Regularization of ill-posed problems in Banach spaces – approximative source conditions and convergence rates results

TORSTEN HEIN *

December 12, 2007

Abstract

In this preprint we deal with convergence rates for regularizing linear and nonlinear ill-posed problems with operators mapping from a Hilbert space into a Banach space. Thereby we deal with so-called distance functions which quantify the violation of a reference source condition. With the aid of these functions we present error bounds and convergence rates for regularized solutions of linear and non-linear problems when the reference source condition is not satisfied. The way of applying distance functions transfers the idea of considering generalized source conditions in Hilbert spaces to inverse problems in Banach spaces in a natural way. Introducing this topic for linear ill-posed problems we additionally show that this theory can be easily extended to nonlinear problems as well as to more general penalty terms using Bregman distances. Moreover, the application of the discrepancy principle as a posteriori choice strategy of the regularization parameter is discussed for both linear and nonlinear problems.

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

MSC 2000 Subject Classification: 47A52, 47J06, 49N45

1 Introduction

Introducing, let $A : \mathcal{X} \longrightarrow \mathcal{Y}$ be a linear and bounded operator between the Hilbert space $\mathcal{X}$ and the reflexive Banach space $\mathcal{Y}$ with non-closed range, i.e. $\mathcal{R}(A) \neq \overline{\mathcal{R}(A)}$. We

*Technische Universität Chemnitz, Fakultät für Mathematik, 09107 Chemnitz, Deutschland. Email: Torsten.Hein@mathematik.tu-chemnitz.de

consider convergence rates for a non-quadratic regularization approach for approximate solving the operator equation

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

for given noisy data $y^\delta \in \mathcal{Y}$. Let $y = y^0$ be the exact data and δ denotes the noise level, i.e. $\|y - y^\delta\| \le \delta$. For simplicity, we assume A to be injective and there exists a solution $x^\dagger := A^{-1}y$ of (1) for the exact data.

Using standard Tikhonov regularization, equation (1) is replaced by the quadratic minimization problem

$$\|A\,x - y^\delta\|^2 + \alpha\|x\|^2 \to \min \quad \text{subject to } x \in \mathcal{X}. \tag{2}$$

The problem is well-studied if $\mathcal{X}$ and $\mathcal{Y}$ are Hilbert spaces, see e.g. [4] and the references therein. Let us – as usual – denote a solution of (2) with x_α^δ for noisy data and with x_α in the noiseless case. In order to prove convergence rates for $x_\alpha^\delta \to x^\dagger$ and $x_\alpha \to x^\dagger$ for $\alpha, \delta \to 0$ an additional source condition has to be satisfied, for example

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

for some index function $\varphi(t)$, $0 \le t \le \|A\|^2$, see also [7, 9, 10] and [14] for some newer results. In this context, index functions are continuous and strictly increasing functions with $\varphi(0) = 0$. For $\varphi(t) = \sqrt{t}$, condition (3) is equivalent to

$$x^\dagger = A^*\tilde{\omega}, \qquad \tilde{\omega} \in \mathcal{Y}, \tag{4}$$

with $\|\tilde{\omega}\| = \|\omega\|$. On the other hand, for functions $\varphi(t)$ with

$$\lim_{t\to 0} \frac{t^{\frac{1}{2}}}{\varphi(t)} = 0,$$

the condition (3) can be interpreted as a weakening of the source condition (4). In particular, power type source conditions $\varphi(t) = t^\eta$ for $0 < \eta < \frac{1}{2}$ are standard examples for such weaker conditions. This approach has the drawback, that it cannot be used in our situation. Since $\mathcal{Y}$ is now a Banach space we cannot apply A^* to elements belonging to the range $\mathcal{R}(A) \subset \mathcal{Y}$ of A without further assumptions. In [6] an alternative concept for proving convergence rates in Hilbert spaces was presented, when the source condition (4) is injured. It is based on the decay rate of so-called distance functions which can be interpreted as a measure for an allowed violation of the source condition (4). It points out, that this approach can be generalized to Banach spaces. This is the main purpose of the present paper.

We also wanted to mention the papers [11] and [12] with deal with convergence rates for regularizing linear and nonlinear operator equations in Banach spaces. There, the operators were mapping from a Banach space into a Hilbert space, which is the opposite situation as discussed here. Convergence rates for regularizing operator equations with operators mapping between two Banach spaces were recently presented in [8].

Additionally, we apply a non-quadratic regularization approach. Instead of (2) we search for a minimizer of the non-quadratic problem

$$\|A\,x - y^\delta\|^p + \alpha\|x\|^2 \to \min \quad \text{subject to } x \in \mathcal{X}. \tag{5}$$

The motivation arises by two aspects. Since $\mathcal{Y}$ is now a Banach space there are only little practical reasons for dealing with a squared norm (besides the application of duality maps, which, however, are only a mathematical tool). In particular, if we have $\mathcal{Y} = L^q(\Omega)$ with domain $\Omega \subset \mathbb{R}^n$ the choice $p = q$ would be the canonical one. The other reason can be found in [2]. There, the case $p = 1$ was considered, which is also known as exact penalization model. Then a convergence rates result is presented (in Hilbert spaces) with the probably simplest strategy for choosing the regularization parameter: one obtains an optimal convergence rate for $\delta \to 0$ by choosing $\alpha > 0$ sufficiently small, but independently of the noise level δ. Note, that this result do not contradicts the Bakhushinsky veto, see [4, Theorem 3.3], since (5) defines a nonlinear regularization method. Unfortunately this result holds true only if the source condition (4) is satisfied, which will be shown in the present paper.

Again, we denote a solution of (5) with x_α^δ for given noisy data and with x_α in the noiseless case. The existence of such solution, the stability and the convergence $x_\alpha^\delta \to x^\dagger$ (and $x_\alpha \to x^\dagger$) for $\delta \to 0$ and $\alpha = \alpha(\delta)$ converging to zero sufficiently slow, can be easily proven by applying standard arguments, see e.g. [3] or [13] in more general situations. Since the main focus of the paper is the presentation of convergence rates, we omit the proofs here. Note, that the results presented here can be easily extended to the problem (2).

The paper is organized as follows: in section 2 we present error bounds for regularized solutions of (5), when the exact solution $x^\dagger$ belongs to some sets $\mathcal{M}_d(R)$ which will be defined later. Using these results we then derive convergence rate results by the aid of distance functions. The relation between distance functions and generalized source conditions in Hilbert spaces is demonstrated in section 4 by an example. In section 5 possible generalizations of the results of section 2 and 3 to non-linear inverse problems are indicated. Section 6 deals with the a posteriory choice of regularization parameter α based on the discrepancy principle of Morozov in the linear case as well as in the nonlinear situation. Finally, in Section7 we examine convergence rates for more general penalty terms by the aid of Bregman distances.

2 Error bounds

For discussing error bounds for $\|x_\alpha^\delta - x^\dagger\|$ and $\|x_\alpha - x^\dagger\|$, a source condition of the form

$$x^\dagger = A^*\omega, \qquad \omega \in \mathcal{Y}^*, \tag{6}$$

as reference source condition is assumed. Here, $A^* : \mathcal{Y}^* \longrightarrow \mathcal{X}$ denotes the dual operator of A, i.e. the relation

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

holds. Thereby $\langle \cdot,\cdot\rangle_{\mathcal{Y}^*,\mathcal{Y}}$ stands for the duality product in $\mathcal{Y}$ and $\langle \cdot,\cdot\rangle_{\mathcal{X}}$ denotes the scalar product in $\mathcal{X}$. Here, we have already identified the dual space $\mathcal{X}^*$ with the space $\mathcal{X}$ itself.

We replace (6) by an approximate source condition. Therefore we define the following sets.

Definition 2.1 *Let $\rho : (0,\infty) \longrightarrow \mathbb{R}$ be a nonnegative decreasing function with*

$$\lim_{R\to\infty} \rho(R) = 0.$$

For each $R > 0$ we define the sets

$$\mathcal{M}_\rho(R) := \{x \in \mathcal{X} \,:\, x = A^*\omega + r,\ \|\omega\| \le R,\ \|r\| \le \rho(R)\}\,. \tag{7}$$

For given $x^\dagger \in \mathcal{X}$ the number $\rho = \rho(R)$ describes the maximal violation of the (reference) source condition (6) which is allowed, when the norm of the source element $\omega \in \mathcal{Y}^*$ is bounded by some constant $R > 0$. The relation of two sets $\mathcal{M}_\rho(R_1)$ and $\mathcal{M}_\rho(R_2)$ for $R_1 < R_2$ is not clear. For increasing R the number of admissible source elements $\omega \in \mathcal{Y}^*$ is growing but the allowed violation $\rho(R)$ becomes smaller. So it is an open problem if we can find conditions for the decrease of the function $\rho(R)$ which guarantees the implication $\mathcal{M}_\rho(R_1) \subseteq \mathcal{M}_\rho(R_2)$ for $R_1 < R_2$.

Now we can present error bounds for $\|x_\alpha - x^\dagger\|$ and $\|x_\alpha^\delta - x^\dagger\|$ for exact solutions $x^\dagger \in \mathcal{X}$ belonging to some set $\mathcal{M}_\rho(R)$. We start with the case $p > 1$. The main idea of the proof bases on Young's inequality with ε, i.e. for $a, b, \varepsilon > 0$ and $p > 0$ we have

$$a\,b \le \varepsilon\, a^p + \frac{b^q}{(\varepsilon\, p)^{\frac{q}{p}} q} \quad \text{with} \quad q = (1-p^{-1})^{-1} = \frac{p}{p-1}.$$

Moreover, the implication

$$a, b, c > 0,\ a^2 \le b^2 + ac \ \Rightarrow\ a \le b + c \tag{8}$$

is used frequently. Now we present the first result.

