Optimal convergence rates for Tikhonov regularization in Besov scales

D A Lorenz and D Trede
Zentrum für Technomathematik, University of Bremen, D-28334 Bremen, Germany
E-mail: dlorenz@math.uni-bremen.de and trede@math.uni-bremen.de

Received 5 June 2008, in final form 18 July 2008
Published 26 August 2008
Online at stacks.iop.org/IP/24/055010

Abstract
In this paper we deal with linear inverse problems and convergence rates for Tikhonov regularization. We consider regularization in a scale of Banach spaces, namely the scale of Besov spaces. We show that regularization in Banach scales differs from regularization in Hilbert scales in the sense that it is possible that stronger source conditions may lead to weaker convergence rates and vice versa. Moreover, we present optimal source conditions for regularization in Besov scales.

1. Introduction
Regularization of inverse problems formulated in Banach spaces has been of interest recently. On the one hand there are several theoretical regularization results such as convergence rates in a general Banach space setting, see e.g. [5, 6, 14, 15, 22, 23], or convergence rates for special sequence spaces such as $\ell^p$, $1 \leq p < 2$, see e.g. [18, 20]. On the other hand there are results which deal with solving inverse problems formulated in Banach spaces, such as Landweber-like iterations or minimization methods for Tikhonov functionals, see e.g. [1–4, 11, 13, 21, 26, 27], respectively.

The interest in Banach spaces is due to the fact that in many situations a Banach space is better suited to model the data under consideration than a Hilbert space. In the context of image processing, for example, the Banach space $BV$ of functions of bounded variation is used to model images with discontinuities along lines [5, 24, 28]. Moreover, two examples in which the use of Banach spaces is necessary for a thorough formulation of the problem are presented in [15]. Another class of Banach spaces are the Besov spaces $B^{s}_{p,q}$ which play an important role in inverse problems related to image processing, see e.g. [7, 8, 17].

In this paper we make a first attempt to analyze inverse problems in scales of Banach spaces generalizing classical Hilbert scales [19]. The easiest scale of Banach spaces is the scale of Sobolev spaces $W^r_0$. However, we are going to use Besov spaces $B^{s}_{p}$ since they coincide with the Sobolev scale in most cases if the integrability indices coincide. Moreover,
they come with a characterization in terms of wavelet coefficients which make them easy to use for our purposes. We apply previous convergence rate regularization results from [5, 15] in the scale of Besov spaces and develop optimal convergence rates. To this end, we derive the source conditions that lead to a convergence rate of $O(\sqrt{\delta})$ in a certain Sobolev space.

Consider the equation

$$ Fu^\dagger = v, $$

(1)

where $F$ is a linear continuous operator:

$$ F : B_D \rightarrow L_2, $$

between the Besov space $B_D := B^{s_D}_{p_D, p_D}, s_D \in \mathbb{R}, p_D > 1$, and the Lebesgue space $L_2$. In general, these function spaces contain functions or distributions defined on the subset $\Omega \subset \mathbb{R}^d$. For clarity we omit $\Omega$ in the following. The Besov spaces $B_{p,p}$ are subspaces of the space of tempered distributions $S'$ and, in contrast to $S'$, they are Banach spaces for $p \geq 1$ [25].

Different from classical approaches we use the domain $B_D$—often a superset of $L_2$—and not $L_2$ itself. That may be of interest in some applications, e.g. mass-spectrometry where the data consist of delta peaks (see [10, 16]) which are not the elements of $L_2$.

If we assume that only noisy data $v^\delta$ with noise level $\|v - v^\delta\| \leq \delta$ are available, the solution of (1) could be unstable and has to be stabilized by regularization methods. We use regularization with a Besov constraint, i.e. we regularize by minimizing a not necessarily quadratic functional $T_\alpha : B_D \rightarrow [0, \infty]$ defined by

$$ T_\alpha(u) := \|Fu - v^\delta\|_{L_2}^2 + \alpha \|u\|_{B_{R}}^p, $$

(2)

where $B_R := B^{s_R}_{p_R, p_R}$ is a Besov space, not necessarily equal to $B_D$. Since $T_\alpha$ shall be defined on $B_D$ we define

$$ \|u\|_{B_{R}} = \infty $$

for $u \notin B_R$.

In this paper we will investigate regularization properties and convergence rates of the regularization method consisting of the minimization of (2), i.e. $u_{\alpha, \delta} \in \operatorname{argmin}T_\alpha(u)$. The proceeding is as follows:

(i) In section 2 we introduce the notation and collect preliminary results.

(ii) In section 3 we apply convergence rate results for Banach spaces [5, 15]. With the constraints on $p_R$ and $s_R$ in mind and the parameter rule $\alpha = \delta$, we will get a stable approximation, i.e.,

$$ \|u_{\alpha, \delta} - u^\dagger\|_{B_R} \rightarrow 0, \quad \delta \rightarrow 0, $$

and a convergence rate in the Sobolev space $H^{\sigma}$,

$$ \|u_{\alpha, \delta} - u^\dagger\|_{H^{\sigma}} = O(\sqrt{\delta}), $$

with $\sigma$ depending on $s_R$ and $p_R$ (theorems 3.1 and 3.2). These results restrict the choice of possible regularization spaces $B_R$. Using Besov space embeddings in section 4 we will get a generalization of the first result. We find a convergence rate—also formulated in a Sobolev space—which holds for a larger set of Besov space penalties $\|\cdot\|_{B_R}^p$ (theorem 4.1).

