Nonstationary Iterated Tikhonov Regularization

Martin Hanke and C.W. Groetsch

Abstract

A convergence rate is established for nonstationary iterated Tikhonov regularization, applied to ill-posed problems involving closed, densely defined linear operators, under general conditions on the iteration parameters. It is also shown that an order-optimal accuracy is attained when a certain a posteriori stopping rule is used to determine the iteration number.

1 Introduction

Many inverse problems in the physical sciences may be posed in the form

$$Tx = y \tag{1}$$

where T is a linear operator on a Hilbert space having an unbounded (generalized) inverse, y is a given "data" vector, and x is a desired solution (e.g., [9], [6], [13]). Because the generalized inverse is discontinuous, problem (1) is ill-posed, that is, the solution x depends in an unstable way on the data y.

---

† *MR Classification:* 65J20, 47A50, 65J15

This work was partially supported by NATO (grant CRG 930044). CWG gratefully acknowledges support from the Charles Phelps Taft Memorial Fund and a Raybould Visiting Fellowship at the University of Queensland.

A natural way to alleviate this instability is to replace (1) with an approximating well-posed problem. The best known way of accomplishing this is by Tikhonov regularization, that is, instead of (1) one solves

$$(T^*T + \alpha I)x_\alpha = T^*y \tag{2}$$

where α is a positive "regularization" parameter and T^* is the adjoint of T. It is easy to show that as $\alpha \to 0$ the unique solution x_α of (2) converges to the minimal norm least squares solution of (1) whenever it exists. There is a well-developed convergence theory for (2) (e.g., [5], [13]), an important ingredient of which is a strategy for relating the regularization parameter to perturbed data in such a way that as the error level diminishes to zero the approximations converge to the desired solution. In this respect the method (2) is deficient – the rate of convergence of (2), with respect to the error level δ in the data, cannot in general exceed a certain "saturation level" of $O\left(\delta^{2/3}\right)$ [5]. It is well known that this rate may be improved in an iterated version of (2) given by

$$(T^*T + \alpha I)x_n = \alpha x_{n-1} + T^*y \tag{3}$$

(see, e.g., [11], [10], [4]). Brill and Schock [2] have investigated a nonstationary version of (3), namely

$$(T^*T + \alpha_n I)x_n = \alpha_n x_{n-1} + T^*y \tag{4}$$

for the case of a compact operator T (see also [16]). A special case of (3), namely $\alpha = 1$, has been analyzed by Lardy [12] for the case of a closed densely defined *unbounded* operator T. In ordinary Tikhonov regularization a number of authors (e.g. [9, p. 96]), [17, p. 92]) have advocated a successive geometric choice of the regularization parameter, which in turn suggests the choice $\alpha_n = \alpha q^{n-1}(0 < q < 1)$ in the nonstationary iterative Tikhonov method. For a certain adaptive choice of α_n Brakhage [1] has established a linear convergence rate.

The convergence rate for (4) derived in [2] assumes a condition on the parameters $\{\alpha_n\}$ that is not satisfied for the stationary method (3) (and in particular for Lardy's method) nor for the geometric choice of parameters. One of our purposes in this paper is to establish the Brill/Schock convergence rate for the nonstationary method under conditions that are flexible enough to cover a wide range of iteration parameters. We also establish a convergence rate for the nonstationary method with perturbed data when the iteration number is selected by a discrepancy principle, and we illustrate our results for the geometric choice of regularization parameters mentioned above.

2 Convergence Rates: Linear Operators

Suppose T is a closed linear operator defined on a dense domain $\mathcal{D}(T)$ in a Hilbert space H and that $y \in R(T)$, the range of T. Let $x^\dagger$ be the normal solution of (1), that is, $x^\dagger$ is the unique vector satisfying

$$x^\dagger \in \mathcal{D}(T) \cap N(T)^\perp \text{ and } Tx^\dagger = y,$$

where $N(T)^\perp$ is the orthogonal complement of the nullspace of T. For a given sequence of positive numbers $\{\alpha_n\}$, take (for simplicity) $x_0 = 0$ and define x_n by

$$x_n = \alpha_n(T^*T + \alpha_n I)^{-1}x_{n-1} + T^*(TT^* + \alpha_n I)^{-1}y. \tag{5}$$

We note that both of the operators $(T^*T + \alpha_n I)^{-1}$ and $T^*(TT^* + \alpha_n I)^{-1}$ are everywhere defined and bounded [15, p. 307] (with $\|(T^*T + \alpha_n I)^{-1}\| \leq \alpha_n^{-1}$), hence for each fixed n, $x_n \in \mathcal{D}(T)$ is stable with respect to perturbations in y.

