Regularization in Banach spaces – optimal convergence rates results

Torsten Hein

May 8, 2008

Abstract

In this preprint we deal with convergence rates for a Tikhonov-like regularization approach for linear and non-linear ill-posed problems in Banach spaces. Therefore we deal with so-called distance functions which quantify the violation of a (non-linear) reference source condition. Under validity of this reference source condition we derive convergence rates which are optimal in a Hilbert space situation. In the linear case we additionally present error bounds and convergence rates which base on the decay rate of the distance functions when the reference source condition is violated.

MSC2000 classification scheme numbers: 47A52, 47J06, 49N45

Keywords: inverse problem, nonlinear ill-posed problem, regularization, Bregman distance, convergence rates

1 Introduction

Let $\mathcal{X}$ and $\mathcal{Y}$ denote reflexive Banach spaces. Introducing, let $A : \mathcal{X} \longrightarrow \mathcal{Y}$ describes a linear and bounded operator with non-closed range, i.e. $\overline{\mathcal{R}(A)} \neq \mathcal{R}(A)$. We consider the linear ill-posed equation

$$A\,x = y^\delta, \quad x \in \mathcal{X}, \tag{1}$$

with noisy data $y^\delta \in \mathcal{Y}$. Here, only the estimate $\|y - y^\delta\| \leq \delta$, $\delta \geq 0$, is known, when the exact data is denoted by $y \in \mathcal{Y}$. We assume that there exists a solution $x^\dagger \in \mathcal{X}$ of (1) for given exact data, i.e. the equation $A\,x^\dagger = y$ holds.

For a stable approximate solution of (1) we deal with modified Tikhonov regularization

$$\frac{1}{p}\|A\,x - y^\delta\|^p + \alpha\,P(x) \to \min \quad \text{subject to } \ x \in \mathcal{D}(P), \tag{2}$$

where $P : \mathcal{D}(P) \subseteq \mathcal{X} \longrightarrow \mathcal{Y}$ defines a nonnegative convex stabilizing functional and $p \geq 1$ is a given parameter. The problem is well-studied if $p = 2$, $\mathcal{X}$ and $\mathcal{Y}$ are Hilbert spaces and $P(x) := \frac{1}{2}\|x\|^2$, see e.g. [4] and the references therein. Let us – as usual – denote a solution of (2) with x_α^δ, if it exists. In order to prove convergence rates $x_\alpha^\delta \to x^\dagger$ for an parameter choice $\alpha = \alpha(\delta) \to 0$ and $\delta \to 0$ an additional smoothness condition has to be satisfied, for example

$$x^\dagger = \varphi(A^*A)\omega, \qquad \omega \in \mathcal{X}, \tag{3}$$

with strictly increasing function $\varphi(t)$, $0 \le t \le \|A\|^2$, and $\varphi(0) = 0$, see also [8, 12, 14] and [18] for some newer results. It is also well-established, that $\varphi(t) = t$, $t \ge 0$, and a parameter choice strategy $\alpha(\delta) \sim \delta^{\frac{2}{3}}$ lead to the optimal convergence rate $\|x_\alpha^\delta - x^\dagger\| \sim \delta^{\frac{2}{3}}$ for Tikhonov regularization. On the other hand, for function $\varphi(t)$, $t \ge 0$, with $t\,\varphi(t) \to 0$ for $t \to 0$, the condition (3) can be interpreted as weakening of the condition $x^\dagger \in \mathcal{R}(A^*A)$ leading to lower convergence rates. However, the convergence rates theory essentially bases on spectral calculus of selfadjoint operators in Hilbert spaces, see e.g. [4, section 2.3]. Since we now deal with non-Hilbert spaces this theory cannot be applied in our situation.

In [7] an alternative concept for proving convergence rates in Hilbert spaces were presented. Here, the violation of a reference source condition (3) with fixed function $\varphi(t)$, $t \ge 0$, is measured by so-called distance functions $d = d(R)$, $R \ge 0$. Based on the decay rate of these distance functions for $R \to \infty$, convergence rates can be proved in a similar way as with source condition (3) and arbitrary function $\varphi(t)$, $t \ge 0$. It turns out, that the idea of distance function can be generalized to Banach spaces by replacing condition (3) by an appropriate reference source condition in Banach spaces. This is the main purpose of the present paper. We also refer to [6] for some first results where a reference source condition was assumed, which do not propose optimal convergence rate.

We also want to mention the papers [15] and [16] which deal with convergence rates for regularizing linear and nonlinear operator equations with operators mapping from a Banach space into a Hilbert space. Convergence rates for regularizing operator equations with operator mapping between two Banach spaces were recently presented in [9].

The paper is organized as follows: in section 2 basic notations and assumptions were introduced. In section 3 we present error bounds in terms of Bregman distances for regularized solutions of (2) under additional (nonlinear) smoothness conditions on the exact solution $x^\dagger \in \mathcal{X}$ of (1). Introducing distance functions in section 4 we formulate first convergence rates based on an a-priori choice of the regularization parameter α. Under an additional convexity condition on the penalty functional $P(x)$ improved convergence rates were derived in section 5. Since the results of section 4 and 5 depend on the choice of the parameter $p > 1$ we present conditions under which we obtain an unified convergence rate result for arbitrary $p > 1$. Sections 7 shows that the derived convergence rates are of optimal order if the spaces $\mathcal{X}$ and $\mathcal{Y}$ are Hilbert spaces. Afterwards, an a-posteriori parameter choice is presented leading to optimal convergence rates. Finally, in section 9, convergence rates for nonlinear operator equations were shown under validity of the proposed smoothness condition.

2 Basic Assumptions and Notations

Throughout the paper the spaces $\mathcal{X}$ and $\mathcal{Y}$ are assumed to be reflexive Banach spaces with dual spaces $\mathcal{X}^*$ and $\mathcal{Y}^*$, respectively. For a linear operator $A : \mathcal{X} \longrightarrow \mathcal{Y}$ we denote with $A^\star : \mathcal{Y}^* \longrightarrow \mathcal{X}^*$ the dual operator of A, i.e.

$$\langle v, A\,x\rangle_{\mathcal{Y}^*,\mathcal{Y}} = \langle A^\star v, x\rangle_{\mathcal{X}^*,\mathcal{X}}, \qquad \forall\, x \in \mathcal{X},\ \forall\, v \in \mathcal{Y}^*$$

hold. Here, $\langle\cdot,\cdot\rangle_{\mathcal{X}^*,\mathcal{X}}$ and $\langle\cdot,\cdot\rangle_{\mathcal{Y}^*,\mathcal{Y}}$ stay for the duality products in $\mathcal{X}$ and $\mathcal{Y}$, respectively. Since we also deal with nonlinear problems, we replace (1) by the nonlinear equation

$$F(x) = y^\delta, \qquad x \in \mathcal{D}(F), \tag{4}$$

where $F : \mathcal{D}(F) \subseteq \mathcal{X} \longrightarrow \mathcal{Y}$ describes a nonlinear operator with domain $\mathcal{D}(F)$. For solving (4) approximately in a stable manner we consider the minimization problem

$$J_\alpha(x) := \frac{1}{p}\|F(x) - y^\delta\|^p + \alpha P(x) \to \min \quad \text{subject to } \ x \in \mathcal{D}(F) \cap \mathcal{D}(P), \tag{5}$$

with stabilizing functional $P : \mathcal{D}(P) \subseteq \mathcal{X} \longrightarrow \mathbb{R}$. In our context, we need the following assumptions:

(A1) For any sequence $\{x_n\} \subset \mathcal{D}(F)$ with weak convergence $x_n \rightharpoonup x \in \mathcal{D}(F)$ and $\{F(x_n)\}$ bounded we have $x \in \mathcal{D}(F)$ and weak convergence $F(x_n) \rightharpoonup F(x)$.

(A2) The nonnegative functional P is convex and weakly lower semi-continuous.

(A3) The set $\mathcal{D}(F) \cap \mathcal{D}(P) \neq \emptyset$ is weakly closed.

(A4) The level sets

$$\mathcal{S}_\alpha(M) := \{x \in \mathcal{D}(F) \cap \mathcal{D}(P) \ :\ J_\alpha(x) \le M\}$$

are bounded for each $\alpha > 0$ and each $M > 0$.

(A5) We have $\delta \in [0, \delta_{max}]$ and $\alpha \in (0, \alpha_{max}]$.

We additional make use of a further assumption. In particular, the existence of a solution of equation (4) is supposed for given exact data $y \in \mathcal{Y}$. We recall, that a P-minimizing solution $x^\dagger$ of equation (4) with $\delta = 0$ satisfies

$$P(x^\dagger) := \min\{P(x)\ :\ F(x) = y\}. \tag{6}$$

Note, that if (4) has a solution then it has also a P-minimizing solution, see e.g. [9, Theorem 3.4]. Under the conditions stated above there exists a solution $x_\alpha^\delta \in \mathcal{D}(F) \cap \mathcal{D}(P)$ of (5). Moreover, the solution x_α^δ depends stable on the given data in the weak sense: assume, that the solution $x^\dagger$ of (4) with $\delta = 0$ is unique. Then, for $y^\delta \to y$ and solutions x_α^δ of (6) with $\alpha = \alpha(\delta)$ chosen such that

$$\alpha \to 0,\ \frac{\delta^p}{\alpha} \to 0 \quad \text{for} \quad \delta \to 0$$

we can conclude $x_\alpha^\delta \rightharpoonup x^\dagger$, see e.g. [9, Theorem 3.5]. Under the assumptions stated above we cannot conclude strong convergence $x_\alpha^\delta \to x^\dagger$, see e.g. [17] for further conditions ensuring strong convergence. On the other hand,

$$\begin{aligned}
\frac{1}{p}\|F(x_\alpha^\delta) - y^\delta\|^p + \alpha P(x_\alpha^\delta) &= J_\alpha(x_\alpha^\delta) \\
&\leq J_\alpha(x^\dagger) \\
&= \frac{1}{p}\|F(x^\dagger) - y^\delta\|^p + \alpha P(x^\dagger) \\
&\leq \frac{\delta_{max}^p}{p} + \alpha_{max} P(x^\dagger) =: M = M(\delta_{max}, \alpha_{max}).
\end{aligned}$$

Hence, $x_\alpha^\delta, x^\dagger \in \mathcal{S}_\alpha(M)$ for each $\delta \in [0, \delta_{max}]$ and $\alpha \in (0, \alpha_{max}]$. By (A4), there exists a constant $K > 0$ such that

$$\|x_\alpha^\delta - x^\dagger\| \leq K \qquad \forall \delta \in [0, \delta_{max}],\ \forall \alpha \in (0, \alpha_{max}], \tag{7}$$

holds. Without further assumptions, this is the only known estimate in the norm-topology.

Additionally we define Bregman distances which has been well-established for presenting convergence rates for general stabilizing functionals $P(x)$, see e.g. [2].

Definition 2.1 *Let $P : \mathcal{D}(P) \subset \mathcal{X} \longrightarrow [0, \infty)$ denotes a convex functional with sub-differential $\partial P(x)$ for $x \in \mathcal{D}(P)$. The Bregman distance of two elements $\tilde{x}, x \in \mathcal{D}(P)$ and $\xi \in \partial P(x)$ is defined as*

$$D_P(\tilde{x}, x) := P(\tilde{x}) - P(x) - \langle \xi, \tilde{x} - x \rangle_{\mathcal{X}^*, \mathcal{X}} \geq 0.$$

Then we can state the following assumption.

(A6) There exist a P-minimizing element $x^\dagger \in \mathcal{D}(F) \cap \mathcal{D}(P)$ and an element $0 \neq \xi^\dagger \in \partial P(x^\dagger)$.

In particular, the sub-differential $\partial P(x^\dagger)$ should be non-empty.

We need some further assumptions. With $J_Y : \mathcal{Y} \longrightarrow \mathcal{Y}^*$ we denote the duality map in the space $\mathcal{Y}$, i.e.

$$J_Y(y) := \left(\frac{1}{2}\|y\|^2\right)', \qquad y \in \mathcal{Y}.$$

Then the following properties are supposed.

(A7) There exists a constant $C_1 > 0$ such that

$$D_P(x, x^\dagger) \leq C_1 \|x - x^\dagger\|^2, \qquad \forall\, x \in \mathcal{B}_\rho(x^\dagger),$$

with ball $\mathcal{B}_\rho(x^\dagger)$ around $x^\dagger$, $\rho \geq K > 0$ chosen sufficiently large.

(A8) The duality map J_Y is differentiable in y and there exists a constant $C_2 > 0$ such that

$$\|J_Y'(y)\| \leq C_2, \qquad \forall\, y \in \mathcal{B}_{\tilde{\rho}}(0),$$

with radius $\tilde{\rho} > 0$ sufficiently large.

Note, that e.g. in L^q-spaces for $q \geq 2$ the term $\|J_Y'(y)\|$ is uniformly bounded, i.e. (A8) holds with $\rho = \infty$ for some constant $C_2 > 0$.

3 Some preliminary estimates

Let us now return the linear equation (1). For deriving error bounds and convergence rates the element $\xi^\dagger \in \partial P(x^\dagger)$ of (A6) has to fulfill an additional smoothness condition. In [16], e.g., a source condition

$$\xi^\dagger = A^\star \omega, \qquad \omega \in \mathcal{Y}^*,\ \|\omega\| \le R, \tag{8}$$

for some $R \ge 0$ is supposed. However, this condition does not provides optimal convergence rates in Hilbert spaces. For improved convergence rates we introduce the stronger condition

$$\xi^\dagger = A^\star J_Y(A\,\omega), \qquad \omega \in \mathcal{X},\ \|\omega\| \le R. \tag{9}$$

Note, that (9) is in general a nonlinear source condition. Moreover, we follow a more general strategy. We weaken condition (9) by assuming an approximative source condition

$$\xi^\dagger = A^\star J_Y(A\,\omega) + \upsilon, \qquad \omega \in \mathcal{X},\ \upsilon \in \mathcal{X}^*,\ \|\omega\| \le R,\ \|\upsilon\| \le d. \tag{10}$$

for some $R, d \ge 0$. Of course, the representation elements ω and υ of the approximative source condition (10) are not unique for given $\xi \in \mathcal{X}^*$. On the other hand, this observation will be used later in section 4 for deriving convergence rates. For presenting a unified framework we introduce the sets

$$\mathcal{M}(R,d) := \{\xi \in \mathcal{X}^* \ :\ \xi = A^\star J_Y(A\,\omega) + \upsilon,\ \omega \in \mathcal{X},\ \upsilon \in \mathcal{X}^*,\ \|\omega\| \le R,\ \|\upsilon\| \le d\}\,.$$

for each $R, d \ge 0$. The interpretation is quite simple: for given $\xi \in \mathcal{X}^*$ the number $d \ge 0$ describes the maximal violation of the (reference) source condition (9) which is allowed, when the norm of the source element $\omega \in \mathcal{X}$ is bounded by some constant $R > 0$. Clearly, $\xi \in \mathcal{M}(R,0)$ is equivalent to the availability of the source condition (9).

