

A SIMPLIFIED L-CURVE METHOD AS ERROR ESTIMATOR*

STEFAN KINDERMANN[†] AND KEMAL RAIK[†]

Abstract. The L-curve method is a well known heuristic method for choosing the regularization parameter for ill-posed problems by selecting it according to the maximal curvature of the L-curve. In this article, we propose a simplified version that replaces the curvature essentially by the derivative of the parameterization on the y -axis. This method shows a similar behaviour to the original L-curve method, but unlike the latter, it may serve as an error estimator under typical conditions. Thus, we can accordingly prove convergence for the simplified L-curve method.

Key words. heuristic parameter choice, L-curve method, regularization

AMS subject classifications. 47A52, 65F22, 65J20, 65J22

1. Introduction. The L-curve criterion is one of the best-known heuristic methods for choosing the regularization parameter in various regularization methods for ill-posed problems. One of the first instances of an L-curve graph appeared in the book by Lawson and Hanson [27], although it was not related to a parameter choice procedure. That it can be the basis for a parameter choice method was originally suggested by Hansen [15], further analyzed by Hansen and O’Leary [19], and popularized by Hansen, e.g., in [17].

The methodology is well known: Suppose that we are faced with the problem of solving an ill-posed problem of the form

$$(1.1) \quad Ax = y,$$

where y are data and $A : X \rightarrow Y$ is a continuous linear operator between Hilbert spaces which lacks a continuous inverse. Moreover, we assume that only noisy data

$$y_\delta = y + e, \quad \|e\| \leq \delta, \quad y = Ax^\dagger,$$

are available, where x^\dagger denotes the “true” unknown solution (or, more precisely, the minimal-norm solution). Here, e denotes an unknown error, and its norm is called the noise-level δ . In the case of heuristic parameter choice rules, which the L-curve method is an example of, this noise-level is considered unavailable.

As the inverse of A is not bounded, the problem (1.1) cannot be solved by classical inversion algorithms, rather, a regularization scheme has to be applied [8]. That is, one constructs a one-parametric family of continuous operators $(R_\alpha)_\alpha$, with $\alpha > 0$, that in some sense approximates the inverse of A for $\alpha \rightarrow 0$.

An approximation to the true solution of (1.1), denoted as x_α^δ , is computed by means of the regularization operators:

$$x_\alpha^\delta = R_\alpha y_\delta.$$

A delicate issue in regularization schemes is the choice of the regularization parameter α , and the standard methods make use of the noise-level δ . However, in situations when this is not available, so-called *heuristic* parameter choice methods [20] are proposed. The L-curve method selects an α corresponding to the corner point of the graph $(\log(\|Ax_\alpha^\delta - y_\delta\|), \log(\|x_\alpha^\delta\|))$ parameterized by α .

*Received August 27, 2019. Accepted December 29, 2020. Published online on January 30, 2020. Recommended by Giuseppe Rodriguez. This work was supported by the Austrian Science Fund (FWF) project P 30157-N31.

[†]Industrial Mathematics Institute, Johannes Kepler University Linz, Altenbergerstraße 69, 4040 Linz, Austria (`{kindermann,kemal.raik}@indmath.uni-linz.ac.at`).

Recently, [20, 21, 23] a convergence theory for certain heuristic parameter choice rules has been developed. Essential in this analysis is a restriction on the noise that rules out noise that is “too regular”. Such restrictions in the form of Muckenhoupt-type conditions were used in [20, 23] and are currently the standard tool in the analysis of heuristic rules. If these conditions hold, then several well-known heuristic parameter choice rules serve as error estimators for the total error in typical regularization schemes, and convergence and convergence rate results follow.

The L-curve method, however, does not seem to be accessible to such an analysis, although some of its properties were investigated, for instance, by Hansen [15, 18] and Reginska [35]. Nevertheless, it does not appear that it can be related to any error estimators directly.

There are various suggestions for efficient practical implementations of the L-curve method, like Krylov-space methods [6, 36] or model functions [28]. Note that the method is also implemented in Hansen’s Regularization Tools [16]. A generalization of the L-curve method in the form of the Q-curve method was recently suggested by Raus and Hämarik [32]. Other simplifications or variations are the V-curve [9] or the U-curve [26]. Some overview and comparisons of other heuristic and non-heuristic methods are given in [2, 12, 13] and the PhD. thesis of Palm [30].