(iii) The convergence result gets stronger as $\sigma$ increases since for $\theta > 0$ it holds $H^{\sigma+\theta} \subset H^\sigma$. Since $\sigma$ depends on $s_R$ and $p_R$, we address the question of how to choose $B_R$ in a way such that $\sigma$ is maximal. We will find the regularization penalty $\|\cdot\|_{B_R}^p$, which gives the best estimate with respect to $\sigma$.

(iv) In section 5 we apply these results to some operators defined in Sobolev and Besov spaces to demonstrate the differences.
2. Notation and basic Besov space properties

As already mentioned, the Besov spaces $B^s_{p,q}$ are subspaces of the space of tempered distributions $S'$. They coincide with special cases of traditional smoothness function spaces such as Hölder and Sobolev spaces. Note, e.g., that $B^s_{p,p} = W^s_p$ for $s \not\in \mathbb{Z}$ and $p \geq 1$ and even $B^s_{2,2} = H^s := W^s_2$ for all $s \in \mathbb{R}$. As from now on we use the term Sobolev space only for the Hilbert spaces $H^s$. This clarifies the characterization that the Besov space $B^s_{p,q}$ contains functions with $s$ derivatives in the $L_p$ norm. The second integrability index $q$ declares a finer nuance of smoothness. In the following we omit the second integrability index $q$ of the Besov spaces, which is always equal to the corresponding first one $p$.

There are several ways of defining Besov spaces. Most commonly they are defined via the modulus of smoothness, a way to model differential properties. For a detailed introduction of Besov spaces via moduli of smoothness in conjunction with other smoothness spaces see e.g. [12, section 4.5].

Another way of defining Besov spaces is based on wavelet coefficients. According to [9] for all $s \in \mathbb{R}$, $p > 0$, there exists a wavelet basis $\{\psi_\lambda\}_{\lambda \in \Lambda}$ such that

$$
\|u\|_{B^s_p}^p \asymp \sum_{\lambda \in \Lambda} 2^{sp(\frac{s}{2} + \frac{d}{p} - \frac{1}{2} + \frac{1}{q}))|\lambda|} |u_\lambda|^p,
$$

(3)

where $u_{\lambda} = \langle u, \psi_\lambda \rangle = \int u \psi_\lambda \, dx$ are the wavelet coefficients of $u$. The notation $A \asymp B$ means that there exist constants $c, C > 0$ such that $cA \leq B \leq CA$. We will use this equivalent norm throughout this paper.

An important ingredient in the analysis of the regularization method (2) is the embedding result (cf [25]).

**Proposition 2.1.** Let $B^s_{p_1}, B^s_{p_2}$ be Besov spaces. If

$$
s_1 = \frac{d}{p_1} > s_2 = \frac{d}{p_2} \quad \text{and} \quad p_1 \leq p_2,
$$

(4)

then $B^s_{p_1} \subset B^s_{p_2}$ continuously. The term $s = \frac{d}{p}$ is called the differential dimension of $B^s_p$.

The embedding of Besov spaces is often visualized with the help of the DeVore diagram [12] where one plots the smoothness $s$ against $1/p$, see figure 1. By $B^{s_1}_{p_1} \subset B^{s_2}_{p_2}$, in the following, we denote not only the set-theoretical embedding but also the continuous embedding.
We are going to use the following Besov spaces:

- \(B_D := B_D^{p_D} \) for the domain of \(F\).
- \(B_R := B_R^{p_R} \) for the space in which we regularize.
- \(B_S := B_S^{p_S} \) for the source condition.
- \(B_G := B_G^{p_G} \) for the range of \(F^* \) which models the smoothing properties of \(F\).

As stated above, the smoothing properties of the operator \(F : B_D \rightarrow L_2\) are modeled by assuming that the range of its adjoint is small, namely 
\[ \text{rg} F^* = B_G \subset B_D^{p_D\ast} , \]
where \(p^\ast\) is defined via \(\frac{1}{p} + \frac{1}{p^\ast} = 1\), hence \(p^\ast = \frac{p}{p-1}\). Consequently, we have 
\[ s_G - \frac{d}{p_G} > -s_D - \frac{d}{p_D^\ast} , \]
\[ \frac{1}{p_G} \geq \frac{1}{p_D^\ast} . \]

3. Convergence and convergence rates

The first result we need is a regularization result in the regularization space \(B_R\).

**Theorem 3.1.** Let \(B_R \subset B_D\) and let \(u^\dagger\) be a minimum-\(\| \cdot \|_{B_R}\)-solution of \(Fu = v\). Then, for each minimizer \(u_{\alpha,\delta}\) of 
\[ T_{\alpha}(u) = \|Fu - v\|^2 + \alpha \|u\|^p_{B_R}, \]
and the parameter rule \(\alpha \asymp \delta\) we get convergence
\[ \|u_{\alpha,\delta} - u^\dagger\|_{B_R} \rightarrow 0, \quad \delta \rightarrow 0. \]

**Proof of theorem 3.1.** We equip \(B_D\) and \(L_2\) with the weak topologies and want to use theorem 3.5 from [15]. To do so, we need the following to be fulfilled:

(i) The norm \(\| \cdot \|_{L_2}\) is weakly lower semicontinuous in \(L_2\).
(ii) \(F : B_D \rightarrow L_2\) is weakly continuous.
(iii) \(\| \cdot \|_{B_R}\) is proper, convex and weakly lower semicontinuous in \(B_D\).
(iv) The sets \(A_{\alpha} = \{u | \|Fu - v\|^2 + \alpha \|u\|^p_{B_R} < M\}\) are weakly sequentially compact in \(B_D\).

The first point is obvious and the second point is fulfilled by the assumption that \(F\) is linear and continuous. For the forth point note that due to the continuous embedding we have \(\| \cdot \|_{B_D} \leq C \| \cdot \|_{B_R}\) and hence, the sets \(A_{\alpha}\) are bounded in \(B_D\) which implies weak sequential compactness due to the reflexivity of \(B_D\).

