Implicit iteration methods in Hilbert scales under general smoothness conditions

Qinian Jin\textsuperscript{1} and Ulrich Tautenhahn\textsuperscript{2}

\textsuperscript{1} Department of Mathematics, Virginia Tech Blacksburg, VA 24061-0123, USA
\textsuperscript{2} Department of Mathematics, University of Applied Sciences Zittau/Görlitz, PO Box 1455, 02755 Zittau, Germany

E-mail: qnjin@math.vt.edu and u.tautenhahn@hs-zigr.de

Received 20 September 2010, in final form 11 January 2011
Published 17 March 2011
Online at stacks.iop.org/IP/27/045012

Abstract

For solving linear ill-posed problems, regularization methods are required when the right-hand side is with some noise. In this paper regularized solutions are obtained by implicit iteration methods in Hilbert scales. By exploiting operator monotonicity of certain functions and interpolation techniques in variable Hilbert scales, we study these methods under general smoothness conditions. Order optimal error bounds are given in case the regularization parameter is chosen either \textit{a priori} or \textit{a posteriori} by the discrepancy principle. For realizing the discrepancy principle, some fast algorithm is proposed which is based on Newton’s method applied to some properly transformed equations.

1. Introduction

In this paper we are interested in solving ill-posed problems

\[ Ax = y, \]

where \( A \in L(X, Y) \) is a linear, injective and bounded operator with non-closed range \( \mathcal{R}(A) \) and \( X, Y \) are Hilbert spaces with corresponding inner products \((\cdot, \cdot)\) and norms \( \| \cdot \| \). Throughout we assume that \( y \in \mathcal{R}(A) \) so that (1.1) has a unique solution \( x^\dagger \in X \). We further assume that \( y \) is unknown and \( y^\delta \in Y \) is the available noisy right-hand side with

\[ \| y - y^\delta \| \leq \delta. \]

In the recent literature many aspects of treating ill-posed problems with a noisy right-hand side have been studied. For an overview, see, e.g., [3, 38]. The numerical treatment of ill-posed problems (1.1) with noisy data \( y^\delta \) requires the application of special regularization methods. In this paper we study \textit{implicit iteration methods in Hilbert scales}, in which regularized solutions \( x^\delta_k \) are obtained by

\[ x^\delta_k = x^\delta_{k-1} - (A^* A + \alpha_k B^2)^{-1} A^* (A x^\delta_{k-1} - y^\delta), \quad k = 1, 2, \ldots, n, \quad x^\delta_0 = x_0 \]
where $B : D(B) \subset X \to X$ is some unbounded densely defined self-adjoint strictly positive definite operator, $\alpha_k > 0$ are properly chosen real numbers, $s$ is some generally nonnegative number that controls the smoothness to be introduced into the regularization procedure and $x_0$ is some properly chosen starting value. In these regularization methods, the positive number

$$\sigma_n := \sum_{k=1}^{n} \frac{1}{\alpha_k}$$  \hspace{1cm} (1.3)

plays the role of the regularization parameter. For results of extending this method to nonlinear ill-posed problems, we recommend the recent paper [13]. For the special case $s = 0$, see [6] in the linear case and [5, 10, 12, 14, 30] in the nonlinear case.

Method (1.2) with $n = 1$ and $x_0 = 0$ is the method of Tikhonov regularization in Hilbert scales which has been studied by Natterer [27]. From this paper we know that under the assumptions $\|B^{-s}x\| \sim \|Ax\|$ and $\|B^p x\| \leq E$ the Tikhonov regularized solution $x_\delta^0$ of problem (1.2) with $n = 1, x_0 = 0$ and $\alpha_1 = \alpha$ guarantees order optimal error bounds

$$\|x_\delta^0 - x^\dagger\| = O(\delta^{p/(a+s)}) \quad \text{for} \quad p \leq 2s + a$$  \hspace{1cm} (1.4)

in case $\alpha$ is chosen a priori by $\alpha \sim \delta^{2(s+a)/(a+p)}$. This result has been extended

(i) to the case of choosing $\alpha$ a posteriori by the discrepancy principle, see, e.g., [28, 32–34],
(ii) to a general regularization scheme, see, e.g., [3, 34],
(iii) to the case of general source conditions including infinitely smoothing operators $A$, see, e.g., [9, 17, 21, 22, 24], and
(iv) to the case of nonlinear ill-posed problems, see, e.g., [3, 11, 15, 29, 32, 35].

Our paper is organized as follows. In section 2, we collect some preliminaries on properties of the implicit iteration methods in Hilbert scales, formulate our general smoothness conditions and give some consequences that follow from the general smoothness conditions by exploiting either operator monotonicity or interpolation in variable Hilbert scales. In section 3, we treat the case of a priori parameter choice of the regularization parameter $\sigma_n$ and in section 4 we treat the case of choosing $\sigma_n$ a posteriori by the discrepancy principle. In section 5 we discuss practical issues of choosing the starting value $x_0$ and the parameters $s, n$ and $\alpha_k, k = 1, \ldots, n$. In particular, some fast globally convergent algorithm for realizing the discrepancy principle is proposed which is based on Newton’s method applied to some properly transformed equations. For testing the algorithm, numerical experiments are performed in section 6.

2. Preliminaries

2.1. Properties of the regularization method

In our further study, instead with $B$, we will work with the inverse $G = B^{-1}$. The following proposition gives us some equivalent representation for $x_\delta^n$ defined by (1.2) along with some preliminary properties which will be useful for deriving order optimal error bounds.

**Proposition 2.1.** Let $T = AG'$, $G = B^{-1}$ and $\sigma_n$ be defined by (1.3). Then, the regularized solution $x_\delta^n$ defined by (1.2) possesses the representation

$$x_\delta^n - x_0 = G^s g_n(T^* T)^s (y^\delta - A x_0) \quad \text{with} \quad g_n(\lambda) = -\lambda \left( 1 - \prod_{k=1}^{n} \frac{\alpha_k}{\lambda + \alpha_k} \right).$$  \hspace{1cm} (2.1)
In addition, the function $g_n : (0, c] \to (0, \infty)$ with $c = \|T\|^2$ and the corresponding residual function $r_n(\lambda) := 1 - \lambda g_n(\lambda)$ obey the properties

\begin{align}
\text{(i) } & g_n(\lambda) \leq \sigma_n, \\
\text{(ii) } & \lambda g_n(\lambda) \leq 1, \\
\text{(iii) } & \lambda r_n(\lambda) \leq \sigma_n^{-1}, \\
\text{(iv) } & r_n(\lambda) \leq \sigma_n^{-1} g_n(\lambda).
\end{align}

\tag{2.2}

**Proof.** The proof of representation (2.1) is standard. For the proof of (i) we follow the paper \cite{6} and observe that the function $r_n(\lambda) = \prod_{k=1}^{n} \frac{1}{\alpha_k/(\lambda + \alpha_k)}$ is monotonically decreasing and convex with $r_n(0) = 1$. From these properties we conclude that $r_n(\lambda) \geq r_n(0) + r'_n(0)\lambda$. Since $r_n(0) = 1$ and $r'_n(0) = -\sigma_n$ we obtain $r_n(\lambda) \geq 1 - \lambda \sigma_n$, which is equivalent to (i). The proof of (ii) follows from the representation $\lambda g_n(\lambda) = 1 - \prod_{k=1}^{n} \frac{1}{\alpha_k/(\lambda + \alpha_k)}$. For the proof of (iv), we multiply (iv) by $\lambda \sigma_n / r_n(\lambda)$ and obtain the equivalent inequality

$$\lambda \sum_{k=1}^{n} \frac{1}{\alpha_k} \leq \prod_{k=1}^{n} \left(1 + \frac{\lambda}{\alpha_k}\right) - 1.$$ 

This inequality, however, always holds true since the left-hand side is the first order term of the polynomial in $\lambda$ on the right-hand side. For the proof of (iii) we use (iv) and (ii) and obtain $r_n(\lambda) \leq \sigma_n^{-1} g_n(\lambda) \leq \sigma_n^{-1} / \lambda$, which is equivalent to (iii). \hfill \Box

**Remark 2.2.** Note that in our forthcoming analysis we will also exploit the fact that $r_n(\lambda) \leq 1$ for $\lambda \in (0, c]$ which is a consequence of the nonnegativity of $g_n(\lambda)$. Further note that property (iii) of the above proposition states that the regularization method (2.1) has at least a qualification of $p_0 = 1$. For the concept of qualification, see [38]. Finally we note that our analysis does not require the full strength of properties (i)–(iii) of proposition 2.1. Indeed, property (i) will be exploited for the $\lambda$-range $\lambda \leq \sigma_n^{-1}$, and properties (ii) and (iii) will be exploited for the $\lambda$-range $\lambda \geq \sigma_n^{-1}$.

For deriving order optimal error bounds for $\|x_n^\delta - x^\dagger\|$ with $x_n^\delta$ defined by (2.1) we introduce the regularized solution $x_n$ with exact data by

$$x_n - x_0 = G^t g_n(T^*T)T^*(y - Ax_0).$$

It can easily be checked that the error representations

$$x_n^\delta - x_0 = G^t g_n(T^*T)T^*(y^\delta - y) \quad \text{and} \quad x^\dagger - x_n = G^t r_n(T^*T)G^{-t}(x^\dagger - x_0) \tag{2.3}$$

are valid. From these error representations we see that in the $\|Ax\|$-norm and in the $X_p$-norm $\|x\|_p := \|G^{-p}x\|$ we have

\begin{align}
\|Ax_n^\delta - Ax_n\| &= \|T g_n(T^*T)T^*(y^\delta - y)\|, \tag{2.4} \\
\|Ax^\dagger - Ax_n\| &= \|T r_n(T^*T)G^{-t}(x^\dagger - x_0)\|, \tag{2.5} \\
\|x_n^\delta - x_n\|_p &= \|G^{-p} g_n(T^*T)T^*(y^\delta - y)\|, \tag{2.6} \\
\|x^\dagger - x_n\|_p &= \|G^{-p} r_n(T^*T)G^{-t}(x^\dagger - x_0)\|. \tag{2.7}
\end{align}
2.2. Smoothness assumptions

We formulate our smoothness assumptions in terms of some densely defined unbounded self-adjoint strictly positive operator $B : X \to X$. We introduce a Hilbert scale $(X_r)_{r \in \mathbb{R}}$ induced by the operator $B$ which is the completion of $\bigcap_{k \in \mathbb{N}} D(B^k)$ with respect to the Hilbert space norm $\|x\|_r = \|B^r x\|$, $r \in \mathbb{R}$. For technical reasons, instead of $B$ we will work with the inverse $G := B^{-1}$, which is a bounded linear injective and self-adjoint operator with non-closed range $\mathcal{R}(G)$. Note that the above Hilbert space norm $\| \cdot \|_r$ may be represented by $$\|x\|_r = \|B^r x\| = \|G^{-r} x\|, \quad r \in \mathbb{R}.$$ In addition, according to [9, 19] we call a function $\varphi : (0, a] \to (0, b]$ an index function if it is continuous and strictly increasing with $\varphi(0+) = 0$ and assume

**Assumption A1.** Constants $M \geq m > 0$ and some index function $\varphi : (0, a] \to (0, b]$ with $a = \|G\|$ and $b = \varphi(a)$ exist such that

(i) $m \|\varphi(G)x\| \leq \|Ax\|$ for all $x \in X$,

(ii) $\|Ax\| \leq M \|\varphi(G)x\|$ for all $x \in X$.

**Assumption A2.** For some positive constants $E$ and $p$ we assume the solution smoothness $x^\dagger - x_0 = G\psi v$ with $v \in X$ and $\|v\| \leq E$. That is, $$x^\dagger \in M_{p,E} = \{x \in X : \|x - x_0\|_p \leq E\}.$$ 

Assumption A1 characterizes the smoothing properties of the operator $A$ relative to the operator $G$ and allows the study of problems with finitely and infinitely smoothing operators $A$ in a unique manner. Typical index functions in applications are power-type index functions $\varphi(t) = t^a$ for problems (1.1) with finitely smoothing operators $A$ and index functions $\varphi(t) = \exp(t^{-a})$ where the inverse $\varphi^{-1}$ is of logarithmic type for problems with infinitely smoothing operators $A$. Such problems appear, e.g., in inverse heat conduction. Assumption A2 characterizes the smoothness of the element $x^\dagger - x_0$ in the Hilbert scale $(X_r)_{r \in \mathbb{R}}$. By using assumption A2 we can study different smoothness situations for $x^\dagger - x_0$.

