Weierstraß-Institut
für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Preprint ISSN 0946 – 8633

Heuristic parameter selection based on functional minimization: Optimality and model function approach

Shuai Lu [1], Peter Mathé [2]

submitted: April 22, 2009

[1] Johann Radon Institute
for Computational and Applied Mathematics
Altenbergerstrasse 69,
A-4040 Linz,
Austria
e-mail: shuai.lu@oeaw.ac.at

[2] Weierstrass Institute
for Applied Analysis and Stochastics
Mohrenstraße 39,
D-10117 Berlin,
Germany
e-mail: mathe@wias-berlin.de

No. 1413
Berlin 2009

2000 *Mathematics Subject Classification.* 65J20 secondary: 47A52.

Key words and phrases. parameter choice, L-curve, model function.

This work has been conducted during the **Special Semester on Stochastics with Emphasis on Finance**, September 3–December 5, 2008, organized by RICAM (Johann Radon Institute for Computational and Applied Mathematics, Austrian Academy of Science) Linz, Austria. The manuscript was completed while the first named author visited the Weierstrass Institute for Applied Analysis and Stochastics, Berlin. The authors want to thank for the hospitality..

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Mohrenstraße 39
10117 Berlin
Germany

Fax: + 49 30 2044975
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Abstract. We analyze some parameter choice strategies in regularization of inverse problems, in particular the (modified) L-curve method and a variant of the Hanke–Raus rule. These are heuristic rules, free of the noise level, and they are based on minimization of some functional. We analyze these functionals, and we prove some optimality results under general smoothness conditions. We also devise some numerical approach for finding the minimizers, which uses model functions. Numerical experiments indicate that this is an efficient numerical procedure.

1. Problem formulation and theoretical background

The choice of the regularization parameter in inverse problems is an important issue, and the key to successfully solving such problems. Among the many parameter choice strategies there are those with a sound mathematical foundation, and others, which are heuristic. Nevertheless, some of these heuristic rules are widely used, as for instance the L-curve method. It is known that such heuristic parameter choice rules cannot be convergent, i.e., reliable under any circumstances, as they do not use the noise level. Here we discuss the (modified) L-curve and a variant of the Hanke-Raus rules. Both are based on minimizing a corresponding (non-linear) functional of the regularization parameter, and both use the (realized) discrepancy to judge a given regularization parameter.

We establish a mathematical calculus, which allows to prove optimality for both parameter choices, provided that the corresponding functional has an interior global minimum, an assumption which cannot be guaranteed, in general. Moreover, we use model functions to devise a numerical procedure for obtaining a minimizer, approximately.

The analysis is mainly restricted to Tikhonov regularization of linear problems in Hilbert space. Precisely, suppose that we are given noisy data

$$y_\delta = Ax + \delta\xi, \tag{1}$$

where $A\colon X \to Y$ is a (compact) linear operator acting between (real) Hilbert spaces, and ξ is noise, norm–bounded by one. The factor δ represents the noise level, which in most of this study is supposed to be unknown. In order to approximately solve this equation for the unknown element $x \in X$ one must use regularization. Here we consider linear regularization g_α, thus we consider the one-parameter family

$$x_{\alpha,\delta} = g_\alpha(A^*A)A^*y_\delta, \quad \alpha > 0, \tag{2}$$

as candidates for the approximate solution. Details on the construction and properties are postponed to Section 3.

There are two important quantities, relevant in this context, and influencing the choice of the regularization parameter. These are

$$f(\alpha) := \|x_{\alpha,\delta}\|^2, \quad \alpha > 0, \quad \text{and} \tag{3}$$

$$\rho(\alpha) := \|Ax_{\alpha,\delta} - y_\delta\|^2, \quad \alpha > 0. \tag{4}$$

The former is the (squared) solution norm. The latter is called the (squared) discrepancy, and it admits the alternative representation

$$\rho(\alpha) = \|r_\alpha(AA^*)y_\delta\|^2, \quad \alpha > 0, \tag{5}$$

with the residual function $r_\alpha(t) := 1 - tg_\alpha(t)$, see Section 3 for details. It remains to choose the regularization parameter α. Under smoothness assumptions, say φ, the discrepancy as a function of α typically has the magnitude of $\sqrt{\alpha}\varphi(\alpha) + \delta$. In the L-curve parameter choice this contribution was weighted by the norm of $x_{\alpha,\delta}$. As was pointed out in [11], under the framework of Tikhonov regularization, this principle is related to the following modification, which aims at minimizing the functional

$$\Psi_\mu(\alpha) := \|Ax_{\alpha,\delta} - y_\delta\|^2 \|x_{\alpha,\delta}\|^{2\mu} \longrightarrow \text{MIN!} \tag{6}$$

The tangent slope value is $-1/\mu$ when the regularization parameter is chosen by the L-curve method at

$$\left(\log \|Ax_{\alpha,\delta} - y_\delta\|^2, \log \|x_{\alpha,\delta}\|^2\right).$$

Actually, in loose terms this minimization corresponds to minimizing $\varphi(\alpha) + \delta/\sqrt{\alpha}$. But this can also be achieved by directly weighting the discrepancy by the factor $1/\sqrt{\alpha}$. The resulting heuristic parameter choice was presented in [3, Sect. 4.5 (4.115)]. It requires to minimize the function

$$\Psi(\alpha) := \frac{\|Ax_{\alpha,\delta} - y_\delta\|^2}{\alpha} \longrightarrow \text{MIN!} \tag{7}$$

Remark 1. We mention that the discrepancy principle can also be derived from minimizing the functional

$$\Psi_D(\alpha) := \left|\rho(\alpha) - c^2\delta^2\right|^2 \longrightarrow \text{MIN},$$

where the constant $c > 1$ is chosen appropriately. The present model function approach also applies to this principle, however, the analysis was presented in [8, 12], and we do not repeat it here.

The theoretical results which are related to such choices of regularization parameter use the *realized discrepancy*. Precisely, if by some method the parameter α_* has been chosen, then we let

$$\delta_* := \|Ax_{\alpha_*,\delta} - y_\delta\| \tag{8}$$

be the *realized discrepancy* according to this parameter choice. In principle, this quantity can be observed from the computations, and this knowledge can be used for predicting error bounds.

For applying any of the above parameter choice we must address the following questions.

(1) Does such minimizer α_* from (6) or (7), respectively, always exist?
(2) If it exists, what is the quality of the chosen reconstruction?
(3) Is the realized discrepancy δ_* of similar size as the error level δ?

Therefore our subsequent analysis is as follows. We first provide some preliminary analysis which points at the existence of minimizers for the functionals of interest. Next we provide theoretical error bounds in terms of the relation between the true error level δ and the realized discrepancy δ_*, granted that some minimizer exists. We then turn to the *model function* approach to speed up the search for a regularization parameter α_*, and we indicate that this is a reasonable way. At the same time we observe that the use of the rule from (7) is limited in real applications. This is shown by relating the true noise level to the realized discrepancy.

We conclude the study with a simple one–step proposal, which arises from numerical experiments.

2. Preliminary analysis for Tikhonov regularization

Tikhonov regularization is derived from minimizing the functional

$$T_\alpha(x) := \|Ax - y_\delta\|^2 + \alpha\|x\|^2 \longrightarrow \text{MIN}, \tag{9}$$

for given parameter $\alpha > 0$. In Hilbert space this is a smooth convex functional and the following is known, see e.g. [8]. The unique minimizer $x_{\alpha,\delta}$ of (9) is given as

$$x_{\alpha,\delta} := \left(A^*A + \alpha I\right)^{-1} A^* y_\delta, \quad \alpha > 0. \tag{10}$$

Moreover, the minimization problem obeys the following variational form

$$\langle Ax, Ag\rangle + \alpha\langle x, g\rangle = \langle y_\delta, Ag\rangle, \quad g \in X, \tag{11}$$

and $x_{\alpha,\delta}$ is the unique solution to this.

2.1. The fundamental function $\mathbf{h}(\alpha)$.

It turns out to be interesting to study the Tikhonov functional (9) at the minimizer $x_{\alpha,\delta}$, i.e., we let

$$J(\alpha) := T_\alpha(x_{\alpha,\delta}) = \|Ax_{\alpha,\delta} - y_\delta\|^2 + \alpha\|x_{\alpha,\delta}\|^2, \quad \alpha > 0. \tag{12}$$

Functionals of this type were used to derive a model function approach in [8, 12], and we will recourse to this idea in Section 4.

