Estimating Solution Smoothness and Data Noise with Tikhonov Regularization

Daniel Gerth and Ronny Ramlaub

ABSTRACT
A main drawback of classical Tikhonov regularization is that often the parameters required to apply theoretical results, e.g., the smoothness of the sought-after solution and the noise level, are unknown in practice. In this paper we investigate in new detail the residuals in Tikhonov regularization viewed as functions of the regularization parameter. We show that the residual carries, with some restrictions, the information on both the unknown solution and the noise level. By calculating approximate solutions for a large range of regularization parameters, we can extract both parameters from the residual given only one set of noisy data and the forward operator. The smoothness in the residual allows to revisit parameter choice rules and relate a-priori, a-posteriori, and heuristic rules in a novel way that blurs the lines between the classical division of the parameter choice rules. All results are accompanied by numerical experiments.

1. Introduction
In this paper we will revisit classical Tikhonov regularization. In this setting, we are interested in the solution of operator equations of the form

\[ y = Ax, \]

where \( A : X \to Y \) is a bounded, linear, and compact operator between Hilbert spaces \( X \) and \( Y \). Compactness implies that \( A \) has a non-closed range, \( \mathcal{R}(A) \neq \overline{\mathcal{R}(A)} \), and thus that (1.1) is ill-posed. While there are operators with non-closed range that are not compact (strictly singular operators), we confine ourselves here to compactness since it allows to use the singular system of \( A \) for analysis, and constitutes a natural limit for finite dimensional (and thus necessarily compact) approximations to \( A \) used in numerical computation. For notational convenience we will consider (1.1) to be scaled such that \( \|A\| = 1 \). Instead of the exact data \( y \) we have access...
only to a noisy datum $y^\delta$, for which we use the additive noise model
\[ y^\delta = y + \epsilon \text{ with } ||\epsilon|| = ||y-y^\delta|| = \delta \] (1.2)
for a (typically unknown) noise level $\delta > 0$. Tikhonov regularization approximates the unknown solution $x^\dagger$ to (1.1) by solving
\[ x^\delta_a = \arg \min \left\{ \frac{1}{2} ||Ax - y^\delta||^2 + \frac{a}{2} ||x||^2 \right\}, \] (1.3)
for $a > 0$. The optimal solution is known to be given as
\[ x^\delta_a = (A^*A + aI)^{-1}A^*y^\delta. \]

Later we will sometimes consider noise-free data. In this case, we drop the superscript and denote the approximate solutions by $x_a$. Since the minimization problem is easily solved, the main task is to find an appropriate value of the regularization parameter $a > 0$. If $a$ is too small, $x^\delta_a$ will be dominated by the amplification of the noise, whereas for $a$ too large the solutions will be too smooth and thus also too far from $x^\dagger$. Naturally, one would like to find the best possible approximation of $x^\dagger$, i.e., minimize the reconstruction error $||x^\delta_a - x^\dagger||$. This is quantified by convergence rates, which is a term for estimates of the form
\[ ||x^\delta_a - x^\dagger|| \leq \varphi(\delta), \quad 0 < \delta \leq \delta_0 \] (1.4)
with some index function $\varphi$, i.e., $\varphi : [0, \infty) \rightarrow \mathbb{R}_+$ is continuous and monotonically increasing with $\varphi(0) = 0$. With no further restriction on $x^\dagger$, no such $\varphi$ exists [1, Proposition 3.11]. A classical assumption on $x^\dagger$ is a source condition, postulating the existence of a parameter $\mu > 0$ such that
\[ x^\dagger \in \mathcal{R}((A^*A)^\mu). \] (1.5)

Using the source condition, one can show that the optimal choice for the regularization parameter, given $\mu$ and $\delta$, is
\[ a = c\delta^2_{\mu+1} \] (1.6)
with a suitable constant $c > 0$, and yields the convergence rate
\[ ||x^\delta_a - x^\dagger|| \leq C\delta^{2\mu}_{\mu+1} \] (1.7)
for $0 < \mu \leq 1$, where the exponent can not be reduced further. For more details on the convergence theory we refer to, e.g. [1, 2]. A common remark on the theory sketched above is that the underlying assumptions are often difficult to verify in practice. The parameter choice (1.6) and the estimate of the reconstruction error (1.7) require the knowledge of the smoothness parameter $\mu$ from (1.5) and the noise level $\delta$ (1.2). While there are statistical methods that can potentially estimate the noise level, the source condition requires the knowledge of the smoothness of $x^\dagger$, which is
generally unavailable. Further, the constants $c$ and $C$ in (1.6), (1.7) can be made explicit in theory (see, e.g. [2, Satz 3.4.3]), but require additionally the knowledge of $\|(A^*A)^{-\beta}x^\dagger\|$, which is unknown in practice.

The main result of the paper is that, in the absence of noise, the asymptotics of the residual of the Tikhonov-regularized approximation to $x^\dagger$ are equivalent to the source smoothness, namely, $\|(A(x^\dagger-x^\dagger))\| = O(x^\mu+\frac{1}{2})$ for $0 < \mu < \frac{1}{2}$ as $\alpha \to 0$ if and only if a variant of the source condition holds. By calculating approximations $x^\alpha$ for various $\alpha$, we can trace the residual curve and, provided $x^\dagger$ fulfills a source condition with $\mu < \frac{1}{2}$, extract the smoothness parameter by regression. We further show that once $\alpha$ is small enough such that the residual reaches the noise level, i.e., $\|Ax^\delta-y^\delta\| \approx \delta$, decreasing $\alpha$ further will not significantly change the residual, which in turn allows inferring the noise level $\delta$ from the residual curve. Both parameter estimations, for $\mu$ and $\delta$, can be carried out for a single given datum $y^\delta$, making it applicable for any practical measurement.

A method for determining the smoothness parameter $\mu$ was first demonstrated in [3], based on exploiting a Kurdyka-Łojasiewicz inequality implied by a source condition. However, the algorithm seemed unstable, and some numerical observations remained unexplained. By using Tikhonov regularization instead of the Landweber method, we can calculate approximate solutions $x^\alpha$ and the corresponding residuals for any regularization parameter $\alpha > 0$, instead of being restricted to the discrete iteration steps, which increases the accuracy significantly.

As a byproduct of the noise level estimation we find a novel parameter choice rule for the regularization parameter $\alpha$, which we compare with other established parameter choice rules. Using our main result on the connection between residual asymptotics and solution smoothness, we can shed new light on the relation of parameter choice rules. In particular, we show that if $x^\dagger$ satisfies a source condition with $0 < \mu < \frac{1}{2}$, the a-priori parameter choice (1.6), the discrepancy principle as an a-posteriori parameter choice rule, and two heuristic rules, the heuristic discrepancy principle and our new method, differ only in the constant. Hence, all four yield order-optimal convergence rates, and we can blur the line between the three categories of parameter choice rules.

Since the source condition is difficult to verify in practice, we set up a simple model problem, for which we know the source condition of $x^\dagger$. We use this model problem to illustrate our results.

**Model Problem:** Let $X = Y = \ell^2$, the space of square-summable sequences. We consider for $\beta > 0$ the operator $A : \ell^2 \to \ell^2$, $[Ax]_i = i^{-\beta}x_i$, $i \in \mathbb{N}$. This is a compact operator with $\sigma_i = i^{-\beta}$, $i \in \mathbb{N}$, where the singular functions $\nu_i$, $u_i$ are the unit vectors in $\ell^2$. We set our exact solution $x^\dagger = \{i^{-\eta}\}_{i=1}^\infty$ for some $\eta > \frac{1}{2}$. This yields $y = Ax^\dagger = \{i^{-\eta-\beta}\}_{i=1}^\infty$. Then, with $\mu^* = \frac{2\eta-1}{4\beta}$,
The remainder of the paper is structured as follows. Our main result is contained in Section 2, where we show a converse result connecting source condition and the residual. We proceed by studying the noise in the residual in Section 3. The method for the estimation of solution smoothness and noise is summarized in Section 4. The converse result holds only for Hölder-type source conditions. In Section 5 we discuss the case of higher and lower solution smoothness, which we show to be detectable in principle. A case study for real sets of tomographic data is presented in Section 6. Finally, we discuss parameter choice rules in Section 7.

