

Convergence Rates for the Iteratively Regularized Landweber Iteration in Banach Space

Barbara Kaltenbacher

► **To cite this version:**

Barbara Kaltenbacher. Convergence Rates for the Iteratively Regularized Landweber Iteration in Banach Space. Dietmar Hömberg; Fredi Tröltzsch. 25th System Modeling and Optimization (CSMO), Sep 2011, Berlin, Germany. Springer, IFIP Advances in Information and Communication Technology, AICT-391, pp.38-48, 2013, System Modeling and Optimization. <10.1007/978-3-642-36062-6_4>. <hal-01347519>

HAL Id: hal-01347519

<https://hal.inria.fr/hal-01347519>

Submitted on 21 Jul 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Convergence rates for the iteratively regularized Landweber iteration in Banach space

Barbara Kaltenbacher*

University of Klagenfurt,
Universitätsstraße 65–67, 9020 Klagenfurt, Austria
barbara.kaltenbacher@aau.at

Abstract. In this paper we provide a convergence rates result for a modified version of Landweber iteration with a priori regularization parameter choice in a Banach space setting.

Keywords: regularization, nonlinear inverse problems, Banach space, Landweber iteration

An increasing number of inverse problems is nowadays posed in a Banach space rather than a Hilbert space setting, cf., e.g., [2, 6, 13] and the references therein.

An Example of a model problem, where the use of non-Hilbert Banach spaces is useful, is the identification of the space-dependent coefficient function c in the elliptic boundary value problem

$$-\Delta u + cu = f \quad \text{in } \Omega \tag{1}$$

$$u = 0 \quad \text{on } \partial\Omega \tag{2}$$

from measurements of u in $\Omega \subseteq \mathbb{R}^d$, $d \in \{1, 2, 3\}$, where f is assumed to be known. Here e.g., the choices $p = 1$ for recovering sparse solutions, $q = \infty$ for modelling uniformly bounded noise, or $q = 1$ for dealing with impulsive noise are particularly promising, see, e.g., [3] and the numerical experiments in Section 7.3.3 of [13].

Motivated by this fact we consider nonlinear ill-posed operator equations

$$F(x) = y \tag{3}$$

where F maps between Banach spaces X and Y .

In the example above, the forward operator F maps the coefficient function c to the solution of the boundary value problem (1), (2), and is well-defined as an operator

$$F : \mathcal{D}(F) \subseteq L^p(\Omega) \rightarrow L^q(\Omega),$$

* Support by the German Science Foundation DFG under grant KA 1778/5-1 and within the Cluster of Excellence in Simulation Technology (EXC 310/1) at the University of Stuttgart is gratefully acknowledged

where $\mathcal{D}(F) = \{c \in X \mid \exists \hat{c} \in L^\infty(\Omega), \hat{c} \geq 0 \text{ a.e.} : \|c - \hat{c}\|_X \leq r\}$, r sufficiently small, for any

$$\begin{aligned} p, q \in [1, \infty], \quad f \in L^1(\Omega) & \quad \text{if } d \in \{1, 2\} \\ p \in [1, \infty], \quad q \in (\frac{d}{2}, \infty], \quad f \in L^s(\Omega), \quad s > \frac{d}{2} & \quad \text{if } d \geq 3, \end{aligned}$$

see Section 1.3 in [13].

Since the given data y^δ are typically contaminated by noise, regularization has to be applied. We are going to assume that the noise level δ in

$$\|y - y^\delta\| \leq \delta \tag{4}$$

is known and provide convergence results in the sense of regularization methods, i.e., as δ tends to zero. In the following, x_0 is some initial guess and we will assume that a solution x^\dagger to (3) exists.

Variational methods in Banach space have been extensively studied in the literature, see, e.g., [2, 10, 6] and the references therein.

Since these generalizations of Tikhonov regularization require computation of a global minimizer, iterative methods are an attractive alternative especially for large scale problems. After convergence results on iterative methods for nonlinear ill-posed operator equations in Banach spaces had already been obtained in the 1990's (cf. the references in [1]) in the special case $X = Y$, the general case $X \neq Y$ has only been treated quite recently, see e.g. [5], [7], and [9] for an analysis of gradient and Newton type iterations. While convergence rates have already been established for the iteratively regularized Gauss-Newton iteration in [7], the question of convergence rates is still open for gradient type, i.e. Landweber methods. It is the aim of this paper to provide such a result.

In order to formulate and later on analyze the method, we have to introduce some basic notations and concepts.

Consider, for some $q \in (1, \infty)$, the duality mapping $J_q^X(x) := \partial \left\{ \frac{1}{q} \|x\|^q \right\}$, which maps from X to its dual X^* . To analyze convergence rates we employ the Bregman distance

$$D_{j_q}(\tilde{x}, x) = \frac{1}{q} \|\tilde{x}\|^q - \frac{1}{q} \|x\|^q - \langle j_q^X(x), \tilde{x} - x \rangle_{X^*..X}$$

