Electronic Transactions on Numerical Analysis.
Volume 37, pp. 87-104, 2010.

ISSN 1068-9613.

ETNA
Kent State University
http://etna.math.kent.edu

CONVERGENCE RATES FOR REGULARIZATION WITH SPARSITY CONSTRAINTS*

RONNY RAMLAU†‡ AND ELENA RESMERITA†

Abstract. Tikhonov regularization with p-powers of the weighted ℓ_p norms as penalties, with $p \in (1, 2)$, have been employed recently in reconstruction of sparse solutions of ill-posed inverse problems. This paper shows convergence rates for such a regularization with respect to the norm of the weighted spaces by assuming that the solutions satisfy a certain smoothness (source) condition. The meaning of the latter is analyzed in some detail. Moreover, converse results are established: Linear convergence rates for the residual, together with convergence of the approximations to the solution, can be achieved only if the solution satisfies a source condition. Further insights for the particular case of a convolution equation are provided by analyzing the equation both theoretically and numerically.

Key words. ill-posed problem, regularization, Bregman distance, sparsity

AMS subject classifications. 47A52, 65J20

1. Introduction. In this paper, we consider linear ill-posed operator equations

$$\begin{aligned} &\tilde{A}\tilde{u} = \tilde{y} \\ &\tilde{A} \,:\, X_{p,\omega} \to L_2(\Omega)\,. \end{aligned} \tag{1.1}$$

Here, $X_{p,\omega}$ denotes a Banach space which is a subspace of $L_2(\Omega)$, with parameters $p \in (1, 2)$ and $\omega = \{\omega_\lambda\}_{\lambda\in\Lambda}$, where Ω is a bounded open subset of $\mathbb{R}^d$, with $d \geq 1$, and Λ is a set of (possibly tuples of) integer indices. Although one could employ more general separable Hilbert spaces than $L_2(\Omega)$, we consider here the Lebesgue space case for simplicity.

We are in particular interested in the reconstruction of solutions of (1.1) that admit a sparse structure with respect to a given basis in the Banach space $X_{p,\omega}$, that is, only a finite number of the solution coefficients do not vanish. In these cases it is desirable to choose a regularization method that also promotes a sparse reconstruction. For instance, suitable choices for the spaces $X_{p,\omega}$ are the Besov spaces $B^s_{p,p}$ with $p \in (1, 2)$ in case of a sufficiently smooth wavelet basis and properly chosen weights; see, e.g., [6, 19] for detailed discussions.

Instead of solving the above equation in a function space setting, we will transform it into a sequential setting. More precisely, we will work with weighted ℓ_p spaces, where $p \in (1, 2)$. We will consider weights that are bounded away from zero, which ensures that the spaces ℓ_p and the weighted ℓ_p spaces are isomorphic.

Convergence of Tikhonov type regularization methods with Besov norm constraints (which can be transformed into a weighted ℓ_p constraint) has been shown with respect to the ℓ_2 norm in [9]. With respect to the weighted ℓ_p strong topologies with $p \in [1, 2]$, convergence has been established in [20, 19, 12]. Error estimates for regularization have been achieved recently via Bregman distances associated to the penalties; see, for instance, [4, 21, 22, 13, 5]. Note that error estimates for regularization with weighted ℓ_p norms have been obtained in [14] with respect to the ℓ_2 norm. In parallel with our work, interesting quantitative results have been shown also in the Hilbert space norm by [12] in the case that the solution is known to be sparse; see a related discussion at the end of the next section. Our study focuses on

*Received March 17, 2009. Accepted for publication October 9, 2009. Published online April 7, 2010. Recommended by D. Calvetti.

†Institute of Industrial Mathematics, Johannes Kepler University, Altenbergerstrasse 69, 4040 Linz, Austria ({ronny.ramlau,elena.resmerita}@jku.at).

‡Johann Radon Institute for Computational and Applied Mathematics (RICAM), Austrian Academy of Sciences, Altenbergerstrasse 69, 4040 Linz, Austria.

ETNA
Kent State University
http://etna.math.kent.edu

convergence rates in terms of weighted ℓ_p space norms which are stronger than the ℓ_2 norm when $p \in (1,2)$, with emphasis on some kind of necessary and sufficient conditions. We dwell on the realization of the error estimates, rather than on the results themselves, which follow via convexity arguments characterizing the specific Banach spaces we work with.

Due to the useful topological properties of the ℓ_p spaces, $p \in (1,2)$, which transfer to the weighted ℓ_p spaces, error estimates can be established also in terms of the norms of these Banach spaces.

Recall that error estimates are usually obtained under smoothness assumptions on the solutions. For instance, classical assumptions of this kind in case of quadratic Tikhonov stabilization (i.e., $f = \|\cdot\|^2$) in Hilbert spaces are as follows: A solution $\bar{x}$ of the equation $Ax = y$ to be solved is in the range of the operator $(A^*A)^\nu$, $\nu > 0$, where A^* is the adjoint operator. The role of spectral theory is known to be essential in that Hilbert space context. We limit our study in the non-Hilbert space framework to the smoothness assumptions already analyzed in [4, 22], where general convex penalties f were considered. More precisely, we will assume that the derivative of the penalty f at the solution $\bar{x}$ belongs either to the range of the Banach space adjoint operator A^* or, in particular, to the range of the operator A^*A. By focusing on this specific sparsity framework, we will obtain the convergence rate $O(\delta^{1/2})$ in case the first source condition holds and $O(\delta^{p/(p+1)})$ under the second condition. We will also show that linear convergence rates for the residual, together with convergence of the approximations to the solution, can be achieved only if the solution satisfies a source condition. An interpretation of the basic source condition will be discussed in some detail, by allowing the domain of the operator to be a larger weighted Lebesgue space than the domain of the penalty function. We will consider a convolution problem and present necessary and sufficient conditions for a source condition, pointing out the case of sparse solutions which satisfy a source condition with sparse source elements. The numerical results on the reconstruction of a function from its noisy convolution data confirm the derived convergence rates. Our study is done for linear operators, although it can be extended to nonlinear ones, as is briefly discussed.

The paper is organized as follows. Section 2 states the notation and general assumptions. The error estimates and the a priori convergence rates are shown in Section 3, while Section 4 presents some type of converse results for those rates. Section 5 consists of a discussion of the basic source condition. Possible extensions to nonlinear operator equations are considered in Section 6. A convolution problem is analyzed both theoretically and numerically in Section 7.

2. Notation and assumptions.

By choosing a suitable orthonormal basis $\{\Phi_\lambda\}_{\lambda\in\Lambda}$ of the space $L_2(\Omega)$, both $\tilde{u}$ and $\tilde{A}\tilde{u}$ can be expressed with respect to Φ_λ. Thus,

$$\tilde{A}\tilde{u} = \sum_{\lambda'}\sum_{\lambda} \langle \tilde{u}, \Phi_\lambda\rangle \langle \tilde{A}\Phi_\lambda, \Phi_{\lambda'}\rangle \, \Phi_{\lambda'} \ . \tag{2.1}$$

Defining the infinite-dimensional matrix A and vectors u, y by

$$A = (\langle \tilde{A}\Phi_\lambda, \Phi_{\lambda'}\rangle)_{\lambda,\lambda'\in\Lambda},\ u = (\langle \tilde{u}, \Phi_\lambda\rangle)_{\lambda\in\Lambda},\ y = (\langle \tilde{y}, \Phi_\lambda\rangle)_{\lambda\in\Lambda}\ , \tag{2.2}$$

equation (1.1) can be reformulated as an (infinite) matrix-vector multiplication

$$Au = y. \tag{2.3}$$

Now let us specify the spaces $X_{p,\omega}$. For a given orthonormal basis $\{\Phi_\lambda\}_{\lambda\in\Lambda}$ and positive weights $\omega = \{\omega_\lambda\}_{\lambda\in\Lambda}$, we define

$$\tilde{u} \in X_{p,\omega} \Longleftrightarrow \sum_{\lambda} \omega_\lambda |\langle \tilde{u}, \Phi_\lambda\rangle|^p < \infty\ ,$$

ETNA
Kent State University
http://etna.math.kent.edu

i.e., $u = \{\langle \tilde{u}, \Phi_\lambda \rangle\}_{\lambda\in\Lambda}$ belongs to the weighted sequence space $\ell_{p,\omega}$. From now on, let

$$u_\lambda = \langle \tilde{u}, \Phi_\lambda \rangle,$$

$$\ell_{p,\omega} = \left\{ u = \{u_\lambda\}_{\lambda\in\Lambda} : \|u\|_{p,\omega} = \left(\sum_\lambda \omega_\lambda |u_\lambda|^p \right)^{\frac{1}{p}} < \infty \right\}.$$

Since $\ell_p \subset \ell_q$ with $\|u\|_q \le \|u\|_p$ for $p \le q$, one also has $\ell_{p,\omega} \subset \ell_{q,\omega'}$ for $p \le q$ and $\omega' \le \omega$. In particular, if the sequence of weights is positive and bounded from below, i.e., $0 < \rho \le \omega_\lambda$ for some $\rho > 0$, then $\ell_{p,\omega} \subset \ell_2$ for $p \le 2$.