For bounded operators Brill and Schock [2] have proved that the method (5) converges to $x^\dagger$ if and only if $\sum \alpha_n^{-1} = \infty$ and they established a convergence rate under the additional assumption that $\sum \alpha_n^{-2} < \infty$. Our first goal is to establish this rate under a strictly weaker assumption on $\{\alpha_n\}$ that includes as special cases Lardy's method and iterated Tikhonov regularization with geometric parameter scheme.

We begin by noting that, since $y = Tx^\dagger$,

$$T^*(TT^* + \alpha_n I)^{-1}y = T^*T(T^*T + \alpha_n I)^{-1}x^\dagger = x^\dagger - \alpha_n(T^*T + \alpha_n I)^{-1}x^\dagger$$

and hence by (5):

$$x^\dagger - x_n = r_n(T^*T)x^\dagger, \text{ where } r_n(\lambda) = \prod_{i=1}^{n} \frac{\alpha_i}{\lambda + \alpha_i}.$$

If, in addition, $x^\dagger = (T^*T)^\nu w$ for some $w \in \mathcal{D}((T^*T)^\nu)$ and $\nu > 0$ then

$$x^\dagger - x_n = f_{n,\nu}(T^*T)w \tag{6}$$

where

$$f_{n,\nu}(\lambda) := \lambda^\nu r_n(\lambda).$$

The error analysis will hinge on an investigation of the function

$$f_{n,\nu}(\lambda) = \lambda^\nu \prod_{i=1}^{n} \frac{\alpha_i}{\lambda + \alpha_i} \;\;, \;\; \lambda \in [0, \infty)$$

where $\nu > 0$, $\alpha_i > 0$ are given parameters. As we are interested in fixed $\nu > 0$ and $n \to \infty$, we shall assume that $n > \nu$ (note that for $\nu \geq n$, $f_{n,\nu}(\lambda)$ is increasing in λ). An easy calculation shows that $f'(\lambda) = 0$ if and only if

$$g(\mu) := \sum_{j=1}^{n} \frac{1}{1 + \alpha_j \mu} = \nu \tag{7}$$

where $\mu = \lambda^{-1}$. Since $g(0) = n > \nu > 0 = g(\infty)$ and g is strictly decreasing, equation (7) has a unique positive solution, say $\mu = \mu_1$. Furthermore, since $f_{n,\nu}(0) = 0 = f_{n,\nu}(\infty)$,

$$\max_{\lambda \in [0,\infty)} f_{n,\nu}(\lambda) = f_{n,\nu}(\mu_1^{-1}) \leq \mu_1^{-\nu}. \tag{8}$$

Also, the negative solutions of (7) are separated by the vertical asymptotes $\mu = -\alpha_j^{-1}$, and hence, if we denote these negative solutions by $\mu_2, \mu_3, \ldots, \mu_n$, then

$$-\sum_{j=2}^{n} \mu_j \geq \sum_{j=1}^{n-1} \alpha_j^{-1} = \sigma_{n-1} \tag{9}$$

where we have used the notation

$$\sigma_m := \sum_{j=1}^{m} \alpha_j^{-1}.$$

Lemma 1 $\sum_{i=1}^{n} \mu_i = \frac{1-\nu}{\nu}\,\sigma_n$

Proof: Equation (7) may be written in the equivalent form

$$\nu \prod_{k=1}^{n}(1+\alpha_k\mu) - \sum_{j=1}^{n} \prod_{\substack{k=1 \\ k\neq j}}^{n}(1+\alpha_k\mu) = 0.$$

Using the fact that the sum of the roots of a monic polynomial is the negative of the next-to-highest order coefficient, we then obtain

$$\sum_{i=1}^{n} \mu_i = -\sum_{i=1}^{n} \alpha_i^{-1} + \frac{1}{\nu \prod_{k=1}^{n} \alpha_k} \sum_{j=1}^{n} \prod_{\substack{k=1 \\ k\neq j}}^{n} \alpha_k = \frac{1-\nu}{\nu}\,\sigma_n. \Box$$

Lemma 2 *If* $0 < \nu < 1$, *then* $\max_{\lambda\in[0,\infty)} f_{n,\nu}(\lambda) \leq \left(\frac{\nu}{1-\nu}\right)^{\nu} \sigma_n^{-\nu}$.

