PENALTY-BASED SMOOTHNESS CONDITIONS IN CONVEX VARIATIONAL REGULARIZATION

BERND HOFMANN, STEFAN KINDERMANN, AND PETER MATHÉ

ABSTRACT. The authors study Tikhonov regularization of linear ill-posed problems with a general convex penalty defined on a Banach space. It is well known that the error analysis requires smoothness assumptions. Here such assumptions are given in form of inequalities involving only the family of noise-free minimizers along the regularization parameter and the (unknown) penalty-minimizing solution. These inequalities control, respectively, the defect of the penalty, or likewise, the defect of the whole Tikhonov functional. The main results provide error bounds for a Bregman distance, which split into two summands: the first smoothness-dependent term does not depend on the noise level, whereas the second term includes the noise level. This resembles the situation of standard quadratic Tikhonov regularization in Hilbert spaces. It is shown that variational inequalities, as these were studied recently, imply the validity of the assumptions made here. Several examples highlight the results in specific applications.

1. INTRODUCTION

As a mathematical model for a linear inverse problem, we consider the ill-posed operator equation

$$(1) \qquad A\,x \,=\, y\,,$$

where A is a bounded linear operator from an infinite-dimensional Banach space X to an infinite-dimensional Hilbert space H such that $\mathcal{R}(A)$, the range of A, is a non-closed subset of H. Let $x^\dagger \in X$ denote an exact solution of (1) with properties to be particularized later. Unlesss specified otherwise, the norm $\|\cdot\|$ in this study always refers to that in H. We assume that instead of the exact right-hand side $y \in \mathcal{R}(A)$ only noisy data $y^\delta \in H$ satisfying

$$(2) \qquad \|y^\delta - y\| \leq \delta$$

with noise level $\delta \geq 0$ are available. Based on y^δ we try to recover $x^\dagger$ in a stable approximate manner by using variational regularization with general *convex* penalty functionals J.

Precisely, we are going to analyze convergence conditions for minimizers of the Tikhonov functional

$$T_\alpha(x; v) := \frac{1}{2} \|Ax - v\|^2 + \alpha J(x), \quad x \in X,$$

with regularization parameter $\alpha > 0$, for exact right-hand sides $v := y$ and noisy data $v := y^\delta$. In this context, we distinguish regularized solutions

$$(3) \qquad\qquad x_\alpha \in \underset{x \in X}{\operatorname{argmin}}\, T_\alpha(x; y)$$

and

$$(4) \qquad\qquad x_\alpha^\delta \in \underset{x \in X}{\operatorname{argmin}}\, T_\alpha(x; y^\delta),$$

respectively. In case of multiple minimizers we select any family of regularized solutions x_α and x_α^δ for $\alpha > 0$. As will be seen from the subsequent results, in particular from the discussion in Remark 1, the specific choice has no impact on the convergence rate results.

We are interested in estimates of the error between x_α^δ and $x^\dagger$ and in proving corresponding convergence rates. In a Hilbert space X, the error norm is a canonical measure in this context, in particular if the penalty J is of norm square type. For Banach spaces X and general convex penalties J, however, norms are not always appropriate measures, and the study [5] introduced alternatively the Bregman distance
$$(5)$$
$$B_\zeta(z; x) := J(x) - J(z) - \langle \zeta, x - z \rangle \geq 0, \quad x \in X, \quad \zeta \in \partial J(z) \subset X^*,$$

with some subgradient ζ from the subdifferential $\partial J(z)$ of J at the point $z \in X$, as a powerful error measure for regularized solutions of ill-posed problems in Banach spaces, see also, e.g., [17, 25, 28, 29]. We stress the fact that the subgradient ζ is taken at the first argument in the Bregman distance, and we recall that the Bregman distance is not symmetric in its arguments. Therefore, we highlight in (5) the base point z, by indicating the corresponding subgradient, say ζ.

It is a classical result that convergence rates for ill-posed problems require a regularity condition (abstract smoothness condition) for $x^\dagger$ as otherwise convergence can be arbitrarily slow. Within the present study we shall propose regularity conditions based on the penalty or overall Tikhonov functionals.

A common regularity assumption, often called *source condition*, as these are usually employed for linear problems in Hilbert space, is to

assume that $\partial J(x^\dagger) \cap \mathcal{R}(A^*) \neq \emptyset$, i.e., there are a subgradient $\xi^\dagger$ and a *source element* $w \in H$ such that

$$\xi^\dagger \in \partial J(x^\dagger) \qquad \text{with} \qquad \xi^\dagger = A^* w, \ \|w\| \leq R.$$

As we will highlight in Section 3, such source conditions yield variational inequalities of the form

(6) $\qquad J(x^\dagger) - J(x) \leq R \left\| Ax^\dagger - Ax \right\| \quad \text{for all} \quad x \in X.$

This represents a *benchmark variational inequality*. If otherwise $\xi^\dagger \notin \mathcal{R}(A^*)$, then a variational inequality

(7) $\qquad J(x^\dagger) - J(x) \leq \Phi\left(\left\| Ax^\dagger - Ax \right\|\right) \quad \text{for all} \quad x \in X,$

with a sub-linear *index function*[1] Φ holds, and the quotient function $\Phi(t)/t$ is strictly decreasing. Under a more restrictive assumption on Φ (concavity instead of sub-linearity) and for a more general setting this condition was introduced as formula (2.11) in [21].

Under such regularity conditions, as e.g. (7), it is known from ibid. that convergence rates

$$B_{\xi_\alpha^\delta}(x_\alpha^\delta; x^\dagger) = \mathcal{O}(\Phi(\delta)) \quad \text{as} \quad \delta \to 0.$$

can be established.

In the past years, convergence rates under variational source conditions (cf., e.g., [8, 15, 19]) were expressed in terms of the Bregman distance $B_{\xi^\dagger}(x^\dagger; x_\alpha^\delta)$, and hence using the base point $x^\dagger$. In this context it is not clear whether a subgradient $\xi^\dagger \in \partial J(x^\dagger)$ exists, for instance if $x^\dagger$ is not in the interior of $\text{dom}(J) := \{x \in X : J(x) < \infty\}$. Taking as base point the minimizer x_α^δ, this cannot happen and the set $\partial J(x_\alpha^\delta)$ is always non-empty (cf., e.g., [9, Lemma 2.2]). This may be seen as an advantage of the present approach, following the original study [21].

Without further notice we follow the convention from that study: if the subdifferential $\partial J(x_\alpha^\delta)$ is multi-valued, then we take for $B_{\xi_\alpha^\delta}(x_\alpha^\delta; x^\dagger)$ a subgradient ξ_α^δ that satisfies the optimality condition

(8) $\qquad\qquad\qquad A^*(Ax_\alpha^\delta - y^\delta) + \alpha \xi_\alpha^\delta = 0.$

A remarkable feature of the error bounds under smoothness assumptions (7) is the splitting of the error, see also in a more general setting [21, Thm. 3.1], as

(9) $\qquad B_{\xi_\alpha^\delta}(x_\alpha^\delta; x^\dagger) \leq \dfrac{\delta^2}{2\alpha} + \Psi(\alpha) \quad \text{for all} \quad \alpha > 0,$

---

[1]Throughout, we call a function $\varphi\colon [0, \infty) \to [0, \infty)$ index function if it is continuous, strictly increasing and obeys the boundary condition $\varphi(0) = 0$.

where the function Ψ is related to Φ, and typically it will also be an index function.

In this study, see Section 2, we analyze the regularity condition

$$(10) \qquad J(x^\dagger) - J(x_\alpha) \leq \Psi(\alpha) \quad \text{for all} \quad \alpha > 0$$

(cf. Assumption 2, and its alternative, Assumption 2′). Under these conditions a similar error splitting as in (9) is shown as the main result. Notice, that these bounds are required to hold only for the minimizers x_α of the noise-free Tikhonov functional $T_\alpha(x; Ax^\dagger)$. This resembles the situation for linear ill-posed problems in Hilbert spaces, where the error is decomposed into the noise propagation term, usually of the form $\delta/\sqrt{\alpha}$, and some noise-free term depending on the solution smoothness, say $\varphi(\alpha)$, which is called profile function in [18]. We refer to a detailed discussion in Section 2. The error bounds will be complemented by some discussion on the equivalence of Assumptions 2 and 2′. Also, a discussion on necessary conditions for an index function Ψ to serve as an inequality (10) is given. We mention that the existence of an index function Ψ satisfying (10) is an immediate consequence of [10, Thm. 3.2] (see also [9, Remark 2.6]) in combination with the results of Section 3. Precisely, we highlight that the variational inequality (7) implies the validity of (10) for some specific index functions Ψ related to Φ by some convex analysis arguments. Then, in Section 4 we present specific applications of this approach.

In an appendix we give detailed proofs of the main results (Appendix A) and some auxiliary discussion concerning convex index functions (Appendix B).

2. Assumptions and main results

In the subsequent analysis, *convex* index functions will be of particular interest, i.e., index functions φ which obey

$$\varphi\left(\frac{s+t}{2}\right) \leq \frac{1}{2}\left(\varphi(s) + \varphi(t)\right), \quad s, t \geq 0.$$

The inverse of a convex index function is a *concave* index function, and hence the above inequality is reversed. We mention that concave index functions are *sub-linear*, which means that these functions have the property that the quotients $\varphi(\lambda)/\lambda$ are non-increasing.