2. Converse results

In [4], Neubauer showed that solution smoothness not only implies a convergence rate for the reconstruction error, but that also the reverse implication, often called a converse result, holds for Tikhonov regularization, both with noise-free and noisy data. The results were later generalized to other regularization approaches, see for example a generalization to Hilbert spaces [5, 6] or Banach spaces [7, 8], that all showed the equivalence of solution smoothness and convergence rates for the reconstruction error in their respective settings. We pursue here a different type of generalization. Instead of the relation between reconstruction error and solution smoothness alone, we consider the behavior of \( \| (A^*A)^{\nu} (x_\nu - x^\dagger) \| \) for parameters \( \nu \geq 0 \) as \( \nu \to 0 \) and its relation to solution smoothness. Instead of the formulation (1.5) for the source condition, we follow Neubauer [4] and use instead the condition

\[
\sum_{n=k}^{\infty} \langle x^\dagger, v_n \rangle^2 = O(\|x^\dagger\|^2 \sigma_k^{4\mu}), \tag{2.1}
\]

for \( k > k^* \), \( k^* \) sufficiently small, which implies

\[
x^\dagger \in \bigcap_{k < \mu} \mathcal{R}((A^*A)^k). \tag{2.2}
\]

Note that this is the setting of our model problem, see (1.8).

Remark 1. The condition \( k \geq k^* \) in (2.1) with sufficiently small \( k^* > 0 \) is needed in order for the theoretical results to be visible in the numerics.

Since \( A \) is compact, we can use its singular system \( \{ \sigma_i, u_i, v_i \}_{i=1}^{\infty} \) for the analysis. There, the functions \( \{ u_i \}_{i=1}^{\infty} \) form an ONB for \( \mathcal{R}(A) \), \( \{ v_i \}_{i=1}^{\infty} \) form an ONB for \( \mathcal{R}(A^*) \) and the singular values \( \{ \sigma_i \}_{i=1}^{\infty} \) accumulate at zero; provided \( \dim(\mathcal{R}(A)) = \infty \). We recall that the relations \( Av_i = \sigma_i u_i \) and \( A^*u_i = \sigma_i v_i \) hold for all \( i \in \mathbb{N} \). Any \( x \in X \) can be written as \( x = \sum_{i=1}^{\infty} \langle x, v_i \rangle v_i \) and
\[ Ax = \sum_{i=1}^{\infty} \langle Ax, u_i \rangle u_i = \sum_{i=1}^{\infty} \langle x, v_i \rangle Av_i. \]

The following lemma is the basis for our converse result.

**Lemma 2.** Let \( \{\sigma_i, v_i, u_i\}_{i=1}^{\infty} \) be the singular system to \( A \), \( x^t \in X \). Then

\[
\frac{1}{4} \|x^t\|^{q+4\mu} \leq \sum_{i=1}^{\infty} \sigma_i^q \lambda^2 (x^t, v_i)^2 \leq 2 \|x^t\|^{q+4\mu} \tag{2.3}
\]

for \( 0 < q + 4\mu < 2p \) if and only if \( x^t \) satisfies (2.1).

**Proof.** In principle the proof follows that of [4, Theorem 1]. We split

\[
\sum_{i=1}^{\infty} \frac{\sigma_i^q \lambda^2}{(\sigma_i^p + \lambda)^2} (x^t, v_i)^2 = \sum_{\sigma_i^p \leq \lambda} \frac{\sigma_i^q \lambda^2}{(\sigma_i^p + \lambda)^2} (x^t, v_i)^2 + \sum_{\sigma_i^p > \lambda} \frac{\sigma_i^q \lambda^2}{(\sigma_i^p + \lambda)^2} (x^t, v_i)^2. \tag{2.4}
\]

Consider first the small singular values. Noting that \( \frac{1}{4} \leq \frac{\lambda^2}{(\sigma_i^p + \lambda)^2} \leq 1 \) for \( \sigma_i^p < \lambda \), we have

\[
\sum_{\sigma_i^p < \lambda} \frac{\sigma_i^q \lambda^2}{(\sigma_i^p + \lambda)^2} (x^t, v_i)^2 \leq \frac{\lambda^2}{p} \sum_{\sigma_i^p < \lambda} (x^t, v_i)^2 = \mathcal{O}(\|x^t\|^2 \lambda^{q+4\mu}) = \mathcal{O}(\|x^t\|^2 \lambda^{q+4\mu}).
\]

For the term corresponding to the larger singular values, we use that \( \frac{1}{4} \leq \frac{\sigma_i^q}{(\sigma_i^p + \lambda)^2} < 1 \) for \( \lambda \leq \sigma \leq 1 \), which yields

\[
\frac{\lambda^2}{4} \sum_{\sigma_i^p > \lambda} \sigma_i^{q-2p} (x^t, v_i)^2 \leq \sum_{\sigma_i^p > \lambda} \sigma_i^{q-2p} \frac{\lambda^2}{(\sigma_i^p + \lambda)^2} (x^t, v_i)^2 < \lambda^2 \sum_{\sigma_i^p > \lambda} \sigma_i^{q-2p} (x^t, v_i)^2.
\]

It remains to show \( \lambda^2 \sum_{\sigma_i^p > \lambda} \sigma_i^{q-2p} (x^t, v_i)^2 = \mathcal{O}(\|x^t\|^2 \lambda^{q+4\mu}) \). Via induction (see Appendix) one finds with (2.1) that

\[
\sum_{i=1}^{k} \sigma_i^{q-2p} (x^t, v_i)^2 = \mathcal{O}(\sigma_k^{q+4\mu-2p}).
\]

Now we choose \( k \) such that \( \sigma_k^p = \mathcal{O}(\lambda) \). With \( q + 4\mu - 2p \leq 0 \) it follows

\[
\lambda^2 \sum_{\sigma_i^p > \lambda} \sigma_i^{q-2p} (x^t, v_i)^2 = \lambda^2 \mathcal{O}(\|x^t\|^2 \sigma_k^{q+4\mu-2p}) = \lambda^2 \mathcal{O}(\|x^t\|^2 \lambda^{q+4\mu-2}) = \mathcal{O}(\|x^t\|^2 \lambda^{q+4\mu}).
\]
This, together with the upper bound for the first summand in (2.4), yields the claim.

**Theorem 3.** Let \( 0 < \mu + \nu < 1 \). Then

\[
\|(A^*A)^\nu(x_a - x^\dagger)\|^2 = \mathcal{O}(\|x^\dagger\|^2 x^{2(\nu+\mu)})
\]

if and only if \( x^\dagger \) satisfies (2.1).

**Proof.** It is

\[
\|(A^*A)^\nu(x_a - x^\dagger)\|^2 = \sum_{i} \frac{\sigma_i^{4\nu} x^2 (x^\dagger, v_i)^2}{(\sigma_i^2 + \alpha)^2}. \]

Hence we apply Lemma 2 with \( q = 4\nu \) and \( p = 2 \).

Note that the proof of Lemma 2, similar to Neubauer’s original proof [4, Theorem 2.1], requires that the asymptotics for the source condition become relevant for small values of \( k \) already, optimally for \( k = 1 \). This is the reason we require \( k^* \) small in the definition of the source condition (2.1).

Neubauer [4], as well as Scherzer et al. [5] also provide converse results for noisy data. Since we assume that we only have one set of data \( y^\delta \) with fixed \( \delta \) available, we will not pursue convergence rates for noisy data further. Due to the additive noise model and Theorem 3, the equivalence between (2.1) and order optimal convergence rate would be no surprise.

Theorem 3 states that the solution smoothness in Tikhonov regularization is preserved under the application of certain powers of \( A^*A \). Hence, the smoothness of the powers \( (A^*A)^\nu(x_a - x^\dagger) \) can be used to assess the solution smoothness. In practice, most values of \( \nu \) are still not observable. However, for \( \nu = \frac{1}{2} \) we find the residual \( \|(A^*A)^{\frac{1}{2}}(x_a - x^\dagger)\| = \|A(x_a - x^\dagger)\| \), and for \( \nu = 1 \) we obtain the gradient. For Tikhonov regularization, all solution smoothness is lost in the gradient, since \( \nu + \mu < 1 \) is a requirement of Theorem 3. The loss of information in the gradient can also be seen from the first order condition

\[
A^*A(x_a - x^\dagger) = -\alpha x_a,
\]

which enforces \( \|A^*A(x_a - x^\dagger)\| = \alpha \|x_a\| \) unconditionally. Because \( x_a \to x^\dagger \) as \( \alpha \to 0 \) (note that we are in the noise-free scenario), \( \|x_a\| \to \|x^\dagger\| \) and hence \( \|A^*A(x_a - x^\dagger)\| \sim \alpha \).