Using the above discretization, we now consider the operator equation

$$\begin{aligned} &Au = y \\ &A : \ell_{p,\omega} \longrightarrow \ell_2, \end{aligned} \tag{2.4}$$

where A is a linear and bounded operator.

We are interested in investigating convergence rates for Tikhonov regularization with sparsity constraints, where the approximation of the solution is obtained as a solution of the problem

$$\min \left\{ \frac{1}{2}\|Au - y^\delta\|^2 + \alpha f(u) \right\}, \tag{2.5}$$

with regularization parameter $\alpha > 0$ and penalty

$$f(u) = \|u\|_{p,\omega}^p = \sum_\lambda \omega_\lambda |u_\lambda|^p. \tag{2.6}$$

Throughout this paper, we assume that $p \in (1, 2)$ and the weight ω is bounded away from zero, i.e., there is $\rho > 0$ such that

$$0 < \rho \le \omega_\lambda\ .$$

Moreover, we assume that (2.4) has solutions and that the available noisy data y^δ satisfy

$$\|y^\delta - y\| \le \delta \tag{2.7}$$

for some noise level $\delta > 0$.

Note that the function f is strictly convex since the p powers of the norms are strictly convex. In addition, the function f is Fréchet differentiable.

Denote by $\bar{u}$ the unique solution of the equation that minimizes the penalty functional f and by $D_f(z, x)$ the Bregman distance defined with respect to f, that is,

$$D_f(z, x) = f(z) - f(x) - \langle f'(x), z - x \rangle\ .$$

The reader is referred to [3], for more information on Bregman distances.

3. Error estimation. As mentioned in the introduction, we would like to estimate the distance between the minimizers u_α^δ of (2.5) and the solution $\bar{u}$ of equation (2.4) which minimizes the penalty f. Let A^* denote the Banach space adjoint operator which maps ℓ_2 into the dual of $\ell_{p,\omega}$, i.e.,

$$A^* : \ell_2 \to (\ell_{p,\omega})^*$$

ETNA
Kent State University
http://etna.math.kent.edu

with

$$\langle A^* g, u\rangle = \langle g, Au\rangle$$

for $u \in \ell_{p,\omega}$, $g \in \ell_2$.

PROPOSITION 3.1. *The dual space of $\ell_{p,\omega}$ is given by*

$$(\ell_{p,\omega})^* = \begin{cases} \ell_{\infty,\omega^{-1}} & \text{for } p = 1, \\ \ell_{q,\omega^{-q/p}} & \text{for } p > 1, \end{cases}$$

where q fulfills the equation $1/q + 1/p = 1$.

Proof. See [17, Corollary 13.14]. □

Note that $\langle v, u\rangle = \sum_\lambda u_\lambda v_\lambda$, whenever $u \in \ell_{p,\omega}$ and $v \in (\ell_{p,\omega})^*$. In order to obtain convergence rates we will need the source conditions (SC) and (SC I), given by

(SC) $f'(\bar{u}) = A^* v$, for some $v \in \ell_2$.

(SC I) $f'(\bar{u}) = A^* A\hat{v}$, for some $\hat{v} \in \ell_{p,\omega}$.

The error estimates established by [4, 21, 22] for regularization with a general convex functional read in our framework as follows: If (SC) is fulfilled, then

$$D_f(u_\alpha^\delta, \bar{u}) \le \frac{\alpha\|v\|^2}{2} + \frac{\delta^2}{2\alpha}, \quad \|Au_\alpha^\delta - A\bar{u}\| \le \alpha\|v\| + \delta. \tag{3.1}$$

If (SC I) is fulfilled, then

$$D_f(u_\alpha^\delta, \bar{u}) \le D_f(\bar{u} - \alpha v, \bar{u}) + \frac{\delta^2}{2\alpha}, \tag{3.2}$$

$$\|Au_\alpha^\delta - A\bar{u}\| \le \alpha\|Av\| + \sqrt{2\alpha}[D_f(\bar{u} - \alpha v, \bar{u})]^{1/2} + c\delta,$$

where c is a positive number.

We are further interested in obtaining error estimates with respect to the norm of the $\ell_{p,\omega}$ spaces with $p \in (1, 2)$. For $u \in \ell_{p,\omega}$ consider the sequence $\tilde{u} = \left\{\omega_\lambda^{1/p} \cdot u_\lambda\right\}_{\lambda\in\Lambda}$. Then

$$\|\tilde{u}\|_p^p = \sum_\lambda |\omega_\lambda^{1/p} u_\lambda|^p = \|u\|_{p,\omega}^p < \infty\ ,$$

i.e., $\tilde{u} \in \ell_p$. By defining $\tilde{f}(\tilde{u}) = \|\tilde{u}\|_p^p$, one gets

$$f(u) = \tilde{f}(\tilde{u}). \tag{3.3}$$

Moreover, for any $h \in \ell_{p,\omega}$ and for $\tilde{h} = \left\{\omega_\lambda^{1/p} \cdot h_\lambda\right\}_{\lambda\in\Lambda}$, one has

$$\langle \tilde{f}'(\tilde{u}), \tilde{h}\rangle = \langle f'(u), h\rangle.$$

We point out below several inequalities in ℓ_p; see, e.g., [23, Corollary 2] and [2, Lemma 1.4.8 and proof of Cor. 3.5.9(ii)], respectively.

PROPOSITION 3.2. *If* $p \in (1, 2]$, *then one has for all* $x, z \in \ell_p$

$$\left\langle \tilde{f}'(z) - \tilde{f}'(x), z - x \right\rangle \leq \tilde{c}_p \|z - x\|_p^p, \tag{3.4}$$

$$D_{\tilde{f}}(z, x) \geq c_p \|z - x\|_p^2, \quad \text{for } \|z - x\|_p \text{ small enough}, \tag{3.5}$$

where $\tilde{c}_p$ *and* $c_p = c_p(x)$ *are some positive numbers.*

The symmetric Bregman distance has the following expression,

$$D_{\tilde{f}}(z, x) + D_{\tilde{f}}(x, z) = \left\langle \tilde{f}'(z) - \tilde{f}'(x), z - x \right\rangle.$$

Thus, inequality (3.4) can be rewritten as

$$D_{\tilde{f}}(z, x) + D_{\tilde{f}}(x, z) \leq \tilde{c}_p \|z - x\|_p^p. \tag{3.6}$$

For our purposes, these results have to be extended to the $\ell_{p,\omega}$ spaces.

LEMMA 3.3. *If* $p \in (1, 2)$ *and* f *is defined by* (2.6), *then the following inequality holds for any* $u, v \in \ell_{p,\omega}$ *and for some positive numbers* $\tilde{c}_p$ *and* $c_p = c_p(u)$,

$$D_f(v, u) \leq \tilde{c}_p \|v - u\|_{p,\omega}^p, \tag{3.7}$$

$$D_f(v, u) \geq c_p \|v - u\|_{p,\omega}^2, \quad \text{for } \|v - u\|_{p,\omega} \text{ small enough}. \tag{3.8}$$

Proof. We can use the equality shown previously $\langle f'(u), h\rangle = \langle \tilde{f}'(\tilde{u}), \tilde{h}\rangle$, for all $u, h \in \ell_{p,\omega}$ and corresponding $\tilde{u}, \tilde{h} \in \ell_p$. Thus, one obtains

$$D_f(v, u) = f(v) - f(u) - \langle f'(u), v - u\rangle = \tilde{f}(\tilde{v}) - \tilde{f}(\tilde{u}) - \langle \tilde{f}'(\tilde{u}), \tilde{v} - \tilde{u}\rangle = D_{\tilde{f}}(\tilde{v}, \tilde{u}).$$

By using further (3.5), (3.6), and (3.3), inequalities (3.7) and (3.8) follow immediately. ∎

Based on the previous result, we can now state the error estimates for the above regularization method in terms of the norm of $\ell_{p,\omega}$.

