Variational regularization theory for sparsity promoting wavelet regularization

Dissertation
zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades
„Doctor rerum naturalium“
der Georg-August-Universität Göttingen

im Promotionsprogramm „Mathematical Science“
der Georg-August University School of Science (GAUSS)

vorgelegt von
Philip Miller
aus Gießen

Göttingen 2021

Betreuungsausschuss

Prof. Dr. Thorsten Hohage,
Institut für Numerische und Angewandte Mathematik,
Georg-August-Universität Göttingen

Prof. Dr. Gerlind Plonka-Hoch,
Institut für Numerische und Angewandte Mathematik,
Georg-August-Universität Göttingen

Prof. Dr. Frank Werner,
Institut für Mathematik,
Julius-Maximilians-Universität Würzburg

Mitglieder der Prüfungskommission

Referent: Prof. Dr. Thorsten Hohage

Korreferent: Prof. Dr. Frank Werner

Weitere Mitglieder der Prüfungskommission:

Prof. Dr. Stephan Huckemann,
Institut für Mathematische Stochastik,
Georg-August-Universität Göttingen

Prof. Dr. Russell Luke,
Institut für Numerische und Angewandte Mathematik,
Georg-August-Universität Göttingen

Prof. Dr. Axel Munk,
Institut für Mathematische Stochastik,
Georg-August-Universität Göttingen

Tag der mündlichen Prüfung: 05. August 2021

Acknowledgements

Contents

Chapter One

Introduction

Inverse problems, ill-posedness and regularization

In many scientific and industrial applications the quantity of interest is not directly observable but rather a causal effect or a parameter producing observations measured in an experiment. The goal is to calculate stable approximations of these effects or parameters from given noisy observations. For example, in computed tomography the goal is to calculate images of parts of the interior of a human body from X-ray measurements from different angles. Another concrete example is to compute the location of a sound source in space using microphone arrays. As an another example, measuring the present state of a physical system one wants to draw conclusions on the past.

Usually it is easier to solve the *forward problem*, which consists of predicting the observations from a given cause or parameter. We model the forward problem as a possibly nonlinear mapping $F \colon \operatorname{dom}(F) \to \mathbb{Y}$ between Banach spaces $\mathbb{X}, \mathbb{Y}$ assigning to each cause or parameter $f \in \operatorname{dom}(F) \subset \mathbb{X}$ an observation $F(f) \in \mathbb{Y}$. Hence, the corresponding inverse problem appears as the task of solving the equation $F(f) = g$ for given data $g \in \mathbb{Y}$.

The main challenge in this task is usually *ill-posedness* of the inverse problem. This means, that the unknown f does not continuously depend on the observations g: For F continuous and injective the inverse $F^{-1} \colon \operatorname{im}(F) \to \operatorname{dom}(F)$ exists, but it may lack continuity. For example in the important case of a linear, injective, and compact forward operator F with dense range and $\dim(\operatorname{im}(F)) = \infty$ the inverse operator F^{-1} exists on a dense subset of $\mathbb{Y}$, but fails to be continuous as otherwise $F^{-1}F = \operatorname{Id}_{\mathbb{X}}$ would be compact contradicting $\dim(\mathbb{X}) = \dim(\operatorname{im}(F)) = \infty$.

Due to ill-posedness, a naive application of the inverse of F (if it exists) may therefore amplify noise in the data g indefinitely. *Regularization methods* are used to calculate stable approximations of the true solution f anyways. The basic idea is to use a family of continuous maps $\mathbb{Y} \to \mathbb{X}$ approximating the discontinuous inverse F^{-1}. This is particularly relevant if only noisy measurements $g^{\mathrm{obs}} \in \mathbb{Y}$ of $F(f)$ are available, but due to numerical errors regularization is even necessary if exact data are available as in optimal control problems.

The purpose of this work is to contribute to the mathematical analysis of *generalized Tikhonov regularization* and in particular *sparsity promoting Tikhonov regularization* which are popular examples of regularization methods.

Classical and sparsity promoting Tikhonov regularization

Suppose for a moment that $\mathbb{X}$ and $\mathbb{Y}$ are Hilbert spaces and $F = A \colon \mathbb{X} \to \mathbb{Y}$ is linear and bounded. For noisy observations $g^{\mathrm{obs}} \in \mathbb{Y}$ of Af and $\alpha > 0$ *classical Tikhonov regularization* approximates f by $\hat{f}_\alpha = (A^*A + \alpha I)^{-1} A^* g^{\mathrm{obs}}$. Here the family $(A^*A + \alpha I)^{-1} A^*$ with $\alpha > 0$ is used to approximate A^{-1} in a stable manner. The estimator can also be characterized variationally by

$$\hat{f}_\alpha \in \operatorname*{argmin}_{h \in \mathbb{X}} \left(\frac{1}{2\alpha} \| g^{\mathrm{obs}} - Ah \|_{\mathbb{Y}}^2 + \frac{1}{2} \| h \|_{\mathbb{X}}^2 \right), \tag{1.1}$$

namely as the unique minimizer of the *Tikhonov functional* $\frac{1}{2\alpha} \| g^{\mathrm{obs}} - A \cdot \|_{\mathbb{Y}}^2 + \| \cdot \|_{\mathbb{X}}^2$. The *data fidelity term* $\frac{1}{2} \| g^{\mathrm{obs}} - A \cdot \|_{\mathbb{Y}}^2$ ensures a good approximation of the observation in the image space $\mathbb{Y}$. The *penalty term* $\| \cdot \|_{\mathbb{X}}^2$ stabilizes the reconstruction, and by choosing $\mathbb{X}$ as a space of functions of a certain regularity it incorporates a priori knowledge on the unknown f. The *regularization parameter* α controls the balance between these two terms. The smaller α is chosen, the better $A\hat{f}_\alpha$ approximates g^{obs}, but the more sensitive the reconstruction is to errors in the observed data g^{obs}.

A major disadvantage of classical Tikhonov regularization is the regularity of the estimators $\hat{f}_\alpha$: In many applications the operator A and the adjoint A^* are smoothing integral operators. For instance with $A \colon L^2 \to L^2$ one may have $\operatorname{im}(A^*) = H^s$ for some $s > \frac{1}{2}$. As the reconstruction $\hat{f}_\alpha$ always belongs to $\operatorname{im}(A^*)$ we obtain a continuous estimate $\hat{f}_\alpha$ even if the true unknown function f has jumps. Hence, classical Tikhonov regularization provides good recovery of smooth functions, but it may fail recovering functions with discontinuities. A way to overcome this problems is to use Banach instead of Hilbert space penalties. For example, evoking a wavelet basis and penalties given by wavelet coefficients one achieves good localization in spacial and the frequency domain and estimators more suitable for the reconstruction of functions with singularities (see [90, Sec. 1] and [9, Sec. 5.5]). Using
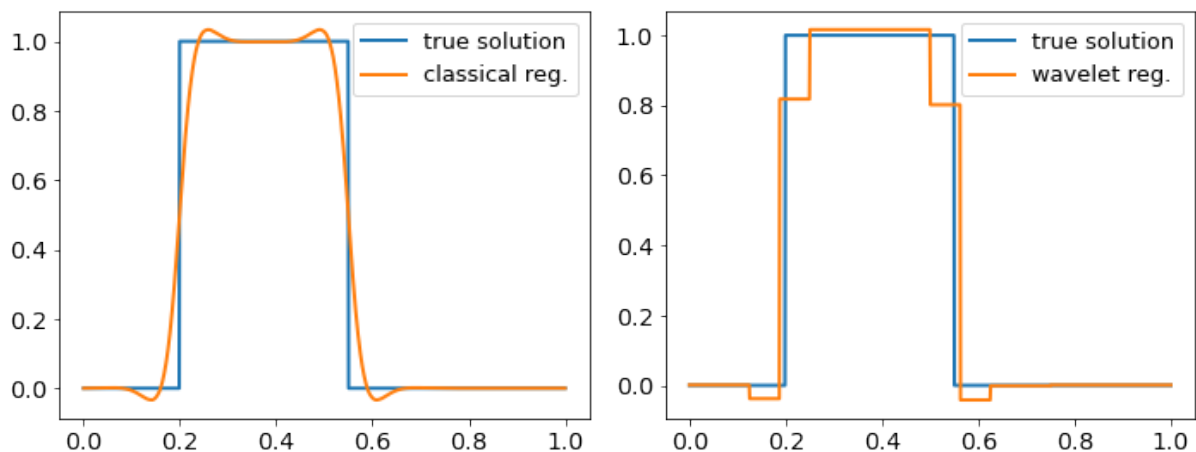

FIGURE 1.1: Noiseless reconstruction of a function with jumps. The left image show the continuous estimator given by classical Tikhonov regularization. The right image shows an estimator by penalization of Haar wavelet coefficients.

an ℓ^1- (or more generally a block ℓ^1-) norm penalty enforces the estimator to be *sparse*

which means that it consists of only finitely many non-vanishing wavelet components. For this reason such schemes are called *sparsity promoting*, and the penalization using 1-norms is referred to as *sparsity constraint.*

Sparsity promoting regularization has been studied intensively. We refer the reader to the survey paper [68] and to the special issue [69] on sparsity regularization for an overview on the use of these methods for a variety of linear and nonlinear inverse problems in different fields of applications. Moreover, the interest in sparsity and wavelet methods and their application in natural science is reflected in the conference series *Wavelets and Sparsity* (see, e.g., the proceeding volume [102]).

Aims and overview of this work

A central task in regularization theory is to provide upper bounds on the distance between the unknown f and the estimators given by a regularization method. The asymptotic behavior of these upper bounds as the noise level tends to zero is called *convergence rate*. However, for ill-posed problems we cannot achieve uniform upper bounds on the domain $\text{dom}(F)$ (see Proposition 2.4). Therefore, upper bounds on the error require regularity conditions on the true solution, which are referred to as *source conditions.*

The aim of this work is to find (quasi-) Banach spaces $\mathbb{X}_s \subset \mathbb{X}$ such that $f \in \mathbb{X}_s \cap \text{dom}(F)$ implies (or even characterizes) convergence rates of a given order. Here s denotes a smoothness index in a scale of function spaces. In simpler words, we seek practical conditions on the unknown in terms of regularity or smoothness measured in terms of function spaces implying (or even characterizing) bounds on the reconstruction error.

To this end, we use penalties given by powers of Banach space norms and formulate assumptions on the forward operator. Most often the latter are Lipschitz conditions in certain function spaces with negative smoothness index. In particular, we focus on convergence rates for sparsity promoting wavelet regularization. Even though the convergence rates theory for ℓ^1-regularization, which we discuss in Section 2.3, is far developed, there are only few results showing concrete convergence rates for non-sparse unknowns. For example with δ denoting the noise level in a deterministic noise model, Hölder type rates $\mathcal{O}(\delta^\nu)$ as $\delta \to 0$ for $0 < \nu < 1$ where ν depends on the smoothness of the unknown measured in some function space are desirable. We will prove such convergence rates for wavelet regularization with Besov space penalties.

This thesis is organized as follows: In Chapter 2 we recall basic regularization theory, motivate and introduce the mathematical framework of generalized Tikhonov regularization and sparsity promoting wavelet regularization with Besov space penalties. We formulate an assumption on the forward operator and discuss its verification in several relevant examples.

The main contribution in Chapter 3 are convergence rate results for sparsity promoting wavelet regularization using variational source conditions as an intermediate step, i.e. characterizations of the smoothness of the unknown in terms of variational inequalities. Such conditions have become popular in regularization theory over the last decade. More precisely, we provide order optimal convergence rates for wavelet regularization with Besov $0, p, 1$-penalties. We discuss the verification of variational source conditions and compare our strategies to the literature on ℓ^1-regularization. Moreover, we demonstrate

by numerical simulations for a parameter identification problem in a differential equation that our theoretical results correctly predict rates of convergence for piecewise smooth unknown coefficients. Besides finitely smoothing operators, we also provide results for exponentially ill-posed problems that furnish convergence rates for the backward heat equation. Furthermore, we address statistical noise models and prove optimal convergence rates for Gaussian white noise.

Converse results, i.e. necessity of certain source conditions, are investigated in Chapter 4. Here we first treat the sparsity promoting regularization schemes considered in Chapter 3 and furnish converse results with respect to approximation rates for exact data. After that we show equivalence of certain variational source conditions and approximation rates in the image space for general convex variational regularization.

Chapter 5 is devoted to oversmoothing in Banach space regularization, i.e. to the case that the unknown solution fails to admit a finite value in the penalty term. Here we present a general theory and demonstrate its flexibility by providing convergence rates results not only for Besov space regularization but also for bounded variation regularization, that is Tikhonov regularization with a BV-norm penalty.

In Chapter 6 we investigate image space approximation rates as new type of source conditions. We prove a nearly minimax theorem showing that up to a constant the modulus of continuity is an upper bound on the reconstruction error. One main advantage of our analysis is the flexibility in the choice of the loss function. The application to Besov space regularization leads to convergence rate results for $0, 2, q$- and $0, p, p$-penalties without restrictions on $p, q \in (1, \infty)$.

Finally, we discuss our results and point out some potential generalizations and open problems in Chapter 7.

Chapter Two

Generalized Tikhonov and wavelet regularization

In this chapter we provide the mathematical framework of this thesis. In the first section we motivate the use of more general penalties in the Tikhonov functional by recalling that ℓ^1-penalization enforces sparsity. After that we introduce generalized Tikhonov regularization with convex penalty terms and provide topological assumptions that ensure the existence of minimizers. In the second section we recall basics from convergence rate theory and the concept of order optimality. In Section 2.3 we provide an overview on the convergence rate theory for sparsity promoting regularization. The following section is devoted to wavelet regularization in Besov spaces. Here we formulate and discuss assumptions on the wavelet system and on the forward operator. At the end of this chapter we briefly touch upon regularization with weighted ℓ^p-penalties.

2.1 Generalized Tikhonov regularization

Sparsity promoting regularization

To further motivate the use of more general penalty terms than squared Hilbert space norms as in (1.1), we compare classical Tikhonov regularization with regularization using a penalty term which is given by the ℓ^1-norm of coefficients in a basis expansion.

Before we do so, let us clarify the following standard notation: For an index set Λ and $p \in (0, \infty]$ we denote $\ell^p(\Lambda) = \{x \in \mathbb{R}^\Lambda \colon \|x\|_p < \infty\}$ with $\|x\|_p^p = \sum_{j \in \Lambda} |x_j|^p$ for $p < \infty$ and $\|x\|_\infty = \sup_{j \in \Lambda} |x_j|$. Moreover, we set $\ell^p := \ell^p(\mathbb{N})$.

Now let, as above, $A \colon \mathbb{X} \to \mathbb{Y}$ be linear and bounded between Hilbert spaces $\mathbb{X}$ and $\mathbb{Y}$. Suppose $\mathbb{X}$ is separable, and $(\psi_n)_{n \in \mathbb{N}}$ is an orthonormal basis in $\mathbb{X}$. Let $\mathcal{S} \colon \ell^2 \to \mathbb{X}$ be the natural unitary operator given by $\mathcal{S} e_n = \psi_n$ for all $n \in \mathbb{N}$ with $(e_n)_{n \in \mathbb{N}}$ the canonical basis of ℓ^2 given by $(e_n)_k = \delta_{nk}$. It is easy to see that the sequence of basis coefficients $\hat{x}_\alpha^{(2)} = \mathcal{S}^{-1} \hat{f}_\alpha$ of $\hat{f}_\alpha$ given by (1.1) minimizes the functional $z \mapsto \frac{1}{2\alpha} \|g^{\mathrm{obs}} - A\mathcal{S}z\|_{\mathbb{Y}}^2 + \frac{1}{2} \|z\|_2^2$ on ℓ^2. Replacing the penalty term $\frac{1}{2} \| \cdot \|_2$ by the norm $\|z\|_1 := \sum_{j=1}^\infty |z_j|$ on the space $\ell^1 \subset \ell^2$ of summable real sequences, we consider

$$\hat{x}_\alpha^{(1)} \in \operatorname*{argmin}_{z \in \ell^1} \left(\frac{1}{2\alpha} \|g^{\mathrm{obs}} - A\mathcal{S}z\|_{\mathbb{Y}}^2 + \alpha \|z\|_1 \right) \tag{2.1}$$

and use $\mathcal{S}\hat{x}_\alpha^{(1)}$ as an estimator in $\mathbb{X}$. This procedure increases the penalization on coefficients with absolute value less than 1 while decreasing it on those with absolute value greater

than 1. Hence, using the 1-norm we put a higher penalty on functions with many small coefficients (see [29, Sec. 1.3] for a similar explanation). We show that this leads to estimators $\hat{x}_\alpha^{(1)}$ consisting of only finitely many non-vanishing components. As already mentioned in the introduction this is the reason why schemes using 1-norms as penalty terms are called *sparsity promoting*. The following result can also be found in [78, Lem. 2.1] (see also [37, Prop. 2.2 (i)] for a slightly more general version).

Proposition 2.1 (Sparsity of the minimizers). *Suppose $\hat{x}_\alpha^{(1)} \in \ell^1$ is given by (2.1). Then there exists $n \in \mathbb{N}$ such that $\left(\hat{x}_\alpha^{(1)}\right)_j = 0$ for all $j \geq n$.*

Proof. Recall that for an element $x \in \ell^1$ the convex subdifferential of the ℓ^1-norm is given by

$$\partial \| \cdot \|_1 (x) = \left\{ \xi \in \ell^\infty \colon \xi_j = 1 \text{ if } x_j > 0 \,, \xi_j = -1 \text{ if } x_j < 0 \text{ and } \xi \in [-1, 1] \text{ if } x_j = 0 \right\}.$$

The first order optimality condition and the sum and chain rules for convex functionals (see [34, Ch. 1.5.]) yield $\xi \in \partial \| \cdot \|_1 (\hat{x}_\alpha^{(1)})$ with $\xi = \frac{1}{\alpha} (A\mathcal{S})^* A\mathcal{S} \left(g^{\mathrm{obs}} - A\hat{x}_\alpha^{(1)} \right)$. Since $\mathrm{im}((A\mathcal{S})^*) \subset \ell^2$, the sequence ξ convergences to 0. Hence, there exists $n \in \mathbb{N}$ such that $|\xi_j| < 1$ and therefore $\left(\hat{x}_\alpha^{(1)}\right)_j = 0$ for all $j \geq n$. $\qquad\square$

In contrast, for classical Tikhonov regularization we cannot expect that $\hat{f}_\alpha$ has only finitely many non vanishing coefficients in general. To see this and to provide some intuition, we study a positive diagonal operator where we can explicitly calculate the estimators for the ℓ^1- and ℓ^2-penalties. To this end, suppose $\mathbb{X} = \mathbb{Y}$ and that there exists a sequence $(\nu_n)_{n \in \mathbb{N}}$ of positive real numbers such that $A\psi_n = \nu_n \psi_n$.
Inserting the unique expansion $g^{\mathrm{obs}} = \sum_{n=1}^\infty g_n \psi_n$, we calculate the coordinates of the minimizer $\hat{f}_\alpha = \sum_{n=1}^\infty \left(\hat{f}_\alpha\right)_n \psi_n$ in (1.1) separately and obtain $\left(\hat{f}_\alpha\right)_n = \frac{\nu_n}{\alpha + \nu_n^2} g_n$. Hence, the support of $\hat{f}_\alpha$ equals the support of g^{obs}.
On the other hand, a routine calculation shows that the minimizer $\hat{x}_\alpha^{(1)}$ in (2.1) has the following components

$$\left(\hat{x}_\alpha^{(1)}\right)_n = \begin{cases} \frac{1}{\nu_n} \left(g_n - \frac{\alpha}{\nu_n}\right) & \text{if } g_n > \frac{\alpha}{\nu_n}, \\ \frac{1}{\nu_n} \left(g_n + \frac{\alpha}{\nu_n}\right) & \text{if } g_n < -\frac{\alpha}{\nu_n}, \\ 0 & \text{if } |g_n| \leq \frac{\alpha}{\nu_n}. \end{cases}$$

Hence, in this case the estimator $\hat{x}_\alpha^{(1)}$ is given by a weighed soft thresholding of g^{obs} which explains once more the finite support, i.e. the sparsity of $\hat{x}_\alpha^{(1)}$.

In the early and influential work [29] the authors establish convergence of estimators with weighed ℓ^1-penalties and propose an iterative thresholding algorithm to compute those. In [7], Beck and Teboulle suggested and analyzed the so-called *Fast Iterative Shrinkage-Thresholding Algorithm (FISTA)* that is based on Nesterov acceleration. Further numerical schemes can be found in the overview paper [16].

Tikhonov regularization with general convex penalty

Motivated by the exposition of sparsity promoting regularization above, we study a generalization of classical Tikhonov regularization with general convex penalties. We prove

existence of minimizers in a framework that grasps the more concrete setups in the sequel. To keep the formulation of the results in this work as self-contained as possible, we avoid standing assumptions. However, we set up assumptions from the text in the following way and refer to them whenever needed.

Assumption 2.2. *Let $\mathbb{X}$ and $\mathbb{Y}$ be Banach spaces. Moreover, let $\tau_{\mathbb{X}}$ be a topology on $\mathbb{X}$ such that $(\mathbb{X}, \tau_{\mathbb{X}})$ is a locally convex Hausdorff space and $\mathcal{R} \colon \mathbb{X} \to (-\infty, \infty]$ a proper, convex function such that the sublevel sets $\{f \in \mathbb{X} \colon \mathcal{R}(f) \leq \lambda\}$ are $\tau_{\mathbb{X}}$-compact for all $\lambda \in \mathbb{R}$. Suppose $F \colon \operatorname{dom}(F) \to \mathbb{Y}$ is a $\tau_{\mathbb{X}}$-to-weak continuous map defined on a $\tau_{\mathbb{X}}$-closed subset $\operatorname{dom}(F) \subset X$ with $\operatorname{dom}(\mathcal{R}) \cap \operatorname{dom}(F) \neq \emptyset$.*

For $\alpha > 0$, $f \in \operatorname{dom}(F)$ and $g \in \mathbb{Y}$ we consider the Tikhonov functional given by

$$T_\alpha(f, g) := \frac{1}{2\alpha} \|g - F(f)\|_{\mathbb{Y}}^2 + \mathcal{R}(f)$$

and denote its set of minimizers by

$$S_\alpha(g) := \operatorname*{argmin}_{f \in \operatorname{dom}(F)} T_\alpha(f, g) \subseteq \operatorname{dom}(\mathcal{R}) \cap \operatorname{dom}(F). \tag{2.2}$$

The next result shows existence of minimizers (see also [93, Thm. 3.22]).

Proposition 2.3 (Existence of minimizers)**.** *Suppose Assumption 2.2 holds true. Let $g \in \mathbb{Y}$. Then $S_\alpha(g)$ is nonempty. Moreover, if $g \in \operatorname{im}(F)$, then there exists a (possibly not unique) $\mathcal{R}$-minimal $f \in \operatorname{dom}(F)$ with $F(f) = g$, i.e. $\mathcal{R}(f) \leq \mathcal{R}(h)$ for all $h \in \operatorname{dom}(F)$ with $F(h) = g$.*
If F is linear and injective with $\operatorname{dom}(F) = \mathbb{X}$, then $S_\alpha(g)$ is a singleton.

Proof. The function $\operatorname{dom}(F) \to \mathbb{Y}$ given by $f \mapsto g - F(f)$ is $\tau_{\mathbb{X}}$-to-weak continuous and $\mathbb{Y} \to [0, \infty)$ given by $y \mapsto \frac{1}{2\alpha}\|y\|^2$ is weakly lower semi-continuous. $\tau_{\mathbb{X}}$-lower semi-continuity of $\mathcal{R}$ follows from the $\tau_{\mathbb{X}}$-compactness of its level sets. Hence, $T_\alpha(\cdot, g) \colon \operatorname{dom}(F) \to (-\infty, \infty]$ is $\tau_{\mathbb{X}}$-lower semi-continuous. Therefore, the sublevel set $L_\lambda := \{f \in \operatorname{dom}(F) \colon T(f, g) \leq \lambda\}$ is $\tau_{\mathbb{X}}$-closed in $\operatorname{dom}(F)$ for all $\lambda \in \mathbb{R}$. As $\operatorname{dom}(F)$ is $\tau_{\mathbb{X}}$-closed in $\mathbb{X}$, the sets L_λ are also $\tau_{\mathbb{X}}$-closed in $\mathbb{X}$. Moreover, we have $L_\lambda \subset \{f \in \mathbb{X} \colon \mathcal{R}(f) \leq \lambda\}$. Hence, L_λ is compact. Using the finite intersection property, $\bigcap_{\lambda \in \mathbb{R}} L_\lambda = \emptyset$ and $L_{\lambda_1} \subset L_{\lambda_2}$ for $\lambda_1 \leq \lambda_2$ implies the existence of $\lambda \in \mathbb{R}$ with $L_\lambda = \emptyset$. Hence, $m := \inf_{f \in \operatorname{dom}(F)} T_\alpha(f, g) > -\infty$. By the assumption $\operatorname{dom}(\mathcal{R}) \cap \operatorname{dom}(F) \neq \emptyset$ we have $m < \infty$. If $\lambda > m$, then $L_\lambda \neq \emptyset$. Therefore, we obtain $S_\alpha(g) = \bigcap_{\lambda > m} L_\lambda \neq \emptyset$ using the finite intersection property once again.
We turn to the proof of the second statement. The set $F^{-1}(\{g\})$ is $\tau_{\mathbb{X}}$-closed in $\operatorname{dom}(F)$ and therefore in $\mathbb{X}$. If $\mathcal{R}(f) = \infty$ for all $f \in \operatorname{dom}(F)$ with $F(f) = g$, then the statement holds true as every element in $F^{-1}(\{g\})$ is $\mathcal{R}$-minimal in this case. Otherwise, the application of the first part to the restriction of F to $F^{-1}(\{g\})$ yields the claim.
The uniqueness statement in the linear, injective case follows from strict convexity of the function $\mathbb{R} \to \mathbb{R}$, $\tau \mapsto \frac{1}{2\alpha}\tau^2$. $\qquad\square$

The above existence result tells us that the method introduced in (2.2) is *well-defined*. Typically *stability* and *convergence* are investigated for inversion methods like S_α. Stability means that the regularized solutions $\hat{f}_\alpha \in S_\alpha(g^{\mathrm{obs}})$ depend (sequentially) continuously on the observed data $g^{\mathrm{obs}} \in \mathbb{Y}$. For a true solution $f \in \operatorname{dom}(F)$ convergence in this

context stands for the convergence of $\hat{f}_\alpha \in S_\alpha(g^{\mathrm{obs}})$ to f whenever the observed data g^{obs} convergences to the true data $F(f)$ and the regularization parameter α is chosen in an appropriate manner. The combination of the three properties, namely well-definedness, stability and convergence, is referred to as *regularizing property.*

As this work is mainly focused on statements on the speed of convergence, the latter two properties are of less significance in the sequel. Nevertheless, we refer to [93, Sec. 3.2] for a treatise of the regularizing property in a setting that is similar to the one in Assumption 2.2.

2.2   Convergence rate theory and source conditions

In the setting of Assumption 2.2 let $f \in \mathrm{dom}(F)$, $\delta > 0$ and $g^{\mathrm{obs}} \in \mathbb{Y}$ an observation satisfying the *deterministic error model* $\|g^{\mathrm{obs}} - F(f)\|_{\mathbb{Y}} \le \delta$. A central aim of regularization theory and of the present work is to find upper bounds on $\|f - \hat{f}_\alpha\|_{\mathbb{X}}$ with $\hat{f}_\alpha \in S_\alpha(g^{\mathrm{obs}})$ in terms of the *noise level δ* under specified parameter choice rules $\alpha = \overline{\alpha}(\delta, g^{\mathrm{obs}})$. As already mentioned in the introduction, such an upper bound which converges to 0 as $\delta \to 0$ is also referred to as a *convergence rate.*

The following abstract result shows that for an ill-posed problem no reconstruction method $R(\delta, g^{\mathrm{obs}}) \in \mathbb{X}$ satisfies a uniform error bound on $\mathrm{dom}(F)$ (see also [35, Prop. 3.11]).

Proposition 2.4. *Suppose $F \colon \mathrm{dom}(F) \to \mathbb{Y}$ is an injective map between Banach spaces $\mathbb{X}$ and $\mathbb{Y}$ and $R \colon (0, \infty) \times \mathbb{Y} \to \mathbb{X}$ is a map. For $\delta > 0$ and a subset $K \subset \mathrm{dom}(F)$ we consider the* worst case error *of R on K given by*

$$\Delta_R(\delta, K) := \sup \left\{ \|f - R(\delta, g^{\mathrm{obs}})\|_{\mathbb{X}} \colon f \in K, g^{\mathrm{obs}} \in \mathbb{Y} \text{ with } \|g^{\mathrm{obs}} - F(f)\|_{\mathbb{Y}} \le \delta \right\}. \quad (2.3)$$

If there exists $\psi \colon (0, \infty) \to (0, \infty)$ with $\psi(\delta) \to 0$ for $\delta \to 0$ and $\Delta_R(\delta, K) \le \psi(\delta)$, then $F^{-1} \colon F(K) \to K$ is uniformly norm-continuous.

Proof. Let $\varepsilon > 0$. There exists $\delta > 0$ such that $\psi(\delta) < \frac{\varepsilon}{2}$. For $g_1, g_2 \in F(K)$ with $\|g_1 - g_2\|_{\mathbb{Y}} \le \delta$ we estimate

$$\|F^{-1}(g_1) - F^{-1}(g_2)\|_{\mathbb{X}} \le \|F^{-1}(g_1) - R(\delta, g_1)\|_{\mathbb{X}} + \|R(\delta, g_1) - F^{-1}(g_2)\|_{\mathbb{X}} \le 2\psi(\delta) < \varepsilon. \;\; \square$$

For an ill-posed operator equation $F(f) = g$, this result shows that we can only hope for error bounds tending to 0 for small noise levels uniformly on sets $K \subsetneq \mathrm{dom}(F)$. In other words, bounds on $\|f - \hat{f}_\alpha\|_{\mathbb{X}}$ require additional assumptions on the true solution f. Those assumptions are referred to as *regularity conditions* or *source conditions* in regularization theory. The name source condition comes from the first such conditions for linear forward operators A between Hilbert spaces: $f = (A^*A)^{\nu/2}\omega$ for some $\nu > 0$, where $\omega \in \mathbb{X}$ is referred to as source generating f. This condition implies the convergence rate $\|f - \hat{f}_\alpha\|_{\mathbb{X}} = \mathcal{O}(\delta^{\frac{\nu}{\nu+1}})$ in the Hilbert space norm that defines the penalty. In [81] convergence rates in Hilbert scales are proven under source conditions of the form $f = \varphi(A^*A)\omega$ for more general functions φ. A generalization of the above source condition for $\nu = 1$ to convex or Banach space penalties is given by *source-wise representations*

$$A^*\omega \in \partial\mathcal{R}(f) \quad \text{for some } \omega \in \mathbb{Y}. \quad (2.4)$$

The latter lead to the convergence rate $\mathcal{O}(\delta)$ in the Bregman divergence of $\mathcal{R}$ (see [13]). We refer to [92] for an analysis that uses the stronger condition $A^*A\omega \in \partial\mathcal{R}(f)$.

Slower rates of convergence in Banach space settings can be shown under *variational source conditions*. We provide details on these conditions and their verification in Chapter 3. Further types are *approximate source conditions* [55, 56] and *approximate variational source conditions* [42] . We refer to [36] for a comparison of those and variational source conditions.

Recently, in [51] convergence rates are shown under the condition $(A^*A)^\nu\omega \in \partial\mathcal{R}(f)$ for convex penalties defined on Hilbert spaces. Upper bounds on $T_\alpha(f, Af) - T_\alpha(f_\alpha, Af)$ in terms of α with $f_\alpha \in S_\alpha(Af)$ (*defect of the Tikhonov functional*) are used in [59] as a regularity condition. In Chapter 6 Hölder-type upper bounds in terms of α on $\|Af - Af_\alpha\|$ will play the role of source conditions.

Optimality

A source condition corresponds to a subset $K \subset \mathrm{dom}(F)$ consisting of all elements in $\mathrm{dom}(F)$ for which the condition is satisfied. In this sense a convergence rate result under the source condition is a bound on the worst case error $\Delta_{R_{\mathrm{Tik}}}(\delta, K)$ given by (2.3) with $R_{\mathrm{Tik}} \colon (0, \infty) \times \mathbb{Y} \to \mathbb{X}$ satisfying $R_{\mathrm{Tik}}(\delta, g^{\mathrm{obs}}) \in S_{\overline{\alpha}(\delta, g^{\mathrm{obs}})}(g^{\mathrm{obs}})$ (see (2.2)) for a prescribed choice of the regularization parameter $\overline{\alpha} \colon (0, \infty) \times \mathbb{Y} \to (0, \infty)$. In this subsection we present the concept of order optimality of such a convergence rate.

Definition 2.5 (Order optimality). *Suppose $F \colon \mathrm{dom}(F) \to \mathbb{Y}$ is an injective map between Banach spaces $\mathbb{X}$ and $\mathbb{Y}$. A map $R \colon (0, \infty) \times \mathbb{Y} \to \mathbb{X}$ is called an* order optimal reconstruction *on $K \subset \mathrm{dom}(F)$ if there exists $C > 0$ and $\delta_0 > 0$ such that*

$$\Delta_R(\delta, K) \le C \inf \left\{ \Delta_{R'}(\delta, K) \colon R' \text{ is a map } (0, \infty) \times \mathbb{Y} \to \mathbb{X} \right\} \quad \text{for all } 0 < \delta < \delta_0$$

with $\Delta_{R'}(\delta, K)$ defined in (2.3).

In [83, Ch. 1] the infimum on the right hand side is called *intrinsic error*. In [33, Sec. 4.3.1.] the term order optimal is referred to as *nearly minimax*. We recall that the *modulus of continuity* of F on a subset $K \subset \mathrm{dom}(F)$ given by

$$\omega(\delta, K) := \sup \left\{ \|f_1 - f_2\|_{\mathbb{X}} \colon f_1, f_2 \in K \quad \text{with } \|F(f_1) - F(f_2)\|_{\mathbb{Y}} \le \delta \right\} \qquad (2.5)$$

provides a lower bound on $\Delta_R(\delta, K)$ that is independent of the reconstruction method R.

Proposition 2.6 (Universal lower bound ([35, Rem. 3.12], [104, Lemma 3.11])). *Suppose $F \colon \mathrm{dom}(F) \to \mathbb{Y}$ is an injective map between Banach spaces $\mathbb{X}$ and $\mathbb{Y}$, $R \colon (0, \infty) \times \mathbb{Y} \to \mathbb{X}$ is a map and $K \subset \mathrm{dom}(F)$ a subset. Then*

$$\omega(2\delta, K) \le 2\Delta_R(\delta, K) \quad \text{for all } \delta > 0.$$

Proof. Let $f_1, f_2 \in K$ with $\|F(f_1) - F(f_2)\|_{\mathbb{Y}} \le 2\delta$. With $g = \frac{1}{2}(F(f_1) + F(f_2))$ we have $\|g - F(f_i)\|_{\mathbb{Y}} \le \delta$ for $i = 1, 2$. Hence

$$\|f_1 - f_2\|_{\mathbb{X}} \le \|f_1 - R(\delta, g)\|_{\mathbb{X}} + \|R(\delta, g) - f_2\|_{\mathbb{X}} \le 2\Delta_R(\delta, K).$$

This implies the claim by definition of $\omega(2\delta, K)$. $\qquad\square$

We refer to [61] for a comprehensive treatment of the modulus of continuity for linear operators in Hilbert spaces.

We finish this section with a practicable criterion to verify order optimality.

Corollary 2.7 (Order optimality via the modulus of continuity)**.** *In the setting of Proposition 2.6 suppose* $\phi \colon (0, \infty) \to [0, \infty)$ *is non-decreasing and that there exists a constant* $c_\omega > 0$ *and* $\delta_0 > 0$ *such that*

$$\Delta_R(\delta, K) \leq \phi(\delta) \quad and \quad c_\omega \phi(\delta) \leq \omega(\delta, K) \quad for\ all\ 0 < \delta < \delta_0.$$

Then R *is an order optimal reconstruction method on* K.

Proof. Let $R' \colon (0, \infty) \times \mathbb{Y} \to \mathbb{X}$ be a map. Let $0 < \delta < \frac{\delta_0}{2}$. Then Proposition 2.6 yields

$$\Delta_R(\delta, K) \leq \phi(\delta) \leq \phi(2\delta) \leq \frac{1}{c_\omega} \omega(2\delta, K) \leq \frac{2}{c_\omega} \Delta_{R'}(\delta, K). \qquad \square$$

2.3 Literature on convergence rates for sparsity promoting regularization

We give a brief overview of the literature on convergence rate theory for ℓ^1-regularization. In the early paper [78] from 2008 the rate $\mathcal{O}(\delta^{1/2})$ in the 1-norm is shown assuming that the unknown solution is sparse (i.e. has only finitely many non-vanishing entries) and that the forward operator is linear. The paper [52] provides the rate $\mathcal{O}(\delta^{1/2})$ for nonlinear operators under a source condition that coincides with (2.4) in the linear case. Furthermore, by additionally requiring sparsity of the unknown the authors achieve the linear rate $\mathcal{O}(\delta)$ and discuss that in contrast to classical Tikhonov regularization, which has the highest possible rate $\mathcal{O}(\delta^{2/3})$, in ℓ^1-regularization no saturation effect occurs. To the best of the author's knowledge the linear rate $\mathcal{O}(\delta)$ was first proven in [14] for a regularization scheme similar to (2.1), which is called *residual method* in [52]. In [50] a linear convergence rate is shown in the more general setting of positively homogeneous functionals under the source condition (2.4) and a mild injectivity type assumption. Furthermore, in [53] it is proven (again under a mild injectivity type assumption) that the condition (2.4) is not only sufficient but even necessary for a linear convergence rate of ℓ^1-regularization. The phenomenon of exact recovery, i.e. the question whether the support of the estimator equals the support of a sparse exact solution, is affirmatively treated in [79].

However, it is usually more realistic to assume that the true solution is only approximately sparse in the sense that it can well be approximated by sparse vectors. Using a variational source condition, convergence rates are shown for non-sparse solutions in [11] for linear forward operators. Therein the analysis is based on the assumption that the unit vectors belong to the range of the adjoint operator. The rates are characterized in terms of the growth of the norms of the preimages of the unit vectors and the speed of decay of the true solution. In [3] the range condition is further discussed and the convergence rate results are extended to ℓ^q-regularization with $q < 1$. We will discuss the latter range condition in more detail in Section 3.2.

In [41] a relaxation of the condition on the unit vectors is introduced, and it is shown

that one also obtains convergence rates if the basis is not smooth enough, i.e. the unit vectors do not belong to the range of the adjoint operator. We will illuminate this relaxed assumptions in more detail in Section 3.3. The approaches of [11] and [41] are put into a common framework in [43]. The relaxed assumptions are revised in [40] where it is shown that they are always satisfied for injective and weak*-to-weak continuous linear operators. Furthermore, in the later paper the linear rate $\mathcal{O}(\delta)$ is proven for sparse true solutions without any further assumptions.

This overview shows that on the abstract level of general ℓ^1-regularization for linear operators the convergence rate theory is far developed. Nevertheless, as already briefly discussed in the introduction, there are only few results providing convergence rates under concrete and practicable assumptions on the forward operator and non-sparse true solution. For example, Hölder type rates $\mathcal{O}(\delta^\nu)$ for $0 < \nu < 1$ where ν depends on smoothness of the unknown measured in some function space are desirable. In the sequel we will prove such convergence rates for wavelet regularization in Besov spaces. The latter will be introduced in the following subsection.

Finally, we touch upon wavelet thresholding methods that have been studied extensively, in particular for statistical noise models. In the special case of a diagonal operator Tikhonov regularization with wavelet penalties coincides with estimators given by these methods. In [32] (near) minimax convergence rates in Besov balls are shown for soft thresholding with a wavelet basis. The application to a class of linear inverse problems (including the Radon transform) is studied in [31]. Here the authors use a wavelet basis in $\mathbb{X}$ and a wavelet vagulette basis in the image space $\mathbb{Y}$ such that the forward operator is diagonal in these bases. The proposed method consists in exact inversion of the forward operator followed by a wavelet shrinkage in $\mathbb{X}$. Minimax optimal convergence rates with L^2-loss functions under white noise are established in many Besov balls. A wavelet vagulette decomposition is also used in [76] for the inversion of the Radon transform. However, the wavelet vagulette decompositions are only available for a small class of linear forward operators.

The further developed wavelet method for inversion of linear inverse problems proposed in [25] relaxes the restrictions on the wavelet decomposition and allows the use of a generic wavelet system in $\mathbb{X}$. In this reference optimal convergence rates in Besov balls for finitely smoothing linear operators with white noise error model are established. However, the degree of smoothing of the forward operator is explicitly used in the inversion method which limits the method to this class of operators. Finally, we refer to [73] for convergence rates results for a two-step procedure consisting of wavelet shrinkage in the data space $\mathbb{Y}$ followed by Tikhonov regularization.

In the following we will consider Tikhonov regularization with penalty terms given by wavelet coefficients. Apart from the fact that it allows for a more flexible penalization, its main advantage compared to the methods above is the applicability to non-linear inverse problems.

2.4   Wavelet regularization in Besov spaces

In the sequel we will consider the situation where the unknown is a function defined on a domain Ω. Our two main examples for Ω are the d-dimensional torus $\mathbb{T}^d := (\mathbb{R}/\mathbb{Z})^d$ and Euclidean space $\mathbb{R}^d$. Our theory may also be applied to bounded Lipschitz domain in

$\mathbb{R}^d$ (see [100, Def. 3.4.(iii)]). We will use a Besov type scale $B^s_{p,q}$ that we will introduce in an abstract and axiomatic way in the sequel. Typically Besov scales come with the parameters: *smoothness* $s \in \mathbb{R}$, *integrabilty* $p \in (0, \infty]$ and *fine index* $q \in (0, \infty]$. The reader unfamiliar with Besov function spaces may find a first orientation in the following relations to classical function spaces (see Proposition A.11):

$$B^0_{p,1}(\Omega) \subset L^p(\Omega) \subset B^0_{p,\infty}(\Omega) \quad \text{for all } p \geq 1$$

and

$$B^s_{2,2}(\Omega) = H^s(\Omega) \quad \text{with equivalent norms for all } s \in \mathbb{R}.$$

In the sequel we shortly introduce Besov function spaces. For a comprehensive treatment of Besov function spaces we refer to [101, 99] and [100]. Furthermore, we recommend [47, Ch. 4] for a readily accessible introduction to the topic and its applications in approximation theory.

The following reasons justify the use of these spaces in regularization theory:

- For certain forward operators it was shown in [66] that convergence rates are equivalent to smoothness measured in Besov spaces. More precisely the spaces $B^s_{2,\infty}$ appear as maxisets of classical Tikhonov regularization of a-times smoothing operators as follows: Let $\mathcal{M}$ be a d-dimensional manifold satisfying the additional requirements in [66, Ass. 6.1], $A \colon \mathbb{X} = L^2(\mathcal{M}) \to \mathbb{Y} = L^2(\mathcal{M})$ be a linear forward operator. With $H^s(\mathcal{M}), s \in \mathbb{R}$ the scale of Sobolev spaces, we assume that A extends or restricts to a Banach space isomorphism $H^s(\mathcal{M}) \to H^{s+a}(\mathcal{M})$ for all $s \in \mathbb{R}$ and some fixed $a > 0$. For $0 < s < 2a$, $f \in L^2$, g^{obs} satisfying a deterministic noise bound with noise level δ and $\hat{f}_\alpha$ given by (1.1) the result [66, Thm. 7.1] yields the following equivalence: There exists a parameter choice rule such that a bound $\|f - \hat{f}_\alpha\|_{\mathbb{X}} = \mathcal{O}(\delta^{\frac{s}{s+a}})$ holds true if and only if $f \in B^s_{2,\infty}(\mathcal{M})$.

- Besov norms provide a precise way to estimate the regularity of stochastic processes. For example, for $p \in [1, \infty)$ Gaussian white noise W on the d-dimensional torus satisfies $W \in B^{-\frac{d}{2}}_{p,\infty}(\mathbb{T}^d)$ and $W \notin B^s_{p,\infty}(\mathbb{T}^d)$ for $s > -\frac{d}{2}$ almost surely (see [103, Thm. 3.4]).

- Suitable wavelet systems give rise to a characterization of Besov spaces in terms of wavelet coefficients. This connection between wavelets and Besov spaces is a crucial point in this work and treated in more detail in the following.

Besov function spaces

We briefly introduce Besov function spaces. The basic ingredient is a smooth dyadic resolution of unity (see [99, Sec. 2.3.1.]) in Fourier space. To this end, let ϕ_0 be a smooth function on $\mathbb{R}^d$ with $\phi(\xi) = 1$ if $\|\xi\|_2 \leq 1$ and $\phi(\xi) = 0$ if $\|\xi\|_2 \geq 2$. For $j \in \mathbb{N}$ we set $\phi_j(\xi) := \phi(2^{-j}\xi) - \phi(2^{-j+1}\xi)$. Then we have

$$\operatorname{supp} \phi_j \leq \{\xi \in \mathbb{R}^d \colon 2^{j-1} \leq \|\xi\|_2 \leq 2^{j+1}\} \quad \text{for all} \quad j \geq 1 \tag{2.6}$$

and $\sum_{j=0}^{\infty} \phi_j(\xi) = 1$ for all $\xi \in \mathbb{R}^d$.

Now we start with $\Omega = \mathbb{R}^d$. Let $\mathcal{S}'(\mathbb{R}^d)$ be the space of tempered distributions, i.e. the

topological dual of the Schwartz space $\mathcal{S}(\mathbb{R}^d)$. The Fourier transform $\mathcal{F}$ extends to a linear isomorphism $\mathcal{F}\colon \mathcal{S}'(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$. With $s \in \mathbb{R}$ and $p, q \in (0, \infty]$ we have

$$B_{p,q}^s(\mathbb{R}^d) := \left\{ f \in \mathcal{S}'(\mathbb{R}^d) \colon \left\| \left(2^{js} \| \mathcal{F}^{-1} \phi_j \mathcal{F} f \|_{L^p(\mathbb{R}^d)} \right)_{j \in \mathbb{N}_0} \right\|_q < \infty \right\} \qquad (2.7)$$

(see [101, Sec. 2.3.1.]).

Next we consider the periodic case. The definition of $B_{p,q}^s(\mathbb{T}^d)$ is similar to the one of $B_{p,q}^s(\mathbb{R}^d)$. Let $\mathcal{D}(\mathbb{T}^d)$ be the space all infinitely differentiable functions on $\mathbb{T}^d$. Then its topological dual $\mathcal{D}'(\mathbb{T}^d)$ is the space of periodic distributions. For $k \in \mathbb{Z}^d$ let $e_k(x) := e^{i\langle k, x \rangle}$. Every $f \in \mathcal{D}'(\mathbb{T}^d)$ has a representation

$$f = \sum_{k \in \mathbb{Z}^d} \hat{f}(k) e_k \quad \text{with Fourier coefficients} \quad \hat{f}(k) = \langle f, e_k \rangle \qquad (2.8)$$

and $\hat{f}(k)$ is of at most polynomial growth. On the other hand, if $(a_k)_{k \in \mathbb{Z}^d}$ is a sequence of at most polynomial growth then there exists $f \in \mathcal{D}'(\mathbb{T}^d)$ such that $\hat{f}(k) = a_k$.

For $j \in \mathbb{N}_0$ we set

$$f_j = \sum_{k \in \mathbb{Z}^d} \hat{f}(k) \phi_j(k) e_k. \qquad (2.9)$$

With $(\phi_j)_{j \in \mathbb{N}_0}$ as above, $p, q \in (0, \infty]$ and $s \in \mathbb{R}$ we have

$$B_{p,q}^s(\mathbb{T}^d) := \left\{ f \in \mathcal{D}'(\mathbb{T}^d) \colon \left\| \left(2^{js} \| f_j \|_{L^p(\mathbb{T}^d)} \right)_{j \in \mathbb{N}_0} \right\|_q < \infty \right\} \qquad (2.10)$$

(see [101, Sec. 9.1.3]).

Before we move on we discuss a simplification in the case $p > 1$.

Remark 2.8 (Lizorkin representations (see [101, Sec. 2.5.4.]))**.** *If $p \in (1, \infty]$ it is possible to replace the smooth resolution of unity $(\phi_j)_{j \in \mathbb{N}_0}$ by the one given by the characteristic function of the unit cube. The characterizations (2.10) and (2.7) remain valid and the definition of the norms with ϕ_j replaced by the characteristic functions yield equivalent norms. This is helpful in the analysis as the overlap in the resolution is omitted which leads to pairwise orthogonality of $(f_j)_{j \in \mathbb{N}}$ given by (2.9). We refer to [101, Sec. 2.5.4.] for details in the case $\Omega = \mathbb{R}^d$. For the d-dimensional torus this leads to the following construction of an equivalent norm (see [95, Sec. 3.5.3.]): Let*

$$\Pi_0 = \{0\} \quad and \quad \Pi_j := \{\xi \in \mathbb{Z}^d \colon 2^{j-1} \le \|\xi\|_\infty < 2^j\} \quad for \ j \ge 1.$$

Then for all $s \in \mathbb{R}$, $p \in (1, \infty)$ and $q \in [1, \infty]$ the following expression

$$\|f | B_{p,q}^s(\mathbb{T}^d)\| := \left\| \left(2^{js} \| \sum_{k \in \Pi_j} \hat{f}(k) e_k \|_{L^p(\mathbb{T}^d)} \right)_{j \in \mathbb{N}_0} \right\|_q \qquad (2.11)$$

defines an equivalent norm on $B_{p,q}^s(\mathbb{T}^d)$.

There are several ways to define Besov function spaces for open subsets $\Omega \subset \mathbb{R}^n$ differing mainly by their treatment of boundary values. We will recall two definitions of Besov spaces for subdomains by restrictions in the sense of the theory of distributions (see [100, Def. 2.1. and Rem. 2.2.]). Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain, $\mathcal{D}(\Omega)$ be the set of all smooth function with compact support contained in Ω and $\mathcal{D}'(\Omega)$ be the space of distributions on Ω. The first way is

$$B^s_{p,q}(\Omega) := \left\{ f \in \mathcal{D}'(\Omega) \colon f = g|_\Omega \text{ for some } g \in B^s_{p,q}(\mathbb{R}^d) \right\} \tag{2.12}$$

with the natural quotient norm given by the infimum over $\|g|B^s_{p,q}(\mathbb{R}^d)\|$ for all extensions g.

The other definition is

$$\tilde{B}^s_{p,q}(\Omega) := \left\{ f \in B^s_{p,q}(\mathbb{R}^d) \colon \operatorname{supp}(f) \subset \overline{\Omega} \right\}. \tag{2.13}$$

Besides basic embedding and interpolation identities the results in this work depend only on the characterization of Besov function spaces by wavelet coefficients that we will address in the following section. Therefore, we use the following general way to introduce Besov scales. This may help the reader mainly interested in regularization theory and not familiar with Besov function spaces as all properties we need in this work are bundled in the following definition. Moreover, the collection shows all properties that have to be met by a scale of function spaces on a domain to make our theory applicable.

Recall that a *quasi-norm* satisfies all properties of a norm except for the triangle inequality, which is only holds true in the weaker form $\|x + y\| \le C(\|x\| + \|y\|)$ with $C > 1$. A complete vector space with a quasi-norm is called *quasi-Banach space*.

Moreover, we recall real interpolation of quasi-Banach spaces via the K-method (see [8, Sec. 3.1 & 3.11]). To this end, let $\mathbb{X}_1$ and $\mathbb{X}_2$ be quasi-Banach spaces with a common ambient Hausdorff topological vector space $\tilde{\mathcal{D}}$, i.e. $\mathbb{X}_i \subset \tilde{\mathcal{D}}$ continuously. Then for $t > 0$ and $f \in \mathbb{X}_1 + \mathbb{X}_2$ the K-functional is defined by

$$K(t, f) := \inf \left(\|f_1\|_{\mathbb{X}_1} + t\|f_2\|_{\mathbb{X}_2} \right) \tag{2.14}$$

where the infimum is taken over all decomposition $f = f_1 + f_2$ with $f_i \in \mathbb{X}_i$. Let $0 < \theta < 1$ and $q \in (0, \infty)$. For $f \in \mathbb{X}_1 + \mathbb{X}_2$ one defines

$$\|f\|_{(\mathbb{X}_1, \mathbb{X}_2)_{\theta,q}} := \left(\int_0^\infty t^{-q\theta} K(t, f)^q \frac{\mathrm{d}t}{t} \right)^{\frac{1}{q}}. \tag{2.15}$$

Furthermore, for $0 \le \theta \le 1$ we have

$$\|f\|_{(\mathbb{X}_1, \mathbb{X}_2)_{\theta,\infty}} = \sup_{t>0} t^{-\theta} K(t, f).$$

We obtain a quasi-Banach space $(\mathbb{X}_1, \mathbb{X}_2)_{\theta,q}$ consisting of all elements $f \in \mathbb{X}_1 + \mathbb{X}_2$ with $\|f\|_{(\mathbb{X}_-, \mathbb{X})_{\theta,q}} < \infty$. For all $0 < \theta < 1$ and $q \in (0, \infty)$ there exists a constant $c_{\theta,q}$ such that the interpolation inequality

$$\|f\|_{(\mathbb{X}_1, \mathbb{X}_2)_{\theta,q}} \le c_{\theta,q} \|f\|_{\mathbb{X}_1}^{1-\theta} \cdot \|f\|_{\mathbb{X}_2}^{\theta}$$

holds true for all $f \in \mathbb{X}_1 \cap \mathbb{X}_2$ (see [98, Sec. 1.3.3.]).

Definition 2.9 (Besov scale). *Let $d \in \mathbb{N}$. A family of quasi-Banach spaces $\mathcal{D} \subset B_{p,q}^s \subset \tilde{\mathcal{D}}$ for $s \in \mathbb{R}$, $p, q \in (0, \infty]$ with common subspace $\mathcal{D}$ and embedded continuously into an ambient Hausdorff topological vector space $\tilde{\mathcal{D}}$ with quasi-norms $\|\cdot \,|\, B_{p,q}^s\|$ is called a* Besov scale *if the following conditions are satisfied.*

(a) *$B_{p,q}^s$ is a Banach space whenever $p, q \in [1, \infty]$. $\mathcal{D}$ is dense in $B_{p,q}^s$ whenever $p, q \in (1, \infty)$.*

(b) *$B_{2,2}^0$ is a Hilbert space. Moreover, suppose that the restriction $\langle \cdot, \cdot \rangle \colon \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ of the inner product on $B_{2,2}^0$ extends to a bounded bilinear pairing $B_{p',q'}^{-s} \times B_{p,q}^s \to \mathbb{R}$ that gives rise to isomorphisms $(B_{p,q}^s)' \simeq B_{p',q'}^{-s}$ for all $s \in \mathbb{R}$, $p, q \in [1, \infty)$ and their Hölder conjugates p', q'.*

(c) *The following inclusions hold true with continuous embeddings:*
  (i) *$B_{p_1,q}^{s_1} \subset B_{p_2,q}^{s_2}$ for $s_1 \geq s_2$, $0 < p_1 \leq p_2 \leq \infty$, $q \in (0, \infty]$ with $s_1 - \frac{d}{p_1} \geq s_2 - \frac{d}{p_2}$,*
  (ii) *$B_{p,q_1}^s \subset B_{p,q_2}^s$ for $s \in \mathbb{R}$, $p \in (0, \infty]$, $0 < q_1 \leq q_2 \leq \infty$,*
  (iii) *$B_{p,\infty}^{s_2} \subset B_{p,q}^{s_1}$ for $s_1 < s_2$, $p, q \in (0, \infty]$.*

(d) *Let $\theta \in (0, 1)$. The following real interpolation identities hold true with equivalent quasi-norms:*
  (i) *$B_{p,q}^s = \left(B_{p,q_1}^{s_1}, B_{p,q_2}^{s_2} \right)_{\theta,q}$ for $s_1 < s_2$ with $s = (1 - \theta)s_1 + \theta s_2$, $p, q, q_1, q_2 \in (0, \infty]$.*
  (ii) *$B_{p,p}^s = \left(B_{p_1,p_1}^{s_1}, B_{p_2,p_2}^{s_2} \right)_{\theta,p}$ for $s_1 < s_2$ with $s = (1 - \theta)s_1 + \theta s_2$, $p, p_1, p_2 \in (0, \infty)$ with $\frac{1}{p} = \frac{1-\theta}{p_1} + \frac{\theta}{p_2}$.*

Moreover, a Besov scale $B_{p,q}^s$ is called p-monotone if $B_{p_2,q}^s \subset B_{p_1,q}^s$ with a bounded embedding whenever $0 < p_1 \leq p_2 \leq \infty$.

If Ω is either $\mathbb{T}^d$, $\mathbb{R}^d$ or a bounded Lipschitz domain in $\mathbb{R}^d$, then the scale of Besov function spaces $\mathcal{D}(\Omega) \subset B_{p,q}^s(\Omega) \subset \mathcal{D}'(\Omega)$ (with $\mathcal{S}$ instead of $\mathcal{D}$ if $\Omega = \mathbb{R}^d$) is a Besov scale in the sense of Definition 2.9 (see e.g. [101], [100] or [95]). If Ω is either $\mathbb{T}^d$ or a bounded Lipschitz domain in $\mathbb{R}^d$ then the scale $B_{p,q}^s(\Omega)$ is p-monotone (see [101, 3.3.1.(9)]). We will frequently use the properties of a Besov scale without further referencing.

We finish this section with a direct consequence of Definition 2.9, namely a version of the mixed embedding in Definition 2.9(c)(i) showing that if the required inequality is strict we may change the fine index q.

Lemma 2.10 (Mixed embedding). *Let $B_{p,q}^s$ be a Besov scale as defined in Definition 2.9, $0 < p_1 \leq p_2 \leq \infty$, $s_1, s_2 \in \mathbb{R}$ with $s_1 - \frac{d}{p_1} > s_2 - \frac{d}{p_2}$ and $q_1, q_2 \in (0, \infty)$. Then there is a continuous embedding $B_{p_1,q_1}^{s_1} \subset B_{p_2,q_2}^{s_2}$.*

Proof. Let $\tilde{s}_2 \in \mathbb{R}$ satisfy $s_1 - \frac{d}{p_1} = \tilde{s}_2 - \frac{d}{p_2}$. Then $s_2 < \tilde{s}_2$ and hence there is a continuous embedding $B_{p_1,q_1}^{s_1} \subset B_{p_2,q_1}^{\tilde{s}_2}$. With this the claim follows from the following chain of continuous embeddings $B_{p_2,q_1}^{\tilde{s}_2} \subset B_{p_2,\infty}^{\tilde{s}_2} \subset B_{p_2,q_2}^{s_2}$. $\qquad\square$

Besov sequence spaces and wavelet systems

We introduce a scale of sequence spaces $b_{p,q}^s$ that allows to characterize Besov function spaces $B_{p,q}^s(\Omega)$ by decay properties of coefficients in wavelet expansions (see also [100, Def. 2.6]).

Let $d \in \mathbb{N}_0$. Let $(\Lambda_j)_{j \in \mathbb{N}_0}$ be either a family of finite sets such that

$$2^{jd} \leq |\Lambda_j| \leq C_\Lambda 2^{jd} \quad \text{for some constant } C_\Lambda \geq 1 \text{ and all } j \in \mathbb{N}_0 \tag{2.16}$$

or a family of countably infinite sets. We refer to Λ_j as the *j-th level* and consider the index set

$$\Lambda := \{(j,k) \colon j \in \mathbb{N}_0, k \in \Lambda_j\}.$$

For a sequence $x = (x_{j,k})_{(j,k) \in \Lambda}$ and a fixed $j \in \mathbb{N}_0$ we denote by $x_j := (x_{j,k})_{k \in \Lambda_j} \in \mathbb{R}^{\Lambda_j}$ the projection onto the j-th level. For $s \in \mathbb{R}$ and $p, q \in (0, \infty]$ let us introduce

$$b_{p,q}^s := \left\{ x \in \mathbb{R}^\Lambda \colon \|x\|_{s,p,q} < \infty \right\} \quad \text{with} \quad \|x\|_{s,p,q} := \left\| \left(2^{js} 2^{jd\left(\frac{1}{2} - \frac{1}{p}\right)} \|x_j\|_p \right)_{j \in \mathbb{N}_0} \right\|_q.$$

If $p, q \in [1, \infty]$, then $b_{p,q}^s$ is a Banach space. For $p \in (0,1)$ or $q \in (0,1)$ the vector spaces $b_{p,q}^s$ are quasi-Banach spaces. Here the smallest possible constant in the replacement of the triangle inequality is $C = 2^{\frac{1}{\min\{p,q\}} - 1}$. The set of finitely supported sequences is a common subset and lies dense in $b_{p,q}^s$ whenever $p \neq \infty$ and $q \neq \infty$. We have $\ell^2(\Lambda) = b_{2,2}^0$.
Furthermore, it is routine to see that all the embedding relations and the statement on the dual pairing in Definition 2.9 are satisfied. An adaption of the proof [101, Thm. 2.4.2.] shows the interpolation identity Definition 2.9(d)(i). Finally, the identity in Definition 2.9(d)(ii) follows from identifying $b_{p,p}^s$ with a weighed ℓ^p-space and using the results in [45]. Hence, the scale $b_{p,q}^s \subset \mathbb{R}^\Lambda$ is a Besov scale in the sense of Definition 2.9. Moreover, it is p-monotone if the levels Λ_j and satisfy (2.16).
We formulate a condition that is typically met by wavelet systems.

Assumption 2.11. *Let $\mathcal{D} \subset B_{p,q}^s \subset \tilde{\mathcal{D}}$ be a Besov scale as defined in Definition 2.9 and $s_{max} > 0$. With $b_{p,q}^s$ as defined above (with d according to the Besov scale) suppose that $(\psi_\lambda)_{\lambda \in \Lambda}$ is a family of elements in $\tilde{\mathcal{D}}$ such that the* synthesis operator

$$\mathcal{S} \colon b_{p,q}^s \to B_{p,q}^s \quad \text{given by } x \mapsto \sum_{\lambda \in \Lambda} x_\lambda \psi_\lambda$$

is well defined and a quasi-norm isomorphism for all $s \in \mathbb{R}$, $p, q \in (0, \infty]$ satisfying $s \in (\sigma_p - s_{max}, s_{max})$ with $\sigma_p = \max\left\{ d\left(\frac{1}{p} - 1\right), 0 \right\}$.

Note that $p \geq 1$ implies $\sigma_p = 0$ and therefore $\mathcal{S}$ is a norm isomorphism whenever $|s| \leq s_{\max}$ in this case.
For Ω either $\mathbb{R}^d$ or $\mathbb{T}^d$ it is shown in [100, Thm. 1.20, Prop. 1.34, Thm. 1.37] that for every $s_{\max} > 0$ there exists an orthonormal wavelet system such that Assumption 2.11 is met with the scale of Besov function spaces $B_{p,q}^s(\Omega)$.
For a similar statement for bounded Lipschitz domains $\Omega \subset \mathbb{R}^d$ we refer to [100, Thm. 2.33, Thm. 3.23, Rem. 3.25]. There a scale of Besov function space is used that mixes the two definitions given in (2.12) and (2.13). In the following two examples we briefly discuss the construction of wavelet systems.

Example 2.12 (Periodic wavelets (cf. [64, Ex. 2.5]))**.** *We sketch the construction of a wavelet system on $\mathbb{T}^d$ starting with a wavelet system on $\mathbb{R}$. Let $s_{max} \in \mathbb{N}$ and $\psi^M \in C^{s_{max}}(\mathbb{R})$ be a wavelet function (or mother wavelet) with corresponding scaling*
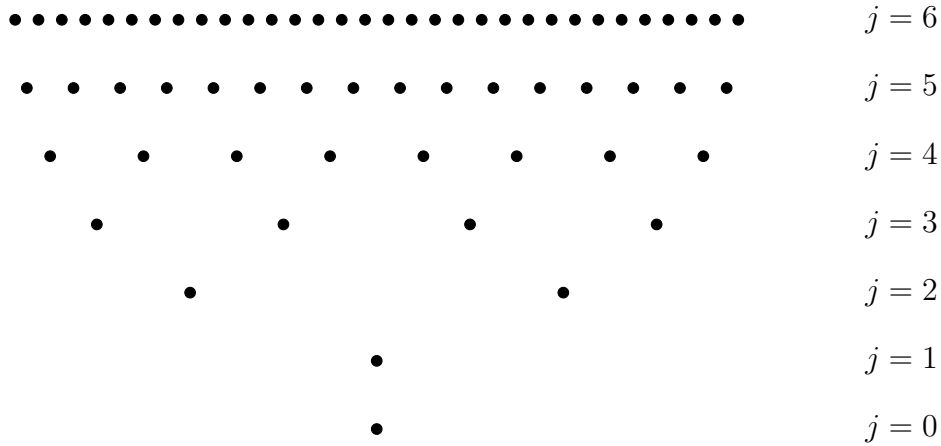
FIGURE 2.1: A schematic illustration of the index set Λ if (2.16) holds true with $d = 1$. Typically wavelet coefficients of the Haar wavelet are stored in this data structure.

function (or father wavelet) $\psi^F \in C^{s_{max}}(\mathbb{R})$ (see e.g. [28] for an introduction to wavelets). For simplicity we assume that $\{\psi_{j,k}\colon j \in \mathbb{N}_0, k \in \mathbb{Z}\}$ given by $\psi_{0,k}(x) := \phi^F(x - k)$ and $\psi_{j+1,k}(x) := 2^{j/2}\psi^M(2^j x - k)$ for $j \in \mathbb{N}_0$ forms an orthonormal basis of $L^2(\mathbb{R})$. We further assume that this wavelet system is s_{max}-regular, in the sense of [47, Def. 4.2.14], which entails in particular that $\int_{\mathbb{R}} \psi^M(x)x^m \, dx = 0$ for all $m \in \mathbb{N}_0$ with $m \le s_{max} - 1$. Examples of such wavelets include Daubechies wavelets [27] of order $N \in \mathbb{N}$, which are supported in an interval of length $2N - 1$ and the regularity can be bounded by $s_{max} \ge 0.2N$ for large N, as well as Meyer-wavelets [82], which are s_{max}-regular for any $s_{max} > 0$.

Using periodization one can construct a 1-periodic wavelet system from a wavelet system on $\mathbb{R}$: For $j \in \mathbb{N}_0$ we consider scaling and periodization $\psi_j^{P,M}(x) := \sum_{k \in \mathbb{Z}} 2^{j/2}\psi^M(2^j(x - k))$ of the mother wavelet. We set $\Lambda_0^P = \{0\}$ and $\psi_{0,0}^P(x) := \sum_{k \in \mathbb{Z}} \psi^F(x - k)$. Moreover, for $j \in \mathbb{N}_0$ and $k \in \Lambda_{j+1}^P := \{0, \ldots, 2^j - 1\}$ we define $\psi_{j,k}^P(x) := \psi_j^{P,M}(x - 2^{-j}k)$. Note that all these functions belong to the space $C^{s_{max}}(\mathbb{T})$ of 1-periodic functions. The system $(\psi_{j,k}^P)_{(j,k) \in \Lambda^P}$ with $\Lambda^P := \{(j,k)\colon j \in \mathbb{N}_0 \text{ and } k \in \Lambda_j^P\}$ constitutes an orthonormal basis of $L^2(\mathbb{T})$. Furthermore, Assumption 2.11 holds true (see [47, Thm. 4.3.26 and (4.137)] and [19, Ch. 3] for a treatment including the case $0 < p < 1$).

Finally to obtain multivariate wavelet systems let us recall the construction of tensor product wavelets (see [19, Sec. 1.4] or [28, Sec. 10.1] for further motivation and explanation). To this end, we also need to scale and shift the father wavelet: $\psi_j^{P,F}(x) := \sum_{k \in \mathbb{Z}} 2^{j/2}\psi^F(2^j(x - k))$ with $j \in \mathbb{N}_0$. Now let $\mathcal{G} := \{F, M\}^d \setminus \{(F, \ldots, F)\}$ and note that $|\mathcal{G}| = 2^d - 1$. We set $\psi_{0,\mathbf{0}}(\mathbf{x}) := \prod_{l=1}^d \psi_0^{P,F}(x_l)$ and $\psi_{j,\mathbf{k},\mathbf{G}} := \prod_{l=1}^d \psi_j^{P,G_l}(x_l - 2^{-j}k_l)$ for $\mathbf{k} \in \{0, \ldots, 2^j - 1\}^d$, $\mathbf{G} \in \mathcal{G}$ and $j \in \mathbb{N}$. Set $\Lambda_0^{P,d} := \{\mathbf{0}\}$, $\Lambda_{j+1}^{P,d} := \{0, \ldots, 2^j - 1\}^d \times \mathcal{G}$ for $j \in \mathbb{N}_0$ and $\Lambda^{P,d} := \{(j,l)\colon j \in \mathbb{N}_0, l \in \Lambda_j^{P,d}\}$. Then $\{\psi_\lambda^{P,d}\colon \lambda \in \Lambda^{P,d}\}$ is an orthonormal basis and Assumption 2.11 remains valid.

Example 2.13 (Wavelets on bounded domains $\Omega \subset \mathbb{R}^d$ (cf. [64, Ex. 2.6]))**.** *In many applications domains with non-periodic boundary conditions occur. To show that this can also be described in our setting, recall the two possibilities (2.12) and (2.12) to define Besov spaces on $\Omega \subset \mathbb{R}^d$. For a bounded Lipschitz domain $\Omega \subset \mathbb{R}^d$ it is possible under certain*

conditions to modify the wavelets whose support intersects with the boundary $\partial\Omega$ (or is close to $\partial\Omega$ relative to its size) such that approximation properties and inverse inequalities of the corresponding wavelet spaces are preserved. This was done in [26] for the symmetric, compactly supported biorthogonal wavelets from [22]. There are different constructions of boundary-adapted wavelets leading to isomorphisms $b_{p,q}^s \cong B_{p,q}^s(\Omega)$ or $b_{p,q}^s \cong \tilde{B}_{p,q}^s(\Omega)$ for certain values of s, p and q. (see e.g. [100, Thm. 2.33., Thm. 3.23., Rem. 3.25], [19, §3.9–3.10], [47, §4.3.5], [5] and [23]).

Wavelet regularization

Let $p, q \in [1, \infty)$, $r \geq 0$ and $u \in [1, \infty)$. Assume Assumption 2.11 holds true with $r < s_{\max}$. Let $F \colon \operatorname{dom}(F) \to \mathbb{Y}$ be a map defined on a subset $\operatorname{dom}(F) \subset \tilde{\mathcal{D}}$ mapping to a Banach space $\mathbb{Y}$. For $g \in \mathbb{Y}$ we consider the following set-valued scheme

$$R_\alpha(g) = \mathcal{S}\hat{x}_\alpha \quad \text{with} \quad \hat{x}_\alpha \in \operatorname*{argmin}_{x \in D} \left(\frac{1}{2\alpha}\|g - F(\mathcal{S}x)\|_{\mathbb{Y}}^2 + \frac{1}{u}\|x\|_{r,p,q}^u \right) \qquad (2.17)$$

with $D := \mathcal{S}^{-1}(B_{p,q}^r \cap \operatorname{dom}(F)) \subseteq b_{p,q}^r$ and $\alpha > 0$ a regularization parameter.

Note that this setting contains regularization with initial guess $\mathcal{S}x_0 = f_0 \in B_{p,q}^r$ as follows: The above scheme with $\|x\|_{r,p,q}$ replaced by $\|x - x_0\|_{r,p,q}$ is the same as (2.17) with $\tilde{F}(f) := F(f + f_0)$ on $\operatorname{dom}(\tilde{F}) := \operatorname{dom}(F) - f_0$.

This kind of regularization scheme with a linear operator is the "primary motivation" for the seminal paper by Daubechies, Devries and DeMol (see [29, Sec. 1.4.1]). In the paper [90] regularizing properties are shown for nonlinear operators.

For linear operators and Besov space wavelet penalty term the source condition (2.4) and the stronger condition below (2.4) are interpreted and used to prove convergence rates in [80] and [91]. Convergence rates with respect to Bregman distances rather than the norm $\|\cdot \mid B_{p,q}^r\|$ as loss function are shown in [12] under Besov smoothness assumptions using approximate source conditions. Finally, we refer to the paper [104] were convergence rates where shown under variational source conditions. In the latter paper optimal convergence rates are achieved for $p \in (1, 2)$ and $q \geq 2$. One central aim of this work is to extend these results to the sparsity promoting penalties with $p \in [1, 2]$ and $q = 1$.

Remark 2.14 (Analysis versus synthesis setting)**.** *Let Ω be either $\mathbb{T}^d$, $\mathbb{R}^d$ or a subdomain of $\mathbb{R}^d$. Some of the references above define the Tikhonov functional on function spaces rather then on sequence spaces and use the analysis operator of a wavelet system $(\psi_\lambda)_{\lambda \in \Lambda} \subset L^2(\Omega)$ to define a Besov space penalty term:*

$$R_\alpha^{(A)}(g) = \operatorname*{argmin}_{f \in B_{p,q}^r \cap \operatorname{dom}(F)} \left(\frac{1}{2\alpha}\|g^{\text{obs}} - F(f)\|_{\mathbb{Y}}^2 + \frac{1}{u}\|(\langle f, \psi_\lambda\rangle)_{\lambda \in \Lambda}\|_{r,p,q}^u \right). \qquad (2.18)$$

We argue that the latter estimators equal the ones in (2.17) whenever $(\psi_\lambda)_{\lambda \in \Lambda}$ satisfies Assumption 2.11 and additionally is an orthonormal basis of $L^2(\Omega)$.

To this end, note that we have $\langle f, \psi_\lambda\rangle = (\mathcal{S}^{-1}f)_\lambda$ first for all test functions $f \in \mathcal{D}(\Omega)$ and therefore for all $f \in B_{p,q}^r$. Hence the penalty in 2.18 equals $\frac{1}{u}\|\mathcal{S}^{-1}f\|_{r,p,q}^u$. From this it follows that $R_\alpha(g) = R_\alpha^{(A)}(g)$ by considering how argmin transforms under composition with a bijective mapping.

Our analysis using the synthesis operator as in Assumption 2.11 and estimators of the

form (2.17) *is restricted to wavelet basis systems. The reason is that for a wavelet frame which is overcomplete the synthesis operator $\mathcal{S}$ and therefore also the composition $F \circ \mathcal{S}$ is not injective. Nevertheless, we will show that the convergence rates results we obtain in the setting of Assumption 2.11 do generalize to wavelet frames if one replaces the estimators given by* (2.17) *by the scheme* (2.18). *(See [3, Sec. 6.2.] for a related discussion.)*

Finitely smoothing operators

We formulate an assumption on the forward operator, discuss several examples and prove existence of minimizers in (2.17).

The following two-sided Lipschitz condition is a well studied mapping property of operators in Besov spaces.

Assumption 2.15. *For a given Besov scale $B_{p,q}^s$ suppose $a > 0$, $\mathrm{dom}(F) \subset B_{2,2}^{-a}$ is closed and $F \colon \mathrm{dom}(F) \to \mathbb{Y}$ for some Banach space $\mathbb{Y}$.*

(a) Assume there exist constants $M_1 > 0$ such that

$$\|f_1 - f_2 \,|\, B_{2,2}^{-a}\| \le M_1 \|F(f_1) - F(f_2)\|_{\mathbb{Y}} \quad \textit{for all } f_1, f_2 \in \mathrm{dom}(F).$$

(b) Assume there exist constants $M_2 > 0$ such that

$$\|F(f_1) - F(f_2)\|_{\mathbb{Y}} \le M_2 \|f_1 - f_2 \,|\, B_{2,2}^{-a}\| \quad \textit{for all } f_1, f_2 \in \mathrm{dom}(F).$$

If the space $B_{2,2}^{-a}$ in Assumption 2.15 agrees with the Besov function space $B_{2,2}^{-a}(\Omega)$ on Ω either $\mathbb{T}^d$, $\mathbb{R}^d$ or a bounded Lipschitz subdomain of $\mathbb{R}^d$, then replacing $B_{2,2}^{-a}$ by $H^{-a}(\Omega)$ in this assumption does only change the constants as the latter spaces agree and with equivalent norms (see Proposition A.11(b)). Assumption 2.15(a) imposes that F is *at most a-times smoothing.* This will be crucial requirement for upper bounds on the reconstruction error. On the other hand, Assumption 2.15(b) states that F is *at least a-times smoothing* and will be used to derive existence of minimizers in (2.17) and lower bounds on the reconstruction error.

We point out that a map F satisfying Assumption 2.15(b) defined initially only on a subset $D_F \subset L^2(\Omega)$ has a unique continuous extension to the closure of D_F in $B_{2,2}^{-a}(\Omega)$ and both inequalities remain valid for this extension.

Before we discuss the verification of Assumption 2.15 in several examples, we prove existence of minimizers in (2.17) under some natural restrictions on the parameters a, d, r and p.

Proposition 2.16 (Existence of minimizers)**.** *Suppose Assumptions 2.11 and 2.15(b) hold true. Let $p, q \in [1, \infty)$, $0 \le r < s_{max}$ and $u \in [1, \infty)$ such that $B_{p,q}^r \cap \mathrm{dom}(F) \ne \emptyset$. Additionally assume that at least one of the following assumptions holds true:*

(a) (2.16) *and $a + r > d\left(\frac{1}{p} - \frac{1}{2}\right)$.*

(b) $a + r \ge d\left(\frac{1}{p} - \frac{1}{2}\right)$, $q < 2$, $\mathrm{dom}(F) \subset B_{2,2}^{-a}$ is weakly closed and F is weak-to-weak continuous.

Then for R_α given by (2.17) *we have $R_\alpha(g) \ne \emptyset$ for all $g \in \mathbb{Y}$.*

Proof.   (a) We verify Assumption 2.2 for $F \circ \mathcal{S}$ on the set $D \subset b_{p,q}^r$ defined below (2.17).

By the assumed inequality there exists $s_{\max} < \tilde{r} < r$ such that $a + \tilde{r} > d\left(\frac{1}{p} - \frac{1}{2}\right)$. We

choose $\tau_{\mathbb{X}}$ to be the subspace topology on $b^r_{p,q}$ induced by the embedding $b^r_{p,q} \subset b^{\tilde{r}}_{p,q}$. Proposition A.7 implies that the sublevel sets of $b^r_{p,q} \to [0, \infty)$ given by $x \mapsto \frac{1}{u}\|x\|_{r,p,q}$ are $\tau_{\mathbb{X}}$-compact.

By Assumption 2.11 the synthesis operator $\mathcal{S}\colon b^{\tilde{r}}_{p,q} \to B^{\tilde{r}}_{p,q}$ is a norm isomorphism. Furthermore, there is a continuous embedding $B^{\tilde{r}}_{p,q} \subset B^{-a}_{2,2}$ (see Lemma 2.10). Hence $\mathcal{S}\colon b^r_{p,q} \to B^{-a}_{2,2}$ is well-defined and $\tau_{\mathbb{X}}$-to-norm continuous. Therefore, D is $\tau_{\mathbb{X}}$-closed. As Assumption 2.15(b) implies norm-to-norm continuity of F we have $\tau_{\mathbb{X}}$-to-norm continuity of $F \circ \mathcal{S}$ which in turn yields $\tau_{\mathbb{X}}$-to-weak continuity of $F \circ \mathcal{S}$. The claim follows as Proposition 2.3 yields the existence of $\hat{x}_\alpha$ in (2.17).

(b) For $\lambda \in \Lambda$ let $e_\lambda \in \mathbb{R}^\Lambda$ be the unit vector given by $(e_\lambda)_{\lambda'} = \delta_{\lambda\lambda'}$ and p', q' the Hölder conjugates of p and q, i.e. $\frac{1}{p} + \frac{1}{p'} = 1$ and likewise for q. Then the norm closure of the linear span of $\{e_\lambda\colon \lambda \in \Lambda\}$ in $b^{-r}_{p',q'}$ is a predual for $b^r_{p,q}$ (see Proposition A.9). We consider the corresponding weak*-topology on $b^r_{p,q}$. Due to the Banach-Alaoglu theorem the sublevel sets of $\frac{1}{u}\|\cdot\|^u_{r,p,q}$ are weak*-compact. By Proposition A.10 the embedding $b^r_{p,q} \subset b^r_{p,2}$ is weak*-to-weak continuous. The operator $\mathcal{S}\colon b^r_{p,2} \to B^r_{p,2}$ and the embedding $B^r_{p,2} \subset B^{-a}_{2,2}$ are bounded and therefore weak-to-weak continuous. Hence $\mathcal{S}\colon b^r_{p,q} \to B^{-a}_{2,2}$ is weak*-to-weak continuous. Therefore, D defined below (2.17) is weak*-closed. Furthermore, $F \circ \mathcal{S}$ is weak*-to-weak continuous. Hence, once again the existence of $\hat{x}_\alpha$ in (2.17) follows from Proposition 2.3. $\qquad\square$

We included the technically more elaborate existence statement under the assumption (*b*) to cover for example the case $a = r = \frac{1}{2}$, $d = 2$ and $p = q = 1$, which is, as we will see, particularly relevant for the inversion of the Radon transform.

Example 2.17 (Periodic differential operators (cf. [64, Ex. 2.7]))**.** *A simple example that satisfies Assumption 2.15 is $F_a := (\mathrm{Id} - \Delta)^{-\frac{a}{2}}\colon B^{-a}_{2,2}(\mathbb{T}^d) \to L^2(\mathbb{T}^d)$. In this case F_a is a norm isomorphim $B^s_{p,q}(\mathbb{T}^d) \to B^{s+a}_{p,q}(\mathbb{T}^d)$ for all $s \in \mathbb{R}$ and $p, q \in [1, \infty]$. More generally, the assumptions hold true for inverses of elliptic differential operators of order a with smooth, periodic coefficients (see [98]). Furthermore, it is fulfilled by periodic convolution operators $Ff := k * f$ for which the Fourier coefficients of the convolution kernels k have the asymptotic behavior $\hat{k}(\mathbf{n}) \sim (1 + |\mathbf{n}|^2)^{-a/2}$, $\mathbf{n} \in \mathbb{Z}^d$.*

Example 2.18 (Radon transform (cf. [64, Ex. 2.7]))**.** *We consider the Radon transform on a bounded domain $\Omega \subset \mathbb{R}^d$, which appears as forward operator in computed tomography (CT) and positron emission tomography (PET), among others. If $S^{d-1} := \{\theta \in \mathbb{R}^d : \|\theta\|_2 = 1\}$ is the unit sphere, $R: L^2(\Omega) \to L^2(S^{d-1} \times \mathbb{R})$ is given by*

$$(Rf)(\theta, t) := \int_{x \cdot \theta = t} f(x)\, \mathrm{d}\beta \quad \text{with } \theta \in S^{d-1}, t \in \mathbb{R}.$$

As a special case of [87, Thm. 5.1] the Radon transform satisfies Assumption 2.15 with $a = \frac{d-1}{2}$ and $\mathbb{Y} = L^2(S^{d-1} \times \mathbb{R})$. Moreover, it follows from [57, Thm. 3.1] that R is bounded with closed range from $B^s_{2,2}(\{x \in \mathbb{R}^d : \|x\|_2 \le 1\})$ to $B^{s+a}_{2,2}(S^{d-1} \times [-1, 1])$ for all $s \in \mathbb{R}$. Note that in order to meet the condition $a + r \ge d\left(\frac{1}{p} - \frac{1}{2}\right)$ in Proposition 2.16(b) with the favorable choice $p = 1$ we need $r \ge \frac{1}{2}$. The weak continuity assumption is automatically satisfied as $F = R$ is linear and bounded. Hence, the choice $r = \frac{1}{2}$, leading to the $B^{\frac{1}{2}}_{1,1}$-penalty that is recommended in [76], is admissible.

As a last linear example we address the Abel integral transformation.

Example 2.19 (Abel integral equation)**.** *Let $a \in \left(0, \frac{1}{2}\right)$. The linear Abel transform $A_a \colon L^2([0,1]) \to L^2([0,1])$ is given by*

$$(Af)(t) = \int_0^t \frac{f(x)}{(t-x)^{1-a}} \, \mathrm{d}x \quad \text{with } t \in [0,1].$$

Then A_α is a-times smoothing in the sense of Assumption 2.15. For a detailed study of the mapping properties and in particular a verification of the latter statement we refer to [48].

We now turn to nonlinear examples (cf.[64, Ex. 2.6]). In the study of nonlinear operator equations in Hilbert scales the following condition, which is closely related to Assumption 2.15, has been investigated ([65, 88]):

$$\frac{1}{\tilde{M}} \|h\|_{\mathbb{X}_{-a}} \leq \|F'[f]h\|_{\mathbb{Y}} \leq \tilde{M} \|h\|_{\mathbb{X}_{-a}} \quad \text{for all } h \in \mathbb{X}_{-a}. \tag{2.19}$$

Here $f \in \mathrm{dom}(F)$, $\tilde{M} > 0$, and $\mathbb{X}_{-a}$ is an element of a Hilbert scale, typically of L^2-based Sobolev spaces $H^{-a}(\Omega)$. We also need the so-called *range invariance condition*: For all $f_1, f_2 \in \mathrm{dom}(F)$ there exists some operator $R(f_1, f_2) : \mathbb{Y} \to \mathbb{Y}$ such that

$$F'[f_1] = R(f_1, f_2)F'[f_2] \qquad \text{and} \qquad \|\mathrm{Id} - R(f_1, f_2)\| \leq \frac{1}{2}. \tag{2.20}$$

(Often a bound $\|I - R(f_1, f_2)\| \leq C\|f_1 - f_2\|$ is shown, which implies (2.20) in a ball of radius $1/(2C)$.) The following lemma shows that (2.19) and (2.20) implies Assumption 2.15 if $\mathbb{X}_{-a} = B_{2,2}^{-a}$ with equivalent norms.

