

HEURISTIC PARAMETER SELECTION BASED ON FUNCTIONAL MINIMIZATION: OPTIMALITY AND MODEL FUNCTION APPROACH

SHUAI LU AND PETER MATHÉ

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 type 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 reconstruction of solutions from indirect measurements is called an inverse problem, and in many cases such problems are ill-posed. In the latter case the solution does not depend continuously on the data, and regularization is required to solve such problems. The choice of the inherent regularization parameter is an important issue, or the key to successfully solving such problems. The choice depends on the unknown solution, typically through its smoothness properties, and hence a data-driven choice is desirable.

Among the many parameter choice strategies there are those with a sound mathematical foundation, and others, which are heuristic, we refer to [5] for an account of parameter choices. Such heuristic parameter choice rules cannot be convergent, i.e., reliable under any circumstances, as they do not use the noise level. This observation was first made in [1], and it is often called *Bakushinski's veto*. Nevertheless, some of these heuristic rules are widely used, as, for instance, the L-curve method, which was introduced in [8, 15]. A deeper understanding of such heuristic parameter choices is desirable. In this paper we discuss the (modified) L-curve method and a variant of the Hanke–Raus type rule. 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 us 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.

Received by the editor March 15, 2010 and, in revised form, October 7, 2011.
2010 *Mathematics Subject Classification*. Primary 65J20; Secondary 47A52.
Key words and phrases. Parameter choice, L-curve, model function.

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

$$(1) \quad y_\delta = Ax + \delta\xi,$$

where $A: X \rightarrow Y$ 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 Tikhonov regularization, where the family of potential solutions is given as

$$(2) \quad x_{\alpha,\delta} := (A^*A + \alpha I)^{-1} A^* y_\delta, \quad \alpha > 0.$$

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

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

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

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

$$(5) \quad \rho(\alpha) = \alpha^2 \|(AA^* + \alpha I)^{-1} y_\delta\|^2, \quad \alpha > 0.$$

It remains to choose the regularization parameter α . If the solution x belongs to the range $\mathcal{R}(\varphi(A^*A))$ of the operator $\varphi(A^*A)$ for some increasing function φ , then 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 penalized by the norm of $x_{\alpha,\delta}$. As was pointed out in [18], under the framework of Tikhonov regularization, this principle is related to the following modification, called the *modified L-curve* method, which aims at minimizing the functional

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

The tangent slope value is $-1/\mu$ when the regularization parameter is chosen by the L-curve method at $(\log \|Ax_{\alpha,\delta} - y_\delta\|^2, \log \|x_{\alpha,\delta}\|^2)$. 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, which shall be called the *Hanke-Raus* type rule, was presented in [5, Sect. 4.5 (4.115)], [7]. It requires minimizing the function

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

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

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

where the constant $c > 1$ is chosen appropriately. The present model function approach also applies to this principle, however, the analysis was presented in [14, 19], and we do not repeat it here.

Recently the mathematical analysis of heuristic parameter choice rules gained importance, and we mention the studies [2, 13]. In these studies properties of the noise and/or solutions which guarantee convergence of several heuristic parameter choices were addressed. This is not the objective of the present investigations. Here, the theoretical results which are related to such choices of regularization parameter focus on the *realized discrepancy*. Precisely, if by some method the parameter α_* has been chosen, then we consider

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

as the *realized discrepancy* according to this parameter choice. In principle, this quantity can be observed from the computations, and we will show that it naturally appears in error bounds.

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

- (1) Does the minimizer α_* of the functional (6) or (7), respectively, always exist within the range $(0, \|A^*A\|)$?
- (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 ?
- (4) Is there an efficient numerical procedure to find the minimizer?

Therefore our subsequent analysis is as follows. We first provide some preliminary analysis which points to 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. Most of these results are valid for general linear regularization schemes. However, in this study we focus on Tikhonov regularization. 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. The model function approach was originally proposed in [12] for its use within the discrepancy principle, and with an important modification in [14, 19]. Its full power was uncovered recently, and we refer to [16] for this new pathway. Here we extend its use to both principles under consideration, the modified L-curve method and the Hanke–Raus type rule. Some of the discussion on the modified L-curve method can also be found in [10]. That study focuses on algorithmical issues and provides a case study in some real application.

In a numerical case study we observe that the use of the Hanke–Raus type rule from (7) is limited in real applications. This is shown by relating the true noise level to the realized discrepancy. We also discuss the efficiency in relation to some fixed point iteration as recently proposed in [4].

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

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

for a given parameter $\alpha > 0$. In Hilbert space this is a smooth convex functional, and the following is known; see e.g. [5, 12, 14]. The unique minimizer $x_{\alpha, \delta}$ of (9) is

given as in (2). Moreover, the minimization problem obeys the following variational form,

$$(10) \quad \langle Ax, Ag \rangle + \alpha \langle x, g \rangle = \langle y_\delta, Ag \rangle, \quad g \in X,$$

and $x_{\alpha,\delta}$ is the unique solution to (10) in the orthogonal complement of the kernel of the operator A .

