Convergence rates analysis of Tikhonov regularization for nonlinear ill-posed problems with noisy operators

SHUAI LU* and JENS FLEMMING†

March 23, 2012

Key words: nonlinear ill-posed problems, noisy right-hand side, noisy operators, Tikhonov regularization, convergence rates
MSC: 65J20, 65M30

Abstract

We investigate convergence rates of Tikhonov regularization for nonlinear ill-posed problems when both the right-hand side and the operator are corrupted by noise. Two models of operator noise are considered, namely uniform noise bounds and point-wise noise bounds. We derive convergence rates for both noise models in Hilbert and in Banach spaces. These results extend existing results where the forward operator is mostly assumed to be linear.

1 Introduction

We are going to investigate convergence rates of Tikhonov regularization for nonlinear ill-posed equations

$$F(x) = y, \quad x \in X, \tag{1.1}$$

when both the right-hand side $y \in Y$ and the operator $F : D(F) \subseteq X \to Y$ are corrupted by some noise. Here F is some nonlinear operator with

*School of Mathematical Sciences, Fudan University, 200433 Shanghai, People's Republic of China, slu@fudan.edu.cn;
†Chemnitz University of Technology, Department of Mathematics, D-09107 Chemnitz, Germany, jens.flemming@mathematik.tu-chemnitz.de. Corresponding author.

domain $D(F)$ acting between Hilbert or Banach spaces X and Y. The inverse problem consists in recovering $x^\dagger \in D(F)$ from observed noisy data $y^\delta \in Y$ near $y = F(x^\dagger)$. In this article we assume $y^\delta = F(x^\dagger) + \delta_y \xi$ where ξ denotes the normalized noise and δ_y is a small positive value measuring the noise level in the right-hand side, for instance,

$$\|y - y^\delta\| \leq \delta_y. \tag{1.2}$$

Such ill-posed inverse problems often arise in many scientific contexts. For applications we refer to [4, 11, 17, 28] and the references therein.

Due to ill-posedness, the solutions of equation (1.1) do not depend continuously on the right-hand side $y \in Y$ (a precise definition of ill-posedness will be given in Section 2). Thus, the presence of noise forces us to apply regularization methods. In this article we are interested in the following nonlinear Tikhonov regularization

$$\|F(x) - y^\delta\|^p + \alpha\Omega(x) \to \min_{x \in D(F)} \tag{1.3}$$

with the constant $p > 1$, with a regularization parameter $\alpha > 0$, and with a convex stabilizing functional $\Omega : X \to (-\infty, \infty]$. In case of Hilbert spaces one often chooses $p = 2$ and the penalty term

$$\Omega(x) = \|x - x_0\|^2 \tag{1.4}$$

with a known initial guess $x_0 \in X$.

Conditions on F, $D(F)$, Ω ensuring the existence of minimizers of (1.3) are given, for example, in [14]. Also the stability of the minimization problem is shown there.

When using regularization techniques one should answer the question how fast the regularized solutions converge to an exact solution of the underlying equation (1.1) if the noise level, i.e. δ_y in (1.2), decreases. Corresponding estimates are usually meant by 'convergence rates'. Classical results on convergence rates with the exactly known forward operator F in Tikhonov regularization (1.3) have been well established in the last decades for both Hilbert and Banach space settings, see for instance [4, 14] and the references therein.

The treatment of problems (1.1) becomes more complex when noise appears in the forward operator F. For example, instead of the exact forward operator F, only a noisy operator F_δ lying 'near' F is known. To retrospect, noise in operators is considered firstly in linear ill-posed problems as discretization noise and operator noise where convergence analysis is carried out in [19, 23, 29] for standard regularization in Hilbert spaces and Hilbert scales. Some other regularization methods based on the (regularized) total least squares and dual regularized total least squares methods are presented in [7, 20, 27, 30] where multi-parameter regularization approaches naturally appear provided with a negative regularization parameter removing the influence of the operator noise. Interests also arise in the stochastic framework where the linear operator is considered in a singular value decomposition form and the operator noise is introduced by adding random noise on individual singular values, see [3, 10, 21].

Though the convergence rates on linear ill-posed problems with operator noise are quite comprehensive, the literatures on nonlinear ill-posed problems with operator noise are quite limited and mostly restrict the operator noise on the discretization error. For instance, in seminal papers [5, 25], the authors considered an extra noise characterizing the influence of the approximation error in the forward operator. In both papers, only convergence results are provided concerning the particular operator noise and no explicit discussion on the convergence rates. Further discussion with convergence rates in view of the discretized operator noise on finite-dimensional nonlinear ill-posed problems is provided in [16, 24] for Tikhonov regularization and Landweber iteration respectively. Recently [26] defines a point-wise noise bound for the noisy operator and discusses the corresponding convergence rate on iteratively regularized Gauss-Newton methods in the Hilbert space setting.

In our framework, instead of the standard Tikhonov regularization (1.3) we consider a modified minimization problem

$$T_\alpha^\delta(x) := \|F_\delta(x) - y^\delta\|^p + \alpha\Omega(x) \to \min_{x \in D(F_\delta)} \tag{1.5}$$

with a known noisy operator $F_\delta : D(F_\delta) \subseteq X \to Y$. Throughout this article the corresponding minimizers will be denoted by

$$x_\alpha^\delta \in \operatorname{argmin}_{x \in D(F_\delta)} T_\alpha^\delta(x).$$

In Banach spaces the discrepancy between regularized solutions x_α^δ and an exact solution $x^\dagger$ can be expressed by the Bregman distance

$$B_{\xi^\dagger}^{\Omega}(x_\alpha^\delta, x^\dagger) := \Omega(x_\alpha^\delta) + \Omega(x^\dagger) - \langle \xi^\dagger, x_\alpha^\delta - x^\dagger \rangle$$

where $\xi^\dagger \in \partial\Omega(x^\dagger)$ is a subgradient of Ω at $x^\dagger$ (see, e.g., [2]). Here, one has to be aware of the fact that in certain 'nonsmooth' Banach spaces, for instance in the sequence space $\ell^1(\mathbb{N})$, the Bregman distance with respect to $\Omega = \|\bullet\|$ contains only few information on the distance between two elements, depending on the chosen subgradient. If in a Hilbert space setting the penalty term $\Omega(x)$ is given by (1.4) then the Bregman distance reduces to the standard Hilbert space norm

$$B^{\Omega}_{\xi^\dagger}(x^\delta_\alpha, x^\dagger) = \|x^\delta_\alpha - x^\dagger\|^2.$$

To obtain convergence rates for Tikhonov regularization with noise free operator, abstract assumptions on the smoothness of the unknown exact solution $x^\dagger$ with respect to the operator F are necessarily formulated (e.g. source conditions), see [4, Section 4.2]. Such assumptions mostly contain a (Fréchet) derivative $F'[x^\dagger]$ of F at $x^\dagger$. In case of noisy operators F_δ usage of these smoothness assumptions with respect to the exact operator implies that the connection between F and F_δ is of high importance, since only F_δ has influence on the regularized solutions x^δ_α. Two possibilities for connecting F with F_δ are proposed in the next section, namely uniform noise bounds and point-wise noise bounds.

The structure of the remaining part of this article is as follows: First we describe and investigate ill-posedness with respect to noisy data and with respect to noisy operators in Section 2. Two proposed noise models for operator noise are introduced in the same section. In Sections 3 and 4, we derive convergence rates for uniform noise bounds and point-wise noise bounds respectively. Finally some conclusions and remarks in Section 5 end the article.

2 Ill-posedness and operator noise

In this section we firstly clarify the term 'ill-posed' and then introduce two noise models for a better understanding of the noisy operators.

2.1 Ill-posedness from a general perspective

In this subsection we show that equations (1.1) which are ill-posed with respect to data noise are also ill-posed with respect to operator noise.

When solving equations (1.1) in practice, one typically has some a priori information at hand which allow to restrict attention to a set $M \subseteq D(F)$ of 'interesting points'. Such a priori information for example could involve properties of the iterates generated by an algorithm. In the context of regularization techniques the set M should contain all possible regularized solutions. For Tikhonov regularization M can be chosen as a sublevel set of the Tikhonov functional (see [14]) or as a sublevel set of the stabilizing functional Ω.

By $S(y) \subseteq \{x \in M : F(x) = y\}$ we denote the set of desired solutions. Typically one is not interested in an arbitrary solution but in one with special properties, e.g. Ω-minimizing solutions. In particular, we assume $S(y) \neq \emptyset$.

An important question in the theory of ill-posed problems is how to express convergence of a sequence of approximate solutions $x_n \in M$ to an exact solution $x^\dagger \in S(y)$ or to the whole set $S(y)$. Note, that we leave out questions on existence and uniqueness if we use the term 'ill-posed'. We are solely interested in the continuity or discontinuity of the 'inverse' of F. In the literature one typically finds results of the type that there are convergent subsequences and that every convergent subsequence converges to some solution $x^\dagger \in S(y)$. Another concept is to consider the convergence $\operatorname{dist}(x_n, S(y)) \to 0$, where $\operatorname{dist}(x_n, S(y)) := \inf_{x \in S(y)} \|x_n - x\|$. We note that the assertions on ill-posedness in this section are true for any kind of convergence in X, since the proofs do not rely on the definition of $\operatorname{dist}(x_n, S(y))$. The following example displays the difference between both convergence concepts and shows that the corresponding notions of ill-posedness differ.

Example 2.1. Let $X := l^2(\mathbb{N})$, $Y := \mathbb{R}$, $F(x) := \|x\|$, and $y := 1$. For simplicity we choose $M := X$ and $S(y) := \{x \in X : \|x\| = 1\}$. Consider the sequence $(x_n)_{n \in \mathbb{N}}$ defined by $x_n := (1 + \frac{1}{n})e_n$, where e_n has a one at position n and zeros else. Then $F(x_n) = 1 + \frac{1}{n} \to 1 = y$, but $(x_n)_{n \in \mathbb{N}}$ neither converges nor it has convergent subsequences. In this sense the equation $F(x) = y$ is ill-posed. Note that in the weak topology the sequence $(x_n)_{n \in \mathbb{N}}$ converges to zero, which is not a solution of $F(x) = y$. If we use $\operatorname{dist}(x_n, S(y))$ for expressing convergence we see $\operatorname{dist}(x_n, S(y)) \to 0$ since $e_n \in S(y)$ and $\|x_n - e_n\| \to 0$. In this weaker but nevertheless meaningful sense the equation $F(x) = y$ is well-posed.

From this simple example one also sees that a sequence of approximate solutions can become arbitrarily close to the set of solutions without converging to one particular solution.

In contrast to data noise, noisy operators are comparably rarely discussed in the literatures. Based on the set M of 'interesting points' we define the set of all admissible noisy operators by

$$\mathcal{N}_M := \left\{ G : M \to Y : \sup_{x \in M} \|G(x)\| < \infty \right\}.$$

Endowed with the norm

$$\|G\|_M := \sup_{x \in M} \|G(x)\|, \quad G \in \mathcal{N}_M,$$

the set $\mathcal{N}_M$ becomes a normed vector space. In the following we identify F with its restriction to M and we assume $F \in \mathcal{N}_M$.

