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

Stephan W Anzengruber[1*] and Ronny Ramlau[1,2]

[1] Johann Radon Institute for Computational and Applied Mathematics, Altenbergerstraße 69, 4040 Linz, Austria

[2] Institute for Industrial Mathematics, Johannes Kepler University, Altenbergerstraße 69, 4040 Linz, Austria

October 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 suitable penalty terms the regularized solutions converge to the true solution in the topology induced by Ψ. It is illustrated that for this parameter choice rule it holds $\alpha \to 0$, $\delta^q/\alpha \to 0$ as the noise level δ goes to 0. 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

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

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.

*Corresponding author: Stephan W Anzengruber. Email: stephan.anzengruber@ricam.oeaw.ac.at

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

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

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

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, 15]. For the case of a nonlinear operator see [16], for Hilbert spaces see [7], and for sparsity promoting penalty terms see [12, 14]. 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 known as 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 a regularized solution obtained by minimizing (2). This strategy has been studied extensively as an option to be used in the classical Tikhonov setting [5, 11, 13, 17], but not until recently has its application to inverse problems with more general penalty terms been investigated further. Tikhonov et al. [19] provide a very rigorous analysis of variational methods for solving extremal problems which cover the case (2) under consideration in this paper. The authors discuss several different parameter choice rules, among them the following version of the discrepancy principle. Let $\alpha = \alpha(\delta, y^\delta)$ be chosen such that

$$\inf_{x_\alpha^\delta} \|F(x_\alpha^\delta) - y^\delta\| \le \delta \le \sup_{x_\alpha^\delta} \|F(x_\alpha^\delta) - y^\delta\|.$$

As opposed to our approach, the selection of the regularized solution x_α^δ corresponding to $\alpha = \alpha(\delta, y^\delta)$ is done depending on properties of arbitrarily chosen minimizers $x_{\alpha_1}^\delta, x_{\alpha_2}^\delta$ of (2) corresponding to $\alpha_1 = \alpha/r, \alpha_2 = r\alpha$ for some $r > 1$ and in a way such that the resulting discrepancy may be smaller than the noise level, i.e.

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

is possible, and no lower bound in terms of δ is available. Thus, in this case it can no longer be ensured that the discrepancy is of the same order as the noise level. Also, considertions regarding convergence are done with respect to the underlying (weak) topology.

First results using Morozov's discrepancy principle as defined in (6) below 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 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 achieve this can be found in [2, 4, 8, 12, 14].

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 regularized solutions x_α^δ converge to a true solution $x^\dagger$ with respect to the penalizing functional Ψ, i.e., that

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

and that the resulting parameter choice rule has the properties $\alpha(\delta, y^\delta) \to 0$ and $\delta^q/\alpha(\delta, y^\delta) \to 0$ as $\delta \to 0$. 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 $F : \mathrm{dom}(F) \subset X \to Y$, with $0 \in \mathrm{dom}(F)$, to be weakly continuous, $q > 0$ to be fixed, and that the penalty term $\Psi(x)$ fulfills the following

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

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

(ii) Ψ is weakly lower semicontinous (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 semicontinuity of Ψ we obtain

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

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 [20, 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

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

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.

$$M_\alpha = \{x_\alpha^\delta \in X \ : \ J_\alpha(x_\alpha^\delta) \leq J_\alpha(x), \ \forall x \in X\} \tag{5}$$

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 positive regularization parameter fulfilling the discrepancy principle. Section 4 contains the main regularization results where we will show in particular that for suitable penalty terms the regularized solutions converge with respect to the penalty term as the noise level goes to zero and that the parameter $\alpha = \alpha(\delta, y^\delta)$ chosen such that (3) holds, satisfies

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

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

Let us start by defining Morozov's discrepancy principle.

Definition 3.1. For $1 < \tau_1 \leq \tau_2$ we choose $\alpha = \alpha(\delta, y^\delta) > 0$ such that

$$\tau_1\delta \leq \|F(x_\alpha^\delta) - y^\delta\| \leq \tau_2\delta \tag{6}$$

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

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

Definition 3.2. Let $y^\delta \in Y$ be fixed. For $\alpha \in (0,\infty)$ and $x_\alpha^\delta \in M_\alpha$ we define

$$G(x_\alpha^\delta) = \|F(x_\alpha^\delta) - y^\delta\|, \tag{7}$$

$$\Omega(x_\alpha^\delta) = \Psi(x_\alpha^\delta), \tag{8}$$

$$m(\alpha) = J_\alpha(x_\alpha^\delta). \tag{9}$$

Remark 3.3. 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 $\Omega(x_\alpha^\delta)$.