2.1. The fundamental function $h(\alpha)$. It turns out to be interesting to study the Tikhonov functional (9) at the minimizer $x_{\alpha,\delta}$, i.e., we define

$$(11) \quad J(\alpha) := T_\alpha(x_{\alpha,\delta}) = \|Ax_{\alpha,\delta} - y_\delta\|^2 + \alpha \|x_{\alpha,\delta}\|^2, \quad \alpha > 0.$$

Functionals of this type were used to derive a model function approach in [14, 19], and we will resort to this idea in Section 4.

Here we make the following observation. If we insert $g := x_{\alpha,\delta}$ in (10) then we obtain that

$$(12) \quad \langle y_\delta, Ax_{\alpha,\delta} \rangle = \|Ax_{\alpha,\delta}\|^2 + \alpha \|x_{\alpha,\delta}\|^2.$$

The latter function will be fundamental and we assign

$$(13) \quad h(\alpha) := \|Ax_{\alpha,\delta}\|^2 + \alpha \|x_{\alpha,\delta}\|^2, \quad \alpha > 0.$$

If, in real Hilbert space, we rewrite the functional J from (11) 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 (12) and the function h from (13) yields the identity

$$(14) \quad \|y_\delta\|^2 = J(\alpha) + h(\alpha), \quad \alpha > 0.$$

The behavior of these functions depends on the given data y_δ , in particular, it proves to be important whether these belong to the domain of the *Moore-Penrose inverse*, denoted by A^+ , or not. To describe this we let $Q: Y \rightarrow Y$ 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. [14, Lemma 2.3] and we recall the following.

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 true 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

$$\rho'(\alpha) = -\alpha f'(\alpha), \quad \alpha > 0.$$

Proof. We know from (14) that $J'(\alpha) = -h'(\alpha)$. We also know from [14] 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. \square

The following consequences are important; cf. [14, 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 [14, 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 the operator AA^* is compact and has singular values $t_1 \geq t_2 \geq \dots \geq 0$, and corresponding singular functions u_1, u_2, \dots . Let us abbreviate $y_j := |\langle y_\delta, u_j \rangle|$. Then we can take the representation of the solution from (2) to derive the explicit form for the function h from (13) as

$$(15) \quad h(\alpha) = \sum_{t_j > 0} \frac{t_j^2}{(t_j + \alpha)^2} y_j^2 + \sum_{t_j > 0} \frac{\alpha t_j}{(t_j + \alpha)^2} y_j^2 = \sum_{t_j > 0} \frac{t_j}{t_j + \alpha} y_j^2.$$

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*

$$(16) \quad 2(-h'(\alpha))^2 \leq h(\alpha)h''(\alpha), \quad \alpha > 0.$$

Proof. Using the Cauchy-Schwarz inequality, we have that

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

Squaring this inequality yields (16), 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 following functional, given for some fixed parameter $\mu > 0$, as in (6), that

$$\|Ax_{\alpha,\delta} - y_\delta\|^2 \|x_{\alpha,\delta}\|^{2\mu} = \|r_\alpha(AA^*)y_\delta\|^2 \|x_{\alpha,\delta}\|^{2\mu}, \quad \alpha > 0,$$

where the function $r_\alpha(t) := \alpha/(t + \alpha)$, $t > 0$ is the residual function for Tikhonov regularization; cf. (5). Properties of this function 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*

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

This can be translated into necessary and sufficient conditions for minimal values. For the smooth function Ψ_μ , the minimizer α_* must necessarily be a stationary point, i.e., $\Psi_\mu(\alpha_*) = 0$, which is called *first order condition*.

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

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

Any such stationary value α_* is a minimum provided that

$$-(\mu + 1)\alpha_* f'(\alpha_*) < f(\alpha_*).$$

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

Lemma 5. *For Tikhonov regularization we have that*

$$(17) \quad \Psi'(\alpha) = -\frac{1}{\alpha^2} (\rho(\alpha) + \alpha^2 f'(\alpha)),$$

$$(18) \quad \Psi''(\alpha) = \frac{2}{\alpha^3} (\rho(\alpha) + \alpha^2 f'(\alpha)) - \frac{1}{\alpha} (f'(\alpha) + \alpha f''(\alpha)).$$

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*

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

Any such stationary value α_* is a minimum provided that

$$\alpha_* f''(\alpha_*) < -f'(\alpha_*).$$

We turn to our first discussion of whether minimizers exist, and we start from the observation first pointed out in [18, Lemma 6]. We formulate this criterion for the functional Ψ_μ from (6), only.