For the third point note that there exists a wavelet basis \(\{\psi_k\}\) which is an unconditional basis for both \(B_R\) and \(B_D\). Now, let \(u_k \rightarrow u\) weakly in \(B_D\) and \(u_k \in B_R\). Since \(\psi_k\) are also elements of the dual spaces \(B_R^\ast\) and \(B_D^\ast\) it holds for all \(\lambda\)
\[ \langle u_k, \psi_k \rangle \xrightarrow{k \rightarrow \infty} \langle u, \psi_k \rangle \]
and hence, a sequence \(u_k\) bounded in \(B_R\) converges weakly to \(u\) in \(B_R\) if it does in the larger space \(B_D\), because the duality pairing is the same in both spaces and \(\{\psi_k\}\) is an unconditional basis in \(B_R\). This shows the weak lower semicontinuity of \(\| \cdot \|_{B_R}\) on \(B_R\)-bounded sets in \(B_D\) (which is sufficient for theorem 3.5 from [15] to hold).
Now, by theorem 3.5 from [15] it follows that $u^{\alpha, \delta} \rightharpoonup u^\dagger$ weakly in $B_D$ (and by the above considerations also in $B_R$) and moreover $\|u^{\alpha, \delta}\|_{B_R} \to \|u^\dagger\|_{B_R}$. Since $B_R$ is uniformly convex, this implies $u^{\alpha, \delta} \to u^\dagger$ strongly in $B_R$. □

Now we formulate a theorem on the rate of convergence which follows from the general results on regularization in Banach spaces [5]. We assume that certain knowledge on the true solution $u^\dagger$ is available, i.e. a certain source condition is fulfilled. The source condition is formulated in terms of Besov smoothness. This assumption, together with the assumptions on the range of $F^*$, leads to a regularization term for which a certain convergence rate in the Sobolev norm can be proven.

**Theorem 3.2.** Let $u^\dagger \in B_S \subset B_D$ with $p_S \leq p_G$. Then, for each minimizer $u^{\alpha, \delta}$ of the Tikhonov functional

$$T_\alpha(u) = \|Fu - v_\delta\|^2 + \alpha\|u\|_{B_R}^{p_R},$$

with

$$p_R = \frac{p_S + p_G}{p_G}, \quad s_R = \frac{p_S s_S - p_G s_G}{p_S + p_G}$$

(6)

and the parameter rule $\alpha \asymp \delta$ we get the convergence rate

$$\|u^{\alpha, \delta} - u^\dagger\|_{H^\sigma} = O(\sqrt{\delta}),$$

(7)

where $\sigma := s_R + d\left(\frac{1}{2} - \frac{1}{p_S}\right)$.

**Remark 3.3.** Note that in general the convergence statements in theorems 3.1 and 3.2 correspond to different Besov spaces $B_R$ and $H^\sigma$. The spaces coincide if and only if $p_S = p_G$, otherwise we cannot give any information of inclusions, because the differential dimensions are equal:

$$\sigma - d = s_R - \frac{d}{p_R}.$$

**Remark 3.4.** The definitions of $p_R$ and $s_R$ in theorem 3.2 imply that $B_S \subset B_R$. Otherwise the statement would not be meaningful, since if $B_S \nsubseteq B_R$,

$$\exists u^\dagger \in B_S : \|u^\dagger\|_{B_R}^{p_R} = \infty.$$

To see this, note that due to $B_G \subset B_D^*$ the inequality $\frac{1}{p_G} \geq \frac{p_G - p_D}{p_D}$ holds and because $B_S \subset B_D$ we get $\frac{1}{p_S} \geq \frac{1}{p_D}$ and hence,

$$p_R = \frac{p_S + p_G}{p_G} = p_S \left(\frac{1}{p_G} + \frac{1}{p_S}\right) \geq p_S \left(\frac{p_D - 1}{p_D} + \frac{1}{p_D}\right) = p_S.$$

To see the inequality for the differential dimension of $B_R$ and $B_S$ note that $B_S \subset B_D \subset B_G^*$, and hence

$$-(s_S + s_G) < d \left(\frac{p_G - 1}{p_G} - \frac{1}{p_S}\right),$$

which leads to

$$-(s_S + s_G)p_G p_S + d(p_G + p_S) - d(p_G p_S)$$

$$< d \left(\frac{p_G - 1}{p_G} - \frac{1}{p_S}\right) p_G p_S + d(p_G + p_S) - d(p_G p_S) = 0.$$
Applying this to the constraints (6) for $p_R$ and $s_R$ yields

$$s_R - \frac{d}{p_R} \leq \frac{p_S s_S - p_G s_G}{p_S + p_G} - d \frac{p_G}{p_S + p_G}$$

$$= s_S - \frac{d}{p_S} + \frac{-(s_S + s_G)p_G p_S + d(p_G + p_S) - d(p_G p_S)}{(p_S + p_G)p_S} < s_S - \frac{d}{p_S}.$$  

For the proof of theorem 3.2 we need a property of the mapping $\| \cdot \|_{p_R}^p : B_S \to [0, \infty)$.