The aim of this article is to propose a simplified version of the L-curve method by dropping several terms in the expression for the curvature of the L-graph. We argue that this simplified version does not alter the original method significantly, and, moreover, we prove that the simplified L-curve has error estimating capabilities similar to several other well-known heuristic methods. This allows us to state conditions under which we can verify convergence of the simplified L-curve method.

1.1. The L-Curve method and its simplification for Tikhonov regularization. We use a standard setting of an ill-posed problem of the form (1.1). Although not necessary for our analysis and only used for clarity, we assume that A is a compact operator and also (again just for simplicity) that A is injective (the nullspace of A is 0). Then the operator A has a singular value decomposition (SVD) $(\sigma_i, v_i, u_i)_{i \in \mathbb{N}}$, with the positive singular values σ_i and the singular functions $v_i \in X, u_i \in Y$ such that

$$Ax = \sum_i \sigma_i \langle v_i, x \rangle u_i, \quad \lambda_i := \sigma_i^2 > 0,$$

where $\langle \cdot, \cdot \rangle$ denotes the scalar product in X (or also in Y). As regularization operator, we employ Tikhonov regularization, which defines a regularized solution to (1.1) (an approximation to the true solution x^\dagger) via

$$(1.2) \quad x_\alpha^\delta := (A^*A + \alpha I)^{-1} A^* y_\delta.$$

Here, $\alpha \in (0, \alpha_{\max})$ is the regularization parameter. For notational purposes we also define the (negative) residual p_α^δ and the regularized solution with exact data x_α :

$$\begin{aligned} p_\alpha^\delta &:= y_\delta - Ax_\alpha^\delta = \alpha(AA^* + \alpha I)^{-1} y_\delta, \\ x_\alpha &:= (A^*A + \alpha I)^{-1} A^* y. \end{aligned}$$

The overall goal of a good parameter choice is always to minimize the total error $\|x_\alpha^\delta - x^\dagger\|$, which can be bounded by the sum of the *stability error* $\|x_\alpha^\delta - x_\alpha\|$ and the *approximation error* $\|x_\alpha - x^\dagger\|$:

$$(1.3) \quad \|x_\alpha^\delta - x^\dagger\| \leq \|x_\alpha^\delta - x_\alpha\| + \|x_\alpha - x^\dagger\|.$$

It is well known that the approximation error can in general decay arbitrarily slowly. In order to establish bounds for it and thus derive convergence rates, one has to postulate a certain smoothness condition on x^\dagger in the form of a source condition: Here, we focus on Hölder source conditions, i.e., such a source condition holds if x^\dagger can be expressed as

$$(1.4) \quad x^\dagger = (A^*A)^\mu \omega, \quad \|\omega\| \leq C, \quad \mu > 0.$$

In terms of the SVD, x^\dagger satisfies (1.4) if (and only if)

$$\sum_i \frac{|\langle x^\dagger, v_i \rangle|^2}{\lambda_i^{2\mu}} < \infty.$$

If this is the case, then for Tikhonov regularization we have that

$$(1.5) \quad \|x_\alpha - x^\dagger\| \leq C\alpha^\mu, \quad \text{for } 0 \leq \mu \leq 1,$$

and consequently the convergence rate

$$\|x_\alpha^\delta - x^\dagger\| \leq C\delta^{\frac{2\mu}{2\mu+1}}, \quad \text{for } 0 \leq \mu \leq 1,$$

which is known to be the optimal order of convergence under (1.4) (for $\mu \leq 1$). Observe the saturation effect of Tikhonov regularization, which means that the rates do not improve for higher source conditions beyond $\mu > 1$; see, e.g., [8].