In the sequel, it will be useful to employ convex duality. For any convex function φ, we may define its Fenchel conjugate

$$(11) \qquad \varphi^*(t) := \sup_{s \geq 0}\left(st - \varphi(s)\right), \quad t > 0.$$

Clearly, both functions φ and its conjugate φ^* obey the *Fenchel-Young Inequality*

$$(12) \qquad st \leq \varphi(s) + \varphi^*(t), \quad s, t \geq 0.$$

Additional considerations concerning convex index functions are collected in Appendix B.

The proofs of the results in this section are technical, and hence they are postponed to Appendix A.

2.1. **Assumptions.** Throughout this study we impose, e.g., along the lines of [29], the following standard assumptions on the penalty.

Assumption 1 (Penalty). *The function $J : X \to [0, \infty]$ is a proper, convex functional defined on a Banach space X, which is lower semicontinuous with respect to weak (or weak*) sequential convergence. Additionally, we assume that J is a stabilizing (weakly coercive) penalty functional, i.e., the sublevel sets $\mathcal{M}_c := \{x \in X : J(x) \leq c\}$ of J are for all $c \geq 0$ weakly (or weak*) sequentially compact. Moreover, we assume that at least one solution $x^\dagger$ of (1) with finite penalty value $J(x^\dagger) < \infty$ exists.*

Consequently, for all $\alpha > 0$ and $v \in H$, the sublevel sets of $T_\alpha(., v)$ are weakly (or weak*) sequentially compact. This ensures the existence and stability of regularized solutions x_α and x_α^δ which are the corresponding minimizers for $v = y$ and $v = y^\delta$, respectively. In the sequel, we use the symbol $x^\dagger$ only for the always existing J-minimizing solutions of (1), i.e. $J(x^\dagger) = \min\limits_{x \in X : Ax = y} J(x)$.

The fundamental regularity condition is given as follows. To this end, let x_α be defined as in (3). This assumption controls the deviation of the penalty at the minimizers from the one at the J-minimizing solution $x^\dagger$.

Assumption 2 (Defect for penalty). *There is an index function Ψ such that*

$$(13) \qquad J(x^\dagger) - J(x_\alpha) \leq \Psi(\alpha) \quad \text{for all} \quad \alpha > 0.$$

It is not difficult to conclude from the minimizing property of x_α,

$$(14) \qquad \frac{1}{2} \left\| Ax_\alpha - Ax^\dagger \right\|^2 + \alpha J(x_\alpha) \leq \alpha J(x^\dagger),$$

that the left-hand side of (13) is nonnegative and hence that

$$(15) \quad \lim_{\alpha \to 0} J(x_\alpha) = J(x^\dagger) \quad \text{and} \quad \frac{1}{2\alpha} \left\| Ax_\alpha - y \right\|^2 \leq J(x^\dagger) - J(x_\alpha),$$

such that Assumption 2 also yields the estimate

$$(16) \qquad \frac{1}{2\alpha} \left\| Ax_\alpha - y \right\|^2 \leq \Psi(\alpha) \quad \text{for all} \quad \alpha > 0.$$

Instead of controlling the defect for the penalty J one might control the defect for the overall Tikhonov functional as follows.

Assumption 2′ (Defect for Tikhonov functional). *There is an index function Ψ such that*

$$(17) \qquad \frac{1}{\alpha} \left(T_\alpha(x^\dagger; Ax^\dagger) - T_\alpha(x_\alpha; Ax^\dagger) \right) \leq \Psi(\alpha) \quad \text{for all} \quad \alpha > 0.$$

By explicitly writing the left-hand side in (17) we see that

$$\frac{1}{\alpha} \left(T_\alpha(x^\dagger; Ax^\dagger) - T_\alpha(x_\alpha; Ax^\dagger) \right) = J(x^\dagger) - J(x_\alpha) - \frac{1}{2\alpha} \left\| Ax_\alpha - y \right\|^2,$$

and hence Assumption 2 is stronger than Assumption 2′, as stated above. One advantage of Assumption 2′ is that it is invariant with respect to the choice of the minimizers x_α. This is not clear for Assumption 2. As a remarkable fact we state that both assumptions are basically equivalent.

Proposition 1. *Assumption 2′ yields that*

$$(18) \qquad J(x^\dagger) - J(x_\alpha) \leq 2\Psi(\alpha) \quad \text{for all} \quad \alpha > 0.$$

Hence Assumption 2 is fulfilled with Ψ replaced by 2Ψ.

Remark 1. The above result has an important impact, and we return to the choice of the minimizers $x_\alpha, x_\alpha^\delta$ from (3) and (4), respectively. As mentioned before, the functional on the left-hand side of (17) is independent of the choice of the minimizers x_α, due to the uniqueness of the value of the Tikhonov functional at the minimizers (cf., e.g., [20, Sec. 3.2]). Thus, if Assumption 2′ is fulfilled for one selection $x_\alpha, \alpha > 0$, then this holds true for arbitrary selections. Since Assumption 2′ implies Assumption 2 (at the expense of a factor 2) the latter will be fulfilled for any selection. Conversely, if Assumption 2 holds for some selection $x_\alpha, \ \alpha > 0$, then this yields the validity of Assumption 2′, but then extends to any other choice of minimizers. Again, by the above proposition this implies that any other choice of minimizers will obey Assumption 2, by losing a factor 2 at most.

We finally discuss which index functions may serve as upper bounds in either of the assumptions 2 or 2′, respectively. We formulate this as follows.

Proposition 2. *Suppose Assumption 2 holds with index function* Ψ. *Then the following is true:*
Either $\ J(x^\dagger) = \min\limits_{x \in X} J(x),$
and then $J(x_\alpha) = J(x^\dagger)$ *for each* $\alpha > 0$, *and any index function* Ψ *is a valid bound in* (13),
or $\ J(x^\dagger) > \min\limits_{x \in X} J(x),$
and then Ψ *increases near zero at most linearly.*

We shall call the first case *singular*. In this case, where $J(x^\dagger) = \min_{x \in X} J(x)$, the choice of the regularization parameter loses importance, which is also the case if the phenomenon of exact penalization occurs (see [5] and more recently in [3]).

2.2. **Main results.** We turn to stating the main results, which highlight the impact of Assumption 2 and Assumption $2'$ on the overall error, measured by the Bregman distance.

Theorem 1. *Under Assumption 2 we have that*

$$B_{\xi_\alpha^\delta}(x_\alpha^\delta; x^\dagger) \le \frac{\delta^2}{2\alpha} + \Psi(\alpha) \quad \text{for all} \quad \alpha > 0.$$

The proof of Theorem 1 is a simple consequence of the following result, which may be also of its own interest.

Theorem 1$'$. *Suppose that Assumption $2'$ is satisfied with an index function* Ψ. *Then an error estimate of the type*

$$(19) \qquad B_{\xi_\alpha^\delta}(x_\alpha^\delta; x^\dagger) \le \frac{\delta^2}{2\alpha} + \Psi(\alpha) \quad \text{for all} \quad \alpha > 0$$

holds.

Since, as mentioned above, Assumption 2 is stronger than Assumption $2'$ it is enough to prove Theorem 1$'$.

2.3. **Discussion.** Resulting from Theorems 1 and Theorem 1$'$, the best possible bound for the Bregman distance between the regularized solutions and $x^\dagger$ as a function of $\delta > 0$ takes place in both cases if $\alpha = \alpha_* > 0$ is chosen such that the right-hand side $\frac{\delta^2}{2\alpha} + \Psi(\alpha)$ is minimized, i.e.,

$$(20) \qquad B_{\xi_{\alpha_*}^\delta}\left(x_{\alpha_*}^\delta; x^\dagger\right) \le \inf_{\alpha > 0}\left\{\frac{\delta^2}{2\alpha} + \Psi(\alpha)\right\},$$

which determines, from this perspective, the best possible convergence rate of $B_{\xi_{\alpha_*}^\delta}\left(x_{\alpha_*}^\delta; x^\dagger\right)$ to zero as $\delta \to 0$. Consequently, this convergence rate is the higher, the faster the decay rate of $\Psi(\alpha) \to 0$ as $\alpha \to 0$ is. As

expressed in the non-singular case of Proposition 2, the function Ψ cannot increase from zero super-linearly, and the limiting case is obtained for $\Psi(\alpha) \sim \alpha$, $\alpha \to 0$. From this perspective, the maximally described rate is $B_{\xi_{\alpha_*}^\delta}\left(x_{\alpha_*}^\delta; x^\dagger\right) \sim \delta$ as $\delta \to 0$, which is obtained whenever for example the regularization parameter is chosen as $\alpha_* = \alpha(\delta) \sim \delta$. For linear ill-posed equations in Hilbert spaces and using the standard penalty $J(x) = \|x\|_X^2$ (see Section 4.1), this results in the error rate $B_{\xi_{\alpha_*}^\delta}\left(x_{\alpha_*}^\delta; x^\dagger\right) = \left\|x_{\alpha_*}^\delta - x^\dagger\right\|_X^2 = \mathcal{O}(\delta)$. However, resulting from Theorems 1 and Theorem 1$'$ the overall best possible convergence rate $\left\|x_{\alpha_*}^\delta - x^\dagger\right\|_X^2 = \mathcal{O}(\delta^{4/3})$ attainable for Tikhonov regularization cannot be obtained, and indeed our analysis is confined to the low rate case expressed be the range-type source condition $x^\dagger \in \mathcal{R}(A^*)$. This is also the case for all other approaches which are based on the minimizing property $T_\alpha(x_\alpha^\delta; y^\delta) \leq T_\alpha(x^\dagger; y^\delta)$ only, including approaches using variational source conditions (see Section 3 below). For alternative techniques leading to enhanced convergence rates we refer to [24, 25, 26], [29, Sect. 4.2.4] and references therein.