It might happen that noisy data y^δ does not belong to the range of F or that y does not belong to the range of a noisy operator. Therefore one has to seek for approximate solutions. The approximation should become better if the noise level is smaller. For example in case of noisy data and exact operator a sequence of approximate solutions $(x_n)_{n\in\mathbb{N}}$ corresponding to a sequence of noisy right-hand sides $(y_n)_{n\in\mathbb{N}}$ with $y_n \to y$ should satisfy $F(x_n) - y_n \to 0$.

We now propose a definition of ill-posedness and then show that this definition covers ill-posedness with respect to data noise as well as ill-posedness with respect to operator noise.

Definition 2.2. Equation (1.1) is *locally ill-posed in* $y \in \mathcal{R}(F)$ if there is a sequence $(x_n)_{n\in\mathbb{N}}$ in M such that $F(x_n) \to y$ but $\operatorname{dist}(x_n, S(y)) \nrightarrow 0$.

This definition of local ill-posedness is different from the one given in [15].

Proposition 2.3. *Equation* (1.1) *is* locally ill-posed in $y \in \mathcal{R}(F)$ *if and only if one of the following three equivalent assertions is true:*

(i) There are sequences $(y_n)_{n\in\mathbb{N}}$ *in* Y *and* $(x_n)_{n\in\mathbb{N}}$ *in* M *such that* $y_n \to y$ *and* $F(x_n) - y_n \to 0$, *but* $\operatorname{dist}(x_n, S(y)) \nrightarrow 0$ *(local ill-posedness w.r.t. data noise).*

(ii) There are sequences $(F_n)_{n\in\mathbb{N}}$ *in* $\mathcal{N}_M$ *and* $(x_n)_{n\in\mathbb{N}}$ *in* M *such that* $\|F_n - F\|_M \to 0$ *and* $F_n(x_n) - y \to 0$, *but* $\operatorname{dist}(x_n, S(y)) \nrightarrow 0$ *(local ill-posedness w.r.t. operator noise).*

(iii) There are sequences $(y_n)_{n\in\mathbb{N}}$ *in* Y, $(F_n)_{n\in\mathbb{N}}$ *in* $\mathcal{N}_M$, *and* $(x_n)_{n\in\mathbb{N}}$ *in* M *such that* $y_n \to y$, $\|F_n - F\|_M \to 0$, *and* $F_n(x_n) - y_n \to 0$, *but* $\operatorname{dist}(x_n, S(y)) \nrightarrow 0$ *(local ill-posedness w.r.t. combined data and operator noise).*

Proof. Ill-posedness (Definition 2.2) obviously implies (ii) (set $F_n := F$). From (ii) we obtain (iii) by defining $y_n := y$. That item (i) follows after (iii) can be seen from the estimate

$$\begin{aligned}\|F(x_n) - y_n\| &\leq \|F(x_n) - F_n(x_n)\| + \|F_n(x_n) - y_n\| \\ &\leq \|F - F_n\|_M + \|F_n(x_n) - y_n\| \to 0.\end{aligned}$$

Finally, (i) implies ill-posedness since

$$\|F(x_n) - y\| \leq \|F(x_n) - y_n\| + \|y_n - y\| \to 0.$$

□

The assertion and the proof of the proposition remain valid if the set $\mathcal{N}_M$ of admissible noisy operators is restricted to a smaller class of mappings. For example we could assume that the original operator F and all possible noisy operators are (weakly) continuous. This is reasonable because most regularization techniques require some kind of continuity.

If F is bounded and linear one might assume that a noisy version of this operator is also bounded and linear. In this case the set M of 'interesting points' is typically bounded. Therefore, without loss of generality we assume $M = \{x \in X : \|x\| \leq 1\}$. Then the norm $\|\bullet\|_M$ coincides with the operator norm and the space $\mathcal{N}_M$ of admissible noisy operators is simply the normed vector space of bounded linear operators mapping X into Y. As an example for linear noisy operators one could consider linear convolution operators F. Noisy operators then appear if the kernel is not modeled correctly or if the kernel is constructed from measurements.

To highlight the crucial influence of operator noise we show that even in case of exactly solvable noisy equations the ill-posedness effect remains.

Proposition 2.4. *If equation* (1.1) *is locally ill-posed in* $y \in \mathcal{R}(F)$*, then there are sequences* $(F_n)_{n\in\mathbb{N}}$ *in* $\mathcal{N}_M$ *and* $(x_n)_{n\in\mathbb{N}}$ *in* M *such that* $\|F_n - F\|_M \to 0$ *and* $F_n(x_n) = y$*, but* $\operatorname{dist}(x_n, S(y)) \not\to 0$.

This assertion remains true if F *is bounded and linear,* $M = \{x \in X : \|x\| \leq 1\}$*, and the set* $\mathcal{N}_M$ *of admissible noisy operators contains only bounded linear operators.*

Proof. By the definition of ill-posedness there is a sequence $(x_n)_{n\in\mathbb{N}}$ such that $F(x_n) \to y$, but $\operatorname{dist}(x_n, S(y)) \not\to 0$.

If the set $\mathcal{N}_M$ of admissible noisy operators is not restricted to linear operators we define $F_n \in \mathcal{N}_M$ by

$$F_n(x) := F(x) + y - F(x_n), \quad x \in M. \tag{2.1}$$

Then $F_n(x_n) = y$ and $\|F_n - F\|_M = \|y - F(x_n)\| \to 0$.

If $A := F$ is linear and $\mathcal{N}_M$ is restricted to bounded linear operators, definition (2.1) cannot be used. Instead we proceed as follows. If $x_n = 0$ for all sufficiently large n then choosing $A_n := A$ and observing $y = 0$ proves the assertion. Otherwise we may assume $x_n \neq 0$ for all $n \in \mathbb{N}$ (take a suitable subsequence). For each x_n there is a bounded linear functional ξ_n on X such that $\langle \xi_n, x_n \rangle = \|x_n\|$ and $\|\xi_n\| = 1$. Defining bounded linear operators $A_n : X \to Y$ by

$$A_n x := Ax - \frac{\langle \xi_n, x \rangle}{\|x_n\|}(Ax_n - y), \quad x \in X,$$

we immediately see $A_n x_n = y$. Since

$$\begin{aligned}\|A_n - A\|_M = \sup_{x \in M} \|A_n x - Ax\| &= \left(\sup_{x \in M} \frac{|\langle \xi_n, x \rangle|}{\|x_n\|} \right) \|Ax_n - y\| \\ &= \|Ax_n - y\| \to 0,\end{aligned}$$

the assertion is thus proven. □

2.2 Two models for operator noise

As stated in the introductory section, one usually assumes a uniform bound of the data noise in the sense that $\|y - y^\delta\| \leq \delta_y$. In view of the definition on ill-posedness in the previous subsection, we firstly propose a uniform operator noise in an analogue way. Such uniform noise could for example appear if the kernel of a convolution opertator is not known exactly. Recall that $M \subseteq D(F)$ is the set of 'interesting points' and that $S(y) \subseteq \{x \in M : F(x) = y\}$ in the noise-free operator setting. For the noisy operator framework we additionally assume $M \subseteq D(F) \cap D(F_\delta)$. The uniform operator noise is thus imposed on the whole set M such that there holds

$$\sup_{x \in M} \|F(x) - F_\delta(x)\| \leq \delta_F^M \tag{2.2}$$

with a known constant δ_F^M referring to the operator noise level. Note, that the δ in the symbol F_δ does not denote the noise level. The noisy operator under consideration is denoted by F_δ and the corresponding noise level is δ_F^M. In principle, assumption (2.2) can be realized as a generalization of the approximation operator in the existing literatures, for instance (2.2) in [24]. Similar to [16, 24], the uniform noise bound assumption allows us to obtain convergence rates in terms of δ_y and δ_F^M in both Hilbert and Banach space settings which are provided in Section 3.

Next to uniform noise bounds other models for operator noise could be suitable in real applications. For example one could ask for point-wise bounds connection F and F_δ. Such an idea is recently proposed in [26] for Fréchet differentiable operators F and F_δ where the authors assume

$$\|F(x^\dagger) - F_\delta(x^\dagger)\| \leq \delta_F, \quad \|F'[x^\dagger] - F'_\delta[x^\dagger]\| \leq \delta_{F'}.$$

We mention that a particular nonlinear ill-posed problem which satisfies the corresponding operator noise assumptions can be found in the same literature. In our current work, we also slightly change the assumptions in a similar manner such that

$$\begin{cases} \|F(x^\dagger) - F_\delta(x^\dagger)\| \leq \delta_F, \\ \|F'[x^\dagger]^* F'[x^\dagger] - F'_\delta[x^\dagger]^* F'_\delta[x^\dagger]\| \leq \delta^2_{F'}. \end{cases} \tag{2.3}$$

Note that in the forthcoming analysis, we need the following result of operator monotonicity whose definition can be found in [1, V.1, Thm. X.1.1] and [22] for finite-dimension and infinite-dimension cases respectively.

Theorem 2.5. *[1, 22] Let f be operator monotone on $(0, \infty)$ with $f(0) = 0$. For any pair A, B of non-negative self-adjoint operators in the Hilbert space we have*

$$\|f(A) - f(B)\| \leq f(\|A - B\|).$$

A special consequence of operator monotonicity is

$$\|A^\mu - B^\mu\| \leq \|A - B\|^\mu, \quad \mu \in (0, 1].$$

One can observe that the second inequality in (2.3) immediately implies

$$\|F'[x^\dagger] - F'_\delta[x^\dagger]\| \leq \delta_{F'}$$

(choose $\mu = \frac{1}{2}$ in Theorem 2.5). Notice that the assumption (2.3) only holds true at the exact solution $x^\dagger$ which is referred as point-wise noise bounds. Convergence rates in terms of δ_y, δ_F, and $\delta_{F'}$ are shown in Section 4.

3 Convergence rates for uniform noise bounds

In case of uniform noise bounds (2.2) for the operator noise we have the following two estimates. The first estimate in Banach spaces bases on a variational inequality, which includes nonlinearity assumptions on F and smoothness assumptions with respect to F on the exact solution $x^\dagger$ (cf. [14]). The second estimate in Hilbert spaces bases on a standard source condition and explicit nonlinearity assumptions (cf. [4]). In principle, the nonlinearity and smoothness assumptions carrying out the convergence rates for uniform noise bounds are the same as those with exact operators.