1.2. The L-curve. The L-curve is a plot of the (logarithm of the) residual against the (logarithm of the) norm of the regularized solution. Define the following curve parameterized by the regularization parameter α :

$$\kappa(\alpha) = \log(\|p_\alpha^\delta\|^2) = \log(\|Ax_\alpha^\delta - y_\delta\|^2), \quad \chi(\alpha) = \log(\|x_\alpha^\delta\|^2).$$

Then a plot of the curve

$$(1.6) \quad \alpha \rightarrow \begin{bmatrix} \kappa(\alpha) \\ \chi(\alpha) \end{bmatrix},$$

yields a graph, which often resembles the shape of an “L”, hence its name L-curve. The idea of the L-curve method is to choose α as the curve parameter that corresponds to the *corner point* of the “L”. Since a corner has large curvature, the operational definition of the parameter selection by the L-curve is that of the maximizer (over the selected range of α) of the curvature of the L-graph, i.e., $\alpha =: \alpha^*$ is selected as

$$\alpha^* = \operatorname{argmax}_\alpha \gamma(\alpha),$$

with the signed curvature defined as (see, e.g., [18]),

$$\gamma(\alpha) = \frac{\chi''(\alpha)\kappa'(\alpha) - \chi'(\alpha)\kappa''(\alpha)}{(\chi'(\alpha)^2 + \kappa'(\alpha)^2)^{3/2}}.$$

Here, a prime $'$ denotes differentiation with respect to α . For Tikhonov regularization and many other methods, it is not difficult to realize that $\kappa(\alpha)$ is strictly monotonically decreasing in α , hence, the L-curve can be considered as a graph of a function $f = \chi(\kappa^{-1})$.

As already observed by several authors [14, 18, 37], for Tikhonov regularization, the curvature can be expressed without second-order derivatives and can be reduced to

$$(1.7) \quad \gamma(\alpha) = \frac{\eta\rho}{|\eta'|} \frac{\rho\eta + \alpha\eta'\rho + \alpha^2\eta'\eta}{(\rho^2 + \alpha^2\eta^2)^{\frac{3}{2}}},$$

where

$$\eta = \eta(\alpha) := \|x_\alpha^\delta\|^2, \quad \rho = \rho(\alpha) := \|p_\alpha^\delta\|^2.$$

The following lemma investigates this expression:

LEMMA 1.1. *We have*

$$(1.8) \quad \gamma(\alpha) = \frac{\eta}{\alpha|\eta'|} \frac{\zeta^2}{(\zeta^2 + 1)^{\frac{3}{2}}} - \frac{\zeta(1 + \zeta)}{(\zeta^2 + 1)^{\frac{3}{2}}}, \quad \text{with } \zeta = \zeta(\alpha) := \frac{\rho}{\alpha\eta}$$

$$(1.9) \quad =: \frac{\eta}{\alpha|\eta'|} c_1(\zeta) - c_2(\zeta),$$

where

$$0 \leq c_1(\zeta) \leq \frac{2}{3\sqrt{3}}, \quad 0 \leq c_2(\zeta) \leq \frac{1}{\sqrt{2}}.$$

Proof. The expression (1.7) can easily be rewritten as (1.8) with

$$c_1(\zeta) = \frac{\zeta^2}{(\zeta^2 + 1)^{\frac{3}{2}}}, \quad c_2(\zeta) = \frac{\zeta(1 + \zeta)}{(\zeta^2 + 1)^{\frac{3}{2}}}.$$

By elementary calculus, we may find the maxima for c_1 at $\zeta = \sqrt{2}$ and for c_2 at $\zeta = 1$ yielding the upper bounds. \square

According to the rationale for the L-curve method, we are searching for a corner of the L-graph, i.e., by definition a point where $\gamma(\alpha)$ has a *large positive* value. (An ideal corner has infinite curvature.) Thus, according to (1.9), the only term in the previous lemma that could contribute to large values is $\frac{\eta}{\alpha|\eta'|}$. Hence, backed by Lemma 1.1, we propose to remove the ζ -dependent term and, instead of (1.6), maximize the functional

$$\alpha^* = \operatorname{argmax}_\alpha \frac{\eta}{\alpha|\eta'|},$$

which leads to the simplified L-curve methods of this article. Instead of maximization, we may equivalently consider minimizing the reciprocal, and as $\eta' \leq 0$, we may replace $|\eta'|$ by $-\eta'$. Moreover, we propose two versions of the simplified method (the factor $\frac{1}{2}$ below is introduced for notational purposes and is irrelevant for the analysis and the method):

