Johann Radon Institute for
Computational and Applied Mathematics
Austrian Academy of Sciences (ÖAW)

Morozov's discrepancy principle for Tikhonov-type functionals with non-linear operators

S. Anzengruber, R. Ramlau

RICAM-Report 2009-13

Morozov's Discrepancy Principle for Tikhonov-type functionals with non-linear operators

Stephan W Anzengruber* Ronny Ramlau*†

June 5, 2009

Abstract

In this paper we deal with Morozov's discrepancy principle as an a-posteriori parameter choice rule for Tikhonov regularization with general convex penalty terms Ψ for non-linear inverse problems. It is shown that a regularization parameter α fulfilling the discprepancy principle exists, whenever the operator F satisfies some basic conditions, and that for this parameter choice rule holds $\alpha \to 0, \delta^q/\alpha \to 0$ as the noise level δ goes to 0. It is illustrated that for suitable penalty terms this yields convergence of the regularized solutions to the true solution in the topology induced by Ψ. Finally, we establish convergence rates with respect to the generalized Bregman distance and a numerical example is presented.

1 Introduction

We will be concerned with the computation of approximate solutions x of an ill-posed problem of the form

$$F(x) = y, \tag{1}$$

where $F : X \to Y$ is a (non-linear) operator between reflexive Banach spaces X, Y. Additionally we assume that only noisy data y^δ with

$$\|y^\delta - y\| \le \delta,$$

is available. The mathematical formulation of a large variety of technical and physical problems – such as, for example, medical imaging and inverse scattering – result in inverse problems that are of this type, where the noise in the data usually appeares due to inaccuracies in the measurement process.

Since we are dealing with ill-posed problems, some form of regularization technique is needed to stabilize the inversion of F, see [5] for more details. One way to achieved this is by minimizing a Tikhonov-type functional

$$J_{\alpha,q}(x) = \|F(x) - y^\delta\|^q + \alpha\Psi(x), \tag{2}$$

*Radon Institute for Computational and Applied Mathematics (RICAM), Austrian Academy of Sciences, Altenbergerstraße 69, 4040 Linz, Austria, email: stephan.anzengruber@ricam.oeaw.ac.at

†Institute for Industrial Mathematics, Johannes Kepler University, Altenbergerstraße 69, 4040 Linz, Austria

with $\alpha > 0$ and $q > 1$.

In this paper we will – for the most part – make only very general assumptions regarding the penalty term Ψ, which will allow for a wide range of possible choices to be made according to specific properties required of the solution, such as, e.g., sparsity promoting functionals (usually weighted ℓ_p norms on the coefficients with respect to some orthonormal basis or frame), functionals related to the total variation for suitable spaces, but also classical Tikhonov regularization, where $\Psi(x)$ is the square of the Banach space norm.

For the quality of the reconstructed solution obtained by minimizing Tikhonov-type functionals as in (2), the choice of the regularization parameter α is crucial. Various results regarding the convergence – and also the rate of convergence – of regularized solutions to a true solution using general penalty terms in Banach spaces for linear operators can be found in [3, 14]. For the case of a nonlinear operator see [15], for Hilbert spaces see [7], and for sparsity promoting penalty terms see [11, 13]. In these papers the parameter is generally assumed to be chosen according to an *a priori* choice rule, which means that the choice only depends on the noise level δ and not on the actually available data y^δ. Moreover, in many cases convergence rates are proven under the additional assumption that the choice of the parameter involves knowledge of certain properties of the searched-for solution $x^\dagger$, such as its smoothness. In most practical applications such knowledge will not be at hand.

An example of an *a posteriori* parameter choice rule, i.e. a rule to determine α which incorporates the data, is Morozov's Discrepancy principle. Here we are interested in choosing $\alpha = \alpha(\delta, y^\delta)$ such that for constants $1 < \tau_1 \le \tau_2$,

$$\tau_1\delta \le \|F(x_\alpha^\delta) - y^\delta\| \le \tau_2\delta$$

holds, where x_α^δ denotes the regularized solution obtained by minimizing (2). This strategy has been studied extensively as an option to be used in the classical Tikhonov setting [5, 12, 16], but not until very recently has its application to inverse problems with more general penalty terms been investigated further. First results were obtained for denoising, where the operator is the identity, when using the L^1-norm as the penalty term, see [10]. Bonesky [1] considered linear inverse problems combined with the discrepancy principle and showed regularization properties and convergence rates in the Bregman distance.

Once the regularization parameter has been chosen, it remains to compute the related regularized solution as the minimizer of the Tikhonov-type functional (2). Different methods to achieved this can be found in [2, 4, 8, 11, 13].

In this paper, we will analyze under what circumstances the discrepancy principle can be applied to the non-linear inverse problem (1). We will see that the resulting parameter choice rule has the property that $\alpha(\delta, y^\delta) \to 0$ and $\delta^q/\alpha(\delta, y^\delta) \to 0$ as $\delta \to 0$. As a consequence it will follow – for suitable penalty terms – that the regularized solutions x_α^δ converge to a true solution $x^\dagger$ in the topology induced by the penalizing functional Ψ, i.e. that

$$\Psi(x_\alpha^\delta - x^\dagger) \to 0 \quad \text{as} \quad \delta \to 0. \tag{3}$$

In addition, we will prove convergence rates of order $\mathcal{O}(\delta)$ with respect to the generalized Bregman distance induced by the penalty term under the assumption that a source condition and a non-linearity condition are satisfied.

2 Preliminaries

Throughout this paper we assume the operator $F : \mathcal{D}(F) \subset X \to Y$, with $0 \in \mathcal{D}(F)$, to be weakly continuous, $q > 1$ to be fixed, and that the penalty term $\Psi(x)$ fulfills the following

Condition 2.1. Let $\Psi : \mathcal{D}(\Psi) \subset X \to \mathbb{R}^+$, with $0 \in \mathcal{D}(\Psi)$, be a convex functional such that

(i) $\Psi(x) = 0$ if and only if $x = 0$,

(ii) Ψ is weakly lower semi continous (w.r.t. the Banach space topology on X),

(iii) Ψ is weakly coercive, i.e. $\|x_n\| \to \infty \implies \Psi(x_n) \to \infty$,

The following consequence of the above conditions will be needed later on.

Lemma 2.2. *If Ψ satisfies Condition 2.1, then for any sequence $\{x_n\} \subset X$ with $\Psi(x_n) \to 0$ it holds that $x_n \rightharpoonup 0$.*

Proof. Take an arbitrary subsequence of $\{x_n\}$ – again denoted by $\{x_n\}$ for simplicity – then $\{\Psi(x_n)\}$ is bounded, and because of the weak coercivity of Ψ, so is $\{x_n\}$. Therefore we can extract a weakly convergent subsequence, $x_{n'} \rightharpoonup \bar{x}$. Due to the weak lower semi-continuity of Ψ we obtain

$$0 \le \Psi(\bar{x}) \le \liminf \Psi(x_{n'}) = 0,$$

which according to condition 2.1 (i) only holds for $\bar{x} = 0$. Altogether we have shown that any subsequence of $\{x_n\}$ has a subsequence that converges weakly to 0 and therefore the same holds true for the entire sequence.

□

In the preceding proof we have used a well known convergence principle in Banach spaces, which can be found in [18, Proposition 10.13]. We now state a slightly different version of this convergence principle, which we will repeatedly use throughout this paper.

Lemma 2.3. *Let $\{x_n\} \subset X$ and a functional $f : X \to \mathbb{R}$ be such that every subsequence $\{x_{n'}\}$ of $\{x_n\}$ has, in turn, a subsequence $\{x_{n''}\}$ such that $f(x_{n''}) \to c \in \mathbb{R}$ as $n'' \to \infty$, then $f(x_n) \to c$ as $n \to \infty$.*

Proof. If $f(x_n) \to c$ does not hold then there is a subsequence $\{x_{n'}\}$ such that $|f(x_{n'}) - c| > \varepsilon$ for some $\varepsilon > 0$. This contradicts the assumption that $\{x_{n'}\}$ has a subsequence $\{x_{n''}\}$ such that $f(x_{n''}) \to c$.

□

Definition 2.4. Our regularized solutions will be the minimizers x_α^δ of the Tikhonov-type functionals

$$J_\alpha(x) = \begin{cases} \|F(x) - y^\delta\|^q + \alpha\Psi(x) & \text{if} \quad x \in \mathcal{D}(\Psi) \cap \mathcal{D}(F) \\ +\infty & \text{otherwise} \end{cases}. \tag{4}$$

For non-linear operators the minimizer of (4) will in general not be unique and for fixed y^δ, we denote the set of all minimizers by M_α, i.e.

$$M_\alpha = \{x_\alpha^\delta \ : \ x_\alpha^\delta = \arg\min_x J_\alpha(x)\} \tag{5}$$

We call a solution $x^\dagger$ of equation (1) an Ψ-minimizing solution if

$$\Psi(x^\dagger) = \min \{\Psi(x) : F(x) = y\},$$

and denote the set of all Ψ-minimizing solutions by $\mathcal{L}$. Throughout this paper we assume that $\mathcal{L} \neq \emptyset$.