Here we make the following observation. If we insert $g := x_{\alpha,\delta}$ in (11) then we obtain that

$$\langle y_\delta, Ax_{\alpha,\delta}\rangle = \|Ax_{\alpha,\delta}\|^2 + \alpha\|x_{\alpha,\delta}\|^2. \tag{13}$$

The latter function will be fundamental and we assign

$$h(\alpha) := \|Ax_{\alpha,\delta}\|^2 + \alpha\|x_{\alpha,\delta}\|^2, \quad \alpha > 0. \tag{14}$$

If, in real Hilbert space, we rewrite the functional J from (12) then we obtain that

$$J(\alpha) = \|Ax_{\alpha,\delta}\|^2 - 2\langle y_\delta, Ax_{\alpha,\delta}\rangle + \|y_\delta\|^2 + \alpha\|x_{\alpha,\delta}\|^2.$$

Using the identity (13) and the function h from (14) yields the identity

$$\|y_\delta\|^2 = J(\alpha) + h(\alpha), \quad \alpha > 0. \tag{15}$$

The behavior of these functions depends on the given data y_δ, in particular it proves important whether these belong to the domain of the *Moore-Penrose inverse* or not. To describe this we let $Q\colon Y \to Y$ denote the orthogonal projection onto $\overline{\mathcal{R}(A)}$, the closure of the range of the operator A. It is well known that the data y_δ belong to the domain of the Moore–Penrose inverse exactly if $Qy_\delta \in \mathcal{R}(A)$.

Properties of the function h can be derived from known properties of the functional J, cf. [8, Lemma 2.3].

Lemma 1. *The function $h(\alpha),\ \alpha > 0$, is decreasing and convex. If $Qy_\delta \neq 0$ then it is strictly decreasing and strictly convex.*

The impact for the functions f and ρ is summarized as follows.

Lemma 2. *The following representations hold for the functions f and ρ from (3) and (4).*

(1) $f(\alpha) = -h'(\alpha),\ \alpha > 0$, *and*
(2) $\rho(\alpha) = \|y_\delta\|^2 - (h(\alpha) - \alpha h'(\alpha)),\ \alpha > 0$.

Moreover, it holds that

$$\rho'(\alpha) = -\alpha f'(\alpha), \quad \alpha > 0. \tag{16}$$

Proof. We know from (15) that $J'(\alpha) = -h'(\alpha)$. We also know from [8] that $J'(\alpha) = \|x_{\alpha,\delta}\|^2 = f(\alpha)$, which yields the first assertion. The second assertion follows from $\rho(\alpha) = J(\alpha) - \alpha J'(\alpha)$. The final assertion follows from differentiating the representation in the second item. □

The following consequences are important, cf. [8, Lemma 2.3].

Corollary 1. *If $Qy_\delta \neq 0$ then $f'(\alpha) < 0$ and $\rho'(\alpha) > 0,\ \alpha > 0$. Thus, the function f is strictly decreasing and the function ρ is strictly increasing in this case.*

Proof. We know from [8, Lemma 2.3] that the functional J is strictly increasing and strictly concave if $Qy_\delta \neq 0$. Hence, the function h must be strictly decreasing and strictly convex, in this case. Consequently, we have that $h'(\alpha) < 0$ and $h''(\alpha) > 0$. The proof can be completed using Lemma 2 and its consequence that $\rho'(\alpha) = \alpha h''(\alpha)$. □

Remark 2. We stress that all the above results could have been obtained using *spectral calculus*, and we sketch this in case that the operator A^*A is compact and has singular numbers $t_1 \geq t_2 \geq \cdots \geq 0$, and corresponding singular functions $u_1, u_2, \ldots$. Let us abbreviate $y_j := |\langle y_\delta, u_j\rangle|$. Then we can take the representation of the solution from (10) to derive the explicit form for the function h from (14) as

$$h(\alpha) = \sum_{t_j>0} \frac{t_j^2}{(t_j+\alpha)^2} y_j^2 + \sum_{t_j>0} \frac{\alpha t_j}{(t_j+\alpha)^2} y_j^2 = \sum_{t_j>0} \frac{t_j}{t_j+\alpha} y_j^2. \tag{17}$$

So, for historical reasons we found it useful to derive the above properties without using spectral calculus, but we will use this in our subsequent analysis, occasionally.

We add the following technical assertion, where we make use of spectral calculus.

Lemma 3. *It holds true that*

$$2\left(-h'(\alpha)\right)^2 \leq h(\alpha)h''(\alpha), \quad \alpha > 0. \tag{18}$$

Proof. Using the Cauchy Schwartz inequality, we have that

$$
\begin{aligned}
-h'(\alpha) &= \sum_{t_j>0} \frac{t_j}{(t_j+\alpha)^2} y_j^2 = \sum_{t_j>0} \frac{t_j^{1/2}}{(t_j+\alpha)^{1/2}} y_j \frac{t_j^{1/2}}{(t_j+\alpha)^{3/2}} y_j \\
&\le \left(\sum_{t_j>0} \frac{t_j}{t_j+\alpha} y_j^2 \right)^{1/2} \left(\sum_{t_j>0} \frac{t_j}{(t_j+\alpha)^3} y_j^2 \right)^{1/2} = (h(\alpha))^{1/2} \left(\frac{h''(\alpha)}{2} \right)^{1/2}.
\end{aligned}
$$

Squaring this inequality yields (18), and completes the proof. □

2.2. **Minimizers: existence and properties.** We turn to analyzing the functionals from (6) and (7), respectively. The modified L-curve method is based on minimizing the functional in (6), which, for given fixed parameter $\mu > 0$, reads as

$$
\Psi_\mu(\alpha) := \|r_\alpha(AA^*)y_\delta\|^2 \|x_{\alpha,\delta}\|^{2\mu}, \quad \alpha > 0, \tag{19}
$$

and properties of this functions are important. With the functions f and ρ from (3) and (4), respectively, we have that $\Psi_\mu(\alpha) = \rho(\alpha) f^\mu(\alpha)$.

Lemma 4. *For Tikhonov regularization we have the representation*

$$
\begin{aligned}
\Psi'_\mu(\alpha) &= f^{\mu-1}(\alpha) f'(\alpha) \left(\mu\rho(\alpha) - \alpha f(\alpha) \right), \\
\Psi''_\mu(\alpha) &= \left(f^{\mu-1}(\alpha) f'(\alpha) \right)' \left(\mu\rho(\alpha) - \alpha f(\alpha) \right) - f^{\mu-1}(\alpha) f'(\alpha) \left((\mu+1)\alpha f'(\alpha) + f(\alpha) \right).
\end{aligned}
$$

This can be translated into necessary and sufficient conditions for minimal values.

Corollary 2. *Any minimizer α_* of the functional (6) must satisfy the first order condition*

$$
\mu \|r_{\alpha_*}(AA^*)y_\delta\|^2 = \alpha_* \|x_{\alpha_*,\delta}\|^2. \tag{20}
$$

Any such stationary value α_* is a minimum provided that

$$
-(\mu+1)\alpha_* f'(\alpha_*) < f(\alpha_*). \tag{21}
$$

We turn to the second approach using the functional from (7).

Lemma 5. *For Tikhonov regularization we have that*

$$
\Psi'(\alpha) = -\frac{1}{\alpha^2} \left(\rho(\alpha) + \alpha^2 f'(\alpha) \right), \tag{22}
$$

$$
\Psi''(\alpha) = \frac{1}{\alpha^3} \left(\rho(\alpha) + \alpha^2 f'(\alpha) \right) - \left(f'(\alpha) + \alpha f''(\alpha) \right). \tag{23}
$$

Again, this results in necessary and sufficient conditions for minimizers.

Corollary 3. *Any minimizer α_* of the functional (7) must satisfy the first order condition*

$$
\|r_{\alpha_*}(AA^*)y_\delta\|^2 = -\alpha_*^2 f'(\alpha_*). \tag{24}
$$

Any such stationary value α_* is a minimum provided that

$$
\alpha_* f''(\alpha_*) < -f'(\alpha_*). \tag{25}
$$

We turn to a first discussion whether minimizers exist, and we start from the observation first pointed out in [11, Lemma 6]. We formulate this criterion for the functional Ψ_μ from (6), only.

Theorem 1. *Let* $0 < \alpha_1 < \alpha_0 < \infty$. *There is a minimizer* α_* *for* Ψ_μ *in* (α_1, α_0) *if*

$$\mu\rho(\alpha_1) > \alpha_1 f(\alpha_1) \quad \textit{and} \quad \mu\rho(\alpha_0) < \alpha_0 f(\alpha_0). \tag{26}$$

Proof. The function Ψ_μ is continuous, and hence must have a minimum, say α_* in $[\alpha_1, \alpha_0]$. We claim that this must be an interior point. To this end, a look into Lemma 4 reveals that both conditions from (26) are equivalent to $\Psi'_\mu(\alpha_1) < 0$ and $\Psi'_\mu(\alpha_0) > 0$. In particular there must be points $\alpha_0 > \beta_1 > \alpha_1$ and $\alpha_1 < \beta_0 < \alpha_0$ with $\Psi_\mu(\beta_1) < \Psi_\mu(\alpha_1)$, $\Psi_\mu(\beta_0) < \Psi_\mu(\alpha_0)$, and the end points cannot be minimal values. $\square$

If one wants to use the functionals in numerical implementation, then one has to find an initial parameter α_0, possibly large, where to start searching for the stationary value α_*. In general this may be difficult.

Theorem 2. *For Tikhonov regularization there is no initial parameter* $\alpha_0 > 0$ *such that*

$$\mu \|r_{\alpha_0}(AA^*)y_\delta\|^2 < \alpha_0 \|x_{\alpha_0,\delta}\|^2, \tag{27}$$

which does not depend on the data y_δ.