Lemma 2.1 *Assume $x^\dagger \in \mathcal{M}_\rho(R)$ for some $R > 0$ and a function $\rho = \rho(R)$. If $p > 1$ then the estimate*

$$\|x_\alpha - x^\dagger\| \le \sqrt{c(p)}\alpha^{\frac{1}{2(p-1)}} R^{\frac{p}{2(p-1)}} + 2\,\rho(R). \tag{9}$$

holds for given exact data y, whereas in the case of noisy data y^δ

$$\|x_\alpha^\delta - x^\dagger\| \le \sqrt{\frac{2\delta^p}{\alpha} + 2c(p)\alpha^{\frac{1}{p-1}} R^{\frac{p}{p-1}}} + 2\rho(R) \tag{10}$$

yields. Here, $c(p)$ is a positive constant depending only on p.

Proof. Since $x^\dagger \in \mathcal{M}_\rho(R)$ we can set $x^\dagger = A^*\omega + r$ with $\|\omega\| \le R$ and $\|r\| \le \rho(R)$. We have by the minimizing property of x_α

$$\|A\,x_\alpha - y\| + \alpha\|x_\alpha\|^2 \le \|A\,x^\dagger - y\| + \alpha\|x^\dagger\|^2 \le \alpha\|x^\dagger\|^2.$$

Hence

$$\begin{aligned}
\|A\,x_\alpha - y\|^p + \alpha\|x_\alpha - x^\dagger\|^2 &\le 2\alpha\langle x^\dagger - x_\alpha, x^\dagger\rangle \\
&= 2\alpha\langle x^\dagger - x_\alpha, A^*\omega + r\rangle \\
&\le 2\alpha\left[\|A\,x_\alpha - y\|\|\omega\| + \|x^\dagger - x_\alpha\|\|r\|\right]
\end{aligned}$$

Applying Young's inequality with $\varepsilon := \frac{1}{2\alpha}$ we have

$$\|A\,x_\alpha - y\|\|\omega\| \le \frac{1}{2\alpha}\|A\,x_\alpha - y\|^p + (2\alpha)^{\frac{1}{p-1}} \frac{\|\omega\|^{\frac{p}{p-1}}}{p^{\frac{1}{p-1}}\frac{p}{p-1}}.$$

Hence, with

$$c(p) := 2^{1+\frac{1}{1-p}} \frac{p-1}{p^{1+\frac{1}{p-1}}} = 2^{\frac{2-p}{1-p}}(p-1)p^{\frac{p}{1-p}} = 2^{\frac{p-2}{p-1}}(p-1)p^{\frac{p}{1-p}}$$

we deduce

$$\|A\,x_\alpha - y\|^p + \alpha\|x_\alpha - x^\dagger\|^2 \le \|A\,x_\alpha - y\|^p + \alpha c(p)\alpha^{\frac{1}{p-1}}\|\omega\|^{\frac{p}{p-1}} + 2\alpha\|x^\dagger - x_\alpha\|\|r\|$$

and further

$$\|x_\alpha - x^\dagger\|^2 \le c(p)\alpha^{\frac{1}{p-1}}\|\omega\|^{\frac{p}{p-1}} + 2\|x^\dagger - x_\alpha\|\|r\|.$$

Now with $\|\omega\| \le R$, $\|r\| \le \rho(R)$ and (8) the estimate (9) yields.

In the case of given noisy data we obtain

$$\begin{aligned}
\|A\,x_\alpha^\delta - y^\delta\|^p + \alpha\|x_\alpha^\delta - x^\dagger\|^2 &\le \delta^p + 2\alpha\langle x^\dagger - x_\alpha^\delta, x^\dagger\rangle \\
&= \delta^p + 2\alpha\langle x^\dagger - x_\alpha^\delta, A^*\omega + r\rangle \\
&= \delta^p + 2\alpha\left[\langle y - y^\delta + y^\delta - A\,x_\alpha^\delta, \omega\rangle + \langle x^\dagger - x_\alpha^\delta, r\rangle\right] \\
&\le \delta^p + 2\alpha\left[\delta\|\omega\| + \|A\,x_\alpha^\delta - y^\delta\|\|\omega\| + \|x^\dagger - x_\alpha^\delta\|\|r\|\right].
\end{aligned}$$

Using Young's inequality with $\varepsilon := \frac{1}{2\alpha}$ twice we derive

$$\begin{aligned}
\|A\,x_\alpha^\delta - y^\delta\|^p + \alpha\|x_\alpha^\delta - x^\dagger\|^2 \le\ & 2\delta^p + \|A\,x_\alpha^\delta - y^\delta\|^p + 2\alpha c(p)\alpha^{\frac{1}{p-1}}\|\omega\|^{\frac{p}{p-1}} \\
& + 2\alpha\|x^\dagger - x_\alpha^\delta\|\|r\|
\end{aligned}$$

which proves (10). ■

The case $p = 1$ plays a singular role, since we cannot apply Young's inequality. Moreover, it turns out that for giving error bounds we have to distinguish between $R > \frac{1}{2\alpha}$ and $R \le \frac{1}{2\alpha}$. The corresponding results can be found in the following two lemmas. The first statement deals with the noise-free case.

Lemma 2.2 *Assume $x^\dagger \in \mathcal{M}_\rho(R)$ for some $R > 0$ and a function $\rho = \rho(R)$. Then, for $p = 1$ he have*

$$\left\|x_\alpha - x^\dagger\right\| \le \begin{cases} 2\rho(R), & R \le \dfrac{1}{2\alpha}, \\ \sqrt{2\alpha R - 1}\|x^\dagger\| + 2\rho(R), & R > \dfrac{1}{2\alpha}. \end{cases} \tag{11}$$

Proof. Again we set $x^\dagger = A^*\omega + r$ with $\|\omega\| \le R$ and $\|r\| \le \rho(R)$. We have by the minimizing property of x_α

$$\|A\,x_\alpha - y\| + \alpha\|x_\alpha\|^2 \le \|A\,x^\dagger - y\| + \alpha\|x^\dagger\|^2 \le \alpha\|x^\dagger\|^2.$$

Hence $\|A\,x_\alpha - y\| \le \alpha \|x^\dagger\|^2$. Following the lines of the proof of Lemma 2.1 we have

$$\begin{aligned}\|A\,x_\alpha - y\| + \alpha\|x_\alpha - x^\dagger\|^2 &\le 2\alpha\left(\|y - A\,x_\alpha\|\,\|\omega\| + \|x^\dagger - x_\alpha\|\,\|r\|\right)\\ &\le 2\alpha R\|y - A\,x_\alpha\| + 2\alpha\,\rho(R)\|x^\dagger - x_\alpha\|\end{aligned}$$

For $R \le \frac{1}{2\alpha}$ we skip the terms including $\|A\,x_\alpha - y\|$ and obtain the first estimate. For $R > \frac{1}{2\alpha}$ we have

$$\begin{aligned}\alpha\|x_\alpha - x^\dagger\|^2 &\le (2\alpha R - 1)\|y - A\,x_\alpha\| + 2\alpha\,\rho(R)\|x^\dagger - x_\alpha\|\\ &\le (2\alpha R - 1)\alpha\|x^\dagger\|^2 + 2\alpha\,\rho(R)\|x^\dagger - x_\alpha\|.\end{aligned}$$

By the implication (8) we derive (11). ■

In the case of noisy data we get the following result.

Lemma 2.3 *Assume $x^\dagger \in \mathcal{M}_\rho(R)$ for some $R > 0$ and a function $\rho = \rho(R)$. Then, for $p = 1$ we have*

$$\left\|x_\alpha^\delta - x^\dagger\right\| \le \begin{cases} \sqrt{\delta\left(\dfrac{1}{\alpha} + 2R\right)} + 2\rho(R), & R \le \dfrac{1}{2\alpha},\\ \sqrt{4\delta R + (2\alpha R - 1)\|x^\dagger\|^2} + 2\rho(R), & R > \dfrac{1}{2\alpha}.\end{cases} \tag{12}$$

Proof. Again we set $x^\dagger = A^*\omega + r$. By

$$\|A\,x_\alpha^\delta - y^\delta\| + \alpha\|x_\alpha^\delta\|^2 \le \|A\,x^\dagger - y^\delta\| + \alpha\|x^\dagger\|^2 \le \delta + \alpha\|x^\dagger\|^2,$$

we obtain $\|A\,x_\alpha^\delta - y^\delta\| \le \delta + \alpha\|x^\dagger\|^2$ and

$$\begin{aligned}\|A\,x_\alpha^\delta - y^\delta\| + \alpha\|x_\alpha^\delta - x^\dagger\|^2 &\le \delta + 2\alpha\left[\delta\|\omega\| + \|A\,x_\alpha^\delta - y^\delta\|\|\omega\| + \|x^\dagger - x_\alpha^\delta\|\|r\|\right]\\ &\le \delta(1 + 2\alpha R) + 2\alpha R\|A\,x_\alpha^\delta - y^\delta\| + 2\alpha\,\rho(R)\|x^\dagger - x_\alpha^\delta\|.\end{aligned}$$

For $R \le \frac{1}{2\alpha}$ we skip the terms including $\|A\,x_\alpha^\delta - y^\delta\|$ and obtain the first result. For $R > \frac{1}{2\alpha}$ we have

$$\begin{aligned}\alpha\|x_\alpha^\delta - x^\dagger\|^2 &\le \delta(1 + 2\alpha R) + (2\alpha R - 1)\|A\,x_\alpha^\delta - y^\delta\| + 2\alpha\,\rho(R)\|x^\dagger - x_\alpha^\delta\|\\ &\le \delta(1 + 2\alpha R) + (2\alpha R - 1)(\delta + \alpha\|x^\dagger\|^2) + 2\alpha\,\rho(R)\|x^\dagger - x_\alpha^\delta\|\\ &= 4\delta\alpha R + (2\alpha R - 1)\|x^\dagger\|^2 + 2\alpha\,\rho(R)\|x^\dagger - x_\alpha^\delta\|\end{aligned}$$

which proves the second estimate by the same argument as above. ■

On the other hand, if $x^\dagger$ fulfills the source condition (6) we immediately obtain similar results as in [2, Theorem 4 and 5] by setting $d(R) = 0$ for $R \ge \|\omega\|$.