Now we can prove a first estimate for the case $p > 1$.

Lemma 3.1 *Assume (A1)-(A6) and $\xi^\dagger \in \mathcal{X}^*$ satisfies (10) for some $\omega \in \mathcal{X}$ and $\upsilon \in \mathcal{X}^*$. If $p > 1$, $\gamma := \alpha^{\frac{1}{p-1}} \|A\,\omega\|^{\frac{2-p}{p-1}}$, then*

$$D_P(x_\alpha^\delta, x^\dagger) \le D_P(x^\dagger - \gamma\omega, x^\dagger) + \frac{1}{\alpha} D_Y(D_Y(y^\delta - A(x^\dagger - \gamma\omega), \gamma A\,\omega) + \|\upsilon\|\,(K + \gamma\|\omega\|)\,, \tag{11}$$

where $K > 0$ is the constant of (7).

Proof. By assumption, the approximative source condition (10) holds. For $\gamma > 0$ we have

$$\begin{aligned}
P(x_\alpha^\delta) - P(x^\dagger - \gamma\omega) &= P(x_\alpha^\delta) - P(x^\dagger) + P(x^\dagger) - P(x^\dagger - \gamma\omega) \\
&= \langle \xi^\dagger, x_\alpha^\delta - x^\dagger\rangle + D_P(x_\alpha^\delta, x^\dagger) - \langle \xi^\dagger, -\gamma\omega\rangle - D_P(x^\dagger - \gamma\omega, x^\dagger) \\
&= \langle J_Y(A\,\omega), A(x_\alpha^\delta + \gamma\omega - x^\dagger)\rangle + D_P(x_\alpha^\delta, x^\dagger) - D_P(x^\dagger - \gamma\omega, x^\dagger) \\
&\quad + \langle \upsilon, x_\alpha^\delta + \gamma\omega - x^\dagger\rangle \\
&= \langle J_Y(A\,\omega), A(x_\alpha^\delta - x^\dagger)\rangle + D_P(x_\alpha^\delta, x^\dagger) - D_P(x^\dagger - \gamma\omega, x^\dagger) \\
&\quad + \gamma\|A\,\omega\|^2 + \langle \upsilon, x_\alpha^\delta + \gamma\omega - x^\dagger\rangle.
\end{aligned}$$

Moreover, we obtain

$$\begin{aligned}\frac{1}{p}\|A(x^\dagger-\gamma\omega)-y^\delta\|^p &= \frac{\gamma^p}{p}\|A\,\omega\|^p+\gamma^{p-1}\|A\,\omega\|^{p-2}\langle J_Y(A\,\omega),y^\delta-A\,x^\dagger\rangle\\ &\quad +D_Y(y^\delta-A(x^\dagger-\gamma\omega),\gamma A\,\omega).\end{aligned}$$

We set $\gamma^{p-1}:=\alpha\|A\,\omega\|^{2-p}$. We conclude by the minimizing property of x_α^δ

$$\frac{1}{p}\|A\,x_\alpha^\delta-y^\delta\|^p+\alpha\left(P(x_\alpha^\delta)-P(x^\dagger-\gamma\omega)\right)\le\frac{1}{p}\|A(x^\dagger-\gamma\omega)-y^\delta\|^p.$$

Hence we can conclude

$$\begin{aligned}\frac{1}{p}\|A\,x_\alpha^\delta-y^\delta\|^p+\alpha D_P(x_\alpha^\delta,x^\dagger) &\le \alpha D_P(x^\dagger-\gamma\omega,x^\dagger)+D_Y(y^\delta-A(x^\dagger-\gamma\omega),\gamma A\,\omega)\\ &\quad +\frac{\gamma^p}{p}\|A\,\omega\|^p+\gamma^{p-1}\|A\,\omega\|^{p-2}\langle J_Y(A\,\omega),y^\delta-A\,x^\dagger\rangle\\ &\quad -\alpha\gamma\|A\,\omega\|^2+\alpha\langle J_Y(A\,\omega),A\,x^\dagger-A\,x_\alpha^\delta\rangle\\ &\quad +\alpha\|v\|\|x_\alpha^\delta+\gamma\omega-x^\dagger\|\\ &= \alpha D_P(x^\dagger-\gamma\omega,x^\dagger)+D_Y(y^\delta-A(x^\dagger-\gamma\omega),\gamma A\,\omega)\\ &\quad +\frac{\gamma\alpha}{p}\|A\,\omega\|^2-\alpha\gamma\|A\,\omega\|^2+\alpha\langle J_Y(A\,\omega),y^\delta-A\,x_\alpha^\delta\rangle\\ &\quad +\alpha\|v\|\left(\gamma\|\omega\|+K\right)\\ &\le \alpha D_P(x^\dagger-\gamma\omega,x^\dagger)+D_Y(y^\delta-A(x^\dagger-\gamma\omega),\gamma A\,\omega)\\ &\quad +\frac{\gamma\alpha}{p}\|A\,\omega\|^2-\alpha\gamma\|A\,\omega\|^2+\frac{1}{p}\|A\,x_\alpha^\delta-y^\delta\|^p+\frac{(\alpha\|A\,\omega\|)^q}{q}\\ &\quad +\alpha\|v\|\left(\gamma\|\omega\|+K\right)\end{aligned}$$

with $p^{-1}+q^{-1}=1$ or $q=\frac{p}{p-1}$. Since

$$\alpha^{q-1}=\alpha^{\frac{1}{p-1}}=\gamma\|A\,\omega\|^{\frac{p-2}{p-1}}\quad\text{and}\quad\|A\,\omega\|^{q-2}=\|A\,\omega\|^{\frac{2-p}{p-1}}$$

we obtain

$$\alpha^q\|A\,\omega\|^q=\alpha\|A\,\omega\|^2\alpha^{q-1}\|A\,\omega\|^{q-2}=\alpha\gamma\|A\,\omega\|^2$$

and hence

$$\begin{aligned}\frac{1}{p}\|A\,x_\alpha^\delta-y^\delta\|^p+\alpha D_P(x_\alpha^\delta,x^\dagger) &\le \alpha D_P(x^\dagger-\gamma\omega,x^\dagger)+D_Y(y^\delta-A(x^\dagger-\gamma\omega),\gamma A\,\omega)\\ &\quad +\frac{1}{p}\|A\,x_\alpha^\delta-y^\delta\|^p+\gamma\alpha\|A\,\omega\|^2\left(\frac{1}{p}+\frac{1}{q}-1\right)\\ &\quad +\alpha\|v\|\left(\gamma\|\omega\|+K\right).\quad\blacksquare\end{aligned}$$

For deriving error bounds we have to find bounds for the residuals $D_P(x^\dagger-\gamma\,\omega,x^\dagger)$ and $D_Y(y^\delta-A(x^\dagger-\gamma\omega),\gamma\,A\,\omega)$. Here, we have to distinguish between the cases $p>2$ and $p<2$: depending on $p\neq 2$ a lower or an upper bound on $\|A\,\omega\|$ is needed. The correct statements we present in the following two lemmas. We start with estimates for the term $D_P(x^\dagger-\gamma\,\omega,x^\dagger)$.

Lemma 3.2 *Assume (A7), $0 \neq \xi^\dagger \in \mathcal{X}^*$ satisfies (10) for some $\omega \in \mathcal{X}$ and $\upsilon \in \mathcal{X}^*$ and $\gamma := \alpha^{\frac{1}{p-1}} \|A\,\omega\|^{\frac{2-p}{p-2}}$.*

(i) If $1 < p \leq 2$ then
$$D_P(x^\dagger - \gamma\omega, x^\dagger) \leq C_3 \alpha^{\frac{2}{p-1}} \|\omega\|^{\frac{2}{p-1}},$$
where $C_3 > 0$ is a constant which does not depend on ω, υ, δ and α.

(ii) If $p > 2$ and $\|\upsilon\| \leq c\|\xi^\dagger\|$ for some constant $0 < c < 1$, then
$$D_P(x^\dagger - \gamma\omega, x^\dagger) \leq \tilde{C}_3 \alpha^{\frac{2}{p-1}} \|\omega\|^2.$$
where $\tilde{C}_3 > 0$ is a constant which does not depend on ω, υ, δ and α.

Proof. From (A7) we conclude
$$D_P(x^\dagger - \gamma\omega, x^\dagger) \leq C_1 \gamma^2 \|\omega\|^2 = C_1 \alpha^{\frac{2}{p-1}} \|A\,\omega\|^{\frac{4-2p}{p-1}} \|\omega\|^2$$
If $1 < p \leq 2$ we estimate $\|A\,\omega\| \leq \|A\|\,\|\omega\|$ which provides
$$D_P(x^\dagger - \gamma\omega, x^\dagger) \leq C_1 \|A\|^{\frac{2-p}{p-1}} \alpha^{\frac{2}{p-1}} \|\omega\|^{2+\frac{4-2p}{p-1}} = C_1 \|A\|^{\frac{2-p}{p-1}} \alpha^{\frac{2}{p-1}} \|\omega\|^{\frac{2}{p-1}}.$$
Hence (i) holds with $C_3 := C_1 \|A\|^{\frac{2-p}{p-1}}$. For $p > 2$ and the additional assumption we conclude by the inverse triangle inequality
$$\frac{\|\xi^\dagger\|}{1-c} \leq \|A^* J_Y(A\,\omega)\| \leq \|A^*\|\,\|J_Y(A\,\omega)\| = \|A\|\,\|A\,\omega\|$$
which implies $\|A\,\omega\| \geq \|\xi^\dagger\| [(1-c)\|A\|]^{-1}$. Hence
$$D_P(x^\dagger - \gamma\omega, x^\dagger) \leq C_1 \alpha^{\frac{2}{p-1}} \|\omega\|^2 \left(\frac{\|\xi^\dagger\|}{(1-c)\|A\|} \right)^{\frac{2-p}{p-1}},$$
which proves (ii) with $\tilde{C}_3 := C_1 \left(\frac{\|\xi^\dagger\|}{(1-c)\|A\|} \right)^{\frac{2-p}{p-1}}$. ■

For the second residual $D_Y(y^\delta - A(x^\dagger - \gamma\omega), \gamma\,A\,\omega)$ we can find the following estimates.

Lemma 3.3 *Assume (A8) and $0 \neq \xi^\dagger \in \mathcal{X}^*$ satisfies (10) for some $\omega \in \mathcal{X}$ and $\upsilon \in \mathcal{X}^*$ and $\gamma := \alpha^{\frac{1}{p-1}} \|A\,\omega\|^{\frac{2-p}{p-2}}$. Moreover, if $p \neq 2$, there exists a constant $0 < \tilde{c} < 1$ such that $\delta \leq \tilde{c}\,\gamma \|A\,\omega\|$.*

(i) If $1 < p < 2$ and $\|\upsilon\| \leq c\|\xi^\dagger\|$ for some constant $0 < c < 1$, then
$$D_Y(y^\delta - A(x^\dagger - \gamma\omega), \gamma\,A\,\omega) \leq \tilde{C}_4 \delta^2 \alpha^{1+\frac{1}{1-p}},$$
where $\tilde{C}_4 > 0$ is a constant which does not depend on ω, υ, δ and α.

(ii) If $p \geq 2$ then
$$D_Y(y^\delta - A(x^\dagger - \gamma\omega), \gamma\,A\,\omega) \leq C_4 \delta^2 \alpha^{1+\frac{1}{1-p}} \|\omega\|^{\frac{p-2}{p-1}},$$
where $C_4 >$ is a constant which does not depend on ω, υ, δ and α.

PROOF. For $f(y) := \frac{1}{p}\|y\|^p$ and $0 \neq y \in \mathcal{Y}$ we have

$$f'(y) = \|y\|^{p-2} J_Y(y) \quad \text{and} \quad f''(y) = (p-2)\|y\|^{p-4} \left[J_Y(y)\cdot\right] \left[J_Y(y)\cdot\right] + \|y\|^{p-2} J_Y'(y).$$

Hence, since $\|J_Y(y)\| = \|y\|$,

$$\|f''(y)\| \leq |p-2| \|y\|^{p-4} \|J_Y(y)\|^2 + C_2 \|y\|^{p-2} = \left(|p-2| + C_2\right) \|y\|^{p-2}.$$

Moreover for $\tau \in (0,1)$ we have $(1-\tilde{c})\gamma\|A\,\omega\| \leq \|\tau(y^\delta - A\,x^\dagger) + \gamma\,A\,\omega\| \leq (1+\tilde{c})\gamma\|A\,\omega\|$. Since

$$D_Y(y^\delta - A(x^\dagger - \gamma\omega), \gamma\,A\,\omega) = f''(\tau(y^\delta - A\,x^\dagger) + \gamma\,A\,\omega)(y^\delta - A\,x^\dagger, y^\delta - A\,x^\dagger)$$

for some $\tau \in (0,1)$ we conclude

$$\begin{aligned} D_Y(y^\delta - A(x^\dagger - \gamma\omega), \gamma\,A\,\omega) &\leq C\,\delta^2 \gamma^{p-2} \|A\,\omega\|^{2-p} \\ &= C\,\delta^2 \gamma^{p-1} \|A\,\omega\|^{2-p} \frac{1}{\gamma} \\ &= C\,\delta^2 \alpha^{1+\frac{1}{1-p}} \|A\,\omega\|^{\frac{p-2}{p-1}}. \end{aligned}$$

where $C := (|p-2| + C_2)(1+\tilde{c})^{p-2}$ for $p \geq 2$ and $C := (|p-2| + C_2)(1-\tilde{c})^{p-2}$ for $1 < p < 2$. The continuation with the same arguments as in the previous lemma yields the assertions. ■

The application of the estimate for $D_Y(y^\delta - A(x^\dagger - \gamma\omega), \gamma\,A\,\omega)$ is the more sensitive one. Proposing a parameter choice strategy $\alpha = \alpha(\delta)$ we have first to ensure that the conditions of Lemma 3.3 are not injured. Based on the results above we are now able to present our first error bound result.