The remainder of the paper is organized as follows. In Section 3 we will analyze Morozov's discrepancy principle for non-linear operators and general penalty terms fulfilling Condition 2.1. As in the well studied case of classical Tikhonov regularization, we will be able to show that standard conditions on the operator F suffice to guarantee the existence of a regularization parameter fulfilling the discrepancy principle. Section 4 contains the main regularization results where we will show in particular that the parameter $\alpha = \alpha(\delta, y^\delta)$ chosen such that (3) holds, satisfies

$$\alpha(\delta, y^\delta) \to 0 \quad \text{and} \quad \delta^q/\alpha(\delta, y^\delta) \to 0 \quad \text{as} \quad \delta \to 0,$$

and that for suitable penalty terms this yields convergence of the regularized solutions in the topology induced by the penalty term as the noise level goes to zero. Additionally, we will see in Section 5 that the generalized Bregman distance between the regularized solution and a Ψ-minimizing solution goes to zero with the same order as the noise level δ. Finally, we will present a numerical example in Section 6, where the theoretically established results are verified.

3 The Discrepancy Principle

For Morozov's discrepancy principle we have to choose the regularization parameter $\alpha = \alpha(\delta, y^\delta)$ such that for constants $1 < \tau_1 \leq \tau_2$,

$$\tau_1\delta \leq \|F(x_\alpha^\delta) - y^\delta\| \leq \tau_2\delta.$$

To analyze when this is possible, we will be using the following functionals frequently.

Definition 3.1. Let $y^\delta \in Y$ be fixed. For $\alpha \in (0, \infty)$ and $x_\alpha^\delta \in M_\alpha$ we define

$$G(x_\alpha^\delta) = \|F(x_\alpha^\delta) - y^\delta\| \tag{6}$$

$$\Omega\big(x_\alpha^\delta\big) = \Psi(x_\alpha^\delta) \tag{7}$$

$$m(\alpha) = J_\alpha(x_\alpha^\delta). \tag{8}$$

Remark 3.2. Since all minimizers $x_\alpha^\delta \in M_\alpha$ have the same value of $J_\alpha(x_\alpha^\delta)$, the value of $m(\alpha)$ does not depend on the particular choice of $x_\alpha^\delta \in M_\alpha$. This is in general not true, however, for $G(x_\alpha^\delta)$ and $\Omega\big(x_\alpha^\delta\big)$.

We will now prove a statement on the monotonicity of G, Ω, and m with respect to α.

Lemma 3.3. *The functional $\Omega(x_\alpha^\delta)$ is non-increasing with respect to $\alpha \in (0, \infty)$ and the functionals $G(x_\alpha^\delta), m(\alpha)$ are non-decreasing on $(0, \infty)$ for any $q > 0$.*

Proof. Let $0 < \alpha < \beta$ and choose $x_\alpha^\delta \in M_\alpha, x_\beta^\delta \in M_\beta$ arbitrary but fixed, then

$$\begin{aligned} G(x_\alpha^\delta)^q + \alpha\Omega(x_\alpha^\delta) &\leq G(x_\beta^\delta)^q + \alpha\Omega(x_\beta^\delta) \\ G(x_\beta^\delta)^q + \beta\Omega(x_\beta^\delta) &\leq G(x_\alpha^\delta)^q + \beta\Omega(x_\alpha^\delta). \end{aligned}$$

Combining these inequalities we get

$$\frac{1}{\alpha}\Big(G(x_\alpha^\delta)^q - G(x_\beta^\delta)^q\Big) \leq \Omega(x_\beta^\delta) - \Omega(x_\alpha^\delta) \leq \frac{1}{\beta}\Big(G(x_\alpha^\delta)^q - G(x_\beta^\delta)^q\Big).$$

Since $0 < 1/\beta < 1/\alpha$ we may conclude that $G(x_\alpha^\delta)^q - G(x_\beta^\delta)^q \leq 0$ and obtain

$$G(x_\alpha^\delta) \leq G(x_\beta^\delta) \quad \text{and} \quad \Omega(x_\beta^\delta) \leq \Omega(x_\alpha^\delta).$$

Finally, the estimate

$$m(\alpha) = G(x_\alpha^\delta)^q + \alpha\Omega(x_\alpha^\delta) \leq G(x_\beta^\delta)^q + \alpha\Omega(x_\beta^\delta) \leq G(x_\beta^\delta)^q + \beta\Omega(x_\beta^\delta) = m(\beta)$$

completes the proof.

□

Remark 3.4. From the proof of Lemma 3.3 we conclude that whenever $\alpha < \beta$, then

$$\sup_{x_\alpha^\delta \in M_\alpha} G(x_\alpha^\delta) \leq \inf_{x_\beta^\delta \in M_\beta} G(x_\beta^\delta) \quad \text{and} \quad \inf_{x_\alpha^\delta \in M_\alpha} \Omega(x_\alpha^\delta) \geq \sup_{x_\beta^\delta \in M_\beta} \Omega(x_\beta^\delta).$$

Moreover, the values of $G(x_\alpha^\delta)^q$ and $\Omega(x_\alpha^\delta)$ are bounded by $m(\alpha)$, which is again uniformely bounded by

$$m(\alpha) \leq J_\alpha(0) = \|F(0) - y^\delta\|^q < \infty, \tag{9}$$

where the last value is finite because we assumed that $0 \in \mathcal{D}(F) \cap \mathcal{D}(\Psi)$. Due to the monotonicity of G, Ω which was established in the preceding Lemma, jumps in these functions can therefore only occur at countably many values $\{\alpha_j : j \in \mathbb{N}\}$ and with uniformly bounded heights. Indeed for every discontinuity point α of $G(x_\alpha^\delta)$ the open interval $\big(\inf_{x_\alpha^\delta \in M_\alpha} G(x_\alpha^\delta),\ \sup_{x_\alpha^\delta \in M_\alpha} G(x_\alpha^\delta)\big)$ is nonempty and thus contains a rational number. Since these intervals are disjoint for different discontinuities and there are only countably many rationals the statment follows. The same argument works also for Ω. A similar result can be found in [9] for linear operators.

We will now see that the jump heights of $G(x_\alpha^\delta)^q$ and $\Omega(x_\alpha^\delta)$ even cancel each other out, making $m(\alpha)$ continuous. Similar results have been proven in [1, 9] for linear operators and under slightly different assumptions on the penalty term, and in [12] for non-linear operators F and classical Tikhonov regularization, i.e. the specific choice $\Psi(x) = \|x\|^2$.

Proposition 3.5. *Let $\delta > 0$ and y^δ such that $\|y - y^\delta\| \le \delta$ be fixed. Furthermore, let $\{\alpha_n\}$ be a positive sequence with $\alpha_n \to \bar{\alpha} > 0$, and let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence of corresponding minimizers of $J_{\alpha_n}(x)$ with data y^δ. Then there exists a subsequence $\{x_{n'}\}_{n'\in\mathbb{N}}$ and a minimizer $\bar{x}$ of $J_{\bar{\alpha}}(x)$ such that*

$$x_{n'} \rightharpoonup \bar{x}.$$

Moreover, it holds that $m(\alpha_{n'}) \to m(\bar{\alpha})$, where $m(\alpha)$ is defined as in (8).

Proof. As we have seen in Lemma 3.3 $\Omega\big(x_\alpha^\delta\big)$ is non-increasing. The sequence $\{\alpha_n\}$ must be from bounded below away from zero. Thus the sequence $\{\Psi(x_n)\}$ is bounded and by the weak coercivity of Ψ, so is $\{x_n\}$. Therefore we can extract a subsequence $\{x_{n'}\}$ with $x_{n'} \rightharpoonup \bar{x}$. We now show that $\bar{x}$ is indeed a minimizer of $J_{\bar{\alpha}}$. To this end we first note by the weak continuity of F, also $F(x_{n'}) \rightharpoonup F(\bar{x})$ holds. Moreover, the functionals $\|\cdot - y^\delta\|$ and Ψ are weakly lower semi-continuous. Using the minimizing property of the $x_{n'}$ we get

$$\begin{aligned}
\|F(\bar{x}) - y^\delta\|^q + \bar{\alpha}\Psi(\bar{x}) &\le \liminf_{n'\to\infty} \Big(\|F(x_{n'}) - y^\delta\|^q + \alpha_{n'}\Psi(x_{n'})\Big) \\
&\le \limsup_{n'\to\infty} \Big(\|F(x_{n'}) - y^\delta\|^q + \alpha_{n'}\Psi(x_{n'})\Big) \\
&\le \lim_{n'\to\infty} \Big(\|F(x) - y^\delta\|^q + \alpha_{n'}\Psi(x)\Big) \qquad \forall x \in X \\
&= \|F(x) - y^\delta\|^q + \bar{\alpha}\Psi(x) \qquad \forall x \in X,
\end{aligned}$$

which proves that $\bar{x}$ is a minimizer of $J_{\bar{\alpha}}$. The last statement of the Proposition is also an immediate consequence of the preceding chain of inequalities.