The residual, obtained with \( \nu = \frac{1}{2} \), on the other hand, contains smoothness information. Since Theorem 3 requires \( \mu + \nu < 1 \), smoothness with \( \mu < \frac{1}{2} \) is preserved and hence can be detected. It is no coincidence that this matches with the well-known fact that the discrepancy principle for the
choice of $x$ yields order-optimal convergence rates for $0 < \mu < \frac{1}{2}$, as discussed in more detail in [9].

Another interesting observation is that the saturation of Tikhonov regularization, i.e., the best obtainable convergence rate is $||x^\delta - x^\dagger|| = \mathcal{O}(\delta^\beta)$ for $\mu \geq 1$, is due to the effect that low-frequency components of the solutions cannot be approximated well: The saturation follows from the condition $\mu + \nu < 1$ in Theorem 3, which we have used solely to evaluate the low frequencies $\sigma_i^2 > x$.

### 3. Data noise

So far we have not considered noise in the data. Traditionally, the analysis is focused on the propagation of the noise to the reconstruction error. Instead, here we focus again on the residuals.

To model the noise we will assume a source condition similar to (2.1). Take $\epsilon$ from (1.2) and assume that there is $0 \leq \kappa \leq \frac{1}{2}$ such that

$$\sum_{n=k}^{\infty} \langle \epsilon, u_i \rangle^2 = \mathcal{O}(\delta^2 \sigma_k^{4\kappa})$$

for $k \to \infty$. In the finite dimensional setting we approximate this by requiring that the condition holds for the first $k^*$ singular vectors, i.e.,

$$\sum_{n=k}^{N} \langle \epsilon, u_i \rangle^2 = \mathcal{O}(\delta^2 \sigma_k^{4\kappa})$$

for $k > k^* > N$ where $N$ is the discretization level and $k^*$ sufficiently large.

We remark here that noise modeling on the interface of infinite dimensional theory and finite dimensional numerical calculations is not trivial. For example, the most common discrete model is a Gaussian one, but a rigorously defined Gaussian random variable $g$ in a Hilbert space $X$ will have $||g||_X = \infty$ almost surely, see, e.g., [10]. Therefore, we must make some compromises and allow for some inaccuracies. If $N$ is large enough, then $\kappa = 0$ in (3.2) yields discrete Gaussian noise, and (3.1) somewhat generalizes it to the infinite dimensional setting. For $\kappa > 0$ we obtain noise with a certain decay rate related to the decay of the singular values of $A$. A similar assumption was made, e.g., in [11], where a noise condition $\langle \epsilon, u_i \rangle^2 \geq \delta^2 \sigma_i^{4-\beta}$, $\beta > 1$, was used. If, for example, $A$ is mildly ill-posed, i.e., $\sigma_i \approx i^{-\beta}$ with some $\beta > 0$, then (3.1) is satisfied for $\kappa = \frac{\beta-1}{4\beta} > 0$. It can be shown using the techniques of Lemma 2 that the noise condition (3.1) fulfills the Muckenhoupt-type noise condition of [11], which here reads

$$\exists C > 0 : \forall k \geq 1 \quad \sigma_k^1 \sum_{i=1}^{k} \sigma_i^{-2} \langle \epsilon, u_i \rangle^2 \leq C \sum_{i=k+1}^{\infty} \sigma_i^2 \langle \epsilon, u_i \rangle^2$$

(3.3)
for $\kappa \leq \frac{1}{2}$. As we will discuss below, the failure of condition (3.3) for $\kappa > \frac{1}{2}$ coincides with the observation that then the noise is no longer clearly separated from the noise-free data in the inverse problem. We also remark that one may alter the definition (3.1) to cover weakly bounded noise $\| (A^*A)^{\rho} (y - y^\delta) \| \leq \delta < \infty$ for some $0 < \rho < \frac{1}{2}$ as in [12] by assuming $\sum_{n=k}^{\infty} \sigma_n^2 \langle \epsilon, u_i \rangle^2 = O(\delta^2 \sigma_n^{4\kappa})$ with appropriate $\kappa$, but in this case the estimation of the noise level is not feasible as our method relies on finiteness in the $Y$-norm.

With these noise assumptions we can now estimate the behavior of the residuals in case of noisy data. Note that we have

$$Ax^\delta - y^\delta = \sum_{i=1}^{\infty} \left( \frac{\sigma_i^2}{\sigma_i^2 + \alpha} - 1 \right) \langle y^\delta, u_i \rangle u_i = \sum_{i=1}^{\infty} \frac{\alpha}{\sigma_i^2 + \alpha} \langle y, u_i \rangle u_i + \sum_{i=1}^{\infty} \frac{\alpha}{\sigma_i^2 + \alpha} \langle \epsilon, u_i \rangle u_i. \quad (3.4)$$

The first, noise-free, sum is evaluated in Theorem 2. For the second one, we have the following result as consequence of Lemma 2:

**Lemma 4.** It is

$$\sum_{i=1}^{\infty} \frac{\alpha^2}{(\sigma_i^2 + \alpha)^2} \langle \epsilon, u_i \rangle^2 = O(\delta^2 \alpha^{2\kappa}) \quad (3.5)$$

if and only if $\epsilon$ from (1.2) satisfies (3.1) with some $0 < \kappa < 1$.

**Proof.** We apply Lemma 2 with $q = 0$ and $p = 2$. \qed

Lemma 4 shows that the noise itself has a certain regularity when $\alpha$ is reduced. This effect can be used to separate the residuals $\| Ax^\delta - y^\delta \|$ into a “noise-free” and a “noisy” part as we describe in the following.

Employing the reverse triangle inequality for the residuals yields

$$\| || Ax^\delta - y^\delta || - || y - y^\delta || \| \leq || Ax^\delta - y^\delta ||,$$

i.e., whenever $|| y^\delta || > || Ax^\delta - y^\delta || \gg || y - y^\delta || = \delta$ then the residuals are dominated by the approximation of the noise-free data, and we have from Theorem 3 with $\nu = \frac{1}{2}$ that

$$|| Ax^\delta - y^\delta || = O(\| x^\dagger \| \alpha^{\mu+\frac{1}{2}})$$

under the source condition (2.1) with $0 < \mu < \frac{1}{2}$. Since the smoothness-parameter $\mu$ is in the exponent of the residuals, we can infer it from a regression of residuals as function of several regularization parameters $\alpha$. This is described in detail in Section 4 below. As $\alpha$ is reduced further and consequently the residuals decrease, the noise will become dominant if $\kappa$ is not
too large. Namely, due to (3.4) and Lemma 4 we have

$$||Ax^\delta - y^\delta|| = O(\delta x^\kappa)$$

if $\kappa < \mu + \frac{1}{2}$, $\delta$ is not too large and $\alpha$ is sufficiently small such that the curves $O(\delta x^\kappa)$ and $O(||x^\delta||^\alpha x^{\alpha+\frac{1}{2}})$ have intersected. If $\kappa \leq \frac{1}{2}$ and we observe $||Ax^\delta - y^\delta|| = O(x^\kappa)$, then we can be sure that we observe the noise since the noise-free residual, for any $\mu > 0$ in the source condition (2.1), falls as $o(x^{\frac{1}{2}})$ (3.6). For $\frac{1}{2} \leq \kappa < \mu + \frac{1}{2}$, $0 < \mu < \frac{1}{2}$, the noise still dominates the residuals $||Ax^\delta - y^\delta||$ if $\alpha$ is sufficiently small, but the rate $O(x^\kappa)$ could as well be explained through data $y = Ax$, where $\tilde{x}$ satisfies a source condition (2.1) with $\tilde{\mu} = \kappa - \frac{1}{2}$, so one can no longer with absolute certainty distinguish between noise and noise-free data. For $\kappa > \frac{1}{2}$, the residuals are dominated by the approximation $||Ax^\delta - y^\delta|| \approx ||Ax - y||$ for all $\alpha > 0$ and the problem behaves essentially noise-free.