In the following we state basic properties of G, Ω and m. The proofs can be found, e.g., in [19, Section 2.6].

Lemma 3.4. *The functional $\Omega(x_\alpha^\delta)$ is non-increasing and the functionals $G(x_\alpha^\delta), m(\alpha)$ are non-decreasing with respect to $\alpha \in (0,\infty)$ for any $q > 0$ in the sense that if $0 < \alpha < \beta$ then*

$$\sup_{x_\alpha^\delta \in M_\alpha} G(x_\alpha^\delta) \le \inf_{x_\beta^\delta \in M_\beta} G(x_\beta^\delta),$$

$$\inf_{x_\alpha^\delta \in M_\alpha} \Omega(x_\alpha^\delta) \ge \sup_{x_\beta^\delta \in M_\beta} \Omega(x_\beta^\delta),$$

$$m(\alpha) \le m(\beta).$$

Lemma 3.5. *The functional $m(\alpha)$ defined in* (9) *is continuous on* $(0,\infty)$.

Lemma 3.6. *The set*

$$\{\alpha > 0 \mid \inf_{x_\alpha^\delta \in M_\alpha} G(x_\alpha^\delta) < \sup_{x_\alpha^\delta \in M_\alpha} G(x_\alpha^\delta)\}$$

is at most countable and the functional G is continuous everywhere else with respect to α. The same holds true for Ω and the respective sets of discontinuity points coincide.

Lemma 3.7. *To each $\bar\alpha > 0$ there exist $x_1, x_2 \in M_{\bar\alpha}$ such that*

$$\lim_{\alpha \to \bar\alpha^-} G(x_\alpha^\delta) = G(x_1) = \inf_{x \in M_{\bar\alpha}} G(x) \quad \textit{and} \quad \lim_{\alpha \to \bar\alpha^+} G(x_\alpha^\delta) = G(x_2) = \sup_{x \in M_{\bar\alpha}} G(x).$$

The next Proposition generalizes a result for classical Tikhonov regularization from [13].

Proposition 3.8. *Assume that $\|F(0) - y^\delta\| > \tau_2\delta$, then we can find $\underline{\alpha}, \bar\alpha \in \mathbb{R}^+$ 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

$$\|F(0) - y^\delta\| \le \liminf_{n\to\infty} \|F(x_n) - y^\delta\|, \tag{10}$$

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

□

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 3.1.

Condition 3.9. Assume that y^δ satisfies

$$\|y - y^\delta\| \le \delta < \tau_2\delta < \|F(0) - y^\delta\|, \tag{11}$$

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

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

For the following theorem compare [13, Theorem 2.5] where the same result is proven for the special case of classical Tikhonov regularization.

Theorem 3.10. *If Condition 3.9 is fulfilled, then there are $\alpha = \alpha(\delta, y^\delta) > 0$ and $x_\alpha^\delta \in M_{\alpha(\delta,y^\delta)}$ such that* (6) *holds.*

Proof. Assume that no α fulfilling (6) 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 (6) would hold or Condition 3.9 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 3.8 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 Lemma 3.7 that

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

This is a contradiction since we chose $\tau_1 \le \tau_2$.

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:

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

□

4 Regularization properties

In [19, Section 2.5] Tikhonov *et al* show that the following conditions ensure weak convergence of the regularized solutions to the set $\mathcal{L}$ of Ψ-minimizing solutions. The proof is repeated here for the convenience of the reader in the special case that is of interest in this paper.

Lemma 4.1. *Let $x^\dagger \in \mathcal{L}, \delta_n \to 0$ and assume that $\{x_n\} \subset \operatorname{dom}(F) \cap \operatorname{dom}(\Psi)$ satisfies*

$$\lim_{n\to\infty} \|F(x_n) - y^{\delta_n}\| = 0 \quad \text{and} \quad \limsup_{n\to\infty} \Psi(x_n) \le \Psi(x^\dagger), \tag{12}$$

then x_n converges weakly to the set $\mathcal{L}$ of Ψ-minimizing solutions and $\Psi(x_n) \to \Psi(x^\dagger)$.