□

Corollary 3.6. *The functional $m(\alpha)$ defined in* (8) *is continuous on* $(0, \infty)$.

Proof. Let $\alpha_n \to \bar{\alpha} > 0$ be a positive sequence and denote for short $x_n = x_{\alpha_n}^\delta$. To any subsequence of $\{\alpha_n\}$ we can apply Proposition 3.5 which yields another subsequence such that $m(\alpha_{n'}) \to m(\bar{\alpha})$. Using the convergence principle described in Lemma 2.3 we obtain that $m(\alpha_n) \to m(\bar{\alpha})$.

□

As mentioned earlier, the value of $G(x_\alpha^\delta)$ will in general depend on the choice of $x_\alpha^\delta \in M_\alpha$. Therefore, when viewed as a function of α for fixed $y^\delta \in Y$, G may be multi-valued for some α. Next we prove, however, that G is left and right continuous with respect to $\alpha \in (0, \infty)$ (compare [1, 17] for the linear and [12] for the non-linear classical Tikhonov case). Consequently, G is continuous (with respect to α) whenever it is constant on M_α (cf. Corollary 3.9).

Proposition 3.7. *Let $\{\alpha_n\}$ be a positive, strictly increasing sequence converging to $\bar{\alpha} \in \mathbb{R}^+$, and $\{x_n\}$ be a corresponding sequence of minimizers, $x_n \in M_{\alpha_n}$, then*

$$G(x_n) \to \inf_{x\in M_{\bar{\alpha}}} G(x).$$

Similarly, if $\{\alpha_n\}$ is strictly decreasing with limit $\underline{\alpha} > 0$, then

$$G(x_n) \to \sup_{x\in M_{\underline{\alpha}}} G(x).$$

Proof. We first consider the case where $\alpha_n \uparrow \bar{\alpha}$. From Proposition 3.5 we obtain that $\{x_n\}$ has a subsequence, denoted by $\{x_{n'}\}$, such that $x_{n'} \rightharpoonup \bar{x} \in M_{\bar{\alpha}}$, and since F is weakly continuous, also $F(x_{n'}) \rightharpoonup F(\bar{x})$. It then follows from the lower semi-continuity of $\| \, . - y^\delta\|$ and the monotonicity of G w.r.t. α (cf. Lemma 3.3) that

$$G(\bar{x}) \leq \liminf_{n' \to \infty} G(x_{n'}) \leq \limsup_{n' \to \infty} G(x_{n'}) \leq \inf_{x \in M_{\bar{\alpha}}} G(x) \leq G(\bar{x})$$

where the last inequatlity holds since $\bar{x} \in M_{\bar{\alpha}}$. Together this implies that

$$G(x_{n'}) \to G(\bar{x}) = \inf_{x \in M_{\bar{\alpha}}} G(x).$$

The same reasoning can be applied to any subsequence of the original sequence $\{x_n\}$ (i.e. to any subsequence of $\{G(x_n)\}$) , which shows that $G(x_n) \to \inf_{x \in M_{\bar{\alpha}}} G(x)$ (cf. Lemma 2.3).

In the case where $\{\alpha_n\}$ is a decreasing sequence, we will use the same argument, but applied to Ω. As above we may choose a subsequence $\{x_{n'}\}$ of $\{x_n\}$ such that $x_{n'} \rightharpoonup \underline{x} \in M_{\underline{\alpha}}$. Now, the lower semi-continuity of Ψ and the monotonicity of Ω w.r.t. α (cf. Corollary 3.3) imply that

$$\Omega\big(\underline{x}\big) \leq \liminf_{n' \to \infty} \Omega\big(x_{n'}\big) \leq \liminf_{n' \to \infty} \Omega\big(x_{n'}\big) \leq \inf_{x \in M_{\underline{\alpha}}} \Omega\big(x\big) \leq \Omega\big(\underline{x}\big),$$

which means that $\Omega\big(x_{n'}\big) \to \Omega\big(\underline{x}\big) = \inf_{x \in M_{\underline{\alpha}}} \Omega\big(x\big)$, since we know that $\underline{x} \in M_{\underline{\alpha}}$. Again these arguments can be applied to any subsequence of $\{x_n\}$ (and therefore $\{\Omega\big(x_n\big)\}$), and the convergence principle in Lemma 2.3 yields that

$$\Omega\big(x_n\big) \to \Omega\big(\underline{x}\big) = \inf_{x \in M_{\underline{\alpha}}} \Omega\big(x\big). \tag{10}$$

Using the continuity of $m(\alpha)$ from Corollary 3.6, equation (10) yields

$$\lim_{n \to \infty} G(x_n) = \lim_{n \to \infty} \Big(m(\alpha_n) - \alpha_n \Omega\big(x_n\big) \Big)^{1/q} = \Big(m(\underline{\alpha}) - \underline{\alpha} \Omega\big(\underline{x}\big) \Big)^{1/q} = G(\underline{x}).$$

Moreover, it is an immediate consequence of (10) and $m(\alpha)$ being single valued that for any $x \in M_{\underline{\alpha}}$ we have

$$G(\underline{x})^q = m(\underline{\alpha}) - \underline{\alpha} \Omega\big(\underline{x}\big) \geq m(\underline{\alpha}) - \underline{\alpha} \Omega\big(x\big) = G(x)^q.$$

Therefore

$$\lim_{n \to \infty} G(x_n) = G(\underline{x}) = \sup_{x \in M_{\underline{\alpha}}} G(x),$$

which completes the proof.

□

Remark 3.8. From the proof of Proposition 3.7 it can be seen that to each $\alpha > 0$ we can find $x_1, x_2 \in M_\alpha$ such that

$$\|F(x_1) - y^\delta\| = \inf_{x \in M_\alpha} \|F(x) - y^\delta\|$$
$$\|F(x_2) - y^\delta\| = \sup_{x \in M_\alpha} \|F(x) - y^\delta\|.$$

If $G(x_\alpha^\delta)$ takes the same value for all $x_\alpha^\delta \in M_\alpha$ (e.g. if M_α has only one element), then it follows from Proposition 3.7 that G is continuous. We have thus proven the following corollary.

Corollary 3.9. *If for all $\alpha > 0$ the functional G is constant on M_α, then G is continuous.*

Let us now define the discrepancy principle.

Definition 3.10. For $1 < \tau_1 \le \tau_2$ we choose $\alpha = \alpha(\delta, y^\delta) > 0$ such that

$$\tau_1\delta \le G(x_\alpha^\delta) \le \tau_2\delta \tag{11}$$

holds for some $x_\alpha^\delta \in M_\alpha$.

The following Proposition generalizes a result for classical Tikhonov regularization from [12].

Proposition 3.11. *Assume that $\|F(0) - y^\delta\| > \tau_2\delta$, then we can find $\underline{\alpha}, \bar{\alpha}$ such that*

$$G(x_{\underline{\alpha}}^\delta) < \tau_1\delta \le \tau_2\delta < G(x_{\bar{\alpha}}^\delta).$$

Proof. Let us first consider a sequence $\{\alpha_n\}$ converging to 0 and a corresponding sequence of minimizers $x_n \in M_{\alpha_n}$, then for $x^\dagger \in \mathcal{L}$ (cf. Definition 2.4) we get

$$G(x_n)^q \le m(\alpha_n) \le J_{\alpha_n}(x^\dagger) \le \delta^q + \alpha_n\Psi(x^\dagger) \to \delta^q < \tau_1^q\delta^q.$$

This proves $G(\underline{x}) < \tau_1\delta$ if we choose $\underline{\alpha} = \alpha_N$ for N large enough.

On the other hand, if $\alpha_n \to \infty$ and $\{x_n\}$ as before, then

$$\Omega(x_n) \le \frac{1}{\alpha_n} m(\alpha_n) \le \frac{1}{\alpha_n}\|F(0) - y^\delta\| \to 0 \quad \text{as} \quad n \to \infty.$$

This shows that $\Psi(x_n) \to 0$, which according to Lemma 2.2 implies $x_n \rightharpoonup 0$ and also $F(x_n) \rightharpoonup F(0)$ by the weak continuity of F. Using the lower semi-continuity of the norm we obtain

$$\|F(0) - y^\delta\| \le \liminf \|F(x_n) - y^\delta\|, \tag{12}$$

Together with the assumption $\tau_2\delta < \|F(0) - y^\delta\|$ this yields the existence of $\bar{\alpha}$ such that $\|F(x_{\bar{\alpha}}^\delta) - y^\delta\| = G(\bar{x}) > \tau_2\delta$.

□

Remark 3.12. Taking a sequence $\alpha_n \to \infty$ and keeping in mind that $G(x_n)^q \le m(\alpha_n)$, it is a consequence of (9) and (12) that

$$G(x_n) \to \|F(0) - y^\delta\|.$$

We are now ready to prove that the following condition is sufficient to ensure the existence of a regularization parameter α chosen according to the discrepancy principle in Definition 3.10.