Before we show a numerical example for the residual curve $||Ax^\delta - y^\delta||$, we remark that numerically, (3.2) will not be a proper characterization of the smoothness any more for very small $\alpha$, and the residuals will decrease quickly, since for $\alpha \ll \sigma_N$, $\sigma_N$ being the smallest discrete singular value,

$$||Ax^\delta - y^\delta||^2 = \sum_{i=1}^{N} \frac{\alpha^2}{(\sigma^2 + \alpha)^2} \langle y^\delta, u_i \rangle^2 \leq \alpha^2 \frac{1}{(\sigma_N^2 + \alpha)^2} \sum_{i=1}^{N} \langle y^\delta, u_i \rangle^2 = O(\alpha^2).$$

(3.7)

If the regularization parameter is that small, then the problem is essentially unregularized and the reconstruction error will be dominated by the noise.

In a last remark on data noise we mention that in practical applications one will often encounter another form of noise: the modeling error. The practical measurement setups are to some extent idealized in the mathematical model. For example ray sources and detectors are modeled as points, although they cover a small area in practice. Also their position cannot be measured to arbitrary precision, and other physical effects may be neglected, e.g., scattering. That means a practical measurement $y^{meas}$ will not coincide with $y^\delta$ as expected from the modeling. Even more, the case $y^{meas} \notin \overline{R(A)}$ is to be expected. The resulting difference $\delta^{meas} := ||y^{meas} - y^\delta||$ thus can not be explained by the model and remains as a limit of the residual, $||Ax^\delta - y^{meas}|| \geq \delta^{meas}$ for all $\alpha \to 0$. In Section 6 we illustrate this numerically.

4. Estimation of smoothness parameter and noise level

We now summarize our results to estimate solution smoothness and noise level using Tikhonov regularization. We will focus on noise satisfying (3.1)
with $\kappa = 0$ as this allows the estimation of the correct noise level $\delta$ in (1.2). We briefly comment on the case $\kappa > 0$ at the end of the section.

The method for the estimation of the source condition is based on Theorem 3. As consequence of the condition $\nu + \mu < 1$ (and using $\nu = \frac{1}{2}$) we can only identify the smoothness parameter from the source condition if $x^\dagger \in \mathcal{R}((A^*A)^\mu)$ with $\mu < \frac{1}{2}$. The method works as follows. We employ Tikhonov regularization for several magnitudes of regularization parameters, and store the residuals

$$r(\alpha) := ||Ax_\alpha^\delta - y^\delta||.$$

We also compute the derivative

$$dr(\alpha) := \frac{\partial}{\partial \log(\alpha)} \log r(\alpha).$$

Both curves as function of $\alpha$ can be characterized in different intervals or stages, that can be used to estimate the smoothness parameter $\mu$ and the point to extract an estimate of the noise level. An example for the Model Problem with $\eta = \beta = 2$ and $\delta = 0.005||y||$ is given in Figure 2.

When the regularization parameters are too large, we have $||x_\alpha^\delta|| \approx 0$, hence $r(\alpha) \approx ||y^\delta||$ and $dr(\alpha) \approx 0$. Lowering $\alpha$ and going through a transition stage, we arrive at the approximation phase. Due to (3.6) it is then $r(\alpha) \approx ||Ax_\alpha - y||$. The residual is dominated by the approximation of the exact data and the noise has little impact on the residual, which therefore carries the information on solution smoothness. Theorem 3 with $\nu = \frac{1}{2}$ yields $r(\alpha) \sim x^{\frac{\mu}{2} + \frac{1}{2}}$ if and only if (2.1). Hence, we can make a regression for the ansatz $r(\alpha) = cx^\kappa$. If the regression yields a good fit to the residual curve with $\frac{1}{2} < \kappa < 1$, we can extract the solution smoothness $x^\dagger \in \mathcal{R}((A^*A)^\mu)$ with $\mu^* = \kappa - \frac{1}{2}$. In the derivative $dr(\alpha)$, we can immediately read off $\mu$: if $dr(\alpha) = \kappa$ for some $\frac{1}{2} < \kappa < 1$, then $x^\dagger$ must satisfy the source condition (2.1) with $\mu = \kappa - \frac{1}{2}$. If the regression yields $r(\alpha) \sim \alpha$, then one can conclude that $x^\dagger$ must be smoother than a source condition (2.1) with $\mu = \frac{1}{2}$ implies, but due to the restriction $\mu + \nu < 1$ in Theorem 3 one can no longer estimate $\mu$. One may even have, for example, an exponential source condition as we demonstrate in Section 5 below.

After another transition, we are in the noise phase, where $r(\alpha) \approx \delta$, namely, the middle term in (3.4) still behaves as $||Ax_\alpha - y|| = \mathcal{O}(x^{\mu + \frac{1}{2}})$ if (2.1) holds, whereas the right-hand term, according to Lemma 4, is of order $\mathcal{O}(\delta^2)$, such that in total $||Ax_\alpha^\delta - y^\delta|| \approx \delta$ for a wide range of parameters $\alpha$. The noise level can therefore be read off the flat part of the residual, cf. Figures 1 and 3. From the figures we also see that the residual curve has almost a saddle point in the flat plateau, which we can use to find it and estimate the noise level algorithmically. We look for
\[ x^* = \arg \min \left( dr(x) = \arg \min \frac{\partial}{\partial \log(x)} \log (||Ax^\delta - y^\delta||) \right) \]  

and estimate \( \delta \approx ||Ax^\delta - y^\delta||. \) We discuss (4.1) as a rule to choose the regularization parameter \( x \) in Section 7.

In theory, the noise phase is then active for all \( x \to 0 \) and, depending on the decay of the noise components \( \langle y - y^\delta, u_i \rangle, ||Ax^\delta - y^\delta|| \) goes to zero slowly. In practice, due to discretization, we observe one or two more stages as \( x \) is decreased further. As noted in Section 3, we can expect to see \( r(x) \approx x \) and thus \( dr(x) \approx 1 \) for sufficiently small \( x \). It might happen that, when reducing the regularization parameter even further, the residuals become chaotic, likely due to numerical errors such as round-off errors and the amplification thereof due to the ill-posed nature of the problem.

We finally note that in order for this estimation to work we require that the noise level be not too high, i.e., \( \delta \ll ||y^\delta|| \). If it is, then the approximation stage is too short or even non-existent, such that the solution smoothness is completely hidden in the noise. It is not clear how much smaller \( \delta \) has to be compared to \( ||y^\delta|| \). One should be able to clearly see the approximation stage (see Figure 2), for example by looking for the plateau in the first derivative as shown in the second image of Figure 2.
The algorithm for the estimation of $\mu$ and $\delta$ is summarized below in Algorithm 1. Note that the closer $q$ is to one, the clearer the expected results are.

**Algorithm 1.** Algorithm for the estimation of $\mu$ and $\delta$.

Input: $A$, $y^\delta$, $z_0 > 0$ such that $r(z_0) \approx ||y^\delta||$, $0 < q < 1$, $N \in \mathbb{N}$

for $i = 1, 2, ..., N$ do

  set $z_i = z_0 q^{i-1}$
  calculate $x^\delta_{z_i} = (A^* A + z_i I)^{-1} A^* y^\delta$
  store $r(z_i) = ||Ax^\delta_{z_i} - y^\delta||$
  calculate $dr(z_i) \approx \frac{\log(r(z_i)) - \log(r(z_{i-1}))}{\log(z_i) - \log(z_{i-1})}$

end for

If $dr(z_i) \approx \kappa$ with $\frac{1}{2} < \kappa < \mu$ for sufficiently many $z_i$, estimate $\mu \approx \kappa - \frac{1}{2}$.

find $x^* = \arg\min dr(z_i)$, estimate $\delta \approx ||Ax^\delta_{x^*} - y^\delta||$

For smoother noise, i.e., $\kappa > 0$ in (3.1), one will see similar stages as for the Gaussian case, but in the noise stage the residuals will not be constant. Instead we will see a behavior $||Ax^\delta_{x^*} - y^\delta|| \sim \delta x^\kappa$ (see Lemma 4). Assume that we observe $||Ax^\delta_{x^*} - y^\delta|| \sim \delta x^\kappa$ for $0 < \kappa < \mu + \frac{1}{2}$ and $z \leq z^*$, then we can estimate $\delta \approx \frac{||Ax^\delta_{x^*} - y^\delta||}{\sqrt{2x^\kappa}}$, i.e., due to the decay of the noise we will not observe $\delta$, but only a fraction of it which has to be corrected for. Indeed, in the case of decaying noise ($\kappa > 0$), we will show in a section below that the effective noise level can drop below $\delta$, which leads to improved convergence rates.