Proof. We rewrite both sides by spectral calculus and obtain the equivalent inequality

$$\mu \sum_{t_j>0} \frac{\alpha_0^2}{(t_j+\alpha_0)^2} y_j^2 < \sum_{t_j>0} \frac{\alpha_0 t_j}{(t_j+\alpha_0)^2} y_j^2.$$

If this is to hold for any square summable sequence y_δ, then, by letting successively y_j be the jth unit vector, it must necessarily hold that

$$\mu \frac{\alpha_0^2}{(t_j+\alpha_0)^2} < \frac{\alpha_0 t_j}{(t_j+\alpha_0)^2}, \quad \text{for all } j = 1, 2, \dots,$$

but the latter holds only for $t_j > \mu\alpha_0$, and hence α_0 must equal zero, which is a contradiction. $\square$

Similarly, one can prove

Theorem 3. *For Tikhonov regularization there is no initial parameter* $\alpha_0 > 0$ *such that*

$$\|r_{\alpha_0}(AA^*)y_\delta\|^2 < -\alpha_0^2 f'(\alpha_0), \tag{28}$$

which does not depend on the data y_δ.

In a similar fashion one can verify that, for both functionals Ψ_μ and Ψ, there is no parameter α_1 in the range $(0, \|A^*A\|)$ which obeys the reverse inequalities from (27) and (28), respectively, regardless of the data y_δ. These results already exhibit the difficulties for numerical use, and this is in conformance with *Bakushinski's veto*, which asserts that principles which do not use the noise level δ, explicitly, must fail, at least for some instances. The applicability of the functionals from (6) and (7) thus depends on additional behavior of the singular values $t_j > 0$ of the operator A^*A and (in relation to) the data y_δ.

2.3. **The reciprocal solution norm.** Here we briefly analyze the function f from (3), or more specifically its companion

$$g(\alpha) := \frac{1}{\|x_{\alpha,\delta}\|}, \quad \alpha > 0. \tag{29}$$

Lemma 6. *For Tikhonov regularization the function g from (29) is a concave increasing function. It holds that $\lim_{\alpha\searrow 0} g(\alpha) = 0$ if and only if $Qy_\delta \notin \mathcal{R}(A)$.*

Proof. For Tikhonov regularization we can rewrite f, using spectral calculus, see Remark 2, and obtain the explicit representation as

$$f(\alpha) = \sum_{t_j>0} \frac{t_j}{(t_j+\alpha)^2} y_j^2. \tag{30}$$

Term-wise differentiation provides us with

$$f'(\alpha) = -2 \sum_{t_j>0} \frac{t_j}{(t_j+\alpha)^3} y_j^2, \quad \alpha > 0. \tag{31}$$

The latter function is finite for all $\alpha > 0$, such that we may interchange differentiation and summation. Observe that (31) shows that the function f is strictly decreasing, and its reciprocal will thus be increasing.

By the definition, it yields that $g = 1/\sqrt{f}$, and differentiation provides us with the first and second derivatives

$$g' = -\frac{f'}{2f^{3/2}}, \tag{32}$$

$$g'' = \frac{3}{4f^{5/2}} \left((f')^2 - \frac{2}{3} f f'' \right). \tag{33}$$

Concavity of g is established if the difference in the brackets is non-positive. This will be shown by using the Cauchy Schwartz Inequality, similar to the proof of Lemma 3. To this end we observe that

$$f''(\alpha) = 6 \sum_{t_j>0} \frac{t_j}{(t_j+\alpha)^4} y_j^2, \quad \alpha > 0.$$

Hence we can bound

$$-f'(\alpha) = 2 \sum_{t_j>0} \frac{t_j}{(t_j+\alpha)^3} y_j^2 = 2 \sum_{t_j>0} \frac{\sqrt{t_j}}{t_j+\alpha} y_j \cdot \frac{\sqrt{t_j}}{(t_j+\alpha)^2} y_j \leq 2 f^{1/2} \left(\frac{f''}{6} \right)^{1/2}.$$

Squaring the above inequality yields concavity of g.

We see from (30) that $f(\alpha) \to 0$, and hence that $g(\alpha) \to \infty$, as $\alpha \to \infty$. It remains to prove that $\lim_{\alpha\searrow 0} g(\alpha) = 0$ if and only if $y_\delta \notin \mathcal{R}(A)$. By monotonicity the limit exists, and it is bounded away from zero exactly if $\{\|x_{\alpha,\delta}\|,\ \alpha > 0\}$ is bounded. The fundamental dichotomy from regularization theory, see [3, Prop. 3.6] asserts that this is the case if and only if $Qy_\delta \in \mathcal{R}(A)$, and the proof is complete. □

The following consequence is interesting.

Corollary 4. *Suppose that $Qy_\delta \notin \mathcal{R}(A)$. For Tikhonov regularization the inverse function*

$$s \longrightarrow g^{-1}(s), \quad s \geq 0,$$

is convex, increasing, and obeys $\lim_{s \searrow 0} g^{-1}(s) = 0$.

Remark 3. Increasing and continuous functions, say φ, which obey $\lim_{t \searrow 0} \varphi(t) = 0$ are called *index functions*, as these are used to represent smoothness in terms of general source conditions, see (36).

3. Theoretical error bounds

Here we neglect the difficulties for finding non-trivial minimizers of the functionals, and we assume that we have found some. Then we may ask for certain *optimality properties* of the chosen parameter, say α_*.

In addition, we present the following theoretical error bounds for more general classes of regularization, where the solutions are given as $x_{\alpha,\delta} := g_\alpha(A^*A)A^*y_\delta$, where g_α is a one-parameter family of complex functions, used as operator functions via spectral calculus, and we refer to [3] for a monograph. Tikhonov regularization is given via $g_\alpha(t) := 1/(t+\alpha),\ \alpha > 0$.

One crucial property of such general linear regularization schemes is their *qualification*, i.e., the ability to react to given smoothness. In this context the residual function $r_\alpha(t) := 1 - tg_\alpha(t),\ \alpha > 0$, is important, and by its very definition we have that $\rho(\alpha) := \|r_\alpha(A^*A)y_\delta\|^2$, which was already stated in Section 1. Several method dependent constants are relevant. We let

$$\gamma_0 := \sup_{\alpha>0} \sup_{0<t\leq\|A^*A\|} |r_\alpha(t)| = \sup_{\alpha>0} \sup_{\|y_\delta\|\leq 1} \|r_\alpha(A^*A)y_\delta\|^2. \tag{34}$$

Furthermore, noise propagation requires to know

$$\gamma_* := \|g_\alpha(A^*A)A^* \colon Y \to Y\|. \tag{35}$$

For smoothness given in terms of *general source conditions* of the type

$$x \in H_\varphi := \{x, \quad x = \varphi(A^*A)v,\ \|v\| \leq 1\}\,, \tag{36}$$

the notion of qualification has been introduced in [10]. Here the function φ is supposed to be an index function as described in Remark 3, and we recall

Definition 1. A regularization is said to have qualification ψ (with constant $\gamma \geq 1$) if

$$|r_\alpha(t)|\,\psi(t) \leq \gamma\psi(\alpha), \quad \alpha > 0.$$

For Tikhonov regularization we have that $r_\alpha(t) = \alpha/(t+\alpha)$, and hence $\gamma_0 = 1$. Its (maximal up to constants) qualification is known to be $\psi(t) = t,\ t > 0$, see [10], again.

3.1. **Bounding the bias.** The following results generalize previous analysis from [4]. In approaches which are based on using the discrepancy one always has a bound for the bias in terms of the realized discrepancy, and we shall first establish this. To do so we will use the *modulus of continuity*, which is given as

$$\omega(A^+, H_\varphi, \delta) := \sup\{\|x\|, \quad x \in H_\varphi, \ \|Ax\| \leq \delta\}), \quad \delta > 0,$$

see [7] for details.

Proposition 1. *Let g_α be a regularization (with constant γ_0), using data (1). For any parameter choice α_* let δ_* be the* realized discrepancy *according to this parameter choice.*

If the true solution obeys $x \in H_\varphi$ then

$$\|r_{\alpha_*}(A^*A)x\| \leq (1+\gamma_0)\,\omega(A^+, H_\varphi, \max\{\delta, \delta_*\}). \tag{37}$$

Consequently, it holds true that

$$\|r_{\alpha_*}(A^*A)x\| \leq 4\gamma_0\varphi(\Theta^{-1}(\max\{\delta, \delta_*\})). \tag{38}$$

Proof. Using the triangle inequality we bound

$$\|r_\alpha(AA^*)y\| \leq \|r_\alpha(AA^*)y_\delta\| + \|r_\alpha(AA^*)(y - y_\delta)\| \leq \delta_* + \gamma_0\delta.$$

Let $z := \frac{1}{\gamma_0} r_\alpha(A^*A)x$. Notice that we can rewrite the above bound as

$$\|Az\| \leq \delta + \frac{\delta_*}{\gamma_0} \leq (1 + \frac{1}{\gamma_0}) \max\{\delta, \delta_*\}.$$

Moreover, under $x \in H_\varphi$ with $x = \varphi(A^*A)v$, it holds that

$$z = \varphi(A^*A)\frac{r_\alpha(A^*A)}{\gamma_0}v,$$

and $\|\frac{r_\alpha(A^*A)}{\gamma_0}v\| \leq 1$, hence $z \in H_\varphi$.