Proof: Because the roots $\mu_2, \ldots, \mu_n$ are negative, we have by Lemma 1,

$$\mu_1 \geq \sum_{i=1}^{n} \mu_i = \frac{1-\nu}{\nu}\,\sigma_n$$

and hence

$$\mu_1^{-\nu} \leq \left(\frac{1-\nu}{\nu}\right)^{-\nu} \sigma_n^{-\nu}$$

and the result follows from (8). $\Box$

For the case $0 < \nu \leq 1$, we note that Xi [20] has obtained the stronger estimate: $f_{n,\nu}(\lambda) \leq \nu^{\nu}\sigma_n^{-\nu}$. In order to handle the case $\nu \geq 1$, we will need to assume an additional condition on the parameters $\{\alpha_n\}$. Specifically, we shall assume that there is a positive constant c such that

$$\frac{1}{\alpha_n} \leq c\sigma_{n-1} \tag{10}$$

for all n sufficiently large. Note that (10) is strictly weaker (considering the necessary condition, $\sigma_n \to \infty$, for convergence) than the condition $\sum \alpha_k^{-2} < \infty$ of [2], and that (10) is satisfied for the stationary method (in particular, Lardy's method) and the method with $\alpha_n = \alpha q^{n-1} (0 < q < 1)$, cf. Section 4.

Lemma 3 *If $0 < \nu < n$ and condition (10) is satisfied, then*

$$\max_{\lambda \in [0,\infty)} f_{n,\nu}(\lambda) \leq c_\nu \sigma_n^{-\nu}$$

where $c_\nu = (2\nu(c+1))^\nu$ for $0 < \nu \leq 1$ and $c_\nu = (2\nu(c+1)^\nu)^\nu$ for $\nu > 1$.

Proof: For $0 < \nu \leq 1/2$, we have $\left(\dfrac{\nu}{1-\nu}\right)^\nu \leq c_\nu$ and the result follows from Lemma 2. On the other hand, if $1/2 < \nu \leq 1$, then by Lemma 1 and (9) we have

$$\mu_1 \geq -\sum_{i=2}^{n} \mu_i \geq \sigma_{n-1}$$

and, by (10),

$$\sigma_n = \frac{1}{\alpha_n} + \sigma_{n-1} \leq (c+1)\sigma_{n-1}. \tag{11}$$

It then follows from (8) that

$$\max_{\lambda \in [0,\infty)} f_{n,\nu}(\lambda) = f_{n,\nu}(\mu_1^{-1}) \leq \mu_1^{-\nu} \leq \sigma_{n-1}^{-\nu} \leq (c+1)^\nu \sigma_n^{-\nu} \leq c_\nu \sigma_n^{-\nu}.$$

The case $\nu > 1$ will be handled by an inductive argument. We suppose that

$$f_{n,\nu}(\mu_1^{-1}) \leq c_\nu \sigma_n^{-\nu} \tag{12}$$

holds for all ν with $0 < \nu \leq \nu_0$, for some $\nu_0 \geq 1$. Take $\nu \in (\nu_0, \nu_0 + 1]$ and $n > \nu$. We will show that (12) holds for ν. By Lemma 1 and (9) we have

$$\begin{aligned} \mu_1 &= -\sum_{i=2}^{n} \mu_i + \frac{1-\nu}{\nu}\sigma_n \\ &\geq \sigma_{n-1} + \frac{1-\nu}{\nu}\sigma_n = \frac{1}{\nu}\sigma_{n-1} - \frac{\nu-1}{\nu}\frac{1}{\alpha_n}. \end{aligned}$$

Consider now two cases. If $\frac{1}{\alpha_n} \leq \frac{1}{2(\nu-1)}\sigma_{n-1}$, then

$$\mu_1 \geq \frac{1}{\nu}\sigma_{n-1} - \frac{\nu-1}{\nu}\frac{1}{2(\nu-1)}\sigma_{n-1} = \frac{1}{2\nu}\sigma_{n-1},$$

and hence by (8) and (11),

$$f_{n,\nu}(\mu_1^{-1}) \leq \mu_1^{-\nu} \leq (2\nu)^{\nu}\sigma_{n-1}^{-\nu} \leq (2\nu(c+1))^{\nu}\sigma_n^{-\nu} \leq c_\nu \sigma_n^{-\nu}.$$