DEFINITION 1.2. *The simple-L method selects the regularization parameter α as the minimizer (over a range of α -values) of the simple-L functional:*

$$(1.10) \quad \alpha^* = \operatorname{argmin}_\alpha \psi_{SL}(\alpha),$$

$$\psi_{SL}(\alpha) := \left(-\frac{1}{2} \alpha \eta'(\alpha) \right)^{\frac{1}{2}} = \left(-\left\langle x_\alpha^\delta, \alpha \frac{\partial}{\partial \alpha} x_\alpha^\delta \right\rangle \right)^{\frac{1}{2}}.$$

The simple-L ratio method selects α as the minimizer (over a range of α -values) of

$$\alpha^* = \operatorname{argmin}_\alpha \psi_{SLR}(\alpha),$$

$$\psi_{SLR}(\alpha) := \left(-\frac{1}{2} \frac{\alpha \eta'(\alpha)}{\eta(\alpha)} \right)^{\frac{1}{2}} = \left(\frac{-\left\langle x_\alpha^\delta, \alpha \frac{\partial}{\partial \alpha} x_\alpha^\delta \right\rangle}{\|x_\alpha^\delta\|^2} \right)^{\frac{1}{2}} = \frac{\psi_{SL}(\alpha)}{\|x_\alpha^\delta\|}.$$

The main advantage that these simplified L-curve methods hold is that under certain conditions, they serve as error estimators and convergence of the associated parameter choice methods can be proven in contrast to the original L-curve method.

Another reason for using the simplified functionals is that ψ_{SL} resembles and can be compared with several other heuristic parameter choice functionals, which are known to have error-estimating properties. For instance the quasi-optimality (QO) principle defines α as the minimizer of

$$\psi_{QO}(\alpha) := \left\| \alpha \frac{\partial}{\partial \alpha} x_{\alpha}^{\delta} \right\|,$$

while the heuristic discrepancy (HD) principle defines it as the minimizer of

$$\psi_{HD}(\alpha) := \frac{\|p_{\alpha}^{\delta}\|}{\sqrt{\alpha}}.$$

An improvement of the HD-rule is the Hanke-Raus (HR) rule, which is defined by

$$\psi_{HR}(\alpha) := \left(\frac{1}{\alpha} \langle p_{\alpha}^{\delta}, p_{\alpha}^{\delta II} \rangle \right)^{\frac{1}{2}},$$

where $p_{\alpha}^{\delta II}$ is the second Tikhonov iterate; for details, see, e.g., [20].

For Tikhonov regularization, the ψ_{SL} functional can be written in terms of the singular value decomposition as

$$\psi_{SL}(\alpha)^2 = \sum_i \frac{\alpha \lambda_i}{(\alpha + \lambda_i)^3} | \langle y_{\delta}, u_i \rangle |^2.$$

We observe that, for Tikhonov regularization, the ψ -functionals of all the four rules, HD, HR, QO, and the simple-L rule, can be written in a common form as (1.11)

$$\psi(\alpha)^2 = \sum_i \frac{\alpha^{n-1} \lambda_i^k}{(\alpha + \lambda_i)^{n+k}} | \langle y_{\delta}, u_i \rangle |^2, \quad \text{with } n, k \in \mathbb{N} : \begin{array}{c|cc} & n=2 & n=3 \\ \hline k=0 & \text{HD} & \text{HR} \\ k=1 & \text{SL} & \text{QO} \end{array}.$$

This indicates a strong structural similarity of the four rules. Note that an analogous similarity of the δ -based variants of the QO- and HR-rules has led Raus to define the so-called R1-family of rules [31].

We may also stress the resemblance of the four rules above by expressing them in terms of the iteration operators (cf., e.g., [33, 34]) for Tikhonov regularization

$$B_{\alpha} := \alpha^{\frac{1}{2}} (\alpha I + AA^*)^{-\frac{1}{2}} \quad \text{and} \quad D_{\alpha} := A^* (\alpha I + AA^*)^{-\frac{1}{2}}.$$