5. Low smoothness and high smoothness

The estimation of the solution smoothness works best when the classical Hölder-type source conditions (1.5) or (2.2) describe the smoothness of $x^\dagger$, as Theorem 3 can be applied. More general, for each $x^\dagger \in X$ there exists an index function $\phi$ and $w \in X$ such that

$$x^\dagger = \phi(A^* A)w,$$

see [13]. We distinguish two cases, depending on whether $\phi(t)$ decays slower or faster to zero than the Hölder-type functions $t^\kappa$, $0 < \kappa < \frac{1}{2}$. In analogy to (2.1), we consider the generalized source conditions of the form

$$\sum_{n=k}^{\infty} \langle x^\dagger, v_n \rangle^2 = O(\phi(\sigma_k^2)^2).$$

We have seen in Theorem 3 that the residual carries no solution smoothness information when $x^\dagger$ fulfills a source condition (2.2) with $\mu \geq \frac{1}{2}$. This
remains the case for functions smoother than the Hölder powers, as then asymptotically $\|Ax_\alpha - Ax^\dagger\| = O(\alpha)$. A numerical example with an exponential source condition $\varphi(t) = \exp(-t^\gamma)$ with $\gamma = 2$ is shown in Figure 3. In absence of noise the residual curve is, for larger regularization parameters, concave in the log-log plot but for small enough $\alpha$ it is of order $\alpha$ as expected. For noisy data, the latter phase is completely masked by the noise. Still, one can easily spot the noise level. The visible part of the residual curve for larger $\alpha$ (i.e., the approximation phase) corresponds to
the concave part of the noise-free residual, which appears to be the indicator for the high smoothness case.

The opposite appears to be the case in the low smoothness setting, i.e. when $\varphi$ in (5.1) decays slower than a power function, where concavity...
appears to be the indicator. To demonstrate this, we consider a generalized source condition (5.2) with \( u(t) = \frac{\log(t)}{C_0} \) for \( c = 1.5 \), see Figure 4.

It was shown in [14] that \( \|x_\mathbf{a} - x^\dagger\| \leq C(\log(\delta)^{-c}) \). In this case one can further show that

\[
\|Ax_\mathbf{a} - Ax^\dagger\| \leq C\sqrt{a(\log(a))^{-c}}
\]

(5.3)

for small enough \( a \). In the noise free case, this is plausible in the experiment. For noisy data this asymptotic is masked by the noise. What remains to extract information are the large regularization parameters. As usual, when \( a \) is too large, the residual changes little and then starts to drop. In the low smoothness case the residual curve is convex after the initial drop in the log-log-plot. This behavior is implied by the upper bound (7.3), since

\[
\log(\sqrt{a(\log(a))^{-c}}) = \frac{1}{2} \log(\log(a)) - \gamma \log(\log(a)).
\]

Substituting \( x := \log(a) \), we have \( R(x) = \frac{1}{2} x - \gamma \log(x) \), and differentiating twice yields \( \frac{d^2}{dx^2} R(x) = \frac{\gamma}{x^2} > 0 \) for all \( x > 0 \), i.e., \( r(x) \) is convex in the log-log-plot. This can also be seen in Figure 4.

The above observations indicate that it is possible to at least detect solution smoothness lower and higher than the Hölder type source condition (2.2) with \( 0 < \mu < \frac{1}{2} \). In practice, however, this is difficult. Due to noise one has, in general, no access to the approximation rate \( \|Ax_\mathbf{a} - y\| \) for sufficiently small \( a \) where the asymptotic rates become visible. Instead one will often be restricted to large regularization parameters, where a power-type
regression often almost holds. Here one must find the minuscule differences and carefully inspect the deviation of the residual curve from the regression curve. This is exemplified in Figure 5. With this observation we can also understand why high and low smoothness are difficult to handle in practice. In the range of regularization parameters one would expect in practice they may behave almost like power-type source conditions.

Figure 6. Residuals and its derivative for the large Lotus dataset. Except for the smallest $\alpha$, the plots have much similarity with the ones from simulated data, cf. Figure 2, and we can observe the same phases. In particular, for $\alpha$ between approximately 0.01 and 0.5, we have the approximation phase that can be used to assess the solution smoothness. We zoom into this area in Figure 7. For the smallest $\alpha$, we no longer have drop in residual but it stays constant, likely due to modeling errors. However, for $\alpha \approx 10^{-3}$ we observe the saddle point in the residual characteristic for the noise level. This residual with value 0.04, yields the noise level $\delta \approx 0.04$, or, together with $\|y^0\| = 1.84$ and $\frac{\delta}{\|y^0\|} = \frac{0.04}{1.84 - 0.04} = 0.022$, an estimated noise level of 2.2%.

regression often almost holds. Here one must find the minuscule differences and carefully inspect the deviation of the residual curve from the regression curve. This is exemplified in Figure 5. With this observation we can also understand why high and low smoothness are difficult to handle in practice. In the range of regularization parameters one would expect in practice they may behave almost like power-type source conditions.
6. Experiments on tomographic data

Moving away from the simulated data, we now apply the method to real data sets. This had been already done in [3] using the Landweber algorithm to compute approximate solutions, but the results were difficult to interpret. We now apply our proposed Tikhonov-regularization approach.

We use two samples from the tomographic X-ray data set collection provided by the Finnish Inverse Problems Society (FIPS), namely the data of a stuffed lotus root [15] and the walnut data [16].

As for simulated data, we compute the approximate solutions by solving

\[(A^T A + \alpha I)x = A^T y^\delta\]

for various values of \(\alpha\). The normal equation is solved using a Conjugate Gradient method. In particular we make use of the structure by interpreting \(\alpha I\) as shifts of the matrix \(A^T A\). This way only one Krylov subspace has to be built independent of the shifts \(\alpha\), and for each fixed \(\alpha\) the computations are cheap. This algorithm, which is described in detail in [17, Algorithm 6], allows to use our method for high-dimensional problems and a large range of regularization parameters at low computational cost.

As our first example we consider the Lotus data set. For the largest data set LotusData256.mat, consisting of a matrix \(A_{mn} \in \mathbb{R}^{51480 \times 65536}\) and measurements \(y^\delta \in \mathbb{R}^{429 \times 120}\). The results are shown in Figure 6. We find that, as for simulated data, for large regularization parameters the residual is constant. In a second phase we find \(\|Ax^\delta - y^\delta\| \sim \alpha^{0.615}\) with the regression approach, see Figure 7 for a zoomed-in plot. This implies that \(x^\dagger\) fulfills a source-condition (1.5) with \(\mu \approx 0.115\). Reducing \(\alpha\) further yields a saddle-point like structure in the residual, from which we estimate the noise level.
\( \delta \approx 0.04 \) and a relative noise level of approximately 2.2%. After that, the residual drops slightly before it remains constant for all smaller regularization parameters, which we attribute to the modeling error.

We repeat the experiment for the largest data set of the walnut, see Figures 8 and 9. In the approximation phase we find through regression that \( ||Ax^\delta - y^\delta|| \approx \alpha^{0.52} \), which would correspond to a source smoothness with \( \mu = 0.02 \). However, we see that the residual curve oscillates around the regression line, which would not be the case if a Hölder source
condition would hold. Hence, we conclude that the walnut data does not fulfill a H"older-type source condition (2.2) for any $\mu > 0$.