**Proposition 3.5.** Let $u \in B_S$ and let $s_R$ and $p_R$ fulfill (6). Then

$$\partial(\|u\|_{p_R}^p) = \{ \nabla\|u\|_{p_R}^p \} \subset B_G.$$  

Proof. Let $u \in B_S$. Since $p_S, p_G > 0$, $p_R = 1 + \frac{p_S}{p_G} > 1$, we get

$$\partial(\|u\|_{p_R}^p) = \{ \nabla\|u\|_{p_R}^p \} = \left\{ p_R \sum_{\lambda \in \Lambda} 2^{p_R(\frac{s_R d}{p_S} + \frac{1}{p_S} - \frac{1}{p_R})} |\lambda| \sign(u_\lambda)|u_\lambda|^{p_R} \right\},$$  

and hence,

$$\|\nabla\|u\|_{p_R}^p\|_{B_G} = p_R \sum_{\lambda \in \Lambda} 2^{p_R(\frac{s_R d}{p_S} + \frac{1}{p_S} - \frac{1}{p_R}) + p_G p_S} |\lambda| |u_\lambda|^{p_R} \lesssim p_R \|u\|_{B_S}^{p_R} < \infty,$$

since $u \in B_S$.

Now we are able to do the proof of theorem 3.2.

**Proof of theorem 3.2.** In [5] it is proved that the source condition

$$\exists w \in L_2 : \quad F^* w \in \partial(\|u\|_{p_R}^p)$$  

leads to the estimate for the so-called Bregman distance

$$D_{\|u\|_{p_R}^p} (u^{\alpha, \delta}, u) = \mathcal{O}(\delta)$$

for minimizers $u^{\alpha, \delta}$ of the Tikhonov functional (2) and $\alpha \asymp \delta$. Here by assumption the range of the adjoint operator $F^*$ is $B_G$, and hence, with proposition 3.5, we get

$$\partial(\|u\|_{p_R}^p) \ni \nabla\|u\|_{p_R}^p \subset B_G = \text{rg}(F^*),$$

since $u \in B_S$. \qed
thus the source condition (9) is fulfilled. Further, we get with (8)
\[
D_{\nabla|u^\lambda|_{p_R}^r}(u^{a,\delta}, u^1) = \|u^{a,\delta}\|_{p_R}^r - \|u^1\|_{p_R}^r - \langle \nabla \|u^1\|_{p_S}^r, u^{a,\delta} - u^1 \rangle
\]
\[
= \sum_{\lambda \in \Lambda} 2^{p_S(s_R + d(\frac{1}{2} - \frac{1}{p_S}))|\lambda|} \left( |u^{a,\delta}_\lambda|_{p_S}^r - |u^1_\lambda|_{p_S}^r \right)
\]
\[
- p_R \text{ sign}(u^1_\lambda) |u^{a,\delta}_\lambda|_{p_S}^{r-1} (u^{a,\delta}_\lambda - u^1_\lambda)
\]
\[
= O(\delta).
\]
For \(a, b \in \mathbb{R}, C > |a|, |b - a| < L, 1 < p \leq 2\) by [3, lemma 4.7],
\[|b|^p - |a|^p - p \text{ sign}(a)|a|^{p-1}(b-a) \geq k(p, C, L)|b - a|^2,
\]
where \(k(p, C, L)\) is a positive constant which depends on \(p, C\) and \(L\). Since by remark 3.4 it holds
\[
\|u^{a,\delta} - u^1\|_{p_S} \to 0 \quad \text{for } \delta \to 0
\]
we get according to theorem 3.1
\[
\exists C > 0 : \int_2^{(1/s + d(\frac{1}{2} - \frac{1}{p_S}))|\lambda|} |u^{a,\delta}_\lambda| \leq C, \quad \lambda \in \Lambda.
\]
Furthermore, since \(\|u^{a,\delta} - u^1\|_{p_S} \to 0\), we get according to theorem 3.1
\[
\exists L > 0 : \int_2^{(1/s + d(\frac{1}{2} - \frac{1}{p_S}))|\lambda|} |u^{a,\delta}_\lambda - u^1_\lambda| \leq L, \quad \lambda \in \Lambda.
\]
Applying this with \(a = 2^{(1/s + d(\frac{1}{2} - \frac{1}{p_S}))|\lambda|} u^{a,\delta}_\lambda, b = 2^{(1/s + d(\frac{1}{2} - \frac{1}{p_S}))|\lambda|} u^1_\lambda,\) and \(p = p_R \in (1, 2),\) since \(p_G \geq p_S,\) we get
\[
D_{\nabla|u^\lambda|_{p_R}^r}(u^{a,\delta}, u^1) \geq K \sum_{\lambda \in \Lambda} 2^{2(s + d(\frac{1}{2} - \frac{1}{p_S}))|\lambda|} \|u^{a,\delta}_\lambda - u^1_\lambda\|^2
\]
\[
\geq K \|u^{a,\delta}_\lambda - u^1_\lambda\|^2_{H^s(p_S, p_S, \frac{1}{p_S})},
\]
because of the norm equivalence (3) and the fact that \(H^s = B^s_{2,2}\) for all \(s\).

Finally, this gives
\[
\|u^{a,\delta} - u^1\|_{H^s} = O(\sqrt{\delta}),
\]
where \(\sigma := s_R + d\left(\frac{1}{2} - \frac{1}{p_S}\right)\).

4. Source condition weakening

In the setup of theorem 3.2 we assumed that a source condition in terms of Besov smoothness is known, i.e. \(u^1 \in B_{p}^s\). From that a regularization penalty \(\|\cdot\|_{p_S}^r\) was derived which leads to a convergence rate in a certain Sobolev space.