Corollary 2.1 *Assume, that the source condition (6) holds. Then, for $\alpha \le \frac{1}{2\|\omega\|}$ we have*

$$x_\alpha = x^\dagger \qquad \textit{and} \qquad \|x_\alpha^\delta - x^\dagger\| \le \sqrt{\delta\left(\frac{1}{\alpha} + 2\|\omega\|\right)} \le \sqrt{\frac{2\delta}{\alpha}}.$$

For any fixed choice $0 < \alpha \le \frac{1}{2\|\omega\|}$ we obtain a convergence rate $\|x_\alpha^\delta - x^\dagger\| \sim \mathcal{O}(\sqrt{\delta})$.

This highlights the exceptional position of the case $p = 1$: as opposite to $p > 1$ we can here achieve convergence rates by choosing the regularization parameter α independently of the noise level δ if the source condition (6) is satisfied. It turns also out, that the choice $\alpha := \frac{1}{2\|\omega\|}$ yields the best error estimate.

3 Convergence rates

The considerations above present a suitable way for formulating convergence rates, if the source condition (6) is violated. To do this, we define distance functions, originally introduced in [1, Theorem 6.8].

Definition 3.1 *For given $x \in \mathcal{X}$ the distance function $d(\cdot; x) : [0, \infty) \longrightarrow \mathbb{R}$ is defined as*

$$d(R; x) := \inf \{\|x - A^*\omega\| \,:\, \omega \in \mathcal{Y}^*,\ \|\omega\| \le R\}\,. \tag{13}$$

Note, that the nonnegative function $d(R; x)$ is well-defined. It is non-increasing with $d(R; x) \to 0$ for $R \to \infty$, if $x \in \overline{\mathcal{R}(A^*)}$. Moreover, $d(R) > 0$ for all $R \ge 0$ if $x \notin \mathcal{R}(A^*)$ and $d(R; x) = 0$ for all $R \ge \|\omega\|$ if $x = A^*\omega$ and $\|\omega\| = R$. In fact, the distance function $d(R; x)$ gives us a quantity to measure the violation of the source condition (6).

Now we can prove convergence rates based on the decay rate of the distance function $d(R) := d(R; x^\dagger)$ of the exact solution $x^\dagger \in \mathcal{X}$.

Theorem 3.1 *Let $x^\dagger \in \overline{\mathcal{R}(A^*)} \setminus \mathcal{R}(A^*)$ with distance function $d = d(R)$ and $\Theta(R) := R^{-1}d(R)^2$. Then, for $p > 1$ and $\tilde{\Theta}(R) := R^{-p}d(R)^{2(p-1)}$ the convergence rate*

$$\|x_\alpha - x^\dagger\| \sim \mathcal{O}\left(d\left(\tilde{\Theta}^{-1}(\alpha)\right)\right). \tag{14}$$

yields. For an a-priori choice $\alpha := \Psi^{-1}(\delta)$ with $\Psi(\alpha) := \left(\alpha\, d\left(\tilde{\Theta}^{-1}(\alpha)\right)^2\right)^{\frac{1}{p}}$ the convergence rate

$$\|x_\alpha^\delta - x^\dagger\| \sim \mathcal{O}\left(d\left(\Theta^{-1}(\delta)\right)\right) \tag{15}$$

holds.

Proof. Since $\mathcal{Y}$ is reflexive, there exists $\omega = \omega(R)$ with

$$d(R) = \|x^\dagger - A^*\omega(R)\| \qquad \text{for all } R > 0,$$

see e.g. [15, Theorem 38.A]. By the definition of the distance function we have $x^\dagger \in \mathcal{M}_d(R)$ for all $R > 0$. By balancing we get

$$\alpha^{\frac{1}{2(p-1)}} = R^{\frac{p}{2(1-p)}}\, d(R) \quad \Leftrightarrow \quad \alpha = R^{-p}d(R)^{2(p-1)} =: \tilde{\Theta}(R) \quad \Leftrightarrow \quad R = \tilde{\Theta}^{-1}(\alpha).$$

which gives the estimate

$$\|x_\alpha - x^\dagger\| \le \left(\sqrt{c(p)} + 2\right) d\left(\tilde{\Theta}^{-1}(\alpha)\right).$$

Setting $R := \tilde{\Theta}^{-1}(\alpha)$ we have

$$\|x_\alpha^\delta - x^\dagger\| \le \sqrt{\frac{2\delta^p}{\alpha} + 2c(p)d\left(\Theta^{-1}(\alpha)\right)^2} + 2d\left(\tilde{\Theta}^{-1}(\alpha)\right).$$

Balancing these terms we set

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

Moreover,

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

which gives $R = \Theta^{-1}(\delta)$. Hence, choosing $\alpha := \Psi^{-1}(\delta)$ we obtain the estimate

$$\begin{aligned} \|x_\alpha^\delta - x^\dagger\| &\leq \left(\sqrt{2+2c(p)}+2\right)\sqrt{\frac{\delta^p}{\Psi^{-1}(\delta)}} \\ &= \left(\sqrt{2+2c(p)}+2\right) d\left(\tilde{\Theta}^{-1}\left(\Psi^{-1}(\delta)\right)\right) \\ &= \left(\sqrt{2+2c(p)}+2\right) d\left(\Theta^{-1}(\delta)\right). \end{aligned}$$

The proof is complete. ■

Note, since $d(R)$ is a non-increasing function and $p > 1$ we see, that $\tilde{\Theta}(R)$ is strictly decreasing on $(0, \infty)$ with

$$\lim_{R\to 0} \tilde{\Theta}(R) = \infty \quad \text{and} \quad \lim_{R\to\infty} \tilde{\Theta}(R) = 0.$$

Hence $\tilde{\Theta}^{-1}(\alpha)$ is well-defined and strictly decreasing. The same holds true for $\Theta(R)$. Moreover, we observe that $d(\tilde{\Theta}^{-1}(\alpha))$ is an increasing function, which shows that $\Psi(\alpha)$ is strictly increasing with

$$\lim_{\alpha\to 0} \Psi(\alpha) = 0 \quad \text{and} \quad \lim_{\alpha\to\infty} \Psi(\alpha) = \infty.$$

Hence, $\Psi^{-1}(\delta)$ is also well-defined and strictly increasing.

We also want to point out, that the convergence rate (15) does not depend on the choice of p. The rate (15) only depends on the distance function $d(R)$. This observation does not seem to be surprisingly. In linear regularization theory, a optimal convergence rates do not depend on the choice of the regularization method but on satisfied source conditions as long as the qualification of the regularization method is not reached. The same holds true here, although we cannot prove the optimality of the rate (15) for general Banach space $\mathcal{Y}$.

Remark 3.1 *The existence of an element $\omega = \omega(R)$ with $\|\omega\| \leq R$ and $d(R) = \min \|x^\dagger - A^*\omega\|$ is not a necessary condition in the proof. If the infimum of (13) is not attained, then clearly $x^\dagger \in \mathcal{M}_{\tilde{d}}(R)$ with $\tilde{d}(R) := d(R) + \eta$ for all $R > 0$ and all $\eta > 0$. Taking the limit $\eta \to 0$ we again obtain the estimates (14) and (15).*

The case $p = 1$ can be treated similarly. It turns out, that for balancing the choice of R only depends on the regularization parameter α.

Theorem 3.2 *Let $x^\dagger \in \overline{\mathcal{R}(A^*)} \setminus \mathcal{R}(A^*)$ with distance function $d = d(R)$ and let $\Theta(R) := R^{-1}d(R)^2$. Then*

$$\|x_\alpha - x^\dagger\| \le 2d\left(\frac{1}{2\alpha}\right). \tag{16}$$

An a priori parameter choice $\alpha := \Psi^{-1}(\delta)$ with $\Psi(\alpha) := \Theta\left(\frac{1}{2\alpha}\right)$ gives the convergence rate (15).

PROOF. Again we notice $x^\dagger \in \mathcal{M}_d(R)$ for all $R > 0$. We set $R := \frac{1}{2\alpha}$. The estimate (16) follows directly from (11). For given noisy data we derive

$$\|x_\alpha^\delta - x^\dagger\| \le \sqrt{4\delta R} + 2d(R) = 2\left(\sqrt{\delta R} + d(R)\right) = \sqrt{\frac{2\delta}{\alpha}} + 2d\left(\frac{1}{2\alpha}\right).$$

Balancing both terms in the brackets we have

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

which gives

$$\|x_\alpha^\delta - x^\dagger\| \le 4\, d\left(\Theta^{-1}(\delta)\right).$$

This proves the convergence rate (15). ■

Hence, both theorems indicate, that for all $p \ge 1$ we obtain the same (optimal) convergence rate (15).