7. Choice of the regularization parameter

A main task for the regularization of inverse problems in general and Tikhonov regularization in particular is the choice of the regularization parameter. Over time, so many parameter choice rules have been proposed such that it is not easy to keep a full overview. This is also not the purpose of this section. We will, on one hand, return to the optimization task (4.1),

Figure 9. Residuals for the large walnut dataset in the approximation phase, zoomed in from Figure 8. While the regression suggests a value $\mu \approx 0.02$, it can be seen that the residual oscillates around the regression lines, indicating the the source condition (1.5) does not hold for the walnut data.

Figure 10. Comparison of the RDM functional with the reconstruction error and the heuristic discrepancy principle.
which we had introduced to estimate the noise level, and view it as a parameter choice rule. On the other hand, we will discuss the consequences of Theorem 3 for analysis of parameter choice rules in general. For the most part, we will again only consider noise with $\kappa = 0$ in (3.1). Only at the end of the section we will show how $\kappa > 0$ allows for parameter choice rules that yield improved convergence rates.

### 7.1. Relating parameter choice rules

Parameter choice rules are usually divided into three subgroups: a-priori choices $\alpha = \alpha(\delta)$, a-posteriori choices $\alpha = \alpha(\delta, y^\delta)$, and heuristics $\alpha = \alpha(y^\delta)$. A-priori choices, for Tikhonov regularization (1.6), are often regarded as the theoretical optimum that is infeasible in practice due to the lack of the necessary parameters $\mu$ and $\delta$. We have described above how both can be estimated, such that the a-priori choice can be carried out. However, as we have also noted, in particular the estimation of $\mu$ can be difficult. Therefore, we discuss in the following the impact of Theorem 3 on other parameter choice rules.

The (Morozov) discrepancy principle is the most prominent a-posteriori principle. It only requires the knowledge of $\delta$ to select

$$\alpha^* = \sup \{ \alpha : \| Ax_\alpha^\delta - y^\delta \| \leq \tau \delta \}$$

for some $\tau > 1$. Of course, this requires the knowledge of $\delta$. Technically one also requires $\mu < \frac{1}{2}$ in the source conditions (2.1) due to the restrictions of the residual as shown in Theorem 3. Heuristic parameter choice rules are often the only alternative in practical situations when neither $\mu$ nor the noise level are known. Naturally, most of these use the residual in some way or the other. Examples are the heuristic discrepancy principle, see, e.g., [1], where the functional

$$f(\alpha) = \frac{\| Ax_\alpha^\delta - y^\delta \|^2}{\alpha}$$

is minimized, or the L-curve [18], and for the sake of brevity we will not discuss other methods.

To discuss the relation between the parameter choice rules, we start again with the a-priori choice (1.6), which is superior in the sense that it is applicable for $0 < \mu \leq 1$ w.r.t a source condition (1.5) for $x^\dagger$. Since the residual is non-informative for the interval $\frac{1}{2} < \mu < 1$ we consider only $0 < \mu < \frac{1}{2}$ in the following. The a-priori choice is due to the decomposition

$$\| x_\alpha^\delta - x^\dagger \| ^2 \leq \| x_\alpha - x^\dagger \| ^2 + \| x_\alpha^\delta - x_\alpha \| ^2$$

$$\leq c \alpha^{2\mu} + \frac{\delta^2}{\alpha},$$
see, e.g., [1], and (1.6) is obtained by optimizing over \( x \), i.e., minimizing

\[
 f_{ap}(x) = c x^2 + \frac{\delta^2}{x}.
\]

Using the analogous decomposition for the residual, we have

\[
 ||Ax^\delta-y^\delta||^2 \leq ||Ax_\mu-Ax^\tau||^2 + ||Ax^\delta-Ax_\mu||^2 + ||Ax^\tau-y^\delta||^2 \leq \tilde{c} x^{2\mu + 1} + \delta^2 + \delta^2,
\]

where for the middle term the estimate

\[
 ||Ax^\delta-Ax_\mu||^2 = ||A(A^*A + \alpha I)^{-1}A^*(y-y^\delta)|| \leq ||A(A^*A + \alpha I)^{-1}A^*|| \ ||y-y^\delta|| \leq \delta
\]

is used. Inserting this into the heuristic discrepancy principle (7.1) yields

\[
 f(x) = \frac{||Ax^\delta-y^\delta||^2}{x} \leq \tilde{c} x^{2\mu} + 2 \frac{\delta^2}{x},
\]

i.e., up to constants this is identical to the a-priori functional (7.2). Hence, asymptotically they yield the same convergence rate, and hence the heuristic discrepancy principle yields the optimal convergence rate \( ||x^\delta-x^\tau|| \leq c \delta^{2\mu+1} \) whenever \( 0 < \mu < \frac{1}{2} \). Of course this convergence result is well-known (see, e.g., [1]), but this proof is simpler than the standard argument.

For Morozov’s discrepancy principle we use, for simplicity, the slightly different version to choose \( \alpha \) such that \( ||Ax^\delta-Ax^\tau|| = \tau \delta \) for \( \tau > 1 \). Considering that when \( ||Ax^\delta-y^\delta|| > \delta \) we have shown that

\[
 ||Ax^\delta-y^\delta|| \approx ||Ax_\mu-Ax^\tau|| = O(\alpha^{\mu+\frac{1}{2}}),
\]

the discrepancy principle corresponds approximately to solving

\[
 \tau \delta = ||Ax_\mu-Ax^\tau|| = \tilde{c} x^{\mu+\frac{1}{2}}
\]

with solution \( \alpha^* = c(\tau, \mu) \delta^{\frac{2}{2\mu+1}} \). This is, up to constants, the solution to the optimization problems for the a-priori choice (7.2) and the heuristic discrepancy principle (7.1). Therefore, all 3 methods differ only in constants and yield the same convergence rate. However, due to the differing constants, they do not yield the same regularization parameters in practice.

We now discuss the L-curve as one of the most popular heuristic parameter choice rules. For the L-curve method Tikhonov-approximations \( x^\delta \) are calculated for many \( \alpha \), similar to Algorithm 1. The corresponding logarithms \( \log(||x^\delta||) \) are then plotted against the logarithm of residuals \( \log(||Ax^\delta-y^\delta||) \). This typically yields a curve resembling the letter L, and the approximate solutions corresponding to the corner of the L often are the closest to \( x^\dagger \).
As already noted e.g. in [1], the “vertical” part of the L is due to noise in the data, whereas the “horizontal” part is due to approximation properties of the regularization method. More precisely, for Tikhonov regularization, we have that \(|x_2^\delta| = |A^*(Ax_2^\delta - y^\delta)|\). As long as the residual is above the noise level, \(|A^*(Ax_2^\delta - y^\delta)| \approx |A^*(Ax_2 - y)| = \alpha\), which follows directly from the first-order condition for noise free data (2.6). Therefore in this case \(|x_2^\delta| \approx \text{const}\). This yields the horizontal arm of the L. Once the noise level is reached, \(|Ax_2^\delta - y^\delta| \approx \delta\) for several magnitudes of \(\alpha\) as explained in Section 3. Therefore, as \(\alpha \to 0\),

\[|x_2^\delta| = \frac{|A^*(Ax_2^\delta - y^\delta)|}{\alpha} \approx \frac{|A^*(y - y^\delta)|}{\alpha} \to \infty\]

fast. This forms the vertical arm of the L. The potential advantage of the L-curve over the other methods is that it essentially uses the gradient because \(|x_2^\delta| = \frac{1}{\alpha} |A^*(Ax_2^\delta - y^\delta)|\). The gradients have a distinct behavior based on whether the residuals \(Ax_2^\delta - y^\delta\) are dominated by the approximation of \(y\) or on the noise. Because for Tikhonov-regularization, the gradient does not depend on the smoothness of \(x^\dagger\), the L-curve does not require solution smoothness. Hence, it is more broadly applicable. On the other hand, the absence of smoothness information makes it more difficult to show convergence rates.