Now we return to the error estimate (19) for general convex penalties J. Since the upper bound with respect to $\alpha > 0$ is decomposed into a sum of a continuous decreasing function $\delta^2/(2\alpha)$ and an increasing continuous function $\Psi(\alpha)$, the minimizer always exists. Given Ψ, let us assign the companion $\Theta(\alpha) := \sqrt{\alpha \Psi(\alpha)}$, $\alpha > 0$. If we then let α_* be obtained from calibrating both summands as

$$(21) \qquad \alpha_* = \alpha_*(\delta) := \left(\Theta^2\right)^{-1}\left(\frac{\delta^2}{2}\right) = \Theta^{-1}\left(\frac{\delta}{\sqrt{2}}\right),$$

then we find that

$$(22) \qquad B_{\xi_{\alpha_*}^\delta}(x_{\alpha_*}^\delta; x^\dagger) \leq 2\Psi\left(\Theta^{-1}\left(\frac{\delta}{\sqrt{2}}\right)\right),$$

and the optimality of this bound will be discussed in the examples presented below in Section 4.

It is interesting to separately discuss the singular case, i.e., when $J(x^\dagger) = \min_{x \in X} J(x)$. We claim that then $B_{\xi_\alpha^\delta}(x_\alpha^\delta; x^\dagger) = 0$ when the subdifferential $\xi_\alpha^\delta = \partial J(x_\alpha^\delta)$ obeys the optimality condition (8), i.e., we have $\alpha \xi_\alpha^\delta = A^*\left(y^\delta - Ax_\alpha^\delta\right)$. If we now look at the minimizing property of x_α^δ then we see that

$$\frac{1}{2}\left\|Ax_\alpha^\delta - y^\delta\right\|^2 + \alpha J(x_\alpha^\delta) \leq \alpha J(x^\dagger),$$

which, in the singular case, requires to have that $\left\| Ax_\alpha^\delta - y^\delta \right\| = 0$, and hence that $\xi_\alpha^\delta = 0$. This yields for the Bregman distance that

$$B_{\xi_\alpha^\delta}\left(x_\alpha^\delta; x^\dagger\right) = J(x^\dagger) - J(x_\alpha^\delta) + \langle \xi_\alpha^\delta, x_\alpha^\delta - x^\dagger \rangle \leq 0,$$

such that the Bregman distance equals zero in the singular case.

We already emphasized that the upper estimate of the error measure $B_{\xi_\alpha^\delta}(x_\alpha^\delta; x^\dagger)$ in (19) consists of two terms, the first δ-dependent noise propagation, and the second δ-independent term which expresses the smoothness of the solution $x^\dagger$ with respect to the forward operator A. In the study [18] such a decomposition was comprehensively analyzed for general linear regularization methods applied to (1) in a Hilbert space setting, i.e., for linear mappings $y^\delta \mapsto x_\alpha^\delta$, and for the norm as an error measure the δ-independent term was called *profile function* there, because this term completely determines the error profile. For the current setting, the index function Ψ plays a similar role, although the mapping $y^\delta \mapsto x_\alpha^\delta$ is *nonlinear* for general convex penalties J different from norm squares in Hilbert space X. This shows the substantial meaning of the right-hand function Φ in the inequality (13) of Assumption 2.

3. Relation to variational inequalities

In this section we shall exhibit that other regularity conditions, such as source conditions or variational inequalities yield bounds as required in Assumptions 2 and 2′, respectively. As already mentioned in the introduction source conditions yield a variational inequality, and we recall the following result.

Lemma 1 (see [9, Lemma 3.2]). *Suppose that $\partial J(x^\dagger) \cap \mathcal{R}(A^*) \neq \emptyset$. Then there is $R \geq 0$ such that*

$$(23) \qquad J(x^\dagger) - J(x) \leq R \left\| Ax^\dagger - Ax \right\| \quad \text{for all} \quad x \in X.$$

If otherwise $\xi^\dagger \notin \mathcal{R}(A^*)$, then a variational inequality

$$(24) \qquad J(x^\dagger) - J(x) \leq \Phi \left(\left\| Ax^\dagger - Ax \right\| \right) \quad \text{for all} \quad x \in X$$

with a sub-linear index function Φ holds, and the quotient function $\Phi(t)/t$ is strictly decreasing.

Proof. For the convenience of the reader we briefly sketch the proof. First, the non-negativity of the Bregman distance from (5) is an immediate consequence of the definition of the subgradient. Hence, for all convex functionals J we have for $\xi^\dagger \in \partial J(x^\dagger)$ and for all $x \in X$ the inequality

$$J(x^\dagger) - J(x) \leq \langle \xi^\dagger, x^\dagger - x \rangle.$$

If in addition $\xi^\dagger \in \mathcal{R}(A^*)$, then we can find a source element $w \in H$ with $\xi^\dagger = A^*w$ and $\|w\| \leq R$ such that we conclude from the above bound that

$$J(x^\dagger) - J(x) \leq \langle \xi^\dagger, x^\dagger - x \rangle = \langle w, A(x^\dagger - x) \rangle \leq R \left\| Ax^\dagger - Ax \right\|.$$

If such linear bound fails then by the method of *approximate variational source conditions* (cf. [12] and more comprehensively [8]), we consider the strictly positive and decreasing *distance function*

$$d(R) := \sup_{x \in X} \left\{ J(x^\dagger) - J(x) - R \left\| Ax^\dagger - Ax \right\| \right\}, \quad R > 0.$$

We find that $\lim_{R \to \infty} d(R) = 0$, and the decay rate to zero as $R \to \infty$ measures the degree of violation of the benchmark variational inequality (6). Plainly, with this function we find that

$$J(x^\dagger) - J(x) \leq \inf_{R > 0} \left\{ d(R) + R \left\| Ax^\dagger - Ax \right\| \right\}, \quad x \in X.$$

It is readily seen that the function

$$\Phi(t) := \inf_{R > 0} \left\{ d(R) + Rt \right\}, \quad t > 0,$$

being the infimum of affine functions, constitutes a (concave) sub-linear index function. Then, following [9, Lemma 3.2] a variational inequality of type (24) holds with

$$(25) \qquad \Phi(\alpha) = 2d(\Theta^{-1}(\alpha)), \quad \text{where} \quad \Theta(R) := d(R)/R.$$

Also, this function Φ is a sub-linear index function such that the quotient function $\Phi(\alpha)/\alpha$ is decreasing for all $\alpha > 0$. $\qquad \square$

From the estimates above, it is not difficult to verify that a source condition $\mathcal{R}(A^*) \cap \partial J(x^\dagger) \neq \emptyset$ also implies Assumptions 2 and 2′ with $\Psi(\alpha) \sim \alpha$. Indeed, as the source condition yields (23), we have by Young's inequality and (15) that

$$J(x^\dagger) - J(x_\alpha) \leq R\sqrt{\alpha} \frac{1}{\sqrt{\alpha}} \left\| Ax^\dagger - Ax_\alpha \right\| \leq R^2 \alpha + \frac{1}{4\alpha} \left\| Ax^\dagger - Ax_\alpha \right\|^2$$

$$\leq R^2 \alpha + \frac{1}{2} \left(J(x^\dagger) - J(x_\alpha) \right),$$

which implies Assumption 2 (and consequently also Assumption 2′) with $\Psi(\alpha) = 2R^2\alpha$. The obtained rate for the total error $B_{\xi_\alpha^\delta}(x_\alpha^\delta; x^\dagger) \sim \delta$ agrees with the classical results for this case.

Now, similarly to Proposition 2 we highlight that the choice of functions Φ in (24) is not arbitrary.

Proposition 3. *Suppose that a variational inequality* (24) *holds with an index function* Φ. *The following is true:*
Either $J(x^\dagger) = \min\limits_{x \in X} J(x)$,
and then any index function Φ *is a valid bound in* (24),
or $J(x^\dagger) > \min\limits_{x \in X} J(x)$,
and then Φ *increases near zero at most linearly.*

Proof. First, if $J(x^\dagger) = \min\limits_{x \in X} J(x)$ then the left-hand side in (24) is non-positive, and hence any non-negative upper bound is valid. Otherwise, suppose that $\Phi(t)/t$ decreases to zero as $t \to 0$. The inequality (24) taken at the point $x^\dagger + t(x - x^\dagger)$, $0 < t < 1$, attains the form

$$J(x^\dagger) - J((1-t)x^\dagger + tx) \leq \Phi(t\|Ax - Ax^\dagger\|),$$

where we can estimate from below the left-hand side as

$$J(x^\dagger) - J((1-t)x^\dagger + tx) \geq J(x^\dagger) - (1-t)J(x^\dagger) - tJ(x) = t(J(x^\dagger) - J(x)),$$

because J is a convex functional. From this we directly derive

$$J(x^\dagger) - J(x) \leq \frac{\Phi(t\|Ax - Ax^\dagger\|)}{t} = \|Ax - Ax^\dagger\| \frac{\Phi(t\|Ax - Ax^\dagger\|)}{t\|Ax - Ax^\dagger\|},$$

where under the assumption of the lemma the right-hand side tends to zero as $t \to 0$. Consequently, we have $J(x^\dagger) \leq J(x)$ for all $x \in X$. This completes the proof. $\qquad\square$

The main result in this section reads as follows:

Proposition 4. *Suppose that a variational inequality* (24) *holds for some index function* Φ. *Let us consider the related index function* $\tilde{\Phi}(t) := \Phi(\sqrt{t})$, $t > 0$. *Then the following assertions hold true.*

(1) *The condition* (17) *is valid with a function*

$$(26) \qquad \Psi(\alpha) = \sup_{t > 0} \left[\Phi(t) - \frac{t^2}{2\alpha} \right],$$

which is increasing for all $\alpha > 0$ *but that may take values* $+\infty$.
(2) *If the function* $\tilde{\Phi}$ *is concave then the function* Ψ *from* (26) *has the representation*

$$\Psi(\alpha) := \frac{\tilde{\Phi}^{-*}(2\alpha)}{2\alpha}, \quad \alpha > 0,$$

where $\tilde{\Phi}^{-*}$ *is the Fenchel conjugate to the convex index function* $\tilde{\Phi}^{-1}$ *(cf. its definition in* (11)*).*

(3) *Finally, if moreover the quotient function $s^2/\Phi(s)$, is an index function and hence strictly increasing for all $0 < s < \infty$, then Ψ also constitutes an index function. Theorem 1' yields the error estimate (19).*

Proof. For the first assertion we find that

$$\frac{1}{\alpha}\left(T_\alpha(x^\dagger; y) - T_\alpha(x_\alpha; y)\right) = J(x^\dagger) - J(x_\alpha) - \frac{1}{2\alpha}\|Ax_\alpha - Ax^\dagger\|^2$$

$$\leq \Phi(\|Ax_\alpha - Ax^\dagger\|) - \frac{1}{2\alpha}\|Ax_\alpha - Ax^\dagger\|^2.$$

Setting $t := \left\|Ax_\alpha^\delta - Ax^\dagger\right\|$ yields the function Ψ as stated.