On the other hand, if $\frac{1}{\alpha_n} > \frac{1}{2(\nu-1)}\sigma_{n-1}$, that is, if $\alpha_n < 2(\nu-1)\sigma_{n-1}^{-1}$, then by (12) and (11), (note that $n-1 > \nu-1$, by assumption),

$$\begin{aligned} f_{n,\nu}(\mu_1^{-1}) &= \mu_1^{-\nu}\frac{\alpha_n\mu_1}{1+\alpha_n\mu_1}\prod_{i=1}^{n-1}\frac{\alpha_i\mu_1}{1+\alpha_i\mu_1} \leq \alpha_n\mu_1^{-(\nu-1)}\prod_{i=1}^{n-1}\frac{\alpha_i\mu_1}{1+\alpha_i\mu_1} \\ &\leq \alpha_n c_{\nu-1}\sigma_{n-1}^{-(\nu-1)} \leq 2(\nu-1)(c+1)^{\nu}c_{\nu-1}\sigma_n^{-\nu}. \end{aligned}$$

Now, if $\nu > 2$, $c_{\nu-1} = (2(\nu-1)(c+1)^{\nu-1})^{\nu-1} \leq (2\nu(c+1)^{\nu})^{\nu-1}$ and hence $f_{n,\nu}(\mu_1^{-1}) \leq 2\nu(c+1)^{\nu}c_{\nu-1}\sigma_n^{-\nu} \leq c_\nu\sigma_n^{-\nu}$. While if $1 < \nu \leq 2$, then

$$\begin{aligned} f_{n,\nu}(\mu_1^{-1}) &\leq 2\nu(c+1)^{\nu}c_{\nu-1}\sigma_n^{-\nu} \leq 2\nu(c+1)^{\nu}(2\nu(c+1))^{\nu-1}\sigma_n^{-\nu} \\ &\leq 2\nu(c+1)^{\nu}(2\nu(c+1)^{\nu})^{\nu-1}\sigma_n^{-\nu} = c_\nu\sigma_n^{-\nu}.\Box \end{aligned}$$

The following theorem now follows directly from (6) and Lemma 3.

Theorem 1 *If $x^\dagger = (T^*T)^\nu w$ for some $\nu > 0$ with some $w \in \mathcal{D}((T^*T)^\nu)$, and if $\{\alpha_n\}$ satisfies (10), then $\|x_n - x^\dagger\| \leq c_\nu \sigma_n^{-\nu} \|w\|$.*

We note the crucial role that an estimate of the type

$$\max_{\lambda \in [0,\infty)} \lambda^\nu r_n(\lambda) = O(\sigma_n^{-\nu}) \quad \text{as} \quad n \to \infty \tag{13}$$

(Lemmas 2 and 3) played in establishing the convergence rate in Theorem 1. As (13) was established for $\nu \geq 1$ on the basis of condition (10), the question of the necessity of this condition naturally arises. The next theorem addresses this question.

Theorem 2 *If $\nu > 1$ and $f_{n,\nu}(\lambda) = \lambda^\nu r_n(\lambda) \leq c_\nu \sigma_n^{-\nu}$ for some $c_\nu > 0$ and all $n \in \mathbb{N}$, then $\{a_n\}$ satisfies condition (10) for some $c > 0$.*

Proof. Since $r_{n-1}(0) = 1$, $r'_{n-1}(0) = -\sigma_{n-1}$, and $r_{n-1}(\lambda)$ is convex for $\lambda \geq 0$, we have

$$r_{n-1}(\lambda) \geq 1 - \sigma_{n-1}\lambda, \quad \text{for} \quad \lambda \geq 0.$$

Therefore,

$$\begin{aligned} c_\nu \sigma_n^{-\nu} &\geq \lambda^\nu r_n(\lambda) = \alpha_n \frac{\lambda}{\alpha_n + \lambda} \lambda^{\nu-1} r_{n-1}(\lambda) \\ &\geq \frac{\alpha_n}{\alpha_n + \lambda} \lambda^\nu (1 - \sigma_{n-1}\lambda), \quad \text{for} \quad \lambda \geq 0. \end{aligned}$$

