparse Solutions

We have seen above that appropriate source conditions imply convergence rates of type $\sqrt{\delta}$. These rates in fact can be improved considerably, if the $\mathcal{R}_q$-minimizing solution $u^\dagger$ is sparse with respect to $(\phi_i)_{i\in\mathbb{N}}$ in the sense that the set

$$J := \big\{ i \in \mathbb{N} : \langle u^\dagger, \phi_i \rangle \ne 0 \big\}$$

is finite.

Assumption 13. *Assume that the following hold:*

1. *The operator equation $F(u) = v$ has an $\mathcal{R}_q$-minimizing solution $u^\dagger$ that is sparse with respect to $(\phi_i)_{i\in\mathbb{N}}$.*

2. *The operator F is Gâteaux differentiable at $u^\dagger$, and for every finite set $J \subset \mathbb{N}$ the restriction of its derivative $F'(u^\dagger)$ to* $\operatorname{span}\{\phi_j : j \in J\}$ *is injective.*

3. *There exist γ_1, $\gamma_2 > 0$, $\sigma > 0$, and $\rho > \mathcal{R}_q(u^\dagger)$ such that*

$$\mathcal{R}_q(u) - \mathcal{R}_q(u^\dagger) \geq \gamma_1 \big\| F(u) - F(u^\dagger) - F'(u^\dagger)(u-u^\dagger) \big\| - \gamma_2 \big\| F(u) - F(u^\dagger) \big\| \tag{12}$$

   for all $u \in \operatorname{dom}(F)$ satisfying $\mathcal{R}_q(u) < \rho$ and $\|F(u) - F(u^\dagger)\| < \sigma$.

We first derive a convergence rates result of order $\delta^{1/q}$ for $q > 1$.

Theorem 14 ($q > 1$). *Let $1 < q \leq 2$ and assume that Assumption 13 holds. Then for a parameter choice strategy $\alpha \sim \delta^{p-1}$ we obtain the convergence rate*

$$\|u_\alpha^\delta - u^\dagger\| = O(\delta^{1/q}) \, .$$

Proof. We verify Assumption 1 with $r = q$ and appropriate constants β_1, $\beta_2 > 0$. Then the assertion follows from Proposition 8.

Let therefore $u \in U$ satisfy $\mathcal{R}_q(u) < \rho$ and $\|F(u) - F(u^\dagger)\| < \sigma$.

Define $J := \{i \in \mathbb{N} : \langle u^\dagger, \phi_i \rangle \neq 0\}$ and $W := \operatorname{span}\{\phi_j : j \in J\}$. Since $u^\dagger$ is sparse, the set J is finite. Therefore, the restriction of $F'(u^\dagger)$ to W is injective, which implies the existence of a constant $C > 0$ such that

$$C \|F'(u^\dagger)\, w\| \geq \|w\| \qquad \text{for all } w \in W \, .$$

Now denote by π_W, $\pi_W^\perp \colon U \to U$ the projections

$$\pi_W u := \sum_{j \in J} \langle \phi_j, u \rangle\, \phi_j \, , \qquad \pi_W^\perp u := \sum_{j \notin J} \langle \phi_j, u \rangle\, \phi_j \, .$$

Note that by assumption $\langle \phi_j, u^\dagger \rangle = 0$ for every $j \notin J$, which implies that $u^\dagger = \pi_W u^\dagger$ and $\pi_W^\perp u^\dagger = 0$. By means of the inequality

$$(a+b)^q \leq 2^{q-1}(a^q + b^q) \leq 2(a^q + b^q) \qquad \text{for every } a,\, b > 0$$

it therefore follows that

$$\begin{aligned} \|u - u^\dagger\|^q &\leq 2 \, \big\| \pi_W(u - u^\dagger) \big\|^q + 2 \big\| \pi_W^\perp u \big\|^q \\ &\leq 2\, C^q \big\| F'(u^\dagger) \big(\pi_W(u - u^\dagger) \big) \big\|^q + 2 \, \big\| \pi_W^\perp u \big\|^q \\ &\leq 4\, C^q \big\| F'(u^\dagger)(u - u^\dagger) \big\|^q + 2 \big(1 + 2 C^q \|F'(u^\dagger)\|^q \big) \big\| \pi_W^\perp u \big\|^q \, . \end{aligned} \tag{13}$$