Lemma 3.4 *Assume (A1)-(A8) and $0 \neq \xi^\dagger \in \mathcal{M}(R,d)$ for some $R, d \geq 0$. Moreover, if $p \neq 2$, there exist two constants $0 < c, \tilde{c} < 1$ such that $d \leq c\|\xi^\dagger\|$ and the regularization parameter α is chosen such that $\delta^{p-1} \leq \tilde{c}\alpha\|\xi^\dagger\|[(1-c)\|A\|]^{-1}$.*

(i) If $1 < p < 2$, then

$$D_P(x_\alpha^\delta, x^\dagger) \leq C_3 \alpha^{\frac{2}{p-1}} \|\omega\|^{\frac{2}{p-1}} + \tilde{C}_4 \delta^2 \alpha^{\frac{1}{1-p}} + d\left(K + C_5 \alpha^{\frac{1}{p-1}} R^{\frac{1}{p-1}}\right).$$

(ii) If $p = 2$, then

$$D_P(x_\alpha^\delta, x^\dagger) \leq C_3 \alpha^2 R^2 + C_4 \frac{\delta^2}{\alpha} + d\left(K + \alpha\,R\right).$$

(iii) If $p > 2$, then

$$D_P(x_\alpha^\delta, x^\dagger) \leq \tilde{C}_3 \alpha^{\frac{2}{p-1}} \|\omega\|^2 + C_4 \delta^2 \alpha^{\frac{1}{1-p}} \|\omega\|^{\frac{p-2}{p-1}} + d\left(K + \tilde{C}_5 \alpha^{\frac{1}{p-1}}\right).$$

PROOF. First, for $p \neq 2$, we note,

$$\gamma\|A\,\omega\| = \alpha^{\frac{1}{p-1}} \|A\,\omega\|^{\frac{2-p}{p-1}+1} = \alpha^{\frac{1}{p-1}} \|A\,\omega\|^{\frac{1}{p-1}} \geq \alpha^{\frac{1}{p-1}} \left(\frac{\|\xi^\dagger\|}{(1-c)\|A\|}\right)^{\frac{1}{p-1}}.$$

This allows us the application of Lemma 3.3. Then the proof is an immediate consequence of the previous lemmas by noticing

$$\gamma := \alpha^{\frac{1}{p-1}} \|A\,\omega\|^{\frac{2-p}{p-1}} \leq \begin{cases} \alpha^{\frac{1}{p-1}} \|A\|^{\frac{2-p}{p-1}} R^{\frac{2-p}{p-1}}, & p < 2, \\ \alpha^{\frac{1}{p-1}} \left(\dfrac{\|\xi^\dagger\|}{(1-c)\|A\|} \right)^{\frac{2-p}{p-1}} & p > 2, \end{cases}$$

and introducing $C_5 := \|A\|^{\frac{2-p}{p-1}}$ and $\tilde{C}_5 := \left(\frac{\|\xi^\dagger\|}{(1-c)\|A\|} \right)^{\frac{2-p}{p-1}}$. ■

We present a first convergence rate result, provided the element $\xi^\dagger \in \mathcal{X}^*$ satisfies the source condition (9).

Theorem 3.1 *Assume (A1)-(A8) and $\xi^\dagger \in \mathcal{X}^*$ satisfies the source condition (9) for some $R > 0$. Then, an a-priori choice $\alpha := \delta^{\frac{2(p-1)}{3}}$ provides an error estimate*

$$D_P(x_\alpha^\delta, x^\dagger) \leq C\,\delta^{\frac{4}{3}}. \tag{12}$$

for some constant $C > 0$ which does not depend on δ and α.

Proof. We set $\upsilon := 0$. Balancing

$$\delta^2 \alpha^{\frac{1}{1-p}} = \alpha^{\frac{2}{p-1}} \;\Leftrightarrow\; \delta^2 = \alpha^{\frac{3}{p-1}}$$

we choose $\alpha := \delta^{\frac{2(p-1)}{3}}$ which proves (12). ■

We now deal with the case $p = 1$, which plays a singular role. Here, we have

$$\begin{aligned} \|A\,x_\alpha^\delta - y^\delta\| + \alpha D_P(x_\alpha^\delta, x^\dagger) &\leq \gamma\|A\,\omega\| + \frac{1}{\|A\,\omega\|}\langle J_Y(A\,\omega), y^\delta - A\,x^\dagger\rangle \\ &\quad + D_Y(y^\delta - A(x^\dagger - \gamma\,\omega), \gamma A\,\omega) + \alpha D_P(x^\dagger - \gamma\,\omega, x^\dagger) \\ &\quad + \alpha\langle J_Y(A\,\omega), A\,x^\dagger - A\,x_\alpha^\delta\rangle - \alpha\gamma\|A\,\omega\|^2 \\ &\quad + \alpha\|\upsilon\|\,(K + \gamma\|\omega\|) \\ &\leq \gamma\|A\,\omega\|(1 - \alpha\|A\,\omega\|) + D_Y(y^\delta - A(x^\dagger - \gamma\,\omega), \gamma A\,\omega) \\ &\quad + \alpha D_P(x^\dagger - \gamma\,\omega, x^\dagger) + \alpha\|A\,\omega\|\|A\,x_\alpha^\delta - y^\delta\|^2 \\ &\quad + \left(\|A\,\omega\|^{-1} - \alpha\right)\langle J_Y(A\,\omega), y^\delta - A\,x^\dagger\rangle \\ &\quad + \alpha\|\upsilon\|\,(K + \gamma\|\omega\|)\,. \end{aligned}$$

We present the following convergence rate result.

Theorem 3.2 *Assume (A1)-(A8) and $\xi^\dagger \in \mathcal{X}^*$ satisfies the source condition (9). If the regularization parameter $\alpha > 0$ is chosen such that $0 \leq 1 - \alpha\|A\,\omega\| \leq \delta^{\frac{2}{3}}$, then*

$$D_P(x_\alpha^\delta, x^\dagger) \sim \mathcal{O}\left(\delta^{\frac{4}{3}}\right).$$

Proof. We assume δ to be sufficiently small, i.e. $\delta < \min\{1, \|A\,\omega\|^3\}$. We set $\gamma := \delta^{\frac{2}{3}}$. Then

$$D_P(x^\dagger - \gamma\omega, x^\dagger) \leq C_1 \delta^{\frac{4}{3}} \|\omega\|^2$$

holds. Moreover, since $\delta < \|A\,\omega\|^3$ we have $\delta < \delta^{\frac{2}{3}}\|A\,\omega\| = \gamma\|A\,\omega\|$ which allows us to apply Lemma 3.3 . Hence, we derive

$$D_Y(y^\delta - A(x^\dagger - \gamma\omega), \gamma\, A\,\omega) \le C_2 \frac{\delta^2}{\gamma\|A\,\omega\|} = \frac{C_2}{\|A\,\omega\|}\delta^{\frac{4}{3}}.$$

This gives

$$\begin{aligned} D_P(x_\alpha^\delta, x^\dagger) &\le \gamma\frac{\|A\,\omega\|(1-\alpha\|A\,\omega\|)}{\alpha} + \delta(1-\alpha\|A\,\omega\|) + C_1\delta^{\frac{4}{3}}\|\omega\|^2 + \frac{C_2}{\alpha\|A\,\omega\|}\delta^{\frac{4}{3}} \\ &\le \delta^{\frac{4}{3}}\left(\frac{\|A\,\omega\|}{\alpha} + C_1\|\omega\|^2 + \frac{C_2}{\alpha\|A\,\omega\|}\right) + \delta^{\frac{5}{3}}\frac{1}{\alpha}, \end{aligned}$$

which proves the lemma. We note by the smallness of δ, that $\alpha > (1-\delta^{\frac{2}{3}})\|A\,\omega\|^{-1} > 0$ and $1 \ge \alpha\|A\,\omega\| \ge 1-\delta^{\frac{2}{3}} > 0$. ■

Note, that the choice of the regularization parameter α depends essentially on $\|A\,\omega\|$ which might be rather unusual. However, the same effect was already observed in [2] for deriving a weaker convergence rate. Moreover, this dependency is the reason, that presenting convergence rates in terms of approximate source conditions does not seems to be promising for the case $p = 1$. For a given choice $R = R(\alpha)$, $\|\omega\| \le R$ we have to ensure the condition $0 \le 1 - \alpha\|A\,\omega\| \le \delta^{\frac{2}{3}}$ which does not seem to be possible under general conditions.

4 Convergence rates

For presenting convergence rates in terms of approximative source conditions we apply the idea of distance functions, see e.g. [7].

Definition 4.1 *For given $\xi \in \mathcal{X}^*$ the distance function $d(\cdot;\xi) : [0,\infty) \longrightarrow \mathbb{R}$ is defined as*

$$d(R;\xi) := \inf\left\{\|\xi - A^\star J_Y(A\,\omega)\| \;:\; \omega \in \mathcal{Y}^*,\ \|\omega\| \le R\right\}, \qquad R \ge 0.$$

Additionally, we introduce the set

$$\mathcal{M} := \{\xi \in \mathcal{X}^* \;:\; \xi = A^\star J_Y(A\,\omega),\ \omega \in \mathcal{X}\}$$

of all element satisfying the source condition (9) for some $\omega \in \mathcal{X}$. Note, that the nonnegative function $d(R,\xi)$ is well-defined for each $\xi \in \mathcal{X}^*$. Moreover, by [20, Theorem 38.A], for each $R \ge 0$ there exists an element $\omega = \omega(R)$ with $d(R;\xi) = \|\xi - A^\star J_Y(A\,\omega)\|$ and $\|\omega\| \le R$. The distance functions are non-increasing with $d(R;\xi) \to 0$ for $R \to \infty$ if $\xi \in \overline{\mathcal{M}}$. We have $d(R;\xi) > 0$ for all $R \ge 0$ if $\xi \notin \mathcal{M}$ and $d(R;\xi) = 0$ for all $R \ge \|\omega\|$ if $\xi = A^\star J_Y(A\,\omega)$ and $\|\omega\| = R$. Altogether, the distance functions $d(R;\xi)$ gives us a quantity for measuring the violation of the source condition (9).

We now are able to present convergence rates, if the source condition (9) is not satisfied, but $\xi^\dagger \in \overline{\mathcal{M}}$. This additional assumption seems to be very restrictive at first view. On the other hand, the following statement yields.

Lemma 4.1 *Assume the operators A and $A^\star$ to be injective. Then $\overline{\mathcal{M}} = \mathcal{X}^*$.*

PROOF. By [19, Satz III.4.5] we have $\overline{\mathcal{R}(A^\star)} = \mathcal{X}^*$ and $\overline{\mathcal{R}(A)} = \mathcal{Y}$. Assume $\xi \notin \mathcal{M}$. On the other hand, for each $\varepsilon > 0$ there exists an element $\tilde{y}_\varepsilon \in \mathcal{Y}^*$ with $\|\xi - A^\star y_\varepsilon\| < \varepsilon$. Since J_Y is bijective, there exists $y_\varepsilon \in \mathcal{Y}$ with $J_Y(y_\varepsilon) = \tilde{y}_\varepsilon$. Finally, there exists $\omega_\varepsilon \in \mathcal{X}$ with $\|y_\varepsilon - A\,\omega_\varepsilon\| < \varepsilon$. Hence,

$$
\begin{aligned}
\|\xi - A^\star J_Y(A\,\omega_\varepsilon)\| &\leq \|\xi - A^\star J_Y(y_\varepsilon)\| + \|A^\star\|\,\|J_Y(A\,\omega_\varepsilon) - J_Y(y_\varepsilon)\| \\
&< \varepsilon + \|A^\star\|\,\|J_Y(A\,\omega_\varepsilon) - J_Y(y_\varepsilon)\|
\end{aligned}
$$

Since J_Y is continuously we conclude $\|J_Y(A\,\omega_\varepsilon) - J_Y(y_\varepsilon)\| \to 0$ for $\varepsilon \to 0$ which implies $A^\star J_Y(\omega_\varepsilon) \to \xi$ for $\varepsilon \to 0$. Consequently, $\xi \in \overline{\mathcal{M}}$ yields. ■

We now present convergence rates results proposing an appropriate a-priori parameter choice of the regularization parameter $\alpha = \alpha(\delta)$. We start with the case $p = 2$.

Theorem 4.1 *Assume (A1)-(A8), $\xi^\dagger \in \overline{\mathcal{M}} \setminus \mathcal{M}$ has distance function $d(R) := d(R;\xi^\dagger)$. Let $p = 2$, $\Psi_2(R) := d(R)^{\frac{1}{2}}R^{-1}$, $\Theta_2(\alpha) := \left(\alpha\, d\left(\Psi_2^{-1}(\alpha)\right)\right)^{\frac{1}{2}}$ and $\Phi(R) := d(R)^{\frac{3}{4}}R^{-\frac{1}{2}}$. Then, the a-priori choice $\alpha := \Theta_2^{-1}(\delta)$ yields the convergence rate*

$$
D_P(x_\alpha^\delta, x^\dagger) \sim \mathcal{O}\left(d\left(\Phi^{-1}(\delta)\right)\right).
$$

PROOF. By definition, $\xi^\dagger \in \mathcal{M}(R, d(R))$ for all $R \geq 0$. Hence, there exist $\omega = \omega(R)$ and $\upsilon = \upsilon(R)$ with $\|\omega\| \leq R$ and $\|\upsilon(R)\| \leq d(R)$ such that $\xi^\dagger = A^\star\omega(R) + \upsilon(R)$. In particular, Lemma 3.1 holds for all $R \geq 0$. We consider only the first three terms, since the term $\gamma d(R)R$ decays faster to zero than $d(R)$. By balancing

$$
\alpha^2 R^2 = d(R) \;\Leftrightarrow\; \alpha = \frac{d(R)^{\frac{1}{2}}}{R} =: \Psi_2(R) \;\Leftrightarrow\; R = \Psi_2^{-1}(\alpha).
$$

Moreover,

$$
\delta^2\alpha^{-1} = d(R) \;\Leftrightarrow\; \delta = \left(\alpha\, d\left(\Psi_2^{-1}(\alpha)\right)\right)^{\frac{1}{2}} =: \Theta_2(\alpha).
$$

Hence $\alpha := \Theta_2^{-1}(\delta)$ is the optimal parameter choice. Finally

$$
d(R) = \delta^2\alpha^{-1} = \delta^2\frac{R}{d(R)^{\frac{1}{2}}} \;\Leftrightarrow\; \delta = \frac{d(R)^{\frac{3}{4}}}{R^{\frac{1}{2}}} =: \Phi(R)
$$

which provides the convergence rate $D_P(x_\alpha^\delta, x^\dagger) \sim \mathcal{O}\left(d\left(\Phi^{-1}(\delta)\right)\right)$. ■

For $p \neq 2$ we need an additional restriction to the decay of the distance function $d(R)$. In particular, in order to apply Lemma 3.3, we have to suppose, that the decay $d(R) \to 0$ for $R \to \infty$ is sufficiently fast. We first consider the case $1 < p < 2$.