Condition 3.13. Assume that y^δ is chosen such that

$$\|y - y^\delta\| \leq \delta < \tau_2\delta < \|F(0) - y^\delta\|, \tag{13}$$

and that there is no $\alpha > 0$ with minimizers $x_1, x_2 \in M_\alpha$ such that

$$\|F(x_1) - y^\delta\| < \tau_1\delta \leq \tau_2\delta < \|F(x_2) - y^\delta\|.$$

For the following theorem compare [12, Theorem 2.5] where the same result is proven for the special case of classical Tikhonov regularization.

Theorem 3.14. *If Condition 3.13 is fulfilled, then there are* $\alpha = \alpha(\delta, y^\delta)$ *and* $x_\alpha^\delta \in M_{\alpha(\delta,y^\delta)}$ *such that*

$$\tau_1\delta \leq \|F(x_\alpha^\delta) - y^\delta\| \leq \tau_2\delta. \tag{14}$$

Proof. Assume that no α fulfilling (14) exists, and define

$$S := \{\alpha : \ \|F(x_\alpha^\delta) - y^\delta\| < \tau_1\delta \quad \text{for some} \quad x_\alpha^\delta \in M_\alpha\}$$
$$\tilde{S} := \{\alpha : \ \|F(x_\alpha^\delta) - y^\delta\| > \tau_2\delta \quad \text{for some} \quad x_\alpha^\delta \in M_\alpha\}.$$

Note that for $\alpha \in S$ it must actually hold that $\|F(x_\alpha^\delta) - y^\delta\| < \tau_1\delta$ for all $x_\alpha^\delta \in M_\alpha$ since otherwise either (14) would hold or condition 3.13 would be violated. The same way we obtain $\|F(x_\alpha^\delta) - y^\delta\| > \tau_2\delta$ for all $x_\alpha^\delta \in M_\alpha$ whenever $\alpha \in \tilde{S}$. Therefore it must hold that

$$S \cap \tilde{S} = \emptyset \quad \text{and} \quad S \cup \tilde{S} = \mathbb{R}^+.$$

If we set $\bar{\alpha} = \sup S$ then it follows from Proposition 3.11 and the monotonicity of $G(x_\alpha^\delta)$ with respect to α that $0 < \bar{\alpha} < \infty$, and therefore $\bar{\alpha}$ must belong to either S or $\tilde{S}$. We consider these two cases separately.

If $\bar{\alpha} \in S$ then we choose $\alpha_n \downarrow \bar{\alpha}$ and $x_n \in M_{\alpha_n}$. Since all α_n must belong to $\tilde{S}$ it follows from Proposition 3.7 that

$$\tau_2\delta \leq \lim_{n\to\infty} \|F(x_n) - y^\delta\| = \sup_{x\in M_{\bar{\alpha}}} \|F(x) - y^\delta\| < \tau_1\delta,$$

where the strict inequality holds because of Remark 3.8. This is a contradiction since we chose $\tau_1 \leq \tau_2$.

Similarly, if $\bar{\alpha} \in \tilde{S}$ then we choose $S \ni \alpha_n \uparrow \bar{\alpha}$ and x_n as before and again obtain a contradiction:

$$\tau_2\delta < \inf_{x\in M_{\bar{\alpha}}} \|F(x) - y^\delta\| = \lim_{n\to\infty} \|F(x_n) - y^\delta\| \leq \tau_1\delta.$$

□

4 Regularization properties

Additionally to the general assumptions on the operator $F(x)$ and the functional $\Psi(x)$, which were taken in Section 2, and to Condition 3.13, which ensures the existence of a regularization parameter $\alpha = \alpha(\delta, y^\delta)$ chosen according to the discrepancy principle, we will need the following condition.

Condition 4.1. For all $x^\dagger \in \mathcal{L}$ (cf. Definition 2.4) we assume that

$$\liminf_{t\to 0^+} \frac{\|F((1-t)x^\dagger) - y\|^q}{t} = 0. \tag{15}$$

We now show that differentiability of F is sufficient for Condition 4.1 to hold.

Lemma 4.2. *Let X be a Hilbert space. If $F(x)$ is Gâteaux differentiable in the directions $x^\dagger \in \mathcal{L}$ and the derivatives are bounded in a neigbourhood of $x^\dagger$, then Condition 4.1 is satisfied.*

Proof. It holds for any y (as long as it admits a Ψ-minimizing solution $x^\dagger$) that

$$\frac{d}{dt}\|F((1-t)x^\dagger) - y\|^q = q\,\|F((1-t)x^\dagger) - y\|^{q-2}\cdot \\ \langle F'((1-t)x^\dagger)\cdot x^\dagger, F((1-t)x^\dagger) - y\rangle$$

and due to the boundedness of $F'((1-t)x^\dagger)$ near $t=0$ this can be estimated by

$$\left|\frac{d}{dt}\|F((1-t)x^\dagger) - y\|^q\right| \le q\,\|F((1-t)x^\dagger) - y\|^{q-1}\,\|F'(1-t)x^\dagger)\cdot x^\dagger\| \overset{t\to 0^+}{\longrightarrow} 0,$$

since $\|F(x^\dagger) - y\| = 0$ by assumption. Together this yields

$$\lim_{t\to 0^+}\frac{\|F((1-t)x^\dagger) - y\|^q}{t} = \frac{d}{dt}\|F((1-t)x^\dagger) - y\|^q\Big|_{t=0} = 0.$$

□

Lemma 4.3. *Assume that Condition 4.1 is satisfied and that there exist $\alpha > 0$ and a solution $x^\star$ of $F(x) = y$ such that*

$$x^\star = \arg\min_x \Big\{\|F(x) - y\|^q + \alpha\Psi(x)\Big\},$$

then $x^\star = 0$.

Proof. Since $x^\star$ is a minimizer of J_α with exact data y, we obtain for all $x^\dagger \in \mathcal{L}$

$$\alpha\Psi(x^\star) \le \alpha\Psi(x^\dagger),$$

and this implies that $x^\star \in \mathcal{L}$. Due to the convexity of Ψ and to the fact that $0, x^\star \in \mathcal{D}(\Psi)$ with $\Psi(0) = 0$, it holds for $t \in [0,1)$ that

$$\Psi((1-t)x^\star) = \Psi((1-t)x^\star + t\cdot 0) \le (1-t)\Psi(x^\star) + t\ \Psi(0) = (1-t)\Psi(x^\star).$$

As $x^\star \in M_\alpha$, we thus get

$$\alpha\Psi(x^\star) = J_\alpha(x^\star) \le J_\alpha((1-t)x^\star) \le \|F((1-t)x^\star) - y\|^q + \alpha(1-t)\Psi(x^\star)$$

and therefore

$$\alpha\ t\Psi(x^\star) \le \|F((1-t)x^\star) - y\|^q.$$

Altogether this implies

$$0 \le \alpha\Psi(x^\star) \le \liminf_{t\to 0^+} \frac{\|F((1-t)x^\dagger) - y\|^q}{t} = 0,$$

which yields $\Psi(x^\star) = 0$. But according to Condtion 2.1 (i) this only holds if $x^\star = 0$.

□

Remark 4.4. To illustrate that Lemma 4.3 does not hold for arbitrary F and y, we give a continuous one dimensional counter example. Let

$$F(x) = 1 + \sqrt{|1-x|}, \qquad x \in \mathbb{R},$$

then the derivative of F is unbounded in $x = 1$, i.e. Lemma 4.2 cannot be applied. For the choices $y = 1, q = 2$ and $\Psi(x) = |x|$ the unique solution of $F(x) = y$ is $x^\dagger = 1$ and it holds that

$$\lim_{t\to 0^+} \frac{\left|F((1-t)x^\dagger) - y\right|^2}{t} = \lim_{t\to 0^+} \frac{|t|}{t} = 1.$$

Therefore Condition 4.1 is violated and indeed for $\alpha = 1$ we find that

$$J_1(x) = (F(x) - y)^2 + \Psi(x) = |1-x| + |x| \ge 1 = J_1(x^\dagger) \qquad \forall x \in \mathbb{R},$$

which shows that the Ψ-minimizing solution $x^\dagger \neq 0$ is also a minimizer of the Tikhonov functional for $\alpha > 0$. The same example also works in the classical Tikhonov case, choosing $\Psi(x) = x^2$. Note that for a choice $y > 1$, Condition 4.1 is always satisfied, so that (15) truly depends on the exact data y. This is no longer an issue, however, if the Gâteux derivative of $F(x)$ is locally bounded (cf. Lemma 4.2).

We now state the main theorem of this paper.