4 An illustrating example

We will show that presenting convergence rates in terms of distance functions seems to be a natural generalization of formulating convergence rates for elements satisfying a generalized source condition in Hilbert spaces. In fact, there is a close relation between the distance function and generalized source conditions in Hilbert spaces. The following result can be found in [7, Theorem 1].

Proposition 4.1 *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 $x^\dagger \notin \mathcal{R}(A^*)$ be the solution of $Ax = y$ with distance function $d(R)$. If $x^\dagger \in \mathcal{R}\left((A^*A)^{\eta/2}\right)$, $0 < \eta < 1$, then the estimate*

$$d(R) \le K\, R^{\frac{\eta}{\eta - 1}}, \quad R > 0, \tag{17}$$

holds for some constant $K > 0$.

A carful reading of the proof show, that the estimate (17) is of optimal order. Let us therefore assume, that the distance function is given by $d(R) := K\, R^{\frac{\eta}{\eta-1}}$ for $K > 0$ and $0 < \eta < 1$. From (16) we derive

$$\|x_\alpha - x^\dagger\| \le 2K(2\alpha)^{\frac{\eta}{1-\eta}} = 2^{\frac{1}{1-\eta}} K \alpha^{\frac{\eta}{1-\eta}}.$$

Moreover we have

$$\tilde{\Theta}(R) = K^2 R^{\frac{2\eta}{\eta-1}-1} = K^2 R^{\frac{\eta+1}{\eta-1}}$$

and hence

$$\Theta(\alpha) = K^2 2^{\frac{\eta+1}{1-\eta}} \alpha^{\frac{\eta+1}{1-\eta}} =: C\alpha^{\frac{\eta+1}{1-\eta}} \quad \text{and} \quad \alpha := \Theta(\delta)^{-1} = \tilde{C}\delta^{\frac{1-\eta}{1+\eta}}$$

for two constants $C, \tilde{C} > 0$. With

$$\tilde{\Theta}(\delta)^{-1} = \left(\frac{\delta}{K^2}\right)^{\frac{\eta-1}{\eta+1}}$$

we derive

$$d\left(\tilde{\Theta}^{-1}(\delta)\right) = K\left[\left(\frac{\delta}{K^2}\right)^{\frac{\eta-1}{\eta+1}}\right]^{\frac{\eta}{\eta-1}} = K^{\frac{1-\eta}{1+\eta}}\delta^{\frac{\eta}{\eta+1}},$$

which gives

$$\|x_\alpha^\delta - x^\dagger\| \sim \mathcal{O}\left(\delta^{\frac{\eta}{\eta+1}}\right). \tag{18}$$

The convergence rate (18) is known to be the optimal one for linear regularization methods such as Tikhonov regularization if $x^\dagger \in \mathcal{R}\left((A^*A)^{\frac{\eta}{2}}\right)$ for $0 \le \eta \le 1$, see e.g. [4, Section 5.1]. Hence, generalized source conditions and distance functions provide the same (optimal) convergence rates. Moreover, if $\mathcal{Y}$ is supposed to be a Hilbert space then we can obtain the same convergence rates of Theorem 3.1 and 3.2 by applying interpolation inequalities. This, however, is again a mathematical tool in Hilbert spaces.

For the limit $\eta \to 1$ we have

$$\frac{\eta}{1-\eta} \to \infty \quad \text{and} \quad \frac{\eta}{1+\eta} \to \frac{1}{2}.$$

Hence, for $\eta \to 1$ the convergence $x_\alpha \to x^\dagger$ becomes faster than any power type decay in the noiseless case, whereas for noisy data the rate converges to the reference rate $\mathcal{O}(\sqrt{\delta})$, which is obtained, when the source condition (6) is fulfilled.

Example 4.1 *We consider the classical case $p = 2$ and suppose $\mathcal{Y}$ to be a Hilbert space. Moreover, we assume $x^\dagger \in \mathcal{R}\left((A^*A)^{\frac{\eta}{2}}\right)$ for $0 < \eta < 1$. Then, for the distance function $d(R)$ the estimate $d(R) \le K\, R^{\frac{\eta}{\eta-1}}$ holds. Hence, by balancing*

$$\alpha = R^{-2}\left[R^{\frac{\eta}{\eta-1}}\right]^{2(2-1)} = R^{\frac{2}{\eta-1}} =: \Theta(R).$$

which gives $R := \Theta^{-1}(\alpha) = \alpha^{\frac{\eta-1}{2}}$ and finally the convergence rate

$$\|x_\alpha - x^\dagger\| \sim \left[\alpha^{\frac{\eta-1}{2}}\right]^{\frac{\eta}{\eta-1}} = \alpha^{\frac{\eta}{2}}.$$

For noisy data we still balance

$$\delta = \sqrt{\alpha}\alpha^{\frac{\eta}{2}} = \alpha^{\frac{\eta+1}{2}} =: \Psi(\alpha)$$

which provides $\alpha := \Psi^{-1}(\delta) = \delta^{\frac{2}{\eta+1}}$. *Hence the convergence rate*

$$\|x_\alpha^\delta - x^\dagger\| \sim \frac{\delta}{\sqrt{\Psi^{-1}(\delta)}} = \delta^{1-\frac{1}{\eta+1}} = \delta^{\frac{\eta}{\eta+1}}$$

holds. These are exactly the same convergence rates results obtained by standard linear regularization theory under the corresponding source condition.

The example shows that by dealing with approximative source conditions we can extend classical convergence rate results to linear (and nonlinear) operators mapping from a Hilbert space into a reflexive Banach space.

Remark 4.1 *This example shows also the drawback of the idea of distance functions. The reference source condition (6) implies a limit convergence rate* $\mathcal{O}(\sqrt{\delta})$. *We cannot prove faster convergence rates in that case. In order to achieve faster convergence rates we have to replace (6) somehow by a stronger source condition, which has to represent an equivalent condition to* $x^\dagger \in \mathcal{R}(A^*A)$ *in Hilbert spaces.*

5 Extension to the nonlinear case

We replace (1) by the nonlinear equation

$$F(x) = y^\delta, \qquad x \in \mathcal{X},\ y^\delta \in \mathcal{Y}, \tag{19}$$

and consider the problem

$$\|F(x) - y^\delta\|^p + \alpha\|x - x^*\|^2 \to \min \quad \text{subject to } x \in \mathcal{X}. \tag{20}$$

for given noisy data y^δ and an a priori guess $x^* \in \mathcal{Y}$. Thereby $F : \mathcal{D}(F) \subseteq \mathcal{X} \longrightarrow \mathcal{Y}$ denotes a nonlinear operator with domain $\mathcal{D}(F)$. We assume F to be continuous and weakly sequentially closed. Then, by simple modifications, we can apply the analysis of [3] which guarantee existence, stability and convergence results. Again we denote with x_α and x_α^δ solutions of (20) for given noiseless data $y = y^0$ and noisy data y^δ. For simplicity we assume again, that there exists a unique solution $x^\dagger \in \mathcal{D}(F)$ of (19) in the case of noiseless data.

In order to prove convergence rates we need a further assumption which restrict the nonlinearity of the operator F.

Assumption 5.1 *There exists a linear and bounded operator* $G \in L(\mathcal{X}, \mathcal{Y})$ *such that*

$$\|F(x) - F(x^\dagger) - G(x - x^\dagger)\| \le \gamma\|F(x) - F(x^\dagger)\|, \qquad \forall\, x \in \mathcal{B}_\varrho(x^\dagger) \cap \mathcal{D}(F) \tag{21}$$

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

If $x^\dagger$ is an interior point of $\mathcal{D}(F)$ and there exists the Fréchet-derivative $F'(x^\dagger)$ we can set $G := F'(x^\dagger)$, which is common practice. Conditions of the form (21) where originally introduced in [5] for proving convergence rates for Landweber iteration. As in opposite to [5] no upper bound to γ is necessary in our situation. It turns out, that the standard assumption

$$\|F(x) - F(x^\dagger) - G(x - x^\dagger)\| \le \frac{L}{2}\|x - x^\dagger\|^2, \qquad \forall\, x \in \mathcal{B}_\varrho(x^\dagger) \cap \mathcal{D}(F)$$

with a constant $L > 0$ for obtaining convergence rates in nonlinear Tikhonov regularization, see e.g. [3], is not strong enough for proving convergence rates for weaker source conditions than

$$x^\dagger - x^* = G^*\omega \quad \text{for some } \omega \in \mathcal{Y}^*.$$

This does not seem surprisingly, since this issue was already observed while considering generalized source conditions of power type for nonlinear Tikhonov regularization in Hilbert spaces, see [4, Theorem 10.7].

We now present error bounds. We start with the noiseless case.

Lemma 5.1 *Let the sets $\mathcal{M}_\rho(R)$ be defined by (7) with G^* instead of A^* and some function $\rho = \rho(R)$. Assume (21) to be hold and $x^\dagger - x^* \in \mathcal{M}_\rho(R)$ for some $R > 0$. Then, for $p = 1$,*

$$\|x_\alpha - x^\dagger\| \le \begin{cases} 2\,\rho(R), & R \le \dfrac{1}{2\alpha(1+\gamma)}, \\ \sqrt{2\alpha R(1+\gamma) - 1}\,\|x^\dagger - x^*\| + 2\,\rho(R), & R > \dfrac{1}{2\alpha(1+\gamma)}. \end{cases} \tag{22}$$

and in the case $p > 1$

$$\|x_\alpha - x^\dagger\| \le \sqrt{\tilde c(p)}\,\alpha^{\frac{1}{2(p-1)}} R^{\frac{p}{2(p-1)}} + 2\,\rho(R), \tag{23}$$

with a constant $\tilde c(p) > 0$, depending only on p (and γ).