In particular, setting $\lambda = \frac{1}{2}\sigma_{n-1}^{-1}$, we find that

$$c_\nu \sigma_n^{-\nu} \geq \frac{\alpha_n \sigma_{n-1}}{1 + 2\alpha_n \sigma_{n-1}} \left(\frac{1}{2}\right)^\nu \sigma_{n-1}^{-\nu}$$

and therefore,

$$\frac{\alpha_n \sigma_{n-1}}{1+2\alpha_n \sigma_{n-1}} \le 2^\nu c_\nu \left(\frac{\sigma_{n-1}}{\sigma_n}\right)^\nu = 2^\nu c_\nu \left(\frac{\alpha_n \sigma_{n-1}}{1+\alpha_n \sigma_{n-1}}\right)^\nu .$$

It follows that

$$\frac{1}{2} < \frac{1+\alpha_n \sigma_{n-1}}{1+2\alpha_n \sigma_{n-1}} \le 2^\nu c_\nu \left(\frac{\alpha_n \sigma_{n-1}}{1+\alpha_n \sigma_{n-1}}\right)^{\nu-1} \le 2^\nu c_\nu (\alpha_n \sigma_{n-1})^{\nu-1}$$

and hence

$$\frac{1}{\alpha_n} \le c_\nu^{\frac{1}{\nu-1}} 2^{\frac{\nu+1}{\nu-1}} \sigma_{n-1},$$

that is, (10) holds. □

Finally, we remark that the "O" estimate of (13) cannot be improved to a "o" estimate. Indeed if

$$\max_{\lambda \in [0,\infty)} \lambda^\nu r_n(\lambda) = o(\sigma_n^{-\nu}), \qquad n \to \infty,$$

then since $1 - \sigma_n \lambda \le r_n(\lambda)$, we would obtain by setting $\lambda = \sigma_n^{-1}/2$,

$$0 < 2^{-\nu-1} \le \sigma_n^\nu \max_{\lambda \in [0,\infty)} \lambda^\nu r_n(\lambda) = o(1), \qquad n \to \infty,$$

which is a contradiction.

3 Perturbed Data: Stopping Criteria

From (5) we see that x_n may be expressed as

$$x_n = T^* q_n(TT^*) y$$

where $q_n(\lambda)$ is generated by:

$$q_n(\lambda) = (\lambda + \alpha_n)^{-1}(\alpha_n q_{n-1}(\lambda) + 1), \quad q_0(\lambda) = 0.$$

It follows that

$$1 - \lambda q_n(\lambda) = \frac{\alpha_n}{\lambda + \alpha_n}(1 - \lambda q_{n-1}(\lambda)),$$

that is, $q_n(\lambda) = \frac{1-r_n(\lambda)}{\lambda}$ where $r_n(\lambda) = \prod_{i=1}^{n} \frac{\alpha_i}{\lambda+\alpha_i}$. We now find that for $\lambda \geq 0$,

$$0 \leq \lambda q_n(\lambda) = 1 - r_n(\lambda) \leq 1, \tag{14}$$

and, by the convexity of r_n,

$$q_n(\lambda) = \frac{1 - r_n(\lambda)}{\lambda} \leq -r_n'(0) = \sigma_n.$$

Finally, since $q_n(\lambda) \to \sigma_n$ as $\lambda \to 0+$, we have

$$\max_{\lambda \in [0,\infty)} q_n(\lambda) = \sigma_n. \tag{15}$$

These estimates can now be used to derive a stability estimate for the approximations x_n. Suppose y^δ is an approximation to the data y with $\|y - y^\delta\| \leq \delta$. Let $\{x_n^\delta\}$ be the sequence generated by (5) using the data y^δ, i.e., $x_n^\delta = T^* q_n(TT^*) y^\delta$. Since $x_n, x_n^\delta \in \mathcal{D}(T)$, we have by (14) and (15),

$$\begin{aligned} \|x_n - x_n^\delta\|^2 &= (TT^* q_n(TT^*)(y - y^\delta), q_n(TT^*)(y - y^\delta)) \\ &\leq \delta^2 \sigma_n \end{aligned}$$

and hence

$$\|x_n - x_n^\delta\| \leq \delta \sigma_n^{1/2}. \tag{16}$$

A sufficient condition for regularity of the approximations is therefore that the iteration number be chosen in terms of the error level, say $n = n(\delta)$, so that the condition

$$\delta\sigma_{n(\delta)}^{1/2} \to 0 \text{ as } \delta \to 0$$

is satisfied (see [2] and [16] for somewhat different formulations of regularity conditions). From (14) we also obtain the stability result

$$\|Tx_n - Tx_n^\delta\| = \|TT^*q_n(TT^*)(y - y^\delta)\| \leq \delta. \tag{17}$$