Then we have the representations

$$\begin{aligned} \psi_{HD}(\alpha) &= \frac{\|p_{\alpha}^{\delta}\|}{\sqrt{\alpha}}, & \psi_{HR}(\alpha) &= \frac{\|B_{\alpha} p_{\alpha}^{\delta}\|}{\sqrt{\alpha}}, \\ \psi_{SL}(\alpha) &= \frac{\|D_{\alpha} p_{\alpha}^{\delta}\|}{\sqrt{\alpha}}, & \psi_{QO}(\alpha) &= \frac{\|D_{\alpha} B_{\alpha} p_{\alpha}^{\delta}\|}{\sqrt{\alpha}}. \end{aligned}$$

Besides the relation to the L-curve method, the above structural congruence with other known rules is a strong motivation for the definition and analysis of $\psi_{SL}(\alpha)$, which now fills a gap in the list of classical parameter choice rules.

As for the other rules (see, e.g., [20, 29]), one may also extend the definition of ψ_{SL} (and ψ_{SLR}) to more general regularization schemes: If R_α is defined by a filter function $g_\alpha(\lambda)$ of the form

$$R_\alpha y_\delta := g_\alpha(A^*A)A^*y_\delta, \quad \text{with} \quad r_\alpha(\lambda) := 1 - \lambda g_\alpha(\lambda),$$

then we may extend the definition of ψ_{SL} as

$$(1.12) \quad \psi_{SL}(\alpha) = \|\rho_\alpha(A^*A)^{\frac{1}{2}}y_\delta\|,$$

$$(1.13) \quad \rho_\alpha(\lambda) = \lambda g_\alpha(\lambda)^2 |r_\alpha(\lambda)|.$$

This definition agrees with the one for Tikhonov regularization, where $g_\alpha(\lambda) = \frac{1}{\alpha + \lambda}$ and

$$(1.14) \quad \rho_\alpha(\lambda) = \frac{\alpha\lambda}{(\alpha + \lambda)^3}.$$

REMARK 1.3. From the preceding comparison with other rules, we may conclude (by looking column-wise at the table in (1.11)) that the simple-L rule is related to the QO-rule in a similar way to how the HD-rule is related to the HR-rule. We will observe that both the HD-rule and the simple-L rule suffer from an *early saturation effect*. Thus, the QO-rule is the “cure” of the simple-L rule from early saturation just as the HR-rule is for the HD-rule. Note, however, that for heuristic rules an early saturation is sometimes beneficial as the rules may perform optimal when the smoothness class is at the saturation index without requiring additional conditions on the exact solution (cf. [20] and Theorem 2.10 below with $\mu = \frac{1}{2}$). On the other hand, the row-wise similarity in the table in (1.11) can be illustrated by the results below, where both the HD- and HR-rules require the same *Muckenhoupt-type condition* MC_1 , while for the simple-L and the QO-rule, we need the condition MC_2 .

Hence, the mentioned similarity of rules indicated in the table in (1.11) is also strikingly reflected in the theoretical results of the convergence theory in this paper:

	early saturation	late saturation
noise condition MC_1	HD	HR
noise condition MC_2	SL	QO

REMARK 1.4. There is a strong relation between heuristic rules and classical δ -based parameter choice rules. If for the above ψ -functional one considers the δ -based rule of the form: find α such that

$$\sqrt{\alpha}\psi(\alpha) \sim \delta,$$

then (up to technical details), this gives the discrepancy principle for $\psi = \psi_{HD}$, the modified discrepancy principle for $\psi = \psi_{HR}$, and the R1-Rule (or balancing principle) for $\psi = \psi_{QO}$. It is hence natural to define for the simple-L rule an accompanying δ -based rule, where α is chosen such that

$$\sqrt{\alpha} \left(-\langle x_\alpha^\delta, \alpha \frac{\partial}{\partial \alpha} x_\alpha^\delta \rangle \right)^{\frac{1}{2}} = \tau \delta,$$

with $\tau > 1$ fixed. To the knowledge of the authors, such a rule has not yet been investigated. From the previous remark, however, it can be concluded that it will suffer from an early saturation like the discrepancy principle.