(where $j_q^X(x)$ denotes a single valued selection of $J_q^X(x)$) or its shifted version

$$D_q^{x_0}(\tilde{x}, x) := D_{j_q}(\tilde{x} - x_0, x - x_0).$$

Throughout this paper we will assume that X is smooth, which means that the duality mapping is single-valued, and moreover, that X is q -convex, i.e.,

$$D_{j_q}(x, y) \geq c_q \|x - y\|^q \tag{5}$$

for some constant $c_q > 0$. As a consequence, X is reflexive and we also have

$$D_{j_{q^*}}(x^*, y^*) \leq C_{q^*} \|x^* - y^*\|^{q^*}, \tag{6}$$

for some C_{q^*} . Here q^* denotes the dual index $q^* = \frac{q}{q-1}$. Moreover, the duality mapping is bijective and $J_q^{-1} = J_{q^*}^{X^*}$, the latter denoting the (by q -convexity also single-valued) duality mapping on X^* . We will also make use of the identities

$$D_{j_q}(x, y) = D_{j_q}(x, z) + D_{j_q}(z, y) + \langle J_q^X(z) - J_q^X(y), x - z \rangle_{X^*, X} \quad (7)$$

and

$$D_{j_q}(y, x) = D_{j_{q^*}}(J_q^X(x), J_q^X(y)). \quad (8)$$

For more details on the geometry of Banach spaces we refer, e.g., to [12] and the references therein.

We here consider the iteratively regularized Landweber iteration

$$\begin{aligned} J_q^X(x_{n+1}^\delta - x_0) &= (1 - \alpha_n)J_q^X(x_n^\delta - x_0) - \mu_n A_n^* j_p^Y(F(x_n^\delta) - y^\delta), \\ x_{n+1}^\delta &= x_0 + J_{q^*}^{X^*}(J_q^X(x_{n+1}^\delta - x_0)), \quad n = 0, 1, \dots \end{aligned} \quad (9)$$

where we abbreviate

$$A_n = F'(x_n^\delta),$$

which, for an appropriate choice of the sequence $\{\alpha_n\}_{n \in \mathbb{N}} \in (0, 1]$, has been shown to be convergent with rates under a source condition

$$x^\dagger - x_0 \in \mathcal{R}(F'(x^\dagger)^* F'(x^\dagger))^{\nu/2}, \quad (10)$$

with $\nu = 1$ in a Hilbert space setting in [11]. Since the linearized forward operator $F'(x)$ typically has some smoothing property (reflecting the ill-posedness of the inverse problems) condition (10) can often be interpreted as a regularity assumption on the initial error $x^\dagger - x_0$, which is stronger for larger ν .

In the Hilbert space case the proof of convergence rates for the plain Landweber iteration (i.e., (9) with $\alpha_n = 0$) under source conditions (10) relies on the fact that the iteration errors $x_n^\delta - x^\dagger$ remain in the range of $(F'(x^\dagger)^* F'(x^\dagger))^{\nu/2}$ and their preimages under $(F'(x^\dagger)^* F'(x^\dagger))^{\nu/2}$ form a bounded sequence (cf., Proposition 2.11 in [8]). Since carrying over this approach to the Banach space setting would require more restrictive assumptions on the structure of the spaces even in the special case $\nu = 1$, we here consider the modified version with an appropriate choice of $\{\alpha_n\}_{n \in \mathbb{N}} \in (0, 1]$.

In place of the Hilbert space source condition (10), we consider variational inequalities

$$\begin{aligned} \exists \beta > 0 \quad \forall x \in \mathcal{B}_\rho^D(x^\dagger) : \\ |\langle J_q^X(x^\dagger - x_0), x - x^\dagger \rangle_{X^* \times X}| \leq \beta D_q^{x_0}(x^\dagger, x)^{\frac{1-\nu}{2}} \|F'(x^\dagger)(x - x^\dagger)\|^\nu, \end{aligned} \quad (11)$$

cf., e.g., [4], where

$$\mathcal{B}_\rho^D(x^\dagger) = \{x \in X \mid D_q^{x_0}(x^\dagger, x) \leq \rho^2\}$$

with $\rho > 0$ such that $x_0 \in \mathcal{B}_\rho^D(x^\dagger)$. Using the interpolation and the Cauchy-Schwarz inequality, it is readily checked that in the Hilbert space case (10)

implies (11). For more details on such variational inequalities we refer to Section 3.2.3 in [13] and the references therein.

The assumptions on the forward operator besides a condition on the domain

$$\mathcal{B}_\rho^D(x^\dagger) \subseteq \mathcal{D}(F) \quad (12)$$

include a structural condition on its degree of nonlinearity (cf. [4])

$$\begin{aligned} \|(F'(x^\dagger + v) - F'(x^\dagger))v\| &\leq K \|F'(x^\dagger)v\|^{c_1} D_q^{x_0}(x^\dagger, v + x^\dagger)^{c_2}, \\ v \in X, x^\dagger + v &\in \mathcal{B}_\rho^D(x^\dagger), \end{aligned} \quad (13)$$

whose strength depends on the smoothness index in (11). Namely, we assume that

$$c_1 = 1 \text{ or } c_1 + c_2 p > 1 \text{ or } (c_1 + c_2 p \geq 1 \text{ and } K \text{ is sufficiently small}) \quad (14)$$

$$c_1 + c_2 \frac{2\nu}{\nu+1} \geq 1, \quad (15)$$

so that in case $\nu = 1$, a Lipschitz condition on F' , corresponding to $(c_1, c_2) = (0, 1)$ is sufficient.