We now derive an estimate for $\|\pi_W^\perp u\|^q$. Using (4) we see that

$$\big\| \pi_W^\perp u \big\|^q = \Big(\sum_{i \notin J} |\langle \phi_i, u \rangle|^2 \Big)^{q/2} \leq \sum_{i \notin J} |\langle \phi_i, u \rangle|^q \leq w_{\min}^{-1} \sum_{i \notin J} w_i |\langle \phi_i, u \rangle|^q \, . \tag{14}$$

Since $q > 1$, the inequality

$$|\langle \phi_i, u \rangle|^q - |\langle \phi_i, u^\dagger \rangle|^q - q |\langle \phi_i, u^\dagger \rangle|^{q-1} \operatorname{sgn}\big(|\langle \phi_i, u^\dagger \rangle| \big) \langle \phi_i, u - u^\dagger \rangle \geq 0$$

holds for all $i \in \mathbb{N}$. Consequently,

$$\begin{aligned}
&\sum_{i \notin J} w_i |\langle \phi_i, u\rangle|^q \\
&\quad = \sum_{i \notin J} w_i \Big[|\langle \phi_i, u\rangle|^q - |\langle \phi_i, u^\dagger\rangle|^q - q|\langle \phi_i, u^\dagger\rangle|^{q-1} \operatorname{sgn}\big(|\langle \phi_i, u^\dagger\rangle|\big) \langle \phi_i, u - u^\dagger\rangle \Big] \\
&\quad \le \sum_{i \in \mathbb{N}} w_i \Big[|\langle \phi_i, u\rangle|^q - |\langle \phi_i, u^\dagger\rangle|^q - q|\langle \phi_i, u^\dagger\rangle|^{q-1} \operatorname{sgn}\big(|\langle \phi_i, u^\dagger\rangle|\big) \langle \phi_i, u - u^\dagger\rangle \Big] \\
&\quad = \mathcal{R}_q(u) - \mathcal{R}_q(u^\dagger) - \big\langle \partial \mathcal{R}_q(u^\dagger), u - u^\dagger \big\rangle \, .
\end{aligned} \tag{15}$$

From (12) we obtain by considering $u = u^\dagger + t\tilde{u}$, dividing by t, and passing to the limit $t \to 0$ that

$$\big\langle \partial \mathcal{R}_q(u^\dagger), \tilde{u} \big\rangle \ge -\gamma_2 \|F'(u^\dagger)\tilde{u}\| \qquad \text{for all } \tilde{u} \in U \, . \tag{16}$$

Together with (12) this implies the inequality

$$\begin{aligned}
&\mathcal{R}_q(u) - \mathcal{R}_q(u^\dagger) - \langle \partial \mathcal{R}_q(u^\dagger), u - u^\dagger \rangle \\
&\qquad \le \mathcal{R}_q(u) - \mathcal{R}_q(u^\dagger) + \gamma_2 \|F'(u^\dagger)(u - u^\dagger)\| \\
&\qquad \le \mathcal{R}_q(u) - \mathcal{R}_q(u^\dagger) + \gamma_2 \|F(u) - F(u^\dagger)\| \\
&\qquad\qquad\qquad + \gamma_2 \|F(u) - F(u^\dagger) - F'(u^\dagger)(u - u^\dagger)\| \\
&\qquad \le (1 + \gamma_2/\gamma_1)\big(\mathcal{R}_q(u) - \mathcal{R}_q(u^\dagger)\big) + \gamma_2(1 + \gamma_2/\gamma_1)\|F(u) - F(u^\dagger)\| \, .
\end{aligned} \tag{17}$$

Combination of estimates (14)–(17) yields

$$w_{\min} \big\|\pi_W^\perp u\big\|^q \le (1 + \gamma_2/\gamma_1)\big(\mathcal{R}_q(u) - \mathcal{R}_q(u^\dagger)\big) + \gamma_2(1 + \gamma_2/\gamma_1)\|F(u) - F(u^\dagger)\| \, . \tag{18}$$