Theorem 3.1. *Assume* (1.2) *and* (2.2). *Further, let* $\beta > 0$ *and let* $\varphi : [0,\infty) \to [0,\infty)$ *be monotonically increasing and concave. If*

$$\beta B_{\xi^\dagger}^{\Omega}(x, x^\dagger) \le \Omega(x) - \Omega(x^\dagger) + \varphi(\|F(x) - F(x^\dagger)\|^p) \quad \text{for all } x \in M$$

then for all $\alpha > 0$, $\delta_y \ge 0$, $\delta_F^M \ge 0$ *such that* $x_\alpha^\delta \in M$ *the estimate*

$$B_{\xi^\dagger}^{\Omega}(x_\alpha^\delta, x^\dagger) \le \frac{4}{\beta}\frac{(\delta_y + \delta_F^M)^p}{\alpha} + \frac{1}{\beta}(-\varphi)^*\left(\frac{-1}{4^{p-1}\alpha}\right)$$

holds true. Here $(-\varphi)^*$ *denotes the Fenchel conjugate function of* $-\varphi$ *given by* $(-\varphi)^*(s) = \sup_{t\ge 0}(st + \varphi(t))$.

Proof. By the minimizing property of x_α^δ we see

$$\begin{aligned}
\Omega(x_\alpha^\delta) - \Omega(x^\dagger) &= \frac{1}{\alpha}\big(\|F_\delta(x_\alpha^\delta) - y^\delta\|^p + \alpha\Omega(x_\alpha^\delta)\big) - \Omega(x^\dagger) - \frac{1}{\alpha}\|F_\delta(x_\alpha^\delta) - y^\delta\|^p \\
&\le \frac{1}{\alpha}\|F_\delta(x^\dagger) - y^\delta\|^p - \frac{1}{\alpha}\|F_\delta(x_\alpha^\delta) - y^\delta\|^p.
\end{aligned}$$

The triangle inequality yields

$$\begin{aligned}
\|F(x_\alpha^\delta) - F(x^\dagger)\|^p &\le 2^{p-1}\|F_\delta(x_\alpha^\delta) - F_\delta(x^\dagger)\|^p \\
&\qquad + 2^{p-1}\|F(x_\alpha^\delta) - F_\delta(x_\alpha^\delta) + F_\delta(x^\dagger) - F(x^\dagger)\|^p \\
&\le 4^{p-1}\|F_\delta(x_\alpha^\delta) - y^\delta\|^p + 4^{p-1}\|y^\delta - F_\delta(x^\dagger)\|^p \\
&\qquad + 4^{p-1}\|F(x_\alpha^\delta) - F_\delta(x_\alpha^\delta)\|^p + 4^{p-1}\|F_\delta(x^\dagger) - F(x^\dagger)\|^p \\
&\le 4^{p-1}\|F_\delta(x_\alpha^\delta) - y^\delta\|^p + 4^{p-1}\|y^\delta - F_\delta(x^\dagger)\|^p + 2^{2p-1}\big(\delta_F^M\big)^p.
\end{aligned}$$

Applying both estimates to the variational inequality we then obtain

$$\begin{aligned}
\beta B_{\xi^\dagger}^{\Omega}(x, x^\dagger) &\le \frac{1}{\alpha}\|F_\delta(x^\dagger) - y^\delta\|^p - \frac{1}{\alpha}\|F_\delta(x_\alpha^\delta) - y^\delta\|^p \\
&\qquad + \varphi\big(4^{p-1}\|F_\delta(x_\alpha^\delta) - y^\delta\|^p + 4^{p-1}\|y^\delta - F_\delta(x^\dagger)\|^p + 2^{2p-1}\big(\delta_F^M\big)^p\big) \\
&= \frac{2}{\alpha}\|F_\delta(x^\dagger) - y^\delta\|^p + \frac{2}{\alpha}\big(\delta_F^M\big)^p \\
&\qquad - \frac{1}{4^{p-1}\alpha}\big(4^{p-1}\|F_\delta(x_\alpha^\delta) - y^\delta\|^p + 4^{p-1}\|F_\delta(x^\dagger) - y^\delta\|^p + 2^{2p-1}\big(\delta_F^M\big)^p\big) \\
&\qquad + \varphi\big(4^{p-1}\|F_\delta(x_\alpha^\delta) - y^\delta\|^p + 4^{p-1}\|F_\delta(x^\dagger) - y^\delta\|^p + 2^{2p-1}\big(\delta_F^M\big)^p\big) \\
&\le \frac{2}{\alpha}(\delta_y + \delta_F^M)^p + \frac{2}{\alpha}\big(\delta_F^M\big)^p + \sup_{t\ge 0}\left(\frac{-1}{4^{p-1}\alpha}t + \varphi(t)\right) \\
&\le 4\frac{(\delta_y + \delta_F^M)^p}{\alpha} + (-\varphi)^*\left(\frac{-1}{4^{p-1}\alpha}\right).
\end{aligned}$$

□

Choosing the regularization parameter similar to [6, Section 4.2] the theorem provides the convergence rate

$$B_{\xi^\dagger}^{\Omega}(x_\alpha^\delta, x^\dagger) = \mathcal{O}\big(\varphi((\delta_y + \delta_F^M)^p)\big) \quad \text{as } \delta_y + \delta_F^M \to 0.$$

The same rate can be obtained by applying the discrepancy principle with noise level $\delta_y + \delta_F^M$ for choosing the regularization parameter, but with a proof slightly different from the one given above. Next, we establish the convergence rate in Hilbert spaces under uniform noise bounds.

Theorem 3.2. *Let X and Y be Hilbert spaces. Assume that the nonlinear operators F, F_δ are weakly closed and that assumptions* (1.2) *and* (2.2) *are satisfied. Moreover, assume that $D(F) \cap D(F_\delta)$ is convex, F is Fréchet differentiable and there exists a Lipschitz constant $L > 0$ such that*

$$\|F'[x_1] - F'[x_2]\| \leq L\|x_1 - x_2\|$$

for all $x_1,\ x_2 \in D(F) \cap D(F_\delta)$. If there exists some $v \in Y$ such that the source condition

$$\begin{aligned} x^\dagger - x_0 &= F'[x^\dagger]^* v, \\ L\|v\| &\leq 1 \end{aligned}$$

is satisfied, then all global minimizers x_α^δ fulfill the error bounds

$$\begin{aligned} \|x_\alpha^\delta - x^\dagger\| &\leq \sqrt{\frac{2}{1 - L\|v\|}} \left(\frac{\delta_y + \delta_F^M}{\sqrt{\alpha}} + \sqrt{\alpha}\|v\| \right) \\ \|F_\delta(x_\alpha^\delta) - y^\delta\| &\leq \sqrt{2}\left(\delta_y + \delta_F^M\right) + (\sqrt{2} + 1)\alpha\|v\|. \end{aligned}$$

If we choose $\alpha = \delta_y + \delta_F^M$, the estimations imply

$$\begin{aligned} \|x_\alpha^\delta - x^\dagger\| &\leq \sqrt{\frac{2}{1 - L\|v\|}} \left(1 + \|v\|\right) \sqrt{\delta_y + \delta_F^M}, \\ \|F_\delta(x_\alpha^\delta) - y^\delta\| &\leq \left(\sqrt{2} + (\sqrt{2} + 1)\|v\|\right)\left(\delta_y + \delta_F^M\right). \end{aligned}$$

The same convergence rates hold true if one applies the discrepancy principle.

Proof. The first several steps are the same as the classic proof in [5]. Note the Lipschitz continuity of the Fréchet derivative F' implies that

$$F(x_\alpha^\delta) = F(x^\dagger) + F'[x^\dagger](x_\alpha^\delta - x^\dagger) + r_\alpha^\delta$$

holds with

$$\|r_\alpha^\delta\| \leq \frac{L}{2}\|x_\alpha^\delta - x^\dagger\|^2.$$

The minimizer x_α^δ satisfies

$$\|F_\delta(x_\alpha^\delta) - y^\delta\|^2 + \alpha\|x_\alpha^\delta - x_0\|^2 \leq \|F_\delta(x^\dagger) - y^\delta\|^2 + \alpha\|x^\dagger - x_0\|^2.$$

By adding $\alpha\|x_\alpha^\delta - x^\dagger\|^2 - \alpha\|x_\alpha^\delta - x_0\|^2$ in both sides, we obtain

$$\|F_\delta(x_\alpha^\delta) - y^\delta\|^2 + \alpha\|x_\alpha^\delta - x^\dagger\|^2 \leq \|F_\delta(x^\dagger) - y^\delta\|^2 + 2\alpha\langle x^\dagger - x_0, x^\dagger - x_\alpha^\delta\rangle.$$

The first and second terms in the right-hand side can be estimated as follows by using the triangle inequality and Lipschitz continuity property,

$$\|F_\delta(x^\dagger) - y^\delta\|^2 \leq 2\left(\|F_\delta(x^\dagger) - F(x^\dagger)\|^2 + \|F(x^\dagger) - y^\delta\|^2\right) \leq 2\left({\delta_F^M}^2 + \delta_y^2\right),$$

$$\begin{aligned}
2\alpha\langle x^\dagger - x_0, x^\dagger - x_\alpha^\delta\rangle &= 2\alpha\langle v, F'[x^\dagger](x^\dagger - x_\alpha^\delta)\rangle \\
&\leq \alpha L\|v\|\|x_\alpha^\delta - x^\dagger\|^2 + 2\alpha\|v\|\|F(x_\alpha^\delta) - F_\delta(x_\alpha^\delta)\| \\
&\qquad + 2\alpha\|v\|\|F_\delta(x_\alpha^\delta) - y^\delta\| + 2\alpha\|v\|\|y^\delta - F(x^\dagger)\| \\
&\leq \alpha L\|v\|\|x_\alpha^\delta - x^\dagger\|^2 + 2\alpha\|v\|\delta_F^M \\
&\qquad + 2\alpha\|v\|\|F_\delta(x_\alpha^\delta) - y^\delta\| + 2\alpha\|v\|\delta_y.
\end{aligned}$$

One then obtains

$$\begin{aligned}
\left(\|F_\delta(x_\alpha^\delta) - y^\delta\| - \alpha\|v\|\right)^2 + \alpha(1 - L\|v\|)\|x_\alpha^\delta - x^\dagger\|^2 &\leq 2\left(\delta_F^M + \frac{\alpha\|v\|}{2}\right)^2 \\
&\quad + 2\left(\delta_y + \frac{\alpha\|v\|}{2}\right)^2.
\end{aligned}$$

The desired results then follow after some simple calculations. With respect to the discrepancy principle

$$c_1(\delta_y + \delta_F^M) \leq \|F_\delta(x_\alpha^\delta) - y^\delta\| \leq c_2(\delta_y + \delta_F^M)$$

with $1 \leq c_1 \leq c_2$ we use the fact that

$$\|x_\alpha^\delta - x^\dagger\|^2 \leq 2\langle x^\dagger - x_0, x^\dagger - x_\alpha^\delta\rangle.$$

□

4 Convergence rates for point-wise noise bounds

4.1 Low order rates in Banach spaces

Throughout this subsection we assume that F and F_δ are Fréchet differentiable at $x^\dagger$ and we denote the corresponding Fréchet derivatives by $F'[x^\dagger]$ and $F_\delta'[x^\dagger]$.