Proof. From the second inequality in (12) it is clear that the sequence $\{\Psi(x_n)\}$ is bounded. Thus, the same holds true for $\{x_n\}$ due to the weak coercivity of Ψ and we can extract a subsequence $x_{n'} \rightharpoonup \bar{x}$. Using (12) and the weak lower semi-continuity of $\|F(\cdot) - y\|$ and Ψ, respectively, we get

$$\begin{aligned}
\|F(\bar{x}) - y\| &\le \liminf_{n'\to\infty} \|F(x_{n'}) - y\| \le \liminf_{n'\to\infty} \{ \|F(x_{n'}) - y^{\delta_{n'}}\| + \delta_{n'}\} = 0, \\
\Psi(\bar{x}) &\le \liminf_{n'\to\infty} \Psi(x_{n'}) \le \limsup_{n'\to\infty} \Psi(x_{n'}) \le \Psi(x^\dagger).
\end{aligned}$$

But $x^\dagger$ was chosen to be a Ψ-minimizing solution and therefore $\Psi(\bar{x}) = \Psi(x^\dagger)$ whence it follows that $\bar{x} \in \mathcal{L}$ and $\Psi(x_{n'}) \to \Psi(x^\dagger)$.

The same reasoning applies to any subsequence of $\{x_n\}$ and yields a subsequence weakly converging to $\mathcal{L}$. Therefore the whole sequence weakly converges to $\mathcal{L}$ and $\Psi(x_n) \to \Psi(x^\dagger)$.

□

In the following proof of weak convergence of regularized solutions found through Morozov's discrepancy principle we use techniques similar to [19].

Corollary 4.2. *Let $\delta_n \to 0$. If $\alpha_n = \alpha(\delta_n, y^{\delta_n})$ and $x_n \in M_{\alpha_n}$ satisfy (6), then the sequence $\{x_n\}$ weakly converges to $\mathcal{L}$ and $\Psi(x_n) \to \Psi(x^\dagger)$.*

Proof. From (6) we know that

$$\lim_{n\to\infty} \|F(x_n) - y^{\delta_n}\| \le \lim_{n\to\infty} \tau_2 \delta_n = 0$$

and also that for $x^\dagger \in \mathcal{L}$ it holds true

$$\tau_1^q \delta_n^q + \alpha_n \Psi(x_n) \le \|F(x_n) - y^{\delta_n}\|^q + \alpha_n \Psi(x_n) \le \delta_n^q + \alpha_n \Psi(x^\dagger).$$

Therefore, we obtain

$$0 \le (\tau_1^q - 1)\frac{\delta_n^q}{\alpha_n} \le \Psi(x^\dagger) - \Psi(x_n) \tag{13}$$

whence it follows that

$$\limsup_{n\to\infty} \Psi(x_n) \le \Psi(x^\dagger).$$

Altogether we have shown that the family $\{x_n\}$ satisfies the assumptions of Lemma 4.1 which gives the assertion.

□

For certain penalty terms one can even show convergence with respect to Ψ.

Condition 4.3. Let $\{x_n\} \subset X$ be such that $x_n \rightharpoonup \bar{x} \in X$ and $\Psi(x_n) \to \Psi(\bar{x}) < \infty$, then x_n converges to $\bar{x}$ with respect to Ψ, i.e.,

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

Remark 4.4. 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.

$$\Psi_{p,w}(x) := \|\mathbf{x}\|_{\mathbf{w},p} = \Big(\sum_{\lambda\in\Lambda} w_\lambda\, |\langle x, \phi_\lambda\rangle|^p\Big)^{1/p}, \qquad 1 \le p \le 2, \tag{14}$$

where $0 < w_{\min} \le w_\lambda$, satisfies Condition 4.3. Therefore the same trivially holds for $\Psi_{p,w}(x)^p$. Note that these choices also fulfill all the assumptions in Condition 2.1.

Corollary 4.5. *Let $\delta_n \to 0$ and F, Ψ satisfy the Conditions 2.1, 4.3. Assume that y^{δ_n} fulfills Condition 3.9 and choose $\alpha_n = \alpha(\delta_n, y^{\delta_n})$, $x_n \in M_{\alpha_n}$ such that* (6) *holds, then x_n converges to $\mathcal{L}$ with respect to Ψ.*

Proof. The sequence $\{x_n\}$ satisfies the assumptions of Corollary 4.2 and hence also of Lemma 4.1. From the proof of Lemma 4.1 we see that $\{x_n\}$ has a subsequence $x_{n'} \rightharpoonup x^\dagger \in \mathcal{L}$. According to Corollary 4.2 also $\Psi(x_{n'}) \to \Psi(x^\dagger)$ holds and we obtain from Condition 4.3

$$\Psi(x_{n'} - x^\dagger) \to 0.$$

The same reasoning applies to any subsequence of $\{x_n\}$ and yields a subsequence converging to $\mathcal{L}$ w.r.t. Ψ. Therefore the whole sequence Ψ-converges to $\mathcal{L}$.