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

$$(19) \quad \mu\rho(\alpha_1) > \alpha_1 f(\alpha_1) \quad \text{and} \quad \mu\rho(\alpha_0) < \alpha_0 f(\alpha_0).$$

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 (19) 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. □

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 $\alpha_0 > 0$ independent of the data y_δ such that*

$$(20) \quad \mu \|r_{\alpha_0}(AA^*)y_\delta\|^2 < \alpha_0 \|x_{\alpha_0,\delta}\|^2.$$

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 j th 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 $\alpha_0 > 0$ independent of the data y_δ such that*

$$(21) \quad \|r_{\alpha_0}(AA^*)y_\delta\|^2 < -\alpha_0^2 f'(\alpha_0).$$

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 (20) and (21), 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,

$$(22) \quad g(\alpha) := \frac{1}{\|x_{\alpha,\delta}\|}, \quad \alpha > 0.$$

Lemma 6. *For Tikhonov regularization the function g from (22) 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, as outlined in Remark 2, and obtain the explicit representation as

$$(23) \quad f(\alpha) = \sum_{t_j > 0} \frac{t_j}{(t_j + \alpha)^2} y_j^2.$$

Termwise differentiation provides us with

$$(24) \quad f'(\alpha) = -2 \sum_{t_j > 0} \frac{t_j}{(t_j + \alpha)^3} y_j^2, \quad \alpha > 0.$$

The latter function is finite for all $\alpha > 0$, such that we may interchange differentiation and summation. Observe that (24) 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:

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

Concavity of g is established if the difference in the brackets is non-positive. This will be shown by using the Cauchy–Schwarz 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

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

Squaring the above inequality yields concavity of g .

We see from (23) that $f(\alpha) \rightarrow 0$, and hence that $g(\alpha) \rightarrow \infty$, as $\alpha \rightarrow \infty$. 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 [5, Prop. 3.6]) asserts that this is the case if and only if $Qy_\delta \in \mathcal{R}(A)$, and the proof is complete. \square

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 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 (25).

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_* . The following results are valid for more general regularization schemes, but we restrict the analysis to Tikhonov regularization, as this is the main objective throughout this study. We recall, that its residual function is given as $r_\alpha(t) = \alpha/(t + \alpha)$, $t, \alpha > 0$. We allow for general smoothness assumptions, i.e., when smoothness is given in terms of *general source conditions* of the type

$$(25) \quad x \in H_\varphi := \{x, \quad x = \varphi(A^*A)v, \quad \|v\| \leq 1\},$$

for an increasing continuous function $\varphi: (0, \|A^*A\|]$ with $\varphi(0+) = 0$.

Optimality properties for the Hanke–Raus type rule were studied in [5, 7]. We also mention [6], where optimality in terms of *oracle type* results is established. Here we restrict to order optimality under general smoothness conditions, as outlined above. Our approach is based on some general bound for the bias through the modulus of continuity and the relation between true noise level δ and realized discrepancy δ_* .

3.1. Bounding the bias. The following results generalize previous analysis from [7]. 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* of the Moore–Penrose inverse A^+ , which is given as

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

see [11] for details. The following companion Θ to the smoothness function φ proves to be important, and we let

$$(26) \quad \Theta(t) := \sqrt{t}\varphi(t), \quad 0 < t \leq \|A^*A\|.$$

Proposition 1. *Let $x_{\alpha,\delta}$ be obtained from Tikhonov regularization, as in (2). 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

$$(27) \quad \|r_{\alpha_*}(A^*A)x\| \leq 2\omega(A^+, H_\varphi, \max \{ \delta, \delta_* \}).$$

Consequently, it holds true that

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

Proof. Using the triangle inequality we bound

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

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

$$\|Az\| \leq \delta + \delta_* \leq 2 \max \{ \delta, \delta_* \}.$$

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

$$z = \varphi(A^*A)r_\alpha(A^*A)v,$$

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

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

$$\|z\| \leq \omega(A^+, H_\varphi, 2 \max \{ \delta, \delta_* \}),$$

which implies (27) using monotonicity. 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 cutoff. We refer to [11] for details. \square

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 [17].

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 cutoff may be used to bound the modulus; we refer to [11] for details.

3.2. Controlling the overall error. 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 extend 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 $x \in H_\varphi$, and Tikhonov regularization is used, then the qualification is required to be at least Θ from (26). Therefore, the maximal smoothness for which optimality of the discrepancy principle can be guaranteed is (a multiple of) $\varphi(t) = \sqrt{t}$, $t > 0$, although Tikhonov regularization respects the maximal smoothness $\varphi(t) = t$, $t > 0$.

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 (22).

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

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

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 values of AA^* and the data y_δ .