Besov spaces are embedded into each other via the nonlinear intricate properties (4) of proposition 2.1. Considering this, the question arises which penalties \(\|\cdot\|_{p_S}^r\) and convergence rates (i.e. which \(\sigma\)) follow from a weakened source-condition \(u^1 \in B_{p}^s\) with \(B_S \subset B_{p}^s \subset B_{p}^D\). In addition to the embedding properties (4) for the application of theorem 3.2 one has to ensure \(p \leq p_G\). This yields to the following set of possible weaker source conditions \(u^1 \in B_{p}^s\) such that:

\[
B_S \subset B_{p}^s, \quad \text{i.e. } s_S - \frac{d}{p_S} > s - \frac{d}{p}, \quad \text{(10)}
\]
\[
\frac{1}{p_S} \geq \frac{1}{p}, \quad \text{(11)}
\]
\[
B_{p}^s \subset B_{p}^D, \quad \text{i.e. } s - \frac{d}{p} > s_D - \frac{d}{p_D}, \quad \text{(12)}
\]
(10)–(14) gives the following theorem.

**Proof.** From (11), (13) and (14) it follows that theorem 3.2 is applicable for $p_D < p_G$ i.e. equations (10) and (12), one has to choose $p_R$ while trying to find the regularization penalty $\| \cdot \|_{B_R}$. The convergence result in theorem 4.1 gets stronger as $\varepsilon > 0$.

The direct application of theorem 3.2 to the idea of weakening the source condition with $p_R$ as introduced in (14) gives the following theorem.

**Theorem 4.1.** Let $u^1 \in B_S \subset B_D$ with $p_S \leq p_G$. Further let $p > 0$ with $p_S \leq p \leq \min\{p_D, p_G\}$ and $\varepsilon > 0$. Then, for each minimizer $u^{\delta, \varepsilon}$ of the Tikhonov functional

$$T_\varepsilon(u) = \| Fu - v^\delta \|^2 + \alpha \| u \|_{B_R}^2,$$

with

$$p_R := \frac{p + p_G}{p G}, \quad s_R \leq \frac{p S - p S G}{p + p G} = \frac{d}{\frac{1}{p} + \frac{p}{p G}},$$

and the parameter rule $\alpha \asymp \delta$ we get the convergence rate

$$\| u^{\delta, \varepsilon} - u^1 \|_{H^2} = O(\sqrt{\delta}),$$

where $\sigma := s_R + d(\frac{1}{2} - \frac{1}{p G})$.

**Proof.** From (11), (13) and (14) it follows that theorem 3.2 is applicable for $p > 0$ with $p_S \leq p \leq \min\{p_D, p_G\}$. To ensure the embedding properties for the differential dimension, i.e. equations (10) and (12), one has to choose $s \in \mathbb{R}$ with

$$s_S - \frac{d}{p_S} > s = \frac{d}{p} > s_D - \frac{d}{p_D}.$$ 

With that the application of theorem 3.2 yields (15) and hence the convergence in $H^2$. $\square$.

The convergence result in theorem 4.1 gets stronger as $\sigma$ increases. Since $\sigma$ depends on $s_R$ and $p_R$ we address the question of how to choose $B_R$ in a way such that $\sigma$ is maximal. We try to find the regularization penalty $\| \cdot \|_{B_R}^2$, which gives the best estimate with respect to $\sigma$ (while $F$ and the spaces $B_S, B_G$ and $B_D$ are fixed). Since $\sigma$ depends strictly monotone on $s_R$. 

Figure 2 illustrates the set of weaker source conditions, i.e. the equalities (10)–(14), graphically for $p_D < p_G$, which ensures $p \leq p_G$.
we have to choose $s_R$ as large as possible so that we have to solve the following optimization problem, which depends only on $p$:

$$\max \frac{ps_S - pG_sG}{p + pG} \cdot d \left[ \frac{1}{2} - \frac{1}{p + pG} \left( \frac{pG}{pG} + \frac{p}{ps} \right) \right] - \varepsilon \left( \frac{p}{ps} - 1 \right) \right)$$

such that $p_S \leq p \leq \min\{p_D, p_G\}$.

Since $\varepsilon > 0$ can be arbitrarily small we neglect the term $\varepsilon (\frac{p}{ps} - 1)$ and hence, we have to find the maximum of

$$\hat{\sigma}(p) := \frac{ps_S - pG_sG}{p + pG} \cdot d \left[ \frac{1}{2} - \frac{1}{p + pG} \left( \frac{pG}{pG} + \frac{p}{ps} \right) \right] - \varepsilon \left( \frac{p}{ps} - 1 \right) \right).$$

The function $\hat{\sigma}$ is monotonically increasing in $p$, since for $p_1 > p_2$ we get

$$\hat{\sigma}(p_1) - \hat{\sigma}(p_2) = \left( \frac{p_1}{p_1 + pG} - \frac{p_2}{p_2 + pG} \right) \left( \frac{d}{ps} \right) - \varepsilon \left( \frac{p}{ps} - 1 \right) \right).$$

Corollary 4.2. Let $u^\dagger \in B_S \subset B_D, p_S \leq p_G$ and $\varepsilon > 0$ be sufficiently small. Then the Tikhonov regularization $T_\alpha$ with the parameter rule $\alpha \asymp \delta$, penalty according to (15) with $p := \min\{p_D, p_G\}$

gives the strongest convergence.

(i) If $p_G \geq p_D$, we get with $\tilde{\varepsilon} := \varepsilon (\frac{p}{ps} - 1)$

$$p_R = \frac{pD + pG}{pG},$$

$$s_R = \frac{psps - pG_sG}{pD + pG} = d \left( \frac{1}{pD + pG} \left( \frac{pD}{ps} - 1 \right) \right) - \tilde{\varepsilon},$$

a convergence rate result (7) in $H^s$ with

$$\sigma = \frac{psps - pG_sG}{pD + pG} \cdot d \left[ \frac{1}{2} - \frac{1}{pD + pG} \left( \frac{pG}{pG} + \frac{p}{ps} \right) \right] - \tilde{\varepsilon}.$$

(ii) If $p_G < p_D$, we obtain with

$$p_R = 2,$$

$$s_R = \frac{1}{2} \left( sG - \frac{d}{ps} - \left( sG - \frac{d}{pG} \right) - \tilde{\varepsilon},$$

convergence rate (7) in $B_R = H^{s_R}$.