Now suppose that the function $\tilde\Phi$ is a concave index function. Then its inverse is a convex index function, and by the definition of the Fenchel conjugate, see (11), we find

$$\sup_{t>0}\left[\Phi(t) - \frac{t^2}{2\alpha}\right] = \sup_{t>0}\left[\tilde\Phi(t^2) - \frac{t^2}{2\alpha}\right]$$

$$= \frac{1}{2\alpha}\sup_{s>0}\left[2\alpha s - \tilde\Phi^{-1}(s)\right] = \frac{\tilde\Phi^{-*}(2\alpha)}{2\alpha},$$

which proves the second assertion. It remains to establish that this function is an index function with the property as stated. To this end we aim at applying Corollary B.1 with $f(t) := \tilde\Phi^{-1}(t)$, $t > 0$. We observe, after substituting $t := \tilde\Phi(s^2)$, that

$$\frac{\tilde\Phi^{-1}(t)}{t} = \frac{s^2}{\Phi(s)}, \quad s > 0,$$

which was supposed to be strictly increasing from 0 to ∞. Thus Corollary B.1 applies, and the proof is complete. □

Under the conditions of item (2) of Proposition 4 we can immediately derive a convergence rate for the Bregman distance as error measure.

Proposition 5. *If the function Φ in (24) is such that $\tilde\Phi(t) := \Phi(\sqrt{t})$ is a concave index function, and with an appropriately selected α, the following convergence rate holds*

(27) $$B_{\xi_\alpha^\delta}(x_\alpha^\delta; x^\dagger) = \mathcal{O}(\Phi(\delta)) \quad as \quad \delta \to 0.$$

Proof. In the Fenchel-Young inequality (12), used for $f := \tilde\Phi^{-1}$, assigning $s := \tilde\Phi(\delta^2)$ and $t := 2\alpha$ we obtain

$$\Phi(\delta) = \tilde\Phi(\delta^2) \leq \frac{\delta^2}{2\alpha} + \frac{\tilde\Phi^{-*}(2\alpha)}{2\alpha}$$

Taking $2\alpha \in \partial\tilde{\Phi}^{-1}(\delta^2)$, which exists by continuity of $\tilde{\Phi}^{-1}$, yields equality in the Fenchel-Young and in the above inequality, thus, with such a choice and by (19)

$$B_{\xi_\alpha^\delta}(x_\alpha^\delta; x^\dagger) \leq \frac{\delta^2}{2\alpha} + \frac{\tilde{\Phi}^{-*}(2\alpha)}{2\alpha} = \Phi(\delta).$$

$\square$

We highlight the previous findings in case that the function Φ in (24) is a monomial.

Example 1. Let us prototypically consider the case that the function Φ is of power type, i.e., $\Phi(t) := t^\mu$, $t > 0$ for some $0 < \mu < \infty$. Then the function $\tilde{\Phi}$ is $\tilde{\Phi}(t) = t^{\mu/2}$. This function is concave whenever $0 < \mu \leq 2$. In that range also the quotients $s^2/\Phi(s)$, $s > 0$ are strictly increasing. For $\mu > 2$ the function Ψ is infinite for all $\alpha > 0$ and for $\mu = 2$ it is a positive constant. For $0 < \mu < 2$, however, Ψ is an index function.

Namely, the inverse of $\tilde{\Phi}$ equals $\tilde{\Phi}^{-1}(t) = t^{2/\mu}$, $t > 0$. By using the simple identity that $(cf)^*(t) = cf^*(t/c)$, $t > 0$, for a convex function f and $c > 0$ we see that the Fenchel conjugate function is for all $0 < \mu < 2$

$$\tilde{\Phi}^{-*}(t) = \frac{2 - \mu}{\mu}(\mu t/2)^{2/(2-\mu)}, \ t > 0.$$

Then the quotient

$$\frac{\tilde{\Phi}^{-*}(2\alpha)}{2\alpha} = \frac{2-\mu}{2}(\mu\alpha)^{\frac{\mu}{2-\mu}}, \quad \alpha > 0,$$

is a strictly increasing index function as predicted by the proposition. This function is sub-linear for $\mu/(2 - \mu) \leq 1$, i.e., for $0 < \mu \leq 1$, and hence may serve as a bound in Assumption 2, including the benchmark case $\Phi(t) = ct$, $t > 0$, in which case the corresponding function Ψ is also linear.

Remark 2. We know from Proposition 3 that in the non-singular case the function Φ is at most linear, i.e., the function $\Phi(s)/s$ is bounded away from zero. In particular this holds for concave index functions. In this particular case the function $s/\Phi(s)$ is non-decreasing, and hence the function $s(s/\Phi(s)) = s^2/\Phi(s)$ is an index function. Thus item (3) of Proposition 4 applies and yields that Assumption 2$'$ holds. Hence Theorem 1$'$ applies and gives a convergence rate.

Let us point out that the stated conditions not only imply rates in the Bregman distance but also in the strict metric. In [14], the strict metric

$$d_{st}(u, v) := \|Au - Av\| + |J(u) - J(v)|$$

was proposed as an alternative to the Bregman distance to measure convergence. Especially for ℓ^1- and BV-regularization, the strict metric has preferable properties compared to $B_\zeta(u; v)$.

We state the following result, from which bounds and convergence rates for the strict metric can easily be derived.

Proposition 6. *Suppose that Assumptions 2 or 2′ are satisfied. Then we have with some constant C,*

$$\left| J(x_\alpha^\delta) - J(x^\dagger) \right| \le C \left(\frac{\delta^2}{\alpha} + \Psi(\alpha) \right) \qquad \forall \alpha > 0$$

$$\| Ax_\alpha^\delta - Ax^\dagger \|^2 \le C \left(\delta^2 + \alpha \Psi(\alpha) \right) \qquad \forall \alpha > 0$$

Proof. At first we prove

(28) $$\| Ax_\alpha - Ax_\alpha^\delta \| \le \delta.$$

From the optimality conditions and the fact that the subgradient is a monotone operator, it follows with $\xi_\alpha^\delta \in \partial J(x_\alpha^\delta), \xi_\alpha \in \partial J(x_\alpha)$ that

$$\begin{aligned}
\| Ax_\alpha - Ax_\alpha^\delta \|^2 &= \langle A^* A(x_\alpha - x_\alpha^\delta), x_\alpha - x_\alpha^\delta \rangle = \langle A^* A(x_\alpha - x_\alpha^\delta), x_\alpha - x_\alpha^\delta \rangle \\
&= \langle A^*(y - y^\delta), x_\alpha - x_\alpha^\delta \rangle - \alpha \langle \xi_\alpha - \xi_\alpha^\delta, x_\alpha - x_\alpha^\delta \rangle \\
&\le \langle y - y^\delta, Ax_\alpha - Ax_\alpha^\delta \rangle \le \| y^\delta - y \| \| Ax_\alpha - Ax_\alpha^\delta \|,
\end{aligned}$$

which proves (28).

By definition of the Bregman distance and $\xi_\alpha^\delta = \frac{1}{\alpha} A^*(y^\delta - Ax_\alpha^\delta)$, we obtain

$$\begin{aligned}
J(x^\dagger) - J(x_\alpha^\delta) &= B_{\xi_\alpha^\delta}(x_\alpha^\delta; x^\dagger) + \tfrac{1}{\alpha} \langle y^\delta - Ax_\alpha^\delta, Ax^\dagger - Ax_\alpha^\delta \rangle \\
&\le B_{\xi_\alpha^\delta}(x_\alpha^\delta; x^\dagger) + \tfrac{1}{\alpha} \| y^\delta - y \| \| y - Ax_\alpha^\delta \| + \tfrac{1}{\alpha} \| y - Ax_\alpha^\delta \|^2 \\
&\le B_{\xi_\alpha^\delta}(x_\alpha^\delta; x^\dagger) + \frac{\delta^2}{2\alpha} + \tfrac{3}{2\alpha} \| y - Ax_\alpha^\delta \|^2,
\end{aligned}$$

where we used the Cauchy-Schwarz inequality in the last step. This bound and (16) yields

$$\frac{1}{\alpha} \| y - Ax_\alpha^\delta \|^2 \le \frac{2}{\alpha} \| y - Ax_\alpha \|^2 + \frac{2}{\alpha} \| Ax_\alpha - Ax_\alpha^\delta \| \le \Psi(\alpha) + \frac{2\delta^2}{\alpha},$$

from which (together with Theorems 1 and 1′) the second statement in the theorem is obtained. Since $J(x_\alpha^\delta) - J(x^\dagger) \le \frac{\delta^2}{2\alpha}$ (cf. [19, Lemma 1]), the desired bound for $|J(x_\alpha^\delta) - J(x^\dagger)|$ follows as well. □

4. Examples

Here we shall highlight the applicability of the main results in special situations. We start with the standard penalty in a Hilbert space context and then analyze other penalties as these are used in specific applications.