Let us give some comments on order optimal convergence rates for identifying $x^\dagger$ from noisy data $y^\delta \in Y$ under the link assumption A1 and the smoothness assumption A2. Let $\mathcal{R} : Y \to X$ be an arbitrary method and $\mathcal{R}y^\delta$ be an approximate solution for $x^\dagger$. Then, the quantity $$\Delta(\delta, \mathcal{R}) = \sup\{\|\mathcal{R}y^\delta - x^\dagger\| : y^\delta \in Y, \|y - y^\delta\| \leq \delta, x^\dagger \in M_{p,E}\}$$ is called the worst case error of the method $\mathcal{R}$ on the set $M_{p,E}$. An optimal method $\mathcal{R}_{\text{opt}}$ is characterized by $\Delta(\delta, \mathcal{R}_{\text{opt}}) = \inf_{\mathcal{R}} \Delta(\delta, \mathcal{R})$, and this quantity is called the best possible worst case error on the set $M_{p,E}$. Under assumption A1 (ii) the best possible worst case error can be estimated by

$$\inf_{\mathcal{R}} \Delta(\delta, \mathcal{R}) \geq E \left[\psi_p^{-1}\left(\frac{\delta}{ME}\right)\right]^p \text{ with } \psi_p(t) := t^p \varphi(t)$$

provided $\delta/(ME)$ is an element of the spectrum of the operator $\psi_p(G)$, see [36] and [24, proof of theorem 2.2]. This lower bound will serve us as benchmark for the best possible accuracy for identifying $x^\dagger$ from noisy data $y^\delta \in Y$ under the link assumption A1 and the smoothness assumption A2.
2.3. Exploiting operator monotonicity

Operator monotonicity has been applied before in different papers for deriving order optimal error bounds in regularization under general smoothness assumptions, see, e.g., \[2, 21, 26\]. In this section we are going to derive some consequences of assumption A1 by using operator monotonicity of certain functions. Considering the error representations (2.5)–(2.7) we see that two types of estimates are helpful for deriving error bounds for the noise amplification error and the regularization error in the \(\|Ax\|\)-norm and the \(\|x\|_p\)-norm, namely estimates of the type

\[
\|f_1(T^*T)G^{-r}x\| \leq \|G^{-p}x\| \quad \text{and} \quad \|G^{s-p}x\| \leq \|f_2(T^*T)x\|
\]

with certain functions \(f_1, f_2\) and some constant \(p > 0\) from assumption A2. We will derive such estimates from assumption A1 by using the concept of operator monotone functions which is based on the concept of semiordering. Note that for two nonnegative, self-adjoint bounded linear operators \(S_1, S_2 \in \mathcal{L}(X)\) the semiordering \(S_1 \leq S_2\) is defined by \((S_1x, x) \leq (S_2x, x)\) for all \(x \in X\), or equivalently, by \(\|S_1^{1/2}x\| \leq \|S_2^{1/2}x\|\) for all \(x \in X\).

**Definition 2.3.** An index function \(f : (0, a] \to \mathbb{R}\) is called operator monotone if and only if for any pair of self-adjoint linear operators \(S_1, S_2\) with spectra in \((0, a]\), the relation \(S_1 \leq S_2\) implies the relation \(f(S_1) \leq f(S_2)\).

Properties and examples for operator monotone functions may be found in \[1, 20, 37\].

Our further study is based on several functions. The first function is

\[
\psi_r(\lambda) = \lambda^r \varphi(\lambda), \quad \psi_r : (0, a] \to (0, a^r \varphi(a))
\]

with \(\varphi\) from assumption A1, \(a = \|G\|\) and arbitrary constant \(r\) for which \(\lambda^r \varphi(\lambda)\) is monotonically increasing. Two other functions \(h\) and \(w\) are

\[
h(t) = \left[\psi_s^{-1}(\sqrt{t})\right]^{p-s}, \quad w(t) = 1/h(t)
\]

with constant \(s\) from (2.1) and \(p\) from assumption A2.

**Remark 2.4.** The function \(h\) defined by (2.9) possesses the following properties.

1. Due to the identity \(\sqrt{th}(t) = \psi_p((\psi_s^{-1})'(t))\), the function \(t \to \sqrt{th}(t)\) is an index function and hence monotonically increasing in both cases \(s \geq p\) and \(s \leq p\).
2. In the case of high order regularization with \(s \geq p\), the function \(h\) is non-increasing. Hence, for \(s \geq p\), the function \(t \to h(t)/\sqrt{t}\) is always monotonically decreasing.

In our next proposition we derive some estimates by using assumption A1.

**Proposition 2.5.** Let \(h\) and \(w\) be defined by (2.9). Then,

\[
\|w(T^*T/M^2)G^{-r}x\| \leq \|G^{-p}x\| \quad \text{under A1(ii) and} \quad w^2 \text{ operator monotone,}
\]

\[
\|w(T^*T/m^2)G^{-r}x\| \leq \|G^{-p}x\| \quad \text{under A1(i) and} \quad h^2 \text{ operator monotone,}
\]

\[
\|G^{s-p}x\| \leq \|w(T^*T/m^2)x\| \quad \text{under A1(i) and} \quad w^2 \text{ operator monotone,}
\]

\[
\|G^{s-p}x\| \leq \|w(T^*T/M^2)x\| \quad \text{under A1(ii) and} \quad h^2 \text{ operator monotone.}
\]

**Proof.** First, let us prove (2.10). It follows from assumption A1(ii) that

\[
\|AG^sx\| = \|T_x\| = \|(T^*T)^{1/2}x\| \leq M\|\varphi(G)G^sx\|.
\]
which be written in the equivalent form $T^*T/M^2 \leq \varrho^2(G)G^{2r}$. By using the function $\psi_2$ defined by (2.8), this estimate can be written as $T^*T/M^2 \leq \varrho^2_2(G)$. Since $w^3 := 1/h^2$ is assumed to be operator monotone and since $w^3(\varrho^2_2(G)) = G^{2r-2p}$ we obtain that $w^2(T^*T/M^2) \leq G^{2r-2p}$ which gives $\|w(T^*T/M^2)x\| \leq \|G^{r-r/2}x\|$ and hence (2.10). Second, let us prove (2.11). The link condition A1(i) may be written as $\varrho^2_2(G) \leq T^*T/m^2$. Since $h^2$ is assumed to be operator monotone and since $h^2(\varrho^2_2(G)) = G^{2r-2p}$ we obtain that $G^{2p-2s} \leq h^2(T^*T/m^2)$. Since $t \to -1/t$ is operator monotone, it follows that $w^2(T^*T/m^2) \leq G^{2r-2p}$, which gives (2.11). The proof of estimates (2.12) and (2.13) is similar.

Example 2.6 (The finitely smoothing case). Let us assume that operators $A^*A$ and $G$ are related by $A^*A = G^{2r}$ where $a$ is some positive constant. In this case both assumptions A1(i) and A1(ii) hold true as equality with $\varrho(\lambda) = \lambda^a$, $m = 1$ and $M = 1$. We easily see that the function $\varrho$ is an index function and that the function $\varrho(x)$ defined in (2.8) attains the form $\varrho_2(x) = \lambda^x$. Since $\varrho^{-1}(\sqrt{t}) = t^{1/(2r+a)}$ we obtain that the functions $h$ and $w$ defined in (2.9) possess the representations

$$h(t) = t^{\frac{a}{2r+a}}, \quad w(t) = t^{\frac{a}{2r}}.$$

Power functions $t^s$ are operator monotone for $0 \leq s \leq 1$, see [20]. Hence, under the natural side conditions $p \geq 0$, $a > 0$ and $s > -a$, we obtain

(i) that $w^2$ is an operator monotone function for $s \geq p$,

(ii) that $h^2$ is an operator monotone function for $s \leq p \leq 2s + a$.

2.4. Interpolation in variable Hilbert scales

By interpolation in variable Hilbert scales we can estimate the intermediate norm $\|x\|$ if estimates for some weaker norm $\|\varrho(G)x\|$ and some stronger norm $\|x\|_r$ are known. Variable Hilbert scale inequalities have been introduced by Hegland, see [8, 9]. Such inequalities which extend the classical interpolation inequality became a powerful tool in the analysis of regularity under general smoothness conditions, see, e.g., [18, 19, 24, 25, 31, 36]. In our paper we aim to combine special variable Hilbert scale inequalities with tools from operator monotonicity.

Proposition 2.7. Assume $r \geq 0$, $\|x\|_r \leq c_1$ and $\|\varrho(G)x\| \leq c_2$ with some index function $\varrho$ and constants $c_1, c_2$. Let $\xi_r(t) := \varrho^{-1}(t^{1/2r})$ be convex where $\varrho_r$ is given by (2.8). Then,

$$\|x\| \leq c_1 \left[ \varrho^{-1} \left( \frac{c_2}{c_1} \right) \right]^{1/2r}.$$  \hspace{1cm} (2.14)

Proof. Let $E_r$ be the spectral family of $G^{-2r}$. Since $\varrho_r$ is convex we may employ Jensen’s inequality and obtain

$$\xi_r \left( \frac{\|x\|}{\|x\|_r} \right) \leq \int \xi_r (\lambda^{-1} \lambda d \|E_r x\|) = \frac{\|\varrho(G)x\|}{\|x\|_r}$$

or, equivalently, $\|x\| \cdot \varrho_r \left( \|x\|^{1/r}/\|x\|_r^{1/r} \right) \leq \|\varrho(G)x\|$. Since $\varrho(t) = t^{r/2} \psi_r(t)$ is increasing we obtain that $t \to t^{r/2} \psi_r(t)$ is decreasing. Hence,

$c_1 \cdot \varrho_r \left( \frac{\|x\|^{1/r}}{c_1^{1/r}} \right) \leq \|x\|_r \cdot \varrho_r \left( \frac{\|x\|^{1/r}}{\|x\|_r^{1/r}} \right) \leq \|\varrho(G)x\| \leq c_2$.

Rearranging the terms gives (2.14).
In our next proposition we provide a further estimate which is based on interpolation arguments.

**Proposition 2.8.** Let \( x_n \) be the regularized solution (2.1) with exact data \( y \), let assumption A2 hold and assume \( \psi \) be defined by (2.8) and \( f(t) := \psi_p^2(t^{(1/2p−2s)} \) be convex. Then, for all regularization parameters \( \sigma_n \),

\[
\|x_n - x^*\|_{2s−p} \leq E \left[ \psi_p^{-1} \left( \frac{\|\psi(G)(x_n - x^*)\|}{E} \right) \right]^{2p−2s}.
\] (2.15)

**Proof.** Let us introduce the abbreviation \( z = x^* - x_n \). From (2.3) we have the identity \( G^{-1}z = r_n(T^*T)G^{-1}(x^* - x_0) \), and due to \( r_n(\lambda) \leq 1 \) we have the estimate \( \|r_n^{1/2}(T^*T)\| \leq 1 \). We use these properties and obtain due to the Cauchy–Schwarz inequality and assumption A2 that

\[
\|z\|^2 = \|r_n(T^*T)G^{-1}(x^* - x_0)\|^2 \\
\leq \|r_n(T^*T)G^{-1}(x^* - x_0)\|^2 \\
= (G^{p−2s}z, G^{-p}(x^* - x_0)) \\
\leq E\|z\|_{2s−p}.
\] (2.16)

In the case \( s = p \), (2.15) follows from (2.16). In the case \( 0 \leq s < p \), our next aim consists in deriving a second estimate that relates the intermediate norm \( \|z\|_{2s−p} \) with the weaker norm \( \|\psi(G)z\| \) and the stronger norm \( \|z\| \). We derive this estimate by interpolation in variable Hilbert scales. Since \( f \) is convex we may employ Jensen’s inequality and have

\[
f \left( \frac{\|z\|_{2s−p}}{\|z\|} \right) = f \left( \frac{\int \lambda^{s−p} \cdot \lambda^s d\|E_\lambda z\|^2}{\int \lambda^s d\|E_\lambda z\|^2} \right) \leq \int \frac{f(\lambda^{s−p})\lambda^s d\|E_\lambda z\|^2}{\int \lambda^s d\|E_\lambda z\|^2}
\]

where \( E_\lambda \) is the spectral family of \( G^{-2} \). Since \( f(\lambda^{s−p})\lambda^s = \sigma^2(\lambda^{−1/2}) \) we obtain

\[
f \left( \frac{\|z\|_{2s−p}}{\|z\|} \right) \leq \int \frac{\sigma^2(\lambda^{−1/2}) d\|E_\lambda z\|^2}{\|\|} = \frac{\|\psi(G)z\|^2}{\|z\|^2}.
\] (2.17)