Theorem 4.5. *Let F, Ψ satisfy the Conditions 2.1, 4.1. Moreover, assume that data y^δ, $\delta \in (0, \delta^*)$, are given such that Condition 3.13 holds, where $\delta^* > 0$ is an arbitrary upper bound. Then the regularization parameter $\alpha = \alpha(\delta, y^\delta)$ obtained from Morozov's discrepancy principle (see Definition 3.10) satisfies*

$$\alpha(\delta, y^\delta) \to 0 \qquad \textit{and} \qquad \frac{\delta^q}{\alpha(\delta, y^\delta)} \to 0 \qquad \textit{as} \quad \delta \to 0.$$

Proof. Let $\delta_n \to 0, \{y^{\delta_n}\}$ s.t. $\|y - y^{\delta_n}\| \le \delta_n$ and $\alpha_n = \alpha(\delta_n, y^{\delta_n})$ be chosen according to the discrepancy principle. As a shorthand we write $x_n = x_{\alpha_n}^{\delta_n}$.

We will first show that $\alpha_n \to 0$. Assume that there is a subsequence of $\{\alpha_n\}$, denoted again by $\{\alpha_n\}$, and a constant $\underline{\alpha}$ such that $0 < \underline{\alpha} \le \alpha_n\ \forall n$. If we denote the minimizers of $J_{\underline{\alpha}}$ with data y^{δ_n} by

$$\underline{x}_n = \arg\min_x \left\{ \|F(x) - y^{\delta_n}\|^q - \underline{\alpha}\Psi(x) \right\}$$

we obtain by using Lemma 3.3 and (11) that

$$\|F(\underline{x}_n) - y\| \le \|F(\underline{x}_n) - y^{\delta_n}\| + \delta_n \le \|F(x_n) - y^{\delta_n}\| + \delta_n \le (\tau_2 + 1)\delta_n \to 0. \tag{16}$$

Moreover, since $\delta_n \to 0$ there is a constant $C > 0$ such that

$$\|F(\underline{x}_n) - y^{\delta_n}\|^q + \underline{\alpha}\Psi(\underline{x}_n) \le \delta_n^q + \underline{\alpha}\Psi(x^\dagger) \le C$$

holds for all n. By the coercivity of Ψ, the sequence $\{\underline{x}_n\}$ is bounded and we can extract a subsequence, again denoted by $\{\underline{x}_n\}$, that converges weakly to some $\underline{x}$. Due to (16) and the weak lower semicontinuity of $\|F(\cdot) - y\|$ we have

$$\|F(\underline{x}) - y\| \le \liminf_n \|F(\underline{x}_n) - y\| = 0,$$

and thus $F(\underline{x}) = y$. On the other hand, because of the weak lower semi-continuity of Ψ, it holds that

$$\begin{aligned}
\|F(\underline{x}) - y\| + \underline{\alpha}\Psi(\underline{x}) &\le \liminf_n \Big(\|F(\underline{x}_n) - y^{\delta_n}\| + \underline{\alpha}\Psi(\underline{x}_n)\Big) \\
&\le \liminf_n \Big(\|F(x) - y^{\delta_n}\| + \underline{\alpha}\Psi(x)\Big) \qquad \forall x \in X \\
&= \|F(x) - y\| + \underline{\alpha}\Psi(x) \qquad \forall x \in X,
\end{aligned}$$

which shows that $\underline{x}$ is a minimizer of $J_{\underline{\alpha}}$ with exact data y. Therefore $\underline{x}$ satisfies the assumptions of Lemma 4.3 and it follows that $\underline{x} = 0$, which in turn means that $y = F(0)$. This violates (13) in Condition 3.13, and we have reached a contradiction.

To see that $\delta_n^q/\alpha_n \to 0$, we observe that for any $x^\dagger \in \mathcal{L}$ (see Definition 2.4)

$$\begin{aligned}
\tau_1^q \delta_n^q + \alpha_n \Psi(x_n) &\le \|F(x_n) - y^\delta\|^q + \alpha_n \Psi(x_n) \\
&\le \|F(x^\dagger) - y^{\delta_n}\|^q + \alpha_n \Psi(x^\dagger) \\
&\le \delta_n^q + \alpha_n \Psi(x^\dagger),
\end{aligned}$$

since x_n is a minimizer of $J_{\alpha_n}(x)$ and it follows that

$$0 < (\tau_1^q - 1)\frac{\delta_n^q}{\alpha_n} \le \Psi(x^\dagger) - \Psi(x_n). \tag{17}$$

As an immediate consequence, the sequence $\{\Psi(x_n)\}$ is bounded. Since Ψ is coercive it follows that also $\{x_n\}$ is bounded. We can thus extract a subsequence, denoted again as $\{x_n\}$, which converges weakly to some $\bar{x} \in X$. It holds that

$$\|F(x_n) - y\| \le \|F(x_n) - y^{\delta_n}\| + \delta_n \le (\tau_2 + 1)\delta_n \to 0,$$

and by the weak continuity of F and the uniqueness of the weak limit we obtain that $F(\bar{x}) = y$. As Ψ is weakly lower semi-continuous, we get by using (17)

$$\Psi(\bar{x}) \le \liminf_n \Psi(x_n) \le \limsup_n \Psi(x_n) \le \Psi(x^\dagger), \tag{18}$$

which altogether shows that actually equality must hold in (18), since otherwise the Ψ-minimizing property of $x^\dagger$ would be violated. This reasoning can be applied to any subsequence of the original sequence $\{x_n\}$, where the weak limits may vary, but we always obtain that for some subsequence $\{x_{n'}\}$

$$\lim_{n'} \Psi(x_{n'}) = \Psi(x^\dagger)$$

holds. Using the convergence principle described in Lemma 2.3, the above statement holds for the whole sequence $\{\Psi(x_n)\}$ and thus from (17) it follows that $\delta_n^q/\alpha_n \to 0$ and the proof is complete.

□

Remark 4.6. In the proof of Theorem 4.5 we have used that $\|F(0) - y\| > 0$, which is an immediate consequence of (13). On the other hand, whenever $\|F(0) - y\| > 0$ we can choose

$$0 < \delta^* \leq \frac{1}{\tau_2 + 1} \|F(0) - y\|$$

and for all $0 < \delta < \delta^*$ and y^δ satisfying $\|y - y^\delta\| \leq \delta$ we obtain

$$\|F(0) - y^\delta\| \geq \|F(0) - y\| - \|y - y^\delta\| \geq \|F(0) - y\| - \delta > \tau_2 \delta,$$

which is (13). Therefore (13) can be fulfilled for all δ smaller than some $\delta^* > 0$, whenever $y \neq F(0)$.

We will now see that for certain penalty terms Theorem 4.5 yields convergence of (a subsequence of) the regularized solutions with respect to the topolgy induced by Ψ.

Condition 4.7. Let $\{x_n\} \subset X$ such that $x_n \rightharpoonup \bar{x} \in X$ and $\Psi(x_n) \to \Psi(\bar{x}) < \infty$, then

$$\Psi(x_n - \bar{x}) \to 0.$$

Remark 4.8. It has been shown in [7, Lemma 2] that choosing weighted ℓ_p-norms of the coefficients with respect to some frame $\{\phi_\lambda\}_{\lambda \in \Lambda} \subset X$ as the penalty term, ie.

$$\Psi_{p,w}(x) := \|\mathbf{x}\|_{\mathbf{w},p} = \Big(\sum_{\lambda \in \Lambda} w_\lambda \ |\langle x, \phi_\lambda \rangle|^p \Big)^{1/p}, \qquad 1 \leq p \leq 2, \tag{19}$$

where $0 < w_{\min} \leq w_\lambda$, satisfies Condition 4.7. Therefore the same trivially holds for $\Psi_{p,w}(x)^p$. Note that these choices also fulfills all the assumptions in Condition 2.1.

The proof of the following corollary is carried out along similar lines as [7, Proposition 7], where the authors are concerned with the concrete setting mentioned in Remark 4.8. In this special case the same result has first been proven in [11] for $q = 2$.

Corollary 4.9. *Let $\delta_n \to 0$ and F, Ψ, y^{δ_n} be such that Condition 4.7 and the assumptions of Theorem 4.5 are fulfilled. Furthermore, let $\alpha_n = \alpha(\delta_n, y^{\delta_n})$ be chosen according to the discrepancy principle in Definition 3.10 and $x_n \in M_{\alpha_n}$ (cf. Definition 2.4), then the sequence $\{x_n\}$ has a subsequence $\{x_{n'}\}$ such that*

$$\Psi(x_{n'} - x^\dagger) \to 0 \qquad \text{as} \quad n' \to \infty,$$

where $x^\dagger \in \mathcal{L}$. Moreover, if the solution $x^\dagger \in \mathcal{L}$ is unique, then we have

$$\Psi(x_n - x^\dagger) \to 0 \qquad \text{as} \quad n \to \infty.$$

Proof. For $x^\dagger \in \mathcal{L}$ we obtain using the minizing property of x_n that

$$\Psi(x_n) \leq \delta_n^q / \alpha_n + \Psi(x^\dagger). \tag{20}$$

Accoring to Theorem 4.5 the first term goes to zero, which yields that $\{\Psi(x_n)\}$ is bounded. Since Ψ is assumed to be weakly coercive the same holds for $\{x_n\}$ and we can extract a weakly convergent subsequence $x_{n'} \rightharpoonup \bar{x}$. Due to the weak lower semicontinuity of $\|F(\cdot) - y\|$ and Ψ we obtain using (11) and (13)

$$\|F(\bar{x}) - y\| \leq \liminf_{n'} \left\{ \|F(x_{n'}) - y^{\delta_n}\| + \|y^{\delta_n} - y\| \right\} \leq \lim_{n'} (\tau_2 + 1)\delta_{n'} = 0,$$

and using (20)

$$\Psi(\bar{x}) \leq \liminf_{n'} \Psi(x_n) \leq \limsup_{n'} \Psi(x_n) \leq \limsup_{n'} \left\{ \delta_{n'}^q / \alpha_{n'} + \Psi(x^\dagger) \right\} = \Psi(x^\dagger).$$

Hence $\bar{x} \in \mathcal{L}$ and $\Psi(x_{n'}) \to \Psi(\bar{x})$ and since Ψ was assumed to satisfy Condition 4.7 this yields $\Psi(x_{n'} - \bar{x}) \to 0$ as $n' \to \infty$.