If point-wise noise bounds (2.3) for the operator noise are valid then we have to control the nonlinearity of F_δ. This is contrary to the previous section where the nonlinearity of F is controlled by a variational inequality and no explicit nonlinearity assumptions on F_δ are required. Controlling the nonlinearity of F_δ but assuming smoothness of $x^\dagger$ with respect to F or $F'[x^\dagger]$ implies that in case of a point-wise noise bound variational inequalities are not an appropriate tool for obtaining convergence rates. The problem is that variational inequalities combine nonlinearity and solution smoothness into one condition which either has to hold for F or for F_δ. But assuming a variational inequality for F does not influence the nonlinearity of F_δ and assuming a variational inequality for all possible noisy operators F_δ is a too strong assumption. In the latter case on the one hand one would implicitly assume that $x^\dagger$ is smooth with respect to many different operators and on the other hand the variational inequality would depend on noise considerations violating the idea of an universal sufficient condition for convergence rates.

Instead we use the concept of approximate source conditions introduced in [12] for Hilbert space problems and extended to Banach spaces in [8, 9]. Before we go into the details we provide a first convergence rate result based on the typical source condition $\xi^\dagger = F'[x^\dagger]^*\eta^\dagger$ in Banach spaces, where $\xi^\dagger \in \partial\Omega(x^\dagger)$ and $\eta^\dagger \in Y^*$. Additionally we assume the property

$$\|x - x^\dagger\|^q \leq c_q B^\Omega_{\xi^\dagger}(x, x^\dagger) \quad \text{for all } x \in M \tag{4.1}$$

with some $q > 1$ and $c_q > 0$ and a sufficiently large set M. The proofs also work without this q-coercivity assumption but then we have to assume that all regularized solutions x_α^δ lie in a bounded set M and the obtained convergence rates will be slower. In case of Hilbert spaces with $\Omega(x) = \|x - x_0\|^2$ we have $q = 2$. More details on the Hilbert space setting are provided in the next subsection.

Theorem 4.1. *Assume* (1.2), (2.3), (4.1), *and*

$$\|F_\delta(x) - F_\delta(x^\dagger) - F_\delta'[x^\dagger](x - x^\dagger)\| \leq c_{\mathrm{NL}} B^\Omega_{\xi^\dagger}(x, x^\dagger) \quad \textit{for all } x \in M \tag{4.2}$$

with $c_{\mathrm{NL}} > 0$ *and a sufficiently large set* M*. Further let there be some* $\eta^\dagger \in Y^*$ *with* $c_{\mathrm{NL}}\|\eta^\dagger\| < 1$ *such that* $\xi^\dagger = F'[x^\dagger]^*\eta^\dagger$*. Then*

$$B^{\Omega}_{\xi^\dagger}(x^\delta_\alpha, x^\dagger) \leq \frac{4}{1 - c_{\mathrm{NL}}\|\eta^\dagger\|}\frac{(\delta_y + \delta_F)^p}{\alpha} + \frac{4(p-1)}{1 - c_{\mathrm{NL}}\|\eta^\dagger\|}\left(\frac{\|\eta^\dagger\|}{p}\right)^{\frac{p}{p-1}} \alpha^{\frac{1}{p-1}}$$
$$+ (q-1)c_q^{\frac{1}{q-1}}\left(\frac{2\|\eta^\dagger\|}{q(1 - c_{\eta^\dagger}\|\eta^\dagger\|)}\right)^{\frac{q}{q-1}} \delta_{F'}^{\frac{q}{q-1}}.$$

Proof. First we observe

$$\begin{aligned}
-\langle \xi^\dagger, x^\delta_\alpha - x^\dagger\rangle &\leq \|\eta^\dagger\|\|F'[x^\dagger](x^\delta_\alpha - x^\dagger)\| \\
&\leq \|\eta^\dagger\|\|(F'[x^\dagger] - F'_\delta[x^\dagger])(x^\delta_\alpha - x^\dagger)\| + \|\eta^\dagger\|\|F_\delta(x^\delta_\alpha) - F_\delta(x^\dagger)\| \\
&\quad + \|\eta^\dagger\|\|F_\delta(x^\delta_\alpha) - F_\delta(x^\dagger) - F'_\delta[x^\dagger](x^\delta_\alpha - x^\dagger)\| \\
&\leq \delta_{F'}\|\eta^\dagger\|\|x^\delta_\alpha - x^\dagger\| + \|\eta^\dagger\|\|F_\delta(x^\delta_\alpha) - F_\delta(x^\dagger)\| \\
&\quad + c_{\mathrm{NL}}\|\eta^\dagger\|B^{\Omega}_{\xi^\dagger}(x^\delta_\alpha, x^\dagger).
\end{aligned}$$

By implementing (4.1) and Young's inequality $ab \leq \frac{1}{q}a^q + \frac{q-1}{q}b^{\frac{q}{q-1}}$ with

$$a := \left(\frac{q}{2}(1 - c_{\mathrm{NL}}\|\eta^\dagger\|)B^{\Omega}_{\xi^\dagger}(x^\delta_\alpha, x^\dagger)\right)^{\frac{1}{q}} \quad \text{and} \quad b := \frac{c_q^{\frac{1}{q}}\|\eta^\dagger\|\delta_{F'}}{\left(\frac{q}{2}(1 - c_{\eta^\dagger}\|\eta^\dagger\|)\right)^{\frac{1}{q}}}$$

we obtain

$$\begin{aligned}
\delta_{F'}\|\eta^\dagger\|\|x^\delta_\alpha - x^\dagger\| &\leq \delta_{F'}\|\eta^\dagger\|c_q^{\frac{1}{q}}B^{\Omega}_{\xi^\dagger}(x^\delta_\alpha, x^\dagger)^{\frac{1}{q}} \\
&\leq \frac{1 - c_{\mathrm{NL}}\|\eta^\dagger\|}{2}B^{\Omega}_{\xi^\dagger}(x^\delta_\alpha, x^\dagger) + \frac{(q-1)c_q^{\frac{1}{q-1}}\|\eta^\dagger\|^{\frac{q}{q-1}}\delta_{F'}^{\frac{q}{q-1}}}{q^{\frac{q}{q-1}}\left(\frac{1}{2}(1 - c_{\eta^\dagger}\|\eta^\dagger\|)\right)^{\frac{1}{q-1}}}
\end{aligned}$$

and therefore

$$-\langle \xi^\dagger, x^\delta_\alpha - x^\dagger\rangle \leq \frac{1 + c_{\mathrm{NL}}\|\eta^\dagger\|}{2}B^{\Omega}_{\xi^\dagger}(x^\delta_\alpha, x^\dagger) + c\delta_{F'}^{\frac{q}{q-1}} + \|\eta^\dagger\|\|F_\delta(x^\delta_\alpha) - F_\delta(x^\dagger)\|$$

with

$$c := \frac{(q-1)c_q^{\frac{1}{q-1}}\|\eta^\dagger\|^{\frac{q}{q-1}}}{q^{\frac{q}{q-1}}\left(\frac{1}{2}(1 - c_{\eta^\dagger}\|\eta^\dagger\|)\right)^{\frac{1}{q-1}}} > 0.$$

Consequently we derive

$$\begin{aligned}
\frac{1 - c_{\mathrm{NL}}\|\eta^\dagger\|}{2}B^{\Omega}_{\xi^\dagger}(x^\delta_\alpha, x^\dagger) &= \Omega(x^\delta_\alpha) - \Omega(x^\dagger) - \langle \xi^\dagger, x^\delta_\alpha - x^\dagger\rangle - \frac{1 + c_{\mathrm{NL}}\|\eta^\dagger\|}{2}B^{\Omega}_{\xi^\dagger}(x^\delta_\alpha, x^\dagger) \\
&\leq \Omega(x^\delta_\alpha) - \Omega(x^\dagger) + \|\eta^\dagger\|\|F_\delta(x^\delta_\alpha) - F_\delta(x^\dagger)\| + c\delta_{F'}^{\frac{q}{q-1}}.
\end{aligned}$$

Noticing the minimization property of the Tikhonov functional such that

$$\begin{aligned}\Omega(x_\alpha^\delta) - \Omega(x^\dagger) &= \frac{1}{\alpha}\big(\|F_\delta(x_\alpha^\delta) - y^\delta\|^p + \alpha\Omega(x_\alpha^\delta)\big) - \Omega(x^\dagger) - \frac{1}{\alpha}\|F_\delta(x_\alpha^\delta) - y^\delta\|^p \\ &\le \frac{1}{\alpha}\|F_\delta(x^\dagger) - y^\delta\|^p - \frac{1}{\alpha}\|F_\delta(x_\alpha^\delta) - y^\delta\|^p,\end{aligned}$$

we obtain

$$\begin{aligned}\frac{1 - c_{\mathrm{NL}}\|\eta^\dagger\|}{2} B_{\xi^\dagger}^\Omega(x_\alpha^\delta, x^\dagger) \le{}& \frac{1}{\alpha}\|F_\delta(x^\dagger) - y^\delta\|^p - \frac{1}{\alpha}\|F_\delta(x_\alpha^\delta) - y^\delta\|^p \\ &+ \|\eta^\dagger\|\|F_\delta(x_\alpha^\delta) - y^\delta\| + \|\eta^\dagger\|\|F_\delta(x^\dagger) - y^\delta\| + c\delta_{F'}^{\frac{q}{q-1}}\end{aligned}$$

and applying Young's inequality $ab \le \frac{1}{p}a^p + \frac{p-1}{p}b^{\frac{p}{p-1}}$ twice with

$$a := \left(\frac{p}{\alpha}\right)^{\frac{1}{p}} \|F_\delta(\ldots) - y^\delta\| \quad \text{and} \quad b := \left(\frac{\alpha}{p}\right)^{\frac{1}{p}} \|\eta^\dagger\|$$

shows

$$\frac{1 - c_{\mathrm{NL}}\|\eta^\dagger\|}{2} B_{\xi^\dagger}^\Omega(x_\alpha^\delta, x^\dagger) \le 2\frac{(\delta_y + \delta_F)^p}{\alpha} + 2(p-1)\left(\frac{\|\eta^\dagger\|}{p}\right)^{\frac{p}{p-1}} \alpha^{\frac{1}{p-1}} + c\delta_{F'}^{\frac{q}{q-1}}.$$

□

If we choose $\alpha \sim (\delta_y + \delta_F)^{p-1}$ the theorem implies the convergence rate

$$B_{\xi^\dagger}^\Omega(x_\alpha^\delta, x^\dagger) = \mathcal{O}\left(\delta_y + \delta_F + \delta_{F'}^{\frac{q}{q-1}}\right) \quad \text{as } \delta_y + \delta_F + \delta_{F'}^{\frac{q}{q-1}} \to 0.$$

The same rate can be obtained via the discrepancy principle but with a slightly different proof.