Now, let us eliminate \( \|z\| \) in estimate (2.17). We write estimate (2.16) in the equivalent form

\[
\|z\|_{2s−p} / E^{1/2} \leq \|z\|_{2s−p} / \|z\|_s
\] (2.18)

and introduce two auxiliary functions \( g \) and \( r \) by

\[
g(t) := t^{−2} f(t^{1/2}) \quad \text{and} \quad r(t) := t f(t) = \psi_p^2(r^{(1/2p−2s)}).
\]

Since \( f \) is convex and \( f(0) = 0 \), \( g \) is monotonically increasing. Hence, by (2.18), the monotonicity of \( g \) and (2.17),

\[
g \left( \frac{\|z\|_{2s−p}^{1/2}}{E^{1/2}} \right) \leq g \left( \frac{\|z\|_{2s−p}}{\|z\|} \right) = \frac{\|z\|_p^2}{\|z\|_{2s−p}} f \left( \frac{\|z\|_{2s−p}}{\|z\|} \right) \leq \frac{\|\psi(G)z\|^2}{\|z\|_{2s−p}^2}.
\]

Multiplying by \( \|z\|_{2s−p}^2 / E^2 \) gives

\[
r \left( \frac{\|z\|_{2s−p}}{E} \right) \leq \frac{\|\psi(G)z\|^2}{E^2}.
\]

Since the inverse \( r^{-1} \) has the form \( r^{-1}(\lambda) = \left[ \psi_p^{-1}(\sqrt{\lambda}) \right]^{2p−2s} \), we obtain (2.15). \( \Box \)
3. A priori parameter choice

In this section we make use of proposition 2.5 for estimating the total error in different norms in case the regularization parameter \( \sigma_n \) from (1.3) is chosen a priori by

\[
\sigma_n^{-1} = \frac{\delta^2}{E^2} \left[ \frac{\psi^{-1}_p}{mE} \right]^{2(s-p)}.
\]  

We note that in the finitely smoothing case of example 2.6, the a priori parameter choice (3.1) attains the form

\[
\sigma_n^{-1} = m^2 \left( \frac{1}{mE} \right)^{2(s-p)/(a+p)}.
\]

3.1. Error bounds in the \( \|Ax\| \)-norm

We start by providing error bounds for arbitrary \( \sigma_n > 0 \).

**Proposition 3.1.** Let \( x^\delta_n \) be defined by (2.1), \( h \) and \( w \) be defined by (2.9) and assume the solution smoothness A2.

(i) High order regularization (\( s \geq p \)): if \( w^2 := 1/h^2 \) is operator monotone, then under the link condition A1(ii),

\[
\|Ax^\delta_n - Ax^\dag\| \leq \delta + E \sqrt{\sigma_n^{-1} h(\sigma_n^{-1}/M^2)}.
\]  

(ii) Low order regularization (\( s \leq p \)): if \( h^2 \) is operator monotone and if \( h(t)/\sqrt{t} \) is decreasing, then under the link condition A1(i),

\[
\|Ax^\delta_n - Ax^\dag\| \leq \delta + E \sqrt{\sigma_n^{-1} h(\sigma_n^{-1}/m^2)}.
\]

**Proof.** For estimating \( \|Ax^\delta_n - Ax^\dag\| \) in the high order case \( s \geq p \) we use the error representation (2.4) and obtain due to \( \lambda g_n(\lambda) \leq 1 \), see (2.2) (ii), the estimate

\[
\|Ax^\delta_n - Ax^\dag\| = \|Tg_n(T^*T)T^*(y^\delta - y)\| \leq \delta \sup_\lambda |\lambda g_n(\lambda)| \leq \delta.
\]  

For estimating \( \|Ax_n - Ax^\dag\| \) in the high order case \( s \geq p \) we use the error representation (2.5), exploit assumption A2 and estimate (2.10) which requires operator monotonicity of \( w^2 \) and the second link condition A1(ii) and obtain

\[
\|Ax_n - Ax^\dag\| = \|Tr_n(T^*T)h(T^*T/M^2)\| \cdot E.
\]  

For estimating the norm term in (3.5) we distinguish two cases \( \lambda \leq \sigma_n^{-1} \) and \( \lambda \geq \sigma_n^{-1} \). In the first case \( \lambda = \sigma_n^{-1} \) we use \( g_n(\lambda) \geq 0 \) or equivalently \( r_n(\lambda) \leq 1 \), exploit the increasing behavior of \( \sqrt{\lambda}h(\lambda/M^2) \) that holds true due to remark 2.4 (i) and obtain

\[
\sqrt{\lambda}r_n(\lambda)h(\lambda/M^2) \leq \sqrt{\lambda}h(\lambda/M^2) \leq \sqrt{\sigma_n^{-1} h(\sigma_n^{-1}/M^2)}.
\]

In the second case \( \lambda = \sigma_n^{-1} \) we use \( \lambda r_n(\lambda) = \sigma_n^{-1} \), see (2.2) (iii), exploit the decreasing behavior of \( h(t)/\sqrt{t} \) that holds true due to remark 2.4 (ii) and obtain

\[
\sqrt{\lambda}r_n(\lambda)h(\lambda/M^2) \leq \sigma_n^{-1} h(\lambda/M^2)/\sqrt{\lambda} \leq \sqrt{\sigma_n^{-1} h(\sigma_n^{-1}/M^2)}.
\]

From both cases we obtain that (3.5) attains the form

\[
\|Ax_n - Ax^\dag\| \leq E \sqrt{\sigma_n^{-1} h(\sigma_n^{-1}/M^2)}.
\]

From this estimate and (3.4) we obtain (3.2). For the proof of part (ii) we proceed analogously by exploiting (2.11) instead of (2.10). \( \square \)
Remark 3.2. Let us discuss the monotonicity condition in part (ii) of proposition 3.1 for the finitely smoothing case of example 2.6. For this example we have
\[ h(t)/\sqrt{t} = t^{e^{-(s+a)}}. \]
Hence, \( h(t)/\sqrt{t} \) is decreasing for \( p \leq 2s + a \). This coincides with Natterer’s side condition in (1.4).

Corollary 3.3. Let the assumptions of proposition 3.1 be satisfied and let \( \sigma_n \) be chosen by (3.1). Then, in both cases (i) and (ii) of proposition 3.1 we have
\[
\begin{align*}
\text{(i) } & \|Ax_n^d - Ax^\dagger\| \leq (1 + M/m)\delta, \\
\text{(ii) } & \|Ax_n^d - Ax^\dagger\| \leq 2\delta.
\end{align*}
\]

Proof. Let us prove estimate (3.6) for the high order case (i). The a priori parameter choice (3.1) can be written in the equivalent form
\[ E\sqrt{\sigma_n^{-1}} = \delta w(\sigma_n^{-1}/m^2). \]
Hence, by (3.2) and (3.8),
\[ \|Ax_n^d - Ax^\dagger\| \leq \delta + \delta h(\sigma_n^{-1}/M^2)w(\sigma_n^{-1}/m^2). \]

Since \( \sigma_n^{-1}/m^2 \leq \sigma_n^{-1}/m^2 \) and since \( \sqrt{t}h(t) \) is increasing we have
\[ h(\sigma_n^{-1}/M^2) \leq \frac{M}{m}h(\sigma_n^{-1}/m^2). \]
From this estimate and (3.9) we obtain (3.6). The proof for estimate (3.7) for the low order case (ii) is analogous. \( \square \)

3.2. Error bounds in \( X_p \)

We start by providing error bounds with respect to the \( \|\cdot\|_p \) norm.

Proposition 3.4. Let \( x_n^d \) be defined by (2.1), \( h \) and \( w \) be defined by (2.9) and assume the link condition A1 and the solution smoothness A2.
\[
\begin{align*}
\text{(i) High order regularization (} s \geq p \text{): if } w^2 := 1/h^2 \text{ is operator monotone,} \\
\|x_n^d - x\|_p \leq \delta \sqrt{\sigma_n}w(\sigma_n^{-1}/m^2) + E \cdot M/m. \tag{3.10}
\end{align*}
\]
\[
\begin{align*}
\text{(ii) Low order regularization (} s \leq p \text{): if } h^2 \text{ is operator monotone and if } h(t)/\sqrt{t} \text{ is decreasing,} \\
\|x_n^d - x\|_p \leq \delta \sqrt{\sigma_n}w(\sigma_n^{-1}/M^2) + E \cdot M/m. \tag{3.11}
\end{align*}
\]

Proof. Let us consider the high order case (i). For estimating \( \|x_n^d - x_n\|_p \) we use the error representation (2.6) and obtain from estimate (2.12) of proposition 2.5
\[ \|x_n^d - x_n\|_p = \|G^{\dagger}g_n(T^*T)y^d - y\| \leq \delta \left\| w \left( \frac{1}{m^2}T^*T \right) g_n(T^*T)T^* \right\|. \tag{3.12} \]

For estimating the norm term in (3.12) we distinguish two cases \( \lambda \leq \sigma_n^{-1} \) and \( \lambda \geq \sigma_n^{-1} \). In the first case \( \lambda \leq \sigma_n^{-1} \) we use \( g_n(\lambda) \leq \sigma_n \), see (2.2) (i), exploit that \( \sqrt{\lambda}w(\lambda) \) is increasing (see remark 2.4 (ii)) and obtain
\[ w(\lambda/m^2)g_n(\lambda)\sqrt{\lambda} \leq w(\lambda/m^2)\sqrt{\lambda}\sigma_n \leq w(\sigma_n^{-1}/m^2)\sqrt{\sigma_n}. \]
In the second case $\lambda \geq \sigma_n^{-1}$ we use $\lambda g_n(\lambda) \leq 1$, exploit that $w(t)/\sqrt{t}$ is decreasing (see remark 2.4 (i)) and obtain

$$w(\lambda/m^2)g_n(\lambda)\sqrt{\lambda} \leq w(\lambda/m^2)/\sqrt{\lambda} \leq w(\sigma_n^{-1}/m^2)\sqrt{\sigma_n}.$$ 

From both cases we obtain that (3.12) attains the form

$$\|x_n^4 - x_n\|_p \leq \delta \sqrt{\sigma_n}w(\sigma_n^{-1}/m^2).$$ 

(3.13)

For estimating $\|x_n - x^1\|_p$ in the high order case (i) we use the error representation (2.7), exploit the estimate (2.10) of proposition 2.5, use in addition (2.12) and obtain due to $\|x^1 - x_0\|_p \leq E$ and $g_n(\lambda) \geq 0$, or equivalently $r_n(\lambda) \leq 1$, the estimate

$$\|x^1 - x_n\| = \|G^{\gamma-P}r_n(T^*T)G^{-\gamma}(x^1 - x_0)\| \leq \|w(T^*T/m^2)r_n(T^*T)G^{-\gamma}(x^1 - x_0)\| \leq E\|w(T^*T/m^2)r_n(T^*T)h(T^*T/M^2)\| \leq E \sup_{\lambda} |h(\lambda/M^2)w(\lambda/m^2)|.$$ 

(3.14)

Due to remark 2.4 (i) the function $\sqrt{\gamma h(t)}$ is increasing or, equivalently, $w(t)/\sqrt{t}$ is decreasing. Hence, from $\lambda/M^2 \leq \lambda/m^2$ we have $w(\lambda/m^2) \leq \frac{M}{m}w(\lambda/M^2)$, and (3.14) attains the form

$$\|x^1 - x_n\|_p \leq E \cdot M/m.$$ 

(3.15)

From (3.13) and (3.15) we obtain (3.10). For the proof of part (ii) we proceed analogously by exploiting (2.11) and (2.13) instead of (2.10) and (2.12).