We exemplify this as follows. Assumption 1, given for the function g^{-1} , translates to the following assumption for the function f from (23), which states that there is $M > 4$ such that

$$(28) \quad f(Mt) \leq \frac{1}{4} f(t), \quad t > 0.$$

If both, the singular values t_j of the Operator AA^* decay polynomially, say like $t_j = j^{-2r}$ for some $r > 1$, and if the given data y_δ have coefficients $y_j := \langle y_\delta, u_j \rangle = j^{-1}$ with respect to the singular basis of AA^* , then the following choice yields the validity of Assumption 1. Let $k \in \mathbb{N}$ be such that $k \geq 2^{1/(r-1)}$, and let $M := k^{2r}$. Then we have that $y_j^2/y_{kj}^2 = k^2 \leq M/4$, and

$$\begin{aligned} f(Mt) &= \sum_j \frac{t_j}{(t_j + Mt)^2} y_j^2 = \frac{1}{M} \sum_j \frac{t_j/M}{(t_j/M + t)^2} y_j^2 \\ &= \frac{1}{M} \sum_j \frac{t_{kj}}{(t_{kj} + t)^2} y_{kj}^2 \frac{y_j^2}{y_{kj}^2} \leq \frac{1}{4} f(t), \quad t > 0. \end{aligned}$$

In terms of the constant C_2 this translates to $C_2 \geq r/(r-1)$. This analysis may easily be extended along the same lines, and it shows that Assumption 1 is fulfilled in “typical” cases, i.e., polynomial decay of the singular values and data y_δ which have spectrum “everywhere”.

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 function Θ from (26) is concave, then

$$(29) \quad \|x - x_{\alpha_*, \delta}\| \leq \left(4 + \left(2^{\mu+1} \frac{\delta}{\delta_*} \right)^{\frac{C_2}{\mu}} \right) \varphi(\Theta^{-1}(\max\{\delta, \delta_*\})),$$

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

Proof. We use the natural error decomposition to deduce that

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

The bias was bounded in Proposition 1, and yields that

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

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

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

where we used the concavity of Θ to deduce that $\|r_{\hat{\alpha}}(AA^*)Ax\| \leq \Theta(\hat{\alpha})$. Therefore, using the realized discrepancy, we have that

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

which implies that

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

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

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

This allows us to bound the noise propagation as

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

and it gives the error bound from (29). The proof is complete. □

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 method and original L-curve method in [18, 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 function Θ from (26) is concave, then*

$$(30) \quad \|x - x_{\alpha_*, \delta}\| \leq 2 \left(1 + \frac{\delta}{\delta_*}\right) \varphi(\Theta^{-1}(\max\{\delta, \delta_*\})),$$

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

Proof. We use the natural error decomposition to deduce that

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

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

$$(31) \quad \frac{\|Ax_{\alpha_*, \delta} - y_\delta\|}{\sqrt{\alpha_*}} \leq 2\varphi(\Theta^{-1}(\delta)).$$

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

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

which proves (31). In a second step we bound the error term as

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

Overall this results in (30) and thus completes the proof. \square

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 [7].

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 for the modified L-curve method; we refer to [3, 4]. This approach uses the observation captured in Theorem 6, below. Specifically, if for some α_0 the assumption (34) is fulfilled, then the function

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

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 rule, based on Theorem 7. We shall compare the performance of both, the iterative approach from [3] 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 [12], and later furnished in [14, 19]. While these studies consider the use of model functions for finding the regularization parameter based on the discrepancy principle, the recent study [10] proposes to use this in context of the modified L-curve method 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 (15) to one term, i.e., replace the function h by a two-parameter family,

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

Remark 8. Notice, that in our approach to the model function, the sign of C is positive, in contrast to the function introduced in [19].

The parameters C and T are determined *pointwise*, i.e., at a 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

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

Thus, for any 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

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

We shall first analyze C and T as functions of the parameter α . The following monotonicity properties were shown in [19, Lemmas 3.1–3.3]. Recall the definition of the projection Q from Section 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 $\alpha \rightarrow \frac{C(\alpha)}{T(\alpha)}$ is non-increasing.*
- (3) *It holds that $\frac{C(\alpha)}{T(\alpha)} < \|Qy_\delta\|^2$ 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 (13) we have that $C = h^2/f$ and $T + \alpha = h/f$. Differentiating this yields that

$$\begin{aligned} C' &= \frac{h}{f} (-2f^2 + h(-f')), \text{ and} \\ T' &= (T + \alpha)' - 1 = \frac{-2f^2 + h(-f')}{f^2}. \end{aligned}$$

In both cases the numerators 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

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

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

$$h^2(\alpha) < (h(\alpha) + \alpha h'(\alpha)) \|Qy_\delta\|^2.$$

But this follows from the Cauchy-Schwarz inequality as

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

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

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

from which we complete the proof of the lemma. □

To simplify the presentation we introduce the auxiliary functions

$$\begin{aligned} f_m(\alpha) &= -m'(\alpha), \quad \text{and} \\ \rho_m(\alpha) &:= \|y_\delta\|^2 - (m(\alpha) + \alpha f_m(\alpha)), \quad \alpha > 0, \end{aligned}$$

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

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

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 [10]. Precisely, at a given instance α_0 , and with the corresponding pair of constants $C(\alpha_0), T(\alpha_0)$, we replace Ψ_μ from (6) and Ψ from (7) by their model counterparts

$$(32) \quad \Psi_\mu^{\text{model}}(\alpha) := \rho_m(\alpha) (f_m(\alpha))^\mu, \quad \alpha > 0, \text{ and}$$

$$(33) \quad \Psi^{\text{model}}(\alpha) := \frac{\rho_m(\alpha)}{\alpha}, \quad \alpha > 0.$$

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 [10, Fig. 1]. Nonetheless, the derivatives are almost identical for small values α_0 (see Figure 6) and this gives rise to the one-step proposal in §5.4.