Remark 4.3. If $p_S < \min\{p_D, p_G\}$ the convergence rate in corollary 4.2 is better than in theorem 3.2. Note that the Tikhonov functionals do not coincide.

A curiosity of theorem 3.2, i.e. of the straightforward application of the Banach space regularization results [5, 15], is that a more restrictive source condition $B_T \subset B_S$ does not necessarily enforce a better convergence rate. As the following counterexample shows, sometimes the converse may happen.
Counterexample 4.4. Let $\varepsilon > 0$ be sufficiently small and $\eta > 0$. Further let $F$ be an operator with

$$F : H^{-\eta} \rightarrow L_2, \quad \text{rg} F^* = H^\eta.$$  

(i) With the loose-source condition $u^\dagger \in B_{S} = H^\eta$ theorem 3.2 yields a convergence rate in the Lebesgue space $L_2$. (The choice $B_{S} = H^\eta$ leads to $B_{R} = L^2$.)

(ii) If we tighten the condition to $u^\dagger \in B_{S}^{\eta + d/6} \subset H^\eta$ we get the regularization space $B_{R} = B_{S}^{\eta + d/6}$ and a convergence rate in the Sobolev space $H^{-\frac{\eta}{2} + \varepsilon}$, which is larger than in $L_2$ for small $\varepsilon$.

In contrast to that the usage of Besov space embeddings, i.e. corollary 4.2, rewards a tighter source condition with a stronger convergence rate: let $B_{T} \subset B_{S}$, i.e.,

$$s_T = \frac{d}{p_T} > s_S = \frac{d}{p_S}, \quad p_T \leq p_S.$$  

Then we get consequently for case (i) ($p_G \geq p_D$)

$$\sigma(B_T) - \sigma(B_S) = \frac{p_D}{p_D + p_G} \left( s_T - \frac{d}{p_T} - \left( s_S - \frac{d}{p_S} \right) \right) > 0,$$

and for case (ii) ($p_G < p_D$)

$$\sigma(B_T) - \sigma(B_S) = \frac{1}{2} \left( s_T - \frac{d}{p_T} - \left( s_S - \frac{d}{p_S} \right) \right) > 0.$$  

5. Examples

In the following we will illustrate the convergence-rate results with a few examples. With the first one we want to show that the choice of the parameter $p$, resp., the choice of the source condition $B_{S} \subset B_{S}^p \subset B_{D}$ (cf (10)–(14)) influences the convergence rate significantly.

Example 5.1 (smoothing in the Sobolev scale). Let $d = 1, \eta > \frac{1}{2}$ and consider the operator

$$F : H^{-\eta} \rightarrow L_2, \quad \text{rg} F^* = H^\eta,$$

i.e. we consider smoothing of order $\eta$ in the Sobolev scale. Moreover, we assume that the source condition $u^\dagger \in B_{S}^{\eta} \subset H^{-\eta}$ holds. Theorem 4.1 yields convergence rates for Tikhonov penalties $\| \cdot \|_{p^\dagger(p)}$ with $p_S \leq p \leq p_G = p_D$. Since $\sigma$ is monotone in $p$, cf solution of the optimization problem, we just investigate the two boundary values here. For $p = p_S$ we get the Tikhonov functional

$$T_\alpha(u) = \| Fu - v^\delta \|^2 + \alpha \| u \|_{p_S}^2.$$  

With that worst parameter choice resp. worst source condition, theorem 4.1 yields

$$\sigma = \frac{2\eta - 2\eta + 1}{3} - \frac{2}{3} = -\frac{1}{6}$$

and hence, the convergence rate occurs in a Sobolev space $H^\eta$ with negative smoothness.

Next let us check the rate with an optimal parameter $p = p_G$. Here we get for the Tikhonov functional

$$T_\alpha(u) = \| Fu - v^\delta \|^2 + \alpha \| u \|_{p_G}^2.$$
the convergence rate in the Sobolev space $H^\sigma$ with smoothness

$$\sigma = \frac{4\eta - 2\eta}{4} + \frac{1}{2} - \frac{1}{4}(1 + 2) - \epsilon = \frac{\eta}{2} - \frac{1}{4} - \epsilon,$$

which is greater than zero for small $\epsilon$, since $\eta > \frac{1}{2}$. Hence, we get a convergence rate in a Sobolev space with positive smoothness.

The first example may lead to the conclusion that a penalty formulated in a Sobolev space gives the best convergence rate. This impression may be intensified, because the optimal source also lives in Sobolev spaces, i.e. $p = p_D = 2$. As we will see now with the following two examples with operators formulated in Banach scales, this guess is not true. Moreover, the following examples illustrate the difference between the cases $p_S = \min\{p_D, p_G\}$ and $p_S < \min\{p_D, p_G\}$. In the first case theorem 4.1 yields a convergence rate for only one Tikhonov functional resp. no optimization is possible, cf example 5.2. In the second we get a set of allowed Tikhonov penalties depending on $p$, $p_S \leq p \leq \min\{p_D, p_G\}$.