PROPOSITION 3.4. *Assume that noisy data* y^δ *fulfilling* $\|y - y^\delta\| \leq \delta$ *are given.*

i) If $f'(\bar{u}) = A^* v$ *for some* $v \in \ell_2$, *then the following error estimates hold for the minimizer of* (2.5), *when* $\|u_\alpha^\delta - \bar{u}\|_{p,\omega}$ *is small enough:*

$$\|u_\alpha^\delta - \bar{u}\|_{p,\omega} \leq \frac{1}{c_p^{1/2}} \left(\frac{\sqrt{\alpha}\|v\|}{\sqrt{2}} + \frac{\delta}{\sqrt{2\alpha}} \right), \quad \|A u_\alpha^\delta - A\bar{u}\| \leq \alpha \|v\| + \delta.$$

ii) If $f'(\bar{u}) = A^* A \hat{v}$ *for some* $\hat{v} \in \ell_{p,\omega}$, *then the following error estimates hold for the minimizer of* (2.5), *when* $\|u_\alpha^\delta - \bar{u}\|_{p,\omega}$ *is small enough:*

$$\|u_\alpha^\delta - \bar{u}\|_{p,\omega} \leq m_p \alpha^{\frac{p}{2}} + \frac{\delta}{\sqrt{2 c_p \alpha}}, \quad \|A u_\alpha^\delta - A\bar{u}\| \leq \alpha \|A\hat{v}\| + \delta,$$

where $m_p = \frac{\tilde{c}_p}{c_p} \|\hat{v}\|_{p,\omega}^{\frac{p}{2}}$.

Proof. i) Follows immediately from (3.1) and (3.8).

ETNA
Kent State University
http://etna.math.kent.edu

ii) Inequalities (3.2), (3.8), and (3.7) imply

$$\|u_\alpha^\delta - \bar{u}\|_{p,\omega}^2 \le \frac{1}{c_p} D_f(\bar{u} - \alpha\hat{v}, \bar{u}) + \frac{\delta^2}{2c_p\alpha}$$
$$\le \frac{\tilde{c}_p}{c_p}\alpha^p \|\hat{v}\|_{p,\omega}^p + \frac{\delta^2}{2c_p\alpha},$$

which, together with the inequality $\sqrt{a^2+b^2} \le a+b$ for $a, b > 0$ yields the result. ☐

COROLLARY 3.5. *i') If the assumptions of Proposition 3.4, part i), hold and $\alpha \sim \delta$, then*

$$\|u_\alpha^\delta - \bar{u}\|_{p,\omega} = O(\delta^{\frac{1}{2}}), \quad \|Au_\alpha^\delta - y\| = O(\delta),$$

for δ small enough.

ii') If the assumptions of Proposition 3.4, part ii), hold and $\alpha \sim \delta^{\frac{2}{p+1}}$, then

$$\|u_\alpha^\delta - \bar{u}\|_{p,\omega} = O(\delta^{\frac{p}{p+1}}), \quad \|Au_\alpha^\delta - y\| = O(\delta),$$

for δ small enough.

The recent work [12] shows the convergence rate $O(\delta^{\frac{1}{p}})$ for $p \in [1, 2)$ (thus, up to $O(\delta)$) with respect to the ℓ_2 norm of $u_\alpha^\delta - \bar{u}$ (which is weaker than the $\ell_{p,\omega}$ norm for $p < 2$), in the case that $\bar{u}$ is sparse and (SC) holds. These rates are already higher, when $p < 1.5$, than the "superior limit" of $O(\delta^{\frac{2}{3}})$ established for quadratic regularization. This also shows that the assumption of sparsity is a very strong source condition. It would be interesting to find out whether these rates could be further improved if the stronger assumption (SC I) is fulfilled. We remark that the numerical experiment we perform under the assumption (SC I) and which is described at the end of the last section shows that better rates than $O(\delta^{\frac{p}{p+1}})$ for $u_\alpha^\delta - \bar{u}$ in the stronger $\ell_{p,\omega}$ norm seem to be achieved—that is, at least $O(\delta^{\frac{1}{p}})$—even if the solution is not sparse.

4. Converse results. Next we prove some kinds of converse results regarding the first type of source condition. To this end, one needs even less than strong convergence of the approximants u_α^δ to $\bar{u}$. More precisely, one can show that a linear convergence rate for the residual and (even weak) convergence of the approximants to the solution $\bar{u}$ ensure that $f'(\bar{u})$ is in the range of the adjoint operator, no matter how fast the approximants converge to the solution. This interesting fact is due, as we will see below, to a special property of the duality mapping J_p in ℓ_p (the derivative of the p-th power of the ℓ_p norm).

We first deal with the noiseless data case, where u_α is the minimizer of (2.5) corresponding to exact data y.

PROPOSITION 4.1. *If $\|Au_\alpha - y\| = O(\alpha)$ and u_α converges to $\bar{u}$, as $\alpha \to 0$, in the $\ell_{p,\omega}$ weak topology, then $f'(\bar{u})$ belongs to the range of the adjoint operator A^*.*

Proof. We use a technique from [10]. Let $\{\alpha_n\}$ be a positive sequence which converges to zero as $n \to \infty$. Let

$$v_n = \frac{1}{\alpha_n}(y - Au_{\alpha_n}).$$

This sequence is bounded in ℓ_2, so there exists a subsequence $\{v_k\}$ which converges weakly to some $v \in \ell_2$, as $k \to \infty$. Since the Banach space adjoint operator A^* is linear and bounded, it maps weakly convergent sequences into weakly convergent sequences; see, e.g., [1, Propositions 2.8-2.9, p. 37]. It follows that A^*v_k converges weakly to A^*v in $(\ell_{p,\omega})^*$, as $k \to \infty$.

ETNA
Kent State University
http://etna.math.kent.edu

On one hand, the first order optimality condition for the minimization problem (2.5) is

$$A^*(Au_\alpha - y) + \alpha f'(u_\alpha) = 0,$$

which implies that $A^* v_k = f'(u_k)$ with $u_k = u_{\alpha_k}$. Consequently, one obtains that $f'(u_k)$ converges weakly to $A^* v$ in $(\ell_{p,\omega})^*$ as $k \to \infty$. On the other hand, u_k converges weakly to $\bar{u}$ by the above assumption. Since any duality mapping on ℓ_p is weakly sequentially continuous (see [7], p. 73), this property is inherited also by the derivative f' on $\ell_{p,\omega}$ with respect to the corresponding weak topologies. Thus, one gets that $f'(u_k) = A^* v_k$ converges also weakly to $f'(\bar{u})$ as $k \to \infty$. Therefore, $f'(\bar{u}) = A^* v$. ∎

Consider now the case of noisy data y^δ which satisfy (2.7).

PROPOSITION 4.2. *If $\|y - y^\delta\| \le \delta$, the rate $\|Au_\alpha^\delta - y\| = O(\delta)$ holds and u_α^δ converges to $\bar{u}$ in the $\ell_{p,\omega}$ weak topology as $\delta \to 0$ and $\alpha \sim \delta$, then $f'(\bar{u})$ belongs to the range of the adjoint operator A^*.*

Proof. The first order optimality condition for the optimization problem (2.5) now reads as

$$A^*(Au_\alpha^\delta - y^\delta) + \alpha f'(u_\alpha^\delta) = 0.$$

One can proceed as in the previous proposition, by using the additional assumption (2.7). ∎

5. Interpretation of the source condition (SC). The aim of this section is to derive conditions on sequences fulfilling (SC). We will use the notation,

$$\omega^t = \{\omega_\lambda^t\}_{\lambda\in\Lambda}, \quad t \in \mathbb{R}.$$

Now let us assume that

$$A : \ell_{p',\omega'} \to \ell_2 \,, \tag{5.1}$$

and as penalty we take the functional f given by (2.6). Note that p, p' and ω, ω' are allowed to be different, respectively. This, however, makes sense only if the penalty enforces the solution to belong to a smaller space than $D(A) = \ell_{p',\omega'}$, i.e., if $\ell_{p,\omega} \subset \ell_{p',\omega'}$. This is the case, e.g., for

$$p \le p' \,, \qquad \omega' \le \omega \,. \tag{5.2}$$

We will assume that the weights ω' are also bounded away from zero, $\omega'_\lambda \ge \rho > 0$. In the sequel, the dual exponents to the given p, p' will be denoted by q, q'. Consider first the case $p, p' > 1$.