4.1. Model function approach for the modified L-curve method. Here, the search for a parameter starts at any α_0 which obeys (20), and minimizes the substitute Ψ_μ^{model} for the functional (6). If the update still obeys (20) 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 method.

Theorem 6. *Suppose that α_0 obeys (20), i.e., it holds that*

$$(34) \quad \mu \|r_{\alpha_0}(AA^*)y_\delta\|^2 < \alpha_0 \|x_{\alpha_0, \delta}\|^2.$$

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 $\mu\rho_m(\underline{\alpha}) = \underline{\alpha} f_m(\underline{\alpha})$.*

Remark 9. A similar result can be found in [10], i.e., Lemmas 3.4, 3.5., ibid. in connection with the algorithm which was discussed there. Here we provide a simplified proof based on Lemmas 4 and 8.

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

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

from which we derive the last assertion.

At α_0 the functions ρ_m and f_m coincide with the original functions, as was shown in Lemma 8. Hence, since $f_m' < 0$, we have that $(\Psi_\mu^{\text{model}})'(\alpha_0) > 0$, if condition (34) holds. This proves the first assertion.

To prove the second assertion we use Lemma 8 to see that $\mu\rho_m(\alpha) \rightarrow \|y_\delta\|^2 - C(\alpha_0)/T(\alpha_0) > 0$, while $\alpha f_m(\alpha) \rightarrow 0$, as $\alpha \searrow 0$. Hence, under (34) there must be $0 < \alpha_1 < \alpha_0$ for which the inequality holds. The right-hand side converges to C/T as $\alpha \rightarrow 0$, 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. \square

Remark 10. 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 (34) (with $\underline{\alpha}$ instead of α_0) holds. This gives rise to the algorithm depicted in Figure 1, below.

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*

$$(35) \quad \rho(\alpha_0) < -\alpha_0^2 m''(\alpha_0).$$

The following assertions hold true.

- (1) *The functional Ψ^{model} is increasing at α_0 .*

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 $x_{\alpha_*, \delta}$.

FIGURE 1. The model function algorithm for the modified L-curve method (Algorithm 1), cf. [10, Algorithm 1].

- (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 11. The analysis for the functional Ψ is less intuitive, since the condition (35) 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 [9] 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 be performed between the different methods.

We first recall, from Corollary 1, that the function ρ is strictly increasing. It actually shows the possibility that the realized discrepancy may be comparably large when α_* is too far away from a reasonable choice $\hat{\alpha}$ which obeys $\Theta(\hat{\alpha}) = \delta$. This is confirmed by simple numerical tests within the regularization tool box from [9]. We apply both parameter choice rules for the function `shaw(64)` along a geometric sequence of 100 noise levels in the range $[10^{-5}, 0.018]$. For the chosen 100 noise levels the realized discrepancies were computed. Below we draw the resulting ratios δ_*/δ for both, the modified L-curve method and the Hanke–Raus type rule, respectively.

As can be seen from Figure 2, both methods provide a good approximation of the realized discrepancy with respect to the exact noise level. The modified L-curve method performs more stably compared with the Hanke–Raus type rule in a reasonable range of noise levels.

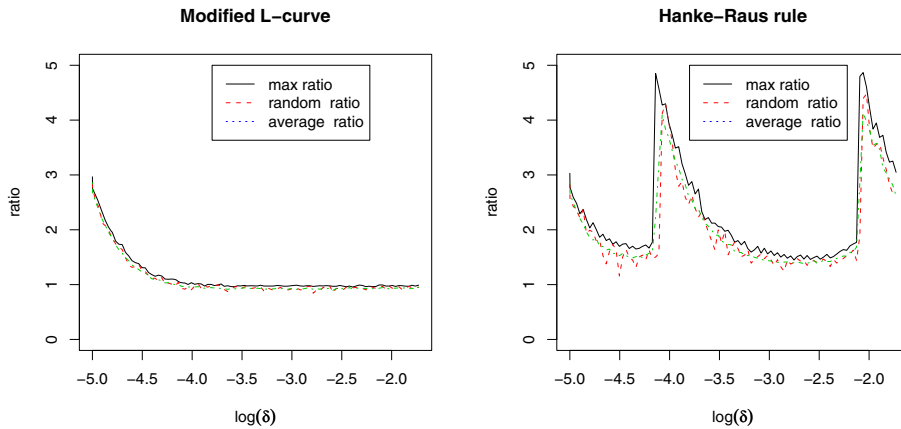


FIGURE 2. The ratio between the realized discrepancies and the noise levels. The left picture for the modified L-curve method and the right picture for the Hanke–Raus type rule. The solid line presents a worst ratio; the dotted line presents an average ratio and the dashed line presents a random ratio with respect to 20 times tests for each noise level.