□

Remark 4.6. If instead of Condition 4.3 the penalty term $\Psi(x)$ satisfies the Kadec property, i.e., $x_n \rightharpoonup \bar{x} \in X$ and $\Psi(x_n) \to \Psi(\bar{x}) < \infty$ imply $\|x_n - \bar{x}\| \to 0$, then the convergence in Corollary 4.5 holds with respect to the norm.

We will now be concerned with the question if Morozov's discrepancy principle is a regularization method according to the natural generalization of [5, Definition 3.1]. What remains to show is that $\alpha = \alpha(\delta, y^\delta) \to 0$ as $\delta \to 0$. This is in general not necessarily true, but as we will see the following condition is sufficient for that matter.

Condition 4.7. For all $x^\dagger \in \mathcal{L}$ (cf. Definition 2.4) we assume that

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

The following Lemma provides more insight as to the nature of Condition 4.7.

Lemma 4.8. *Let X be a Hilbert space and $q > 1$. If $F(x)$ is differentiable in the directions $x^\dagger \in \mathcal{L}$ and the derivatives are bounded in a neigbourhood of $x^\dagger$, then Condition 4.7 is satisfied.*

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

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

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 4.9. *Assume that Condition 4.7 is satisfied and that there exist* $\alpha > 0$ *and a solution* $x^\star$ *of* $F(x) = y$ *such that*

$$x^\star = \operatorname*{arg\,min}_{x\in X} \left\{ \|F(x) - y\|^q + \alpha\Psi(x) \right\},$$

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 $0, x^\star \in \mathrm{dom}(\Psi)$ 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 4.10. To illustrate that Lemma 4.9 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 4.8 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 4.7 is violated and indeed for $\alpha = 1$ we find that

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

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 4.7 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 4.8).

Theorem 4.11. *Let F, Ψ satisfy the Conditions 2.1, 4.7. Moreover, assume that data y^δ, $\delta \in (0, \delta^*)$, are given such that Condition 3.9 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 3.1) satisfies*

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

Proof. Let $\delta^* > \delta_n \to 0$ and $\alpha_n = \alpha(\delta_n, y^{\delta_n})$ be chosen according to the discrepancy principle. As a shorthand we write $x_n = x_{\alpha_n}^{\delta_n}$ for the corresponding regularized solutions satisfying (6).

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

$$\bar{x}_n = \operatorname*{arg\,min}_{x \in X} \left\{ \|F(x) - y^{\delta_n}\|^q - \underline{\alpha}\Psi(x) \right\}$$

we obtain using Lemma 3.4 and (6) that

$$\begin{aligned}
&\|F(\bar{x}_n) - y^{\delta_n}\| \leq \|F(x_n) - y^{\delta_n}\| \leq \tau_2 \delta_n \to 0, \\
&\limsup_{n\to\infty} \bar{\alpha}\Psi(\bar{x}_n) \leq \limsup_{n\to\infty} \{ \|F(\bar{x}_n) - y^{\delta_n}\| + \bar{\alpha}\Psi(\bar{x}_n)\} \leq \bar{\alpha}\Psi(x^\dagger).
\end{aligned}$$

Therefore, $\{\bar{x}_n\}$ satisfies the assumptions of Lemma 4.1 and we can extract a subsequence $x_{n'} \rightharpoonup x^\dagger \in \mathcal{L}$. Because of the weak lower semi-continuity of Ψ and $\|F(\cdot) - y\|$, it holds that

$$\begin{aligned}
\|F(x^\dagger) - y\|^q + \bar{\alpha}\Psi(x^\dagger) &\leq \liminf_{n'\to\infty} \Big(\|F(\bar{x}_{n'}) - y^{\delta_{n'}}\|^q + \bar{\alpha}\Psi(\bar{x}_{n'}) \Big) \\
&\leq \liminf_{n'\to\infty} \Big(\|F(x) - y^{\delta_{n'}}\|^q + \bar{\alpha}\Psi(x) \Big) && \forall x \in X \\
&= \|F(x) - y\|^q + \bar{\alpha}\Psi(x) && \forall x \in X,
\end{aligned}$$

which shows that $x^\dagger$ is also a minimizer of $J_{\bar{\alpha}}$ with exact data y. Therefore, $x^\dagger$ satisfies the assumptions of Lemma 4.9 and it follows that $x^\dagger = 0$, which in turn means that $y = F(0)$. This violates (11) in Condition 3.9, and we have reached a contradiction.