Lemma 2.20 (cf.[64, Lem. 2.9])**.** *Let $\mathbb{X}_{-a}$ and $\mathbb{Y}$ be Banach spaces and suppose that $\mathrm{dom}(F) \subset \mathbb{X}_{-a}$ is convex, $F : \mathrm{dom}(F) \subset \mathbb{X}_{-a} \to \mathbb{Y}$ is Fréchet differentiable and satisfies (2.19) and (2.20). Then*

$$\frac{1}{2\tilde{M}} \|f_1 - f_2\|_{\mathbb{X}_{-a}} \leq \|F(f_1) - F(f_2)\|_{\mathbb{Y}} \leq \frac{3}{2}\tilde{M}\|f_1 - f_2\|_{\mathbb{X}_{-a}} \quad \text{for all} \quad f_1, f_2 \in \mathrm{dom}(F).$$

Proof. By the mean value theorem and (2.20) we have

$$F(f_1) - F(f_2) = \int_0^1 F'[f_2 + t(f_1 - f_2)](f_1 - f_2) \, dt$$

$$= \left(\int_0^1 (R(f_2 + t(f_1 - f_2), f) - \mathrm{Id}) \, dt + \mathrm{Id} \right) F'[f](f_1 - f_2).$$

As $\|\int_0^1 (R(f_2 + t(f_1 - f_2), f) - \mathrm{Id}) \, dt\| \leq \frac{1}{2}$, we obtain

$$\frac{1}{2}\|F'[f](f_1 - f_2)\|_{\mathbb{Y}} \leq \|F(f_1) - F(f_2)\|_{\mathbb{Y}} \leq \frac{3}{2}\|F'[f](f_1 - f_2)\|_{\mathbb{Y}}.$$

Together with (2.19) this yields the assertion. $\qquad \square$

The conditions (2.19) and (2.20) have been verified for the following nonlinear inverse problems:

Example 2.21 (Identification of a reaction coefficient c (cf.[64, Ex. 2.8] and [85, Ex. 2])). *Let $\Omega \subset \mathbb{R}^d$, $d \in \{1, 2, 3\}$ be a bounded Lipschitz domain, $f : \Omega \to [0, \infty)$ and $g : \partial\Omega \to (0, \infty)$ be smooth functions. For $c \in L^\infty(\Omega)$ with $c \geq 0$ almost everywhere, we define $F(c) := u$ where u solves the elliptic boundary value problem*

$$
\begin{aligned}
-\Delta u + cu &= f & &\text{in } \Omega, \\
u &= g & &\text{on } \partial\Omega.
\end{aligned}
\tag{2.21}
$$

For this problem (2.19) with $\mathbb{Y} = L^2(\Omega)$ and $a = 2$ has been shown in [65, Thm. 4.5], and (2.20) in [54, Ex. 4.2]. Hence Assumption 2.15 with $a = 2$ holds true in some L^2-neighborhood of a reference solution $c_0 \in L^\infty(\Omega)$, $c_0 \geq 0$. (Note that for coefficients c with arbitrary negative values uniqueness in the boundary value problem (2.21) may fail and every L^2-ball contains functions with negative values on a set of positive measure, well-posedness of (2.21) can still be established for all c in a sufficiently small L^2-ball centered at c_0. This can be achieved by Banach's fixed point theorem applied to $u = u_0 + (-\Delta + c_0)^{-1}(u(c_0 - c))$ where $u_0 := G(c_0)$ and $(-\Delta + c_0)^{-1}\tilde{f}$ solves (2.21) with $c = c_0$, $f = \tilde{f}$ and $g = 0$, using the fact that $(-\Delta + c_0)^{-1}$ maps boundedly from $L^1(\Omega) \subset H^{-2}(\Omega)$ to $L^2(\Omega)$ for $d \leq 3$.)

Example 2.22 (Identification of a diffusion coefficient ρ (cf.[64, Ex. 2.8])). *Given $\rho \in L^\infty((0,1))$ with $\rho \geq \underline{\rho} > 0$, we define $F(\rho) := u$ where u solves the boundary value problem*

$$
\begin{aligned}
-(\rho u')' &= f & &\text{in } (0,1), \\
u(0) &= g_0, \quad u(1) = g_1.
\end{aligned}
\tag{2.22}
$$

Here (2.19) with $a = 1$ and $\mathbb{Y} = L^2([0,1])$ has been verified in [65, Thm. 5.4] (in a Hilbert scale of Sobolev spaces with mean 0 that was shifted for technical reasons). Moreover, (2.20) was shown in [54, Ex. 4.3].

Example 2.23 (Hammerstein integral equations (cf.[64, Ex. 2.8])). *The forward operator is defined by*

$$
(F(f))(t) := \int_0^t \phi(f(s)) \, ds
$$

where $\phi \in C^{2,1}(I)$ on all intervals $I \subset \mathbb{R}$. In this case (2.19) with $a = 1$ and $\mathbb{Y} = L^2([0,1])$ is shown in [89, §4], and (2.20) in [54, Ex. 4.1].

Example 2.24 (Counterexample: backward heat equation). *For fixed $\bar{t} > 0$ we introduce the heat equation on $\mathbb{T}^d$ given by*

$$
\begin{aligned}
\partial_t u &= \Delta u & &\text{on } \mathbb{T}^d \times (0, \bar{t}) \\
u(\cdot, 0) &= f & &\text{on } \mathbb{T}^d.
\end{aligned}
$$

We consider the linear forward operator given by $A_{BH}f = u(\cdot, \bar{t})$.
We will see that with $\mathbb{Y} = L^2(\mathbb{T}^d)$ there is no $a > 0$ such that A satisfies Assumption 2.15(a) while it satisfies Assumption 2.15(b) for all $a > 0$.
With e_k and $\hat{f}$ as in (2.8) it is routine to see that for a right hand side $f \in \mathcal{D}'(\mathbb{T}^d)$ the function

$$
u(x, t) = \sum_{k \in \mathbb{Z}^d} e^{-t\|k\|_2^2} \hat{f}(k) e_k(x)
$$

solves the initial value problem. Since $\left(e^{-\bar{t}\|k\|_2^2}\hat{f}(k)\right)_{k\in\mathbb{Z}^d} \in \ell^2(\mathbb{Z}^d)$ *for all* $f \in \mathcal{D}'(\mathbb{T}^d)$ *we obtain* $A_{BH}\colon \mathcal{D}'(\mathbb{T}^d) \to L^2(\mathbb{T}^d)$ *with*

$$A_{BH}f = \sum_{k\in\mathbb{Z}^d} e^{-\bar{t}\|k\|_2^2}\hat{f}(k)e_k. \tag{2.23}$$

Now let $a > 0$. *First we show that Assumption 2.15(a) is violated. For* $n \in \mathbb{N}$ *let us consider* $f_n = n^a e_{(n,\dots,0)}$. *Then on the one hand* $(f_n)_{n\in\mathbb{N}_0}$ *is bounded in* $B_{2,2}^{-a}(\mathbb{T}^d)$. *This can be seen either using the norm in (2.11) or the norm on* $H^{-a}(\mathbb{T}^d) = B_{2,2}^{-a}(\mathbb{T}^d)$ *briefly recalled in the appendix (see above Proposition A.11). On the other hand, we have*

$$\|A_{BH}f_n\|_{L^2(\mathbb{T}^d)} = n^a e^{-\bar{t}n^2} \longrightarrow 0 \quad \text{for} \quad n \to \infty.$$

This excludes the existence of a constant $M_1 > 0$ *such that* $F = A_{BH}$ *satisfies Assumption 2.15(a).*

Moreover, we prove that A_{BH} *maps* $H^{-a}(\mathbb{T}^d) = B_{2,2}^{-a}(\mathbb{T}^d)$ *boundedly to* $L^2(\Omega)$, *i.e. it satisfies Assumption 2.15(b) for all* $a > 0$. *To this end, let* $c_a \geq 1$ *such that* $e^{-2\bar{t}x} \leq c_a(1+x)^{-\frac{a}{2}}$ *for all* $x \geq 0$. *Then we obtain*

$$\|A_{BH}f\|_{L^2(\mathbb{T}^d)}^2 = \sum_{k\in\mathbb{Z}^d} e^{-2\bar{t}\|k\|_2^2}|\hat{f}(k)|^2 \leq c_a \sum_{k\in\mathbb{Z}^d} (1+\|k\|_2^2)^{-\frac{a}{2}} = \|f\|_{H^{-a}(\mathbb{T}^d)}^2$$

for all $f \in H^{-a}(\mathbb{T}^d)$. *Therefore, in contrast to the examples above, one calls* A_{BH} *an infinitely smoothing operator.*

Remark 2.25 (Globality of the assumptions). *In the literature on iterative regularization methods for nonlinear problems similar conditions are typically formulated as a local condition (see e.g. [71, (2.4)]). However, for Tikhonov regularization conditions on the global behavior seem to be necessary. Here we require the two-sided Lipschitz condition globally on the domain* dom(F). *This entails that we have to assume (2.19) uniformly for all* $f \in$ dom(F) *with a joint constant* $\tilde{M}$.

Remark 2.26 (Degree of ill-posedness (cf.[63, Ex. 3])). *In [63] the authors introduce the local degree of ill-posedness of a nonlinear map* $F\colon D_F \to \mathbb{Y}$ *mapping from a subset* D_F *of a Hilbert space* $\mathbb{X}$ *to a Hilbert space* $\mathbb{Y}$ *and* $f \in D_F$. *By [63, Prop. 7] the local degree of ill-posedness of* F *in* f *coincides with the speed of decay of the singular values of a compact linear operator* G *if*

$$K_1\|G(h-f)\|_{\mathbb{Y}} \leq \|F(h) - F(f)\|_{\mathbb{Y}} \leq K_2\|G(h-f)\|_{\mathbb{Y}} \tag{2.24}$$

holds true for some constants $K_1, K_2 > 0$ *and for all* h *in a neighborhood of* f.

Suppose $\mathbb{Y} = \mathbb{X} = L^2(\Omega)$ *for* Ω *the d-dimensional torus or a bounded Lipschitz domain* $\Omega \subset \mathbb{R}^d$ *and* $F\colon D_F \to \mathbb{Y}$ *is a forward operator such that both inequalities in Assumption 2.15 hold true for some* $a > 0$ *and* $B_{2,2}^{-a} = B_{2,2}^{-a}(\Omega)$. *With* $(\psi_{j,k})_{(j,k)\in\Lambda}$ *a wavelet system satisfying Assumption 2.11 with* $s_{max} > a$ *and (2.16) we define*

$$G\colon \mathbb{X} \to \mathbb{Y} \quad \text{by} \quad f \mapsto \sum_{(j,k)\in\Lambda} 2^{-ja}\langle f, \psi_{j,k}\rangle \psi_{j,k}.$$

Then G *is compact and satisfies* $\|G \cdot\|_{\mathbb{Y}} \sim \|\cdot \,| B_{2,2}^{-a}\|$. *Therefore, condition (2.24) holds true. The operator* G *has the singular value* 2^{-ja} *with multiplicity* $|\Lambda_j|$. *Numerating these singular values we obtain* $(\lambda_n)_{n\in\mathbb{N}}$ *with* $|\{n \in \mathbb{N}\colon \lambda_n = 2^{-ja}\}| = |\Lambda_j|$ *and using* $|\Lambda_j| \sim 2^{jd}$ *we see that the singular values decay as* $\lambda_n = \mathcal{O}(n^{-\frac{a}{d}})$ *(see (2.16)). Hence, the local degree of ill-posedness of* F *is* $\frac{a}{d}$ *for all* $f \in D_F$.

2.5   Weighed ℓ^p-regularization

In the case $r \geq 0$ and $p = q$ the norm

$$\|x\|_{r,p,p}^p = \sum_{j=0}^{\infty} 2^{jrp} 2^{jdp\left(\frac{1}{2}-\frac{1}{p}\right)} \|x_j\|_p^p = \sum_{(j,k)\in\Lambda} 2^{jrp} 2^{jdp\left(\frac{1}{2}-\frac{1}{p}\right)} |x_{j,k}|^p$$

on the Besov sequence space $b_{p,p}^r$ is a weighed ℓ^p-norm. The particular sequence of weights is not of structural importance in most of the results in this work. As it also allows a better comparison to the literature we introduce weighed ℓ^p-penalties and provide the theory for those with the case of Besov space regularization with $p = q$ as the main application.

To fix the notation we start by introducing weighed ℓ^p-spaces. Let Λ be a countable index set, $p \in (0,\infty)$ and $\overline{\omega} = (\overline{\omega}_j)_{j\in\Lambda}$ a sequence of positive reals. We consider $\ell_{\overline{\omega}}^p$ defined by

$$\ell_{\overline{\omega}}^p = \left\{ x \in \mathbb{R}^\Lambda \colon \|x\|_{\overline{\omega},p} < \infty \right\} \quad \text{with} \quad \|x\|_{\overline{\omega},p} = \left(\sum_{j\in\Lambda} \overline{\omega}_j^p |x_j|^p \right)^{\frac{1}{p}}.$$

For $p \in [1,\infty)$ the vector spaces $\ell_{\overline{\omega}}^p$ are Banach spaces and for $p \in (0,1)$ quasi-Banach spaces. In the latter case the best possible constant in the generalized triangle inequality is $C = 2^{\frac{1}{p}-1}$.

Now let $p, u \in [1,\infty)$, $\overline{r} = (\overline{r}_j)_{j\in\Lambda}$ a sequence of positive weights and $G \colon \mathrm{dom}(G) \to \mathbb{Y}$ a map defined on a set of sequences $\mathrm{dom}(G) \subset \mathbb{R}^\Lambda$ with values in a Banach space $\mathbb{Y}$. For $g \in \mathbb{Y}$ we consider the following set-valued scheme

$$S_\alpha(g) = \operatorname*{argmin}_{x\in\ell_r^p\cap\mathrm{dom}(G)} \left(\frac{1}{2\alpha} \|g - G(x)\|_{\mathbb{Y}}^2 + \frac{1}{u} \|x\|_{\overline{r},p}^u \right). \tag{2.25}$$

We formulate an assumption on the forward operator G and discuss its relation to Assumption 2.15.

Assumption 2.27. *Let $\mathbb{Y}$ be a Banach space, $\overline{a} = (\overline{a}_j)_{j\in\Lambda}$ a sequence of positive real numbers, $\mathrm{dom}(G) \subseteq \ell_{\overline{a}}^2$ closed and $G \colon \mathrm{dom}(G) \to \mathbb{Y}$ a map. We assume that there exist constants $M_1, M_2 > 0$ with*

$$\frac{1}{M_1} \|z_1 - z_2\|_{\overline{a},2} \leq \|G(z_1) - G(z_2)\|_{\mathbb{Y}} \leq M_2 \|z_1 - z_2\|_{\overline{a},2} \quad \textit{for all } z_1, z_2 \in \mathrm{dom}(G).$$

For a finitely smoothing map F and a wavelet system which is regular enough we show that the composition $F \circ \mathcal{S}$ meets Assumption 2.27.

Remark 2.28 (Connection to Besov space regularization). *Suppose Assumptions 2.11 and 2.15 hold true with $a < s_{max}$. Let Λ as in Section 2.4. We set $\overline{a}_{(j,k)} := 2^{-ja}$. Then $b_{2,2}^{-a} = \ell_{\overline{a}}^2$ with equal norms. The wavelet synthesis operator is an isomorphism $\mathcal{S} \colon b_{2,2}^{-a} \to B_{2,2}^{-a}$. Hence $\mathrm{dom}(G) := \mathcal{S}^{-1}(\mathrm{dom}(F))$ is closed in $b_{2,2}^{-a}$. Denoting S_+ the operator norm of $\mathcal{S}$ and S_- the operator norm of its inverse we transfer Assumption 2.15 to*

$$\frac{1}{M_1 S_-} \|z_1 - z_2\|_{-a,2,2} \leq \|F(\mathcal{S}z_1) - F(\mathcal{S}z_2)\|_{\mathbb{Y}} \leq M_2 S_+ \|z_1 - z_2\|_{-a,2,2} \tag{2.26}$$

for all $z_1, z_2 \in \mathrm{dom}(G)$. Hence with the composition $G := F \circ \mathcal{S}$ Assumption 2.27 holds true.

Remark 2.29 (Existence of minimizers). *Under Assumption 2.27 existence of minimizers in (2.25) for $p \in [1,2]$ can be shown if $\ell^p_{\bar{r}} \cap \mathrm{dom}(G) \neq \emptyset$ and $\bar{a}_j \bar{r}_j \to 0$ meaning that for every $\varepsilon > 0$ all but finitely many $j \in \Lambda$ satisfy $\bar{a}_j \bar{r}_j^{-1} \leq \varepsilon$. Alternatively one can also require $\bar{a}_j \bar{r}_j^{-1}$ is bounded above and additional assume that $\mathrm{dom}(G)$ is weakly closed and G is weak-to-weak continuous. The proof of the existence can be carried out with the same ideas as in Proposition 2.3. As our main application is Besov space wavelet regularization with $p = q$, where we can bear on Proposition 2.3, we refrain of going into detail here.*

Chapter Three

Error bounds based on variational source conditions

This chapter focuses on the verification of variational source conditions for sparsity promoting wavelet regularization and the convergence rates that we can deduce from those. In a first section we will recall the concept of variational source conditions and see that they imply approximation and convergence rates. Next we show a first strategy for the verification of variational source conditions and apply it to the situation of wavelet regularization of a finitely smoothing operator with a wavelet system that is sufficiently regular in relation to the degree of smoothing of the forward operator. Generalizing the strategy of verification allows to drop this regularity condition. We provide minimax order optimality on balls in $B_{p,\infty}^s$ of our achieved error bounds on these sets. Moreover, we show convergence rates for weighed ℓ^1-regularization under smoothness conditions given by weak sequence spaces and show that the application to wavelet regularization with Besov $0, 1, 1$-penalties improves the convergence rates for piecewise smooth functions with kinks or jumps. We confirm our theoretical results with numerical simulations with a nonlinear parameter identification problem in an elliptic differential equation.

We extend our theory to show that it can also be applied to infinitely smoothing forward operators. As an example we provide a convergence rate theorem for the backward heat equation. Furthermore, we use the purely deterministic variational source conditions to provide error bounds also for statistical noise models. Finally, in the last section of this chapter, we show that in the our theory can be adapted to regularization with wavelet frames.

3.1   Variational source conditions

We recall the concept of variational source conditions. To do so, we revisit generalized Tikhonov regularization with regularizers given by

$$S_\alpha(g) = \operatorname*{argmin}_{z \in D} \left(\frac{1}{2\alpha} \|g - G(z)\|_{\mathbb{Y}}^2 + \mathcal{R}(z) \right).$$

Throughout this subsection we assume that $\mathbb{X}$ and $\mathbb{Y}$ are Banach spaces, $\mathcal{R} \colon \mathbb{X} \to [0, \infty]$ is convex and $G \colon D \to \mathbb{Y}$ is a map defined on a subset $D \subset \mathbb{X}$. The reason for the change of notation in comparison to the first chapter is that we mostly apply the following results to Besov sequence spaces and maps $G = F \circ \mathcal{S}$ that are compositions of forward maps defined on function spaces and wavelet synthesis operators.

Let $L\colon \mathbb{X} \times \mathbb{X} \to [0,\infty)$ be a map and $\varphi\colon [0,\infty) \to [0,\infty)$ concave and continuous. We say that a true solution $x \in D$ satisfies a *variational source condition* with index function φ and with respect to the *loss function L* if

$$L\left(x,z\right) + \mathcal{R}(x) - \mathcal{R}(z) \leq \varphi\left(\|G(x) - G(z)\|_{\mathbb{Y}}^2\right) \quad \text{for all } z \in D. \tag{3.1}$$

The first variant of such a condition using the Bregman distance as loss function appears in [58]. Moreover, we refer to [93, Sec. 3.2] and [96, Sec. 3.2.3] for further readings on the topic. In [66] it was shown that variational source conditions are not only sufficient but even necessary for convergence rates in a Hilbert space setting.

In order to recall that a variational source condition implies a convergence rate we introduce the function

$$\breve{\varphi}\colon (0,\infty) \to [0,\infty] \quad \text{given by} \quad \breve{\varphi}(\alpha) = \sup_{\tau \geq 0}\left(\varphi(\tau) - \frac{1}{2\alpha}\tau\right) \tag{3.2}$$

that will serve as an upper bound on the reconstruction error.

We consider two parameter choice rules for the regularization parameter α. An a priori rule requiring prior knowledge of the function φ in (3.1) characterizing the regularity of the unknown x, and the *discrepancy principle* as most well-known a posteriori rule. The following result is a variant of [49, Thm. 3.3, Cor. 3.4], [105, Thm. 3.3] or [67, Thm. 2.3]. Convergence rates for the discrepancy principle under variational source conditions can also be found in [4]. The result we present here for this a posteriori rule is similar to [40, Prop. 14].

Proposition 3.1 (Convergence rate under variational source condition). *Assume $x \in D$ satisfies (3.1). Suppose $\delta \geq 0$, $g^{\mathrm{obs}} \in \mathbb{Y}$ with $\|g^{\mathrm{obs}} - G(x)\|_{\mathbb{Y}} \leq \delta$. Let $\alpha > 0$ and $\hat{x}_\alpha \in S_\alpha(g^{\mathrm{obs}})$.*

(a) (error bounds) The following bounds

$$L(x,\hat{x}_\alpha) \leq \frac{\delta^2}{\alpha} + \breve{\varphi}(4\alpha) \quad and \tag{3.3}$$

$$\|G(x) - G(\hat{x}_\alpha)\|_{\mathbb{Y}} \leq 2\sqrt{2}\left(\delta + \sqrt{\alpha\breve{\varphi}(4\alpha)}\right) \tag{3.4}$$

hold true.

(b) (rates with discrepancy principle) Let $1 \leq c_D \leq C_D$. If $c_D\delta \leq \|g^{\mathrm{obs}} - G(\hat{x}_\alpha)\|_{\mathbb{Y}} \leq C_D\delta$ then

$$L(x,\hat{x}_\alpha) \leq \varphi\left((1+C_D)^2\delta^2\right) \quad and \tag{3.5}$$

$$\|G(x) - G(\hat{x}_\alpha)\|_{\mathbb{Y}} \leq (1+C_D)\delta. \tag{3.6}$$

Proof. (a) As $\hat{x}_\alpha$ minimizes the Tikhonov functional we obtain

$$\frac{1}{2\alpha}\|g^{\mathrm{obs}} - G(\hat{x}_\alpha)\|_{\mathbb{Y}}^2 + \mathcal{R}(\hat{x}_\alpha) \leq \frac{\delta^2}{2\alpha} + \mathcal{R}(x). \tag{3.7}$$

Using the basic inequality $(a + b)^2 \leq 2a^2 + 2b^2$ for $a, b \in \mathbb{R}$ and the inequalities (3.7), (3.1) we bound

$$L(x, \hat{x}_\alpha) + \frac{1}{4\alpha} \|G(x) - G(\hat{x}_\alpha)\|_{\mathbb{Y}}^2 \leq L(x, \hat{x}_\alpha) + \frac{\delta^2}{2\alpha} + \frac{1}{2\alpha} \|g^{\mathrm{obs}} - G(\hat{x}_\alpha)\|_{\mathbb{Y}}^2$$

$$\leq \frac{\delta^2}{\alpha} + L(x, \hat{x}_\alpha) + \mathcal{R}(x) - \mathcal{R}(\hat{x}_\alpha)$$

$$\leq \frac{\delta^2}{\alpha} + \varphi\left(\|G(x) - G(\hat{x}_\alpha)\|_{\mathbb{Y}}^2\right).$$

Subtracting $\frac{1}{8\alpha} \|G(x) - G(\hat{x}_\alpha)\|_{\mathbb{Y}}^2$ on both sides and the definition of $\breve{\varphi}$ yield

$$L(x, \hat{x}_\alpha) + \frac{1}{8\alpha} \|G(x) - G(\hat{x}_\alpha)\|_{\mathbb{Y}}^2 \leq \frac{\delta^2}{\alpha} + \breve{\varphi}(4\alpha).$$

From this (3.3) follows immediately from neglecting $\frac{1}{8\alpha} \|G(x) - G(\hat{x}_\alpha)\|_{\mathbb{Y}}^2$ and (3.4) follows from neglecting $L(x, \hat{x}_\alpha)$, multiplying by α and using the basic inequality $a^2 + b^2 \leq (a + b)^2$ for $a, b \geq 0$.

(b) We obtain (3.6) using the triangle inequality

$$\|G(x) - G(\hat{x}_\alpha)\|_{\mathbb{Y}} \leq \delta + \|g^{\mathrm{obs}} - G(\hat{x}_\alpha)\|_{\mathbb{Y}} \leq (1 + C_D)\delta.$$

Inserting the left inequality of the assumption into (3.7) we obtain

$$c_D^2 \frac{\delta^2}{2\alpha} + \mathcal{R}(\hat{x}_\alpha) \leq \frac{\delta^2}{2\alpha} + \mathcal{R}(x).$$

Hence, $\mathcal{R}(\hat{x}_\alpha) \leq \mathcal{R}(x)$ and (3.1) yields

$$L(x, \hat{x}_\alpha) \leq L(x, \hat{x}_\alpha) + \mathcal{R}(x) - \mathcal{R}(\hat{x}_\alpha) \leq \varphi\left(\|G(x) - G(\hat{x}_\alpha)\|_{\mathbb{Y}}^2\right).$$

As the function φ is non-decreasing (see Lemma A.1) we obtain (3.5) from (3.6) together with the last inequality. $\qquad\square$

In the following we specialize Proposition 3.1 to Hölder-type upper bounds.

Corollary 3.2 (Convergence rate under Hölder-type variational source condition). *Assume* $x \in D$ *satisfies* (3.1) *with* $\varphi(\tau) = c\tau^\nu$ *for* $c > 0$ *and* $\nu \in (0, \frac{1}{2}]$. *There exits a constant* $k_\nu > 0$ *depending only on* ν, *such that the following statements hold true.*

(a) (bias bounds) Let $\alpha > 0$ *and* $x_\alpha \in S_\alpha(G(x))$. *Then*

$$L(x, x_\alpha) \leq k_\nu c^{\frac{1}{1-\nu}} \alpha^{\frac{\nu}{1-\nu}} \quad and \tag{3.8}$$

$$\|G(x) - G(x_\alpha)\|_{\mathbb{Y}} \leq k_\nu c^{\frac{1}{2-2\nu}} \alpha^{\frac{1}{2-2\nu}}. \tag{3.9}$$

Suppose $\delta > 0$, $g^{\mathrm{obs}} \in \mathbb{Y}$ *with* $\|g^{\mathrm{obs}} - G(x)\|_{\mathbb{Y}} \leq \delta$. *Let* $\alpha > 0$ *and* $\hat{x}_\alpha \in S_\alpha(g^{\mathrm{obs}})$.

(b) (a priori rule) Let $0 < c_l \leq c_r$. *If* $c_l c^{-1} \delta^{2-2\nu} \leq \alpha \leq c_r c^{-1} \delta^{2-2\nu}$ *then*

$$L(x, \hat{x}_\alpha) \leq k_\nu \left(c_l^{-1} + c_r^{\frac{\nu}{1-\nu}}\right) c\delta^{2\nu} \quad and \tag{3.10}$$

$$\|G(x) - G(\hat{x}_\alpha)\|_{\mathbb{Y}} \leq k_\nu \left(1 + c_r^{\frac{1}{2-2\nu}}\right) \delta. \tag{3.11}$$

(c) (discrepancy principle) Let $1 \leq c_D \leq C_D$. If $c_D \delta \leq \|g^{\mathrm{obs}} - G(\hat{x}_\alpha)\|_{\mathbb{Y}} \leq C_D \delta$ then

$$L(x, \hat{x}_\alpha) \leq (1 + C_D)^{2\nu} c \delta^{2\nu} \quad and \tag{3.12}$$

$$\|G(x) - G(\hat{x}_\alpha)\|_{\mathbb{Y}} \leq (1 + C_D)\delta. \tag{3.13}$$

Proof. An easy calculus shows $\breve{\varphi}(\alpha) = k'_\nu c^{\frac{1}{1-\nu}} \alpha^{\frac{\nu}{1-\nu}}$ with $k'_\nu = (2\nu)^{\frac{\nu}{1-\nu}} - \frac{1}{2}(2\nu)^{\frac{1}{1-\nu}}$. Proposition 3.1(a) yields *(a)* by inserting $\delta = 0$ and *(b)* by inserting the parameter choice rule. Statement *(c)* follows immediately from Proposition 3.1(b). $\qquad\square$

Remark 3.3 (The restriction $\nu \leq \frac{1}{2}$ (cf. [59, Prop. 3]))**.** *Let us comment on the restriction $\nu \leq \frac{1}{2}$ in Corollary 3.2. To this end, we assume that G satisfies a Lipschitz condition $\|G(z_1) - G(z_2) \leq M\|z_1 - z_2\|_{Lip}$ for some norm $\|\cdot\|_{Lip}$ defined on X and that D is convex with $0 \in D$. Suppose $x \in \mathbb{X}$ satisfies (3.1) with a concave and continuous function φ decaying faster than the square root function, i.e. we assume that $\lim_{\tau \to 0} \tau^{-\frac{1}{2}} \varphi(\tau) = 0$. We repeat an argument that can be found in [59, Prop. 3] to show that we obtain $x \in \mathrm{argmin}_{z \in D} \mathcal{R}(z)$. First note that neglecting the loss function in (3.1) and the Lipschitz condition yield*

$$\mathcal{R}(x) - \mathcal{R}(z) \leq \varphi\left(M^2 \|x - z\|_{Lip}^2\right) \quad for\ all\ z \in D.$$

Assuming $z \neq x$ and inserting a convex combination $\theta z + (1 - \theta)x$ with $\theta \in (0, 1)$ we wind up with

$$\mathcal{R}(x) - \mathcal{R}(z) \leq \frac{1}{\theta} \varphi\left(\theta^2 M^2 \|x - z\|_{Lip}^2\right) = \frac{M\|x - z\|_{Lip}}{\theta M\|x - z\|_{Lip}} \varphi\left(\theta^2 M^2 \|x - z\|_{Lip}^2\right)$$

by convexity of $\mathcal{R}$. As the right hand side vanishes for $\theta \to 0$ we conclude $\mathcal{R}(x) \leq \mathcal{R}(z)$. Having in mind that in our examples $\mathcal{R}$ is a norm defined on $\mathbb{X}$, this shows that 0 is the only element satisfying a variational source condition (3.1) with a function φ decaying faster than the square root. Hence, the restriction to $\nu \leq \frac{1}{2}$ does not exclude any relevant unknown solutions.

Variational source conditions for ℓ^1-type regularization

In this chapter we focus on the choices $\mathcal{R}(z) = \|z\|_{\mathbb{X}}$ and $L(z_1, z_2) = \beta\|z_1 - z_2\|_{\mathbb{X}}$ for some $\beta \in (0, 1]$ such that (3.1) becomes

$$\beta\|x - z\|_{\mathbb{X}} + \|x\|_{\mathbb{X}} - \|z\|_{\mathbb{X}} \leq \varphi\left(\|G(x) - G(z)\|_{\mathbb{Y}}^2\right) \quad \text{for all } z \in D. \tag{3.14}$$

Variational source conditions of the form (3.14) are used in the references [11, 41, 43, 40] which we already discussed in more detail in Section 2.2. Moreover, the author made use of them in the articles [64, 85].

Remark 3.4 (preasymptotic error bounds)**.** *Let $x \in \mathbb{X}$. Then the triangle inequality yields*

$$\beta\|x - z\|_{\mathbb{X}} + \|x\|_{\mathbb{X}} - \|z\|_{\mathbb{X}} \leq (\beta + 1)\|x\|_{\mathbb{X}} + (\beta - 1)\|z\|_{\mathbb{X}} \leq (\beta + 1)\|x\|_{\mathbb{X}} \tag{3.15}$$

for all $z \in \mathbb{X}$. This will be advantageous in the proofs later in the following way: Suppose (3.14) holds true with some $\beta \in (0, 1]$ a concave and continuous function φ. Then (3.15)

implies that (3.14) *holds true with* φ *replaced by* $\overline{\varphi}(\tau) := \min\{\varphi(\tau), (\beta + 1)\|x\|_{\mathbb{X}}\}$. *Here* φ *and* $\overline{\varphi}$ *share the same asymptotic behavior near* 0.
For instance suppose φ *satisfies a Hölder-type bound near zero, i.e. there exists* $\nu \in (0, 1/2]$ *and* τ_0, $c > 0$ *such that* $\varphi(\tau) \leq c\tau^\nu$ *for all* $0 \leq \tau \leq \tau_0$. *Then applying Proposition 3.1 directly yields Hölder-type error bounds only for* δ *small enough. As* $\overline{\varphi}$ *is bounded we may increase the constant* c *such that we obtain* $\overline{\varphi}(\tau) \leq \overline{c}\tau^\nu$ *globally for all* $\tau \geq 0$. *Therefore, Corollary 3.2 yields preasymptotic error bounds.*

3.2 Smooth wavelet basis

The central result of this section is a convergence rate for 1-homogeneous wavelet regularization in Besov spaces. We analyze the regularization scheme introduced in Section 2.4 with $u = q = 1$, $0 \leq r < s_{\max}$ and $p \in [1, 2]$, i.e. for an observation $g \in \mathbb{Y}$ the estimators we consider are given by

$$R_\alpha(g) = \mathcal{S}\hat{x}_\alpha \quad \text{with} \quad \hat{x}_\alpha \in \operatorname*{argmin}_{x \in D} \left(\frac{1}{2\alpha} \|g - F(\mathcal{S}x)\|_{\mathbb{Y}}^2 + \|x\|_{r,p,1} \right) \tag{3.16}$$

with $D := \mathcal{S}^{-1}(B_{p,1}^r \cap \operatorname{dom}(F)) \subseteq b_{p,1}^r$. Note that the penalty term $\|x\|_{r,p,1}$ is a weighed sum of level-wise p-norms of wavelet coefficients. In the special case that F is linear and allows a wavelet vaguelette decomposition and the norm in $\mathbb{Y}$ is chosen as ℓ^2-norm of vaguelette coefficients, the estimators for $p = 1$ in (3.16) coincide with wavelet shrinkage estimators studied in [31] and [76].
For an at most a-times smoothing operator (see Assumption 2.15(a)) and in the case $r = 0$ we prove the convergence rate $\|f - \hat{f}_\alpha\|_{L^p(\Omega)} = \mathcal{O}\left(\delta^{\frac{s}{s+a}}\right)$ for $f \in B_{p,\infty}^s$ and $\hat{f}_\alpha \in R_\alpha(g^{\mathrm{obs}})$ with $\|g^{\mathrm{obs}} - F(f)\|_{\mathbb{Y}} \leq \delta$.
In this section we consider wavelet systems that are smooth enough to characterize Besov spaces with smoothness $-a$ by decay properties of wavelet expansions, i.e. with the notation of Assumption 2.15 and 2.11 we assume $a < s_{\max}$. This restriction is also required in the results of the article [64]. We will overcome this restriction by a generalization in the next section.

The proof of the convergence rate result is based on a variational source condition of the form (3.14) with $\mathbb{X} = b_{p,1}^r$ the and $G = F \circ \mathcal{S}$ the composition of the forward map F and the wavelet synthesis operator $\mathcal{S}$ (see Assumption 2.11). From the latter condition we obtain an error bound in the space of wavelet coefficients $b_{p,1}^r$ which then in turn translates into a bound measured in $B_{p,1}^r$. If $B_{p,q}^s = B_{p,q}^s(\Omega)$ for some suitable domain Ω, then we obtain bounds in an $L^{\overline{p}}(\Omega)$-norm with $p \leq \overline{p} \leq 2$ by real interpolation.
Before we go into the details, let us sketch the basic idea of the verification of condition (3.14). We introduce the projection $P_n \colon b_{p,1}^r \to b_{p,1}^r$ onto the first $n + 1$ levels given by $(P_n(x))_j = x_j$ if $0 \leq j \leq n$ and $(P_n(x))_j = 0$ if $j > n$. Then the Pythagoras type equality

$$\|z\|_{r,p,1} = \|P_n z\|_{r,p,1} + \|(I - P_n)z\|_{r,p,1}$$

holds true for all $z \in b_{p,1}^r$ and $n \in \mathbb{N}_0$.
Let $x, z \in b_{p,1}^r$ and $n \in \mathbb{N}_0$.

The starting point is the following inequality that has been also used in [11, Lemma 5.1]

$$\|x - z\|_{r,p,1} + \|x\|_{r,p,1} - \|z\|_{r,p,1} \le 2\left(\|P_n(x - z)\|_{r,p,1} + \|(I - P_n)x\|_{r,p,1}\right). \qquad (3.17)$$

From there we proceed with a Bernstein type inequality $\|P_n\cdot\|_{r,p,1} \lesssim 2^{n(a+r)}\|\cdot\|_{-a,p,\infty}$ to handle the first summand. The Jackson type inequality $\|(I - P_n)\cdot\|_{r,p,1} \lesssim 2^{n(r-s)}\|\cdot\|_{s,p,\infty}$ provides a bound on the second summand. Minimizing over n leads to

$$\|x - z\|_{r,p,1} + \|x\|_{r,p,1} - \|z\|_{r,p,1} \le c\|x\|_{s,p,\infty}^{\frac{a+r}{s+a}} \cdot \|x - z\|_{-a,p,\infty}^{\frac{s-r}{s+a}}. \qquad (3.18)$$

Note that this inequality is a variational source condition for the embedding operator $b_{p,1}^r \subset b_{p,\infty}^{-a}$. As we assume that the wavelet system is smooth enough to characterize the Besov function spaces with smoothness index $-a$ Assumption 2.15(a) directly transfers to an inequality in terms of wavelet coefficients, namely $\|x - z\|_{-a,p,\infty} \lesssim \|F(\mathcal{S}x) - F(\mathcal{S}z)\|$ and we obtain the desired variational source condition from (3.18).

Before we present the discussed result in detail, we provide a general scheme for the verification of condition (3.18).

Verification of variational source conditions

In [66, Thm. 2.1] and [104, Thm. 3.3] a general strategy for the verification of variational source conditions with Bregman loss has been proposed. The following proposition provides an analog for conditions of the form (3.14). An essential prerequisite is the Pythagoras type equality (3.19a) with exponent 1. The latter restricts the choice penalty to ℓ^1-type norms, i.e. to the choice $q = u = 1$ in (2.17).

Recall that a bounded linear operator $P\colon \mathbb{X} \to \mathbb{X}$ is called *projection* if $P^2 = P$.

Proposition 3.5. *Let $\mathbb{X}$, $\mathbb{Y}$ be Banach spaces and $G\colon D \to \mathbb{Y}$ a map defined on a subset $D \subset \mathbb{X}$. Let $x \in D$ and J be an index set. Suppose $(P_j\colon \mathbb{X} \to \mathbb{X})_{j \in J}$ is a family of projections and $(\nu_j)_{j \in J}$ a family of non-negative real numbers such that the following conditions are satisfied:*

$$\|z\|_{\mathbb{X}} = \|P_j z\|_{\mathbb{X}} + \|(I - P_j)z\|_{\mathbb{X}} \quad \text{for all } z \in D, j \in J, \qquad (3.19a)$$

$$\|P_j(z_1 - z_2)\|_{\mathbb{X}} \le \nu_j\|G(z_1) - G(z_2)\|_{\mathbb{Y}} \quad \text{for all } z_1, z_2 \in D, j \in J. \qquad (3.19b)$$

Then (3.14) holds true with $\beta = 1$ and the continuous, concave function $\varphi\colon [0,\infty) \to [0,\infty)$ given by

$$\varphi(\tau) = 2\inf_{j \in J}\left(\nu_j\sqrt{\tau} + \|(I - P_j)x\|_{\mathbb{X}}\right).$$

Proof. The function φ is concave and upper semi-continuous functions as it is the pointwise infimum of functions possessing these properties. Lemma A.1 provides continuity of φ. Next we turn to the proof of (3.14) (see also [11, Lemma 5.1, Theorem 5.2]). Let $z \in D$ and $j \in J$. With (3.19a) we split every $\|\cdot\|_{\mathbb{X}}$-term into $\|P_j\cdot\|_{\mathbb{X}} + \|(I - P_j)\cdot\|_{\mathbb{X}}$ and use the triangle equality for $\|(I - P_j)\cdot\|_{\mathbb{X}}$-terms and the reverse triangle inequality for $\|P_j\cdot\|_{\mathbb{X}}$-terms to achieve

$$\|x - z\|_{\mathbb{X}} + \|x\|_{\mathbb{X}} - \|z\|_{\mathbb{X}} \le 2\left(\|P_j(x - z)\|_{\mathbb{X}} + \|(I - P_j)x\|_{\mathbb{X}}\right). \qquad (3.20)$$

Taking the infimum over $j \in J$ after inserting (3.19b) on the right hand side yields the result. $\qquad \square$

The condition (3.19b) resembles a *Bernstein inequality*. In the sequel we will see that for a linear operator $G = A$ the latter condition is equivalent to $\mathrm{im}(P_j^*) \subset \mathrm{im}(A^*)$. This explains the connection to smoothness of the basis system as for smoothing operators (e.g. operators satisfying Assumption 2.15) the space $\mathrm{im}(A^*)$ consists of function with a certain regularity, and in the following we will mostly choose P_j as a projection onto certain wavelet coefficients.

The function φ in Proposition 3.5 satisfies $\varphi(0) = 0$ if and only if $\inf_{j \in J} \|(I - P_j)x\|_{\mathbb{X}} = 0$. Moreover, note that (3.19a) implies that the operator norms of P_j and of $(I - P_j)$ are bounded by 1.

The asymptotic behavior of φ is determined by two (generalized) sequences: $(\nu_j)_{j \in J}$ and $(\|(I - P_j)x\|_{\mathbb{X}})_{j \in J}$. Using a simple example we illustrate that the growth of the numbers $(\nu_j)_{j \in J}$ depends on the degree of ill-posedness of the operator and that the decay $(\|(I - P_j)x\|_{\mathbb{X}})_{j \in J}$ can be interpreted as a measure of smoothness of the true solution x.

Example 3.6 (Diagonal operator on sequences). *Suppose $\mathbb{X} = \ell^1 = \ell^1(\mathbb{N})$ and $\mathbb{Y} = \ell^2 = \ell^2(\mathbb{N})$ and $G = D \colon \ell^1 \to \ell^2$ is a diagonal operator, i.e. D is linear and bounded and there exists a sequence $(\mu_n)_{n \in \mathbb{N}}$ of real numbers such that $De_n = \mu_n e_n$ where e_n denotes the unit sequence with a 1 in the n'th entry and 0's else.*

Let $P_n \colon \mathbb{X} \to \mathbb{X}$ be the projection onto the first n entries, that is $(P_n x)_j = x_j$ if $1 \leq j \leq n$ and $(P_n x)_j = 0$ if $j > n$. Then the family $(P_n)_{n \in \mathbb{N}}$ fulfills condition (3.19a). Using the Cauchy-Schwarz inequality we obtain

$$\|P_n z\|_1 = \sum_{j=1}^n |\mu_j|^{-1} |\mu_j| |z_j| \leq \left(\sum_{j=1}^n |\mu_j|^{-2} \right)^{\frac{1}{2}} \left(\sum_{j=1}^n |\mu_j|^2 |z_j|^2 \right)^{\frac{1}{2}} \leq \left(\sum_{j=1}^n |\mu_j|^{-2} \right)^{\frac{1}{2}} \|Dz\|_2$$

(3.21)

for all $z \in \mathbb{X}$. This inequality is sharp for the choice $z \in \mathbb{X}$ with $z_j = \mu_j^{-2}$ if $1 \leq j \leq n$ and $z_j = 0$ for $j > n$. Hence, the smallest possible choice of ν_n such that (3.19b) holds true is $\nu_n = \left(\sum_{j=1}^n |\mu_j|^{-2} \right)^{\frac{1}{2}}$.

Therefore, we see that the growth of the numbers ν_n is determined by the asymptotic behavior of the singular values μ_n. The latter in turn is a measure of ill-posedness of the opertor D (see Remark 2.26 and [63]). Moreover, the asymptotic behavior of $\|(I - P_n)x\|_1 = \sum_{j=n+1}^\infty |x_j|$ measures how fast the entries of x converge to 0. Interpreting x as Fourier coefficients or more generally as the coefficients in some wavelet basis this tail behavior is related to the smoothness of the solution. We refer to [78, Rem. 3.5] for related considerations.

Next we show that for $\mathbb{X} = \ell^1$ the condition (3.19b) is equivalent to a Lipschitz continuity assumption on G^{-1} with respect to a weighed ℓ^∞-space. We refer to [11, Rem. 2.5] for a similar statement in the case of a linear forward operator $G = A$. There, the authors show that the condition $e_j \in \mathrm{im}(A^*)$ implies a lower bound of the form $\|\cdot\|_{\omega,\infty} \leq \|A \cdot \|_{\mathbb{Y}}$ where ω is given by the reciprocals of the norms of the source elements $\psi_i \in \mathbb{Y}^*$ with $A^* \psi_i = e_i$.

Proposition 3.7. *Let $D \subset \mathbb{X} = \ell^1$, $G \colon D \to \mathbb{Y}$ a map and $(P_n)_{n \in \mathbb{N}}$ as in Example 3.6. The following statements are equivalent:*

(i) *There exists a sequence $\omega = (\omega_n)_{n \in \mathbb{N}_0} \subset [0, \infty)$ and $L \geq 0$ such that*

$$\|z_1 - z_2\|_{\omega, \infty} \leq L\|G(z_1) - G(z_2)\|_{\mathbb{Y}} \quad \text{for all } z_1, z_2 \in D. \tag{3.22}$$

(ii) *There exists a sequence of positive real numbers $(\nu_n)_{n \in \mathbb{N}_0}$ such that (3.19b) holds true.*

More precisely, (i) implies (ii) with $\nu_n = L \sum_{j=1}^{n} \omega_j^{-1}$ and (ii) implies (i) with $\omega_n = \nu_n^{-1}$ and $L = 1$.

Proof.

$(i) \Rightarrow (ii)$: With a variant of the estimation (3.21) we prove the following Bernstein type inequality

$$\|P_n z\|_1 = \sum_{j=1}^{n} \omega_j^{-1} \omega_j |z_j| \leq \left(\sum_{j=1}^{n} \omega_j^{-1} \right) \|z\|_{\omega, \infty} \quad \text{for all } z \in \ell_\omega^\infty. \tag{3.23}$$

The implication follows from inserting (3.22) into (3.23) with $z = z_1 - z_2$.

$(ii) \Rightarrow (i)$: Taking the supremum on the left hand side of

$$\nu_n^{-1} |(z_1)_n - (z_2)_n| \leq \nu_n^{-1}\|P_n(z_1 - z_2)\|_1 \leq \|G(z_1) - G(z_2)\|_{\mathbb{Y}}$$

provides this implication. $\square$

The next remark substantiates that the strategy captured in Proposition 3.5 is reserved to penalties that are homogeneous of degree less or equal to 1. Similar ideas as in Proposition 3.5 for non convex regularization with a weighed ℓ^q-norm with $0 < q < 1$ can be found in [3, Sec. 4].

Remark 3.8. *Proposition 3.5 is built upon the inequality (3.20) which seems not to have a straightforward generalization for families of projections $(P_j \colon \mathbb{X} \to \mathbb{X})_{j \in J}$ satisfying a Pythagoras type equality*

$$\|x\|_{\mathbb{X}}^q = \|P_j x\|_{\mathbb{X}}^q + \|(I - P_j)x\|_{\mathbb{X}}^q \quad \text{for all } x \in X, j \in J \tag{3.24}$$

for some $q > 1$ instead of (3.19a). First note that one gets back (3.19a) from (3.20) by fixing $z \in \mathbb{X}$ and choosing $x := P_j z$.
An obvious inequality generalizing (3.20) would be

$$c\|x - z\|_{\mathbb{X}}^q + \|x\|_{\mathbb{X}}^q - \|z\|_{\mathbb{X}}^q \leq C \left(\|P_j(x - z)\|_{\mathbb{X}}^q + \|(I - P_j)x\|_{\mathbb{X}}^q \right) \tag{3.25}$$

for all $x, z \in \mathbb{X}, j \in J$ with some constants $0 < c < C$. Assuming J is directed with $\lim_{j \to \infty} \|(I - P_j)x\|_{\mathbb{X}} = 0$ and $x \neq 0$ we will show that this inequality leads to a contradiction. Let $0 < \lambda < 1$. Inserting $z = \lambda x$ into (3.25) yields

$$c \left((1 - \lambda)^q + 1 - \lambda^q \right) \|x\|_{\mathbb{X}}^q \leq C \left((1 - \lambda)^q \|P_j x\|_{\mathbb{X}}^q + \|(I - P_j)x\|_{\mathbb{X}}^q \right).$$

Letting j tend to ∞ using (3.24), dividing by $\|x\|_{\mathbb{X}}$ and rearranging terms leads to

$$c(1 - \lambda^q) \leq (C - c)(1 - \lambda)^q,$$

which is impossible for λ approaching 1 from below.
Another generalization would be of the form

$$c\|x - z\|_{\mathbb{X}} + \|x\|_{\mathbb{X}} - \|z\|_{\mathbb{X}} \leq C \left(\|P_j(x - z)\|_{\mathbb{X}} + \|(I - P_j)x\|_{\mathbb{X}} \right) \tag{3.26}$$

for all $x, z \in \mathbb{X}, j \in J$ with some constants $0 < c < C$. We show that this inequality is false for all $x \neq 0$ with $P_j x = x$ for some $j \in \mathbb{N}_0$. Hence, it is also impossible to show (3.26) for all x belonging to some subspace of $\mathbb{X}$ characterized by the speed of decay of $\|(I - P_j) \cdot \|_{\mathbb{X}}$. Let $\xi \in \ker(P_j)$ with $\|\xi\|_{\mathbb{X}} = 1$ and $\lambda > 0$. Inserting $z = x + \lambda \xi$ in (3.26) implies

$$c\lambda + \|x\|_{\mathbb{X}} \leq \|x + \lambda\xi\|_{\mathbb{X}} = \left(\|x\|_{\mathbb{X}}^q + \lambda^q \right)^{\frac{1}{q}}.$$

This leads to a contradiction if λ approaches 0 from above.

Error bounds for wavelet regularization in Besov spaces with smooth basis

In this subsection we turn to the convergence rate result that we already mentioned at the beginning of this section.

For $n \in \mathbb{N}_0$ we recall the projection $P_n \colon b_{p,1}^r \to b_{p,1}^r$ onto the first $n + 1$ levels given by $(P_n(x))_j = x_j$ if $0 \leq j \leq n$ and $(P_n(x))_j = 0$ if $j > n$. With $\mathbb{X} = b_{p,1}^r$ the family $(P_n)_{n \in \mathbb{N}_0}$ satisfies (3.19a). The next lemma provides a Bernstein type inequality that yields condition (3.19b).

Lemma 3.9 (Bernstein inequality)**.** *Suppose $p \in [1, \infty]$, $r \geq 0$, $a > 0$ and $n \in \mathbb{N}_0$. Then*

$$\|P_n z\|_{r,p,1} \leq C_B 2^{n(a+r)} \|z\|_{-a,p,\infty} \quad \text{for all } z \in b_{p,\infty}^{-a}$$

with $C_B := 2^{a+r} \left(2^{a+r} - 1 \right)^{-1}$.

Proof. We set $\omega = 2^{d\left(\frac{1}{2} - \frac{1}{p} \right)}$ and estimate

$$\|P_n z\|_{r,p,1} = \sum_{j=0}^{n} 2^{j(a+r)} 2^{-ja} \omega^j \|z_j\|_p \leq \left(\sum_{j=0}^{n} 2^{j(a+r)} \right) \|z\|_{-a,p,\infty}.$$

The claim follows from estimating the geometric sum. $\square$

To estimate the $(I - P_n)$-term in the definition of φ in Proposition 3.5 in terms of the the regularity of x, we use the following Jackson type inequality.

Lemma 3.10 (Jackson inequality)**.** *Let $r, s \in \mathbb{R}$ with $r < s$ and $p \in [1, \infty]$. Then*

$$\|(I - P_n) x\|_{r,p,1} \leq C_J 2^{n(r-s)} \|x\|_{s,p,\infty} \quad \text{for all } x \in b_{p,\infty}^s$$

with $C_J = 2^{r-s} \left(1 - 2^{r-s} \right)^{-1}$.

Proof. The claim follows from estimating the geometric series in

$$\|(I - P_n) x\|_{r,p,1} = \sum_{j=n+1}^{\infty} 2^{j(r-s)} 2^{js} 2^{jd\left(\frac{1}{2} - \frac{1}{p} \right)} \|x_j\|_p \leq \left(\sum_{j=n+1}^{\infty} 2^{j(r-s)} \right) \|x\|_{s,p,\infty}. \quad \square$$

Now we are in position to prove a variational source condition for the embedding operator $b_{p,1}^r \subset b_{p,\infty}^{-a}$.

Lemma 3.11 (Variational source condition for embedding operator). *Let $p \in [1,\infty]$, $a > 0$ and $0 \le r < s$. Then there exists a constant $c > 0$ depending only on a, r and s such that*

$$\|x - z\|_{r,p,1} + \|x\|_{r,p,1} - \|z\|_{r,p,1} \le c\|x\|_{s,p,\infty}^{\frac{a+r}{s+a}} \cdot \|x - z\|_{-a,p,\infty}^{\frac{s-r}{s+a}}. \quad \text{for all } x \in b_{p,\infty}^s, z \in b_{p,1}^r.$$

Proof. With a variation of the argument in Lemma 3.10 one proves that there is a continuous embedding $b_{p,\infty}^s \subset b_{p,1}^r$ with an embedding constant c_1 depending only on r and s. With C_J and C_B as above we set $c = 2\max\{c_1, C_B + 2^{s-r}C_J\}$. Now assuming $x \in b_{p,\infty}^s$ and $z \in b_{p,1}^r$ we set $\varrho := \|x\|_{s,p,\infty}$ and $\vartheta = \|x - z\|_{-a,p,\infty}$. If $\varrho \le \vartheta$, then the triangle inequality (see also Remark 3.4) yields

$$\|x - z\|_{r,p,1} + \|x\|_{r,p,1} - \|z\|_{r,p,1} \le 2\|x\|_{r,p,1} \le c\varrho \le c\varrho^{\frac{a+r}{s+a}}\vartheta^{\frac{s-r}{s+a}}.$$

Therefore, we may assume $\vartheta < \varrho$ in the following.
Lemma 3.9 provides $\|P_n(x-z)\|_{r,p,1} \le C_B 2^{n(a+r)}\vartheta$ and from Lemma 3.10 we obtain $\|(I - P_n)x\|_{r,p,1} \le C_J 2^{n(r-s)}\varrho$. Therefore, Proposition 3.5 with G the embedding operator $b_{p,1}^r \subset b_{p,\infty}^{-a}$ and $(P_n)_{n\in\mathbb{N}_0}$ defined above yields

$$\|x - z\|_{r,p,1} + \|x\|_{r,p,1} - \|z\|_{r,p,1} \le 2\inf_{n\in\mathbb{N}_0}\left(C_B 2^{n(a+r)}\vartheta + C_J 2^{n(r-s)}\varrho\right). \quad (3.27)$$

Since $\vartheta < \varrho$ there exists $N \in \mathbb{N}_0$ such that $2^N \le (\varrho/\vartheta)^{\frac{1}{s+a}} \le 2^{N+1}$. By inserting N we estimate the infimum on the right hand side

$$2\inf_{n\in\mathbb{N}_0}\left(C_B 2^{n(a+r)}\vartheta + C_J 2^{n(r-s)}\varrho\right) \le 2C_B 2^{N(a+r)}\vartheta + 2C_J 2^{N(r-s)}\varrho \le c\varrho^{\frac{a+r}{s+a}}\vartheta^{\frac{s-r}{s+a}}.$$

Inserting into (3.27) yields the claim. $\qquad\square$

Now a variational source condition for finitely smoothing operators is an easy consequence.

Proposition 3.12 (Variational source condition). *Suppose Assumptions 2.11 and 2.15(a) hold true with $a < s_{max}$. Let $p \in [1,2]$ and $r \ge 0$. If $p \in [1,2)$, then we require that the scale $B_{p,q}^s$ is p-monotone (see Definition 2.9). Moreover, let $s > r$ and $\varrho > 0$ and assume $x \in b_{p,\infty}^s \cap \mathcal{S}^{-1}(\mathrm{dom}(F))$ with $\|x\|_{s,p,\infty} \le \varrho$. Then*

$$\frac{1}{2}\|x - z\|_{r,p,1} + \|x\|_{r,p,1} - \|z\|_{r,p,1} \le \varphi\left(\|F(\mathcal{S}x) - F(\mathcal{S}z)\|_{\mathbb{Y}}^2\right) \quad (3.28)$$

for all $z \in b_{p,1}^r \cap \mathcal{S}^{-1}(\mathrm{dom}(F))$ with $\varphi(\tau) = C_\varphi \varrho^{\frac{a+r}{s+a}}\tau^{\frac{s-r}{2s+2a}}$ for a constant C_φ independent of f, ϱ and τ.

Proof. By the assumption $a < s_{\max}$ the synthesis operator $\mathcal{S}\colon b_{p,\infty}^{-a} \to B_{p,\infty}^{-a}$ is a Banach space isomorphism. Let c_1 be the operator norm of its inverse.
There is a continuous embedding $B_{p,2}^{-a} \subset B_{p,\infty}^{-a}$. Moreover, in the case $p \in [1,2)$ the p-monotonicity yields a continuous embedding $B_{2,2}^{-a} \subset B_{p,2}^{-a}$. Therefore, we have a continuous

embedding $B_{2,2}^{-a} \subset B_{p,\infty}^{-a}$ and denote its operator norm by c_2. Along with Assumption 2.15(a) we conclude

$$\|x - z\|_{-a,p,\infty} \le c_2 \|x - z\|_{-a,2,2} \le c_1 c_2 \|\mathcal{S}x - \mathcal{S}z \mid B_{2,2}^{-a}\| \le c_1 c_2 M_1 \|F(\mathcal{S}x) - F(\mathcal{S}z)\|_{\mathbb{Y}}.$$
$$(3.29)$$

With this the claim follows from Lemma 3.11. □

Remark 3.13 (Weakening of the assumptions). *Note that Proposition 3.12 remains true if one replaces the space $B_{2,2}^{-a}$ in Assumption 2.15(a) by $B_{2,\infty}^{-a}$ or even by $B_{p,\infty}^{-a}$. Nevertheless, we refrained from formulating this weaker Lipschitz assumption as we will need a Lipschitz bound with respect to $B_{2,2}^{-a}$ later.*

Remark 3.14. *The idea of the verification of a variational source condition in Proposition 3.12 generalizes as follows: Suppose $\mathbb{X}_s, \mathbb{X}, \mathbb{X}_-$ are Banach spaces with continuous embeddings $\mathbb{X}_s \subset \mathbb{X} \subset \mathbb{X}_-$, $(P_j\colon \mathbb{X} \to \mathbb{X})_{j\in J}$ is a family of projections such that (3.19a) holds true. Assume that one has the following Bernstein and Jackson type inequalities*

$$\|P_j \cdot\|_{\mathbb{X}} \le \nu_j \| \cdot \|_{\mathbb{X}_-} \quad and \quad \|(I - P_j) \cdot\|_X \le \kappa_j \| \cdot \|_{\mathbb{X}_s}. \qquad (3.30)$$

Moreover, suppose $G\colon D \to \mathbb{Y}$ is a map defined on a subset $D \subset \mathbb{X}$ mapping to a Banach space $\mathbb{Y}$ such that a Lipschitz condition $\|z_1 - z_2\|_{\mathbb{X}_-} \le L\|G(z_1) - G(z_2)\|_{\mathbb{Y}}$ for all $z_1, z_2 \in D$ holds true.
Then Proposition 3.5 implies that every $x \in \mathbb{X}_s$ with $\|x\|_{\mathbb{X}_s} \le \varrho$ satisfies a variational source condition (3.14) with $\beta = 1$ and $\varphi(\tau) = 2\inf_{j\in J}\left(L\nu_j\sqrt{\tau} + \kappa_j\varrho\right)$.
This consideration shows similarities to the discussion in [104, Rem. 4.8.]. A difference is that in the latter reference Jackson and Bernstein estimates with respect to dual spaces are required and smoothness is measured in terms of the tail behavior of the subdifferential of the true solution.

Together with Corollary 3.2 the variational source condition provides the following result for the reconstruction R_α given by (3.16). Optimal convergence rates for Besov space regularization with $p \in (1, 2]$ and $q \ge 2$ are shown in [104, Thm. 4.2].

Theorem 3.15 (Error bounds for smooth wavelet systems). *Suppose Assumptions 2.11 and 2.15(a) hold true with $a < s_{max}$. Let $p \in [1, 2]$, $0 \le r < s < s_{max}$ and $\varrho > 0$. If $p \in [1, 2)$, then we require that the scale $B_{p,q}^s$ is p-monotone (see Definition 2.9). Suppose $f \in B_{p,\infty}^s \cap \operatorname{dom}(F)$ with $\|f \mid B_{p,\infty}^s\| \le \varrho$. Let $\overline{p} := \frac{2p(a+r)}{2a+pr}$.*

(a) (Bias bounds) There exists a constant C_b independent of f and ϱ such that such that for all $\alpha > 0$ and $f_\alpha \in R_\alpha(F(f))$ (see (3.16)) the following bounds hold true:

$$\|f - f_\alpha \mid B_{p,1}^r\| \le C_b \varrho^{\frac{2a+2r}{s+2a+r}} \alpha^{\frac{s-r}{s+2a+r}}, \qquad (3.31)$$

$$\|f - f_\alpha \mid B_{2,2}^{-a}\| \le C_b \varrho^{\frac{a+r}{s+2a+r}} \alpha^{\frac{s+a}{s+2a+r}} \quad and \qquad (3.32)$$

$$\|f - f_\alpha \mid B_{\overline{p},\overline{p}}^0\| \le C_b \varrho^{\frac{2a+r}{s+2a+r}} \alpha^{\frac{s}{s+2a+r}}. \qquad (3.33)$$

(b) (Convergence rates) Let $0 < c_l \le c_r$ and $1 \le c_D \le C_D$. Suppose $\delta > 0$, $g^{\mathrm{obs}} \in \mathbb{Y}$ with $\|g^{\mathrm{obs}} - F(f)\|_{\mathbb{Y}} \le \delta$. Let $\alpha > 0$ and $\hat{f}_\alpha \in R_\alpha(g^{\mathrm{obs}})$. Then there exists a constant C_r independent of $f, g^{\mathrm{obs}}, \varrho$ and δ such that either of the conditions

$$c_l \varrho^{-\frac{a+r}{s+a}} \delta^{\frac{s+2a+r}{s+a}} \le \alpha \le c_r \varrho^{-\frac{a+r}{s+a}} \delta^{\frac{s+2a+r}{s+a}} \quad and \quad c_D \delta \le \|g^{\mathrm{obs}} - F(\hat{f}_\alpha)\|_{\mathbb{Y}} \le C_D \delta$$

implies the following bounds

$$\|f - \hat{f}_\alpha \mid B_{p,1}^r\| \leq C_r \varrho^{\frac{a+r}{s+a}} \delta^{\frac{s-r}{s+a}}, \tag{3.34}$$

$$\|f - \hat{f}_\alpha \mid B_{2,2}^{-a}\| \leq C_r \delta \quad and \tag{3.35}$$

$$\|f - \hat{f}_\alpha \mid B_{\bar{p},\bar{p}}^0\| \leq C_r \varrho^{\frac{a}{s+a}} \delta^{\frac{s}{s+a}}. \tag{3.36}$$

Proof.   (a) Let x_α be a minimizer in (3.16) with $f_\alpha = \mathcal{S}x_\alpha$. The smoothness assumption on f implies that $x := \mathcal{S}^{-1}f \in b_{p,\infty}^s$ and $\|x\|_{s,p,\infty} \leq c_1 \varrho$ with c_1 denoting the operator norm of $\mathcal{S}^{-1} \colon B_{p,\infty}^s \to b_{p,\infty}^s$. Let c_2 be the operator norm of $\mathcal{S} \colon b_{p,1}^r \to B_{p,1}^r$. Proposition 3.12 and Corollary 3.2(a) with $\nu = \frac{s-r}{2s+2a}$ yield

$$\|f - f_\alpha \mid B_{p,1}^r\| \leq c_2 \|x - x_\alpha\|_{r,p,1} \leq 2c_2 k_\nu C_\varphi^{\frac{2s+2a}{s+2a+r}} c_1^{\frac{2a+2r}{s+2a+r}} \varrho^{\frac{2a+2r}{s+2a+r}} \alpha^{\frac{s-r}{s+2a+r}}.$$

Moreover, Corollary 3.2 and Assumption 2.15 provide

$$\|f - f_\alpha \mid B_{2,2}^{-a}\| \leq M_1 \|F(\mathcal{S}x) - F(\mathcal{S}x_\alpha)\|_{\mathbb{Y}} \leq M_1 k_\nu C_\varphi^{\frac{s+a}{s+2a+r}} c_1^{\frac{a+r}{s+2a+r}} \varrho^{\frac{a+r}{s+2a+r}} \alpha^{\frac{s+a}{s+2a+r}}.$$

The continuous embedding $B_{p,1}^0 \subseteq B_{p,p}^0$ implies (3.33) in the case $r = 0$. In the case $r > 0$ we combine the latter embedding with the real interpolation identity $\left(B_{2,2}^{-a}, B_{p,p}^r\right)_{\frac{a}{a+r},\bar{p}} = B_{\bar{p},\bar{p}}^0$ (see Definition 2.9(d)(i) if $p = 2$ and (ii) if $p < 2$) to see that there is a constant c_3 such that

$$\|\cdot \mid B_{\bar{p},\bar{p}}^0\| \leq c_3 \|\cdot \mid B_{2,2}^{-a}\|^{\frac{r}{a+r}} \|\cdot \mid B_{p,1}^r\|^{\frac{a}{a+r}}. \tag{3.37}$$

Inserting $f - f_\alpha$ we obtain (3.33) from (3.32) and (3.31).

(b) The proof is along the lines of the proof of (*a*) using Corollary 3.2(*b*) and (*c*) instead of Corollary 3.2(*a*).   □

In the case that $B_{\bar{p},\bar{p}}^0$ is a Besov function space we conclude error bounds in the $L^{\bar{p}}$-norm.

Corollary 3.16 (Error bounds in $L^{\bar{p}}$)**.** *In the setting of Theorem 3.15 suppose that* $B_{\bar{p},\bar{p}}^0 = B_{\bar{p},\bar{p}}^0(\Omega)$ *with* Ω *either* $\mathbb{T}^d, \mathbb{R}^d$ *or a bounded Lipschitz domain in* $\mathbb{R}^d$. *Then we have the upper bounds*

$$\|f - f_\alpha\|_{L^{\bar{p}}(\Omega)} \leq C_{\bar{p}} \varrho^{\frac{2a+r}{s+2a+r}} \alpha^{\frac{s}{s+2a+r}} \quad and$$

$$\|f - \hat{f}_\alpha\|_{L^{\bar{p}}(\Omega)} \leq C_{\bar{p}} \varrho^{\frac{a}{s+a}} \delta^{\frac{s}{s+a}}.$$

with a constant $C_{\bar{p}}$ *independent of* $f, g^{\mathrm{obs}}, \varrho$ *and* δ.

Proof. By Proposition A.11(a) we have a continuous embedding $B_{\bar{p},\bar{p}}^0 \subset L^{\bar{p}}(\Omega)$. Hence the inequalities in the $L^{\bar{p}}(\Omega)$-norm follow from the corresponding inequalities in the $B_{\bar{p},\bar{p}}^0$-norm in Theorem 3.15.   □

The following example illuminates which Besov integrabilty and smoothness is assigned to piecewise smooth functions with kinks or jumps.

Example 3.17 (Functions with jumps or kinks (cf.[85, Ex. 30]))**.** *We consider functions* $f^{\mathrm{jump}}, f^{\mathrm{kink}} : [0,1] \to \mathbb{R}$ *which are* C^∞ *everywhere with uniform bounds on all derivatives except at a finite number of points in* $[0,1]$*, and* $f^{\mathrm{kink}} \in C^{0,1}([0,1])$*. In other words,* $f^{\mathrm{jump}}, f^{\mathrm{kink}}$ *are piecewise smooth,* f^{jump} *has a finite number of jumps, and* f^{kink} *has a finite number of kinks. Then for* $p \in (0, \infty)$*,* $q \in (0, \infty]$*, and* $s \in \mathbb{R}$ *with* $s > \sigma_p$ *with* σ_p *as in Assumption 2.11 we have*

$$f^{\mathrm{jump}} \in B^s_{p,q}((0,1)) \;\Leftrightarrow\; s < \frac{1}{p}, \qquad f^{\mathrm{kink}} \in B^s_{p,q}((0,1)) \;\Leftrightarrow\; s < 1 + \frac{1}{p}$$

if $q < \infty$ *and*

$$f^{\mathrm{jump}} \in B^s_{p,\infty}((0,1)) \;\Leftrightarrow\; s \le \frac{1}{p}, \qquad f^{\mathrm{kink}} \in B^s_{p,\infty}((0,1)) \;\Leftrightarrow\; s \le 1 + \frac{1}{p}.$$

To see this, we can use the classical definition of Besov spaces in terms of the modulus of continuity $\|\Delta_h^m f\|_{L^p}$ *where* $(\Delta_h f)(x) := f(x+h) - f(x)$ *and* $\Delta_h^{m+1} f := \Delta_h(\Delta_h^m f)$*, see, e.g. [100, eq. (1.23)]. Elementary computations show that* $\|\Delta_h^m f^{\mathrm{jump}}\|_{L^p}$ *decays of the order* $h^{1/p}$ *as* $h \searrow 0$ *if* $m \ge 1/p$*, and* $\|\Delta_h^m f^{\mathrm{kink}}\|_{L^p}$ *decays as* $h^{1/p+1}$ *if* $m \ge 2/p$*.*
Therefore, the regularity of f^{jump} *or* f^{kink} *measured by the smoothness parameters* s *for which these functions belong to* $B^s_{p,q}((0,1))$ *depends on* p*. A smaller choice of* p *allows for a larger value of* s *and hence a due to Theorem 3.15 for a faster convergence rate.*

We further discuss our result in a series of remarks.

Remark 3.18 (No saturation(cf. [64, Rem. 4.4.]))**.** *With* Ω *as in Corollary 3.16 and in the case* $p = 2$ *and we obtain an error bound*

$$\|f^\dagger - \hat{f}_\alpha\|_{L^2(\Omega)} = \mathcal{O}\left(\varrho^{\frac{a}{s+a}} \delta^{\frac{s}{s+a}}\right). \tag{3.38}$$

This allows to compare the sparsity promoting regularization (3.16) *to standard Tikhonov-regularization with the quadratic penalty term* $\alpha\|f\|^2_{L^2(\Omega)}$*: Whereas for quadratic Tikhonov regularization the error bound* (3.38) *is only valid for* $s \le 2a$ *(see [35, Prop. 5.3.]), the sparsity promoting regularization* (2.17) *obeys this error bound for arbitrarily high smoothness parameters* $s > 0$ *given sufficient smoothness of the wavelet system, i.e. large enough* s_{max} *in Assumption 2.11.*
Figure 3.1 shows that this effect can also be observed in practice. In contrast to classical Tikhonov regularization, which saturates at the rate $\mathcal{O}(\delta^{2/3})$*, we see that the reconstruction error using* $b^0_{2,1}$*-penalty is linear in* δ*. Here we use the linear forward operator* $(I - \partial_x^2)^{-\frac{1}{2}} : B^{-1}_{2,2}(\mathbb{T}) \to L^2(\mathbb{T})$ *(see Example 2.17). We implement the latter operator using the fast Fourier transform on a grid with* 2^{13} *nodes. Moreover, we penalize with a wavelet system of high regularity (Daubechies wavelets of order* 28*). As true solution we take a sine function. The staircase behavior of the reconstruction error for classical Tikhonov regularization is due to the sequential discrepancy principle.*

Remark 3.19 (Boundedness of Ω not required if $p = 2$)**.** *As we do not need require monotonicity of the Besov scale in the case* $p = 2$ *the stated convergence rates results hold also true for* $B^s_{p,q} = B^s_{p,q}\left(\mathbb{R}^d\right)$*.*
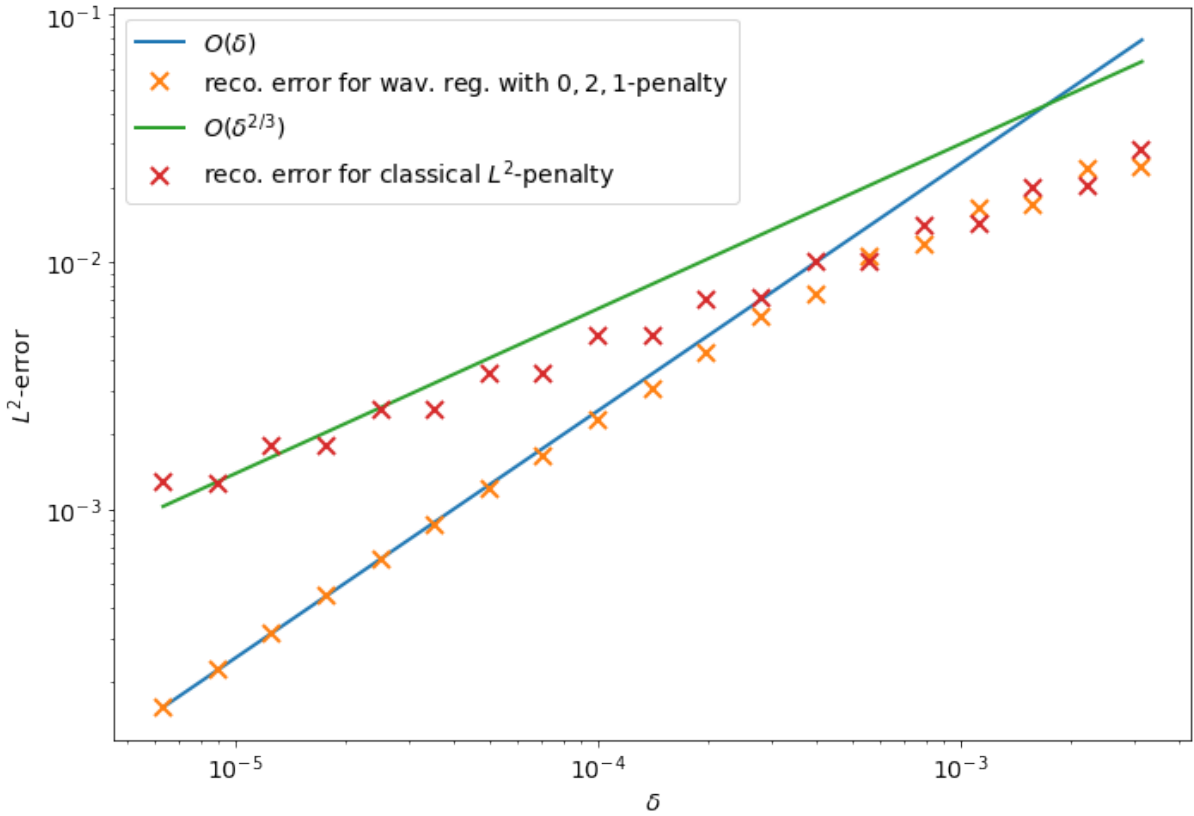
FIGURE 3.1: Wavelet regularization with $b_{2,1}^0$-penalty outperforms classical Tikhonov regularization in reconstructing a smooth function.

Remark 3.20 (The assumption $p \leq 2$)**.** *Suppose $B_{p,q}^s$ is p-monotone. The assumption $p \leq 2$ enters as we use the embedding $B_{2,2}^{-a} \subset B_{p,2}^{-a}$ in the proof of Proposition 3.12. This embedding is no longer true if we allow $p > 2$. Replacing the $p = q = 2$ in Assumption 2.15 by a greater index $\tilde{p}$ would make the theory work for all $p \leq \tilde{p}$. Nevertheless, such a replacement would force $\mathbb{Y}$ to be an $L^{\tilde{p}}$-space.*
Besides the promotion of sparsity another point why we think the case $p \leq 2$ is more interesting than $p > 2$ is that for $p \in [1, 2]$ the source set $B_{p,\infty}^s$ leading to the rate $\mathcal{O}(\delta^{\frac{s}{s+a}})$ with a Besov $0, p, 1$-penalty term is larger than the source set $B_{2,\infty}^s$ leading to the same rate using classical Tikhonov regularization with L^2-penalty.

Remark 3.21 (Converse result to variational source condition)**.** *In the setting of Proposition 3.12 we additionally assume $B_{p,1}^r \subset \text{dom}(F)$ and Assumption 2.15(b). Note that we have $B_{p,1}^r \subset \text{dom}(F)$ for example if F is linear with $\text{dom}(F) = B_{2,2}^{-a}$ and $a + r \geq d(1/p - 1/2)$. We investigate the validity of a converse result to Proposition 3.12 as follows: Suppose $x \in b_{p,1}^r$ such that (3.28) holds true with $\varphi(\tau) = K\tau^{\frac{s-r}{2s+2a}}$. Assumption 2.15(b) and Assumption 2.11 yield*

$$\|F(\mathcal{S}x) - F(\mathcal{S}z)\|_{\mathbb{Y}} \leq M_2'\|x - z\|_{-a,2,2} \quad \text{for all } z \in b_{p,1}^r \cap \mathcal{S}^{-1}(D_F)$$

with M_2' the product of M_2 and the operator norm of $\mathcal{S} : b_{2,2}^{-a} \to B_{2,2}^{-a}$. Using this Lipschitz

bound and leaving out the loss function in (3.28) we obtain

$$\|x\|_{r,p,1} - \|z\|_{r,p,1} \leq KM_2' \|x - z\|_{-a,2,2}^{\frac{s-r}{s+a}} \quad \text{for all } z \in b_{p,1}^r. \tag{3.39}$$

Let $j \in \mathbb{N}_0$. Inserting $z \in b_{p,1}^r$ given by $z_k = x_k$ for all $k \neq j$ and $z_j = 0$ we obtain

$$2^{jr} 2^{jd\left(\frac{1}{2} - \frac{1}{p}\right)} \|x_j\|_p \leq KM_2' 2^{-\frac{ja(s-r)}{s+a}} \|x_j\|_2^{\frac{s-r}{s+a}}.$$

As $p \leq 2$ we may use $\| \cdot \|_2 \leq \| \cdot \|_p$ on the right hand side. Rearranging the obtained inequality leads to

$$\|x_j\|_p^{\frac{r+a}{s+a}} \leq KM_2' 2^{-js\frac{r+a}{s+a}} 2^{jd\left(\frac{1}{p} - \frac{1}{2}\right)}.$$

In the case $p = 2$ this resolves to $\|x_j\|_2 \leq (KM_2')^{\frac{s+a}{r+a}} 2^{-js}$. Hence,

$$x \in b_{2,\infty}^s \quad \text{with} \quad \|x\|_{s,2,\infty} \leq (KM_2')^{\frac{s+a}{r+a}} 2^{-js}.$$

Therefore, we have proven the equivalence of the following statements in the case $p = 2$:

(i) There exists a constant $K_{vsc} \geq 0$ such that (3.28) holds true with $\varphi(\tau) = K_{vsc}\tau^{\frac{s-r}{2s+2a}}$.

(ii) There exists a constant $K \geq 0$ such that (3.39) holds true.

(iii) $x \in b_{2,\infty}^s$.

This result does not generalize to $p < 2$: In Section 3.5 we characterize the largest space on which a variational source condition of order $\varphi(\tau) = \mathcal{O}(\tau^{\frac{s-r}{2s+2a}})$ holds true in the case of $p = 1$.

Remark 3.22 (The parameter r)**.** *While the smoothness assumption $f \in B_{p,\infty}^s$ is independent of the parameter r which controls the weights in the penalty term the latter parameter has an impact on the parameter choice. That is reasonable as for larger r the stronger penalty requires a smaller parameter α to handle to offset between penalization and data fidelity. It is worth noting that in the case $p < 2$ the strength of the $L^{\bar{p}}$-norm in which the error is measured in Corollary 3.16 grows with r. That could be a benefit of a choice $r > 0$ in this case. Another advantage in particular in higher dimensional problems is that it weakens the condition $a \geq d\left(\frac{1}{p} - \frac{1}{2}\right)$ to guarantee existence of minimizers to $a + r \geq d\left(\frac{1}{p} - \frac{1}{2}\right)$ (see Proposition 2.16).*

Remark 3.23 (Limiting case: $r = s$)**.** *In the setting of Theorem 3.15 let us assume $f \in B_{p,1}^r$ with $\|f \mid B_{p,1}^r\| \leq \varrho$. Hence, we assume that f has no additional smoothness than needed for the true solutions wavelet coefficients x to have a finite value in the penalty term. By inserting x into the Tikhonov functional in (3.16) it is easy to see and well-known that a parameter choice $\alpha \sim \delta^2$ as well as the discrepancy principle as in Theorem 3.15(b) lead to bounds $\|\hat{f}_\alpha \mid B_{p,1}^r\| \lesssim \|\hat{x}_\alpha\|_{r,p,1} \lesssim \varrho$ and $\|f - \hat{f}_\alpha \mid B_{2,2}^{-a}\| \lesssim \|F(f) - F(\hat{f}_\alpha)\|_\mathbb{Y} \lesssim \delta$. With this the interpolation inequality (3.37) yields $\|f - \hat{f}_\alpha \mid B_{\bar{p},\bar{p}}^0\| \lesssim \varrho^{\frac{a}{a+r}} \delta^{\frac{r}{a+r}}$.*

As the same arguments apply also in the noiseless case we see that Theorem 3.15 remains true in the case $r = s$ if we replace $B_{p,\infty}^r$ by $B_{p,1}^r$.

Remark 3.24 (Oversmoothing: $r < s$)**.** *Note that in view of the assumption $r < s$ in Theorem 3.15 it is natural to ask what happens for $0 < s \leq r$ (and for $r = s$ without the replacement of ∞ by 1 in the last remark). Note that in this case the wavelet coefficients x*

of the true solution may not belong to $b_{p,1}^r$ such that the value of the penalty term in (2.17) at x is infinite. Therefore, the concept of variational source conditions does not apply here and other ideas are needed to tackle this question.

Situations where the true solution does not admit a finite penalty are referred to as oversmoothing in the literature and we refer to the papers [60, 62, 86, 17] for convergence rates theory in this case.

In Chapter 5 we provide a general convergence rates theory for oversmoothing in Banach space regularization. An application of this theory to wavelet regularization in Besov spaces will provide error bounds in the case $0 < s \leq r$.

Numerical validation

We confirm the theoretical results in Theorem 3.15 with numerical experiments for the nonlinear parameter identification problem in Example 2.21. In one dimension it becomes the reconstruction of the coefficient c in the elliptic boundary value problem

$$
\begin{aligned}
-u'' + cu &= f \quad \text{in } (0,1), \\
u(0) = u(1) &= 1.
\end{aligned}
\tag{3.40}
$$

The forward operator in the function space setting is $F(c) := u$ for the fixed right hand side $f(\cdot) = \sin(4\pi\cdot) + 2$. For this problem the verification of Assumption 2.15 with $a = 2$ is discussed in Example 2.21. The true coefficient c^{jump} or c^{kink} is given by a piecewise smooth function with either finitely many jumps or kinks (see Example 3.17).

Numerical simulations were carried out in Python. To solve the boundary value problem (3.40) we used quadratic finite elements and an equidistant grid containing 127 finite elements. The coefficient c was sampled on an equidistant grid with 1024 points.

For the penalization by a $b_{p,1}^r$-norm we used the discrete wavelet decomposition from `PyWavelets` [75] with Daubechies wavelets of order 7, whose regularity satisfies $s_{\max} > 2.15$ (see [27, Table 2]).

The minimization problem in (3.16) was solved via the Gauß-Newton-type method $c_{k+1} = \mathcal{S}x_{k+1}$,

$$
x_{k+1} \in \operatorname*{argmin}_x \left[\frac{1}{2} \|F'[\mathcal{S}x_k](\mathcal{S}x - \mathcal{S}x_k) + F(\mathcal{S}x_k) - u\|_{\mathbb{Y}}^2 + \alpha \|x - x_0\|_{r,p,1} \right].
$$

with a constant initial guess $c_0 = 1$. In each Gauß-Newton step these linearized minimization problems were solved with the *Fast Iterative Shrinkage- Thresholding Algorithm* (FISTA) proposed and analyzed by Beck & Teboulle in [7]. We used the *inertial parameter* as in [15, Sec. 4].

We did not impose a constraint on the size of $\|c - c_0\|$, which is required by our theory if Assumption 2.15 does not hold true globally. However, the size of the domain of validity of this assumption is difficult to assess, and such a constraint is likely to be never active for a sufficiently good initial guess c_0.

The regularization parameter α was chosen by a sequential discrepancy principle with $\tau_1 = 1$ and $\tau_2 = 2$ on a grid $\alpha_j = 2^{-j}\alpha_0$.

To simulate worst case errors, we computed for each noise level δ reconstructions for several data errors $u^\delta - G(c^\dagger)$, $\|u^\delta - G(c^\dagger)\|_{L^2} = \delta$, which were given by sin functions with

different frequencies.

Here we focused on the choices $p = 1$ or $p = 2$ as in these cases the corresponding proximity operator is given by simple explicit operations. In Figures 3.2 and 3.3, that is for $p = 2$ and $r = 0$, we see that the asymptotic behavior of the reconstruction error agrees with impressive accuracy with our predicted convergence rate. Whereas the error diagrams in Figures 3.4 and 3.5, i.e. for $p = 1$ and $r = 0$, suggest that our predicted rate could be improved.

That is in line with the statements in Remark 3.21 as follows: By Example 3.17 the smoothness parameters $s = \frac{1}{2}$ for c^{jump} and $s = \frac{3}{2}$ for c^{kink} are the maximal ones such that these function belong to $B_{2,\infty}^s((0,1))$. Therefore, as Remark 3.21 characterizes these spaces by variational source conditions, we see that s is also the maximal exponent such that a variational source condition of the form (3.28) with $\varphi(\tau) = \mathcal{O}\left(\tau^{\frac{s}{2s+2a}}\right)$ holds true. This suggests, that the convergence rate we state in this case cannot be improved. In the next chapter we will even characterize the spaces $B_{p,\infty}^s$ by the bias bound in the case of exact data which consolidates this suggestion.