Nevertheless, if in particular \(x^\dagger\) satisfies a source condition, the corner of the \(L\) corresponds to the regularization parameters where the residuals transitions from the asymptotic \(|Ax_2^\delta - y^\delta| \sim \alpha^{\mu + \frac{1}{2}}\) to \(|Ax_2^\delta - y^\delta| \approx \delta\). Therefore the L curve will yield a regularization parameter comparable to the discrepancy principle, and hence comparable to the other parameter choice rules as discussed above. Since the solution smoothness is irrelevant for the L-curve, more precise convergence results using our approach requires a more precise description of the noise, which is out of the scope of this paper.

7.2. A new method and comparison

The rule (4.1) as crucial step in the estimation of the noise level can be used as a heuristic parameter choice rule. In our experiments, a slightly different variant appeared more favorable as the minimum was less flat, and the functional resembled the curve for the reconstruction error much closer. Therefore, instead of (4.1), we search for the minimizer of

\[f_{RD}(\alpha) = \arg\min_{\alpha} \frac{\partial}{\partial \alpha} |Ax_2^\delta - y^\delta|.\]  

(7.3)
We have shown in the previous sections that in the absence of noise the residual carries solution smoothness information and due to this behaves as $||Ax - y|| \sim \mu^{\frac{1}{2}}$. As long as this is above the noise level, we also have $||Ax^\delta - y^\delta|| \sim \mu^{\frac{1}{2}}$. We have also shown in Section 3 that $||Ax^\delta - y^\delta|| \approx \delta$ for several magnitudes of $\alpha$ once it reaches the noise level. In between the two stages there is a transition phase. Therefore, the idea is to track the change in the residual when $\alpha$ decreases, which is done by (7.3). We call this the residual differential method (RDM). A property separating it from other heuristic methods such as the heuristic discrepancy principle is that it tends to be optimistic, i.e., chooses regularization parameters lower than the optimal one, instead of being pessimistic, i.e., choosing too large regularization parameters. Therefore the approximate solutions appear slightly less smooth: Figure 10 shows that the RDM functional approximates the minimum of the reconstruction error curve well in the example. For noise fulfilling (3.1) with $\kappa = 0$, we know that the residual will stagnate around $||Ax^\delta - y^\delta|| \approx \delta$, and hence $\partial \partial \alpha ||Ax^\delta - y^\delta|| = 0$ for sufficiently small $\alpha$, whereas for larger $\alpha$ we have shown $||Ax^\delta - y^\delta|| = O(\alpha^{\mu+\frac{1}{2}})$, hence $\partial \partial \alpha ||Ax^\delta - y^\delta|| \approx \mu + \frac{1}{2}$. Additionally, from (3.7) follows that for extremely small $\alpha$ and finite discretization, $||Ax^\delta - y^\delta|| = O(\alpha)$, thus $\partial \partial \alpha ||Ax^\delta - y^\delta|| \approx 1$. Therefore the method will yield a residual $||Ax^\delta - y^\delta|| \approx \delta$, such that convergence follows in principle by the same argument as the discrepancy principle. Because we cannot guarantee $||Ax^\delta - y^\delta|| > \delta$ and the residual also depends on the discretization, we refrain here from attempting a rigorous proof of convergence.

We compare some parameter choice rules in Figure 11. There we show the reconstruction errors $||x^\delta - x^\dagger||$ for our Model Problem with $\eta = \beta = 2$.
yielding \( \mu = 0.375 \) in (2.2)) for a large range of regularization parameters \( \alpha \). 0.5% Gaussian noise were added to the simulated exact data. We compare the following parameter choice rules: a-priori choice \( \alpha = (\frac{2}{\rho})^{\frac{1}{2}} \) from [1, Equation (4.29)], where \( \rho = ||w|| \) in the (numerical) source representation \( x^\dagger = (A^*A)^{\alpha}w \); Morozovs discrepancy principle with \( \tau = 1.01 \) and \( \tau = 1.1 \); the heuristic discrepancy principle; the practically infeasible parameter \( \alpha_{opt} = \arg\min_\alpha ||x^\dagger - x^\dagger|| \), the L-curve, and finally the RDM method (7.3). The performance of the parameter choice rules is also compared in Table 1. Of course, the results are just a snapshot, and depending on the exact structure of the noise \( \epsilon \) in (1.2) and the unknown solution \( x^\dagger \), the quality of the reconstructions for each parameter choice rule may vary slightly. The main point here was to demonstrate that, since they all work based on the change of the behavior of the residual as function of \( \alpha \) from predominantly approximating \( y \) to being dominated by noise, they yield comparable results. In particular, this yields a rather simple explanation why heuristic parameter choice rules often do well in practice. A more detailed investigation and a relation to the theory of heuristic parameter choice rules (see, e.g., [19–21]) is left as future work.

7.3. Noise with \( \kappa > 0 \)

Without reiterating all the parameter choice rules discussed above, we show that \( \kappa > 0 \) in (3.1), i.e., smooth noise, allows to use smaller regularization parameters without amplifying the noise, which in turn improves the convergence rate.

**Theorem 5.** Let \( x^\dagger \) fulfill (2.1) with \( 0 < \mu < 1 \). Let \( \epsilon \) fulfill (3.1) with \( 0 \leq \kappa < \frac{1}{2} \). Then

\[
||x^\dagger - x^\dagger|| \leq C\delta^{\frac{2\mu}{2\kappa - 2\kappa + 1}}
\]

with some constant \( C > 0 \) independent of \( \delta \), provided the regularization parameter \( \alpha \) is chosen via

\[
\alpha \asymp \delta^{\frac{2}{2\kappa - 2\kappa + 1}}.
\]
Proof. We split
\[ ||x^\delta_x - x^\dagger|| \leq ||x_x - x^\dagger|| + ||x^\delta_x - x_x||. \] (7.4)

It is well-known, and also follows from Theorem 3 with \( \nu = 0 \), that for \( 0 < \mu < 1 \)
\[ ||x_x - x^\dagger|| \leq c_\mu x^\mu. \]

Provided \( \kappa < \frac{1}{2} \), the noise term can be estimated with Lemma 2 with \( q = 2 \) and \( p = 2 \), as we have
\[ ||x^\delta_x - x_x||^2 = \sum_{i=1}^{\infty} \left( \frac{\sigma_i}{\sigma_i^2 + \alpha} \right)^2 \langle y^\delta - y^\dagger, u^\dagger \rangle^2 \]
\[ = \alpha^{-2} \sum_{i=1}^{\infty} \frac{\sigma_i^2 y^2}{(\sigma_i^2 + \alpha)^2} \langle y^\delta - y^\dagger, u^\dagger \rangle^2 \]
\[ \leq \alpha^{-2} c_\kappa \delta^2 \alpha^{1+2\kappa}, \]
which yields
\[ ||x^\delta_x - x_x|| \leq c_\kappa \delta^{\kappa-\frac{1}{2}}, \]
such that
\[ ||x^\delta_x - x^\dagger|| \leq c_\mu x^\mu + c_\kappa \delta^{\kappa-\frac{1}{2}}. \]

Equating the two terms for \( \alpha \) then yields the parameter choice and inserting that the convergence rate.

That means that for \( \kappa > 0 \) one can achieve better convergence rates than the worst-case rate \( \delta^{2\mu+1} \). In particular, the theorem suggests that for \( \kappa \geq \frac{1}{2} \) one may achieve the linear convergence rate known from well-posed problems. While it is not yet clear whether this is always the case, the linear convergence rate is guaranteed if the operator is sufficiently ill-posed. Namely, it is
\[ ||A^\dagger (y^\delta - y^\dagger)||^2 = \sum_{i=1}^{\infty} \sigma_i^{-2} \langle e, u_i \rangle^2 \leq \sum_{i=1}^{\infty} \sigma_i^{-2} \sum_{j=i}^{\infty} \langle e, u_i \rangle^2 \leq \sum_{i=1}^{\infty} C \delta^2 \sigma_i^{-2+4\kappa} \]
\[ = \sum_{i=1}^{\infty} C \delta^2 \sigma_i^{-2+4\kappa}. \]

Therefore, if \( \sum_{i=1}^{\infty} \sigma_i^{-2+4\kappa} < \infty \), which, for example holds if \( \sigma_i = \mathcal{O}(i^{-\beta}) \) for some \( \beta > 0 \) and \( \kappa > \frac{1}{2} + \frac{1}{4\beta} \), then the inverse problem behaves like a well-posed one as
\[ ||A^\dagger y^\delta - x^\dagger|| \leq c\delta. \]
8. Conclusion and outlook

We have shown that the classical source conditions \( x^+ \in \mathcal{R}((A^*A)^\mu) \) with \( 0 < \mu < 1 \) are not only equivalent to approximation rates \( \|x_\delta - x^+\| \sim \delta^\mu \) under classical Tikhonov regularization, but also to a rate \( \|Ax_\delta - y\| \sim \delta^{1/2 + \mu} \) for \( 0 < \mu < \frac{1}{2} \). This result allows to extract the solution smoothness from the residuals, making this information accessible in practical computation. We have demonstrated how larger and higher smoothness are detectable from the residual, although not quantifiable. We have demonstrated that an estimate of the noise level can be read off the residual curve. Because the residual carries so much information, we were able to relate parameter choice rules in a novel way. There are several open topics for future work. One possible topic is the extension of the smoothness estimation to modern Banach space regularization methods, which at first requires to find an appropriate way to measure solution smoothness, since source conditions are in general not applicable. A second path is to investigate the noise even further and to categorize different levels of noise smoothness. This could be combined with revisiting the heuristic parameter choice and clarify the open questions.