The second part of the theorem is an immediate consequence of (13) and the assertion $\Psi(x_\alpha^\delta) \to \Psi(x^\dagger)$ in Lemma 4.2.

□

Remark 4.12. In the proof of Theorem 4.11 we have used that $\|F(0) - y\| > 0$, which is an immediate consequence of (11). 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 $\|y - y^\delta\| \le \delta$ we obtain

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

which is (11). Therefore (11) can be fulfilled for all δ smaller than some $\delta^* > 0$, whenever $y \neq F(0)$.

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 [16]. 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

$$F'(x^\dagger)^* w \in \partial\Psi(x^\dagger). \tag{16}$$

Throughout the remainder of this section let $w \in Y^*$ be arbitrary but fixed fulfilling (16) and $\xi \in \partial\Psi(x^\dagger)$ be defined as

$$\xi = F'(x^\dagger)^* w. \tag{17}$$

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

$$\langle w, F(x) - F(z) - F'(z)(x - z) \rangle \le c\,\|w\|_{Y^*}\,\|F(x) - F(z)\|. \tag{18}$$

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

$$\|F(x) - F(x^\dagger) - F'(x^\dagger)(x - x^\dagger)\| \le c\, D_\Psi^\xi(x, x^\dagger), \tag{19}$$

and it holds that

$$c\,\|w\|_{Y^*} < 1. \tag{20}$$

Using the above condition we are now ready to prove the same convergence rates for Morozov's discrepancy principle which were established in [3, 16] 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 Condtions 2.1 and 5.2 hold. For all $0 < \delta < \delta^*$ assume that the data y^δ fulfill Condition 3.9, and choose $\alpha = \alpha(\delta, y^\delta)$ according to the discrepancy principle in Definition 3.1. Then*

$$\|F(x_\alpha^\delta) - F(x^\dagger)\| = \mathcal{O}(\delta), \qquad D_\Psi^\xi(x_\alpha^\delta, x^\dagger) = \mathcal{O}(\delta). \tag{21}$$

Proof. It is an immediate consequence of (6) and (11) that

$$\|F(x_\alpha^\delta) - y\| \leq \|F(x_\alpha^\delta) - y^\delta\| + \|y - y^\delta\| \leq (\tau_2 + 1)\delta, \tag{22}$$

which proves the first part of (21). In order to show the second part we use

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

which shows that $\Psi(x_\alpha^\delta) \leq \Psi(x^\dagger)$. Assume now that Condition 5.2 (i) holds, then using (18) we get for $D_\Psi^\xi(x_\alpha^\delta, x^\dagger) \in D_\Psi(x_\alpha^\delta, x^\dagger)$ 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 (19) and (22) 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 (20) this yields

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

□

Remark 5.4. In the special case where X is a Hilbert space and $\Psi(x) = \|x\|^2$ it holds that $D_\Psi^\xi(x_\alpha^\delta, x^\dagger) = \|x_\alpha^\delta - x^\dagger\|^2$. Thus the convergence rate $\mathcal{O}(\delta)$ with respect to the Bregman distance in Theorem 5.3 corresponds to the well known rate $\mathcal{O}(\sqrt{\delta})$ in norm.

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.

This operator is of importance, for example, in stochastics, where it describes the density of the sum of two independent and identically distributed random variables, or also in spectroscopy (see [6] and the references therein for further details).

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

$$F(f)(s) = (f * f)(s) = \int_0^s f(s-t)f(t)dt. \tag{23}$$

Autoconvolution has been studied in some detail in [6], where the authors showed that for the choice

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

the 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

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

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

$$\|f * h\| \leq \|f\| \, \|h\|. \tag{24}$$

For a proof see [6, Theorem 2, Lemma 4]. Moreover, $F'(\cdot)$ is linear and due to (24) 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.

The adjoint of the Fréchet derivative $F'(f)^*h$ for $f, h \in X$ evaluates to

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

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

$$\begin{aligned}[F'(f)^*h](t) &= 2\int_t^1 f(s-t)\ h(s)\ ds \\ &= 2\int_0^{1-t} f(s)\ \tilde{h}((1-t)-s)\ ds \\ &= 2\ (f * \tilde{h})\ (1-t) = 2\ (f * \tilde{h})\tilde{}\ (t).\end{aligned} \tag{25}$$

In this example we will be reconstructing a solution f of

$$F(f) = g \tag{26}$$

from given noisy data g^δ, $\|g - g^\delta\| \le \delta$. Among all possible solutions f we are interested in choosing the one with the sparsest representation in the Haar wavelets, which form an orthonormal basis of the Hilbert space X. Finding sparse solutions using wavelets is commonly used, e.g., in signal compression, but can also be of importance to obtain a good resolution of discontinuities in the solution. We define