In the case $p = 1$ the values $s = 1$ for c^{jump} and $s = 2$ for c^{kink} are also maximal such that these functions belong the $B_{1,\infty}^s((0,1))$. The difference to the above situation is that these spaces are not characterized by the corresponding variational source conditions. Therefore, at this point we cannot exclude that there is a larger s allowing for such a condition with φ as above. Indeed, we will later prove a variational source condition for any $s < \frac{4}{3}$ in case of c^{jump} and for any $s < \frac{8}{3}$ for c^{kink}. The convergence rates we are going to conclude from the latter conditions explain the numerical simulations as accurate as in the case $p = 2$.

Characterization as source type condition

In the case of a linear forward operator A we characterize the condition (3.19b) in terms of the range of the adjoint operator A^*. For $\mathbb{X} = \ell^1$ we show that the existence of a sequence $(\nu_n)_{n \in \mathbb{N}}$ such that (3.19b) holds true is equivalent to the condition $e_j \in \text{im}(A^*)$ for all $j \in \mathbb{N}$. The latter source type condition is used in [11] to prove convergence rates.

We start with a lemma that characterizes (3.19b) for a single projection.

Lemma 3.25. *Suppose $A\colon \mathbb{X} \to \mathbb{Y}$ is linear and bounded, $P\colon \mathbb{X} \to \mathbb{X}$ is a projection with operator norm 1 and $\nu \geq 0$. The following statements are equivalent:*

(i) $\|Pz\|_{\mathbb{X}} \leq \nu \|Az\|_{\mathbb{Y}}$ for all $z \in \mathbb{X}$.

(ii) For every $\xi \in \text{im}(P^*)$ with $\|\xi\|_{\mathbb{X}'} \leq 1$, there exists $\psi \in \mathbb{Y}'$ with $\|\psi\|_{\mathbb{Y}'} \leq \nu$ such that $A^*\psi = \xi$.

Proof.

$(i) \Rightarrow (ii)$: As $\xi \in \text{im}(P^*)$ we have $P^*\xi = \xi$. Hence, with $z \in \mathbb{X}$ we obtain

$$\langle \xi, z \rangle = \langle \xi, Pz \rangle \leq \|Pz\|_{\mathbb{X}} \leq \nu \|Az\|_{\mathbb{Y}}.$$

Therefore, Proposition A.2 yields (ii).

$(ii) \Rightarrow (i)$: Let $z \in \mathbb{X}$. By the Hahn-Banach theorem there is $\xi \in \mathbb{X}'$ with $\|\xi\|_{\mathbb{X}'} = 1$ and $\langle \xi, Pz \rangle = \|Pz\|_{\mathbb{X}}$. As P has operator norm 1 so does P^*. Hence, $\|P^*\xi\| \leq 1$ and (ii)
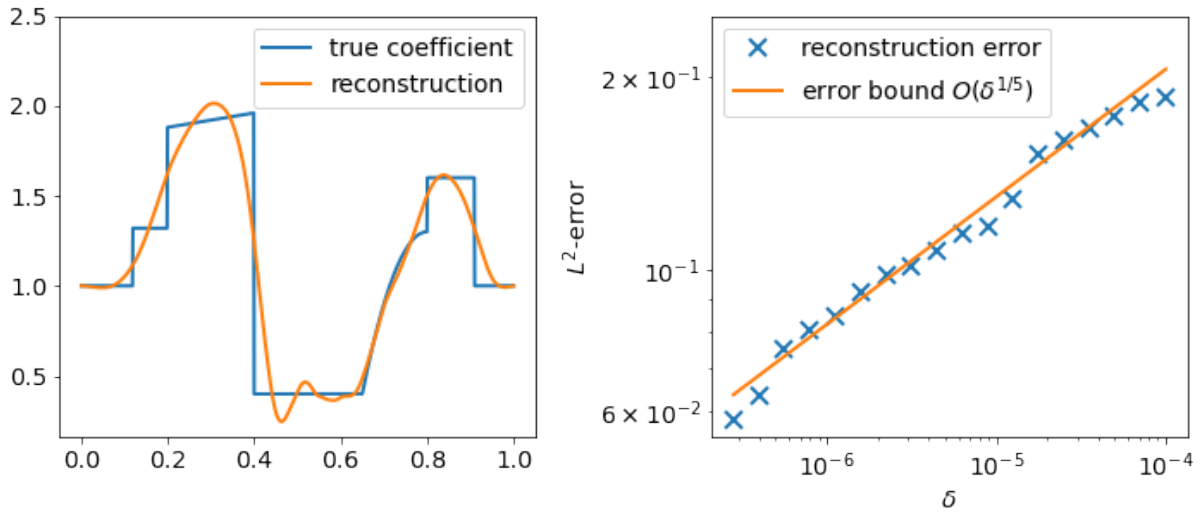
FIGURE 3.2: *Left:* true coefficient c^{jump} with jumps in the boundary value problem (3.40) together with a typical reconstruction at noise level $\delta = 2.5 \cdot 10^{-5}$. *Right:* Reconstruction error using $b_{2,1}^0$-penalization, the rate $\mathcal{O}(\delta^{1/5})$ the L^2-norm predicted by Corollary 3.16 as Example 3.17 yields $c^{\text{jump}} \in B_{2,\infty}^{1/2}((0,1))$.
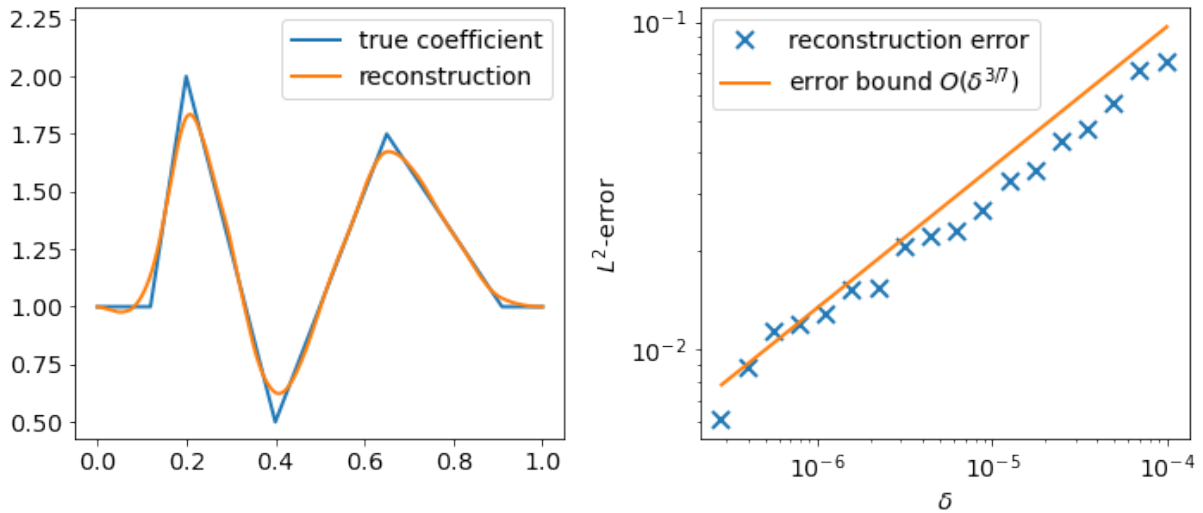
FIGURE 3.3: *Left:* true coefficient c^{kink} with kinks in the boundary value problem (3.40) together with a typical reconstruction at noise level $\delta = 2.5 \cdot 10^{-5}$. *Right:* Reconstruction error using $b_{2,1}^0$-penalization, the rate $\mathcal{O}(\delta^{3/7})$ the L^2-norm predicted by Corollary 3.16 as Example 3.17 yields $c^{\text{kink}} \in B_{2,\infty}^{3/2}((0,1))$.
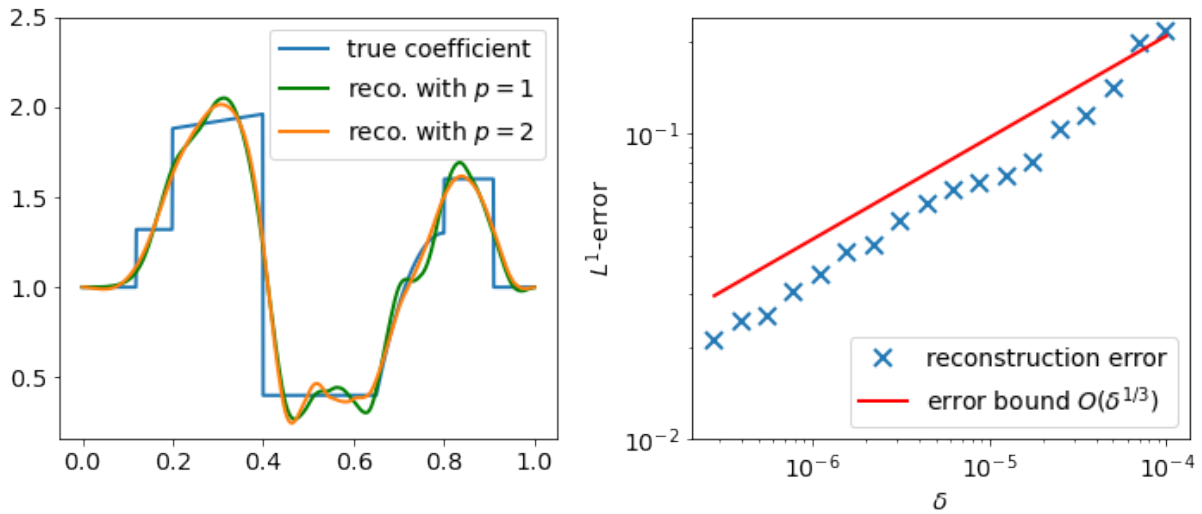
FIGURE 3.4: *Left:* true coefficient c^{jump} with jumps in the boundary value problem (3.40) together with reconstructions for $b_{1,1}^0$- and $b_{2,1}^0$-penalization at noise level $\delta = 2.5 \cdot 10^{-5}$ for the same data. *Right:* Reconstruction error using $b_{1,1}^0$-penalization, the rate $\mathcal{O}(\delta^{1/3})$ in the L^1-norm predicted by Corollary 3.16 as Example 3.17 yields $c^{\text{jump}} \in B_{1,\infty}^1((0,1))$.
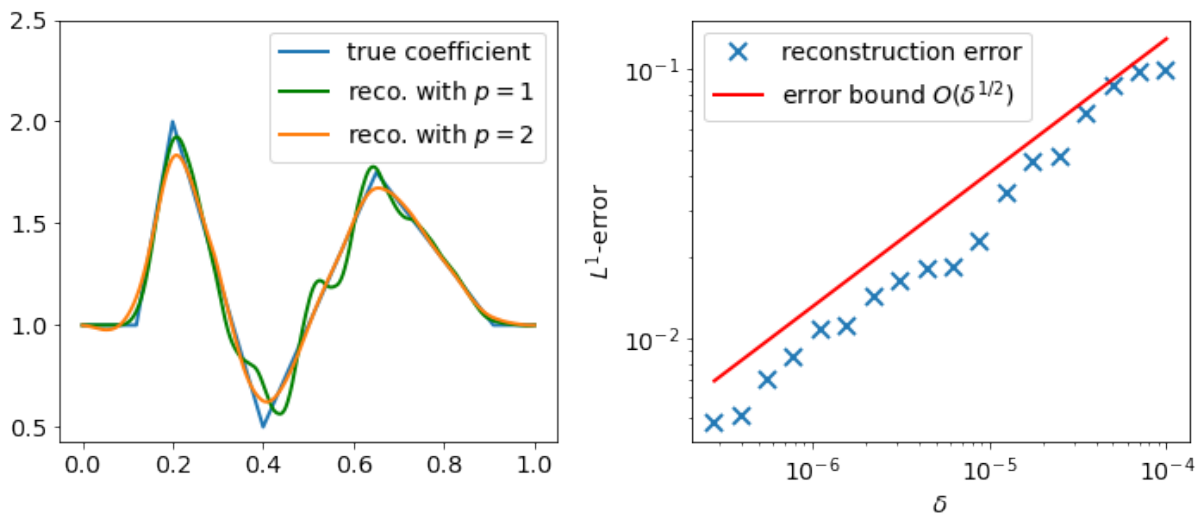
FIGURE 3.5: *Left:* true coefficient c^{kinks} with kinks in the boundary value problem (3.40) together with reconstructions for $b_{1,1}^0$- and $b_{2,1}^0$-penalization at noise level $\delta = 2.5 \cdot 10^{-5}$ for the same data. *Right:* Reconstruction error using $b_{1,1}^0$-penalization, the rate $\mathcal{O}(\delta^{1/2})$ in the L^1-norm predicted by Corollary 3.16 as Example 3.17 yields $c^{\text{kink}} \in B_{1,\infty}^2((0,1))$.

provides $\psi \in \mathbb{Y}'$ with $\|\psi\|_{\mathbb{Y}'} \leq \nu$ such that $A^*\psi = P^*\xi$. Using Proposition A.2 again we conclude

$$\|Pz\|_{\mathbb{X}} = \langle P^*\xi, z \rangle \leq \nu \|Az\|_{\mathbb{Y}}. \qquad \square$$

Remark 3.26. *Note that the statement of Lemma 3.25 remains true, if ξ in* (*ii*) *ranges only over a set whose convex hull is $\{\xi \in \mathrm{im}(P^*) : \|\xi\|_{\mathbb{X}'} \leq 1\}$. Hence, by the Krein–Milman theorem it is enough to require statement* (*ii*) *for the extreme points of the latter set whenever $\mathrm{im}(P^*)$ is finite dimensional.*

**Corollary 3.27.** *Suppose $G = A : \mathbb{X} \to \mathbb{Y}$ is linear and bounded, $(P_j : \mathbb{X} \to \mathbb{X})_{j \in J}$ a family of projections satisfying* (3.19a) *and $(\nu_j)_{j \in J} \subset [0, \infty)$. Then* (3.19b) *holds true if and only if for every for all $j \in J$ and $\xi \in \mathrm{im}(P_j^*)$ with $\|\xi\|_{\mathbb{X}'} \leq 1$ there exists $\psi \in \mathbb{Y}$ with $\|\psi\|_{\mathbb{Y}'} \leq \nu_j$ and $A^*\psi = \xi$.*

Proof. By linearity the inequality in (3.19b) for all $z_1, z_2 \in \mathbb{X}$ is equivalent to statement (*i*) in Lemma 3.25. $\qquad \square$

Applying the considerations to a linear operator on ℓ^1 we obtain the following characterization of condition (3.19b).

Proposition 3.28. *Let $\mathbb{X} = \ell^1$, $G = A : \ell^1 \to \mathbb{Y}$ bounded and linear and $(P_n)_{n \in \mathbb{N}}$ as in Example 3.6. Then there is a sequence $(\nu_n)_{n \in \mathbb{N}}$ such that* (3.19b) *holds true if and only if $e_n \in \mathrm{im}(A^*)$ for all $n \in \mathbb{N}$. In this case the minimal ν_n satisfying* (3.19b) *is given by*

$$\max_{\lambda \in \{-1,1\}^n} \inf \left\{ \|\psi\|_{\mathbb{Y}'} : \psi \in \mathbb{Y}' \text{ with } A^*\psi = \sum_{j=1}^n \lambda_j e_j \right\}.$$

Proof. Let $\underline{\nu}_n$ denote the maximum defined in the last line. Suppose $(\nu_n)_{n \in \mathbb{N}}$ is a sequence such that (3.19b) holds true. The first implication in Lemma 3.25 for $P = P_n$ yields $e_n \in \mathrm{im}(A^*)$. To show that $\underline{\nu}_n \leq \nu_n$ let $\lambda \in \{-1,1\}^n$. Then $\|\lambda_1 e_1 + \ldots + \lambda_n e_n\|_\infty = 1$. Lemma 3.25 yields $\psi \in \mathbb{Y}^*$ with $\|\psi\|_{\mathbb{Y}'} \leq \nu_n$ and $A^*\psi = \lambda_1 e_1 + \ldots + \lambda_n e_n$. Hence, the infimum in the definition of $\underline{\nu}_n$ is less or equal ν_n. Taking the maximum over λ yields $\underline{\nu}_n \leq \nu_n$.

Vice versa, we assume $e_n \in \mathrm{im}(A^*)$ and verify (*ii*) with $\nu_n = \underline{\nu}_n$ in Lemma 3.25. Let $\varepsilon > 0$. By linearity for every $\lambda \in \{-1,1\}^n$ there is ψ as in the definition of $\underline{\nu}_n$ with $\|\psi\|_{\mathbb{Y}'} \leq \underline{\nu} + \varepsilon$. Since $\{\lambda_1 e_1 + \ldots + \lambda_n e_n : \lambda \in \{-1,1\}^n\}$ is the set of extreme points of $\mathrm{im}(P_n^*)$, Remark 3.26 yields (3.19b) with $\nu_n = \underline{\nu}_n + \varepsilon$, and the statement follows from taking the limit $\varepsilon \to 0$. $\qquad \square$

3.3 Non-smooth wavelet basis

In this section we generalize the results of Section 3.2 to the case where the regularity $s_{\max}$ of the wavelet system is lower then the smoothing parameter a of the forward operator. We will validate our results numerically using the same example as above with Daubechies wavelets of order 3 instead of order 7.

The restriction to $a < s_{\max}$ in Section 3.2 as well as in the papers [104] and [64] is a disadvantage, because for instance wavelet regularization of a twice smoothing operator

as in Example 2.21 with a non-smooth wavelet system like the Haar wavelets or other Daubechies wavelets of low order is not covered.

Another motivation to study the case $s_{\max} < a$ is that it is a first step towards the treatment of exponentially ill-posed problems. Indeed, we apply some of the presented ideas later to obtain a convergence rates result for the backward heat equation.

Before we present the details we briefly sketch the basic modification of the verification of the variational source condition. Recall that the starting point in Section 3.2 was the inequality

$$\|x - z\|_{r,p,1} + \|x\|_{r,p,1} - \|z\|_{r,p,1} \leq 2\left(\|P_n(x - z)\|_{r,p,1} + \|(I - P_n)x\|_{r,p,1}\right) \tag{3.41}$$

where P_n projects onto the first $n + 1$ levels and that we used $\|P_n \cdot\|_{r,p,1} \lesssim 2^{n(a+r)}\|\cdot\|_{-a,p,\infty}$ to estimate the first summand on the right hand side. We have to modify the latter Bernstein type inequality since the Besov norm with smoothness index $-a$ on its right hand side cannot be transferred to function spaces as in (3.29) if $s_{\max} < a$.

The key idea is to use the interpolation inequality $\|\cdot \mid B_{p,\infty}^{-\tilde{a}}\| \lesssim \|\cdot \mid B_{p,\infty}^{-a}\|^{\theta}\|\cdot \mid B_{p,1}^{r}\|^{1-\theta}$ with $\tilde{a} < s_{\max}$. Then the norm on the left hand side $\|\cdot \mid B_{p,\infty}^{-\tilde{a}}\|$ is characterized by wavelet coefficients. Combining with the Bernstein type inequality with a replaced by $\tilde{a}$ we obtain

$$\|P_n \cdot\|_{r,p,1} \lesssim 2^{n(\tilde{a}+r)}\|\mathcal{S}\cdot \mid B_{p,\infty}^{-a}\|^{\theta}\|\mathcal{S}\cdot \mid B_{p,1}^{r}\|^{1-\theta}.$$

From there we translate the second factor $\|\mathcal{S}\cdot \mid B_{p,1}^{r}\| \lesssim \|\cdot\|_{r,p,1}$ back to wavelet coefficients and apply Young's inequality to wind up with $\|P_n \cdot\|_{r,p,1} \leq C2^{n(a+r)}\|\cdot \mid B_{p,\infty}^{-a}\| + \frac{1}{4}\|\cdot\|_{r,p,1}$. Inserting into (3.41) we absorb the additional summand $\frac{1}{4}\|x - z\|_{r,p,1}$ into $\|x - z\|_{r,p,1}$ on the left hand side and use the same arguments as in the last section to conclude a variational source condition with loss function $\frac{1}{2}\|x - z\|_{r,p,1}$.
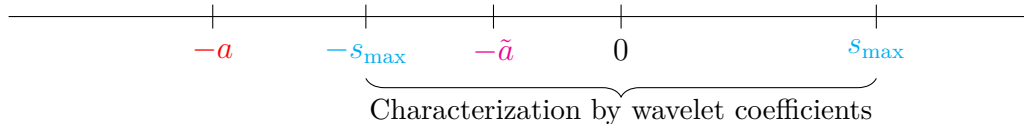
FIGURE 3.6: Smoothness indices: Replacing a by $\tilde{a} < s_{\max}$ in order to access via wavelet coefficients.

The structure of this section is as above: First we will present a general strategy for the verification of variational source conditions that is a generalization of Proposition 3.5 with (3.19b) replaced by a weaker assumption allowing for an additional summand as above. In a second subsection we will apply the considerations to sparsity promoting wavelet regularization to obtain the desired generalization of Theorem 3.15. Finally, we interpret the replacement of (3.19b) in the linear case as a source type condition involving the adjoint of the forward operator.

Verification of variational source conditions

Proposition 3.29. *Let $\mathbb{X}$, $\mathbb{Y}$ be Banach spaces and $G: D \to \mathbb{Y}$ a map defined on a subset $D \subset \mathbb{X}$, $x \in D$ and J be an index set. Suppose $\kappa \in [0, \frac{1}{2})$, $(P_j: \mathbb{X} \to \mathbb{X})_{j \in J}$ is a family*

of projections and $(\nu_j)_{j \in J}$ a family of positive real numbers such that (3.19a) and the following condition are satisfied:

$$\|P_j(z_1 - z_2)\|_{\mathbb{X}} \leq \nu_j \|G(z_1) - G(z_2)\|_{\mathbb{Y}} + \kappa \|z_1 - z_2\|_{\mathbb{X}} \quad \textit{for all } z_1, z_2 \in D, j \in J. \quad (3.42)$$

Then the variational source condition (3.14) holds true with $\beta = 1 - 2\kappa$ and

$$\varphi(\tau) = 2 \inf_{j \in J} \left(\nu_j \sqrt{\tau} + \|(I - P_j)x\|_{\mathbb{X}} \right).$$

Proof. The same argument as in Proposition 3.5 yields continuity of φ.
To prove (3.14) let $z \in D$ and $j \in J$. Inserting condition (3.42) into (3.20) and subtracting $2\kappa \|z_1 - z_2\|_{\mathbb{X}}$ on both sides yields

$$(1 - 2\kappa)\|x - z\|_{\mathbb{X}} + \|x\|_{\mathbb{X}} - \|z\|_{\mathbb{X}} \leq 2 \left(\nu_j \|G(z_1) - G(z_2)\|_{\mathbb{Y}} + \|(I - P_j)x\|_{\mathbb{X}} \right).$$

Taking the infimum over $j \in J$ on the right hand side yields the result. $\qquad \square$

Example 3.30. *Let $(\nu_j)_{j \in J}$ be a family of positive real numbers and $\theta \in (0, 1)$. We consider the following interpolation-type condition*

$$\|P_j(z_1 - z_2)\|_{\mathbb{X}} \leq \nu_j^\theta \|G(z_1) - G(z_2)\|_{\mathbb{Y}}^\theta \|z_1 - z_2\|_{\mathbb{X}}^{1-\theta} \quad \textit{for all } z_1, z_2 \in D. \quad (3.43)$$

Assuming (3.43) Young's inequality implies

$$\|P_j(z_1 - z_2)\|_{\mathbb{X}} \leq \theta(1-\theta)^{\frac{1-\theta}{\theta}} \kappa^{\frac{\theta-1}{\theta}} \nu_j \|G(z_1) - G(z_2)\|_{\mathbb{Y}} + \kappa \|z_1 - z_2\|_{\mathbb{X}} \quad (3.44)$$

for all $z_1, z_2 \in D$. and $\kappa \in (0, \infty)$. In particular (3.43) implies condition (3.42) for every $\kappa \in (0, \frac{1}{2})$. Note that the choice of κ does not influence the asymptotic behavior of the numbers in front of $\|G(z_1) - G(z_2)\|_{\mathbb{Y}}$ in (3.44), which enter the index function in Proposition 3.29.
Vice versa, assuming (3.44) holds true for all $z_1, z_2 \in D$ and all $\kappa \in (0, \infty)$, we regain (3.43) up to a constant by choosing $\kappa = \nu_j^\theta \|G(z_1) - G(z_2)\|_{\mathbb{Y}}^\theta \|z_1 - z_2\|_{\mathbb{X}}^{-\theta}$.

In the sequel we show that for a linear, injective and weak*-to-weak continuous operator every $x \in \mathbb{X}$ satisfies a variational source condition (3.14). For $\mathbb{X} = \ell^1$ this is the main result of the article [40]. Here we present a simpler proof based on (3.42) without evoking the adjoint of the forward operator.

Proposition 3.31. *Assume $\mathbb{X}$ is separable and has a predual. Suppose $G = A \colon \mathbb{X} \to \mathbb{Y}$ is linear, injective and weak*-to-weak continuous. Let $(P_j \colon \mathbb{X} \to \mathbb{X})_{j \in J}$ be a family of weak*-to-norm continuous projections satisfying (3.19a) and $\inf_{j \in J} \|(I - P_j)x\|_{\mathbb{X}} = 0$ for all $x \in \mathbb{X}$. Then for every $\kappa \in (0, \frac{1}{2})$ there exists $(\nu_j)_{j \in J}$ such that (3.42) holds true.*

Proof. By linearity it suffices to prove that for every $j \in J$ there exists $\nu_j > 0$ such that

$$\|P_j z\|_{\mathbb{X}} \leq \nu_j \|Az\|_{\mathbb{Y}} + \kappa \|z\|_{\mathbb{X}} \quad \text{for all } z \in \mathbb{X}. \quad (3.45)$$

Assuming the contrary, let $j \in J$ be such that for every $m \in \mathbb{N}$ there exists $x_m \in \mathbb{X}$ such that

$$\|P_j x_m\|_{\mathbb{X}} = 1 \quad \text{and} \quad m\|Ax_m\|_{\mathbb{Y}} + \kappa \|x_m\|_{\mathbb{X}} < 1.$$

Then $(x_m)_{m \in \mathbb{N}}$ is bounded and $\|Ax_m\|_{\mathbb{Y}} \to 0$. The Banach-Alaoglu theorem implies the existence of a subsequence $(x_{m_k})_{k \in \mathbb{N}}$ converging to some $x \in \mathbb{X}$ in the weak*-topology. The weak*-to-weak continuity of A yields weak convergence of Ax_{m_k} to Ax as $k \to \infty$. Since $\|Ax_{m_k}\|_{\mathbb{Y}} \to 0$ we obtain $Ax = 0$. Injectivity implies $x = 0$. The weak*-to-norm continuity of P_j leads to the contradiction

$$1 = \lim_{k \to \infty} \|P_j x_{m_k}\|_{\mathbb{X}} = \|P_j x\|_{\mathbb{X}} = 0. \qquad \square$$

Corollary 3.32. *In the setting of Proposition 3.31 for every $x \in \mathbb{X}$ and $\beta \in (0,1)$ there exists a concave and continuous function $\varphi \colon [0, \infty) \to [0, \infty)$ with $\varphi(0) = 0$ such that (3.14) holds true.*

Proof. Proposition 3.31 together with Proposition 3.29 yield the claim. $\qquad \square$

Example 3.33. *The space $\mathbb{X} = \ell^1$ is separable and the space c_0 of real null sequences equipped with the supremum norm is a predual. The weak*-convergence in $\mathbb{X}$ implies componentwise convergence. Therefore, the projections P_n defined in Example 3.6 are weak*-to-norm continuous. Furthermore, they satisfy (3.19a) and*

$$\inf_{n \in \mathbb{N}} \|(I - P_n)x\|_{\mathbb{X}} = \lim_{n \to \infty} \|(I - P_n)x\|_{\mathbb{X}} = 0 \quad \text{for all } x \in \mathbb{X}.$$

Error bounds for wavelet regularization in Besov Spaces with non-smooth basis

In this subsection we show that Theorem 3.15 remains valid without the assumption $a < s_{\max}$. Let $P_n \colon b^r_{p,1} \to b^r_{p,1}$ be the projection onto the first $n+1$ levels as defined in Section 3.2. We start with the replacement of the Bernstein type inequality in Lemma 3.9.

Lemma 3.34 (Replacement of Bernstein inequality). *Suppose Assumption 2.11. Let $a > 0$ and $0 \le r < s_{max}$. Then there exists a constant $C_{BR} > 0$ such that*

$$\|P_n z\|_{r,p,1} \le C_{BR} 2^{n(a+r)} \|\mathcal{S}z \mid B^{-a}_{p,\infty}\| + \frac{1}{4}\|z\|_{r,p,1} \quad \text{for all } z \in b^r_{p,1} \text{ and } n \in \mathbb{N}_0.$$

Proof. Let $\tilde{a} = \min\left\{\frac{s_{\max}}{2}, a\right\}$. Then $\tilde{a} < s_{\max}$. If $\tilde{a} < a$ then by Definition 2.9(d)(i) there exists a constant $c_I > 0$ such that

$$\|f \mid B^{-\tilde{a}}_{p,\infty}\| \le c_I \|f \mid B^{-a}_{p,\infty}\|^{\frac{\tilde{a}+r}{a+r}} \|f \mid B^r_{p,1}\|^{\frac{a-\tilde{a}}{a+r}} \quad \text{for all} \quad f \in B^r_{p,1}.$$

We may assume $c_I \ge 1$ such that this inequality remains valid for $\tilde{a} = a$.
Let $C_{\tilde{a}}$ and C_r, respectively, denote the operator norms of $\mathcal{S}^{-1} \colon B^{-\tilde{a}}_{p,\infty} \to b^{-\tilde{a}}_{p,\infty}$ and $\mathcal{S} \colon b^r_{p,1} \to B^r_{p,1}$. The above interpolation inequality with $f = \mathcal{S}z$ and Lemma 3.9 yield

$$\begin{aligned} \|P_n z\|_{r,p,1} &\le C_B 2^{n(\tilde{a}+r)} \|z\|_{-\tilde{a},p,\infty} \\ &\le C_B C_{\tilde{a}} 2^{n(\tilde{a}+r)} \|\mathcal{S}z \mid B^{-\tilde{a}}_{p,\infty}\| \\ &\le C_B C_{\tilde{a}} c_I C_r^{\frac{a-\tilde{a}}{a+r}} 2^{n(\tilde{a}+r)} \|\mathcal{S}z \mid B^{-a}_{p,\infty}\|^{\frac{\tilde{a}+r}{a+r}} \|z\|_{r,p,1}^{\frac{a-\tilde{a}}{a+r}}. \end{aligned}$$

As in Example 3.30 an application of Young's inequality yields the result. $\qquad \square$

Remark 3.35. *Note that the particular choice of $\tilde{a}$ in the proof of Lemma 3.34 is not relevant as long as $\tilde{a} < s_{max}$ and $\tilde{a} \leq a$. Moreover, the prefactor $\frac{1}{4}$ could be replaced by any $\kappa > 0$ (see also Example 3.30).*

With Proposition 3.29 and Lemma 3.34 we are in position to drop the assumption $a < s_{\max}$ in Proposition 3.12.

Proposition 3.36 (Variational source condition)**.** *Suppose Assumptions 2.11 and 2.15(a) hold true. Let $p \in [1, 2]$ and $0 \leq r < s_{max}$. If $p \in [1, 2)$, then we require that the scale $B_{p,q}^s$ is p-monotone (see Definition 2.9). Moreover, let $r < s$ and $\varrho > 0$ and assume $x \in b_{p,\infty}^s \cap \mathcal{S}^{-1}(\mathrm{dom}(F))$ with $\|x\|_{s,p,\infty} \leq \varrho$. Then*

$$\frac{1}{2}\|x - z\|_{r,p,1} + \|x\|_{r,p,1} - \|z\|_{r,p,1} \leq \varphi\left(\|F(\mathcal{S}x) - F(\mathcal{S}z)\|_{\mathbb{Y}}^2\right)$$

for all $z \in b_{p,1}^r \cap \mathcal{S}^{-1}(\mathrm{dom}(F))$ with $\varphi(\tau) = C'_\varphi \varrho^{\frac{a+r}{s+a}} \tau^{\frac{s-r}{2s+2a}}$ for a constant C'_φ independent of f, ϱ and τ.

Proof. We verify the assumptions of Proposition 3.29 for $G := F \circ \mathcal{S}$ on $D := b_{p,1}^r \cap \mathcal{S}^{-1}(\mathrm{dom}(F))$. As in the proof of Proposition 3.12 we see that there is a continuous embedding $B_{2,2}^{-a} \subset B_{p,\infty}^{-a}$. Let c_1 be its operator norm. Together with Assumption 2.15(a) we obtain

$$\|f_1 - f_2 \mid B_{p,\infty}^{-a}\| \leq c_1 M_1 \|F(f_1) - F(f_2)\|_{\mathbb{Y}} \quad \text{for all } f_1, f_2 \in \mathrm{dom}(F).$$

Let $z_1, z_2 \in D$. Inserting the last inequality with $f_i = \mathcal{S}z_i$ into the inequality in Lemma 3.34 we wind up with

$$\|P_n(z_1 - z_2)\|_{r,p,1} \leq \tilde{C}_{BR} 2^{n(a+r)} \|G(z_1) - G(z_2)\|_{\mathbb{Y}} + \frac{1}{4}\|z_1 - z_2\|_{r,p,1} \quad \text{for all } n \in \mathbb{N}_0$$

with $\tilde{C}_{BR} := C_{BR} c_1 M_1$. That is (3.42) with $\nu_n = \tilde{C}_{BR} 2^{n(a+r)}$ and $\kappa = \frac{1}{4}$.
Moreover, Lemma 3.10 yields

$$\|(I - P_n)x\|_{r,p,1} \leq C_J 2^{n(r-s)} \varrho.$$

Now for $z \in b_{p,1}^r \cap \mathcal{S}^{-1}(\mathrm{dom}(F))$ we set $\vartheta = \|F(\mathcal{S}x) - F(\mathcal{S}z)\|_{\mathbb{Y}}$. We apply Proposition 3.29 to obtain

$$\frac{1}{2}\|x - z\|_{r,p,1} + \|x\|_{r,p,1} - \|z\|_{r,p,1} \leq \tilde{\varphi}\left(\vartheta^2\right) \tag{3.46}$$

with $\tilde{\varphi}(\tau) = 2 \inf_{n \in \mathbb{N}_0} \left(\tilde{C}_{BR} 2^{n(a+r)} \sqrt{\tau} + C_J 2^{n(r-s)} \varrho\right)$ for $\tau > 0$.
We set $C'_\varphi := \max\{\frac{3}{2}c_2, 2(\tilde{C}_{BR} + 2^{s-r}C_J)\}$ with c_2 the embedding constant of the embedding $b_{p,\infty}^s \subset b_{p,1}^r$. With the same routine as at the end of the proof of Lemma 3.11 we see that $\tilde{\varphi}(\vartheta^2) \leq \varphi(\vartheta^2)$ whenever $\vartheta < \varrho$. Hence it remains to prove the result in the case $\vartheta \geq \varrho$. But here the triangle inequality (see also Remark 3.4) provides

$$\frac{1}{2}\|x - z\|_{r,p,1} + \|x\|_{r,p,1} - \|z\|_{r,p,1} \leq \frac{3}{2}\|x\|_{r,p,1} \leq C'_\varphi \varrho \leq C'_\varphi \varrho^{\frac{a+r}{s+a}} \vartheta^{\frac{s-r}{s+a}} = \varphi(\vartheta^2). \qquad \square$$

With Proposition 3.36 at hand the same arguments as in Section 3.2 lead to the same error bounds for the reconstruction R_α in (3.16).

Theorem 3.37 (Error bounds for non-smooth wavelet systems)**.** *Theorem 3.15 and Corollary 3.16 remain valid without the assumption $a < s_{max}$.*

Proof. The proofs work along the lines of the proof of Theorem 3.15 with Proposition 3.12 replaced by Proposition 3.36. □

Remark 3.38 ($p, q \geq 1$)**.** *Note that so far we have not used any spaces $b_{p,q}^s$ and $B_{p,q}^s$ with $p < 1$ or $q < 1$ such that it is sufficient to require Assumption 2.11 for $p, q \in [1, \infty]$.*

Numerical validation

We use the same setup and code as in the last section with Daubechies wavelets of order 3 instead of order 7 to reconstruct the coefficient c^{jump} with jumps in the elliptic boundary value problem (3.40). The regularity $s_{\max}$ of the order 3 wavelet system is estimated by 1.0878 (see [28, p. 239]). So we do not reconstruct c^{kink} here as the wavelet system is not capable of measuring Besov smoothness larger then 1.0878 and point out that we have $a = 2 > 1.0878 \sim s_{\max}$.

The lower regularity of the wavelet functions is clearly visible in the reconstructions. As expected the predicted rate agrees with the one in the experiment accurately in the case $p = 2$ (see Figure 3.7). Furthermore, in Figure 3.8 we see once more that our theory in the case $p = 1$ has to be revised as the error bound is to pessimistic. This motivates the further investigation in weighed ℓ^1-regularization in the next but one section.

FIGURE 3.7: *Left:* true coefficient c^{jump} with jumps in the boundary value problem (3.40) together with reconstructions for $b_{2,1}^0$-penalization with *db*3 and *db*7 at noise level $\delta = 2.5 \cdot 10^{-5}$ for the same data. *Right:* Reconstruction error using $b_{2,1}^0$-penalization, the rate $\mathcal{O}(\delta^{1/5})$ in the L^2-norm predicted by Theorem 3.37 as Example 3.17 yields $c^{\mathrm{jump}} \in B_{2,\infty}^{1/2}$.
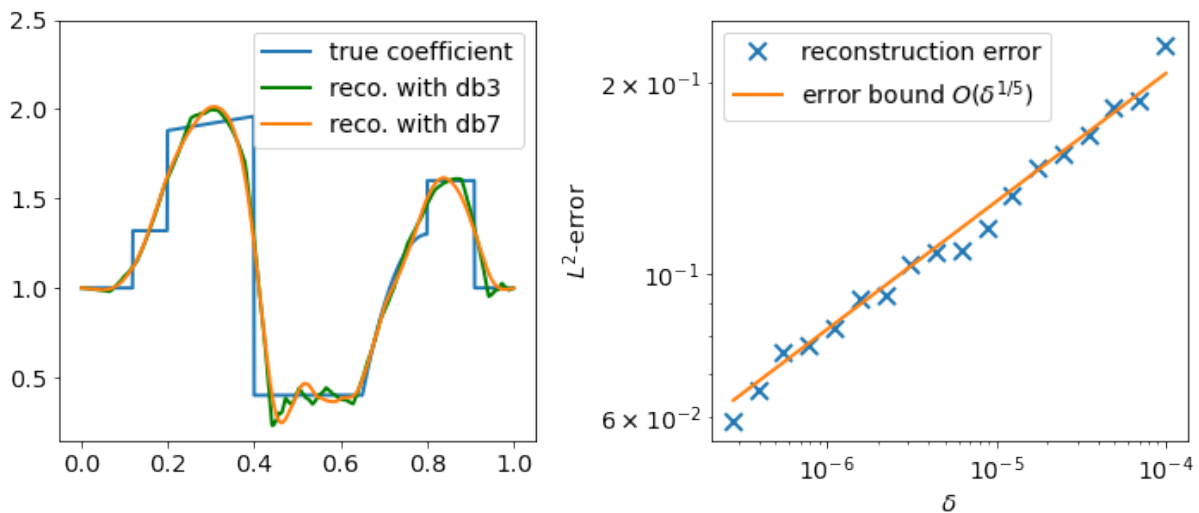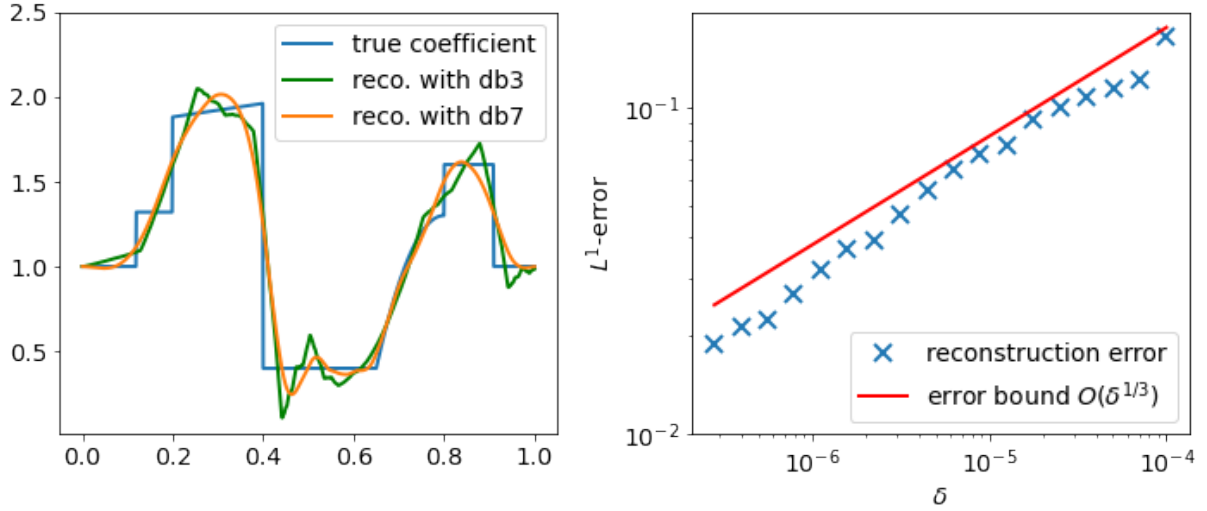
FIGURE 3.8: *Left:* true coefficient c^{jump} with jumps in the boundary value problem (3.40) together with reconstructions for $b^0_{1,1}$-penalization with *db*3 and *db*7 at noise level $\delta = 2.5 \cdot 10^{-5}$ for the same data. *Right:* Reconstruction error using $b^0_{1,1}$-penalization, the rate $\mathcal{O}(\delta^{1/3})$ in the L^1-norm predicted by Theorem 3.37 as Example 3.17 yields $c^{\text{jump}} \in B^1_{1,\infty}$.

Characterization as approximative source type condition

The purpose of this subsection is a characterization of (3.42) in terms of the range of the adjoint operator A^* in the case of a linear forward operator. Furthermore, we compare (3.42) to a source type condition that can be found in [41], [43], and [40].
We start with a lemma that can be seen as an approximative version of Lemma 3.25.

Lemma 3.39. *Suppose $A\colon \mathbb{X} \to \mathbb{Y}$ is linear and bounded, $P\colon \mathbb{X} \to \mathbb{X}$ is a projection with operator norm 1, $\kappa \in \left[0, \frac{1}{2}\right)$ and $\nu \geq 0$. Then the following statements are equivalent:*

(i) *$\|Pz\|_{\mathbb{X}} \leq \nu\|Az\|_{\mathbb{Y}} + \kappa\|z\|_{\mathbb{X}}$ for all $z \in \mathbb{X}$.*
(ii) *For every $\xi \in \text{im}(P^*)$ with $\|\xi\|_{\mathbb{X}'} \leq 1$, there exists $\psi \in \mathbb{Y}'$ with $\|\psi\|_{\mathbb{Y}'} \leq \nu$ such that $\|\xi - A^*\psi\|_{\mathbb{X}'} \leq \kappa$.*

If in addition P satisfies $\|z\|_{\mathbb{X}} = \|Pz\|_{\mathbb{X}} + \|(I - P)z\|_{\mathbb{X}}$ for all $z \in \mathbb{X}$, then we have equivalence of $(i), (ii)$ and the statement

(iii) *For every $\xi \in \text{im}(P^*)$ with $\|\xi\|_{\mathbb{X}'} \leq 1$, there exists $\psi \in \mathbb{Y}'$ with $\|\psi\|_{\mathbb{Y}'} \leq \frac{\nu}{1-\kappa}$ such that $P^*A^*\psi = \xi$ and $\|(I - P^*)A^*\psi\|_{\mathbb{X}'} \leq \frac{\kappa}{1-\kappa}$.*

Proof. We prove the equivalence of (i) and (ii) by tracing it back to Lemma 3.25. Consider the Banach space $\left(\mathbb{Z}_{(ii)}, \|\cdot\|_{(ii)}\right)$ given by $\mathbb{Z}_{(ii)} = \mathbb{Y} \times \mathbb{X}$ and $\|(y, z)\|_{(ii)} = \nu\|y\|_{\mathbb{Y}} + \kappa\|z\|_{\mathbb{X}}$ together with the bounded linear operator $S_{(ii)}\colon \mathbb{X} \to \mathbb{Z}_{(ii)}$ defined by $z \mapsto (Az, z)$. Then (i) is equivalent to

$$\|Pz\|_{\mathbb{X}} \leq \|Sz\|_{(ii)} \quad \text{for all} \quad z \in \mathbb{X}. \tag{3.47}$$

Introducing the norm $\|(\psi, \xi)\|_{(ii)'} := \max\left(\nu^{-1}\|\psi\|_{\mathbb{Y}'}, \kappa^{-1}\|\xi\|_{\mathbb{X}'}\right)$ on $\mathbb{Y}' \times \mathbb{X}'$ it is routine to see that the pairing $\langle(\psi, \xi), (y, x)\rangle := \langle\psi, y\rangle + \langle\xi, x\rangle$ gives rise to an isometric isomorphism

$\mathbb{Y}' \times \mathbb{X}' \cong \mathbb{Z}'_{(ii)}$. Hence, we may view the adjoint operator $S^*_{(ii)} \colon \mathbb{Y}' \times \mathbb{X}' \to \mathbb{X}'$, and a routine computation shows that it is given by $S^*_{(ii)}(\psi, \xi) = A^*\psi + \xi$.

Lemma 3.25 provides equivalence of (3.47) and the following statement: For every $\xi \in \operatorname{im}(P^*)$ there exists $\psi \in \mathbb{Y}'$ and $\xi_e \in \mathbb{X}'$ with $\|\psi\|_{\mathbb{Y}'} \le \nu$ and $\|\xi_e\|_{\mathbb{X}} \le \kappa$ such that $A^*\psi + \xi_e = \xi$. The latter statement is in turn equivalent to (ii).

Assuming $\|z\|_{\mathbb{X}} = \|Pz\|_{\mathbb{X}} + \|(I-P)z\|_{\mathbb{X}}$ we add or subtract $\kappa\|Pz\|_{\mathbb{X}}$ and divide or multiply by $1 - \kappa$ on both sides of the inequality in (i) to see that (i) is equivalent to

$$\|Pz\|_{\mathbb{X}} \le \frac{\nu}{1-\kappa}\|Az\|_{\mathbb{Y}} + \frac{\kappa}{1-\kappa}\|(I-P)z\|_{\mathbb{X}} \quad \text{for all} \quad z \in \mathbb{X}. \qquad (3.48)$$

Consider $\left(\mathbb{Z}_{(iii)}, \|\cdot\|_{(iii)}\right)$ given by $\mathbb{Z}_{(iii)} = \mathbb{Y} \times \mathbb{X}$ and $\|(y, z)\|_{(iii)} = \frac{\nu}{1-\kappa}\|y\|_{\mathbb{Y}} + \frac{\kappa}{1-\kappa}\|z\|_{\mathbb{X}}$ together with $S_{(iii)} \colon \mathbb{X} \to \mathbb{Z}_{(iii)}$ defined by $z \mapsto (Az, (1-P)z)$. The same arguments as above provide the equivalence of (3.48) and the following statement: For every $\xi \in \operatorname{im}(P^*)$ there exists $\psi \in \mathbb{Y}'$ and $\xi_e \in \mathbb{X}'$ with $\|\psi\|_{\mathbb{Y}'} \le \frac{\nu}{1-\kappa}$ and $\|\xi_e\|_{\mathbb{X}} \le \frac{\kappa}{1-\kappa}$ such that $A^*\psi + (1-P^*)\xi_e = \xi$. It is easy to see that the latter statement is equivalent to (iii). $\qquad\square$

The condition (ii) in Lemma 3.39 appears as a natural generalization of condition (ii) in Lemma 3.25. In Lemma 3.25 the elements in $\operatorname{im}(P^*)$ are required to be exactly realized as $A^*\psi$. In contrast, Lemma 3.39 allows an error of size κ.

We included statement (iii) for the purpose of comparing (3.42) to source conditions in the literature (see [41, 43, 40] and Remark 3.44). Here the requirement is an exact representation of ξ in the range of P^* while allowing for an error in the range of $I - P^*$.

Remark 3.40. *As in Remark 3.26 the statement of Lemma 3.39 remains true, if ξ in (ii) or (iii) ranges only over a set whose convex hull is $\{\xi \in \operatorname{im}(P^*) \colon \|\xi\|_{\mathbb{X}'} \le 1\}$.*

Corollary 3.41. *Suppose $G = A \colon \mathbb{X} \to \mathbb{Y}$ is linear and bounded, $(P_j \colon \mathbb{X} \to \mathbb{X})_{j \in J}$ a family of projections satisfying (3.19a). Let $\kappa \in \left[0, \frac{1}{2}\right)$. Then there exists a family $(\nu_j)_{j \in J}$ of positive real numbers such that (3.42) holds true if and only if for all $j \in J$ and $\xi \in \operatorname{im}(P^*_j)$ with $\|\xi\|_{\mathbb{X}'} \le 1$, there exists $\psi \in \mathbb{Y}'$ with $\|\psi\|_{\mathbb{Y}'} \le \nu_j$ such that $\|\xi - A^*\psi\|_{\mathbb{X}'} \le \kappa$.*

Proof. By linearity the inequality in (3.42) for all $z_1, z_2 \in \mathbb{X}$ is equivalent to statement (i) with $P = P_j$ in Lemma 3.39. $\qquad\square$

In the case of a linear operator on ℓ^1 the considerations lead to the following characterization of condition (3.42).

Proposition 3.42. *Let $\mathbb{X} = \ell^1$, $G = A \colon \ell^1 \to \mathbb{Y}$ linear, bounded and weak*-to-weak continuous, $(P_n)_{n \in \mathbb{N}}$ as in Example 3.6 and $\kappa \in \left[0, \frac{1}{2}\right)$. Then the minimal ν_n such that (3.42) holds true is given by*

$$\max_{\lambda \in \{-1,1\}^n} \inf \left\{ \|\psi\|_{\mathbb{Y}'} : \psi \in \mathbb{Y}' \text{ with } \left\|\sum_{j=1}^n \lambda_j e_j - A^*\psi\right\|_\infty \le \kappa \right\}.$$

Proof. Let $\underline{\nu}_n$ denote the maximum defined in the last line. Suppose $(\nu_n)_{n \in \mathbb{N}_0}$ is a sequence such that (3.42) holds true. We show that $\underline{\nu}_n \le \nu_n$. To this end, let $\lambda \in \{-1, 1\}^n$. Then $\|\sum_{j=1}^n \lambda_j e_j\|_\infty = 1$. By Lemma 3.39 there exists $\psi_\lambda \in \mathbb{Y}'$ with $\|\psi_\lambda\|_{\mathbb{Y}'} \le \nu_n$ and

$\left\| \sum_{j=1}^{n} \lambda_j e_j - A^* \psi \right\|_{\mathbb{X}'} \leq \kappa$. Hence, the infimum in the definition of $\underline{\nu}_n$ is less or equal ν_n. Taking the maximum over λ yields $\underline{\nu}_n \leq \nu_n$.

It remains to prove that (3.42) holds true with $\underline{\nu}_n \leq \nu_n$. Let $\varepsilon > 0$. For every $\lambda \in \{-1, 1\}^n$ there exists ψ with $\|\psi\|_{\mathbb{Y}'} \leq \underline{\nu}_n + \varepsilon$ and $\left\| \sum_{j=1}^{n} \lambda_j e_j - A^* \psi \right\|_{\mathbb{X}'} \leq \kappa$. Since $\left\{ \sum_{j=1}^{n} \lambda_j e_j \colon \lambda \in \{-1, 1\}^n \right\}$ is the set of extreme points of the unit ball in $\mathrm{im}(P_n^*)$, Remark 3.40 yields (3.42) with $\nu_n = \underline{\nu}_n + \varepsilon$, and the statement follows from taking the limit $\varepsilon \to 0$. $\qquad\square$

Remark 3.43. *In particular Proposition 3.42 together with Proposition 3.31 shows that every linear, bounded and weak*-to-weak continuous operator $A \colon \ell^1 \to \mathbb{Y}$ satisfies $e_j \in \overline{\mathrm{im}(A^*)}$ for all $j \in \mathbb{N}$. Hence, $c_0 \subset \overline{\mathrm{im}(A^*)}$. Moreover, the weak*-to-weak continuity yields $\mathrm{im}(A^*) \subset c_0$ (see [40, Lem. 3]). Therefore, we obtain $c_0 = \overline{\mathrm{im}(A^*)}$. The latter statement is proven in [40, Prop. 8] as an intermediate step to obtain the result in Corollary 3.32.*

Remark 3.44 (Answers to questions posed in [41] and [43])**.** *Lemma 3.39 and Remark 3.40 prove that for a sequence $(\nu_n)_{n \in \mathbb{N}} \subset [0, \infty)$ of non-negative numbers and a constant $\kappa \in [0, \frac{1}{2})$ condition (3.42) holds true if and only if for every $n \in \mathbb{N}$ and $\xi \in \ell^\infty$ with $\xi_i \in \{-1, 1\}$ for $i \leq n$ and $\xi_i = 0$ for $i > n$ there exists $\psi \in \mathbb{Y}'$ such that $\|\psi\|_{\mathbb{Y}'} \leq \frac{\nu_n}{1-\kappa}$, $P_n^* A^* \psi = \xi$ and $\|(I - P_n^*) A^* \psi\|_\infty \leq \frac{\kappa}{1-\kappa}$.*
Up to an irrelevant reparameterization the latter condition matches [40, Condition 4] which is the starting point of the analysis in [40]. We refer also to [41, Def. 3.1; Ass. 3.2] and [43, Ass. 2.1] for variants of the latter condition. For a linear operator the condition (3.42) is equivalent to this source condition. This answers a question posed in [41, Ch. 7]. As (3.42) does not evoke the adjoint of the forward operator, it is qualified for the generalization to nonlinear mappings.
In [40, Ch. 5] the authors point to the question of the influence of the constant κ on the sequence ν_n. In Example 3.30 we saw that an interpolation type inequality of the form

$$\|P_n z\|_1 \leq \nu_j^\theta \|Az\|_{\mathbb{Y}}^\theta \|z\|_1^{1-\theta} \quad \text{for all } z \in \ell^1$$

implies (3.42) and therefore Condition 4 in [40] and the constant κ influences the sequence ν_n only by a constant factor. In the proof of Lemma 3.34 we saw that such an inequality holds true for finitely smoothing operators on Besov spaces. Hence, at least for these operators the constant κ does not influence the asymptotic behavior of the sequence ν_n (see also Example 3.30 and Remark 3.35)

3.4 Optimality

Here we show that the upper bounds on the reconstruction error in Theorem 3.15 and therefore also in Theorem 3.37 are order optimal.

To do so we provide a lower bound on the modulus of continuity on Besov balls

$$K_{p,q}^s(\overline{f}, \varrho) := \left\{ f \in B_{p,q}^s \colon \|f - \overline{f} \mid B_{p,q}^s\| \leq \varrho \right\}$$

with $s > 0, p, q \in (0, \infty), \overline{f} \in B_{p,q}^s$ and $\varrho > 0$. Similar results (with more restrictions on the involved indices) can be found in [104, Prop. 3.21] and [64, Thm. 4.5].

Theorem 3.45 (Lower bound on the reconstruction error). *Let $B_{p,q}^s$ be a Besov scale (see Definition 2.9) such that for every $s_{max} > 0$ there exists a wavelet system meeting Assumption 2.11. Suppose Assumption 2.15(b) holds true. Let $s > 0$ and $p, q \in (0, \infty]$ and assume that $\overline{f} \in B_{p,q}^s$ and $\varrho > 0$ such that $K_{p,q}^s(\overline{f}, \varrho) \subset \mathrm{dom}(F)$. If $\overline{p}, \overline{q} \in [1, \infty]$ with a continuous embedding $B_{p,q}^s \subset B_{\overline{p},\overline{q}}^0$. Then there is a constant $c > 0$ independent of $\overline{f}$ and ϱ such that the worst case error on $K_{p,q}^s(\overline{f}, \varrho)$ given by*

$$
\Delta_R(\delta, K_{p,q}^s(\overline{f}, \varrho))
$$
$$
= \sup \left\{ \|f - R(\delta, g^{\mathrm{obs}}) \,|\, B_{\overline{p},\overline{q}}^0\| \colon f \in K_{p,q}^s(\overline{f}, \varrho),\ g^{\mathrm{obs}} \in \mathbb{Y} \text{ with } \|g^{\mathrm{obs}} - F(f)\|_{\mathbb{Y}} \le \delta \right\}
$$

of any reconstruction method given by a map $R \colon (0, \infty) \times \mathbb{Y} \to B_{\overline{p},\overline{q}}^0$ satisfies the lower bound

$$
c \varrho^{\frac{a}{s+a}} \delta^{\frac{s}{s+a}} \le \Delta_R(\delta, K_\varrho) \quad \text{for all } \delta < \frac{\varrho}{2}.
$$

Proof. We show that the modulus of continuity

$$
\omega\left(\delta, K_{p,q}^s(\overline{f}, \varrho)\right) = \sup \left\{ \|f_1 - f_2 \,|\, B_{\overline{p},\overline{q}}^0\| \colon f_1, f_2 \in K_{p,q}^s(\overline{f}, \varrho) \text{ with } \|F(f_1) - F(f_2)\|_{\mathbb{Y}} \le \delta \right\}
$$

satisfies

$$
\overline{c} \varrho^{\frac{a}{s+a}} \delta^{\frac{s}{s+a}} \le \omega\left(\delta, K_{p,q}^s(\overline{f}, \varrho)\right) \quad \text{for all } \delta < \varrho
$$

with a constant $\overline{c}$ independent of $\overline{f}$ and ϱ. Then the result is an immediate consequence of the universal lower bound on the worst case error in Proposition 2.6.

Let $\varrho > 0$ and $\delta > 0$. For $n \in \mathbb{N}_0$ we define $\beta_n := \min\{2^{-ns}\varrho, 2^{na}\delta\}$. Suppose for a moment that $(f_n)_{n \in \mathbb{N}_0} \subset B_{p,q}^s$ is a sequence with

$$
\|f_n \,|\, B_{p,q}^s\| \le \varrho, \quad \|F(\overline{f} + f_n) - F(\overline{f})\|_{\mathbb{Y}} \le \delta \text{ and } \quad \|f_n \,|\, B_{\overline{p},\overline{q}}^0\| \ge c_1 \beta_n \tag{3.49}
$$

for all $n \in \mathbb{N}_0$ with a constant c_1 independent $\overline{f}$, ϱ_0, ϱ and δ.

Then for $\delta < \varrho$ there exists a minimal natural number $n_0 \ge 1$ with $(\varrho/\delta)^{\frac{1}{s+a}} \le 2^{n_0}$. With this $\beta_{n_0} = 2^{-n_0 s}\varrho$ and $2^{n_0} < 2\,(\varrho/\delta)^{\frac{1}{s+a}}$ imply

$$
\omega\left(\delta, K_{p,\infty}^s(\overline{f}, \varrho), \|\cdot\|_{0,p,1}\right) \ge \|f_{n_0} \,|\, B_{\overline{p},\overline{q}}^0\| \ge c_1 \beta_{n_0} = c_1 2^{-n_0 s}\varrho > c_1 2^{-s}\varrho^{\frac{a}{s+a}}\delta^{\frac{s}{s+a}}.
$$

Hence we are left with constructing a sequence satisfying (3.49). To this end, choose a wavelet system with $s_{\max} > 0$ large enough such that $\mathcal{S}$ acts as a Banach space isomorphism between the three spaces $b_{p,q}^s$, $b_{\overline{p},\overline{q}}^0$, $b_{2,2}^{-a}$ and $B_{p,q}^s$, $B_{\overline{p},\overline{q}}^0$, $B_{2,2}^{-a}$ respectively. With c_2 the maximum of the operator norms of these three isomorphism and their inverses we set $c_3 := c_2^{-1} \min\{1, M_2^{-1}\}$.

Now fix $n \in \mathbb{N}_0$. Choose a subset $\overline{\Lambda}_j \subset \Lambda_j$ with $|\overline{\Lambda}_j| = 2^{jd}$. Let $x \in \mathbb{R}^\Lambda$ given by $x_{j,k} = c_3 2^{-\frac{1}{2}nd}\beta_n$ if $j = n$ and $k \in \overline{\Lambda}_j$ and $x_{j,k} = 0$ else. We define $f_n := \mathcal{S}x$ and estimate

$$
\|f_n \,|\, B_{p,q}^s\| \le c_2 \|x\|_{s,p,q} = c_2 c_3 2^{ns}\beta_n \le \varrho,
$$
$$
\|f_n \,|\, B_{2,2}^{-a}\| \le c_2 \|x\|_{-a,2,2} = c_2 c_3 2^{-na}\beta_n \le M_2^{-1}\delta \quad \text{and}
$$
$$
\|f_n \,|\, B_{\overline{p},\overline{q}}^0\| \ge c_2^{-1}\|x\|_{0,\overline{p},\overline{q}} = c_2^{-1} c_3 \beta_n.
$$

Using the Lipschitz continuity Assumption 2.15(b) we see that the obtained sequence $(f_n)_{n \in \mathbb{N}_0}$ satisfies the requirements in (3.49). $\qquad\square$

As already noted below Assumption 2.11 all Besov function spaces $B_{p,q}^s(\Omega)$ with $\Omega \in \{\mathbb{R}^d, \mathbb{T}^d\}$ can be characterized by wavelet systems. Hence in this case the assumption on the Besov scale in Theorem 3.45 is satisfied.

This result shows that under Assumption 2.15 the upper bounds on the reconstruction error (3.36) is order optimal, i.e. the wavelet reconstruction method given by R_α in (3.16) is order optimal in the sense of Section 2.2 for both the a priori choice rule as well as for the discrepancy principle.

Note that this is no contradiction to the claims in the last sections on insufficiency of the predictions we have for piecewise smooth functions with kinks or jumps with a $b_{1,1}^r$-penalty: The stated optimality holds for suprema over balls in $B_{p,\infty}^s$. However, Theorem 3.45 does not rule out the existence of elements in such balls (such as piecewise smooth functions) for which faster convergence rates can be achieved (even if s is maximal for these elements).

Remark 3.46 (Extension to $L^{\overline{p}}$-loss)**.** *In view of Theorem 3.45 we also get that the bounds in the $L^{\overline{p}}$-norm (see Corollary 3.16) cannot be improved up to a change of the constant. To this end, let $\overline{p} \in [1, \infty)$. Suppose that $B_{\overline{p},\infty}^0 = B_{\overline{p},\infty}^0(\Omega)$ with Ω either $\mathbb{T}^d$, $\mathbb{R}^d$ or a bounded Lipschitz domain in $\mathbb{R}^d$. Then there is a continuous embedding $L^{\overline{p}}(\Omega) \subset B_{\overline{p},\infty}^0(\Omega)$ (see Proposition A.11). Hence Theorem 3.45 remains valid if one replaces the space $B_{\overline{p},\overline{q}}^0(\Omega)$ by $L^{\overline{p}}(\Omega)$.*

Remark 3.47. *Finally we like to point out that Theorem 3.45 also generalizes to loss functions given by $B_{\overline{p},\overline{q}}^r$-norms with $r \geq 0$. We obtain lower bounds $\varrho^{\frac{a+r}{s+a}} \delta^{\frac{s-r}{s+a}} \lesssim \omega\left(\delta, K_{p,q}^s(\overline{f}, \varrho)\right)$. This shows that the upper bound (3.34) is order optimal as well.*

3.5 Weighted ℓ^1-regularization

We have seen in our numerical simulation displayed in Figures 3.4 and 3.5 that our error bounds in the case $p = 1$ for piecewise smooth functions with jumps or kinks are too pessimistic. Therefore, the aim of this section is to characterize maximal sets on which certain approximation rates are achieved.

Using weak sequence spaces to measure the regularity of the true solution we will prove convergence rates for weighted ℓ^1-regularization as introduced in Section 2.5. The results of this section can also be found in the article [85] to which the author contributed. For some data $g \in \mathbb{Y}$ we consider the following set-valued scheme

$$S_\alpha(g) = \operatorname*{argmin}_{x \in \operatorname{dom}(G) \cap \ell_r^1} \left(\frac{1}{2\alpha} \|g - G(x)\|_{\mathbb{Y}}^2 + \|x\|_{\overline{r},1} \right). \tag{3.50}$$

The structure of this section is as follows: First we will introduce weak sequence spaces and show that these are characterized by approximation rates of certain thresholding operators. The latter thresholding procedures will play a central role in the analysis. Next we verify a variational source condition on bounded sets of these weak sequence spaces. To this end, we introduce a family $(P_\alpha)_{\alpha>0}$ adaptively to the true solution x consisting of projections onto those indices where the coefficients of x lie above some threshold. Providing a Bernstein type inequality and using the characterization by thresholding mentioned above we apply Proposition 3.5 to obtain a variational source condition of the

form (3.14). As usual this condition implies convergence rates.

We finish this section with an application to wavelet regularization with penalties given by $b_{1,1}^r$-norms where we use similar arguments as in Section 3.3 to provide a convergence rate result also for non-smooth wavelet systems.

Weak sequence spaces

As announced we introduce spaces of sequences whose bounded sets will provide the source sets for the convergence analysis in the sequel. We define a specific thresholding map and analyze its approximation properties.

As in Section 2.5 we let Λ be a countable set and $\overline{r} = (\overline{r}_j)_{j \in \Lambda}$, $\overline{a} = (\overline{a}_j)_{j \in \Lambda}$ sequences of positive weights. First we introduce a scale of spaces, part of which interpolates between the spaces $\ell_{\overline{r}}^1$ and $\ell_{\overline{a}}^2$ involved in our setting. For $t \in (0, 2]$ we define weights

$$(\overline{\omega}_t)_j = (\overline{a}_j^{2t-2} \overline{r}_j^{2-t})^{\frac{1}{t}}. \tag{3.51}$$

Note that $\overline{\omega}_1 = \overline{r}$ and $\overline{\omega}_2 = \overline{a}$. The next proposition captures interpolation inequalities we will need later.

Proposition 3.48 (Interpolation inequality)**.** *Let $u, v, t \in (0, 2]$ and $\theta \in (0, 1)$ with $\frac{1}{t} = \frac{1-\theta}{u} + \frac{\theta}{v}$. Then*

$$\|x\|_{\overline{\omega}_t, t} \leq \|x\|_{\overline{\omega}_u, u}^{1-\theta} \|x\|_{\overline{\omega}_v, v}^{\theta} \quad \text{for all } x \in \ell_{\overline{\omega}_u}^u \cap \ell_{\overline{\omega}_v}^v.$$

Proof. We use Hölder's inequality with the conjugate exponents $\frac{u}{(1-\theta)t}$ and $\frac{v}{\theta t}$:

$$\|x\|_{\overline{\omega}_t, t}^t = \sum_{j \in \Lambda} \left(\overline{a}_j^{2u-2} \overline{r}_j^{2-u} |x_j|^u\right)^{\frac{(1-\theta)t}{u}} \left(\overline{a}_j^{2v-2} \overline{r}_j^{2-v} |x_j|^v\right)^{\frac{\theta t}{v}} \leq \|x\|_{\overline{\omega}_u, u}^{(1-\theta)t} \|x\|_{\overline{\omega}_v, v}^{\theta t}. \qquad \square$$

Remark 3.49. *In the setting of Proposition 3.48 real interpolation theory yields the stronger statement $\ell_{\overline{\omega}_t}^t = (\ell_{\overline{\omega}_u}^u, \ell_{\overline{\omega}_v}^v)_{\theta, t}$ with equivalent quasi-norms (see, e.g. [45, Theorem 2]). The stated interpolation inequality is a consequence.*

For $t \in (0, 2)$ we define a weak version of the space $\ell_{\overline{\omega}_t}^t$.

Definition 3.50 (Source sets)**.** *Let $t \in (0, 2)$. We define $k_t := \{x \in \mathbb{R}^\Lambda : \|x\|_{k_t} < \infty\}$ with*

$$\|x\|_{k_t} = \sup_{\alpha > 0} \alpha \left(\sum_{j \in \Lambda} \overline{a}_j^{-2} \overline{r}_j^2 \mathbb{1}_{\{\overline{a}_j^{-2} \overline{r}_j \alpha < |x_j|\}}\right)^{\frac{1}{t}}.$$

Remark 3.51. *The functions $\|\cdot\|_{k_t}$ are quasi-norms. The quasi-Banach spaces k_t are weighted Lorenz spaces. They appear as real interpolation spaces between weighted L^p-spaces. To be more precise [45, Theorem 2] yields $k_t = (\ell_{\overline{\omega}_u}^u, \ell_{\overline{\omega}_v}^v)_{\theta, \infty}$ with equivalence of quasi-norms for u, v, t and θ as in Proposition 3.48.*

Remark 3.52. *Remark 3.49 and Remark 3.51 predict an embedding*

$$\ell_{\overline{\omega}_t}^t = (\ell_{\overline{\omega}_u}^u, \ell_{\overline{\omega}_v}^v)_{\theta, t} \subset (\ell_{\overline{\omega}_u}^u, \ell_{\overline{\omega}_v}^v)_{\theta, \infty} = k_t.$$

Indeed, the Markov-type inequality

$$\alpha^t \sum_{j \in \Lambda} \overline{a}_j^{-2} \overline{r}_j^2 \mathbb{1}_{\{\overline{a}_j^{-2}\overline{r}_j\alpha < |x_j|\}} \leq \sum_{j \in \Lambda} \overline{a}_j^{2t-2} \overline{r}_j^{2-t} |x_j|^t = \|x\|_{\overline{\omega}_t, t}^t$$

proves $\| \cdot \|_{k_t} \leq \|\cdot\|_{\overline{\omega}_t, t}$.

For $\overline{a}_j = \overline{r}_j = 1$ we obtain the weak-ℓ_t spaces $k_t = \ell_{t,\infty}$ that appear in nonlinear approximation theory (see e.g. [24], [21]).

Next we define a specific nonlinear thresholding procedure depending on r and a whose approximation theory is characterized by the spaces k_t. This characterization is the core for the proofs of the convergence results in this section. The statement is [24, Theorem 7.1] for weighed sequence space. For sake of completeness we present an elementary proof based on a partition trick that is perceivable in the proof of [24, Theorem 4.2].

Let $\alpha > 0$. We consider the map

$$T_\alpha \colon \mathbb{R}^\Lambda \to \mathbb{R}^\Lambda \quad \text{by} \quad T_\alpha(x)_j := \begin{cases} x_j & \text{if } \overline{a}_j^{-2}\overline{r}_j\alpha < |x_j| \\ 0 & \text{else.} \end{cases} \tag{3.52}$$

Let $x \in \ell_{\overline{a}}^2$. Then

$$\|x - T_\alpha(x)\|_{\overline{a},2}^2 = \sum_{j \in \Lambda} \overline{a}_j^2 |x_j|^2 \mathbb{1}_{\{|x_j| \leq \overline{a}_j^{-2}\overline{r}_j\alpha\}} \longrightarrow 0 \quad \text{for } \alpha \to 0$$

by Lebesgue dominated convergence. Moreover, note that

$$\alpha^2 \sum_{j \in \Lambda} \overline{a}_j^{-2} \overline{r}_j^2 \mathbb{1}_{\{\overline{a}_j^{-2}\overline{r}_j\alpha < |x_j|\}} \leq \|T_\alpha(x)\|_{\overline{a},2}^2 \leq \|x\|_{\overline{a},2}^2.$$

If $\overline{a}_j \overline{r}_j^{-1}$ is bounded above then $\overline{a}_j^{-2} \overline{r}_j^2$ is bounded away from zero. Hence in this case we see that the set of $j \in \Lambda$ with $\overline{a}_j^{-2}\overline{r}_j\alpha < |x_j|$ is finite, i.e. $T_\alpha(x)$ has only finitely many non vanishing coefficients whenever $x \in \ell_{\overline{a}}^2$.

Lemma 3.53 (Approximation rates for T_α). *Let $0 < t < p \leq 2$ and $x \in \mathbb{R}^\Lambda$. Then $x \in k_t$ if and only if $\eta(x) := \sup_{\alpha>0} \alpha^{\frac{t-p}{p}} \|x - T_\alpha(x)\|_{\overline{\omega}_p, p} < \infty$.*
More precisely we show bounds

$$\eta(x) \leq 2 \left(2^{p-t} - 1\right)^{-\frac{1}{p}} \|x\|_{k_t}^{\frac{t}{p}} \quad \text{and} \quad \|x\|_{k_t} \leq 2^{\frac{p}{t}} (2^t - 1)^{-\frac{1}{t}} \eta(x)^{\frac{p}{t}}.$$

Proof. We use a partitioning to estimate

$$\|x - T_\alpha(x)\|_{\overline{\omega}_p, p}^p = \sum_{j \in \Lambda} \overline{a}_j^{2p-2} \overline{r}_j^{2-p} |x_j|^p \mathbb{1}_{\{|x_j| \leq \overline{a}_j^{-2}\overline{r}_j\alpha\}}$$

$$= \sum_{k=0}^{\infty} \sum_{j \in \Lambda} \overline{a}_j^{2p-2} \overline{r}_j^{2-p} |x_j|^p \mathbb{1}_{\{\overline{a}_j^{-2}\overline{r}_j 2^{-(k+1)}\alpha < |x_j| \leq \overline{a}_j^{-2}\overline{r}_j 2^{-k}\alpha\}}$$

$$\leq \alpha^p \sum_{k=0}^{\infty} 2^{-pk} \sum_{j \in \Lambda} \overline{a}_j^{-2} \overline{r}_j^2 \mathbb{1}_{\{\overline{a}_j^{-2}\overline{r}_j 2^{-(k+1)}\alpha < |x_j|\}}$$

$$\leq \alpha^{p-t} \|x\|_{k_t}^t 2^t \sum_{k=0}^{\infty} (2^{t-p})^k$$

$$= \alpha^{p-t} 2^p \left(2^{p-t} - 1\right)^{-1} \|x\|_{k_t}^t.$$

A similar estimation yields the second inequality:

$$
\begin{aligned}
\sum_{j\in\Lambda} \overline{a}_j^{-2}\overline{r}_j^2 \mathbb{1}_{\{\overline{a}_j^{-2}\overline{r}_j\alpha<|x_j|\}} &= \sum_{k=0}^{\infty}\sum_{j\in\Lambda} \overline{a}_j^{-2}\overline{r}_j^2 \mathbb{1}_{\{\overline{a}_j^{-2}\overline{r}_j 2^k\alpha<|x_j|\leq \overline{a}_j^{-2}\overline{r}_j 2^{k+1}\alpha\}} \\
&\leq \alpha^{-p}\sum_{k=0}^{\infty} 2^{-kp}\sum_{j\in\Lambda} \overline{a}_j^{2p-2}\overline{r}_j^{2-p}|x_j|^p \mathbb{1}_{\{|x_j|\leq \overline{a}_j^{-2}\overline{r}_j 2^{k+1}\alpha\}} \\
&= \alpha^{-p}\sum_{k=0}^{\infty} 2^{-kp}\|x - T_{2^{k+1}\alpha}(x)\|_{\overline{\omega}_p,p}^p \\
&\leq \alpha^{-t}\eta(x)^p 2^{p-t}\sum_{k=0}^{\infty}(2^{-t})^k \\
&= \alpha^{-t}\eta(x)^p 2^p \left(2^t - 1\right)^{-1}. \qquad\qquad \square
\end{aligned}
$$

Corollary 3.54. *Suppose $\overline{a}_j\overline{r}_j^{-1}$ is bounded above. Let $0 < t < p \leq 2$. Then $k_t \subset \ell_{\overline{\omega}_p}^p$. More precisely there is a constant $M > 0$ depending on t, p and $\sup_{j\in\Lambda}\overline{a}_j\overline{r}_j^{-1}$ such that $\|\cdot\|_{\overline{\omega}_p,p} \leq M\|\cdot\|_{k_t}$.*

Proof. Let $x \in k_t$. The assumption implies the existence of a constant $c > 0$ with $c \leq \overline{a}_j^{-2}\overline{r}_j^2$ for all $j \in \Lambda$. Let $\alpha > 0$. Then

$$
c\sum_{j\in\Lambda} \mathbb{1}_{\{\overline{a}_j^{-2}\overline{r}_j\alpha<|x_j|\}} \leq \sum_{j\in\Lambda} \overline{a}_j^{-2}\overline{r}_j^2 \mathbb{1}_{\{\overline{a}_j^{-2}\overline{r}_j\alpha<|x_j|\}} \leq \|x\|_{k_t}^t \alpha^{-t}.
$$

Inserting $\overline{\alpha} := 2\|x\|_{k_t}c^{-\frac{1}{t}}$ implies $\overline{a}_j^{-2}\overline{r}_j\overline{\alpha} \geq |x_j|$ for all $j \in \Lambda$. Hence, $T_{\overline{\alpha}}(x) = 0$. With $C = 2\left(2^{p-t} - 1\right)^{-\frac{1}{p}}$ Lemma 3.53 yields

$$
\|x\|_{\overline{\omega}_p,p} = \|x - T_{\overline{\alpha}}(x)\|_{\overline{\omega}_p,p} \leq C\|x\|_{k_t}^{\frac{t}{p}}\overline{\alpha}^{\frac{p-t}{p}} = 2^{\frac{p-t}{p}}Cc^{\frac{t-p}{tp}}\|x\|_{k_t}. \qquad \square
$$

Error bounds on bounded sets in k_t

For $t \in (0,1)$ and $x \in k_t$ we prove a variational source condition using Proposition 3.5. The key difference to the verifications in Section 3.2 and 3.3 is that we choose the family of projections involved in Proposition 3.5 adaptively to x. This possibility is already mentioned, but not further exploited in [43, Remark 2.4] and [40, Ch. 5].

We introduce coordinate projections onto the support of $T_\alpha(x)$ and provide a Bernstein type inequality for them.

Lemma 3.55 (Bernstein inequality). *Let $t \in (0,2)$, $p \in (1,2]$, $\alpha, \varrho > 0$ and $x \in k_t$ with $\|x\|_{k_t} \leq \varrho$. We consider*

$$
\Lambda_\alpha = \{j \in \Lambda\colon \overline{a}_j^{-2}\overline{r}_j\alpha < |x_j|\}
$$

and the coordinate projection $P_\alpha\colon \mathbb{R}^\Lambda \to \mathbb{R}^\Lambda$ onto Λ_α given by $(P_\alpha z)_j = z_j$ if $j \in \Lambda_\alpha$ and $(P_\alpha z)_j = 0$ else. Then

$$
\|P_\alpha z\|_{\overline{r},1} \leq \varrho^{\frac{t(p-1)}{p}}\alpha^{-\frac{t(p-1)}{p}}\|z\|_{\overline{\omega}_p,p} \quad \textit{for all} \quad z \in \ell_{\overline{\omega}_p}^p.
$$

Proof. We prove the claim using Hölder's inequality as follows

$$\|P_\alpha z\|_{\overline{r},1} = \sum_{j \in \Lambda} \left((\overline{\omega}_p)_j^{-1} \, \overline{r}_j \mathbb{1}_{\{\overline{a}_j^{-2}\overline{r}_j\alpha < |x_j|\}} \right) \left((\overline{\omega}_p)_j \, |z_j| \right)$$

$$\leq \left(\sum_{j \in \Lambda} \overline{a}_j^{-2}\overline{r}_j^2 \mathbb{1}_{\{\overline{a}_j^{-2}\overline{r}_j\alpha < |x_j|\}} \right)^{\frac{p-1}{p}} \left(\sum_{j \in \Lambda} (\overline{\omega}_p)_j^p \, |z_j|^p \right)^{\frac{1}{p}} \leq \|x\|_{k_t}^{\frac{t(p-1)}{p}} \alpha^{-\frac{t(p-1)}{p}} \|z\|_{\overline{\omega}_p,p}. \square$$

Proposition 3.56 (Variational source condition). *Assume Assumption 2.27. Let $t \in (0,1)$ and $\varrho > 0$. Suppose $x \in k_t \cap \mathrm{dom}(G)$ with $\|x\|_{k_t} \leq \varrho$. Then*

$$\|x - z\|_{\overline{r},1} + \|x\|_{\overline{r},1} - \|z\|_{\overline{r},1} \leq \varphi \left(\|G(x) - G(z)\|_\mathbb{Y}^2 \right) \quad \text{for all} \quad z \in \ell_{\frac{1}{\overline{r}}}^1 \cap \mathrm{dom}(G) \quad (3.53)$$

with $\varphi(\tau) = C_\varphi \varrho^{\frac{t}{2-t}} \tau^{\frac{1-t}{2-t}}$. Here C_φ denotes a constant depending only on t and M_1.

Proof. Let $J := (0, \infty)$ and $(P_\alpha)_{\alpha \in J}$ be as introduced in Lemma 3.55. Then Equation (3.19a) is satisfied. Using Lemma 3.55 with $p = 2$ and Assumption 2.27 we obtain

$$\|P_\alpha(z_1 - z_2)\|_{\overline{r},1} \leq M_1 \varrho^{\frac{t}{2}} \alpha^{-\frac{t}{2}} \|G(z_1) - G(z_2)\|_\mathbb{Y} \quad \text{for all } z_1, z_2 \in \ell_{\frac{1}{\overline{r}}}^1 \cap \mathrm{dom}(G).$$

Note that $P_\alpha x = T_\alpha(x)$. Hence, Lemma 3.53 yields

$$\|(I - P_\alpha)x\|_{\overline{r},1} = \|x - T_\alpha(x)\|_{\overline{r},1} \leq 2(2^{1-t} - 1)^{-1} \varrho^t \alpha^{1-t}.$$

We set $\tilde{C} := 2\max\left\{ M_1, 2(2^{1-t} - 1)^{-1} \right\}$ and apply Proposition 3.5 to obtain (3.53) with

$$\varphi(\tau) = \tilde{C} \inf_{\alpha > 0} \left(\varrho^{\frac{t}{2}} \alpha^{-\frac{t}{2}} \sqrt{\tau} + \varrho^t \alpha^{1-t} \right) \leq 2\tilde{C} \varrho^{\frac{t}{2-t}} \tau^{\frac{1-t}{2-t}}.$$

where the last step follows from inserting $\alpha = \tau^{\frac{1}{2-t}} \varrho^{-\frac{t}{2-t}}$. Redefining $\varphi(\tau) := C_\varphi \varrho^{\frac{t}{2-t}} \tau^{\frac{1-t}{2-t}}$ with $C_\varphi = 2\tilde{C}$ yields the claim. $\square$

Remark 3.57 (Lipschitz condition for G^{-1} suffices). *Note that as in Proposition 3.12 and Proposition 3.36 we only need the first inequality in Assumption 2.27 for the verification of the variational source condition.*

Remark 3.58 (Converse result to variational source condition). *In the setting of Proposition 3.56 we additionally assume $\ell_{\frac{1}{\overline{r}}}^1 \subset \mathrm{dom}(G)$ to investigate into a converse statement to Proposition 3.56. To this end, suppose $x \in \ell_{\frac{1}{\overline{r}}}^1$ satisfies 3.53 with $\varphi(\tau) = K\tau^{\frac{1-t}{2-t}}$. We remove the loss function and use Assumption 2.27 to obtain*

$$\|x\|_{\overline{r},1} - \|z\|_{\overline{r},1} \leq KM_2 \|x - z\|_{\overline{a},2}^{\frac{2-2t}{2-t}} \quad \text{for all } x \in \ell_{\frac{1}{\overline{r}}}^1. \quad (3.54)$$

As in Remark 3.21 we insert specific elements z into (3.54) to conclude $x \in k_t$: Let $\alpha > 0$. We define

$$z_j := \begin{cases} x_j & \text{if } |x_j| \leq \overline{a}_j^{-2}\overline{r}_j\alpha \\ x_j - \overline{a}_j^{-2}\overline{r}_j\alpha & \text{if } x_j > \overline{a}_j^{-2}\overline{r}_j\alpha \\ x_j + \overline{a}_j^{-2}\overline{r}_j\alpha & \text{if } x_j < -\overline{a}_j^{-2}\overline{r}_j\alpha \end{cases}.$$

Then $|z_j| \le |x_j|$ for all $j \in \Lambda$. Hence, $z \in \ell^1_{\overline{r}}$. With (3.54) we estimate

$$\alpha \sum_{j \in \Lambda} \overline{a}_j^{-2} \overline{r}_j^2 \mathbb{1}_{\{\overline{a}_j^{-2} \overline{r}_j \alpha < |x_j|\}} = \|x\|_{\overline{r},1} - \|z\|_{\overline{r},1} \le K M_2 \|x - z\|_{\overline{a},2}^{\frac{2-2t}{2-t}}$$

$$= K M_2 \alpha^{\frac{2-2t}{2-t}} \left(\sum_{j \in \Lambda} \overline{a}_j^{-2} \overline{r}_j^2 \mathbb{1}_{\{\overline{a}_j^{-2} \overline{r}_j \alpha < |x_j|\}} \right)^{\frac{1-t}{2-t}}.$$

Rearranging terms yields

$$\sum_{j \in \Lambda} \overline{a}_j^{-2} \overline{r}_j^2 \mathbb{1}_{\{\overline{a}_j^{-2} \overline{r}_j \alpha < |x_j|\}} \le (K M_2)^{2-t} \alpha^{-t}.$$

Hence, $\|x\|_{k_t} \le (K M_2)^{\frac{2-t}{t}}$.
In summary we have equivalence of the following statements:

(i) There exists a constant $K_{vsc} \ge 0$ such that (3.53) holds true with $\varphi(\tau) = K_{vsc} \tau^{\frac{1-t}{2-t}}$.
(ii) There exists a constant $K \ge 0$ such that (3.54) holds true.
(iii) $x \in k_t$.