5.2. Original function versus model function. Next, we exhibit the closeness of the model functions Ψ_μ^{model} (with $\mu = 1$) to the original functions Ψ_μ , at an initial parameter point $\alpha = 0.1$, again for the problem mentioned above, and at different noise levels. The comparison between Ψ^{model} and Ψ , respectively, at an initial parameter point $\alpha = 10^{-5}$ are also shown.

Model function approach for the modified L-curve method. The properties of the model function and the original modified L-curve method, i.e., monotonicity and convexity can be partially observed in [10, Fig. 1]. The minimization of the model function based modified L-curve method is carried out as in Algorithm 1. The realized discrepancy coincides with Figure 2.

Model function approach for the Hanke–Raus type rule. 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 univariate function on a given bounded interval, i.e., $\alpha \in [10^{-4}, 0.5]$, for the minimization of (33). More information will be given in §5.4. The comparison between the model function based and the original Hanke–Raus type functional can be found in Figure 3, which shows the reliability of the model function based Hanke–Raus type functional which inherits the properties of the original Hanke–Raus type functional.

The above numerical simulations show that the minimization of the model function based error free principles (32) and (33) are actually quite reliable, and in comparison with the original error free principles (6) and (7). Meanwhile, both principles generate a comparatively small realized discrepancy. Therefore, the error bounds in Theorems 4 and 5 will also be comparatively small.

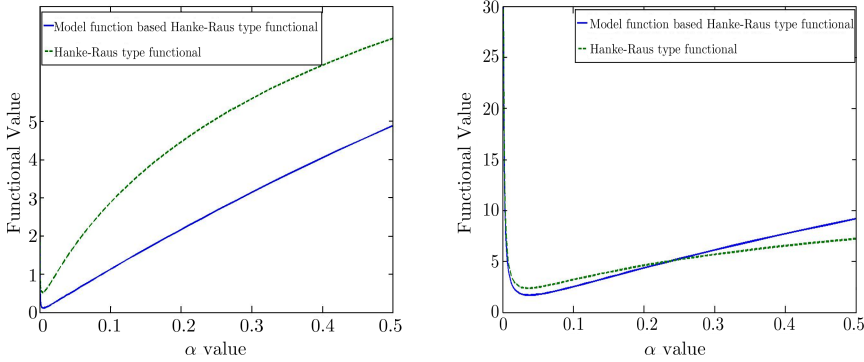


FIGURE 3. Functional value $\Psi^{\text{model}}(\alpha)$ in (33) and $\Psi(\alpha)$ in (7) on the model function based versus original Hanke–Raus type functional. (left) $\delta = 0.019$, $\delta_* = 0.0461$ (Hanke–Raus type rule) vs. 0.0515 (model function based); (right) $\delta = 0.1864$, $\delta_* = 0.2877$ (Hanke–Raus type rule) vs. 0.2997 (model function based).

5.3. Comparison with the fixed-point iteration from [3]. 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 [3]. We carried out the same experiments as in [3, Example 1], and we shall use the same notation as there.

Precisely, the problem `shaw(64)` was solved 500 times with different random noise. The notation is the same as in [3], i.e., we let

- SR: successful runs;
- NL : noise level defined as $\|y_\delta - y\|_2 / \|y\|_2$;
- $\bar{\lambda}$: average values of the computed regularization parameter; for comparison we let $\bar{\lambda} := \sqrt{\bar{\alpha}_*}$;
- 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.

In this example all runs were successful, so we skip the row of Successful Runs. We summarize the findings in Table 1. We observe that Algorithm 1 has a similar quality as the 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 happened 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.

5.4. 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 of Corollary 2, and the related model function approach from Theorem 7 we propose to start searching for a minimizer with some “large” value

TABLE 1. Comparison of the proposed model function approach with the fixed-point iteration from [3].

	$NL = 1\%$		$NL = 5\%$	
	Alg. 1	FP [3]	Alg. 1	FP [3]
E_{\max}	0.2628	0.2985	0.2818	0.2867
E_{\min}	0.0436	0.0475	0.0945	0.0939
\bar{E}	0.1168	0.1213	0.1735	0.1728
$\sigma(E)$	0.0410	0.0410	0.0333	0.0335
$\bar{\lambda}$	0.0222	0.0221	0.1155	0.1155
I_{\max}	6	11	6	12
I_{\min}	3	5	3	7
\bar{I}	4.4	-	4.0	-

of α , in order to make sure that (35) 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 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 functional has a minimizer to the right of the small initial guess α_0 . Again, simulations were carried out for the function `shaw(64)`.