From proposition 3.4 we have along the line of corollary 3.3 the following.

**Corollary 3.5.** Let the assumptions of proposition 3.4 be satisfied and let $\sigma_n$ be chosen by (3.1). Then, in both cases (i) and (ii) of proposition 3.4 we have

(i) $\|x_n^4 - x^1\|_p \leq E \cdot (1 + M/m),$ 

(ii) $\|x_n^4 - x^1\|_p \leq 2E \cdot M/m.$

### 3.3. Error bounds in X

For deriving order optimal error bounds for the total error $\|x_n^4 - x^1\|$ with $\sigma_n$ chosen *a priori* by (3.1) we employ interpolation techniques from proposition 2.7 and use the results of corollary 3.5 which provides a bound for $\|x_n^4 - x^1\|_p$ and the results of corollary 3.3 which together with the first link condition A1(i) provides a bound for $\|\rho(G)(x_n^4 - x^1)\|$. 

**Theorem 3.6.** Let the assumptions of proposition 3.1 and 3.4 be satisfied and let $\sigma_n$ be chosen *a priori* by (3.1). If the function $\xi_p(t) := \psi_p^2(t^{1/2p})$ is convex, then $x_n^4$ is order optimal on the set $M_{p,E}$ in both cases of high order regularization $s \geq p$ and low order regularization $s \leq p$. In fact, in both cases,

$$\|x_n^4 - x^1\| \leq c_1 \left[\psi_p^{-1}(c_2\delta)\right]^p$$ 

(3.16)

with some constants $c_1$ and $c_2$ which can be extracted from the proof.

**Proof.** Due to corollary 3.5, corollary 3.3 and assumption A1(i), in both cases of high- and low order regularization we have

$$\|x_n^4 - x^1\|_p \leq k_1$$ 

and

$$\|\rho(G)(x_n^4 - x^1)\| \leq k_2\delta$$

with some constants $k_1$ and $k_2$. Using the interpolation estimate (2.14) of proposition 2.7 yields (3.16).

Note that for the finitely smoothing case of example 2.3 we have $\xi_p(t) = t^{(a+p)/p}$, which is convex for arbitrary $p > 0.$
3.4. Revisiting the low order case

The error bounds given in subsection 3.3 require in both cases of high order and low order regularization both link conditions A1(i) and A1(ii). We will show in this subsection that in the case of low order regularization \( s \leq p \) order optimal error bounds can be obtained without the second link condition A1(ii). However, this will only be possible for \( s \geq 0 \). We exploit in our study the property

\[
\sqrt{\lambda}g_n(\lambda) \leq \sqrt{s_\lambda} \quad \text{for} \quad \lambda \in (0, \|T\|^2]
\]

which follows from both properties (2.2) (i) and (2.2) (ii) of proposition 2.1 and start by providing some error bound in the \( \| \cdot \|_s \)-norm for arbitrary \( s_n > 0 \).

**Proposition 3.7.** Let \( x_n^\delta \) be defined by (2.1), \( h \) be defined by (2.9) and assume the solution smoothness A2. If \( h^2 \) is operator monotone and \( h(t)/t \) is decreasing, then under the link condition A1(i),

\[
\| x_n^\delta - x^\delta \|_s \leq \delta \sqrt{s_\lambda} + Eh(\sigma_n^{-1}/m^2).
\]

**Proof.** For estimating \( \| x_n^\delta - x^\delta \|_s \) in the low order case \( s \leq p \) we use the error representation (2.3) and obtain due to \( \sqrt{\lambda}g_n(\lambda) \leq \sqrt{s_\lambda} \), see (3.17), the estimate

\[
\| x_n^\delta - x^\delta \|_s \leq \| g_n(T^*T)T^*(y^\delta - y) \| \leq \delta \sqrt{s_\lambda}.
\]

For estimating \( \| x_n^\delta - x^\delta \|_s \), we use the error representation (2.3), exploit the estimate (2.11) of proposition 2.5 and obtain due to \( \| x^\delta - x_0 \|_p \leq E \) the estimate

\[
\| x^\delta - x_0 \|_p = \| r_n(T^*T)G^{-1}(x^\delta - x_0) \| \leq E \| r_n(T^*T)h(T^*T/m^2) \|.
\]

For estimating the norm term in (3.20) we distinguish two cases \( \lambda \leq \sigma_n^{-1} \) and \( \lambda \geq \sigma_n^{-1} \). In the first case \( \lambda \leq \sigma_n^{-1} \) we use \( r_n(\lambda) \leq 1 \) or, equivalently, \( g_n(\lambda) \geq 0 \) exploit the increasing behavior of \( h(t) \) which is always satisfied since \( h^2 \) is operator monotone and obtain

\[
r_n(\lambda)h(\lambda/m^2) \leq h(\lambda/m^2) = h(\sigma_n^{-1}/m^2).
\]

In the second case \( \lambda \geq \sigma_n^{-1} \) we use \( \lambda r_n(\lambda) \leq \sigma_n^{-1} \), exploit the decreasing behavior of \( h(t)/t \) and obtain

\[
r_n(\lambda)h(\lambda/m^2) \leq \sigma_n^{-1}h(\lambda/m^2)/\lambda \leq h(\sigma_n^{-1}/m^2).
\]

From both cases we obtain that (3.20) attains the form \( \| x^\delta - x_0 \|_s \leq Eh(\sigma_n^{-1}/m^2) \). From this estimate and (3.19) we obtain (3.18).

Since the parameter choice (3.1) can be written in the equivalent form \( \delta \sqrt{s_\lambda} = Eh(\sigma_n^{-1}/m^2) \) we obtain from proposition 3.7 the following.

**Corollary 3.8.** Let the assumptions of proposition 3.7 be satisfied and let \( \sigma_n \) be chosen a priori by (3.1). Then,

\[
\| x_n^\delta - x^\delta \|_s \leq 2E \left[ \psi_p^{-1}\left( \frac{\delta}{mE} \right) \right]^{p-s}.
\]

**Proof.** From (2.8) we have \( \varrho(\lambda) = \lambda^{-p}\psi_p(\lambda) \) and \( \varrho(\lambda) = \lambda^{-p}\psi_p(\lambda) \). Consequently,

\[
\psi_p(\lambda) = \lambda^{p-s}\psi_p(\lambda) \quad \text{and} \quad \psi_p(\lambda) = \lambda^{p-s}\psi_p(\lambda).
\]

We use the first equation of (3.22), substitute \( \lambda = \psi_p^{-1}(t) \) and obtain \( \psi_p(\psi_p^{-1}(t)) = t[\psi_p^{-1}(t)]^{p-s} \). From this equation we conclude that the a priori parameter choice (3.1), which is
equivalent to $\sigma_n^{-1/2}/m = \frac{1}{mE} [\psi_p^{-1}(\frac{\delta}{mE})]^{1-p}$, can be rewritten as $\sigma_n^{-1/2}/m = \psi_s(\psi_p^{-1}(\frac{\delta}{mE}))$ or, equivalently,

$$\psi_s^{-1}\left(\frac{\sigma_n^{-1/2}}{m}\right) = \psi_p^{-1}\left(\frac{\delta}{mE}\right).$$  \hfill (3.23)

Clearly, (3.23) is equivalent to $\psi_p(\psi_s^{-1}(\sigma_n^{-1/2}/m)) = \frac{\delta}{mE}$. We use the second equation of (3.22) and write this equation in the form

$$\left[\psi_s^{-1}\left(\frac{\sigma_n^{-1/2}}{m}\right)\right]^{p-s} \frac{\sigma_n^{-1/2}}{m} = \frac{\delta}{mE}.$$

From this equation and the definition of $h$ by (2.9) we see that the parameter choice (3.1) can be written in the equivalent form

$$\delta \sqrt{\sigma n} = E h\left(\psi_s^{-1}(\sigma_n^{-1/2}/m)\right).$$

Due to this equation, estimate (3.18) of proposition 3.7 attains the form

$$\|x_n^d - x\|_8 \leq 2\delta \sqrt{\sigma n}.$$  \hfill (3.24)

From this estimate and the parameter choice (3.1), we obtain (3.21).

□

Remark 3.10. The \textit{a priori} parameter choice (3.1) requires to choose $n$ and the sequence $(\alpha_k)_n$ which can be done in different ways. Our convergence rate analysis is flexible enough to cover different regularization methods such as

(i) the method of ordinary Tikhonov regularization with differential operators, which is characterized by (1.2) with $n = 1$ and $\alpha_1 = \alpha$. In this method, the parameter choice (3.1) for $\alpha$ attains the form $\alpha = \frac{\delta}{E}\psi_p^{-1}(\frac{\delta}{mE})^{2(\alpha-p)}$.
(ii) the method of iterated Tikhonov regularization of order \( n \) with differential operators which is characterized by (1.2) with some fixed \( n > 1 \) and \( \alpha_k = \alpha \) for \( k = 1, \ldots, n \). In this method, the parameter choice (3.1) for \( \alpha \) attains the form \( \alpha = n \cdot \frac{E^2}{\varphi_p} \left( \psi_p^{-1} \left( \frac{1}{mE} \right) \right)^{\frac{1}{2(p-s)}} \);

(iii) the implicit iteration scheme with differential operators which is characterized by (1.2) with some fixed \( \alpha_k = \alpha \) for \( k = 1, \ldots, n \). In this method, the a priori parameter choice (3.1) for the iteration number \( n \) attains the form \( n = \left[ \alpha \cdot \frac{E^2}{\varphi_p} \left( \psi_p^{-1} \left( \frac{1}{mE} \right) \right)^{\frac{1}{2(p-s)}} \right] \) where \( [c] \) is the largest integer which is less or equal to \( c \). As a special case, Lardy’s method is characterized by \( \alpha = 1 \);

(iv) the nonstationary version of the implicit iteration scheme with differential operators which is characterized by (1.2) with a geometric sequence \((\alpha_k)_{k=1}^{\infty}, \alpha_k = \alpha q^{k-1}\) and some \( q \in (0, 1) \). In this method with \( q = \frac{1}{s} \), the parameter choice (3.1) for \( n \) attains the form \( n = \left[ \ln \left( 1 + \alpha \cdot \frac{E^2}{\varphi_p} \left( \psi_p^{-1} \left( \frac{1}{mE} \right) \right)^{\frac{1}{2(p-s)}} \right) / \ln 2 \right] \).

In the finitely smoothing case of example 2.6, the rate of convergence in the above methods (i)–(iv) with respect to the noise level \( \delta \) cannot exceed the best possible rate \( \| x^\delta - x^* \| = O(\delta^{1/(p+\alpha)}) \) with \( p = 2s + \alpha \) in general. This restriction is due to the assumption \( h^2 \) to be operator monotone which leads to the side condition \( p \leq 2s + \alpha \). In the case of commuting operators \( A^*A \) and \( G \), however, the assumption on operator monotonicity of \( h^2 \) is not required (compare [2]), and it can be shown that the best possible rate for Tikhonov’s method (i) is of order \( O(\delta^{1/(p+\alpha)}) \) with \( p = 2s + 2\alpha \), and that this rate may be improved in the iterated versions (ii)–(iv) in smoothness cases \( p > 2s + 2\alpha \). From the viewpoints of small computational costs and best possible convergence rates it therefore makes sense to choose \( n = 1 \) in smoothness situations \( p \leq 2s + 2\alpha \) and to increase \( n \) for increasing \( p \).

4. Discrepancy principle

If the constants \( m \) and \( E \) in the a priori parameter choice (3.1) are unknown, then the parameter choice \( \sigma_n^{-1} = \left( \frac{1}{\lambda} \right)^{\frac{1}{2}} \left[ \psi_p^{-1} \left( \frac{1}{mE} \right) \right]^{\frac{1}{2(p-s)}} \) may be used where \( c_1 \) and \( c_2 \) are positive constants guessing \( m \) and \( E \), respectively. For this parameter choice, the order optimality results of theorems 3.6 and 3.9 still hold true. In case of rough estimates for \( m \) and \( E \), and in particular in cases where \( p \) and \( \psi_p \) are unknown, a posteriori rules for choosing \( \sigma_n \) have to be used. In the discrepancy principle (see [23]) the regularization parameter \( \sigma_n \) is chosen as the solution of the nonlinear equation

\[
d(\sigma_n) := \| Ax^\delta_n - y^\delta \| = C\delta
\]

with a constant \( C \geq 1 \). For practical reasons it makes sense to choose \( \sigma_n \) such that

\[
C_1\delta \leq d(\sigma_n) \leq C_2\delta
\]

with some constants \( C_1, C_2 \) that obey \( 1 \leq C_1 \leq C_2 \). In computations it makes sense to choose \( C_2 \) with \( C_2 > C_1 \).