If $x^\dagger \in \mathcal{L}$ is unique, the convergence of the original sequence follows since the same reasoning can be applied to any subsequence of $\{x_n\}$.

□

5 Convergence rates

Our quantitative estimates on the distance between the regularized solutions and a Ψ-minimizing solution $x^\dagger$ will be given with respect to the generalized Bregman distance, which is defined as follows.

Definition 5.1. Let $\partial\Psi(x)$ denote the subgradient of Ψ at $x \in X$. The generalized Bregman distance with respect to Ψ of two elements $x, z \in X$ is defined as

$$D_\Psi(x, z) := \{D_\Psi^\xi(x, z) : \xi \in \partial\Psi(z) \neq \emptyset\},$$

where

$$D_\Psi^\xi(x, z) := \Psi(x) - \Psi(z) - \langle \xi, x - z \rangle$$

denotes the Bregman distance with respect to Ψ and $\xi \in \partial\Psi(z)$. We remark that throughout this section $\langle \cdot, \cdot \rangle$ denotes the dual pairing in X^*, X or Y^*, Y and not the inner product on a Hilbert space. Moreover $\|\cdot\|_{Y^*}$ denotes the norm on Y^* and, in accordance with our previous notations, we write $\|\cdot\|$ for the norms in the Banach spaces X and Y.

Convergence rates with respect to Bregman distances for Tikhonov-type functionals with convex penalty terms have first been proven by Burger and Osher [3], who focused mainly on the case of linear operators, but also proposed a non-linear generalization of their results, and by Resmerita and Scherzer in [15]. The following non-linearity and source conditions were introduced in the respective works.

Condition 5.2. Let $x^\dagger$ be an arbitrary but fixed Ψ-minimizing solution of $F(x) = y$. Assume that the operator $F : X \to Y$ is Gâteaux differentiable and that there is $w \in Y^*$ such that

$$\xi := F'(x^\dagger)^* w \in \partial\Psi(x^\dagger). \tag{21}$$

Moreover, assume that one of the two following non-linearity conditions holds:

(i) There is $c > 0$ such that for all $x, z \in X$ it holds that

$$\langle w, F(x) - F(z) - F'(z)(x - z)\rangle \leq c \, \|w\|_{Y^*} \, \|F(x) - F(z)\|. \tag{22}$$

(ii) There are $\rho > 0, c > 0$ such that for all $x \in \mathcal{D}(F) \cap \mathcal{B}_\rho(x^\dagger)$,

$$\|F(x) - F(x^\dagger) - F'(x^\dagger)(x - x^\dagger)\| \leq c \, D_\Psi^\xi(x, x^\dagger), \tag{23}$$

and it holds that

$$c \, \|w\|_{Y^*} \leq 1. \tag{24}$$

Using the above condition we are now ready to prove the same convergence rates for Morozov's discrepancy principle which were established in [3, 15] for a-priori parameter choice rules. In the case of linear operators a similar result has been shown in [1].

Theorem 5.3. *Let the operator F and the penalty term Ψ be such that Conditions 2.1 and 5.2 hold. For all $0 < \delta < \delta^*$ assume that the data y^δ fulfill Condition 3.13, and choose $\alpha = \alpha(\delta, y^\delta)$ according to the discrepancy principle in Definition 3.10. Then*

$$\|F(x_\alpha^\delta) - F(x^\dagger)\| = \mathcal{O}(\delta), \qquad D_\Psi^\xi(x_\alpha^\delta, x^\dagger) = \mathcal{O}(\delta). \tag{25}$$

Proof. It is an immediate consequence of (11) and (13) that

$$\|F(x_\alpha^\delta) - y\| \leq \|F(x_\alpha^\delta) - y^\delta\| + \|y - y^\delta\| \leq (\tau_2 + 1)\delta, \tag{26}$$

which proves the first part of (25). In order to show the second part, as in the steps leading to (17) in the proof of Theorem 4.5, we obtain

$$\tau_1^q \delta^q + \alpha\Psi(x_\alpha^\delta) \leq \|F(x_\alpha^\delta) - y^\delta\| + \alpha\Psi(x_\alpha^\delta) \leq \delta^q + \alpha\Psi(x^\dagger),$$

which shows that $\Psi(x_\alpha^\delta) \leq \Psi(x^\dagger)$. Assume now that Condition 5.2 (i) holds, then using (22) we get that

$$\begin{aligned}
D_\Psi^\xi(x_\alpha^\delta, x^\dagger) &\leq \Psi(x_\alpha^\delta) - \Psi(x^\dagger) - \langle F'(x^\dagger)^* w, x_\alpha^\delta - x^\dagger\rangle \\
&\leq -\langle w, F'(x^\dagger)(x_\alpha^\delta - x^\dagger)\rangle \\
&\leq c \, \|w\|_{Y^*} \, \|F(x_\alpha^\delta) - y\| + \left|\langle w, F(x_\alpha^\delta) - y\rangle\right| \\
&\leq (c+1)\tau_2 \, \|w\|_{Y^*} \delta.
\end{aligned}$$

If on the other hand Condition 5.2 (ii) is satisfied, then using (23) and (26) it follows

$$\begin{aligned}
D_\Psi^\xi(x_\alpha^\delta, x^\dagger) &\leq \Psi(x_\alpha^\delta) - \Psi(x^\dagger) - \langle F'(x^\dagger)^* w, x_\alpha^\delta - x^\dagger\rangle \\
&\leq \left|\langle w, F'(x^\dagger)(x_\alpha^\delta - x^\dagger)\rangle\right| \\
&\leq \|w\|_{Y^*}(c \, D_\Psi^\xi(x_\alpha^\delta, x^\dagger) + \|F(x_\alpha^\delta) - y\|) \\
&\leq c \, \|w\|_{Y^*} D_\Psi^\xi(x_\alpha^\delta, x^\dagger) + (\tau_2 + 1) \, \|w\|_{Y^*} \delta.
\end{aligned}$$

Due to (24) this yields

$$D_\Psi^\xi(x_\alpha^\delta, x^\dagger) \leq \frac{\tau_2 + 1}{1 - c \, \|w\|_{Y^*}} \, \|w\|_{Y^*} \delta.$$

□

6 A numerical example

To illustrate the theoretical results of the previous sections, we will analyze a specific example that meets the imposed conditions, namely the autoconvolution operator over a finite interval.

Let $X = Y = L^2[0,1]$, and define for $f \in \mathcal{D}(F) \subset X$ and $s \in [0,1]$ the operator $F : \mathcal{D}(F) \subset X \to Y$ through

$$F(f)(s) := (f * f)(s) = \int_0^s f(s-t)f(t)dt, \tag{27}$$

which has been studied in some detail in [6]. The authors showed that for the choice

$$\mathcal{D}(F) = D^+ := \{f \in L^2[0,1] : f(t) \geq 0 \quad \text{a.e. in} \quad [0,1]\}$$

the autoconvolution operator is weakly continuous and, since D^+ is weakly closed in $L^2[0,1]$, also weakly sequentially closed.

The Fréchet derivative of F at the point $f \in X$ is given by the bounded, linear operator $F'(f) : X \to X$ defined as

$$[F'(f)h](s) = 2\int_0^s f(s-t)h(t)dt \qquad 0 \leq s \leq 1.$$

Indeed, we have

$$F(f+h) - F(f) - F'(f)h = F(h)$$

and therefore

$$\|F(f+h) - F(f) - F'(f)h\| = \|F(h)\|_X \leq \|h\|^2,$$

where $\|\cdot\|$ denotes the Hilbert space norm on $L^2[0,1]$, and the last inequality holds since for all $f, h \in X$

$$\|f * h\| \leq \|f\| \, \|h\|. \tag{28}$$

For a proof see [6, Theorem 2, Lemma 4]. Moreover, $F'(\cdot)$ is linear and due to (28) holds

$$\|F'(f-g)\| = \sup_{\|h\|=1} \|F'(f-g)h\| = \sup_{\|h\|=1} \|2(f-g) * h\| \leq 2\,\|f-g\|,$$

which shows that F' is Lipschitz continuous.