Here F' denotes the Gateaux derivative of F , hence a Taylor remainder estimate

$$\begin{aligned} &\|F(x_n^\delta) - F(x^\dagger) - F'(x^\dagger)(x_n^\delta - x^\dagger)\| \quad (16) \\ &= \|g(1) - g(0) - F'(x^\dagger)(x_n^\delta - x^\dagger)\| \\ &= \left\| \int_0^1 g'(t) dt - F'(x^\dagger)(x_n^\delta - x^\dagger) \right\| \\ &= \left\| \int_0^1 F'(x^\dagger + t(x_n^\delta - x^\dagger))(x_n^\delta - x^\dagger) dt - F'(x^\dagger)(x_n^\delta - x^\dagger) \right\| \\ &\leq K \|F'(x^\dagger)(x_n^\delta - x^\dagger)\|^{c_1} D_q^{x_0}(x^\dagger, x_n^\delta)^{c_2} \end{aligned} \quad (17)$$

where $g : t \mapsto F(x^\dagger + t(x_n^\delta - x^\dagger))$, follows from (13).

We will assume that in each step the step size $\mu_n > 0$ in (9) is chosen such that

$$\mu_n \frac{1 - 3C(c_1)K}{3(1 - C(c_1)K)} \|F(x_n^\delta) - y^\delta\|^p - 2^{q^*+q-2} C_{q^*} \mu_n^{q^*} \|A_n^* j_p^Y(F(x_n^\delta) - y^\delta)\|^{q^*} \geq 0 \quad (18)$$

where $C(c_1) = c_1^{c_1} (1 - c_1)^{1-c_1}$, and c_1, K are as in (13), which is possible, e.g.,

by a choice $0 < \mu_n \leq C_\mu \frac{\|F(x_n^\delta) - y^\delta\|^{\frac{q-p}{q-1}}}{\|A_n\|^{q^*}} =: \bar{\mu}_n$ with $C_\mu := \frac{2^{2-q^*-q}}{3} \frac{1-3C(c_1)K}{(1-C(c_1)K)C_{q^*}}$. If

$$p \geq q \quad (19)$$

and F, F' are bounded on $\mathcal{B}_\rho^D(x^\dagger)$, it is possible to bound $\bar{\mu}_n$ away from zero

$$\bar{\mu}_n \geq C_\mu \left(\sup_{x \in \mathcal{B}_\rho^D(x^\dagger)} (\|F(x) - y\| + \bar{\delta})^{p-q} \|F'(x)\|^q \right)^{-1/(q-1)} =: \underline{\mu} \quad (20)$$

for $\delta \in [0, \bar{\delta}]$, provided the iterates remain in $\mathcal{B}_\rho^D(x^\dagger)$ (which we will show by induction in the proof of Theorem 1). Hence, there exist $\underline{\mu}, \bar{\mu} > 0$ independent of n and δ such that we can choose

$$0 < \underline{\mu} \leq \mu_n \leq \bar{\mu}, \quad (21)$$

(e.g., by simply setting $\mu_n \equiv \underline{\mu}$).

Moreover, we will use an a priori choice of the stopping index n_* according to

$$n_*(\delta) = \min\{n \in \mathbb{N} : \alpha_n^{\frac{\nu+1}{p(\nu+1)-2\nu}} \leq \tau\delta\}, \quad (22)$$

and of $\{\alpha_n\}_{n \in \mathbb{N}}$ such that

$$\left(\frac{\alpha_{n+1}}{\alpha_n}\right)^{\frac{2\nu}{p(\nu+1)-2\nu}} + \frac{1}{3}\alpha_n - 1 \geq c\alpha_n \quad (23)$$

for some $c \in (0, \frac{1}{3})$ independent of n , where $\nu \in [0, 1]$ is the exponent in the variational inequality (11).

Remark 1. A possible choice of $\{\alpha_n\}_{n \in \mathbb{N}}$ satisfying (23) and smallness of α_{\max} is given by

$$\alpha_n = \frac{\alpha_0}{(n+1)^x}$$

with $x \in (0, 1]$ such that $3x\theta < \alpha_0$ sufficiently small, since then with $c := \frac{1}{3} - \frac{x\theta}{\alpha_0} > 0$, using the abbreviation $\theta = \frac{2\nu}{p(\nu+1)-2\nu} \in [0, \frac{1}{p-1}]$ we get by the Mean Value Theorem

$$\begin{aligned} & \left(\frac{\alpha_{n+1}}{\alpha_n}\right)^\theta + \left(\frac{1}{3} - c\right)\alpha_n - 1 \\ &= \frac{\alpha_n}{\alpha_0} \left\{ \alpha_0 \left(\frac{1}{3} - c\right) - \frac{(n+2)^{x\theta} - (n+1)^{x\theta}}{(n+2)^{x\theta}} (n+1)^x \right\} \\ &= \frac{\alpha_n}{\alpha_0} \left\{ \alpha_0 \left(\frac{1}{3} - c\right) - \frac{x\theta(n+1+t)^{x\theta-1}}{(n+2)^{x\theta}} (n+1)^x \right\} \\ &\geq \frac{\alpha_n}{\alpha_0} \left\{ \alpha_0 \left(\frac{1}{3} - c\right) - x\theta \frac{(n+1)^x}{n+1+t} \right\} \geq 0, \end{aligned}$$

for some $t \in [0, 1]$.