PROPOSITION 5.1. *Let $p, p' > 1$, the operator A and the penalty f be given by* (5.1), (2.6), *and assume that* (5.2) *holds. Then a solution $\bar{u}$ of $Au = y$ fulfilling $A^* v = f'(\bar{u})$ satisfies*

$$\bar{u} \in \ell_{(p-1)q',(\omega')^{-q'/p'}\cdot\omega^{q'}} \,. \tag{5.3}$$

Proof. For $p, p' > 1$, the source condition reads

$$A^* v = \{p\omega_\lambda \operatorname{sgn}(\bar{u}_\lambda)|\bar{u}_\lambda|^{p-1}\}_{\lambda\in\Lambda}.$$

Since $A^* : \ell_2 \to \ell_{q',\omega'^{-q'/p'}}$ we have $A^* v \in \ell_{q',\omega'^{-q'/p'}}$ and thus the condition

$$\sum_\lambda {\omega'}_\lambda^{-q'/p'} \omega_\lambda^{q'} |\bar{u}_\lambda|^{(p-1)q'} < \infty$$

ETNA
Kent State University
http://etna.math.kent.edu

has to be fulfilled, i.e., $\bar{u} \in \ell_{(p-1)q',(\omega')^{-q'/p'}\cdot\omega^{q'}}$. $\square$

The previous result states only a necessary condition; e.g., for the special case $p = p',\ q = q',\ \omega_\lambda = \omega'_\lambda$ one gets no additional information:

REMARK 5.2. For the case $p = p',\ q = q', \omega_\lambda = \omega'_\lambda$, condition (5.3) reduces to

$$\bar{u} \in \ell_{p,\omega} = D(A)\ .$$

In order to characterize the smoothness condition in terms of spaces of sequences, we relate the spaces to $R(A^*)$.

PROPOSITION 5.3. *Let $p, p' > 1$, the operator A and the penalty f be given by* (5.1), (2.6), *and assume that* (5.2) *holds. Moreover, assume that*

$$R(A^*) = \ell_{\tilde{q},\tilde{\omega}^{-\tilde{q}/\tilde{p}}} \subset \ell_{q',\omega'^{-q'/p'}}$$

for some $\tilde{p},\ \tilde{q} > 1$. Then each sequence

$$\bar{u} \in \ell_{(p-1)\tilde{q},\tilde{\omega}^{-\tilde{q}/\tilde{p}}\cdot\omega^{\tilde{q}}} \tag{5.4}$$

fulfills the smoothness condition (SC).

Proof. Based on (5.4), one can verify that

$$\left\{p\omega_\lambda \operatorname{sgn}(\bar{u}_\lambda)|\bar{u}_\lambda|^{p-1}\right\}_{\lambda\in\Lambda} \in \ell_{\tilde{q},\tilde{\omega}^{-\tilde{q}/\tilde{p}}},$$

which implies that $\bar{u}$ satisfies (SC) due to the equality $R(A^*) = \ell_{\tilde{q},\tilde{\omega}^{-\tilde{q}/\tilde{p}}}$. $\square$

Let us now consider the (special) case $p' > 1,\ p = 1$, i.e., the case when the penalty is neither differentiable nor strictly convex. In this situation, a solution $\bar{u}$ which minimizes the penalty might not be unique. Moreover, the source condition reads as $A^*v = \{\omega_\lambda \operatorname{sgn}(\bar{u}_\lambda)\}_{\lambda\in\Lambda}$, where $\operatorname{sgn}(\bar{u}_\lambda)$ equals 1 if $\bar{u}_\lambda > 0$, equals -1 if $\bar{u}_\lambda < 0$, and belongs to $[-1, 1]$ otherwise; see, e.g., [5].

PROPOSITION 5.4. *Let $p' > 1,\ p = 1,\ \omega' \le \omega$, and let the operator A and the penalty f be given by* (5.1), (2.6). *Then the source condition $A^*v = \{\omega_\lambda \operatorname{sgn}(\bar{u}_\lambda)\}_{\lambda\in\Lambda}$ only can be fulfilled if the solution $\bar{u}$ is sparse.*

Proof. Let $\Lambda_{\bar{u}} = \{\lambda : |\bar{u}_\lambda| > 0\}$. For $p' > 1$ one has $R(A^*) \subset \ell_{q',\omega'^{-q'/p'}}$, and from the source condition follows the condition

$$\sum_\lambda {\omega'_\lambda}^{-q'/p'} \cdot \omega_\lambda^{q'} |\operatorname{sgn}(\bar{u}_\lambda)| < \infty\ .$$

Since $\omega' \le \omega$ and $0 < \rho \le \omega'_\lambda$, we conclude further that

$$\begin{aligned}
\sum_\lambda {\omega'_\lambda}^{-q'/p'} \cdot \omega_\lambda^{q'} |\operatorname{sgn}(\bar{u}_\lambda)| &\ge \sum_{\lambda\in\Lambda_{\bar{u}}} {\omega'_\lambda}^{-q'/p'} \cdot {\omega'_\lambda}^{q'} \\
&= \sum_{\lambda\in\Lambda_{\bar{u}}} {\omega'_\lambda}^{\frac{q'}{p'}(p'-1)} \\
&\ge \sum_{\lambda\in\Lambda_{\bar{u}}} \rho^{\frac{q'}{p'}(p'-1)}\ ,
\end{aligned} \tag{5.5}$$

and the sum in (5.5) converges only if $\Lambda_{\bar{u}}$ is finite, i.e., if the solution $\bar{u}$ is sparse. $\square$

We further note that for $p = p' = 1,\ \omega' = \omega$, the source condition reads

$$A^*v = \{\omega_\lambda \operatorname{sgn}(\bar{u}_\lambda)\}_{\lambda\in\Lambda} \in \ell_{\infty,\omega^{-1}}\ ,$$

ETNA
Kent State University
http://etna.math.kent.edu

and, similarly as in Remark 5.2, no further conclusions can be drawn.

The above derived conditions on sequences fulfilling a source condition (SC) mean in principle that the sequence itself has to converge to zero fast enough. They can also be interpreted in terms of smoothness of an associated function: If the function system $\{\Phi_\lambda\}_{\lambda\in\Lambda}$ in (2.1), (2.2) is formed by a wavelet basis, then the norm of a function in the Besov space $B^s_{p,p}$ coincides with a weighted ℓ_p norm of its wavelet coefficients and properly chosen weights [8]. In this sense, the source condition requires the solution to belong to a certain Besov space. The assumption on $R(A^*)$ in Proposition 5.3 then means that the range of the dual operator equals a Besov space. Similar assumptions were used for the analysis of convergence rates for Tikhonov regularization in Hilbert scales; see [18, 16, 15].

6. Nonlinear operators. A similar analysis as in the linear case can be carried out for nonlinear operator equations $F(u) = y$ under several conditions on the operator F, which allow reducing the study to the linear operator case. For instance, differentiability of F on a ball around the solution, the source conditions $f'(\bar{u}) = F'(\bar{u})^* v$ and $f'(\bar{u}) = F'(\bar{u})^* F'(\bar{u})\bar{v}$ with $v \in \ell_2$, $\bar{v} \in \ell_{p,\omega}$, smallness conditions on the source elements v and $\bar{v}$, and the inequality

$$\|F(u) - F(\bar{u}) - F'(\bar{u})(u - \bar{u})\| \le \eta(u, \bar{u}),$$

for any u sufficiently close to $\bar{u}$, guarantee that estimates similar to those in Corollary 3.5 hold also when regularizing the ill-posed problem $F(u) = y$. A couple of choices that have been used so far for $\eta(u, \bar{u})$ are as follows (see, e.g., [11, 4, 22]):

$$\eta(u, \bar{u}) = c\|F(u) - F(\bar{u})\|, \tag{6.1}$$

$$\eta(u, \bar{u}) = cD_f(u, \bar{u}),$$

for some number $c > 0$. Since working with (6.1) is quite restrictive regarding the nonlinearity of the operator F (see, for a discussion, [11, Chapter 11]) and since f is strictly convex and ensures that $D_f(u, \bar{u}) \neq 0$ if $u \neq \bar{u}$, we believe that the second condition represents a better choice for the analysis.

We note again that the most recent study of sparse regularization for nonlinear equations, including convergence and error estimates for the method with respect to the ℓ_2 norm is [12], as far as we know.

7. Reconstruction of a function from its convolution data. This section presents a numerical example which illustrates our analytical results on the convergence rates. We intend to show that the convergence rates established above can be obtained if appropriate source conditions are fulfilled. Two difficulties arise: The numerics should be accurate enough so that the numerical errors do not dominate the reconstruction error of the regularization method as $\delta \to 0$ and one should be able to construct functions that fulfill a source condition exactly. Of course, the operator equation under consideration should not be trivial. All of these requirements can be met by choosing the convolution operator,

$$y(\tau) = (Au)(\tau) = \int_{-\pi}^{\pi} r(\tau - t)u(t)\,dt =: (r * u)(\tau), \tag{7.1}$$

where u, r and Au are 2π-periodic functions belonging to $L_2((-\pi, \pi))$.