In this example we will be reconstructing a true solution that is known to be sparse w.r.t. the Haar wavelets, which form an orthonormal basis of the Hilbert space X. We therefore define

$$\varphi(t) = \begin{cases} 1 & \text{if} \quad 0 \leq t < 1/2 \\ -1 & \text{if} \quad 1/2 \leq t < 1, \end{cases}$$

$$\psi(t) = 1 \qquad 0 \leq t < 1$$

and for $j \in \mathbb{N}_0$ and $k \in \{0, \ldots, 2^j - 1\}$

$$\varphi_{j,k}(t) = 2^{j/2}\varphi(2^j t - k).$$

The coefficient vector of $f \in X$ will be denoted by $x = \{x_\lambda\}_{\lambda\in\Lambda}$, where

$$\Lambda = \{1\} \cup \{(j,k) : j \in \mathbb{N}_0, 0 \le k \le 2^j - 1\},$$
$$\varphi_\lambda = \begin{cases} \psi & \text{if} \quad \lambda = 1 \\ \varphi_{j,k} & \text{if} \quad \lambda = (j,k) \end{cases} \qquad \text{and}$$
$$x_\lambda = \langle f, \varphi_\lambda \rangle \qquad \forall \lambda \in \Lambda$$

In this setting the operator F can be expressed as follows,

$$\begin{aligned} F(f)(s) &= F(\sum_{\lambda\in\Lambda} x_\lambda \varphi_\lambda)(s) \\ &= \int_0^s \sum_{\lambda\in\Lambda} x_\lambda \varphi_\lambda(s-t) \sum_{\mu\in\Lambda} x_\mu \varphi_\mu(t) dt \\ &= \sum_{\lambda,\mu\in\Lambda} x_\lambda x_\mu \ \varphi_\lambda * \varphi_\mu(s) \end{aligned}$$

Since $F(f) \in Y = X$ we can also represent the image of f with respect to the Haar wavelet basis and obtain

$$y_\eta = \langle F(f), \varphi_\eta \rangle = \sum_{\lambda,\mu\in\Lambda} x_\lambda x_\mu \ \langle \varphi_\lambda * \varphi_\mu, \varphi_\eta \rangle = x^T K_\eta \ x, \tag{29}$$

where

$$K_\eta := \Big\{ \langle \varphi_\lambda * \varphi_\mu, \varphi_\eta \rangle \Big\}_{\lambda,\mu\in\Lambda} \qquad \forall \eta \in \Lambda. \tag{30}$$

The above representation allows us to reformulate the problem in the coefficient space. Since there is a 1-1 correspondence between each element of the Hilbert space X and its coefficient vector w.r.t. the orthonormal wavelet basis, we denote the transformed operator again by $F : \ell_2 \to \ell_2$, with

$$F(x) = \{y_\eta\}_{\eta\in\Lambda}, \quad y_\eta := x^T K_\eta \ x. \tag{31}$$

For the iterative computation of the regularized solutions in (32) we will need the Fréchet derivative of F to be evaluated directly through the coefficients of an element in the Hilbert space. We will see now, how it can indeed be expressed through the matrizes K_η as defined in (30) and the coeffictient vectors x, z of f, h, respectively. To this end we first take a closer look at $F'(f)^* h$ for $f, h \in X$. It is

$$\begin{aligned} \langle F'(f)v, h \rangle &= \int_0^1 2 \int_0^s f(s-t) v(t) dt \ h(s) ds \\ &= \int_0^1 \int_t^1 2 f(s-t) h(s) ds \ v(t) dt \\ &= \langle v, F'(f) * h \rangle \end{aligned}$$

Writing $\tilde{h}(t) = h(1-t)$ for any $h \in X$ we get

$$\begin{aligned} [F'(f)^* h](t) &= 2 \int_t^1 f(s-t) \ h(s) \ ds \\ &= 2 \int_0^{1-t} f(s) \ \tilde{h}((1-t)-s) \ ds \\ &= 2 \ (f * \tilde{h}) \ (1-t). \end{aligned}$$

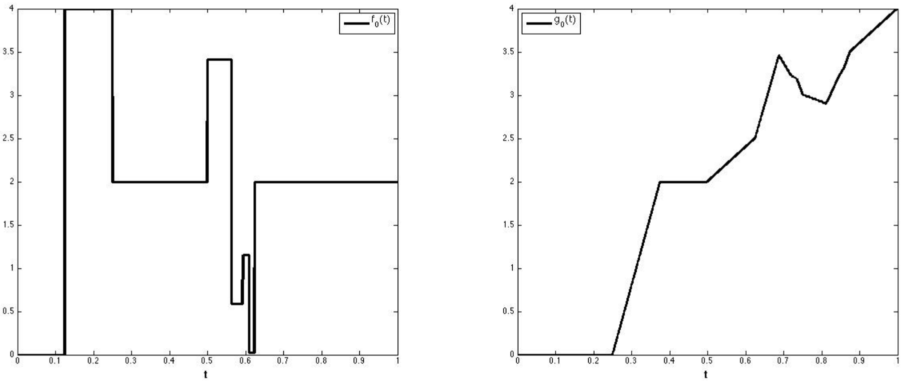

Figure 1: The Ψ-minimizing solution $f^\dagger$ (left), and the true data $g = F(f^\dagger)$ (right).

The coefficients $\tilde{z}$ of $\tilde{h}(t)$ can be computed according to the formulae

$$\tilde{z}_1 = \langle \tilde{h}, \psi \rangle = \langle h, \tilde{\psi} \rangle = \langle h, \psi \rangle = z_1,$$
$$\tilde{z}_{(j,k)} = \langle \tilde{h}, \varphi_{(j,k)} \rangle = \langle h, \tilde{\varphi}_{(j,k)} \rangle = \langle h, -\varphi_{(j,k)'} \rangle = -z_{(j,k)'},$$

where $(j,k)' := (j, 2^j - 1 - k)$. Adapting the steps that were used in (29) we obtain

$$\langle F'(f)^* h, \varphi_\eta \rangle = 2\ (x^T K_\eta\ \tilde{z})\tilde{}.$$

Due to these formulations there will be no need to use a discretization in the spacial variable during the reconstruction. Instead it suffices to discretize by choosing a maximal level $J \in \mathbb{N}$ and considering only those wavelet coefficients where $j \leq J$.

In order to reconstruct a sparse solutions, we choose the penalty term $\Psi(f)$ as

$$\Psi(f) = \sum_{\lambda \in \Lambda} |\langle f, \varphi_\lambda \rangle| = \|x\|_1.$$

Here as before, $x \in \ell_2$ denotes the coefficient vector associated to $f \in X$. The regularized solutions f_α^δ are the minimizers of the Tikhonov-type functionals

$$J_\alpha(f) = \|F(f) - g^\delta\|^2 + \alpha \Psi(f).$$

By the virtue of (31) we can alternatively consider the transformed functional

$$J_\alpha(x) = \|F(x) - y^\delta\|^2 + \alpha \|x\|_1,$$

where from now on y, y^δ denote the coefficient vectors of g, g^δ, respectively. In this notation it holds that $\|g - g^\delta\| = \|y - y^\delta\|$ and therefore the condition for the noise level can be equivalently stated in the ℓ_2 framework. It simply reads

$$\|y - y^\delta\| \leq \delta.$$

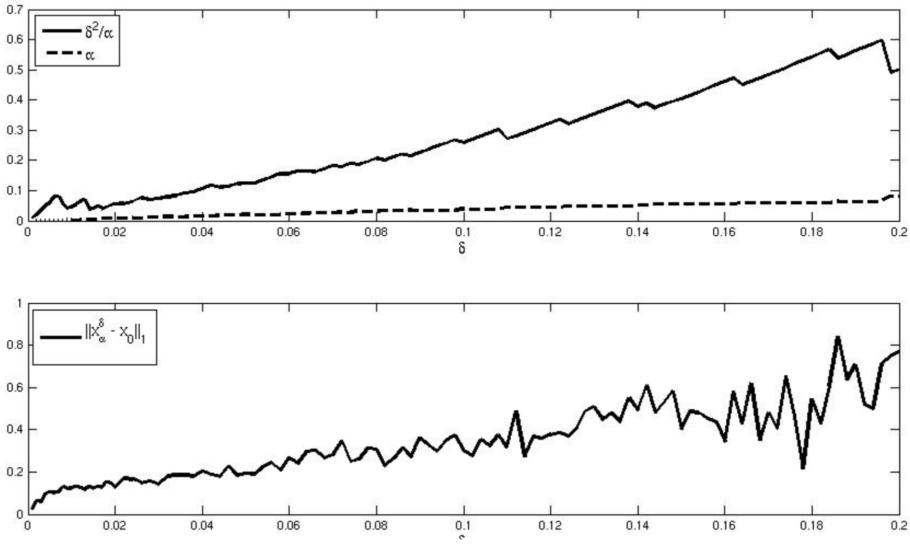

Figure 2: Top: Graph of $\delta \mapsto \delta^2/\alpha(\delta, y^\delta)$ (solid) and of $\delta \mapsto \alpha(\delta, y^\delta)$ (dashed). Bottom: Graph of $\delta \mapsto \|x^\delta_{\alpha(\delta,y^\delta)} - x^\dagger\|_1$. The noise was chosen uniformly random at each step.