Therefore, by the definition of the modulus of continuity we have that

$$\|z\| \leq \omega(A^+, H_\varphi, (1 + \frac{1}{\gamma_0})\max\{\delta, \delta_*\}),$$

which implies (37) using monotonicity, and multiplying by γ_0. The final consequence is the well known bound for the modulus of continuity, as $\omega(A^+, H_\varphi, \delta) \leq 2\varphi(\Theta^{-1}(\delta)$, which follows from spectral cut–off. □

Remark 4. The above bound strongly suggests to have δ_* of similar size as the noise level δ. In case the latter is known to us this goal is achieved by the discrepancy principle.

Notice also, that the bound on the bias was obtained without any requirement on the qualification of the chosen regularization, a drawback which is responsible for saturation, and we refer to [9].

The final assertion uses the fact that the modulus of continuity is a lower bound for *any* method of reconstruction, and thus the bound obtained for spectral cut–off may be used to bound the modulus, we refer to [7] for details.

3.2. **Controlling the overall error.** Thus, if we want to bound the overall error for a chosen parameter choice rule, we must assess the noise propagation. For both the above rules from (6) and (7) this can be done, and we present the following results which extends previous bounds to the situation of general source conditions. We emphasize, that parameter choice rules which use the discrepancy require higher qualification than smoothness of the true solution. Precisely, if $x \in H_\varphi$, then the qualification is required to be at least

$$\Theta(t) := \sqrt{t}\varphi(t),\ 0 < t \leq \|A^*A\|. \tag{39}$$

For the modified L-curve functional from (6) we need an additional assumption, called Δ_2-condition, for the inverse function g^{-1} corresponding to (29).

Assumption 1. There is a constant $C_2 \geq 1$ such that

$$g^{-1}(2s) \leq 2^{C_2} g^{-1}(s), \quad s > 0. \tag{40}$$

Remark 5. Notice that in Corollary 4 and for Tikhonov regularization the function g^{-1} was shown to be convex and increasing, and hence the constant C_2 cannot be less than 1, in this case. Furthermore, and more importantly, the validity of such Δ_2-condition depends on the interplay between the singular numbers of A^*A and the data y_δ.

Theorem 4. *Fix any $\mu > 0$, and let α_* be a global minimizer of (6). Suppose that $x \in H_\varphi$, and that the function g^{-1} from Corollary 4 obeys Assumption 1.*

If the regularization g_α has qualification Θ with constant γ then

$$e(x_{\alpha_*,\delta}, x, \delta) \leq \left(4\gamma_0 + \gamma_* \left(2^{\mu+1}\gamma\frac{\delta}{\delta_*}\right)^{\frac{C_2}{\mu}}\right)\varphi(\Theta^{-1}(\max\{\delta, \delta_*\})), \tag{41}$$

where $\delta_* > 0$ denotes the realized discrepancy as in (8).

Proof. We use the natural error decomposition to deduce that

$$e(x_{\alpha_*,\delta}, x, \delta) \leq \|r_{\alpha_*}(A^*A)x\| + \gamma_*\frac{\delta}{\sqrt{\alpha_*}}.$$

The bias was bounded in Proposition 1, and yields that

$$\|r_{\alpha_*}(A^*A)x\| \leq 4\gamma_0\varphi(\Theta^{-1}(\max\{\delta, \delta_*\})).$$

We turn to bounding the noise term. Being a minimizer of Ψ_μ, we also have the inequality $\Psi_\mu(\alpha_*) \leq \Psi_\mu(\hat\alpha)$, where $\hat\alpha$ is chosen from $\Theta(\hat\alpha) = \delta$. We bound

$$\begin{aligned}\|Ax_{\hat\alpha} - y_\delta\| &\leq (\|Ax_{\hat\alpha} - Ax\| + \|Ax - y_\delta\|)\\ &\leq (\|r_\alpha(AA^*)Ax\| + \delta) \leq \gamma(\Theta(\hat\alpha) + \delta) = 2\gamma\delta,\end{aligned}$$

where we used that $\gamma \geq 1$. Therefore, using the realized discrepancy, we have that

$$\|x_{\alpha_*,\delta}\|^{2\mu}\delta_*^2 \leq (2\gamma\delta)^2\,\|x_{\hat\alpha,\delta}\|^{2\mu},$$

which implies that

$$g(\hat\alpha) \leq \left(2\gamma\frac{\delta}{\delta_*}\right)^{1/\mu} g(\alpha_*).$$

If $(2\gamma\delta/\delta_*)^{1/\mu} \le 1$ then this implies that $\hat{\alpha} \le \alpha_*$, since g is increasing. Otherwise, let κ be the smallest integer such that 2^κ is larger than or equal to $(2\gamma\delta/\delta_*)^{1/\mu} > 1$. Iterating the Δ_2 condition κ times yields

$$\hat{\alpha} \le 2^{C_2\kappa}\alpha_* \le 2^{C_2}2^{C_2(\kappa-1)}\alpha_* \le \left(2^{\mu+1}\gamma\frac{\delta}{\delta_*}\right)^{\frac{C_2}{\mu}}\alpha_*.$$

This allows to bound the noise propagation as

$$\gamma_*\frac{\delta}{\sqrt{\alpha_*}} \le \gamma_*\left(2^{\mu+1}\gamma\frac{\delta}{\delta_*}\right)^{\frac{C_2}{\mu}}\frac{\delta}{\sqrt{\hat{\alpha}}} = \gamma_*\left(2^{\mu+1}\gamma\frac{\delta}{\delta_*}\right)^{\frac{C_2}{\mu}}\varphi(\Theta^{-1}(\delta)).$$

This gives the error bound from (41) and completes the proof. □

Remark 6. To the best of our knowledge, Theorem 4 is the first optimality result for the modified L-curve method (6) within the deterministic noise model. Since this holds true for any chosen parameter $\mu > 0$, we may use the equivalency analysis of the modified L-curve and original L-curve method in [11, Thm. 1], to conclude that for the (true but unknown) parameter μ_0, a similar result is valid for the original L-curve method.

A similar and simpler reasoning applies for the second criterion (7) and provides us with

Theorem 5. *Suppose that $x \in H_\varphi$, and let α_* be a minimizer of (7). If the regularization g_α has qualification Θ with constant γ then*

$$e(x_{\alpha_*,\delta}, x, \delta) \le 2\gamma_0\left(2 + \gamma_*\frac{\delta}{\delta_*}\right)\varphi(\Theta^{-1}(\max\{\delta, \delta_*\})), \tag{42}$$

where $\delta_* > 0$ denotes the realized discrepancy as in (8).

Proof. We use the natural error decomposition to deduce that

$$e(x_{\alpha_*,\delta}, x, \delta) \le \|r_{\alpha_*}(A^*A)x\| + \gamma_*\frac{\delta}{\sqrt{\alpha_*}}.$$

The bias was bounded in Proposition 1, and we turn to bounding the noise term. We first claim that

$$\frac{\|Ax_{\alpha_*,\delta} - y_\delta\|}{\sqrt{\alpha_*}} \le 2\gamma\varphi(\Theta^{-1}(\delta)). \tag{43}$$

Indeed, for $\hat{\alpha}$ with $\Theta(\hat{\alpha}) = \delta$ we have by the choice of α_* as minimizer that

$$\begin{aligned}\frac{\|Ax_{\alpha_*,\delta} - y_\delta\|}{\sqrt{\alpha_*}} &\le \frac{\|Ax_{\hat{\alpha},\delta} - y_\delta\|}{\sqrt{\hat{\alpha}}} \le \frac{\|r_{\hat{\alpha}}(AA^*)Ax\| + \delta\|r_{\hat{\alpha}}(AA^*)\|}{\sqrt{\hat{\alpha}}} \\ &\le \frac{1}{\sqrt{\hat{\alpha}}}\left(\gamma\Theta(\hat{\alpha}) + \gamma\delta\right) = 2\gamma\frac{\delta}{\sqrt{\Theta^{-1}(\delta)}} = 2\gamma\varphi(\Theta^{-1}(\delta)),\end{aligned}$$

which proves (43). In a second step we bound the error term as

$$\delta\|g_{\alpha_*}(A^*A)A^*\| \le \gamma_*\frac{\delta}{\sqrt{\alpha_*}} = \gamma_*\frac{\delta}{\delta_*}\frac{\|Ax_{\alpha_*,\delta} - y_\delta\|}{\sqrt{\alpha_*}} \le 2\gamma\gamma_*\frac{\delta}{\delta_*}\varphi(\Theta^{-1}(\delta)).$$

Overall this results in the error bound

$$\begin{aligned} e(x_{\alpha_*,\delta}, x, \delta) &\leq (1+\gamma_0)\omega(A^+, H_\varphi, \max\{\delta, \delta_*\}) + 2\gamma_0\gamma_* \frac{\delta}{\delta_*}\varphi(\Theta^{-1}(\delta)) \\ &\leq 2\gamma_0 \left(\omega(A^+, H_\varphi, \max\{\delta, \delta_*\}) + \gamma_* \frac{\delta}{\delta_*}\varphi(\Theta^{-1}(\delta))\right) \\ &\leq 2\gamma_0 \left(2\varphi(\Theta^{-1}(\max\{\delta, \delta_*\})) + \gamma_* \frac{\delta}{\delta_*}\varphi(\Theta^{-1}(\max\{\delta, \delta_*\}))\right) \\ &\leq 2\gamma_0 \left(2 + \gamma_* \frac{\delta}{\delta_*}\right)\varphi(\Theta^{-1}(\max\{\delta, \delta_*\})), \end{aligned}$$

which is (42) and thus completes the proof. □

Remark 7. For power type source conditions, i.e., when $\varphi(t) := t^p$, for some $p > 0$, such result was obtained in the original study [4].