The source condition above provides only one fixed rate in the sense that either the obtained rate can be shown or not. The second convergence rate result is based on approximate source conditions and thus provides a wide range of possible convergence rates, that is, the rate is adapted to the (abstract) smoothness of the exact solution. This concept relies on distance functions

$$d(r) := \inf\{\|F'[x^\dagger]^*\eta - \xi^\dagger\| : \eta \in Y^*, \|\eta\| \le r\}, \quad r \ge 0,$$

measuring the smoothness of $x^\dagger$ w.r.t. $F'[x^\dagger]^*$. For noisy operators we define

$$d_\delta(r) := \inf\{\|F_\delta'[x^\dagger]^*\eta - \xi^\dagger\| : \eta \in Y^*, \|\eta\| \le r\}, \quad r \ge 0. \tag{4.3}$$

Then we have a trivial estimate

$$d_\delta(r) \le \inf_{\|\eta\|\le r} \left(\|F'_\delta[x^\dagger]^* - F'[x^\dagger]^*\|\|\eta\| + \|F'[x^\dagger]^*\eta - \xi^\dagger\|\right) \le \delta_{F'} r + d(r) \quad (4.4)$$

for all $r \ge 0$.

Note that the nonlinearity condition (4.2) seems to be unsuitable for a solution smoothness under $F'[x^\dagger]$ (up to now there are no convergence rate results for this case). Therefore we assume

$$\|F'_\delta[x^\dagger](x - x^\dagger)\| \le c_s \|F_\delta(x) - F_\delta(x^\dagger)\|^s \quad \text{for all } x \in M \qquad (4.5)$$

with some $s \in (0, p)$.

Theorem 4.2. *Assume that* (1.2), (2.3), (4.1), *and* (4.5) *hold. Further we assume* $d(r) \le c_\nu r^{-\nu}$ *for all* $r > 0$ *with some* $\nu > 0$. *Then the following error estimate holds true*

$$B^\Omega_{\xi^\dagger}(x^\delta_\alpha, x^\dagger) \le 4\frac{(\delta_y + \delta_F)^p}{\alpha} + 4\bar{c}\alpha^\gamma + 4\bar{c}\delta_{F'}^\kappa$$

for sufficiently small $\delta_y, \delta_F, \delta_{F'}, \alpha$, *where*

$$\gamma := \min\left\{\frac{\frac{s\nu q}{q-1}}{p + \frac{(p-s)\nu q}{q-1}}, \frac{s\nu}{(p-s)(\nu+1)}\right\}$$

and

$$\kappa := \min\left\{\frac{\frac{\nu q}{q-1}}{\frac{p(q-1)}{(p-s)q} + \nu}, \frac{q\nu}{(q-1)(\nu+1)}\right\}.$$

The constant $\bar{c} > 0$ *is independent of* $\delta_y, \delta_F, \delta_{F'}, \alpha$.

Proof. By (4.5) for all $\eta \in Y^*$ with $\|\eta\| \le r$ we have

$$\begin{aligned}-\langle \xi^\dagger, x^\delta_\alpha - x\rangle &= \langle F'_\delta[x^\dagger]\eta - \xi^\dagger, x^\delta_\alpha - x^\dagger\rangle + \langle \eta, F'_\delta[x^\dagger](x^\delta_\alpha - x^\dagger)\rangle \\ &\le \|F'_\delta[x^\dagger]\eta - \xi^\dagger\|\|x^\delta_\alpha - x^\dagger\| + c_s r\|F_\delta(x^\delta_\alpha) - F_\delta(x^\dagger)\|^s\end{aligned}$$

and therefore (take the infimum over all η)

$$\begin{aligned}\frac{1}{2}B^\Omega_{\xi^\dagger}(x^\delta_\alpha, x^\dagger) &= \Omega(x^\delta_\alpha) - \Omega(x^\dagger) - \frac{1}{2}B^\Omega_{\xi^\dagger}(x^\delta_\alpha, x^\dagger) - \langle \xi^\dagger, x^\delta_\alpha - x^\dagger\rangle \\ &\le \Omega(x^\delta_\alpha) - \Omega(x^\dagger) - \frac{1}{2}B^\Omega_{\xi^\dagger}(x^\delta_\alpha, x^\dagger) + d_\delta(r)\|x^\delta_\alpha - x^\dagger\| + c_s r\|F_\delta(x^\delta_\alpha) - F_\delta(x^\dagger)\|^s\end{aligned}$$

for all $r \geq 0$. By using (4.1) we derive

$$d_\delta(r)\|x_\alpha^\delta - x^\dagger\| \leq c_q^{\frac{1}{q}} d_\delta(r) B_{\xi^\dagger}^\Omega(x_\alpha^\delta, x^\dagger)^{\frac{1}{q}}$$

and Young's inequality $ab \leq \frac{1}{q}a^q + \frac{q-1}{q}b^{\frac{q}{q-1}}$ with

$$a := \left(\frac{q}{2}\right)^{\frac{1}{q}} B_{\xi^\dagger}^\Omega(x_\alpha^\delta, x^\dagger)^{\frac{1}{q}} \quad \text{and} \quad b := \left(\frac{2c_q}{q}\right)^{\frac{1}{q}} d_\delta(r)$$

yields

$$d_\delta(r)\|x_\alpha^\delta - x^\dagger\| \leq \frac{1}{2} B_{\xi^\dagger}^\Omega(x_\alpha^\delta, x^\dagger) + \frac{(q-1)(2c_q)^{\frac{1}{q-1}}}{q^{\frac{q}{q-1}}} d_\delta(r)^{\frac{q}{q-1}}.$$

Thus, the following estimation holds

$$\frac{1}{2} B_{\xi^\dagger}^\Omega(x_\alpha^\delta, x^\dagger) \leq \Omega(x_\alpha^\delta) - \Omega(x^\dagger) + c_s r\|F_\delta(x_\alpha^\delta) - F_\delta(x^\dagger)\|^s + c d_\delta(r)^{\frac{q}{q-1}}$$

for all $r \geq 0$ with

$$c := \frac{(q-1)(2c_q)^{\frac{1}{q-1}}}{q^{\frac{q}{q-1}}} > 0.$$

By the minimizing property of x_α^δ we have

$$\begin{aligned}
\Omega(x_\alpha^\delta) - \Omega(x^\dagger) &= \frac{1}{\alpha}\big(\|F_\delta(x_\alpha^\delta) - y^\delta\|^p + \alpha\Omega(x_\alpha^\delta)\big) - \Omega(x^\dagger) - \frac{1}{\alpha}\|F_\delta(x_\alpha^\delta) - y^\delta\|^p \\
&\leq \frac{1}{\alpha}\|F_\delta(x^\dagger) - y^\delta\|^p - \frac{1}{\alpha}\|F_\delta(x_\alpha^\delta) - y^\delta\|^p
\end{aligned}$$

and therefore

$$\begin{aligned}
\frac{1}{2} B_{\xi^\dagger}^\Omega(x_\alpha^\delta, x^\dagger) \leq{}& \frac{2}{\alpha}(\delta_y + \delta_F)^p - \frac{1}{2^{p-1}\alpha}\left(2^{p-1}\|F_\delta(x_\alpha^\delta) - y^\delta\|^p + 2^{p-1}\|y^\delta - F_\delta(x^\dagger)\|^p\right) \\
&+ c_s r\left(2^{p-1}\|F_\delta(x_\alpha^\delta) - y^\delta\|^p + 2^{p-1}\|y^\delta - F_\delta(x^\dagger)\|^p\right)^{\frac{s}{p}} + c d_\delta(r)^{\frac{q}{q-1}}.
\end{aligned}$$

Young's inequality $\frac{s}{p}a^{\frac{p}{s}} + \frac{p-s}{p}b^{\frac{p}{p-s}}$ with

$$a := \left(\frac{p}{2^{p-1}s\alpha}\left(2^{p-1}\|F_\delta(x_\alpha^\delta) - y^\delta\|^p + 2^{p-1}\|y^\delta - F_\delta(x^\dagger)\|^p\right)\right)^{\frac{s}{p}} \quad \text{and} \quad b := c_s r\left(\frac{2^{p-1}s\alpha}{p}\right)^{\frac{s}{p}}$$

now yields

$$\frac{1}{2}B^{\Omega}_{\xi^\dagger}(x^\delta_\alpha, x^\dagger) \le \frac{2}{\alpha}(\delta_y+\delta_F)^p + \frac{(p-s)c_s^{\frac{p}{p-s}}(2^{p-1}s)^{\frac{s}{p-s}}}{p^{\frac{p}{p-s}}}\alpha^{\frac{s}{p-s}} r^{\frac{p}{p-s}} + c d_\delta(r)^{\frac{q}{q-1}}.$$

We use (4.4) to obtain

$$c d_\delta(r)^{\frac{q}{q-1}} \le c\left(\delta_{F'} r + d(r)\right)^{\frac{q}{q-1}} \le 2^{\frac{1}{q-1}} c \delta_{F'}^{\frac{q}{q-1}} r^{\frac{q}{q-1}} + 2^{\frac{1}{q-1}} c c_\nu^{\frac{q}{q-1}} r^{\frac{-\nu q}{q-1}}$$

for all $r \ge 0$. Therefore

$$\frac{1}{2}B^{\Omega}_{\xi^\dagger}(x^\delta_\alpha, x^\dagger) \le \frac{2}{\alpha}(\delta_y+\delta_F)^p + \tilde{c}\left(\alpha^{\frac{s}{p-s}} r^{\frac{p}{p-s}} + \delta_{F'}^{\frac{q}{q-1}} r^{\frac{q}{q-1}} + r^{\frac{-\nu q}{q-1}}\right)$$

for all $r \ge 0$ with some constant $\tilde{c} > 0$ independent of $\delta_y, \delta_F, \delta_{F'}, \alpha, r$. If $r \ge 1$ then

$$\frac{1}{2}B^{\Omega}_{\xi^\dagger}(x^\delta_\alpha, x^\dagger) \le \frac{2}{\alpha}(\delta_y+\delta_F)^p + \tilde{c}\left(\left(\alpha^{\frac{s}{p-s}} + \delta_{F'}^{\frac{q}{q-1}}\right) r^{\max\left\{\frac{q}{q-1},\frac{p}{p-s}\right\}} + r^{\frac{-\nu q}{q-1}}\right)$$

and choosing r such that

$$\left(\alpha^{\frac{s}{p-s}} + \delta_{F'}^{\frac{q}{q-1}}\right) r^{\max\left\{\frac{q}{q-1},\frac{p}{p-s}\right\}} = r^{\frac{-\nu q}{q-1}},$$

that is

$$r = \left(\alpha^{\frac{s}{p-s}} + \delta_{F'}^{\frac{q}{q-1}}\right)^{-\min\left\{\frac{1}{\frac{p}{p-s}+\frac{\nu q}{q-1}},\frac{q-1}{q(\nu+1)}\right\}},$$