Theorem 4.2 *Assume (A1)-(A8), $0 \neq \xi^\dagger \in \overline{\mathcal{M}} \setminus \mathcal{M}$ has distance function $d(R) := d(R;\xi^\dagger)$ with $d(R)R^{\frac{2}{p-1}} \to 0$ for $R \to \infty$. Let $1 < p < 2$, $\Psi_p(R) := d(R)^{\frac{p-1}{2}}R^{-1}$, $\Theta_p(\alpha) := \left(\alpha^{\frac{1}{p-1}} d\left(\Psi_p^{-1}(\alpha)\right)\right)^{\frac{1}{2}}$ and $\Phi_p(R) := d(R)^{\frac{3}{4}}R^{-\frac{1}{2(p-1)}}$. Then, the a-priori choice $\alpha := \Theta_p^{-1}(\delta)$ yields the convergence rate*

$$
D_P(x_\alpha^\delta, x^\dagger) \sim \mathcal{O}\left(d\left(\Phi_p^{-1}(\delta)\right)\right).
$$

PROOF. First, we assume that the condition on the choice of the parameter α of Lemma 3.3 is satisfied. Then we choose $R = R(\alpha)$ such that

$$d(R) = \alpha^{\frac{2}{p-1}} R^{\frac{2}{p-1}} \ \Leftrightarrow \ \alpha = \frac{d(R)^{\frac{p-1}{2}}}{R} =: \Psi_p(R)$$

or $R := \Psi_p^{-1}(\alpha)$. Moreover,

$$\delta^2 \alpha^{\frac{1}{1-p}} = d(R) = d\left(\Psi_p^{-1}(\alpha)\right),$$

which gives

$$\delta = \left(d\left(\Psi_p^{-1}(\alpha)\right) \alpha^{\frac{1}{p-1}}\right)^{\frac{1}{2}} =: \Theta_p(\alpha).$$

Hence $\alpha := \Theta_p^{-1}(\delta)$ is the optimal parameter choice. Moreover,

$$\begin{aligned} d(R) = \delta^2 \alpha^{\frac{1}{p-1}} &= \delta^2 \frac{d(R)^{\frac{p-1}{2(1-p)}}}{R^{\frac{1}{p-1}}} \\ &= \delta^2 R^{\frac{1}{p-1}} d(R)^{-\frac{1}{2}} \end{aligned}$$

which provides

$$\delta = \frac{d(R)^{\frac{3}{4}}}{R^{\frac{1}{2(p-1)}}} =: \Phi_p(R)$$

Hence $d\left(\Phi_p^{-1}(\delta)\right)$ describes the convergence rate. Finally we note, that

$$\frac{\delta}{\alpha^{\frac{1}{p-1}}} = \frac{d(R)^{\frac{3}{4}}}{\alpha^{\frac{1}{p-1}} R^{\frac{1}{2(p-1)}}} = \frac{d(R)^{\frac{3}{4}-\frac{1}{2}}}{R^{\frac{1}{2(p-1)}-\frac{1}{p-1}}} = d(R)^{\frac{1}{4}} R^{\frac{1}{2(p-1)}} \to 0$$

for $\delta \to 0$ which allows us the application of Lemma 3.3 if δ is sufficiently small. ■

For $p > 2$ we can achieve the following result by similar calculations.

Theorem 4.3 *Assume (A1)-(A8), $0 \neq \xi^\dagger \in \overline{\mathcal{M}} \setminus \mathcal{M}$ has distance function $d(R) := d(R;\xi^\dagger)$ with $d(R)R^{\frac{2}{p-1}} \to 0$ for $R \to \infty$. Let $p > 2$, $\Psi_p(R) := d(R)^{\frac{p-1}{2}} R^{1-p}$, $\Theta_p(\alpha) := \alpha^{\frac{2p-3}{2(p-1)^2}} d\left(\Psi_p^{-1}(\alpha)\right)^{\frac{p}{4(p-1)}}$ and $\Phi_p(R) := d(R)^{\frac{3}{4}} R^{-\frac{2p-3}{2(p-1)}}$. Then, the a-priori choice $\alpha := \Theta_p^{-1}(\delta)$ yields the convergence rate*

$$D_P(x_\alpha^\delta, x^\dagger) \sim \mathcal{O}\left(d\left(\Phi_p^{-1}(\delta)\right)\right).$$

PROOF. Here, by the same arguments as above, we have

$$d(R) = \alpha^{\frac{2}{p-1}} R^2 \ \Leftrightarrow \ \alpha = \frac{d(R)^{\frac{p-1}{2}}}{R^{p-1}} =: \Psi_p(R) \ \Leftrightarrow \ R = \Psi_p^{-1}(\alpha).$$

Moreover,

$$\begin{aligned} \delta^2 \alpha^{\frac{1}{1-p}} R^{\frac{p-2}{p-1}} &= \delta^2 \alpha^{\frac{1}{1-p}} \left(\frac{d(R)^{\frac{1}{2}}}{\alpha^{\frac{1}{p-1}}}\right)^{\frac{p-2}{p-1}} \\ &= \delta^2 d(R)^{\frac{p-2}{2(p-1)}} \alpha^{\frac{1}{1-p}-\frac{p-2}{(p-1)^2}} \\ &= \delta^2 d(R)^{\frac{p-2}{2(p-1)}} \alpha^{-\frac{2p-3}{(p-1)^2}} \end{aligned}$$

which implies

$$\delta = \left(d(R)^{1-\frac{p-2}{2(p-1)}}\alpha^{\frac{2p-3}{(p-1)^2}}\right)^{\frac{1}{2}} = \left(d(R)^{\frac{p}{2(p-1)}}\alpha^{\frac{2p-3}{(p-1)^2}}\right)^{\frac{1}{2}} =: \Theta_p(\alpha)$$

which provides the parameter choice $\alpha := \Theta_p^{-1}(\delta)$. Hence

$$d(R) = \delta^2\alpha^{\frac{1}{1-p}}R^{\frac{p-2}{p-1}} = \delta^2\frac{R}{d(R)^{\frac{1}{2}}}R^{\frac{p-2}{p-1}} = \delta^2 d(R)^{-\frac{1}{2}}R^{\frac{2p-3}{p-1}}$$

or

$$\delta = \frac{d(R)^{\frac{3}{4}}}{R^{\frac{2p-3}{2(p-1)}}} =: \Phi_p(R)$$

which describes the convergence rate $d\left(\Phi_p^{-1}(\delta)\right)$ again. Moreover,

$$\frac{\delta}{\alpha^{\frac{1}{p-1}}} = \frac{d(R)^{\frac{3}{4}}}{\alpha^{\frac{1}{p-1}}R^{\frac{2p-3}{2(p-1)}}} = \frac{d(R)^{\frac{3}{4}-\frac{1}{2}}}{R^{\frac{2p-3}{2(p-1)}-1}} = d(R)^{\frac{1}{4}}R^{\frac{1}{2(p-1)}} \to 0$$

for $\delta \to 0$ which allows us again the application of Lemma 3.3 if δ is sufficiently small. ■

Note, that the achieved convergence rates depend on the choice of the parameter p. In particular, for $p = 2$ the fastest rate of convergence could be achieved. However, this seems to have technical reasons. We can find an upper bound for the term $\|A\,\omega\|$ depending on $\|\omega\| \geq 0$ but there is no lower bound for $\|A\,\omega\|$ which depends on the choice of ω. As we will see in further considerations, an additional lower bound on $\|A\,\omega\|$ with respect to $\|\omega\|$ will lead to convergence rates which do not depend on the parameter p anymore.

5 Improved convergence rates

For violated source condition (9) the bounds seems to be not of optimal order. Assume $\mathcal{X}$ to be a Hilbert space and $P(x) := \frac{1}{2}\|x\|^2$. Then $\xi^\dagger = 2x^\dagger$ and $D_P(x_\alpha^\delta, x^\dagger) = \frac{1}{2}\|x_\alpha^\delta - x^\dagger\|^2$. Hence, e.g. for $p = 2$, the estimate

$$\begin{aligned}\frac{1}{2}\|x_\alpha^\delta - x^\dagger\|^2 &\leq C_3\alpha^2R^2 + C_4\frac{\delta^2}{\alpha} + d\|x_\alpha^\delta - x^\dagger\| + d\,\alpha\,R \\ &\leq \left(C_3 + \frac{1}{2}\right)\alpha^2R^2 + C_4\frac{\delta^2}{\alpha} + \frac{d^2}{2} + d\|x_\alpha^\delta - x^\dagger\|\end{aligned}$$

holds. By the implication

$$a, b, c > 0,\ a^2 \leq b^2 + ac \ \Rightarrow\ a \leq b + c \tag{13}$$

we conclude

$$\begin{aligned}\|x_\alpha^\delta - x^\dagger\| &\leq \sqrt{2}\left(\sqrt{C_3 + \frac{1}{2}}\,\alpha\,R + \sqrt{C_4}\frac{\delta}{\sqrt{\alpha}} + d\left(\sqrt{2} + \frac{1}{\sqrt{2}}\right)\right) \\ &= \sqrt{2C_3 + 1}\,\alpha\,R + \sqrt{2\,C_4}\frac{\delta}{\sqrt{\alpha}} + 3\,d.\end{aligned}$$

The improved result will provide better convergence rates. For applying the idea we need an additional convexity condition on the stabilizing functional $P(x)$.

(A9) The functional $P(x)$ is strongly convex in $x^\dagger$, i.e. there exists a constant $\eta > 0$ such that

$$D_P(x, x^\dagger) = P(x) - P(x^\dagger) - \langle \xi^\dagger, x - x^\dagger \rangle \geq \eta \|x - x^\dagger\|^2$$

for all $x \in \mathcal{B}_\rho(x^\dagger)$ with radius $\rho \geq K > 0$ sufficiently large.

First we present an error bound result.

Lemma 5.1 *Assume (A1)-(A7) and (A9) and $\xi^\dagger$ satisfies (10) for some $\omega \in \mathcal{X}$ and $\upsilon \in \mathcal{X}^*$. If $p > 1$, $\gamma := \alpha^{\frac{1}{p-1}} \|A\,\omega\|^{\frac{2-p}{p-1}}$ then*

$$\|x_\alpha^\delta - x^\dagger\| \leq C_6\, \gamma\, \|\omega\| + \sqrt{\frac{D_Y(y^\delta - A(x^\dagger - \gamma\omega), \gamma A\,\omega)}{\eta\, \alpha}} + C_7 \|\upsilon\|$$

for two positive constants $C_6, C_7 > 0$.

Proof. From Lemma 3.1 we conclude

$$\begin{aligned}
\eta \|x_\alpha^\delta - x^\dagger\|^2 &\leq D_P(x_\alpha^\delta, x^\dagger) \\
&\leq \left(C_1 + \frac{1}{2}\right) \gamma^2 \|\omega\|^2 + \frac{1}{\alpha} D_Y(y^\delta - A(x^\dagger - \gamma\omega), \gamma A\,\omega) \\
&\quad + \frac{\|\upsilon\|^2}{2} + \|\upsilon\|\, \|x_\alpha^\delta - x^\dagger\|.
\end{aligned}$$

By the implication (13) we conclude

$$\sqrt{\eta} \|x_\alpha^\delta - x^\dagger\| \leq \sqrt{C_1 + \frac{1}{2}}\, \gamma \|\omega\| + \sqrt{\frac{D_Y(y^\delta - A(x^\dagger - \gamma\omega), \gamma A\,\omega)}{\alpha}} + \|\upsilon\| \left(\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{\eta}}\right),$$

which shows the lemma with $C_6 := \sqrt{\eta^{-1}(C_1 + 2^{-1})}$ and $C_7 := (2\eta)^{-\frac{1}{2}} + \eta^{-1}$. ∎

We present the improved error bounds.

Lemma 5.2 *Assume (A1)-(A9) and $0 \neq \xi^\dagger \in \mathcal{M}(R, d)$ for some $R, d \geq 0$. Moreover, if $p \neq 2$, there exist two constants $0 < c, \tilde{c} < 1$ such that $d < c\|\xi^\dagger\|$ and the regularization parameter α is chosen such that $\delta^{p-1} \leq \tilde{c}\alpha \|\xi^\dagger\| [(1-c)\|A\|]^{-1}$.*

(i) If $1 < p < 2$, then

$$\|x_\alpha^\delta - x^\dagger\| \leq C_8 \alpha^{\frac{1}{p-1}} \|\omega\|^{\frac{1}{p-1}} + \tilde{C}_8 \delta \alpha^{\frac{1}{2(1-p)}} + C_7 \|\upsilon\|.$$

(ii) If $p = 2$, then

$$\|x_\alpha^\delta - x^\dagger\| \leq C_8 \alpha \|\omega\| + C_9 \frac{\delta}{\sqrt{\alpha}} + C_7 \|\upsilon\|.$$

(iii) If $p > 2$, then

$$\|x_\alpha^\delta - x^\dagger\| \leq \tilde{C}_8 \alpha^{\frac{1}{p-1}} \|\omega\| + C_9 \delta \alpha^{\frac{1}{2(1-p)}} \|\omega\|^{\frac{p-2}{2(p-1)}} + C_7 \|\upsilon\|.$$

Here, the constants $C_8, \tilde{C}_8, C_9$ and $\tilde{C}_9$ do not depend on R, d, α and δ.