It remains to find an estimate for $\|F'(u^\dagger)(u - u^\dagger)\|^q$. Since by assumption $\mathcal{R}_q(u^\dagger)$, $\mathcal{R}_q(u) < \rho$, and $\|F(u) - F(u^\dagger)\| < \sigma$, it follows from (12) that

$$\begin{aligned}
&\|F'(u^\dagger)(u - u^\dagger)\|^q \\
&\quad \le 2^{q-1} \big\|F(u) - F(u^\dagger) - F'(u^\dagger)(u - u^\dagger)\big\|^q + 2^{q-1} \|F(u) - F(u^\dagger)\|^q \\
&\quad \le \frac{2^{q-1}}{\gamma_1^q} \Big(\mathcal{R}_q(u) - \mathcal{R}_q(u^\dagger) + \gamma_2 \big\|F(u) - F(u^\dagger)\big\| \Big)^q + 2^{q-1} \big\|F(u) - F(u^\dagger)\big\|^q \\
&\quad \le \frac{2^{q-1}(\rho + \gamma_2\sigma)^{q-1}}{\gamma_1^q} \Big(\mathcal{R}_q(u) - \mathcal{R}_q(u^\dagger) \Big) \\
&\qquad\qquad + \Big(2^{q-1}\sigma^{q-1} + \frac{2^{q-1}(\rho + \gamma_2\sigma)^{q-1}\gamma_2}{\gamma_1^q} \Big) \big\|F(u) - F(u^\dagger)\big\| \, .
\end{aligned} \tag{19}$$

Combining the inequalities (13), (18), and (19), we obtain the assertion. □

The argumentation in the proof of Theorem 14 cannot be applied directly to the case $q = 1$. The main difficulty is that here the estimate (16) does not follow from (12), since the subgradient of $\mathcal{R}_1$ is not single valued. Therefore it is necessary to postulate the existence of a subgradient element $\xi \in \partial \mathcal{R}_1(u^\dagger)$ for which such an inequality holds.

Theorem 15 ($q = 1$). *Let $q = 1$ and assume that Assumption 13 holds. In addition we assume the existence of $\xi \in \partial\mathcal{R}_1(u^\dagger)$ and $\gamma_3 > 0$ such that*

$$\mathcal{R}_1(u) - \mathcal{R}_1(u^\dagger) \geq -\gamma_3\big\langle \xi, u - u^\dagger\big\rangle - \gamma_2\big\|F(u) - F(u^\dagger)\big\| \tag{20}$$

for all $u \in \mathrm{dom}(F)$ with $\mathcal{R}_1(u) < \rho$ and $\|F(u) - F(u^\dagger)\| < \sigma$.

Then it follows for a parameter choice strategy $\alpha \sim \delta^{p-1}$ that

$$\|u_\alpha^\delta - u^\dagger\| = O(\delta)\ .$$

Proof. We show that Assumption 1 holds with $r = 1$. Then the result follows from Proposition 8.

Define $J := \{i \in \mathbb{N} : |\langle \phi_i, \xi\rangle| \geq w_{\min}\}$ and $W := \mathrm{span}\{\phi_j : j \in J\}$. Since $\xi \in U$, it follows that J is a finite set. Therefore there exists $C > 0$ such that $C\|F'(u^\dagger)w\| \geq \|w\|$ for all $w \in W$.

By assumption we have that $\langle \phi_i, u^\dagger\rangle = 0$ for every $i \notin J$. Proceeding as in the proof of Theorem 14, we obtain that

$$\|u - u^\dagger\| \leq C\big\|F'(u^\dagger)(u - u^\dagger)\big\| + \big(1 + C\|F'(u^\dagger)\|\big)\|\pi_W^\perp u\|\ .$$

Denote now $m := \max\big\{|\langle \phi_i, \xi\rangle| : i \notin J\big\}$, which is well-defined, as $\big(\langle \phi_i, \xi\rangle\big)_{i\in\mathbb{N}} \in l^2$ and therefore converges to zero. Using the inequalities $0 \leq m < w_{\min}$ and $\langle \phi_i, \xi\rangle \leq m$, the assumption $\xi \in \partial\mathcal{R}_1(u^\dagger)$, and (20), we can therefore estimate