This above equivalence is a success message: k_t is the maximal set of elements satisfying a variational source condition with a given Hölder rate. Note the analogy of the statement to Remark 3.21 where we characterized variational source conditions for Besov $b^r_{2,1}$-penalization. Therefore, we are on the right track to describe the convergence rates for penalties given by $b^r_{1,1}$-norms more accurately.

The following convergence rates result for the regularization scheme $S_\alpha(g)$ given in (3.50) is a consequence of Proposition 3.56.

Theorem 3.59 (Error bounds for weighted ℓ^1-penalties). *Suppose Assumption 2.27. Let $t \in (0,1)$, $p \in (1,2)$ and $\varrho, \delta > 0$. Suppose $x \in k_t \cap \operatorname{dom}(G)$ with $\|x\|_{k_t} \le \varrho$.*

(a) (Bias bounds) There exists a constant C_b depending only on t and M_1 such that such that for all $\alpha > 0$ and $x_\alpha \in S_\alpha(G(x))$ (see (3.50)) the following bounds hold true:

$$\|x - x_\alpha\|_{\overline{r},1} \le C_b \varrho^t \alpha^{1-t}, \tag{3.55}$$

$$\|x - x_\alpha\|_{\overline{a},2} \le C_b \varrho^{\frac{t}{2}} \alpha^{\frac{2-t}{2}} \quad and \tag{3.56}$$

$$\|x - x_\alpha\|_{\overline{\omega}_p,p} \le C_b \varrho^{\frac{t}{p}} \alpha^{\frac{p-t}{p}}. \tag{3.57}$$

(b) (Convergence rates) Let $0 < c_l \le c_r$ and $1 \le c_D \le C_D$. Suppose $\delta > 0$, $g^{\mathrm{obs}} \in \mathbb{Y}$ with $\|g^{\mathrm{obs}} - F(f)\|_{\mathbb{Y}} \le \delta$. Let $\alpha > 0$ and $\hat{x}_\alpha \in S_\alpha(g^{\mathrm{obs}})$. There is a constant C_r independent of $x, g^{\mathrm{obs}}, \varrho$ and δ such that either of the conditions

$$c_l \varrho^{-\frac{t}{2-t}} \delta^{\frac{2}{2-t}} \le \alpha \le c_r \varrho^{-\frac{t}{2-t}} \delta^{\frac{2}{2-t}} \quad and \quad c_D \delta \le \|g^{\mathrm{obs}} - G(\hat{x}_\alpha)\|_{\mathbb{Y}} \le C_D \delta$$

implies the following bounds

$$\|x - \hat{x}_\alpha\|_{\overline{r},1} \le C_r \varrho^{\frac{t}{2-t}} \delta^{\frac{2-2t}{2-t}}, \tag{3.58}$$

$$\|x - \hat{x}_\alpha\|_{\overline{a},2} \le C_r \delta \quad and \tag{3.59}$$

$$\|x - \hat{x}_\alpha\|_{\overline{\omega}_p,p} \le C_r \varrho^{\frac{t}{p} \frac{2-p}{2-t}} \delta^{\frac{2}{p} \frac{p-t}{2-t}}. \tag{3.60}$$

Proof. In analogy to the proof of Theorem 3.37: The first two bounds in (*i*) and (*ii*) follow from Corollary 3.2, Proposition 3.56 and Assumption 2.27. The interpolation inequality $\|\cdot\|_{\overline{\omega}_p,p} \leq \|\cdot\|_{\overline{a},2}^{\frac{2p-2}{p}} \|\cdot\|_{\overline{r},1}^{\frac{2-p}{p}}$ (see Proposition 3.48) provides the third bound in (*i*) and (*ii*). $\square$

Remark 3.60 (Limit $t \to 1$). *Let us consider the limiting case $t = 1$ by assuming only $x \in \ell_{\frac{1}{r}}^1 \cap D$. Then as in Remark 3.23 we see that the parameter choice $\alpha \sim \delta^2$ as well the discrepancy principle as in Theorem 3.59(b) lead to bounds $\|x - \hat{x}_\alpha\|_{\overline{r},1} \leq C\|x\|_{\overline{r},1}$ and $\|F(x) - F(\hat{x}_\alpha)\|_{\mathbb{Y}} \leq C\delta$. Assumption 2.27 allows to transfer to a bound $\|x - \hat{x}_\alpha\|_{\overline{a},2} \leq \tilde{C}\delta$. Interpolating as in the proof of Theorem 3.59 yields*

$$\|x - \hat{x}_\alpha\|_{\overline{\omega}_p,p} \leq \tilde{C}\|x\|_{\overline{r},1}^{\frac{2-p}{p}} \delta^{\frac{2p-2}{p}}.$$

Remark 3.61 (Limit $t \to 0$). *Note that in the limit $t \to 0$ the convergence rates get arbitrarily close to the linear convergence rate. As already mentioned in Section 2.2, the formal limiting rate for $t \to 0$, i.e. a linear convergence rate in δ occurs if and only if x is sparse as shown by different methods in [53].*

We finish this subsection by showing that the rates in Theorem 3.59 are optimal in the sense of Section 2.2.

Proposition 3.62 (Optimality). *Suppose that Assumption 2.27 holds true. Assume furthermore that there are $c_0 > 0$, $q \in (0,1)$ such that for every $\eta \in (0, c_0]$ there is $j \in \Lambda$ satisfying $q\eta \leq \overline{a}_j \overline{r}_j^{-1} \leq \eta$. Let $p \in [1,2]$, $t \in (0,p)$ and $\rho > 0$. Suppose $\mathrm{dom}(G)$ contains all $x \in k_t$ with $\|x\|_{k_t} \leq \varrho$. Consider an arbitrary reconstruction method described by a mapping $R : (0,\infty) \times \mathbb{Y} \to \ell_{\overline{\omega}_p}^p$. Then the worst case error under the a priori information $\|x\|_{k_t} \leq \varrho$ is bounded below by*

$$\sup\left\{ \|x - R(\delta, g^{\mathrm{obs}})\|_{\overline{\omega}_p,p} : \|x\|_{k_t} \leq \varrho, \ g^{\mathrm{obs}} \in \mathbb{Y} \text{ with } \|F(x) - g^{\mathrm{obs}}\|_{\mathbb{Y}} \leq \delta \right\} \geq c\varrho^{\frac{t}{p}\frac{2-p}{2-t}} \delta^{\frac{2}{p}\frac{p-t}{2-t}}$$

for all $\delta \leq M_2 \varrho c_0^{\frac{2-t}{t}}$ with $c = q^{\frac{2p-2t}{pt}} M_2^{\frac{2}{p}\frac{t-p}{2-t}}$.

Proof. By Proposition 2.6 the worst case error is bounded below by $\frac{1}{2}\omega(2\delta, \varrho)$ with the modulus of continuity

$$\omega(\delta, \varrho) = \sup\left\{ \|z_1 - z_2\|_{\overline{\omega}_p,p} : \|z_1\|_{k_t}, \|z_2\|_{k_t} \leq \varrho, \ \|F(z_1) - F(z_2)\|_{\mathbb{Y}} \leq \delta \right\}.$$

Using Assumption 2.27 we obtain

$$\omega(\delta, \varrho) \geq \sup\{\|z\|_{\overline{\omega}_p,p} : \|z\|_{k_t} \leq \rho, \|z\|_{\overline{a},2} \leq M_2^{-1}\delta\}.$$

By assumption there exists $j_0 \in \Lambda$ such that

$$q\left(M_2^{-1}\delta\varrho^{-1}\right)^{\frac{t}{2-t}} \leq \overline{a}_{j_0}\overline{r}_{j_0}^{-1} \leq \left(M_2^{-1}\delta\varrho^{-1}\right)^{\frac{t}{2-t}}.$$

Choosing $z_{j_0} = \varrho a_{j_0}^{\frac{2-2t}{t}} r_{j_0}^{\frac{t-2}{t}}$ and $z_j = 0$ if $j \neq j_0$ we obtain $\|z\|_{k_t} = \varrho$ and $\|z\|_{\overline{a},2} \leq M_2^{-1}\delta$ and estimate

$$\|z\|_{\overline{\omega}_p,p} = \varrho\left(\overline{a}_{j_0}\overline{r}_{j_0}^{-1}\right)^{\frac{2p-2t}{pt}} \geq q^{\frac{2p-2t}{pt}} M_2^{\frac{2}{p}\frac{t-p}{2-t}} \varrho^{\frac{t}{p}\frac{2-p}{2-t}} \delta^{\frac{2}{p}\frac{p-t}{2-t}}. \qquad \square$$

Note that for $\Lambda = \mathbb{N}$ the additional assumption in Proposition 3.62 is satisfied if $\overline{a}_j\overline{r}_j^{-1} \sim \tilde{q}^j$ for $\tilde{q} \in (0,1)$ or if $\overline{a}_j\overline{r}_j^{-1} \sim j^{-\kappa}$ for $\kappa > 0$, but violated if $\overline{a}_j\overline{r}_j^{-1} \sim \exp(-j^2)$.

Wavelet regularization with Besov $r, 1, 1$-penalties

In this section we apply the results of the last subsections to obtain convergence rates for wavelet regularization with Besov $r, 1, 1$-norm penalties. For fixed $r \geq 0$ we consider (3.16) with $p = 1$, i.e. for an observation $g \in \mathbb{Y}$ we analyze

$$R_\alpha(g) = \mathcal{S}\hat{x}_\alpha \quad \text{with} \quad \hat{x}_\alpha \in \operatorname*{argmin}_{x \in D} \left(\frac{1}{2\alpha} \|g - F(\mathcal{S}x)\|_\mathbb{Y}^2 + \|x\|_{r,1,1} \right) \tag{3.61}$$

with $D := \mathcal{S}^{-1}(B_{1,1}^r \cap \operatorname{dom}(F)) \subseteq b_{1,1}^r$.

Let $a > 0$ and Λ as in Section 2.4 we set $\overline{a}_{(j,k)} := 2^{-ja}$ and $\overline{r}_{(j,k)} := 2^{j(r-\frac{d}{2})}$. Then we have $b_{1,1}^r = \ell_{\overline{r}}^1$ and $b_{2,2}^{-a} = \ell_{\overline{a}}^2$ with equal norms. For $s \in [-a, \infty)$ we set

$$t_s := \frac{2a + 2r}{s + 2a + r}. \tag{3.62}$$

Note that $t_s \in (0, 2]$ and for $s > r$ we have $t_s < 1$. Moreover, we obtain $b_{t_s,t_s}^s = \ell_{\overline{\omega}_{t_s}}^{t_s}$ isometrically for $\overline{\omega}_{t_s}$ given by (3.51). The next proposition introduces function spaces K_s that will replace $B_{1,\infty}^s$ in Theorem 3.15 in the case $p = 1$. Motivated by Remark 3.51 we define K_s in correspondence to k_{t_s} as real interpolation space with fine index ∞.

Proposition 3.63 (Spaces K_s). *Suppose $B_{p,q}^s$ is a Besov scale as in Definition 2.9. Let $a, s > 0$. For $v > s$ the real interpolation space*

$$K_s := \left(B_{2,2}^{-a}, B_{t_v,t_v}^v \right)_{\theta,\infty} \quad \text{with} \quad \theta = \frac{a + s}{a + v}$$

does not depend on the choice of v. We have $B_{t_s,t_s}^s \subset K_s$ continuously.
If the scale $B_{p,q}^s$ is p-monotone, then we have a continuous embedding $K_s \subset B_{t_v,\infty}^s$. Hence,

$$K_s \subset \bigcap_{t < t_s} B_{t,\infty}^s.$$

Proof. Let $-a < u < v$. Then

$$\left(B_{2,2}^{-a}, B_{t_v,t_v}^v \right)_{\xi,t_u} = B_{t_u,t_u}^u \quad \text{with} \quad \xi = \frac{a + u}{a + v}. \tag{3.63}$$

Hence, if $s < u < v$ and $\theta_v = \frac{a+s}{a+v}$ and $\theta_u = \frac{a+s}{a+u}$ then $\theta_u \xi = \theta_v$ and the reiteration theorem for real interpolation (see A.6) yields

$$\left(B_{2,2}^{-a}, B_{t_u,t_u}^u \right)_{\theta_u,\infty} = \left(B_{2,2}^{-a}, B_{t_v,t_v}^v \right)_{\theta_v,\infty}.$$

This shows that K_s is independent of the choice of v.
Using (3.63) with $u = s$ and [8, Thm. 3.4.1.(b)] we obtain the first continuous embedding

$$B_{t_s,t_s}^s = \left(B_{2,2}^{-a}, B_{t_v,t_v}^v \right)_{\theta,t_s} \subset \left(B_{2,2}^{-a}, B_{t_v,t_v}^v \right)_{\theta,\infty} = K_s.$$

For a *p*-monotone scale $t_v \leq \frac{2a+2r}{2a+r} \leq 2$ yields the embeddings $B_{2,2}^{-a} \subset B_{2,\infty}^{-a} \subset B_{t_v,\infty}^{-a}$. Therefore, the interpolation identity $B_{t_v,\infty}^s = (B_{t_v,\infty}^{-a}, B_{t_v,t_v}^v)_{\theta,\infty}$ provides the second embedding using [101, 2.4.1 Rem. 4]. Finally, the last statement follows from $t_v \to t_s$ for $v \searrow s$. $\quad\square$

The following lemma shows that a function belongs to K_s if and only if its wavelet coefficients belong to k_{t_s}. As spaces $b_{p,q}^s$ and $B_{p,q}^s$ with $p < 1$ are involved let us first argue that within the scale B_{t_s,t_s}^s for $s > 0$ the extra condition $\sigma_{t_s} - s_{\max} < s$ in Assumption 2.11 is always satisfied if we assume $a + r \geq \frac{d}{2}$. To this end, let $0 < s < s_{\max}$. Then

$$\sigma_{t_s} = d\left(\frac{1}{t_s} - 1\right) = \frac{d(s-r)}{2a+2r} \leq s - r \leq s < s_{\max}. \tag{3.64}$$

Hence $\sigma_{t_s} - s_{\max} < 0 < s$.

Note that the condition $a + r \geq \frac{d}{2}$ is reasonable as it already appears in Proposition 2.3 for $p = 1$ as a requirement for the existence of minimizers in (3.61).

Lemma 3.64. *Suppose Assumption 2.11 holds true. Let $a > 0$ with $a + r \geq \frac{d}{2}$. If $s \in (0, s_{max})$ then $\mathcal{S}: k_{t_s} \to K_s$ is an isomorphism of quasi-Banach spaces.*

Proof. Let $v \in \mathbb{R}$ with $s < v < s_{\max}$. If $a < s_{\max}$, then we obtain the result by functor properties of real interpolation (see [98, 1.3.3. Rem. 1]) as $k_{t_s} = \left(b_{2,2}^{-a}, b_{t_v,t_v}^v\right)_{\theta,\infty}$ with $\theta = \frac{a+s}{a+v}$ by Remark 3.51 and $\mathcal{S}$ provides an isomorphism $b_{2,2}^{-a} \to B_{2,2}^{-a}$ and $b_{t_v,t_v}^v \to B_{t_v,t_v}^v$, where the latter isomorphism holds since $a + r \geq \frac{d}{2}$ (see (3.64)).

In general (in particular for $a \geq s_{\max}$) we may pick $\tilde{a} > 0$ such that $\tilde{a} \leq a$ and $\tilde{a} \in (-s_{\max}, s_{\max})$. Then (3.63) with $u = -\tilde{a}$ and the reiteration theorem (see [8, Thm. 3.11.5]) yield $K_s = \left(B_{t_{-\tilde{a}},t_{-\tilde{a}}}^{-\tilde{a}}, B_{t_v,t_v}^v\right)_{\eta,\infty}$ for $\eta = \frac{s+\tilde{a}}{v+\tilde{a}}$. As Remark 3.51 provides the corresponding representation of k_{t_s}, we obtain the result by the same argument as above. $\qquad\square$

In case of a smooth enough wavelet basis (i.e. $a < s_{\max}$) we apply Theorem 3.59 to obtain the following convergence rate result for the method R_α given in (3.61).

Theorem 3.65 (Error bounds for $r, 1, 1$-penalties with smooth wavelet systems)**.** *Suppose Assumptions 2.11 and 2.15 hold true with $a < s_{max}$. Let $0 \leq r < s < s_{max}$ such that $a + r \geq \frac{d}{2}$. Suppose $\varrho > 0$ and $f \in K_s \cap \mathrm{dom}(F)$ (see Proposition 3.63) with $\|f\|_{K_s} \leq \varrho$. Let $\overline{p} := \frac{2a+2r}{2a+r}$.*

(a) (Bias bounds) There exists a constant C_b independent of f and ϱ such that such that for all $\alpha > 0$ and $f_\alpha \in R_\alpha(F(f))$ (see (3.61)) the following bounds hold true:

$$\|f - f_\alpha \mid B_{1,1}^r\| \leq C_b \varrho^{\frac{2a+2r}{s+2a+r}} \alpha^{\frac{s-r}{s+2a+r}}, \tag{3.65}$$

$$\|f - f_\alpha \mid B_{2,2}^{-a}\| \leq C_b \varrho^{\frac{a+r}{s+2a+r}} \alpha^{\frac{s+a}{s+2a+r}} \quad and \tag{3.66}$$

$$\|f - f_\alpha \mid B_{\overline{p},\overline{p}}^0\| \leq C_b \varrho^{\frac{2a+r}{s+2a+r}} \alpha^{\frac{s}{s+2a+r}}. \tag{3.67}$$

(b) (Convergence rates) Let $0 < c_l \leq c_r$ and $1 \leq c_D \leq C_D$. Suppose $\delta > 0$, $g^{\mathrm{obs}} \in \mathbb{Y}$ with $\|g^{\mathrm{obs}} - F(f)\|_{\mathbb{Y}} \leq \delta$. Let $\alpha > 0$ and $\hat{f}_\alpha \in R_\alpha(g^{\mathrm{obs}})$. There is a constant C_r independent of $f, g^{\mathrm{obs}}, \varrho$ and δ such that either of the conditions

$$c_l \varrho^{-\frac{a+r}{s+a}} \delta^{\frac{s+2a+r}{s+a}} \leq \alpha \leq c_r \varrho^{-\frac{a+r}{s+a}} \delta^{\frac{s+2a+r}{s+a}} \quad and \quad c_D \delta \leq \|g^{\mathrm{obs}} - F(\hat{f}_\alpha)\|_{\mathbb{Y}} \leq C_D \delta$$

implies the following bounds

$$\|f - \hat{f}_\alpha \mid B_{1,1}^r\| \leq C_r \varrho^{\frac{a+r}{s+a}} \delta^{\frac{s-r}{s+a}}, \tag{3.68}$$

$$\|f - \hat{f}_\alpha \mid B_{2,2}^{-a}\| \leq C_r \delta \quad and \tag{3.69}$$

$$\|f - \hat{f}_\alpha \mid B_{\overline{p},\overline{p}}^0\| \leq C_r \varrho^{\frac{a}{s+a}} \delta^{\frac{s}{s+a}}. \tag{3.70}$$

Proof. By Remark 2.28 $G = F \circ \mathcal{S}$ satisfies Assumption 2.27. Due to $s > r$ we have $t_s \in (0, 1)$. With Lemma 3.64 we obtain $f = \mathcal{S}x$ with $x \in k_{t_s}$ and there exists a constant c_1 independent of f and ϱ such that $\|x\|_{k_t} \le c_1 \varrho$. For $\hat{f}_\alpha \in R_\alpha(g)$ (with $g = F(f)$ in (i) or $g = g^{\mathrm{obs}}$ in (ii)) there is $\hat{x}_\alpha$ a minimizer in (3.61) with $\mathcal{S}\hat{x}_\alpha = \hat{f}_\alpha$. Using the boundedness of $\mathcal{S}$ the results in Theorem 3.59 are transferred to the claimed error bounds. Here we obtain the third bound in (i) and (ii) as $\overline{p} = t_0$. $\qquad\square$

As in Section 3.2 we use the embedding $B^0_{\overline{p}, \overline{p}} \subset L^{\overline{p}}(\Omega)$ to conclude a bound in the $L^{\overline{p}}$-norm.

Corollary 3.66 (error bound in $L^{\overline{p}}$)**.** *In the setting of Theorem 3.65 suppose that* $B^0_{\overline{p}, \overline{p}} = B^0_{\overline{p}, \overline{p}}(\Omega)$ *with* Ω *either* $\mathbb{T}^d, \mathbb{R}^d$ *or a bounded Lipschitz domain in* $\mathbb{R}^d$*. Then we have the upper bounds*

$$\|f - f_\alpha\|_{L^{\overline{p}}(\Omega)} \le C_{\overline{p}} \varrho^{\frac{2a+r}{s+2a+r}} \alpha^{\frac{s}{s+2a+r}} \quad and$$
$$\|f - \hat{f}_\alpha\|_{L^{\overline{p}}(\Omega)} \le C_{\overline{p}} \varrho^{\frac{a}{s+a}} \delta^{\frac{s}{s+a}}$$

with a constant $C_{\overline{p}}$ *independent of* f, ϱ *and* δ*.*

Proof. See the proof of Corollary 3.16. $\qquad\square$

Example 3.67 (Functions with kinks or jumps)**.** *Let us revisit functions* $f^{\mathrm{jump}}, f^{\mathrm{kink}}$: $[0, 1] \to \mathbb{R}$ *which are piecewise smooth,* f^{jump} *has a finite number of jumps, and* f^{kink} *has a finite number of kinks (see Example 3.17). Let* $r \ge 0$ *and* $a = 2$*. Then we have*

$$f^{\mathrm{jump}} \in K_s \quad for \ s < \frac{4+r}{3+2r} \quad and \quad f^{\mathrm{kink}} \in K_s \quad for \ s < \frac{8+3r}{3+2r}. \tag{3.71}$$

To see this, note that $s < \frac{4+r}{3+2r}$ *is equivalent to* $s < \frac{1}{t_s}$ *with* t_s *defined in (3.62). Hence Example 3.17 implies* $f_{jump} \in B^s_{t_s, t_s}$ *which yields the statement for* f_{jump} *by the embedding* $B^s_{t_s, t_s} \subset K_s$ *shown in Proposition 3.63. The same consideration leads to the statement for the function* f^{kink}*. Hence, in the case* $r = 0$ *Corollary 3.66 predicts the rate* $\mathcal{O}(\delta^e)$ *for all* $e < \frac{2}{5}$ *measured in the* L^1*-norm for* f_{jump} *and* $\mathcal{O}(\delta^e)$ *for all* $e < \frac{4}{7}$ *for* f_{kink}*.*
Therefore, describing the regularity of f^{jump} *or* f^{kink} *in the scale* K_s *as in Theorem 3.65 allows for a larger value of* s *and hence a faster convergence rate than describing the regularity of these functions in the Besov spaces* $B^s_{1,\infty}$ *as in Theorem 3.15. Recall that* $s = \frac{1}{2}$ *is maximal with* $f^{\mathrm{jump}} \in B^s_{1,\infty}$ *and* $s = \frac{3}{2}$ *with* $f^{\mathrm{jump}} \in B^s_{1,\infty}$*. In other words, the previous analysis provided only suboptimal rates of convergence for this important class of functions in the case* $p = 1$*.*
Figures 3.9 and 3.10 show that this can also be observed numerically. Here we see the same error diagrams as in Section 3.2 for $p = 1$ *using Daubechies wavelets of order* 7*. Once more we see that the old error bounds are too pessimistic. The new error bounds obtained from Corollary 3.66 are in line with the numerical experiment with a remarkable precision.*

Remark 3.68 (Order optimality)**.** *The bounds in Theorem 3.65 and Corollary 3.66 are order optimal in the sense of Section 2.2. This can be seen either with Proposition 3.62 and the identifications below (3.62) or directly from Theorem 3.45 using the embedding* $B^s_{t_s, t_s} \subset K_s$ *(see Proposition 3.63).*

FIGURE 3.9: *Left:* true coefficient c^{jump} with jumps in (3.40) together with a reconstruction for $b_{1,1}^0$-*db*7-penalization at noise level $\delta = 2.5 \cdot 10^{-5}$. *Right:* Reconstruction error using $b_{1,1}^0$-penalization, the rate $\mathcal{O}(\delta^{2/5})$ predicted by Corollary 3.66 (see Example 3.67), and the rate $\mathcal{O}(\delta^{1/3})$ from the previous analysis in Section 3.2.

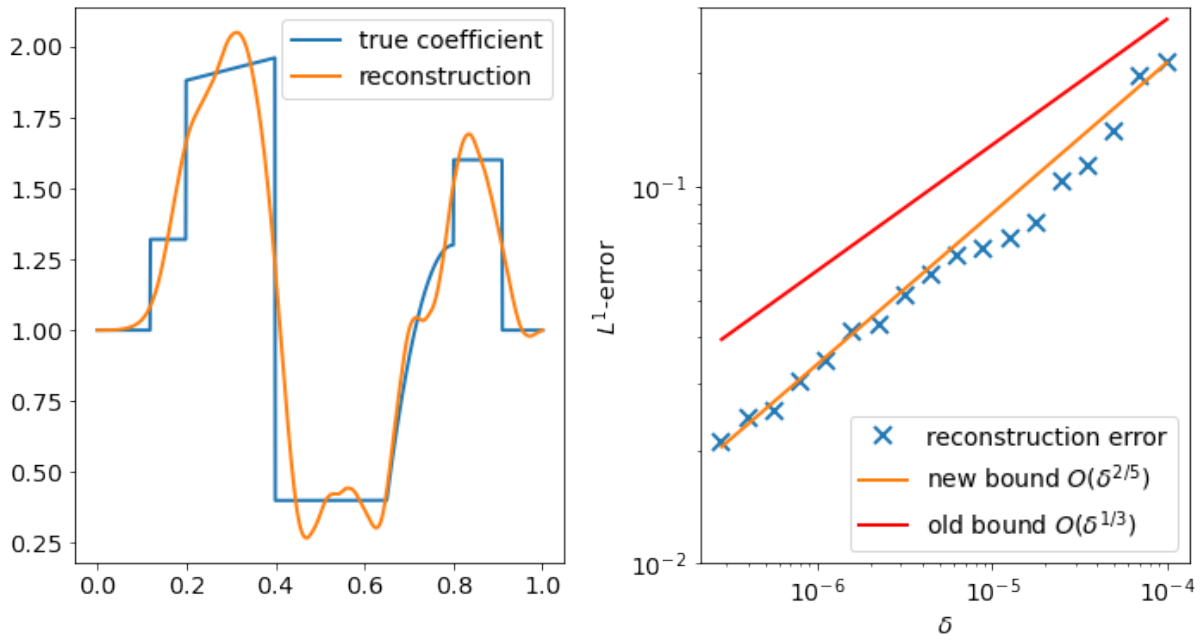FIGURE 3.10: *Left:* true coefficient c^{kink} with kinks in (3.40) together with a reconstruction for $b_{1,1}^0$-*db*7-penalization at noise level $\delta = 2.5 \cdot 10^{-5}$. *Right:* Reconstruction error using $b_{1,1}^0$-penalization, the rate $\mathcal{O}(\delta^{4/7})$ predicted by Corollary 3.66 (see Example 3.67), and the rate $\mathcal{O}(\delta^{1/2})$ from the previous analysis in Section 3.2.
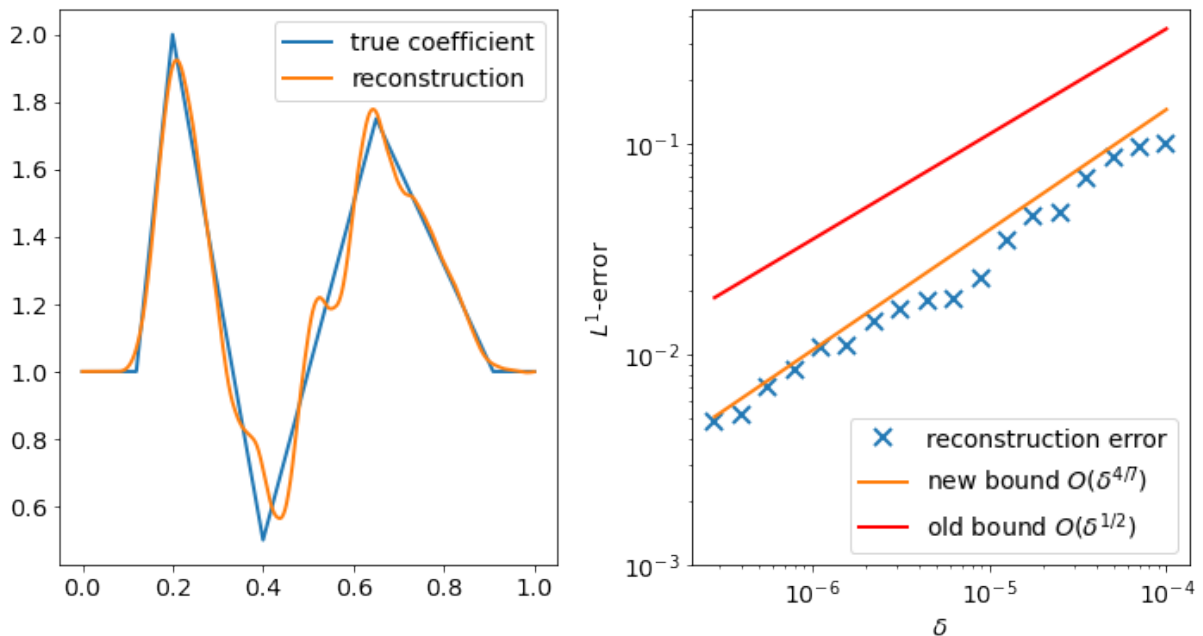
In the sequel we investigate into a generalization of Theorem 3.65 for non-smooth wavelet systems, i.e. we drop the assumption $a < s_{\max}$ to include cases with $a \geq s_{\max}$. Using the same strategy as in Section 3.3 we start with a replacement of the Bernstein type inequality Lemma 3.55.

Lemma 3.69 (Replacement of Bernstein inequality). *Suppose Assumption 2.11. Let $a > 0$ and $0 \leq r < s_{max}$ and $\varrho > 0$. Suppose $s > -a$, $x \in k_{t_s}$ with $\|x\|_{k_{t_s}} \leq \varrho$ and Λ_α and P_α as in Lemma 3.55. Then there is a constant $C_{BR} > 0$ such that*

$$\|P_\alpha z\|_{r,p,1} \leq C_{BR} \varrho^{\frac{a+r}{s+2a+r}} \alpha^{-\frac{a+r}{s+2a+r}} \|\mathcal{S}z \mid B_{2,2}^{-a}\| + \frac{1}{4}\|z\|_{r,1,1} \quad \text{for all } z \in b_{p,1}^r \text{ and } \alpha > 0.$$

Proof. As in the proof of Lemma 3.34 we consider $\tilde{a} := \min\left\{\frac{s_{\max}}{2}, a\right\}$. Assuming $\tilde{a} < a$, the interpolation identity (3.63) with $u = -\tilde{a}$ and $v = r$ implies an interpolation inequality

$$\|f \mid B_{t-\tilde{a},t-\tilde{a}}^{-\tilde{a}}\| \leq c_I \|f \mid B_{2,2}^{-a}\|^{\frac{\tilde{a}+r}{a+r}} \|f \mid B_{1,1}^r\|^{\frac{a-\tilde{a}}{a+r}} \quad \text{for all} \quad f \in B_{1,1}^r$$

with a constant c_I. We may assume $c_I \geq 1$ such that this inequality remains valid for $\tilde{a} = a$.

Let $C_{\tilde{a}}$ and respectively C_r denote the operator norm of $\mathcal{S}^{-1} \colon B_{t-\tilde{a},t-\tilde{a}}^{-\tilde{a}} \to b_{t-\tilde{a},t-\tilde{a}}^{-\tilde{a}}$ and $\mathcal{S} \colon b_{1,1}^r \to B_{1,1}^r$. The above interpolation inequality with $f = \mathcal{S}z$ and Lemma 3.55 with $p = t_{-\tilde{a}}$ yield

$$\begin{aligned}
\|P_\alpha z\|_{r,1,1} &\leq \varrho^{\frac{\tilde{a}+r}{s+2a+r}} \alpha^{-\frac{\tilde{a}+r}{s+2a+r}} \|z\|_{-\tilde{a},t-\tilde{a},t-\tilde{a}} \\
&\leq C_{\tilde{a}} \varrho^{\frac{\tilde{a}+r}{s+2a+r}} \alpha^{-\frac{\tilde{a}+r}{s+2a+r}} \|\mathcal{S}z \mid B_{t-\tilde{a},t-\tilde{a}}^{-\tilde{a}}\| \\
&\leq C_{\tilde{a}} c_I C_r^{\frac{a-\tilde{a}}{a+r}} \varrho^{\frac{\tilde{a}+r}{s+2a+r}} \alpha^{-\frac{\tilde{a}+r}{s+2a+r}} \|\mathcal{S}z \mid B_{2,2}^{-a}\|^{\frac{\tilde{a}+r}{a+r}} \|z\|_{r,1,1}^{\frac{a-\tilde{a}}{a+r}}.
\end{aligned}$$

An application of Young's inequality as in Example 3.30 yields the claim. □

Proposition 3.70 (Variational source condition). *Suppose Assumptions 2.11 and 2.15(a). Let $0 \leq r < s_{max}$ with $a + r \geq \frac{d}{2}$. Moreover, let $r < s < s_{max}$ and $\varrho > 0$. Assume $x \in k_{t_s} \cap \mathcal{S}^{-1}(\mathrm{dom}(F))$ with $\|x\|_{k_{t_s}} \leq \varrho$. Then*

$$\frac{1}{2}\|x - z\|_{r,1,1} + \|x\|_{r,1,1} - \|z\|_{r,1,1} \leq \varphi\left(\|F(\mathcal{S}x) - F(\mathcal{S}z)\|_{\mathbb{Y}}^2\right) \text{ for all } z \in b_{1,1}^r \cap \mathcal{S}^{-1}(\mathrm{dom}(F))$$

with $\varphi(\tau) = C'_\varphi \varrho^{\frac{a+r}{s+a}} \tau^{\frac{s-r}{2s+2a}}$ for a constant C'_φ independent of x and ϱ.

Proof. We use the Proposition 3.29 for $G := F \circ \mathcal{S}$ on $D := b_{p,1}^r \cap \mathcal{S}^{-1}(\mathrm{dom}(F))$ with $J := (0,\infty)$ and $(P_\alpha)_{\alpha \in J}$ as in Lemma 3.55. Then (3.19a) is satisfied. Lemma 3.69 and Assumption 2.15 yield (3.42) with $\nu_\alpha = M_1 C_{BR} \varrho^{\frac{a+r}{s+2a+r}} \alpha^{-\frac{a+r}{s+2a+r}}$ and $\kappa = \frac{1}{4}$.
The same routine as in the proof of Theorem 3.59 yields a bound on $\|(I - P_\alpha)x\|_{r,1,1}$ and an estimation of the infimum involved in the definition of φ in Proposition 3.29. □

The variational source condition in Proposition 3.70 leads to the following result.

Theorem 3.71 (Error bounds for $r, 1, 1$-penalties with non-smooth wavelet systems). *Theorem 3.65 and Corollary 3.66 remain valid without the assumption $a < s_{max}$.*

Proof. With Lemma 3.64 the smoothness assumption on f transfers to the corresponding sequence of wavelet coefficients x with $\mathcal{S}x = f$. Then the claim follows from Proposition 3.70 using the same arguments as in the proof of Theorem 3.15. $\square$

Remark 3.72 (boundedness of Ω not required)**.** *Note that the only place within this section where we require p-monotonicity is for the superset of K_s in Proposition 3.63. Hence the stated convergence rates results hold also true if the domain Ω of the functions in the domain of the forward operator is $\mathbb{R}^d$.*

3.6   Infinitely smoothing operators

This section is devoted to error bounds for infinitely smoothing operators, i.e. operators with range contained in the set of infinitely differentiable functions on some domain. As seen for the backward heat equation in Example 2.24 we cannot expect that Assumption 2.15 holds true for this kind of operators: Recall that the first inequality in Assumption 2.15(a) means that F is at most a-times smoothing. Hence for infinitely smoothing operators there is no $a > 0$ such that the first inequality holds true whereas the second one is satisfied for all $a > 0$.

As a first step we provide a modification of the statement Proposition 3.29 for the verification of variational source conditions for 1 homogeneous wavelet regularization in Besov spaces. The key amendment is that it allows to pick a family of operators $(\mathcal{P}_j)_{j \in J}$ acting on the space $B^r_{p,1}$ in the Besov scale instead of a family of projections $(P_n)_{n \in \mathbb{N}_0}$ acting on the Besov sequence $b^r_{p,1}$ as in the earlier sections.

For example, the gained flexibility allows to choose $\mathcal{P}_j$ as a high cut filter, i.e. a multiplication with a compact supported function in Fourier space. Due to the representation of the forward operator of the backward heat equation as Fourier multiplier this choice allows us to prove a convergence rate theorem for the backward heat equation.

Verification of variational source condition

To state a general scheme for the verification of a variational source condition we need some preliminaries. The first one concerns the subdifferential of ℓ^1-type norms.

Lemma 3.73. *Let $r \geq 0$, $p \in [1, 2]$ and $x \in b^r_{p,1}$. With p' the Hölder conjugate of p we obtain $\xi \in \partial \|\cdot\|_{r,p,1}(x)$ if and only if*

$$\|\xi\|_{-r,p',\infty} \leq 1 \quad and \quad \langle \xi_j, x_j \rangle = 2^{jr} 2^{jd\left(\frac{1}{2} - \frac{1}{p}\right)} \|x_j\|_p \quad for \ all \ j \in \mathbb{N}_0.$$

Here we denote $\langle \xi_j, x_j \rangle = \sum_{k \in \Lambda_j} \xi_{j,k} x_{j,k}$.

Proof. By [94, Prop. 4.6.2] and in view of Proposition A.8 we identify

$$\partial \|\cdot\|_{r,p,1}(x) = \{\xi \in b^{-r}_{p',\infty} \colon \|\xi\|_{-r,p',\infty} \leq 1 \text{ and } \langle \xi, x \rangle = \|x\|_{r,p,1}\}.$$

Hence $\langle \xi, x \rangle = \|x\|_{r,p,1}$ is equivalent to

$$\sum_{j=0}^{\infty} \langle \xi_j, x_j \rangle = \sum_{j=0}^{\infty} 2^{jr} 2^{jd\left(\frac{1}{2} - \frac{1}{p}\right)} \|x_j\|_p. \tag{3.72}$$

The claim follows as Hölder's inequality and $\|\xi\|_{-r,p',\infty} \leq 1$ imply

$$\langle \xi_j, x_j \rangle \leq \|\xi\|_{p'} \|x_j\|_p \leq 2^{jr} 2^{jd\left(\frac{1}{2} - \frac{1}{p}\right)} \|x_j\|_p. \qquad \square$$

The next lemma provides a technical tool to estimate index functions given as an infimum over simple concave functions as in Proposition 3.29.

Lemma 3.74. *Let* $(\nu_n)_{n\in\mathbb{N}_0}, (b_n)_{n\in\mathbb{N}_0}$ *be sequences of of positive real numbers. Suppose* $(\nu_n)_{n\in\mathbb{N}_0}$ *is strictly increasing with* $\lim_{n\to\infty} \nu_n = \infty$ *and* $\inf_{n\in\mathbb{N}_0} b_n = 0$. *Then for every* $k \in \mathbb{N}_0$ *there exists* $\tau_0 > 0$ *such that*

$$\inf_{n\in\mathbb{N}_0} \left(\nu_n \sqrt{\tau} + b_n \right) = \inf_{n\in\mathbb{N}_0} \left(\nu_{n+k} \sqrt{\tau} + b_{n+k} \right) \quad \textit{for all} \quad 0 \leq \tau \leq \tau_0.$$

Proof. Consider the function $\varphi \colon [0,\infty) \to [0,\infty)$ given by $\varphi(\tau) = \inf_{n\in\mathbb{N}_0} (\nu_n \sqrt{\tau} + b_n)$. Let $\tau > 0$. Then we have $\nu_n \sqrt{\tau} + b_n \to \infty$ as $n \to \infty$. Therefore, the infimum in the definition of $\varphi(\tau)$ is attained. Hence we may define $I \colon (0,\infty) \to \mathbb{N}_0$ such that for all $\tau > 0$ the integer $I(\tau)$ is minimal with $\varphi(\tau) = \nu_{I(\tau)} \sqrt{\tau} + b_{I(\tau)}$. Note that the claim follows from $I(\tau) \to \infty$ for $\tau \searrow 0$ which we prove in the following.
At first we show that I is non-increasing. To this end, let $0 < \tau_1 < \tau_2$. Then $I(\tau_1) \geq I(\tau_2)$ follows from

$$\begin{aligned}
\varphi(\tau_2) &\leq \nu_{I(\tau_1)} \sqrt{\tau_2} + b_{I(\tau_1)} \\
&= \nu_{I(\tau_1)} \left(\sqrt{\tau_2} - \sqrt{\tau_1} \right) + \varphi(\tau_1) \\
&< \nu_n \left(\sqrt{\tau_2} - \sqrt{\tau_1} \right) + \varphi(\tau_1) \\
&\leq \nu_n \sqrt{\tau_2} + b_n
\end{aligned}$$

for all $n > I(\tau_1)$.
Moreover, the function φ is concave and upper semi-continuous as it is the pointwise infimum of functions possessing these properties. Lemma A.1 provides continuity of φ. In particular,

$$\lim_{\tau \searrow 0} \varphi(\tau) = \varphi(0) = \inf_{n\in\mathbb{N}_0} b_n = 0.$$

Hence I cannot be bounded as this would imply

$$\lim_{\tau \searrow 0} \varphi(\tau) = \lim_{\tau \searrow 0} \left(\nu_k \sqrt{\tau} + b_k \right) = b_k > 0. \qquad \square$$

The following general scheme requires an orthonormal basis system. Note that if $(\psi_\lambda)_{\lambda\in\Lambda}$ an orthonormal basis of $B_{2,2}^0$ satisfying Assumption 2.11, then $\mathcal{S} \colon \ell^2(\Lambda) = b_{2,2}^0 \to B_{2,2}^0$ is a unitary operator.

Proposition 3.75 (Variational source condition)**.** *Suppose Assumption 2.11 holds true and that* $(\psi_\lambda)_{\lambda\in\Lambda}$ *is an orthonormal basis of* $B_{2,2}^0$. *Let* $F \colon \operatorname{dom}(F) \to \mathbb{Y}$ *be a function with* $\operatorname{dom}(F) \subset \tilde{\mathcal{D}}$ *(see Definition 2.9) and* $\mathbb{Y}$ *a Banach space. Let* $p \in [1,2]$, $0 \leq r < s < s_{max}$, $\varrho > 0$ *and* $x \in b_{p,\infty}^s$ *with* $\|x\|_{s,p,\infty} \leq \varrho$. *Assume* $\left(\mathcal{P}_n \colon B_{p,1}^r \to B_{p,1}^r \right)_{j\in\mathbb{N}_0}$ *is a family of bounded linear operators,* $(\nu_n)_{n\in\mathbb{N}_0}$ *is a strictly increasing sequence of positive real numbers and* $c_J > 0$ *is a constant such that the following conditions are satisfied:*

$$\|\mathcal{P}_n(f_1 - f_2) \,|\, B_{p,1}^r\| \leq \nu_n \|F(f_1) - F(f_2)\|_{\mathbb{Y}} \quad \textit{for all } f_1, f_2 \in \operatorname{dom}(F), n \in \mathbb{N}_0, \quad \text{(3.73a)}$$

$$\|(I - \mathcal{P}_n^*)f \,|\, B_{p',\infty}^{-r}\| \leq c_J 2^{-n(s+r)} \|f \,|\, B_{p',\infty}^s\| \quad \textit{for all } f \in B_{p',\infty}^s, n \in \mathbb{N}_0. \quad \text{(3.73b)}$$

Then the variational source condition

$$\frac{1}{2}\|x - z\|_{r,p,1} + \|x\|_{r,p,1} - \|z\|_{r,p,1} \leq \varphi\left(\|F(\mathcal{S}x) - F(\mathcal{S}z)\|_{\mathbb{Y}}^2\right)$$

for all $z \in b_{p,1}^r \cap \mathcal{S}^{-1}(\mathrm{dom}(F))$ *holds true with a concave and continuous* φ *satisfying*

$$\varphi(\tau) \leq C \inf_{n \in \mathbb{N}_0}\left(\nu_n \sqrt{\tau} + \varrho 2^{n(r-s)}\right) \quad \text{for all } \tau \leq \tau_0 \tag{3.74}$$

for some $\tau_0 > 0$ *and a constant* C *that depends on* r, s, c_J *and operator norms of* $\mathcal{S}$ *only.*

Proof. Let $P_n\colon b_{p,1}^r \to b_{p,1}^r$ be the projection onto the first $n + 1$ levels as defined in Section 3.2. At first we use (3.73b) to show that

$$\|P_n z\|_{r,p,1} \leq c_1 \|\mathcal{P}_{n+k}\mathcal{S}z \mid B_{p,1}^r\| + \frac{1}{4}\|z\|_{r,p,1} \quad \text{for all} \quad z \in b_{p,1}^r \text{ and } n \in \mathbb{N}_0 \tag{3.75}$$

with a constant c_1 specified below and k the smallest integer such that $c_2 c_3 c_J 2^{-k(s+r)} \leq \frac{1}{4}$ where c_2 denotes the operator norm of $\mathcal{S}\colon b_{p,1}^r \to B_{p,1}^r$ and c_3 the one of $\mathcal{S}\colon b_{p',\infty}^s \to B_{p',\infty}^s$. To this end, let $z \in b_{p,1}^r$ and $n \in \mathbb{N}_0$. Let $\xi_n \in \partial\|\cdot\|_{r,p,1}(P_n z)$. By Lemma 3.73 we may assume $(\xi_n)_j = 0$ for all $j > n$. Therefore, we obtain

$$\|P_n z\|_{r,p,1} = \langle \xi_n, P_n z \rangle = \langle \xi_n, z \rangle = \langle \mathcal{S}\xi_n, \mathcal{S}z \rangle = \langle \mathcal{S}\xi_n, \mathcal{P}_{n+k}\mathcal{S}z \rangle + \langle (I - \mathcal{P}_{n+k}^*)\mathcal{S}\xi_n, \mathcal{S}z \rangle. \tag{3.76}$$

Here the third identity needs a further justification: Since $r < s_{\max}$ the bilinear map $b_{p',\infty}^{-r} \times b_{p,1}^r \to$ given by $(\xi, z) \mapsto \langle \mathcal{S}\xi, \mathcal{S}x \rangle$ is continuous (see Definition 2.9 (b)). Moreover, since $\mathcal{S}$ is a unitary operator $b_{2,2}^0 \to B_{2,2}^0$ the latter bilinear map agrees with $(\xi, z) \mapsto \langle \xi, z \rangle$ on the set of finitely supported sequences. Therefore, the identity holds since $z \in b_{p,1}^r$ can be approximated by finitely supported sequences.

In the following we separately estimate the two summands on the right hand side of (3.76). Lemma 3.73 yields $\|\mathcal{S}\xi_n \mid B_{p',\infty}^{-r}\| \leq c_4$ with c_4 the operator norm of $\mathcal{S}\colon b_{p',\infty}^{-r} \to B_{p',\infty}^{-r}$. By 2.9(b) there is a constant c_5 such that

$$\langle \mathcal{S}\xi_n, \mathcal{P}_{n+k}\mathcal{S}z \rangle \leq c_5 \|\mathcal{S}\xi_n \mid B_{p',\infty}^{-r}\| \cdot \|\mathcal{P}_{n+k}\mathcal{S}z \mid B_{p,1}^r\| \leq c_4 c_5 \|\mathcal{P}_{n+k}\mathcal{S}z \mid B_{p,1}^r\|.$$

For the second summand we first observe that by the definition of the norms in Besov sequence spaces and once again by Lemma 3.73 we obtain

$$\|\xi_n\|_{s,p',\infty} \leq 2^{n(s+r)}\|\xi_n\|_{-r,p',\infty} \leq 2^{n(s+r)}.$$

Making use of (3.73b) we estimate

$$\begin{aligned}
\langle (I - \mathcal{P}_{n+k}^*)\mathcal{S}\xi_n, \mathcal{S}z \rangle &\leq \|(I - \mathcal{P}_{n+k}^*)\mathcal{S}\xi_n \mid B_{p',\infty}^{-r}\| \cdot \|\mathcal{S}z \mid B_{p,1}^r\| \\
&\leq c_J 2^{-(n+k)(s+r)}\|\mathcal{S}\xi_n \mid B_{p',\infty}^s\| \cdot \|\mathcal{S}z \mid B_{p,1}^r\| \\
&\leq c_2 c_3 c_J 2^{-(n+k)(s+r)}\|\xi_n\|_{s,p',\infty} \cdot \|z\|_{r,p,1} \\
&\leq c_2 c_3 c_J 2^{-k(s+r)}\|z\|_{r,p,1} \leq \frac{1}{4}\|z\|_{r,p,1}.
\end{aligned}$$

Inserting into (3.76) proves (3.75).

Now let $z_1, z_2 \in b_{p,1}^r$. Inequality (3.75) for $z = z_1 - z_2$ together with (3.73a) for $f_i = \mathcal{S}z_i$ yield

$$\|P_n(z_1 - z_2)\|_{r,p,1} \leq c_1 \nu_{n+k}\|F(\mathcal{S}z_1) - F(\mathcal{S}z_2)\|_{\mathbb{Y}} + \frac{1}{4}\|z_1 - z_2\|_{r,p,1} \quad \text{for all } n \in \mathbb{N}_0.$$

Moreover, Lemma 3.10 provides $\|(I - P_n)\,x\|_{r,p,1} \leq c_4 2^{n(r-s)}\varrho$ with $c_4 > 0$ depending only on r and s. Therefore, Proposition 3.29 yields the stated variational source condition with

$$\varphi(\tau) = 2 \inf_{n \in \mathbb{N}_0} \left(c_1 \nu_{n+k}\sqrt{\tau} + c_4 2^{n(r-s)}\varrho\right).$$

Finally with $c_5 = 2 \max\{c_1, c_4 2^{k(s-r)}\}$ the estimation

$$\varphi(\tau) \leq c_5 \inf_{n \in \mathbb{N}_0} \left(\nu_{n+k}\sqrt{\tau} + 2^{(n+k)(r-s)}\varrho\right)$$

together with Lemma 3.74 implies (3.74). □

Backward heat equation

Now we turn to the periodic heat equation as described in Example 2.24. Recall $A_{BH}: \mathcal{D}'(\mathbb{T}^d) \to L^2(\mathbb{T}^d)$ with

$$A_{BH}f = \sum_{k \in \mathbb{Z}^d} e^{-\bar{t}\|k\|_2^2}\hat{f}(k)e_k$$

for some fixed $\bar{t} > 0$. For an observation $g \in L^2(\mathbb{T}^d)$ we consider the regularization scheme

$$R_\alpha(g) = \mathcal{S}\hat{x}_\alpha \quad \text{with} \quad \hat{x}_\alpha \in \operatorname*{argmin}_{x \in b_{p,1}^r} \left(\frac{1}{2\alpha}\|g - A_{BH}\mathcal{S}x\|_{L^2(\mathbb{T}^d)}^2 + \|x\|_{0,2,1}\right). \tag{3.77}$$

Here we restrict ourselves to $r = 0$ and $p = 2$ for simplicity. Optimal convergence rates for Besov space regularization with $p \in (1, 2]$ and $q \geq 2$ can be found in [104, Thm. 5.1]. In Example 2.24 we showed that A_{BH} maps $B_{2,2}^{-a}(\mathbb{T}^d)$ boundedly to $L^2(\mathbb{T}^d)$ for every $a > 0$. Hence we obtain existence of minimizers in (3.77) by Proposition 2.16.

Proposition 3.76 (Variational source condition)**.** *Suppose Assumption 2.11 holds true with an orthonormal wavelet system on $\mathbb{T}^d$. Let $0 < s < s_{max}$ and assume $x \in b_{2,\infty}^s$ with $\|x\|_{s,2,\infty} \leq \varrho$. Then*

$$\frac{1}{2}\|x - z\|_{0,2,1} + \|x\|_{0,2,1} - \|z\|_{0,2,1} \leq \varphi\left(\|A_{BH}\mathcal{S}(x - z)\|_{L^2(\mathbb{T}^d)}^2\right) \quad \text{for all } z \in b_{2,1}^0$$

holds true with a concave and continuous $\varphi: [0, \infty) \to [0, \infty)$ satisfying

$$\varphi(\tau) \leq C\varrho \ln\left(\frac{\varrho}{\sqrt{\tau}}\right)^{-\frac{s}{2}} \quad \text{for all } \tau \leq \tau_0$$

for a constant C that does not depend on f and τ and some $\tau_0 > 0$.

Proof. Here we always use the norms $\|\cdot \mid B_{p,q}^s(\mathbb{T}^d)\|$ given by (2.11). For $n \in \mathbb{N}_0$ and $f \in \mathcal{D}'$ we define $\mathcal{P}_n f := \sum_{j=0}^n \sum_{k \in \Pi_j} \hat{f}(k)e_k$ (see Section 2.4 for clarification of the notations). Then $\mathcal{P}_n$ is self adjoint, we have $\mathcal{P}_n B_{2,1}^0(\mathbb{T}^d) \subset B_{2,1}^0(\mathbb{T}^d)$ and

$$\|(I - \mathcal{P}_n)f \mid B_{2,\infty}^0(\mathbb{T}^d)\| = \sup_{j>n} \left\| \sum_{k \in \Pi_j} \hat{f}(k)e_k \right\|_{L^2(\mathbb{T}^d)} \leq 2^{-(n+1)s} \|f \mid B_{2,\infty}^s(\mathbb{T}^d)\|.$$

Hence (3.73b) is satisfied with $c_J = 2^{-s}$.
In the following we verify (3.73a). First note that we have $\|k\|_2 \leq \sqrt{d}\|k\|_\infty < \sqrt{d}2^j$. With $a = \bar{t}d$ we obtain

$$\sum_{k \in \Pi_j} |\hat{f}(k)|^2 \leq e^{2a4^j} \sum_{k \in \Pi_j} e^{-2\bar{t}\|k\|_2^2} |\hat{f}(k)|^2 \quad \text{for all } j \in \mathbb{N}_0.$$

We set $A_j(f) := \sum_{k \in \Pi_j} e^{-2\bar{t}\|k\|_2^2} |\hat{f}(k)|^2$. Then $\sum_{j=0}^\infty A_j(f) = \|A_{BH}f\|_{L^2(\Omega)}^2$, and the Cauchy-Schwarz inequality provides

$$\|\mathcal{P}_n f \mid B_{2,1}^0(\mathbb{T}^d)\| = \sum_{j=0}^n \left(\sum_{k \in \Pi_j} |\hat{f}(k)|^2 \right)^{\frac{1}{2}} \leq \sum_{j=0}^n e^{a4^j} A_j(f)^{\frac{1}{2}} \leq \left(\sum_{j=0}^n e^{2a4^j} \right)^{\frac{1}{2}} \|A_{BH}f\|_{L^2(\Omega)}. \tag{3.78}$$

Let $d_j := e^{2a4^j}$. Then for $j \geq 1$ we have $\frac{d_{j-1}}{d_j} = e^{-6a4^{j-1}} \leq e^{-6a} < 1$. Therefore, there is a constant c_1 such that $\sum_{j=0}^n d_j \leq c_1 d_n$ for all $n \in \mathbb{N}_0$. With $c_2 = c_1^{\frac{1}{2}}$ we update

$$\|\mathcal{P}_n f \mid B_{2,1}^0(\mathbb{T}^d)\| \leq c_2 e^{a4^n} \|A_{BH}f\|_{L^2(\Omega)}.$$

This shows that condition (3.73a) is satisfied with $\nu_n := c_2 e^{a4^n}$.
Hence Proposition 3.75 implies the desired variational source condition with a φ satisfying

$$\varphi(\tau) \leq C \inf_{n \in \mathbb{N}_0} \left(e^{a4^n} \sqrt{\tau} + \varrho 2^{-ns} \right) \quad \text{for all } \tau \leq \tau_0.$$

Note that we have absorbed c_2 into C. It remains to estimate the infimum on the right hand side. For τ small enough we choose $4^n \sim \frac{1}{2a} \ln\left(\frac{\varrho}{\sqrt{\tau}}\right) = \frac{1}{a} \ln\left(\left(\frac{\sqrt{\tau}}{\varrho}\right)^{-\frac{1}{2}}\right)$ and obtain a constant c_3 with

$$\varphi(\tau) \leq c_3 \varrho \left(\left(\frac{\sqrt{\tau}}{\varrho}\right)^{\frac{1}{2}} + \ln\left(\frac{\varrho}{\sqrt{\tau}}\right)^{-\frac{s}{2}} \right).$$

Finally, note that for small τ the second summand dominates the first such that we wind up with $\varphi(\tau) = \mathcal{O}\left(\varrho \ln\left(\frac{\varrho}{\sqrt{\tau}}\right)^{-\frac{s}{2}} \right)$ for τ small enough. $\qquad \square$

The verification of (3.73a) in the above proof is an adaption of a similar argument in the proof of [104, Thm. 5.1]. From 3.76 we deduce the following *logarithmic* convergence rate for the inversion method (3.77).

Theorem 3.77 (Convergence rate for the backward heat problem)**.** *Suppose Assumption 2.11 holds true with an orthonormal wavelet system on $\mathbb{T}^d$. Let $0 < s < s_{max}$ and assume $f \in B^s_{2,\infty}$ with $\|f \mid B^s_{2,\infty}\| \leq \varrho$. Moreover, let $\delta > 0$, $1 \leq c_D \leq C_D$ and suppose $g^{\mathrm{obs}} \in L^2(\mathbb{T}^d)$ satisfies $\|g^{\mathrm{obs}} - A_{BH}f\|_{L^2(\mathbb{T}^d)} \leq \delta$. Let $\alpha > 0$ and $\hat{f}_\alpha \in R_\alpha(g^{\mathrm{obs}})$ (see (3.77)). There exists a constant C independent of f, g^{obs}, ϱ and δ and some δ_0 such that for $\delta \leq \delta_0$ the parameter choice $c_D\delta \leq \|g^{\mathrm{obs}} - A_{BH}\|_{L^2(\mathbb{T}^d)} \leq C_D\delta$ implies*

$$\|f - \hat{f}_\alpha\|_{L^2(\mathbb{T}^d)} \leq C\varrho \ln\left(\frac{\varrho}{\delta}\right)^{-\frac{s}{2}}.$$

Proof. As in the proof of Theorem 3.15 the $b^s_{2,\infty}$-norm of the wavelet coefficients x of f is bounded by a multiple of ϱ. Hence x satisfies the variational source condition given in Proposition 3.76. From Proposition 3.1(b) we obtain

$$\|x - \hat{x}_\alpha\|_{0,2,1} \leq \varphi\left((1+c_D)^2\delta^2\right) \leq (1+c_D)^2\varphi\left(\delta^2\right)$$

with $\hat{x}_\alpha$ given by (3.77). Here the second inequality is due to concavity and $\varphi(0) = 0$. Using the synthesis operator and the embedding $B^0_{2,1}(\mathbb{T}^d) \subset B^0_{2,2}(\mathbb{T}^d) = L^2(\mathbb{T}^d)$ (see Proposition A.11) we obtain the error bound with respect to the L^2-norm. $\qquad\square$

Remark 3.78 (Optimality)**.** *In [104, Thm. 5.3] it is shown that the above convergence rate is order optimal in the sense of Section 2.2.*

Remark 3.79 (A priori choice rule)**.** *Optimizing α in the right hand side of the error bound in Proposition 3.1(a) one can show that there always exists an a priori choice rule such that one obtains an error bound $L(x, \hat{x}_\alpha) = \mathcal{O}(\varphi(\delta^2))$ (see [67, Thm. 2.3]). Hence the variational source condition in Proposition 3.76 implies the same convergence rate also for some a priori rule.*

3.7 Error bounds for statistical noise models

In this section we consider a statistical model: We assume that the the forward operator $F : \mathrm{dom}(F) \to \mathbb{Y}$ with $\mathbb{Y} = L^2(\Omega_{\mathbb{Y}})$ for $\Omega_{\mathbb{Y}}$ either the d−dimensional torus or a bounded Lipschitz domain in $\mathbb{R}^d$. We will write $\|\cdot\|_{L^2}$ for the norm in $\mathbb{Y}$ and $\|\cdot \mid B^s_{p,q}\|$ for Besov norms on $\Omega_{\mathbb{Y}}$. The scale $B^s_{p,q}(\Omega_{\mathbb{Y}})$ satisfies all properties in Definition 2.9 including p-monotonicity.

For some unknown $f \in \mathrm{dom}(F)$ we model the data by

$$g^{\mathrm{obs}} = F(f) + \varepsilon Z \tag{3.79}$$

with statistical noise level $\varepsilon > 0$ and some noise process Z on $L^2(\Omega)$ with white noise as prominent example.

This section is structured as follows. In the first part we show existence of minimizers and statistical bounds on the reconstruction error under variational source conditions. In the second part we use the variational source conditions verified in the last sections to conclude statistical error bounds for a-times smoothing operators. Finally, we show a statistical convergence rate for the backward heat equation.

We will make the following assumption on the noise process.

Assumption 3.80. *We assume that Z is a random variable with*

$$Z \in B_{u,\infty}^{-d/2}(\Omega_{\mathbb{Y}}) \quad \text{almost surely}$$

for all $u \in [1, \infty)$.

This assumption is satisfied for Gaussian white noise on $\Omega_{\mathbb{Y}} = \mathbb{T}^d$ by [103, Thm. 3.4]. For Gaussian white noise and the Brownian bridge process on $\Omega_{\mathbb{Y}} = [0, 1]$ the assumption is satisfied by [47, Thm. 4.4.3].

If the Besov norm of the noise additionally satisfies a a *large deviation inequality* of the form

$$\mathbb{P}\left[\|Z \mid B_{u,\infty}^{-d/2}\| > M_Z + \tau\right] \leq \exp(-C_Z \tau^\mu) \quad \text{for all } \tau > 0. \tag{3.80}$$

with constants $C_Z, M_Z, \mu > 0$ we will show error bounds on the expectation of arbitrary moments of the error. By [103, Cor. 3.7] such an inequality is satisfied in particular for Gaussian white noise on $\Omega_{\mathbb{Y}} = \mathbb{T}^d$.

Now suppose Assumption 2.11 holds true and $\mathrm{dom}(F) \subset \mathcal{D}$ (see Definition 2.9). As for example white noise does not belong to $L^2(\Omega_{\mathbb{Y}})$ with probability 1 we expand the square in (3.16) and omit the term $\frac{1}{2}\|g\|_{L^2}^2$ which has no influence on the minimizer. This yields

$$R_\alpha(g^{\mathrm{obs}}) = \mathcal{S}\hat{x}_\alpha \quad \text{with} \quad \hat{x}_\alpha \in \underset{x \in D}{\mathrm{argmin}} \left(\frac{1}{2\alpha} \|F(\mathcal{S}x)\|_{L^2}^2 - \frac{1}{\alpha} \langle g^{\mathrm{obs}}, F(\mathcal{S}x) \rangle + \|x\|_{r,p,1} \right) \tag{3.81}$$

with $D := \mathcal{S}^{-1}(B_{p,1}^r \cap \mathrm{dom}(F)) \subseteq b_{p,1}^r$.

Existence of minimizers and error bounds

Following the lines of [104, Prop. 6.3], we prove existence of minimizers in (3.81). In the sequel we will frequently use inequalities

$$\langle g_1, g_2 \rangle \leq c_t \|g_1 \mid B_{t',\infty}^{-d/2}\| \cdot \|g_2 \mid B_{t,1}^{d/2}\| \tag{3.82}$$

for t' the Hölder conjugate of $t \in (1, 2]$, without further mentioning.

Proposition 3.81 (Existence of minimizers)**.** *Suppose Assumptions 3.80, 2.11 and 2.15(b) hold true and that g^{obs} is given by (3.79). Let $r \geq 0$ and $p \in [1, 2]$ with $B_{p,1}^r \cap \mathrm{dom}(F) \neq \emptyset$ and $a + r > \frac{d}{2}$. Moreover, assume that there exists $t \in (1, 2]$, $\theta \in (0, 1)$ and $C_I > 0$ such that the following interpolation type inequality*

$$\|F(f_1) - F(f_2) \mid B_{t,1}^{d/2}\| \leq C_I \|F(f_1) - F(f_2)\|_{L^2}^{1-\theta} \|f_1 - f_2 \mid B_{p,1}^r\|^\theta \tag{3.83}$$

holds true for all $f_1, f_2 \in B_{p,1}^r \cap \mathrm{dom}(F)$. Finally, suppose that there exists $\overline{f} \in B_{p,1}^r \cap \mathrm{dom}(F)$ such that $F(\overline{f}) \in B_{t,1}^{d/2}$. Then $R_\alpha(g^{\mathrm{obs}}) \neq \emptyset$ (see (3.81)) almost surely.

Proof. We set $N := c_t \|g^{\mathrm{obs}} \mid B_{t',\infty}^{-d/2}\|$ and $A := c_t N \|F(\overline{f}) \mid B_{t,1}^{d/2}\|$. Then $A, N < \infty$ almost surely by Assumption 3.80 and (3.82). Let $h \in B_{p,1}^s \cap \mathrm{dom}(F)$. We use Young's inequality

$\beta\gamma \le \frac{1}{4}\beta^{\frac{2}{1-\theta}} + c\gamma^{\frac{2}{1+\theta}}$ and the inequality $(\beta+\gamma)^2 \le 2\beta^2 + 2\gamma^2$ for all $\beta,\gamma \ge 0$ to obtain

$$
\begin{aligned}
\langle g^{\mathrm{obs}}, F(h)\rangle &= \langle g^{\mathrm{obs}}, F(h) - F(\overline{f})\rangle + \langle g^{\mathrm{obs}}, F(\overline{f})\rangle \\
&\le N\|F(h) - F(\overline{f}) \mid B_{t,1}^{d/2}\| + A \\
&\le NC_I\|F(h) - F(\overline{f})\|_{L^2}^{1-\theta}\|h - \overline{f} \mid B_{p,1}^r\|^{\theta} + A \\
&\le \frac{1}{4}\|F(h) - F(\overline{f})\|_{L^2}^2 + cN^{\frac{2}{1-\theta}}\|h - \overline{f} \mid B_{p,1}^r\|^{\frac{2\theta}{1+\theta}} + A \\
&\le \frac{1}{2}\|F(h)\|_{L^2}^2 + \frac{1}{2}\|F(\overline{f})\|_{L^2}^2 + cN^{\frac{2}{1-\theta}}\|h - \overline{f} \mid B_{p,1}^r\|^{\frac{2\theta}{1+\theta}} + A
\end{aligned}
$$

with a generic constant c. Plugging this into the product of the Tikhonov functional and α yields

$$
\frac{1}{2}\|F(\mathcal{S}x)\|_{L^2}^2 - \langle g^{\mathrm{obs}}, F(\mathcal{S}x)\rangle + \alpha\|x\|_{r,p,1} \ge -B - cN^{\frac{2}{1-\theta}}\|\mathcal{S}x - \overline{f} \mid B_{p,1}^r\|^{\frac{2\theta}{1+\theta}} + \alpha\|x\|_{r,p,1}
$$

for $B = A + \frac{1}{2}\|F(\overline{f})\|_{L^2}^2$. Using $\frac{2\theta}{1+\theta} < 1$ and Assumption 2.11, we see that the right hand side tends to ∞ if $\|x\|_{r,p,1} \to \infty$. Hence, a minimizing sequence in (3.81) has to stay bounded in $b_{p,1}^r$ almost surely. Therefore, using the same topology $\tau_{\mathbb{X}}$ on $b_{p,1}^r$ as in the proof of Proposition 2.16(a) (note that $a + r > \frac{d}{2} \ge d\left(\frac{1}{p} - \frac{1}{2}\right)$) we see that a minimizing sequence in (3.81) has a $\tau_{\mathbb{X}}$-convergent subsequence $x_n \to x \in b_{p,1}^r$. As D is $\tau_{\mathbb{X}}$-closed and $F \circ \mathcal{S}$ is $\tau_{\mathbb{X}}$-to-norm continuous we obtain $\|F(\mathcal{S}x_n) - F(\mathcal{S}x)\|_{L^2} \to 0$. Together with boundedness of x_n in $b_{p,1}^r$ we also obtain $\|F(\mathcal{S}x_n) - F(\mathcal{S}x) \mid B_{t',1}^{d/2}\| \to 0$ by (3.83). Using $\tau_{\mathbb{X}}$-lower semi-continuity we conclude that x is a minimizer in (3.81). Hence, $\mathcal{S}x \in R_\alpha(g)$. $\qquad\square$

Note that the additional requirement on the existence of $\overline{f}$ is always met if $F(0) = 0$. In the articles [105, 67] statistical error bounds are proven under variational source conditions. The following proposition is an adaption of [104, Thm. 2.6]. Relying on the interpolation type inequality (3.83) we show that a variational source condition (3.14) implies statistical convergence rates.

Proposition 3.82 (Statistical error bounds under variational source conditions). *Suppose Assumptions 3.80, 2.11 and 2.15(b) hold true and that g^{obs} is given by (3.79). Let $r \ge 0$ and $p \in [1,2]$ and assume that there exists $t \in (1,2)$, $\theta \in (0,1)$ and $C_I > 0$ such that (3.83) holds true for all $f_1, f_2 \in B_{p,1}^r \cap \mathrm{dom}(F)$. Suppose that $f \in B_{p,1}^r \cap \mathrm{dom}(F)$ and the sequence of wavelet coefficients $x \in b_{p,1}^r$ with $\mathcal{S}x = f$ satisfies a variational source condition*

$$
\frac{1}{2}\|x - z\|_{r,p,1} + \|x\|_{r,p,1} - \|z\|_{r,p,1} \le \varphi\left(\|F(\mathcal{S}x) - F(\mathcal{S}z)\|_{L^2}^2\right) \quad \text{for all } z \in D
$$

with $\varphi\colon [0,\infty) \to [0,\infty)$ concave and continuous. Then every $\hat{f}_\alpha \in R_\alpha(g^{\mathrm{obs}})$ (see (3.81)) satisfies the following bounds

$$
\|F(f) - F(\hat{f}_\alpha)\|_{L^2} \le C_1\|\varepsilon Z \mid B_{t',\infty}^{-d/2}\|^{\frac{1}{1-\theta}}\alpha^{-\frac{\theta}{1-\theta}} + \sqrt{8\alpha\breve{\varphi}(4\alpha)} \quad \text{and} \tag{3.84}
$$

$$
\|f - \hat{f}_\alpha \mid B_{p,1}^r\| \le C_2\|\varepsilon Z \mid B_{t',\infty}^{-d/2}\|^{\frac{2}{1-\theta}}\alpha^{-\frac{1+\theta}{1-\theta}} + C_3\breve{\varphi}(4\alpha) \tag{3.85}
$$

hold true for constants $C_1, C_2, C_3 > 0$ depending only on C_I, θ and the operator norm of $\mathcal{S}\colon b_{p,1}^r \to B_{p,1}^r$. Here $\breve{\varphi}$ denotes the function given in (3.2).

Proof. For $h \in \mathrm{dom}(F)$ we define $K(h) := \frac{1}{2\alpha}\|F(h)\|_{L^2}^2 - \frac{1}{\alpha}\langle g^{\mathrm{obs}}, F(h)\rangle$ the data fidelity term in (3.81). Then

$$
\frac{1}{2\alpha}\|F(f) - F(\hat{f}_\alpha)\|_{L^2}^2 = K(\hat{f}_\alpha) - K(f) + \frac{1}{\alpha}\langle g^{\mathrm{obs}} - F(f), F(\hat{f}_\alpha) - F(f)\rangle
$$
$$
= K(\hat{f}_\alpha) - K(f) + \frac{1}{\alpha}\langle \varepsilon Z, F(\hat{f}_\alpha) - F(f)\rangle. \tag{3.86}
$$

Here the first equality is verified by expanding the mixed term on the right hand side. Such an expansion is also used in [104, Sec. 2.3].
Now let $\hat{x}_\alpha$ be a minimizer in (3.81) such that $\hat{f}_\alpha = \mathcal{S}\hat{x}_\alpha$. Then

$$
K(\hat{f}_\alpha) + \|\hat{x}_\alpha\|_{r,p,1} \le K(f) + \|x\|_{r,p,1}.
$$

Denoting c_1 the operator norm of $\mathcal{S}\colon b_{p,1}^r \to B_{p,1}^r$ and $N := c_t\|\varepsilon Z \,|\, B_{t',\infty}^{-d/2}\|$ we combine (3.86), the last inequality and the variational source condition to

$$
\frac{1}{2c_1}\|f - \hat{f}_\alpha \,|\, B_{p,1}^r\| + \frac{1}{2\alpha}\|F(f) - F(\hat{f}_\alpha)\|_{L^2}^2
$$
$$
\le \frac{1}{2}\|x - \hat{x}_\alpha\|_{r,p,1} + \|x\|_{0,p,1} - \|\hat{x}_\alpha\|_{0,p,1} + \frac{1}{\alpha}\langle \varepsilon Z, F(\hat{f}_\alpha) - F(f)\rangle
$$
$$
\le \varphi\left(\|F(\mathcal{S}x) - F(\mathcal{S}\hat{x}_\alpha)\|_{L^2}\right) + \frac{1}{\alpha}N\|F(f) - F(\hat{f}_\alpha) \,|\, B_{t,1}^{d/2}\|. \tag{3.87}
$$

Next we evoke (3.83) and Young's inequality $\beta\gamma\delta \le c_2\beta^{\frac{2}{1-\theta}} + \frac{1}{4}\gamma^{\frac{2}{1-\theta}} + \frac{1}{4c_1}\delta^{\frac{1}{\theta}}$ for all $\beta, \gamma, \delta \ge 0$ with c_2 depending only on θ, c_1 to estimate the second summand on the right hand side

$$
\frac{1}{\alpha}N\|F(f) - F(\hat{f}_\alpha) \,|\, B_{t,1}^{d/2}\| \le \left(c_1^\theta C_I N \alpha^{-\frac{1+\theta}{2}}\right)\left(a^{-\frac{1}{2}}\|F(f) - F(\hat{f}_\alpha)\|_{L^2}\right)^{1-\theta}\|f - \hat{f}_\alpha \,|\, B_{p,1}^r\|^\theta
$$
$$
\le c_3 N^{\frac{2}{1-\theta}}\alpha^{-\frac{1+\theta}{1-\theta}} + \frac{1}{4\alpha}\|F(f) - F(\hat{f}_\alpha)\|_{L^2}^2 + \frac{1}{4c_1}\|f - \hat{f}_\alpha \,|\, B_{p,1}^r\|
$$

with $c_3 := c_2(c_1^\theta C_I)^{\frac{2}{1-\theta}}$.
Plugging into (3.87) and rearranging terms we see

$$
\frac{1}{4c_1}\|f - \hat{f}_\alpha \,|\, B_{p,1}^r\| + \frac{1}{4\alpha}\|F(f) - F(\hat{f}_\alpha)\|_{L^2}^2 \le c_3 N^{\frac{2}{1-\theta}}\alpha^{-\frac{1+\theta}{1-\theta}} + \varphi\left(\|F(\mathcal{S}x) - F(\mathcal{S}\hat{x}_\alpha)\|_{L^2}^2\right).
$$

Subtracting $\frac{1}{8\alpha}\|F(f) - F(\hat{f}_\alpha)\|_{L^2}^2$ on both sides and using the definition of $\breve{\varphi}$ we conclude

$$
\frac{1}{4c_1}\|f - \hat{f}_\alpha \,|\, B_{p,1}^r\| + \frac{1}{8\alpha}\|F(f) - F(\hat{f}_\alpha)\|_{L^2}^2 \le c_3 N^{\frac{2}{1-\theta}}\alpha^{-\frac{1+\theta}{1-\theta}} + \breve{\varphi}(4\alpha).
$$

This proves (3.85) and (3.84) follows from the basic inequality $\beta^2 + \gamma^2 \le (\beta + \gamma)^2$ for all $\beta, \gamma \ge 0$. $\qquad\qquad\square$

Finitely smoothing operators

We apply Proposition 3.82 to finitely smoothing operators. In order to verify the interpolation inequality (3.83) for a forward map F satisfying Assumption 2.15, we need an

additional assumption, namely Lipschitz continuity of F as a map from $B^{\tilde{s}}_{\tilde{p},\tilde{p}} \to B^{a+\tilde{s}}_{\tilde{p},\tilde{p}}(\Omega_{\mathbb{Y}})$ for $\tilde{s}$ and $\tilde{p}$ specified below. The basic example of periodic differential operators (see Example 2.17) shows that such an assumption is reasonable.

We formulate two theorems. The first provides statistical error bounds for (3.81) with $p \in (1, 2]$ under the smoothness assumption given by bounded sets in $B^s_{p,\infty}$ as in Sections 3.2 and 3.3. This result agrees with [64, Thm. 6.4]. A similar result on Besov space regularization with $p \in (1, 2]$ and $q \geq 2$ is [104, Thm. 4.11]. A small but anyway noteworthy difference to this results is that we keep on using the additional parameter $r > 0$ which allows to loosen the restriction $a > \frac{d}{2}$ on the degree of smoothing.

The second theorem is devoted to the case $p = 1$. In this case the techniques in [64] could not furnish the optimal convergence rate because the value $u = \infty$ is not reasonable in Assumption 3.80. Here we choose $t \in (1, 2)$ and prove (3.83) to obtain the optimal upper bound even under the weaker smoothness assumption given by bounded sets in K_s as in Section 3.5.

Theorem 3.83 (Statistical error bounds for $p > 1$). *Suppose Assumptions 2.11, 3.80, 2.15 and that g^{obs} is given by (3.79). Let $p \in (1, 2]$, $0 \leq r < s < s_{max}$ with $a + r > \frac{d}{2}$ and set $\overline{p} = \frac{2p(a+r)}{2a+pr}$. If $p < 2$ then we require that the scale $B^s_{p,q}$ (domain) is p-monotone (see Definition 2.9). Suppose that there exists $M_3 > 0$ such that the following Lipschitz condition holds true*

$$\|F(f_1) - F(f_2) \mid B^{a+r}_{p,p}\| \leq M_3 \|f_1 - f_2 \mid B^r_{p,p}\| \quad \text{for all } f_1, f_2 \in B^r_{p,p} \cap \mathrm{dom}(F). \quad (3.88)$$

If $f \in B^s_{p,\infty} \cap \mathrm{dom}(F)$ with $\|f \mid B^s_{p,\infty}\| \leq \varrho$ for some $\varrho > 0$ and the parameter α is chosen such that

$$\alpha \sim \varrho^{-\frac{a+r-d/2}{s+a+d/2}} \varepsilon^{\frac{s+2a+r}{s+a+d/2}}$$

then there exists a constant C_r independent of f, Z, ϱ and ε such that every $\hat{f}_\alpha \in R_\alpha(g^{\mathrm{obs}})$ (see (3.81)) satisfies

$$\|f - \hat{f}_\alpha \mid B^{-a}_{2,2}\| \leq C_r \left(1 + \|Z \mid B^{-d/2}_{p',\infty}\|^{\frac{a+r}{a+r-d/2}}\right) \varrho^{\frac{d/2}{s+a+d/2}} \varepsilon^{\frac{s+a}{s+a+d/2}},$$

$$\|f - \hat{f}_\alpha \mid B^r_{p,1}\| \leq C_r \left(1 + \|Z \mid B^{-d/2}_{p',\infty}\|^{\frac{2a+2r}{a+r-d/2}}\right) \varrho^{\frac{a+r+d/2}{s+a+d/2}} \varepsilon^{\frac{s-r}{s+a+d/2}} \quad \text{and}$$

$$\|f - \hat{f}_\alpha \mid B^0_{\overline{p},\overline{p}}\| \leq C_r \left(1 + \|Z \mid B^{-d/2}_{p',\infty}\|^{\frac{2a+r}{a+r-d/2}}\right) \varrho^{\frac{a+d/2}{s+a+d/2}} \varepsilon^{\frac{s}{s+a+d/2}}.$$

Proof. Starting with (3.83) we verify all assumptions of Proposition 3.82. By the real interpolation identity $B^{d/2}_{p,1}(\Omega_{\mathbb{Y}}) = \left(B^0_{p,2}(\Omega_{\mathbb{Y}}), B^{a+r}_{p,p}(\Omega_{\mathbb{Y}})\right)_{\frac{d}{2a+2r},1}$ and the continuous embedding $L^2(\Omega_{\mathbb{Y}}) = B^0_{2,2}(\Omega_{\mathbb{Y}}) \subset B^0_{p,2}(\Omega_{\mathbb{Y}})$ (see Definition 2.9 and Proposition A.11) there exists a constant c_1 such that $\|\cdot \mid B^{d/2}_{p,1}\|c_1 \leq \|\cdot\|^{1-\frac{d}{2a+2r}}_{L^2} \|\cdot \mid B^{a+r}_{p,p}\|^{\frac{d}{2a+2r}}$. Moreover, the continuous embedding $B^r_{p,1} \subset B^r_{p,p}$ yields a constant c_2 with $\|\cdot \mid B^r_{p,p}\| \leq c_2 \|\cdot \mid B^r_{p,1}\|$. We insert $F(f_1) - F(f_2)$ into the above interpolation inequality and use (3.88) to wind up with

$$\|F(f_1) - F(f_2) \mid B^{d/2}_{p,1}\| \leq c_1 M_3^{\frac{d}{2a+2r}} \|F(f_1) - F(f_2)\|^{1-\frac{d}{2a+2r}}_{L^2} \|f_1 - f_2 \mid B^r_{p,p}\|^{\frac{d}{2a+2r}} \quad (3.89)$$

$$\leq c_1 (M_3 c_2)^{\frac{d}{2a+2r}} \|F(f_1) - F(f_2)\|^{1-\frac{d}{2a+2r}}_{L^2} \|f_1 - f_2 \mid B^r_{p,1}\|^{\frac{d}{2a+2r}}.$$

That is (3.83) with $t = p$ and $\theta = \frac{d}{2a+2r}$.

The smoothness assumption on f implies a bound $\|x\|_{s,p,q} \leq c_3 \varrho$ with x the wavelet coefficients of f and with c_3 the operator norm of $\mathcal{S} \colon b_{p,\infty}^s \to B_{p,\infty}^s$. Hence Proposition 3.36 yields a variational source condition with $\varphi(\tau) = c_4 \varrho^{\frac{a+r}{s+a}} \tau^{\frac{s-r}{2s+2a}}$.

Therefore, the bounds in Proposition 3.82 are available. We have $\breve{\varphi}(\alpha) = c_5 \varrho^{\frac{2a+2r}{s+2a+r}} \alpha^{\frac{s-r}{s+2a+r}}$ for a constant depending on c_4 that is made explicit in the proof of Corollary 3.2. Moreover, with Assumption 2.15(a) we may replace $\|F(f) - F(\hat{f}_\alpha)\|_{L^2}$ by $\|f - \hat{f}_\alpha \,|\, B_{2,2}^{-a}\|$ in (3.84) to obtain

$$\|f - \hat{f}_\alpha \,|\, B_{2,2}^{-a}\| \leq c_6 \|\varepsilon Z \,|\, B_{p',\infty}^{-d/2}\|^{\frac{a+r}{a+r-d/2}} \alpha^{-\frac{d/2}{a+r-d/2}} + c_6 \varrho^{\frac{a+r}{s+2a+r}} \alpha^{\frac{s+a}{s+2s+r}} \quad \text{and}$$

$$\|f - \hat{f}_\alpha \,|\, B_{p,1}^{r}\| \leq c_6 \|\varepsilon Z \,|\, B_{p',\infty}^{-d/2}\|^{\frac{2a+2r}{a+r-d/2}} \alpha^{-\frac{a+r+d/2}{a+r-d/2}} + c_6 \varrho^{\frac{2a+2r}{s+2a+r}} \alpha^{\frac{s-r}{s+2a+r}}$$

for a constant c_6 independent of f, Z, ϱ and ε. Inserting the parameter choice rule yields

$$\|f - \hat{f}_\alpha \,|\, B_{2,2}^{-a}\| \leq c_7 \max\left\{1, \|Z \,|\, B_{p',\infty}^{-d/2}\|\right\}^{\frac{a+r}{a+r-d/2}} \varrho^{\frac{d/2}{s+a+d/2}} \varepsilon^{\frac{s+a}{s+a+d/2}} \quad \text{and}$$

$$\|f - \hat{f}_\alpha \,|\, B_{p,1}^{r}\| \leq c_7 \max\left\{1, \|Z \,|\, B_{p',\infty}^{-d/2}\|\right\}^{\frac{2a+2r}{a+r-d/2}} \varrho^{\frac{a+r+d/2}{s+a+d/2}} \varepsilon^{\frac{s-r}{s+a+d/2}}.$$

A bound in $\|\cdot \,|\, B_{\overline{p},\overline{p}}^{0}\|$ follows from interpolating the latter two inequalities with

$$\|\cdot \,|\, B_{\overline{p},\overline{p}}^{0}\| \lesssim \|\cdot \,|\, B_{2,2}^{-a}\|^{\frac{r}{a+r}} \|\cdot \,|\, B_{p,1}^{r}\|^{\frac{a}{a+r}}$$

(see (3.37)). Finally, we use $\max\{1, N\}^\gamma \leq 1 + N^\gamma$ for $N, \gamma \geq 0$ to finish the proof. $\qquad\square$

Before we state and prove our result for $p = 1$ let us discuss an alternative to condition (3.88) that already reveals the basic idea we use to handle $p = 1$.