Proof. In the noiseless case we have $\|F(x_\alpha) - y\|^p \le \alpha\|x^\dagger - x^*\|^2$. Moreover, with the notation $R(x, x^\dagger) := F(x) - F(x^\dagger) - G(x - x^\dagger)$ for $x \in \mathcal{D}(F)$ we conclude

$$\begin{aligned}
\|F(x_\alpha) - y\|^p + \alpha\|x_\alpha - x^\dagger\|^2 &\le 2\alpha\langle x^\dagger - x_\alpha, x^\dagger - x^*\rangle \\
&= 2\alpha\langle x^\dagger - x_\alpha, G^*\omega + r\rangle \\
&= 2\alpha\left(\langle G(x^\dagger - x_\alpha), \omega\rangle + \langle x^\dagger - x_\alpha, r\rangle\right) \\
&= 2\alpha\left(\langle F(x_\alpha) - F(x^\dagger) + R(x_\alpha, x^\dagger), \omega\rangle + \langle x^\dagger - x_\alpha, r\rangle\right) \\
&\le 2\alpha\left((1+\gamma)\|y - F(x_\alpha)\|\,\|\omega\| + \|x^\dagger - x_\alpha\|\,\|r\|\right) \\
&\le 2\alpha(1+\gamma)\|y - F(x_\alpha)\|\,R + 2\alpha\,\rho(R)\|x^\dagger - x_\alpha\|.
\end{aligned}$$

Let us now assume $p > 1$. By applying Young's inequality with $\tilde\varepsilon := \frac{1}{2\alpha(1+\gamma)}$ we derive

$$\|F(x_\alpha) - y\|^p + \alpha\|x_\alpha - x^\dagger\|^2 \le \|F(x_\alpha) - y\|^p + \alpha\,\tilde c(p)\alpha^{\frac{1}{p-1}} R^{\frac{p}{p-1}} + 2\alpha\,\rho(R)\|x^\dagger - x_\alpha\|.$$

with

$$\tilde c(p) := [2(1+\gamma)]^{\frac{p-2}{p-1}}\,(p-1)p^{\frac{p}{1-p}}.$$

Hence

$$\|x_\alpha - x^\dagger\|^2 \le \tilde{c}(p)\alpha^{\frac{1}{p-1}} R^{\frac{p}{p-1}} + 2\,\rho(R)\|x^\dagger - x_\alpha\|,$$

which by the application of (8) provides

$$\|x_\alpha - x^\dagger\| \le \sqrt{\tilde{c}(p)}\alpha^{\frac{1}{2(p-1)}} R^{\frac{p}{2(p-1)}} + 2\,\rho(R).$$

For the case $p=1$ we have to differentiate again. If $R \le \frac{1}{2\alpha(1+\gamma)}$ we follow the argumentation in the proof of Lemma 2.2, which shows $\|x_\alpha - x^\dagger\| \le 2\,\rho(R)$. For $R > \frac{1}{2\alpha(1+\gamma)}$ we have

$$\begin{aligned}
\alpha\|x_\alpha - x^\dagger\|^2 &\le (2\alpha(1+\gamma)\,R - 1)\|y - F(x_\alpha)\| + 2\alpha\,\rho(R)\|x^\dagger - x_\alpha\| \\
&\le \alpha(2\alpha(1+\gamma)\,R - 1)\|x^\dagger - x^*\|^2 + 2\alpha\,\rho(R)\|x^\dagger - x_\alpha\|.
\end{aligned}$$

Dividing by α an using (8) we obtain the second result. ■

Lemma 5.2 *Let the sets $\mathcal{M}_\rho(R)$ be defined by (7) with G^* instead of A^* and some function $\rho = \rho(R)$. Assume (21) to be hold and $x^\dagger - x^* \in \mathcal{M}_\rho(R)$ for some $R > 0$. Then, for $p=1$,*

$$\|x_\alpha^\delta - x^\dagger\| \le \begin{cases} \sqrt{\delta\left(\dfrac{1}{\alpha} + 2(1+\gamma)R\right)} + 2\,\rho(R), & R \le \dfrac{1}{2\alpha(1+\gamma)}, \\ \sqrt{4\delta R(1+\gamma) + (2\alpha R(1+\gamma) - 1)\|x^\dagger - x^*\|^2} + 2\,\rho(R), & R > \dfrac{1}{2\alpha(1+\gamma)}. \end{cases} \tag{24}$$

and in the case $p > 1$

$$\|x_\alpha^\delta - x^\dagger\| \le \sqrt{\frac{2\delta^p}{\alpha} + 2\tilde{c}(p)\alpha^{\frac{1}{p-1}} R^{\frac{p}{p-1}}} + 2\,\rho(R). \tag{25}$$

Proof. In the noisy case we have $\|F(x_\alpha^\delta) - y^\delta\|^p \le \delta^p + \alpha\|x^\dagger - x^*\|^2$. Following the lines of the proof above we deduce

$$\begin{aligned}
\|F(x_\alpha^\delta) - y^\delta\|^p + \alpha\|x_\alpha^\delta - x^\dagger\|^2 &\le \delta^p + 2\alpha\langle x^\dagger - x_\alpha^\delta, x^\dagger - x^*\rangle \\
&= 2\delta^p + \alpha\langle x^\dagger - x_\alpha^\delta, G^*\omega + r\rangle \\
&= \delta^p + 2\alpha\left(\langle G(x^\dagger - x_\alpha^\delta), \omega\rangle + \langle x^\dagger - x_\alpha^\delta, r\rangle\right) \\
&= \delta^p + 2\alpha\langle F(x_\alpha^\delta) - F(x^\dagger) + R(x_\alpha^\delta, x^\dagger), \omega\rangle \\
&\quad\; + 2\alpha\langle x^\dagger - x_\alpha^\delta, r\rangle \\
&= \delta^p + 2\alpha\langle F(x_\alpha^\delta) - y^\delta + y^\delta - y + R(x_\alpha^\delta, x^\dagger), \omega\rangle \\
&\quad\; + 2\alpha\langle x^\dagger - x_\alpha^\delta, r\rangle \\
&\le \delta^p + 2\alpha(1+\gamma)\left(\|y^\delta - F(x_\alpha^\delta)\|\,\|\omega\| + \delta\|\omega\|\right) \\
&\quad\; + 2\alpha\|x^\dagger - x_\alpha^\delta\|\,\|r\| \\
&\le \delta^p + 2\alpha\,\delta R + 2\alpha(1+\gamma)\|y^\delta - F(x_\alpha^\delta)\|\,R \\
&\quad\; + 2\alpha\,\rho(R)\|x^\dagger - x_\alpha^\delta\|.
\end{aligned}$$

Here we used

$$\|R(x_\alpha^\delta, x^\dagger)\| \le \gamma\|F(x_\alpha^\delta) - F(x^\dagger)\| \le \gamma\left(\|F(x_\alpha^\delta) - y^\delta\| + \delta\right).$$

For $p>1$ we apply Young's inequality with $\varepsilon := \frac{1}{2\alpha(1+\gamma)}$ twice for obtaining

$$2\alpha(1+\gamma)\,\delta R \le \delta^p + \alpha\,\tilde{c}(p)\alpha^{\frac{1}{p-1}} R^{\frac{p}{p-1}}.$$

and

$$2\alpha(1+\gamma)\|y^\delta - F(x_\alpha^\delta)\|\,R \le \|y^\delta - F(x_\alpha^\delta)\|^p + \alpha\,\tilde{c}(p)\alpha^{\frac{1}{p-1}} R^{\frac{p}{p-1}}.$$

Using these estimate we derive

$$\alpha\|x_\alpha^\delta - x^\dagger\|^2 \le 2\delta^p + 2\alpha\tilde{c}(p)\alpha^{\frac{1}{p-1}} R^{\frac{p}{p-1}} + 2\alpha\,\rho(R)\|x^\dagger - x_\alpha^\delta\|,$$

which provides

$$\|x_\alpha^\delta - x^\dagger\| \le \sqrt{\frac{2\delta^p}{\alpha} + 2\tilde{c}(p)\alpha^{\frac{1}{p-1}} R^{\frac{p}{p-1}}} + 2\,\rho(R).$$

For $p=1$ and $R \le \frac{1}{2\alpha(1+\gamma)}$ we have

$$\alpha\|x_\alpha^\delta - x^\dagger\|^2 \le \delta + 2\alpha(1+\gamma)\delta R + 2\alpha\,\rho(R)\|x_\alpha^\delta - x^\dagger\|,$$

which implicates the first part of (24). For $R > \frac{1}{2\alpha(1+\gamma)}$ we deduce

$$\begin{aligned}
\alpha\|x_\alpha^\delta - x^\dagger\|^2 &\le \delta + 2\alpha(1+\gamma)\delta R + [2\alpha(1+\gamma)R - 1]\left(\delta + \alpha\|x^\dagger - x^*\|^2\right) \\
&\qquad + 2\alpha\,\rho(R)\|x_\alpha^\delta - x^\dagger\| \\
&= 4\alpha\,\delta R(1+\gamma) + \alpha\,[2\alpha(1+\gamma)R - 1]\,\|x^\dagger - x^*\|^2 + 2\alpha\,\rho(R)\|x_\alpha^\delta - x^\dagger\|,
\end{aligned}$$

which proves the second estimate. ■

Hence, by choosing $R := \frac{1}{2\alpha(1+\gamma)}$ for $p=1$ we can prove convergence rates in the same way as in the linear situation.