To compute the regularized solutions

$$x^\delta_\alpha = \arg\min_{x \in \ell_2} J_\alpha(x)$$

we will use the iterative soft-shrinkage algorithm for non-linear inverse problems from [11], which is based on the surrogate functional approach introduced in [4]. We denote the soft-shrinkage operator with threshold $\alpha > 0$ by S_α, i.e. for $x = \{x_\lambda\}_{\lambda \in \Lambda} \in \ell_2$

$$(S_\alpha(x))_\lambda = \begin{cases} x_\lambda - \alpha & \text{if} \quad x_\lambda > \alpha \\ x_\lambda + \alpha & \text{if} \quad x_\lambda < -\alpha \\ 0 & \text{if} \quad |x_\lambda| \le \alpha. \end{cases}$$

It has been shown in [11] that for arbitrary $x^0 \in \ell_2$ and with

$$\begin{aligned} x^{n,0} &= x_n, \\ x^{n,k+1} &= S_{\alpha/2C}\big(x^{n,0} + \frac{1}{C} F'(x^{n,k})(y^\delta - F(x^{n,0}))\big), \\ x^{n+1} &= \lim_{k\to\infty} x^{n,k} \end{aligned} \tag{32}$$

the resulting sequence x_n converges at least to a critical point of J_α. Choosing the starting value x^0 reasonably close to the true solution and C large enough, we observed that the iteration even converges to a minimizer.

Moreover, choosing the right hand side g of our inverse problem

$$F(f) = g \tag{33}$$

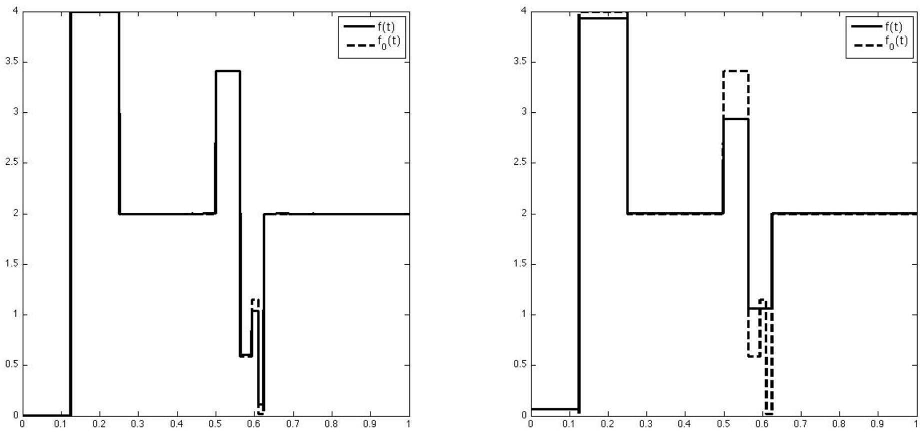

Figure 3: Regularized solutions with noise level $\delta = 0.001$ (left) and $\delta = 0.1$ (right) as solid lines together with the Ψ-minimizing solution $f^\dagger$ (dashed).

such that

$$g \in R_\varepsilon^+ := \{g \in C[0,1] \;:\; g \geq 0,\ \varepsilon = \max\{s : g(\xi) = 0 \ \forall \xi \in [0,s]\}\}$$

for some $\varepsilon > 0$, it has been shown in [6] that any $f \in D(F)$ fulfilling (33) possesses the form

$$f(t) = \begin{cases} 0 & \text{a.e. in} \quad t \in [0, \varepsilon/2] \\ \text{uniquely determined by } g & \text{a.e. in} \quad t \in [\varepsilon/2, 1-\varepsilon/2] \\ \text{arbitrarily non-negative} & \text{in} \quad t \in [1-\varepsilon/2, 1] \end{cases} \tag{34}$$

For our test, we have chosen $g \in R_\varepsilon^+$ with $\varepsilon = 1/4$ as shown in Figure 1. Due to the representation of the possible solutions in (34), we know that they are uniquely determined on $[0, 7/8]$, but differ on the complementary set. Also, the right hand side was chosen such that it actually permits a sparse solution, which is the Ψ-minimizing solution $f^\dagger$ seen in Figure 1. Note that in this example the solution $f^\dagger$ is different from the minimum norm solution, which would have the constant value zero on the set where the solutions may differ.

Since we have seen that the operator F and the penatly term Ψ fulfill all the conditions needed to apply Theorem 4.5 and Corollary 4.9 (a parameter $\alpha(\delta, y^\delta)$ could indeed be found for all choices of the noisy data, ensuring that the discrepancy principle is applicable), and since moreover the Ψ-minimizing solution $x^\dagger$ is unique, it can be expected that for $\delta \to 0$ we observe

$$\alpha(\delta, y^\delta) \to 0, \qquad \frac{\delta^2}{\alpha(\delta, y^\delta)} \to 0, \quad \text{and} \quad \Psi(x_\alpha^\delta - x^\dagger) = \|x_\alpha^\delta - x^\dagger\|_1 \to 0.$$

Our numerical experiments confirm these expectations and for a sample case with $J = 5, \delta \in (0, 0.2]$, and noise chosen uniformly random such that $\|y - y^\delta\| = \delta$ at each step, we show the results in Figure 2. The reconstructed solutions with noise level $\delta = 0.001$ and $\delta = 0.1$ can be seen in Figure 3, where the regularization parameters $\alpha = 0.0001$ and $\alpha = 0.03872$, respectively, were chosen according to Morozov's discrepancy principle.

Acknowledgement

The authors have been supported by the Austrian Science Fund (FWF) within the Austrian Academy of Sciences, project P19496-N18, and would like to thank Andreas Neubauer and Stefan Kindermann (University of Linz, Austria) for their helpful remarks.

References

[1] Bonesky T 2009 Morozov's discrepancy principle and Tikhonov-type functionals. *Inverse Problems* **25** 015015 doi: 10.1088/0266-5611/25/1/015015

[2] Bredies K, Lorenz D A and Maass P 2008 A generalized conditional gradient method and its connection to an iterative shrinkage method. *Comp. Optim. Appl.* (doi: 10.007/s10589-007-9083-3)

[3] Burger M and Osher S 2004 Convergence rates of convex variational regularization. *Inverse Problems* **20** 1411-1421

[4] Daubechies I, Defries M and DeMol C 2004 An iterative thresholding algorithm for linear inverse problems with a sparsity constraint *Comm. Pure Appl. Math.* **51** 1413-1541

[5] Engl H W, Hanke M and Neubauer A 1996. *Regularization of Inverse Problems* (*Mathematics and its Application* vol 375) (Kluwer Academic Publishers)

[6] Gorenflo R and Hofmann B 1994 On autoconvolution and regularization. *Inverse Problems* **10** 353-373

[7] Grasmair M, Haltmeier M and Scherzer O 2008 Sparse regularization with ℓ^q penalty term. *Inverse Problems* **24** (5) 1-13

[8] Griesse R and Lorenz D A 2008 A semismooth Newton method for Tikhonov functionals with sparsity constraints. *Inverse Problems* **24** 3, 035007

[9] Ito K, Jin B and Zou J A new choice rule for regularization parameters in Tikhonov regularization *Numer. Math.* submitted

[10] Justen L and Ramlau R 2009 A general framework for soft-shrinkage with applications to blind deconvolution and wavelet denoising. *Appl. and Comp. Harmonic Anal.* **26** (1) 43-63

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

[12] Ramlau R 2002 Morozov's discrepancy principle for Tikhonov regularization of non-linear operators. *Journal for Numer. Funct. Anal. and Opt.* **23** 147-172

[13] R Ramlau and G Teschke 2006 A Tikhonov-based projection iteration for non-linear ill-posed problems with sparsity constraints. *Numer. Math.* **104** (2) 177-203

[14] Resmerita E 2005 Regularization of ill-posed problems in Banach spaces: convergence rates. *Inverse Problems* **21** 1303-1314

[15] Resmerita E and Scherzer O 2006 Error estimates for non-quadratic regularization and the relation to enhancement. *Inverse Problems* **22** 801-814

[16] Scherzer O 1993 The use of Morozov's discrepancy principle for Tikhonov regularization for solving non-linear ill-posed problems. *SIAM J. Numer. Anal.,* **30** (6) 1796-1838

[17] Tikhonov A N and Arsenin V Y 1977 *Solutions of Ill-posed Problems.* V. H. Winston and Sonst, Washington D. C., Group, Dordrecht

[18] Zeidler E 1986 *Nonlinear Functional Analysis and its Applications I: Fixed Point Theorems* (Springer-Verlag, New York, Berlin, Heidelberg, Tokyo)