REMARK 1.5. Let us also mention that the simple-L and simple-L ratio methods have some similarities with the V-curve method [9], which is defined as minimizer of the speed of the parameterization of the L-curve on a logarithmic grid. Thus, the minimization functional for the V-curve is for Tikhonov regularization (using the identity $\rho' = -\alpha\eta'$; cf. [18])

$$\begin{aligned}
 \psi_V(\alpha) &= \left\| \begin{bmatrix} \alpha\kappa'(\alpha) \\ \alpha\chi'(\alpha) \end{bmatrix} \right\| = \left\| \begin{bmatrix} \alpha\frac{\rho'}{\rho} \\ \alpha\frac{\eta'}{\eta} \end{bmatrix} \right\| \\
 &= \alpha|\eta'| \sqrt{\frac{\alpha^2}{\rho^2} + \frac{1}{\eta^2}} = \psi_{SL}(\alpha)^2 \sqrt{\frac{\alpha^2}{\rho^2} + \frac{1}{\eta^2}} \\
 &= \psi_{SLR}(\alpha)^2 \sqrt{\frac{\alpha^2|\eta|^2}{\rho^2} + 1} = \psi_{SLR}(\alpha)^2 \sqrt{\frac{1}{\zeta^2} + 1}.
 \end{aligned}$$

Thus, the V-curve is essentially a weighted form (with weight $\sqrt{\frac{1}{\zeta^2} + 1} \geq 1$) of our simple L-ratio functional ψ_{SLR} . It is obvious that the simple-L functional equals the derivative of the parameterization of the y -axis of the non-logarithmic L-curve $(\rho(\alpha), \eta(\alpha))$ weighted with α , which also equals the derivative of the x -axis parameterization as $\rho'(\alpha) = -\alpha\eta'$.

Another related method is the so-called composite residual and smoothing operator method (CRESO-method) [7]. It defines the regularization parameter by an argmax of the function

$$C(\alpha) := \|x_\alpha^\delta\|^2 + 2\alpha \frac{\partial}{\partial \alpha} \|x_\alpha^\delta\|^2 = \eta + 2\alpha\eta'.$$

Since maximizing $C(\alpha)$ is the same as minimizing $-C(\alpha)$, we observe that the method minimizes the functional

$$-C(\alpha) = \eta(\alpha)(2\psi_{SLR}(\alpha, y_\delta)^2 - 1).$$

Since $\eta(\alpha)$ is bounded from below (and approaches $\|x^\dagger\|^2$ for the optimal choice of α), we may regard the CRESO method essentially as a variant of the simple-L ratio method.

It is worth mentioning that the expression denoted by ζ in the curvature in Lemma 1.1 also has a relation to existing parameter choice functionals. In fact, in the simplest case, the Brezinski-Rodriguez-Seatzu rule [4, 5] is defined as the minimizer of $\frac{\|Ax_\alpha^\delta - y_\delta\|^2}{\alpha\|x_\alpha^\delta\|}$, which in our notation equals $\|x_\alpha^\delta\|\zeta$.

2. Convergence theory for Tikhonov regularization. The convergence theory for error-estimating heuristic methods is based on the idea that a “surrogate” functional $\psi(\alpha)$ behaves in a similar way to the total error $\|x_\alpha^\delta - x^\dagger\|$. Hence, minimizing $\psi(\alpha)$ could be expected to give a small total error and thus a successful parameter choice. For verifying this, we have to estimate the functionals against the approximation and stability errors, which can be expressed in terms of the SVD as follows:

$$\begin{aligned}
 \|x_\alpha^\delta - x_\alpha\|^2 &= \sum_i \frac{\lambda_i}{(\lambda_i + \alpha)^2} |\langle y_\delta - y, u_i \rangle|^2, \\
 \|x_\alpha - x^\dagger\|^2 &= \sum_i \frac{\alpha^2}{(\lambda_i + \alpha)^2} |\langle x^\dagger, v_i \rangle|^2.
 \end{aligned}$$

As usual, the total error $\|x_\alpha^\delta - x^\dagger\|$ can be bounded by the stability error and the approximation error as in (1.3).