Theorem 5.1 *Let $x^\dagger \in \overline{\mathcal{R}(G^*)} \setminus \mathcal{R}(G^*)$ with distance function $d = d(R)$, which is defined by (13) with G^* instead of A^* and let $\Theta(R) := R^{-1}d(R)^2$. Then, for $p=1$,*

$$\|x_\alpha - x^\dagger\| \le 2d\left(\frac{1}{2\alpha(1+\gamma)}\right). \tag{26}$$

An a priori parameter choice $\alpha := \Psi^{-1}(\delta)$ with $\Psi(\alpha) := \Theta\left(\frac{1}{2\alpha(1+\gamma)}\right)$ gives the convergence rate

$$\|x_\alpha^\delta - x^\dagger\| \sim \mathcal{O}\left(d\left(\Theta^{-1}(\delta)\right)\right). \tag{27}$$

Proof. We set $R := \frac{1}{2\alpha(1+\gamma)}$. Then (26) follows directly from (22). In the noisy case we deduce

$$\|x_\alpha^\delta - x^\dagger\| \le 2\left(\sqrt{(1+\gamma)\delta\,R} + d(R)\right) = \sqrt{\frac{2\delta}{\alpha}} + 2\,d\left(\frac{1}{2\alpha(1+\gamma)}\right).$$

Following the proof of Theorem 3.2, we derive

$$\|x_\alpha^\delta - x^\dagger\| \le 2\left(\sqrt{1+\gamma} + 1\right) d\left(\Theta^{-1}(\delta)\right),$$

which proves the convergence rate (27). ■

For $p>1$ the convergence rate result is similar as in the linear case as well.

Theorem 5.2 *Let $x^\dagger \in \overline{\mathcal{R}(G^*)} \setminus \mathcal{R}(G^*)$ with distance function $d = d(R)$, which is defined by (13) with G^* instead of A^* and let $\Theta(R) := R^{-1}d(R)^2$. Then, for $p > 1$ and $\tilde{\Theta}(R) := R^{-p}d(R)^{2(p-1)}$,*

$$\|x_\alpha - x^\dagger\| \sim \mathcal{O}\left(d\left(\tilde{\Theta}^{-1}(\alpha)\right)\right). \tag{28}$$

For an a priori choice $\alpha := \Psi^{-1}(\delta)$ with $\Psi(\alpha) := \left(\alpha\, d\left(\tilde{\Theta}^{-1}(\alpha)\right)^2\right)^{\frac{1}{p}}$ the convergence rate (27) holds.

Proof. We have $x^\dagger \in \mathcal{M}_d(R)$ for all $R > 0$. By balancing we get

$$\alpha^{\frac{1}{2(p-1)}} = R^{\frac{p}{2(1-p)}}\, d(R) \quad \Leftrightarrow \quad \alpha = R^{-p}d(R)^{2(p-1)} =: \tilde{\Theta}(R) \quad \Leftrightarrow \quad R = \tilde{\Theta}^{-1}(\alpha).$$

which gives the estimate

$$\|x_\alpha - x^\dagger\| \le \left(\sqrt{\tilde{c}(p)} + 2\right) d\left(\tilde{\Theta}^{-1}(\alpha)\right).$$

Setting $R = \tilde{\Theta}^{-1}(\alpha)$ we have

$$\|x_\alpha^\delta - x^\dagger\| \le \sqrt{\frac{2\delta^p}{\alpha} + 2\tilde{c}(p)d\left(\tilde{\Theta}^{-1}(\alpha)\right)^2} + 2d\left(\tilde{\Theta}^{-1}(\alpha)\right).$$

Balancing these terms we set

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

or in terms of R (see Theorem 3.1)

$$\delta = \frac{d(R)^2}{R} = \Theta(R) \quad \Leftrightarrow \quad R = \Theta^{-1}(\delta).$$

Hence, choosing $\alpha := \Psi^{-1}(\delta)$ we obtain the estimate

$$\begin{aligned}
\|x_\alpha^\delta - x^\dagger\| &\le \left(\sqrt{2 + 2\tilde{c}(p)} + 2\right)\sqrt{\frac{\delta^p}{\Psi^{-1}(\delta)}} \\
&= \left(\sqrt{2 + 2\tilde{c}(p)} + 2\right) d\left(\tilde{\Theta}^{-1}\left(\Psi^{-1}(\delta)\right)\right) \\
&= \left(\sqrt{2 + 2\tilde{c}(p)} + 2\right) d\left(\Theta^{-1}(\delta)\right).
\end{aligned}$$

The proof is complete. ■

6 On discrepancy principles

Finally we want to show, that we can apply the discrepancy principle of Morozov as an a posteriori parameter choice strategy at leat in the linear situation. In particular, we choose the parameter $\alpha > 0$ such that for the regularized solution x_α^δ of (5) the equality

$$\|A\, x_\alpha^\delta - y^\delta\| = \delta. \tag{29}$$

holds. Then we can prove the convergence rate result.

Theorem 6.1 *Let $x^\dagger \in \overline{\mathcal{R}(A^*)} \setminus \mathcal{R}(A^*)$ with distance function $d = d(R)$ and assume $\|y^\delta\| > \delta$. Then there exists a regularization parameter $\alpha > 0$ such that (29) holds. Moreover, for $\delta \to 0$ we have the convergence rate (15).*

Proof. For the existence we consider the minimization problem

$$\|x\|^2 \to \min \quad \text{subject to} \quad \|A\,x - y^\delta\| \le \delta. \tag{30}$$

with corresponding Lagrange functional

$$L(x, \lambda; p) := \lambda \left(\|A\,x - y^\delta\|^p - \delta^p \right) + \|x\|^2,$$

after possible reformulation of the contraint. By [15, Theorem 47.E] there exist $x^\delta \in \mathcal{X}$ and a Lagrange multiplier $\lambda \ge 0$ such that x^δ is a the solution of the minimization problem and the pair (x^δ, λ) forms a saddle point of the Lagrange functional. Since $\|A\,x - y^\delta\| \le \delta < \|y^\delta\|$, the constraint has to be active in x^δ which provides $\lambda > 0$. By setting $\alpha := \lambda^{-1}$ and $x_\alpha^\delta := x^\delta$ we proved the existence, by noticing that x_α^δ is also a minimizer of (5) applying the saddle point property. Moreover, we have

$$\|A\,x_\alpha^\delta - y^\delta\|^p + \alpha \|x_\alpha^\delta\|^2 \le \|A\,x^\dagger - y^\delta\|^p + \alpha \|x^\dagger\|^2 \le \delta^p + \alpha \|x^\dagger\|^2.$$

By applying $\|A\,x_\alpha^\delta - y^\delta\| = \delta$ and dividing by α we conclude

$$\|x_\alpha^\delta\|^2 \le \|x^\dagger\|^2.$$

Following the proof of Lemma 2.3 we derive

$$\|x_\alpha^\delta - x^\dagger\| \le 2 \left(\sqrt{\delta\, R} + d(R) \right),$$

independently of the choice of p. By setting $R := \Theta^{-1}(\delta)$ we obtain the convergence rate (15). ■

As we see the choice of the power p does not have any effect on the applicability of the discrepancy principle. Moreover, if we suppose, that the solution of (30) is unique, then the regulaized solution x_α^δ of (1) does not depend on the choice of p. However, the choice of p may influence the efficience of numerical agorithms for solving (30).

The application of the discrepancy principle for nonlinear problems is more crucial. We cannot guarantee in general the existence of a regularization parameter $\alpha > 0$ such that for a solution x_α^δ of (20) the equality

$$\|F(x_\alpha^\delta) - y^\delta\| = \delta \tag{31}$$

holds. On the other hand, if such parameter $\alpha = \alpha(\delta)$ exists for each $\delta > 0$, we can prove convergence rates in an analogous way as in the linear case. The corresponding statement is presented below.

Corollary 6.1 *Let $x^\dagger \in \overline{\mathcal{R}(G^*)} \setminus \mathcal{R}(G^*)$ with distance function $d = d(R)$, which is defined by (13) with G^* instead of A^* and assume the existence of a regularization parameter $\alpha = \alpha(\delta) > 0$ such that (31) holds for each $\delta > 0$ sufficiently small. Then, for $\delta \to 0$, we have the convergence rate (27).*

7 Convergence rates and Bregman distances

Now we consider a more general situation. Let $\mathcal{X}$ and $\mathcal{Y}$ be Banach spaces which are not necessarily reflexive. Instead of (20) we deal with the more general regularization approach

$$J_\alpha(x) := \|F(x) - y^\delta\|^p + \alpha P(x) \to \min \tag{32}$$

with stabilization functional $P : \mathcal{D}(P) \subset \mathcal{X} \longrightarrow [0, \infty)$. According to [8], we need the following assumptions:

(A1) $\mathcal{X}$ and $\mathcal{Y}$ are Banach spaces with associated topologies $\tau_\mathcal{X}$ and $\tau_\mathcal{Y}$, which are weaker than the norm topologies.

(A2) $\|\cdot\|_\mathcal{Y}$ is sequentially lower semi-continuous with respect to $\tau_\mathcal{Y}$.

(A3) The operator $F : \mathcal{D}(F) \subset \mathcal{X} \longrightarrow \mathcal{Y}$ is continuous with respect to the topologies $\tau_\mathcal{X}$ and $\tau_\mathcal{Y}$.

(A4) The functional $P : \mathcal{D}(P) \subset \mathcal{X} \longrightarrow [0, \infty)$ is convex and lower semi-continuous with respect to $\tau_\mathcal{X}$.

(A5) The set $\mathcal{D}(F)$ is closed with respect to $\tau_\mathcal{X}$ and $\mathcal{D}(F) \cap \mathcal{D}(P) \neq \emptyset$.