The proof is essentially the same as in the previous section.

We now present improved convergence rates. Again, we have to distinguish between the cases $p < 2$, $p = 2$ and $p > 2$. We first deal with $p = 2$.

Theorem 5.1 *Assume (A1)-(A9), $\xi^\dagger \in \overline{\mathcal{R}(A^\star)} \setminus \mathcal{R}(A^\star)$ has distance function $d(R) := d(R;\xi^\dagger)$. Let $p = 2$, $\tilde{\Psi}_2(R) := d(R)R^{-1}$, $\tilde{\Theta}_2(\alpha) := \alpha^{\frac{1}{2}} d\left(\Psi_2^{-1}(\alpha)\right)$ and $\tilde{\Phi}(R) := d(R)^{\frac{3}{2}} R^{-\frac{1}{2}}$. Then, the a-priori choice $\alpha := \tilde{\Theta}_2^{-1}(\delta)$ yields the convergence rate*

$$\|x_\alpha^\delta - x^\dagger\| \sim \mathcal{O}\left(d\left(\tilde{\Phi}^{-1}(\delta)\right)\right).$$

Proof. By balancing the estimate of Lemma 5.1

$$\alpha\, R = d(R) \;\Leftrightarrow\; \alpha = \frac{d(R)}{R} =: \tilde{\Psi}_2(R) \;\Leftrightarrow\; R = \tilde{\Psi}_2^{-1}(\alpha).$$

Moreover,

$$\delta\alpha^{-\frac{1}{2}} = d(R) \;\Leftrightarrow\; \delta = \alpha^{\frac{1}{2}} d\left(\tilde{\Psi}_2^{-1}(\alpha)\right) =: \Theta_2(\alpha).$$

Hence $\alpha := \Theta_2^{-1}(\delta)$ is the optimal parameter choice. Finally

$$d(R) = \delta\alpha^{-\frac{1}{2}} = \delta\frac{R^{\frac{1}{2}}}{d(R)^{\frac{1}{2}}} \;\Leftrightarrow\; \delta = \frac{d(R)^{\frac{3}{2}}}{R^{\frac{1}{2}}} =: \tilde{\Phi}(R)$$

which provides the convergence rate $\|x_\alpha^\delta - x^\dagger\| \sim \mathcal{O}\left(d\left(\tilde{\Phi}^{-1}(\delta)\right)\right)$. ■

For $p \neq 2$ we obtain similar results by assuming a sufficient decay rate of the distance function $d(R)$. We present the corresponding rates in the following two statements.

Theorem 5.2 *Assume (A1)-(A9), $0 \neq \xi^\dagger \in \overline{\mathcal{M}} \setminus \mathcal{M}$ has distance function $d(R) := d(R;\xi^\dagger)$ with $d(R)R^{\frac{1}{p-1}} \to 0$ for $R \to \infty$. Let $1 < p < 2$, $\tilde{\Psi}_p(R) := d(R)^{p-1}R^{-1}$, $\tilde{\Theta}_p(\alpha) := \alpha^{\frac{1}{2(p-1)}} d\left(\Psi_p^{-1}(\alpha)\right)$ and $\tilde{\Phi}_p(R) := d(R)^{\frac{3}{2}} R^{-\frac{1}{2(p-1)}}$. Then, the a-priori choice $\alpha := \Theta_p^{-1}(\delta)$ yields the convergence rate*

$$\|x_\alpha^\delta - x^\dagger\| \sim \mathcal{O}\left(d\left(\tilde{\Phi}_p^{-1}(\delta)\right)\right).$$

Proof. Here,

$$d(R) = \alpha^{\frac{1}{p-1}} R^{\frac{1}{p-1}} \;\Leftrightarrow\; \alpha = \frac{d(R)^{p-1}}{R} =: \tilde{\Psi}_p(R) \;\Leftrightarrow\; R := \tilde{\Psi}_p^{-1}(\alpha).$$

Moreover

$$\delta = \alpha^{\frac{1}{2(p-1)}} d\left(\tilde{\Psi}_p^{-1}(\alpha)\right) =: \tilde{\Theta}_p(\alpha) \;\Leftrightarrow\; \alpha := \tilde{\Theta}_p^{-1}(\delta)$$

is the optimal parameter choice. On the other hand

$$d(R) = \delta\alpha^{\frac{1}{2(1-p)}} = \delta\frac{d(R)^{\frac{p-1}{2(1-p)}}}{R^{\frac{1}{2(1-p)}}} = \delta\, d(R)^{-\frac{1}{2}} R^{\frac{1}{2(p-1)}}$$

which gives the improved convergence rate

$$\delta = \frac{d(R)^{\frac{3}{2}}}{R^{\frac{1}{2(p-1)}}} =: \tilde{\Phi}_p(R).$$

or $\|x_\alpha^\delta - x^\dagger\| \sim d\left(\tilde{\Phi}_p^{-1}(\delta)\right)$. In order to apply Lemma 3.3 we consider

$$\frac{\delta}{\alpha^{\frac{1}{p-1}}} = \frac{d(R)^{\frac{3}{2}}}{\alpha^{\frac{1}{p-1}} R^{\frac{1}{2(p-1)}}} = \frac{d(R)^{\frac{3}{2}-1}}{R^{\frac{1}{2(p-1)}-\frac{1}{p-1}}} = d(R)R^{\frac{1}{p-1}}$$

which provides us the decay of the left hand side. ■

Theorem 5.3 *Assume (A1)-(A9), $0 \neq \xi^\dagger \in \overline{\mathcal{M}} \setminus \mathcal{M}$ has distance function $d(R) := d(R;\xi^\dagger)$ with $d(R)R^{\frac{1}{p-1}} \to 0$ for $R \to \infty$. Let $p > 2$, $\tilde{\Psi}_p(R) := d(R)^{p-1}R^{1-p}$, $\tilde{\Theta}_p(\alpha) := \alpha^{\frac{2p-3}{2(p-1)^2}} d\left(\tilde{\Psi}_p^{-1}(\alpha)\right)^{\frac{p}{2(p-1)}}$ and $\tilde{\Phi}_p(R) := d(R)^{\frac{3}{2}}R^{-\frac{2p-3}{2(p-1)}}$. Then, the a-priori choice $\alpha := \tilde{\Theta}_p^{-1}(\delta)$ yields the convergence rate*

$$\|x_\alpha^\delta - x^\dagger\| \sim \mathcal{O}\left(d\left(\tilde{\Phi}_p^{-1}(\delta)\right)\right).$$

Proof. First,

$$d(R) = \alpha^{\frac{1}{p-1}} R \;\Leftrightarrow\; \alpha = \frac{d(R)^{p-1}}{R^{p-1}} =: \tilde{\Psi}_p(R) \;\Leftrightarrow\; R := \Psi_p^{-1}(\alpha).$$

Then,

$$\begin{aligned} d(R) = \delta\alpha^{\frac{1}{2(1-p)}} R^{\frac{p-2}{2(p-1)}} &= \delta\alpha^{\frac{1}{2(1-p)}} \left(\frac{d(R)}{\alpha^{\frac{1}{p-1}}}\right)^{\frac{p-2}{2(p-1)}} \\ &= \delta\, d(R)^{\frac{p-2}{2(p-1)}} \alpha^{\frac{1}{2(1-p)} - \frac{p-2}{2(p-1)^2}} \\ &= \delta\, d(R)^{\frac{p-2}{2(p-1)}} \alpha^{-\frac{2p-3}{2(p-1)^2}} \end{aligned}$$

which implies with

$$\delta = d(R)^{1-\frac{p-2}{2(p-1)}} \alpha^{\frac{2p-3}{2(p-1)^2}} = d(R)^{\frac{p}{2(p-1)}} \alpha^{\frac{2p-3}{2(p-1)^2}} =: \tilde{\Theta}_p(\delta)$$

the parameter choice $\alpha := \tilde{\Theta}_p^{-1}(\alpha)$. On the other hand,

$$d(R) = \delta\alpha^{\frac{1}{2(1-p)}} R^{\frac{p-2}{2(p-1)}} = \delta\frac{R^{\frac{1}{2}}}{d(R)^{\frac{1}{2}}} R^{\frac{p-2}{2(p-1)}} = \delta\, d(R)^{-\frac{1}{2}} R^{\frac{2p-3}{2(p-1)}}$$

implies

$$\delta = \frac{d(R)^{\frac{3}{2}}}{R^{\frac{2p-3}{2(p-1)}}} =: \tilde{\Phi}_p(R).$$

Here we have

$$\frac{\delta}{\alpha^{\frac{1}{p-1}}} = \frac{d(R)^{\frac{3}{2}}}{\alpha^{\frac{1}{p-1}} R^{\frac{2p-3}{2(p-1)}}} = \frac{d(R)^{\frac{3}{2}-1}}{R^{\frac{2p-3}{2(p-1)}-1}} = d(R)R^{\frac{1}{p-1}}$$

which describes the necessary decay rate of the distance function $d(R)$. ■

6 Optimal rates

In the following section we will show that we can achieve convergence rates which do not depend on the parameter p provided we find upper and lower bounds for the term $\|A\,\omega\|$ depending on $\|\omega\| \geq 0$. Therefore, we need a further assumption.

(A10) There exists a reflexive Banach space $\mathcal{Z} \supset \mathcal{X}$ such that

(i) The space $\mathcal{X}$ is continuously embedded in $\mathcal{Z}$.

(ii) It holds
$$\langle z, x\rangle_{\mathcal{X}^*,\mathcal{X}} = \langle z, x\rangle_{\mathcal{Z}^*,\mathcal{Z}}, \qquad \forall z \in \mathcal{Z}^*,\ \forall x \in \mathcal{X}.$$

(iii) There exist two positive constants c_1, c_2 such that
$$c_1\|\omega\|_{\mathcal{Z}} \leq \|A\,\omega\| \leq c_2\|\omega\|_{\mathcal{Z}}, \qquad \forall\,\omega \in \mathcal{X},$$

(iv) Assumption (A7) holds with $\|\cdot\|_{\mathcal{X}}$ replaced by $\|\cdot\|_{\mathcal{Z}}$.

(v) The element $\xi^\dagger$ belongs to (the smaller space) $\mathcal{Z}^*$.

Condition (iii) implies that a solution of equation (1) depends continuously on the given data in the (weaker) $\mathcal{Z}$-norm. Analogously to the previous calculations we introduce the sets

$$\mathcal{M}_{\mathcal{Z}}(R,d) := \{\xi \in \mathcal{Z}^* \,:\, \xi := A^\star J_Y(A\,\omega) + \upsilon,\ \omega \in \mathcal{X},\ \upsilon \in \mathcal{Z}^*,\ \|\omega\|_{\mathcal{Z}} \leq R,\ \|\upsilon\|_{\mathcal{Z}^*} \leq d\}\,.$$

and
$$\mathcal{M}_{\mathcal{Z}} := \{\xi \in \mathcal{Z}^* \,:\, \xi := A^\star J_Y(A\,\omega),\, \omega \in \mathcal{X}\}\,.$$

Then we can achieve convergence rates which do not depend on p anymore, when we define the distance function in the space $\mathcal{Z}^*$. Hence, for $\xi \in \mathcal{Z}^*$ we introduce

$$d_{\mathcal{Z}}(R;\xi) := \inf\{\|\xi - A^\star J_Y(A\,\omega)\|_{\mathcal{Z}^*} \,:\, \omega \in \mathcal{X},\ \|\omega\|_{\mathcal{Z}} \leq R\}\,.$$

First we present an error bound result which is an immediate consequence of the previous calculations.

Lemma 6.1 *Assume (A1)-(A6), (A8), (A10) and $0 \neq \xi^\dagger \in \mathcal{Z}^*$ satisfies (10) for some $\omega \in \mathcal{X}$ and $\upsilon \in \mathcal{X}^*$. Then*

$$D_P(x_\alpha^\delta, x^\dagger) \leq C_{10}\alpha^{\frac{2}{p-1}}\|\omega\|_{\mathcal{Z}}^{\frac{2}{p-1}} + C_{11}\delta^2\alpha^{\frac{1}{1-p}}\|\omega\|_{\mathcal{Z}}^{\frac{p-2}{p-1}} + C_{12}\|\upsilon\|_{\mathcal{Z}^*}.$$

If - additionally - (A9) holds, then

$$\|x_\alpha^\delta - x^\dagger\| \leq \tilde{C}_{10}\alpha^{\frac{1}{p-1}}\|\omega\|_{\mathcal{Z}}^{\frac{1}{p-1}} + \tilde{C}_{11}\delta\alpha^{\frac{1}{2(1-p)}}\|\omega\|_{\mathcal{Z}}^{\frac{p-2}{2(p-1)}} + \tilde{C}_{12}\|\upsilon\|_{\mathcal{Z}^*}.$$

Here, the constants $C_k, \tilde{C}_k$ do not depend on ω, υ, α and δ.

Now we present convergence rates.