4.1. Quadratic Tikhonov regularization in Hilbert spaces.

Suppose we are in the classical context of Tikhonov regularization in Hilbert spaces X and Y, where the penalty is given as $J(x) := \frac{1}{2} \|x\|^2$, $x \in X$. In this case, which has been comprehensively discussed in the literature (cf., e.g., [7, Chap. 5] and [1, 2, 13, 16]), we can explicitly calculate the terms under consideration.

First, let $g_\alpha(\lambda) := 1/(\alpha + \lambda)$ be the filter from Tikhonov regularization and its companion $r_\alpha(\lambda) = \alpha/(\alpha + \lambda)$. With these short-hands, we see that $x_\alpha - x^\dagger = r_\alpha(A^*A)x^\dagger$, and also $A(x_\alpha - x^\dagger) = Ar_\alpha(A^*A)x^\dagger$. This yields

$$(29) \quad \frac{\left\| Ax_\alpha - Ax^\dagger \right\|^2}{2\alpha} = \frac{\left\| Ar_\alpha(A^*A)x^\dagger \right\|^2}{2\alpha} = \frac{\left\| r_\alpha(A^*A)\,(A^*A)^{1/2}\,x^\dagger \right\|^2}{2\alpha}.$$

We also see that

$$T_\alpha(x_\alpha; Ax^\dagger) = \frac{1}{2}\left(\left\| r_\alpha(A^*A)\,(A^*A)^{1/2}\,x^\dagger \right\|^2 + \alpha\,\|x_\alpha\|^2 \right)$$

$$= \frac{1}{2} \int \left[\frac{\alpha^2\lambda}{(\alpha + \lambda)^2} + \frac{\alpha\lambda^2}{(\alpha + \lambda)^2} \right] dE_\lambda\|x^\dagger\|^2$$

$$= \frac{1}{2} \int \frac{\alpha\lambda}{(\alpha + \lambda)} dE_\lambda\|x^\dagger\|^2,$$

which in turn yields

$$(30) \quad \begin{aligned} \frac{1}{\alpha}\left(T_\alpha(x^\dagger; y) - T_\alpha(x_\alpha; y) \right) &= \frac{1}{2} \int \frac{\alpha}{(\lambda + \alpha)} dE_\lambda\|x^\dagger\|^2 \\ &= \frac{1}{2} \left\| r_\alpha^{1/2}(A^*A)x^\dagger \right\|^2. \end{aligned}$$

Finally, we bound

$$(31) \quad \begin{aligned} J(x^\dagger) - J(x_\alpha) &= \frac{1}{2}\left(\left\|x^\dagger\right\|^2 - \|x_\alpha\|^2 \right) = \frac{1}{2}\langle x^\dagger - x_\alpha, x^\dagger + x_\alpha \rangle \\ &= \frac{1}{2}\langle r_\alpha(A^*A)x^\dagger, \left(I + (\alpha + A^*A)^{-1}\,A^*A \right)x^\dagger \rangle \\ &\leq \left\| r_\alpha^{1/2}(A^*A)x^\dagger \right\|^2. \end{aligned}$$

We observe that the right-hand sides in (30) and (31) differ by a factor $\frac{1}{2}$, as predicted in Proposition 1.

In the classical setup of Tikhonov regularization, a regularity condition is usually imposed by a source-condition. Thus, let us now assume that the element $x^\dagger$ obeys a source-wise representation

$$(32) \qquad\qquad x^\dagger = \varphi(A^*A)v, \qquad \|v\| \le 1,$$

for an index function φ. Then the estimate (31) reduces to bounding

$$(33) \qquad \begin{aligned} \left\| r_\alpha^{1/2}(A^*A)x^\dagger \right\|^2 &\le \left\| r_\alpha^{1/2}(A^*A)\varphi(A^*A) \right\|^2 \\ &\le \left\| r_\alpha(A^*A)\varphi^2(A^*A) \right\|, \end{aligned}$$

where we used the estimate $\left\| H^{1/2} \right\| \le \|H\|^{1/2}$ for a self-adjoint non-negative operator H.

Then, if the function φ^2 is sub-linear, we find that

$$\left\| r_\alpha(A^*A)\varphi^2(A^*A) \right\| \le \varphi^2(\alpha).$$

Hence, in the notation of [22], φ^2 is a qualification for Tikhonov regularization, and Assumption 2 holds true with the index function $\Psi(\alpha) = \varphi^2(\alpha)$. In particular, the rate (20), which is obtained by equilibrating both summands by letting the parameter α_* be given as solution to the equation $\Theta(\alpha_*) = \delta/\sqrt{2}$, yields the convergence rate

$$\left\| x^\dagger - x_{\alpha_*}^\delta \right\| \le 2\varphi\left(\Theta^{-1}(\delta/\sqrt{2}) \right),$$

which is known to be optimal in the "low smoothness" case, i.e., $x^\dagger \in \mathcal{R}(A^*)$.

Under the same condition on φ we can also bound the right-hand side in (29) as

$$\frac{\left\| r_\alpha(A^*A)\,(A^*A)^{1/2}\,x^\dagger \right\|^2}{2\alpha} \le \frac{(\sqrt{\alpha}\varphi(\alpha))^2}{2\alpha} = \frac{1}{2}\varphi^2(\alpha),$$

which verifies (16).

We finally turn to discussing the maximal rate at which the function Ψ may tend to zero as $\alpha \to 0$, provided that $x^\dagger \ne 0$. Considering the ratio $\Psi(\alpha)/\alpha$ we find

$$\begin{aligned} \frac{\Psi(\alpha)}{\alpha} &\ge \frac{\left\| r_\alpha(A^*A)\,(A^*A)^{1/2}\,x^\dagger \right\|^2}{2\alpha^2} = \frac{1}{2\alpha^2}\int \frac{\alpha^2\lambda}{(\lambda+\alpha)^2} dE_\lambda\|x^\dagger\|^2 \\ &\ge \frac{1}{8}\int_{\lambda \ge \alpha} \frac{1}{\lambda} dE_\lambda\|x^\dagger\|^2 = \frac{1}{8}\left\| \chi_{[\alpha,\infty)}(A^*A)\,(A^*A)^{-1/2}\,x^\dagger \right\|^2. \end{aligned}$$

This shows that either $x^\dagger \in \mathcal{D}\,(A^*A)^{-1/2}$, and hence that $x^\dagger \in \mathcal{R}\,(A^*)$, in which case the right-hand side is bounded away from zero (if $x^\dagger \ne 0$), or we have that $x^\dagger \notin \mathcal{R}\,(A^*)$, and the right-hand side diverges. Hence,

for nonzero $x^\dagger$ the best attainable rate near zero of the function Ψ is linear as also predicted in Proposition 2.

4.2. **ROF-Filter.** We consider the celebrated ROF-filter in image processing [27]: Let $y^\delta \in L^2(\mathbb{R}^2)$ represent an noisy image. Then a filtered version $x \in L^2(\mathbb{R}^2) \cap BV(\mathbb{R}^2)$ is computed by minimizing the Tikhonov functional

$$T_\alpha(x; y^\delta) = \frac{1}{2} \left\| x - y^\delta \right\|^2_{L^2(\mathbb{R}^2)} + \alpha |x|_{TV},$$

where $J(x) := |x|_{TV}$ denotes the total variation of x on $\mathbb{R}^2$. Obviously, this can be put into our framework with A being the embedding operator from $BV(\mathbb{R}^2)$ to $L^2(\mathbb{R}^2)$.

For some special cases, where $x^\dagger$ is the characteristic function of simple geometric shapes, the minimizers can be computed explicitly. Denote by $B_{x,R}$ a ball with center x and radius R. Consider first the case when $x^\dagger$ is the characteristic function of a ball $B_{0,R}$:

$$x^\dagger(s) = \chi_{B_{0,R}}(s) := \begin{cases} 1 & \text{if } \|s\|_{\mathbb{R}^2} \le 1, \\ 0 & \text{else.} \end{cases}$$

The minimizer x_α of $T_\alpha(.; y)$ with exact data is given by, e.g., [23]

$$x_\alpha = \max\{1 - \tfrac{2\alpha}{R}, 0\} \chi_{B_{0,R}}(s).$$

Calculating the index function in Assumption 2 is now a simple task as $|\chi_{B_{0,R}}|_{TV} = 2\pi R$

$$J(x^\dagger) - J(x_\alpha) = \Psi(\alpha) = 2\pi R \left(1 - \max\{1 - \tfrac{2\alpha}{R}, 0\}\right)$$

$$= 2\pi R \min\{\frac{2\alpha}{R}, 1\} = 4\pi\alpha \quad \text{if } \alpha < R/2.$$

For a comparison, we may compute the Bregman distances. For the asymptotically interesting case, $\alpha < \frac{R}{2}$ we find that

$$B_{\xi_\alpha}(x_\alpha; x^\dagger) = B_{\xi_\alpha}(x^\dagger; x_\alpha) = 0 \qquad \forall \alpha < \frac{R}{2},$$

which yields a trivial rate, but of course, does not violate the upper bound $\Phi(\alpha)$ in (19) for $\delta = 0$. The squared norm of the residual for $\alpha < \frac{R}{2}$ is given by

$$\|Ax_\alpha - y\|^2 = 4\pi\alpha^2,$$

hence, (16) clearly holds. We also observe that a variational inequality of the form (7), or (24) below, holds with $\Phi(s) \sim s$.