**Example 5.2** (smoothing in the Besov scale, $p_S = \min\{p_D, p_G\}$). Let $d = 1$, $\eta > 0$ arbitrary, $0 < \theta \leq 1$ small and consider the operator

$$F : B_{1+\theta}^{-\eta} \rightarrow L_2,$$

with $\operatorname{rg} F^* = B_{\frac{\theta}{\theta+1}}^{\eta}$, which models smoothing in the scale of Besov spaces. Further let $u^\dagger \in B_{1+\theta}^{-\eta+\theta}$ be the source condition. Note that $B_{1+\theta}^{-\eta+\theta} \subset B_{1+\theta}^{-\eta}$ and

$$1 + \theta \leq \frac{1 + \theta}{\theta}, \quad \text{for} \quad \theta \leq 1,$$

and hence, we can guarantee $p_S \leq p_G$. Due to $p_S = \min\{p_D, p_G\} = p_D$ it follows from theorem 4.1 that only the Tikhonov functionals with $p = p_S = p_D$, i.e.

$$T_q(u) = \| Fu - v^\delta \|_2^2 + \alpha \| u \|_{B_{\frac{\theta}{\theta+1}}}^p,$$

with $p = \frac{p_\theta p^*}{p^*} = p = \theta + 1$ and

$$s_R \leq p(-\eta + \theta) + p^*(-\eta) = -\eta + \frac{\theta^2}{\theta + 1}$$

yield a convergence rate in the Sobolev space $H^\sigma$. The maximal smoothness $\sigma$ is obtained by the penalty with $s_R = -\eta + \frac{\theta^2}{\theta + 1}$ and it reads as

$$\sigma = \frac{p(-\eta + \theta) + p^*(-\eta)}{p + p^*} + \frac{1}{2} = -\eta + \frac{1}{2} - \theta + 1.$$

To put it roughly. For the operator

$$F : B_{1}^{-\eta} \rightarrow L_2,$$

the source condition $u^\dagger \in B_1^{-\eta}$ and the Tikhonov functional

$$T_q(u) = \| Fu - v^\delta \|_2^2 + \alpha \| u \|_{B_1^{-\eta}}^p,$$

we get a convergence rate

$$\| u^{a,\delta} - u^\dagger \|_{H^{-\eta+\frac{\theta}{\theta+1}}} = O(\sqrt{\delta}).$$

In the above Besov scale example no optimization of the convergence rate was possible.

In the following example there is a set of allowed Tikhonov regularizations and hence an optimal one.
Example 5.3 (smoothing in the Besov scale, $p_S < \min\{p_D, p_G\}$). Let $d = 1, \eta > 0$ arbitrary, $0 < \theta \leq \frac{1}{2}$ small and consider the operator

$$F : B_{\frac{1}{2}}^{-\eta-\theta} \rightarrow L_2, \quad \text{with} \quad \text{rg} F^* = B_{\frac{1}{2}}^\eta = (B_{\frac{1}{2}}^{-\eta})^*.$$ 

Further let $u^\dagger \in B_{\frac{1}{2}}^{-\eta+1} \subset B_{\frac{1}{2}}^{-\eta-\theta}$ be the source condition. Note that since $B_{\frac{1}{2}} \subset B_D$ and $p_D < p_G$ we can guarantee the second assumption of theorem 4.1, $p_S < p_G$. Here theorem 4.1 yields convergence rates for a set of Tikhonov penalties with $p_S < p \leq p_D$. We will just investigate the two boundary values here again.

With the worst parameter choice, i.e. $p = p_S$, theorem 4.1 yields with the Tikhonov functional

$$T_{\alpha}(u) = \| Fu - v^\delta \|^2 + \alpha \| u \|^p_{B_{\frac{1}{2}}^\eta},$$

with $p_R = \frac{4d}{3}$ and $s_R = \frac{p_S - p_G}{p_S + p_G} = -\eta + \frac{2\theta + 1}{\theta + 4}$ a convergence rate in $H^\sigma$ with

$$\sigma = -\eta + \frac{1}{2} + \frac{\theta - 2}{\theta + 4}.$$

For the optimal parameter $p = p_D$ we get a penalty $p_R = p_D = \frac{3}{2}$ and

$$s_R = \frac{1}{p_G} s_S - \frac{1}{p_D} s_G - \left( \frac{1}{p_D + p_G} \left( \frac{p_D}{p_S} - 1 \right) \right) - \tilde{\varepsilon} = -\eta + \frac{1}{3} \cdot \frac{1}{9} \left( \frac{2\theta - 1}{\theta + 1} \right) - \tilde{\varepsilon}.$$

Theorem 4.1 yields a convergence rate with

$$\sigma = -\eta + \frac{1}{6} + \frac{1}{9} \left( \frac{2\theta - 1}{\theta + 1} \right) - \tilde{\varepsilon}$$

with small $\tilde{\varepsilon} := \varepsilon (\frac{p_D}{p_S} - 1) > 0$. To put it roughly. Consider the operator

$$F : B_{\frac{1}{2}}^{-\eta-\theta} \rightarrow L_2, \quad F^* : L_2 \rightarrow B_{\frac{1}{2}}^\eta,$$

and the source condition $u^\dagger \in B_{\frac{1}{2}}^{-\eta+1}$. For the worst choice $p = p_S$ we get for the Tikhonov functional

$$T_{\alpha}(u) = \| Fu - v^\delta \|^2 + \alpha \| u \|^p_{B_{\frac{1}{2}}^{-\eta-1}},$$

a convergence rate

$$\| u^{*,\delta} - u^\dagger \|_{H^{-\eta}} = O(\sqrt{\delta}).$$

The optimal choice $p = p_D$ yields with the Tikhonov functional

$$T_{\alpha}(u) = \| Fu - v^\delta \|^2 + \alpha \| u \|^p_{B_{\frac{1}{2}}^{-\eta-\theta}},$$

a convergence rate

$$\| u^{*,\delta} - u^\dagger \|_{H^{-\eta+1}} = O(\sqrt{\delta}).$$
6. Conclusion

The aim of this paper was to make a first attempt to analyze scales of Banach spaces for Tikhonov regularization. We used Besov spaces to model the smoothing properties of the operator, the regularization term and the source condition. The convergence rate results were obtained in the Hilbert scale of Sobolev spaces. In comparison to regularization in Hilbert scales initiated in [19] the relation between these spaces is more complicated. Of particular interest is the fact that on one hand tighter source conditions may not lead to stronger convergence rates and on the other hand a less tight source condition may result in a stronger result.