Theorem 6.1 *Assume (A1)-(A6), (A8), (A10),* $0 \neq \xi^\dagger \in \overline{\mathcal{M}_\mathcal{Z}} \setminus \mathcal{M}_\mathcal{Z}$ *has distance function* $d_\mathcal{Z}(R) := d_\mathcal{Z}(R;\xi^\dagger)$, $\Psi_p(R) := d_\mathcal{Z}(R)^{\frac{p-1}{2}} R^{-1}$, $\hat{\Theta}_p(\alpha) := d_\mathcal{Z}\left(\Psi_p^{-1}(\alpha)\right)^{\frac{4-p}{4}} \alpha^{\frac{p}{2(p-1)}}$ *and* $\Phi(R) := d_\mathcal{Z}(R)^{\frac{3}{4}} R^{-\frac{1}{2}}$*. Then, the a-priori choice* $\alpha := \hat{\Theta}_p^{-1}(\delta)$ *yields the convergence rate*

$$D_P(x_\alpha^\delta, x^\dagger) \sim \mathcal{O}\left(d_\mathcal{Z}\left(\Phi^{-1}(\delta)\right)\right).$$

Proof. We have $\xi^\dagger \in \mathcal{M}_\mathcal{Z}(R, d_\mathcal{Z}(R)+\varepsilon)$ for each $R \geq 0$ and $\varepsilon \to 0$. Taking the limit $\varepsilon \to 0$ we can apply the above error bounds. The same calculations as in the previous sections lead to

$$\alpha = d_\mathcal{Z}(R)^{\frac{p-1}{2}} R^{-1} =: \Psi_p(R) \;\Leftrightarrow\; R = \Psi_p^{-1}(\alpha).$$

Moreover,

$$\begin{aligned} d_\mathcal{Z}(R) = \delta^2 \alpha^{\frac{1}{1-p}} R^{\frac{p-2}{p-1}} &= \delta^2 \alpha^{\frac{1}{1-p}} \left(\frac{d_\mathcal{Z}(R)^{\frac{1}{2}}}{\alpha^{\frac{1}{p-1}}}\right)^{p-2} \\ &= \delta^2 d_\mathcal{Z}(R)^{\frac{p-2}{2}} \alpha^{\frac{1}{1-p} - \frac{p-2}{p-1}} \\ &= \delta^2 d_\mathcal{Z}(R)^{\frac{p-2}{2}} \alpha^{-\frac{p}{p-1}} \end{aligned}$$

which gives

$$\delta = d_\mathcal{Z}(R)^{\frac{4-p}{4}} \alpha^{\frac{p}{2(p-1)}} =: \hat{\Theta}_p(\alpha)$$

and the parameter choice $\alpha := \hat{\Theta}_p^{-1}(\delta)$. Moroeover

$$\begin{aligned} d_\mathcal{Z}(R) = \delta^2 \alpha^{\frac{1}{1-p}} R^{\frac{p-2}{p-1}} &= \delta^2 R^{\frac{p-2}{p-1}} \frac{R^{\frac{1}{p-1}}}{d_\mathcal{Z}(R)^{\frac{1}{2}}} \\ &= \delta^2 d_\mathcal{Z}(R)^{-\frac{1}{2}} R, \end{aligned}$$

which implies $\delta = R^{-\frac{1}{2}} d_\mathcal{Z}(R)^{\frac{3}{4}}$ and the corresponding convergence rate. ■

Under the additional convexity condition (A9) we can present the following improved convergence rate.

Theorem 6.2 *Assume (A1)-(A6), (A8)-(A10),* $0 \neq \xi^\dagger \in \overline{\mathcal{M}_\mathcal{Z}} \setminus \mathcal{M}_\mathcal{Z}$ *has distance function* $d_\mathcal{Z}(R) := d_\mathcal{Z}(R;\xi^\dagger)$, $\tilde{\Psi}_p(R) := d_\mathcal{Z}(R)^{p-1} R^{-1}$, $\bar{\Theta}_p(\alpha) := d_\mathcal{Z}\left(\tilde{\Psi}_p^{-1}(\alpha)\right)^{\frac{4-p}{4}} \alpha^{\frac{1}{2}}$ *and* $\bar{\Phi}(R) := d_\mathcal{Z}(R)^{\frac{3}{2}} R^{-\frac{1}{2}}$*. Then, the a-priori choice* $\alpha := \hat{\Theta}_p^{-1}(\delta)$ *yields the convergence rate*

$$\|x_\alpha^\delta - x^\dagger\| \sim \mathcal{O}\left(d_\mathcal{Z}\left(\bar{\Phi}^{-1}(\delta)\right)\right).$$

Proof. Here we have

$$d_\mathcal{Z}(R) = \alpha^{\frac{1}{p-1}} R^{\frac{1}{p-1}} \;\Leftrightarrow\; \alpha = d_\mathcal{Z}(R)^{p-1} R^{-1} = \tilde{\Psi}_p(R) \;\Leftrightarrow\; R = \tilde{\Psi}_p^{-1}(\alpha)$$

and

$$\begin{aligned} d_\mathcal{Z}(R) = \delta \alpha^{\frac{1}{2(1-p)}} R^{\frac{p-2}{2(p-1)}} &= \delta \alpha^{\frac{1}{2(1-p)}} \left(\frac{d_\mathcal{Z}(R)}{\alpha^{\frac{1}{p-1}}}\right)^{\frac{p-2}{2}} \\ &= \delta\, d_\mathcal{Z}(R)^{\frac{p-2}{2}} \alpha^{-\frac{1}{2}} \end{aligned}$$

or

$$\delta = d_{\mathcal{Z}}(R)^{\frac{4-p}{4}} \alpha^{\frac{1}{2}} =: \bar{\Theta}_p(\alpha).$$

Finally

$$d_{\mathcal{Z}}(R) = \delta \alpha^{\frac{1}{2(p-1)}} R^{\frac{p-2}{2(p-1)}} = \delta\, R^{\frac{p-2}{2(p-1)}} \frac{R^{\frac{1}{2(p-1)}}}{d_{\mathcal{Z}}(R)^{\frac{1}{2}}} = \delta\, R^{\frac{1}{2}} d_{\mathcal{Z}}(R)^{-\frac{1}{2}}$$

or $\delta = d_{\mathcal{Z}}(R)^{\frac{3}{2}} R^{-\frac{1}{2}} =: \bar{\Phi}(R)$ which implies the improved convergence rate. ■

Here, we also do not need a restriction to the decay rate of the distance function $d_{\mathcal{Z}}(R)$ because we have

$$\gamma \|A\,\omega\| = \alpha^{\frac{1}{p-1}} \|A\,\omega\|^{\frac{2-p}{p-1}+1} = \alpha^{\frac{1}{p-1}} \|A\,\omega\|^{\frac{1}{p-1}} \geq C\,\alpha^{\frac{1}{p-1}} \|\omega\|_{\mathcal{Z}}^{\frac{1}{p-1}}$$

Hence we need

$$\frac{\delta}{\alpha^{\frac{1}{p-1}} R^{\frac{1}{p-1}}} \to 0 \quad \text{for} \quad \delta \to 0.$$

In Lemma 3.3 we have

$$\frac{\delta}{\alpha^{\frac{1}{p-1}} R^{\frac{1}{p-1}}} = \frac{d_{\mathcal{Z}}(R)^{\frac{3}{4}}}{R^{\frac{1}{2}} d_{\mathcal{Z}}(R)^{\frac{1}{2}}} = d_{\mathcal{Z}}(R)^{\frac{1}{4}} R^{-\frac{1}{2}} \to 0$$

for $R \to \infty$. In the second case we derive

$$\frac{\delta}{\alpha^{\frac{1}{p-1}} R^{\frac{1}{p-1}}} = \frac{d_{\mathcal{Z}}(R)^{\frac{3}{2}}}{R^{\frac{1}{2}} d_{\mathcal{Z}}(R)} = d_{\mathcal{Z}}(R)^{\frac{1}{2}} R^{-\frac{1}{2}} \to 0$$

for $R \to \infty$ again. Hence, Lemma 3.3 holds, provided δ is sufficiently small.

7 Application in Hilbert spaces

We will show that presenting convergence rates in terms of distance functions in Banach spaces describes a natural generalization of formulating convergence rates for elements satisfying a general source condition in Hilbert spaces. In fact, there seems to be a close relation between the distance function and general source conditions in Hilbert spaces. The following result can be found in [10, Theorem 3.2].

Proposition 7.1 *Let $\mathcal{X}$ and $\mathcal{Y}$ be Hilbert spaces and $A \in L(\mathcal{X}, \mathcal{Y})$ be injective and compact. For given $y \in \mathcal{R}(A)$, let $x^\dagger \notin \mathcal{R}(A^*A)$ be the solution of $A\,x = y$ with distance function $d(R)$. If $x^\dagger \in \mathcal{R}\left((A^*A)^\nu\right)$, $0 < \nu < 1$, then the estimate*

$$d(R) \leq \kappa\, R^{\frac{\nu}{\nu-1}}, \quad R > 0, \tag{14}$$

holds for some constant $\kappa > 0$.

A careful reading of the proof shows, that the estimate (14) is of optimal order. Let us therefore assume, that the distance function is given by $d(R) := \kappa\, R^{\frac{\nu}{\nu-1}}$ for some $\kappa > 0$ and $0 < \nu < 1$. Moreover, we assume $P(x) := \frac{1}{2}\|x\|^2$, which implies $\xi = 2x^\dagger$ and $D_P(x, x^\dagger) = \frac{1}{2}\|x - x^\dagger\|^2$ for $x \in \mathcal{X}$.

Let us consider the case $p = 2$. We have for the function $\Psi_2(R)$ in Theorem 4.1

$$\Psi_2(R) := d(R)^{\frac{1}{2}} R^{-1} = \kappa^{\frac{1}{2}} R^{\frac{\nu}{2(\nu-1)}-1} = \kappa^{\frac{1}{2}} R^{\frac{2-\nu}{2(\nu-1)}} =: \alpha$$

which implies

$$R = \left(\frac{\alpha}{\sqrt{\kappa}}\right)^{\frac{2(\nu-1)}{2-\nu}} = \Psi_2^{-1}(\alpha).$$

Hence

$$\Theta_2(\alpha) = \sqrt{\alpha\, d\left(\Psi_2^{-1}(\alpha)\right)} \sim \alpha^{\frac{1}{2}\left(1+\frac{\nu}{\nu-1}\frac{2(\nu-1)}{2-\nu}\right)} = \alpha^{\frac{2+\nu}{2(2-\nu)}}$$

which provieds a parameter choice $\alpha \sim \delta^{\frac{2(2-\nu)}{2+\nu}}$. Moreover

$$\Phi(R) = d(R)^{\frac{3}{4}} R^{-\frac{1}{2}} = \kappa^{\frac{3}{4}} R^{\frac{3\nu}{4(\nu-1)}-\frac{1}{2}} = \kappa^{\frac{3}{4}} R^{\frac{\nu+2}{4(\nu-1)}}$$

and

$$d\left(\Phi^{-1}(\delta)\right) \sim \delta^{\frac{\nu}{\nu-1}\frac{4(\nu-1)}{\nu+2}} = \delta^{\frac{4\nu}{\nu+2}}$$

hold. This yields a convergence rate $\|x_\alpha^\delta - x^\dagger\| \sim \delta^{\frac{2\nu}{\nu+2}}$, which is not the optimal one. On the other hand, in Theorem 5.1 we introduced

$$\tilde{\Psi}_2(R) = d(R)R^{-1} = \kappa\, R^{\frac{\nu}{\nu-1}-1} = \kappa\, R^{\frac{1}{\nu-1}} =: \alpha$$

or $R = \kappa^{1-\nu}\alpha^{\nu-1}$. Hence

$$\tilde{\Theta}_2(\alpha) = \alpha^{\frac{1}{2}} d\left(\tilde{\Psi}_2^{-1}(\alpha)\right) \sim \alpha^{\frac{1}{2}+\frac{\nu}{\nu-1}(\nu-1)} = \alpha^{\frac{2\nu+1}{2}}$$

which implies the known optimal parameter choice $\alpha \sim \delta^{\frac{2}{2\nu+1}}$. Finally

$$\tilde{\Phi}(R) = d(R)^{\frac{3}{2}} R^{-\frac{1}{2}} = \kappa^{\frac{3}{2}} R^{\frac{3\nu}{2(\nu-1)}-\frac{1}{2}} = \kappa^{\frac{3}{2}} R^{\frac{2\nu+1}{2(\nu-1)}} =: \delta$$

which provides

$$R = \left(\frac{\delta}{\kappa^{\frac{3}{2}}}\right)^{\frac{2(\nu-1)}{2\nu+1}} = \tilde{\Phi}^{-1}(\delta).$$

Then we derive the convergence rate

$$\|x_\alpha^\delta - x^\dagger\| \sim d\left(\tilde{\Phi}^{-1}(\delta)\right) = \kappa\left(\frac{\delta}{\kappa^{\frac{3}{2}}}\right)^{\frac{2\nu}{2\nu+1}} \sim \delta^{\frac{2\nu}{2\nu+1}}. \tag{15}$$

The rate (15) is known to be the optimal one for linear regularization methods such as Tikhonov regularization if $x^\dagger \in \mathcal{R}\left((A^*A)^\nu\right)$ for $0 < \nu \leq 1$, see e.g. [4, Section 5.1]. Hence, generalized source conditions and distance functions provide the same (optimal) convergence rates. The example shows that by dealing with approximative source conditions we can extend classical convergence rate results to linear (and nonlinear) operators mapping between reflexive Banach spaces.

8 On an a-posteriori parameter choice

If the stabilizing functional $P(x)$ is strongly convex in $x^\dagger$, i.e. $P(x)$ satisfies (A9), and the choice $p = 2$ we can apply an a-posteriori parameter strategy for the regularization parameter α, which is also known as Lepskij- or balancing principle, see e.g. [11], [13] and [1]. This strategy has been well-established in the recent years since it is easy to implement and applicable under relatively weak technical assumptions. In particular, it can also be applied in Banach spaces.

For given (sufficiently small) $\alpha_0 > 0$, a real number $q > 1$ and maximal index $j_{max} > 0$ we define the (finite) sequence

$$\{\alpha_j := q^j \alpha_0 \ : \ 0 \le j \le j_{max}\}. \tag{16}$$

The maximal index j_{max} is chosen such that $\alpha_{j_{max}} \le \alpha_{max}$. Then we can present the following a-posteriori choice of the regularization parameter α.

Definition 8.1 (Lepskij-Principle) *Let the sequence $\{\alpha_j\}$ be defined by (16). We calculate solutions $\{x^\delta_{\alpha_j}\}$ of (2) and choose the regularization parameter $\alpha_L := \alpha_{j_L}$ such that*

$$j_L := \max\left\{ j \le j_{max} \ : \ \|x^\delta_{\alpha_i} - x^\delta_{\alpha_j}\| \le 4\sqrt{\frac{C_2}{\eta}}\,\frac{\delta}{\sqrt{\alpha_i}},\ \forall\, i \le j \right\}, \tag{17}$$

where C_2 is the constant of (A8) and η is the constant of (A9). Then $x^\delta_\alpha := x^\delta_{\alpha_L}$ is chosen as regularized solution of (1).