Remark 3.84 (Alternative Lipschitz condition if $r > 0$). *In the setting of Theorem 3.83 with $r > 0$ and $a > \frac{d}{2}$ we can require the following Lipschitz condition*

$$\|F(f_1) - F(f_2) \,|\, B_{p,p}^{a}\| \leq M_3 \|f_1 - f_2 \,|\, B_{p,p}^{0}\| \quad \textit{for all } f_1, f_2 \in B_{p,p}^{0} \cap \mathrm{dom}(F)$$

instead of (3.88) *to verify the interpolation inequality* (3.83). *To this end, first use* (3.89) *with $r = 0$, i.e.*

$$\|F(f_1) - F(f_2) \,|\, B_{p,1}^{d/2}\| \lesssim \|F(f_1) - F(f_2)\|_{L^2}^{1-\frac{d}{2a}} \|f_1 - f_2 \,|\, B_{p,p}^{0}\|^{\frac{d}{2a}}.$$

Now in a second step we use the interpolation identity $B_{p,p}^{0} = \left(B_{p,2}^{-a}, B_{p,1}^{r}\right)_{\frac{a}{a+r},p}$. Together with

$$\|f_1 - f_2 \,|\, B_{p,2}^{-a}\| \lesssim \|F(f_1) - F(f_2)\|_{L^2},$$

which besides Assumption 2.15(a) and in case $p < 2$ follows from p-monotonicity, we wind up with

$$\|F(f_1) - F(f_2) \,|\, B_{p,1}^{d/2}\| \lesssim \|F(f_1) - F(f_2)\|_{L^2}^{1-\frac{d}{2a+2r}} \|f_1 - f_2 \,|\, B_{p,1}^{r}\|^{\frac{d}{2a+2r}}.$$

But this is up to a change of the constants the same inequality we used in the proof of Theorem 3.83. Note that this replacement has the disadvantage that we have to require the stronger inequality $a > \frac{d}{2}$ instead of $a + r > \frac{d}{2}$.

Using the same idea with a mixed interpolation identity (see Definition 2.9 (d) (ii)) in the case $p = 1$ allows us to start with an integrabilty index $t > 1$. In this way we avoid the critical value $u = p' = \infty$ in Assumption 3.80.

Now we state the theorem specialized to the case $p = 1$.

Theorem 3.85 (Statistical error bounds for $p = 1$). *Suppose Assumptions 2.11, 3.80 and 2.15 hold true with $a + r > \frac{d}{2}$ and that g^{obs} is given by (3.79). Let $p = 1$, $0 \leq r < s < s_{max}$ and set $\overline{p} = \frac{2a+2r}{2a+r}$. We assume that there exists $e \in \left(0, a + r - \frac{d}{2}\right)$ such that with $t := \frac{a+r}{a+r-e/2}$ the following Lipschitz condition holds true*

$$\|F(f_1) - F(f_2)\,|\,B_{t,t}^{a+r-e}\| \leq M_3 \|f_1 - f_2\,|\,B_{t,t}^{r-e}\| \quad \text{for all } f_1, f_2 \in B_{t,t}^{r-e} \cap \mathrm{dom}(F) \quad (3.90)$$

for a constant $M_3 > 0$. If $f \in K_s \cap \mathrm{dom}(F)$ (see Proposition 3.63) with $\|f\|_{K_s} \leq \varrho$ for some $\varrho > 0$ and the parameter α is chosen such that

$$\alpha \sim \varrho^{-\frac{a+r-d/2}{s+a+d/2}} \varepsilon^{\frac{s+2a+r}{s+a+d/2}}$$

then there exists a constant C_r independent of f, Z, ϱ and ε such that every $\hat{f}_\alpha \in R_\alpha(g^{\mathrm{obs}})$ (see (3.81)) satisfies

$$\|f - \hat{f}_\alpha\,|\,B_{2,2}^{-a}\| \leq C_r \left(1 + \|Z\,|\,B_{t',\infty}^{-d/2}\|^{\frac{a+r}{a+r-d/2}}\right) \varrho^{\frac{d/2}{s+a+d/2}} \varepsilon^{\frac{s+a}{s+a+d/2}},$$

$$\|f - \hat{f}_\alpha\,|\,B_{1,1}^{r}\| \leq C_r \left(1 + \|Z\,|\,B_{t',\infty}^{-d/2}\|^{\frac{2a+2r}{a+r-d/2}}\right) \varrho^{\frac{a+r+d/2}{s+a+d/2}} \varepsilon^{\frac{s-r}{s+a+d/2}} \quad \text{and}$$

$$\|f - \hat{f}_\alpha\,|\,B_{\overline{p},\overline{p}}^{0}\| \leq C_r \left(1 + \|Z\,|\,B_{t',\infty}^{-d/2}\|^{\frac{2a+r}{a+r-d/2}}\right) \varrho^{\frac{a+d/2}{s+a+d/2}} \varepsilon^{\frac{s}{s+a+d/2}}.$$

Proof. As in the proof for $p > 1$ let us start with the verification of (3.83). Due to (3.89) with p, r replaced by $t, r - e$ respectively (note $a + (r - e) > \frac{d}{2}$ and $t > 1$) we have

$$\|F(f_1) - F(f_2)\,|\,B_{t,1}^{d/2}\| \leq c_1 \|F(f_1) - F(f_2)\|_{L^2}^{1 - \frac{d}{2a+2r-2e}} \cdot \|f_1 - f_2\,|\,B_{t,t}^{r-e}\|^{\frac{d}{2a+2r-2e}} \quad (3.91)$$

with a constant c_1 specifiable as in the proof of Theorem 3.83. The interpolation inequality that belongs to the identity $B_{t,t}^{r-e} = \left(B_{2,2}^{-a}, B_{1,1}^{r}\right)$ and the assumption $\|f_1 - f_2\,|\,B_{2,2}^{-a}\| \leq M_1 \|F(f_1) - F(f_2)\|_{L^2}$ yield

$$\|F(f_1) - F(f_2)\,|\,B_{t,1}^{d/2}\| \leq c_2 \|F(f_1) - F(f_2)\|_{L^2}^{1 - \frac{d}{2a+2r}} \cdot \|f_1 - f_2\,|\,B_{1,1}^{r}\|^{\frac{d}{2a+2r}}. \quad (3.92)$$

That is (3.83) with $\theta = \frac{d}{2a+2r}$. By Lemma 3.64 the smoothness assumption on f transfers to $\|x\|_{k_{t_s}} \leq c_3 \varrho$. Hence Proposition 3.70 provides a variational source condition. From here the same arguments as in the proof of Theorem 3.83 apply and finish the proof. $\square$

Remark 3.86 (Existence of minimizers). *Note that in the setting of the last two theorems if there exists $\overline{f} \in B_{p,1}^{r}$ such that $F(\overline{f}) \in B_{t,1}^{\frac{d}{2}}$ then by Proposition 3.81 the interpolation type inequalities (3.89) and (3.92) provide existence of the estimators in (3.81) almost surely.*

Theorem 3.83 and Theorem 3.85 yield not only rates in terms of the noise level parameter ε, but also a bound of the distribution of the error in terms of the distribution of a negative Besov norm of the noise. In particular, we obtain bounds on the expectation of arbitrary moments of the error if the Besov norm of the noise satisfies a deviation inequality of the form (3.80).

Corollary 3.87 (Bound on moments (cf. [64, Cor. 6.5])). *If (3.80) holds true in addition to the assumptions of Theorem 3.83 or Theorem 3.85, then for any $\sigma \geq 1$ there exists a constant C_m independent of f, ϱ and ε and such that $\hat{f}_\alpha \in R_\alpha(g^{\mathrm{obs}})$ (see (3.81)) satisfies*

$$\mathbb{E}\left(\|f - \hat{f}_\alpha \mid B^0_{\overline{p},\overline{p}}\|^\sigma\right)^{\frac{1}{\sigma}} \leq C_m \varrho^{\frac{a+d/2}{s+a+d/2}} \varepsilon^{\frac{s}{s+a+d/2}}.$$

In particular, this inequality holds true if Z is Gaussian white noise.

Proof. It suffices to show that $\mathbb{E}\left[(1 + N^t)^\sigma\right] < \infty$ for all $t > 0$ where N denotes one of the Besov norms of the noise process appearing in Theorem 3.83 or Theorem 3.85. Due to the inequality $(\beta + \gamma)^\sigma \leq 2^{\sigma-1}(\beta^\sigma + \gamma^\sigma)$ for $\beta, \gamma \geq 0$ this reduces to showing that $\mathbb{E}\left[N^t\right] < \infty$ for all $t > 0$. This can be deduced from (3.80) as follows:

$$\mathbb{E}\left[N^t\right] \leq M_Z^t \mathbb{P}\left[N \leq M_Z\right] + \sum_{j=1}^{\infty}(M_Z + j)^t \mathbb{P}\left[j - 1 < \|Z\| - M_Z \leq j\right]$$

$$\leq M_Z^t + \sum_{j=1}^{\infty}(M_Z + j)^t \exp(-C_Z(j-1)^\mu) < \infty. \qquad \square$$

Remark 3.88 (Lower bounds). *The upper bound in the $\|\cdot \mid B^0_{\overline{p},\overline{p}}\|$-norm in Corollary 3.87 for $\sigma = 1$ coincide up to a constant with lower bounds on the expected reconstruction error for any reconstruction method shown in [104, Cor. 4.14] and [64, Prop. 6.6.] in the case of Gaussian white noise. In this sense the presented error bounds are order optimal.*

Finally, as in Corollary 3.16 the embedding $B^0_{\overline{p},\overline{p}}(\Omega) \subset L^{\overline{p}}(\Omega)$ allows to transfer the bounds on the $B^0_{\overline{p},\overline{p}}$-norm in Theorem 3.83, Theorem 3.85 and Corollary 3.87 to error bounds measured in the $L^{\overline{p}}$-norm.

Corollary 3.89 (Statistical bound in $L^{\overline{p}}$). *In the setting of Theorem 3.83 or Theorem 3.85 suppose that $B^0_{\overline{p},\overline{p}} = B^0_{\overline{p},\overline{p}}(\Omega)$ with Ω either $\mathbb{T}^d, \mathbb{R}^d$ or a bounded Lipschitz domain in $\mathbb{R}^d$. Then also the following bounds hold true*

$$\|f - \hat{f}_\alpha\|_{L^{\overline{p}}(\Omega)} \leq C'_r \max\left\{1, \|Z \mid B^{-d/2}_{p',\infty}\|\right\}^{\frac{2a+r}{a+r-d/2}} \varrho^{\frac{a+d/2}{s+a+d/2}} \varepsilon^{\frac{s}{s+a+d/2}}$$

with a constant C'_r independent of f, Z, ϱ and ε.
If in addition a large deviation inequality (3.80) is available as it is for Gaussian white noise, then

$$\mathbb{E}\left(\|f - \hat{f}_\alpha\|^\sigma_{L^{\overline{p}}(\Omega)}\right)^{\frac{1}{\sigma}} \leq C'_m \varrho^{\frac{a+d/2}{s+a+d/2}} \varepsilon^{\frac{s}{s+a+d/2}}$$

for a constant C'_m independent of f, ϱ and ε.

Backward heat equation

Now we return to the backward heat equation and show a logarithmic bound on the reconstruction error for random noise. The main ingredients are Proposition 3.76 and Proposition 3.82. We refer to [104, Thm. 5.5] for the same result for Besov space regularization with $q \geq 2$.

Theorem 3.90 (Statistical error bound for the backward heat problem)**.** *Suppose Assumptions 2.11 and 3.80 hold true with an orthonormal wavelet system on $\mathbb{T}^d$ and that g^{obs} is given by (3.79). Let $0 < s < s_{max}$ and assume $f \in B_{2,\infty}^s$ with $\|f \mid B_{2,\infty}^s\| \leq \varrho$. For $\alpha > 0$ there exists $\hat{f}_\alpha \in R_\alpha(g^{\mathrm{obs}})$ in (3.81) with $F = A_{BH}$ on $D = b_{2,1}^0$ almost surely and if α is chosen by $\alpha \sim \varepsilon$ then we have the following asymptotic behavior*

$$\|f - \hat{f}_\alpha\|_{L^2(\mathbb{T}^d)} \leq \varepsilon \|Z \mid B_{2,\infty}^{-d/2}\| + C\varrho \ln\left(\frac{\varrho}{\varepsilon}\right)^{-\frac{s}{2}} \quad \textit{for } \varepsilon \to 0.$$

If moreover a large deviation inequality (3.80) holds true as for Gaussian white noise, then we obtain

$$\mathbb{E}\left(\|f - \hat{f}_\alpha\|_{L^2(\Omega)}^\sigma\right)^{\frac{1}{\sigma}} = \mathcal{O}\left(\varrho \ln\left(\frac{\varrho}{\varepsilon}\right)^{-\frac{s}{2}}\right) \quad \textit{for } \varepsilon \to 0$$

for all $\sigma \geq 1$.

Proof. By the smoothness assumption, the $b_{2,\infty}^s$-norm of the wavelet coefficients x of f is bounded by a multiple of ϱ. In view of Proposition 3.76 it remains to find an interpolation type inequality of the form (3.83) in order to apply Proposition 3.82. By the same argument we used in Example 2.24 to show that A_{BH} maps $B_{2,2}^{-a}$ boundedly to $L^2(\mathbb{T}^d) = B_{2,2}^0(\mathbb{T}^d)$ (see Proposition A.11) one shows that it maps $B_{2,2}^0(\mathbb{T}^d)$ boundedly to $B_{2,2}^d(\mathbb{T}^d)$. Together with the continuous embedding $B_{2,1}^0(\mathbb{T}^d) \subset B_{2,2}^0(\mathbb{T}^d)$ we obtain a constant c_1 such that $\|A_{BH}f \mid B_{2,2}^d(\mathbb{T}^d)\| \leq c_1 \|f \mid B_{2,1}^0(\mathbb{T}^d)\|$. Moreover, we use the identity $B_{2,1}^{\frac{d}{2}}(\mathbb{T}^d) = \left(B_{2,2}^0, B_{2,2}^d\right)_{\frac{1}{2},1}$ to wind up with a constant c_2 such that

$$\|A_{BH}f \mid B_{2,1}^{\frac{d}{2}}(\mathbb{T}^d)\| \leq c_2 \|A_{BH}f\|_{L^2(\mathbb{T})}^{\frac{1}{2}} \cdot \|f \mid B_{2,1}^0(\mathbb{T}^d)\| \quad \text{for all } f \in B_{2,1}^0.$$

Inserting $f = f_1 - f_2$ we obtain (3.83) with $t = 2$ and $\theta = \frac{1}{2}$.
Hence existence of minimizers follows from 3.81 as $A_{BH}0 = 0$.
Furthermore, the second error bound in Proposition 3.82 and once again the continuous embedding $B_{2,1}^0(\mathbb{T}^d) \subset B_{2,2}^0(\mathbb{T}^d)$ provide

$$\|f - \hat{f}_\alpha\|_{L^2(\mathbb{T}^d)} \leq c_3 \|\varepsilon Z \mid B_{2,\infty}^{-d/2}\|^4 \alpha^{-3} + c_3 \breve{\varphi}(4\alpha)$$

with $\breve{\varphi}$ belonging to φ in Proposition 3.76.
In [36] it is shown that if $\gamma \colon [0,\infty) \to [0,\infty)$ is concave and satisfies $\gamma(\tau) \leq (-\ln \tau)^{-\frac{s}{2}}(1 + o(1))$ as $\tau \to 0$ then

$$\sup_{\tau \geq 0} \left(\gamma(\tau) - \frac{\tau}{\alpha}\right) \leq (-\ln \alpha)^{-\frac{s}{2}}(1 + o(1)) \text{ as } \alpha \to 0. \tag{3.93}$$

We use this to provide an asymptotic bound on $\breve{\varphi}(4\alpha)$. To this end, we first observe, that φ satisfies

$$\varphi(\tau) \leq C\varrho \ln\left(\frac{\varrho}{\sqrt{\tau}}\right)^{-\frac{s}{2}} = c_1 \varrho \ln\left(\frac{\varrho^2}{\tau}\right)^{-\frac{s}{2}} \quad \text{for all } \tau \leq \tau_0$$

for $c_1 = 2^{\frac{s}{2}} C$ with C and τ_0 from Proposition 3.76. We set $\gamma(\tau) := (c_1 \varrho)^{-1} \varphi(\varrho^2 \tau)$ and use (3.93) to wind up with

$$\check{\varphi}(4\alpha) = \sup_{\tau \geq 0} \left(\varphi(\tau) - \frac{\tau}{8\alpha} \right)$$

$$= c_1 \varrho \sup_{\tau \geq 0} \left(\gamma(\tau) - \frac{\varrho \tau}{8 c_1 \alpha} \right) \leq c_1 \varrho \left(- \ln \left(\frac{8 c_1 \alpha}{\varrho} \right) \right)^{-\frac{s}{2}} (1 + o(1)) \text{ as } \alpha \to 0.$$

Therefore,

$$\check{\varphi}(4\alpha) = \mathcal{O} \left(\varrho \ln \left(\frac{\varrho}{\alpha} \right)^{-\frac{s}{2}} \right) \quad \text{for } \alpha \to 0.$$

Together with the the parameter choice $\alpha \sim \varepsilon$ we obtain

$$\|f - \hat{f}_\alpha\|_{L^2(\mathbb{T}^d)} \leq \varepsilon \|Z \mid B_{2,\infty}^{-d/2}\| + C \varrho \ln \left(\frac{\varrho}{\varepsilon} \right)^{-\frac{s}{2}}.$$

If (3.80) holds true then we use the same argument as in the proof of Corollary 3.87 to wind up with

$$\mathbb{E} \left(\|f - \hat{f}_\alpha\|_{L^2(\Omega)}^\sigma \right)^{\frac{1}{\sigma}} = \mathcal{O} \left(\varepsilon + \varrho \ln \left(\frac{\varrho}{\varepsilon} \right)^{-\frac{s}{2}} \right) = \mathcal{O} \left(\varrho \ln \left(\frac{\varrho}{\varepsilon} \right)^{-\frac{s}{2}} \right) \quad \text{for } \varepsilon \to 0. \qquad \square$$

3.8 Extension to frames

To stably decompose and reconstruct signals one sometimes gains flexibility by using an overcomplete wavelet system instead of a Riesz basis. Appending more functions to a basis could also be a benefit in sparsity regularization as more functions admit a sparse representation (see [3, Sec. 6.2.]). On the other hand, if the wavelet system is not linearly independent, then the functions decomposition is not unique, i.e. the wavelet synthesis operator $\mathcal{S}$ is not longer injective. In this case, there seems to be no straightforward extension of our theory as for example (3.29), and also the inequality in Lemma 3.34 cannot hold true if $\mathcal{S}$ has a nontrivial kernel.

The situation changes when we penalize with the analysis operator as in (2.18). Here the forward operator is F itself and the Tikhonov functional is defined on a function space rather then a sequence space.

Suitable wavelet frames $\phi_{j,k} \in L^2(\Omega)$ characterize Besov spaces via their analysis operators as follows: f belongs to $B_{p,q}^s(\Omega)$ if and only if the sequence of frame coefficients $(\langle f, \phi_{j,k} \rangle)_{(j,k) \in \Lambda}$ belongs to $b_{p,q}^s$. Moreover, there are $c, C > 0$ possibly depending on s, p and q such that

$$c\|f \mid B_{p,q}^s\| \leq \|(\langle f, \phi_\lambda \rangle)_{\lambda \in \Lambda}\|_{r,p,q} \leq C\|f \mid B_{p,q}^s\|.$$

This is equivalently formulated with the analysis operator $\mathcal{T} \colon B_{p,q}^s(\Omega) \to b_{p,q}^s$ given by $\mathcal{T}(f) = (\langle f, \phi_\lambda \rangle)_{\lambda \in \Lambda}$ being bounded with closed range. We formulate the above frame condition in the context of Besov scales as defined in Definition 2.9.

Assumption 3.91. *Let $\mathcal{D} \subset B_{p,q}^s \subset \tilde{\mathcal{D}}$ be a Besov scale as defined in Definition 2.9 and $s_{max} > 0$. With $b_{p,q}^s$ defined as in Section 2.4 (with d according to the Besov scale) assume*

that $\mathcal{T}\colon \mathcal{D} \to \mathbb{R}^\Lambda$ is a linear map that extends and co-restricts to $\mathcal{T}\colon B^s_{p,q} \to b^s_{p,q}$ such that there exist constants $c_{spq}, C_{spq} > 0$ with

$$c_{spq}\|f \mid B^s_{p,q}\| \leq \|\mathcal{T}f\|_{s,p,q} \leq C_{spq}\|f \mid B^s_{p,q}\|$$

for all $s \in \mathbb{R}$, $p,q \in (0,\infty]$ satisfying $s \in (\sigma_p - s_{max}, s_{max})$ with $\sigma_p = \max\left\{d\left(\frac{1}{p}-1\right),0\right\}$.

In the setting of this assumption, let $F\colon \operatorname{dom}(F) \to \mathbb{Y}$ with $\operatorname{dom}(F) \subset \tilde{\mathcal{D}}$, $p \in [0,1]$ and $r \geq 0$. Then we consider the reconstruction method

$$R_\alpha(g) = \operatorname*{argmin}_{h \in B^r_{p,1} \cap \operatorname{dom}(F)} \left(\frac{1}{2\alpha}\|g^{\mathrm{obs}} - F(h)\|^2_{\mathbb{Y}} + \|\mathcal{T}h\|_{r,p,1}\right). \tag{3.94}$$

In the sequel we show exemplarily how to extend the results of the chapter to this situation. The following proposition is the analogue of Proposition 3.36 providing a variational source for finitely smoothing operators under a $B^s_{p,\infty}$ smoothness assumption.

Proposition 3.92 (Variational source condition)**.** *Suppose Assumptions 3.91 and 2.15(a) hold true. Let $p \in [1,2]$ and $0 \leq r < s < s_{max}$. If $p \in [1,2)$, then we require that the scale $B^s_{p,q}$ is p-monotone (see Definition 2.9). Moreover, let $\varrho > 0$ and assume $f \in B^s_{p,\infty} \cap \operatorname{dom}(F)$ with $\|f \mid B^s_{p,\infty}\| \leq \varrho$. Then*

$$\frac{c_{rp1}}{2}\|f - h \mid B^r_{p,1}\| + \|Tf\|_{r,p,1} - \|Th\|_{r,p,1} \leq \varphi\left(\|F(f) - F(h)\|^2_{\mathbb{Y}}\right)$$

for all $h \in B^r_{p,1} \cap \operatorname{dom}(F)$ with $\varphi(\tau) = C\varrho^{\frac{a+r}{s+a}}\tau^{\frac{s-r}{2s+2a}}$ for a constant C independent of f, ϱ and τ.

Proof. For sake of a better readability of the proof let us first consider the case $a < s_{\max}$. Here we simply use the variational source condition for the embedding operator found in Lemma 3.11 with $x = \mathcal{T}f$ and $z = \mathcal{T}h$ to obtain

$$\|\mathcal{T}(f-h)\|_{r,p,1} + \|\mathcal{T}f\|_{r,p,1} - \|\mathcal{T}h\|_{r,p,1} \leq c_1\|\mathcal{T}f\|^{\frac{a+r}{s+a}}_{s,p,\infty} \cdot \|\mathcal{T}(f-h)\|^{\frac{s-r}{s+a}}_{-a,p,\infty}. \tag{3.95}$$

The smoothness assumption on f yields $\|\mathcal{T}f\|_{s,p,\infty} \leq C_{sp\infty}\varrho$. The frame condition together with the embedding $B^{-a}_{2,2} \subset B^{-a}_{p,\infty}$, which requires p-monotonicity if $p < 2$, and Assumption 2.15(a) provide

$$\|\mathcal{T}(f-h)\|_{-a,p,\infty} \lesssim \|f - h \mid B^{-a}_{p,\infty}\| \lesssim \|f - h \mid B^{-a}_{2,2}\| \lesssim \|F(f) - F(h)\|_{\mathbb{Y}}. \tag{3.96}$$

Finally, the loss function in (3.95) is bounded below by $c_{rp1}\|f - h \mid B^r_{p,1}\|$. Updating (3.95) yields the claim, and we pass over to the case $a \geq s_{\max}$.
In this case we use interpolation involving the surrogate $\tilde{a} = \frac{s_{\max}}{2}$ visualized in Section 3.3. Again we start with the variational source condition for the embedding operator with the insertions as in (3.95) but with a replaced by $\tilde{a}$. We use the smoothness assumption and bound the loss function from below as above to wind up with

$$c_{rp1}\|f - h \mid B^r_{p,1}\| + \|\mathcal{T}f\|_{r,p,1} - \|\mathcal{T}h\|_{r,p,1} \leq c_2\varrho^{\frac{\tilde{a}+r}{s+\tilde{a}}}\|\mathcal{T}(f-h)\|^{\frac{s-r}{s+\tilde{a}}}_{-\tilde{a},p,\infty}. \tag{3.97}$$

We estimate the right hand side. Starting with the frame condition condition

$$\|\mathcal{T}(f-h)\|_{-\tilde{a},p,\infty} \leq C_{-\tilde{a}p\infty}\|f-h \,|\, B_{p,\infty}^{-\tilde{a}}\|$$

we move to the Besov space $B_{p,\infty}^{-\tilde{a}}$. Here we use the interpolation identity $B_{p,\infty}^{-\tilde{a}} = \left(B_{p,2}^{-a}, B_{p,1}^r\right)_{\frac{a-\tilde{a}}{a+r},1}$ and with a simple calculation of the exponents we wind up with

$$c_2 \varrho^{\frac{\tilde{a}+r}{s+\tilde{a}}} \|\mathcal{T}(f-h)\|_{-\tilde{a},p,\infty}^{\frac{s-r}{s+\tilde{a}}} \leq \left(c_3 \varrho^{\frac{a+r}{s+a}} \|f-h \,|\, B_{p,\infty}^{-a}\|^{\frac{s-r}{s+a}}\right)^{1-\vartheta} \|f-h \,|\, B_{p,1}^0\|^{\vartheta}$$

for $\vartheta = \frac{a-\tilde{a}}{a+r}\frac{s-r}{s+\tilde{a}} \in (0,1)$ and a suitable constant c_3. Note $1-\vartheta = \frac{\tilde{a}+r}{a+r}\frac{s+a}{s+\tilde{a}}$. Using Young's inequality $\beta\gamma \leq c_4 \beta^{\frac{1}{1-\vartheta}} + \frac{c_{rp1}}{2}\gamma^{\frac{1}{\vartheta}}$ on the right hand side and inserting into (3.97) we obtain

$$c_{rp1}\|f-h \,|\, B_{p,1}^r\| + \|\mathcal{T}f\|_{r,p,1} - \|\mathcal{T}h\|_{r,p,1} \leq c_5 \varrho^{\frac{a+r}{s+a}} \|f-h \,|\, B_{p,\infty}^{-a}\|^{\frac{s-r}{s+a}} + \frac{c_{rp1}}{2}\|f-h \,|\, B_{p,1}^0\|.$$

We move the last summand to the right hand side and settle $\|f-h \,|\, B_{p,\infty}^{-a}\|$ as in (3.96) to finish the proof. $\qquad\square$

Remark 3.93 (proof concept)**.** *The simpler proof concept of the latter proposition also applies in the synthesis setting. We decided to write Section 3.3 in the way it is because it introduces Proposition 3.29 which allows the comparison to the literature and is used in Section 3.6.*

Using the above variational source condition we see that Theorem 3.65 and Corollary 3.16 hold true with Assumption 2.11 replaced by Assumption 3.91 for the reconstruction method given in (3.94). As Proposition 3.92 does not require $a < s_{\max}$ the error bounds hold also true if $a \geq s_{\max}$ as in Theorem 3.37.

Remark 3.94 (Transfer of other results)**.** *All the other results in this chapter can be transferred in a similar spirit. For example we obtain an analogue of Proposition 3.70 with similar ideas as in Proposition 3.92 using the scale B_{t_s,t_s}^s instead of $B_{p,\infty}^s$ and mixed interpolation as in Definition 2.9(d)(ii) instead of the one with fixed integrabilty as in (i). Therefore, also the improved convergence rates theory in the case $p = 1$ holds true for (3.94).*

Chapter Four

Converse results and sparsity bounds

In Chapter 3 most of our theorems provide sufficient conditions, namely smoothness assumptions on the unknown, leading to certain upper bounds on the reconstruction error. We also proved optimality of the upper bounds uniformly over the class of functions satisfying the regularity assumptions. Nevertheless, we have not judged whether our assumptions are minimal, i.e. hitherto we cannot exclude that a certain rate of convergence holds true on a larger class than the one given by our smoothness conditions. This motivates the investigation of *converse results*, which state that a certain condition on the unknown is not only sufficient but even necessary for a certain asymptotic upper bound on the reconstruction error.

We elaborate this motivation by exemplary recalling the course of development of the convergence rate theory for wavelet regularization with $b_{2,1}^0$- and $b_{1,1}^0$-penalty in the last chapter. The first smoothness assumptions we required to obtain Hölder-type convergence rates were given by bounded sets in $B_{2,\infty}^s$ for the $b_{2,1}^0$-penalty and by bounded sets in $B_{1,\infty}^s$ for the $b_{1,1}^0$-penalty (see Theorem 3.15). To predict a convergence speed for functions with kinks or jumps we assigned the maximal parameters s such that these functions belong to $B_{2,\infty}^s$ or to $B_{1,\infty}^s$. In the case $p = 2$ we saw an accurate accordance of theory and practice, whereas for $p = 1$ it seemed that there is still room for sharper error bounds. Moreover, we saw that bounded sets in $B_{2,\infty}^s$ are precisely the class of functions allowing for certain variational source conditions and such a characterization seemed unlikely for $p = 1$. Hence, we developed a further theory dedicated to weighed ℓ^1-penalties and discovered that we can considerably weaken the regularity condition in the case $p = 1$, namely to smoothness measured in spaces K_s, which are slightly larger than B_{t_s,t_s}^s with $t_s < 1$. The new source sets at hand, we then could assign larger smoothness indices to functions with jumps and kinks such that we obtained faster convergence rates. In the first two sections of this chapter we will prove necessity in both situations: A certain approximation rate in $b_{2,1}^0$-regularization implies that the unknown belongs to $B_{2,\infty}^s$ and a certain approximation rate in $b_{1,1}^0$-regularization implies that the unknown belongs to K_s. These converse results show that for both scenarios there are no weaker source conditions leading to certain Hölder-type rates.

The ideas to prove these results differ significantly from each other. In the case $p = 2$ a main ingredient are *sparsity bounds*, which quantify the sparsity of the estimators and are of interest on their own. In the case $p = 1$ the strategy is to compare the minimizers with surrogate minimizers given by soft thresholding.

In Section 4.3 we prove that Hölder type image space approximation rates are equivalent to certain variational source conditions with vanishing loss function. This result is shown in the general framework of regularization with a convex penalty. In passing we establish

some new techniques which may be of independent interest in variational regularization theory.

4.1 Sparsity bound and converse results for Besov $r, 2, 1$-penalties

In Section 2.1 we recalled that estimators given by Tikhonov regularization with ℓ^1-penalization on coefficients in same basis are sparse in the sense that their decomposition is a finite linear combination of basis elements. In this section we will use the two sided Lipschitz condition in Assumption 2.15 to quantify the sparsity of the minimizers

$$\hat{x}_\alpha \in \operatorname*{argmin}_{x \in D} \left(\frac{1}{2\alpha} \|g - F(\mathcal{S}x)\|_\mathbb{Y}^2 + \|x\|_{r,p,1} \right) \tag{4.1}$$

with $D = \mathcal{S}^{-1}(B_{p,1}^r \cap \operatorname{dom}(F))$. We start with a first order optimality condition for minimality in (4.1).

Proposition 4.1 (First order condition). *Suppose Assumptions 2.11 and 2.15(b) hold true. With $\operatorname{dom}(F)^\circ$ the interior of $\operatorname{dom}(F)$ in $B_{2,2}^{-a}$ we further assume that $F : \operatorname{dom}(F)^\circ \to \mathbb{Y}$ is continuously Fréchet differentiable. Let $p \in [1, 2]$ and $0 \le r < s_{max}$ with $a + r \ge d\left(\frac{1}{p} - \frac{1}{2}\right)$. If $g^{\mathrm{obs}} \in \mathbb{Y}$, $\alpha > 0$ and $\hat{x}_\alpha$ given by (4.1) satisfies $\mathcal{S}\hat{x}_\alpha \in \operatorname{dom}(F)^\circ$ then*

$$\frac{1}{\alpha} \mathcal{S}^* F'[\mathcal{S}\hat{x}_\alpha]^* \left(g^{\mathrm{obs}} - F(\mathcal{S}\hat{x}_\alpha) \right) \in \partial \|\cdot\|_{r,p,1}(\hat{x}_\alpha).$$

Proof. The synthesis operator $\mathcal{S} \colon b_{p,1}^r \to B_{p,1}^r$ is a norm isomorphism and there is a continuous embedding $B_{p,1}^r \subset B_{2,2}^{-a}$. Therefore, the set $\tilde{D} := \mathcal{S}^{-1}(\operatorname{dom}(F)^\circ \cap B_{p,1}^r)$ is open in $b_{p,1}^r$ and together with Assumption 2.15(b) we obtain Lipschitz continuity of $F \circ \mathcal{S} \colon \tilde{D} \to \mathbb{Y}$.

This implies that the Tikhonov functional $\frac{1}{2}\|g^{\mathrm{obs}} - F(\mathcal{S}\cdot)\|_\mathbb{Y}^2 + \alpha\|\cdot\|_{r,p,1}$ is locally Lipschitz continuous on $\tilde{D}$. By [18, 2.3.2.] we obtain

$$0 \in \partial_C \left(\frac{1}{2}\|F(\mathcal{S}\cdot) - g^{\mathrm{obs}}\|_\mathbb{Y}^2 + \alpha\|\cdot\|_{r,p,1} \right)(\hat{x}_\alpha), \tag{4.2}$$

where ∂_C denotes Clarke's generalized gradient (see [18, Sec. 2.1]).

By [18, Prop. 2.2.4] the set $\partial_C \left(\frac{1}{2}\|F(\mathcal{S}\cdot) - g^{\mathrm{obs}}\|_\mathbb{Y}^2 \right)(\hat{x}_\alpha)$ is a singleton containing the Fréchet derivative. Due to the chain rule the latter is

$$\left(\frac{1}{2}\|F(\mathcal{S}\cdot) - g^{\mathrm{obs}}\|_\mathbb{Y}^2 \right)'[\hat{x}_\alpha] = \mathcal{S}^* F'[\mathcal{S}\hat{x}_\alpha]^* \left(F(\mathcal{S}\hat{x}_\alpha) - g^{\mathrm{obs}} \right).$$

For convex functionals Clarke's generalized gradient agrees with the usual subdifferential for convex functions (see [18, Prop. 2.2.7]). The claim follows as Clarke's generalized gradient satisfies a sum rule (see [18, Prop. 2.3.1]) and is compatible with scalar multiplication (see [18, Prop. 2.3.3]). $\qquad\square$

Using this first order optimality we can quantify the size of the support of the estimators in (4.1). More precisely we give an inequality that besides the known parameters a, r, d, p involves only α, the image space residual in the image space $\|g^{\mathrm{obs}} - F(\mathcal{S}\hat{x}_\alpha)\|_{\mathbb{Y}}$ and j, and claim that $(\hat{x}_\alpha)_j = 0$ whenever this inequality is satisfied. The following result is similar to [64, Thm. 5.1] to which the author contributed, but therein the restriction $a < s_{\max}$ is required.

Theorem 4.2. *Suppose Assumptions 2.11 and 2.15(b) hold true and additionally that $F: \mathrm{dom}(F)^{\mathrm{o}} \to \mathbb{Y}$ is continuously Fréchet differentiable. Let $p \in [1, 2]$ and $0 \leq r < s_{max}$ with $a + r \geq d\left(\frac{1}{p} - \frac{1}{2}\right)$. Then there exists a constant $C_s > 0$ such that for $g^{\mathrm{obs}} \in \mathbb{Y}$, $\alpha > 0$ and $\hat{x}_\alpha$ given by (4.1) with $\mathcal{S}\hat{x}_\alpha \in \mathrm{dom}(F)^{\mathrm{o}}$ we obtain*

$$(\hat{x}_\alpha)_j = 0 \quad \text{for all } j \in \mathbb{N}_0 \text{ satisfying} \quad \frac{C_s}{\alpha}\|g^{\mathrm{obs}} - F(\mathcal{S}\hat{x}_\alpha)\|_{\mathbb{Y}} < 2^{j(a+r)}2^{jd\left(\frac{1}{2} - \frac{1}{p}\right)}.$$

Proof. Assumption 2.15(b) implies that the operator norm of the Fréchet derivative $F[\mathcal{S}\hat{x}_\alpha]: B_{2,2}^{-a} \to \mathbb{Y}$ is bounded by M_2. By Definition 2.9(b) we may identify the dual of $B_{2,2}^{-a}$ with $B_{2,2}^a$. Hence, the operator norm of the adjoint operator $F[\mathcal{S}\hat{x}_\alpha]^*: \mathbb{Y} \to B_{2,2}^a$ is also bounded by M_2. Furthermore, we set $\tilde{a} = a + d\left(\frac{1}{2} - \frac{1}{p}\right)$. Then $\tilde{a} = a - d\left(\frac{1}{2} - \frac{1}{p'}\right)$ with p' the Hölder conjugate of p. Let c_1 be the embedding constant of the continuous embedding $B_{2,2}^a \subset B_{p',2}^{\tilde{a}}$. Therefore, $\xi := F'[\mathcal{S}\hat{x}_\alpha]^*\left(g^{\mathrm{obs}} - F(\mathcal{S}\hat{x}_\alpha)\right)$ satisfies

$$\|\xi \mid B_{p',2}^{\tilde{a}}\| \leq c_1\|\xi \mid B_{2,2}^a\| \leq c_1 M_2\|g^{\mathrm{obs}} - F(\hat{x}_\alpha)\|_{\mathbb{Y}}. \tag{4.3}$$

Let us first assume $\tilde{a} < s_{\max}$. Note that $\tilde{a} \geq -r > -s_{\max}$. Hence, $\mathcal{S}: b_{p,2}^{-\tilde{a}} \to B_{p,2}^{-\tilde{a}}$ is bounded in this case. Let c_2 be the operator norm of $\mathcal{S}$. In view of Definition 2.9(b) we obtain $\|\mathcal{S}^*\cdot\|_{\tilde{a},p',2} \leq c_2\|\cdot \mid B_{p',2}^{\tilde{a}}\|$. We apply Proposition 4.1 and Lemma 3.73 to estimate

$$2^{jr}2^{jd\left(\frac{1}{2} - \frac{1}{p}\right)}\| (\hat{x}_\alpha)_j \|_p = \frac{1}{\alpha}\langle(\mathcal{S}^*\xi)_j, (\hat{x}_\alpha)_j\rangle$$

$$\leq \frac{1}{\alpha}\| (\mathcal{S}^*\xi)_j \|_{p'}\| (\hat{x}_\alpha)_j \|_p$$

$$\leq \frac{c_2}{\alpha}2^{-j\tilde{a}}2^{jd\left(\frac{1}{p'} - \frac{1}{2}\right)}\|\xi \mid B_{p',2}^{\tilde{a}}\|\| (\hat{x}_\alpha)_j \|_p$$

$$\leq \frac{c_1 c_2 M_2}{\alpha}2^{-j\tilde{a}}2^{jd\left(\frac{1}{p'} - \frac{1}{2}\right)}\|g^{\mathrm{obs}} - F(\mathcal{S}\hat{x}_\alpha)\|_{\mathbb{Y}}\| (\hat{x}_\alpha)_j \|_p.$$

Since $\frac{1}{2} - \frac{1}{p} = \frac{1}{p'} - \frac{1}{2}$ this implies the claim with $C_s = c_1 c_2 M_2$ in this case.

For $\tilde{a} \geq s_{\max}$, we use real interpolation to modify (4.3) such that it becomes accessible by the wavelet system as follows: Since $r < s_{\max}$ the map $\mathcal{S}: b_{p,1}^r \to B_{p,1}^r$ is a norm isomorphism. Hence, there is a constant $c_3 > 0$, such that $\|\cdot \mid B_{p',\infty}^{-r}\| \leq c_3\|\mathcal{S}^*\cdot\|_{-r,p',\infty}$. Together with Lemma 3.73 we obtain

$$\|\xi \mid B_{p',\infty}^{-r}\| \leq c_3\|\mathcal{S}^*\xi\|_{-r,p',\infty} \leq c_3\alpha. \tag{4.4}$$

Let $\overline{a} = \frac{s_{\max}}{2}$. The interpolation inequality $\|\cdot \mid B_{p',\infty}^{\overline{a}}\| \leq c_4\|\cdot \mid B_{p',\infty}^{-r}\|^{\frac{\tilde{a} - \overline{a}}{\tilde{a} + r}}\|\cdot \mid B_{p',2}^{\tilde{a}}\|^{\frac{\overline{a} + r}{\tilde{a} + r}}$ allows us to combine (4.3) and (4.4) to

$$\|\xi \mid B_{p',\infty}^{\overline{a}}\| \leq c_5\alpha^{\frac{\tilde{a} - \overline{a}}{\tilde{a} + r}}\|g^{\mathrm{obs}} - F(\mathcal{S}\hat{x}_\alpha)\|_{\mathbb{Y}}^{\frac{\overline{a} + r}{\tilde{a} + r}}$$

with $c_5 := c_4 c_3^{\frac{\tilde{a}-\bar{a}}{\bar{a}+r}} (c_1 M_2)^{\frac{\bar{a}+r}{\tilde{a}+r}}$.

With c_6 the operator norm of $\mathcal{S}\colon b_{p,1}^{-\bar{a}} \to B_{p,1}^{-\bar{a}}$ we obtain $\|\mathcal{S}^* \cdot\|_{\bar{a},p',\infty} \le c_6 \|\cdot \mid B_{p',\infty}^{\bar{a}}\|$ and we conclude as above

$$
\begin{aligned}
2^{jr} 2^{jd\left(\frac{1}{2}-\frac{1}{p}\right)} \| (\hat{x}_\alpha)_j \|_p &\le \frac{1}{\alpha} \| (\mathcal{S}^* \xi)_j \|_{p'} \| (\hat{x}_\alpha)_j \|_p \\
&\le \frac{c_6}{\alpha} 2^{-j\bar{a}} 2^{jd\left(\frac{1}{p'}-\frac{1}{2}\right)} \|\xi \mid B_{p',2}^{\bar{a}}\| \| (\hat{x}_\alpha)_j \|_p \\
&\le c_5 c_6 \alpha^{-\frac{\bar{a}+r}{\tilde{a}+r}} 2^{-j\bar{a}} 2^{jd\left(\frac{1}{p'}-\frac{1}{2}\right)} \|g^{\mathrm{obs}} - F(\mathcal{S}\hat{x}_\alpha)\|_{\mathbb{Y}}^{\frac{\bar{a}+r}{\tilde{a}+r}} \| (\hat{x}_\alpha)_j \|_p.
\end{aligned}
$$

Hence, in this case the claim follows with $C_s = (c_5 c_6)^{\frac{\tilde{a}+r}{\bar{a}+r}}$.   □

With Theorem 4.2 at hand we conclude a converse result to the bound (3.31) in Theorem 3.15, i.e. we show that the smoothness condition $f \in B_{2,\infty}^s$ is necessary for the bound $\|f - f_\alpha \mid B_{2,1}^r\| = \mathcal{O}(\alpha^{\frac{s-r}{s+2a+r}})$. The key idea is that $\|x_j\|_2$ agrees with $\|x_j - (\hat{x}_\alpha)_j\|_2$ if j is large enough.

Theorem 4.3 (Converse result for $r, 2, 1$-penalties). *Suppose Assumptions 2.11 and 2.15 hold true and additionally that $F\colon \mathrm{dom}(F)^\mathrm{o} \to \mathbb{Y}$ is continuously Fréchet differentiable. Let $p = 2$, $0 \le r < s < s_{max}$. Assume that $f \in B_{2,1}^r \cap \mathrm{dom}(F)$ and $(f_\alpha)_{\alpha>0}$ with $f_\alpha \in R_\alpha(F(f)) \cap \mathrm{dom}(F)^\mathrm{o}$ (see (3.16) with $p = 2$). Then the following statements are equivalent:*
   (i) $f \in B_{2,\infty}^s$.
   (ii) $\gamma := \sup_{\alpha>0} \alpha^{-\frac{s-r}{s+2a+r}} \|f - f_\alpha \mid B_{2,1}^r\| < \infty$.

More precisely, we can bound $\gamma \le C \|f \mid B_{2,\infty}^s\|^{\frac{2a+2r}{s+2a+r}}$ and $\|f \mid B_{2,\infty}^s\| \le C \gamma^{\frac{s+2a+r}{2a+2r}}$ with a constant C independent of f and $(f_\alpha)_{\alpha>0}$.

Proof.

$(i) \Rightarrow (ii)$: See Theorem 3.15 (a) if $a < s_{\max}$ and Theorem 3.37 the generalization without this restriction.

$(ii) \Rightarrow (i)$: Let x_α be a minimizer in (4.1) with $\mathcal{S}x_\alpha = f_\alpha$. Since $r < s_{\max}$, there exists $x \in b_{2,1}^r$ with $\mathcal{S}x = f$. Furthermore, let c_1 be the operator norm of the inverse of $\mathcal{S}\colon b_{2,1}^r \to B_{2,1}^r$. Inserting x into the Tikhonov functional we obtain

$$
\frac{1}{2} \|F(f) - F(f_\alpha)\|_{\mathbb{Y}}^2 + \alpha \|x_\alpha\|_{r,2,1} \le \alpha \|x\|_{r,2,1}.
$$

We subtract $\alpha \|x_\alpha\|_{r,2,1}$ on both sides and apply the reverse triangle inequality to wind up with

$$
\frac{1}{2} \|F(f) - F(f_\alpha)\|_{\mathbb{Y}}^2 \le \alpha \|x\|_{r,2,1} - \alpha \|x_\alpha\|_{r,2,1} \le \alpha \|x - x_\alpha\|_{r,2,1} \le c_1 \alpha \|f - f_\alpha \mid B_{2,1}^r\|.
$$

Using the hypothesis, multiplying by 2 and taking the square roots on both sides yields

$$
\|F(f) - F(f_\alpha)\|_{\mathbb{Y}} \le (2c_1 \gamma)^{\frac{1}{2}} \alpha^{\frac{s+a}{s+2a+r}}. \tag{4.5}
$$

We apply Theorem 4.2 to obtain

$$(x_\alpha)_j = 0 \quad \text{for all } j \in \mathbb{N}_0 \text{ satisfying} \quad K\gamma^{\frac{1}{2}}\alpha^{-\frac{a+r}{s+2a+r}} < 2^{j(a+r)}$$

with $K := (2c_1)^{\frac{1}{2}}C_s$. Let $j \in \mathbb{N}_0$. We choose $\alpha = K^{\frac{s+2a+r}{a+r}}\gamma^{\frac{s+2a+r}{2a+2r}}2^{-j(s+2a+r)}$ and use the hypothesis once again to obtain

$$\|x_j\|_2 = \|x_j - (x_\alpha)_j\|_2 \leq 2^{-jr}\|x - x_\alpha\|_{r,2,1} \leq c_1 2^{-jr}\gamma\alpha^{\frac{s-r}{s+2a+r}} = c_1 K^{\frac{s-r}{a+r}}\gamma^{\frac{s+2a+r}{2a+2r}}2^{-js}. \quad (4.6)$$

This shows $x \in b_{2,\infty}^s$ with $\|x\|_{s,2,\infty} \leq c_1 K^{\frac{s-r}{a+r}}\gamma^{\frac{s+2a+r}{2a+2r}}$. With c_2 the operator norm of $\mathcal{S}: b_{2,\infty}^s \to B_{2,\infty}^s$ we obtain $\|f \mid B_{2,\infty}^s\| \leq c_1 c_2 K^{\frac{s-r}{a+r}}\gamma^{\frac{s+2a+r}{2a+2r}}$. $\qquad\square$

Remark 4.4 (Interior point). *We show that the condition on the estimators being interior points can be dropped if we require that f is an interior point of the domain.*
Suppose Assumptions 2.11 and 2.15 hold true and additionally that $F : \mathrm{dom}(F)^\circ \to \mathbb{Y}$ is continuously Fréchet differentiable. Let $p = 2$, $0 \leq r < s < s_{max}$. Assume that $f \in B_{2,1}^r \cap \mathrm{dom}(F)^\circ$ and that (ii) in Theorem 4.3 holds true. Then there exists $\delta > 0$ such that the $B_{2,2}^{-a}$ ball with radius δ around f is contained in $\mathrm{dom}(F)^\circ$. As in the proof of Theorem 4.3 we obtain (4.5) and together with the Lipschitz condition in 2.15(a) we get

$$\|f - f_\alpha \mid B_{2,2}^{-a}\| \leq c_3 \gamma^{\frac{1}{2}}\alpha^{\frac{s+a}{s+2a+r}}$$

with $c_3 := M_1(2c_1)^{\frac{1}{2}}$. Therefore, we have $f_\alpha \in \mathrm{dom}(F)^\circ$ for all $\alpha < \alpha_0 := \left(\frac{\delta}{c_3\gamma^{1/2}}\right)^{\frac{s+2a+r}{s+a}}$. Now let $j \in \mathbb{N}_0$. If $c_3 K^{\frac{s+a}{a+r}}\gamma^{\frac{s+2a+r}{2a+2r}}2^{-j(s+a)} < \delta$, then the choice of α in the fifth line from the bottom of the above proof satisfies $\alpha < \alpha_0$ and we obtain $2^{js}\|x_j\|_2 \leq c_1 K^{\frac{s-r}{a+r}}\gamma^{\frac{s+2a+r}{2a+2r}}$ in this case. On the other hand, if $c_3 K^{\frac{s+a}{a+r}}\gamma^{\frac{s+2a+r}{2a+2r}}2^{-j(s+a)} \geq \delta$ then

$$2^{js}\|x_j\|_2 \leq 2^{j(s+a)}\|x \mid B_{2,2}^{-a}\| \leq c_3 K^{\frac{s+a}{a+r}}\gamma^{\frac{s+2a+r}{2a+2r}}\|x \mid B_{2,2}^{-a}\|\delta^{-1}.$$

Putting the bound on $2^{js}\|x_j\|_2$ for small and large j together we obtain $x \in B_{2,\infty}^s$ with

$$\|x \mid B_{2,\infty}^s\| \leq c_4 \gamma^{\frac{s+2a+r}{2a+2r}}\max\{1, \|x \mid B_{2,2}^{-a}\|\delta^{-1}\}$$

with c_4 independent of f and δ. From here we conclude $f \in B_{2,\infty}^s$ with $\|f \mid B_{2,\infty}^s\| \leq c_2 c_4 \gamma^{\frac{s+2a+r}{2a+2r}}\max\{1, \|x \mid B_{2,2}^{-a}\|\delta^{-1}\}$.

In the following two remarks we provide converse results also for the other two bias bounds, namely $\|f - f_\alpha \mid B_{2,2}^{-a}\| = \mathcal{O}(\alpha^{\frac{s+a}{s+2a+r}})$ and $\|f - f_\alpha \mid B_{2,2}^0\| = \mathcal{O}(\alpha^{\frac{s}{s+2a+r}})$, which we proved in Theorem 3.15 under the condition $f \in B_{2,\infty}^s$.

Remark 4.5 (Converse result for image space bound). *In the setting of Theorem 4.3 we additionally assume $a < s_{max}$ and investigate a converse to the bias bound in the $B_{2,2}^{-a}$-norm (see (3.32)). The idea is simply to replace the $B_{2,1}^r$-norm in (4.6) by the $B_{2,2}^{-a}$-norm. In the following we provide the details:*
Assume that $\|f - f_\alpha \mid B_{2,2}^{-a}\| \leq \gamma_\mathbb{Y}\alpha^{\frac{s+a}{s+2a+r}}$ holds true for some constant $\gamma_\mathbb{Y} > 0$. Then we

obtain $\|F(f) - F(f_\alpha)\|_{\mathbb{Y}} \le M_2 \gamma_{\mathbb{Y}} \alpha^{\frac{s+a}{s+2a+r}}$ with M_2 from Assumption 2.15(b). We apply Theorem 4.2 to obtain

$$(x_\alpha)_j = 0 \quad \text{for all } j \in \mathbb{N}_0 \text{ satisfying} \quad K\gamma_{\mathbb{Y}} \alpha^{-\frac{a+r}{s+2a+r}} < 2^{j(a+r)}$$

with $K := M_2 C_s$.
Since we assume $a < s_{max}$, there exists a constant c such that $\|\cdot\|_{-a,2,2} \le c\|\mathcal{S}\cdot \mid B_{2,2}^{-a}\|$. With the choice $\alpha = (K\gamma_{\mathbb{Y}})^{\frac{s+2a+r}{a+r}} 2^{-j(s+2a+r)}$ we estimate

$$\|x_j\|_2 = \|x_j - (x_\alpha)_j\|_2 \le 2^{ja}\|x - x_\alpha\|_{-a,2,2} \le c2^{ja}\gamma_{\mathbb{Y}}\alpha^{\frac{s+a}{s+2a+r}} = cK^{\frac{s+a}{a+r}}\gamma_{\mathbb{Y}}^{\frac{s+2a+r}{a+r}} 2^{-js}.$$

This implies $f \in B_{2,\infty}^s$ with $\|f \mid B_{2,\infty}^s\| \le cK^{\frac{s+a}{a+r}}\gamma_{\mathbb{Y}}^{\frac{s+2a+r}{a+r}}$. Hence together with the statement of Theorem 3.15(a), we obtain the equivalence of the following statements.

  (i) $f \in B_{2,\infty}^s$.
  (ii) $\gamma_{\mathbb{Y}} := \sup_{\alpha>0} \alpha^{-\frac{s+a}{s+2a+r}}\|f - f_\alpha \mid B_{2,2}^{-a}\| < \infty$.

Here we can bound $\gamma_{\mathbb{Y}} \le C\|f \mid B_{2,\infty}^s\|^{\frac{a+r}{s+2a+r}}$ and $\|f \mid B_{2,\infty}^s\| \le C\gamma_{\mathbb{Y}}^{\frac{s+2a+r}{a+r}}$ with a constant C independent of f and $(f_\alpha)_{\alpha>0}$.

Remark 4.6 (Converse result for interpolated bound). *Again in the setting of Theorem 4.3 and in view of Theorem 3.15 it remains to examine whether also the bias bound in the $B_{2,2}^0$-norm is sufficient to reclaim $f \in B_{2,\infty}^s$ (note that $\overline{p} = 2$ whenever $p = 2$). Here we need the following additional assumption on the forward operator:*

$$\|f_1 - f_2 \mid B_{2,\infty}^{-r-2a}\| \le M_3\|F'[f_0]^* (F(f_1) - F(f_2)) \mid B_{2,\infty}^{-r}\| \tag{4.7}$$

for all $f_0 \in \text{dom}(F)^\circ$ and $f_1, f_2 \in \text{dom}(F)$ with a constant $M_3 > 0$. This assumption is natural as the adjoint of an a-times smoothing operator is again a-times smoothing and the composition of two a-times smoothing operators is $2a$-times smoothing (see Section 2.4, especially Example 2.17 and (2.19)).
Since $r < s_{max}$ there exists a constant $c_1 > 0$ such that $\|\cdot \mid B_{2,\infty}^{-r}\| \le c_1\|\mathcal{S}^*\cdot\|_{-r,2,\infty}$. With Proposition 4.1 and Lemma 3.73 and (4.7) we obtain

$$\|f - f_\alpha \mid B_{2,\infty}^{-r-2a}\| \le c_1 M_3\|\mathcal{S}^* F'[f_\alpha]^* (F(f) - F(f_\alpha))\|_{-r,2,\infty} \le c_1 M_3 \alpha. \tag{4.8}$$

Now let us assume $\|f - f_\alpha \mid B_{2,2}^0\| \le \gamma_2 \alpha^{\frac{s}{s+2a+r}}$ holds true with a constant $\gamma_2 > 0$. By the interpolation identity $B_{2,2}^{-a} = \left(B_{2,\infty}^{-r-2a}, B_{2,2}^0\right)_{\frac{a+r}{2a+r},2}$ there exists a constant c_2 with

$$\|\cdot \mid B_{2,2}^{-a}\| \le c_2\|\cdot \mid B_{2,\infty}^{-r-2a}\|^{\frac{a}{2a+r}} \cdot \|\cdot \mid B_{2,2}^0\|^{\frac{a+r}{2a+r}}.$$

We combine the bound in the $B_{2,2}^0$-norm with (4.8) and wind up with

$$\|f - f_\alpha \mid B_{2,2}^{-a}\| \le c_3\gamma_2^{\frac{a+r}{2a+r}}\alpha^{\frac{s+a}{s+2a+r}} \quad \text{with} \quad c_3 := c_2 (c_1 M_3)^{\frac{a}{2a+r}}.$$

In the case $a < s_{max}$ we may apply the converse result in Remark 4.5 to obtain $\|f \mid B_{2,\infty}^s\| \le Cc_3^{\frac{s+2a+r}{a+r}}\gamma_2^{\frac{s+2a+r}{2a+r}}$. If the basis is not smooth enough to satisfy $a < s_{max}$ a further modification of (4.6) using the $b_{2,2}^0$-norm instead of the $b_{2,1}^r$-norm yields the same kind of bound on $\|f \mid B_{2,\infty}^s\|$. All in all, we obtain the equivalence of the following statements:

(i) $f \in B_{2,\infty}^s$.

(ii) $\gamma_2 := \sup_{\alpha > 0} \alpha^{-\frac{s}{s+2a+r}} \| f - f_\alpha \,|\, B_{2,2}^0 \| < \infty$.

Here we can bound $\gamma_2 \leq C_2 \| f \,|\, B_{2,\infty}^s \|^{\frac{2a+r}{s+2a+r}}$ *and* $\| f \,|\, B_{2,\infty}^s \| \leq C_2 \gamma_2^{\frac{s+2a+r}{2a+r}}$ *with a constant* C_2 *independent of* f *and* $(f_\alpha)_{\alpha>0}$.

4.2   Converse result for weighted ℓ^1-regularization

As in Section 3.5 we switch to the setting of weighed ℓ^1-regularization where the forward operator G is defined on sequence spaces and the estimators are given by

$$S_\alpha(g) = \underset{z \in \ell_r^1 \cap \mathrm{dom}(G)}{\mathrm{argmin}} \left(\frac{1}{2\alpha} \| g - G(z) \|_{\mathbb{Y}}^2 + \| z \|_{\bar{r},1} \right). \tag{4.9}$$

The aim of this section are converse results to the bias bounds in Theorem 3.59: We prove that the regularity condition $x \in k_t$ is necessary and sufficient for the Hölder type approximation rate $\| x - x_\alpha \|_{\bar{r},1} = \mathcal{O}(\alpha^{1-t})$. Whereas the argument that $\| x - x_\alpha \|_{\bar{r},1} = \mathcal{O}(\alpha^{1-t})$ implies an image space approximation rate $\| G(x) - G(x_\alpha) \|_{\mathbb{Y}} = \mathcal{O}(\alpha^{\frac{2-t}{2}})$ is the same as above the strategy to reclaim the smoothness condition $x \in k_t$ differs significantly from the arguments in the last section where the converse result was built upon a first order condition. Here we will see that the latter image space approximation rate yields a certain quality of the approximation of $T_\alpha(x)$ for x and conclude with Lemma 3.53. This result appears also in [85].

Theorem 4.7 (Converse result for weighted ℓ^1-penalties)**.** *Suppose Assumption 2.27 holds true and let* $\bar{r}$ *such that* $\bar{a}_j \bar{r}_j^{-1}$ *is bounded above. Suppose* $x \in \ell_{\bar{r}}^1 \cap \mathrm{dom}(G)^\circ$, $t \in (0,1)$, *and* $(x_\alpha)_{\alpha>0}$ *minimizers of* (4.9), *i.e.* $x_\alpha \in S_\alpha(G(x))$. *Then the following statements are equivalent:*

(i) $x \in k_t$.

(ii) $\gamma := \sup_{\alpha>0} \alpha^{-1-t} \| x - x_\alpha \|_{\bar{r},1} < \infty$.

(iii) $\gamma_{\mathbb{Y}} := \sup_{\alpha>0} \alpha^{-\frac{2-t}{2}} \| G(x) - G(x_\alpha) \|_{\mathbb{Y}} < \infty$.

More precisely, we can bound $\gamma \leq c \| x \|_{k_t}^t$, $\gamma_{\mathbb{Y}} \leq \sqrt{2\gamma}$ *and* $\| x \|_{k_t} \leq c \max \left\{ \gamma_{\mathbb{Y}}^{\frac{2}{t}}, \delta^{\frac{2-2t}{t}} \| x \|_{\bar{r},1}^{\frac{2-t}{t}} \right\}$ *with* $\delta > 0$ *such that the ball with radius* δ *around* x *in* $\ell_{\bar{a}}^2$ *is contained in* $\mathrm{dom}(G)$ *and a constant* $c > 0$ *that depends on* M_1, M_2 *and* t *only.*

Proof.

$(i) \Rightarrow (ii)$: See Theorem 3.59(a).

$(ii) \Rightarrow (iii)$: As in the proof of Theorem 4.3, inserting x into the Tikhonov functional and the triangle inequality provide

$$\frac{1}{2} \| G(x) - G(x_\alpha) \|_{\mathbb{Y}}^2 \leq \alpha \left(\| x \|_{\bar{r},1} - \| x_\alpha \|_{\bar{r},1} \right) \leq \alpha \| x - x_\alpha \|_{\bar{r},1} \leq C_2 \alpha^{2-t}.$$

Multiplying by 2 and taking square roots on both sides yields (iii).

$(iii) \Rightarrow (i)$: Here the strategy is to prove that $\|G(x) - G(x_\alpha)\|_\mathbb{Y}$ is an upper bound on $\|x - T_\alpha(x)\|_{\bar{a},2}$ with T_α as defined in (3.52) up to a constant and a linear change of α and then proceed using Lemma 3.53.

First note that the statement is trivial if $x = 0$ as we have $x_\alpha = 0$ in this case. Therefore, we assume $x \neq 0$. As an intermediate step we first consider

$$z_\alpha \in \operatorname*{argmin}_{z \in \ell_{\frac{1}{r}}^1} \left(\frac{1}{2} \|x - z\|_{\bar{a},2}^2 + \alpha \|z\|_{\bar{r},1} \right). \tag{4.10}$$

The minimizer can be calculated in each coordinate separately by

$$\begin{aligned}
(z_\alpha)_j &= \operatorname*{argmin}_{z \in \mathbb{R}} \left(\frac{1}{2} \bar{a}_j^2 |x_j - z|^2 + \alpha \bar{r}_j |z| \right) \\
&= \operatorname*{argmin}_{z \in \mathbb{R}} \left(\frac{1}{2} |x_j - z|^2 + \alpha \bar{a}_j^{-2} \bar{r}_j |z| \right).
\end{aligned}$$

Hence

$$(z_\alpha)_j = \begin{cases} x_j - \bar{a}_j^{-2} \bar{r}_j \alpha & \text{if } x_j > \bar{a}_j^{-2} \bar{r}_j \alpha \\ x_j + \bar{a}_j^{-2} \bar{r}_j \alpha & \text{if } x_j < -\bar{a}_j^{-2} \bar{r}_j \alpha \\ 0 & \text{if } |x_j| \leq \bar{a}_j^{-2} \bar{r}_j \alpha \end{cases}.$$

Comparing z_α with $T_\alpha(x)$ yields $|x - T_\alpha(x)_j| \leq |x_j - (z_\alpha)_j|$ for all $j \in \Lambda$. Therefore, we have

$$\|x - T_\alpha(x)\|_{\bar{a},2} \leq \|x - z_\alpha\|_{\bar{a},2} \quad \text{for all } \alpha > 0. \tag{4.11}$$

Next we find a bound on $\|x - z_\alpha\|_{\bar{a},2}$ in terms of $\|G(x) - G(x_\alpha)\|_\mathbb{Y}$. To this end, we first assure that $z_\alpha \in \operatorname{dom}(G)$ for α small enough: Since x belongs to the interior of $\operatorname{dom}(G)$ there exist $\delta > 0$ such that $z \in \operatorname{dom}(G)$ whenever $\|x - z\|_{\bar{a},2} \leq \delta$. Comparing the values of z_α and x in (4.10) we obtain

$$\|x - z_\alpha\|_{\bar{a},2}^2 \leq 2\alpha \|x\|_{\bar{r},1} \quad \text{for all } \alpha > 0. \tag{4.12}$$

Hence $z_\alpha \in \operatorname{dom}(G)$ for all $\alpha \leq \frac{\delta^2}{2\|x\|_{\bar{r},1}}$.

Now let $0 < \alpha \leq \alpha_0 := \frac{\delta^2}{2\|x\|_{\bar{r},1}}$. We set $\beta := 2M_2^2 \alpha$ and insert into (4.10) once more to see

$$\frac{1}{2} \|x - z_\alpha\|_{\bar{a},2}^2 + \alpha \|z_\alpha\|_{\bar{r},1} \leq \frac{1}{2} \|x - x_\beta\|_{\bar{a},2}^2 + \alpha \|x_\beta\|_{\bar{r},1}.$$

Using the first inequality in Assumption 2.27 and subtracting $\alpha \|z_\alpha\|_{\bar{r},1}$ yield

$$\frac{1}{2} \|x - z_\alpha\|_{\bar{a},2}^2 \leq \frac{M_1^2}{2} \|G(x) - G(x_\beta)\|_\mathbb{Y}^2 + \alpha \left(\|x_\beta\|_{\bar{r},1} - \|z_\alpha\|_{\bar{r},1} \right). \tag{4.13}$$

As x_β is a minimizer of (4.9) we obtain

$$\beta \|x_\beta\|_{\bar{r},1} \leq \frac{1}{2} \|G(x) - G(x_\beta)\|_\mathbb{Y}^2 + \beta \|x_\beta\|_{\bar{r},1} \leq \frac{1}{2} \|G(x) - G(z_\alpha)\|_\mathbb{Y}^2 + \beta \|z_\alpha\|_{\bar{r},1}.$$

Using the second inequality in Assumption 2.27, subtracting $\beta\|z_\alpha\|_{\bar{r},1}$ and dividing by β on both sides we end up with

$$\|x_\beta\|_{\bar{r},1} - \|z_\alpha\|_{\bar{r},1} \leq \frac{M_2^2}{2\beta}\|x - z_\alpha\|_{\bar{a},2}^2 = \frac{1}{4\alpha}\|x - z_\alpha\|_{\bar{a},2}^2.$$

We insert the last inequality into (4.13), subtract $\frac{1}{4}\|x - z_\alpha\|_{\bar{a},2}^2$, multiply by 4 and take the square root and get $\|x - z_\alpha\|_{\bar{a},2} \leq \sqrt{2}M_1\|G(x) - G(x_\beta)\|_{\mathbb{Y}}$. Together with (4.11), the hypothesis (*iii*) and the definition of β we achieve

$$\alpha^{-\frac{2-t}{2}}\|x - T_\alpha(x)\|_{\bar{a},2} \leq \alpha^{-\frac{2-t}{2}}\|G(x) - G(x_\beta)\|_{\mathbb{Y}} \leq c_1\gamma_{\mathbb{Y}} \quad \text{for all } \alpha \leq \alpha_0 \qquad (4.14)$$

with $c_1 := \sqrt{2}M_1(2M_2^2)^{\frac{2-t}{2}}$. From (4.11) and (4.12) we obtain

$$\alpha^{-\frac{2-t}{2}}\|x - T_\alpha(x)\|_{\bar{a},2} \leq (2\|x\|_{\bar{r},1})^{\frac{1}{2}}\,\alpha^{-\frac{1-t}{2}} \leq (2\|x\|_{\bar{r},1})^{\frac{1}{2}}\,\alpha_0^{-\frac{1-t}{2}} = \delta^{1-t}\,(2\|x\|_{\bar{r},1})^{\frac{2-t}{2}}$$

for all $\alpha > \alpha_0$. We combine with (4.14) and use Lemma 3.53 to conclude $x \in k_t$ with

$$\|x\|_{k_t} \leq c_2 \left(\sup_{\alpha>0}\alpha^{-\frac{2-t}{2}}\|x - T_\alpha(x)\|_{\bar{a},2}\right)^{\frac{2}{t}} \leq c_3 \max\left\{\gamma_{\mathbb{Y}}^{\frac{2}{t}}, \delta^{\frac{2-2t}{t}}\|x\|_{\bar{r},1}^{\frac{2-t}{t}}\right\}.$$

with $c_3 := c_2 \max\{c_1^{\frac{2}{t}}, 2^{\frac{2-t}{t}}\}$ and c_2 given explicitly in Lemma 3.53. $\qquad\qquad\square$

Application to Besov $r, 1, 1$-penalties

We apply the above result to wavelet regularization with $b_{1,1}^r$-penalty to see that K_s is the maximal source set leading to the approximation rate $\|f - f_\alpha \,|\, B_{1,1}^r\| = \mathcal{O}(\alpha^{\frac{s-r}{s+2a+r}})$ for the estimator for the estimators $f_\alpha \in R_\alpha(F(f))$ given in (3.61).

Theorem 4.8 (Converse result for Besov $r, 1, 1$-penalties)**.** *Suppose Assumptions 2.11 and 2.15 hold true with $a < s_{max}$. Let $0 \leq r < s < s_{max}$ with $a + r \geq \frac{d}{2}$. Assume that $f \in B_{1,1}^r \cap \mathrm{dom}(F)^\circ$ and $(f_\alpha)_{\alpha>0}$ is a selection of minimizers $f_\alpha \in R_\alpha(F(f))$ (see (3.61)). Then the following statements are equivalent:*

(i) $f \in K_s$.
(ii) $\gamma := \alpha^{-\frac{s-r}{s+2a+r}}\|f - f_\alpha \,|\, B_{1,1}^r\| < \infty$.
(iii) $\gamma_{\mathbb{Y}} := \alpha^{-\frac{s+a}{s+2a+r}}\|f - f_\alpha \,|\, B_{2,2}^{-a}\| < \infty$.

More precisely, we can bound

$$\gamma \leq C\|f\|_{K_s}^{\frac{2a+2r}{s+2a+r}}, \quad \gamma_{\mathbb{Y}} \leq C\gamma^{\frac{1}{2}} \quad \text{and} \quad \|f\|_{K_s} \leq C\max\left\{\gamma_{\mathbb{Y}}^{\frac{s+2a+r}{a+r}}, \delta^{\frac{s-r}{a+r}}\|f \,|\, B_{1,1}^r\|^{\frac{a+s}{a+r}}\right\}$$

with $\delta > 0$ such that the ball with radius δ around f in $B_{2,2}^{-a}$ is contained in $\mathrm{dom}(F)$ and a constant C independent of f and $(f_\alpha)_{\alpha>0}$.

Proof.

$(i) \Rightarrow (ii)$: See Theorem 3.65(a).

$(ii) \Rightarrow (iii)$: Let $x_\alpha, x \in b_{1,1}^r$ with $f_\alpha = \mathcal{S}x_\alpha$ and $f = \mathcal{S}x$. With c_1 such that $\|\cdot\|_{r,1,1} \leq c_1 \|\cdot \,|\, B_{1,1}^r\|$, the Lipschitz condition in 2.15(a) and the same idea as in the corresponding implication in the proof of Theorem 4.3 we see

$$\frac{1}{2M_1^2} \|f - f_\alpha \,|\, B_{2,2}^{-a}\|^2 \leq \frac{1}{2} \|F(f) - F(f_\alpha)\|_{\mathbb{Y}}^2 \leq c_1 \alpha \|f - f_\alpha \,|\, B_{1,1}^r\| \leq c_1 \gamma \alpha^{\frac{2s+2a}{s+2a+r}}. \quad (4.15)$$

Multiplying by $2M_1^2$ and taking the square root yields the desired implication.

$(iii) \Rightarrow (i)$: Let $\overline{a}, \overline{r}, \mathrm{dom}(G)$ and G be as in Remark 2.28 As $\mathcal{S} \colon b_{2,2}^{-a} \to B_{2,2}^{-a}$ is a norm isomorphism we have $x \in \mathrm{dom}(G)^\circ$. With 2.15(b) and the hypothesis we get

$$\|G(x) - G(x_\alpha)\|_{\mathbb{Y}} = \|F(f) - F(f_\alpha)\|_{\mathbb{Y}} \leq M_2 \|f - f_\alpha \,|\, B_{2,2}^{-a}\| \leq M_2 \gamma_{\mathbb{Y}} \alpha^{\frac{s+a}{s+2a+r}} = \gamma_{\mathbb{Y}} \alpha^{\frac{2-t_s}{2}}$$

with t_s given by (3.62). By Theorem 4.7 we obtain

$$x \in k_{t_s} \quad \text{with} \quad \|x\|_{k_{t_s}} \leq c_2 \max\left\{ \gamma_{\mathbb{Y}}^{\frac{s+2a+r}{a+r}}, \delta^{\frac{s-r}{a+r}} \|x\|_{r,1,1}^{\frac{a+s}{a+r}} \right\}$$

with $\delta > 0$ such that the ball with radius δ around x in $b_{2,2}^{-a}$ is contained in $\mathrm{dom}(G)$ and a constant c_2 depending on M_1, M_2 and t_s only. With Lemma 3.64 we conclude $f \in K_s$ and the claimed bound on $\|f\|_{K_s}$. $\qquad\square$

4.3 Image space approximation rates and variational source conditions

In this section we revisit the general setting of a convex penalty introduced in Assumption 2.2. We recall the minimizers

$$S_\alpha(g) := \underset{f \in \mathrm{dom}(F)}{\mathrm{argmin}} \; T_\alpha(f, g) \subseteq \mathrm{dom}(\mathcal{R}) \cap \mathrm{dom}(F) \qquad (4.16)$$

with the generalized Tikhonov functional

$$T_\alpha(f, g) = \frac{1}{2\alpha} \|g - F(f)\|_{\mathbb{Y}}^2 + \mathcal{R}(f).$$

In this general setting where the only conditions on the forward operator and the penalty term are the topological ones in Assumption 2.2 it seems unlikely to obtain convergence rates or converse results under smoothness assumptions given by balls in concrete function spaces as above. Nevertheless, the results of this section reveal some structural insights to generalized Tikhonov regularization.

The main result of this section is the equivalence of a Hölder-type image space approximation rate, i.e. a bound of the form $\|Af - Af_\alpha\|_{\mathbb{Y}} = \mathcal{O}(\alpha^\nu)$ with $\nu \in \left(\frac{1}{2}, 1\right]$ and a Hölder-type variational source condition with vanishing loss function

$$\mathcal{R}(f) - \mathcal{R}(h) \leq \varphi(\|Af - Ah\|_{\mathbb{Y}}^2) \quad \text{for all } h \in \mathbb{X} \qquad (4.17)$$

with $\varphi(\tau) = \mathcal{O}\left(\tau^{\frac{2\nu-1}{2\nu}}\right)$ under the assumption that the forward operator $F = A$ is linear and $\mathbb{Y}$ is a Hilbert space.

In [72] condition (4.17) is used to prove convergence rates with respect to the twisted Bregman distance of $\mathcal{R}$, and it is shown that the source condition (2.4) implies (4.17) with $\phi \sim \sqrt{\cdot}$. In [38] necessity of (4.17) for convergence rates with respect to the twisted Bregman distance under a fixed parameter choice rule is proven. In Remarks 3.21 and 3.58 we saw that variational source conditions with vanishing loss function are equivalent to those with norm loss functions exemplary in two concrete setups. The results of this section can also be found in the author's article [84].

This section is structured as follows: As a first intermediate step we show that bounds on the *defect of the Tikhonov functional*, used as a source condition in [59], give rise to a characterization of condition (4.17). Thereafter we study the *minimal value* of the Tikhonov functional and show that it is differentiable in the regularization parameter α. This implies differentiability of the defect of the Tikhonov functional and allows us to expose a connection between bounds on the defect of the Tikhonov functional and image space approximation rates. We end this section with a converse result for exact data using the Bregman distance of the penalty as a loss function.

Defect function and its link to variational source conditions

As already mentioned above, we introduce the *defect of the Tikhonov functional*. Whereas we will assume linearity of the forward operator in the sequel, the result of this section applies also in the nonlinear case.

Definition 4.9 (Defect of the Tikhonov functional)**.** *In the setting of Assumption 2.2 let* $f \in \mathrm{dom}(\mathcal{R}) \cap \mathrm{dom}(F)$. *We define the defect of the Tikhonov functional* $\sigma_f \colon (0, \infty) \to [0, \infty)$ *by*

$$\sigma_f(\alpha) := T_\alpha(f, F(f)) - T_\alpha(f_\alpha, F(f)) = \mathcal{R}(f) - \mathcal{R}(f_\alpha) - \frac{1}{2\alpha} \|F(f) - F(f_\alpha)\|_{\mathbb{Y}}^2$$

for a minimizer $f_\alpha \in S_\alpha(F(f))$ *(see (4.16)).*

Note that σ_f is well-defined as the value of $\sigma_f(\alpha)$ does not depend on the particular choice of $f_\alpha \in S_\alpha(F(f))$.

In the following lemma we show that $\sigma_f(\alpha) \to 0$ for $\alpha \searrow 0$ provided that f is $\mathcal{R}$-minimal in $F^{-1}(\{F(f)\})$, which is automatically satisfied if F is injective. This kind of statement is covered by the usual convergence results for generalized Tikhonov regularization. For the sake of completeness we yet decided to include a proof.

Lemma 4.10 (Limit of the defect)**.** *Suppose Assumption 2.2 holds true and let* $f \in \mathrm{dom}(\mathcal{R}) \cap \mathrm{dom}(F)$. *Then* σ_f *is non-decreasing. If additionally* $\mathcal{R}(f) \leq \mathcal{R}(h)$ *is satisfied for all* $h \in \mathrm{dom}(F)$ *with* $F(h) = F(f)$, *then* $\sigma_f(\alpha) \to 0$ *as* $\alpha \searrow 0$.

Proof. We denote $\vartheta(\alpha) := \inf_{h \in \mathrm{dom}(F)} \frac{1}{2\alpha} \|F(f) - F(h)\|_{\mathbb{Y}}^2 + \mathcal{R}(h)$ for the minimal value of the Tikhonov functional. Then ϑ is non-increasing as a pointwise infimum over non-increasing functions. Therefore, $\sigma_f(\alpha) = \mathcal{R}(f) - \vartheta(\alpha)$ is non-decreasing.
We set $m = \sup_{\alpha > 0} \vartheta(\alpha)$. Since ϑ is non-increasing we obtain $\vartheta(\alpha) \to m$ and therefore $\sigma_f(\alpha) \to \mathcal{R}(f) - m$ as $\alpha \searrow 0$. Thereby, it remains to prove $m = \mathcal{R}(f)$.
From $\vartheta(\alpha) = \inf_{h \in \mathrm{dom}(F)} T_\alpha(h, F(f)) \leq T_\alpha(f, F(f)) = \mathcal{R}(f)$ for all $\alpha > 0$ we obtain

$m \leq \mathcal{R}(f)$. In particular $m < \infty$ as $f \in \mathrm{dom}(\mathcal{R})$.

For every α let $f_\alpha \in S_\alpha(F(f))$ (see (4.16)). We obtain $\vartheta(\alpha) = \frac{1}{2\alpha}\|F(f) - F(f_\alpha)\|_{\mathbb{Y}}^2 + \mathcal{R}(f_\alpha)$. First this implies

$$\|F(f) - F(f_\alpha)\|_{\mathbb{Y}} \leq \sqrt{2\alpha m} \to 0 \quad \text{as} \quad \alpha \searrow 0. \tag{4.18}$$

Second $(f_\alpha)_{\alpha>0}$ is contained in the $\tau_{\mathbb{X}}$-compact set $\{h \in \mathrm{dom}(F) \colon \mathcal{R}(h) \leq m\}$. Hence there exists $f_0 \in \mathrm{dom}(F)$ with $\mathcal{R}(f_0) \leq m$ such that every $\tau_{\mathbb{X}}$-neighborhood of f_0 contains f_α for arbitrarily small α. By $\tau_{\mathbb{X}}$-to-weak continuity of F this implies that every weak-neighborhood of $F(f_0)$ contains $F(f_\alpha)$ for arbitrarily small α. In view of (4.18) this forces $F(f) = F(f_0)$. Finally, assuming that f satisfies the stated $\mathcal{R}$-minimality this implies $\mathcal{R}(f) \leq \mathcal{R}(f_0) \leq m$. $\qquad\square$

We write

$$\sigma_f(\alpha) = \sup_{h \in \mathrm{dom}(F)} \left(\mathcal{R}(f) - \mathcal{R}(h) - \frac{1}{2\alpha}\|F(f) - F(h)\|^2 \right) \tag{4.19}$$

to note a similarity to the *distance functions* used in [42, (4.1)], [36, Ch. 12], [38, (3.1)] and [59, Ch. 3]. In [59, Prop. 4] it is shown for linear forward operators that a variational source condition (4.17) implies a bound on the defect function σ_f.

Before we state and prove a sharp connection between bounds on the defect function and variational source conditions we need a preliminary result showing that a concave and continuous function $\varphi \colon [0,\infty) \to [0,\infty)$ can be recovered from the function $\breve{\varphi}$ that we have already used in Section 3.1.

Lemma 4.11 (Index function calculus)**.** *Let $\varphi \colon [0,\infty) \to [0,\infty)$ be concave and continuous. Recall the function $\breve{\varphi} \colon (0,\infty) \to [0,\infty)$ given by $\breve{\varphi}(\alpha) = \sup_{\tau \geq 0} \left(\varphi(\tau) - \frac{1}{2\alpha}\tau \right)$ introduced in (3.2). We have*

$$\varphi(\tau) = \inf_{\alpha > 0} \left(\breve{\varphi}(\alpha) + \frac{1}{2\alpha}\tau \right) \quad \text{for all } \tau \in [0,\infty).$$

Proof. The function

$$g_\varphi \colon \mathbb{R} \to (-\infty, \infty] \text{ given by } g_\varphi(t) = \begin{cases} -\varphi(t) & \text{if } t \geq 0 \\ \infty & \text{if } t < 0 \end{cases}$$

is convex, proper and lower semi-continuous. Its Fenchel conjugate (see [6, Def. 13.1]) is given by $g_\varphi^*(\xi) = \sup_{\tau \geq 0} (\varphi(\tau) + \xi\tau)$. Non-negativity of φ implies $g_\varphi^*(\xi) = \infty$ for $\xi > 0$. Moreover, g_φ^* is non-decreasing and together with lower semi-continuity we obtain $g_\varphi^*(0) = \lim_{\xi \nearrow 0} g_\varphi^*(\xi)$. Furthermore, we note that $\breve{\varphi}(\alpha) = g_\varphi^*\left(-\frac{1}{2\alpha}\right)$ for all $\alpha > 0$.

Let $\tau \geq 0$. The biconjugation theorem (see [6, Thm. 13.32]) yields

$$\varphi(\tau) = -g_\varphi^{**}(\tau) = -\sup_{\xi \leq 0} \left(\xi\tau - g_\varphi^*(\xi) \right) = \inf_{\xi \leq 0} \left(g_\varphi^*(\xi) - \xi\tau \right). \tag{4.20}$$

As $g_\varphi^*(0) = \lim_{\xi \nearrow 0} g_\varphi^*(\xi) - \xi\tau$ we may leave out $\xi = 0$ in the infimum in (4.20) and obtain

$$\varphi(\tau) = \inf_{\xi < 0} \left(g_\varphi^*(\xi) - \xi\tau \right) = \inf_{\alpha > 0} \left(\breve{\varphi}(\alpha) + \frac{1}{2\alpha}\tau \right). \qquad\square$$

Example 4.12 (Hölder-type index functions). *Let $\varphi(\tau) = c_1\tau^\nu$ with $c_1 > 0$ and $\nu \in \left(0, \frac{1}{2}\right]$. As in the proof of Corollary 3.2 an easy calculus shows*

$$\breve{\varphi}(\alpha) = kc_1^{\frac{1}{1-\nu}}\alpha^{\frac{\nu}{1-\nu}} \quad \text{with } k = (2\nu)^{\frac{\nu}{1-\nu}} - \frac{1}{2}(2\nu)^{\frac{1}{1-\nu}}.$$

Vice versa let us consider $\sigma(\alpha) = c_2\alpha^{\frac{\nu}{1-\nu}}$. Then $\varphi_\sigma(\tau) := k^{\nu-1}c_2^{1-\nu}\tau^\nu$ satisfies $\breve{\varphi}_\sigma = \sigma$.

As announced we prove a connection between a variational source condition

$$\mathcal{R}(f) - \mathcal{R}(h) \leq \varphi(\|F(f) - F(h)\|_{\mathbb{Y}}^2) \quad \text{for all} \quad h \in \operatorname{dom}(F) \tag{4.21}$$

for a concave and continuous function $\varphi\colon [0, \infty) \to [0, \infty)$ and a bound on σ_f.

Proposition 4.13. *Suppose Assumption 2.2 holds true. Let $f \in \operatorname{dom}(\mathcal{R}) \cap \operatorname{dom}(F)$ and $\varphi\colon [0, \infty) \to [0, \infty)$ be concave and continuous. Then (4.21) holds true if and only if $\sigma_f(\alpha) \leq \breve{\varphi}(\alpha)$ for all $\alpha > 0$.*

Proof. Assuming (4.21) we estimate

$$\sigma_f(\alpha) \leq \varphi(\|F(f) - F(f_\alpha)\|_{\mathbb{Y}}^2) - \frac{1}{2\alpha}\|F(f) - F(f_\alpha)\|_{\mathbb{Y}}^2 \leq \breve{\varphi}(\alpha).$$

This proves the first implication.
Let $h \in \operatorname{dom}(F)$. With

$$T_\alpha(f_\alpha, F(f)) \leq \frac{1}{2\alpha}\|F(f) - F(h)\|_{\mathbb{Y}}^2 + \mathcal{R}(h)$$

we obtain

$$\mathcal{R}(f) - \mathcal{R}(h) = \sigma_f(\alpha) + T_\alpha(f_\alpha, F(f)) - \mathcal{R}(h) \leq \sigma_f(\alpha) + \frac{1}{2\alpha}\|F(x) - F(h)\|_{\mathbb{Y}}^2. \tag{4.22}$$

Inserting $\sigma_f(\alpha) \leq \breve{\varphi}(\alpha)$ on the right hand side and taking the infimum over α yields $\mathcal{R}(f) - \mathcal{R}(h) \leq \inf_{\alpha>0} \breve{\varphi}(\alpha) + \frac{1}{2\alpha}\tau$ with $\tau = \|F(x) - F(h)\|_{\mathbb{Y}}^2$. Using Lemma 4.11 we obtain (4.21) which proves the reverse implication. $\qquad\square$

Corollary 4.14 (Application to Hölder-type bounds). *Suppose Assumption 2.2 holds true and let $f \in \operatorname{dom}(\mathcal{R}) \cap \operatorname{dom}(F)$ and $\nu \in \left(0, \frac{1}{2}\right]$. Then the following statements are equivalent:*

(i) There exists $c_1 \geq 0$ such that (4.21) holds true with $\varphi(\tau) = c_1\tau^\nu$.
(ii) There exists $c_2 \geq 0$ such that $\sigma_f(\alpha) \leq c_2\alpha^{\frac{\nu}{1-\nu}}$.