ETNA
Kent State University
http://etna.math.kent.edu

7.1. The system matrix and the interpretation of the source condition. In (7.1), the operator A is defined between function spaces. Therefore, we have to transform the operator A to a matrix, such that the application of the operator can be expressed as a matrix-vector multiplication; cf. (2.1)-(2.3). The appropriate discretization of this linear operator is apparently given by the Fourier coefficients of the periodic functions. It is well known that a periodic function on $[-\pi,\pi]$ can be either expressed via the orthonormal bases formed by

$$\left\{\frac{1}{\sqrt{2\pi}}e^{ikt}\right\}_{k\in\mathbb{Z}} \text{ or } \left\{\frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}}\cos(kt), \frac{1}{\sqrt{\pi}}\sin(kt)\right\}_{k\in\mathbb{N}} .$$

It turns out that, by the convolution theorem, a discretization via the exponential basis leads to a diagonal system matrix. However, by using the exponential basis one has to work with complex valued matrices and vectors, which is not covered by our theory. Therefore, we have to use the trigonometric basis, which leads to real valued matrices and vectors. That is,

$$u(t) = a_0 + \sum_{k\in\mathbb{N}} a_k\cos(kt) + b_k\sin(kt)$$

with coefficients

$$a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} u(t)\,dt$$

$$a_k = \frac{1}{\pi}\int_{-\pi}^{\pi} u(t)\cos(kt)\,dt$$

$$b_k = \frac{1}{\pi}\int_{-\pi}^{\pi} u(t)\sin(kt)\,dt .$$

In the following, the Fourier coefficients of a function u will be collected in the vector $\mathbf{u} = [a_0^u, a_1^u, b_1^u, \cdots, a_k^u, b_k^u, \cdots]$. Using the Fourier convolution theorem for the exponential basis and transformation formulas between the exponential and trigonometric bases, the Fourier coefficients $\mathbf{y}$ of $y = Au$ can be computed as

$$\mathbf{y} = \mathbf{A}\mathbf{u},$$

with $\mathbf{A}$ given by

$$\mathbf{A} = \pi\begin{bmatrix} 2\cdot a_0^r & 0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & a_1^r & -b_1^r & 0 & 0 & \cdots & \cdots & \cdots \\ 0 & b_1^r & a_1^r & 0 & 0 & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & a_2^r & -b_2^r & 0 & 0 & \cdots \\ 0 & 0 & 0 & b_2^r & a_2^r & 0 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix},$$

where $a_0^r, a_k^r,\ b_k^r$ denote the Fourier coefficients of the kernel function r.

For given $\mathbf{u}$, the weighted ℓ_p norm is defined by

$$\|\mathbf{u}\|_{p,\omega}^p := \omega_0|a_0^u|^p + \sum_{k=1}^{\infty}(\omega_k^a|a_k^u|^p + \omega_k^b|b_k^u|^p) .$$

ETNA
Kent State University
http://etna.math.kent.edu

For $p = 2$ and $\omega_k = \pi$, the norm so defined coincides with the usual norm on $L_2(-\pi, \pi)$. In order to get convergence rates, the solution needs to fulfill a source condition involving the (Banach) adjoint of A. We have

$$\mathbf{A}^* = \mathbf{A}^T \,,$$

i.e., the adjoint operator is the transposed matrix. We get the following result.

COROLLARY 7.1. *A sequence $\bar{u} \in \ell_2$ fulfills the source condition (SC) for $f(\bar{\mathbf{u}}) = \|\bar{\mathbf{u}}\|_{p,\omega}^p$ if there exists a sequence $\mathbf{v} = \{a_0^v, a_1^v, b_1^v, \cdots\} \in \ell_2$ with*

$$\begin{bmatrix} a_k^{sc} \\ b_k^{sc} \end{bmatrix} = \begin{bmatrix} p\omega_k^a \operatorname{sgn}(a_k^{\bar{u}})|a_k^{\bar{u}}|^{p-1} \\ p\omega_k^b \operatorname{sgn}(a_k^{\bar{u}})|b_k^{\bar{u}}|^{p-1} \end{bmatrix} = \mathbf{A}_k^* \begin{bmatrix} a_k^v \\ b_k^v \end{bmatrix} \tag{7.2}$$

for each $k > 0$ and, for $k = 0$,

$$p\omega_0^a \operatorname{sgn}(a_0^{\bar{u}})|a_0^{\bar{u}}|^{p-1} = 2 \cdot \pi \cdot a_0^r \cdot a_0^v$$

holds, where $\mathbf{A}_k^*$ is given by

$$\mathbf{A}_k^* = \pi \begin{bmatrix} a_k^r & b_k^r \\ -b_k^r & a_k^r \end{bmatrix}. \tag{7.3}$$

Proof. The proof is straightforward and follows from $\mathbf{A}^* = \mathbf{A}^T$, the fact that $\mathbf{A}$ can be formed as a "diagonal" matrix, $\mathbf{A} = \operatorname{diag}[2a_0^r, \mathbf{A}_1, \mathbf{A}_2, \cdots]$, and the definition of f. ☐

Thus, for a given $\mathbf{v} \in \ell_2$, one can compute $\mathbf{A}^*\mathbf{v}$ and therefore construct $\bar{\mathbf{u}}$ that fulfills the source condition (SC). On the other hand, one can check if a given function fulfills the source condition: For given $\bar{\mathbf{u}}$, the left hand side $(a_k^{sc}, b_k^{sc})^T$ of (7.2) can be computed. Thus, in order to find $\mathbf{v}$, one has to invert $\mathbf{A}_k^*$ for each k. It is easy to see that

$$(\mathbf{A}_k^*)^{-1} = \frac{1}{\pi} \frac{1}{(a_k^r)^2 + (b_k^r)^2} \begin{bmatrix} \bar{a}_k^r & -b_k^r \\ b_k^r & a_k^r \end{bmatrix},$$

which is well defined as long as $(a_k^r)^2 + (b_k^r)^2 \neq 0$. As $\mathbf{A}_k^*$ is invertible in this case, a source condition is formally fulfilled for every given solution, i.e., $\mathbf{u} = \mathbf{A}^*\mathbf{v}$. However, the source condition is only fulfilled if $\mathbf{v} \in \ell_2$, which requires the coefficients of $\mathbf{u}$ to converge to zero fast enough. We have the following result.

COROLLARY 7.2. *Assume that the coefficients of the kernel function fulfill $(a_k^r)^2 + (b_k^r)^2 > 0$ for all $k \in \mathbb{N}$. Then each sparse solution $\bar{\mathbf{u}}$ fulfills the source condition (SC) with a sparse source element $\mathbf{v}$.*

Proof. Let $\bar{\mathbf{u}} = \{a_0^{\bar{u}}, a_1^{\bar{u}}, b_1^{\bar{u}}, \cdots\}$. As $\mathbf{A}_k^*$ is invertible, one can find $\mathbf{v} = \{a_0^v, a_1^v, b_1^v, \cdots\}$ with

$$\begin{bmatrix} p\omega_k^a \operatorname{sgn}(a_k^{\bar{u}})|a_k^{\bar{u}}|^{p-1} \\ p\omega_k^b \operatorname{sgn}(a_k^{\bar{u}})|b_k^{\bar{u}}|^{p-1} \end{bmatrix} = \mathbf{A}_k^* \begin{bmatrix} a_k^v \\ b_k^v \end{bmatrix}.$$

As $\bar{\mathbf{u}}$ is sparse, there exists $k_0 \in \mathbb{N}$ such that for all $k > k_0$, $a_k^{\bar{u}} = b_k^{\bar{u}} = 0$ holds. By the above equation this also yields $a_k^v = b_k^v = 0$, i.e., $\mathbf{v}$ is sparse and thus belongs to ℓ_2. ☐

In the following we will characterize the source condition in terms of the decay rate of the coefficients of the solution. In order to simplify the notation, we will write $a \sim b$ if $|a| \leq C|b|$ holds for some constant C independent of a and b.