We summarize the most important facts. The main idea of the balancing principle is based on the decomposition of the approximation error of regularized solutions into two parts which both depend on the regularization parameter α. We state the assumption in detail below.

Assumption 8.1 *For each $0 < \alpha \le \|A\|^2$ and given data y^δ let x^δ_α denotes any regularized solution of (1) satisfying*

$$\|x^\delta_\alpha - x^\dagger\| \le \frac{1}{2}\left(\psi(\alpha) + \phi(\alpha)\right) \tag{18}$$

for a known non-increasing function $\psi(\alpha)$, which can depend on δ and an unknown non-decreasing (index) function $\phi(\alpha)$.

Now we can establish the theoretical main results of the balancing principle, see also [13, Proposition 2 and Corollary 1].

Proposition 8.1 *Let $\alpha_0 > 0$ be chosen such that $\phi(\alpha_0) < \psi(\alpha_0)$, $\{\alpha_j\}$ is given by (16) and the index j_L satisfies (17) when the bound $4\sqrt{C_2\eta^{-1}}\frac{\delta}{\sqrt{\alpha_i}}$ is replaced by $2\psi(\alpha_i)$. Moreover, define $\hat{j} := \max\{j \ : \ \phi(\alpha_j) \le \psi(\alpha_j)\}$ and $\tilde{j} := \max\{j \ : \ \|x^\delta_{\alpha_i} - x^\dagger\| \le \psi(\alpha_i),\ \forall\, i \le j\}$. Then, under Assumption 8.1,*

$$j_L \ge \tilde{j} \ge \hat{j} \ge 0 \quad \textit{and} \quad \|x^\delta_{\alpha_L} - x^\dagger\| \le 3\,\psi(\tilde{\alpha}) \le 3\,\psi(\hat{\alpha}).$$

If – in addition – there exists a constant $1 < D < \infty$ *such that* $\psi(\alpha_j) \le D\,\psi(\alpha_{j+1})$, $0 \le j < j_{max}$*, then*

$$\|x^\delta_{\alpha_L} - x^\dagger\| \le 6D\,\min\{\psi(\alpha_i) + \phi(\alpha_i),\ 0 \le j \le j_{max}\}.$$

Assume, $\xi^\dagger \in \mathcal{X}^*$ satisfies the source condition (9) for some $\omega \in \mathcal{X}$ with $\|\omega\| = R$. Then, by Lemma 3.4

$$\eta\|x^\delta_\alpha - x^\dagger\|^2 \le D_P(x^\delta_\alpha, x^\dagger) \le C_2\frac{\delta^2}{\alpha} + C_3\alpha^2R^2$$

respectively

$$2\|x^\delta_\alpha - x^\dagger\| \le 2\sqrt{\frac{C_2}{\eta}}\frac{\delta}{\sqrt{\alpha}} + 2\sqrt{\frac{C_3}{\eta}}R\,\alpha =: \psi(\alpha) + \phi(\alpha)$$

holds. Hence, Assumption 8.1 is satisfied. On the other hand, for $\xi^\dagger \in \overline{\mathcal{M}} \setminus \mathcal{M}$ we have for all $R \ge 0$

$$\begin{aligned}
2\|x^\delta_\alpha - x^\dagger\| &\le 2\left(C_8\alpha\,R + \sqrt{\frac{C_2}{\eta}}\frac{\delta}{\sqrt{\alpha}} + C_7 d(R)\right)\\
&= 2\sqrt{\frac{C_2}{\eta}}\frac{\delta}{\sqrt{\alpha}} + 2\,(C_7 + C_8)\,d\left(\tilde{\Psi}_2^{-1}(\alpha)\right) =: \psi(\alpha) + \phi(\alpha),
\end{aligned}$$

which $\tilde{\Psi}_2(R) := d(R)R^{-1}$ which is the first balancing step in the proof of Theorem 5.1. Hence, in both cases (18) holds and we can apply Proposition 8.1. Of main interest is the following consequence.

Corollary 8.1 *Assume all conditions of Proposition 8.1 to be satisfied and* α_L *is chosen by (17). Moreover, assume* $\hat{j} < j_{max}$*.*

(i) If $\xi^\dagger = A^\star J_Y(A\,\omega)$ *for some* $\omega \in \mathcal{X}$ *with* $\|\omega\| \le R$*, then*

$$\|x^\delta_\alpha - x^\dagger\| \le 6\sqrt{q\,\eta^{-1}}\max\left\{\sqrt{C_2}, \sqrt{C_3}R\right\}\delta^{\frac{2}{3}}.$$

(ii) If $\xi^\dagger \in \overline{\mathcal{M}} \setminus \mathcal{M}$ *with distance function* $d(R) = d(R;\xi^\dagger)$*, then*

$$\|x^\delta_\alpha - x^\dagger\| \le 6\sqrt{q}\max\left\{\sqrt{C_2\eta^{-1}}, C_7 + C_8\right\} d\left(\tilde{\Phi}^{-1}(\delta)\right),$$

where $\tilde{\Phi}(R) := d(R)^{\frac{3}{2}}R^{-\frac{1}{2}}$*.*

Proof. We can apply Proposition 8.1 with $\psi(\alpha) := 2\sqrt{C_2\eta^{-1}}\frac{\delta}{\sqrt{\alpha}}$ and $\phi(\alpha) := 2\sqrt{C_3\eta^{-1}}R\,\alpha$ in the first and $\phi(\alpha) := 2(C_7 + C_8)d(\tilde{\Psi}_2^{-1}(\alpha))$ in the second case. We introduce the notation $\bar{\alpha} > 0$ which satisfies $\psi(\bar{\alpha}) = \phi(\bar{\alpha})$. Obviously $\bar{\alpha} < \alpha_{\hat{j}+1} \le j_{max}$ holds. Hence, by monotonicity

$$\|x^\delta_{\alpha_L} - x^\dagger\| \le 3\,\psi(\hat{\alpha}) \le 3D\,\psi(\alpha_{\hat{j}+1}) \le 3D\,\psi(\bar{\alpha}) = 3D\,\phi(\bar{\alpha}).$$

Moreover, since

$$\psi(\alpha_i) = 2\sqrt{\frac{C_2}{\eta}}\frac{\delta}{\sqrt{\alpha_i}} = 2\sqrt{\frac{C_2}{\eta}}\sqrt{q}\frac{\delta}{\sqrt{q\alpha_i}} = \sqrt{q}\psi(\alpha_{i+1})$$

the additional condition of Proposition 8.1 is satisfied with $D := \sqrt{q}$. We now consider the first case. Let $\alpha_* > 0$ satisfy $\frac{\delta}{\sqrt{\alpha_*}} = \alpha_* \Leftrightarrow \alpha_* = \delta^{\frac{2}{3}}$. Assume $C_2 \leq C_3 R$. Then $\psi(\alpha_*) \leq \phi(\alpha_*)$ which implies $\alpha_* \geq \bar{\alpha}$. Hence

$$\|x^\delta_{\alpha_L,h} - x^\dagger\| \leq 3\sqrt{q}\phi(\bar{\alpha}) \leq 3\sqrt{q}\phi(\alpha_*) = 6\sqrt{q\,C_3\eta^{-1}}R\alpha_* = 6\sqrt{q\,C_3\eta^{-1}}R\,\delta^{\frac{2}{3}}.$$

On the other hand, if $C_2 \geq C_3 R$ then $\phi(\alpha_*) \geq \phi(\alpha_*)$ and hence $\alpha_* \leq \bar{\alpha}$ holds. This provides

$$\|x^\delta_{\alpha_L} - x^\dagger\| \leq 3\sqrt{q}\psi(\bar{\alpha}) \leq 3\sqrt{q}\psi(\alpha_*) = 6\sqrt{q\,C_2\eta^{-1}}\frac{\delta}{\sqrt{\alpha_*}} = 6\sqrt{q\,C_2\eta^{-1}}\alpha_* = 6\sqrt{q\,C_2\eta^{-1}}\delta^{\frac{2}{3}}.$$

The second case can be treated analogously when the constant $\sqrt{C_3\eta^{-1}}R$ is replaced by $C_7 + C_8$. ■

9 Nonlinear equations

We now deal with the nonlinear equation (4). In order to restrict the nonlinearity of the operator F by the following assumption

(A11) There exists a linear operator $G \in L(\mathcal{X}, \mathcal{Y})$ such that

$$\|F(x) - F(x^\dagger) - G(x - x^\dagger)\| \leq L\,D_P(x, x^\dagger), \qquad \forall\, x \in \mathcal{B}_\rho(x^\dagger) \cap \mathcal{D}(F) \tag{19}$$

with ball $\mathcal{B}_\rho(x^\dagger)$ around $x^\dagger$, $\rho \geq K > 0$ chosen sufficiently large, and a constant $L > 0$.

This restriction is a modification of the original condition introduced in [3] for deriving convergence rates for Tikhonov regularization of nonlinear ill-posed problems in Hilbert spaces. In particular, if F is Fréchet-differentiable in $x^\dagger \in \mathrm{int}\mathcal{D}(F)$, we choose $G = F'(x^\dagger)$. For $\gamma > 0$ we define the residuals

$$R_1 := F(x^\delta_\alpha) - F(x^\dagger) - G\,(x^\delta_\alpha - x^\dagger) \quad \text{and} \quad R_2 := F(x^\dagger - \gamma\omega) - F(x^\dagger) + \gamma\,G\,\omega.$$

Hence, the estimates

$$\|R_1\| \leq L\,D_P(x^\delta_\alpha, x^\dagger) \quad \text{and} \quad \|R_2\| \leq L\,D_P(x^\dagger - \gamma\,\omega, x^\dagger) \leq L\,C_1\gamma^2\|\omega\|^2$$

hold. We present the convergence rate result.

Theorem 9.1 *Assume $p > 1$, (A1)-(A8), $\xi^\dagger \in \mathcal{X}^*$ satisfies the source condition (9) for some $\omega \in \mathcal{X}$. Moreover, (A11) holds for some $L \in \mathbb{R}$ with $L\,\|G\,\omega\| < 1$. Then, an a-priori choice $\alpha \sim \delta^{\frac{2(p-1)}{3}}$ leads to the convergence rate*

$$D_P(x^\delta_\alpha, x^\dagger) \sim \mathcal{O}\left(\delta^{\frac{4}{3}}\right).$$

PROOF. Here, we estimate

$$\begin{aligned}
\frac{1}{p}\|F(x^\dagger - \gamma\omega) - y^\delta\|^p &= \frac{1}{p}\|F(x^\dagger) - \gamma\, G\,\omega + R_2 - y^\delta\|^p \\
&= \frac{\gamma^p}{p}\|G\,\omega\|^p + \gamma^{p-1}\|G\,\omega\|^{p-2}\langle J_Y(G\,\omega), y^\delta - R_2 - F(x^\dagger)\rangle \\
&\qquad + D_Y(y^\delta - F(x^\dagger - \gamma\omega), \gamma\, G\,\omega) \\
&= \frac{\alpha\gamma}{p}\|G\,\omega\|^2 + \alpha\langle J_Y(G\,\omega), y^\delta - R_2 - F(x^\dagger)\rangle \\
&\qquad + D_Y(y^\delta - F(x^\dagger - \gamma\omega), \gamma\, G\,\omega),
\end{aligned}$$

by setting $\gamma^{p-1} := \alpha\|G\,\omega\|^{2-p}$. Moreover,

$$\begin{aligned}
\frac{1}{p}\|F(x_\alpha^\delta) - y^\delta\|^p + \alpha D_P(x_\alpha^\delta, x^\dagger) &\le \alpha D_P(x^\dagger - \gamma\omega, x^\dagger) + D_Y(y^\delta - F(x^\dagger - \gamma\omega), \gamma\, G\,\omega) \\
&\quad + \alpha\langle J_Y(G\,\omega), G\,(x^\dagger - x_\alpha^\delta)\rangle - \alpha\gamma\|G\,\omega\|^2 \\
&\quad + \frac{\alpha\gamma}{p}\|G\,\omega\|^2 + \alpha\langle J_Y(G\,\omega), y^\delta - R_2 - F(x^\dagger)\rangle \\
&= \alpha D_P(x^\dagger - \gamma\omega, x^\dagger) + D_Y(y^\delta - F(x^\dagger - \gamma\omega), \gamma\, G\,\omega) \\
&\quad + \frac{\gamma\alpha}{p}\|G\,\omega\|^2 - \alpha\gamma\|G\,\omega\|^2 \\
&\quad + \alpha\langle J_Y(G\,\omega), y^\delta - F(x^\dagger) - G\,(x_\alpha^\delta - x^\dagger) - R_2\rangle \\
&\le \alpha D_P(x^\dagger - \gamma\omega, x^\dagger) + D_Y(y^\delta - F(x^\dagger - \gamma\omega), \gamma\, G\,\omega) \\
&\quad + \frac{\gamma\alpha}{p}\|G\,\omega\|^2 - \alpha\gamma\|G\,\omega\|^2 + \frac{1}{p}\|F(x_\alpha^\delta) - y^\delta\|^p \\
&\quad + \frac{(\alpha\|G\,\omega\|)^q}{q} + \alpha\|G\,\omega\|\left(\|R_1\| + \|R_2\|\right).
\end{aligned}$$

Same calculations as above leads to

$$\begin{aligned}
D_P(x_\alpha^\delta, x^\dagger) &\le D_P(x^\dagger - \gamma\omega, x^\dagger) + \frac{1}{\alpha}D_Y(y^\delta - F(x^\dagger - \gamma\omega), \gamma\, G\,\omega) \\
&\quad + \|G\,\omega\|\left(L\, D_P(x_\alpha^\delta, x^\dagger) + L\, D_P(x^\dagger - \gamma\omega, x^\dagger)\right) \\
&= (1 + L\,\|G\,\omega\|)D_P(x^\dagger - \gamma\omega, x^\dagger) + \frac{1}{\alpha}D_Y(y^\delta - F(x^\dagger - \gamma\omega), \gamma\, G\,\omega) \\
&\quad + L\,\|G\,\omega\|D_P(x_\alpha^\delta, x^\dagger).
\end{aligned}$$