Remark 4.1. For realizing the discrepancy principle (4.1) or (4.2) approximately, one practical way is as follows. We start with some large \( \alpha_1 \) in (1.2), use a decreasing \( \alpha \)-sequence and iterate as long as the discrepancy is in the magnitude of the noise level. More accurately, we consider the decreasing sequence \( \Delta = [\alpha_k]_{k=1}^{\infty} \) and choose \( n \) as the first integer for which

\[
\| Ax^\delta_n - y^\delta \| \leq C\delta < \| Ax^\delta_k - y^\delta \|, \quad 0 \leq k < n
\]

with some \( C > 1 \). Some care is required for the final iteration step in which one has to take care that the discrepancy does not become too small and remains in the magnitude of \( \delta \). This can be guaranteed by assuming that the final \( \alpha_n \) is not too small and obeys

\[
\frac{1}{\alpha_n} \leq c\sigma_{n-1}
\]
with some positive constant \(c\). For the geometric sequence \(\Delta = \{q^{k-1}a_1\}_{k=1}^{\infty}\) with some \(q < 1\), assumption (4.4) is satisfied with \(c = 1/q\), see [5]. We show in subsection 4.3 that for version (4.3) of the discrepancy principle, analogous convergence rate results to that of the a posteriori rule (4.2) hold true.

4.1. Properties

Due to (2.1), the discrepancy \(y^\delta - Ax_n^\delta\) can be represented by

\[
y^\delta - Ax_n^\delta = r_n(TT^*)(y^\delta - Ax_0) = \left(\prod_{k=1}^{n} \alpha_k(TT^* + \alpha_k I)^{-1}\right) (y^\delta - Ax_0). \tag{4.5}
\]

From this representation we conclude that the discrepancy is monotonically decreasing with respect to the iteration number, that is,

\[
\|y^\delta - Ax_n^\delta\| < \|y^\delta - Ax_{k-1}^\delta\|, \quad k = 1, 2, \ldots.
\]

For \(\sigma_n \to \infty\) we have \(r_n(\lambda) \to 0\), and for \(\sigma_n \to 0\) we have \(r_n(\lambda) \to 1\). Therefore, by (4.5), we have the two limit relations

\[
\lim_{\sigma_n \to \infty} \|y^\delta - Ax_n^\delta\| = 0 \quad \text{and} \quad \lim_{\sigma_n \to 0} \|y^\delta - Ax_n^\delta\| = \|y^\delta - Ax_0\|.
\]

From both limit relations we conclude that under the condition \(\|y^\delta - Ax_0\| > C\delta \sigma_n\) (not necessarily unique) exists that obeys rule (4.1) or rule (4.3), respectively, and that under the condition \(\|y^\delta - Ax_0\| > C_2\delta\), \(\sigma_n\) exists that obeys rule (4.2).

Now we assume that for some given \(\sigma_n\) we have \(\|Ax_{n-1}^\delta - y^\delta\| > C\delta\).

(i) For \(\sigma_n \to 0\) we have the limit relation \(\lim_{\sigma_n \to 0} d(\alpha_n) = 0\).

(ii) For \(\sigma_n \to \infty\) we have \(\lim_{\sigma_n \to \infty} d(\alpha_n) = \|Ax_{n-1}^\delta - y^\delta\| > C\delta\).

(iii) The function \(d(\alpha_n)\) is continuous and strictly monotonically increasing.

As a consequence, \(\alpha_n^\ast\) exists with \(d(\alpha_n^\ast) = C\delta\).

The following proposition gives us some monotonicity property for the error \(\|x_n^\delta - x^\ast\|_s\) with respect to the \(X_s\)-norm which states that the iteration (1.2) should not be stopped as long as \(\|Ax_n^\delta - y^\delta\| \geq \delta\) holds. In the special case \(s = 0\), such monotonicity property may be found in [4].

**Proposition 4.2.** Let \(x^\ast \in X_s\), let \(x_n^\delta\) be defined by the iteration (1.2) and let \(\|Ax_n^\delta - y^\delta\| \geq \delta\). Then,

\[
\|x_n^\delta - x^\ast\|_s < \|x_{n-1}^\delta - x^\ast\|_s.
\]

**Proof.** The iteration (1.2) can be rewritten as

\[x_n^\delta = x_{n-1}^\delta + B^{-2s}A^*z_{n-1} \quad \text{with} \quad z_{n-1} = (TT^* + \alpha_n I)^{-1}(y^\delta - Ax_{n-1}^\delta).\]

Consequently, for \(d := \|x_n^\delta - x^\ast\|_s^2 - \|x_{n-1}^\delta - x^\ast\|_s^2\) we have

\[
d = \|x_{n-1}^\delta + B^{-2s}A^*z_{n-1} - x^\ast\|_s^2 - \|x_{n-1}^\delta - x^\ast\|_s^2
\]

\[
= (2x_{n-1}^\delta - 2x^\ast + B^{-2s}A^*z_{n-1}, B^{-2s}A^*z_{n-1})_s
\]

\[
= \left((x_{n-1}^\delta + x_n^\delta - 2x^\ast, B^{-2s}A^*z_{n-1})_s
\]

\[
= \left((Ax_{n-1}^\delta + Ax_n^\delta - 2y, z_{n-1})
\]

\[
= (2(y - y) + (Ax_{n-1}^\delta - y^\delta) + (Ax_n^\delta - y^\delta), z_{n-1})
\]

\[
\leq 2\|z_{n-1}\| \left(\delta - \frac{(y^\delta - Ax_{n-1}^\delta) + (y^\delta - Ax_n^\delta), z_{n-1})}{2\|z_{n-1}\|}\right).
\]
Let \( r_n := y^\delta - Ax_n^\delta \). Then, from (4.5) we have \( r_n = \alpha_n( TT^* + \alpha_n I)^{-1} r_{n-1} \). Hence, the element \( z_{n-1} \) can be written as \( z_{n-1} = \alpha_n^{-1} r_n \). Consequently,
\[
\| x_n^\delta - x^\|_s^2 - \| x_n^\delta - x^\|_s^2 \leq \frac{2\| r_n \|}{\alpha_n} \left( \delta - \frac{(r_{n-1} + r_n, r_n)}{2\| r_n \|} \right). \tag{4.6}
\]

We use again the identity \( r_n = \alpha_n( TT^* + \alpha_n I)^{-1} r_{n-1} \) or, equivalently, \( r_{n-1} = \alpha_n^{-1}( TT^* + \alpha_n I)r_n \), multiply by \( r_n \) and obtain
\[
(r_{n-1}, r_n) = \alpha_n^{-1}\| T^* r_n \|^2 + \| r_n \|^2 > \| r_n \|^2. \tag{4.7}
\]

From (4.6), (4.7) and \( \| r_n \| \geq \delta \) we obtain \( \| x_n^\delta - x^\|_s^2 - \| x_n^\delta - x^\|_s^2 < 0. \) \( \square \)

4.2. Error bounds in \( X \)

In this subsection we show that for \( \sigma_n \) chosen by (4.1) or (4.2), respectively, the order optimal error bound (3.16) holds true under analogous assumptions of theorem 3.6. In a first proposition we provide some estimates for the regularization parameter \( \sigma_n \) chosen by (4.2).

**Proposition 4.3.** Let \( x_n^\delta \) be defined by (2.1), \( h \) and \( w \) be defined by (2.9), \( \sigma_n \) be chosen by the discrepancy principle (4.2) with \( 1 < C_1 \leq C_2 \) and assume the solution smoothness A2.

(i) High order regularization (\( s \geq p \)): if \( w^2 := 1/h^2 \) is operator monotone, then under the link condition A1(ii),
\[
(C_1 - 1)\delta \leq E\sqrt{\sigma_n^{-1} h(\sigma_n^{-1}/M^2)}. \tag{4.8}
\]

(ii) Low order regularization (\( s \leq p \)): if \( h^2 \) is operator monotone and if \( h(t)/\sqrt{t} \) is decreasing, then under the link condition A1(i),
\[
(C_1 - 1)\delta \leq E\sqrt{\sigma_n^{-1} h(\sigma_n^{-1}/m^2)}. \tag{4.9}
\]

**Proof.** Let us prove part (i). From (2.1) we have \( y^\delta - Ax_n^\delta = r_n( TT^*) (y^\delta - Ax_0) \). Due to rule (4.2), the identity \( y - Ax_n = r_n( TT^*) (y - Ax_0) \) and the estimate \( \| r_n( TT^*) \| \leq 1 \) we obtain that
\[
C_1 \delta \leq \| r_n( TT^*) (y^\delta - Ax_0) \|
\leq \| r_n( TT^*) (y - Ax_0) \| + \| r_n( TT^*) (y - y^\delta) \|
\leq \| y - Ax_n \| + \delta. \tag{4.10}
\]

From the proof of proposition 3.1 we have that
\[
\| y - Ax_n \| \leq E\sqrt{\sigma_n^{-1} h(\sigma_n^{-1}/M^2)}.
\]

This estimate and (4.10) provide the desired estimate (4.8) and the proof of part (i) is complete. For the proof of part (ii) we proceed in an analogous way but use instead of (2.10) the estimate (2.11) which requires the link condition A1(i) and the operator monotonicity of the function \( h^2 \). \( \square \)

From propositions 3.4 and 4.3 we obtain that the total error \( x_n^\delta - x^\) is bounded in the \( \| \cdot \|_s \)-norm for the \textit{a posteriori} parameter choice \( \sigma_n \) chosen by the discrepancy principle (4.2).

**Proposition 4.4.** Let \( x_n^\delta \) be defined by (2.1), let \( \sigma_n \) be chosen by the discrepancy principle (4.2) with \( 1 < C_1 \leq C_2 \) and let the assumptions A1 and A2 be valid.
Inverse Problems 27 (2011) 045012
Q Jin and U Tautenhahn

(i) High order regularization ($s \geq p$): if $w^2 := 1/h^2$ is operator monotone,
\[
\|x_n^\delta - x^\dagger\|_p \leq \frac{E}{C_1 - 1} \cdot \frac{M}{m} + E \cdot \frac{M}{m}.
\] (4.11)

(ii) Low order regularization ($s \leq p$): if $h^2$ is operator monotone and $h(t)/\sqrt{t}$ is decreasing,
then
\[
\|x_n^\delta - x^\dagger\|_p \leq \frac{E}{C_1 - 1} \cdot \frac{M}{m} + E \cdot \frac{M}{m}.
\] (4.12)

**Proof.** In case (i) we exploit the increasing behavior of $\sqrt{th(t)}$ and conclude from
\[
\sigma_{\min}^{-1} n / M^2 \leq \sigma_{\min}^{-1} / m^2
\]
that
\[
h\left(\frac{\sigma_{\min}^{-1} n / M^2}{\sigma_{\min}^{-1} / m^2}\right) \leq M m h\left(\frac{\sigma_{\min}^{-1} n / M^2}{\sigma_{\min}^{-1} / m^2}\right),
\]
which together with part (i) of the two propositions 3.4 and 4.3 provides (4.11). In case (ii) we exploit the decreasing
behavior of $h(t)/\sqrt{t}$, or equivalently the increasing behavior of $\sqrt{tw(t)}$, and conclude from
\[
\sigma_{\min}^{-1} n / M^2 \leq \sigma_{\min}^{-1} / m^2
\]
that
\[
w\left(\frac{\sigma_{\min}^{-1} n / M^2}{\sigma_{\min}^{-1} / m^2}\right) \leq M m w\left(\frac{\sigma_{\min}^{-1} n / M^2}{\sigma_{\min}^{-1} / m^2}\right),
\]
which together with part (ii) of the two propositions 3.4 and 4.3 provides (4.12).

Now we are in a position to prove the main result of this section. In our next theorem
we will see that order optimal error bounds can be guaranteed in case $\sigma_n$ is chosen by the
discrepancy principle (4.2) with $1 < C_1 \leq C_2$.