Notice, that we need to require higher qualification, and this leads to early saturation if the underlying smoothness is close to the maximal covered by the regularization. This can be overcome by considering a different functional, as e.g.

$$\frac{1}{\sqrt{\alpha}} \|r_\alpha^2(A^*A)y_\delta\| \longrightarrow \text{MIN!} \tag{44}$$

This corresponds to one times iterated regularization. It may overcome the early saturation if only the underlying regularization has qualification of $\rho(t) = \sqrt{t}$, at least. Such modifications may be obtained in more generality, the most prominent one is called *Hanke–Raus* parameter choice, see [4], and we refer to [3, Sect. 4.5].

4. Use of model functions to speed up parameter search

The theoretical results from Section 3 are non-constructive, and it remains to design efficient methods for finding minimizers of the functions Ψ_μ, and Ψ from (6) and (7), respectively. Recent work to accomplish this task uses some fixed–point iteration, we refer to [1, 2]. This approach uses the observation captured in Theorem 6, below. Specifically, if for some α_0 the assumption (53) is fulfilled then the function

$$\Phi_\mu(\alpha) := \mu \frac{\rho(\alpha)}{f(\alpha)}, \quad 0 < \alpha \leq \alpha_0, \tag{45}$$

obeys that $\alpha_1 := \Phi_\mu(\alpha_0) < \alpha_0$, and this gives rise to an iterative procedure for finding a stationary point. We add that a similar iterative procedure could be designed for the Hanke–Raus type method, based on Theorem 7. We shall compare the performance of both, the iterative approach from [1] and the model function approach, outlined next.

Here we use some results on the use of model functions for the determination of the regularization parameter. This approach was first proposed in [8], and later improved in [12]. While these studies consider the use of model functions for finding the regularization parameter based on the discrepancy principle, the recent study [6] proposes to use this in context of the modified L-curve criterion from (6) which is the first attempt on some error free criterion. In this section, we will revisit the model function approach and introduce a new presentation of the model function under the framework of (6) and (7).

Within the model function approach we reduce the above sum in (17) to one term, i.e., replace the function h by a two-parameter family

$$h(\alpha) \sim m(\alpha) = \frac{C}{T+\alpha}.$$

Remark 8. Notice, that in our approach to the model function, the sign of C is positive, in contrast to the function introduced in [12].

The parameters C and T are determined *point-wise*, i.e., at given parameter α the values of $h(\alpha)$ and its first derivative $h'(\alpha)$ coincide with the values $m(\alpha)$ and its first derivative $m'(\alpha)$. Formally we let

$$m(\alpha) = m_{C,T}(\alpha) := \frac{C}{T+\alpha}, \quad \alpha > 0, \tag{46}$$

Thus, for fixed value of α, this results in the system

$$\begin{aligned} \frac{C}{T+\alpha} &= h(\alpha) = \|Ax_{\alpha,\delta}\|^2 + \alpha\|x_{\alpha,\delta}\|^2, \\ \frac{C}{(T+\alpha)^2} &= -h'(\alpha) = f(\alpha) = \|x_{\alpha,\delta}\|^2, \end{aligned}$$

for the determination of C, T. The solution can be seen to equal

$$C = \frac{\left(\|Ax_{\alpha,\delta}\|^2 + \alpha\|x_{\alpha,\delta}\|^2\right)^2}{\|x_{\alpha,\delta}\|^2}, \text{ and } T = \frac{\|Ax_{\alpha,\delta}\|^2}{\|x_{\alpha,\delta}\|^2}.$$

We shall first analyze C and T as functions of the parameter α. The following monotonicity properties were shown in [12, Lemma 3.1–3.3]. Recall the definition of the projection Q from § 2.

Lemma 7. *Suppose that* $\|Qy_\delta\| > 0$.

(1) *The functions* $C = C(\alpha)$ *and* $T = T(\alpha)$ *are positive and non-decreasing.*
(2) *The quotient* $\alpha \to \frac{C(\alpha)}{T(\alpha)}$ *is non-increasing.*
(3) *It holds* $\frac{C(\alpha)}{T(\alpha)} < \|Qy_\delta\|^2$, *and* $\lim_{\alpha\searrow 0} \frac{C(\alpha)}{T(\alpha)} = \|Qy_\delta\|^2$.

Proof. With the help of the functions f and h from (3) and (14) we have that $C = h^2/f$ and $T + \alpha = h/f$. Differentiating this yields that

$$C' = \frac{h}{f}\left(-2f^2 + h(-f')\right),$$

$$\text{and} \quad T' = (T+\alpha)' - 1 = \frac{-2f^2 + h(-f')}{f^2}.$$

In both cases the enumerators are non-negative by Lemma 3, which yields monotonicity of both the functions C and T.

To prove the second assertion we rewrite

$$\frac{C(\alpha)}{T(\alpha)} = \frac{h^2(\alpha)}{h(\alpha)+\alpha h'(\alpha)}.$$

Notice that $h(\alpha)+\alpha h'(\alpha) = \sum_{t_j>0} \frac{t_j^2}{(t_j+\alpha)^2} y_j^2 > 0$. Differentiation yields

$$\left(\frac{C(\alpha)}{T(\alpha)}\right)' = \frac{\alpha h(\alpha)\left(2\left(h'(\alpha)\right)^2 - h(\alpha)h''(\alpha)\right)}{\left(h(\alpha)+\alpha h'(\alpha)\right)^2},$$

hence Lemma 3 applies and completes the proof of the second assertion.

The upper bound for the quotient is proved, once we have shown that

$$h^2(\alpha) < \left(h(\alpha)+\alpha h'(\alpha)\right)\|Qy_\delta\|^2. \tag{47}$$

But this follows from the Cauchy Schwartz inequality as

$$\sum_{t_j>0} \frac{t_j}{t_j+\alpha} y_j^2 = \sum_{t_j>0} \frac{t_j}{t_j+\alpha} y_j y_j \le \left(\sum_{t_j>0} \frac{t_j^2}{(t_j+\alpha)^2} y_j^2\right)^{1/2} \left(\sum_{t_j>0} y_j^2\right)^{1/2}.$$

Equality is not possible, since $\frac{t_j}{t_j+\alpha} < 1$. Finally, it holds that

$$\lim_{\alpha\searrow 0} h(\alpha) = \|Qy_\delta\|^2 \text{ as well as } \lim_{\alpha\searrow 0}\left(h(\alpha)+\alpha h'(\alpha)\right) = \|Qy_\delta\|^2,$$

from which we complete the proof of the Lemma. □

To simplify the presentation we introduce the auxiliary functions

$$f_m(\alpha) = -m'(\alpha), \tag{48}$$

$$\text{and } \rho_m(\alpha) := \|y_\delta\|^2 - \left(m(\alpha)+\alpha f_m(\alpha)\right), \quad \alpha > 0, \tag{49}$$

which should be compared with Lemma 2.

Lemma 8. *Fix any value $\alpha_0 > 0$, and let m be the model function at this point. Then $f_m(\alpha_0) = f(\alpha_0)$ and $\rho_m(\alpha_0) = \rho(\alpha_0)$.*

Moreover, it holds that

$$\lim_{\alpha\searrow 0} \rho_m(\alpha) = \|y_\delta\|^2 - \frac{C(\alpha_0)}{T(\alpha_0)} > 0. \tag{50}$$

Proof. Since, at any given α_0, it holds that $f(\alpha_0) = -h'(\alpha_0)$, the value $f_m(\alpha_0)$ coincides with $f(\alpha_0)$. Moreover, since the discrepancy ρ is expressed in terms of the value and the first derivative of the model function we also have that $\rho_m(\alpha_0) = \rho(\alpha_0)$. The last assertion follows since $\lim_{\alpha\searrow 0} m(\alpha) = C(\alpha_0)/T(\alpha_0)$, and $\lim_{\alpha\searrow 0} \alpha m'(\alpha) = 0$. □

The model function approach is algorithmically described in [6]. Precisely, at a given instance α_0, and with corresponding pair of constants $C(\alpha_0), T(\alpha_0)$, we replace Ψ_μ

from (6) and Ψ from (7) by their model counterparts

$$\Psi_\mu^{\text{model}}(\alpha) := \rho_m(\alpha)\left(f_m(\alpha)\right)^\mu, \quad \alpha > 0, \tag{51}$$

$$\text{and} \quad \Psi^{\text{model}}(\alpha) := \frac{\rho_m(\alpha)}{\alpha}, \quad \alpha > 0. \tag{52}$$

As seen from Lemma 8, the model functions will coincide with the original ones at the given parameter α_0. However, this will not be the case for the derivatives, which can be seen from Figure 3. Nonetheless, the derivatives are almost identical for small values α_0, see Figure 6, and this gives rise to the one-step proposal in § 5.3.