Accordingly, we may split the functional ψ_{SL} in a similar way, into a noise-dependent term and an x^\dagger -dependent one:

$$\begin{aligned}\psi_{SL}(\alpha, e)^2 &:= \|\rho_\alpha(AA^*)^{\frac{1}{2}}(y_\delta - y)\|^2 = \sum_i \frac{\alpha\lambda_i}{(\lambda_i + \alpha)^3} |\langle y_\delta - y, u_i \rangle|^2, \\ \psi_{SL}(\alpha, x^\dagger)^2 &:= \|\rho_\alpha(AA^*)^{\frac{1}{2}}y\|^2 = \|\rho_\alpha(AA^*)^{\frac{1}{2}}Ax^\dagger\|^2 = \sum_i \frac{\alpha\lambda_i^2}{(\lambda_i + \alpha)^3} |\langle x^\dagger, v_i \rangle|^2.\end{aligned}$$

Obviously, we have the bound

$$(2.1) \quad \psi_{SL}(\alpha) \leq \psi_{SL}(\alpha, e) + \psi_{SL}(\alpha, x^\dagger).$$

Convergence is based on the following theorem which is proven in [20]:

THEOREM 2.1. *Let a ψ -functional be given as in (1.12) by some nonnegative continuous function $\rho_\alpha(\lambda)$ defined on the spectrum of A^*A . Let α^* be selected as*

$$\alpha^* = \operatorname{argmin}_\alpha \psi(\alpha) = \operatorname{argmin}_\alpha \|\rho_\alpha(A^*A)^{\frac{1}{2}}y_\delta\|.$$

Assume that

$$\|\rho_\alpha(A^*A)^{\frac{1}{2}}Ax^\dagger\| \leq B(\alpha), \quad \|\rho_\alpha(A^*A)^{\frac{1}{2}}(y_\delta - y)\| \leq V(\alpha),$$

where $B(\alpha)$ is monotonically increasing and $V(\alpha)$ is monotonically decreasing. Furthermore, assume that the following lower bounds involving the stability and approximation errors hold:

$$(2.2) \quad \|x_\alpha^\delta - x_\alpha\| \leq C_0 \|\rho_\alpha(A^*A)^{\frac{1}{2}}(y_\delta - y)\|,$$

$$(2.3) \quad \|x_\alpha - x^\dagger\| \leq \Phi \left(\|\rho_\alpha(A^*A)^{\frac{1}{2}}Ax^\dagger\| \right),$$

with some increasing function Φ . Then the total error can be bounded by

$$\|x_{\alpha^*}^\delta - x^\dagger\| \leq \Phi \left(2 \inf_\alpha \{B(\alpha) + V(\alpha)\} \right) + 2C_0 \inf_\alpha \{B(\alpha) + V(\alpha)\}.$$

If a heuristic parameter choice functional ψ is (under certain circumstances) a good estimator for both the approximation error and the stability error, i.e., both the lower bounds hold and the upper bounds B, V are close to the approximation and stability error, then the corresponding parameter choice is usually a successful one in the sense that it yields the optimal order of convergence.

2.1. Upper bounds for ψ_{SL} . At first we provide an upper bound for $\psi_{SL}(\alpha, y_\delta - y)$. Since $\rho_\alpha(\lambda) \leq \frac{\lambda}{(\lambda + \alpha)^2}$, the next result follows immediately:

LEMMA 2.2. *We have that*

$$(2.4) \quad \psi_{SL}(\alpha, e) \leq V(\alpha) := \|x_\alpha^\delta - x_\alpha\| \leq \frac{\delta}{\sqrt{\alpha}}.$$

The term $\psi_{SL}(\alpha, x^\dagger)$ can be bounded in the following way:

LEMMA 2.3. *We have*

$$(2.5) \quad \psi_{SL}(\alpha, x^\dagger) \leq B(\alpha) := \left(\sum_i \frac{\alpha}{(\lambda_i + \alpha)} |\langle v_i, x^\dagger \rangle|^2 \right)^{\frac{1}{2}} = \langle x^\dagger - x_\alpha, x^\dagger \rangle^{\frac{1}{2}},$$

and $B(\alpha)$ is monotonically increasing in α . Moreover, if a source condition (1.4) is satisfied, then

$$B(\alpha) \leq C\alpha^\mu \quad \text{for } \mu \leq \frac{1}{2}.$$

Proof. Noting the definition of $\rho_\alpha(\lambda)$ in (1.14) and that $\frac{\lambda}{(\lambda+\alpha)} \leq 1$, we have that $\rho_\alpha(\lambda)^2 \lambda \leq \frac{\alpha}{(\lambda+\alpha)}$, which verifies the result. The fact that the last expression is monotone and allows for the stated convergence rates is standard. \square