**Theorem 4.5.** Let the assumptions of proposition 4.4 be satisfied and assume in addition that $\xi_p(t) := \psi_p^2(1/(2p))$ is convex. Then,
\[
\|x_n^\delta - x^\dagger\|_p \leq c_1 \left[\psi_p^{-1}(c_2 \delta)\right]^p
\]
with some constants $c_1, c_2$ which can be extracted from the proof.

**Proof.** Due to proposition 4.4, in both cases (i) and (ii) of high and low order regularization
the total error obeys
\[
\|x_n^\delta - x^\dagger\|_p \leq c E
\] (4.13)
with some $c \geq 1$ and $\sigma_n$ chosen by the discrepancy principle (4.2) with $1 < C_1 \leq C_2$. From
(4.2) and the triangle inequality we have
\[
\|Ax_n^\delta - Ax^\dagger\| \leq \|Ax_n^\delta - y^\delta\| + \|y - y^\delta\| \leq (C_2 + 1)\delta.
\]
Using in addition the link condition A1(i) yields
\[
\rho(G)(x_n^\delta - x^\dagger) \leq (C_2 + 1)\delta/m.
\] (4.14)
Now the result of the theorem follows from (4.13), (4.14) and proposition 2.7.

4.3. Error bounds for rule (4.3)

For the *a posteriori* rule (4.3) of choosing the regularization parameter $\sigma_n$, analogous order
optimal error bounds to that of theorem 4.5 can be obtained.

**Theorem 4.6.** Let $x_n^\delta$ be defined by (2.1), let $\sigma_n$ be chosen by rule (4.3) where $\alpha_n$ obeys (4.4),
let both assumptions A1 and A2 hold and assume that $\xi_p(t) := \psi_p^2(t^{1/(2p)})$ be convex. Assume
further

(i) in the case of high order regularization ($s \geq p$) that $w^2 := 1/h^2$ is operator monotone,
(ii) in the case of low order regularization ($s \leq p$) that $h^2$ is operator monotone and $h(t)/\sqrt{t}$ is decreasing.

\[
\|x_n^\delta - x^\dagger\|_p \leq c E
\]
Then, the regularized solution \( x^\delta_n \) obeys the order optimal error bound
\[
\| x^\delta_n - x^\dagger \| \leq c_1 \left[ h_p^{-1}(c_2 \delta) \right]^p
\]  
(4.15)
with some constants \( c_1, c_2 \) which can be extracted from the proof.

Proof. We give the proof for the high order case \( s \geq p \); the proof for the low order case \( s \leq p \) is similar. In the first step of our proof we proceed according to the proof of proposition 4.3, exploit that \( C \delta \leq \| n_{n-1}(TT^*)^T(y^\dagger - Ax_0) \| \) and obtain
\[
(C - 1) \delta \leq E \sqrt{\sigma_{n-1}^{-1} h(\sigma_{n-1}^{-1}/M^2)}.
\]
From this estimate and (3.10) we obtain
\[
\| x^\delta_n - x^\dagger \| \leq \frac{E}{C - 1} \cdot \sqrt{\sigma_{n-1}^{-1} h(\sigma_{n-1}^{-1}/M^2)} + E \cdot M/m.
\]  
(4.16)
Now we consider two cases. In the first case with \( \sigma_{n-1}^{-1}/M^2 \leq \sigma_{n-1}^{-1}/m^2 \) we use the increasing behavior of \( \sqrt{th(t)} \) and obtain from (4.16) the estimate
\[
\| x^\delta_n - x^\dagger \| \leq \frac{E}{C - 1} \cdot \frac{M}{m} + E \cdot M/m.
\]
In the second case with \( \sigma_{n-1}^{-1}/M^2 \geq \sigma_{n-1}^{-1}/m^2 \) we use the decreasing behavior of \( h \), exploit in addition that due to (4.4) we have \( \sigma_n = 1/\alpha_n + \sigma_{n-1} \leq (c + 1)\sigma_{n-1} \), or equivalently, \( \sigma_{n-1} \leq (c + 1)\sigma_{n-1}^{-1} \), and obtain from (4.16) the estimate
\[
\| x^\delta_n - x^\dagger \| \leq \frac{E}{C - 1} \cdot \sqrt{c+1} + E \cdot M/m.
\]
From both cases we have that \( \| x^\delta_n - x^\dagger \|_p \) can be estimated by
\[
\| x^\delta_n - x^\dagger \|_p \leq \frac{E}{C - 1} \cdot \max\{M/m, \sqrt{c+1}\} + E \cdot M/m.
\]  
(4.17)
In the second step we proceed according to the proof of (4.14) and obtain
\[
\| \rho(G)(x^\delta_n - x^\dagger) \| \leq (C + 1)\delta/m.
\]  
(4.18)
In the final third step of the proof we use both estimates (4.17) and (4.18), apply proposition 2.7 and obtain (4.15). \( \square \)

4.4. Discrepancy principle revisited

The error bounds given in subsection 4.2 require in both cases of high order and low order regularization both link conditions A1(i) and A1(ii), and the assumption \( C_1 > 1 \) in the discrepancy principle (4.2). We will show in this subsection that in the case of low order regularization \( s \leq p \) order optimal error bounds can be obtained without the second link condition A1(ii). Our estimate in theorem 4.8 shows that \( C_1 = C_2 = 1 \) in the discrepancy principle (4.2) is best possible in the sense of minimal error bounds. We start our study with some important inequality.

Proposition 4.7. For \( 0 \leq s \leq p \), the regularized solution \( x^\delta_n \) defined by (2.1) obeys the estimate
\[
\| Ax^\delta_n - y^\dagger \|^2 + \sigma_{n-1}^{-1} \| x^\delta_n - x^\dagger \|^2 + \sigma_n^{-1} \| n_{n-1/2}(T^*T)G^{-s}(x^\dagger - x_0) \|^2 + \| y - y^\delta \|^2.
\]  
(4.19)
Proof. Let \( A : X_s \to Y \) be the restriction of \( A \) to \( X_s \subset X \) and \( A^*_s : Y \to X_s \) its adjoint. Due to the valid identity \((Ax, y) = (x, A^*_s y), (x, A^*_s y) = (x, G^{2s} A^*_s y), \) for all \( x \in X_s \) and

Q Jin and U Tautenhahn
Theorem 4.8. Let \( x_0 \) be defined by (2.1) and \( \sigma_n \) be chosen by the discrepancy principle (4.2) with \( 1 \leq C_1 \leq C_2 \), and assume the link condition \( A1(i) \), the solution smoothness \( A2 \) and

\[
G^2 g_n(T^*T) = g_n(A_n^*A)G^2.
\]

Consequently, the regularized solution (2.1) which is an element of the space \( X_i \) can be written in the equivalent form

\[
x_n^y - x_0 = g_n(A_n^*)A_n^*(y^\delta - Ax_0).
\]

From the valid identity \( x_n^y - x^1 = -r_n(A_n^*)A(x^1 - x_0) + g_n(A_n^*)A_n^*(y^\delta - Ax^1) \) and the identity \( g_n(A_n^*)A_n^* = A_n^*g_n(AA_n^*) \) we obtain

\[
\|x_n^y - x^1\|^2 = \|r_n(A_n^*)A(x^1 - x_0)\|^2 + \|g_n(A_n^*)A_n^*(y^\delta - Ax^1)\|^2
\]

\[
- 2(Ag_n(A_n^*)r_n(A_n^*)A(x^1 - x_0), y^\delta - Ax^1). \tag{4.21}
\]

We introduce the abbreviations

\[
R_n := g_n(A_n^*)r_n(AA_n^*) \quad \text{and} \quad \gamma_0^1 := y^\delta - Ax_0,
\]

decompose \( \gamma_0^1 \) into the sum \( A(x^1 - x_0) \) plus \( y^\delta - Ax^1 \) and obtain the equality

\[
(R_n(y_0^0, \gamma_0^1) = (R_nA(x^1 - x_0), A(x^1 - x_0)) + (R_n(Ax^1 - y^\delta), Ax^1 - y^\delta)
\]

\[
+ 2(R_nA(x^1 - x_0), y^\delta - Ax^1). \tag{4.22}
\]

Addition of equations (4.21) and (4.22) yields

\[
\begin{aligned}
(R_n(y_0^0, \gamma_0^1) + \|x_n^y - x^1\|^2 &= \|r_n(A_n^*)A(x^1 - x_0)\|^2 + \|g_n(A_n^*)A_n^*(y^\delta - Ax^1)\|^2 + \|A_n^*R_nA(x^1 - x_0), (x^1 - x_0)_s + (R_n(Ax^1 - y^\delta), Ax^1 - y^\delta).
\end{aligned}
\]

We use the valid identities

\[
r_n^2(A_n^*)A_n^* + A_n^*R_nA = r_n(A_n^*)A, \quad Ag_n^2(A_n^*)A_n^* + R_n = g_n(AA_n^*)
\]

and obtain from the above equation

\[
\begin{aligned}
(R_n(y_0^0, \gamma_0^1) + \|x_n^y - x^1\|^2 &= (r_n(A_n^*)A(x^1 - x_0), x^1 - x_0)_s
\end{aligned}
\]

\[
+ (g_n(AA_n^*)A(x^1 - y^\delta), Ax^1 - y^\delta). \tag{4.23}
\]

By exploiting properties (i) and (iv) of proposition 2.1, we obtain

(a) \( \sigma_n^{-1}(g_n(AA_n^*)A(x^1 - y^\delta), Ax^1 - y^\delta) \leq \|Ax^1 - y^\delta\|^2 \),

(b) \( \sigma_n^{-1}(R_n(y_0^0, \gamma_0^1) + \|x_n^y - x^1\|^2 \leq \|r_n(AA_n^*)A_n^*(y^\delta - Ax^1)\|^2 + \|A_n^*R_nA(x^1 - x_0), (x^1 - x_0)_s
\]

\[
+ (g_n(AA_n^*)A(x^1 - y^\delta), Ax^1 - y^\delta). \tag{4.23}
\]

We multiply (4.23) by \( \sigma_n^{-1} \), use the estimates (a) and (b) and obtain

\[
\|Ax^1 - y^\delta\|^2 + \|x_n^y - x^1\|^2 \leq \|Ax^1 - y^\delta\|^2 + \sigma_n^{-1}(r_n(A_n^*)A(x^1 - x_0), x^1 - x_0)_s. \tag{4.24}
\]

Finally we observe that due to (4.20) we have

\[
(r_n(A_n^*)A(x^1 - x_0), x^1 - x_0)_s = \|r_n^2(T^*T)G^{-1}(x^1 - x_0)\|^2.
\]

From this identity and (4.24) we obtain (4.19).

From propositions 4.7 and 2.8 we obtain the main result of this subsection.