More precisely, with $k = (2\nu)^{\frac{\nu}{1-\nu}} - \frac{1}{2}(2\nu)^{\frac{1}{1-\nu}}$ statement (i) implies (ii) with $c_2 = kc_1^{\frac{1}{1-\nu}}$ and (ii) implies (i) with $c_1 = k^{\nu-1}c_2^{1-\nu}$.

Proof. Assuming (i) the first implication in Proposition 4.13 and Example 4.12 yield $\sigma_f(\alpha) \leq \breve{\varphi}(\alpha) = kc_1^{\frac{1}{1-\nu}}\alpha^{\frac{\nu}{1-\nu}}$.
Assuming (ii) we set $\varphi(\tau) = k^{\nu-1}c_2^{1-\nu}\tau^\nu$. Then we obtain $\sigma_f(\alpha) \leq c_2\tau^{\frac{\nu}{1-\nu}} = \breve{\varphi}(\alpha)$ by Example 4.12. Therefore, the reverse implication in Proposition 4.13 yields (1) with $c_1 = k^{\nu-1}c_2^{1-\nu}$. $\qquad\square$

Note that due to Remark 3.3 the restriction to $\nu \in \left(0, \frac{1}{2}\right]$ in Corollary 4.14 does not exclude any relevant unknown solutions.

Proposition 4.15 (Minimal index function satisfying a vsc)**.** *Suppose Assumption 2.2 holds true. Let $f \in \operatorname{dom}(\mathcal{R}) \cap \operatorname{dom}(F)$. Then*

$$\varphi_f \colon [0, \infty) \to [0, \infty) \quad \text{given by} \quad \varphi_f(\tau) = \inf_{\alpha > 0} \left(\sigma_f(\alpha) + \frac{1}{2\alpha} \tau \right)$$

is concave and continuous with $\breve{\varphi}_f = \sigma_f$. We have

$$\mathcal{R}(f) - \mathcal{R}(h) \le \varphi_f(\|F(f) - F(h)\|_{\mathbb{Y}}^2) \quad \text{for all} \quad h \in \operatorname{dom}(F). \tag{4.23}$$

Furthermore, (4.21) holds true if and only if $\varphi_f \le \varphi$.

Proof. The function φ_f is concave and upper semi-continuous functions as it is the pointwise infimum of functions possessing these properties. Lemma A.1 provides continuity of φ_f. Taking the infimum over α on the right hand side of (4.22) yields (4.23). Hence $\sigma_f(\alpha) \le \breve{\varphi}_f(\alpha)$ due to Proposition 4.13. Let $\alpha > 0$ and $\tau \ge 0$. By definition of φ_f we have $\varphi_f(\tau) - \frac{1}{2\alpha}\tau \le \sigma_f(\alpha)$. Taking the supremum over τ on the left hand side yields $\breve{\varphi}_f(\alpha) \le \sigma_f(\alpha)$.
Now suppose (4.21) holds true. Then Proposition 4.13 implies $\breve{\varphi}_f = \sigma_f \le \breve{\varphi}$. With Lemma 4.11 we conclude

$$\varphi_f(\tau) = \inf_{\alpha > 0} \left(\breve{\varphi}_f(\alpha) + \frac{1}{2\alpha} \tau \right) \le \inf_{\alpha > 0} \left(\breve{\varphi}(\alpha) + \frac{1}{2\alpha} \tau \right) = \varphi(\tau).$$

Vice verse assuming $\varphi_f \le \varphi$ we obtain (4.21) immediately from (4.23). $\square$

Remark 4.16 (Existence of variational source condition)**.** *In particular, Proposition 4.15 shows that for every $f \in \operatorname{dom}(\mathcal{R}) \cap \operatorname{dom}(F)$ there is a function φ such that (4.21) is satisfied. Note that if f is $\mathcal{R}$-minimal in $F^{-1}(\{F(f)\})$, then Lemma 4.10 yields $\varphi_f(0) = \inf_\alpha \sigma_f(\alpha) = 0$.*
A similar statement is proven in [39]. In this context we also refer to Corollary 3.32 and [40] for the existence of variational source conditions for ℓ^1-regularization.

Remark 4.17. *Inequality (4.23) is sharp for all $h = f_\alpha \in S_\alpha(g)$, $\alpha > 0$ (see (4.16)) . To see this note that by definition $\varphi_f(\tau) \le \sigma_f(\alpha) + \frac{1}{2\alpha}\tau$ for all $\tau \ge 0$ and $\alpha > 0$. Hence we have*

$$\begin{aligned}
\mathcal{R}(f) - \mathcal{R}(f_\alpha) &\le \varphi_f \left(\|F(f) - F(f_\alpha)\|_{\mathbb{Y}}^2 \right) \\
&\le \sigma_f(\alpha) + \frac{1}{2\alpha} \|F(f) - F(f_\alpha)\|_{\mathbb{Y}}^2 \\
&= \mathcal{R}(f) - \mathcal{R}(f_\alpha).
\end{aligned}$$

In order to link Hölder-type bounds on the defect of the Tikhonov functional as in condition (*ii*) in Corollary 4.14 to Hölder-type image space approximation we need some preliminary results which we establish in the following two subsections. Besides that we also need them in other places in this work. They contain some new techniques which may be of some independent interest in variational regularization theory.

Characterization of $A \circ S_\alpha$ as proximity mapping

In the setting of Assumption 2.2 with a linear forward operator $F = A$ defined on $\mathrm{dom}(A) = \mathbb{X}$ and a Hilbert space $\mathbb{Y}$ we introduce a convex function $\mathcal{Q}$ on $\mathbb{Y}$ that can be seen as a push forward of $\mathcal{R}$ through the linear operator A. We show that the proximity mapping of $\alpha\mathcal{Q}$ equals $A \circ S_\alpha$. Recall that for a convex, proper and lower semi-continuous function $\mathcal{Q} \colon \mathbb{Y} \to (-\infty, \infty]$ and $g \in \mathbb{Y}$ there is a unique minimizer $\mathrm{Prox}_\mathcal{Q}(g)$ of the function $y \mapsto \frac{1}{2}\|g - y\|_\mathbb{Y}^2 + \mathcal{Q}(y)$. The single-valued mapping

$$\mathrm{Prox}_\mathcal{Q} \colon \mathbb{Y} \to \mathbb{Y} \quad \text{given by} \quad g \mapsto \mathrm{Prox}_\mathcal{Q}(g) := \underset{y \in \mathbb{Y}}{\mathrm{argmin}} \left(\frac{1}{2}\|g - y\|_\mathbb{Y}^2 + \mathcal{Q}(y) \right)$$

is called *proximity mapping* of Q (see [6, 11.4, Def. 12.23]).

Lemma 4.18. *In the setting of Assumption 2.2 with a linear forward operator $F = A$, $\mathrm{dom}(A) = \mathbb{X}$ and $\mathbb{Y}$ a Hilbert space we define*

$$\mathcal{Q} \colon \mathbb{Y} \to (-\infty, \infty] \quad \text{by} \quad \mathcal{Q}(g) := \inf\{\mathcal{R}(f) : f \in \mathbb{X} \text{ with } Af = g\}$$

with $\inf \emptyset = \infty$. Then $\mathcal{Q}$ is convex, proper and lower semi-continuous, and we have $\mathrm{dom}(\mathcal{Q}) = A(\mathrm{dom}(\mathcal{R}))$.

Proof. Let $\lambda \in \mathbb{R}$. First we prove that $L_\lambda := \{g \in \mathbb{Y} : \mathcal{Q}(g) \leq \lambda\}$ satisfies

$$L_\lambda = A(\{f \in \mathbb{X} : \mathcal{R}(f) \leq \lambda\}).$$

To this end, let $g \in L_\lambda$. By Proposition 2.3 there exists $f \in \mathbb{X}$ with $Af = g$ and $\mathcal{R}(f) \leq \mathcal{R}(h)$ for all $h \in \mathbb{X}$ with $Ah = g$. Then $\mathcal{R}(f) = \mathcal{Q}(g) \leq \lambda$. On the other hand, if $f \in \mathbb{X}$ with $\mathcal{R}(f) \leq \lambda$ then $\mathcal{Q}(Af) \leq \mathcal{R}(f) \leq \lambda$.
Taking union over $\lambda \in \mathbb{R}$ yields $\mathrm{dom}(\mathcal{Q}) = A(\mathrm{dom}(\mathcal{R}))$. Hence $\mathcal{Q}$ is proper as $\mathcal{R}$ is proper. The sublevel sets L_λ are convex as the image of a convex set under a linear map and closed as the image of a $\tau_\mathbb{X}$-compact set under a $\tau_\mathbb{X}$-to-weak continuous map. Hence $\mathcal{Q}$ is convex and lower semi-continuous. $\square$

Remark 4.19. *Note that in the case of an injective forward operator A, the map $\mathcal{Q}$ is given by $\mathcal{Q}(g) = \mathcal{R}(A^{-1}g)$ if $g \in \mathrm{im}(A)$ and $\mathcal{Q}(g) = \infty$ if $g \in \mathbb{Y} \setminus \mathrm{im}(A)$ where $A^{-1} \colon \mathrm{im}(A) \to \mathbb{X}$ denotes the inverse map of A.*

Proposition 4.20. *Suppose Assumption 2.2 holds true with a linear forward operator $F = A$, $\mathrm{dom}(A) = \mathbb{X}$ and $\mathbb{Y}$ a Hilbert space. Let $g \in \mathbb{Y}$ and $\alpha > 0$. Then*

$$A\hat{f}_\alpha = \mathrm{Prox}_{\alpha\mathcal{Q}}(g) \quad \text{and} \quad \mathcal{R}(\hat{f}_\alpha) = \mathcal{Q}(\mathrm{Prox}_{\alpha\mathcal{Q}}(g)) \quad \text{for all} \quad \hat{f}_\alpha \in S_\alpha(g) \quad \text{(see (4.16))}.$$

In particular $A \circ S_\alpha = \mathrm{Prox}_{\alpha\mathcal{Q}}$ is single-valued. Hence $A\hat{f}_\alpha$ and $\mathcal{R}(\hat{f}_\alpha)$ do not depend on the particular choice of $\hat{f}_\alpha \in S_\alpha(g)$.

Proof. Let $v \in \mathrm{dom}(\mathcal{Q})$. Then by Lemma 4.18 we have $v \in \mathrm{im}(A)$. Due to Proposition 2.3 there exists $h \in \mathbb{X}$ with $Ah = v$ and $\mathcal{R}(h) \leq \mathcal{R}(y)$ for all $y \in \mathbb{X}$ with $Ay = v$. By definition

of $\mathcal{Q}$ that is $\mathcal{R}(h) = \mathcal{Q}(v)$. The first identity follows from

$$
\begin{aligned}
\frac{1}{2\alpha}\|g - A\hat{f}_\alpha\|_{\mathbb{Y}}^2 + \mathcal{Q}(A\hat{f}_\alpha) &\leq \frac{1}{2\alpha}\|g - A\hat{f}_\alpha\|_{\mathbb{Y}}^2 + \mathcal{R}(\hat{f}_\alpha) \\
&\leq \frac{1}{2\alpha}\|g - Ah\|_{\mathbb{Y}}^2 + \mathcal{R}(h) \\
&= \frac{1}{2\alpha}\|g - v\|_{\mathbb{Y}}^2 + \mathcal{Q}(v).
\end{aligned}
$$

Inserting $v = A\hat{f}_\alpha$ yields $\mathcal{R}(\hat{f}_\alpha) = \mathcal{Q}(A\hat{f}_\alpha) = \mathcal{Q}\left(\mathrm{Prox}_{\alpha\mathcal{Q}}(g)\right)$. $\qquad\square$

The statement in Proposition 4.20 can be read as follows: the function $\mathcal{Q}$ on $\mathbb{Y}$ stores all relevant information on $\mathcal{R}$ and A to recover the mapping $A \circ S_\alpha$ in one object.

Remark 4.21. *Suppose $f \in \mathrm{dom}(\mathcal{R})$, $\alpha > 0$ and $f_\alpha \in S_\alpha(Ax)$. In [59] the authors study upper bounds on $\mathcal{R}(f) - \mathcal{R}(f_\alpha)$ (defect for penalty) and on $\sigma_f(\alpha)$ (defect for Tikhonov functional) in terms of α. The first quantity bounds the second and it is bounded by the double of the second (see [59, Prop. 2.4]). In [59, Rem. 2.5] the authors rely on this nesting to argue that changing the selection of minimizers changes the defect for penalty at most by a factor of 2. Proposition 4.20 actually shows that the defect for penalty is independent of the choice of $f_\alpha \in S_\alpha(Ax)$.*

Differentiability of the minimal value function

As an important preliminary we introduce the minimal value function and prove its differentiability.

Definition 4.22 (minimal value function)**.** *In the setting of Assumption 2.2 with a linear forward operator $F = A$, $\mathrm{dom}(A) = \mathbb{X}$ and $\mathbb{Y}$ a Hilbert space we let $g \in \mathbb{Y}$ and define*

$$
\vartheta_g \colon (0,\infty) \to \mathbb{R} \quad by \quad \vartheta_g(\alpha) = \inf_{h \in \mathrm{dom}(\mathcal{R})} T_\alpha(h,g) = \frac{1}{2\alpha}\|g - A\hat{f}_\alpha\|_{\mathbb{Y}}^2 + \mathcal{R}(\hat{f}_\alpha)
$$

independent of the choice $\hat{f}_\alpha \in S_\alpha(g)$ in (4.16).

The main result of this subsection is the differentiability in α of ϑ_g for all $g \in \mathbb{Y}$. The approximation error $\|g - A\hat{f}_\alpha\|_{\mathbb{Y}}$ is represented by calculus rules of ϑ_g.
Recall that the Moreau envelope function of some function $\mathcal{Q}\colon \mathbb{Y} \to (-\infty, \infty]$ for $\alpha > 0$ is given by

$$
\mathcal{Q}_\alpha(g) = \inf_{y \in \mathbb{Y}} \left(\frac{1}{2\alpha}\|g - y\|_{\mathbb{Y}}^2 + \mathcal{Q}(y) \right)
$$

and the infimum is uniquely attained at $\mathrm{Prox}_{\alpha\mathcal{Q}}(g) \in \mathbb{Y}$. The key ingredient is the following result by T. Strömberg:

Lemma 4.23. *(see [97, Prop. 3(iii)]) Let $\mathcal{Q}\colon \mathbb{Y} \to (-\infty, \infty]$ be convex, proper and lower semi-continuous. The family of Moreau envelope functions $\mathcal{Q}_\alpha\colon \mathbb{Y} \to \mathbb{R}$, $\alpha > 0$ satisfies*

$$
\frac{\partial}{\partial \alpha}\mathcal{Q}_\alpha(g) = -\frac{1}{2}\|(\nabla\mathcal{Q}_\alpha)(g)\|_{\mathbb{Y}}^2.
$$

We apply Lemma 4.23 to the function $\mathcal{Q}$ defined in Lemma 4.18. Note that due to Proposition 4.20 we have

$$\mathcal{Q}_\alpha(g) = \frac{1}{2\alpha}\|g - \mathrm{Prox}_{\alpha\mathcal{Q}}(g)\|_\mathbb{Y}^2 + \mathcal{Q}(\mathrm{Prox}_{\alpha\mathcal{Q}}(g)) = \vartheta_g(\alpha). \tag{4.24}$$

Proposition 4.24. *Suppose Assumption 2.2 holds true with a linear forward operator $F = A$, $\mathrm{dom}(A) = \mathbb{X}$ and $\mathbb{Y}$ a Hilbert space. Let $g \in \mathbb{Y}$ and $\hat{f}_\alpha \in S_\alpha(g), \alpha > 0$ any selection of minimizers. The function ϑ_g is convex, non-increasing and continuously differentiable with*

$$\vartheta_g'(\alpha) = -\frac{1}{2\alpha^2}\|g - A\hat{f}_\alpha\|_\mathbb{Y}^2.$$

Proof. The Moreau envelope function $\mathcal{Q}_\alpha$ is convex, real valued and continuous with the Fenchel conjugate $(\mathcal{Q}_\alpha)^* = \mathcal{Q}^* + \frac{\alpha}{2}\|\cdot\|_\mathbb{Y}^2$ (see [6, Prop. 12.15; Prop. 13.21]). The biconjugation theorem implies

$$\vartheta_g(\alpha) = \mathcal{Q}_\alpha(g) = \left(\mathcal{Q}^* + \frac{\alpha}{2}\|\cdot\|_\mathbb{Y}\right)^*(g) = \sup_{v\in\mathbb{Y}}\left(\langle g, v\rangle - \mathcal{Q}^*(v) - \frac{\alpha}{2}\|v\|_\mathbb{Y}^2\right).$$

Hence ϑ_g is convex and non-increasing being the supremum of affine non-increasing functions.

By [6, Prop. 12.29] $\mathcal{Q}_\alpha$ is Fréchet differentiable with $\nabla\mathcal{Q}_\alpha = \frac{1}{\alpha}(\mathrm{Id}_\mathbb{Y} - \mathrm{Prox}_{\alpha\mathcal{Q}})$. Lemma 4.23 yields differentiability of $\alpha \mapsto \mathcal{Q}_\alpha(g)$ with derivative $-\frac{1}{2}\|(\nabla\mathcal{Q}_\alpha)(g)\|^2$ for all $g \in \mathbb{Y}$. Therefore, ϑ_g is differentiable and we conclude with Proposition 4.20

$$\vartheta_g'(\alpha) = -\frac{1}{2}\|(\nabla\mathcal{Q}_\alpha)(g)\|^2 = -\frac{1}{2\alpha^2}\|g - \mathrm{Prox}_{\alpha\mathcal{Q}}(g)\|_\mathbb{Y} = -\frac{1}{2\alpha^2}\|g - A\hat{x}_\alpha\|_\mathbb{Y}.$$

Finally, ϑ_g' is continuous as ϑ_g is convex and differentiable. $\qquad\square$

As a corollary we obtain non increasingness of $\alpha \mapsto \frac{1}{\alpha}\|g - A\hat{f}_\alpha\|_\mathbb{Y}$.

Corollary 4.25. *In the setting of Proposition 4.24 the map $(0,\infty) \to [0,\infty)$ given by $\alpha \mapsto \frac{1}{\alpha}\|g - A\hat{f}_\alpha\|_\mathbb{Y}$ is non-increasing.*

Proof. By convexity of ϑ_g the derivative $\vartheta_g'(\alpha) = -\frac{1}{2\alpha^2}\|g - A\hat{f}_\alpha\|_\mathbb{Y}^2$ is non decreasing. $\quad\square$

Remark 4.26. *As it fits in the context we recall the well-known facts that the function $(0,\infty) \to \mathbb{R}$ given by $\alpha \mapsto \mathcal{R}(\hat{f}_\alpha)$ is non increasing and the function $(0,\infty) \to [0,\infty)$ given by $\alpha \mapsto \|g - A\hat{f}_\alpha\|_\mathbb{Y}$ is non decreasing. To see this let $\alpha < \beta$ and set $m = \frac{1}{2}\|g - A\hat{f}_\alpha\|_\mathbb{Y}^2 - \frac{1}{2}\|g - A\hat{f}_\beta\|_\mathbb{Y}^2$. From $T_\alpha(\hat{f}_\alpha, g) \le T_\alpha(\hat{f}_\beta, g)$ and $T_\beta(\hat{f}_\beta, g) \le T_\beta(\hat{f}_\alpha, g)$ we obtain*

$$m \le \alpha\left(\mathcal{R}(\hat{f}_\beta) - \mathcal{R}(\hat{f}_\alpha)\right) \le \frac{\alpha}{\beta}m.$$

Hence $m \le 0$.

Remark 4.27 (An expression for $\mathcal{R}(f_\alpha)$)**.** *Besides the approximation error $\|g - A\hat{f}_\alpha\|_\mathbb{Y}$ also the value of the penalty at the minimizers $\mathcal{R}(f_\alpha)$ is expressed by a calculus rule of ϑ_g. To this end, we use the product rule and Proposition 4.24 to see*

$$(\alpha\vartheta_g(\alpha))' = \vartheta_g(\alpha) + \alpha\vartheta_g'(\alpha) = \mathcal{R}(f_\alpha).$$

We end this subsection by recording the limit behavior of ϑ_g.

Lemma 4.28 (Limits of the minimal value)**.** *Suppose Assumption 2.2 holds true with a linear forward operator $F = A$, $\mathrm{dom}(A) = \mathbb{X}$ and $\mathbb{Y}$ a Hilbert space. Let $g \in \mathbb{Y}$. Then*

$$\lim_{\alpha \searrow 0} \vartheta_g(\alpha) = \mathcal{Q}(g) \quad and \quad \lim_{\alpha \to \infty} \vartheta_g(\alpha) = \inf_{h \in \mathbb{X}} \mathcal{R}(h) > -\infty.$$

In particular, ϑ_g is bounded if and only if $g \in A(\mathrm{dom}(\mathcal{R}))$.

Proof. By [6, Prop. 12.32] the Moreau envelope functions have the following limits:

$$\lim_{\alpha \searrow 0} \mathcal{Q}_\alpha(g) = \mathcal{Q}(g) \quad \text{and} \quad \lim_{\alpha \to \infty} \mathcal{Q}_\alpha(g) = \inf_{y \in \mathbb{Y}} \mathcal{Q}(y).$$

Hence the stated limits follows from $\mathcal{Q}_\alpha(g) = \vartheta_g(\alpha)$ (see (4.24)) together with the observation $\inf_{v \in \mathbb{Y}} \mathcal{Q}(v) = \inf_{h \in \mathbb{X}} \mathcal{R}(h)$ (see Lemma 4.18). Using similar arguments as in the proof of Proposition 2.3 one obtains $-\infty < \inf_{h \in \mathbb{X}} \mathcal{R}(h)$ from Assumption 2.2. $\qquad\square$

Link between defect function and image space approximation

The result of this subsection is that σ_f from Definition 4.9 and hence also the smallest index function φ_f allowing for a variational source condition from Proposition 4.15 depend only on the net $(\|Ax - Ax_\alpha\|_\mathbb{Y})_{\alpha>0}$. Further we will exploit a condition when a bound $\|Ax - Ax_\alpha\|_\mathbb{Y} \leq \psi(\alpha)$ implies a bound on the defect function σ_f.
Specializing these findings to Hölder-type bounds in the next subsection leads to the equivalence result between Hölder-type image space approximation rates and variational source conditions of the form (4.17).

Lemma 4.29. *Suppose Assumption 2.2 holds true with $F = A$ linear, $\mathrm{dom}(A) = \mathbb{X}$ and $\mathbb{Y}$ a Hilbert space. Let $f \in \mathrm{dom}(\mathcal{R})$, $f_\alpha \in S_\alpha(Af)$ (see (4.16)) for $\alpha > 0$ any selection of minimizers and assume $\mathcal{R}(f) \leq \mathcal{R}(h)$ for all $h \in \mathbb{X}$ with $Ah = Af$. Then*

$$\sigma_f(\alpha) = \int_0^\alpha \frac{1}{2\beta^2} \|Af - Af_\beta\|_\mathbb{Y}^2 \, \mathrm{d}\beta \quad for \ all \ \alpha > 0.$$

Proof. Note that $\sigma_f(\alpha) = \mathcal{R}(f) - \vartheta_{Af}(\alpha)$ with the minimal value function ϑ_{Af} from Definition 4.22. Therefore, Proposition 4.24 provides continuous differentiability of σ_f with

$$\sigma_f'(\alpha) = \frac{1}{2\alpha^2} \|Af - Af_\alpha\|_\mathbb{Y}^2. \tag{4.25}$$

Let $0 < \varepsilon < \alpha$. The fundamental theorem of calculus yields

$$\sigma_f(\alpha) - \sigma_f(\varepsilon) = \int_\varepsilon^\alpha \sigma_f'(\beta) \, \mathrm{d}\beta = \int_\varepsilon^\alpha \frac{1}{2\beta^2} \|Af - Af_\alpha\|_\mathbb{Y}^2 \, \mathrm{d}\beta.$$

In view of Lemma 4.10 the expression for σ_f follows by taking the limit $\varepsilon \to 0$. $\qquad\square$

Remark 4.30. *In the setting of Lemma 4.29 we compute the limit $\alpha \to \infty$ to wind up with an interesting identity showing that the value $\mathcal{R}(f)$ is determined by the net $(\|Af - Af_\alpha\|_{\mathbb{Y}})_{\alpha>0}$. By Lemma 4.28 we have*

$$\lim_{\alpha\to\infty} \sigma_f(\alpha) = \mathcal{R}(f) - \lim_{\alpha\to\infty} \vartheta_{Af}(\alpha) = \mathcal{R}(f) - \inf_{h\in\mathbb{X}} \mathcal{R}(h).$$

Hence the identity in Lemma 4.29 implies

$$\mathcal{R}(f) = \int_0^\infty \frac{1}{2\beta^2} \|Af - Af_\beta\|_{\mathbb{Y}}^2 \, \mathrm{d}\beta + \inf_{h\in\mathbb{X}} \mathcal{R}(h). \tag{4.26}$$

Proposition 4.31 (Image space approximation)**.** *Suppose Assumption 2.2 holds true with $F = A$ linear, $\mathrm{dom}(A) = \mathbb{X}$ and $\mathbb{Y}$ a Hilbert space. Let $f \in \mathrm{dom}(\mathcal{R})$ and assume $\mathcal{R}(f) \leq \mathcal{R}(h)$ for all $h \in \mathbb{X}$ with $Ah = Af$.*

(a) We have

$$\|Af - Af_\alpha\|_{\mathbb{Y}} \leq \sqrt{2\alpha\sigma_f(\alpha)} \quad \text{for all} \quad \alpha > 0 \text{ and } f_\alpha \in S_\alpha(Af) \quad (\text{see (4.16)}).$$

(b) Let $\psi \colon [0,\infty) \to [0,\infty)$ be continuous. Assume that there is a constant $C_\psi > 0$ with

$$\int_0^\alpha \frac{1}{\beta} \psi(\beta) \, \mathrm{d}\beta \leq C_\psi \psi(\alpha) \quad \text{for all } \alpha > 0. \tag{4.27}$$

Then a bound $\|Af - Af_\alpha\|_{\mathbb{Y}} \leq \sqrt{2\alpha\psi(\alpha)}$ for all $\alpha > 0$ with some selection $f_\alpha \in S_\alpha(Ax)$ implies $\sigma_f(\alpha) \leq C_\psi \psi(\alpha)$ for all $\alpha > 0$.

Proof. (a) By Proposition 4.24 the function $\sigma_f(\alpha) = \mathcal{R}(x) - \vartheta_{Af}(\alpha)$ is concave and by Lemma 4.10 it extends continuously to $[0,\infty)$. Together with (4.25) the claim follows from

$$\frac{1}{2\alpha^2} \|Af - Af_\alpha\|_{\mathbb{Y}}^2 = \sigma_f'(\alpha) \leq \frac{1}{\alpha} \sigma_f(\alpha).$$

(b) Using Lemma 4.29 and (4.27) we obtain

$$\sigma_f(\alpha) = \int_0^\alpha \frac{1}{2\beta^2} \|Af - Af_\beta\|_{\mathbb{Y}}^2 \, \mathrm{d}\beta \leq \int_0^\alpha \frac{1}{\beta} \psi(\beta) \, \mathrm{d}\beta \leq C_\psi \psi(\alpha). \qquad \square$$

Remark 4.32. *Note that the bound in Proposition 4.31 can also be derived using the first order condition $\xi_\alpha := \frac{1}{\alpha} A^*(Af - Af_\alpha) \in \partial\mathcal{R}(f_\alpha)$. Namely, the definition of the subdifferential yields*

$$\mathcal{R}(f) - \mathcal{R}(f_\alpha) \leq \langle \xi_\alpha, f - f_\alpha \rangle = \frac{1}{\alpha} \|Af - Af_\alpha\|_{\mathbb{Y}}^2.$$

Subtracting $\frac{1}{2}\|Af - Af_\alpha\|_{\mathbb{Y}}$ and solving the resulting inequality for $\|Af - Af_\alpha\|_{\mathbb{Y}}$ proves the desired inequality.

Equivalence results for Hölder-type image space approximation rates

Now we are in position to prove the announced equivalence results.

Theorem 4.33. *Suppose Assumption 2.2 holds true with $F = A$ linear, $\mathrm{dom}(A) = \mathbb{X}$ and $\mathbb{Y}$ a Hilbert space. Let $\nu \in (\frac{1}{2}, 1]$, $f \in \mathrm{dom}(\mathcal{R})$ and $f_\alpha \in S_\alpha(Af)$ (see (4.16)) for $\alpha > 0$ any selection of a minimizers for exact data. Assume $\mathcal{R}(f) \leq \mathcal{R}(h)$ for all $h \in \mathbb{X}$ with $Ah = Af$. Then the following statements are equivalent:*

(i) There exists a constant $c_1 > 0$ with $\|Af - Af_\alpha\|_{\mathbb{Y}} \leq c_1 \alpha^\nu$ for all $\alpha > 0$.

(ii) There exists a constant $c_2 > 0$ such that the variational source condition

$$\mathcal{R}(f) - \mathcal{R}(h) \leq \varphi(\|Af - Ah\|_{\mathbb{Y}}^2) \quad \text{for all } h \in \mathbb{X}$$

holds true with $\varphi(t) = c_2 t^{\frac{2\nu-1}{2\nu}}$.

More precisely (i) implies (ii) with $c_2 = (2\nu - 1)^{-\frac{1}{2\nu}} c_1^{\frac{1}{\nu}}$, and (ii) implies (i) with $c_1 = c_2^\nu$.

Proof.

(i)$\Rightarrow$(ii): Consider the continuous function

$$\psi \colon [0, \infty) \to [0, \infty) \quad \text{given by } \psi(\alpha) = \frac{1}{2} c_1^2 \alpha^{2\nu-1}.$$

Then $c_1 \alpha^\nu = \sqrt{2\alpha\psi(\alpha)}$ for all $\alpha > 0$. We have

$$\int_0^\alpha \frac{1}{\beta} \psi(\beta) \, \mathrm{d}\beta = \frac{1}{2} c_1^2 \int_0^\alpha \beta^{2\nu-2} \, \mathrm{d}\beta = \frac{1}{2\nu - 1} \psi(\alpha).$$

Hence (4.27) is satisfied with $C_\psi = \frac{1}{2\nu-1}$. Therefore, Proposition 4.31.(b) implies $\sigma_f(\alpha) \leq \frac{c_1^2}{4\nu-2} \alpha^{2\nu-1}$ with σ_f defined in Definition 4.9. With φ_f as defined in Proposition 4.15 we estimate

$$\varphi_f(\tau) \leq \inf_{\alpha>0} \left(\frac{c_1^2}{4\nu - 2} \alpha^{2\nu-1} + \frac{1}{2\alpha} \tau \right) \leq (2\nu - 1)^{-\frac{1}{2\nu}} c_1^{\frac{1}{\nu}} \tau^{\frac{2\nu-1}{2\nu}}$$

by inserting $\alpha = (2\nu - 1)^{\frac{1}{2\nu}} \tau^{\frac{1}{2\nu}} c_1^{-\frac{1}{\nu}}$. Therefore, Proposition 4.15 implies (ii) for the desired constant $c_2 = (2\nu - 1)^{-\frac{1}{2\nu}} c_1^{\frac{1}{\nu}}$.

(ii)$\Rightarrow$(i): (see also [59, proof of Prop. 4]) With the first order condition

$$\xi_\alpha := \frac{1}{\alpha} A^* A(f - f_\alpha) \in \partial \mathcal{R}(f_\alpha)$$

condition (ii) provides

$$\frac{1}{\alpha} \|Af - Af_\alpha\|_{\mathbb{Y}}^2 = \langle \xi_\alpha, f - f_\alpha \rangle \leq \mathcal{R}(f) - \mathcal{R}(f_\alpha) \leq c_2 \|Af - Af_\alpha\|_{\mathbb{Y}}^{\frac{2\nu-1}{\nu}}.$$

Solving for $\|Af - Af_\alpha\|_{\mathbb{Y}}$ yields (i) with $c_1 = c_2^\nu$. $\square$

Theorem 4.33 is dedicated to true solutions contained in the domain of the penalty and to image space approximation rates $\|Af - Af_\alpha\|_{\mathbb{Y}} = \mathcal{O}(\alpha^\nu)$ with $\nu > \frac{1}{2}$. In the following we investigate low order image approximation rates that occur in the oversmoothing situation where $f \notin \mathrm{dom}(\mathcal{R})$. We will further motivate and analyze oversmoothing in the next chapter. Here, we show that approximation rates $\|Af - Af_\alpha\|_{\mathbb{Y}} = \mathcal{O}(\alpha^\nu)$ with $\nu < \frac{1}{2}$ are characterized by upper bounds on the minimal value $\vartheta_{Af}(\alpha)$ in terms of α. Whereas it is easy to see that upper bounds on ϑ_{Af} imply bounds on $\|Af - Af_\alpha\|_{\mathbb{Y}}$ the converse relies on the differentiability of ϑ_{Af}.

Theorem 4.34. *Suppose Assumption 2.2 holds true with $F = A$ linear, $\mathrm{dom}(A) = \mathbb{X}$ and $\mathbb{Y}$ a Hilbert space. Let $\nu \in [0, \frac{1}{2})$, $f \in \mathbb{X}$ and $f_\alpha \in S_\alpha(Af)$ (see (4.16)) for $\alpha > 0$ any selection of a minimizers for exact data. Then the following statements are equivalent:*

(i) There exists a constant $c_1 > 0$ such that $\|Af - Af_\alpha\|_{\mathbb{Y}} \le c_1 \alpha^\nu$ for all $\alpha > 0$.

(ii) There exists a constant $c_2 > 0$ such that the minimal value (see Definition 4.22) satisfies $\vartheta_{Af}(\alpha) - \inf_{h \in \mathbb{X}} \mathcal{R}(h) \le c_2 \alpha^{2\nu - 1}$ for all $\alpha > 0$.

More precisely, we show that (i) implies (ii) with $c_2 = \frac{c_1^2}{2(1-2\nu)}$ and (ii) implies (i) with $c_1 = \sqrt{2c_2}$.

Proof.

(i)⇒(ii): Suppose (i) holds true and let $0 < \alpha < \varepsilon$. We use Proposition 4.24 and the fundamental theorem of calculus to obtain

$$
\begin{aligned}
\vartheta_{Af}(\alpha) - \vartheta_{Af}(\varepsilon) &= \int_\alpha^\varepsilon \frac{1}{2\beta^2} \|Af - Af_\beta\|_{\mathbb{Y}}^2 \, \mathrm{d}\beta \\
&\le \frac{c_1^2}{2} \int_\alpha^\varepsilon \beta^{2\nu - 2} \, \mathrm{d}\beta \\
&= \frac{c_1^2}{2(1-2\nu)} \left(\alpha^{2\nu - 1} - \varepsilon^{2\nu - 1} \right).
\end{aligned}
$$

By Lemma 4.28 taking the limit $\varepsilon \to \infty$ yields (ii) with $c_2 = \frac{c_1^2}{2(1-2\nu)}$.

(ii)⇒(i): We estimate

$$
\frac{1}{2\alpha} \|Af - Af_\alpha\|_{\mathbb{Y}}^2 \le \frac{1}{2\alpha} \|Af - Af_\alpha\|_{\mathbb{Y}}^2 + \mathcal{R}(f_\alpha) - \inf_{h \in \mathbb{X}} \mathcal{R}(h) = \vartheta_{Af}(\alpha) - \inf_{h \in \mathbb{X}} \mathcal{R}(h). \quad (4.28)
$$

Therefore, condition (ii) implies (i) with $c_1 = \sqrt{2c_2}$. $\qquad \square$

Remark 4.35 (The limiting case $\nu = \frac{1}{2}$)**.** *Neither Theorem 4.33 nor Theorem 4.34 provides a characterization of the image space approximation rate $\|Af - Af_\alpha\|_{\mathbb{Y}} = \mathcal{O}\left(\alpha^{\frac{1}{2}}\right)$. As in Remarks 3.23 and 3.60 we see that the condition $f \in \mathrm{dom}(\mathcal{R})$ is sufficient for the latter approximation rate as $T_\alpha(f_\alpha, Af) \le T_\alpha(f, Af)$ implies $\|Af - Af_\alpha\|_{\mathbb{Y}} \le c\alpha^{\frac{1}{2}}$ for $c = (2\mathcal{R}(f) - 2\inf_{h \in \mathbb{X}} \mathcal{R}(h))^{\frac{1}{2}}$. Due to the identity (4.26) we cannot expect that $f \in \mathrm{dom}(\mathcal{R})$ is necessary for the latter approximation rate. Additionally we provide a concrete example where $f \notin \mathrm{dom}(\mathcal{R})$ but anyway satisfies the approximation rate $\|Af - Af_\alpha\|_{\mathbb{Y}} = \mathcal{O}\left(\alpha^{1/2}\right)$: Let $\mathbb{X} = \mathbb{Y} = \ell^2(\mathbb{N})$, A the identity map $\ell^2 \to \ell^2$, $\mathcal{R}(x) = \|x\|_1$ and $f = (f_n)_{n \in \mathbb{N}} \in \ell^2$ given*

by $f_n = \frac{1}{n}$. Then $f \notin \mathrm{dom}(\mathcal{R})$. As in the last chapter the minimizers f_α are given by soft thresholding, i.e. $(f_\alpha)_n = \mathrm{sgn}(f_n) \cdot \max\left(0, |f_n| - \alpha\right)$. We see

$$\|f - f_\alpha\|_2^2 = \alpha^2 \# \left\{ n < \frac{1}{\alpha} \right\} + \sum_{n=1}^{\infty} \frac{1}{n^2} \mathbb{1}_{\{n \geq \frac{1}{\alpha}\}}.$$

The first summand is bounded by α and for $\alpha \leq \frac{1}{2}$ we estimate the second by an integral comparison

$$\sum_{n=1}^{\infty} \frac{1}{n^2} \mathbb{1}_{\{n \geq \frac{1}{\alpha}\}} \leq \int_{\frac{1}{\alpha}-1}^{\infty} \frac{1}{x^2} \, \mathrm{d}x \leq \frac{\alpha}{1-\alpha} \leq 2\alpha.$$

Moreover, for $\alpha > \frac{1}{2}$ we have $\|f - f_\alpha\|_2 \leq \|f\|_2 \leq \sqrt{2}\|f\|_2 \alpha^{\frac{1}{2}}$. Therefore, we have a bound $\|f - f_\alpha\|_2 = \mathcal{O}\left(\alpha^{1/2}\right)$ for all $\alpha > 0$.

Note that a variational source condition always implies $f \in \mathrm{dom}(\mathcal{R})$ and that condition (ii) in Theorem 4.34 for $\nu = \frac{1}{2}$ yields boundedness of ϑ_{Af} which entails that there exists $\tilde{f} \in \mathrm{dom}(\mathcal{R})$ with $Af = A\tilde{f}$. This shows that neither variational source conditions nor bounds on the minimal value are appropriate to give a characterization of the Hölder type image space precision with exponent $\frac{1}{2}$.

Converse Result for approximation rates with respect to the Bregman distance

We finish this section with a further equivalence result: A variational source condition with vanishing loss function is necessary and sufficient for approximation rates measured in the skewed Bregman distance of $\mathcal{R}$. This result, even for Banach space valued forward operators, can be found in [44]. Moreover, in [38] it is shown that convergence rates with skewed Bregman loss for deterministic noise under a fixed parameter choice rule implies a variational source condition with vanishing loss function. Here we illustrate how such a result can be obtained from the theory developed above.

Let us first recall the *Bregman distance*. Let $\mathcal{R}\colon \mathbb{X} \to (-\infty, \infty]$ be as in Assumption 2.2. Then for $h \in \mathrm{dom}(\mathcal{R})$ such that $\partial\mathcal{R}(h) \neq \emptyset$ we pick $\xi \in \partial\mathcal{R}(h)$ and define

$$\Delta_\xi(f, h) := \mathcal{R}(f) - \mathcal{R}(h) - \langle \xi, f - h \rangle \quad \text{for all } f \in \mathbb{X}.$$

In [72] the *skewed Bregman distance* is used to measure the reconstruction error. In the case of exact data $g = F(f)$ for some $f \in \mathbb{X}$ the latter is given by $\Delta_{\xi_\alpha}(f, f_\alpha)$ with $f_\alpha \in S_\alpha(g)$ and $\xi_\alpha := \frac{1}{\alpha} A^*(Af - Af_\alpha) \in \partial\mathcal{R}(f_\alpha)$ due to the first order optimality condition.

Theorem 4.36. *Suppose Assumption 2.2 holds true with $F = A$ linear, $\mathrm{dom}(A) = \mathbb{X}$ and $\mathbb{Y}$ a Hilbert space. Let $f \in \mathrm{dom}(\mathcal{R})$ and $f_\alpha \in S_\alpha(g)$, $\alpha > 0$ (see (4.16)). Then the following statements hold true:*

(a) If

$$\mathcal{R}(f) - \mathcal{R}(h) \leq \varphi\left(\|Af - Ah\|_{\mathbb{Y}}^2\right) \quad \text{for all } h \in \mathbb{X} \tag{4.29}$$

with concave and continuous $\varphi\colon [0, \infty) \to [0, \infty)$, then $\Delta_{\xi_\alpha}(f, f_\alpha) \leq \breve{\varphi}(\alpha)$ for all $\alpha > 0$.

(b) *Let* $\psi \colon (0, \infty) \to [0, \infty)$ *be continuous such there exists a constant $C > 0$ with*

$$\int_\alpha^\infty \frac{1}{\beta^2} \psi(\beta) \, \mathrm{d}\beta \leq C \frac{1}{\alpha} \psi(\alpha) \quad \textit{for all } \alpha > 0. \tag{4.30}$$

If $\Delta_{\xi_\alpha}(f, f_\alpha) \leq \psi(\alpha)$ *for all $\alpha > 0$, then there exists $\varphi : [0, \infty) \to [0, \infty)$ concave and continuous such that (4.29) holds true and $\breve{\varphi}(\alpha) \leq C\psi(\alpha)$ for all $\alpha > 0$.*

Proof. With σ_f as defined in Definition 4.9 we have $\sigma'_f(\alpha) = \frac{1}{2\alpha^2} \|Af - Af_\alpha\|_{\mathbb{Y}}^2$ (see (4.25)). Therefore, the skewed Bregman distance has the following representation:

$$\Delta_{\xi_\alpha}(f, f_\alpha) = \mathcal{R}(x) - \mathcal{R}(x_\alpha) - \frac{1}{\alpha} \|Ax - Ax_\alpha\|^2 = \sigma_f(\alpha) - \alpha\sigma'_f(\alpha). \tag{4.31}$$

Here the first identity is due to $\langle \xi_\alpha, f - f_\alpha \rangle = \frac{1}{\alpha} \|Ax - Ax_\alpha\|^2$. With this we prove both statements separately.

(a) Due to (4.31) and Proposition 4.13 we have $\Delta_{\xi_\alpha}(f, f_\alpha) \leq \sigma_f(\alpha) \leq \breve{\varphi}(\alpha)$.

(b) Due to Proposition 4.15 it suffices to show

$$\sigma_f(\alpha) \leq C\psi(\alpha) \quad \text{for all } \alpha > 0 \tag{4.32}$$

as then $\breve{\varphi}_f(\alpha) = \sigma_f(\alpha) \leq C\psi(\alpha)$.

We define $\gamma(\beta) := -\frac{1}{\beta}\sigma_f(\beta)$ and use (4.31) to obtain

$$\gamma'(\beta) = \frac{1}{\beta^2} \left(\sigma_f(\beta) - \alpha\sigma'_f(\beta) \right) = \frac{1}{\beta^2} \Delta_{\xi_\beta}(f, f_\beta) \leq \frac{1}{\beta^2} \psi(\beta).$$

Since $\sigma_f(\beta) = \mathcal{R}(f) - \vartheta_{Af}(\beta)$ with the minimal value function ϑ_{Af} from Definition 4.22 we have $\lim_{\beta \to \infty} \sigma_f(\beta) = \mathcal{R}(f) - \inf_{h \in \mathbb{X}} \mathcal{R}(h) < \infty$ (see Lemma 4.28). Hence $\lim_{\beta \to \infty} \gamma(\beta) = 0$ and with (4.30) we conclude

$$\frac{\sigma_f(\alpha)}{\alpha} = \int_\alpha^\infty \gamma'(\beta) \, \mathrm{d}\beta \leq \int_\alpha^\infty \frac{1}{\beta^2} \psi(\beta) \, \mathrm{d}\beta \leq C \frac{1}{\alpha} \psi(\alpha).$$

Multiplying with α yields (4.32). $\qquad\qquad \square$

Remark 4.37 (Hölder type bounds). *The condition on ψ in Theorem 4.36(b) is valid for Hölder type functions $\psi(\alpha) := c_\psi \alpha^\nu$ with $c_\psi > 0$ and $\nu \in [0, 1)$. More precisely, with $C := \frac{1}{1-\nu}$ we obtain equality in (4.30) in this case.*
It is violated for $\psi(\alpha) := c_\psi \alpha$. The following example shows that the assumption on ψ is not superfluous: Consider $\mathbb{X} = \mathbb{Y} = \ell^2$, $\mathcal{R} = \|\cdot\|_1$ and $A = \mathrm{Id}$. If $f \in \ell^1$ then with the explicit expression for f_α given in Remark 4.35 an elementary calculation results in

$$\Delta_{\xi_\alpha}(f_\alpha, f) = \sum_{|f_k| \leq \alpha} \left(|f_k| - \frac{1}{\alpha} |f_k|^2 \right). \tag{4.33}$$

Now suppose f is finitely supported. The function $\psi(\alpha) := \Delta_{\xi_\alpha}(f_\alpha, f)$ is continuous and we have $\psi(\alpha) = 0$ for all $\alpha \leq \min\{|f_k| \colon f_k \neq 0\}$.
But on the other hand, assuming a variational source condition (4.29) holds true, then a further computation and Proposition 4.13 yield

$$\breve{\varphi}(\alpha) \geq \sigma_f(\alpha) = \sum_{|f_k| \leq \alpha} \left(|f_k| - \frac{1}{2\alpha} |f_k|^2 \right) + \frac{\alpha}{2} \#\{|f_k| > \alpha\} > 0$$

for all $\alpha > 0$. This excludes the existence of a constant $C > 0$ with $\breve{\varphi}(\alpha) \leq C\psi(\alpha)$. Hence the approximation rate $\Delta_{\xi_\alpha}(x_\alpha, x) = \mathcal{O}(\psi(\alpha))$ is not obtainable by a variational source condition in this example.

Remark 4.38 (Bounds of logarithmic type)**.** *We verify* (4.30) *for a logarithmic function of the form $\psi(\alpha) := \left(\ln\left(e + \frac{1}{\alpha}\right)\right)^{-\nu}$ for $\nu > 0$. To this end, first note that ψ is increasing and maps $(0, \infty)$ bijective onto $(0, 1)$. Moreover, strict concavity yields that $\alpha \mapsto \frac{1}{\alpha}\psi(\alpha)$ is decreasing.*
Let $c \in (0, 1)$ such that $\nu c^{\frac{1}{\nu}} \leq \frac{1}{2}$ and $\alpha_0 := \psi^{-1}(c)$. For $\alpha \geq \alpha_0$ we obtain

$$\int_\alpha^\infty \frac{1}{\beta^2}\psi(\beta)\,\mathrm{d}\beta \leq \int_\alpha^\infty \frac{1}{\beta^2}\,\mathrm{d}\beta = \frac{1}{\alpha} = c^{-1}\frac{1}{\alpha}\psi(\alpha_0) \leq c^{-1}\frac{1}{\alpha}\psi(\alpha). \qquad (4.34)$$

Therefore, in the following we assume $\alpha < \alpha_0$. We estimate

$$\psi'(\alpha) = \frac{\nu}{2\alpha^2 + \alpha}\psi(\alpha)^{\frac{\nu+1}{\nu}} \leq \frac{\nu c^{\frac{1}{\nu}}}{\alpha}\psi(\alpha) \leq \frac{1}{2\alpha}\psi(\alpha).$$

Let $\eta(\alpha) := -\frac{1}{\alpha}\psi(\alpha)$. Then we have

$$\eta'(\alpha) = \frac{1}{\alpha}\left(\frac{1}{\alpha}\psi(\alpha) - \psi'(\alpha)\right) \geq \frac{1}{2\alpha^2}\psi(\alpha)$$

and conclude

$$\frac{1}{\alpha}\psi(\alpha) - \frac{1}{\alpha_0}\psi(\alpha_0) = \int_\alpha^{\alpha_0} \eta'(\beta)\,\mathrm{d}\beta \geq \frac{1}{2}\int_\alpha^{\alpha_0} \frac{1}{\beta^2}\psi(\beta)\,\mathrm{d}\beta.$$

Together with (4.34) *and $\frac{1}{\alpha_0}\psi(\alpha_0) \leq \frac{1}{\alpha}\psi(\alpha)$ and by splitting the integral $\int_\alpha^\infty = \int_\alpha^{\alpha_0} + \int_{\alpha_0}^\infty$ we finish up with*

$$\int_\alpha^\infty \frac{1}{\beta^2}\psi(\beta)\,\mathrm{d}\beta \leq \frac{2}{\alpha}\psi(\alpha) + (c^{-1} - 2)\frac{1}{\alpha_0}\psi(\alpha_0) \leq \max(2, c^{-1})\frac{1}{\alpha}\psi(\alpha).$$

Chapter Five

Convergence rates for oversmoothing Banach space regularization

In this chapter we provide convergence rates results in the case of oversmoothing, i.e. we show upper bounds on the reconstruction error if the unknown solution $f \in \mathbb{X}$ fails to admit a finite value in the penalty term: $\mathcal{R}(f) = \infty$. For wavelet regularization with $b_{p,q}^r$-norm penalties as in (2.17), this case may be rather unlikely for $r = 0$, except maybe for delta peaks. However, for example in case of the Radon transform and the particularly interesting choice $p = 1$, our theory requires to choose $r \geq \frac{1}{2}$ (see Example 2.18), and more generally, mildly ill-posed problems in higher spatial dimensions require larger values of r (see Proposition 2.16). In the situation of a strong penalty term, i.e. a choice $r > 0$, it becomes much more likely that the penalty term fails to be finite at the exact solution. Therefore, it is desirable to derive error bounds also for this situation.

So far, however, this case has only rarely been considered in variational regularization theory. In the oversmoothing case one difficulty is that neither variational source conditions nor source conditions based on the range of the adjoint operator are applicable. Whereas oversmoothing in Hilbert scales has been analyzed in numerous papers (see, e.g., [60, 62, 86]), the literature on oversmoothing for more general variational regularization is sparse. The special case of diagonal operators in ℓ^1-regularization has been discussed in [46]. In a very recent work, which has been drafted in parallel, Chen, Hofmann & Yousept [17] have studied oversmoothing for finitely smoothing operators in scales of Banach spaces generated by sectorial operators. Although the aims and the type of results of this paper are similar to ones presented in this chapter we believe that our approach is considerably simpler and more flexible. In particular, we can also treat BV-regularization.

This section is structured as follows: First we introduce the setup of oversmoothing in Banach space regularization and formulate abstract assumptions we need on the occurring spaces. As in most of the scenarios in this work we suppose that the forward operator satisfies a two-sided Lipschitz condition.

The main result of the second section is a convergence rate result in the abstract setting. The analysis relies on spaces given by real interpolation, part of which also provide the smoothness assumptions.

In a third section we apply our theory to regularization of finitely smoothing operators with $B_{p,q}^r$-norm penalties for $r > 0$. Here the results append smoothness assumptions given by Besov smoothness s with $s \leq r$ to the convergence results in Chapter 3. As above, we numerically confirm our result in this setup using the nonlinear parameter identification problem in (3.40).

Finally, we prove a convergence rate result for BV-regularization. In particular, this last

example shows the wide applicability of our theory.

5.1   Framework and strategy

The following mathematical framework requires four Banach spaces:

- $\mathbb{X}_{\mathcal{R}}$ provides the penalty term given by $\mathcal{R}(f) = \frac{1}{u}\|f\|_{\mathbb{X}_{\mathcal{R}}}^u$ for some fixed $u \in [1, \infty)$,
- $\mathbb{X}_-$ carries a weaker norm, and we will assume that the forward map satisfies a two-sided Lipschitz condition with respect to the norm on $\mathbb{X}_-$,
- the image space $\mathbb{Y}$ is the range of the possibly nonlinear forward map,
- the reconstruction error is measured in the norm on $\mathbb{X}_L$ (the subscript L stands for 'loss').

The minus sign in $\mathbb{X}_-$ indicates that $\mathbb{X}_-$ may be a function space with negative smoothness index. For wavelet regularization in Besov spaces as considered in Section 2.4 we have $\mathbb{X}_{\mathcal{R}} = B_{p,q}^r$ with $r > 0$ and a norm given by wavelet coefficients, $X_- = B_{2,2}^{-a}$ and $\mathbb{Y}$ a Hilbert space, typically an L^2-space and $\mathbb{X}_L = B_{\bar{p},\bar{q}}^0$.

The following assumption collects the precise requirements on these spaces, the forward operator and their interaction.

Assumption 5.1. *Suppose $\mathbb{X}_{\mathcal{R}}, \mathbb{Y}$ are a Banach spaces, $\tilde{D}_F \subset \mathbb{X}_{\mathcal{R}}$, $F \colon \tilde{D}_F \to \mathbb{Y}$ a map, and $\mathcal{R}(f) = \frac{1}{u}\|f\|_{\mathbb{X}_{\mathcal{R}}}^u$ for $u \in [1, \infty)$. Moreover, we assume that $\mathbb{X}_{\mathcal{R}}$ continuously embeds into a Banach space $\mathbb{X}_-$ with*

$$\frac{1}{M_1}\|f_1 - f_2\|_{\mathbb{X}_-} \le \|F(f_1) - F(f_2)\|_{\mathbb{Y}} \le M_2\|f_1 - f_2\|_{\mathbb{X}_-} \quad \text{for all } f_1, f_2 \in \tilde{D}_F$$

for some constants $M_1, M_2 > 0$. Finally, let $\xi \in [0, 1)$. If $\xi \in (0, 1)$, let $\mathbb{X}_L$ be a Banach space and suppose that there exists a continuous embedding

$$(\mathbb{X}_-, \mathbb{X}_{\mathcal{R}})_{\xi,1} \subset \mathbb{X}_L.$$

If $\xi = 0$, we set $\mathbb{X}_L := \mathbb{X}_-$.

Note that under Assumption 5.1 F has a unique continuous extension to the norm closure of $\tilde{D}_F$ in $\mathbb{X}_-$ which we denote by $\mathrm{dom}(F)$.

For an observation $g \in \mathbb{Y}$ we consider

$$R_\alpha(g) := \underset{h \in \tilde{D}_F}{\mathrm{argmin}} \left(\frac{1}{2\alpha}\|g - F(h)\|_{\mathbb{Y}}^2 + \frac{1}{u}\|h\|_{\mathbb{X}_{\mathcal{R}}}^u \right) \subset \tilde{D}_F. \tag{5.1}$$

5.2   Error bounds

An important ingredient of most proofs in variational convergence rate theory is a comparison of the Tikhonov functional at the minimizer and at the exact solution. In the oversmoothing case such a comparison is obviously not useful. As a substitute, one may use a family of approximations of the unknown at which the penalty functional is finite. See also [60] and [62] where this idea is used and the approximations are called auxiliary elements. We start with a lemma that introduces auxiliary elements based on real interpolation theory and provides their approximation properties we need later.

Lemma 5.2 (Auxiliary elements). *Suppose Assumption 5.1 holds true. Let $\theta \in (\xi, 1]$ and $\varrho > 0$. Suppose $f \in (\mathbb{X}_-, \mathbb{X}_\mathcal{R})_{\theta,\infty}$ with $\|f\|_{(\mathbb{X}_-, \mathbb{X}_\mathcal{R})_{\theta,\infty}} \leq \varrho$. Then there exists a net $(f_t)_{t>0} \subset \mathbb{X}_\mathcal{R}$ such that the following bounds hold true*

$$\|f - f_t\|_{\mathbb{X}_-} \leq 2\varrho t^\theta \tag{5.2}$$

$$\|f - f_t\|_{\mathbb{X}_L} \leq C_L \varrho t^{\theta - \xi} \tag{5.3}$$

$$\|f_t\|_{\mathbb{X}_\mathcal{R}} \leq 2\varrho t^{\theta - 1}. \tag{5.4}$$

Here $C_L > 0$ denotes a constant that is independent of ϱ, t and f.

Proof. Recall that $\|f\|_{(\mathbb{X}_-, \mathbb{X}_\mathcal{R})_{\theta,\infty}} \leq \varrho$ implies $K(t, f) \leq \varrho t^\theta$ for $t > 0$ with the K-functional

$$K(t, f) = \inf_{h \in \mathbb{X}_\mathcal{R}} \left(\|f - h\|_{\mathbb{X}_-} + t\|h\|_{\mathbb{X}_\mathcal{R}} \right)$$

(see (A.1)). Hence, for every $t > 0$ there exists $f_t \in \mathbb{X}_\mathcal{R}$ such that

$$\|f - f_t\|_{\mathbb{X}_-} + t\|f_t\|_{\mathbb{X}_\mathcal{R}} \leq 2K(t, f) \leq 2\varrho t^\theta.$$

We neglect the first summand on the left hand side to see (5.4) and the second to obtain (5.2). This finishes the proof for $\xi = 0$ and we now turn to the case $\xi \in (0, 1)$.

As an intermediate step to (5.3) we first prove that $\|f - f_t\|_{(\mathbb{X}_-, \mathbb{X}_\mathcal{R})_{\theta,\infty}} \leq 3\varrho$ for all $t > 0$. To this end, we first consider $s \geq t$ and insert $h = 0$ into the K-functional to wind up with

$$K(s, f - f_t) = \inf_{h \in \mathbb{X}_\mathcal{R}} \left(\|f - f_t - h\|_{\mathbb{X}_-} + s\|h\|_{\mathbb{X}_\mathcal{R}} \right) \leq \|f - f_t\|_{\mathbb{X}_-} \leq 2\varrho t^\theta \leq 2\varrho s^\theta.$$

For $s \leq t$ we substitute $h = h' - f_t$ and use the triangle inequality in $\mathbb{X}_\mathcal{R}$ to estimate

$$K(s, f - f_t) = \inf_{h' \in \mathbb{X}_\mathcal{R}} \left(\|f - h'\|_{\mathbb{X}_-} + s\|h' - f_t\|_{\mathbb{X}_\mathcal{R}} \right)$$
$$\leq K(s, f) + s\|f_t\|_{\mathbb{X}_\mathcal{R}} \leq \varrho s^\theta + 2\varrho s t^{\theta - 1} \leq 3\varrho s^\theta.$$

From the last two inequalities we conclude.

$$\|f - f_t\|_{(\mathbb{X}_-, \mathbb{X}_\mathcal{R})_{\theta,\infty}} = \sup_{s > 0} s^{-\theta} K(f - f_t, s) \leq 3\varrho.$$

By the reiteration theorem (see Proposition A.6) we have

$$\mathbb{X}_L \supset (\mathbb{X}_-, \mathbb{X}_\mathcal{R})_{\xi,1} = \left(\mathbb{X}_-, (\mathbb{X}_-, \mathbb{X}_\mathcal{R})_{\theta,\infty} \right)_{\frac{\xi}{\theta}, 1}$$

with equivalent norms of the latter two spaces. Hence, Lemma A.4 provides an interpolation inequality $\| \cdot \|_{\mathbb{X}_L} \leq c \| \cdot \|_{\mathbb{X}_L}^{1 - \frac{\xi}{\theta}} \cdot \| \cdot \|_{(\mathbb{X}_-, \mathbb{X}_\mathcal{R})_{\theta,\infty}}^{\frac{\xi}{\theta}}$. Inserting $f - f_t$ we finalize

$$\|f - f_t\|_{\mathbb{X}_L} \leq c \left(2\varrho t^\theta \right)^{1 - \frac{\xi}{\theta}} (3\varrho)^{\frac{\xi}{\theta}} \leq 3c\varrho t^{\theta - \xi}. \qquad \square$$

With the auxiliary elements at hand we are in position to prove the main result of this section.

Theorem 5.3 (Error bounds). *Suppose Assumption 5.1 holds true. Let $\theta \in (\xi, 1]$ and $\varrho > 0$. Assume that $f \in (\mathbb{X}_-, \mathbb{X}_\mathcal{R})_{\theta,\infty}$ with $\|f\|_{(\mathbb{X}_-, \mathbb{X}_\mathcal{R})_{\theta,\infty}} \leq \varrho$ and moreover that $\mathrm{dom}(F)$ contains an $\mathbb{X}_L$-ball with radius $\tau > 0$ around f.*

(a) (Bias bounds) There exits a constant C_b independent of f, ϱ and τ such that

$$\|f - f_\alpha\|_{\mathbb{X}_-} \leq C_b \varrho^{\frac{u}{(1-\theta)u+2\theta}} \alpha^{\frac{\theta}{(1-\theta)u+2\theta}}, \tag{5.5}$$

$$\|f - f_\alpha\|_{\mathbb{X}_L} \leq C_b \varrho^{\frac{(1-\xi)u+2\xi}{(1-\theta)u+2\theta}} \alpha^{\frac{\theta-\xi}{(1-\theta)u+2\theta}} \quad and$$

$$\|f_\alpha\|_{\mathbb{X}_\mathcal{R}} \leq C_b \varrho^{\frac{2}{(1-\theta)u+2\theta}} \alpha^{\frac{\theta-1}{(1-\theta)u+2\theta}} \tag{5.6}$$

holds true for all $0 < \alpha < \varrho^{-\frac{(1-\theta)u+2\xi}{\theta-\xi}} \tau^{\frac{(1-\theta)u+2\theta}{\theta-\xi}}$ and $f_\alpha \in R_\alpha(F(f))$ (see (5.1)).

(b) (Rates with a priori choice) Let $0 < c_l \leq c_r$. Suppose $0 < \delta < \varrho^{-\frac{\xi}{\theta-\xi}} \tau^{\frac{\theta}{\theta-\xi}}$, $g^{\mathrm{obs}} \in \mathbb{Y}$ with $\|g^{\mathrm{obs}} - F(f)\|_\mathbb{Y} \leq \delta$. Let $\alpha > 0$ and $\hat{f}_\alpha \in R_\alpha(g^{\mathrm{obs}})$. There exists a constant C_a independent of $f, g^{\mathrm{obs}}, \varrho, \tau$ and δ such that

$$c_l \varrho^{-\frac{u}{\theta}} \delta^{\frac{(1-\theta)u+2\theta}{\theta}} \leq \alpha \leq c_r \varrho^{-\frac{u}{\theta}} \delta^{\frac{(1-\theta)u+2\theta}{\theta}}$$

implies the following bounds

$$\|f - \hat{f}_\alpha\|_{\mathbb{X}_-} \leq C_a \delta,$$

$$\|f - \hat{f}_\alpha\|_{\mathbb{X}_L} \leq C_a \varrho^{\frac{\xi}{\theta}} \delta^{\frac{\theta-\xi}{\theta}} \quad and$$

$$\|\hat{f}_\alpha\|_{\mathbb{X}_\mathcal{R}} \leq C_a \varrho^{\frac{1}{\theta}} \delta^{\frac{\theta-1}{\theta}}.$$

(c) (Rates with discrepancy principle) Let $1 < c_D \leq C_D$. Suppose $0 < \delta < \varrho^{-\frac{\xi}{\theta-\xi}} \tau^{\frac{\theta}{\theta-\xi}}$, $g^{\mathrm{obs}} \in \mathbb{Y}$ with $\|g^{\mathrm{obs}} - F(f)\|_\mathbb{Y} \leq \delta$. Let $\alpha > 0$ and $\hat{f}_\alpha \in R_\alpha(g^{\mathrm{obs}})$. There exists a constant C_d independent of $f, g^{\mathrm{obs}}, \varrho$ and δ such that

$$c_D \delta \leq \|g^{\mathrm{obs}} - F(\hat{f}_\alpha)\|_\mathbb{Y} \leq C_D \delta$$

implies the following bounds

$$\|f - \hat{f}_\alpha\|_{\mathbb{X}_L} \leq C_d \varrho^{\frac{\xi}{\theta}} \delta^{\frac{\theta-\xi}{\theta}} \quad and$$

$$\|\hat{f}_\alpha\|_{\mathbb{X}_\mathcal{R}} \leq C_d \varrho^{\frac{1}{\theta}} \delta^{\frac{\theta-1}{\theta}}.$$

Proof. Let $(f_t)_{t>0}$ be as in Lemma 5.2.

(a) We choose

$$t = C_L^{-\frac{1}{\theta-\xi}} \varrho^{\frac{u-2}{(1-\theta)u+2\theta}} \alpha^{\frac{1}{(1-\theta)u+2\theta}}$$

with C_L from Lemma 5.2. Inequality (5.3) yields

$$\|f - f_t\|_{\mathbb{X}_L} \leq C_L \varrho t^{\theta-\xi} = \varrho^{\frac{(1-\xi)u+2\xi}{(1-\theta)u+2\theta}} \alpha^{\frac{\theta-\xi}{(1-\theta)u+2\theta}} < \tau. \tag{5.7}$$

Hence $f_t \in \mathrm{dom}(F)$, i.e. we may insert f_t into the Tikhonov functional and use the Lipschitz condition of F, (5.2) and (5.4) to wind up with

$$
\frac{1}{2\alpha}\|F(f) - F(f_\alpha)\|_{\mathbb{Y}}^2 + \frac{1}{u}\|f_\alpha\|_{\mathbb{X}_{\mathcal{R}}}^u \leq \frac{1}{2\alpha}\|F(f) - F(f_t)\|_{\mathbb{Y}}^2 + \frac{1}{u}\|f_t\|_{\mathbb{X}_{\mathcal{R}}}^u
$$
$$
\leq \frac{M_2^2}{2\alpha}\|f - f_t\|_{\mathbb{X}_-}^2 + \frac{1}{u}\|f_t\|_{\mathbb{X}_{\mathcal{R}}}^u
$$
$$
\leq \frac{2M_2^2}{\alpha}\varrho^2 t^{2\theta} + \frac{2^u}{u}\varrho^u t^{(\theta-1)u}
$$
$$
= c_1 \varrho^{\frac{2u}{(1-\theta)u+2\theta}} \alpha^{\frac{(\theta-1)u}{(1-\theta)u+2\theta}}
$$

with $c_1 := 2M_2^2 C_L^{-\frac{2\theta}{\theta-\xi}} + \frac{2^u}{u} C_L^{\frac{(1-\theta)u}{\theta-\xi}}$. We neglect the penalty term and use the Lipschitz condition of the inverse of F to obtain the first bound

$$
\|f - f_\alpha\|_{\mathbb{X}_-} \leq M_1\|F(f) - F(f_\alpha)\|_{\mathbb{Y}} \leq (2c_1)^{\frac{1}{2}} M_1 \varrho^{\frac{u}{(1-\theta)u+2\theta}} \alpha^{\frac{\theta}{(1-\theta)u+2\theta}}.
$$

Together with (5.2) we record

$$
\|f_t - f_\alpha\|_{\mathbb{X}_-} \leq \|f - f_t\|_{\mathbb{X}_-} + \|f - f_\alpha\|_{\mathbb{X}_-} \leq c_2 \varrho^{\frac{u}{(1-\theta)u+2\theta}} \alpha^{\frac{\theta}{(1-\theta)u+2\theta}}
$$

with $c_2 = 2C_L^{-\frac{\theta}{\theta-\xi}} + (2c_1)^{\frac{1}{2}} M_1$.

Neglecting the data fidelity term in the above estimation of the Tikhonov functional provides

$$
\|f_\alpha\|_{\mathbb{X}_{\mathcal{R}}} \leq (c_1 u)^{\frac{1}{u}} \varrho^{\frac{2}{(1-\theta)u+2\theta}} \alpha^{\frac{\theta-1}{(1-\theta)u+2\theta}}.
$$

Furthermore, we see that $\|f_t\|_{\mathbb{X}_{\mathcal{R}}}$ satisfies the same upper bound. With the triangle inequality in $\mathbb{X}_{\mathcal{R}}$ we combine

$$
\|f_t - f_\alpha\|_{\mathbb{X}_{\mathcal{R}}} \leq \|f_t\|_{\mathbb{X}_{\mathcal{R}}} + \|f_\alpha\|_{\mathbb{X}_{\mathcal{R}}} \leq 2(c_1 u)^{\frac{1}{u}} \varrho^{\frac{2}{(1-\theta)u+2\theta}} \alpha^{\frac{\theta-1}{(1-\theta)u+2\theta}}.
$$

Next, the interpolation inequality $\|\cdot\|_{\mathbb{X}_L} \leq c_3 \|\cdot\|_{\mathbb{X}_-}^{1-\xi} \cdot \|\cdot\|_{\mathbb{X}_{\mathcal{R}}}^{\xi}$ (see Lemma A.4) furnishes

$$
\|f_t - f_\alpha\|_{\mathbb{X}_L} \leq c_4 \varrho^{\frac{(1-\xi)u+2\xi}{(1-\theta)u+2\theta}} \alpha^{\frac{\theta-\xi}{(1-\theta)u+2\theta}}
$$

with $c_4 := 2^\xi c_3 c_2^{1-\xi}(c_1 u)^{\frac{\xi}{u}}$. Together with (5.7) we finalize with

$$
\|f - f_\alpha\|_{\mathbb{X}_L} \leq \|f - f_t\|_{\mathbb{X}_L} + \|f_t - f_\alpha\|_{\mathbb{X}_L} \leq (1 + c_4)\varrho^{\frac{(1-\xi)u+2\xi}{(1-\theta)u+2\theta}} \alpha^{\frac{\theta-\xi}{(1-\theta)u+2\theta}}.
$$

(b) Taking $t = C_L^{-\frac{1}{\theta-\xi}} \varrho^{-\frac{1}{\theta}} \delta^{\frac{1}{\theta}}$ we have

$$
\|f - f_t\|_{\mathbb{X}_L} \leq C_L \varrho t^{\theta-\xi} = \varrho^{\frac{\xi}{\theta}} \delta^{\frac{\theta-\xi}{\theta}} < \tau. \tag{5.8}
$$

This ensures $f_t \in \mathrm{dom}(F)$. We insert into the Tikhonov functional, use the elementary inequality $(a+b)^2 \leq 2a^2 + 2b^2$ for $a, b \geq 0$, (5.2), (5.4), the Lipschitz condition of F and the choice of α to estimate

$$
\frac{1}{2\alpha}\|g^{\mathrm{obs}} - F(\hat{f}_\alpha)\|_{\mathbb{Y}}^2 + \frac{1}{u}\|\hat{f}_\alpha\|_{\mathbb{X}_{\mathcal{R}}}^u \leq \frac{1}{2\alpha}\|g^{\mathrm{obs}} - F(f) + F(f) - F(f_t)\|_{\mathbb{Y}}^2 + \frac{1}{u}\|f_t\|_{\mathbb{X}_{\mathcal{R}}}^u
$$
$$
\leq \frac{\delta^2}{\alpha} + \frac{M_2^2}{\alpha}\|f - f_t\|_{\mathbb{X}_-}^2 + \frac{2^u}{u}\varrho^u t^{(\theta-1)u}
$$
$$
\leq (1 + 4M_2^2 C_L^{-\frac{2\theta}{\theta-\xi}})\frac{\delta^2}{\alpha} + \frac{2^u}{u} C_L^{\frac{(1-\theta)u}{\theta-\xi}} \varrho^{\frac{u}{\theta}} \delta^{\frac{(\theta-1)u}{\theta}}
$$
$$
\leq c_5 \varrho^{\frac{u}{\theta}} \delta^{\frac{(\theta-1)u}{\theta}}
$$

with $c_5 := c_l^{-1}(1 + 4M_2^2 C_L^{-\frac{2\theta}{\theta-\xi}}) + \frac{2^u}{u} C_L^{\frac{(1-\theta)u}{\theta-\xi}}$. Now we follow the argument in (a): From the last inequality and the triangle inequality in $\mathbb{Y}$ we get

$$\begin{aligned}
\|f - \hat{f}_\alpha\|_{\mathbb{X}_-} &\leq M_1\|F(f) - g^{\mathrm{obs}} + g^{\mathrm{obs}} - F(\hat{f}_\alpha)\|_{\mathbb{Y}} \\
&\leq M_1\delta + M_1(2c_5)^{\frac{1}{2}}\alpha^{\frac{1}{2}}\varrho^{\frac{u}{2\theta}}\delta^{\frac{(\theta-1)u}{2\theta}} \\
&\leq M_1(1 + (2c_5c_r)^{\frac{1}{2}})\delta
\end{aligned}$$

which together with (5.2) implies $\|f_t - \hat{f}_\alpha\|_{\mathbb{X}_-} \leq c_6\delta$ with

$$c_6 := 2C_L^{-\frac{\theta}{\theta-\xi}} + M_1(1 + (2c_5c_r)^{\frac{1}{2}}).$$

Moreover, $\|\hat{f}_\alpha\|_{\mathbb{X}_\mathcal{R}}, \|f_t\|_{\mathbb{X}_\mathcal{R}} \leq (c_5u)^{\frac{1}{u}}\varrho^{\frac{1}{\theta}}\delta^{\frac{\theta-1}{\theta}}$. Hence $\|f_t - \hat{f}_\alpha\|_{\mathbb{X}_\mathcal{R}} \leq 2(c_5u)^{\frac{1}{u}}\varrho^{\frac{1}{\theta}}\delta^{\frac{\theta-1}{\theta}}$ by the triangle inequality in $\mathbb{X}_\mathcal{R}$.

We use the above interpolation inequality to combine the last two inequalities to $\|f_t - \hat{f}_\alpha\|_{\mathbb{X}_L} \leq c_7\varrho^{\frac{\xi}{\theta}}\delta^{\frac{\theta-\xi}{\theta}}$ with $c_7 := c_3c_6^{1-\xi}2^\xi(c_5u)^{\frac{\xi}{u}}$. With (5.8) we conclude

$$\|f - \hat{f}_\alpha\|_{\mathbb{X}_L} \leq (1 + c_7)\varrho^{\frac{\xi}{\theta}}\delta^{\frac{\theta-\xi}{\theta}}.$$

(c) We set $\varepsilon := \min\left\{\frac{c_D^2-1}{2}, 4M_2C_L^{-\frac{2\theta}{\theta-\xi}}\right\}$. Then $\varepsilon > 0$. Furthermore, we take

$$t = \left(\frac{(4M_2)^{-1}\varepsilon}{1 + \varepsilon^{-1}}\right)^{\frac{1}{2\theta}}\varrho^{-\frac{1}{\theta}}\delta^{\frac{1}{\theta}}.$$

Then (5.2) reads as

$$\|f - f_t\|_{\mathbb{X}_-} \leq 2\varrho t^\theta = \left(\frac{\varepsilon}{1 + \varepsilon^{-1}}\right)^{\frac{1}{2}}M_2^{-\frac{1}{2}}\delta. \tag{5.9}$$

Due to (5.3) we obtain

$$\|f - f_t\|_{\mathbb{X}_L} \leq C_L\varrho t^{\theta-\xi} \leq C_L\left((4M_2)^{-1}\varepsilon\right)^{\frac{\theta-\xi}{2\theta}}\varrho^{\frac{\xi}{\theta}}\delta^{\frac{\theta-\xi}{\theta}} \leq \varrho^{\frac{\xi}{\theta}}\delta^{\frac{\theta-\xi}{\theta}} < \tau \tag{5.10}$$

which provides $f_t \in \mathrm{dom}(F)$.

In the following we use the elementary inequality $(a + b)^2 \leq (1 + \varepsilon)a^2 + (1 + \varepsilon^{-1})b^2$ for all $a, b \geq 0$, which is proven by expanding the square and applying Young's inequality on the mixed term, and (5.9) to estimate

$$\begin{aligned}
\|g^{\mathrm{obs}} - F(f_t)\|_{\mathbb{Y}}^2 &\leq (1 + \varepsilon)\delta^2 + (1 + \varepsilon^{-1})\|F(f) - F(f_t)\|_{\mathbb{Y}}^2 \\
&\leq (1 + \varepsilon)\delta^2 + (1 + \varepsilon^{-1})M_2\|f - f_t\|_{\mathbb{X}_-}^2 \\
&\leq (1 + 2\varepsilon)\delta^2 \leq c_D^2\delta^2 \leq \|g^{\mathrm{obs}} - F(\hat{f}_\alpha)\|_{\mathbb{Y}}^2.
\end{aligned}$$

Therefore, a comparison of the Tikhonov functional taken at $\hat{f}_\alpha$ and f_t, and (5.4) yield

$$\|\hat{f}_\alpha\|_{\mathbb{X}_\mathcal{R}} \leq \|f_t\|_{\mathbb{X}_\mathcal{R}} \leq 2\varrho t^{\theta-1} = c_8\varrho^{\frac{1}{\theta}}\delta^{\frac{\theta-1}{\theta}} \quad \text{with } c_8 := 2\left(\frac{(4M_2)^{-1}\varepsilon}{1 + \varepsilon^{-1}}\right)^{\frac{\theta-1}{2\theta}}.$$

Hence $\|f_t - \hat{f}_\alpha\|_{\mathbb{X}_\mathcal{R}} \le 2c_8\varrho^{\frac{1}{\theta}}\delta^{\frac{\theta-1}{\theta}}$. Moreover,

$$\|F(f_t) - F(\hat{f}_\alpha)\|_{\mathbb{Y}} \le \|g^{\mathrm{obs}} - F(f_t)\|_{\mathbb{Y}} + \|g^{\mathrm{obs}} - F(\hat{f}_\alpha)\|_{\mathbb{Y}} \le 2C_D\delta.$$

Therefore, $\|f_t - \hat{f}_\alpha\|_{\mathbb{X}_-} \le M_1 C_D \delta$ by the Lipschitz condition. As above we conclude $\|f_t - \hat{f}_\alpha\|_{\mathbb{X}_L} \le c_9 \varrho^{\frac{\xi}{\theta}} \delta^{\frac{\theta-\xi}{\theta}}$ with $c_9 := c_3 (2C_D)^{1-\xi}(2c_8)^\xi$ and use (5.10) to finish up with

$$\|f - \hat{f}_\alpha\|_{\mathbb{X}_L} \le (1 + c_9)\varrho^{\frac{\xi}{\theta}} \delta^{\frac{\theta-\xi}{\theta}}. \qquad \square$$

We discuss our result in a series of remarks.

Remark 5.4 (Interior point)**.** *The requirement that f is an interior point of the domain in $\mathbb{X}_L$ may be weakened to the requirement that elements f_t satisfying the bounds given in Lemma 5.2 belong to $\mathrm{dom}(F)$ for t small enough. This may be helpful in concrete situations where there is a natural choice of f_t that maybe satisfies some helpful additional requirements like support constraints or further smoothness assumptions.*

Remark 5.5 (Influence of the exponent u)**.** *A strength of the above theorem is that it works for all exponents u independent of the space $\mathbb{X}_\mathcal{R}$. Note that the choice of u does not influence the rate of convergence while it does influence the bias bounds and the parameter choice rule. An inspection of the a priori rule shows that a larger u allows for a larger choice of the parameter α. The flexibility in the choice of u in our theory is a remarkable difference to many other variational convergence theories. For example we saw in Remark 3.8 that the theory in Chapter 3 is designed to the choice $u = 1$. Also in [104] the authors have to pick a specific exponent u in order to obtain error bounds in the norm instead of the Bregman distance.*

Remark 5.6 (Equivalent norms)**.** *The presented theory relies on a purely Banach space theoretic framework: As we do not appeal to any metric or convex notions like subdifferentials or convexity the result in Theorem 5.3 stays the same up to a change of the constants if we change the norm on any of the occurring spaces up to equivalence.*
Once again this is a major difference to classical variational regularization theory. For example it is not clear how the subdifferential of a norm involved in the classical source condition (2.4) changes if the norm is replaced by an equivalent one.
This has an important impact on wavelet regularization with Besov space penalties. In many places the analysis in Chapter 3 makes explicit use of the Besov norms given by wavelet coefficients, and we use the equivalence of the initial norm to the norm given by wavelet coefficients several times for different indices. Here we will see that it is enough to require that the wavelet system yields an equivalent norm on the particular space $\mathbb{X}_\mathcal{R} = B^r_{p,q}$ involved in the penalization.

Remark 5.7 (Limiting case $\theta = 1$)**.** *In the case $\theta = 1$ the parameter choice rule in Theorem 5.3 becomes $\alpha \sim \delta^2$. Here the results provides boundedness of the estimators f_α and $\hat{f}_\alpha$ in $\mathbb{X}_\mathcal{R}$. Due to Proposition A.5 we have $\mathbb{X}_\mathcal{R} \subset (\mathbb{X}_-, \mathbb{X}_\mathcal{R})_{1,\infty}$. The latter two spaces agree if $\mathbb{X}_\mathcal{R}$ is reflexive (see [98, 1.3.2. Rem. 2]).*
In Remark 3.23 and Remark 3.60 we have discussed similar limiting cases. Therein, we assumed that x has a finite value in the penalty term to obtain uniform boundedness of the estimators in the norm of the penalty.

Remark 5.8 (Converse result)**.** *If minimizers in* (5.1) *exist and* $\mathrm{dom}(F) = \mathbb{X}_-$, *then one can reclaim* $f \in (\mathbb{X}_-, \mathbb{X}_{\mathcal{R}})_{\theta,\infty}$ *from the bias bound* (5.5) *together with* (5.6)*. To this end, one starts with*

$$K(t, f) \leq \|f - f_\alpha\|_{\mathbb{X}_-} + t\|f_\alpha\|_{\mathbb{X}_{\mathcal{R}}}$$

for $f_\alpha \in R_\alpha(F(f))$*. Then choosing* α *balancing the two summands on the right hand side, yields the desired upper bound on the* K*-functional.*

5.3 Application to Besov space regularization

In this section we apply Theorem 5.3 to regularization of finitely smoothing operators with Besov space penalty term. To this end, let $B_{p,q}^s$ be a Besov scale in the sense of Definition 2.9. We suppose Assumption 2.15 holds true and use $\mathbb{X}_{\mathcal{R}} := B_{p,q}^r$ for some $r > 0$ and $p, q \in [1, \infty]$. We take any norm $\|\cdot\|_{\mathbb{X}_{\mathcal{R}}}$ that is equivalent to the original $\|\cdot\,|\,B_{p,q}^r\|$ and consider the regularization scheme given by

$$R_\alpha(g) = \underset{h \in B_{p,q}^r \cap \mathrm{dom}(F)}{\mathrm{argmin}} \left(\frac{1}{2\alpha}\|g - F(h)\|_{\mathbb{Y}}^2 + \frac{1}{u}\|h\|_{\mathbb{X}_{\mathcal{R}}}^u \right) \subset B_{p,q}^r \cap \mathrm{dom}(F) \qquad (5.11)$$

with $g \in \mathbb{Y}$ and a fixed exponent $u \in [1, \infty)$.

In contrast to the analysis in Chapter 3 and [104], which are restricted to certain choices of r, p, q and u, the only restriction on the parameters we assume here is $r > 0$ and $q \leq p$. Moreover, we will see that the assumption $q \leq p$ can be dropped by some refined argument using a complex interpolation identity. This illustrates the impact and flexibility of Theorem 5.3 as to the best of the authors knowledge this is the first result in Besov space regularization with this flexibility in the parameters p and q.

On the other hand, the result we present in this section is restricted to the case of oversmoothing which depending on the choice of r occurs only for functions of low regularity. In this sense the following result is complementary to the former convergence rates results in Besov spaces but not improving them.

Before we state the convergence rates we show that the setting of wavelet regularization as introduced in Section 2.4 is contained in this setting.

Remark 5.9 (Wavelet regularization)**.** *Suppose* $B_{p,q}^s \subset \tilde{\mathcal{D}}$ *is a Besov scale as in Definition 2.9 and* $(\psi_\lambda)_{\lambda \in \Lambda}$ *is a family of elements in* $\tilde{\mathcal{D}}$ *with* Λ *as in Section 2.4. Assume that* $\mathcal{S}$ *the wavelet synthesis operator as in Assumption 2.11 is a norm isomorphism* $\mathcal{S} \colon b_{p,q}^r \to B_{p,q}^r$ *with* $\mathbb{X}_{\mathcal{R}} = B_{p,q}^r$ *the particular element of the Besov space that provides the penalty term. In this case we define the norm* $\|f\|_{\mathbb{X}_{\mathcal{R}}} := \|\mathcal{S}^{-1}f\|_{r,p,q}$*. Then* $\|\cdot\|_{\mathbb{X}_{\mathcal{R}}}$ *is equivalent to* $\|\cdot\,|\,B_{p,q}^r\|$ *on* $B_{p,q}^r$*. By considering how* argmin *transforms under composition with a bijective mapping one sees that with this the estimators in* (5.11) *agree with the ones in* (2.17)*.*

We first state our theorem under the abstract smoothness condition given by the real interpolation space in Theorem 5.3 and discuss how to find more handsome smoothness conditions afterwards.