In order to apply Lemma 3.3 we consider

$$\begin{aligned}
\|y^\delta - F(x^\dagger - \gamma\omega) - \gamma\, G\,\omega\| &= \|y^\delta - F(x^\dagger) + R_2\| \\
&\le \delta + L\, C_1\,\gamma^2\|\omega\|^2 \\
&= \delta + L\, C_1\gamma\,\alpha^{\frac{1}{p-1}}\|G\,\omega\|^{\frac{2-p}{p-1}}\|\omega\|^2,
\end{aligned}$$

which leads to the condition

$$\delta + L\, C_1\gamma^2\|\omega\|^2 \le \tilde{c}\gamma\|G\,\omega\| = \tilde{c}\,\alpha^{\frac{1}{p-1}}\|G\,\omega\|^{\frac{1}{p-1}}$$

for some constant $0 < \tilde{c} < 1$. Hence, Lemma 3.3 can be applied if $\delta^{p-1}\alpha^{-1}$ and α are sufficiently small. Then

$$\begin{aligned} D_Y(y^\delta - F(x^\dagger - \gamma\omega), \gamma\, G\,\omega) &\leq C_2\left(\delta + L\, D_P(x^\dagger - \gamma\omega, x^\dagger)\right)^2 \gamma^{p-2}\|G\,\omega\|^{p-2} \\ &\leq C_2\left(\delta + L\, C_1\gamma^2\|\omega\|^2\right)^2 \gamma^{p-2}\|G\,\omega\|^{p-2} \\ &\leq 2C_2\delta^2\alpha^{1+\frac{1}{p-1}}\|G\,\omega\|^{\frac{p-2}{p-1}} \\ &\quad +L^2C_1^2\alpha^{\frac{4}{p-1}}\|G\,\omega\|^{\frac{2(2-p)}{p-1}}\|\omega\|^4\alpha^{1+\frac{1}{p-1}}\|G\,\omega\|^{\frac{p-2}{p-1}} \\ &= 2C_2\delta^2\alpha^{1+\frac{1}{p-1}}\|G\,\omega\|^{\frac{p-2}{p-1}} \\ &\quad +L^2C_1^2\alpha^{1+\frac{5}{p-1}}\|G\,\omega\|^{\frac{2-p}{p-1}}\|\omega\|^4. \end{aligned}$$

Hence

$$(1 - L\|G\,\omega\|)\, D_P(x_\alpha^\delta, x^\dagger) \leq 2C_2\|G\,\omega\|^{\frac{p-2}{p-1}}\frac{\delta^2}{\alpha^{\frac{1}{p-1}}} + \tilde{C}\alpha^{\frac{2}{p-1}}$$

with

$$\tilde{C} := L^2C_1^2\alpha^{1+\frac{3}{p-1}}\|G\,\omega\|^{\frac{2-p}{p-1}}\|\omega\|^4 + (1 + L\,\|G\,\omega\|)C_1\|\omega\|^2,$$

which leads to the desired convergence rate result. Note, that the suggested parameter choice provides $\delta^{p-1}\alpha^{-1} \to 0$ for $\delta \to 0$ which holds the validy of Lemma 3.3. ■

On the other hand, smallness conditions contradicts the idea of distance functions. So we cannot derive convergence rates for nonlinear problems under condition (A11) for violated source condition (9). On the other hand, other nonlinearity restrictions, which might allow the application of distance functions turned to be not appropriated for proofing convergence rates in this specific situation.

Finally, the case $p = 1$ is considered. Here we can present the following result.

Theorem 9.2 *Assume $p = 1$, (A1)-(A8), $\xi^\dagger \in \mathcal{X}^*$ satisfies the source condition (9) for some $\omega \in \mathcal{X}$. Moreover, (A11) holds for some $L \in \mathbb{R}$ with $L\,\|G\,\omega\| < 1$. If the regularization parameter α chosen such that $0 \leq 1 - \alpha\|G\,\omega\| \leq \delta^{\frac{2}{3}}$, then*

$$D_P(x^\dagger, x_\alpha^\delta) \sim \mathcal{O}\left(\delta^{\frac{4}{3}}\right).$$

Proof. By the calculations above,

$$\begin{aligned} \|F(x_\alpha^\delta) - y^\delta\| + \alpha D_P(x_\alpha^\delta, x^\dagger) &\leq \alpha D_P(x^\dagger - \gamma\omega, x^\dagger) + D_Y(y^\delta - F(x^\dagger - \gamma\omega), \gamma\, G\,\omega) \\ &\quad +\alpha\langle J_Y(G\,\omega), G\,(x^\dagger - x_\alpha^\delta)\rangle - \alpha\gamma\|G\,\omega\|^2 \\ &\quad +\gamma\|G\,\omega\| + \frac{1}{\|G\,\omega\|}\langle J_Y(G\,\omega), y^\delta - R_2 - F(x^\dagger)\rangle \\ &\leq \alpha D_P(x^\dagger - \gamma\omega, x^\dagger) + D_Y(y^\delta - F(x^\dagger - \gamma\omega), \gamma\, G\,\omega) \\ &\quad +\alpha\langle J_Y(G\,\omega), G\,(x^\dagger - x_\alpha^\delta)\rangle + \alpha\, L\,\|\omega\| D_P(x_\alpha^\delta, x^\dagger) \\ &\quad -\alpha\gamma\|G\,\omega\|^2 + L\, D_P(x^\dagger - \gamma\,\omega, x^\dagger) \\ &\quad +\gamma\|G\,\omega\| + \frac{1}{\|G\,\omega\|}\langle J_Y(G\,\omega), y^\delta - F(x^\dagger)\rangle \\ &\leq (\alpha + L)D_P(x^\dagger - \gamma\omega, x^\dagger) + D_Y(y^\delta - F(x^\dagger - \gamma\omega), \gamma\, G\,\omega) \\ &\quad +\alpha\|G\,\omega\|\,\|y^\delta - F(x_\alpha^\delta)\| + \alpha\, L\,\|G\,\omega\|\, D_P(x_\alpha^\delta, x^\dagger) \\ &\quad +\gamma\|G\,\omega\|(1 - \alpha\|G\,\omega\|) + |1 - \alpha\|G\,\omega\||\,\delta \end{aligned}$$

By Lemma 3.3 we have

$$D_Y(y^\delta - F(x^\dagger - \gamma\omega), \gamma\, G\,\omega) \le C_2 \frac{\left(\delta + L\, D_P(x^\dagger - \gamma\omega, x^\dagger)\right)^2}{\gamma\|G\,\omega\|} \le C_2 \frac{\left(\delta + L\, C_1\, \gamma^2\|\omega\|^2\right)^2}{\gamma\|G\,\omega\|}$$

Setting $\gamma := \delta^{\frac{2}{3}}$ again, we derive

$$\begin{aligned}(1 - L\|G\,\omega\|)\, D_P(x_\alpha^\delta, x^\dagger) \quad &\le \quad \delta^{\frac{4}{3}}(\alpha + L)C_1\|\omega\|^2 + \delta^{\frac{4}{3}} \frac{C_2\left(1 + L\, C_1 \delta^{\frac{1}{3}}\|\omega\|^2\right)^2}{\|G\,\omega\|} \\ &\quad\ +\delta^{\frac{5}{3}}.\end{aligned}$$

This proves the assertion. ■

A Duality maps in L^p-spaces

In order to get a bit more familiar we consider the duality maps in the spaces $\mathcal{X} = L^p(0,1)$, $1 \le p < \infty$. We set

$$P(x) := \frac{1}{2}\|x\|_P^2 = \frac{1}{2}\left(\int_0^1 |x(t)|^p\, dt\right)^{\frac{2}{p}}, \qquad x \in L^p(0,1),$$

and $J_X(x) := P'(x)$. Straightforward calculations shows

$$\begin{aligned}\lim_{\varepsilon \to 0} \frac{P(x + \varepsilon\, h) - P(x)}{\varepsilon} \quad &= \quad \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon}\left(\left(\int_0^1 |(x + \varepsilon\, h)(t) + |^p\, dt\right)^{\frac{2}{p}} - \left(\int_0^1 |x(t)|^p\, dt\right)^{\frac{2}{p}}\right) \\ &= \quad \|x\|_p^{2-p} \int_0^1 |x(t)|^{p-1}\mathrm{sgn}(x(t))h(t)\, dt,\end{aligned}$$

i.e. $J_X(x) = \|x\|_p^{2-p}|x|^{p-1}\mathrm{sgn}(x)$. Note, with $q := \frac{p}{p-1}$ for $p > 1$ we have

$$\begin{aligned}\|J_X(x)\|_q \quad &= \quad \|x\|_p^{2-p}\left(\int_0^1 |x|^{(p-1)q}\, dt\right)^{\frac{1}{q}} \\ &= \quad \|x\|_p^{2-p}\|x\|_p^{\frac{p-1}{p}p} \\ &= \quad \|x\|_p\end{aligned}$$

which shows $J_X(x) \in L^p(0,1)^* = L^q(0,1)$ and $\|J_X(x)\|_q = \|x\|_p$. Moreover, for $p > 2$ (the case $p = 2$ is clear),

$$J_X'(x) = \left(\|x\|_p^{2-p}\right)'|x|^{p-1}\mathrm{sgn}(x) + \|x\|_p^{2-p}(p-1)|x|^{p-2}$$

holds. Here we have

$$\begin{aligned}
\left(\|x\|_P^{2-p}\right)' &= \left[\left(\|x\|_p^2\right)^{\frac{2-p}{2}}\right]' \\
&= \frac{2-p}{2}\left(\|x\|_p^2\right)^{\frac{2-p}{2}-1} 2\, J_X(x) \\
&= (2-p)\|x\|_p^{-p} J_X(x)
\end{aligned}$$

Hence,

$$\begin{aligned}
J_X'(x)(h_1,h_2) &= (2-p)\|x\|_p^{2-2p}\int_0^1 |x(t)|^{p-1}\mathrm{sgn}(x(t))h_1(t)\,dt \int_0^1 |x(t)|^{p-1}\mathrm{sgn}(x(t))h_2(t)\,dt \\
&\quad +(p-1)\|x\|_p^{2-p}\int_0^1 |x(t)|^{p-2}h_1(t)h_2(t)\,dt \\
&=: M_1(h_1,h_2)+M_2(h_1,h_2).
\end{aligned}$$

Since $\||x|^{p-1}\|_q = \|x\|_p^{p-1}$ we conclude by Hölder's inequality

$$\begin{aligned}
|M_1(h_1,h_2)| &\leq (p-2)\|x\|_p^{2-2p}\||x|^{p-1}\|_q\|h_1\|_p\||x|^{p-1}\|_q\|h_2\|_p \\
&= (p_2)\|h_1\|_p\|h_2\|_p
\end{aligned}$$

and

$$|M_2(h_1,h_2)| \leq (p-1)\|x\|_p^{2-p}\|h_1\|_p\|h_2\|_p\||x|^{p-2}\|_r$$

with

$$r = \frac{p\,q}{p-q} = \frac{\frac{p^2}{p-1}}{p-\frac{p}{p-1}} = \frac{p}{p-2}.$$

Finally,

$$\||x|^{p-2}\|_r = \left(\int_0^1 |x(t)|^{(p-2)\frac{p}{p-2}}\,dt\right)^{\frac{p-2}{p}} = \|x\|_p^{p-2},$$

which shows $|M_2(h_1,h_2)| \leq (p-1)\|h_1\|_p\|h_2\|_p$ and finally $\|J_X'(x)\| \leq 2p-3$, i.e. the norm of the second derivative of $P(x)$ does not depend on the element x anymore if $p \geq 2$.

References

[1] F.Bauer and T.Hohage, *A Lepskij-type stopping rule for regularized Newton methods.* Inverse Problems **21**, (2005) 1979-1991.

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

[3] H.W.Engl, K.Kunisch and A.Neubauer, *Convergence rates for Tikhonov regularization of nonlinear ill-posed problems.* Inverse Problems **5** (1989), 523-540.

[4] H.W.Engl, M.Hanke and A.Neubauer, "Regularization of Inverse Problems." Kluwer Academic Publishers, Dortrecht (1996).

[5] M.Hanke, A.Neubauer and O.Scherzer, *A convergence analysis of the Landweber iteration for nonlinear ill-posed problems.* Numer.Math **72** (1995), 21-27.

[6] T.Hein, *Regularization of ill-posed problems in Banach spaces - approximative source conditions and convergence rates results.* Preprint 2007-28, TU Chemnitz, Faculty of Mathematics.

[7] B.Hofmann, *Approximative source conditions in Tikhonov-Phillips regularization and consequences for inverse problems with multiplication operators.* Mathematical Methods in the Applied Sciences **29** (2006), 351-371.

[8] B.Hofmann, D.Düvelmeyer and K.Krumbiegel, *Approximative source conditions in Tikhonov regularization – new analytical results and some numerical studies.* Mathematical Modelling and Analysis **10**, No.1 (2006), 41-56.

[9] B.Hofmann, B.Kaltenbacher, C.Pöschl and O.Scherzer, *A convergence rates result in Banach spaces with non-smooth operators.* Inverse Problems **23** (2007), 987-1010.

[10] B.Hofmann, D.Düvelmeyer and M.Yamamoto, *Range inclusions and approximative source conditions with general benchmark functions.* Preprint 2007-02, TU Chemnitz, Faculty of Mathematics.

[11] O.V.Lepskij, *A problem of adaptive estimation in Gaussian withe noise.* Teor.Veroyatn.Primen. **35** (1990), 459-470.

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

[13] P.Mathé, *The Lepskij principle revisited.* Inverse Problems **22** (2006), L11-L15.

[14] M.T.Nair, E.Schock and U.Tautenhahn, *Morozov's discrepancy principle under general source conditions.* Zeitschrift für Analysis u. Anwendungen **299** (2003), 199-214.

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

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

[17] T.I.Seidman and C.R.Vogel, *Well posedness and convergence of some regularisation methods for non-linear ill-posed problems.* Inverse Problems **5** (1989), 227-238.

[18] U.Tautenhahn, *Optimality for ill-posed problems under general source conditions.* Numer.Funct.Anal.Optim. **19** (1998), 377-398.

[19] D.Werner, "Funktionalanalysis" (in German), Springer-Verlag Berlin (1997).

[20] E.Zeidler, "Nonlinear Functional Analysis and its Applications III", Springer-Verlag New York (1985).