Theorem 1. *Assume that X is smooth and q -convex, that x_0 is sufficiently close to x^\dagger , i.e., $x_0 \in \mathcal{B}_\rho^D(x^\dagger)$, (which by (5) implies that $\|x^\dagger - x_0\|$ is also small), that a variational inequality (11) with $\nu \in (0, 1]$ and β sufficiently small is satisfied, that F satisfies (13) with (14), (15), that F and F' are continuous and uniformly bounded in $\mathcal{B}_\rho^D(x^\dagger)$, that (12) holds and that*

$$q^* \geq \frac{2\nu}{p(\nu+1) - 2\nu} + 1. \quad (24)$$

Let $n_*(\delta)$ be chosen according to (22) with τ sufficiently large. Moreover assume that (19) holds and the sequence $\{\mu_n\}_{n \in \mathbb{N}}$ is chosen such that (21) holds for $0 < \underline{\mu} < \bar{\mu}$ according to (20), and let the sequence $\{\alpha_n\}_{n \in \mathbb{N}} \subseteq [0, 1]$ be chosen such that (23) holds, and $\alpha_{\max} = \max_{n \in \mathbb{N}} \alpha_n$ is sufficiently small.

Then, the iterates x_{n+1}^δ remain in $\mathcal{B}_\rho^D(x^\dagger)$ for all $n \leq n_*(\delta) - 1$ with n_* according to (22). Moreover, we obtain optimal convergence rates

$$D_q^{x_0}(x^\dagger, x_{n_*}) = O(\delta^{\frac{2\nu}{\nu+1}}), \quad \text{as } \delta \rightarrow 0 \quad (25)$$

as well as in the noise free case $\delta = 0$

$$D_q^{x_0}(x^\dagger, x_n) = O\left(\alpha_n^{\frac{2\nu}{p(\nu+1)-2\nu}}\right) \quad (26)$$

for all $n \in \mathbb{N}$.

Remark 2. Note that the rate exponent in (26) $\frac{2\nu}{p(\nu+1)-2\nu} = \frac{2\nu}{\nu+1}(p - \frac{2\nu}{\nu+1})^{-1}$, always lies in the interval $[0, \frac{1}{p-1}]$, since $\frac{2\nu}{\nu+1} \in [0, 1]$.

Moreover, note that Theorem 1 provides a results on rates only, but no convergence result without variational inequality. This corresponds to the situation from [11] in a Hilbert space setting.

Proof. First of all, for $x_n^\delta \in \mathcal{B}_\rho^D(x^\dagger)$, (13) allows us to estimate as follows (see also (16)) in case $c_1 \in [0, 1)$:

$$\begin{aligned} & \|F(x_n^\delta) - F(x^\dagger) - A(x_n^\delta - x^\dagger)\| \\ & \leq K \|A(x_n^\delta - x^\dagger)\|^{c_1} D_q^{x_0}(x^\dagger, x_n^\delta)^{c_2} \\ & \leq C(c_1)K \left(\|A(x_n^\delta - x^\dagger)\| + D_q^{x_0}(x^\dagger, x_n^\delta)^{\frac{c_2}{1-c_1}} \right), \end{aligned} \quad (27)$$

where we have used the abbreviation $A = F'(x^\dagger)$ and the elementary estimate

$$a^{1-\lambda}b^\lambda \leq C(\lambda)(a+b) \text{ with } C(\lambda) = \lambda^\lambda(1-\lambda)^{1-\lambda} \text{ for } a, b \geq 0, \lambda \in (0, 1), \quad (28)$$

and therewith, by the second triangle inequality,

$$\|A(x_n^\delta - x^\dagger)\| \leq \frac{1}{1 - C(c_1)K} \left(\|F(x_n^\delta) - F(x^\dagger)\| + C(c_1)K D_q^{x_0}(x^\dagger, x_n^\delta)^{\frac{c_2}{1-c_1}} \right) \quad (29)$$

as well as analogously

$$\begin{aligned} & \|F(x_n^\delta) - F(x^\dagger) - A_n(x_n^\delta - x^\dagger)\| \\ & \leq 2C(c_1)K \left(\|A(x_n^\delta - x^\dagger)\| + D_q^{x_0}(x^\dagger, x_n^\delta)^{\frac{c_2}{1-c_1}} \right) \\ & \leq \frac{2C(c_1)K}{1 - C(c_1)K} \left(\|F(x_n^\delta) - F(x^\dagger)\| + D_q^{x_0}(x^\dagger, x_n^\delta)^{\frac{c_2}{1-c_1}} \right). \end{aligned} \quad (30)$$

For any $n \leq n_*$ according to (22), by (7) we have

$$\begin{aligned}
 & D_q^{x_0}(x^\dagger, x_{n+1}^\delta) - D_q^{x_0}(x^\dagger, x_n^\delta) \\
 &= D_q^{x_0}(x_n^\delta, x_{n+1}^\delta) + \langle J_q^X(x_n^\delta - x_0) - J_q^X(x_{n+1}^\delta - x_0), x^\dagger - x_n^\delta \rangle_{X^* \times X} \\
 &= D_q^{x_0}(x_n^\delta, x_{n+1}^\delta) - \mu_n \langle j_p^Y(F(x_n^\delta) - y^\delta), A_n(x_n^\delta - x^\dagger) \rangle_{Y^* \times Y} \\
 &\quad + \alpha_n \langle J_q^X(x^\dagger - x_0), x^\dagger - x_n^\delta \rangle_{X^* \times X} \\
 &\quad - \alpha_n \langle J_q^X(x^\dagger - x_0) - J_q^X(x_n^\delta - x_0), x^\dagger - x_n^\delta \rangle_{X^* \times X}
 \end{aligned} \tag{31}$$

where the terms on the right hand side can be estimated as follows.