we obtain

$$\begin{aligned}\frac{1}{2}B^{\Omega}_{\xi^\dagger}(x^\delta_\alpha, x^\dagger) &\le \frac{2}{\alpha}(\delta_y+\delta_F)^p + 2\tilde{c}\left(\alpha^{\frac{s}{p-s}} + \delta_{F'}^{\frac{q}{q-1}}\right)^{\min\left\{\frac{\frac{\nu q}{q-1}}{\frac{p}{p-s}+\frac{\nu q}{q-1}},\frac{\nu}{\nu+1}\right\}} \\ &\le \frac{2}{\alpha}(\delta_y+\delta_F)^p + 2\bar{c}\alpha^{\min\left\{\frac{\frac{s\nu q}{q-1}}{p+\frac{(p-s)\nu q}{q-1}},\frac{s\nu}{(p-s)(\nu+1)}\right\}} + 2\bar{c}\delta_{F'}^{\min\left\{\frac{\frac{\nu q}{q-1}}{\frac{p(q-1)}{(p-s)q}+\nu},\frac{q\nu}{(q-1)(\nu+1)}\right\}}\end{aligned}$$

with some constant $\bar{c} > 0$. □

With the parameter choice

$$\alpha \sim (\delta_y+\delta_F)^{\max\left\{\frac{p+\frac{(p-s)\nu q}{q-1}}{1+\frac{\nu q}{q-1}},\frac{(p-s)(\nu+1)}{\nu+1-\frac{s}{p}}\right\}}$$

the theorem provides the convergence rate

$$B^{\Omega}_{\xi^\dagger}(x^\delta_\alpha, x^\dagger) = \mathcal{O}\left((\delta_y + \delta_F)^{\min\left\{\frac{\frac{s\nu q}{q-1}}{1+\frac{\nu q}{q-1}}, \frac{s\nu}{\nu+1-\frac{s}{p}}\right\}} + \delta_{F'}^{\min\left\{\frac{\frac{\nu q}{q-1}}{\frac{p(q-1)}{(p-s)q}+\nu}, \frac{q\nu}{(q-1)(\nu+1)}\right\}} \right)$$

as $\delta_y + \delta_F + \delta_{F'} \to 0$.

In a Hilbert space setting with $\Omega(x)$ given by (1.4) and linear operator $A := F$ we have $p = 2$, $q = 2$, $s = 1$. Suppose $x^\dagger - x_0$ is in the range of $(A^*A)^\mu$ with $\mu \in (0, \frac{1}{2})$ we then derive $\nu = \frac{2\mu}{1-2\mu}$ (see [13, Theorem 1]). In this case the obtained convergence rate reduces to

$$\|x^\delta_\alpha - x^\dagger\|^2 = \mathcal{O}\left((\delta_y + \delta_F)^{\frac{4\mu}{2\mu+1}} + \delta_{F'}^{4\mu}\right). \tag{4.6}$$

If the benchmark source condition is satisfied, that is, if $d(r) = 0$ for sufficiently large r, then we have

$$B^{\Omega}_{\xi^\dagger}(x^\delta_\alpha, x^\dagger) \leq 4\frac{(\delta_y + \delta_F)^p}{\alpha} + c_1 \alpha^{\frac{s}{p-s}} + c_2 \delta_{F'}^{\frac{q}{q-1}}.$$

With $\alpha \sim (\delta_y + \delta_F)^{p-s}$ the corresponding rate is

$$B^{\Omega}_{\xi^\dagger}(x^\delta_\alpha, x^\dagger) = \mathcal{O}\left((\delta_y + \delta_F)^s + \delta_{F'}^{\frac{q}{q-1}}\right).$$

4.2 High order rates in Hilbert spaces

Since point-wise noise bounds are only recently proposed in [26], it is worthwhile for us to take a close look of such assumptions within the framework of Tikhonov regularization in Hilbert spaces.

The first statement concerns a standard source condition which is the same as in [26]. We also refer to Theorem 4.1 where a similar situation is considered in the Banach space setting.

Theorem 4.3. *Assume that the nonlinear operators F, F_δ are weakly closed, Fréchet differentiable and that assumptions* (1.2), (2.3) *are satisfied at $x^\dagger$. Moreover, we assume that $D(F) \cap D(F_\delta)$ is convex and that there exists a Lipschitz constant $L > 0$ such that*

$$\|F'_\delta[x_1] - F'_\delta[x_2]\| \leq L\|x_1 - x_2\|$$

for all x_1, $x_2 \in D(F) \cap D(F_\delta)$. If a source condition

$$\begin{aligned} x^\dagger - x_0 &= F'[x^\dagger]^* v, \\ (L + \|v\|)\|v\| &\le 1 \end{aligned}$$

for some $v \in Y$ is satisfied, then all global minimizers x_α^δ fulfill the error bounds

$$\begin{aligned} \|x_\alpha^\delta - x^\dagger\| &\le \delta_{F'} + \frac{\delta_F + \delta_y + \alpha\|v\|}{\sqrt{\alpha(1-(L+\|v\|)\|v\|)}} \\ \|F_\delta(x_\alpha^\delta) - y^\delta\| &\le \delta_F + \delta_y + \sqrt{\alpha}\delta_{F'} + 2\alpha\|v\|. \end{aligned}$$

If we choose $\alpha = \delta_F + \delta_y$, the estimations imply

$$\begin{aligned} \|x_\alpha^\delta - x^\dagger\| &\le \delta_{F'} + \frac{1+\|v\|}{\sqrt{1-(L+\|v\|)\|v\|}}\sqrt{\delta_F + \delta_y}, \\ \|F_\delta(x_\alpha^\delta) - y^\delta\| &\le (1+2\|v\|)(\delta_F + \delta_y) + \sqrt{\delta_F + \delta_y}\delta_{F'}. \end{aligned}$$

We omit the proof here but refer to those of Theorems 3.2 and 4.1. One can observe that the same convergence rate holds for the discrepancy principle with noise level $\delta_y + \delta_F$ for choosing the regularization parameter. The saturation of the convergence rate appears when the *a posteriori* parameter choice rule is implemented. We will not touch this topic in detail but refer to the monograph [4]. Finally, we investigate a higher monomial source condition such that

$$x^\dagger - x_0 = (F'[x^\dagger]^* F'[x^\dagger])^\mu w$$

for $\mu \in (1/2, 1]$. The following theorem is the main statement of convergence rates on point-wise noise bounds with a monomial source condition large than $1/2$.

Theorem 4.4. *Assume that the nonlinear operators F, F_δ are weakly closed, Fréchet differentiable and that assumptions (1.2), (2.3) are satisfied at $x^\dagger$. Moreover, we assume that $D(F) \cap D(F_\delta)$ is convex and that there exists a Lipschitz constant $L > 0$ such that*

$$\|F'_\delta[x_1] - F'_\delta[x_2]\| \le L\|x_1 - x_2\|$$

for all x_1, $x_2 \in D(F) \cap D(F_\delta)$. If the source conditions

$$\begin{aligned} x^\dagger - x_0 &= F'[x^\dagger]^* v, \\ x^\dagger - x_0 &= (F'[x^\dagger]^* F'[x^\dagger])^\mu w \end{aligned}$$

are satisfied for $\mu \in (1/2, 1]$ and variables $v \in Y$ and $w \in X$ such that $(L + \|v\|)\|v\| \leq 1$, then with an a priori parameter choice $\alpha \sim (\delta_y + \delta_F + \delta_{F'}^{2\mu+1})^{\frac{2}{2\mu+1}}$ all global minimizers x_α^δ fulfill the error bound

$$\|x_\alpha^\delta - x^\dagger\| = \mathcal{O}\left(\max\left\{(\delta_y + \delta_F + \delta_{F'}^{2\mu+1})^{\frac{2\mu}{2\mu+1}}, \delta_{F'}\right\}\right).$$

The convergence rate of Theorem 4.4 can also be presented in the form

$$\|x_\alpha^\delta - x^\dagger\| = \mathcal{O}\left(\max\left\{(\delta_y + \delta_F)^{\frac{2\mu}{2\mu+1}} + \delta_{F'}^{2\mu}, \delta_{F'}\right\}\right)$$

where the consistence with the results of linear ill-posed problems in Banach spaces can be observed (see i.e. (4.6)) but with a saturation rate on $\delta_{F'}$. The proof of the theorem is based on the classic ones for the convergence rate analysis of nonlinear ill-posed problems in [4, Thm.10.7] and [18] where the forward operator is exactly known. For sake of convenience, we denote $B = F'[x^\dagger]$ and $B_\delta = F'_\delta[x^\dagger]$. Similar to these references, we introduce an auxiliary element

$$z_\alpha = x^\dagger - \alpha({B_\delta}^* B_\delta + \alpha I)^{-1}(x^\dagger - x_0).$$

The following lemma is important.

Lemma 4.5. *Assume that the assumptions in Theorem 4.4 are valid. Then the following estimation for the auxiliary element z_α holds true*

$$\begin{aligned}\|z_\alpha - x^\dagger\| &\leq C_\mu\|w\|\alpha^\mu + \|w\|\delta_{F'}^{2\mu}\\ &\leq C_\mu\|w\|\alpha^\mu + \|w\|(\delta_y + \delta_F + \delta_{F'}^{2\mu+1})^{\frac{2\mu}{2\mu+1}}\end{aligned}$$

with a constant C_μ independent of α, δ_y, δ_F and $\delta_{F'}$.

Proof. By using the operator monotonicity, the estimation follows after the fact that

$$z_\alpha - x^\dagger = -\alpha({B_\delta}^* B_\delta + \alpha I)^{-1}({B_\delta}^* B_\delta)^\mu w - \alpha({B_\delta}^* B_\delta + \alpha I)^{-1}((B^* B)^\mu - ({B_\delta}^* B_\delta)^\mu)w$$

and

$$\|z_\alpha - x^\dagger\| \leq C_\mu\|w\|\alpha^\mu + \|w\|\delta_{F'}^{2\mu}.$$

Here and in what follows C_μ represents the constant which is induced by the Tikhonov regularization for linear ill-posed problems with a monomial smoothness μ. For detailed descriptions, we refer to [23, Definition 1]. □

In addition, we occasionally use the property of the partial isometry U (see also in [18]) that allows the relation

$$(B^*B)^{1/2}Uv = (B^*B)^{\mu}w \Rightarrow Uv = (B^*B)^{\mu-1/2}w$$

since $\mu \in (1/2, 1]$. Now we are ready to prove Theorem 4.4.