Theorem 5.10 (Error bounds for oversmoothing Besov space regularization)**.** *Suppose* $B_{p,q}^s$ *is a Besov scale, Assumption 2.15 holds true,* $q \leq p$ *and that* $B_{p,q}^r \subset B_{2,2}^{-a}$ *with a*

continuous embedding. Let $s \in (0, r]$ and $\varrho > 0$. Assume that

$$f \in \left(B_{2,2}^{-a}, B_{p,q}^{r}\right)_{\theta_s, \infty} \quad for \quad \theta_s := \frac{s+a}{a+r}$$

with $\|f\|_{\left(B_{2,2}^{-a}, B_{p,q}^{r}\right)_{\theta,\infty}} \leq \varrho$. Suppose $\|\cdot\|_{\mathbb{X}_{\mathcal{R}}}$ in (5.11) is equivalent to $\|\cdot\,|\,B_{p,q}^{r}\|$. Let $\overline{p} := \frac{2p(a+r)}{2a+pr}$ and assume that $\operatorname{dom}(F)$ contains a $B_{\overline{p},\overline{p}}^{0}$-ball with radius $\tau > 0$ around f.

(a) (Bias bounds) There exits a constant C_b independent of f, ϱ and τ such that

$$\|f - f_\alpha \,|\, B_{2,2}^{-a}\| \leq C_b \varrho^{\frac{u(a+r)}{(2-u)s+2a+ur}} \alpha^{\frac{s+a}{(2-u)s+2a+ur}},$$

$$\|f - f_\alpha \,|\, B_{\overline{p},\overline{p}}^{0}\| \leq C_b \varrho^{\frac{2a+ur}{(2-u)s+2a+ur}} \alpha^{\frac{s}{(2-u)s+2a+ur}} \quad and$$

$$\|f_\alpha \,|\, B_{p,q}^{r}\| \leq C_b \varrho^{\frac{2a+2r}{(2-u)s+2a+ur}} \alpha^{\frac{s-r}{(2-u)s+2a+ur}}$$

holds true for all $0 < \alpha < \varrho^{-\frac{u(r-s)+2a}{s}} \tau^{\frac{(2-u)s+2a+ur}{s}}$ and $f_\alpha \in R_\alpha(F(f))$ (see (5.11)).

(b) (Convergence rates) Let $0 < c_l \leq c_r$ and $1 < c_D \leq C_D$. Suppose $0 < \delta < \varrho^{-\frac{a}{s}} \tau^{\frac{s+a}{s}}$, $g^{\mathrm{obs}} \in \mathbb{Y}$ with $\|g^{\mathrm{obs}} - F(f)\|_{\mathbb{Y}} \leq \delta$. Let $\alpha > 0$ and $\hat{f}_\alpha \in R_\alpha(g^{\mathrm{obs}})$. There is a constant C_r independent of $f, g^{\mathrm{obs}}, \varrho, \tau$ and δ such that either of the conditions

$$c_l \varrho^{-\frac{u(a+r)}{s+a}} \delta^{\frac{(2-u)s+2a+ur}{s+a}} \leq \alpha \leq c_r \varrho^{-\frac{u(a+r)}{s+a}} \delta^{\frac{(2-u)s+2a+ur}{s+a}} \quad and$$

$$c_D \delta \leq \|g^{\mathrm{obs}} - F(\hat{f}_\alpha)\|_{\mathbb{Y}} \leq C_D \delta$$

implies the following bounds

$$\|f - \hat{f}_\alpha \,|\, B_{2,2}^{-a}\| \leq C_r \delta,$$

$$\|f - \hat{f}_\alpha \,|\, B_{\overline{p},\overline{p}}^{0}\| \leq C_r \varrho^{\frac{a}{s+a}} \delta^{\frac{s}{s+a}} \quad and$$

$$\|\hat{f}_\alpha \,|\, B_{p,q}^{r}\| \leq C_r \varrho^{\frac{a+r}{s+a}} \delta^{\frac{s-r}{s+a}}.$$

Proof. Let $\xi := \frac{a}{a+r}$. Since $q \leq p$ and $1 \leq \overline{p}$ we have the following chain of continuous embeddings:

$$\left(B_{2,2}^{-a}, B_{p,q}^{r}\right)_{\xi,1} \subset \left(B_{2,2}^{-a}, B_{p,q}^{r}\right)_{\xi,\overline{p}} \subset \left(B_{2,2}^{-a}, B_{p,p}^{r}\right)_{\xi,\overline{p}} = B_{\overline{p},\overline{p}}^{0} \tag{5.12}$$

(see [8, Thm. 3.4.1.(b)] for the first embedding and [101, 4.2.1. Rem. 4] for the second). Therefore, $\mathbb{X}_{\mathcal{R}} = B_{p,q}^{r}$, $\mathbb{X}_{-} := B_{2,2}^{-a}$, $\mathbb{Y}$, $\mathbb{X}_{L} := B_{\overline{p},\overline{p}}^{0}$ and $\tilde{D}_F := B_{p,1}^{r} \cap \operatorname{dom}(F)$ satisfy Assumption 5.1 and the result follows from Theorem 5.3. Here one first obtains bounds in the $\|\cdot\|_{\mathbb{X}_{\mathcal{R}}}$-norm on $B_{p,q}^{r}$ which in turn yield the stated bounds in the $B_{p,q}^{r}$-norm by equivalence of the latter two norms. □

We discuss our result in the following remarks.

Remark 5.11 ($L^{\overline{p}}$-loss)**.** *In the setting of Theorem 5.10 suppose that $p \leq 2$ and $B_{\overline{p},\overline{p}}^{0} = B_{\overline{p},\overline{p}}^{0}(\Omega)$ with Ω either $\mathbb{T}^d, \mathbb{R}^d$ or a bounded Lipschitz domain in $\mathbb{R}^d$. Then the continuous embedding $B_{\overline{p},\overline{p}}^{0} \subset L^{\overline{p}}(\Omega)$ (see Proposition A.11(a)) together with (5.12) yields $\left(B_{2,2}^{-a}, B_{p,q}^{r}\right)_{\xi,1} \subset L^{\overline{p}}(\Omega)$. Therefore, Theorem 5.10 remains true word for word if one replaces $B_{\overline{p},\overline{p}}^{0}$ by $L^{\overline{p}}(\Omega)$ in this case.*

Remark 5.12 (Smoothness condition)**.** *Suppose that both $B_{p,q}^r = B_{p,q}^r(\Omega)$ and $B_{2,2}^{-a} = B_{2,2}^{-a}(\Omega)$ with Ω either $\mathbb{T}^d, \mathbb{R}^d$ or a bounded Lipschitz domain in $\mathbb{R}^d$ and let $0 < s < r$. Then due to [101, Thm.3.3.6] the following complex interpolation result with equivalence of norms is available:*

$$\left(B_{2,2}^{-a}(\Omega), B_{p,q}^r(\Omega) \right)_{\theta_s} = B_{p_s,q_s}^s(\Omega) \text{ with } p_s = \frac{2p(a+r)}{s(2-p)+2a+pr}, \ q_s = \frac{2q(a+r)}{s(2-q)+2a+qr}.$$
(5.13)

With this we obtain a continuous embedding $B_{p_s,q_s}^s(\Omega) \subset \left(B_{2,2}^{-a}(\Omega), B_{p,q}^r(\Omega) \right)_{\theta_s,\infty}$ as the complex interpolation space $(\cdot,\cdot)_\theta$ is always continuously embedded in the real interpolation space $(\cdot,\cdot)_{\theta,\infty}$ (see [8, Thm. 4.7.1.]). Hence the statements in Theorem 5.10 remain true if the smoothness assumption on f formulated in terms of $\left(B_{2,2}^{-a}, B_{p,q}^r \right)_{\theta_s,\infty}$ is replaced by $f \in B_{p_s,q_s}^s(\Omega)$ with $\|f \,|\, B_{p_s,q_s}^s\| \leq \varrho$.
In this case the results in Section 3.4 show that the convergence rate measured in the $B_{\bar{p},\bar{p}}^0$-norm or with the $L^{\bar{p}}(\Omega)$-norm as in the last remark is minimax optimal on bounded sets in $B_{t_s,t_s}^s(\Omega)$.

Example 5.13 ($p = 2$)**.** *In the case $p = 2$ we have $\left(B_{2,2}^{-a}, B_{2,q}^r \right)_{\theta_s,\infty} = B_{2,\infty}^s$. Hence for $p = 2$ and $q = 1$ the smoothness condition is in line with the one in Theorems 3.15 and Theorem 3.37 such that Theorem 5.10 complements these results for the case $r < s$. (see also Remark 3.24).*
For a coefficient c^{jump} with finitely many jumps in the boundary value problem (3.40) we numerically confirm the convergence rate predicted by Theorem 5.10 and Remark 5.11. To this end, we recall that $c^{\mathrm{jump}} \in B_{2,\infty}^{\frac{1}{2}}((0,1))$. We use the Besov norm penalty on $B_{2,1}^2((0,1))$ given by the $b_{2,1}^2$-norm of wavelet coefficients with respect to Daubechies wavelets of order 7. As $s = \frac{1}{2}$ is the maximal smoothness index such that $c^{\mathrm{jump}} \in B_{2,\infty}^s((0,1))$ the penalty term is infinite at the true coefficient in this case. In Figure 5.1 we see a good agreement of the reconstruction error in the numerical experiment with the predicted rate $\mathcal{O}(\delta^{1/5})$ measured in the L^2-norm. A possible explanation of the jumpy behavior in the reconstruction error in this case is that our choice of noisy data may not attain the worst case at some noise levels.

Example 5.14 (The case $p = q = 1$)**.** *In the case $p = q = 1$ we have $\left(B_{2,2}^{-a}, B_{1,1}^r \right)_{\theta_s,\infty} = K_s$ with the spaces K_s introduced in Proposition 3.63. Hence Theorem 5.10 complements Theorems 3.65 and 3.71.*
As above we use our numerical example (3.40) to confirm the convergence rate in a concrete example. We use the $b_{1,1}^2$-penalty on db7 wavelet coefficients. Hence, we have $a = 2$ and $r = 2$ which yields $\bar{p} = \frac{4}{3}$ Due to Example 3.67 a true coefficient c^{jump} with finitely many jumps belongs to K_s for $s < \frac{6}{7}$. Therefore, Theorem 5.10 and Remark 5.11 predict the rate $\mathcal{O}(\delta^e)$ for all $e < \frac{3}{10}$. Here, we see a precise agreement with the reconstruction error in the numerical experiment.

Remark 5.15 (Assumption $q \leq p$)**.** *Finally, we comment on the assumption $q \leq p$. Using complex interpolation this restriction can be dropped as follows: As in Remark 5.12 we*
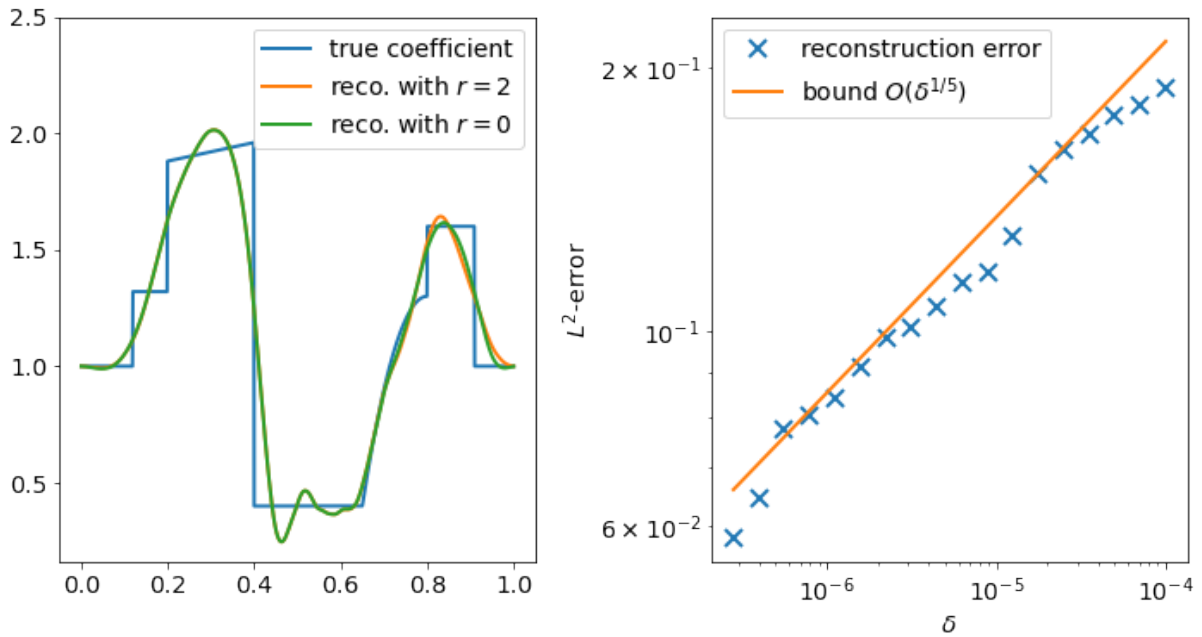
FIGURE 5.1: *Left:* true coefficient c^{jump} with jumps in the boundary value problem (3.40) together with reconstructions for $b_{2,1}^2$- and $b_{2,1}^0$-penalization at noise level $\delta = 2.5 \cdot 10^{-5}$ for the same data. *Right:* Reconstruction error using $b_{2,1}^2$-penalization, the rate $\mathcal{O}(\delta^{1/5})$ in the L^2-norm predicted by Theorem 5.10 (see Example 5.13).

assume $B_{p,q}^r = B_{p,q}^r(\Omega)$ and $B_{2,2}^{-a} = B_{2,2}^{-a}(\Omega)$. Since the real interpolation space $(\cdot, \cdot)_{\theta,1}$ is always continuously embedded in the complex interpolation space (see [8, Thm. 4.7.1.]) identity (5.13) yields a continuous embedding $\left(B_{2,2}^{-a}(\Omega), B_{p,q}^r(\Omega) \right)_{\xi,1} \subset B_{\overline{p},\overline{q}}^0(\Omega)$ for $\overline{p}$ as in Theorem 5.10 and $\overline{q} := \frac{2q(a+r)}{2a+qr}$. Hence the statements in Theorem 5.10 remain true in the case $q > p$ if one replaces $B_{\overline{p},\overline{p}}^0$ by $B_{\overline{p},\overline{q}}^0(\Omega)$.

Remark 5.16 ('Does oversmoothing hurt?'). *To conclude this section we point out a difference in the statements for $p = 2$ and $p = 1$ in Examples 5.13 and 5.14. For $p = 2$ our convergence rate analysis in the case of oversmoothing, i.e. for $r = 2$, yields the same convergence rate $\mathcal{O}(\delta^{\frac{s}{s+a}})$ measured in the same norm, the L^2-norm, under the same smoothness condition given by $B_{2,\infty}^s$ as in the case $r = 0$. Hence the paradigm 'oversmoothing does not hurt' known for Hilbert-space regularization remains true for $B_{2,q}^r$ Banach space penalties with $q < 2$.*
In contrast to $p = 1$ where we see in Example 3.67 that a higher value of r may cause an assignment of a lower smoothness s to a fixed true solution. For example, for the coefficient c^{jump} we assign every $s < \frac{4}{3}$ in the case $r = 0$ but only the values $s < \frac{6}{7}$ when we set $r = 2$. For $r = 0$ we obtain the faster rate $\mathcal{O}(\delta^e)$ for $e < \frac{2}{5}$ measured in the L^1-norm and for $r = 2$ we obtain the slower rate $\mathcal{O}(\delta^e)$ for $e < \frac{3}{10}$ but measured in the stronger $L^{\frac{4}{3}}$-norm.
This indicates that the chosen loss function may have an influence on the observed convergence rate. This motivates the development of a convergence rates theory that is more flexible in the loss function, which is as far as the author knows an open problem. It is not clear how to obtain error bounds with respect to norms whose corresponding spaces $\mathbb{X}_L$ do not relate to the scale spanned by real interpolation of $\mathbb{X}_-$ and $\mathbb{X}_{\mathcal{R}}$ as in Assumption 5.1.
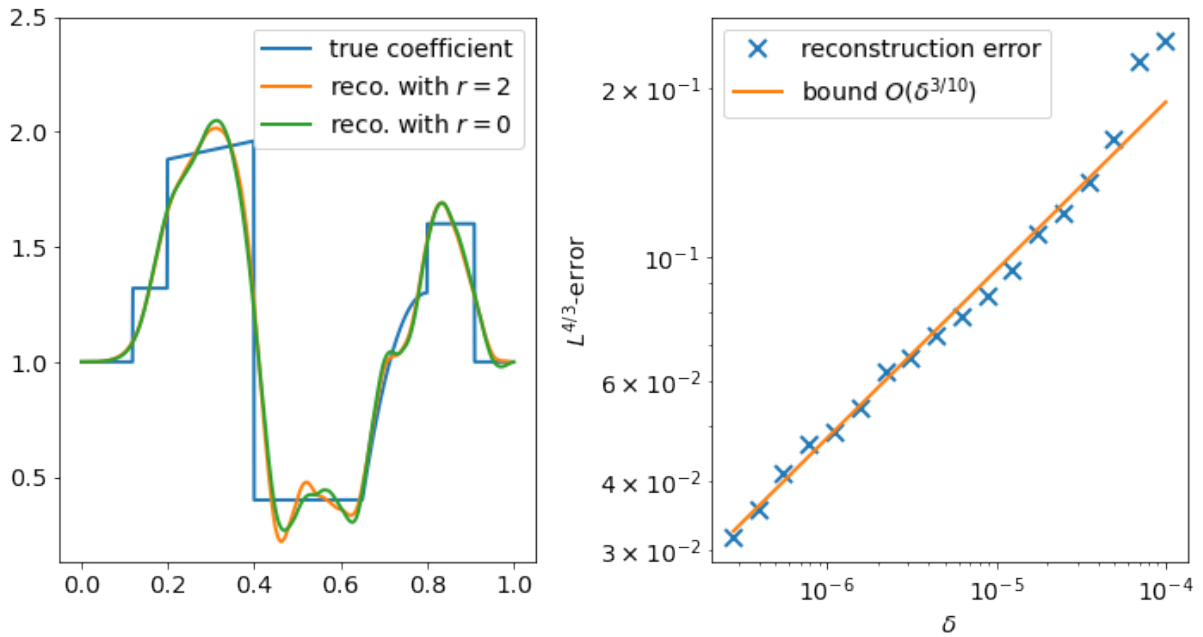
FIGURE 5.2: *Left:* true coefficient c^{jump} with jumps in the boundary value problem (3.40) together with reconstructions for $b_{1,1}^2$- and $b_{1,1}^0$-penalization at noise level $\delta = 2.5 \cdot 10^{-5}$ for the same data. *Right:* Reconstruction error using $b_{1,1}^2$-penalization, the rate $\mathcal{O}(\delta^{3/10})$ in the $L^{4/3}$-norm predicted by Theorem 5.10 (see Example 5.14).

In the next chapter we obtain convergence rates in terms of the modulus of continuity using image space approximation rates as source conditions. Although this approach works in a very general setting its application to Banach space regularization requires the estimation of the latter modulus which in turn forces us again to require an assumption similar to the one on $\mathbb{X}_L$ in Assumption 5.1.

5.4 Application to bounded variation regularization

We close this chapter with an application of Theorem 5.3 to Tikhonov regularization with penalty term given by the BV-norm. Let $d \in \mathbb{N}$ and $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain. A function $f \in L^1(\Omega)$ has *bounded variation* if

$$|f|_{\text{BV}(\Omega)} := \sup\left\{\int_\Omega f(x)\,\text{div}g(x)\,\mathrm{d}x \colon g \in C_c^1(\Omega, \mathbb{R}^d),\ \|g\|_{L^\infty(\Omega,\mathbb{R}^d)} \leq 1\right\} < \infty.$$

Here $\|g\|_{L^\infty(\Omega,\mathbb{R}^d)} := \left\|\left(\sum_{j=1}^d g_i^2\right)^{1/2}\right\|_{L^\infty(\Omega)}$ with $g = (g_1, \ldots g_n)$. Then

$$\text{BV}(\Omega) := \{f \in L^1(\Omega) \colon |f|_{\text{BV}(\Omega)} < \infty\}$$

is a Banach space equipped with $\|\cdot\|_{\text{BV}(\Omega)} := \|\cdot\|_{L^1(\Omega)} + |\cdot|_{\text{BV}(\Omega)}$. We refer to [2] for a detailed study of spaces of bounded variation.

Throughout this section we assume that $F \colon \text{dom}(F) \to \mathbb{Y}$ with $\text{dom}(F) \subset B_{2,2}^{-a}(\Omega)$ closed

for some $a \geq 0$ such that $a \geq \frac{d}{2} - 1$, $\mathrm{BV}(\Omega) \cap \mathrm{dom}(F) \neq \emptyset$ and with a Banach space $\mathbb{Y}$ such that a two-sided Lipschitz condition

$$\frac{1}{M}\|f_1 - f_2 \,|\, B_{2,2}^{-a}\| \leq \|F(f_1) - F(f_2)\|_{\mathbb{Y}} \leq M\|f_1 - f_2 \,|\, B_{2,2}^{-a}\| \quad \text{for all } f_1, f_2 \in \mathrm{dom}(F).$$

for some constant $M \geq 1$.

For $g \in \mathbb{Y}$ we consider

$$R_\alpha(g) = \operatorname*{argmin}_{h \in \mathrm{BV}(\Omega) \cap \mathrm{dom}(F)} \left(\frac{1}{2\alpha}\|g - F(h)\|_{\mathbb{Y}}^2 + \|h\|_{\mathrm{BV}(\Omega)}\right) \subset \mathrm{BV}(\Omega) \cap \mathrm{dom}(F). \quad (5.14)$$

We refer to [1] for this kind of regularization scheme for linear operators including a proof of existence of minimizers and to [30] for a treatment of similar estimators in a statistical setting.

We start by showing that the condition $a \geq \frac{d}{2} - 1$ ensures a continuous embedding $\mathrm{BV}(\Omega) \subset B_{2,2}^{-a}(\Omega)$.

Lemma 5.17 (Embedding). *There is a continuous embedding* $\mathrm{BV}(\Omega) \subset B_{2,2}^{-a}(\Omega)$.

Proof. For all embeddings involving the space $\mathrm{BV}(\Omega)$ in this proof we refer to [2, Cor. 3.49 & Prop. 3.21].

For $d = 1$ there is a continuous embedding $\mathrm{BV}(\Omega) \subset L^2(\Omega)$. By Proposition A.11 we have $L^2(\Omega) = B_{2,2}^0(\Omega) \subset B_{2,2}^{-a}(\Omega)$, which yields the claim in this case.

For $d > 1$ we set $p := \frac{d}{d-1}$. Then $p \in (1, 2]$ and there is a continuous embedding $\mathrm{BV}(\Omega) \subset L^p(\Omega)$. By Proposition A.11 we have a continuous embedding $L^p(\Omega) \subset B_{p,2}^0(\Omega)$. Furthermore, $a + \frac{d}{2} \geq d - 1 = \frac{d}{p}$ yields a continuous embedding $B_{p,2}^0(\Omega) \subset B_{2,2}^{-a}(\Omega)$. Putting together the latter three embeddings yields the claim. $\qquad\square$

Let $s \in (-a, 1)$. The following interpolation identity, based on the result by Cohen et al. in [20, Thm. 1.4], is a crucial ingredient for our convergence rates result

$$B_{t_s,t_s}^s(\Omega) = \left(B_{2,2}^{-a}(\Omega), \mathrm{BV}(\Omega)\right)_{\theta_s,t_s} \quad \text{with } \theta_s := \frac{s+a}{a+1} \text{ and } t_s := \frac{2a+2}{s+2a+1} \quad (5.15)$$

with equivalent norms. In the latter reference the authors show this identity for $\Omega = \mathbb{R}^d$ and from there we conclude the statement in Proposition A.12.

To avoid the abstract smoothness condition in Theorem 5.3 we state our theorem under a slightly stronger smoothness condition and comment on the weaker condition in a remark afterwards.

Theorem 5.18 (Convergence rates for BV-regularization). *Let* $0 < s < 1$ *and* $\varrho > 0$. *Assume that*

$$f \in B_{t_s,t_s}^s(\Omega) \quad \text{with } \|f \,|\, B_{t_s,t_s}^s\| < \varrho.$$

Let $\overline{p} = \frac{2a+2}{2a+1}$ *and suppose that* $\mathrm{dom}(F)$ *contains an* $L^{\overline{p}}(\Omega)$-*ball with radius* τ *around* f.

(a) (Bias bounds) There exits a constant C_b *independent of* f, ϱ *and* τ *such that*

$$\|f - f_\alpha \,|\, B_{2,2}^{-a}\| \leq C_b \varrho^{\frac{a+1}{s+2a+1}} \alpha^{\frac{s+a}{s+2a+1}},$$

$$\|f - f_\alpha\|_{L^{\overline{p}}(\Omega)} \leq C_b \varrho^{\frac{2a+1}{s+2a+1}} \alpha^{\frac{s}{s+2a+1}} \quad \text{and}$$

$$\|f_\alpha\|_{\mathrm{BV}(\Omega)} \leq C_b \varrho^{\frac{2a+2}{s+2a+1}} \alpha^{\frac{s-1}{s+2a+1}}.$$

holds true for all $0 < \alpha < \varrho^{-\frac{1-s+2a}{s}} \tau^{\frac{s+2a+1}{s}}$ *and* $f_\alpha \in R_\alpha(F(f))$ *(see* (5.14)*).*

(b) *(Convergence rates) Let $0 < c_l \leq c_r$ and $1 < c_D \leq C_D$. Suppose $0 < \delta < \varrho^{-\frac{a}{s}}\tau^{\frac{s+a}{s}}$, $g^{\mathrm{obs}} \in \mathbb{Y}$ with $\|g^{\mathrm{obs}} - F(f)\|_{\mathbb{Y}} \leq \delta$. Let $\alpha > 0$ and $\hat{f}_\alpha \in R_\alpha(g^{\mathrm{obs}})$. There is a constant C_r independent of $f, g^{\mathrm{obs}}, \varrho, \tau$ and δ such that either of the conditions*

$$c_l \varrho^{-\frac{a+1}{s+a}}\delta^{\frac{s+2a+1}{s+a}} \leq \alpha \leq c_r \varrho^{-\frac{a+1}{s+a}}\delta^{\frac{s+2a+1}{s+a}} \quad and \quad c_D \delta \leq \|g^{\mathrm{obs}} - F(\hat{f}_\alpha)\|_{\mathbb{Y}} \leq C_D \delta$$

implies the following bounds

$$\|f - \hat{f}_\alpha \mid B_{2,2}^{-a}\| \leq C_r \delta,$$
$$\|f - \hat{f}_\alpha\|_{L^{\bar{p}}(\Omega)} \leq C_r \varrho^{\frac{a}{s+a}}\delta^{\frac{s}{s+a}} \quad and$$
$$\|\hat{f}_\alpha\|_{\mathrm{BV}(\Omega)} \leq C_r \varrho^{\frac{a+1}{s+a}}\delta^{\frac{s-1}{s+a}}.$$

Proof. We show that Assumption 5.1 is satisfied with $\mathbb{X}_{\mathcal{R}} = \mathrm{BV}(\Omega)$, $\mathbb{X}_- = B_{2,2}^{-a}(\Omega)$, $\mathbb{Y}$ and $\mathbb{X}_L = L^{\bar{p}}(\Omega)$ and $\tilde{D}_F = \mathrm{BV}(\Omega) \cap \mathrm{dom}(F)$ and $u = 1$. Due to Lemma 5.17 we have a continuous embedding $\mathrm{BV}(\Omega) \subset B_{2,2}^{-a}$. For $a = 0$, we have $\bar{p} = 2$ and $B_{2,2}^0 = L^2(\Omega)$ (see Proposition A.11). Hence we have Assumption 5.1 with $\xi = 1$ in this case and the result follows from Theorem 5.3.

If $a > 0$, we set $\xi := \frac{a}{a+1}$. Note that $1 < \bar{p} = t_0 < 2$. Hence [8, Thm. 3.4.1.(b)], (5.15) and Proposition A.11 yield the following chain of continuous embeddings

$$\left(B_{2,2}^{-a}(\Omega), \mathrm{BV}(\Omega)\right)_{\xi,1} \subset \left(B_{2,2}^{-a}(\Omega), \mathrm{BV}(\Omega)\right)_{\xi,\bar{p}} = B_{\bar{p},\bar{p}}^0(\Omega) \subset L^{\bar{p}}(\Omega). \qquad (5.16)$$

Finally, [8, Thm. 3.4.1.(b)] and (5.15) yields

$$B_{t_s,t_s}^s(\Omega) = \left(B_{2,2}^{-a}(\Omega), \mathrm{BV}(\Omega)\right)_{\theta_s,t_s} \subset \left(B_{2,2}^{-a}(\Omega), \mathrm{BV}(\Omega)\right)_{\theta_s,\infty}.$$

Hence the smoothness condition on f in the claim implies the smoothness condition in Theorem 5.3. Therefore, the stated result follows from Theorem 5.3. $\qquad \square$

Remark 5.19 (Weaker smoothness condition)**.** *The statements in Theorem 5.18 remain true if the smoothness assumption on f is replaced by $f \in \left(B_{2,2}^{-a}(\Omega), \mathrm{BV}(\Omega)\right)_{\theta_s,\infty}$ with a bound by ϱ on the norm of f therein.*

Remark 5.20 (Similarity to $B_{1,1}^1$-regularization)**.** *We see that the convergence rates and also the smoothness condition for BV-regularization equals the ones for $B_{1,1}^1$-regularization in Theorem 5.10. The reason for that is that the interpolation identity in (5.15) holds true with $\mathrm{BV}(\Omega)$ replaced by $B_{1,1}^1(\Omega)$.*
Whereas for $B_{1,1}^1$-regularization with a norm given by wavelet coefficients we also have a convergence rate result in the non oversmoothing case $s > r$ (see Theorem 3.71) a similar result remains open for BV-regularization.

Remark 5.21 (Comparison to the results by del Alamo)**.** *A major advantage of our result over the ones in [30] is that we do not have to require that the unknown solution f itself is a BV-function. Whereas our result presumes a deterministic noise model the latter work deals with a statistical setup similar to the one in Section 3.7. It remains open to generalize the presented approach to statistical noise models to obtain convergence rates for the estimators in [30] under weaker smoothness conditions on the unknown.*

Chapter Six

Convergence rates based on image space approximation

In this chapter we present an approach to convergence rate results published in the author's article [84]. Avoiding Bregman distances and using image space approximation rates as source conditions we prove a nearly minimax theorem showing that the modulus of continuity is an upper bound on the reconstruction error up to a constant. Applied to Besov space regularization we obtain new convergence rate results for $0, 2, q$- and $0, p, p$-penalties without restrictions on $p, q \in (1, \infty)$.

6.1 Minimax convergence rates

In this section we work in the general setting of linear and bounded forward operators $A \colon \mathbb{X} \to \mathbb{Y}$ between a Banach space $\mathbb{X}$ and a Hilbert space $\mathbb{Y}$. We require the topological assumptions in Assumption 2.2 and use a loss function $L \colon \mathbb{X} \times \mathbb{X} \to [0, \infty]$ to measure the reconstruction error. For the sake of clearness and self-containedness we summarize the precise assumption as follows.

Assumption 6.1. *Let $\mathbb{X}$ be a Banach space and let $\tau_{\mathbb{X}}$ be a further topology such that $(\mathbb{X}, \tau_{\mathbb{X}})$ is a locally convex Hausdorff space and $\mathcal{R} \colon \mathbb{X} \to (-\infty, \infty]$ a proper, convex function. We assume that the sublevel sets $\{x \in \mathbb{X} \colon \mathcal{R}(x) \leq \lambda\}$ are $\tau_{\mathbb{X}}$-compact for all $\lambda \in \mathbb{R}$. Let $\mathbb{Y}$ be a Hilbert space and $A \colon \mathbb{X} \to \mathbb{Y}$ a linear, $\tau_{\mathbb{X}}$-to-weak continuous operator. Finally, suppose that $L \colon \mathbb{X} \times \mathbb{X} \to [0, \infty]$ satisfies the triangle inequality.*

As we will apply the following theory mainly to sequence spaces we denote the elements of $\mathbb{X}$ by x, z. We consider the Tikhonov functional

$$T_\alpha(z, g) = \frac{1}{2\alpha} \|g - Az\|_{\mathbb{Y}}^2 + \mathcal{R}(z)$$

and its set of minimizers

$$S_\alpha(g) = \operatorname*{argmin}_{z \in \operatorname{dom}(\mathcal{R})} T_\alpha(z, g) \subset \operatorname{dom}(\mathcal{R}). \tag{6.1}$$

Due to Proposition 2.3 we have $S_\alpha(g) \neq \emptyset$ for all $g \in \mathbb{Y}$ and $\alpha > 0$. Moreover, $\operatorname{argmin}_{z \in \mathbb{X}} \mathcal{R}(z) \neq \emptyset$ and for every $g \in \operatorname{im}(A)$ there exists a $\mathcal{R}$-minimal solution $x \in \mathbb{X}$, that is $Ax = g$ and $\mathcal{R}(x) \leq \mathcal{R}(z)$ for all $z \in \mathbb{X}$ with $Az = g$. As announced, we introduce

source sets, that is, sets on which we will prove convergence rates, based on image space approximation bounds: Let $\nu \in [0, \infty)$ and $\varrho > 0$. We define

$$\varrho_\nu \colon \mathbb{X} \to [0, \infty] \quad \text{by} \quad \varrho_\nu(x) = \sup\left\{\alpha^{-\nu}\|Ax - Ax_\alpha\|_\mathbb{Y} : \alpha > 0, \, x_\alpha \in S_\alpha(Ax)\right\}, \quad (6.2)$$

$$K_\nu^\varrho = \{x \in \mathbb{X} \colon \varrho_\nu(x) \leq \varrho\} \quad \text{and} \quad K_\nu := \{x \in \mathbb{X} : \varrho_\nu(x) < \infty\}.$$

Recall that Ax_α does not depend on the choice of the minimizer x_α (see Proposition 4.20). Note that $x \in K_\nu$ if and only if a bound $\|Ax - Ax_\alpha\|_\mathbb{Y} \leq c\alpha^\nu$ holds true and $\varrho_\nu(x)$ provides the smallest possible constant $c > 0$.

The first and central result is a uniform bound in K_ν^ϱ on $L(x, \hat{x}_\alpha)$ with $\hat{x}_\alpha \in S_\alpha(g^\delta)$ in terms of the *modulus of continuity*. Recall that the latter is given by

$$\omega(\delta, K, L) := \sup\left\{L(x_1, x_2) : x_1, x_2 \in K \text{ with } \|Ax_1 - Ax_2\|_\mathbb{Y} \leq \delta\right\} \quad (6.3)$$

for a subset $K \subset \mathbb{X}$. As above we will consider both, an a priori rule requiring prior knowledge of the parameter ν such that the unknown belongs to K_ν characterizing the regularity of the unknown x, and the discrepancy principle as most well-known a posteriori rule.

To this end, we follow an idea presented in the seminal paper [33]: Any feasible procedure is nearly minimax (see [33, 4.3.1.]). In our context feasibility means

- image space bounds: $\|Ax - A\hat{x}_\alpha\|_\mathbb{Y} \leq c\delta$,
- regularity of minimizers: $\varrho_\nu(\hat{x}_\alpha) \leq c\varrho_\nu(x)$ for some constant $c > 0$.

After proving feasibility, we use the same argument as in [33, 4.3.1. Prop. 2] to obtain the minimax result.

In the sequel we will first address the sets K_ν and their basic properties. Moreover for $\nu \in (\frac{1}{2}, 1]$, we provide descriptions of the latter sets that do not make explicit use of the minimizers of the Tikhonov functional. After that we will analyze image space bounds and regularity of minimizers separately to conclude the minimax result.

The sets K_ν

We show that K_ν is nontrivial for $\nu \in (0, 1]$ and recall that for $\nu \in (\frac{1}{2}, 1]$ the set K_ν are characterized by variational source conditions. The oversmoothing case corresponds to $\nu \in (0, \frac{1}{2})$ and here the image space rate is characterized by bounds on the minimal value in (6.1). Moreover, K_ν does not change for $\nu > 1$.

Proposition 6.2 (Properties of the K_ν's)**.** *Suppose Assumption 6.1 holds true. The following statements hold true:*

(a) $K_{\nu_2} \subset K_{\nu_1}$ *for* $0 \leq \nu_1 \leq \nu_2$.

(b) $K_0 = \mathbb{X}$.

(c) *Let* $\nu \in (0, \frac{1}{2})$. *Then* $x \in K_\nu$ *if and only if the minimal value* $\vartheta_{Ax}(\alpha)$ *in (6.1) (see Definition 4.22) satisfies*

$$\vartheta_{Ax}(\alpha) - \inf_{z \in \mathbb{X}} \mathcal{R}(z) \leq c_{\text{over}}\alpha^{2\nu-1}$$

for some constant $c_{\text{over}} \geq 0$.

(d) $\text{dom}(\mathcal{R}) + \ker(A) \subseteq K_{1/2}$.

(e) *Let $\nu \in (\frac{1}{2}, 1]$ and assume $x \in \mathbb{X}$ satisfies $\mathcal{R}(x) \leq \mathcal{R}(z)$ for all $z \in \mathbb{X}$ with $Ax = Az$. Then $x \in K_\nu$ if and only if there exists a constant $c_{\mathrm{vsc}} > 0$ such that the variational source condition*

$$\mathcal{R}(x) - \mathcal{R}(z) \leq c_{\mathrm{vsc}} \|Ax - Az\|_{\mathbb{Y}}^{\frac{2\nu-1}{\nu}} \quad \text{for all } z \in \mathbb{X}$$

holds true.

(f) *For $\nu > 1$ we have $K_\nu = \mathrm{argmin}_{z \in \mathbb{X}} \mathcal{R}(z) + \ker(A)$ and $x \in K_\nu$ implies $Ax = Ax_\alpha$ for all $\alpha > 0$ in this case.*

Proof. (a) Suppose $x \in K_{\nu_2}$. Then

$$\|Ax - Ax_\alpha\|_{\mathbb{Y}} = \|Ax - Ax_\alpha\|_{\mathbb{Y}}^{\frac{\nu_1}{\nu_2}} \|Ax - Ax_\alpha\|_{\mathbb{Y}}^{1 - \frac{\nu_1}{\nu_2}} \leq \varrho_{\nu_2}(x)^{\frac{\nu_1}{\nu_2}} \varrho_0(x)^{1 - \frac{\nu_1}{\nu_2}} \alpha^{\nu_1}$$

yields $\varrho_{\nu_1}(x) \leq \varrho_{\nu_2}(x)^{\frac{\nu_1}{\nu_2}} \varrho_0(x)^{1 - \frac{\nu_1}{\nu_2}}$.

(b) Let $x \in \mathbb{X}$. We set

$$D_x := \inf \left\{ \|Ax - Ay\|_{\mathbb{Y}} : y \in \mathrm{argmin}_{z \in \mathbb{X}} \mathcal{R}(z) \right\}.$$

Suppose $y \in \mathrm{argmin}_{z \in \mathbb{X}} \mathcal{R}(z)$, $\alpha > 0$ and $x_\alpha \in S_\alpha(Ax)$. Then

$$\frac{1}{2\alpha} \|Ax - Ax_\alpha\|_{\mathbb{Y}}^2 + \mathcal{R}(x_\alpha) \leq \frac{1}{2\alpha} \|Ax - Ay\|_{\mathbb{Y}}^2 + \mathcal{R}(y).$$

As $\mathcal{R}(y) \leq \mathcal{R}(x_\alpha)$, this implies $\|Ax - Ax_\alpha\|_{\mathbb{Y}} \leq \|Ax - Ay\|_{\mathbb{Y}}$. Hence $\varrho_0(x) \leq D_x < \infty$. Therefore, $x \in K_0$. The inclusion $\mathbb{X} \subset K_0$ holds true by definition.

(c) See Theorem 4.34.

(d) Let $x = y + k \in \mathrm{dom}(\mathcal{R}) + \ker(A)$. From

$$\frac{1}{2\alpha} \|Ax - Ax_\alpha\|_{\mathbb{Y}}^2 \leq T_\alpha(x_\alpha, Ax) \leq T_\alpha(y, Ax) = \mathcal{R}(y)$$

we obtain $\varrho_{1/2}(x) \leq \sqrt{2\mathcal{R}(y)}$.

(e) See Theorem 4.33.

(f) Let $\nu > 1$. Suppose $x \in K_\nu$. From the first order condition [6, Prop. 16.34] and Proposition 4.20 we obtain

$$\eta_\alpha := \frac{1}{\alpha}(Ax - Ax_\alpha) = \frac{1}{\alpha}(Ax - \mathrm{Prox}_{\alpha \mathcal{Q}}(Ax)) \in \partial \mathcal{Q}(Ax_\alpha).$$

Since $\eta_\alpha \to 0$ and $Ax_\alpha \to Ax$ for $\alpha \to 0$ in the norm topology of $\mathbb{Y}$ this implies $0 \in \partial \mathcal{Q}(Ax)$. Hence $Ax \in \mathrm{argmin}_{g \in \mathbb{Y}} \mathcal{Q}(g)$. Now let $y \in \mathbb{X}$ be $\mathcal{R}$-minimal with $Ay = Ax$. Then

$$\mathcal{R}(y) = \mathcal{Q}(Ax) \leq \mathcal{Q}(Az) \leq \mathcal{R}(z) \quad \text{for all } z \in \mathbb{X}.$$

Hence

$$x = y + x - y \in \mathrm{argmin}_{z \in \mathbb{X}} \mathcal{R}(z) + \ker(A).$$

On the other hand, assume $x = y + k \in \mathrm{argmin}_{z \in \mathbb{X}} \mathcal{R}(z) + \ker(A)$. Then

$$\frac{1}{2\alpha} \|Ax - Ax_\alpha\|_{\mathbb{Y}}^2 + \mathcal{R}(x_\alpha) \leq T_\alpha(y, Ax) = \mathcal{R}(y) \leq \mathcal{R}(x_\alpha)$$

yields $\|Ax - Ax_\alpha\|_{\mathbb{Y}} = 0$ and $x \in K_\nu$. $\qquad\square$

Note that the last inequality in the above proof of (*a*) resembles an interpolation inequality. This gives a first hint to a connection to interpolation theory in the case of Banach space regularization.

In Theorems 4.33 and 4.34 we have seen, that powers of the best constants c_{over} and c_{vsc} and (*c*) in (*e*) are lower and upper bounded by the best constant $\varrho_\nu(x)$ in the image space rate (see (6.2)). Moreover, (*f*) entails the following statement that yields positive definiteness of ϱ_ν if $\mathcal{R}$ is positive definite and A is injective.

Corollary 6.3. *Suppose Assumption 6.1 holds true and let $\nu \in [0, \infty)$. We have $\varrho_\nu(x) = 0$ if and only if $x \in \operatorname{argmin}_{z \in \mathbb{X}} \mathcal{R}(z) + \ker(A)$.*

Proof. Suppose $\varrho_\nu(x) = 0$. Then $Ax = Ax_\alpha$ for all $\alpha > 1$. Therefore, $x \in K_2$ and Proposition 6.2(f) yields $x \in \operatorname{argmin}_{z \in \mathbb{X}} \mathcal{R}(z) + \ker(A)$. On the other hand, again by Proposition 6.2(f), $x \in \operatorname{argmin}_{z \in \mathbb{X}} \mathcal{R}(z) + \ker(A)$ implies $Ax = Ax_\alpha$ for all $\alpha > 0$ and $x_\alpha \in S_\alpha(Ax)$. Hence $\varrho_\nu(x) = 0$. $\qquad\square$

Next we present a further characterization of K_ν in the case $\nu = 1$ that will be of central importance in the sequel: K_1 contains precisely the elements satisfying a classical source condition $A^*\omega \in \partial\mathcal{R}(x)$ for some $\omega \in \mathbb{Y}$ based on the subdifferential of the penalty term at the true solution (see also (2.4)).

It is a well-known fact that the latter source condition implies a linear convergence rate in the image space (see e.g. [53, Lem. 3.5]). Also the converse holds true: A linear bound $\|Ax - Ax_\alpha\|_\mathbb{Y} = \mathcal{O}(\alpha)$ implies the classical source condition and the minimal $\mathcal{O}$-constant $\varrho_1(x)$ agrees with the minimal norm $\|\omega\|_\mathbb{Y}$ attained by a source element ω. Similar results can be found in [53, Lem. 4.1] and [91, Prop. 4.1]. For sake of self-containedness we include a proof.

Proposition 6.4 (Linear image space rate)**.** *Suppose Assumption 6.1 holds true and that $x \in \mathbb{X}$ such that $\mathcal{R}(x) \le \mathcal{R}(z)$ for all $z \in \mathbb{X}$ with $Ax = Az$. Let $x_\alpha \in S_\alpha(Ax)$, $\alpha > 0$ any selection of minimizers for exact data (see (6.1)). Then $x \in K_1$ if and only if there exists $\omega \in \mathbb{Y}$ such that $A^*\omega \in \partial\mathcal{R}(x)$. More precisely, we have*

$$\varrho_1(x) = \inf \left\{ \|\omega\|_\mathbb{Y} \colon A^*\omega \in \partial\mathcal{R}(x) \right\}.$$

Moreover, the net $(\frac{1}{\alpha}(Ax - Ax_\alpha))_{\alpha>0}$ convergences weakly for $\alpha \searrow 0$ to the unique $\omega \in \mathbb{Y}$ with $A^*\omega \in \partial\mathcal{R}(x)$ and $\|\omega\|_\mathbb{Y} = \varrho_1(x)$.

Proof. First we assume $\omega \in \mathbb{Y}$ with $A^*\omega \in \partial\mathcal{R}(x)$. The first order optimality condition $\xi_\alpha := \frac{1}{\alpha} A^*A(x - x_\alpha) \in \partial\mathcal{R}(x_\alpha)$ yields

$$\frac{1}{\alpha}\|Ax - Ax_\alpha\|^2 = \langle \xi_\alpha, x - x_\alpha \rangle \le \mathcal{R}(x) - \mathcal{R}(x_\alpha) \le \langle A^*\omega, x - x_\alpha \rangle \le \|\omega\|_\mathbb{Y} \cdot \|Ax - Ax_\alpha\|_\mathbb{Y}.$$

Hence

$$\|Az - Az_\alpha\|_\mathbb{Y} \le \|\omega\|_\mathbb{Y}\alpha. \tag{6.4}$$

Taking the infimum over ω in (6.4) yields

$$\varrho_1(x) \le \inf \left\{ \|\omega\|_\mathbb{Y} \colon A^*\omega \in \partial\mathcal{R}(x) \right\}. \tag{6.5}$$

Next we assume that there exists a bound $\|Ax - Ax_\alpha\|_{\mathbb{Y}} = \mathcal{O}(\alpha)$, which means $\varrho_1(x) < \infty$. Then the net $(\frac{1}{\alpha}(Ax - Ax_\alpha))_{\alpha > 0}$ is norm bounded by $\varrho_1(x)$ in the Hilbert space $\mathbb{Y}$. By the Banach–Alaoglu theorem every null sequence of positive numbers has a subsequence $\alpha_n > 0$ such that $\frac{1}{\alpha_n}(Ax - Ax_{\alpha_n})$ converges weakly to some $\omega \in \mathbb{Y}$ with $\|\omega\|_{\mathbb{Y}} \leq \varrho_1(x)$. Lemma 4.28 and the minimality assumption yield

$$\frac{1}{2\alpha_n}\|Ax - Ax_{\alpha_n}\|_{\mathbb{Y}}^2 + \mathcal{R}(x_{\alpha_n}) \to \mathcal{R}(x).$$

Together with $\|Ax - Ax_{\alpha_n}\|_{\mathbb{Y}} \leq \varrho_1(x)\alpha_n$ we obtain $\mathcal{R}(x_{\alpha_n}) \to \mathcal{R}(x)$. Using the first order optimality condition $\xi_{\alpha_n} \in \mathcal{R}(x_{\alpha_n})$ yields

$$
\begin{aligned}
\mathcal{R}(x) + \langle A^*\omega, z - x \rangle = \mathcal{R}(x) + \langle \omega, A(z - x) \rangle \\
= \lim_{n \to \infty} \mathcal{R}(x_{\alpha_n}) + \langle \frac{1}{\alpha_n} A(x - x_{\alpha_n}), A(z - x_{\alpha_n}) \rangle \\
= \lim_{n \to \infty} \mathcal{R}(x_{\alpha_n}) + \langle \xi_{\alpha_n}, z - x_{\alpha_n} \rangle \leq \mathcal{R}(z)
\end{aligned}
$$

for all $z \in \mathbb{X}$. This shows $A^*\omega \in \partial\mathcal{R}(x)$ and therefore equality in (6.5).

Finally, we turn to the claim's last statement. Being the preimage of the convex set $\partial\mathcal{R}(x)$ under the linear map A^*, the set $\{\omega \in \mathbb{Y} \colon A^*\omega \in \partial\mathcal{R}(x)\}$ is convex. Strict convexity of $\|\cdot\|_{\mathbb{Y}}$ yields uniqueness of ω. In particular this implies the convergence of the net. $\quad\square$

As an example we explicitly calculate ϱ_1 for the embedding operator $\ell^p \subset \ell^2$.

Example 6.5. *Let $p \in [1, 2]$, $\mathbb{X} := \ell^p$, $\mathbb{Y} := \ell^2$, $A \colon \ell^p \to \ell^2$ the embedding operator and $\mathcal{R}$ given by $\mathcal{R}(x) = \frac{1}{p}\|x\|_p$.*
Let $x \in \ell^p$. First we consider the case $p > 1$. Here $\partial\mathcal{R}(x) = \{\xi\}$ with $|\xi_j| = |x_j|^{p-1}$. The adjoint A^* identifies with the embedding operator $\ell^2 \to \ell^{p'}$ with p' the Hölder conjugate of p. Hence $x \in K_1$ if and only if $\|\xi\|_{\ell^2} < \infty$, and we have

$$\varrho_1(x) = \|\xi\|_{\ell^2} = \left(\sum_{j \in \mathbb{N}} |x_j|^{2p-2}\right)^{1/2} = \|x\|_{2p-2}^{p-1}.$$

Note that $\varrho_1(\cdot)^{\frac{1}{p-1}}$ is a quasi-norm.
For $p = 1$ we have $\xi \in \partial\mathcal{R}(x)$ if and only if $\xi_j = 1$ for $x_j > 0$, $\xi_j = -1$ for $x_j < 0$ and $|\xi_j| \leq 1$ for $x_j = 0$. Hence K_1 consists of all elements with finitely many non vanishing coefficients. We have $\varrho_1(x) = \#\{j \in \mathbb{N} \colon x_j \neq 0\}^{1/2}$ and therefore in this case no power of ϱ_1 is a quasi-norm.

Image space bounds

This subsection is devoted to error bounds in the image space $\mathbb{Y}$ in terms of the deterministic noise level and the image space approximation error for exact data. As a first step we exploit firm non-expansiveness (see [6, Def. 4.1]) of the proximal operator $A \circ S_\alpha$.

Lemma 6.6 (Firm non-expansiveness). *Suppose Assumption 6.1 holds true. Let $g, h \in \mathbb{Y}$, $\alpha > 0$, $\hat{x}_\alpha \in S_\alpha(g)$ and $\hat{z}_\alpha \in S_\alpha(h)$. Then*

$$\|(g - A\hat{x}_\alpha) - (h - A\hat{z}_\alpha)\|_{\mathbb{Y}}^2 + \|A\hat{x}_\alpha - A\hat{z}_\alpha\|_{\mathbb{Y}}^2 \leq \|g - h\|_{\mathbb{Y}}^2.$$

Proof. By [6, Prop. 12.27] the proximity operator $\text{Prox}_{\alpha Q}$ is firm non-expansive, that is

$$\|(g - \text{Prox}_{\alpha Q}(g)) - (h - \text{Prox}_{\alpha Q}(h))\|_{\mathbb{Y}}^2 + \|\text{Prox}_{\alpha Q}(g) - \text{Prox}_{\alpha Q}(h)\|_{\mathbb{Y}}^2 \leq \|g - h\|_{\mathbb{Y}}^2$$

for all $g, h \in \mathbb{Y}$. Inserting the first identity in Proposition 4.20 yields the claim. $\qquad \square$

With Lemma 6.6 at hand we obtain bounds on $\|Ax - A\hat{x}_\alpha\|_{\mathbb{Y}}$ and on $\|g^\delta - A\hat{x}_\alpha\|_{\mathbb{Y}}$ of the simplest conceivable shape.

Lemma 6.7. *Suppose Assumption 6.1 holds true. Let $\delta \geq 0$, $x \in \mathbb{X}$ and $g^\delta \in \mathbb{Y}$ with $\|g^\delta - Ax\|_{\mathbb{Y}} \leq \delta$. Then following two inequalities*

$$\|Ax - A\hat{x}_\alpha\|_{\mathbb{Y}} \leq \delta + \|Ax - Ax_\alpha\|_{\mathbb{Y}}, \tag{6.6}$$

$$\|g^\delta - A\hat{x}_\alpha\|_{\mathbb{Y}} \leq \delta + \|Ax - Ax_\alpha\|_{\mathbb{Y}} \tag{6.7}$$

hold true for all $\alpha > 0$, $\hat{x}_\alpha \in S_\alpha(g^\delta)$, $x_\alpha \in S_\alpha(Ax)$.

Proof. Lemma 6.6 with $g = g^\delta$ and $h = Ax$ yields

$$\|(Ax - Ax_\alpha) - (g^\delta - A\hat{x}_\alpha)\|_{\mathbb{Y}}^2 + \|Ax_\alpha - A\hat{x}_\alpha\|_{\mathbb{Y}}^2 \leq \delta^2.$$

We neglect the first summand on the left hand side and obtain

$$\|Ax - A\hat{x}_\alpha\|_{\mathbb{Y}} \leq \|Ax - Ax_\alpha\|_{\mathbb{Y}} + \|Ax_\alpha - A\hat{x}_\alpha\|_{\mathbb{Y}} \leq \delta + \|Ax - Ax_\alpha\|_{\mathbb{Y}}$$

and the second for

$$\|g^\delta - A\hat{x}_\alpha\|_{\mathbb{Y}} \leq \|(Ax - Ax_\alpha) - (g^\delta - A\hat{x}_\alpha)\|_{\mathbb{Y}} + \|Ax - Ax_\alpha\|_{\mathbb{Y}} \leq \delta + \|Ax - Ax_\alpha\|_{\mathbb{Y}}. \quad \square$$

The next proposition has two parts. First we apply the last lemma to obtain an image space error bound uniformly on the set K_ν if the parameter α is chosen not too large. The second shows that the lower bound on the norm of the residual in the discrepancy principle implies a certain lower bound on the choice of the regularization parameter.

Proposition 6.8 (Image space bounds). *Suppose Assumption 6.1 holds true, and let $\alpha, \delta, \varrho > 0$ and $\nu \in (0, 1]$. Suppose $x \in K_\nu^\varrho$ and $g^\delta \in \mathbb{Y}$ with $\|g^\delta - Ax\|_{\mathbb{Y}} \leq \delta$.*

(a) Let $c_r > 0$. If $\alpha \leq c_r \varrho^{-\frac{1}{\nu}} \delta^{\frac{1}{\nu}}$ then

$$\|Ax - A\hat{x}_\alpha\|_{\mathbb{Y}} \leq (1 + c_r^\nu)\delta \quad \text{for all} \quad \hat{x}_\alpha \in S_\alpha(g^\delta).$$

(b) Let $c_D > 1$. If $\hat{x}_\alpha \in S_\alpha(g^\delta)$ satisfies $c_D \delta \leq \|g^\delta - A\hat{x}_\alpha\|$, then

$$(c_D - 1)^{\frac{1}{\nu}} \varrho^{-\frac{1}{\nu}} \delta^{\frac{1}{\nu}} \leq \alpha.$$

Proof. Let $x_\alpha \in S_\alpha(Ax)$.

(a) By (6.6) and the definition of ϱ_ν we obtain

$$\|Ax - A\hat{x}_\alpha\|_{\mathbb{Y}} \leq \delta + \varrho\alpha^\nu \leq (1 + c_r^\nu)\delta.$$

(b) The bound (6.7) implies

$$c_D \delta \leq \delta + \|Ax - Ax_\alpha\|_{\mathbb{Y}} \leq \delta + \varrho\alpha^\nu.$$

Subtracting δ and rearranging yields the claim. $\qquad \square$

Regularity of the minimizers

As already announced the second ingredient of the minimax result is the regularity of the minimizers which we will establish in the following. To show that $x \in K_\nu$ implies $\hat{x}_\alpha \in K_\nu$ we have to estimate $\|A\hat{x}_\alpha - A(\hat{x}_\alpha)_\beta\|_{\mathbb{Y}}$ for $(\hat{x}_\alpha)_\beta \in S_\beta(A\hat{x}_\alpha)$ (see (6.1)).

Lemma 6.9. *Suppose Assumption 6.1 holds true, and let $\alpha, \delta, \varrho > 0$ and $\nu \in (0,1]$. Suppose $x \in K_\nu^\varrho$, $g^\delta \in \mathbb{Y}$ with $\|g^\delta - Ax\|_{\mathbb{Y}} \leq \delta$ and $\hat{x}_\alpha \in S_\alpha(g^\delta)$. Furthermore, let $\beta > 0$, $(\hat{x}_\alpha)_\beta \in S_\beta(A\hat{x}_\alpha)$ and $x_\beta \in S_\beta(Ax)$.*

(a) If $\beta \in (0, \alpha]$, then

$$\|A\hat{x}_\alpha - A(\hat{x}_\alpha)_\beta\|_{\mathbb{Y}} \leq \frac{\beta\delta}{\alpha} + \|Ax - Ax_\beta\|_{\mathbb{Y}}.$$

(b) If $\beta \in [\alpha, \infty)$, then

$$\|A\hat{x}_\alpha - A(\hat{x}_\alpha)_\beta\|_{\mathbb{Y}} \leq \delta + 2\|Ax - Ax_\beta\|_{\mathbb{Y}}.$$

Proof. (a) By the first order optimality condition, the element $\hat{x}_\alpha$ satisfies the classical source condition $A^*\omega \in \partial\mathcal{R}(\hat{x}_\alpha)$ with $\omega = \frac{1}{\alpha}(g^\delta - A\hat{x}_\alpha)$.

By Corollary 4.25 the map $\alpha \mapsto \frac{1}{\alpha}\|Ax - Ax_\alpha\|_{\mathbb{Y}}$ is non-increasing. Together with (6.7) we obtain

$$\|\omega\|_{\mathbb{Y}} = \frac{1}{\alpha}\|g^\delta - A\hat{x}_\alpha\|_{\mathbb{Y}} \leq \frac{\delta}{\alpha} + \frac{1}{\alpha}\|Ax - Ax_\alpha\|_{\mathbb{Y}} \leq \frac{\delta}{\alpha} + \frac{1}{\beta}\|Ax - Ax_\beta\|_{\mathbb{Y}}.$$

Hence (6.5) implies the claim.

(b) We use first Lemma 6.6 with $g = Ax$ and $h = A\hat{x}_\alpha$ then (6.6) and finally non-decreasingness of $\alpha \mapsto \|Ax - Ax_\alpha\|_{\mathbb{Y}}$ (see Remark 4.26) to estimate

$$\|(Ax - Ax_\beta) - (A\hat{x}_\alpha - A(\hat{x}_\alpha)_\beta)\|_{\mathbb{Y}} \leq \|Ax - A\hat{x}_\alpha\|_{\mathbb{Y}}$$
$$\leq \delta + \|Ax - Ax_\alpha\|_{\mathbb{Y}}$$
$$\leq \delta + \|Ax - Ax_\beta\|_{\mathbb{Y}}.$$

The triangle inequality finishes the proof. □

In the next proposition we see that in terms of the regularity condition given by K_ν the minimizers stay as smooth as the true solution, whenever α is chosen large enough.

Proposition 6.10 (Regularity of minimizers)**.** *Let $\nu \in (0,1]$ and $\varrho, c_l, \alpha > 0$. Suppose $x \in K_\nu^\varrho$ and $\hat{x}_\alpha \in S_\alpha(g^\delta)$. If $c_l\varrho^{-\frac{1}{\nu}}\delta^{\frac{1}{\nu}} \leq \alpha$, then $\varrho_\nu(\hat{x}_\alpha) \leq (2 + c_l^{-\nu})\varrho$.*

Proof. Let $\beta \leq \alpha$. With $\delta \leq c_l^{-\nu}\varrho\alpha^\nu$ we estimate

$$\frac{\delta\beta}{\alpha} \leq c_l^{-\nu}\varrho\alpha^{\nu-1}\beta \leq c_l^{-\nu}\varrho\beta^\nu.$$

Furthermore,

$$\delta \leq c_l^{-\nu}\varrho\alpha^\nu \leq c_l^{-\nu}\varrho\beta^\nu \quad \text{for all } \beta \geq \alpha.$$

Together with $\|Ax - Ax_\beta\|_{\mathbb{Y}} \leq \varrho\beta^\nu$ for all $\beta > 0$ and $x_\beta \in S_\beta(Ax)$ the result follows from Lemma 6.9. □

Propositions 6.8 and 6.10 precisely confirm the following intuition for Tikhonov regularizers. The choice of α has to be small enough to obtain good image space approximation and it has to be large enough to get the desired regularity of the minimizers.

Almost minimaxity on the sets K_ν

Now we are in position to state and prove the main result of this section.

Theorem 6.11. *Suppose Assumption 6.1 holds true, and let $\alpha, \delta, \varrho > 0$ and $\nu \in (0,1]$. Suppose $x \in K_\nu^\varrho$, $g^\delta \in \mathbb{Y}$ with $\|g^\delta - Ax\|_\mathbb{Y} \leq \delta$ and $\hat{x}_\alpha \in S_\alpha(g^\delta)$ (see (6.1)).*

    (a) (A priori rule) Let $c_r \geq c_l > 0$. If $c_l \varrho^{-\frac{1}{\nu}} \delta^{\frac{1}{\nu}} \leq \alpha \leq c_r \varrho^{-\frac{1}{\nu}} \delta^{\frac{1}{\nu}}$, then

$$L(x, \hat{x}_\alpha) \leq \omega(c_1 \delta, K_\nu^{c_2 \varrho}, L)$$

    with $c_1 := 1 + c_r^\nu$ and $c_2 := 2 + c_l^{-\nu}$.

    (b) (Discrepancy principle) Let $C_D > c_D > 1$. If $c_D \delta \leq \|g^\delta - A\hat{x}_\alpha\|_\mathbb{Y} \leq C_D \delta$, then

$$L(x, \hat{x}_\alpha) \leq \omega(d_1 \delta, K_\nu^{d_2 \varrho}, L)$$

    with $d_1 := 1 + C_D$ and $d_2 := 2 + (c_D - 1)^{-1}$.

Proof.   (a) By Proposition 6.8 we have $\|Ax - A\hat{x}_\alpha\|_\mathbb{Y} \leq c_1 \delta$ and Proposition 6.10 yields $x, \hat{x}_\alpha \in K_\nu^{c_2 \varrho}$.

    (b) Using the triangle inequality we obtain

$$\|Ax - A\hat{x}_\alpha\|_\mathbb{Y} \leq \delta + \|g^\delta - A\hat{x}_\alpha\|_\mathbb{Y} \leq d_1 \delta.$$

    Proposition 6.8 provides $(c_D - 1)^{\frac{1}{\nu}} \varrho^{-\frac{1}{\nu}} \delta^{\frac{1}{\nu}} \leq \alpha$. Therefore, Proposition 6.10 yields $x, \hat{x}_\alpha \in K_\nu^{d_2 \varrho}$.

  In both cases the claim follows from the definition of the modulus ω (see (6.3)).   □

Under mild assumptions, Theorem 6.11 gives rise to an almost minimax result in the following manner. Recall the worst case error $\Delta_R(\delta, K, L)$ of a reconstruction map $R: \mathbb{Y} \to \mathbb{X}$ on a set $K \subset \mathbb{X}$ given by

$$\Delta_R(\delta, K, L) := \sup \left\{ L\left(f, R(\delta, g^{\text{obs}})\right) : f \in K, g^{\text{obs}} \in \mathbb{Y} \text{ with } \|g^{\text{obs}} - F(f)\|_\mathbb{Y} \leq \delta \right\}. \quad (6.8)$$

(see also (2.3)) and the lower bound $\Omega(2\delta, K, L) \leq 2\Delta_R(\delta, K)$ (see Proposition 2.6 with $\|\cdot - \cdot\|_\mathbb{X}$ replaced by L).

Let $\overline{R}_\alpha: \mathbb{Y} \to \mathbb{X}$ satisfy $\overline{R}_\alpha(g^\delta) \in S_\alpha(g^\delta)$ for all $g^\delta \in \mathbb{Y}$ with either $\alpha = \alpha(\delta)$ satisfying the a priori parameter choice given in (a) or $\alpha = \alpha(\delta, g^\delta)$ satisfying the discrepancy principle in (b). In the case $\omega(\delta, K_\nu^\varrho, L) \sim \varrho^e \delta^f$ for some exponents $e, f > 0$ Theorem 6.11 yields a minimax result

$$\Delta_{\overline{R}_\alpha}(\delta, K_\nu^\varrho, L) \leq C \inf_R \Delta_R(\delta, K_\nu^\varrho, L).$$

This shows that up to a constant C no method can achieve a better approximation uniformly on K_ν^ϱ in this case.

Moreover, we would like to highlight the flexibility in the choice of the loss function L. Many recent works in Banach space or convex regularization theory are restricted to error bounds in the Bregman divergence (see e.g. [72], [59], [38], [105]). In some situations the meaning of the Bregman divergence is unclear and lower bounds on the Bregman distance are required to obtain more tangible statements. In [104] these lower bounds cause a restriction on the parameters s, p, q of the Besov scale. By applying Theorem 6.11 to Besov space regularization we can overcome these restrictions.

We end this section with a characterization of the sets K_ν via approximation by elements in K_1.

K_ν via approximation by elements of K_1

In [13, Prop. 1] the authors point out that the set of elements satisfying the source condition (2.4), i.e. the set K_1 (see Proposition 6.4) equals the set of possible minimizers of the Tikhonov functional. Therefore, one might suggest that the approximation error of $x \in \mathbb{X}$ by $x_\alpha \in S_\alpha(Ax)$ is determined by the approximation error from the family of sets

$$K_1^r := \{x \in \mathbb{X} : \varrho_1(x) \leq r\} \quad \text{with } r \geq 0.$$

This motivates to consider the error of best approximation error

$$\gamma_x \colon [0, \infty) \to [0, \infty) \quad \text{given by } \gamma_x(r) = \inf_{z \in K_1^r} \|Ax - Az\|_{\mathbb{Y}}.$$

The function γ_x is well defined as Corollary 6.3 yields $\emptyset \neq \operatorname{argmin}_{z \in \mathbb{X}} \mathcal{R}(z) \subset K_1^r$ for all $r \geq 0$. Moreover, it is non-increasing as $K_1^s \subseteq K_1^r$ for $s \leq r$.
The following proposition is the starting point to prove equivalence of Hölder-type bounds on γ_x and on $\|Ax - Ax_\alpha\|_{\mathbb{Y}}$.

Proposition 6.12. *Let $x \in \mathbb{X}$, $\alpha > 0$ and $x_\alpha \in S_\alpha(Ax)$. Then*

$$\gamma_x \left(\frac{1}{\alpha} \|Ax - Ax_\alpha\|_{\mathbb{Y}} \right) \leq \|Ax - Ax_\alpha\|_{\mathbb{Y}} \leq 4\gamma_x \left(\frac{1}{4\alpha} \|Ax - Ax_\alpha\|_{\mathbb{Y}} \right).$$

Proof. The first order optimality condition $\frac{1}{\alpha} A^* A(x - x_\alpha) \in \partial \mathcal{R}(x_\alpha)$ together with (6.5) provide $\varrho_1(x_\alpha) \leq \frac{1}{\alpha} \|Ax - Ax_\alpha\|_{\mathbb{Y}}$. This proves the first inequality by definition of γ_x.
To show the second inequality let $z \in K_1^r$. Then there exists a $\mathcal{R}$-minimal $\tilde{z} \in \mathbb{X}$, such that $Az = A\tilde{z}$. Then $\tilde{z} \in K_1^r$, and due to Proposition 6.4 there exists $\omega \in \mathbb{Y}$ with $\|\omega\|_{\mathbb{Y}} \leq r$ and $A^* \omega \in \partial \mathcal{R}(\tilde{z})$. Hence

$$\mathcal{R}(\tilde{z}) - \mathcal{R}(x_\alpha) \leq \langle A^* \omega, \tilde{z} - x_\alpha \rangle \leq r \|A\tilde{z} - Ax_\alpha\|_{\mathbb{Y}}.$$

From $2\alpha T_\alpha(x_\alpha, Ax) \leq 2\alpha T_\alpha(\tilde{z}, Ax)$ and the last inequality we deduce

$$\|Ax - Ax_\alpha\|_{\mathbb{Y}}^2 \leq \|Ax - A\tilde{z}\|_{\mathbb{Y}}^2 + 2\alpha r \|A\tilde{z} - Ax_\alpha\|_{\mathbb{Y}}$$
$$\leq \|Ax - Az\|_{\mathbb{Y}}^2 + 2\alpha r \|Ax - Az\|_{\mathbb{Y}} + 2\alpha r \|Ax - Ax_\alpha\|_{\mathbb{Y}}.$$

Taking the infimum over $z \in K_1^r$ and estimating the third summand using $ab \leq \frac{1}{2}a^2 + \frac{1}{2}b^2$ we obtain

$$\|Ax - Ax_\alpha\|_{\mathbb{Y}}^2 \leq \gamma_x(r)^2 + 2\alpha r \gamma_x(r) + 2\alpha^2 r^2 + \frac{1}{2} \|Ax - Ax_\alpha\|_{\mathbb{Y}}^2$$
$$\leq 2(\gamma_x(r) + \alpha r)^2 + \frac{1}{2} \|Ax - Ax_\alpha\|_{\mathbb{Y}}^2.$$

Hence $\|Ax - Ax_\alpha\|_{\mathbb{Y}} \leq 2\gamma_x(r) + 2\alpha r$ and the choice $r = \frac{1}{4\alpha} \|Ax - Ax_\alpha\|_{\mathbb{Y}}$ yields the second inequality. $\square$

As announced we see equivalence of Hölder-type bounds on γ_x and on $\|Ax - Ax_\alpha\|_{\mathbb{Y}}$ as a consequence.

Proposition 6.13. *Let $\mu \in (0, \infty)$ and $x \in \mathbb{X}$. The following statements are equivalent:*

(i) *There exists a constant $c_1 > 0$ such that $\gamma_x(r) \leq c_1 r^{-\mu}$ for all $r > 0$.*

(ii) *There exists a constant $c_2 > 0$ such that $\|Ax - Ax_\alpha\|_{\mathbb{Y}} \leq c_2 \alpha^{\frac{\mu}{1+\mu}}$ for all $\alpha > 0$ and $x_\alpha \in S_\alpha(Ax)$.*

More precisely (i) implies (ii) with $c_2 = 4 c_1^{\frac{1}{1+\mu}}$ and (ii) implies (i) with $c_1 = c_2^{1+\mu}$.

Proof.

(i)$\Rightarrow$(ii): The second inequality in Proposition 6.12 yields

$$\|Ax - Ax_\alpha\|_{\mathbb{Y}} \leq 4^{1+\mu} c_1 \alpha^\mu \|Ax - Ax_\alpha\|_{\mathbb{Y}}^{-\mu}.$$

Multiplying by $\|Ax - Ax_\alpha\|_{\mathbb{Y}}^\mu$ and taking the power $\frac{1}{1+\mu}$ yields

$$\|Ax - Ax_\alpha\|_{\mathbb{Y}} \leq 4 c_1^{\frac{1}{1+\mu}} \alpha^{\frac{\mu}{1+\mu}}.$$

(ii)$\Rightarrow$(i): Let $r > 0$. For $\alpha = c_2^{1+\mu} r^{-(1+\mu)}$ we obtain $\frac{1}{\alpha} \|Ax - Ax_\alpha\|_{\mathbb{Y}} \leq c_2 \alpha^{-\frac{1}{1+\mu}} = r$. Hence non-increasingness of γ_x and the first inequality in Proposition 6.12 yields

$$\gamma_x(r) \leq \gamma_x\left(\frac{1}{\alpha} \|Ax - Ax_\alpha\|_{\mathbb{Y}}\right) \leq \|Ax - Ax_\alpha\|_{\mathbb{Y}} \leq c_2 \alpha^{\frac{\mu}{1+\mu}} = c_2^{1+\mu} r^{-\mu}. \qquad \square$$

6.2　Banach space regularization

We apply the theory developed in the last section to Banach spaces penalties, i.e. we consider $\mathcal{R}\colon \mathbb{X} \to [0,\infty)$ given by $\mathcal{R}(x) = \frac{1}{u}\|x\|_{\mathbb{X}}^u$ for fixed $u \in (1,\infty)$. To obtain a convergence rate result we need further assumptions that we collect in the following. These involve K_1 and ϱ_1 as defined in (6.2).

Assumption 6.14. *Suppose Assumption 6.1 holds true with $\mathcal{R}(x) = \frac{1}{u}\|x\|_{\mathbb{X}}^u$ for some $u \in (1,\infty)$. Let $\mathbb{X}_-$ be a Banach space with a continuous, dense embedding $\mathbb{X} \subset \mathbb{X}_-$ such that A extends to $A\colon \mathbb{X}_- \to \mathbb{Y}$ and there exists a constant $M \geq 1$ such that*

$$\frac{1}{M}\|x\|_{\mathbb{X}_-} \leq \|Ax\|_{\mathbb{Y}} \leq M\|x\|_{\mathbb{X}_-} \quad \textit{for all} \quad x \in \mathbb{X}_-. \tag{6.9}$$

Moreover, suppose K_1 is a vector space and that there is a quasi-norm $\|\cdot\|_{\mathrm{lin}}$ on K_1 such that $(K_1, \|\cdot\|_{\mathrm{lin}})$ is a quasi-Banach space and that we have

$$\frac{1}{M}\varrho_1(x) \leq \|x\|_{\mathrm{lin}}^{u-1} \leq M\varrho_1(x) \quad \textit{for all } x \in K_1. \tag{6.10}$$

Note that injectivity is necessary for (6.9). On the other hand, injectivity of $A\colon \mathbb{X} \to \mathbb{Y}$ suffices for the existence of a space $\mathbb{X}_-$ such that (6.9) holds with $M = 1$. (Take the Banach completion of $\mathbb{X}$ in the norm $x \mapsto \|Ax\|_{\mathbb{Y}}$).

For example, in Besov space settings we will assume $\mathbb{X}_-$ to be a space with negative smoothness index, and we consider spaces $\mathbb{X}$ with smoothness index $r \geq 0$.

The assumption is motivated by Example 6.5 and the computation of K_1 for the examples in the next section.

Remark 6.15 (Topological assumptions)**.** *If $\mathbb{X}$ is reflexive we choose $\tau_{\mathbb{X}}$ to be the weak topology on $\mathbb{X}$. Then the sublevel sets of $\mathcal{R}$ are $\tau_{\mathbb{X}}$-compact by the Banach–Alaoglu theorem. Moreover, A is weak-to-weak continuous as it is bounded. Therefore, the topological conditions in Assumption 6.1 are automatically satisfied in this case.*
If the embedding $\mathbb{X} \subset \mathbb{X}_-$ is compact, then a further way to verify these assumptions that works also for non-reflexive spaces is to choose $\tau_{\mathbb{X}}$ the topology induced by the norm $\| \cdot \|_{\mathbb{X}_-}$ on $\mathbb{X}$.

K_ν via real interpolation

In the sequel we use Proposition 6.13 to show that under Assumption 6.14 image space approximation rates are characterized by real interpolation spaces. More precisely, we show that K_ν and $(\mathbb{X}_-, K_1)_{\theta,\infty}$ agree for a certain choice of ν depending on u and θ. Here the case of oversmoothing, i.e. the case $\nu < \frac{1}{2}$ is naturally included.

Proposition 6.16 (K_ν as a real interpolation space)**.** *Suppose Assumption 6.14 holds true. Let $\theta \in (0,1)$ and $\nu := \frac{\theta}{(1-\theta)(u-1)+\theta}$. Then we have $K_\nu = (\mathbb{X}_-, K_1)_{\theta,\infty}$ with*

$$C_1 \|x\|_{(\mathbb{X}_-, K_1)_{\theta,\infty}} \le \varrho_\nu(x)^{\frac{(1-\theta)(u-1)+\theta}{u-1}} \le C_2 \|x\|_{(\mathbb{X}_-, K_1)_{\theta,\infty}} \quad \text{for all } x \in (\mathbb{X}_-, K_1)_{\theta,\infty}$$

with constants $C_1, C_2 > 0$ depending only on u, θ and M.

Proof. Assume $\varrho := \varrho_\nu(x) < \infty$. Proposition 6.13 with $\mu = \frac{\theta}{(1-\theta)(u-1)}$ provides the bound

$$\gamma_x(r) \le \varrho^{\frac{(1-\theta)(u-1)+\theta}{(1-\theta)(u-1)}} r^{-\frac{\theta}{(1-\theta)(u-1)}}.$$

Let $t > 0$. We choose $r := \varrho^{(1-\theta)(u-1)+\theta} t^{-(1-\theta)(u-1)}$. If $\varepsilon > 0$, then there exists $z \in K_1$ with $\varrho_1(z) \le r$ and $\|Ax - Az\|_{\mathbb{Y}} \le \gamma_x(r) + \varepsilon$. Therefore, we obtain

$$K(x,t) \le \|x - z\|_{\mathbb{X}_-} + t\|z\|_{\mathrm{lin}} \le M(\gamma_x(r) + \varepsilon) + tM^{\frac{1}{u-1}} r^{\frac{1}{u-1}}.$$

For $\varepsilon \to 0$ we obtain

$$K(x,t) \le M\varrho^{\frac{(1-\theta)(u-1)+\theta}{(1-\theta)(u-1)}} r^{-\frac{\theta}{(1-\theta)(u-1)}} + tM^{\frac{1}{u-1}} r^{\frac{1}{u-1}} = \left(M + M^{\frac{1}{u-1}}\right) \varrho^{\frac{(1-\theta)(u-1)+\theta}{u-1}} t^\theta.$$

This proves the first inequality.
Assume $n := \|x\|_{(\mathbb{X}_-, K_1)_{\theta,\infty}} < \infty$. We prove a bound on γ_x and apply Proposition 6.13 with μ as above. Let $r > 0$. We choose $t := 2M^{\frac{1}{(1-\theta)(u-1)}} n^{\frac{1}{1-\theta}} r^{-\frac{1}{(1-\theta)(1-u)}}$. Since $2^{1-\theta} > 1$, there exists $z \in \mathbb{X}_-$ such that

$$M^{-1}\|Ax - Az\| + tM^{-\frac{1}{u-1}} \varrho_1(z)^{\frac{1}{u-1}} \le \|x - z\|_{\mathbb{X}_-} + t\|z\|_{\mathrm{lin}}$$
$$\le 2^{1-\theta} K(x,t) \le 2^{1-\theta} n t^\theta = tM^{-\frac{1}{u-1}} r^{\frac{1}{u-1}}.$$

Neglecting the first summand on the left hand side we obtain $\varrho_1(z) \le r$. Therefore,

$$\gamma_x(r) \le \|Ax - Az\|_{\mathbb{Y}} \le M2^{1-\theta} n t^\theta = 2M^{\frac{(1-\theta)(u-1)+\theta}{(1-\theta)(u-1)}} n^{\frac{1}{1-\theta}} r^{-\frac{\theta}{(1-\theta)(u-1)}}.$$

Proposition 6.13 yields $\varrho_\nu(x) \le 8M n^{\frac{u-1}{(1-\theta)(u-1)+\theta}}$. $\square$

Remark 6.17. *As already exposed in Example 6.5 we cannot expect that (6.10) holds true for ℓ^1-type norms like $b_{p,1}^r$-norms. Therefore, Proposition 6.16 does not apply in these cases. Nevertheless, one may use Proposition 6.13 directly to characterize the sets K_ν in this case. Applying Theorem 6.11 then reproduces the convergence rate results in the case of linear operators for the $0, 2, 1$-penalty in Theorem 3.15 and for weighed ℓ^1-penalties in Theorem 3.59.*

Remark 6.18 $((\mathbb{X}_-, K_1)_{\theta,\infty}$ extends $(\mathbb{X}_-, \mathbb{X})_{\theta',\infty})$**.** *Note the similarity of the source sets $(\mathbb{X}_-, \mathbb{X})_{\theta',\infty}$ in case of the convergence result for oversmoothing Banach space regularization in Theorem 5.3 to the ones in Proposition 6.16 given by $(\mathbb{X}_-, K_1)_{\theta,\infty}$. We have $(\mathbb{X}_-, \mathbb{X})_{\theta',\infty} = (\mathbb{X}_-, K_1)_{\theta,\infty}$ for $\theta' \in (0,1)$ and $\theta = \frac{u-1}{u}\theta'$. To show this one can argue as follows. In Theorem 5.3 we saw that an element in $(\mathbb{X}_-, \mathbb{X})_{\theta',\infty}$ satisfies a bound on the minimal value of the Tikhonov functional for exact data. Replacing the Tikhonov functional and the K-functional a similar estimation shows the converse statement. Hence, $(\mathbb{X}_-, \mathbb{X})_{\theta',\infty}$ contains precisely the elements satisfying a certain Hölder type bound on the minimal value. In Proposition 6.16 we characterize $(\mathbb{X}_-, K_1)_{\theta,\infty}$ by image space approximation rates. Due to Theorem 4.34 bound on the minimal value of the Tikhonov functional for exact data are equivalent to image space approximation rates.*

Therefore, the scale $(\mathbb{X}_-, K_1)_{\theta,\infty}$ may be viewed as an extension of the scale $(\mathbb{X}_-, \mathbb{X})_{\theta',\infty}$. Nevertheless, the theory we present here cannot be understood as a generalization of the results in Chapter 5 as we do not require linearity and the rather restrictive assumption on K_1 being a quasi-Banach space there.

Error bounds for Banach space regularization

We apply Theorem 6.11 to obtain error bounds measured in the norm of certain Banach spaces $\mathbb{X}_L$ with a continuous embedding $\mathbb{X}_L \subset \mathbb{X}_-$.

To this end, we consider the loss function $L\colon \mathbb{X} \times \mathbb{X} \to [0,\infty]$ given by $L(x_1, x_2) = \|x_1 - x_2\|_{\mathbb{X}_L}$ if $x_1 - x_2 \in \mathbb{X}_L$ and $L(x_1, x_2) = \infty$ if $x_1 - x_2 \notin \mathbb{X}_L$. Before we prove our convergence rate result we state a proposition that characterizes for which spaces $\mathbb{X}_L$ Hölder-type bounds on the modulus of continuity on balls of a given quasi-Banach space $\mathbb{X}_S \subset \mathbb{X}$ are satisfied.

Lemma 6.19 (bound on the modulus)**.** *Suppose Assumption 6.14 holds true. Let $\mathbb{X}_S \subset \mathbb{X}$ be quasi-Banach space, $\mathbb{X}_L$ a Banach space with continuous embeddings $\mathbb{X}_S \subset \mathbb{X}_L \subset \mathbb{X}_-$ and $e \in (0,1)$. For $\varrho > 0$ we denote*

$$K_{\mathbb{X}_S}^\varrho := \left\{ x \in \mathbb{X}_S \colon \|x\|_{\mathbb{X}_S} \le \varrho \right\}.$$

The following statements are equivalent:

  (i) There is a continuous embedding $(\mathbb{X}_-, \mathbb{X}_S)_{e,1} \subset \mathbb{X}_L$.
  (ii) There exists a constant $c > 0$ with $\omega(\delta, K_{\mathbb{X}_S}^\varrho, L) \le c\varrho^e \delta^{1-e}$ for all $\delta, \varrho > 0$.

Proof. By Lemma A.4 statement (i) is equivalent to an interpolation inequality

$$\|z\|_{\mathbb{X}_L} \le c_L \|z\|_{\mathbb{X}_-}^{1-e} \cdot \|z\|_{\mathbb{X}_S}^e \quad \text{for all} \quad z \in \mathbb{X}_S. \tag{6.11}$$

Let $x_1, x_2 \in K_{\mathbb{X}_S}^\varrho$ with $\|Ax_1 - Ax_2\|_{\mathbb{Y}} \le \delta$. The quasi-triangle inequality yields $\|x_1 - x_2\|_{\mathbb{X}_S} \le 2c\varrho$ and from (6.9) we obtain $\|x_1 - x_2\|_{\mathbb{X}_-} \le M\delta$. Hence (6.11) with $z = x_1 - x_2$ yields

$\|x_1 - x_2\|_{\mathbb{X}_L} \le CM^{1-e}(2c)^e \varrho^e \delta^{1-e}$. Taking the supremum over x_1, x_2 yields (*ii*). Assuming a bound on the modulus we obtain (6.11)

$$\|z\|_{\mathbb{X}_L} \le \Omega\left(M\|z\|_{\mathbb{X}_-}, K_{\mathbb{X}_S}^{\|z\|_{\mathbb{X}_S}}, L\right). \qquad \square$$

Now we deduce a convergence rate result as follows: From Theorem 6.11 together with Proposition 6.16 we obtain error bounds in terms of the modulus of continuity ω on bounded subsets of $(\mathbb{X}_-, K_1)_{\theta,\infty}$. Then we conclude with the Hölder type bounds on ω, that we obtain from Lemma 6.19.