Acknowledgement

D. Gerth would like to thank Prof. Oliver Ernst (TU Chemnitz) for the helpful comments.

Funding

D. Gerth was supported by Deutsche Forschungsgemeinschaft (DFG) under grant GE3171/1-1 (Project No. 416552794). R. Ramlau was supported by the Austrian Science Fund (FWF), SFB “Tomography across the scales” F6805-N36.

ORCID

Daniel Gerth http://orcid.org/0000-0003-1270-0162
Ronny Ramlau http://orcid.org/0000-0002-0277-9615

References

[1] Engl, H. W., Hanke, M., Neubauer, A. (1996). Regularization of Inverse Problems. Dordrecht: Kluwer Academic Publishers.
[2] Louis, A. K. (1989). Inverse Und Schlecht Gestellte Probleme. Stuttgart: Teubner.
[3] Gerth, D. (2019). Using Landweber iteration to quantify source conditions - A numerical study. J. Inverse Ill-Pose Probl. 27(3):367–383. DOI: 10.1515/jiip-2018-0071.
[4] Neubauer, A. (1997). On converse and saturation results for Tikhonov regularization of linear ill-posed problems. SIAM J. Numer. Anal. 34(2):517–527. DOI: 10.1137/S0036142993253928.
[5] Albani, V., Elbou, P., de Hoop, M. V., Scherzer, O. (2016). Optimal convergence rates results for linear inverse problems in Hilbert spaces. *Numer. Funct. Anal. Optim.* 37(5):521–540. DOI: 10.1080/01630563.2016.1144070.

[6] Flemming, J., Hofmann, B., Mathé, P. (2011). Sharp converse results for the regularization error using distance functions. *Inverse Probl.* 27(2):025006–0218. DOI: 10.1088/0266-5611/27/2/025006.

[7] Flemming, J. (2018). A converse result for Banach space convergence rates in Tikhonov-type convex regularization of ill-posed linear equations. *J. Inverse Ill-Posed Probl.* 26(5):639–646. DOI: 10.1515/jiip-2017-0116.

[8] Miller, P., Hohage, T. (2020). Maximal spaces for approximation rates in $\ell^1$-regularization. *Numer. Math.* 149:341–374 (2021). DOI: 10.1007/s00211-021-01225-4.

[9] Gerth, D. (2021). A new interpretation of (Tikhonov) regularization. *Inverse Probl.* 37(6):064002. DOI: 10.1088/1361-6420/abfb4d.

[10] Kekkonen, H., Lassas, M., Siltanen, S. (2014). Analysis of regularized inversion of data corrupted by white Gaussian noise. *Inverse Probl.* 30(4):045009. DOI: 10.1088/0266-5611/30/4/045009.

[11] Kindermann, S., Pereverzyev, S., Jr., Pilipenko, A. (2018). The quasi-optimality criterion in the linear functional strategy. *Inverse Probl.* 34:075001. DOI: 10.1088/1361-6420/aae4f.

[12] Kindermann, S., Raik, K. (2019). Heuristic parameter choice rules for Tikhonov Regularisation with weakly bounded noise. *Numer. Funct. Anal. Optim.* 40(12):1373–1394. DOI: 10.1080/01630563.2019.1604546.

[13] Hofmann, B., Mathé, P. (2007). Analysis of profile functions for general linear regularization methods. *SIAM J. Numer. Anal.* 45(3):1122–1141. DOI: 10.1137/060654530.

[14] Hofmann, B., Plato, R. (2020). Convergence results and low-order rates for nonlinear Tikhonov regularization with oversmoothing penalty term. *Electron. Trans. Numer. Anal.* 53:313–328. DOI: 10.1553/etna_vol53s313.

[15] Bubba, T., Hauptmann, A., Houtari, S., Rimpeläinen, J., Siltanen, S. (2016). Tomographic X-ray data of a lotus root slice filled with different chemical elements. [http://arxiv.org/abs/1609.07299](http://arxiv.org/abs/1609.07299).

[16] Hämäläinen, K., Harhanen, L., Kallonen, A., Kujunpää, A., Niemi, E., Siltanen, S. (2015). Tomographic X-ray data of a walnut. [http://arxiv.org/abs/1502.04064](http://arxiv.org/abs/1502.04064).

[17] Kindermann, S., Raik, K. (2020). Convergence of heuristic parameter choice rules for convex Tikhonov regularisation. *SIAM J. Numer. Anal.* 58(3):1773–1800. DOI: 10.1137/19M1263066.

[18] Hansen, P. C., O’Leary, D. (1993). The use of the L-curve in the regularization of discrete ill-posed problems *SIAM. SIAM J. Sci. Comput.* 14(6):1487–1503. DOI: 10.1137/0914086.

[19] Frommer, A., Maass, P. (1999). Fast CG-based methods for Tikhonov–Phillips regularization *SIAM. J. Sci. Comput.* 20(5):1831–1850.

[20] Kindermann, S. (2011). Convergence analysis of minimization-based noise level-free parameter choice rules for linear ill-posed problems. *Electron. Trans. Numer. Anal.* 38:233–257.

[21] Hämarik, U., Palm, R., Raus, T. (2011). Comparison of parameter choices in regularization algorithms in case of different information about noise level. *Calcolo.* 48(1):47–59. DOI: 10.1007/s10092-010-0027-4.
Appendix

We show

\[ \sum_{i=1}^{k} \sigma_i^{q-2p} \langle x^i, v_i \rangle^2 = O(\sigma_k^{q+4\mu-2p}) \]

if \( \sum_{i=k}^{\infty} \langle x, v_i \rangle^2 = O(\sigma_k^{4\mu}) \) and \( q + 4\mu - 2p < 0 \).

For \( k = 1 \) it is

\[ \sum_{i=1}^{k} \sigma_i^{q-2p} \langle x^i, v_i \rangle^2 = \sigma_1^{q-2p} \left( \sum_{i=1}^{\infty} \langle x, v_i \rangle^2 - \sum_{i=2}^{\infty} \langle x, v_i \rangle^2 \right) = O(\sigma_1^{q+4\mu-2p}). \]

Now assume the assertion holds for arbitrary \( k > 0 \). Then

\[ \sum_{i=1}^{k+1} \sigma_i^{q-2p} \langle x^i, v_i \rangle^2 = \sum_{i=1}^{k} \sigma_i^{q-2p} \langle x^i, v_i \rangle^2 + \sigma_{k+1}^{q-2p} \langle x, v_{k+1} \rangle^2 \]

\[ = O(\sigma_k^{q+4\mu-2p}) + \sigma_{k+1}^{q-2p} \left( \sum_{i=k+1}^{\infty} \langle x, v_i \rangle^2 - \sum_{i=k+2}^{\infty} \langle x, v_i \rangle^2 \right) \]

\[ = O(\sigma_k^{q+4\mu-2p}) + O(\sigma_{k+1}^{q+4\mu-2p}) = O(\sigma_{k+1}^{q+4\mu-2p}) \]

because \( q + 4\mu - 2p < 0 \).