**Theorem 4.8.** Let \( x_0^y \) be defined by (2.1) and \( \sigma_n \) be chosen by the discrepancy principle (4.2) with \( 1 \leq C_1 \leq C_2 \) and assume the link condition \( A1(i) \), the solution smoothness \( A2 \) and
\[0 \leq s \leq p. \text{ If } f(t) := \psi_s^2(t^{1/(2p-2s)}) \text{ is convex, and } \xi_s(t) := \psi_s^2(t^{1/(2s)}) \text{ is convex where } \psi_s \text{ is given by } (2.8), \text{ then}
\]
\[\|x^s_n - x^1\| \leq E \left[ \psi_p^{-1} \left( \frac{(C_2 + 1)\delta}{mE} \right) \right]^p.
\] (4.25)

**Proof.** For \(\sigma_n\) chosen by the discrepancy principle (4.2) the estimate (4.19) of proposition 4.7 attains the form
\[C_1^2 \delta^2 + \sigma_n^{-1} \|x^s_n - x^1\|^2 \leq \delta^2 + \sigma_n^{-1} \|r_n^{1/2}(T^*T)G^{-1}(x^1 - x_0)\|^2.
\]
Since \(C_1 \geq 1\), we have \(\|x^s_n - x^1\|^2 \leq \|r_n^{1/2}(T^*T)G^{-1}(x^1 - x_0)\|^2\). We use the representation \(G^{-1}(x^1 - x_n) = r_n(T^*T)G^{-1}(x^1 - x_0)\), see (2.3), use assumption A2 and obtain
\[\|x^s_n - x^1\|^2 \leq \|r_n^{1/2}(T^*T)G^{-1}(x^1 - x_0)\|^2 = (G^{p-2s}(x^1 - x_n), G^{-p}(x^1 - x_0)) \leq E \|x_n - x^1\|_{2\sigma-p},
\]
where \(x_n\) is the regularized solution with exact data. For estimating \(\|x_n - x^1\|_{2\sigma-p}\), we use estimate (2.15) of proposition 2.8 and obtain
\[\|x^s_n - x^1\| \leq E \left[ \psi_p^{-1} \left( \frac{\varrho(G)(x_n - x^1)}{E} \right) \right]^{p-s}.
\] (4.26)

For estimating \(\|Ax_n - Ax^1\|\), we use (2.5), the identity \(r_n(T^*T)(y^3 - Ax_0) = y^3 - Ax^s_n, r_n(\lambda) \leq 1\) and (4.2) and obtain the estimate
\[\|Ax_n - Ax^1\| = \|r_n(T^*T)(y - Ax_0)\| \leq \|r_n(T^*T)(y^3 - Ax_0)\| + \|r_n(T^*T)(y - y^3)\| \leq (C_2 + 1)\delta.
\]

Hence, by using A1(i) we have \(\|\varrho(G)(x_n - x^1)\| \leq (C_2 + 1)\delta/\sigma_m\). Since \(\psi_p^{-1}\) is monotone, we obtain from (4.26) the estimate
\[\|x^s_n - x^1\| \leq E \left[ \psi_p^{-1} \left( \frac{(C_2 + 1)\delta}{mE} \right) \right]^{p-s}.
\] (4.27)

Next, let us estimate \(\|\varrho(G)(x^s_n - x^1)\|\). Using assumption A1(i) and the estimate \(\|Ax^s_n - Ax^1\| \leq \|Ax^s_n - y^3\| + \|y - y^3\| \leq (C_2 + 1)\delta\) yields
\[\|\varrho(G)(x^s_n - x^1)\| \leq (C_2 + 1)\delta/\sigma_m.
\] (4.28)

Now we apply the interpolation estimate (2.14) of proposition 2.7 and obtain by using (4.27), (4.28) and the abbreviation \(\delta_1 := \frac{(C_2 + 1)\delta}{mE}\) that
\[\|x^s_n - x^1\| \leq E \left[ \psi_p^{-1}(\delta_1) \right]^{p-s} : \left[ \psi_p^{-1}(\delta_1) \right]^{p-s} \right]^s.
\] (4.29)

From the first equation of (3.22) we have \(\psi_p^{-1}(\psi_p(\lambda) \cdot \lambda^{s-p}) = \lambda\). Substituting \(\lambda = \psi_p^{-1}(\delta_1)\) yields \(\psi_p^{-1}(\delta_1 \cdot [\psi_p^{-1}(\delta_1)]^{p-s}) = \psi_p^{-1}(\delta_1)\). From this equation and (4.29) we obtain (4.25). \(\square\)
5. Practical implementation

For the practical application of implicit iteration methods in Hilbert scales one has to make different decisions. First, one has to choose the operator $B$, second, one has to fix the number $s$ in the method (1.2), third, one has to choose the starting value $x_0$ and to fix the numbers $\alpha_k$, $k = 1, \ldots, n$, and fourth, one has to effectively realize the discrepancy principle (4.1) with a little number $n$ of iteration steps. The choice of $B$ and $x_0$ should be made depending on the expected smoothness of the element $x^\dagger - x_0$ such that assumption A2 holds true for $p$ sufficiently large, and $s$ should have the magnitude of $p$. In our further study we concentrate on the choice of the numbers $\alpha_k$, $k = 1, \ldots, n$, for effectively realizing the discrepancy principle (4.1), (4.2) or (4.3), respectively, with a little number $n$ of iteration steps. In a first proposition we give an upper bound for the regularization parameter of the discrepancy principle in case $n = 1$ which will serve us as the starting value for iteration (1.2). To the best of our knowledge, so far there have not been upper bounds for the regularization parameter of the discrepancy principle in the literature.

**Proposition 5.1.** Let $n = 1$, let $x^\delta_1$ be the regularized solution (2.1) and let $\alpha_1 = \alpha_D$ be chosen by the discrepancy principle (4.1) with $C \geq 1$. If $\Vert y^\delta - Ax_0 \Vert > C \delta$, then

$$\alpha_D < \frac{C \delta \Vert G^t A^* (y^\delta - Ax_0) \Vert^2}{(\Vert y^\delta - Ax_0 \Vert - C \delta) \Vert y^\delta - Ax_0 \Vert^2}.$$  
(5.1)

**Proof.** Let $x^\delta_0 = x_0 - (A^* A + \alpha G^{-2s})^{-1} A^* (A x_0 - y^\delta)$ and $\alpha = \alpha_D$ be the regularization parameter that obeys the discrepancy principle $\Vert Ax^\delta_0 - y^\delta \Vert = C \delta$. For solving this nonlinear equation, Newton’s method applied to the equivalent equation

$$g(r) = \Vert Ax^\delta_1/r - y^\delta \Vert^{-1} - (C \delta)^{-1} = 0$$

is studied in [16] which results in the iteration

$$r_{k+1} = r_k - \frac{\Vert Ax^\delta_1/r_k - y^\delta \Vert^{-1} - (C \delta)^{-1}}{r_k^{-3} (v^\delta_{1/r_k}, G^{-2s} (v^\delta_{1/r_k} - x_0)) \Vert Ax^\delta_1/r_k - y^\delta \Vert^{-3}}$$  
(5.2)

where $v^\delta_{1/r}$ is given by $v^\delta_{1/r} = (A^* A + r^{-1} G^{-2s})^{-1} G^{-2s} (x^\delta_{1/r} - x_0)$. From [16, theorem 3.5] we know that iteration (5.2) possesses the following properties.

(i) The sequence $(r_k)$ converges globally and monotonically from the left to $r_D$ for any starting values $0 \leq r_0 < r_D := 1/\alpha_D$.

(ii) The speed of convergence is locally quadratic.

For $r \to 0$, both limit relations

$$\lim_{r \to 0^+} x^\delta_{1/r} = x_0 \quad \text{and} \quad \lim_{r \to 0^+} r^{-3} (v^\delta_{1/r}, G^{-2s} (x^\delta_{1/r} - x_0)) = \Vert G^t A^* (y^\delta - Ax_0) \Vert^2$$

are valid. We execute one iteration step of iteration (5.2) with starting value $r_0 = 0$ and obtain due to the above limit relations that

$$r_1 = \frac{\Vert y^\delta - Ax_0 \Vert - C \delta} {C \delta \Vert G^t A^* (y^\delta - Ax_0) \Vert^2}.$$  

Due to the above property (i) we have $r_1 < r_D$. Since $r$ and $\alpha$ are related by $\alpha = 1/r$ we obtain (5.1).

Based on the Newton iteration (5.2) we propose the following strategy for effectively realizing the discrepancy principle (4.3) with a little number $n$ of iteration steps.
Algorithm 1 Global convergent Newton iteration for rule (4.3)

1: Start with initial data $y^δ, A, G, s, δ, C := 1.1$ and $x_0$.
2: if $\|Ax_0 - y^δ\| > Cδ$ then
3: Compute $α$ by the right hand side of (5.1) with $C = 1$.
4: Compute $x := x_0 - (A^*A + αG^{-2s})^{-1}A^*(Ax_0 - y^δ)$ and set $n := 1$.
5: while $\|Ax - y^δ\| > Cδ$ do
6: Compute $v := (A^*A + αG^{-2s})^{-1}G^{-2s}(x - x_0)$.
7: Update $r := \frac{1}{α}\frac{\|Ax - y^δ\|-1}{\|Ax - y^δ\|-1 - δ^{-1}}\frac{\|Ax - y^δ\|-1}{\|Ax - y^δ\|-1}$, $n := n + 1$.
8: $x_0 := x$, $α := \frac{1}{r}$, $x := x_0 - (A^*A + αG^{-2s})^{-1}A^*(Ax_0 - y^δ)$.
9: end while
10: end if

For discussing some properties of algorithm 1, we will work with the notation

\[ x_k^α(α_k) := x_{k-1}^α - (A^*A + αB^{2s})^{-1}A^*(Ax_{k-1}^α - y^δ), \quad k = 1, 2, \ldots, \]

that indicates the dependence of $x_k^α$ defined by (1.2) on the parameter $α$. We start by some monotonicity property of the sequence $(α_k)_{k=1}^n$ in the iteration (1.2).

**Proposition 5.2.** The regularized solutions $x_k^α$, $k = 1, \ldots, n$, obtained by algorithm 1 have the form (1.2). The related sequence $(α_k)_{k=1}^n$ is strictly monotonically decreasing, and all $α_k$ are positive.

**Proof.** In steps 3 and 4 of algorithm 1, $α_1$ and $x_1^α(α_1)$ are computed. Then, the while loop (steps 5–9 of algorithm 1) is executed $n - 1$ times to obtain $α_k$ and $x_k^α(α_k)$ for $k = 2, \ldots, n$. The parameter $α = α_k := 1/r_k$ (see step 7 of algorithm 1) is obtained by performing one Newton step for solving the nonlinear equation

\[ g(r) = \|Ax_{k-1}^α(1/r) - y^δ\|-1 - δ^{-1} = 0 \]

with starting value $r_0 = 1/α_0$. It can be shown (compare [16]) that the function $g$ possesses the following properties.

(i) There hold the two limit relations

\[ \lim_{r \to 0^+} g(r) = \|Ax_{k-2}^α - y^δ\|-1 - δ^{-1} < 0 \quad \text{and} \quad \lim_{r \to \infty} g(r) = +∞. \]

(ii) The function $g: \mathbb{R}^+ \to \mathbb{R}$ is monotonically increasing and concave.

From these properties and $g(r_{k-1}) < 0$ we conclude that $r_{k-1} < r_k < \infty$. It follows that $0 < α_k < α_{k-1}$ for all $k = 2, \ldots, n$.

For discussing further properties of algorithm 1 we consider Tikhonov regularization

\[ x_k^α(β_m) := x_0 - (A^*A + β_mG^{-2s})^{-1}A^*(Ax_0 - y^δ) \quad (5.3) \]

and assume

(i) $β_k = r_{k-1}^{-1}$, $k = 2, \ldots, m$, is obtained by the iteration

\[ r_k = r_{k-1} - \frac{\|Ax_{k-1}^α - y^δ\|^3 - δ^{-1}}{r_{k-1}^3(v_{0}^α)^β_mG^{-2s}(x_{k-1})_{0}^α - x_0^α)}\|Ax_{k-1}^α - y^δ\|^3, \quad (5.4) \]

where $v_{0}^α$ is given by $v_{0}^α = (A^*A + r_{-1}^{-1}G^{-2s})^{-1}G^{-2s}(x_{k-1}^α - x_0^α)$ and $x_{k-1}^α$ is given by $x_{k-1}^α = x_k^α(1/r)$, and
(ii) \( r_1 \) is chosen as \( r_1 = \frac{(y^\delta - A x_0)(y^\delta - A x_0)^T}{\|G A (y^\delta - A x_0)\|^2} \) and the iteration (5.4) is stopped with the first integer \( m \) for which, with \( C := 1.1 \),
\[
\|A x_1^k(\beta_1) - y^\delta\| \leq C \delta < \|A x_1^k(\beta_k) - y^\delta\|, \quad 0 \leq k < m. \tag{5.5}
\]

From [16] we know that iteration (5.4) converges globally and monotonically from the left to the solution of the equation \( g(r) = \|A x_1^k(1/r) - y^\delta\|^{-1} - \delta^{-1} = 0 \), and that in the vicinity of the solution we have quadratic speed of convergence. It follows that by the stopping rule (5.5) a finite number \( m \) of iteration steps is defined. Our next proposition states that algorithm 1 is not slower than iteration (5.4) with stopping rule (5.5).

**Proposition 5.3.** Let \( m \) be the number of iterations of method (5.4) with stopping rule (5.5). Then, \( n \leq m \), where \( n \) is the number of iterations of algorithm 1.