$$\begin{aligned}
\|\pi_W^\perp u\| &= \Big(\sum_{i\notin J} |\langle \phi_i, u\rangle|^2\Big)^{1/2} \leq \sum_{i\notin J} |\langle \phi_i, u\rangle| \\
&\leq \frac{1}{w_{\min} - m}\sum_{i\notin J}(w_i - m)|\langle \phi_i, u\rangle| \\
&\leq \frac{1}{w_{\min} - m}\sum_{i\notin J}\big(w_i|\langle \phi_i, u\rangle| - \langle \phi_i, \xi\rangle\langle \phi_i, u\rangle\big) \\
&\leq \frac{1}{w_{\min} - m}\sum_{i\in\mathbb{N}}\big(w_i|\langle \phi_i, u\rangle| - w_i|\langle \phi_i, u^\dagger\rangle| - \langle \phi_i, \xi\rangle\langle \phi_i, u - u^\dagger\rangle\big) \\
&= \frac{1}{w_{\min} - m}\Big(\mathcal{R}_1(u) - \mathcal{R}_1(u^\dagger) - \big\langle \xi, u - u^\dagger\big\rangle\Big) \\
&\leq \frac{1}{w_{\min} - m}\Big(\big(1 + \gamma_3^{-1}\big)\big(\mathcal{R}_1(u) - \mathcal{R}_1(u^\dagger)\big) + \gamma_2/\gamma_3\big\|F(u) - F(u^\dagger)\big\|\Big)\ .
\end{aligned}$$

Here, the third to last line follows from the definition of the subgradient and the fact that $\langle \phi_i, u^\dagger\rangle = 0$ for $i \notin J$.

For $\|F'(u^\dagger)(u - u^\dagger)\|$ we obtain from (12) the estimate

$$\begin{aligned}
\big\|F'(u^\dagger)(u - u^\dagger)\big\| &\leq \big\|F(u) - F(u^\dagger) - F'(u^\dagger)(u - u^\dagger)\big\| + \big\|F(u) - F(u^\dagger)\big\| \\
&\leq \gamma_1^{-1}\big(\mathcal{R}_1(u) - \mathcal{R}_1(u^\dagger)\big) + (1 + \gamma_2/\gamma_1)\big\|F(u) - F(u^\dagger)\big\|\ .
\end{aligned}$$

Again, the assertion follows by collecting the above inequalities. □

Remark 16. Note that in fact for the convergence rates to hold the injectivity of $F'(u^\dagger)$ is only required on the subspace W defined in the proofs of Theorems 14 and 15. ■

Remark 17. Consider now the special case, where $F\colon U \to V$ is linear and bounded. Then (12) with $1 \le q \le 2$ is equivalent to the source condition

$$\partial\mathcal{R}_q(u^\dagger) \cap \operatorname{range}(F^*) \neq \emptyset \,.$$

Indeed, in this case the operator F equals its differential and therefore (12) reads as

$$\mathcal{R}_q(u) - \mathcal{R}_q(u^\dagger) \ge -\gamma_2 \big\| F(u-u^\dagger) \big\| \,, \tag{21}$$

which is equivalent to the existence of some $\omega \in \partial\mathcal{R}_q(u^\dagger)$ satisfying

$$\big\langle \omega, u - u^\dagger \big\rangle \ge -\gamma_2 \big\| F(u-u^\dagger) \big\| \,.$$

This last inequality is in turn equivalent to the condition $\partial\mathcal{R}_q(u^\dagger) \cap \operatorname{range}(F^*) \neq \emptyset$, which shows the assertion.

In the case $q = 1$ the inequality (20) with $\gamma_3 = 1/2$ follows from (21), since

$$\mathcal{R}_1(u) - \mathcal{R}_1(u^\dagger) + \frac{1}{2}\big\langle \xi, u-u^\dagger \big\rangle \ge \frac{1}{2}\big(\mathcal{R}_1(u) - \mathcal{R}_1(u^\dagger)\big) \ge -\frac{\gamma_2}{2}\big\| F(u-u^\dagger) \big\| \,.$$

As a consequence, the convergence rate $O(\delta^{1/q})$ follows from the range condition $\partial\mathcal{R}_q(u^\dagger) \cap \operatorname{range}(F^*) \neq \emptyset$ and the finite basis injectivity property, which postulates the injectivity of the restriction of F to every subspace of U spanned by a finite number of basis elements ϕ_i. ∎

6 Conclusion

We have studied the application of Tikhonov regularization with l^q type penalty term for $1 \le q \le 2$ to sparse regularization. In general, quadratic and l^q regularization enjoy the same basic properties concerning existence, stability, and convergence of the corresponding approximate solutions. If additionally q is strictly greater than one, then also the same convergence rates can be obtained provided a source condition holds.

For linear operators F this condition requires the subgradient of the penalty term to be contained in the range of the adjoint of F. This assumption implies convergence rates with respect to the Bregman distance, which for non-quadratic functionals in general cannot be compared with the norm on the Hilbert space. In the l^q case, however, such a comparison is possible and leads to convergence rates of order $\sqrt{\delta}$ in the norm.