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

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

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

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

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

We now reformulate the problem in the coefficient space, so that all computations can be executed only on sequences in ℓ_2. To this end we express the operator F 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

$$y_\eta = \langle F(f), \varphi_\eta \rangle = \sum_{\lambda,\mu\in\Lambda} x_\lambda x_\mu \ \langle \varphi_\lambda * \varphi_\mu, \varphi_\eta \rangle = x^T K_\eta \ x, \tag{27}$$

where

$$K_\eta := \Big\{ \langle \varphi_\lambda * \varphi_\mu, \varphi_\eta \rangle \Big\}_{\lambda,\mu\in\Lambda} \qquad \forall \eta \in \Lambda. \tag{28}$$

The above representation allows us to consider $\tilde{F} : \ell_2 \to \ell_2$, with

$$\tilde{F}(x) = \{y_\eta\}_{\eta\in\Lambda}, \quad y_\eta := x^T K_\eta \ x. \tag{29}$$

For the iterative computation of the regularized solutions in (30) below, we will also need to express the Fréchet derivative of F in the coefficient space. Adapting the steps leading up to (27) we obtain from (25)

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

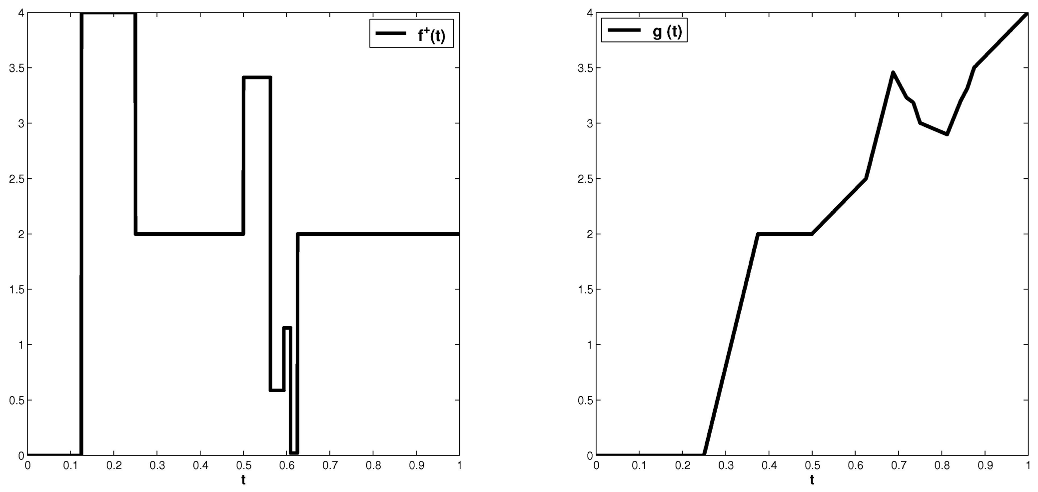

Figure 1: The Ψ-minimizing solution $f^\dagger$ (left), and the exact data $g = F(f^\dagger)$ (right).

where the matrices K_η were defined in (28) and $\tilde{z}$ denotes the coefficients of $\tilde{h}(t) = h(1-t)$. They can be computed according to the formulae

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

where $(j,k)' := (j, 2^j - 1 - k)$.

In order to find a solution that has a sparse representation, we choose the penalty term to be $\Psi(x) = \|x\|_1$ as it is well known that regularization with the ℓ_1 norm promotes sparsity (see, e.g., [4, 7, 10, 12, 14]). We thus consider the Tikhonov-type functionals

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

where from now on y, y^δ denote the coefficient vectors of the exact data g and the noisy data 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 $\|y - y^\delta\| \le \delta$.