Our examples in section 5 show only slight improvements in the Sobolev exponents when the Besov-penalty is optimized. It is questionable if the effect can be observed numerically. However, the effect that looser source conditions lead to tighter convergence results is interesting in its own

We did not use Besov spaces neither for the discrepancy term in the Tikhonov functional nor to measure the convergence rate. Both points are of interest and may lead to more general results. Since this paper is a first attempt in this direction we postpone this analysis for future work.

References

[1] Bonesky T, Bredies K, Lorenz D A and Maass P 2007 A generalized conditional gradient method for nonlinear operator equations with sparsity constraints Inverse Problems 23 2041–58
[2] Bredies K 2008 A forward–backward splitting algorithm for the minimization of non-smooth convex functionals in Banach space Preprint arXiv:0807.0778
[3] Bredies K and Lorenz D A 2008 Iterated hard shrinkage for minimization problems with sparsity constraints SIAM J. Sci. Comput. 30 657–83
[4] Bredies K and Lorenz D A 2008 Linear convergence of iterated soft-thresholding J. Fourier Anal. Appl. at press (arXiv:0709.1598)
[5] Burger M and Osher S 2004 Convergence rates of convex variational regularization Inverse Problems 20 1411–20
[6] Burger M, Resmerita E and He L 2007 Error estimation for Bregman iterations and inverse scale space methods in image restoration Computing 81 109–35
[7] Chambolle A, DeVore R A, Lee N and Lucier B J 1998 Nonlinear wavelet image processing: variational problems, compression and noise removal through wavelet shrinkage IEEE Trans. Image Process. 7 319–35
[8] Chan T F and Shen J 2005 Image Processing and Analysis—Variational, PDE, Wavelet, and Stochastic Methods (Philadelphia, PA: SIAM)
[9] Cohen A 2003 Numerical Analysis of Wavelet Methods (Amsterdam: Elsevier)
[10] Dahlke S, Maass P, Teschke G, Koch K, Lorenz D A, Müller S, Schiffler S, Stämpfli A, Thiele H and Werner M 2008 Multiscale approximation Mathematical Methods in Signal Processing and Digital Image Analysis ed R Dahlhaus, J Kurths, P Maass and J Timmer (Berlin: Springer) pp 75–109
[11] Daubechies I, Defrise M and De Mol C 2004 An iterative thresholding algorithm for linear inverse problems with a sparsity constraint Commun. Pure Appl. Math. 57 1413–57
[12] DeVore R A 1998 Nonlinear approximation Acta Numer. 7 51–150
[13] Griesse R and Lorenz D A 2008 A semismooth Newton method for Tikhonov functionals with sparsity constraints Inverse Problems 24 035007 (19 pp)
[14] Hein T 2008 Convergence rates for regularization of ill-posed problems in Banach spaces by approximate source conditions Inverse Problems 24 045007
[15] Hofmann B, Kaltenbacher B, Poeschl C and Scherzer O 2007 A convergence rate result for Tikhonov regularization in Banach spaces with non-smooth operators Inverse Problems 23 987–1010
[16] Klaar E, Kuhn M, Lorenz D A, Maass P and Thiele H 2007 Shrinkage versus deconvolution Inverse Problems 23 2231–48
[17] Lorenz D A 2007 Solving variational methods in image processing via projections—a common view on TV-denoising and wavelet shrinkage Z. Angew. Math. Phys. 87 247–56
[18] Lorenz D A 2008 Convergence rates and source conditions for Tikhonov regularization with sparsity constraints J. Inverse and Ill-Posed Problems at press (arXiv:0801.1774)

[19] Natterer F 1984 Error bounds for Tikhonov regularization in Hilbert scales Appl. Anal. 18 29–37

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

[21] Ramlau R and Teschke G 2006 A Tikhonov-based projection iteration for nonlinear ill-posed problems with sparsity constraints Numer. Math. 104 177–203

[22] Resmerita E 2005 Regularization of ill-posed problems in Banach spaces: convergence rates Inverse Problems 21 1303–14

[23] Resmerita E and Scherzer O 2006 Error estimates for non-quadratic regularization and the relation to enhancement Inverse Problems 22 801–14

[24] Rudin L, Osher S J and Fatemi E 1992 Nonlinear total variation based noise removal algorithms Physica D 60 259–68

[25] Runst T and Sickel W 1996 Sobolev Spaces of Fractional Order, Nemytskij Operators, and Nonlinear Partial Differential Equations (de Gruyter Series in Nonlinear Analysis and Applications) (Berlin: Walter de Gruyter)

[26] Schöpfer F, Louis A K and Schuster T 2006 Nonlinear iterative methods for linear ill-posed problems in Banach spaces Inverse Problems 22 311–29

[27] Schuster T, Maass P, Bonesky T, Kazimierski K S and Schöpfer F 2008 Minimization of Tikhonov functionals in Banach spaces Abstract Appl. Anal. 192679 (19 pp)

[28] Vogel C R and Acar R 1994 Analysis of bounded variation penalty methods for ill-posed problems Inverse Problems 10 1217–29