Our goal in this section is to establish a convergence rate for an a posteriori stopping criterion for the iteration (5). The criterion is of discrepancy type and relies on monitoring the residual

$$y^\delta - Tx_n^\delta = r_n(TT^*)y^\delta. \tag{18}$$

We assume that $\sigma_n \to \infty$ (as is necessary for convergence [2]) and hence $r_n(\lambda) \to 0$ for every $\lambda > 0$. Therefore, by (18),

$$\lim_{n\to\infty} \|y^\delta - Tx_n^\delta\| = \|Py^\delta\| = \|P(y^\delta - y)\| \leq \delta$$

where P is the projector onto the orthogonal complement of the range of T. Finally we assume that the signal-to-noise ratio of the data is bounded above 1, that is, there is a number $\tau > 1$ such that $\|y^\delta\| > \delta\tau$. There is then a first value of n, say $n = n(\delta) \geq 1$, for which

$$\|y^\delta - Tx_{n(\delta)}^\delta\| \leq \delta\tau. \tag{19}$$

Lemma 4 *If $n(\delta)$ is chosen by (19), then*

$$\frac{\tau-1}{\tau+1}\|y - Tx_{n(\delta)}\| \le (\tau-1)\delta < \|y - Tx_{n(\delta)-1}\|.$$

Proof. Using (18) and the fact that $|r_n(\lambda)| \le 1$, we have

$$\begin{aligned} \|y - Tx_{n(\delta)}\| &= \|y^\delta - Tx^\delta_{n(\delta)} + r_{n(\delta)}(TT^*)(y - y^\delta)\| \\ &\le (\tau+1)\delta. \end{aligned}$$

On the other hand,

$$y - Tx_{n-1} = y^\delta - Tx^\delta_{n-1} - r_{n-1}(TT^*)(y^\delta - y)$$

and therefore

$$\|y - Tx_{n(\delta)-1}\| > \tau\delta - \delta = (\tau-1)\delta. \Box$$

To prove a convergence rate for the iterative method (5) with stopping criterion (19), we will use a specialized moment inequality ([3], [13]) which is proved for convenience in the next Lemma.

Lemma 5 *If $y \in \mathcal{D}((T^*T)^{\nu+1/2})$, for some $\nu \ge 0$, then*

$$\|(T^*T)^\nu y\| \le \|y\|^{\frac{1}{2\nu+1}} \left\|(T^*T)^{\nu+\frac{1}{2}} y\right\|^{\frac{2\nu}{2\nu+1}}.$$

Proof. Let $u = (T^*T)^\nu y$ and let $\{E_\lambda\}_{\lambda\ge 0}$ be a resolution of the identity generated by T^*T. Then

$$\int_0^\infty \lambda^{-2\nu} d\|E_\lambda u\|^2 = \int_0^\infty d\|E_\lambda y\|^2 = \|y\|^2 < \infty$$

and

$$\int_0^\infty \lambda d\|E_\lambda u\|^2 < \infty, \text{ since } u \in \mathcal{D}((T^*T)^{1/2}).$$

Therefore, by Hölder's inequality,

$$\begin{aligned} \|(T^*T)^\nu y\|^2 &= \|u\|^2 = \int_0^\infty (\lambda^{-2\nu})^{\frac{1}{2\nu+1}} (\lambda)^{\frac{2\nu}{2\nu+1}} d\|E_\lambda u\|^2 \\ &\leq \left(\int_0^\infty \lambda^{-2\nu} d\|E_\lambda u\|^2\right)^{\frac{1}{2\nu+1}} \left(\int_0^\infty \lambda d\|E_\lambda u\|^2\right)^{\frac{2\nu}{2\nu+1}} \\ &= \|y\|^{\frac{2}{2\nu+1}} \|(T^*T)^{\nu+1/2} y\|^{\frac{4\nu}{2\nu+1}}. \square \end{aligned}$$

The proof of the next theorem follows that of Vainikko [18] (see also [7]).