To compute the regularized solutions

$$x_\alpha^\delta = \underset{x \in \ell_2}{\arg\min}\, J_\alpha(x)$$

we will use the iterative soft-shrinkage algorithm for non-linear inverse problems from [12], which is based on the surrogate functional approach described in [4]. We denote the soft-shrinkage operator with threshold $\beta > 0$ by S_β, i.e. for $x = \{x_\lambda\}_{\lambda \in \Lambda} \in \ell_2$

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

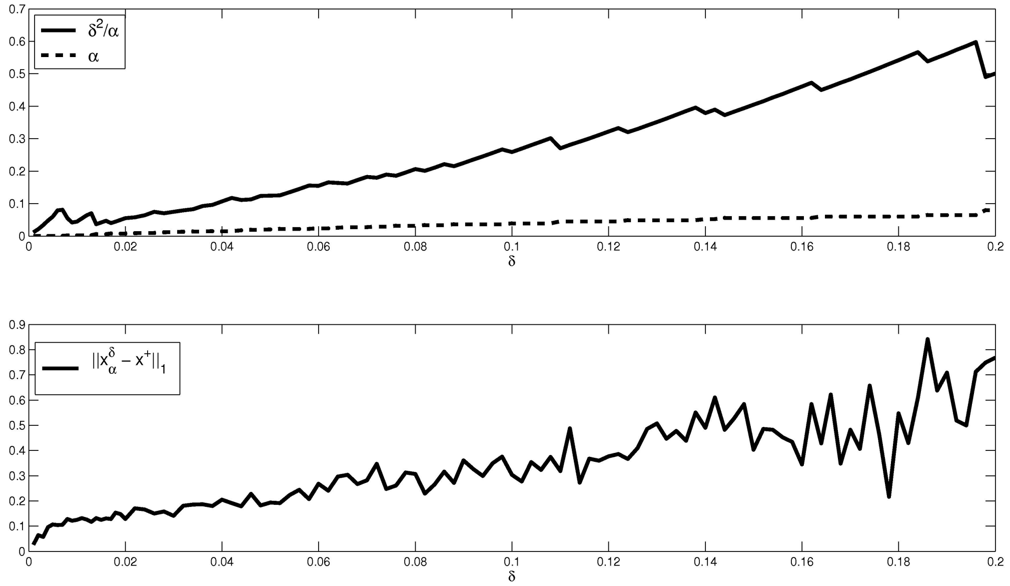

Figure 2: Top: Graph of $\delta \mapsto \delta^2/\alpha(\delta, y^\delta)$ (solid) and of $\delta \mapsto \alpha(\delta, y^\delta)$ (dashed). Bottom: Graph of $\delta \mapsto \|x^\delta_{\alpha(\delta,y^\delta)} - x^\dagger\|_1$. The noise was chosen uniformly random at each step.

It has been shown in [12] that for arbitrary $x^0 \in \ell_2$ and with

$$\begin{aligned} x^{n,0} &= x_n, \\ x^{n,k+1} &= S_{\alpha/2C}\left(x^{n,0} + \frac{1}{C}F'(x^{n,k})^*\left(y^\delta - F(x^{n,0})\right)\right), \\ x^{n+1} &= \lim_{k\to\infty} x^{n,k} \end{aligned} \tag{30}$$

the resulting sequence x_n converges at least to a critical point of $J_\alpha(x)$. Here the constant C has to be choosen large enough for the algorithm to converge (see [12] for details). Choosing the starting value x^0 reasonably close to the true solution we observed numerically that in this case the iteration actually approximates a minimizer.

Moreover, choosing the right hand side g 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 (26) possesses the form

$$f(t) = \begin{cases} 0 & \text{a.e. in } \ t \in [0, \varepsilon/2] \\ \text{uniquely determined by } g & \text{a.e. in } \ t \in [\varepsilon/2, 1-\varepsilon/2] \\ \text{arbitrarily non-negative} & \text{in } \ t \in [1-\varepsilon/2, 1] \end{cases} \tag{31}$$

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 (31), we know that they are uniquely determined on $[0, \frac{7}{8}]$, but differ on $(\frac{7}{8}, 1]$. Also, the right hand side was chosen such that it actually permits a sparse solution. The Ψ-minimizing solution $f^\dagger$ (see Figure 1) has four non-zero coefficients when expressed in the Haar wavelet basis. Note that in this example the solution