By (6) and (8) we have

$$\begin{aligned}
 & D_q^{x_0}(x_n^\delta, x_{n+1}^\delta) \\
 &\leq C_{q^*} \|J_q^X(x_{n+1}^\delta - x_0) - J_q^X(x_n^\delta - x_0)\|^{q^*} \\
 &= C_{q^*} \|\alpha_n J_q^X(x_n^\delta - x_0) + \mu_n A_n^* j_p^Y(F(x_n^\delta) - y^\delta)\|^{q^*} \\
 &\leq 2^{q^*-1} C_{q^*} \left(\alpha_n^{q^*} \|x_n^\delta - x_0\|^q + \mu_n^{q^*} \|A_n^* j_p^Y(F(x_n^\delta) - y^\delta)\|^{q^*} \right) \\
 &\leq 2^{q^*-1} C_{q^*} \left(\alpha_n^{q^*} (2^{q-1} (\|x^\dagger - x_0\|^q + \frac{1}{c_q} D_q^{x_0}(x^\dagger, x_n^\delta)) + \mu_n^{q^*} \|A_n^* j_p^Y(F(x_n^\delta) - y^\delta)\|^{q^*}) \right)
 \end{aligned} \tag{32}$$

where we have used the triangle inequality in X^* and X , the inequality

$$(a + b)^\lambda \leq 2^{\lambda-1}(a^\lambda + b^\lambda) \quad \text{for } a, b \geq 0, \lambda \geq 1, \tag{34}$$

and (5).

For the second term on the right hand side of (31) we get, using (30), (28), (34),

$$\begin{aligned}
 & \langle j_p^Y(F(x_n^\delta) - y^\delta), A_n(x_n^\delta - x^\dagger) \rangle_{Y^* \times Y} \\
 &= \langle j_p^Y(F(x_n^\delta) - y^\delta), F(x_n^\delta) - y^\delta \rangle_{Y^* \times Y} \\
 &\quad - \langle j_p^Y(F(x_n^\delta) - y^\delta), F(x_n^\delta) - y^\delta - A_n(x_n^\delta - x^\dagger) \rangle_{Y^* \times Y} \\
 &\geq \frac{1 - 3C(c_1)K}{1 - C(c_1)K} \|F(x_n^\delta) - y^\delta\|^p \\
 &\quad - \|F(x_n^\delta) - y^\delta\|^{p-1} \left(\frac{2C(c_1)K}{1 - C(c_1)K} D_q^{x_0}(x^\dagger, x_n^\delta)^{\frac{c_2}{1-c_1}} + \frac{1 + C(c_1)K}{1 - C(c_1)K} \delta \right) \\
 &= \frac{1 - 3C(c_1)K}{1 - C(c_1)K} \|F(x_n^\delta) - y^\delta\|^p \\
 &\quad - \left(\frac{1 - 3C(c_1)K}{3C(\frac{p-1}{p})(1 - C(c_1)K)} \|F(x_n^\delta) - y^\delta\|^p \right)^{\frac{p-1}{p}} \left(\frac{(3C(\frac{p-1}{p}))^{p-1}}{(1 - C(c_1)K)} \right)^{\frac{1}{p}} \\
 &\quad \left(2C(c_1)K D_q^{x_0}(x^\dagger, x_n^\delta)^{\frac{c_2}{1-c_1}} + (1 + C(c_1)K)\delta \right)
 \end{aligned}$$

$$\begin{aligned}
&\geq \frac{1-3C(c_1)K}{1-C(c_1)K} \|F(x_n^\delta) - y^\delta\|^p - C\left(\frac{p-1}{p}\right) \left\{ \frac{1-3C(c_1)K}{3C\left(\frac{p-1}{p}\right)(1-C(c_1)K)} \|F(x_n^\delta) - y^\delta\|^p \right. \\
&\quad \left. + \frac{(3C\left(\frac{p-1}{p}\right))^{p-1}}{(1-C(c_1)K)} 2^{p-1} \left((2C(c_1)K)^p D_q^{x_0}(x^\dagger, x_n^\delta)^{\frac{c_2 p}{1-c_1}} + (1+C(c_1)K)^p \delta^p \right) \right\}. \tag{35}
\end{aligned}$$

Using the variational inequality (11), (29), and

$$(a+b)^\lambda \leq (a^\lambda + b^\lambda) \quad \text{for } a, b \geq 0, \lambda \in [0, 1], \tag{36}$$