Theorem 3 *Let* $\{\alpha_n\} \subset \mathbb{R}^+$ *be a sequence of regularization parameters for which (10) holds. If* $x^\dagger \in R((T^*T)^\nu) \cap \mathcal{D}((T^*T)^{1/2})$ *and* $n(\delta)$ *is chosen as in (19), then* $\|x^\delta_{n(\delta)} - x^\dagger\| = O\left(\delta^{\frac{2\nu}{2\nu+1}}\right)$. *Moreover, at the stopping index we have* $\sigma_{n(\delta)} = O\left(\delta^{-\frac{2}{2\nu+1}}\right)$.

Proof. Suppose $x^\dagger = (T^*T)^\nu w$, where $w \in \mathcal{D}((T^*T)^{\nu+\frac{1}{2}})$. Then

$$x_n - x^\dagger = (T^*T)^\nu r_n(T^*T) w.$$

Using Lemma 5 on $y = r_n(T^*T)w$, we find

$$\begin{aligned} \|x_n - x^\dagger\| &\leq \|r_n(T^*T)w\|^{\frac{1}{2\nu+1}} \|(T^*T)^{\nu+1/2} r_n(T^*T)w\|^{\frac{2\nu}{2\nu+1}} \\ &\leq \|w\|^{\frac{1}{2\nu+1}} \|(T^*T)^{1/2}(x_n - x^\dagger)\|^{\frac{2\nu}{2\nu+1}}. \end{aligned}$$

Therefore, by Lemma 4,

$$\|x_{n(\delta)} - x^\dagger\| = O\left(\delta^{\frac{2\nu}{2\nu+1}}\right).$$

Now, by (16),

$$\|x^\delta_{n(\delta)} - x^\dagger\| \le \delta\sigma^{1/2}_{n(\delta)} + O\left(\delta^{\frac{2\nu}{2\nu+1}}\right)$$

and it is sufficient to show that $\sigma^{1/2}_{n(\delta)} = O\left(\delta^{-\frac{1}{2\nu+1}}\right)$. By Lemma 4,

$$\begin{aligned}(\tau-1)\delta &\le \|y - Tx_{n(\delta)-1}\| = \|r_{n(\delta)-1}(TT^*)y\| \\ &= \|T(T^*T)^\nu r_{n(\delta)-1}(T^*T)w\|.\end{aligned}$$

But, by Lemma 3 and condition (10),

$$\begin{aligned}&\|T(T^*T)^\nu r_{n-1}(T^*T)w\|^2 \\ &\quad = ((T^*T)^{\nu+1}r_{n-1}(T^*T)w, (T^*T)^\nu r_{n-1}(T^*T)w) \\ &\quad \le c_\nu c_{\nu+1}\sigma^{-2\nu-1}_{n-1}\|w\|^2 = O(\sigma^{-2\nu-1}_n).\end{aligned}$$

Therefore, $(\tau-1)\delta \le \text{const.} \times \sigma^{-\frac{2\nu+1}{2}}_{n(\delta)}$, and hence $\sigma^{1/2}_{n(\delta)} = O\left(\delta^{-\frac{1}{2\nu+1}}\right)$, giving the result. □

A method of Vainikko [19] can be adapted to show that the parameter strategy (19) is a regularizing scheme, i.e., $x^\delta_{n(\delta)} \to x^\dagger$ as $\delta \to 0$, without additional assumptions on $x^\dagger$. However, we note that a general result of Plato [14, Thm. 2.1] can also be extended to the case considered here to deduce the regularity of the scheme (19).

4 Example

We close with the aforementioned example of a geometric sequence of regularization parameters, i.e.,

$$\alpha_n = \alpha q^{n-1} \qquad \text{with fixed } \alpha > 0 \text{ and } 0 < q < 1^1 .$$

In this case we have

$$\sigma_n = \frac{1}{\alpha}\, q^{1-n} \frac{1-q^n}{1-q} \geq \frac{1}{\alpha}\, q^{1-n} = q\, \frac{1}{\alpha_{n+1}}\,,$$

so that (10) holds with $c = 1/q$. We can therefore apply Theorem 1 and obtain

$$\|x_n - x^\dagger\| = O(\sigma_n^{-\nu}) = O(q^{\nu n})\,,$$

i.e., a linear rate of convergence where the root convergence factor q^ν depends on the "smoothness" of the exact solution $x^\dagger$: The larger is ν, the faster is the convergence.