ETNA
Kent State University
http://etna.math.kent.edu

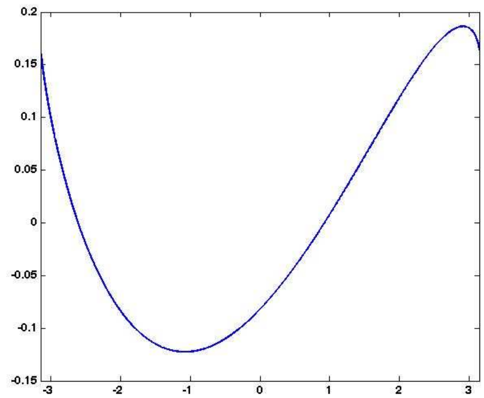

FIG. 7.1. *Kernel function r.*

COROLLARY 7.3. *Assume that the Fourier coefficients $a_k^r,\ b_k^r$, $k \in \mathbb{N}$, of the kernel r fulfill*

$$a_k^r,\ b_k^r \sim k^{-s}, \qquad s > 0\,.$$

Then $\mathbf{u} = \{a_0^u, a_1^u, b_1^u, \cdots\}$ fulfills the source condition only if $\mathbf{u} \in \ell_{2(p-1),k^{2s}\omega_k^2}$.

Proof. As pointed out earlier in this section, every function in $D(\mathbf{A})$ formally fulfills the source condition. Thus we only have to verify that $\mathbf{v}$ with $\mathbf{A}^*\mathbf{v} = f'(\mathbf{u})$ belongs to the space ℓ_2. Denoting the coefficients of $\mathbf{A}^*\mathbf{v}$ by $\{a_0^*, a_1^*, b_1^*, \ldots\}$, one obtains from (7.3),

$$a_k^* \sim (|a_k^v| + |b_k^v|)\, k^{-s},$$
$$b_k^* \sim (|a_k^v| + |b_k^v|)\, k^{-s}.$$

The source condition yields

$$\left|p\omega_k^a \operatorname{sgn}(a_k^u)|a_k^u|^{p-1}\right| \sim (|a_k^v| + |b_k^v|)\, k^{-s},$$
$$\left|p\omega_k^b \operatorname{sgn}(a_k^u)|b_k^u|^{p-1}\right| \sim (|a_k^v| + |b_k^v|)\, k^{-s},$$

or

$$k^{2s}(\omega_k^a)^2|a_k^u|^{2(p-1)} \sim (|a_k^v| + |b_k^v|)^2 \sim |a_k^v|^2 + |b_k^v|^2,$$
$$k^{2s}(\omega_k^b)^2|b_k^u|^{2(p-1)} \sim |a_k^v|^2 + |b_k^v|^2\,.$$

For $\mathbf{v} \in \ell_2$, the right-hand side has to be summable, and thus also the left-hand side, which results in $\mathbf{u} \in \ell_{2(p-1),k^{2s}\omega_k^2}$. □

Thus, the source condition requires the solution to go to zero much faster than would be the case if the solution only belonged to $D(\mathbf{A}) = \ell_{p,\omega}$. In particular, the growth of the factor $k^{2s}\omega_k^a|a_k^u|^{p-2}$ (similar for the coefficient b_k) has to be compensated.

7.2. Numerical results. For the numerical realization, the interval $[-\pi, \pi]$ was divided into 2^{12} equidistant intervals, leading to a discretization of $\mathbf{A}$ as $2^{12} \times 2^{12}$ matrix. The convolution kernel r was defined by its Fourier coefficients $(a_0^r, a_1^r, b_1^r, a_2^r, b_2^r, \cdots)$ with

$$a_0^r = 0,$$
$$a_k^r = (-1)^k \cdot k^{-2},$$
$$b_k^r = (-1)^{k+1} \cdot k^{-2}\,.$$

ETNA
Kent State University
http://etna.math.kent.edu

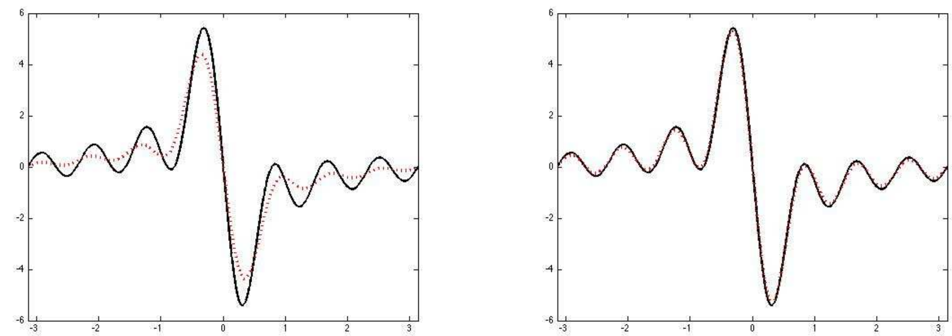

FIG. 7.2. *Example 7.4, solution (solid) and reconstruction (dotted) for different error levels* $\delta = 0.1$ *(left) and* $\delta = 0.0005$ *(right).*

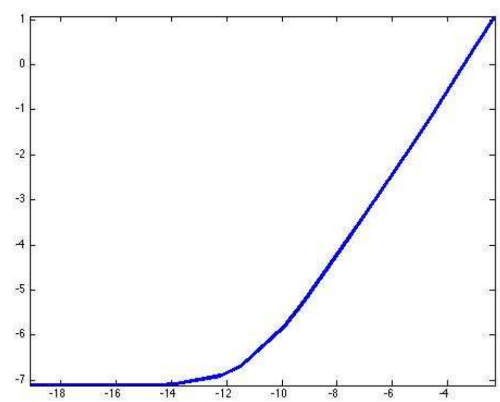

FIG. 7.3. *Example 7.4, log-log plot of the reconstruction error in dependance of the data error for the reconstruction of a sparse solution.*

Notice that r fulfills the decay condition of Corollary 7.3 with $s = 2$. Moreover, we have $(a_k^r)^2 + (b_k^r)^2 > 0$ for all $k \in \mathbb{N}$ and thus all matrices $\mathbf{A}_k^*$ are invertible and Corollary 7.2 applies. A plot of the kernel function can be seen in Figure 7.1.

For our numerical tests, we set $p = 1.1$ in Examples 7.4–7.6 and $p = 1.5$ in Example 7.7.

7.2.1. Minimization of the Tikhonov functional. In order to verify our analytical convergence rate results, the minimizers of the Tikhonov functionals for given $1 < p < 2$ have to be computed. We used an iterative approach based on the minimization of so-called surrogate functionals, i.e., we computed a sequence $\{u_k\}$ as

$$\begin{aligned} u_{k+1} &= \arg\min J_\alpha^s(u, u_k), \\ J_\alpha^s(u, u_k) &= \frac{1}{2}\|Au - y^\delta\|^2 + \alpha f(u) + C\|u - u_k\|^2 - \|A(u - u_k)\|^2. \end{aligned}$$

For a given penalty term $f(x) = \|x\|_{p,\omega}^p$ the minimizer of $J_\alpha^s(u, u_k)$ can be computed explicitly, and the iterates $\{u_k\}$ converge towards the unique minimizer of the functional. For a detailed convergence analysis we refer to [9]. We remark that the algorithm is stable, but also

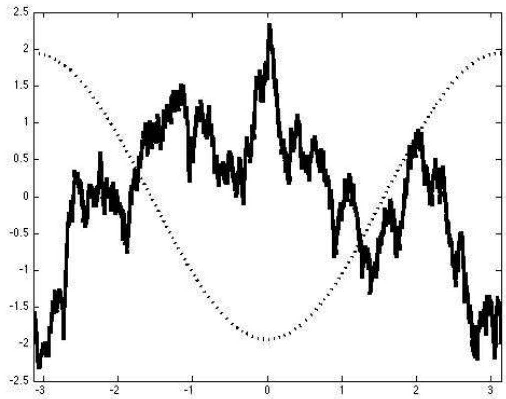

FIG. 7.4. *Example 7.5, source term* $\mathbf{v}$ *(solid) and solution* $\bar{\mathbf{u}}$ *(dotted) with* $f'(\bar{\mathbf{u}}) = \mathbf{A}^*\mathbf{v}$ *for a nonsparse* $\mathbf{v}$ *with a decay of the Fourier coefficients as* $1/k$.

very slowly convergent. In particular, for small data error δ, many iterations are needed in order to achieve the required convergence rate. Another problem lies in the fact that the updates computed by the algorithm are rapidly getting rather small and, due to the limited accuracy of the computer arithmetic, inaccurate. At a certain error level, this leads to a stagnation of the reconstruction accuracy and convergence rates cannot be observed anymore. The error level at which this effect occurs depends on the underlying solution. We remark further that Newton's method, applied to the necessary condition for a minimizer of the Tikhonov functional $J_\alpha(u)$, fails as a reconstruction method. The method involves the second derivative of the penalty $f(u) = \|u\|_p^p$, which is singular at zero for $p < 2$. This causes problems in particular for the reconstruction of a sparse solution.