4.1. **Model function approach for the modified L-curve method.** Here, the search for a parameter starts at any α_0 which obeys (27), and minimizes the substitute Ψ_μ^{model} for the functional (6). If the update still obeys (27) then we continue, and if this is violated then the procedure is stopped. Therefore we have to prove that the update from α_k to the next α_{k+1} points towards zero, i.e, that $\alpha_{k+1} < \alpha_k$.

The following result is crucial for the model function approach for the modified L-curve.

Theorem 6. *Suppose that α_0 obeys (27), i.e., it holds that*

$$\mu \|r_{\alpha_0}(AA^*)y_\delta\|^2 < \alpha_0 \|x_{\alpha_0,\delta}\|^2. \tag{53}$$

The following assertions hold true.

1. *The functional Ψ_μ^{model} is increasing at α_0.*
2. *There is $0 < \alpha_1 < \alpha_0$ such that $\mu\rho_m(\alpha_1) > \alpha_1 f_m(\alpha_1)$. Hence there is a minimizer $\underline{\alpha}$ in the interval (α_1, α_0).*
3. *The minimizer obeys the first order condition $\mu\rho_m(\underline{\alpha}) = \underline{\alpha} f_m(\underline{\alpha})$.*

Proof. Analogously to Lemma 4 we have the expression for the first derivative as

$$\left(\Psi_\mu^{\text{model}}\right)'(\alpha) = f_m^{\mu-1}(\alpha) f_m'(\alpha)\left(\mu\rho_m(\alpha) - \alpha f_m(\alpha)\right).$$

from which we derive the last assertion.

At α_0 the functions ρ_m and f_m coincide with the original functions, see Lemma 8, and hence, since $f_m' < 0$, we have that $\left(\Psi_\mu^{\text{model}}\right)'(\alpha_0) > 0$, if condition (53) holds. This proves the first assertion.

To prove the second assertion we use Lemma 8 to see that $\mu\rho_m(\alpha) \to \|y_\delta\|^2 - C(\alpha_0)/T(\alpha_0) > 0$, while $\alpha f_m(\alpha) \to 0$, as $\alpha \searrow 0$. Hence, under (53) there must be $0 < \alpha_1 < \alpha_0$ for which the inequality holds. The right hand side converges to C/T as $\alpha \to 0$, such that there is α_1 for which $\mu\rho_m(\alpha_1) > \alpha_1 f_m(\alpha_1)$. Finally, as in the proof of Theorem 1 we conclude from this that there is a minimizer inside (α_1, α_0), which completes the proof of the second assertion, and of the theorem. □

Remark 9. From this theorem we obtain that the iterate $\underline{\alpha}$ satisfies $\underline{\alpha} < \alpha_0$, and hence we can continue successive minimization as long as condition (53) (with $\underline{\alpha}$ instead of α_0) holds. We depict this in the Figure 1, below.

Algorithm 1 (Modified L-curve method)

Input: y_δ, search range $[\alpha_{\min}, \alpha_{\max}]$ with
$\mu\|r_{\alpha_{\max}}(AA^*)y_\delta\|^2 < \alpha_{\max}\|x_{\alpha_{\max},\delta}\|^2$.
1: Let $k := 1$; Choose initial guess $\alpha_k := \alpha_{\max}$.
2: **While** ($\mu\|r_{\alpha_k}(AA^*)y_\delta\|^2 < \alpha_k\|x_{\alpha_k,\delta}\|^2$) **do**:
(1) build the model function $m(\alpha_k)$ at α_k;
(2) Let $\alpha_{k+1} := \arg\min \Psi_\mu^{\text{model}}$;
3: Return α_* as last parameter where **while()** condition was fulfilled.
Output: Return $x(\alpha_*)$.

FIGURE 1. The model function algorithm for the modified L-curve approach (Algorithm 1).

4.2. **Model function approach for the functional Ψ.** By the construction, it is natural to use the same model function for searching for a minimizer of the functional Ψ from (7), and the following result corresponds to Theorem 6. The proof is similar and we omit it.

Theorem 7. *Suppose that for some parameter α_0 with corresponding model function m at α_0 obeys*

$$\rho(\alpha_0) < -\alpha_0^2 m''(\alpha_0). \tag{54}$$

The following assertions hold true.

(1) *The functional Ψ^{model} is increasing at α_0.*
(2) *There is $0 < \alpha_1 < \alpha_0$ such that $\mu\rho_m(\alpha_1) > -\alpha_1^2 m''(\alpha_1)$. Hence there is a minimizer $\underline{\alpha}$ in the interval (α_1, α_0).*
(3) *The minimizer obeys the first order condition $\mu\rho_m(\underline{\alpha}) = -\underline{\alpha}^2 m''(\underline{\alpha})$.*

Remark 10. The analysis for the functional Ψ is less intuitive, since the condition (54) cannot be checked “automatically”. However, observe that

$$-\alpha_0^2 m''(\alpha_0) = \frac{2\alpha_0^2 C(\alpha_0)}{(T(\alpha_0) + \alpha_0)^3},$$

which is easy to compute, once the quantities $C(\alpha_0)$ and $T(\alpha_0)$ are computed.

5. OBSERVATIONS FROM NUMERICAL EXPERIMENTS

Here we discuss some numerical experiments, carried out in `matlab`. The tests were performed for the function `shaw(n)` with the discretization level $n = 64$. The readers are encouraged to find the background on `shaw(n)` in the original monograph [5] which will not be repeated within the current context.

5.1. **Realized discrepancy.** As we can see from Section 3, the error bound for each method is essentially dependent on the exact noise level and realized discrepancy, namely $\max\{\delta, \delta_*\}$. It is interesting to investigate how the realized discrepancy will perform for the different methods.

We first recall from Corollary 1 that the function ρ is strictly increasing. This actually shows the possibility that the realized discrepancy may be comparably large when α_* is too far away from an reasonable choice $\hat{\alpha}$ which obeys $\Theta(\hat{\alpha}) = \delta$. This is confirmed by simple numerical tests within the regularization tool box from [5]. We applied both parameter choice rules for the function `shaw(64)` along a sequence of 20 noise levels in the range $[0.0186, 0.1958]$. For the chosen 20 regularization parameters the realized discrepancies were computed. Below we draw the resulting relationship $\delta_* = \delta_*(\delta)$ for both, the modified L-Curve and the Hanke–Raus type rule. This exhibits, that both parameter choices provide amplified discrepancies,

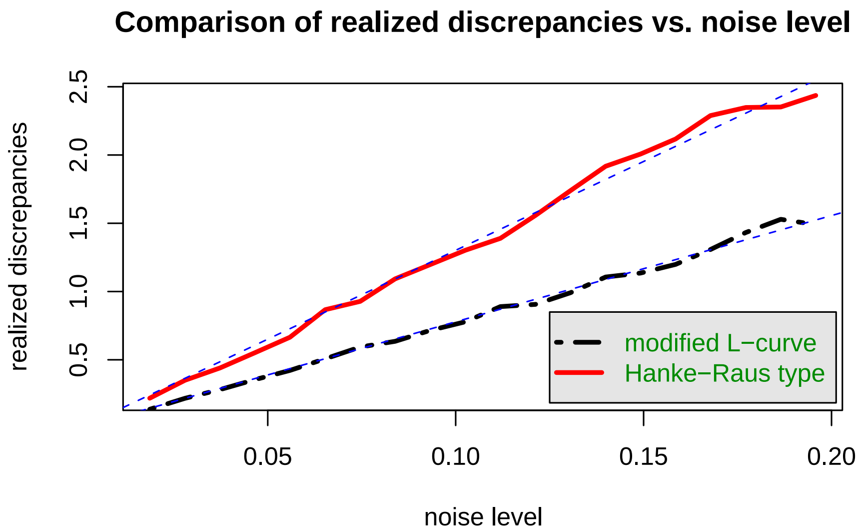

FIGURE 2. Realized discrepancies along the noise levels. The dashed lines are from linear regression $\delta_* \sim \delta$. This gives the values of $C_L = 7.782$ and $C_{HR} = 13.02$, respectively.

but the modified L-curve is better by a factor of almost two.

Remark 11. It is worth to remind that within the classical discrepancy principle, the corresponding realized discrepancy will be $C_{discr} \cdot \delta$, for some constant $C_{discr} > 1$. Our numerical tests in this subsection behave similarly to the discrepancy principle with constants $C_{discr} = 8$ and $C_{discr} = 13$, respectively.