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 $x_{\alpha_*, \delta}$.
-

FIGURE 4. The one-step algorithm

The reliability of this approach was analyzed as follows, and we explain the details of Figure 5. We generated perturbed right-hand sides with uniformly distributed noisy data at 1% and 0.1% noise levels, with respect to the exact ones. The random noise was produced 50 times for each noise level. In Figure 5, 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 2 in Figure 4. The minimization of the original Hanke–Raus type functional is performed on a geometrically increasing discrete set of parameters $\{\alpha_i\}_{i=1}^{100} \in [5 \cdot 10^{-7}, 0.5]$, i.e., with common ratio. Define α_* , which produces the minimal value over $\{\Psi(\alpha_i)\}_{i=1}^{100}$, as the minimizer of the Hanke–Raus type functional, and here $x_{\alpha_*, \delta}$ is the solution at this minimizer. Figure 5 shows that the one-step method (Algorithm 2) is quite

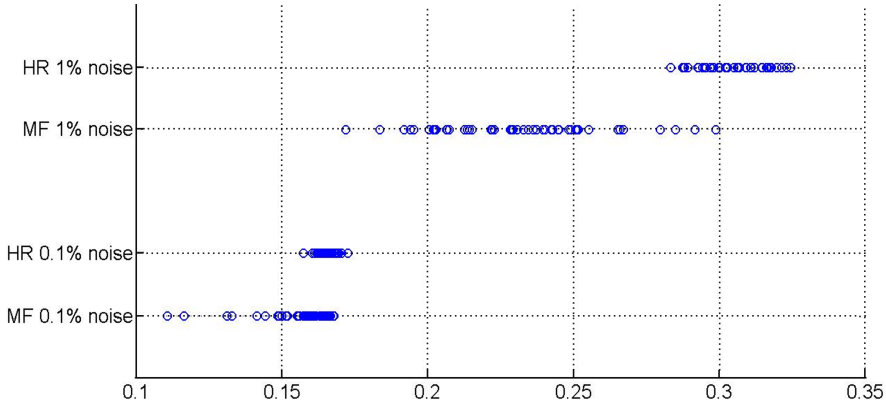


FIGURE 5. The relative error for the simulation with $\text{shaw}(64)$ using different type of functionals. The smaller the relative error the better the reconstruction. **MF** for the modeling function based Hanke Raus functional and **HR** for the original Hanke–Raus functional.

reliable in comparison with the original Hanke–Raus type rule.

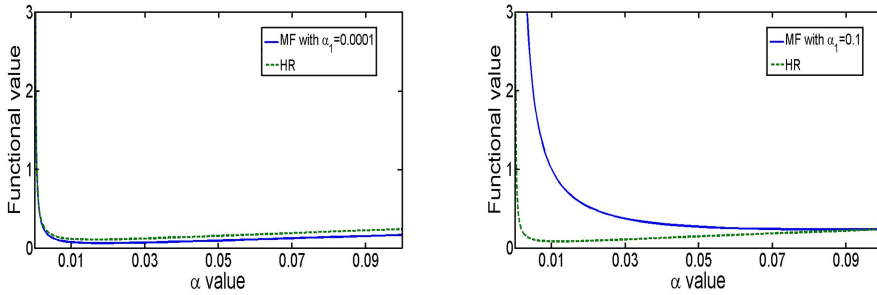


FIGURE 6. Comparison of the functionals $\Psi^{\text{model}}(\alpha)$ from (33) and $\Psi(\alpha)$ from (7) with respect to different initial guess α_1 in Algorithm 2. (left) $\alpha_1 = 0.0001$, (right) $\alpha_1 = 0.1$. **MF** for the modeling function based Hanke Raus functional and **HR** for the original Hanke–Raus functional.

A technical problem arises when we try to use a large α_1 to generate the model function under the framework of the model function based Hanke–Raus type rule. The model function Ψ^{model} at large α_1 does not inherit the properties of the original functional Ψ . Figure 6 shows the numerical simulations, again from $\text{shaw}(64)$, with 0.1% noise level. Here, we generated corresponding functions $\Psi^{\text{model}}(\alpha)$ with two initial $\alpha_1 = 10^{-l}$, $l = 1, 4$. 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, with a large value $\alpha_1 := 0.1$, the “minimal” points of $\Psi^{\text{model}}(\alpha)$ and $\Psi(\alpha)$ differ essentially. Nevertheless, with smaller values α_1 , e.g., $\alpha_1 := 0.0001$, the model function $\Psi^{\text{model}}(\alpha)$ mimics the original $\Psi(\alpha)$ quite well. We also stress, even for a large noise level 1%,

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

ACKNOWLEDGEMENTS

This work was initiated 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 in Berlin.

The first author gratefully acknowledges support from the Shanghai Science and Technology Commission Grants: 11ZR1402800, 11PJ1400800 and National Natural Science Foundation of China Grant: 11101093. He is currently a research fellow of the Alexander von Humboldt Foundation.