EXAMPLE 7.4. *Reconstruction of a sparse solution.* Our numerical tests start with a reconstruction of a sparse solution from the associated noisy data. As solution we chose $\bar{\mathbf{u}} = \{a_0^{\bar{u}}, a_1^{\bar{u}}, b_1^{\bar{u}}, \cdots\}_{k\in\mathbb{N}}$ with

$$a_0^{\bar{u}} = 0,$$
$$a_k^{\bar{u}} = \begin{cases} 10^{-3} \text{ for } k = 1, \cdots, 7, \\ 0 \text{ for } k > 7, \end{cases}$$
$$b_k^{\bar{u}} = \begin{cases} -1 \text{ for } k = 1, \cdots, 7, \\ 0 \text{ for } k > 7\ . \end{cases}$$

For the reconstruction, the parameters in the penalty were set to $p = 1.1$, while the weights were set to $\omega_k^a = \omega_k^b = 1$ for all k. As $\bar{\mathbf{u}}$ is sparse, it fulfills the source condition according to Corollary 7.2, and we expect a convergence rate of $O(\delta^{1/2})$ with respect to the $\ell_{1.1}$ norm. Figure 7.2 shows the reconstructions for two different error levels, and Figure 7.3 shows a log-log plot of the reconstruction error versus data error. Reconstructions were carried out for different noise levels $\delta \in \{0.5, 1\} \cdot 10^{-l}, l = 1, 2, \cdots, 9$. Up to a certain error level, Figure 7.3 shows a linear behavior of the log-log plot, whereas, for the smaller noise levels only small improvements are visible. This is mainly due to the fact that the regularization parameter used for small data error is also very small, which usually yields relatively flat functionals around the associated minimizers. One of the main problems of the iterative methods for the minimization of the Tikhonov functional with sparsity constraints is their slow convergence, which leads to very small updates at the final stages of the iteration. Due

ETNA
Kent State University
http://etna.math.kent.edu

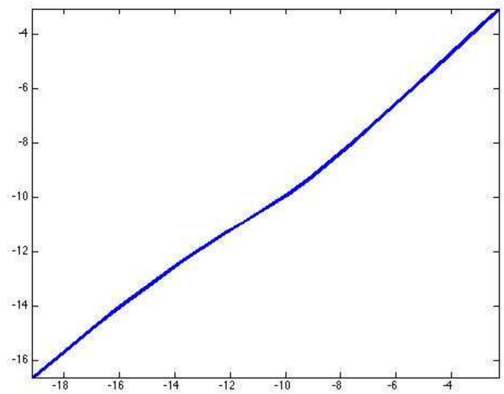

FIG. 7.5. *Example 7.5, log-log plot of the reconstruction error in dependance of the data error level for the reconstruction of a nonsparse solution fulfilling the source condition (SC).*

to numerical errors, the computed descent directions are not correct anymore, and the iteration stalls. Nevertheless, by a linear fitting of the log-log plot over the whole range of the data errors we obtained a convergence rate of $O(\delta^\nu)$ with $\nu = 0.4924$, which is reasonably close to the expected $\nu = 0.5$. If we do restrict the fitting process to the almost linear part of the plot (with smallest error level $\delta = 10^{-6}$), then we obtain a convergence rate of $O(\delta^\nu)$ with $\nu = 0.76$, which is well above the expected rate.

EXAMPLE 7.5. *A nonsparse solution fulfilling the source condition (SC).* In the next test, we used a nonsparse solution that still fulfills the source condition (SC). Again, the parameters were set to $p = 1.1$ and $\omega_k^a = \omega_k^b = 1$ for all k. The solution was constructed by choosing $\mathbf{v} \in \ell_2$ first and computing the solution $\bar{\mathbf{u}}$ with $f'(\bar{\mathbf{u}}) = \mathbf{A}^*\mathbf{v}$ afterwards. The source $\mathbf{v} = \{a_0^v, a_1^v, b_1^v, \cdots\}$ was set to

$$a_k^v = \alpha_k k^{-1},$$
$$b_k^v = \beta_k k^{-1},$$

with $\alpha_k,\ \beta_k \in [-1, 1]$ randomly chosen. Figure 7.4 shows both $\bar{\mathbf{u}}$ and $\mathbf{v}$. The reconstruction was carried out again for $\delta \in \{0.5, 1\} \cdot 10^{-l}, l = 1, 2, \cdots, 9$. In this case, the minimizers of the solution were well reconstructed even for small data error, which is reflected in the almost linear behavior of the log-log plot in Figure 7.5. A possible explanation for the good results of the minimization algorithms might be the structure of the solution. Although the solution is not sparse, the coefficients decay rapidly, which leads to fewer significant coefficients as in Example 7.4. We obtained a convergence rate of $O(\delta^\nu)$ with $\nu = 0.7812$, which is again well above the expected rate.

If we compare the log-log plots of the reconstruction for the sparse solution and for the nonsparse solution, then we observe that, although similar rates are achieved, the absolute error of the reconstruction for the sparse solution is significantly higher than for the non-sparse solution. The explanation for this behavior is quite simple: The source condition reads $f'(\bar{\mathbf{u}}) = \mathbf{A}^*\mathbf{v}$ and the constant in the convergence rate estimate depends on the norm of $\mathbf{v}$. In Example 7.4, the norm of the source element $\|\mathbf{v}\|$ for the sparse solution is approximately 20 times bigger than for the nonsparse solution in Example 7.5, and therefore the absolute reconstruction error is larger in the first case.

EXAMPLE 7.6. *A solution fulfilling no source condition.* For the third test we set $\mathbf{u} = \mathbf{v}$ with $\mathbf{v}$, p and $\omega_k^a,\ \omega_k^b$ defined as in Example 7.5. Since $\mathbf{v}$ decays only as k^{-1}, we have $\mathbf{v} \in \ell_2$

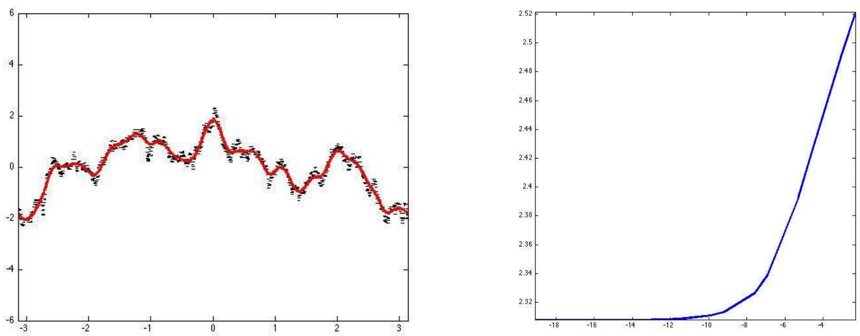

FIG. 7.6. *Example 7.6, reconstruction of a solution fulfilling no source condition (left) and log-log plot of the reconstruction error in dependance of data error level.*

only. Therefore, the solution fulfills no source condition and we do not expect a convergence rate of $O(\delta^{1/2})$. Indeed, the reconstructions for $\delta \in \{0.5, 1\} \cdot 10^{-l}, l = 1, 2, \cdots, 9$ suggest a rate of at most $O(\delta^{\nu})$ with $\nu = 0.01$; see Figure 7.6. Even if we assume that the minimizers were not reconstructed with high accuracy for small error level δ by the iterative method, and take into account only the reconstructions where the plot in Figure 7.6 shows linear behavior, we still obtain a rate of $\nu = 0.0377$ only. Clearly, the convergence is much slower (also in absolute values) than in the previous cases.