Next, we exhibit the closeness of the model functions Ψ_μ^{model} (with $\mu = 1$) and Ψ^{model} to the original functions Ψ_μ, Ψ, at point $\alpha := 0.1$, again for the problem from above, and at different noise levels.

Model function approach for the modified L-curve method. For the modified L-curve this is shown in Figure 3. We see similar properties of the model function, and the original modified L-curve method, i.e. monotonicity and local convexity. Then, the minimization of the model function based modified L-curve method is carried out as in Algorithm 2 in [6]. It is interesting to note that at the 0.1% noise level, the realized discrepancy is almost the same as to the true noise level. For a larger 1%

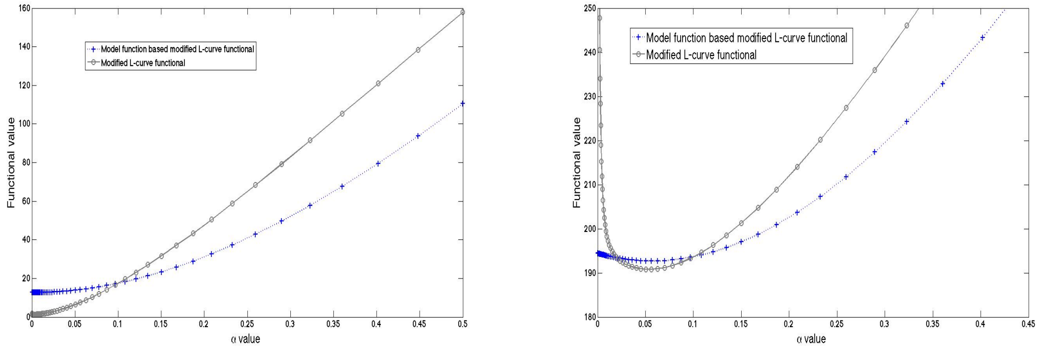

FIGURE 3. Functional value $\Psi_\mu^{\text{model}}(\alpha)$ in (51) and $\Psi_\mu(\alpha)$ in (6) on the model function based versus original modified L-curve functional. (left) $\delta = 0.019$, δ_*: 0.0238 (modified L-curve) vs. 0.0234 (model function based);
(right) $\delta = 0.18$, δ_*: 2.1287 (modified L-curve) vs. 2.0823 (model function based).

noise level, the corresponding realized discrepancy is around 10 times larger as the true noise level. This is in coincidence with Figure 2.

Model function approach for the Hanke–Raus type method. Concerning the Hanke–Raus type functional, the minimization of the model function based Hanke–Raus type functional is by a simple `matlab` function `fminbnd` which finds the minimum of a uni-variate function on a given bounded interval, i.e. $\alpha \in [10^{-4}, 0.5]$, for the minimization of (52). More information will be given in § 5.3. The value on the model function based and the exact Hanke–Raus functional can be found in Figure 4, which shows the reliability the model functional based Hanke–Raus functional which inherits the properties of the original Hanke–Raus functional. However, the realized discrepancy of each method is about 10 times larger than the noise level. The above numerical simulations show that the minimization of the model function based error free principles (51) and (52) are actually quite reliable, and in comparison with the original error free principles (6) and (7). Nevertheless, all these principles generate a comparatively large realized discrepancy. Therefore, the error bounds in Theorems 4 and 5, will also be comparatively large.

5.2. **Comparison with fixed point iteration from** [1]**.** Here we shall briefly compare the performance of the model function approach for the modified L-curve methods with the fixed point iteration presented in [1]. We carried out the same experiments as in [1, Example 1], and we shall use the same notation as there. In addition, Algorithm 2 from [6] was used in this comparison.

Precisely, the problem `shaw(64)` was solved 500 times with random data. The notation is as in [1], i.e., we let

- NL: noise level defined as $\|x_* - x\|_2/\|x\|_2$.

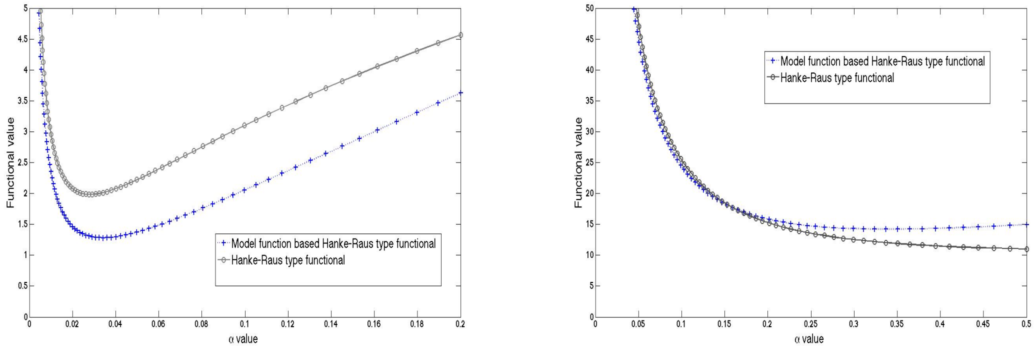

FIGURE 4. Functional value $\Psi^{\text{model}}(\alpha)$ in (52) and $\Psi(\alpha)$ in (7) on the model function based versus original Hanke–Raus type functional. (left) $\delta = 0.019$, δ_*: 0.234 (Hanke–Raus rule) vs. 0.26 (model function based);
(right) $\delta = 0.18$, δ_*: 2.347 (Hanke–Raus rule) vs. 2.01 (model function based).

- $\bar{\lambda}$: average values of the computed regularization parameter; for comparison we let $\bar{\lambda} := \sqrt{\alpha_*}$;
- E_{max}, E_{min}, $\bar{E}$: maximum, minimum, and average errors occurring in all experiments;
- $\sigma(E)$: standard deviation of computed errors;
- I_{max}, I_{min}, $\bar{I}$: maximal, minimal and average number of iterations.

For this example all runs were successful, so we skipped this row. We summarize the findings in Table 1. We observe that Algorithm 1 has a similar quality as the

	$NL = 1\%$			$NL = 5\%$		
	Alg. 1	Alg. 2	FP [1]	Alg. 1	Alg. 2	FP [1]
E_{max}	0.2628	0.2537	0.2985	0.2818	0.3404	0.2867
E_{min}	0.0436	0.0452	0.0475	0.0945	0.0590	0.0939
$\bar{E}$	0.1168	0.1278	0.1213	0.1735	0.1789	0.1728
$\sigma(E)$	0.0410	0.0413	0.0410	0.0333	0.0490	0.0335
$\bar{\lambda}$	0.0222	0.0164	0.0221	0.1155	0.0843	0.1155
I_{max}	6	6	11	6	3	12
I_{min}	3	6	5	3	3	7
$\bar{I}$	4.4	6	?	4.0	3	?

TABLE 1. Comparison of the proposed model function approach with the fixed point iteration from [1].

fixed point iteration. However, this quality is achieved with about half as many iterations, see the average number $\bar{I}$ in the table. Actually, more than 60% of the runs used 4 iterations in the 1% case, whereas in the 5% case this happend in more than 90% of the trials. This points towards superiority of the model function approach, in particular for large scale problems, since each iteration represents the solution of one ill-posed problem, and may be expensive. We mention one additional point. Algorithm 2 has a similar performance as the other algorithms at lower noise level, whereas, for $NL = 5\%$, the maximum relative error E_{max} is larger compared to both, Algorithm 1 and the fixed point iteration. However, Algorithm 2 uses a smaller number of iterations.

5.3. **One-step proposal method.** As we can see from the Section 4, the key point for the model function idea is to find a suitably approximating functional Ψ_μ or Ψ.

Simulation with the model function approach for the Hanke–Raus type functional (7) revealed an interesting *one–step method*, which we will describe, next. While in the analysis in Corollary 2, and the related model function approach from Theorem 7 we propose to start searching for a minimizer with some "large" value of α, in order to make sure that (54) holds, one might as well start with a "small" value of α, construct the corresponding model function Ψ^{model} at this point, and take as α_* the minimizer of Ψ^{model}. This is depicted in the algorithm, below. It may be hard

Algorithm 2 (One step method)

Input: y_δ, search range $[\alpha_{\min}, \alpha_{\max}]$.
1: Choose initial guess $\alpha_1 := \alpha_{\min}$.
2: Calculate the $x(\alpha_1)$ by the Tikhonov regularization; build the model function $m(\alpha)$ at α_1.
3: Search for the minimum of the single-value function Ψ^{model} on the fixed interval $[\alpha_{\min}, \alpha_{\max}]$.
4: Return the minimal point α_* as the chosen regularization parameter.
Output: Return $x(\alpha_*)$.

FIGURE 5. The one–step algorithm

to generally justify this method, but it works well as long as the functional Ψ is "typical", i.e., it has a steep negative slope for small α and becomes "flat" for larger values. As was observed in the simulations, see Figure 6, the slope of the model function Ψ^{model} closely follows the original function, and the function has a minimizer to the right of the small initial guess α_0. Again, simulations were carried out for the function `shaw(64)`.