(A6) For every $\alpha > 0$ and every $M > 0$ the sets $\{x \in \mathcal{D} : J_\alpha(x) \leq M\}$ are $\tau_\mathcal{X}$ sequentially compact, i.e. every sequence $\{x_j\}$ belonging to such set has a subsequence which is convergent in $\mathcal{X}$ with respect to $\tau_\mathcal{X}$.

Under these assumptions the existence of a minimizer $x_\alpha^\delta \in \mathcal{D}(F)$ of (32) can be guaranteed, depending stable on the given data $y^\delta \in \mathcal{Y}$, see e.g. [8, Theorem 3.1+3.2].

Proposition 7.1 *Assume the conditions (A1)-(A6) hold true. Then*

(i) The problem (32) has a solution $x_\alpha^\delta \in \mathcal{D}(F) \cap \mathcal{D}(P)$.

(ii) A solution of (32) depends stable on the given data y^δ*, i.e., for* $y_k \to y^\delta$ *in the norm topology of* $\mathcal{Y}$ *and corresponding solutions* x_k *of (32) there exists at least one subsequence* $\{x_{k_j}\}$ *which converges to* x_α^δ *in the* $\tau_\mathcal{X}$ *topology.*

Under the additional assumption, that equation (19) admitts a solution $x^\dagger \in \mathcal{D}(F)$ for given exact data we can present the following convergence result, see [8, Theorem 3.5].

Proposition 7.2 *Assume the conditions (A1)-(A6) hold true and der exists an element* $x^\dagger \in \mathcal{D}(F)$ *with* $F(x^\dagger) = y$*. Assume a sequence* $\{\delta_k\}$ *with* $\delta_k \to 0$ *for* $k \to \infty$ *and* $\{y_k\} = \{y^{\delta_k}\}$ *with* $\|y - y_k\| \leq \delta_k$*. Moreover,* $\alpha = \alpha(\delta)$ *is chosen such that*

$$\alpha(\delta) \to 0 \quad \textit{and} \quad \frac{\delta^p}{\alpha} \to 0 \quad \textit{for} \quad \delta \to 0.$$

Then a sequence of solutions $\{x_k\}$ *of (32) has at least one subsequence which converges with respect to the* $\tau_\mathcal{X}$ *topology. Each limit of such subseqence is a* P*-minimizing solution. If, in addition, the* P*-minimizing solution* $x^\dagger$ *is uniuqe, then* $x_k \to x^\dagger$ *in the* $\tau_\mathcal{X}$ *topology.*

We recall, that a P-minimizing solution $x^\dagger$ of equation (19) with y^δ replaced by y satisfies

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

We now define Bregman distances which has been established for presenting convergence rates in general Banach spaces.

Definition 7.1 *Let $P : \mathcal{D}(P) \subset \mathcal{X} \longrightarrow [0, \infty)$ denotes a convex functional with subdifferential $\partial P(x)$ for $x \in \mathcal{D}(P)$. The Bregman distance of two elements $\tilde{x}, x \in \mathcal{D}(P)$ is defined as*

$$D_\xi(\tilde{x}, x) := \{P(\tilde{x}) - P(x) - \langle \xi, \tilde{x} - x\rangle_{\mathcal{X}^*,\mathcal{X}}\ :\ \xi \in \partial P(x)\},$$

where $\langle\cdot,\cdot\rangle_{\mathcal{X}^*,\mathcal{X}}$ denotes the duality product with respect to $\mathcal{X}^*$ and $\mathcal{X}$.

(A7) There exists an element $x^\dagger \in \mathcal{D}(F) \cap \mathcal{D}(P)$ with $F(x^\dagger) = y$ and $\partial P(x^\dagger) \neq \emptyset$.

(A8) There exists a linear operator $G \in L(\mathcal{X}, \mathcal{Y})$ satisfying Assumption 5.1.

(A9) There exists $\xi \in \partial P(x^\dagger)$ with $\xi \in \overline{\mathcal{R}(G^*)} \setminus \mathcal{R}(G^*)$ and distance function $d = d(R)$.

(A10) The sets

$$\mathcal{D}_\alpha(K) := \left\{h \in \mathcal{D}(F) \cap \mathcal{D}(P)\ :\ \|F(x^\dagger + h) - y\|^p + \alpha P(x^\dagger + h) \le K\right\}$$

are bounded for all $\alpha > 0$ and all $K > 0$.

(A11) We have $\delta \in [0, \delta_{max}]$ and $\alpha \in (0, \alpha_{max}]$.

We prove some estimates. Here we conly consider with the case of noisy data. Error bounds for regularized solutions x_α for given exact data y can be obtained in a similar way.

Lemma 7.1 *Assume (A1)-(A11) hold true. Then, for $p > 1$,*

$$D_\xi(x_\alpha^\delta, x^\dagger) \le \frac{2\delta^p}{\alpha} + 2\hat{c}(p)\alpha^{\frac{1}{p-1}} R^{\frac{p}{p-1}} + C\, d(R), \tag{33}$$

and for $p = 1$,

$$D_\xi(x_\alpha^\delta, x^\dagger) \le \begin{cases} \delta\left(\dfrac{1}{\alpha} + (1+\gamma)R\right) + C\, d(R), & R \le \dfrac{1}{\alpha(1+\gamma)}, \\ 2\delta R(1+\gamma) + (\alpha R(1+\gamma) - 1)P(x^\dagger) + C\, d(R), & R > \dfrac{1}{\alpha(1+\gamma)}. \end{cases} \tag{34}$$

Thereby the constant $\hat{c}(p)$ depends only on p and γ, whereas C depends only on δ_{max}, α_{max}, p and $x^\dagger$.

Proof. First we note

$$
\begin{aligned}
\|F(x_\alpha^\delta)-y\|^p+\alpha P(x_\alpha^\delta) &\le \big(\|F(x_\alpha^\delta)-y^\delta\|+\delta\big)^p+\alpha P(x_\alpha^\delta)\\
&\le 2^{p-1}\delta^p+2^{p-1}\|F(x_\alpha^\delta)-y^\delta\|^p+\alpha P(x_\alpha^\delta)\\
&\le 2^{p-1}\delta^p+2^{p-1}\left(\|F(x_\alpha^\delta)-y^\delta\|^p+\alpha P(x_\alpha^\delta)\right)\\
&\le 2^{p-1}\delta^p+2^{p-1}\left(\|F(x^\dagger)-y^\delta\|^p+\alpha P(x^\dagger)\right)\\
&\le 2^{p-1}\delta^p+2^{p-1}\left(\delta^p+\alpha P(x^\dagger)\right)\\
&= 2^{p-1}\left(2\delta^p+\alpha P(x^\dagger)\right)=:K(\alpha,\delta).
\end{aligned}
$$

Hence, by (A10), $\|x_\alpha^\delta-x^\dagger\|\le C$ by some positive constant $C>0$. Moreover

$$
\begin{aligned}
P(x_\alpha^\delta)-P(x^\dagger) &= D_\xi(x_\alpha^\delta,x^\dagger)+\langle\xi,x_\alpha^\delta-x^\dagger\rangle_{\mathcal{X}^*,\mathcal{X}}\\
&= D_\xi(x_\alpha^\delta,x^\dagger)+\langle G^*\omega+r,x_\alpha^\delta-x^\dagger\rangle_{\mathcal{X}^*,\mathcal{X}}\\
&= D_\xi(x_\alpha^\delta,x^\dagger)+\langle\omega,G(x_\alpha^\delta-x^\dagger)\rangle_{\mathcal{Y}^*,\mathcal{Y}}+\langle r,x_\alpha^\delta-x^\dagger\rangle_{\mathcal{X}^*,\mathcal{X}}.
\end{aligned}
$$

Then, with $R(x_\alpha^\delta,x^\dagger):=F(x_\alpha^\delta)-F(x^\dagger)-G(x_\alpha^\delta-x^\dagger)=F(x_\alpha^\delta)-y-G(x_\alpha^\delta-x^\dagger)$,

$$
\begin{aligned}
\langle\omega,G(x_\alpha^\delta-x^\dagger)\rangle_{\mathcal{Y}^*,\mathcal{Y}} &= \langle\omega,F(x_\alpha^\delta)-y^\delta+y^\delta-y-R(x_\alpha^\delta,x^\dagger)\rangle_{\mathcal{Y}^*,\mathcal{Y}}\\
&\le \|\omega\|(1+\gamma)\left(\|F(x_\alpha^\delta)-y^\delta\|+\delta\right)
\end{aligned}
$$

Hence, by $\|F(x_\alpha^\delta)-y\|^p+\alpha P(x_\alpha^\delta)-\alpha P(x^\dagger)\le\delta^p$,

$$
\begin{aligned}
\|F(x_\alpha^\delta)-y\|^p+\alpha D_\xi(x_\alpha^\delta,x^\dagger) &\le \delta^p+\alpha\langle\xi,x^\dagger-x_\alpha^\delta\rangle_{\mathcal{X}^*,\mathcal{X}}\\
&\le \delta^p+\alpha\,R(1+\gamma)\left(\delta+\|F(x_\alpha^\delta)-y^\delta\|\right)+\alpha C\left(d(R)+\varepsilon\right)
\end{aligned}
$$

which holds true for all $\varepsilon>0$. Hence, taking the limit $\varepsilon\to 0$, we have

$$
\|F(x_\alpha^\delta)-y\|^p+\alpha D_P(x_\alpha^\delta,x^\dagger)\le\delta^p+\alpha\,R(1+\gamma)\left(\delta+\|F(x_\alpha^\delta)-y^\delta\|\right)+\alpha\,C\,d(R).
$$