For noisy data, x_α^δ cannot be calculated analytically, but our results suggest for such $x^\dagger$ a suitable parameter choice of the form $\alpha = \delta$,

which provides a convergence rate

$$B_{\xi_\alpha^\delta}(x_\alpha^\delta; x^\dagger) \leq (4\pi + \frac{1}{2})\delta.$$

A less simple situation appears when the exact solution is the characteristic function of the unit square

$$x^\dagger(s) = \chi_{[0,1]^2}(s) = \begin{cases} 1 & \text{if } s \in [0,1]^2 \\ 0 & \text{else} \end{cases}.$$

An explicit solution is known here as well [6]. For $R > 0$ define the rounded square

$$C_R := \bigcup_{x:B_{x,R}\subset[0,1]^2} B_{x,R},$$

which has the shape of a square with the four corners cut off and replaced by circular arcs of radius R that meet tangentially the edges of the square. The solution satisfies $0 \leq x_\alpha \leq 1$ and can be characterized by the level sets: for $s \in [0,1]$

$$\{x_\alpha > s\} = \begin{cases} \emptyset & \text{if } s \geq 1 - \frac{\alpha}{R^*} \\ C_{\frac{\alpha}{1-s}} & \text{if } s \leq 1 - \frac{\alpha}{R^*} \end{cases}.$$

Here R^* is a limiting value, which can be computed explicitly. Since we are interested in the asymptotics $\alpha \to 0$, we generally impose the condition $\alpha \leq R^*$ as otherwise $x_\alpha = 0$. The index function Ψ can now be calculated by the coarea formula

$$J(x^\dagger) - J(x_\alpha) = \Psi(\alpha) = 4 - \int_0^{1-\frac{\alpha}{R^*}} |C_{\frac{\alpha}{1-s}}|_{TV} ds$$

The value of $|C_R|_{TV}$ is its perimeter and can be calculated by elementary geometry to $|C_R|_{TV} = 4-2(4-\pi)R$. Thus, evaluating the integral, we obtain

$$\Psi(\alpha) = \frac{4}{R^*}\alpha + 2(4-\pi)\alpha\left(\log\left(\frac{R^*}{\alpha}\right)\right) \qquad \alpha \leq R^*.$$

Thus, in this case,

$$\Psi(\alpha) \sim \alpha\log(1/\alpha) \qquad \text{as } \alpha \to 0.$$

The residual norm is given by

$$\|Ax_\alpha - y\|^2 = \|x_\alpha - x^\dagger\|_{L^2}^2$$
$$= \frac{\alpha^2}{R^{*2}} + 2(4-\pi)\alpha^2\left(\log\left(\frac{R^*}{\alpha}\right)\right) \qquad \alpha < R^*.$$

Obviously, the bound (16) is satisfied. The approximation error in the Bregman distance (with our choice of the subgradient element) is hence given by

$$B_{\xi_\alpha}(x_\alpha; x^\dagger) = J(x^\dagger) - J(x_\alpha^\delta) - \frac{1}{\alpha}\|Ax_\alpha - y\|^2 = \frac{3}{R^{*2}}\alpha \qquad \alpha < R^*.$$

We observe that, for the square, the parameter choice that minimizes the upper bound (19) differs from that for the ball as we have that $\alpha \sim C\frac{\delta}{(\log(1/\delta))^{\frac{1}{2}}}$, which highlights the (well-known) dependence of the parameter choice on the regularity of the exact solution.

Note also, that the decay of the Bregman distance $B_{\xi_\alpha}(x_\alpha; x^\dagger)$ alone does not suit well as a measure of regularity for $x^\dagger$ since the logarithmic factor that appears in the condition in Assumption 2 is not observed for this Bregman distance. On the contrary, when using the strict metric, this factor is included, which indicates that this metric might be more appropriate for BV-regularization than the Bregman distance (compare [14]).

4.3. On ℓ^1-regularization when sparsity is slightly missing.

We consider the *injective* continuous linear operator $A\colon \ell^1 \to \ell^2$ and the penalty $J(x) := \|x\|_{\ell^1}$. Notice that $\ell^1 = c_0^*$, it thus has a predual, and we assume that A is weak*-to-weak continuous, and the penalty J is stabilizing in this sense.

The crucial additional assumption on the operator A is that the unit elements $e^{(k)}$, with $e_k^{(k)} = 1$ and $e_i^{(k)} = 0$ for $i \neq k$, satisfy source conditions $e^{(k)} = A^* f^{(k)}$, $f^{(k)} \in Y$, for all $k \in \mathbb{N}$. Under these assumptions (for overcoming such additional requirements see [11]), and with $x^\dagger = (x_k^\dagger)_{k \in \mathbb{N}} \in X$ from (1), we assign the function

$$(34) \qquad \Phi(t) = 2 \inf_{n \in \mathbb{N}} \left(\sum_{k=n+1}^{\infty} |x_k^\dagger| + t \sum_{k=1}^{n} \|f^{(k)}\|_Y \right), \quad t > 0.$$

Notice that the function Φ from (34) is a concave index function. It was shown in [4] that then a variational inequality of the form

$$\|x - x^\dagger\|_X \le \|x\|_X - \|x^\dagger\|_X + \Phi\left(\left\|Ax^\dagger - Ax\right\|\right) \quad \text{for all} \quad x \in X$$

holds true. This immediately implies the validity of the condition (7) with the same index function Φ, and an application of item (3) of Proposition 4 shows that the error estimate (19) is valid for that Φ.

The behavior of the index function Φ from (34) essentially depends on the decay rate of the tail of $x_k^\dagger \to 0$ of the solution element $x^\dagger$. When sparsity is (slightly) missing, then the function Φ will be strictly

concave. However, if $x^\dagger$ is sparse, i.e., $x_k^\dagger = 0$ for $k > n_{max}$, then the function Φ reduces to the linear function

$$\Phi(t) = \left(\sum_{k=1}^{n_{max}} \|f^{(k)}\|_Y \right) t, \quad t > 0.$$

As Example 1 highlights, this results in a linear companion function Ψ. Thus Theorem 1$'$ applies, and the choice of $\alpha \sim \delta$ yields a rate for the Bregman distance $B_{\xi_\alpha^\delta}(x_\alpha^\delta; x^\dagger) = \mathcal{O}(\delta)$ as $\delta \to 0$ in the sparse case.

5. Outlook to higher order rates

There might be a way for overcoming the limitation of sub-linear functions Ψ in the assumptions 2 or 2$'$. The underlying observation for this is the identity

$$(35) \qquad B_{\xi_\alpha}\left(x_\alpha; x^\dagger\right) = \frac{2}{\alpha}(T_\alpha(x^\dagger, y) - T_\alpha(x_\alpha, y)) - (J(x^\dagger) - J(x_\alpha)).$$

The right-hand side above is again entirely based on noise-free quantities, and its decay could be used as smoothness assumption.

If one could prove that there were an inequality of the form

$$(36) \qquad B_{\xi_\alpha^\delta}\left(x_\alpha^\delta; x^\dagger\right) \leq C_1\, B_{\xi_\alpha}\left(x_\alpha; x^\dagger\right) + C_2\,\delta^2/\alpha, \quad \alpha > 0,$$

with positive constants C_1 and C_2, then this might open the pathway for higher order rates. Indeed, in Hilbert space X and for the standard penalty $J(x) := \frac{1}{2}\|x\|_X^2$, cf. Section 4.1, we find that $B_{\xi_\alpha^\delta}\left(x_\alpha^\delta; x^\dagger\right) = \|x_\alpha^\delta - x^\dagger\|_X^2$, and hence that the inequality (36) is satisfied with $C_1 = 2$ and $C_2 = 1$. Moreover, one can easily verify that

$$\frac{2}{\alpha}(T_\alpha(x^\dagger, y) - T_\alpha(x_\alpha, y)) - \left(J(x^\dagger) - J(x_\alpha)\right) = \frac{1}{2}\left\|r_\alpha(A^*A)x^\dagger\right\|^2,$$

with $r_\alpha(A^*A) = \alpha\,(\alpha I + A^*A)^{-1}$, being the (squared) residual for (standard linear) Tikhonov regularization. This squared residual is known to decay of order up to $\mathcal{O}(\alpha^2)$ as $\alpha \to 0$, which then allows for higher rates $B_{\xi_\alpha^\delta}\left(x_\alpha^\delta; x^\dagger\right) = \mathcal{O}(\delta^{4/3})$, attained under the limiting source condition $x^\dagger = A^*Aw$, $w \in X$, and for the a priori parameter choice $\alpha \sim \delta^{2/3}$. It is thus interesting to see whether and under which additional assumptions an inequality of the form (36) holds.