The reliability of this approach was analyzed as follows, and we explain the details of Figure 7. We generated perturbed right–hands sides with uniformly distributed noisy data at 1% , 0.5% and 0.1% noise levels, with respect to the exact data. This was repeated 50 times for each noise level. In Figure 7, each circle represents the relative error $\frac{\|x_{\alpha_*,\delta} - x\|}{\|x\|}$ on a solution x. For the model function based Hanke–Raus type functional, the element $x_{\alpha_*,\delta}$ is the solution by the Algorithm from Figure 5. The minimization of the original Hanke–Raus type functional is performed on a geometrically increasing discrete set of parameters $\{\alpha_i\}_{i=1}^{100} \in [10^{-4}, 0.5]$, i.e., with common ratio. The minimal value of $\{\Psi(\alpha_i)\}_{i=1}^{100}$ is considered as the minimizer of the Hanke–Raus type functional, and here $x_{\alpha_*,\delta}$ is the solution at this minimizer. Figure 7 shows that the one-step method (Algorithm 2) is quite reliable in comparison with the original Hanke–Raus method.

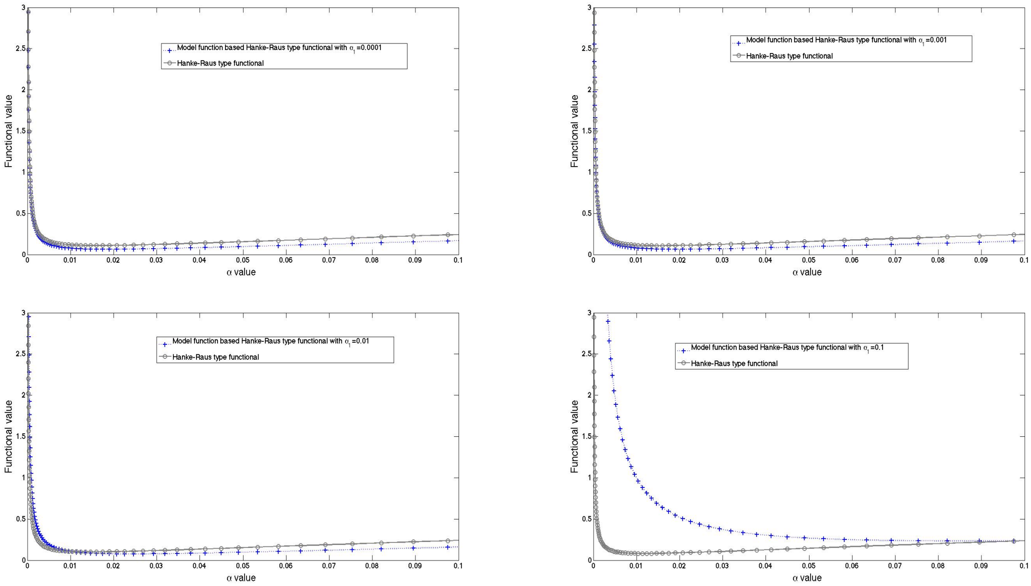

FIGURE 6. Comparison of the functionals $\Psi^{\text{model}}(\alpha)$ from (52) and $\Psi(\alpha)$ from (7).
(upper left) $\alpha_1 = 0.0001$, (upper right) $\alpha_1 = 0.001$,
(bottom left) $\alpha_1 = 0.01$, (bottom right) $\alpha_1 = 0.1$.

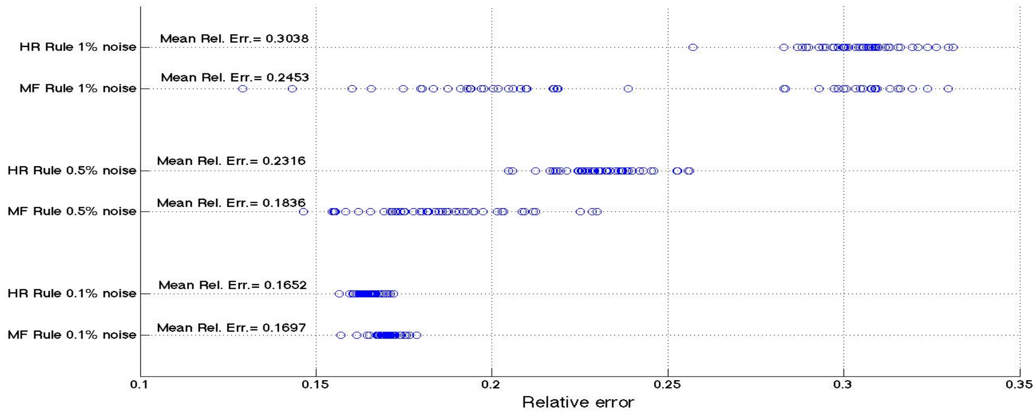

FIGURE 7. Simulation with `shaw(64)` using the model function based (MF) and original Hanke–Raus (HR) type functional. From bottom to top the noise levels are 0.1%, 0.5% and 1% of the exact solution.

A technical problem arises when we try to use a large α to generate the model function under the framework of the model function based Hanke–Raus type principle. The model function Ψ^{model} at large α does not inherit the properties of the original functional Ψ. Figure 6 shows the numerical simulations, again from `shaw(64)`, with 0.1% noise level. Here, we generated corresponding functions $\Psi^{\text{model}}(\alpha)$ with four initial $\alpha_1 = 10^{-l}$, $l = 1, \ldots, 4$. Based on the model function approach, the functional value of $\Psi^{\text{model}}(\alpha)$ and $\Psi(\alpha)$ must coincide at each initial α_1.However, for the large value $\alpha_1 := 0.1$, the "corner" points of $\Psi^{\mathrm{model}}(\alpha)$ and $\Psi(\alpha)$ differ essentially. Nevertheless, for smaller values α_1, as e.g. $\alpha_1 := 0.0001$, the model function $\Psi^{\mathrm{model}}(\alpha)$ models the original $\Psi(\alpha)$ quite well. We also stress, that even for large noise level of 1%, the original and the model functions have a good degree of approximation if α_1 is small.

Remark 12. From the numerical point of view it is not desirable to solve the problem for small values of α, as these are ill–conditioned. However, the advantage is, that this has to be done only once to give the minimizer α_*, of the model function Ψ^{model}. Finally, the problem has to be solved at α_*.

6. Conclusions

In this paper, we revisit two classical error free parameter choice rules, the modified L-curve method (6) and some Hanke–Raus type method (7). A general error analysis is performed for both methods. We emphasize that Theorem 4 is the first optimality result for the L-curve type method. Considerations on the noise level and the realized discrepancy are provided in Section 5. Then, the model function approach is considered for both methods under consideration. Here we present some model function based Hanke–Raus type method which is introduced as one-step method. Numerical simulation for the comparison between this algorithm and the original Hanke–Raus type method (7) shows the reliability of the proposed method.

References

[1] Fermín S. Viloche Bazán. Fixed-point iterations in determining the Tikhonov regularization parameter. *Inverse Problems*, 24(3):035001, 15, 2008.

[2] Fermín S Viloche Bazán and Juliano B Francisco. An improved fixed-point algorithm for determining a Tikhonov regularization parameter. *Inverse Problems*, 25(4):045007 (16pp), 2009.

[3] Heinz W. Engl, Martin Hanke, and Andreas Neubauer. *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1996.

[4] Martin Hanke and Toomas Raus. A general heuristic for choosing the regularization parameter in ill-posed problems. *SIAM J. Sci. Comput.*, 17(4):956–972, 1996.

[5] Per Christian Hansen. Regularization tools: a Matlab package for analysis and solution of discrete ill-posed problems. *Numer. Algorithms*, 6(1-2):1–35, 1994.

[6] Yi Heng, Shuai Lu, Adel Mhamdi and Sergei V. Pereverzev. Model function approach in modified L-curve method for regularization parameter choice. submitted, 2009.

[7] Bernd Hofmann, Peter Mathé, and Matthias Schieck. Modulus of continuity for conditionally stable ill-posed problems in Hilbert space. *J. Inverse Ill-Posed Probl.*, 16:567–585, 2008.

[8] Karl Kunisch and Jun Zou. Iterative choices of regularization parameters in linear inverse problems. *Inverse Problems*, 14(5):1247–1264, 1998.

[9] Peter Mathé. Saturation of regularization methods for linear ill-posed problems in Hilbert spaces. *SIAM J. Numer. Anal.*, 42(3):968–973 (electronic), 2004.

[10] Peter Mathé and Sergei V. Pereverzev. Geometry of linear ill-posed problems in variable Hilbert scales. *Inverse Problems*, 19(3):789–803, 2003.

[11] Teresa Regińska. A regularization parameter in discrete ill-posed problems. *SIAM J. Sci. Comput.*, 17(3):740–749, 1996.

[12] Jianli Xie and Jun Zou. An improved model function method for choosing regularization parameters in linear inverse problems. *Inverse Problems*, 18(3):631–643, 2002.