we get

$$\begin{aligned}
&|\alpha_n \langle J_q^X(x^\dagger - x_0), x^\dagger - x_n^\delta \rangle_{X^* \times X}| \\
&\leq \beta \alpha_n D_q^{x_0}(x^\dagger, x_n^\delta)^{\frac{1-\nu}{2}} \|F'(x^\dagger)(x_n^\delta - x^\dagger)\|^\nu \\
&\leq \beta \alpha_n D_q^{x_0}(x^\dagger, x_n^\delta)^{\frac{1-\nu}{2}} \frac{1}{(1-C(c_1)K)^\nu} \left(\|F(x_n^\delta) - y^\delta\| + \delta + C(c_1)K D_q^{x_0}(x^\dagger, x_n^\delta)^{\frac{c_2}{1-c_1}} \right)^\nu \\
&\leq \beta \alpha_n D_q^{x_0}(x^\dagger, x_n^\delta)^{\frac{1-\nu}{2}} \epsilon^{-\nu} \left(\epsilon \frac{1}{(1-C(c_1)K)^\nu} (\|F(x_n^\delta) - y^\delta\| + \delta) \right)^\nu \\
&\quad + \beta \alpha_n \left(\frac{C(c_1)K}{(1-C(c_1)K)} \right)^\nu D_q^{x_0}(x^\dagger, x_n^\delta)^{\frac{1-\nu}{2} + \frac{\nu c_2}{1-c_1}} \\
&\leq C\left(\frac{\nu}{p}\right) \left\{ \left(\beta \alpha_n D_q^{x_0}(x^\dagger, x_n^\delta)^{\frac{1-\nu}{2}} \epsilon^{-\nu} \right)^{\frac{p}{p-\nu}} + \left(\epsilon \frac{1}{(1-C(c_1)K)^\nu} (\|F(x_n^\delta) - y^\delta\| + \delta) \right)^p \right\} \\
&\quad + \beta \alpha_n \left(\frac{C(c_1)K}{(1-C(c_1)K)} \right)^\nu D_q^{x_0}(x^\dagger, x_n^\delta)^{\frac{1-\nu}{2} + \frac{\nu c_2}{1-c_1}} \\
&= C\left(\frac{\nu}{p}\right) \left\{ \left(\beta \epsilon^{-\nu} \right)^{\frac{p}{p-\nu}} \left(3C\left(\frac{\nu}{p}\right) C\left(\frac{p(1-\nu)}{2(p-\nu)}\right) \right)^{\frac{p(1-\nu)}{2(p-\nu)}} \alpha_n^{\frac{p(1+\nu)}{2(p-\nu)}} \left(\frac{\alpha_n D_q^{x_0}(x^\dagger, x_n^\delta)}{3C\left(\frac{\nu}{p}\right) C\left(\frac{p(1-\nu)}{2(p-\nu)}\right)} \right)^{\frac{p(1-\nu)}{2(p-\nu)}} \right\} \\
&\quad + \left(\epsilon \frac{1}{(1-C(c_1)K)^\nu} (\|F(x_n^\delta) - y^\delta\| + \delta) \right)^p \\
&\quad + \beta \alpha_n \left(\frac{C(c_1)K}{(1-C(c_1)K)} \right)^\nu D_q^{x_0}(x^\dagger, x_n^\delta)^{\frac{1-\nu-c_1+\nu c_1+2\nu c_2}{2(1-c_1)}} \\
&\leq C\left(\frac{\nu}{p}\right) \left\{ C\left(\frac{p(1-\nu)}{2(p-\nu)}\right) \left[\left(\beta \epsilon^{-\nu} \left(3C\left(\frac{\nu}{p}\right) C\left(\frac{p(1-\nu)}{2(p-\nu)}\right) \right)^{\frac{1-\nu}{2}} \right)^{\frac{2p}{p(\nu+1)-2\nu}} \alpha_n^{\frac{p(1+\nu)}{p(\nu+1)-2\nu}} \right. \right. \\
&\quad \left. \left. + \left(\frac{\alpha_n D_q^{x_0}(x^\dagger, x_n^\delta)}{3C\left(\frac{\nu}{p}\right) C\left(\frac{p(1-\nu)}{2(p-\nu)}\right)} \right) \right] \right\}
\end{aligned}$$

$$\begin{aligned}
 & + \left(\epsilon \frac{1}{(1 - C(c_1)K)^\nu} (\|F(x_n^\delta) - y^\delta\| + \delta) \right)^p \Big\} \\
 & + \frac{1}{3} \alpha_n D_q^{x_0}(x^\dagger, x_n^\delta)
 \end{aligned} \tag{37}$$

where we have used (28) two times and $\epsilon > 0$ will be chosen as a sufficiently small number below. Moreover, by (15), the exponent $\frac{1-\nu-c_1+\nu c_1+2\nu c_2}{2(1-c_1)} = 1 + \frac{1+\nu}{2(1-c_1)}(c_1 + \frac{2\nu}{\nu+1}c_2 - 1)$ is larger or equal to one and β is sufficiently small so that $\beta \left(\frac{C(c_1)K}{(1-C(c_1)K)} \right)^\nu \rho^{\frac{1-\nu-c_1+\nu c_1+2\nu c_2}{2(1-c_1)}-1} < \frac{1}{3}$.