Appendix A. Proofs

Let us define the noisy and exact residuals, and the noise term as

$$(37)\quad p_\alpha^\delta := Ax_\alpha^\delta - y^\delta, \; p_\alpha := Ax_\alpha - y = A(x_\alpha - x^\dagger) \quad \text{and } \Delta := y^\delta - y.$$

Notice that all quantities $p_\alpha, p_\alpha^\delta$ as well as Δ belong to the Hilbert space H. The subsequent analysis will be based on the optimality

conditions (recall our convention on the choice of $\xi_\alpha^\delta \in \partial J(x_\alpha^\delta)$ and $\xi_\alpha \in \partial J(x_\alpha)$)

(38) $$\langle Ax_\alpha^\delta - y^\delta, Aw \rangle + \alpha\langle \xi_\alpha^\delta, w \rangle = 0, \qquad \forall w \in X,$$

(39) $$\langle Ax_\alpha - Ax^\dagger, Aw \rangle + \alpha\langle \xi_\alpha, w \rangle = 0, \qquad \forall w \in X.$$

In particular, the optimality conditions lead to the following formulas, by

(1) subtracting (38) from (39) using $w = x_\alpha - x^\dagger$,
(2) using (38) with $w = x_\alpha - x_\alpha^\delta$, and
(3) using (39) with $w = x_\alpha - x^\dagger$,

respectively:

(40) $$\langle \xi_\alpha - \xi_\alpha^\delta, x^\dagger - x_\alpha \rangle = -\frac{1}{\alpha}\langle p_\alpha^\delta - p_\alpha, p_\alpha \rangle$$

(41) $$-\langle \xi_\alpha^\delta, x_\alpha - x_\alpha^\delta \rangle = \frac{1}{\alpha}\langle p_\alpha^\delta, p_\alpha - p_\alpha^\delta - \Delta) \rangle$$

(42) $$-\langle \xi_\alpha, x^\dagger - x_\alpha \rangle = -\frac{1}{\alpha}\|p_\alpha\|^2.$$

The following bounds will be the key for proving Theorem 1$'$.

Lemma 2. *Under Assumption 2$'$ we have*

(1) $B_{\xi_\alpha}(x_\alpha; x^\dagger) \leq \Psi(\alpha) - \frac{1}{2\alpha}\|p_\alpha\|^2.$
(2) $B_{\xi_\alpha^\delta}(x_\alpha^\delta; x_\alpha) \leq \frac{\delta^2}{2\alpha} - \frac{1}{2\alpha}\|p_\alpha\|^2 + \frac{1}{\alpha}\langle p_\alpha^\delta, p_\alpha \rangle$

Proof. Using the optimality condition (42) we find that

$$B_{\xi_\alpha}(x_\alpha; x^\dagger) + \frac{1}{2\alpha}\|p_\alpha\|^2$$

$$= J(x^\dagger) - J(x_\alpha) - \langle \xi_\alpha, x^\dagger - x_\alpha \rangle + \frac{1}{2\alpha}\|p_\alpha\|^2$$

$$= J(x^\dagger) - J(x_\alpha) - \frac{1}{2\alpha}\|p_\alpha\|^2$$

$$= \frac{1}{\alpha}\left(T_\alpha(x^\dagger; y) - T_\alpha(x_\alpha; y)\right) \leq \Psi(\alpha),$$

which proves the first assertion.

For proving the second assertion we use the definition of $B_{\xi_\alpha^\delta}(x_\alpha^\delta; x_\alpha)$ and (41) to find

(43) $$B_{\xi_\alpha^\delta}(x_\alpha^\delta; x_\alpha) = J(x_\alpha) - J(x_\alpha^\delta) + \frac{1}{\alpha}\langle p_\alpha^\delta, p_\alpha - p_\alpha^\delta - \Delta) \rangle.$$

The minimizing property of x_α also yields

$$J(x_\alpha) - J(x_\alpha^\delta) \leq \frac{1}{2\alpha} \left[\left\| Ax_\alpha^\delta - y \right\|^2 - \left\| Ax_\alpha - y \right\|^2 \right]$$

$$= \frac{1}{2\alpha} \left[\left\| p_\alpha^\delta + \Delta \right\|^2 - \left\| p_\alpha \right\|^2 \right]$$

We rewrite

$$\langle p_\alpha^\delta, p_\alpha - p_\alpha^\delta - \Delta) \rangle = - \left\| p_\alpha^\delta \right\|^2 + \langle p_\alpha^\delta, p_\alpha - \Delta \rangle.$$

Using this and plugging the above estimate into (43) gives

$$B_{\xi_\alpha^\delta}(x_\alpha^\delta; x_\alpha) \leq \frac{1}{2\alpha} \left[\left\| p_\alpha^\delta + \Delta \right\|^2 - \left\| p_\alpha \right\|^2 \right] - \frac{1}{\alpha} \left\| p_\alpha^\delta \right\|^2 + \frac{1}{\alpha} \langle p_\alpha^\delta, p_\alpha - \Delta \rangle$$

$$= \frac{1}{2\alpha} \left\| \Delta \right\|^2 - \frac{1}{2\alpha} \left\| p_\alpha \right\|^2 - \frac{1}{2\alpha} \left\| p_\alpha^\delta \right\|^2 + \frac{1}{\alpha} \langle p_\alpha^\delta, p_\alpha \rangle$$

$$\leq \frac{\delta^2}{2\alpha} - \frac{1}{2\alpha} \left\| p_\alpha \right\|^2 + \frac{1}{\alpha} \langle p_\alpha^\delta, p_\alpha \rangle,$$

completing the proof of the second assertion and of the lemma.   □

We are now in a position to give detailed proofs of the results in Section 2.

Proof of Proposition 1. From item (1) of Lemma 2 we know that, for all $\alpha > 0$,

$$\frac{\left\| Ax_\alpha - y \right\|^2}{2\alpha} = \frac{\left\| p_\alpha \right\|^2}{2\alpha} \leq \Psi(\alpha).$$

Therefore, Assumption 2′ implies that

$$J(x^\dagger) - J(x_\alpha) = J(x^\dagger) - J(x_\alpha) - \frac{\left\| Ax_\alpha - y \right\|^2}{2\alpha} + \frac{\left\| p_\alpha \right\|^2}{2\alpha}$$

$$\leq \Psi(\alpha) + \frac{\left\| p_\alpha \right\|^2}{2\alpha} \leq 2\Psi(\alpha),$$

which completes the proof.   □

Proof of Proposition 2. First, if $\inf_{x \in X} J(x) = J(x^\dagger)$ then

$$\alpha J(x^\dagger) = T_\alpha(x^\dagger; Ax^\dagger) \geq T_\alpha(x_\alpha; Ax^\dagger) \geq \alpha J(x_\alpha) \geq \alpha J(x^\dagger),$$

and for all $\alpha > 0$ we have $J(x_\alpha) = J(x^\dagger)$. This allows us to prove the assertion in the first (singular) case. Otherwise, assume to the contrary that there is an index function Ψ, and for a decreasing sequence of regularization parameters $(\alpha_k)_k$ with $\lim_{k \to \infty} \alpha_k = 0$ the limit condition

$$\lim_{k \to \infty} \frac{\Psi(\alpha_k)}{\alpha_k} = 0$$

holds. Consequently, we find from (16) that $\lim_{k\to\infty}\frac{1}{\alpha_k}\|Ax_{\alpha_k}-y\|=0$, with $y=Ax^\dagger$.

Due to the optimality condition (39) for x_α we have that

$$A^*(Ax_{\alpha_k}-y)+\alpha_k\xi_{\alpha_k}=0 \quad \text{for some} \quad \xi_{\alpha_k}\in\partial J(x_{\alpha_k})\subset X^*.$$

This yields

$$\xi_{\alpha_k}=-\frac{1}{\alpha_k}A^*(Ax_{\alpha_k}-y).$$

Since $A^*:H\to X^*$ is a bounded linear operator, we get

$$\|\xi_{\alpha_k}\|_{X^*}\leq\frac{1}{\alpha_k}\|A^*\|_{\mathcal{L}(H,X^*)}\|Ax_{\alpha_k}-y\|\to 0 \quad \text{as} \quad k\to\infty.$$

Since $\xi_{\alpha_k}\in\partial J(x_{\alpha_k})$, and after taking the limit, we find for all $z\in X$ that

$$J(z)\geq\limsup_{k\to\infty}\{J(x_{\alpha_k})+\langle\xi_{\alpha_k},z-x_{\alpha_k}\rangle\}=\limsup_{k\to\infty}J(x_{\alpha_k})=J(x^\dagger),$$

where we used Assumption 2 and $\lim_{k\to\infty}\Psi(\alpha_k)=0$. Thus, we conclude that $\inf_{x\in X}J(x)=J(x^\dagger)$. This contradicts the assumption, and hence the function Ψ cannot decrease to zero super-linearly as $\alpha\to 0$. $\square$

Proof of Theorem 1'. Here we recall the three-point identity (see, e.g., [29]). For $u,v,w\in X$ and $\xi\in\partial J(w)$, $\eta\in\partial J(v)$, we have that

$$(44)\qquad B_\xi(w;u)=B_\eta(v;u)+B_\xi(w;v)+\langle\eta-\xi,u-v\rangle,$$

and this specifies with $u:=x^\dagger$, $v:=x_\alpha$, and $w:=x_\alpha^\delta$ to

$$(45)\quad B_{\xi_\alpha^\delta}(x_\alpha^\delta;x^\dagger)=B_{\xi_\alpha}(x_\alpha;x^\dagger)+B_{\xi_\alpha^\delta}(x_\alpha^\delta;x_\alpha)+\langle\xi_\alpha-\xi_\alpha^\delta,x^\dagger-x_\alpha\rangle,$$

Inserting (40) into (45) gives

$$B_{\xi_\alpha^\delta}(x_\alpha^\delta;x^\dagger)=B_{\xi_\alpha}(x_\alpha;x^\dagger)+B_{\xi_\alpha^\delta}(x_\alpha^\delta;x_\alpha)-\frac{1}{\alpha}\langle p_\alpha^\delta-p_\alpha,p_\alpha\rangle$$