Proof of Theorem 4.4. From the minimization property of the Tikhonov functional, it follows that

$$\begin{aligned} \alpha\|x_\alpha^\delta - x^\dagger\|^2 \le \|F_\delta(z_\alpha) - y^\delta\|^2 - \|F_\delta(x_\alpha^\delta) - y^\delta\|^2 + \alpha\|z_\alpha - x^\dagger\|^2 \\ + 2\alpha\langle z_\alpha - x^\dagger, x^\dagger - x_0\rangle - 2\alpha\langle x_\alpha^\delta - x^\dagger, x^\dagger - x_0\rangle. \end{aligned} \tag{4.7}$$

Note the Lipschitz continuity of the Fréchet derivative $F'_\delta[x^\dagger]$ implies that

$$\begin{aligned} F_\delta(z_\alpha) &= F_\delta(x^\dagger) + B_\delta(z_\alpha - x^\dagger) + s_\alpha, \qquad s.t. \quad \|s_\alpha\| \le \frac{L}{2}\|z_\alpha - x^\dagger\|^2 \\ F_\delta(x_\alpha^\delta) &= F_\delta(x^\dagger) + B_\delta(x_\alpha^\delta - x^\dagger) + r_\alpha, \qquad s.t. \quad \|r_\alpha\| \le \frac{L}{2}\|x_\alpha^\delta - x^\dagger\|^2. \end{aligned}$$

Moreover, we can derive that

$$\begin{aligned} F_\delta(z_\alpha) - y^\delta &= F_\delta(z_\alpha) - F_\delta(x^\dagger) + F_\delta(x^\dagger) - y^\delta \\ &= B_\delta(z_\alpha - x^\dagger) + s_\alpha + F_\delta(x^\dagger) - y^\delta. \end{aligned}$$

The basic source condition $x^\dagger - x_0 = F'[x^\dagger]^*v = B^*v$ yields

$$\begin{aligned} 2\alpha\langle z_\alpha - x^\dagger, x^\dagger - x_0\rangle &= 2\alpha\langle B_\delta(z_\alpha - x^\dagger), v\rangle + 2\alpha\langle (B - B_\delta)(z_\alpha - x^\dagger), v\rangle; \\ -2\alpha\langle x_\alpha^\delta - x^\dagger, x^\dagger - x_0\rangle &= -2\alpha\langle B_\delta(x_\alpha^\delta - x^\dagger), v\rangle - 2\alpha\langle (B - B_\delta)(x_\alpha^\delta - x^\dagger), v\rangle \\ &= -2\alpha\langle F_\delta(x_\alpha^\delta) - y^\delta, v\rangle + 2\alpha\langle r_\alpha, v\rangle - 2\alpha\langle y^\delta - F_\delta(x^\dagger), v\rangle \\ &\quad - 2\alpha\langle (B - B_\delta)(x_\alpha^\delta - x^\dagger), v\rangle. \end{aligned}$$

Substitute these equalities into (4.7), we obtain

$$\begin{aligned} \|x_\alpha^\delta - x^\dagger\|^2 \le{}& \frac{1}{\alpha}\|B_\delta(z_\alpha - x^\dagger) + \alpha v\|^2 + \frac{1}{\alpha}\|s_\alpha + F_\delta(x^\dagger) - y^\delta\|^2 \\ &+ \frac{2}{\alpha}\Big(\langle B_\delta(z_\alpha - x^\dagger), s_\alpha + F_\delta(x^\dagger) - y^\delta\rangle + \alpha\langle F_\delta(x^\dagger) - y^\delta, v\rangle\Big) \\ &+ 2\langle (B - B_\delta)(z_\alpha - x^\dagger), v\rangle + 2\langle r_\alpha, v\rangle - 2\langle (B - B_\delta)(x_\alpha^\delta - x^\dagger), v\rangle \\ &+ \|z_\alpha - x^\dagger\|^2 - \frac{1}{\alpha}\|F_\delta(x_\alpha^\delta) - y^\delta + \alpha v\|^2. \end{aligned} \tag{4.8}$$

The third line in the previous inequality (4.8) can be estimated as follows

$$\begin{aligned}&2\langle (B-B_\delta)(z_\alpha - x^\dagger), v\rangle + 2\langle r_\alpha, v\rangle - 2\langle (B-B_\delta)(x_\alpha^\delta - x^\dagger), v\rangle \\ &\le 2\|v\|\delta_{F'}\|z_\alpha - x^\dagger\| + L\|v\|\|x_\alpha^\delta - x^\dagger\|^2 + 2\|v\|\delta_{F'}\|x_\alpha^\delta - x^\dagger\| \\ &\le (L+\|v\|)\|v\|\|x_\alpha^\delta - x^\dagger\|^2 + \|z_\alpha - x^\dagger\|^2 + (1+\|v\|^2)\delta_{F'}^2.\end{aligned}$$

By omitting the negative term in (4.8) and defining

$$\mathfrak{M} := \langle B_\delta(z_\alpha - x^\dagger), s_\alpha + F_\delta(x^\dagger) - y^\delta\rangle + \alpha\langle v, F_\delta(x^\dagger) - y^\delta\rangle,$$

we obtain the following estimation from (4.8) such that

$$\begin{aligned}\sqrt{1-(L+\|v\|)\|v\|}\|x_\alpha^\delta - x^\dagger\| \le \frac{1}{\sqrt{\alpha}}\|B_\delta(z_\alpha - x^\dagger) + \alpha v\| + \frac{1}{\sqrt{\alpha}}\|s_\alpha + F_\delta(x^\dagger) - y^\delta\| \\ + \sqrt{\frac{2}{\alpha}}\sqrt{\mathfrak{M}} + \sqrt{2}\|z_\alpha - x^\dagger\| + \sqrt{1+\|v\|^2}\delta_{F'}.\end{aligned}$$

In view of Lemma 4.5, we only need to estimate the first three terms in the right-hand side separately.

Notice the first term satisfies

$$\begin{aligned}B_\delta(z_\alpha - x^\dagger) + \alpha v = (-\alpha B_\delta({B_\delta}^* B_\delta + \alpha I)^{-1}{B_\delta}^* + \alpha I)v \\ - \alpha B_\delta({B_\delta}^* B_\delta + \alpha I)^{-1}(B^* - {B_\delta}^*)v.\end{aligned}$$

We then employ the same argument including the partial isometry U in [18] to derive

$$\begin{aligned}\|B_\delta(z_\alpha - x^\dagger) + \alpha v\| &\le \alpha^2\|({B_\delta}^* B_\delta + \alpha I)^{-1}Uv\| + \alpha\|B_\delta({B_\delta}^* B_\delta + \alpha I)^{-1}(B^* - {B_\delta}^*)v\| \\ &\le \alpha^2\|({B_\delta}^* B_\delta + \alpha I)^{-1}({B_\delta}^* B_\delta)^{\mu-1/2}w\| \\ &\quad + \alpha^2\|({B_\delta}^* B_\delta + \alpha I)^{-1}((B^*B)^{\mu-1/2} - ({B_\delta}^* B_\delta)^{\mu-1/2})w\| \\ &\quad + \alpha\|B_\delta({B_\delta}^* B_\delta + \alpha I)^{-1}(B^* - {B_\delta}^*)v\| \\ &\le C_{\mu-1/2}\|w\|\alpha^{\mu+1/2} + \|w\|\alpha\delta_{F'}^{2\mu-1} + C_{1/2}\|v\|\sqrt{\alpha}\delta_{F'}.\end{aligned}$$

It is quite obvious that with the *a priori* choice $\alpha \sim (\delta_y + \delta_F + \delta_{F'}^{2\mu+1})^{\frac{2}{2\mu+1}}$ we can estimate

$$\begin{aligned}\frac{1}{\sqrt{\alpha}}\|B_\delta(z_\alpha - x^\dagger) + \alpha v\| &\le C_{\mu-1/2}\|w\|\alpha^\mu + \|w\|\sqrt{\alpha}\delta_{F'}^{2\mu-1} + C_{1/2}\|v\|\delta_{F'} \\ &\le C_{\mu-1/2}\|w\|\alpha^\mu + \|w\|\sqrt{\alpha}(\delta_y + \delta_F + \delta_{F'}^{2\mu+1})^{\frac{2\mu-1}{2\mu+1}} + C_{1/2}\|v\|\delta_{F'} \\ &= \mathcal{O}\left(\max\left\{(\delta_y + \delta_F + \delta_{F'}^{2\mu+1})^{\frac{2\mu}{2\mu+1}}, \delta_{F'}\right\}\right).\end{aligned}$$

We next estimate the second term $\frac{1}{\sqrt{\alpha}}\|s_\alpha + F_\delta(x^\dagger) - y^\delta\|$ which is quite straightforward such that

$$\frac{1}{\sqrt{\alpha}}\|s_\alpha + F_\delta(x^\dagger) - y^\delta\| \leq \frac{L}{2\sqrt{\alpha}}\|z_\alpha - x^\dagger\|^2 + \frac{1}{\sqrt{\alpha}}\|F_\delta(x^\dagger) - y^\delta\|.$$

By using the proof of Lemma 4.5, $\mu > 1/2$ and without loss of generality we assume $\alpha < 1$ and $\delta_{F'} < 1$, the following estimation holds true

$$\begin{aligned}\frac{L}{2\sqrt{\alpha}}\|z_\alpha - x^\dagger\|^2 + \frac{1}{\sqrt{\alpha}}\|F_\delta(x^\dagger) - y^\delta\| &\leq C_\mu^2 L\|w\|^2\alpha^{2\mu-1/2} + \frac{\delta_{F'}^{4\mu}\|w\|^2}{\sqrt{\alpha}} + \frac{\delta_y + \delta_F}{\sqrt{\alpha}}\\ &\leq C_\mu^2 L\|w\|^2\alpha^{\mu} + \frac{\delta_{F'}^{2\mu+1}\|w\|^2 + \delta_y + \delta_F}{\sqrt{\alpha}}.\end{aligned}$$

That is

$$\frac{1}{\sqrt{\alpha}}\|s_\alpha + F_\delta(x^\dagger) - y^\delta\| = \mathcal{O}\left((\delta_y + \delta_F + \delta_{F'}^{2\mu+1})^{\frac{2\mu}{2\mu+1}}\right)$$

with the *a priori* choice $\alpha \sim (\delta_y + \delta_F + \delta_{F'}^{2\mu+1})^{\frac{2}{2\mu+1}}$.