For $p>1$ we again apply Young's inequality with $\varepsilon:=\frac{1}{\alpha(1+\gamma)}$ twice for deriving

$$
\alpha(1+\gamma)\,R\,\|F(x_\alpha^\delta)-y^\delta\|\le\|F(x_\alpha^\delta)-y^\delta\|^p+\alpha\hat{c}(p)\alpha^{\frac{1}{p-1}}R^{\frac{p}{p-1}}
$$

and

$$
\alpha(1+\gamma)\,R\,\delta\le\delta^p+\alpha\hat{c}(p)\alpha^{\frac{1}{p-1}}R^{\frac{p}{p-1}}
$$

with

$$
\hat{c}(p):=(1+\gamma)^{\frac{p-2}{p-1}}(p-1)p^{\frac{p}{p-1}}.
$$

This gives

$$
D_\xi(x_\alpha^\delta,x^\dagger)\le\frac{2\delta^p}{\alpha}+2\hat{c}(p)\alpha^{\frac{1}{p-1}}R^{\frac{p}{p-1}}+C\,d(R).
$$

For $p=1$ and $R\le\frac{1}{\alpha(1+\gamma)}$ we conclude

$$
D_\xi(x_\alpha^\delta,x^\dagger)\le\delta\left(\frac{1}{\alpha}+R(1+\gamma)\right)+C\,d(R).
$$

For $R > \frac{1}{\alpha(1+\gamma)}$ we derive

$$\begin{aligned} D_\xi(x_\alpha^\delta, x^\dagger) &\leq \frac{\delta}{\alpha} + R(1+\gamma)\delta + \left(R(1+\gamma) - \frac{1}{\alpha}\right)\left(\delta + \alpha P(x^\dagger)\right) + C\, d(R) \\ &= 2R(1+\gamma)\delta + \left(\alpha\, R(1+\gamma) - 1\right) P(x_\alpha^\delta) + C\, d(R), \end{aligned}$$

by using $\|F(x_\alpha^\delta) - y^\delta\| \leq \delta + \alpha P(x^\dagger)$. ■

Now we can prove convergence rates. We start with $p = 1$.

Theorem 7.1 *Assume (A1)-(A11) hold true and* $\Theta(R) := R^{-1} d(R)$. *Then for* $p = 1$ *and an a priori choice* $\alpha := \Psi^{-1}(\delta)$ *with* $\Psi(\alpha) := \Theta\left(\frac{1}{\alpha(1+\gamma)}\right)$ *we have*

$$D_\xi(x_\alpha^\delta, x^\dagger) \sim \mathcal{O}\left(d\left(\Theta^{-1}(\delta)\right)\right). \tag{35}$$

Proof. We set $R := \frac{1}{\alpha(1+\gamma)}$. Then

$$D_\xi(x_\alpha^\delta, x^\dagger) \leq \frac{2\delta}{\alpha} + C\, d\left(\frac{1}{\alpha(1+\gamma)}\right) = 2(1+\gamma)\delta\, R + C\, d(R).$$

By balancing

$$\delta = \frac{d(R)}{R} = \Theta(R) = \Theta\left(\frac{1}{\alpha(1+\gamma)}\right) = \Psi(\alpha) \;\Leftrightarrow\; R = \Theta^{-1}(\delta) \;\Leftrightarrow\; \alpha = \Psi^{-1}(\alpha),$$

we obtain

$$D_\xi(x_\alpha^\delta, x^\dagger) \leq \left(2(1+\gamma) + C\right) d\left(\Theta^{-1}(\delta)\right),$$

which proves the rate. ■

For $p > 1$ we get the following result.

Theorem 7.2 *Assume (A1)-(A11) hold true and* $\Theta(R) := R^{-1} d(R)$. *Then, for* $p > 1$, $\tilde{\Theta}(R) := R^{-p} d(R)^{p-1}$ *and an a priori choice* $\alpha := \Psi^{-1}(\delta)$ *with* $\Psi(\alpha) := \left(\alpha\, d\left(\tilde{\Theta}^{-1}(\alpha)\right)\right)^{\frac{1}{p}}$ *we have the convergence rate (35).*

Proof. We balance first

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

Hence

$$D_\xi(x_\alpha^\delta, x^\dagger) \leq \frac{2\delta^p}{\alpha} + \left(2\hat{c}(p) + C\right) d\left(\Theta^{-1}(\alpha)\right)$$

and further

$$\delta = \left[\alpha\, d\left(\Theta^{-1}(\alpha)\right)\right]^{\frac{1}{p}} = \Psi(\alpha) \;\Leftrightarrow\; \alpha = \Psi^{-1}(\delta).$$

Moreover,

$$\delta = \left[\alpha\, d\left(\Theta^{-1}(\alpha)\right)\right]^{\frac{1}{p}} = \left[R^{-p} d(R)^p\right]^{\frac{1}{p}} = \frac{d(R)}{R} = \Theta(R) \;\Leftrightarrow\; R = \Theta^{-1}(\delta).$$

This gives the estimate

$$\begin{aligned} D_\xi(x_\alpha^\delta, x^\dagger) &\leq (2+2\hat{c}(p)+C)\, d\left(\tilde{\Theta}^{-1}\left(\Psi^{-1}(\delta)\right)\right) \\ &= (2+2\hat{c}(p)+C)\, d\left(\Theta^{-1}(\delta)\right). \end{aligned}$$

This proves the convergence rate. ■

Finally we want to compare the convergence analysis presented here and in [8]. Following the proofs above, the convergence rates result (35) depends on an assumption of the form

- For each $R > 0$ there exists an element $\omega = \omega(R) \in \mathcal{Y}^*$ with $\|\omega\| = R$ such that

$$\langle \xi, x - x^\dagger \rangle_{\mathcal{X}^*,\mathcal{X}} \leq (1+\gamma)\|\omega\|\, \|F(x) - F(x^\dagger)\| + C\, d(\|\omega\|). \tag{36}$$

  for an element $\xi \in \partial P(x^\dagger)$ and all $x \in \mathcal{B}_\varrho(x^\dagger)$ with $\varrho > 0$ chosen sufficiently large.

On the other hand, in [8, Assumption 4.1] we find the following condition:

- There exist functionals $C_1, C_2 : \mathcal{Y}^* \longrightarrow [0,\infty)$ and $\omega \in \mathcal{Y}^*$ with $C_1(\omega) < 1$ and

$$\langle \xi, x - x^\dagger \rangle_{\mathcal{X}^*,\mathcal{X}} \leq C_1(\omega) D_\xi(x, x^\dagger) + C_2(\omega)\|F(x) - F(x^\dagger)\| \tag{37}$$

  for an element $\xi \in \partial P(x^\dagger)$ and all $x \in \mathcal{B}_\varrho(x^\dagger)$ with $\varrho > 0$ chosen sufficiently large.

Following the proof we see, that the condition (36) contains a larger restriction to the nonlinearity of the operator F but the condition to $\xi \in \partial P(x^\dagger)$ is weaker. In particular we assume, that we cannot find an element $\omega \in \mathcal{Y}^*$ in order to satisfy (37). So, we have proved convergence rates for weaker source conditions by assuming a stronger condition on the nonlinearity of the operator F. The necessity of strengthening the restriction of the nonlinearity was already discussed in the more simple situation of Section 5.

By the argumentations in [8, Remark 4.2] we can also modify conditions (A8) and (A9) as presented below.

(A8') There exists a linear operator $G \in L(\mathcal{X}, \tilde{\mathcal{Y}})$ with $\mathcal{Y} \subseteq \tilde{\mathcal{Y}}$ and

$$\langle v, y \rangle_{\mathcal{Y}^*,\mathcal{Y}} = \langle v, y \rangle_{\tilde{\mathcal{Y}}^*,\tilde{\mathcal{Y}}}, \quad \forall\, v \in \tilde{\mathcal{Y}}^*,\ \forall\, y \in \mathcal{Y},$$

satisfying

$$\|F(x) - F(x^\dagger) - G(x - x^\dagger)\|_{\tilde{\mathcal{Y}}} \leq \gamma \|F(x) - F(x^\dagger)\|_{\mathcal{Y}}.$$

(A9') There exists an element $\xi \in \partial P(x^\dagger) \cap \tilde{\mathcal{Y}}^*$ with $\xi \in \overline{\mathcal{R}(G^*)} \setminus \mathcal{R}(G^*)$ and distance function $d = d(R)$, defined with $\tilde{\mathcal{Y}}$ instead of $\mathcal{Y}$.

The interpretation of this modification is given as follows: Since the norm of $\tilde{\mathcal{Y}}$ is weaker than the norm of $\mathcal{Y}$ the restriction (A8) to the nonlinearity of the operator F is weakened. The price of this weakening we pay by assuming a stronger source condition $\omega \in \tilde{\mathcal{Y}}^* \subseteq \mathcal{Y}^*$ with $\omega = \omega(R)$ and $\|\omega\|_{\tilde{\mathcal{Y}}^*} = R$ for each $R > 0$ on the element $\xi \in \partial P(x^\dagger)$.

References

[1] Baumeister, J.: *Stable solution of Inverse Problems.* Vieweg, Braunschweig (1987).

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

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

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

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

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

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

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

[9] Mathé, P.; Pereverzev, S.V.: *Geometry of linear ill-posed problems in variable Hilbert scales.* Inverse Problems 19, 789-803, 2003.

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

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

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

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

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

[15] Zeidler, E.: *Nonlinear Functional Analysis and its Application III* Springer, New York (1985).