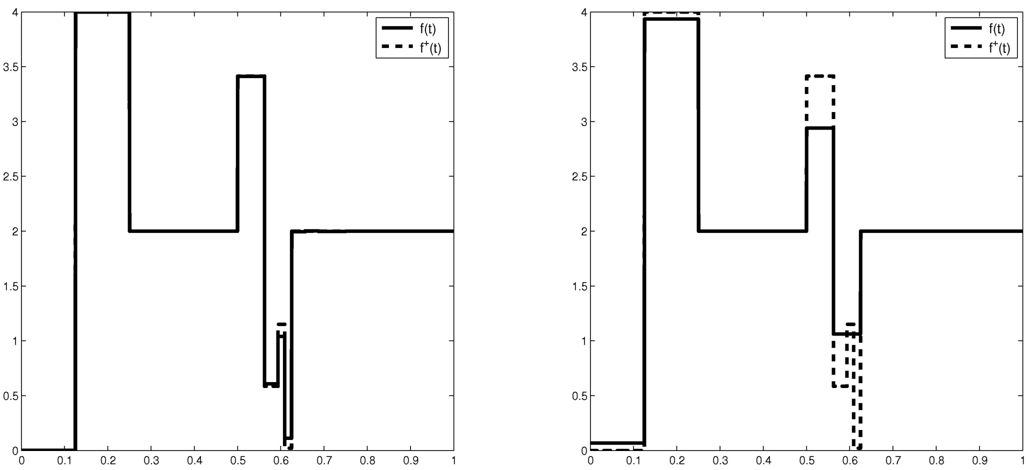

Figure 3: Regularized solutions with noise level $\delta = 0.001$ (left) and $\delta = 0.1$ (right) as solid lines together with the Ψ-minimizing solution $f^\dagger$ (dashed).

$f^\dagger$ is different from the minimum norm solution, which would have the constant value zero on $(\frac{7}{8}, 1]$.

Since we have seen that the operator F and the penatly term Ψ fulfill all the conditions needed to apply Theorem 4.11 and Corollary 4.5 (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. For a sample case with wavelets up to a maximal index level $J = 5$ and noise chosen uniformly random such that $\|y - y^\delta\| = \delta,\ \delta \in (0, 0.2]$ at each step, we show the results in Figure 2. The corresponding values of $\alpha = \alpha(\delta, y^\delta)$ can be found, e.g., using a bisecting trial and error approach exploring the monotonicity of the residual functional $G(x_\alpha^\delta)$ (cf. Lemma 3.4). The reconstructed solutions have between three and seven non-zero coefficients for $10^{-2} \leq \delta \leq 0.2$ and eight to eleven non-zero coefficients for $10^{-3} \leq \delta < 10^{-2}$. As the noise level – and thus also the regularization parameter – approaches zero, the data fit term becomes dominant and the coefficient sequences less sparse. The solutions corresponding to the noise levels $\delta = 10^{-3}$ and $\delta = 10^{-1}$ can be seen in Figure 3, where the regularization parameters $\alpha = 10^{-4}$ and $\alpha = 3.872 \cdot 10^{-2}$, respectively, were chosen according to Morozov's discrepancy principle.

7 Conclusion

We have studied Morozov's discrepancy principle as a parameter choice rule for non-linear inverse problems with Tikhonov type regularization using a general convex penalty term $\Psi(x)$. Adapting results from [19] we showed weak convergence of the regularized solutions to the set of Ψ-minimzing solutions. For a large class of widely used penalty terms (including the ℓ^p norms) this even implies strong convergence or convergence with respect to the penalty term (which may be even stronger than norm convergence). We gave a condition (see

Condition 4.7) which is sufficient for the so chosen regularization parameter $\alpha = \alpha(\delta, y^\delta)$ to satisfy $\alpha \to 0$ and $\delta^q/\alpha \to 0$ as the noise level $\delta \to 0$, where $q > 1$ can be chosen suitable for the problem under consideration (see definition 2.4). A key ingredient here was to show that a Ψ-minimizing solution cannot be a regularized solution minimizing the Tikhonov-type functional with exact data (cf. Lemma 4.9).

Convergence rates of order δ were obtained in the Bregman distance (corresponding to $\sqrt{\delta}$ in the classical Hilbert space setting) under the common smoothness type source condition and the nonlinearity conditions introduced in [3] and [16].

Finally, to substantiate our theoretical findings on a practical example we discussed the autoconvolution operator over a finite interval reconstructing sparse solutions with respect to a wavelet basis.

Acknowledgments

This work has been supported by the Austrian Science Fund (FWF) within the Austrian Academy of Sciences, project P19496-N18.

The authors would like to thank Andreas Neubauer and Stefan Kindermann (University of Linz, Austria) for their helpful remarks and the referees for their valuable suggestions which led to substantial improvements of the paper.

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.* **42** 173-193 (doi: 10.1007/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) (Dordrecht: 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] Morozov V A 1984 *Methods for solving incorrectly posed problems.* (New York: Springer-Verlag)

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

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

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

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

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

[17] 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

[18] Tikhonov A N and Arsenin V Y 1977 *Solutions of Ill-posed Problems.* (Washington, DC: V. H. Winston & Sons)

[19] Tikhonov A N, Leonov A S and Yagola A G 1998 *Nonlinear Ill-posed Problems.* (London: Chapman & Hall)

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