Even better results hold if the true solution $u^\dagger$ of the considered problem is known to have a sparse representation in the chosen basis. Then the l^q regularization method yields rates of order $\delta^{1/q}$, as long as the derivative of the operator F at $u^\dagger$ is injective on the subspace spanned by the non-zero components of $u^\dagger$. For $q = 1$ and an additional assumption concerning the subgradient of the penalty term, this implies linear convergence of the regularized solutions to $u^\dagger$.

Acknowledgement

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, and by the Technology Transfer Office of the University of Innsbruck (transIT).

The authors want to express their thanks to Andreas Neubauer for his careful proofreading of the article and to the referees for their valuable suggestions and comments.

References

[1] T. Bonesky, K. Bredies, D. Lorenz, and P. Maass. A generalized conditional gradient method for nonlinear operator equations with sparsity constraints. *Inverse Probl.*, 23(5):2041–2058, 2007.

[2] T. Bonesky, K. S. Kazimierski, P. Maass, F. Schöpfer, and T. Schuster. Minimization of Tikhonov functionals in Banach spaces. *Abstr. Appl. Anal.*, 2008:Art. ID 192679, 19, 2008.

[3] K. Bredies and D. Lorenz. Iterated hard shrinkage for minimization problems with sparsity constraints. *SIAM J. Sci. Comput.*, 30(2):657–683, 2008.

[4] K. Bredies and D. Lorenz. Linear convergence of iterative soft-thresholding. *arXiv:0709.1598v4, to appear in J. Fourier Anal. Appl.*, 2008.

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

[6] D. Butnariu, A. N. Iusem, and C. Zălinescu. On uniform convexity, total convexity and convergence of the proximal point and outer Bregman projection algorithms in Banach spaces. *J. Convex Anal.*, 10(1):35–61, 2003.

[7] 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.

[8] P. L. Combettes and J.-C. Pesquet. Proximal thresholding algorithm for minimization over orthonormal bases. *SIAM J. Optim.*, 18(4):1351–1376, 2007.

[9] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.*, 4(4):1168–1200 (electronic), 2005.

[10] 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.

[11] I. Daubechies, M. Fornasier, and I. Loris. Accelerated projected gradient methods for linear inverse problems with sparsity constraints. *J. Fourier Anal. Appl.*, 2008. To appear.

[12] V. Dicken. A new approach towards simultaneous activity and attenuation reconstruction in emission tomography. *Inverse Probl.*, 15(4):931–960, 1999.

[13] J. Diestel. *Geometry of Banach spaces—selected topics*. Springer-Verlag, Berlin, 1975. Lecture Notes in Mathematics, Vol. 485.

[14] D. L. Donoho. Compressed sensing. *IEEE Trans. Inf. Theory*, 52(4):1289–1306, 2006.

[15] 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.

[16] H. W. Engl, K. Kunisch, and A. Neubauer. Convergence rates for Tikhonov regularisation of nonlinear ill-posed problems. *Inverse Probl.*, 5(3):523–540, 1989.

[17] M. Figueiredo, R. Nowak, and S. Wright. Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems. *IEEE J. Sel. Topics Signal Process.*, 1(4):586–598, 2007.

[18] R. Griesse and D. Lorenz. A semismooth newton method for tikhonov functionals with sparsity constraints. *Inverse Probl.*, 24(3):035007, 19, 2008.

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

[20] D. Lorenz. Convergence rates and source conditions for Tikhonov regularization with sparsity constraints. *arXiv:0801.1774v1, submitted*, 2008.

[21] A. Neubauer. On converse and saturation results for Tikhonov regularization of linear ill-posed problems. *SIAM J. Numer. Anal.*, 34:517–527, 1997.

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

[23] R. Ramlau and G. Teschke. A Tikhonov-based projection iteration for nonlinear ill-posed problems with sparsity constraints. *Numer. Math.*, 104(2):177–203, 2006.

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

[25] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational Methods in Imaging*. Springer, 2008. To appear.

[26] A. N. Tikhonov. On the regularization of ill-posed problems. *Dokl. Akad. Nauk SSSR*, 153:49–52, 1963.

[27] A. N. Tikhonov. On the solution of ill-posed problems and the method of regularization. *Dokl. Akad. Nauk SSSR*, 151:501–504, 1963.

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