Concerning perturbed data we can employ the discrepancy principle (19) as a stopping rule, and we have

$$\|x^\delta_{n(\delta)} - x^\dagger\| = O(\delta^{\frac{2\nu}{2\nu+1}})$$

according to Theorem 3. Moreover, this theorem shows that at the stopping index $n(\delta)$ we have

$$\frac{q}{\alpha}\, q^{-n(\delta)} \leq \sigma_{n(\delta)} = O(\delta^{\frac{-2}{2\nu+1}})\,,$$

[1] As Robert Plato kindly pointed out to us, this special case of a geometric sequence of parameters α_n can actually be analyzed in a more sophisticated way.

which implies that at most

$$n(\delta) \leq O(|\log \delta|)$$

iterations are necessary to achieve this accuracy.

An efficient numerical implementation of nonstationary iterated Tikhonov regularization is not more expensive than using the same sequence of regularization parameters in a successive way for ordinary Tikhonov regularization. This follows from the fact that the major amount of work stems from the computation of a bidiagonalization of the discretized operator which has to be done in either approach; details are given in the survey [8]. However, as illustrated above, while the computational costs are the same, the convergence properties for the iterated Tikhonov scheme are much better.

References

[1] H. Brakhage, personal communication.

[2] M. Brill and E. Schock, Iterative solution of ill-posed problems – a survey, in "Model Optimization in Exploration Geophysics," (A. Vogel, Ed.), Vieweg, Braunschweig, 1987, pp. 13-38.

[3] R. Dautray and J.-L. Lions, Mathematical Analysis and Numerical Methods for Science and Technology, Vol. 3, Springer-Verlag, Berlin, 1990.

[4] A.G. Fakeev, A class of iterative processes for solving degenerate systems of linear algebraic equations, USSR. Comp. Math. Math. Phys. **21**,3 (1981), 15-22.

[5] C.W. Groetsch, The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind, Pitman, Boston, 1984.

[6] C.W. Groetsch, Inverse Problems in Mathematical Sciences, Vieweg, Braunschweig, 1993.

[7] M. Hanke, Accelerated Landweber iterations for the solution of ill-posed equations, Numer. Math. **60** (1991), 341-375.

[8] M. Hanke and P.C. Hansen, Regularization methods for large-scale problems, Surveys Math. Indust. **3** (1993), 253-315.

[9] B. Hofmann, Regularization for Applied Inverse and Ill-posed Problems, Teubner, Leipzig, 1986.

[10] J.T. King and D. Chillingworth, Approximation of generalized inverses by iterated regularization, Numer. Funct. Anal. Optim. **1** (1979), 499-513.

[11] A.V. Kryanev, An iterative method for solving incorrectly posed problem, USSR. Comp. Math. Math. Phys. **14**,1 (1974), 24-33.

[12] L.J. Lardy, A series representation for the generalized inverse of a closed linear operator, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur., Ser. VIII, **58** (1975), 152-157.

[13] A.K. Louis, Inverse und schlecht gestellte Probleme, Teubner, Stuttgart, 1989.

[14] R. Plato, Optimal algorithms for linear ill-posed problems yield regularization methods, Numer. Funct. Anal. Optim. **11** (1990), 111-118.

[15] F. Riesz and B. Sz.-Nagy, Functional Analysis, Ungar, New York, 1955.

[16] E. Schock, Implicite iterative methods for the approximate solution of ill-posed problems, Bolletino U.M.I., Series 1-B, **7** (1987), 1171-1184.

[17] A.N. Tikhonov and V.Y. Arsenin, Solutions of Ill-posed Problems, Winston, New York, 1977.

[18] G.M. Vainikko, The discrepency principle for a class of regularization methods, USSR. Comp. Math. Math. Phys. **22**,3 (1982), 1-19.

[19] G.M. Vainikko, The critical level of discrepancy in regularization methods, USSR. Comp. Math. Math. Phys. **23**,6 (1983), 1-9.

[20] Z. Xi, Iterated Tikhonov Regularization for Linear Ill-posed Problems, Dissertation, Universität Kaiserslautern, Kaiserslautern, 1993.

Fachbereich Mathematik	and	Department of Mathematical Sciences
Universität Kaiserslautern		University of Cincinnati
D-67653 Kaiserslautern		Cincinnati, OH 45221-0025
Germany		U.S.A.