Finally, we have that

$$\begin{aligned}
 \langle J_q^X(x^\dagger - x_0) - J_q^X(x_n^\delta - x_0), x^\dagger - x_n^\delta \rangle_{X^* \times X} &= D_q^{x_0}(x^\dagger, x_n^\delta) + D_q^{x_0}(x_n^\delta, x^\dagger) \\
 &\geq D_q^{x_0}(x^\dagger, x_n^\delta)
 \end{aligned} \tag{38}$$

Inserting estimates (32)-(38) with $\epsilon = 2^{p-1} \mu_n^{1/p} \left(\frac{1-3C(c_1)K}{3(1-C(c_1)K)} \right)^{1/p} \frac{(1-C(c_1)K)^\nu}{C(\frac{\nu}{p})}$ into (31) and using boundedness away from zero of μ_n and the abbreviations

$$\begin{aligned}
 d_n &= D_q^{x_0}(x^\dagger, x_n^\delta)^{1/2} \\
 C_0 &= 6^{p-1} C(\frac{p-1}{p})^p \frac{(2C(c_1)K)^p}{(1-C(c_1)K)} \\
 C_1 &= 2^{q^*+q-2} \frac{C_{q^*}}{c_q} \\
 C_2 &= C(\frac{\nu}{p}) C(\frac{p(1-\nu)}{2(p-\nu)}) \left(\beta \epsilon^{-\nu} (3C(\frac{\nu}{p}) C(\frac{p(1-\nu)}{2(p-\nu)})^{\frac{1-\nu}{2}}) \right)^{\frac{2p}{p(\nu+1)-2\nu}} \\
 C_3 &= 2^{q^*+q-2} C_{q^*} \|x^\dagger - x_0\|^q \\
 C_4 &= 2^{p-1} C(\frac{\nu}{p}) \bar{\epsilon} \frac{1}{(1-C(c_1)K)^\nu} + 6^{p-1} C(\frac{p-1}{p})^p \frac{(1+C(c_1)K)^p}{1-C(c_1)K} \\
 \underline{\epsilon} &= 2^{p-1} \underline{\mu}^{1/p} \left(\frac{1-3C(c_1)K}{3(1-C(c_1)K)} \right)^{1/p} \frac{(1-C(c_1)K)^\nu}{C(\frac{\nu}{p})} \\
 \bar{\epsilon} &= 2^{p-1} \bar{\mu}^{1/p} \left(\frac{1-3C(c_1)K}{3(1-C(c_1)K)} \right)^{1/p} \frac{(1-C(c_1)K)^\nu}{C(\frac{\nu}{p})}
 \end{aligned}$$

we obtain

$$\begin{aligned}
 d_{n+1}^2 &\leq C_0 d_n^{\frac{2c_2 p}{1-c_1}} + (1 - \frac{1}{3} \alpha_n + C_1 \alpha_n^{q^*}) d_n^2 + C_2 \alpha_n^{\frac{p(1+\nu)}{p(\nu+1)-2\nu}} + C_3 \alpha_n^{q^*} + C_4 \delta^p \\
 &- \left(\mu_n \frac{1-3C(c_1)K}{3(1-C(c_1)K)} \|F(x_n^\delta) - y^\delta\|^p - 2^{q^*+q-2} C_{q^*} \mu_n^{q^*} \|A_n^* J_p^Y(F(x_n^\delta) - y^\delta)\|^{q^*} \right).
 \end{aligned}$$

Here the last term is nonpositive due to the choice (18) of μ_n , so that we arrive at

$$d_{n+1}^2 \leq C_0 d_n^{\frac{2c_2 p}{1-c_1}} + \left(1 - \frac{1}{3}\alpha_n + C_1 \alpha_n^{q^*}\right) d_n^2 + \underbrace{(C_2 + C_3 + C_4 \tau^{-p})}_{=: C_5} \alpha_n^{\frac{p(1+\nu)}{p(\nu+1)-2\nu}} \quad (39)$$

where we have used (24) and the stopping rule (22). Denoting

$$\gamma_n := \frac{d_n^2}{\alpha_n^{\frac{2\nu}{p(\nu+1)-2\nu}}}$$

we get the following recursion

$$\gamma_{n+1} \leq C_0 \left(\frac{\alpha_n}{\alpha_{n+1}}\right)^\theta \alpha_n^{\theta\omega} \gamma_n^\omega + \left(\frac{\alpha_n}{\alpha_{n+1}}\right)^\theta \left(1 - \frac{1}{3}\alpha_n + C_1 \alpha_n^{q^*}\right) \gamma_n + C_5 \left(\frac{\alpha_n}{\alpha_{n+1}}\right)^\theta \alpha_n \quad (40)$$

with

$$\theta = \frac{2\nu}{p(\nu+1)-2\nu} \quad \omega = \frac{c_2 p}{1-c_1},$$

where

$$\omega \geq 1$$

by (14) and

$$\theta\omega = \frac{p}{p - \frac{2\nu}{\nu+1}} \frac{c_2 \frac{2\nu}{\nu+1}}{1-c_1} \geq 1$$

due to assumption (15). Hence as sufficient conditions for uniform boundedness of $\{\gamma_n\}_{n \leq n_*}$ by $\bar{\gamma}$ and for $x_{n+1}^\delta \in \mathcal{B}_\rho^D(x^\dagger)$ we get

$$\bar{\gamma} \leq \rho^2 \quad (41)$$

$$C_0 \alpha_n^{\theta\omega-1} \bar{\gamma}^\omega - \left\{ \left(\frac{\alpha_{n+1}}{\alpha_n}\right)^\theta + \frac{1}{3}\alpha_n - 1 - C_1 \alpha_n^{q^*} \right\} \alpha_n^{-1} \bar{\gamma} + C_5 \leq 0, \quad (42)$$

where by $q^* > 1$, (15) the factors $C_0 \alpha_n^{\theta\omega-1}$, $C_1 \alpha_n^{q^*-1}$ and C_5 can be made small for small α_{\max} , β , $\|x^\dagger - x_0\|$ and large τ . We use this fact to achieve

$$C_0 \alpha_n^{\theta\omega-1} \rho^{\omega-1} + C_1 \alpha_n^{q^*-1} \leq \tilde{c} < c$$

with \tilde{c} independent of n , which together with (23) yields sufficiency of

$$\frac{C_5}{c - \tilde{c}} \leq \bar{\gamma} \leq \rho^2$$

for (41), (42), which for any (even small) prescribed ρ is indeed enabled by possibly decreasing β , $\|x^\dagger - x_0\|$, τ^{-1} , and therewith C_5 .