Theorem 6.20 (error bounds). *Suppose Assumption 6.14 holds true and let $\mathbb{X}_L$ be a Banach space with a continuous embedding $\mathbb{X}_L \subset \mathbb{X}_-$. Let $0 < \xi < \theta < 1$ and $\delta, \varrho, \alpha > 0$ and $c_r \ge c_l > 0$, $C_D > c_D > 1$. Suppose there is a continuous embedding $(\mathbb{X}_-, K_1)_{\xi,1} \subset \mathbb{X}_L$. Assume*

$$x \in (\mathbb{X}_-, K_1)_{\theta,\infty} \quad with \quad \|x\|_{(\mathbb{X}_-, K_1)_{\theta,\infty}} \le \varrho.$$

Let $\hat{x}_\alpha \in S_\alpha(g^\delta)$ (see (6.1)). There exists a constant $C > 0$ independent of $x, g^{\mathrm{obs}}, \delta$ and ϱ such that either of the conditions

$$c_l \varrho^{-\frac{u-1}{\theta}} \delta^{\frac{(1-\theta)(u-1)+\theta}{\theta}} \le \alpha \le c_r \varrho^{-\frac{u-1}{\theta}} \delta^{\frac{(1-\theta)(u-1)+\theta}{\theta}} \quad and \quad c_D \delta \le \|g^\delta - A\hat{x}_\alpha\|_{\mathbb{Y}} \le C_D \delta$$

implies the bound

$$\|x - \hat{x}_\alpha\|_{\mathbb{X}_L} \le C \varrho^{\frac{\xi}{\theta}} \delta^{1-\frac{\xi}{\theta}}.$$

Proof. For ν as in Proposition 6.16 the second inequality therein yields

$$x \in K_{(\mathbb{X}_-, K_1)_{\theta,\infty}}^{\varrho} \subset K_\nu^{\overline{\varrho}}$$

with $\overline{\varrho} = (C_2 \varrho)^{\frac{u-1}{(1-\theta)(u-1)+\theta}}$.

In view of Theorem 6.11 it remains to prove an upper bound on $\omega(c_1\delta, K_\nu^{c_2\overline{\varrho}}, L) \le C\varrho^{\frac{\xi}{\theta}}\delta^{1-\frac{\xi}{\delta}}$ for constants $c_1, c_2 > 0$ given therein. The first inequality in Proposition 6.16 provides

$$K_\nu^{c_2\overline{\varrho}} \subset K_{(\mathbb{X}_-, K_1)_{\theta,\infty}}^{c_3\varrho} \quad with \quad c_3 = C_1^{-1} C_2 c_2^{\frac{(1-\theta)(u-1)+\theta}{u-1}}.$$

The reiteration theorem (see Proposition A.6) yields

$$(\mathbb{X}_-, K_1)_{\xi,1} = \left(\mathbb{X}_-, (\mathbb{X}_-, K_1)_{\theta,\infty}\right)_{\frac{\xi}{\theta},1} \tag{6.12}$$

with equivalent quasi-norms. In particular $(\mathbb{X}_-, K_1)_{\theta,\infty} \subset (\mathbb{X}_-, K_1)_{\xi,1} \subset \mathbb{X}_L \subset \mathbb{X}_-$ (see [8, Thm. 3.4.1 (c)&(d), Sec. 3.11]). Hence Lemma 6.19 with $\mathbb{X}_S = (\mathbb{X}_-, K_1)_{\theta,\infty}$ yields a constant c_4 with

$$\omega\left(c_1\delta, K_\nu^{c_2\overline{\varrho}}, L\right) \le \omega\left(c_1\delta, K_{(\mathbb{X}_-, K_1)_{\theta,\infty}}^{c_3\varrho}, L\right) \le C\varrho^{\frac{\xi}{\theta}}\delta^{1-\frac{\xi}{\theta}}$$

with $C = c_4 c_3^{\frac{\xi}{\theta}} c_1^{1-\frac{\xi}{\theta}}$.

For the discrepancy principle the bound $\omega(d_1\delta, K_\nu^{d_2\overline{\varrho}}, L) \le C\varrho^{\frac{\xi}{\theta}}\delta^{1-\frac{\xi}{\delta}}$ follows by replacing c_1 by d_1 and c_2 by d_2. $\qquad \square$

Remark 6.21 (Limiting case $\theta = 1$)**.** *The statement of the theorem remains valid in the limiting case $\theta = 1$ where the source condition in terms of $(\mathbb{X}_-, K_1)_{\theta,\infty}$ has to be replaced by simply $x \in K_1$ with $\|x\|_{\text{lin}} \leq \varrho$. Here the a priori rule is $\alpha \sim \varrho^{-(u-1)}\delta$. The proof follows along the same lines as above leaving out the step involving the reiteration theorem.*

Remark 6.22. *The relation $(\mathbb{X}_-, K_1)_{\xi,1} \subset \mathbb{X}_L$ is necessary to obtain error bounds as in Theorem 6.20 in the following sense: Assuming $\mathbb{X}_L$ satisfies an error bound*

$$\|x - \hat{x}_\alpha\|_{\mathbb{X}_L} \leq C\varrho^e \delta^{1-e}$$

for some $e \in (0, 1)$ and all $x \in K^\varrho_{(\mathbb{X}_-, K_1)_{\theta,\infty}}$ under some a priori parameter choice $\alpha = \alpha(\delta)$, then the lower bound in Proposition 2.6 yields

$$\frac{1}{2}\omega\left(2\delta, K^\varrho_{(\mathbb{X}_-, K_1)_{\theta,\infty}}, L\right) \leq \Delta_{R_{\alpha(\delta)}}\left(\delta, K^\varrho_{(\mathbb{X}_-, K_1)_{\theta,\infty}}, L\right) \leq C\varrho^e \delta^{1-e}.$$

Thus the converse implication in Lemma 6.19 and the identity (6.12) *provide*

$$(\mathbb{X}_-, K_1)_{\theta e,1} = \left(\mathbb{X}_-, (\mathbb{X}_-, K_1)_{\theta,\infty}\right)_{e,1} \subset \mathbb{X}_L.$$

In the sequel we illustrate the impact of this result by applying it to three more concrete Banach space regularization setups: weighed ℓ^p-, Besov $b^r_{p,p}$- and $b^r_{2,q}$-regularization. Before that we finish this section by providing a tool that helps to find a quasi-norm satisfying (6.10). Note that the density $\mathbb{X} \subset \mathbb{X}_-$ allows us to view the adjoint of the embedding as an embedding $\mathbb{X}'_A \subset \mathbb{X}'$.

Lemma 6.23. *Suppose Assumption 6.14 holds true. We have $x \in K_1$ if and only if $\partial\mathcal{R}(x) \cap \mathbb{X}'_A \neq \emptyset$. The function*

$$\overline{\varrho}_1 \colon \mathbb{X}_- \to [0, \infty] \quad \text{given by} \quad \overline{\varrho}_1(x) = \inf\left\{\|\xi\|_{\mathbb{X}'_-} : \xi \in \partial\mathcal{R}(x) \cap \mathbb{X}'_-\right\}$$

satisfies

$$\frac{1}{M}\varrho_1(x) \leq \overline{\varrho}_1(x) \leq M\varrho_1(x) \quad \text{for all } x \in \mathbb{X}.$$

Proof. Suppose $\xi \in \partial\mathcal{R}(x) \cap \mathbb{X}'_-$. Let $z \in \mathbb{X}_-$, then

$$\langle \xi, z \rangle \leq \|\xi\|_{\mathbb{X}'_-}\|z\|_{\mathbb{X}_-} \leq M\|\xi\|_{\mathbb{X}'_-}\|Az\|_{\mathbb{Y}}.$$

Therefore, Proposition A.2 provides $\omega \in \mathbb{Y}$ with $\|\omega\|_{\mathbb{Y}} \leq M\|\xi\|_{\mathbb{X}'_-}$ and $A^*\omega = \xi \in \partial\mathcal{R}(x)$. With Proposition 6.4 we obtain $\varrho_1(x) \leq M\|\xi\|_{\mathbb{X}'_-}$. This shows the first inequality. Let $\omega \in \mathbb{Y}$ such that $A^*\omega \in \partial\mathcal{R}(x)$. Then

$$\langle A^*\omega, z \rangle = \langle \omega, Az \rangle \leq \|\omega\|_{\mathbb{Y}}\|Az\|_{\mathbb{Y}} \leq M\|\omega\|_{\mathbb{Y}}\|z\|_{\mathbb{X}_-}$$

for all $z \in \mathbb{X}$. Hence $\overline{\varrho}_1(x) \leq \|A^*\omega\|_{\mathbb{X}'_-} \leq M\|\omega\|_{\mathbb{Y}}$. Again by Proposition 6.4 this proves the second inequality. $\qquad\square$

6.3 Weighed ℓ^p-regularization

We will prove a convergence rate result for weighed ℓ^p-regularization with $p > 1$ (see Section 2.5). Throughout this section we suppose $A\colon \ell_{\bar{a}}^2 \to \mathbb{Y}$ is linear and satisfies

$$\frac{1}{M}\|z\|_{\bar{a},2} \le \|Az\|_{\mathbb{Y}} \le M\|z\|_{\bar{a},2} \quad \text{for all } z \in \ell_{\bar{a}}^2$$

for a constant $M > 0$. Moreover, let $p\ in(1,2)$ and $\bar{r} = (\bar{r}_j)_{j\in\Lambda}$ such that $\bar{a}_j\bar{r}_j^{-1}$ is bounded above. Then there is a continuous embedding $\ell_{\bar{r}}^p \subset \ell_{\bar{a}}^2$. We consider

$$S_\alpha(g) = \operatorname*{argmin}_{z\in\ell_p^{\bar{r}}} \left(\frac{1}{2\alpha}\|g - Az\|_{\mathbb{Y}} + \frac{1}{p}\|z\|_{\bar{r},p}^p \right). \tag{6.13}$$

Hence, we have $\mathbb{X}_- := \ell_{\bar{a}}^2$ and $\mathbb{X} := \ell_{\bar{r}}^p$ and $u = p$. As $p > 1$ the space $\ell_{\bar{r}}^p$ is reflexive. Therefore, choosing the weak topology on $\ell_{\bar{r}}^p$ the topological assumptions in Assumption 6.1 are satisfied (see also Remark 6.15). In particular, Proposition 2.3 yields existence of the minimizers, i.e. $S_\alpha(g) \ne \emptyset$ for all $g \in \mathbb{Y}$.

Using Lemma 6.23 we compute the set K_1 of elements satisfying a classical source condition. Similar considerations can also be found in [78, Rem. 3.5].

Lemma 6.24 (K_1 for weighted p-norm penalties)**.** *Let $\bar{s} = \bar{a}^{-\frac{1}{p-1}}\bar{r}^{\frac{p}{p-1}}$. Then $K_1 = \ell_{\bar{s}}^{2p-2}$ with*

$$\frac{1}{M}\varrho_1(x) \le \|x\|_{\bar{s},2p-2}^{p-1} \le M\varrho_1(x) \quad \text{for all } x \in \ell_{\bar{a}}^2.$$

Proof. Let $x \in \ell_{\bar{r}}^p$. Then $\partial\mathcal{R}(x) = \{\xi\}$ with $|\xi_j| = r_j^p|x_j|^{p-1}$. With $\bar{\varrho}_1$ as in Lemma 6.23 and in view of Proposition A.8 we obtain

$$\bar{\varrho}_1(x) = \|\xi\|_{a^{-1},2} = \left(\sum\nolimits_{j\in\Lambda} a_j^{-2}r_j^{2p}|x_j|^{2p-2} \right)^{1/2} = \|x\|_{s,2p-2}^{p-1}.$$

Lemma 6.23 yields the result. $\qquad\square$

Therefore, (6.10) holds true with $\|\cdot\|_{\mathrm{lin}} := \|\cdot\|_{s,2p-2}$.

The real interpolation spaces $(\ell_{\bar{a}}^2, \ell_{\bar{s}}^{2p-2})_{\theta,\infty}$ are weighted weak ℓ^p-spaces which we introduce in the following. Let $\mu = (\mu_j)_{j\in\Lambda}$ and $\nu = (\nu_j)_{j\in\Lambda}$ sequences of positive reals and $t \in (0,\infty)$. Then those are defined by the following quasi-norms

$$\ell_{\mu,\nu}^{t,\infty} = \{x \in \mathbb{R}^\Lambda\colon \|x\|_{\mu,\nu,t} < \infty\} \quad \text{with } \|x\|_{\mu,\nu,t}^t = \sup_{\tau>0}\left(\tau^t \sum\nolimits_{j\in\Lambda} \nu_j \mathbb{1}_{\{\mu_j|x_j|>\tau\}} \right).$$

We apply Theorem 6.20 and obtain the following result.

Theorem 6.25 (error bounds for weighted p-norm penalties)**.** *Let $p \in (1,2)$, $t \in (2p-2, p)$ and $\delta, \varrho, \alpha > 0$ and $c_r \ge c_l > 0$, $C_D > c_D > 1$ and $\mu := (\bar{a}^2\bar{r}^{-p})^{\frac{1}{2-p}}$, $\nu := (\bar{a}^{-1}\bar{r})^{\frac{2p}{2-p}}$. Assume $x \in \ell_{\mu,\nu}^{t,\infty}$ with $\|x\|_{\mu,\nu,t} \le \varrho$, $g^{\mathrm{obs}} \in \mathbb{Y}$ with $\|g^{\mathrm{obs}} - Ax\|_{\mathbb{Y}} \le \delta$ and $\hat{x}_\alpha \in S_\alpha(g^\delta)$ (see (6.13)). There exists a constant $C > 0$ independent of $x, g^{\mathrm{obs}}, \delta$ and ϱ such that whenever α satisfies either*

$$c_l\varrho^{-\frac{t(2-p)}{2-t}}\delta^{\frac{2(2-p)}{2-t}} \le \alpha \le c_r\varrho^{-\frac{t(2-p)}{2-t}}\delta^{\frac{2(2-p)}{2-t}} \quad \text{or} \quad c_D\delta \le \|g^\delta - A\hat{x}_\alpha\|_{\mathbb{Y}} \le C_D\delta$$

the bound

$$\|x - \hat{x}_\alpha\|_{\overline{r},p} \leq C\varrho^{\frac{t(2-p)}{p(2-t)}}\delta^{\frac{2(p-t)}{p(2-t)}}$$

holds true.

Proof. By Lemma 6.24 and [45, Thm. 2, Rem.] we have

$$\ell_{\overline{r}}^p = \left(\ell_{\overline{a}}^2, \ell_{\overline{s}}^{2p-2}\right)_{\xi,p} = (\mathbb{X}_-, K_1)_{\xi,p} \quad \text{with} \quad \xi := \frac{p-1}{p}.$$

By [8, Thm. 3.4.1 (b); Sec. 3.11] there is a continuous embedding $\left(\ell_{\overline{a}}^2, \ell_{\overline{s}}^{2p-2}\right)_{\xi,1} \subset \ell_{\overline{r}}^p$. Hence the choice $\mathbb{X}_L = \ell_{\overline{r}}^p$ satisfies the assumption of Theorem 6.20.
The interpolation spaces $(\mathbb{X}_-, K_1)_{\theta,\infty} = \left(\ell_{\overline{a}}^2, \ell_{\overline{s}}^{2p-2}\right)_{\theta,\infty}$ are characterized by weighted weak ℓ^p-spaces $\ell_{\mu,\nu}^{t,\infty}$ in the following manner:

$$\ell_{\mu,\nu}^{t,\infty} = (\ell_{\overline{a}}^2, \ell_{\overline{s}}^{2p-2})_{\theta,\infty} \text{ with } \frac{1}{t} = \frac{1-\theta}{2} + \frac{\theta}{2p-2}, \ \mu := (\overline{a}^2\overline{r}^{-p})^{\frac{1}{2-p}}, \ \nu := (\overline{a}^{-1}\overline{r})^{\frac{2p}{2-p}}$$

with equivalent quasi-norms (see [45, Thm. 2]).
The application of Theorem 6.20 yields the result. $\qquad\square$

Remark 6.26 (Limiting case $t = 2p-2$). *Due to Remark 6.21 the statement remains valid in the limiting case $t = 2p - 2$ if one replaces $\ell_{\mu,\nu}^{t,\infty}$ by $K_1 = \ell_{\overline{s}}^{2p-2}$ with $\overline{s} = \overline{a}^{-\frac{1}{p-1}}\overline{r}^{\frac{p}{p-1}}$. Here we obtain the rate $\|x - \hat{x}_\alpha\|_{\overline{r},p} \leq C\varrho^{\frac{p-1}{p}}\delta^{\frac{1}{p}}$. In [49] the convergence rate $\mathcal{O}(\delta^{\frac{1}{p}})$ is already proven under a condition similar to (2.4). Here we obtain intermediate convergences rates between $\mathcal{O}(\delta^0)$ and $\mathcal{O}(\delta^{\frac{1}{p}})$. This has the advantage that we obtain statements on the speed of convergences on larger sets.*

Remark 6.27 ($p = 1$). *Theorem 6.25 remains valid word by word in the case $p = 1$ (see Theorem 3.59).*

6.4   Besov r, p, p-Penalties

We consider Besov space wavelet regularization with a finitely smoothing operator as introduced in Section 2.4. Let $B_{p,q}^s$ be a Besov scale that is p-monotone as in Definition 2.9. We assume linearity of the forward operator $A\colon B_{2,2}^{-a} \to \mathbb{Y}$ and that there is a constant $\tilde{M} \geq 1$ such that

$$\frac{1}{\tilde{M}}\|f \mid B_{2,2}^{-a}\| \leq \|Ax\|_{\mathbb{Y}} \leq \tilde{M}\|f \mid B_{2,2}^{-a}\| \quad \text{for all } f \in B_{2,2}^{-a}. \tag{6.14}$$

Moreover, throughout this section we suppose that Assumption 2.11 holds true with $s_{\max} > a$ and that the number of coefficients in each level satisfies (2.16). The wavelet synthesis operator $\mathcal{S}\colon b_{2,2}^{-a} \to B_{2,2}^{-a}$ is a Banach space isomorphism. Hence, there exists a constant $M \geq 1$ such that

$$\frac{1}{M}\|x\|_{-a,2,2} \leq \|A\mathcal{S}x\|_{\mathbb{Y}} \leq M\|x\|_{-a,2,2} \quad \text{for all } x \in b_{2,2}^{-a}. \tag{6.15}$$

With $p \in (1, \infty)$ and $r \in [0, s_{\max})$ such that $a + r \geq d \left(\frac{1}{p} - \frac{1}{2} \right)$ we obtain a continuous embedding $b_{p,p}^r \subset b_{2,2}^{-a}$ (see [101, 3.3.1.(6),(7), 3.2.4.(1)]) and consider

$$R_\alpha(g) = \mathcal{S}\hat{x}_\alpha \quad \text{with} \quad \hat{x}_\alpha \in \operatorname*{argmin}_{z \in b_{p,p}^r} \left(\frac{1}{2\alpha} \|g - A\mathcal{S}z\|_{\mathbb{Y}}^2 + \frac{1}{p} \|z\|_{r,p,p}^p \right). \tag{6.16}$$

We set $\mathbb{X}_- = b_{2,2}^{-a}$ and $\mathbb{X} = b_{p,p}^r$. Recall that we have

$$b_{p,p}^s = \ell_{\overline{\omega}_{s,p}}^p \quad \text{with equal norm for} \quad (\overline{\omega}_{s,p})_{(j,k)} = 2^{j(s + \frac{d}{2} - \frac{d}{p})}.$$

Hence for $p < 2$, this example is a special case of Section 6.3.
As above we use Lemma 6.23 to compute K_1.

Lemma 6.28 (K_1 for for r, p, p-penalties)**.** *Let $\tilde{s} := \frac{a + rp}{p - 1}$ and $\tilde{t} := 2p - 2$. Then $K_1 = b_{\tilde{t}, \tilde{t}}^{\tilde{s}}$ with*

$$\frac{1}{M} \varrho_1(x) \leq \|x\|_{\tilde{s}, \tilde{t}, \tilde{t}}^{p-1} \leq M \varrho_1(x) \quad \text{for all } x \in b_{2,2}^{-a}.$$

Proof. The proof works along the lines of the proof of Lemma 6.24 by identifying the expression for $\|\xi\|_{a,2,2}$ with $\|x\|_{\tilde{s}, \tilde{t}, \tilde{t}}$. $\qquad\square$

For $r < s < \tilde{s}$ we set

$$k_s := \left(b_{2,2}^{-a}, b_{\tilde{t}, \tilde{t}}^{\tilde{s}} \right)_{\theta, \infty} \quad \text{with} \quad \theta = \frac{p-1}{p} \frac{a}{s+a}. \tag{6.17}$$

We apply Theorem 6.20 to obtain convergence rates for functions whose wavelet coefficients belong to k_s. After that we give necessary and sufficient conditions in terms of spaces $B_{p,q}^s$ for the latter smoothness condition.

Theorem 6.29 (Error bound for r, p, p-penalties)**.** *Let $r < s < \tilde{s}$ and $\delta, \varrho, \alpha > 0$, $c_r \geq c_l > 0$, $C_D > c_D > 1$. Assume $f = \mathcal{S}x$ for $x \in k_s$ with $\|x\|_{k_s} \leq \varrho$, $g^{\mathrm{obs}} \in \mathbb{Y}$ with $\|g^{\mathrm{obs}} - Af\|_{\mathbb{Y}} \leq \delta$ and $\hat{f}_\alpha \in R_\alpha(g^{\mathrm{obs}})$. (see (6.16)). Let $\overline{p} := \frac{2p(a+r)}{2a+pr}$. There is a constant $C > 0$ independent of f, g^{obs}, δ and ϱ such that either of the conditions*

$$c_l \varrho^{-\frac{pa}{s+a}} \delta^{\frac{(2-p)s+2a}{s+a}} \leq \alpha \leq c_r \varrho^{-\frac{pa}{s+a}} \delta^{\frac{(2-p)s+2a}{s+a}} \quad \text{and} \quad c_D \delta \leq \|g^\delta - A\hat{f}_\alpha\|_{\mathbb{Y}} \leq C_D \delta$$

implies the bound

$$\|f - \hat{f}_\alpha \,|\, B_{\overline{p}, \overline{p}}^0\| \leq C \varrho^{\frac{a}{s+a}} \delta^{\frac{s}{s+a}}.$$

Proof. We apply Theorem 6.20 with $\mathbb{X}_- = b_{2,2}^{-a}$ and $\mathbb{X} = b_{p,p}^r$. By Lemma 6.28 we have (6.10) with $K_1 = b_{\tilde{t}, \tilde{t}}^{\tilde{s}}$. The identification [45, Thm. 2, Rem.] for $p \neq 2$ and [101, 3.3.6.(9)] for $p = 2$ yield

$$b_{\overline{p}, \overline{p}}^0 = \left(b_{2,2}^{-a}, b_{\tilde{t}, \tilde{t}}^{\tilde{s}} \right)_{\xi, p} = (\mathbb{X}_-, K_1)_{\xi, p} \quad \text{with} \quad \xi = \frac{p-1}{p} \frac{a}{a+r}. \tag{6.18}$$

Hence, the choice $\mathbb{X}_L = b_{\overline{p}, \overline{p}}^0$ satisfies the assumption in Theorem 6.20 as $\overline{p} > 1$ implies that there exists a continuous embedding $(\cdot, \cdot)_{\xi, 1} \subset (\cdot, \cdot)_{\xi, \overline{p}}$ (see [8, Thm. 3.4.1 (b); Sec. 3.11]). We apply Theorem 6.20 and obtain first a bound $\|x - \hat{x}_\alpha\|_{0, \overline{p}, \overline{p}} \leq C \varrho^{\frac{a}{s+a}} \delta^{\frac{s}{s+a}}$. With this the desired bound follows from boundedness of $\mathcal{S} \colon b_{\overline{p}, \overline{p}}^0 \to B_{\overline{p}, \overline{p}}^0$. $\qquad\square$

Remark 6.30 (Limiting case $s = \tilde{s}$). *In the limiting case $s = \tilde{s}$ the result remains valid if one replaces k_s by $K_1 = b_{\tilde{t},\tilde{t}}^{\tilde{s}}$ and we obtain the bound $\|f - \hat{f}_\alpha\|_{0,p,p} \leq C\varrho^{\frac{p-1}{p}}\delta^{\frac{1}{p}}$ (see Remark 6.21).*

Remark 6.31 (Error bound in $L^{\bar{p}}$). *Let $p \leq 2$. As in Corollary 3.16 the error bound measured in the $B_{\bar{p},\bar{p}}^0$-norm implies $\|f - \hat{f}_\alpha\|_{L^{\bar{p}}(\Omega)} = \mathcal{O}(\varrho^{\frac{a}{a+s}}\delta^{\frac{s}{s+a}})$ if $B_{\bar{p},\bar{p}}^0 = B_{\bar{p},\bar{p}}^0(\Omega)$ with Ω being either $\mathbb{T}^d, \mathbb{R}^d$ or a bounded Lipschitz domain in $\mathbb{R}^d$ (see also Proposition A.11).*

Remark 6.32 ($p = 1$). *The above result holds true word for word in the case $p = 1$ (see Theorem 3.65).*

Remark 6.33 (Influence of $r > 0$). *A larger value of $r > 0$ leads to a larger $\tilde{s}$, such that our result is able to guarantee faster convergence rates (under stronger assumptions) in this case.*

For $p = 2$ we have $k_s = b_{2,\infty}^s$ (see [101, 3.3.6.(9)]). Hence if $r < s < \min\{s_{\max}, \tilde{s}\}$ then the smoothness condition in Theorem 6.29 is satisfied if and only if $f \in B_{2,\infty}^s$ in this case. To obtain a similar statement in the case $p \neq 2$ we start by showing a nesting of k_s for $p \neq 2$ by Besov sequence spaces.

Lemma 6.34. *Let $r < s < \tilde{s}$ and $t = \frac{2p(a+r)}{s(2-p)+2a+pr}$.*
(a) For $p < 2$ we have continuous embeddings

$$b_{t,t}^s \subset k_s \subset b_{t-\varepsilon,\infty}^s \quad \text{for all} \quad 0 < \varepsilon < t.$$

(b) For $p > 2$ we have continuous embeddings $b_{t,t}^s \subset k_s \subset b_{2,\infty}^s$.

Proof. Let $\theta = \frac{p-1}{p}\frac{s+a}{a}$. Then $\frac{1}{t} = \frac{1-\theta}{2} + \frac{\theta}{\tilde{t}}$ and $s = (1-\theta)(-a) + \theta\tilde{s}$. With [8, Thm. 3.4.1 (b)] we obtain

$$b_{t,t}^s = \left(b_{2,2}^{-a}, b_{\tilde{t},\tilde{t}}^{\tilde{s}}\right)_{\theta,t} \subset \left(b_{2,2}^{-a}, b_{\tilde{t},\tilde{t}}^{\tilde{s}}\right)_{\theta,\infty} = k_s$$

in both cases.
Suppose $p < 2$. Then $\tilde{t} < 2$ and $t \in (\tilde{t}, 2)$. Let $\varepsilon > 0$ such that $t - \varepsilon \in (\tilde{t}, 2)$. There are $s < s' < \tilde{s}$ and $\theta < \theta' < 1$ such that $b_{t-\varepsilon,t-\varepsilon}^{s'} = \left(b_{2,2}^{-a}, b_{\tilde{t},\tilde{t}}^{\tilde{s}}\right)_{\theta',t-\varepsilon}$. The reiteration theorem (see Proposition A.6) yields $k_s = \left(b_{2,2}^{-a}, b_{t-\varepsilon,t-\varepsilon}^{s'}\right)_{\frac{\theta}{\theta'},\infty}$. From $t - \varepsilon < 2$ we obtain the continuous embeddings $b_{2,2}^{-a} \subset b_{2,\infty}^{-a} \subset b_{t-\varepsilon,\infty}^{-a}$. Together with the interpolation identity $b_{t-\varepsilon,\infty}^s = (b_{t-\varepsilon,\infty}^{-a}, b_{t-\varepsilon,\infty}^r)_{\theta,\infty}$ we get the second inclusion using [101, 2.4.1 Rem. 4]. Due to p-monotonicity we obtain the second inclusion for all $0 < \epsilon < t$.
For $p > 2$ we have $\tilde{t} > 2$. Hence $b_{\tilde{t},\tilde{t}}^{\tilde{s}} \subset b_{2,\tilde{t}}^{\tilde{s}}$. Once again [101, 2.4.1 Rem. 4] yields $k_s \subset \left(b_{2,2}^{-a}, b_{2,\tilde{t}}^{\tilde{s}}\right)_{\theta,\infty} = b_{2,\infty}^s$. $\qquad\square$

We use this to give a sufficient condition on f such that $f = \mathcal{S}x$ with $x \in k_s$.

Corollary 6.35 (Smoothness condition on f). *Let $p \in (1,\infty)$, $r < s < \min\{s_{max}, \tilde{s}\}$, $t = \frac{2p(a+r)}{s(2-p)+2a+pr}$ and $\varrho > 0$. There exists a constant c such that $f \in B_{t,t}^s$ with $\|f \mid B_{t,t}^s\| \leq \varrho$ implies $f = \mathcal{S}x$ with $\|x\|_{k_s} \leq c\varrho$.*

Proof. Due to Assumption 2.11 $\mathcal{S}^{-1}\colon B^s_{t,t} \to b^s_{t,t}$ is bounded. Lemma 6.34 yields a continuous embedding $b^s_{t,t} \subset k_s$. □

Note that from Lemma 6.34 one can similarly also derive necessary conditions for the smoothness condition in Theorem 6.29.

In the article [104] the authors achieve the minimax optimal convergence rate for s, p, q-penalties only under the restriction $q \geq 2$. For $q < 2$ the rates obtained there by variational source conditions and Bregman distances are suboptimal. Here we obtain the optimal convergence rate for all $1 \leq p = q < \infty$.

The same arguments as in Example 3.67 shows that describing the regularity of functions with jumps or kinks via their wavelet expansion in terms of k_s allows for a higher value of s then using $B^p_{s,\infty}$. Therefore, we obtain a faster convergence rate for this class of functions. For $p > 2$ we measure the error in a stronger norm than the L^2-norm. On the other hand, the set on which we obtain convergence rates is smaller than $b^s_{2,\infty}$.

6.5 Besov $r, 2, q$-Penalties

Once again we consider wavelet regularization as in Section 2.4. Let $B^s_{p,q}$ be a Besov scale. Here we do not have to require p-monotonicity. Again we assume $A\colon B^{-a}_{2,2} \to \mathbb{Y}$ to be linear such that (6.14) and Assumption 2.11 hold true with $s_{\max} > a$ (we do not need (2.16) here).

With $q \in (1, \infty)$ and $r \in [0, s_{\max})$ we obtain a continuous embedding $b^r_{2,q} \subset b^{-a}_{2,2}$ and consider

$$R_\alpha(g) = \mathcal{S}\hat{x}_\alpha \quad \text{with} \quad \hat{x}_\alpha \in \operatorname*{argmin}_{z \in b^r_{2,q}} \left(\frac{1}{2\alpha} \|g - A\mathcal{S}z\|^2_{\mathbb{Y}} + \frac{1}{q} \|z\|^q_{r,2,q} \right). \tag{6.19}$$

We set $\mathbb{X}_- = b^{-a}_{2,2}$ and $\mathbb{X} = b^r_{2,q}$. As above (6.14) transfers to a lower and upper bound for $A \circ \mathcal{S}$ (see (6.15)).

Once again we start by computing K_1.

Lemma 6.36 (K_1 for for $r, 2, q$-penalties). *Let $\tilde{s} := \frac{a+rq}{q-1}$ and $\tilde{q} := 2q - 2$. Then $K_1 = b^{\tilde{s}}_{2,\tilde{q}}$ with*

$$\frac{1}{M} \varrho_1(x) \leq \|x\|^{q-1}_{\tilde{s},2,\tilde{q}} \leq M \varrho_1(x) \quad \text{for all } x \in b^{-a}_{2,2}.$$

Proof. If $x \in b^0_{2,q}$, then $\partial \mathcal{R}(x) = \{\xi\}$ with $\xi_{j,k} = 2^{jrq}\|x_j\|^{q-2}|x_{j,k}|$. With $\bar{\varrho}_1$ as in Lemma 6.23 and using Proposition A.8 we obtain $\bar{\varrho}_1(x) = \|\xi\|_{a,2,2} = \|x\|^{q-1}_{\tilde{s},2,\tilde{q}}$. Lemma 6.23 yields the result. □

Next we deduce the convergence rate result from Theorem 6.20.

Theorem 6.37 (Error bounds for $0, 2, q$-penalties). *Let $r < s < \min\{s_{max}, \tilde{s}\}$ and $\delta, \varrho, \alpha > 0$, $c_r \geq c_l > 0$, $C_D > c_D > 1$. Assume $f \in B^s_{2,\infty}$ with $\|f \mid B^s_{2,\infty}\| \leq \varrho$, $g^{\mathrm{obs}} \in \mathbb{Y}$ with $\|g^{\mathrm{obs}} - A\hat{f}_\alpha\|_{\mathbb{Y}} \leq \delta$ and $\hat{f}_\alpha \in R_\alpha(g^{\mathrm{obs}})$ (see (6.19)). There exists a constant $C > 0$ independent of x, g^{obs}, δ and ϱ such that either of the conditions*

$$c_l \varrho^{-\frac{qa}{s+a}} \delta^{\frac{(2-q)s+2a}{s+a}} \leq \alpha \leq c_r \varrho^{-\frac{qa}{s+a}} \delta^{\frac{(2-q)s+2a}{s+a}} \quad \text{and} \quad c_D \delta \leq \|g^\delta - A\hat{x}_\alpha\|_{\mathbb{Y}} \leq C_D \delta$$

implies the bound

$$\|f - \hat{f}_\alpha \mid B^0_{2,2}\| \leq C \varrho^{\frac{a}{s+a}} \delta^{\frac{s}{s+a}}.$$

Proof. We verify Assumption 6.14 with $\mathbb{X}_- = b^{-a}_{2,2}$, $\mathbb{X} = b^r_{2,q}$ and $\mathbb{X}_L = b^0_{2,2}$. As already mentioned above the requirement (6.9) follows as in (6.15). Due to Lemma 6.36 condition (6.10) is satisfied. Moreover, we have

$$b^0_{2,2} = \left(b^{-a}_{2,2}, b^{\tilde{s}}_{2,\tilde{q}}\right)_{\xi,2} \quad \text{with} \quad \xi = \frac{q-1}{q}\frac{a}{a+r}.$$

Therefore, the choice $\mathbb{X}_L = b^0_{2,2}$ satisfies the assumption on $\mathbb{X}_L$ in Theorem 6.20. Finally, for $r < s < \tilde{s}$ we have

$$b^s_{2,\infty} = \left(b^{-a}_{2,2}, b^{\tilde{s}}_{2,\tilde{q}}\right)_{\theta,2} \quad \text{with} \quad \theta = \frac{q-1}{q}\frac{s+a}{a}.$$

Due to Assumption 2.11 we have $x := \mathcal{S}^{-1}f \in b^s_{2,\infty}$ with $\|x\|_{s,2,\infty} \leq c_1\varrho$ with c_1 the operator norm of $\mathcal{S}^{-1} \colon B^s_{2,\infty} \to b^s_{2,\infty}$. Therefore, Theorem 6.20 yields the error bound $\|x - \hat{x}_\alpha\|_{0,2,2} \leq c_2\varrho^{\frac{a}{s+a}}\delta^{\frac{s}{s+a}}$ with $\hat{x}_\alpha$ a minimizer in (6.19). Due to boundedness of $\mathcal{S} \colon b^0_{2,2} \to B^0_{2,2}$ this bound on the wavelet coefficients transfers to the desired error bound. $\qquad\square$

Remark 6.38 (L^2-loss). *As in Corollary 3.16 the error bound measured in the $B^0_{2,2}$-norm equals a bound $\|f - \hat{f}_\alpha\|_{L^2(\Omega)} = \mathcal{O}\left(\varrho^{\frac{a}{a+s}}\delta^{\frac{s}{s+a}}\right)$ if $B^0_{2,2} = B^0_{2,2}(\Omega)$ with Ω either $\mathbb{T}^d, \mathbb{R}^d$ or a bounded Lipschitz domain in $\mathbb{R}^d$ (see also Proposition A.11).*

Remark 6.39 ($q = 1$). *The above result holds true word for word in the case $q = 1$ (see Theorem 3.15).*

Remark 6.40 (Limiting case $s = \tilde{s}$). *In the limiting case $s = \tilde{s}$ the result remains valid if one replaces $b^s_{2,\infty}$ by $K_1 = b^{\tilde{s}}_{2,\tilde{q}}$ with $\tilde{q} = 2q - 2$. Here we obtain $\|f - \hat{f}_\alpha \mid B^0_{2,2}\| \leq C\varrho^{\frac{q-1}{q}}\delta^{\frac{1}{q}}$.*

Here the results for $q \geq 2$ agree with the ones in [104]. Theorem 6.37 proves the order optimal convergence rate for all $q \in (1, \infty)$. In particular for $q < 2$ our result improves the convergence rate in the latter reference.

Note that we measure the error in the $B^0_{2,2}$-norm independent of the value of q, i.e. the error norm is not dictated by the penalty term.

Suppose $r = 0$. Then the smaller q the larger is the region $r < s < \tilde{s} = \frac{a}{q-1}$ of regularity parameters for which we guarantee upper bounds. Furthermore, we see that changing the fine index q while keeping $p = 2$ does not change the set where convergence rates are guaranteed, but it influences the parameter choice rule.

Chapter Seven

Discussion and outlook

In this work we have successfully derived order optimal convergence rates for wavelet regularization in Besov scales. In particular, we believe that with our contributions the convergence rate theory for sparsity promoting $r, p, 1$-penalties has now reached a satisfactory stage in many respects, e.g., concerning minimax convergence rates and converse results for approximation rates. Nevertheless, starting with Chapter 3 we would like to point out some open problems and potential extensions:

- Remark 3.8 shows that we cannot expect that our strategy for the verification of variational source condition (see Proposition 3.5 and 3.29) generalizes to Besov space r, p, q-penalties with $q > 1$. Nevertheless, an extension to the non convex case $0 < q < 1$ seems straightforward by adapting the the arguments in [3, Thm. 4.3.].

- A further interesting extension concerns redundant frames. In Section 3.8 we showed how to extend the convergence rate results in Chapter 3 to redundant frames when we consider the Tikhonov functional on function spaces evoking the analysis operator of the frame to define the penalty term (see Remark 2.14 and the discussion at the beginning of Section 3.8). It remains open how to obtain convergence rates in the synthesis setting given by (2.17) for redundant systems. Note that lacking injectivity the composition of a forward operator in a function space with a synthesis operator of a redundant frame cannot meet (3.19b) or (3.42). Therefore, the mapping properties of the forward operator will have to be described in a different manner. In this context we refer to [37] for convergence rate theory with ℓ^1-penalty and non-injective linear forward operators.

- Shearlets and curvelets are a natural extension of wavelets used e.g. in tomographic imaging [10] or in phase retrieval problems [77]. In [74] so called *shearlet smoothness spaces* are investigated. Although the latter scale of spaces has a close connection to Besov spaces (see [74, Prop. 4.3.]) those systems do not satisfy Assumption 2.11 with the scale of Besov spaces. Anyways, the sequence spaces associated to shearlet smoothness spaces seem to have a similar structure as the spaces of wavelet coefficients $b_{p,q}^s$ used in this work. Therefore, an extension to penalties given by the latter systems seems possible.

- For deterministic noise all our convergence rate results are proven under both an a priori parameter choice rule depending on the smoothness of the unknown and the discrepancy principle as a standard a posteriori rule. In Section 3.7 we successfully used the purely deterministic variational source conditions to treat statistical noise models. Bounds on the moments of the reconstruction error are achieved under an a priori parameter choice rule of α. It is desirable to investigate convergence rates

under an a posteriori choice rules in the statistical setting.

- Most of the results in this work are devoted to finitely smoothing operators as introduced and discussed in Section 2.4. The theory for sparsity promoting wavelet regularization is complemented by our convergence rate results for the backward heat equation (see Theorems 3.77 and 3.90). Proposition 3.75 hopefully paves the way to treat also nonlinear exponentially ill-posed problems like inverse scattering problems or electrical impedance tomography. Moreover, another technical challenge is to extend the results for the backward heat equation to wavelet regularization with $0, 1, 1$-penalty.

In Chapter 4 we have shown that the smoothness condition in terms of $B_{2,\infty}^s$ for $r, 2, 1$-penalties and in terms of K_s for $r, 1, 1$-penalties are not only sufficient but also necessary for Hölder type approximation rates (see Theorems 4.3 and 4.8). Also the other converse results, i.e. Theorems 4.7, 4.33, 4.34, 4.36, Remark 5.8 and Proposition 6.16, provide necessary conditions for bounds on the reconstruction error with exact data. However, converse results for noisy data, that is the question whether a regularization scheme achieves given convergence rates in terms of the noise level on even larger sets, remains open. Although it seems likely that the answer will be negative, a rigorous proof would probably require uniform lower bounds on the maximal effect of data noise.
So far the theory presented in Section 4.3 is restricted to Hölder-type variational source conditions. To also cover exponentially ill-posed problems it is of interest to investigate logarithmic convergence rates and source conditions. At first sight condition (4.27) seems to fail for index functions not of Hölder-type. Thus it remains open whether an equivalence between image space approximation rates and variational source conditions remains valid for more general upper bounds.
An inspection of the proofs in Sections 4.3 and 6.1 shows that the set valued mapping $A \circ S_\alpha$ that assigns minimizers $\hat{f}_\alpha$ of the Tikhonov functional to an observation g^{obs} plays the key role in our contributions for general convex regularization. The discovery that the latter map is a proximal operator (see Proposition 4.20) allowed us to provide new arguments and structural insights with the differentiability result in Proposition 4.24 and the firm non-expansiveness in Lemma 6.6 as two forceful examples.

The simplicity and clearness of the the convergence rate theory for oversmoothing with Banach space penalty in Chapter 5 indicates that this case can be treated more flexibly and is less challenging than classical variational regularization theory with a true solution contained in the penalty's domain. Here we saw a close connection between smoothness conditions and the real interpolation scale spanned by the space associated to the forward operator and the one providing the penalty term. Therefore, as we saw for BV-regularization, it is enough to know the interpolation theory of the latter spaces to conclude a convergence rate result. Concerning converse results for noisy data in Banach space regularization the author believes that the oversmoothing setting may be a good starting point. As already mentioned in Remark 5.21, a treatment of statistical noise models would be desirable.

So far, the presented theory based on image space approximation rates in Chapter 6 is limited to Hilbert space data fidelity terms. It would be interesting to generalize the arguments to Banach spaces $\mathbb{Y}$. A generalization to nonlinear operators seems even more

challenging.

As for approaches using variational source conditions, the fastest convergence rate we are able to prove for a p-homogeneous penalty term is $\mathcal{O}(\frac{1}{p})$ (see Remarks 6.26, 6.30 and 6.40). It seems to be an interesting aim to extend the presented approach to higher order convergence rates. Also the question of saturation in Banach space regularization is hardly understood. The author believes that the the exponent u in a norm penalty of the form $\mathcal{R}(x) = \frac{1}{u}\|x\|_{\mathbb{X}}^{u}$ has an influence on the saturation effect. Another direction is the application to further concrete settings as in the three presented examples in Section 6.2. An idea is to formulate a weaker version of the condition (6.10) by requiring a nesting $\mathbb{X}_{1a} \subseteq K_1 \subseteq \mathbb{X}_{1b}$ with quasi-Banach spaces $\mathbb{X}_{1a}, \mathbb{X}_{1b}$ and try to prove a generalized version of Theorem 6.20. The author believes that this approach would cover e.g. Besov norm penalties with mixed indices p, q with $p \neq 2$.

Finally, we address the choice of the loss function to measure the reconstruction error. In all our convergence rate results we are limited to loss functions associated to spaces that continuously contain a real interpolation space in the scale generated by the forward operator with respect to $\mathbb{Y}$ and the space providing the penalty term (see the condition on $\mathbb{X}_L$ in Assumption 5.1). This restriction impedes a clear comparison of different regularization schemes. Therefore, a theory that is more flexible in the choice of the loss function is desirable (see also Remark 5.16). In this context Remark 6.22 shows that Hölder type convergence rates under an a priori parameter choice rule on source sets containing all elements satisfying some Hölder type image space approximation rate cannot hold true for loss function whose corresponding spaces do not satisfy an embedding relation as above.

In summary, although many important theoretical questions concerning sparsity promoting wavelet regularization could be settled in this thesis, a number of interesting problems remains unsolved and may be subject of further research.

Appendix

A.1 Index functions

Lemma A.1. *Let $\varphi\colon [0,\infty) \to [0,\infty)$ be concave and upper semi-continuous. Then φ is non-decreasing and continuous.*

Proof. To prove non-decreasingness let $\tau_1 < \tau_2$. Set $m := \frac{\varphi(\tau_2)-\varphi(\tau_1)}{\tau_2-\tau_1}$. Using concavity it is routine to check that $\varphi(\tau) \leq m(\tau - \tau_1) + \varphi(\tau_1)$ for all $\tau > \tau_2$. Together with non-negativity of φ this forces $m \geq 0$. Therefore, $\varphi(\tau_1) \leq \varphi(\tau_2)$.

Concavity of φ yields continuity on the open set $(0,\infty)$ (see [6, Cor. 8.30]) and it remains to prove continuity at 0. By non-decreasingness the limit $\lim_{\tau\searrow 0}\varphi(\tau)$ exists. Let $0 < \tau < 1$. Then $\varphi(\tau) \geq (1-\tau)\varphi(0) + \tau\varphi(1)$ yields $\lim_{\tau\searrow 0}\varphi(\tau) \geq \varphi(0)$. Moreover, upper semi-continuity yields $\varphi(0) \geq \liminf_{\tau\searrow 0}\varphi(\tau) = \lim_{\tau\searrow 0}\varphi(\tau)$. Therefore, φ is continuous. $\qquad\square$

A.2 Banach spaces

In this section we collect some functional analytic tools. The first one characterizes the range of the adjoint operator of a linear bounded operator.

Proposition A.2 ([93, Lem. 8.21.])**.** *Let $A\colon \mathbb{X} \to \mathbb{Y}$ be a bounded linear operator between Banach spaces and $\xi \in \mathbb{X}'$. The following statements are equivalent:*

  (i) There exists a constant $c \geq 0$ such that $\langle \xi, x \rangle \leq c\|Ax\|_{\mathbb{Y}}$ for all $x \in \mathbb{X}$.

  (ii) There exists $\omega \in \mathbb{Y}'$ with $\|\omega\|_{\mathbb{Y}'} \leq c$ and $A^*\omega = \xi$.

Next we characterize weak*-to-weak continuity.

Proposition A.3 (weak*-to-weak continuity)**.** *Let $\mathbb{Z}$ and $\mathbb{Y}$ be real Banach spaces and $A\colon \mathbb{Z}' \to \mathbb{Y}$ linear and bounded. Then A is weak*-to-weak continuous if and only if $\operatorname{im}(A^*) \subset \mathbb{Z}$.*
Here we view $\mathbb{Z} \subset \mathbb{Z}''$ canonically embedded.

Proof. First recall that a linear functional $\psi\colon \mathbb{Z}' \to \mathbb{R}$ is weak*-continuous if and only if $\psi \in \mathbb{Z}$.

Suppose A is weak*-to-weak continuous and let $\phi \in \mathbb{Y}'$. Then $A^*\phi = \phi \circ A$ is weak* continuous. Hence $A^*\phi \in \mathbb{Z}$. This proves $\operatorname{im}(A^*) \subset \mathbb{Z}$.

On the other hand, suppose $\operatorname{im}(A^*) \subset \mathbb{Z}$. Then $\phi \circ A = A^*\phi \in \mathbb{Z}$ for all $\phi \in \mathbb{Y}'$. Hence $\phi \circ A$ is weak*-continuous. Therefore, A is weak*-to-weak continuous as the weak topology on $\mathbb{Y}$ is the initial topology with respect to the elements in $\mathbb{Y}'$. $\qquad\square$

We turn to real interpolation theory of Banach spaces. Here we assume that $\mathbb{X}_-$ is a Banach space and $\mathbb{X}$ is a quasi-Banach spaces with a continuous embedding $\mathbb{X} \subset \mathbb{X}_-$. Then we can rewrite the K-functional

$$K(t, f) = \inf_{h \in \mathbb{X}} \left(\|f - h\|_{\mathbb{X}_-} + t\|h\|_{\mathbb{X}} \right) \quad \text{for } t > 0 \text{ and } f \in \mathbb{X}_-. \tag{A.1}$$

Note that this agrees with the definition given in Section 2.4 using $\mathbb{X}_-$ as ambient space: $\tilde{D} = \mathbb{X}_-$ in the notation of Section 2.4). Recall

$$\|f\|_{(\mathbb{X}_-, \mathbb{X})_{\theta, q}} = \left(\int_0^\infty \left(t^{-\theta} K(t, f) \right)^q \frac{dt}{t} \right)^{\frac{1}{q}}$$

for $0 < \theta < 1$ and $q \in [1, \infty)$ and

$$\|f\|_{(\mathbb{X}_-, \mathbb{X})_{\theta, \infty}} = \sup_{t > 0} t^{-\theta} K(t, f)$$

for $0 \le \theta \le 1$. Here we obtain quasi-Banach spaces $(\mathbb{X}_-, \mathbb{X})_{\theta, q}$ consisting of all $f \in \mathbb{X}_-$ with $\|f\|_{(\mathbb{X}_-, \mathbb{X})_{\theta, q}} < \infty$ (see e.g. [8, Sec. 3.11.]).

Lemma A.4 (Interpolation inequality (see [8, Sec. 3.5, Thm. 3.11.4]))**.** *Suppose $\mathbb{X}_L$ is a Banach space with $\mathbb{X}_L \subset \mathbb{X}_-$ and $\xi \in (0, 1)$. Then the following statements are equivalent*

(i) $\mathbb{X}_L$ continuously embeds into $(\mathbb{X}_-, \mathbb{X})_{\xi, 1}$.
(ii) There exists a constant $c_L > 0$ such that

$$\|f\|_{\mathbb{X}_L} \le c_L \|f\|_{\mathbb{X}_-}^{1-\xi} \cdot \|f\|_{\mathbb{X}}^{\xi} \quad \text{for all } f \in \mathbb{X}.$$

Proposition A.5. *We have $\mathbb{X} \subset (\mathbb{X}_-, \mathbb{X})_{1, \infty}$ with a continuous embedding. The embedding constant is 1.*

Proof. Let $f \in \mathbb{X}$. Then we insert $h = f$ in the K-functional (A.1) to obtain

$$K(t, f) \le t\|f\|_{\mathbb{X}} \quad \text{for all } t > 0.$$

Hence $f \in (\mathbb{X}_-, \mathbb{X})_{1, \infty}$ with $\|f\|_{(\mathbb{X}_-, \mathbb{X})_{1, \infty}} \le \|f\|_{\mathbb{X}}$. $\qquad\qquad\square$

Proposition A.6 (Reiteration)**.** *Let $0 < \xi < \theta \le 1$. Then*

$$(\mathbb{X}_-, \mathbb{X})_{\xi, 1} = \left(\mathbb{X}_-, (\mathbb{X}_-, \mathbb{X})_{\theta, \infty} \right)_{\frac{\xi}{\theta}, 1}$$

with equivalent quasi-norms.

Proof. In the notation of [8, Def. 3.5.1] we have that $\mathbb{X}_-$ is of class $\mathcal{C}\left(0, (\mathbb{X}_-, \mathbb{X})\right)$. Moreover, $(\mathbb{X}_-, \mathbb{X})_{\theta, \infty}$ is of class $\mathcal{C}\left(\theta, (\mathbb{X}_-, \mathbb{X})\right)$. If $\theta < 1$ this follows directly from [8, Thm. 3.11.4]). For $\theta = 1$ the definition yields that $(\mathbb{X}_-, \mathbb{X})_{1, \infty}$ is of class $\mathcal{C}_K\left(1, (\mathbb{X}_-, \mathbb{X})\right)$ (see [8, Def. 3.5.1]). Moreover, from Proposition A.5 we see

$$\|f\|_{(\mathbb{X}_-, \mathbb{X})_{1, \infty}} \le \|f\|_{\mathbb{X}} \le t^{-1} \max\{\|f\|_{\mathbb{X}_-}, t\|f\|_{\mathbb{X}}\}.$$

Hence $(\mathbb{X}_-, \mathbb{X})_{1, \infty}$ is of class $\mathcal{C}_J\left(1, (\mathbb{X}_-, \mathbb{X})\right)$ (see again [8, Def. 3.5.1]). Therefore, the result follows from the reiteration theorem [8, Thm. 3.11.5]. $\qquad\qquad\square$

A.3 Besov spaces

In this section we collect some properties of Besov function and related sequence spaces needed in this work. The first one is a compactness statement needed to prove existence of minimizers of the Tikhonov functional in wavelet regularization.

Proposition A.7. *Let $p, q \in [1, \infty]$ and $r, \tilde{r} \in \mathbb{R}$ with $\tilde{r} < r$ and suppose (2.16). The unit ball*

$$B := \{x \in b_{p,q}^r \colon \|x\|_{r,p,q} \leq 1\}$$

is norm compact in $b_{p,q}^{\tilde{r}}$.

Proof. For $n \in \mathbb{N}_0$ we consider $P_n \colon b_{p,q}^r \to b_{p,q}^{\tilde{r}}$ given by $(P_n x)_{(j,k)} = x_{(j,k)}$ if $j \leq n$ and $(P_n x)_{(j,k)} = 0$ for $j > n$. Writing out the definition of the norm we obtain

$$\|x - P_n x\|_{\tilde{r},p,q} \leq 2^{-(n+1)(a+r)} \|x\|_{r,p,q} \quad \text{for all } x \in b_{p,q}^r.$$

This shows that the finite range operator P_n converges strongly to the embedding operator $b_{p,q}^r \subset b_{p,q}^{\tilde{r}}$. Hence the embedding is compact. Therefore, B is relatively compact. It remains to show that B is closed. To this end, let $(x^{(k)})_{k \in \mathbb{N}} \subset B$ be a sequence that converges to some $x \in b_{p,q}^{\tilde{r}}$ in the $\|\cdot\|_{\tilde{r},p,q}$-norm. In particular we obtain $x_j = \lim_k x_j^{(k)}$ for all $j \in \mathbb{N}_0$. As P_n has finite range this implies

$$\|P_n x\|_{r,p,q} = \lim_{k \to \infty} \|P_n x^{(k)}\|_{r,p,q} \leq \limsup_{k \to \infty} \|x^{(k)}\|_{r,p,q} \leq 1$$

for all $n \in \mathbb{N}$. Therefore, $x \in B$. $\qquad\square$

Next we characterize dual spaces of Besov sequence spaces $b_{p,q}^s$ and weighed ℓ^p-spaces. The statement is proven by adapting the well known proofs of the isomorphisms $(\ell^p)' = \ell^{p'}$ with p' the Hölder conjugate of $p \in [1, \infty)$.

Proposition A.8. *(a) Let $p \in [1, \infty)$ with Hölder conjugate p' and $\omega = (\omega_j)_{j \in \Lambda}$ a sequence of positive reals. Let $p' \in (1, \infty]$ with $\frac{1}{p} + \frac{1}{p'} = 1$. Then the pairing*

$$\langle \cdot, \cdot \rangle \colon \ell_{\omega^{-1}}^{p'} \times \ell_\omega^p \to \mathbb{R} \quad \text{given by } \langle \xi, x \rangle = \sum\nolimits_{j \in \Lambda} \xi_j x_j$$

is well defined and gives rise to an isometric isomorphism $(\ell_\omega^p)' \cong \ell_{\omega^{-1}}^{p'}$.
(b) Let $p, q \in [1, \infty)$ with Hölder conjugates p', q' and $s \in \mathbb{R}$. Then the pairing

$$\langle \cdot, \cdot \rangle \colon b_{p',q'}^{-s} \times b_{p,q}^s \to \mathbb{R} \quad \text{given by } \langle \xi, x \rangle = \sum\nolimits_{(j,k) \in \Lambda} \xi_{j,k} x_{j,k}$$

is well defined and gives rise to an isometric isomorphism $(b_{p,q}^s)' \cong b_{p',q'}^{-s}$. (see also [101, 2.11.2 (1)])

The next proposition provides a predual for the sequence spaces $b_{p,q}^s$.

Proposition A.9 (Predual space)**.** *Suppose $s \in \mathbb{R}$ and $p, q \in [1, \infty)$ with Hölder conjugates p' and q'. For $\lambda \in \Lambda$ let $e_\lambda \in \mathbb{R}^\Lambda$ be the unit vector given by $(e_\lambda)_{\lambda'} = \delta_{\lambda\lambda'}$. Then the norm closure E of the linear span of $F := \{e_\lambda \colon \lambda \in \Lambda\}$ in $b_{p',q'}^{-s}$ is a predual for $b_{p,q}^s$, i.e. $E' = b_{p,q}^s$. More precisely the map $\Psi \colon b_{p,q}^s \to E'$ given by $(\Psi(x))(z) = \langle z, x \rangle$ is an isometric isomorphism.*

Proof. Proposition A.8 allows us to view F as a subset of $\left(b_{p,q}^s\right)'$. We show that the closed unit ball $B := \{x \in b_{p,q}^r \colon \|x\|_{r,p,q} \leq 1\}$ is compact in the topology τ_F induced by F. Then the claim follows from [70, Thm. 1] as F obviously separates points of $b_{p,q}^r$.

We consider $\mathbb{R}^\Lambda$ with the product topology. Then it is easy to see that τ_D coincides with the induced topology by $b_{p,q}^s \subset \mathbb{R}^\Lambda$. Let $\omega_j := 2^{jr} 2^{jd\left(\frac{1}{2}-\frac{1}{p}\right)}$. Then $x \in B$ implies

$$|x_{j,k}| \leq \|x_j\|_p \leq \omega_j^{-1} \|x\|_{r,p,q} \leq \omega_j^{-1}.$$

Hence, $B \subset \prod_{(j,k)\in\Lambda}[-\omega_j^{-1}, \omega_j^{-1}]$. As the right hand side of the latter inclusion is compact by Tychonoff's theorem it suffices to show that B is closed in $\mathbb{R}^\Lambda$. To this end, suppose $x \in \mathbb{R}^\Lambda \setminus B$. Then there exists a finite subset $J \subset \Lambda$ and $\epsilon > 0$, such that whenever $z \in \mathbb{R}^\Lambda$ satisfies $z_\lambda \in (x_\lambda - \epsilon, x_\lambda + \epsilon)$ for all $\lambda \in J$ then $\|z\|_{r,p,q} > 1$. Hence, B is closed in $\mathbb{R}^\Lambda$ since its complement is open. $\qquad\square$

Proposition A.10 (weak*-to-weak continuous embedding)**.** *Let $s \in \mathbb{R}$ and $p, q, q_2 \in [1, \infty)$ with $q < q_2$. We endow $b_{p,q}^s$ with the weak*-topology given by the predual in Proposition A.9. Then the embedding $b_{p,q}^s \subset b_{p,q_2}^s$ is weak*-to-weak continuous.*

Proof. With p', q' and q_2' the Hölder conjugates of p, q and q_2 and in view of Proposition A.8 we may view the adjoint operator of the embedding $b_{p,q}^s \subset b_{p,q_2}^s$ as the embedding $b_{p',q_2'}^{-s} \subset b_{p',q'}^{-s}$. By Proposition A.9 and Proposition A.3 it suffices to show that $b_{p',q_2'}^{-s}$ is contained in the closure of the linear span of $F = \{e_\lambda \colon \lambda \in \Lambda\}$ in $b_{p',q'}^{-s}$.

Note that $1 \leq q < q_2$ implies $q_2 < \infty$. Let $x \in b_{p',q_2'}^{-s}$. For $n \in \mathbb{N}$ let $z_n \in \mathbb{R}^\Lambda$ be given by $(z_n)_j = x_j$ if $j \leq n$ and $(z_n)_j = 0$ if $j > n$. Then z_n belongs to the linear span of F. Furthermore, $z_n \in b_{p',q'}^{-s}$ and with $\omega_j = 2^{-js} 2^{jd\left(\frac{1}{2}-\frac{1}{p'}\right)}$ we obtain

$$\|x - z_n\|_{-s,p',q_2'}^{q_2} = \sum_{j=n+1}^\infty w_j^{q_2'} \|x_j\|_{p'}^{q_2'} \longrightarrow 0 \quad \text{for} \quad n \to \infty.$$

This shows that $(z_n)_{n\in\mathbb{N}}$ converges to x in $b_{p',q_2'}^{-s}$. This implies convergence in $b_{p',q'}^{-s}$ and therefore x belongs to the closure of the linear span of D in $b_{p',q'}^{-s}$. $\qquad\square$

Now we turn to Besov function spaces and their relation to other classical function spaces.

Recall that for $s \in \mathbb{R}$ fractional Sobolev spaces on $\mathbb{R}^d$ are given by

$$H^s(\mathbb{R}^d) := \left\{ f \in \mathcal{S}'(\mathbb{R}^d) \colon \|f\|_{H^s(\mathbb{R}^d)} := \left\| (1 + \|\xi\|_2^2)^{\frac{s}{2}} \mathcal{F}(f) \right\|_{L^2(\mathbb{R}^d)} < \infty \right\}$$

with the Fourier transform $\mathcal{F} \colon \mathcal{S}'(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$. Likewise for the d-dimensional torus we have

$$H^s(\mathbb{T}^d) := \left\{ f \in \mathcal{D}'(\mathbb{T}^d) \colon \|f\|_{H^s(\mathbb{T}^d)} := \left(\sum_{k\in\mathbb{Z}^d} (1 + \|k\|_2^2)^{\frac{s}{2}} |\hat{f}(k)|^2 \right)^{\frac{1}{2}} < \infty \right\}$$

with the Fourier coefficients $\hat{f}(k)$ of f. For a bounded Lipschitz domain $\Omega \subset \mathbb{R}^d$ we define $H^s(\Omega)$ by restrictions (in the sense of the theory of distributions):

$$H^s(\Omega) := \left\{ f \in \mathcal{D}'(\Omega) \colon f = g|_\Omega \text{ for some } g \in H^s(\mathbb{R}^d) \right\} \tag{A.2}$$

with the norm given by the infimum over $\|g\|_{H^s(\mathbb{R}^d)}$ for all extensions g.

Proposition A.11 (Relation to L^p and Sobolev spaces). *(a) Let $p \in (1, \infty)$. Then we have continuous embeddings*

$$B^0_{p,\min\{p,2\}}(\Omega) \subset L^p(\Omega) \subset B^0_{p,\max\{p,2\}}(\Omega).$$

For $p = 1$ the following embeddings hold true:

$$B^0_{1,1}(\Omega) \subset L^1(\Omega) \subset B^0_{1,\infty}(\Omega).$$

(b) Let $s \in \mathbb{R}$, then $B^s_{2,2}(\Omega) = H^s(\Omega)$ with equivalent norms.

Proof. First note that as all occurring spaces on bounded Lipschitz domains in $\mathbb{R}^d$ are defined by restriction of the respective spaces on $\mathbb{R}^d$, it suffices to prove the assertions for $\Omega = \mathbb{R}^d$ and for the periodic case $\Omega = \mathbb{T}^d$.

(a) Let $F^s_{p,q}(\Omega)$ be the scale of function spaces as defined in [101, 2.3.1. Def. 2(ii)] for $\Omega = \mathbb{R}^d$ and in [101, 9.1.3. Def. (ii)] for the d-dimensional torus. By [101, 3.2.4.(3)] and [95, 3.5.1.(20)] we have continuous embeddings

$$B^s_{p,\min\{p,q\}} \subset F^s_{p,q}(\Omega) \subset B^s_{p,\max\{p,q\}} \quad \text{for all} \quad p, q \in [1, \infty). \tag{A.3}$$

With this the embeddings for $p > 1$ follow from the identity $F^0_{p,2}(\Omega) = L^p(\Omega)$ with equivalent norms. The latter identity can be found in [101, 2.5.6.] for $\Omega = \mathbb{R}^d$ and in [95, 3.5.4(22)] in to the periodic case.

For the assertion in the case $p = 1$ we refer to [101, 2.5.7.(2)] and to [95, 3.5.4(21)] in the periodic case.

(b) We have $F^s_{2,2}(\Omega) = H^s(\Omega)$ (see [101, Thm. 2.5.6] and [95, Thm. 3.5.4.]). Therefore, the claimed identity follows from (A.3). $\qquad\square$

A.4 Bounded variation functions

Proposition A.12. *Let $a \geq 0$, $s \in (-a, 1)$, and $\Omega \subset \mathbb{R}^d$ a bounded Lipschitz domain. Then*

$$B^s_{t_s, t_s}(\Omega) = \left(B^{-a}_{2,2}(\Omega), \mathrm{BV}(\Omega) \right)_{\theta_s, t_s} \quad \text{with } \theta_s := \frac{s+a}{a+1} \text{ and } t_s := \frac{2a+2}{s+2a+1}$$

with equivalent norms.

Proof. First note that if $f \in \mathrm{BV}(\mathbb{R}^d)$ and $\Omega \subset \mathbb{R}^d$ a subdomain, then $f|_\Omega \in \mathrm{BV}(\Omega)$ with $\|f|_\Omega\|_{\mathrm{BV}(\Omega)} \leq \|f\|_{\mathrm{BV}(\mathbb{R}^d)}$.

Due to [20, Thm. 1.4] to claim holds true for $\Omega = \mathbb{R}^d$. Note that here the condition

$\gamma < 1 - \frac{1}{d}$ from the latter reference on $\gamma := \frac{-(2a+2)}{d} + 1$ is satisfied. Let c_1 be a constant such that the norm in $B^s_{t_s,t_s}(\mathbb{R}^d)$ is bounded by c_1 times the norm in $\left(B^{-a}_{2,2}(\mathbb{R}^d), \mathrm{BV}(\mathbb{R}^d)\right)_{\theta_s,t_s}$ and the other way around.

We transfer this result to bounded Lipschitz domains. To this end, we separately prove both inclusions in the stated identity.

Let $f \in B^s_{t_s,t_s}(\Omega)$. Then there is $\tilde{f} \in B^s_{t_s,t_s}(\mathbb{R}^d)$ with $\tilde{f}|_\Omega = f$ and $\|\tilde{f} \mid B^s_{t_s,t_s}\| \leq 2\|f \mid B^s_{t_s,t_s}\|$. Let $t > 0$ and $\tilde{f} = \tilde{f}_1 + \tilde{f}_2$ with $\tilde{f}_1 \in B^{-a}_{2,2}(\mathbb{R}^d)$ and $\tilde{f}_2 \in \mathrm{BV}(\mathbb{R}^d)$ be a decomposition such that

$$\|\tilde{f}_1 \mid B^{-a}_{2,2}\| + t\|\tilde{f}_2\|_{\mathrm{BV}(\mathbb{R}^d)} \leq 2K(t, \tilde{f})$$

with the K-functional given in (2.14). Then $\tilde{f}_1|_\Omega \in B^{-a}_{2,2}(\Omega)$, $\tilde{f}_2|_\Omega \in \mathrm{BV}(\Omega)$, $f = f_1 + f_2$ and

$$K(t, f) \leq \|\tilde{f}_1|_\Omega \mid B^{-a}_{2,2}\| + t\|\tilde{f}_2|_\Omega\|_{\mathrm{BV}(\Omega)} \leq 2K(t, \tilde{f}).$$

Hence with the definition of the norm on real interpolation spaces (see (2.15)) we obtain

$$\|f\|_{\left(B^{-a}_{2,2}(\Omega), \mathrm{BV}(\Omega)\right)_{\theta_s,t_s}} \leq 2\|\tilde{f}\|_{\left(B^{-a}_{2,2}(\mathbb{R}^d), \mathrm{BV}(\mathbb{R}^d)\right)_{\theta_s,t_s}} \leq 2c_1\|\tilde{f} \mid B^s_{t_s,t_s}\| \leq 4c_1\|f \mid B^s_{t_s,t_s}\|.$$

We turn to the other inclusion. There exists a constant $C_{ext} > 0$ such that for every $f \in B^{-a}_{2,2}(\Omega)$ there exists $\tilde{f} \in B^{-a}_{2,2}(\mathbb{R}^d)$ with $\tilde{f}|_\Omega = f$ and $\|\tilde{f} \mid B^{-a}_{2,2}\| \leq C_e\|f \mid B^{-a}_{2,2}\|$ and likewise for every $f \in \mathrm{BV}(\Omega)$ there exists $\tilde{f} \in \mathrm{BV}(\mathbb{R}^d)$ with $\tilde{f}|_\Omega = f$ and $\|\tilde{f}\|_{\mathrm{BV}(\mathbb{R}^d)} \leq C_{ext}\|f\|_{\mathrm{BV}(\Omega)}$. This holds true by the definition of $B^{-a}_{2,2}(\Omega)$ via restrictions (see (2.12)) and due to [2, Prop. 3.21] for of bounded variation functions. Now suppose $f \in \left(B^{-a}_{2,2}(\Omega), \mathrm{BV}(\Omega)\right)_{\theta_s,t_s}$. Let $f = f_1 + f_2$ with $f_1 \in B^{-a}_{2,2}(\Omega)$ and $f_2 \in \mathrm{BV}(\Omega)$ such that

$$\|f_1 \mid B^{-a}_{2,2}\| + t\|f_2\|_{\mathrm{BV}(\Omega)} \leq 2K(t, f).$$

Let $\tilde{f}_1 \in B^{-a}_{2,2}(\mathbb{R}^d)$ and $\tilde{f}_2 \in \mathrm{BV}(\mathbb{R}^d)$ be extensions as above. Then $\tilde{f} := \tilde{f}_1 + \tilde{f}_2$ satisfies $\tilde{f}|_\Omega = f$, and

$$K(t, \tilde{f}_1 + \tilde{f}_2) \leq \|\tilde{f}_1 \mid B^{-a}_{2,2}\| + t\|\tilde{f}_2\|_{\mathrm{BV}(\mathbb{R}^d)} \leq 2C_e K(t, f).$$

We conclude that

$$\|f \mid B^s_{t_s,t_s}\| \leq \|\tilde{f} \mid B^s_{t_s,t_s}\| \leq c_1\|\tilde{f}\|_{\left(B^{-a}_{2,2}(\mathbb{R}^d), \mathrm{BV}(\mathbb{R}^d)\right)_{\theta_s,t_s}} \leq 2c_1 C_e\|f\|_{\left(B^{-a}_{2,2}(\Omega), \mathrm{BV}(\Omega)\right)_{\theta_s,t_s}}. \qquad \square$$

Bibliography

[1] R. Acar and C. R. Vogel. Analysis of bounded variation penalty methods for ill-posed problems. *Inverse problems*, 10(6):1217, 1994.

[2] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of bounded variation and free discontinuity problems.* Oxford Mathematical Monographs, The Clarendon Press Oxford University Press, New York, 2000.

[3] S. W. Anzengruber, B. Hofmann, and R. Ramlau. On the interplay of basis smoothness and specific range conditions occurring in sparsity regularization. *Inverse Problems*, 29(12):125002, 21, 2013.

[4] S. W. Anzengruber and R. Ramlau. Convergence rates for morozov's discrepancy principle using variational inequalities. *Inverse Problems*, 27(10):105007, Sept. 2011.

[5] P. Auscher. Compactly supported wavelets and boundary conditions. *Journal of Functional Analysis*, 111(1):29–43, Jan. 1993.

[6] H. H. Bauschke and P. L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces.* Springer Publishing Company, Incorporated, 1st edition, 2011.

[7] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.

[8] J. Bergh and J. Löfström. *Interpolation spaces.* Springer Berlin Heidelberg, 1976.

[9] M. Bertero and P. Boccacci. *Introduction to Inverse Problems in Imaging.* Taylor & Francis, 1998.

[10] T. A. Bubba, M. März, Z. Purisha, M. Lassas, and S. Siltanen. Shearlet-based regularization in sparse dynamic tomography. In *Wavelets and Sparsity XVII*, volume 10394, page 103940Y. International Society for Optics and Photonics, 2017.

[11] M. Burger, J. Flemming, and B. Hofmann. Convergence rates in ℓ^1-regularization if the sparsity assumption fails. *Inverse Problems*, 29(2):025013, 2013.

[12] M. Burger, T. Helin, and H. Kekkonen. Large noise in variational regularization. *Transactions of Mathematics and its Applications*, 2(1):1–45, 2018.

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

[14] E. J. Candes, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 59(8):1207–1223, 2006.

[15] A. Chambolle and C. Dossal. On the convergence of the iterates of the "fast iterative shrinkage/thresholding algorithm". *Journal of Optimization theory and Applications*, 166(3):968–982, 2015.

[16] A. Chambolle and T. Pock. An introduction to continuous optimization for imaging. *Acta Numerica*, 25:161–319, 2016.

[17] D.-H. Chen, B. Hofmann, and I. Yousept. Oversmoothing Tikhonov regularization in Banach spaces. *Inverse Problems*, 37(8):085007, jul 2021.

[18] F. Clarke. *Optimization and nonsmooth analysis*. Canadian Mathematical Society series of monographs and advanced texts. A Wiley-Interscience publication, 1983.

[19] A. Cohen. *Numerical Analysis of Wavelet Methods*. Studies in Mathematics and its Applications. Elsevier Science, 2003.

[20] A. Cohen, W. Dahmen, I. Daubechies, R. DeVore, et al. Harmonic analysis of the space bv. *Revista Matematica Iberoamericana*, 19(1):235–263, 2003.

[21] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k-term approximation. *Journal of the American mathematical society*, 22(1):211–231, 2009.

[22] A. Cohen, I. Daubechies, and J.-C. Feauveau. Biorthogonal bases of compactly supported wavelets. *Communications on pure and applied mathematics*, 45(5):485–560, 1992.

[23] A. Cohen, I. Daubechies, and P. Vial. Wavelets on the interval and fast wavelet transforms. *Applied and Computational Harmonic Analysis*, 1(1):54–81, Dec. 1993.

[24] A. Cohen, R. A. DeVore, and R. Hochmuth. Restricted nonlinear approximation. *Constructive Approximation*, 16(1):85–113, 2000.

[25] A. Cohen, M. Hoffmann, and M. Reiß. Adaptive wavelet Galerkin methods for linear inverse problems. *SIAM Journal of Numerical Analysis*, 42(4):1479–1501 (electronic), 2004.

[26] W. Dahmen, A. Kunoth, and K. Urban. Biorthogonal spline wavelets on the interval—stability and moment conditions. *Applied and Computational Harmonic Analysis*, 6(2):132–196, 1999.

[27] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 41(7):909–996, 1988.

[28] I. Daubechies. *Ten Lectures on Wavelets*. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics, 1992.

[29] I. Daubechies, M. Defrise, and C. D. Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, 57(11):1413–1457, 2004.

[30] M. del Álamo. *Multiscale Total Variation Estimators for Regression and Inverse Problems*. PhD thesis, Georg-August University, 2019.

[31] D. L. Donoho. Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Applied and Computational Harmonic Analysis*, 2(2):101–126, 1995.

[32] D. L. Donoho and I. M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the american statistical association*, 90(432):1200–1224, 1995.

[33] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard. Wavelet shrinkage: Asymptopia? *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(2):301–369, 1995.

[34] I. Ekeland and R. Temam. *Convex Analysis and Variational Problems.* North-Holland Publishing Company, Amsterdam, Oxford, 1976.

[35] 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.

[36] J. Flemming. *Generalized Tikhonov regularization and modern convergence rate theory in Banach spaces.* Shaker Verlag, Aachen, 2012.

[37] J. Flemming. Convergence rates for ℓ^1-regularization without injectivity-type assumptions. *Inverse Problems*, 32(9):095001, 19, 2016.

[38] J. Flemming. A converse result for Banach space convergence rates in Tikhonov-type convex regularization of ill-posed linear equations. *Journal of Inverse and Ill-posed Problems*, 26(5):639 – 646, 2018.

[39] J. Flemming. Existence of variational source conditions for nonlinear inverse problems in banach spaces. *Journal of Inverse and Ill-posed Problems*, 26(2):277–286, 2018.

[40] J. Flemming and D. Gerth. Injectivity and weak*-to-weak continuity suffice for convergence rates in ℓ^1-regularization. *Journal of Inverse and Ill-posed Problems*, 26(1):85–94, 2018.

[41] J. Flemming and M. Hegland. Convergence rates in ℓ^1-regularization when the basis is not smooth enough. *Applicable Analysis*, 94(3):464–476, 2015.

[42] J. Flemming and B. Hofmann. A new approach to source conditions in regularization with general residual term. *Numerical Functional Analysis and Optimization*, 31(3):254–284, 2010.

[43] J. Flemming, B. Hofmann, and I. Veselić. A unified approach to convergence rates for ℓ^1-regularization and lacking sparsity. *Journal of Inverse and Ill-posed Problems*, 24(2), Jan. 2016.

[44] L. Frantzen. Notwendige Bedingungen für Konvergenzraten von verallgemeinerter Tikhonov Regularisierung, Georg-August-Universität Göttingen. May 2021. Master Thesis in Mathematics.

[45] D. Freitag. Real interpolation of weighted l^p-spaces. *Mathematische Nachrichten*, 86(1):15–18, 1978.

[46] D. Gerth and B. Hofmanm. Oversmoothing regularization with ℓ^1-penalty term. *AIMS Mathematics*, 4:1223–1247, 2019.

[47] E. Giné and R. Nickl. *Mathematical foundations of infinite-dimensional statistical models*, volume 40. Cambridge University Press, 2015.

[48] R. Gorenflo and M. Yamamoto. Operator theoretic treatment of linear abel integral equations of first kind. *Japan Journal of Industrial and Applied Mathematics*, 16(1):137–161, Feb. 1999.

[49] M. Grasmair. Generalized Bregman distances and convergence rates for non-convex regularization methods. *Inverse Problems*, 26:115014 (16pp), 2010.

[50] M. Grasmair. Linear convergence rates for Tikhonov regularization with positively homogeneous functionals. *Inverse Problems*, 27(7):075014, jun 2011.

[51] M. Grasmair. Source conditions for non-quadratic Tikhonov regularization. *Numerical Functional Analysis and Optimization*, 41(11):1352–1372, June 2020.

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

[53] M. Grasmair, M. Haltmeier, and O. Scherzer. Necessary and sufficient conditions for linear convergence of ℓ^1-regularization. *Communications on Pure and Applied Mathematics*, 64(2):161–182, 2011.

[54] M. Hanke, A. Neubauer, and O. Scherzer. A convergence analysis of the Landweber iteration for nonlinear ill-posed problems. *Numerische Mathematik*, 72:21–37, 1995.

[55] T. Hein. Convergence rates for regularization of ill-posed problems in Banach spaces by approximate source conditions. *Inverse Problems*, 24(4):045007, 2008.

[56] T. Hein and B. Hofmann. Approximate source conditions for nonlinear ill-posed problems—chances and limitations. *Inverse Problems*, 25(3):035003, 2009.

[57] A. Hertle. Continuity of the Radon transform and its inverse on Euclidean spaces. *Mathematische Zeitschrift*, 184:165–192, 1983.

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

[59] B. Hofmann, S. Kindermann, and P. Mathé. Penalty-based smoothness conditions in convex variational regularization. *Journal of Inverse and Ill-posed Problems*, 27(2):283–300, Apr. 2019.

[60] B. Hofmann and P. Mathé. Tikhonov regularization with oversmoothing penalty for non-linear ill-posed problems in Hilbert scales. *Inverse Problems*, 34(1):015007, dec 2018.

[61] B. Hofmann, P. Mathé, and M. Schieck. Modulus of continuity for conditionally stable ill-posed problems in Hilbert space. *Journal of Inverse and Ill-Posed Problems*, 16(6):567–585, 2008.

[62] B. Hofmann and R. Plato. Convergence results and low-order rates for nonlinear Tikhonov regularization with oversmoothing penalty term. *Electronic Transactions on Numerical Analysis*, 53:313–328, 2020.

[63] B. Hofmann and O. Scherzer. Factors influencing the ill-posedness of nonlinear problems. *Inverse Problems*, 10(6):1277, 1994.

[64] T. Hohage and P. Miller. Optimal convergence rates for sparsity promoting wavelet-regularization in Besov spaces. *Inverse Problems*, 35:65005 (27pp), 2019.

[65] T. Hohage and M. Pricop. Nonlinear Tikhonov regularization in Hilbert scales for inverse boundary value problems with random noise. *Inverse Problems and Imaging*, 2:271–290, 2008.

[66] T. Hohage and F. Weidling. Characterizations of variational source conditions, converse results, and maxisets of spectral regularization methods. *SIAM Journal of Numerical Analysis*, 55(2):598–620, 2017.

[67] T. Hohage and F. Werner. Convergence Rates for Inverse Problems with Impulsive Noise. *SIAM Journal on Numerical Analysis*, 52(3):1203–1221, Jan 2014.

[68] B. Jin and P. Maass. Sparsity regularization for parameter identification problems. *Inverse Problems*, 28(12):123001, 2012.

[69] B. Jin, P. Maaß, and O. Scherzer. Sparsity regularization in inverse problems. *Inverse Problems*, 33(6), 2017.

[70] S. Kaijser. A note on dual Banach spaces. *Mathematica Scandinavica*, 41:325–330, 1977.

[71] B. Kaltenbacher, A. Neubauer, and O. Scherzer. *Iterative regularization methods for nonlinear ill-posed problems*, volume 6. Walter de Gruyter, 2008.

[72] S. Kindermann. Convex Tikhonov regularization in Banach spaces: New results on convergence rates. *Journal of Inverse and Ill-posed Problems*, 24(3), Jan. 2015.

[73] E. Klann, P. Maass, and R. Ramlau. Two-step regularization methods for linear inverse problems. *Journal of Inverse and Ill-posed Problems*, 14:583–609, 2006.

[74] D. Labate, L. Mantovani, and P. Negi. Shearlet smoothness spaces. *Journal of Fourier Analysis and Applications*, 19(3):577–611, 2013.

[75] G. R. Lee, R. Gommers, F. Waselewski, K. Wohlfahrt, and A. O'Leary. Pywavelets: A python package for wavelet analysis. *Journal of Open Source Software*, 4(36):1237, 2019.

[76] N.-Y. Lee and B. J. Lucier. Wavelet methods for inverting the Radon transform with noisy data. *IEEE Transactions on Image Processing*, 10(1):79–94, 2001.

[77] S. Loock and G. Plonka. Phase retrieval for fresnel measurements using a shearlet sparsity constraint. *Inverse Problems*, 30(5):055005, 2014.

[78] D. A. Lorenz. Convergence rates and source conditions for Tikhonov regularization with sparsity constraints. *Journal of Inverse and Ill-posed Problems.*, 16(5):463–478, 2008.

[79] D. A. Lorenz, S. Schiffler, and D. Trede. Beyond convergence rates: exact recovery with the Tikhonov regularization with sparsity constraints. *Inverse Problems*, 27(8):085009, 2011.

[80] D. A. Lorenz and D. Trede. Optimal convergence rates for Tikhonov regularization in Besov scales. *Inverse Problems*, 24(5):055010, Aug 2008.

[81] P. Math and S. V. Pereverzev. Geometry of linear ill-posed problems in variable Hilbert scales. *Inverse Problems*, 19(3):789–803, May 2003.

[82] Y. Meyer. *Wavelets and operators*, volume 37 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1992. Translated from the 1990 French original by D. H. Salinger.

[83] C. Michelli and T. Rivlin. *Optimal Estimation in Approximation Theory*. The IBM Research Symposia Series. Springer US, 1977.

[84] P. Miller. Variational regularization theory based on image space approximation rates. *Inverse Problems*, 37(6):065003, May 2021.

[85] P. Miller and T. Hohage. Maximal spaces for approximation rates in ℓ^1-regularization. *Numerische Mathematik*, pages 1–34, 2021.

[86] F. Natterer. Error bounds for Tikhonov regularization in Hilbert scales. *Applicable Analysis*, 18:29–37, 1984.

[87] F. Natterer. *The Mathematics of Computerized Tomography*. Teubner, Stuttgart, 1986.

[88] A. Neubauer. Tikhonov regularization of nonlinear ill-posed problems in Hilbert scales. *Applicable Analysis*, 46(1-2):59–72, 1992.

[89] A. Neubauer. On Landweber iteration for nonlinear ill-posed problems in Hilbert scales. *Numerische Mathematik*, 85(2):309–328, 2000.

[90] R. Ramlau. Regularization properties of Tikhonov regularization with sparsity constraints. *Electronic Transactions on Numerical Analysis*, 30:54–74, 2008.

[91] R. Ramlau and E. Resmerita. Convergence rates for regularization with sparsity constraints. *Electronic Transactions on Numerical Analysis*, 37:87–104, 2010.

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

[93] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational methods in imaging*, volume 167 of *Applied Mathematical Sciences.* Springer, New York, 2009.

[94] W. Schirotzek. *Nonsmooth analysis.* Springer Science & Business Media, 2007.

[95] H.-J. Schmeisser and H. Triebel. *Topics in Fourier Analysis and Function Spaces.* Wiley, New York, 1987.

[96] T. Schuster, B. Kaltenbacher, B. Hofmann, and K. S. Kazimierski. *Regularization methods in Banach spaces.* De Gruyter, July 2012.

[97] T. Strömberg. On regularization in Banach spaces. *Ark. Mat., Volume 34, Number 2, 383-406.*, 1996.

[98] H. Triebel. *Interpolation theory, function spaces, differential operators.* VEB Deutscher Verlag der Wissenschaften, Berlin, 1978.

[99] H. Triebel. *Theory of function spaces II.* Modern Birkhäuser Classics. Springer, Basel, 1992.

[100] H. Triebel. *Function spaces and wavelets on domains*, volume 7 of *EMS Tracts in Mathematics.* European Mathematical Society (EMS), Zürich, 2008.

[101] H. Triebel. *Theory of function spaces.* Modern Birkhäuser Classics. Springer, Basel, reprint. edition, 2010. Reprint of the 1983 edition.

[102] D. Van De Ville, M. Papadakis, and Y. M. Lu. Wavelets and sparsity xviii. *Wavelets and Sparsity XVIII*, 11138, 2019.

[103] M. C. Veraar. Regularity of Gaussian white noise on the d-dimensional torus. *Banach Center Publications*, 95:385–398, 2011.

[104] F. Weidling, B. Sprung, and T. Hohage. Optimal convergence rates for Tikhonov regularization in Besov spaces. *SIAM Journal of Numerical Analysis*, 58:21–47, 2020.

[105] F. Werner and T. Hohage. Convergence rates in expectation for Tikhonov-type regularization of inverse problems with Poisson data. *Inverse Problems*, 28(10):104004, Oct 2012.