EXAMPLE 7.7. *A nonsparse solution fulfilling the source condition (SC I).* Finally, we present a reconstruction from noisy data $\mathbf{y}^{\delta}$, where the associated solution $\bar{\mathbf{u}}$ fulfills the source condition $f'(\bar{\mathbf{u}}) = \mathbf{A}^*\mathbf{A}\mathbf{v}$. In this case, a convergence rate of at least $O(\delta^{\frac{p}{p+1}})$ is expected. For the reconstruction, we used a different setting than in the previous cases. First, we changed the penalty f by using now $p = 1.5$ and weight functions $\omega_k^a = \omega_k^b = k$. Second, we chose a different source term $\mathbf{v} = \{a_0^v, a_1^v, b_1^v, \cdots\}$ with coefficients,

$$\begin{aligned} a_0^w &= 0\,, \\ a_k^w &= k^{-1.4}\,, \\ b_k^w &= -k^{-1.4}\,. \end{aligned}$$

It is easy to see that the source element belongs to $\ell_{p,\omega}$. The solution was then determined as $\bar{\mathbf{u}} = \mathbf{A}^*\mathbf{A}\mathbf{v}$, Figure 7.7 displays both $\mathbf{v}$ and $\bar{\mathbf{u}}$. The reconstruction was again carried out for error levels $\delta = \{0.5, 1\} \cdot 10^{-l}, l = 1, 2, \cdots, 9$, and the reconstruction accuracy was measured in the $\ell_{p,\omega}$ norm; see Figure 7.8 for the results. From the reconstructions we obtained a convergence rate of $O(\delta^{0.66})$ which is slightly higher than the theoretically expected rate of $O(\delta^{0.6})$.

The observed convergence rates for the reconstructions in Examples 7.4-7.7 are summarized in Table 7.1.

Acknowledgments. The second author acknowledges support from the Austrian Science Fund, Elise Richter scholarship (V82-N18 FWF). Part of her work was done while she was with the Radon Institute, Austrian Academy of Sciences. The first author acknowledges support from the Austrian Science Fund, Projects P19496-N18 and P20237-N14. The authors are grateful to Wei Wang (Harbin Institute of Technology) for useful remarks.

ETNA
Kent State University
http://etna.math.kent.edu

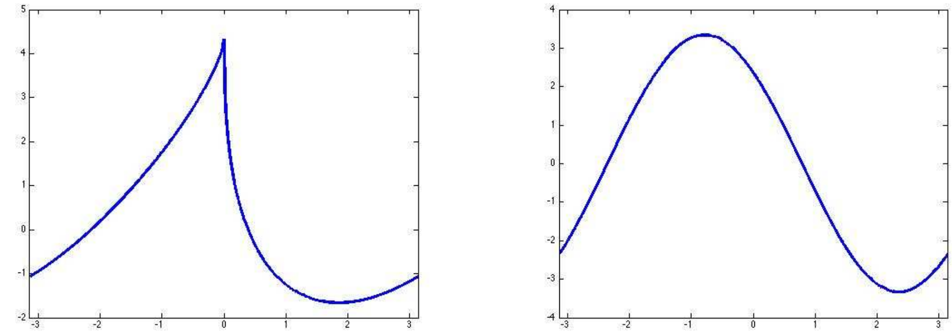

FIG. 7.7. *Example 7.7, source element* $\mathbf{v}$ *(left) and solution* $\bar{\mathbf{u}}$ *with* $f'(\bar{\mathbf{u}}) = \mathbf{A}^*\mathbf{A}\mathbf{v}$ *(right).*

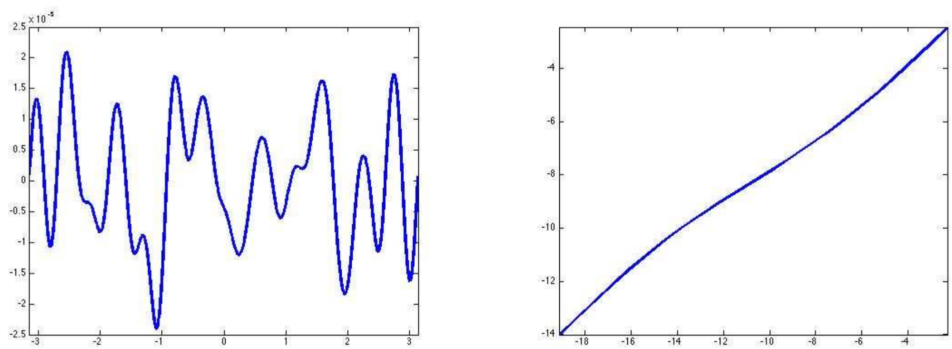

FIG. 7.8. *Example 7.7, plot of* $\bar{\mathbf{u}} - \mathbf{u}_\alpha^\delta$ *for* $\delta = 5 \cdot 10^{-5}$ *(left) and log-log plot of the reconstruction error in dependence of the data error (right).*

REFERENCES

[1] V. BARBU AND T. PRECUPANU, *Convexity and Optimization in Banach Spaces*, Second ed., D. Reidel, Dordrecht, 1986.

[2] D. BUTNARIU AND A. N. IUSEM, *Totally Convex Functions for Fixed Points Computation and Infinite Dimensional Optimization*, Kluwer Academic Publishers, Dordrecht, 2000.

[3] L. BREGMAN, *The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming*, USSR Comput. Math. Math. Phys., 7 (1967), pp. 200–217.

[4] M. BURGER AND S. OSHER, *Convergence rates of convex variational regularization*, Inverse Problems, 20 (2004), pp. 1411–1421.

[5] M. BURGER, E. RESMERITA, AND L. HE, *Error estimation for Bregman iterations and inverse scale space methods in image restoration*, Computing, 81 (2007), pp. 109–135.

[6] A. CHAMBOLLE, R. DEVORE, N. Y. LEE, AND B. LUCIER, *Non-linear wavelet image processing: Variational problems, compression and noise removal through wavelet shrinkage*, IEEE Trans. Image Process., 7 (1998), pp. 319–333.

[7] I. CIORANESCU, *Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems*, Kluwer Academic Publishers, Dordrecht, 1990.

[8] A. COHEN, *Numerical Analysis of Wavelet Methods*, North-Holland, Amsterdam, 2003.

[9] 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 (2004), pp. 1413–1457.

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

[11] H. ENGL, M. HANKE, AND A. NEUBAUER, *Regularization of Inverse Problems*, Kluwer Academic Publish-

TABLE 7.1
Expected and numerically observed convergence rates for different types of solutions $\bar{\mathbf{u}}$

	$f'(\bar{\mathbf{u}}) = \mathbf{A}^*\mathbf{v}$		$f'(\bar{\mathbf{u}}) = \mathbf{A}^*\mathbf{A}\mathbf{v}$	no source condition
	sparse $\bar{\mathbf{u}}$	nonsparse $\bar{\mathbf{u}}$		
expected rate	$O\left(\delta^{0.5}\right)$	$O\left(\delta^{0.5}\right)$	$O\left(\delta^{0.6}\right)$	—
observed rate	$O\left(\delta^{0.49}\right)$	$O\left(\delta^{0.76}\right)$	$O\left(\delta^{0.66}\right)$	$O\left(\delta^{0.0377}\right)$

ers, Dordrecht, 1996.

[12] M. GRASSMAIR, M. HALTMEIER, AND O. SCHERZER, *Sparse regularization with ℓ_q penalty term*, Inverse Problems, 24 (2008), pp. 1–13.

[13] B. HOFMANN, B. KALTENBACHER, C. POESCHL, AND O. SCHERZER, *A convergence rates result in Banach spaces with non-smooth operators*, Inverse Problems, 23 (2007), pp. 987–1010.

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

[15] E. KLANN, P. MAASS, AND R. RAMLAU, *Two-step regularization methods for linear inverse problems*, J. Inverse Ill-Posed Probl., 14 (2006), pp. 583–609.

[16] P. MATHÉ AND S. PEREVERZEV, *Optimal discretization of inverse problems in Hilbert scales. Regularization and self-regularization of projection methods*, SIAM J. Numer. Anal., 38 (2001), pp. 1999–2021.

[17] J. MUSIELAK, *Orlicz Spaces and Modular Spaces*, Lecture Notes in Math., Vol. 1034, Springer, Berlin, 1983.

[18] A. NEUBAUER, *When do Sobolev spaces form a Hilbert scale*, Proc. Amer. Math. Soc., 103 (1988), pp. 557–562.

[19] R. RAMLAU, *Regularization properties of Tikhonov regularization with sparsity constraints*, Electron. Trans. Numer. Anal., 30 (2008), pp. 54–74.
http://etna.math.kent.edu/vol.30.2008/pp54-74.dir.

[20] R. RAMLAU AND G. TESCHKE, *A Projection iteration for nonlinear operator equations with sparsity constraints*, Numer. Math., 104 (2006), pp. 177–203.

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

[22] E. RESMERITA AND O. SCHERZER, *Error estimates for non-quadratic regularization and the relation to enhancing*, Inverse Problems, 22 (2006), pp. 801–814.

[23] H-K. XU, *Inequalities in Banach spaces with applications*, Nonlinear Anal., 16 (1991), pp. 1127–1138.