In case $c_1 = 1$, estimates (29), (30) simplify to

$$\|A(x_n^\delta - x^\dagger)\| \leq \frac{1}{1 - \rho^{2c_2} K} \|F(x_n^\delta) - F(x^\dagger)\| \quad (43)$$

and

$$\|F(x_n^\delta) - F(x^\dagger) - A_n(x_n^\delta - x^\dagger)\| \leq \frac{2\rho^{2c_2} K}{1 - \rho^{2c_2} K} \|F(x_n^\delta) - F(x^\dagger)\|. \quad (44)$$

Therewith, the terms containing $D_q^{x_0}(x^\dagger, x_n^\delta)^{\frac{c_2}{1-c_1}}$ are removed and $C(c_1)$ is replaced by ρ^{2c_2} in (32)-(38), so that we end up with a recursion of the form (40) (with C_0 replace by zero) as before. Hence the remainder of the proof of uniform boundedness of γ_n can be done in the same way as in case $c_1 < 1$.

In case $\delta = 0$, i.e., $n_* = \infty$, uniform boundedness of $\{\gamma_n\}_{n \in \mathbb{N}}$ implies (26). For $\delta > 0$ we get (25) by using (22) in

$$D_q^{x_0}(x^\dagger, x_{n_*}) = \gamma_{n_*} \alpha_{n_*}^{\frac{2\nu}{p(\nu+1)-2\nu}} \leq \bar{\gamma} \alpha_{n_*}^{\frac{2\nu}{p(\nu+1)-2\nu}} \leq \bar{\gamma}(\tau\delta)^{\frac{2\nu}{\nu+1}}$$

Remark 3. In view of estimate (39), an optimal choice of α_n would be one that minimizes the right hand side. At least in the special case that the same power of α_n appears in the last two terms, i.e., $\frac{p(1+\nu)}{p(\nu+1)-2\nu} = q^*$, elementary calculus yields

$$(\alpha_n^{opt})^{\frac{2\nu}{p(\nu+1)-2\nu}} = \frac{D_q^{x_0}(x^\dagger, x_n^\delta)}{3q^*(C_1 D_q^{x_0}(x^\dagger, x_n^\delta) + C_5)},$$

which shows that the obtained relation $D_q^{x_0}(x^\dagger, x_n^\delta) \sim \alpha_n^{\frac{2\nu}{p(\nu+1)-2\nu}}$ is indeed reasonable and probably even optimal.

References

1. A. B. Bakushinsky and M. Yu. Kokurin. *Iterative methods for approximate solution of inverse problems*. Springer, Dordrecht, 2004.
2. M. Burger and S. Osher. Convergence rates of convex variational regularization. *Inverse Problems*, 20(5):1411–1421, 2004.
3. Christian Clason, Bangti Jin: A semi-smooth Newton method for nonlinear parameter identification problems with impulsive noise, *SIAM J. Imaging Sci* 5:505–538, 2012.
4. T. Hein and B. Hofmann. Approximate source conditions for nonlinear ill-posed problems – chances and limitations. *Inverse Problems*, 25:035003 (16pp), 2009.
5. T. Hein and K. Kazimierski. Accelerated Landweber iteration in Banach spaces. *Inverse Problems*, 26:055002 (17pp), 2010.
6. 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.
7. B. Kaltenbacher and B. Hofmann. Convergence rates for the iteratively regularized Gauss-Newton method in Banach spaces. *Inverse Problems*, 26:035007 (21pp), 2010.

8. B. Kaltenbacher, A. Neubauer, and O. Scherzer. *Iterative Regularization Methods for Nonlinear Ill-posed Problems*. de Gruyter, 2007.
9. B. Kaltenbacher, F. Schöpfer, and Th. Schuster. Convergence of some iterative methods for the regularization of nonlinear ill-posed problems in Banach spaces. *Inverse Problems*, 25, 2009. 065003 doi: 10.1088/0266-5611/25/6/065003.
10. A. Neubauer, T. Hein, B. Hofmann, S. Kindermann, and U. Tautenhahn. Improved and extended results for enhanced convergence rates of Tikhonov regularization in Banach spaces. *Appl. Anal.*, 89(11):1729–1743, 2010.
11. O. Scherzer. A modified Landweber iteration for solving parameter estimation problems. *Appl. Math. Optim.*, 38:45–68, 1998.
12. F. Schöpfer, A. K. Louis, and T. Schuster. Nonlinear iterative methods for linear ill-posed problems in Banach spaces. *Inverse Problems*, 22(1):311–329, 2006.
13. T. Schuster, B. Kaltenbacher, B. Hofmann, and K. Kazimierski. *Regularization Methods in Banach Spaces*. de Gruyter, 2007.