Finally, we estimate the third term $\sqrt{\frac{2}{\alpha}}\sqrt{\mathfrak{M}}$ with $\mathfrak{M} := \langle B_\delta(z_\alpha - x^\dagger), s_\alpha + F_\delta(x^\dagger) - y^\delta\rangle + \alpha\langle v, F_\delta(x^\dagger) - y^\delta\rangle$. Notice the facts that

$$\begin{aligned}\mathfrak{M} = &\langle -\alpha B_\delta({B_\delta}^* B_\delta + \alpha I)^{-1}{B_\delta}^* v, s_\alpha + F_\delta(x^\dagger) - y^\delta\rangle + \alpha\langle v, F_\delta(x^\dagger) - y^\delta\rangle\\ &+ \langle -\alpha B_\delta({B_\delta}^* B_\delta + \alpha I)^{-1}(B^* - {B_\delta}^*)v, s_\alpha + F_\delta(x^\dagger) - y^\delta\rangle\end{aligned}$$

and

$$\begin{aligned}&\langle -\alpha B_\delta({B_\delta}^* B_\delta + \alpha I)^{-1}{B_\delta}^* v, F_\delta(x^\dagger) - y^\delta\rangle + \alpha\langle v, F_\delta(x^\dagger) - y^\delta\rangle\\ &\qquad = \langle \alpha^2(B_\delta {B_\delta}^* + \alpha I)^{-1}v, F_\delta(x^\dagger) - y^\delta\rangle.\end{aligned}$$

These equalities yield

$$\begin{aligned}\mathfrak{M} = &\langle -\alpha B_\delta({B_\delta}^* B_\delta + \alpha I)^{-1}{B_\delta}^* v, s_\alpha\rangle + \langle \alpha^2(B_\delta {B_\delta}^* + \alpha I)^{-1}v, F_\delta(x^\dagger) - y^\delta\rangle\\ &+ \langle -\alpha B_\delta({B_\delta}^* B_\delta + \alpha I)^{-1}(B^* - {B_\delta}^*)v, s_\alpha\rangle\\ &+ \langle -\alpha B_\delta({B_\delta}^* B_\delta + \alpha I)^{-1}(B^* - {B_\delta}^*)v, F_\delta(x^\dagger) - y^\delta\rangle.\end{aligned}$$

We use the same arguments as in the previous estimations with the partial isometry U to obtain

$$\begin{aligned}\mathfrak{M} \leq & \frac{L}{2}\|v\|\alpha\|z_\alpha - x^\dagger\|^2 + C_{\mu-1/2}\|w\|\alpha^{\mu+1/2}\|F_\delta(x^\dagger) - y^\delta\| + \|w\|\alpha\delta_{F'}^{2\mu-1}\|F_\delta(x^\dagger) - y^\delta\| \\ & + \frac{L}{2}C_{1/2}\|v\|\sqrt{\alpha}\delta_{F'}\|z_\alpha - x^\dagger\|^2 + C_{1/2}\|v\|\sqrt{\alpha}\delta_{F'}\|F_\delta(x^\dagger) - y^\delta\| \\ \leq & \frac{L}{2}\|v\|\alpha\|z_\alpha - x^\dagger\|^2 + C_{\mu-1/2}\|w\|\alpha^{\mu+1/2}\|F_\delta(x^\dagger) - y^\delta\| + \|w\|\alpha\delta_{F'}^{2\mu-1}\|F_\delta(x^\dagger) - y^\delta\| \\ & + C_{1/2}^2\|v\|^2\alpha\delta_{F'}^2 + \frac{L^2}{8}\|z_\alpha - x^\dagger\|^4 + \frac{1}{2}\|F_\delta(x^\dagger) - y^\delta\|^2.\end{aligned}$$

That implies

$$\begin{aligned}\sqrt{\frac{2}{\alpha}}\sqrt{\mathfrak{M}} \leq & \sqrt{L\|v\|}\|z_\alpha - x^\dagger\| + \sqrt{2C_{\mu-1/2}\|w\|}\sqrt{\alpha^{\mu-1/2}}\|F_\delta(x^\dagger) - y^\delta\|^{1/2} \\ & + \sqrt{2\|w\|}\delta_{F'}^{\mu-1/2}\|F_\delta(x^\dagger) - y^\delta\|^{1/2} + C_{1/2}\|v\|\delta_{F'} \\ & + \frac{L}{2\sqrt{\alpha}}\|z_\alpha - x^\dagger\|^2 + \frac{1}{\sqrt{\alpha}}\|F_\delta(x^\dagger) - y^\delta\|.\end{aligned}$$

We only need to verify the second and the third terms at the *a priori* choice $\alpha \sim (\delta_y + \delta_F + \delta_{F'}^{2\mu+1})^{\frac{2}{2\mu+1}}$ such that

$$\begin{aligned}\sqrt{\alpha^{\mu-1/2}}\|F_\delta(x^\dagger) - y^\delta\|^{1/2} &\sim (\delta_y + \delta_F + \delta_{F'}^{2\mu+1})^{\frac{\mu-1/2}{2\mu+1} + \frac{2\mu+1}{2(2\mu+1)}} \\ &= (\delta_y + \delta_F + \delta_{F'}^{2\mu+1})^{\frac{2\mu}{2\mu+1}}\end{aligned}$$

and

$$\begin{aligned}\delta_{F'}^{\mu-1/2}\|F_\delta(x^\dagger) - y^\delta\|^{1/2} &\sim (\delta_{F'}^{2\mu+1})^{\frac{\mu-1/2}{2\mu+1}}\|F_\delta(x^\dagger) - y^\delta\|^{1/2} \\ &\leq (\delta_y + \delta_F + \delta_{F'}^{2\mu+1})^{\frac{\mu-1/2}{2\mu+1} + \frac{2\mu+1}{2(2\mu+1)}} \\ &= (\delta_y + \delta_F + \delta_{F'}^{2\mu+1})^{\frac{2\mu}{2\mu+1}}.\end{aligned}$$

The theorem is thus proven. □

5 Conclusions

In this article, we investigate the convergence rates of Tikhonov regularization for nonlinear ill-posed problems when both the right-hand side and the operator are corrupted by some noise. After introducing two operator noise models, detailed discussions in Banach and Hilbert spaces are carried out provided with appropriate assumptions. We aim at filling the gap between linear and nonlinear ill-posed problems where comprehensive convergence rates on operator noise have been obtained for the former case.

Acknowledges

This work was initiated while the first author was visiting the Department of Mathematics of the Chemnitz University of Technology. He would like to thank Bernd Hofmann for the kind invitation and nice hospitality. The authors thank the two referees for their helpful comments which improved readability. S. Lu is supported by the National Natural Science Foundation of China (No.11101093) and Shanghai Science and Technology Commission (No.11PJ1400800). J. Flemming is supported by the DFG under grant HO 1454/8-1. Both authors are partially supported by the DFG-NSFC joint project "Regularization of nonlinear ill-posed problems in Banach spaces and conditional stability" (NSFC No.11161130001).

References

[1] R. Bhatia. *Matrix Analysis.* Number 169 in Graduate Texts in Mathematics. Springer, New York, 1997.

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

[3] L. Cavalier and N. Hengartner. Adaptive estimation for inverse problems with noisy operators. *Inverse Problems*, 21:1345–1361, 2005.

[4] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems.* Mathematics and Its Applications. Kluwer Academic Publishers, Dordrecht, 1996.

[5] H. W. Engl, K. Kunisch, and A. Neubauer. Convergence rates for Tikhonov regularisation of non-linear ill-posed problems . *Inverse Problems*, 5:523–540, 1989.

[6] J. Flemming. *Generalized Tikhonov regularization. Basic theory and comprehensive results on convergence rates.* PhD thesis, Chemnitz University of Technology, Chemnitz, Germany, October 2011.

[7] G. H. Golub, P. C. Hansen, and D. P. O'Leary. Tikhonov regularization and total least squares. *SIAM J. Matrix Anal. Appl.*, 21:185–194, 1999.

[8] T. Hein. Convergence rates for regularization of ill-posed problems in Banach spaces by approximate source conditions. *Inverse Problems*, 24(4):045007 (10pp), 2008.

[9] T. Hein and B. Hofmann. Approximate source conditions for nonlinear ill-posed problems—chances and limitations. *Inverse Problems*, 25(3):035033 (16pp), 2009.

[10] M. Hoffmann and M. Reiss. Nonlinear estimation for linear inverse problems with error in the operator. *Ann. Stat.*, 36:310–336, 2008.

[11] B. Hofmann. *Regularization of Applied Inverse and Ill-Posed Problems.* Teubner, Leipzig, 1986.

[12] B. Hofmann. Approximate source conditions in Tikhonov–Phillips regularization and consequences for inverse problems with multiplication operators. *Mathematical Methods in the Applied Sciences*, 29(3):351–371, 2006.

[13] B. Hofmann, D. Düvelmeyer, and K. Krumbiegel. Approximate source conditions in Tikhonov regularization-new analytical results and some numerical studies. *Mathematical Modelling and Analysis*, 11(1):41–56, 2006.

[14] B. Hofmann, B. Kaltenbacher, C. Pöschl, and O. Scherzer. A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators. *Inverse Problems*, 23(3):987–1010, 2007.

[15] B. Hofmann and O. Scherzer. Local ill-posedness and source conditions of operator equations in Hilbert spaces. *Inverse Problems*, 14(5):1189, 1998.

[16] Q. N. Jin and U. Amato. A Discrete Scheme of Landweber Iteration for Solving Nonlinear Ill-Posed Problems. *J. Math. Anal. Appl.*, 253:187–203, 2001.

[17] A. K. Louis. *Inverse und Schlecht Gestellte Probleme.* Teubner, Stuttgart, 1989.

[18] S. Lu, S. V. Pereverzev, and R. Ramlau. An analysis of Tikhonov regularization for nonlinear ill-posed problems under a general smoothness assumption. *Inverse Problems*, 23:217–230, 2007.

[19] S. Lu, S. V. Pereverzev, Y. Shao, and U. Tautenhahn. On the generalized discrepancy principle for Tikhonov regularization in Hilbert scales. *Journal of Integral equations and application*, 22:483–517, 2010.

[20] S. Lu, S. V. Pereverzev, and U. Tautenhahn. Regularized total least squares: Computational aspects and error bounds. *SIAM J. Matrix Anal. Appl.*, 31:918–941, 2009.

[21] C. Marteau. Regularization of inverse problems with unknown operators. *Math. Methods Stat.*, 15:415–443, 2006.

[22] P. Mathé and S. V. Pereverzev. Moduli of continuity for operator valued functions. *Numer. Funct. Anal. Optim.*, 23:623–631, 2002.

[23] P. Mathé and S. V. Pereverzev. Discretization strategy for linear ill-posed problems in variable Hilbert scales. *Inverse Problems*, 19:1263–1277, 2003.

[24] A. Neubauer and O. Scherzer. Finite-dimensional approximation of tikhonov regularized solutions of non-linear ill-posed problems. *Numer. Funct. Anal. Optim.*, 11:85–99, 1990.

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

[26] R. Stück, M. Burger, and T. Hohage. The iteratively regularized Gauss-Newton method with convex constraints and application in 4Pi-Microscopy. *Inverse Problems*, 28:015012 (16pp), 2012.

[27] U. Tautenhahn. Regularization of linear ill-posed problems with noisy right hand side and noisy operator. *J. Inv. Ill-Posed Problems*, 16:507–523, 2008.

[28] A. N. Tikhonov and V. Y. Arsenin. *Solution of Ill-Posed Problems.* Wiley, New York, 1977.

[29] G. M. Vainikko and A. Y. Veretennikov. *Iteration Procedures in Ill-Posed Problems.* Nauka, Moscow, 1986. in Russian.

[30] S. Van Huffel and J. Vandewalle. *The Total Least Squares Problem: Computational Aspects and Analysis.* SIAM, Philadelphia, 1991.