REFERENCES

- [1] A. B. BAKUSHINSKIĬ, *Remarks on choosing a regularization parameter using the quasi-optimality and ration criterion*, Zh. Vychisl. Mat. i Mat. Fiz., 24 (1984), pp. 181–182. MR760252 (86b:65053)
- [2] F. BAUER AND S. KINDERMANN, *The quasi-optimality criterion for classical inverse problems*, Inverse Problems, 24 (2008), 035002, 20. MR2421956 (2009f:65301)
- [3] F. S. V. BAZÁN, *Fixed-point iterations in determining the Tikhonov regularization parameter*, Inverse Problems, 24 (2008), 035001, 15. MR2421955 (2010c:65081)
- [4] F. S. V. BAZÁN AND J. B. FRANCISCO, *An improved fixed-point algorithm for determining a Tikhonov regularization parameter*, Inverse Problems, 25 (2009), 045007 (16pp). MR2482158 (2010h:65053)
- [5] H. W. ENGL, M. HANKE, AND A. NEUBAUER, *Regularization of inverse problems*, vol. 375 of Mathematics and its Applications, Kluwer Academic Publishers Group, Dordrecht, 1996. MR1408680 (97k:65145)
- [6] U. HÄMARIK AND T. RAUS, *On the choice of the regularization parameter in ill-posed problems with approximately given noise level of data*, J. Inverse Ill-Posed Probl., 14 (2006), pp. 251–266. MR2241272 (2007e:65052)
- [7] M. HANKE AND T. RAUS, *A general heuristic for choosing the regularization parameter in ill-posed problems*, SIAM J. Sci. Comput., 17 (1996), pp. 956–972. MR1395358 (97c:65100)
- [8] P. C. HANSEN, *Analysis of discrete ill-posed problems by means of the L-curve*, SIAM Rev., 34 (1992), pp. 561–580. MR1193012 (93k:65035)

- [9] ———, *Regularization tools: a Matlab package for analysis and solution of discrete ill-posed problems*, Numer. Algorithms, 6 (1994), pp. 1–35.
- [10] Y. HENG, S. LU, A. MHAMDI, AND S. V. PEREVERZYEY, *Model functions in the modified L -curve method—case study: The heat flux reconstruction in pool boiling*, Inverse Problems, 26 (2010), 13pp. MR2647148 (2011g:65065)
- [11] B. HOFMANN, P. MATHÉ, AND M. SCHIECK, *Modulus of continuity for conditionally stable ill-posed problems in Hilbert space*, J. Inverse Ill-Posed Probl., 16 (2008), pp. 567–585. MR2458286 (2009m:47026)
- [12] K. ITO AND K. KUNISCH, *On the choice of the regularization parameter in nonlinear inverse problems*, SIAM J. Optim., 2 (1992), pp. 376–404. MR1172497 (93h:65086)
- [13] S. KINDERMANN AND A. NEUBAUER, *On the convergence of the quasioptimality criterion for (iterated) Tikhonov regularization*, Inverse Probl. Imaging, 2 (2008), pp. 291–299. MR2395145 (2009k:47038)
- [14] K. KUNISCH AND J. ZOU, *Iterative choices of regularization parameters in linear inverse problems*, Inverse Problems, 14 (1998), pp. 1247–1264. MR1654635 (2000b:65110)
- [15] C. L. LAWSON, AND R. J. HANSON, *Solving least squares problems*. Prentice-Hall Series in Automatic Computation. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1974. MR0366019 (51:2270)
- [16] S. LU, S. V. PEREVERZEV, AND U. TAUTENHAHN, *Dual regularized total least squares and multi-parameter regularization*, Comput. Meth. in Appl. Math., 8 (2008), pp. 253–262. MR2641318 (2011d:65139)
- [17] P. MATHÉ, *Saturation of regularization methods for linear ill-posed problems in Hilbert spaces*, SIAM J. Numer. Anal., 42 (2004), pp. 968–973 (electronic). MR2112789 (2005m:65093)
- [18] T. REGIŃSKA, *A regularization parameter in discrete ill-posed problems*, SIAM J. Sci. Comput., 17 (1996), pp. 740–749. MR1384263 (96m:65045)
- [19] J. XIE AND J. ZOU, *An improved model function method for choosing regularization parameters in linear inverse problems*, Inverse Problems, 18 (2002), pp. 631–643. MR1910192 (2003j:65058)

SCHOOL OF MATHEMATICAL SCIENCES, FUDAN UNIVERSITY, SHANGHAI 200433, CHINA
E-mail address: slufudan.edu.cn

WEIERSTRASS INSTITUTE FOR APPLIED ANALYSIS AND STOCHASTICS, MOHRENSTRASSE 39, 10117
BERLIN, GERMANY
E-mail address: mathe@wias-berlin.de