An application of the bounds in Lemma 2 provides us with the estimate

$$B_{\xi_\alpha^\delta}(x_\alpha^\delta;x^\dagger)\leq\Psi(\alpha)+\frac{\delta^2}{2\alpha}-\frac{1}{\alpha}\|p_\alpha\|^2+\frac{1}{\alpha}\langle p_\alpha^\delta,p_\alpha\rangle-\frac{1}{\alpha}\langle p_\alpha^\delta-p_\alpha,p_\alpha\rangle$$

$$=\Psi(\alpha)+\frac{\delta^2}{2\alpha},$$

and the proof is complete. $\square$

APPENDIX B. SOME CONVEX ANALYSIS FOR INDEX FUNCTIONS

We shall provide some additional details for convex index functions. First, it is well known that for convex index function f we have that $0 < s \leq t$ yields $f(s)/s \leq f(t)/t$. Indeed, we let $0 < \theta := s/t \leq 1$ and obtain that

$$f(s) = f(\theta t + (1 - \theta)0) \leq \theta f(t) + (1 - \theta)f(0) = \frac{s}{t}f(t),$$

which allows us to prove the assertion. This implies that the limit $g := \lim_{t \to 0} f(t)/t \geq 0$ exists. If $g > 0$ then f is linear near zero, and this case is not interesting in this study. Otherwise, we assume that $g = 0$. In this interesting (sub-linear) case the following result is relevant.

Lemma 3. *Suppose that f is a convex index function. The following assertions are equivalent.*

(1) *The quotient $f(t)/t$, $t > 0$ is a strictly increasing index function.*
(2) *There is a strictly increasing index function φ, and the companion $\Theta(t) := \sqrt{t}\varphi(t)$, $t > 0$ such that the representation $f(t) = \Theta^2\left(\left(\varphi^2\right)^{-1}(t)\right)$, $t > 0$ is valid.*

Proof. Clearly, if f has a representation as in (2) then we find, with letting $\varphi^2(s) = t$, that

$$\frac{f(t)}{t} = \frac{\Theta^2(s)}{\varphi^2(s)} = s \searrow 0,$$

as $s \to 0$.

For the other implication we observe that by assumption we can (implicitly) define the strictly increasing index function φ by

$$(46) \qquad\qquad \varphi\left(\frac{f(t)}{t}\right) := \sqrt{t}, \quad t > 0.$$

This yields that

$$f(t) = t\left(\varphi^2\right)^{-1}(t) = \Theta^2\left(\left(\varphi^2\right)^{-1}(t)\right), \quad t > 0,$$

which completes the proof.   □

As an interesting consequence we mention the following result for the Fenchel conjugate function f^* to the convex (index) function f.

Corollary B.1. *Suppose that f is a convex index function such that the quotient $f(t)/t$, $t > 0$ is a strictly increasing index function. Then*

the Fenchel conjugate function f^* is an index function, and there is a strictly increasing index function φ such that

$$\frac{f^*(t)}{t} \leq \varphi^2(t), \quad t > 0.$$

Proof. First, by Lemma 3 there is a strictly increasing index function φ such that $f(t) = \Theta^2\left(\left(\varphi^2\right)^{-1}(t)\right)$, $t > 0$. Now we use the "poor man's Young Inequality" of the form

$$\varphi^2(x)y \leq \varphi^2(x)x + \varphi^2(y)y, \quad x, y > 0,$$

which is easily seen by considering the cases $x < y$ and $y \leq x$, separately. This in turn, by letting $s := \varphi^2(x)$ and $t := y$, implies

$$st \leq \Theta^2\left(\left(\varphi^2\right)^{-1}(s)\right) + \Theta^2(t), \quad s, t > 0.$$

For the Fenchel conjugate f^* this yields

$$f^*(t) := \sup_{s>0}\left\{st - f(s)\right\} \leq \Theta^2(t).$$

From this bound we conclude that f^* will be an index function for which the quotient $f^*(t)/t$ has the desired bound. $\square$

Acknowledgment

References

[1] V. Albani, P. Elbau, M. V. de Hoop, and O. Scherzer, *Optimal convergence rates results for linear inverse problems in Hilbert spaces*, Numer. Funct. Anal. Optim., 37 (2016), pp. 521–540.

[2] R. Andreev, P. Elbau, M. de Hoop, L. Qiu, and O. Scherzer, *Generalized convergence rates results for linear inverse problems in Hilbert spaces*, Numer. Funct. Anal. Optim., 36 (2015), pp. 549–566.

[3] S. W. Anzengruber, B. Hofmann, and P. Mathé, *Regularization properties of the sequential discrepancy principle for Tikhonov regularization in Banach spaces*, Appl. Anal., 93 (2014), pp. 1382–1400.

[4] M. Burger, J. Flemming, and B. Hofmann, *Convergence rates in ℓ^1-regularization if the sparsity assumption fails*, Inverse Problems, 29 (2013), p. 025013 (16pp).

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

[6] A. Chambolle, V. Caselles, D. Cremers, M. Novaga, and T. Pock, *An introduction to total variation for image analysis*, in Theoretical foundations and numerical methods for sparse recovery, vol. 9 of Radon Ser. Comput. Appl. Math., Walter de Gruyter, Berlin, 2010, pp. 263–340.

[7] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of inverse problems*, vol. 375 of Mathematics and its Applications, Kluwer Academic Publishers Group, Dordrecht, 1996.

[8] J. Flemming, *Generalized Tikhonov Regularization and Modern Convergence Rate Theory in Banach Spaces*, Shaker Verlag, Aachen, 2012.

[9] ———, *A converse result for Banach space convergence rates in Tikhonov-type convex regularization of ill-posed linear equations*, J. Inverse Ill-Posed Probl., 26 (2018). https://doi.org/10.1515/jiip-2017-0116; aop.

[10] J. Flemming, *Existence of variational source conditions for nonlinear inverse problems in Banach spaces*, J. Inverse Ill-Posed Probl., 26 (2018), pp. 277–286.

[11] J. Flemming and D. Gerth, *Injectivity and weak*-to-weak continuity suffice for convergence rates in ℓ^1-regularization*, J. Inv. Ill-Posed Probl., 26 (2018), pp. 85–94.

[12] J. Flemming and B. Hofmann, *A new approach to source conditions in regularization with general residual term*, Numer. Funct. Anal. Optim., 31 (2010), pp. 254–284.

[13] J. Flemming, B. Hofmann, and P. Mathé, *Sharp converse results for the regularization error using distance functions*, Inverse Problems, 27 (2011), pp. 025006, 18.

[14] K. Frick, D. A. Lorenz, and E. Resmerita, *Morozov's principle for the augmented Lagrangian method applied to linear inverse problems*, Multiscale Model. Simul., 9 (2011), pp. 1528–1548.

[15] M. Grasmair, *Generalized Bregman distances and convergence rates for nonconvex regularization methods*, Inverse Problems, 26 (2010).

[16] T. Hohage and F. Weidling, *Characterizations of variational source conditions, converse results, and maxisets of spectral regularization methods*, SIAM J. Numer. Anal., 55 (2017), pp. 598–620.

[17] B. Hofmann, B. Kaltenbacher, C. Pöschl, and O. Scherzer, *A convergence rates result for Tikhonov regularization in Banach spaces with nonsmooth operators*, Inverse Problems, 23 (2007), pp. 987–1010.

[18] B. Hofmann and P. Mathé, *Analysis of profile functions for general linear regularization methods*, SIAM J. Numer. Anal., 45 (2007), pp. 1122–1141.

[19] ———, *Parameter choice in Banach space regularization under variational inequalities*, Inverse Problems, 28 (2012), pp. 104006, 17.

[20] K. Ito and B. Jin, *Inverse Problems: Tikhonov Theory and Algorithms*, vol. 22 of Series on Applied Mathematics, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2015.

[21] S. Kindermann, *Convex Tikhonov regularization in Banach spaces: new results on convergence rates*, J. Inverse Ill-Posed Probl., 24 (2016), pp. 341–350.

[22] P. Mathé and S. V. Pereverzev, *Geometry of linear ill-posed problems in variable Hilbert scales*, Inverse Problems, 19 (2003), pp. 789–803.

[23] Y. Meyer, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*, AMS, Providence, RI, 2001.

[24] A. Neubauer, T. Hein, B. Hofmann, S. Kindermann, and U. Tautenhahn, *Improved and extended results for enhanced convergence rates of Tikhonov regularization in Banach spaces*, Appl. Anal., 89 (2010), pp. 1729–1743.

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

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

[27] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D, 60 (1992), pp. 259–268.

[28] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen, *Variational Methods in Imaging*, vol. 167 of Applied Mathematical Sciences, Springer, New York, 2009.

[29] T. Schuster, B. Kaltenbacher, B. Hofmann, and K. S. Kazimierski, *Regularization methods in Banach spaces*, vol. 10 of Radon Series on Computational and Applied Mathematics, Walter de Gruyter GmbH & Co. KG, Berlin, 2012.

(Bernd Hofmann) Faculty of Mathematics, Chemnitz University of Technology, 09107 Chemnitz, Germany
  E-mail address: `bernd.hofmann@mathematik.tu-chemnitz.de`

(Stefan Kindermann) Industrial Mathematics Institute, Johannes Kepler University Linz, Alternbergergstrasse 69, 4040 Linz
  E-mail address: `kindermann@indmath.uni-linz.ac.at`

(Peter Mathé) Weierstrass Institute, Mohrenstrasse 39, 10117 Berlin, Germany
  E-mail address: `peter.mathe@wias-berlin.de`