**Proof.** Assume that \( \alpha_1 \) and \( x_1(\alpha_1) \) in steps 3 and 4 of algorithm 1 are computed, which coincide with \( \beta_1 \) and \( x_1(\beta_1) \) of iteration (5.4). Then, in the first iteration step of the while-loop (steps 5–9 of algorithm 1) we obtain \( \alpha_2 \) and
\[
x_2^\delta(\alpha_2) = x_1^\delta(\alpha_1) - (A^* A + \alpha_2 G^{-2\gamma})^{-1} A^* (A x_1^\delta(\alpha_1) - y^\delta).
\]
For \( x_2^\delta(\alpha_2) \) computed in this way we have
\[
y^\delta - A x_2^\delta(\alpha_2) = \alpha_2 (T T^* + \alpha_2 I)^{-1} \alpha_1 (T T^* + \alpha_1 I)^{-1} (y^\delta - A x_0). \tag{5.6}
\]
On the other hand, from iteration (5.4) we obtain after the first step the regularization parameter \( \beta_2 = \alpha_2 \) and the regularized solution \( x_1^\delta(\beta_2) \) which obeys
\[
y^\delta - A x_1^\delta(\beta_2) = \beta_2 (T T^* + \beta_2 I)^{-1} (y^\delta - A x_0). \tag{5.7}
\]
Comparing both identities (5.6) and (5.7) and observing that \( \alpha_2 = \beta_2 \) we obtain that
\[
\|y^\delta - A x_2^\delta(\alpha_2)\| \leq \|y^\delta - A x_1^\delta(\beta_2)\|. \tag{5.8}
\]
In a similar way we obtain that
\[
\|y^\delta - A x_2^\delta(\alpha_k)\| \leq \|y^\delta - A x_1^\delta(\beta_k)\|, \quad k = 3, \ldots, n,
\]
where \( (\alpha_k) \) is the sequence generated by algorithm 1 and \( (\beta_k) \) is the sequence generated by (5.4). From this estimate we obtain that algorithm 1 terminates not later than iteration (5.4) with stopping rule (5.5). \( \Box \)

After termination of algorithm 1, different cases can appear.
(i) We have \( \delta < \|A x_n^\delta - y^\delta\| \leq C \delta \) with \( C = 1.1 \). In this case, all three theorems 4.5, 4.6 and 4.8 apply.
(ii) We have \( \|A x_n^\delta - y^\delta\| = \delta \). Then, both theorems 4.6 and 4.8 apply.
(iii) We have \( \|A x_n^\delta - y^\delta\| < \delta \). In this case, theorem 4.6 applies.

Our next proposition states that in all three termination cases (i)–(iii), the additional assumption (4.4) of theorem 4.6 is satisfied with some \( c < 1 \).

**Proposition 5.4.** The regularized solution \( x_n^\delta \) obtained by algorithm 1 has the form (2.1) with some sequence \( (\alpha_k)_{k=1}^n \) that obeys assumption (4.4) with \( c < 1 \).

**Proof.** Consider the final iteration of the while-loop (steps 5–9 of algorithm 1). Starting from \( x = x_{n-1}(\alpha_{n-1}) \) with \( \|A x - y^\delta\| > C \delta, \alpha_n := 1/r_n \) is obtained by performing one Newton step for solving the nonlinear equation
\[
g(r) = \|A x_{n-1}(1/r) - y^\delta\|^{-1} - \delta^{-1} = 0
\]
with starting value $r_{n-1} = 1/\alpha_{n-1}$. As a result, we obtain some $\alpha_n < \alpha_{n-1}$, see proposition 5.2, and the final regularized solution $x_n^\delta$ is obtained by

$$x_n^\delta(\alpha_n) = x_{n-1}^\delta(\alpha_{n-1}) - (A^*A + \alpha_n^2G^{-2s})^{-1}A^*(Ax_{n-1}^\delta(\alpha_{n-1}) - y^\delta).$$

Some formal computations show that $x_n^\delta(\alpha_n)$ can be rewritten as

$$x_n^\delta(\alpha_n) = x_{n-1}^\delta(\alpha_{n-1}) - (A^*A + \alpha_nG^{-2s})^{-1}A^*(Ax_{n-1}^\delta(\alpha_{n-1}) - y^\delta).$$

Since the function $g$ is monotonically increasing and concave and since $g(r_{n-1}) < 0$ we conclude that the element $x_{n-1}^\delta(\alpha_{n})$ obeys $\|Ax_{n-1}^\delta(\alpha_{n}) - y^\delta\| > \delta$. It follows that the final two parameters $\alpha_{n-1}$ and $\alpha_n$ in iteration (1.2) can be interchanged such that we have $\alpha_n > \alpha_{n-1}$. This yields (4.4) with some constant $c < 1$.

6. Numerical experiments

In this section we perform numerical experiments for computing regularized solutions by algorithm 1. We consider Fredholm integral equations

$$[Ax](s) := \int_0^1 K(s, t)x(t)\, dt = y(s), \quad 0 \leq s \leq 1, \quad A : L^2(0, 1) \to L^2(0, 1)$$

(6.1)

and differential operators $B : D \subset L^2(0, 1) \to L^2(0, 1)$ of first order defined by

$$Bx = \sum_{k=1}^\infty k(x, e_k)e_k \quad \text{with} \quad e_k(t) = \sqrt{2}\sin(k\pi t).$$

(6.2)

Example 6.1. Our test example (deriv2 from [7]) is (6.1) with the kernel function

$$K(s, t) = \begin{cases} s(1-t) & \text{for } s \leq t \\ t(1-s) & \text{for } s \geq t. \end{cases}$$

For this kernel function, assumption A1 is satisfied with $m = M = \pi^{-2}$ and $g(t) = t^2$. We consider three subexamples in which the right-hand sides $y(s)$, the corresponding solutions $x^\delta(t)$ and the maximal smoothness parameters $p_0$ for which assumption A2 with $x_0 = 0$ holds true for all $p \in (0, p_0)$ are given by

(i) $y(s) = -\frac{1}{4\pi^2} \sin 2\pi s, \quad x^\delta(t) = \sin 2\pi t, \quad p_0 = \infty$,

(ii) $y(s) = \frac{s}{3}(1 - 2x^2 + x^3), \quad x^\delta(t) = 4t(1 - t), \quad p_0 = \frac{5}{2}$,

(iii) $y(s) = \frac{s}{6}(1 - x^2), \quad x^\delta(t) = t, \quad p_0 = \frac{1}{2}$.

The discretization of (6.1) has been performed by Galerkin approximation as outlined, e.g., in [7, 16], guaranteeing that $\|x^1\|_2 \approx \|x^1(t)\|_{L^2(0,1)}$ and $\|y\|_2 \approx \|y(s)\|_{L^2(0,1)}$ hold. As a discrete approximation of the first order differential operator (6.2) we use the $(m, m)$-matrix

$$B := B_m^{1/2} \quad \text{with} \quad B_m = \frac{(m+1)^2}{\pi^2} \begin{pmatrix} 2 & -1 & \cdots & -1 \\ -1 & \ddots & \ddots & \ddots \\ \vdots & \ddots & \ddots & -1 \\ -1 & \cdots & \cdots & 2 \end{pmatrix}.$$
For modeling noise in the discretized right-hand side $y \in \mathbb{R}^m$, for given nonnegative $\sigma$ we compute

$$y^\delta = y + \sigma \frac{\|y\|_2}{\|e\|_2} e,$$

where $e = (e_i)$ is a random vector with $e_i \sim \mathcal{N}(0, 1)$. In this way of modeling noise we guarantee that for the relative error we have $\|y - y^\delta\|_2/\|y\|_2 = \sigma$. The noise level $\delta$ is then given by $\delta = \sigma \|y\|_2$. Tables 1–3 show our numerical results with $x_0 = 0$, where the letter codes in the leftmost column refer to the following three iteration methods.

- (TI/DP): this is the method of Tikhonov regularization (5.3) with $x_0 = 0$ and $s = 1$, where the regularization parameter obeys (5.5) and is obtained by iteration (5.4) which converges globally and locally quadratically.
- (IIM/A1): this is the implicit iteration method (1.2) with $x_0 = 0$ and $s = 1$, where the sequence $(\alpha_k)_{k=1}^\infty$ is obtained by algorithm 1.
- (IIM/GS): this is the implicit iteration method (1.2) with $x_0 = 0$ and $s = 1$, where the sequence $(\alpha_k)_{k=1}^\infty$ is the geometric sequence $(\sigma^{k-1} \alpha_1)_{k=1}^\infty$ with $q = \frac{1}{2}$ as proposed in [6] and stopping rule (4.3) with $C = 1.1$.

For all three iteration methods our tables contain

1. the number $n$ of required iterations,
2. the final regularization parameter $\alpha_n$,
3. the discrepancy $d_n := \|Ax_n^\delta(\alpha_n) - y^\delta\|_2$ of the final approximation for the iteration method (TI/DP), and the discrepancy $d_n := \|Ax_n^\delta - y^\delta\|_2$ of the final approximation for the iteration methods (IIM/A1) and (IIM/GS),
4. the error $e_n := \|x_n^\delta(\alpha_n) - x^\dagger\|_2$ of the final approximation for the iteration method (TI/DP), and the error $e_n := \|x_n^\delta - x^\dagger\|_2$ of the final approximation for the iteration methods (IIM/A1) and (IIM/GS).
In our experiments, all three iteration methods have been started first with

$$\alpha_1 = \delta \left( \frac{R^{-2s}A^*(y^\delta - Ax_0), A^*(y^\delta - Ax_0)}{\|y^\delta - Ax_0\|^2} \right)$$

(6.3)

with $s = 1$ and $x_0 = 0$, cf (5.1), and second with $\alpha_1 = 1$ as performed in [6]. In order to keep the discretization error small, we have used the dimension number $m = 400$ in all computations.

In our numerical experiments we observed that the accuracy of each individual regularization method in the three test cases of examples 6.1 (i)–(iii) is as predicted by the theory. In tables 1–3 we mainly concentrate on the performance of the three methods and observe the following.

(i) As far as computational expenses are concerned, the iteration method (IIM/A1) performs best. In fact, this method requires the smallest number of iterations compared with the other two methods.

(ii) For the method (IIM/GS), the number of iterations can considerably be reduced by starting with $\alpha_1$ from (6.3) instead of starting with $\alpha_1 = 1$. For the other two methods (TI/DP) and (IIM/A1), the number of iterations differs only by 1 for the two starting values (6.3) and $\alpha_1 = 1$, respectively.

(iii) In all three iteration methods, the $\alpha$-sequence $(\alpha_k)_{k=1}^n$ is decreasing. However, the final regularization parameter $\alpha_n$ is the smallest for the method (TI/DP). Comparing the discrepancies $d_k$ for the individual iterations $k = 1, 2, \ldots$ (which are not contained in the tables) we observed that, for $k \geq 2$, $d_k$ in the method (IIM/A1) is always smaller than $d_k$ in the method (TI/DP).

In tables 1–3, the simulation results show a strong dependence of the number $n$ of iterations on the choice of the starting value $\alpha_1$ and on the chosen sequence $(\alpha_k)_{k=1}^n$. Since the starting value $\alpha_1$ has to obey $\|Ax_1^0 - y^\delta\| \geq \delta$, some random choice seems to be unreasonable and we propose the choice (6.3), which by proposition 5.1 obeys the required property $\|Ax_1^0 - y^\delta\| \geq \delta$. So far, we do not know any better starting point. Starting from this $\alpha_1$, it appears to us that due to the global and local quadratic speed of convergence of Newton’s method realized in algorithm 1, this choice for $(\alpha_k)_{k=1}^n$ should be preferred over the geometric sequence $(q^{k-1} \alpha_1)_{k=1}^n$.

References


[12] Jin Q 2010 On a regularized Levenberg–Marquardt scheme is of optimal order J. Integral Eqns Appl. 22 259–83
[30] Porsnawad P and Böckmann C 2010 Convergence rate analysis of the first-stage Range–Kutta-type regularizations Inverse Problems 26 035005
[37] Uchiyama M 2010 Majorization and some operator monotone functions Linear Algebra Appl. 432 1867–72
[38] Vainikko G M and Veretennikov A Y 1986 Iteration Procedures in Ill-Posed Problems (Moscow: Nauka) (in Russian)