REMARK 2.4. From the previous lemmas we obtain that under a source condition and by (2.1)

$$\inf_{\alpha} \psi_{SL}(\alpha) \leq \inf_{\alpha} (B(\alpha) + V(\alpha)) \leq \inf_{\alpha} \left(C\alpha^\mu + \frac{\delta}{\sqrt{\alpha}} \right) \sim \delta^{\frac{2\mu}{2\mu+1}}, \quad \text{for } \mu \leq \frac{1}{2}.$$

This is the optimal-order rate of the error, but it is only achieved under the restriction that $\mu \leq \frac{1}{2}$. Thus, ψ_{SL} shows *early saturation*, that is, it is only of the same order as the optimal rate for a lower smoothness index, but it shows suboptimal rates for $\mu \geq \frac{1}{2}$. This is akin to the early saturation of the discrepancy principle [8] and the HD-method [20].

2.2. Lower bounds for ψ_{SL} . The main issue in the convergence theory is to find conditions which are sufficient to verify the lower bounds in Theorem 2.1. However, it is well known that due to the so-called Bakushinskii veto [1, 20], a heuristic parameter choice functional *cannot* be a valid estimator for the stability error—in the sense that (2.2) holds—unless the permissible noise $y_\delta - y$ is restricted in some way. Conditions imposing such noise restrictions are at the heart of the convergence theory.

We recall the following classical noise restrictions that were used in [20, 23] denoted as Muckenhoupt-type conditions (MC):

DEFINITION 2.5. *The condition MC₁ is satisfied if there exists a constant C_1 such that for all occurrent errors $e = y_\delta - y$ and for all $0 < \alpha \leq \alpha_{\max}$, it holds that*

$$\sum_{\lambda_i \geq \alpha} \frac{\alpha}{\lambda_i} |\langle e, u_i \rangle|^2 \leq C_1 \sum_{\lambda_i \leq \alpha} |\langle e, u_i \rangle|^2.$$

The condition MC₂ is satisfied if there exists a constant C_2 such that for all occurrent errors $e = y_\delta - y$ and for all $0 < \alpha \leq \alpha_{\max}$, it holds that

$$(2.6) \quad \sum_{\lambda_i \geq \alpha} \frac{\alpha}{\lambda_i} |\langle e, u_i \rangle|^2 \leq C_2 \sum_{\lambda_i \leq \alpha} \frac{\lambda_i}{\alpha} |\langle e, u_i \rangle|^2.$$

It is obvious that MC₂ is slightly stronger than MC₁: MC₂ \implies MC₁. Simply put, these conditions are *irregularity* conditions for the noise in the sense that e should not be smooth (i.e., in the range of A). Meanwhile, they are quite well understood and are satisfied in many cases. Moreover, it has been shown that for mildly ill-posed problems they hold for white and colored noise with probability one [24]. Although MC₂ is slightly stronger, they are often both satisfied.

Here, we show that the error-dependent part of ψ_{SL} is an upper bound for the error propagation term. As mentioned before, for this we require a Muckenhoupt condition:

PROPOSITION 2.6. *Let $y_\delta - y$ satisfy a Muckenhoupt-type condition MC₂ with constant C_2 . Then with $\rho_\alpha(\lambda)$ corresponding to the ψ_{SL} -functional, we have for all $0 < \alpha < \alpha_{\max}$*

$$\|x_\alpha^\delta - x_\alpha\| \leq \sqrt{C_2 + 1} \|\rho_\alpha(AA^*)^{\frac{1}{2}}(y_\delta - y)\|.$$

Proof. As in [20, 23], the idea of the proof is to split the spectral decomposition into terms involving $\lambda \leq \alpha$ and $\lambda > \alpha$: This works because of the estimates

$$(2.7) \quad \frac{1}{2} \begin{cases} \frac{1}{\alpha} & \lambda \leq \alpha \\ \frac{1}{\lambda} & \lambda \geq \alpha \end{cases} \leq \frac{1}{\alpha + \lambda} \leq \begin{cases} \frac{1}{\alpha} & \lambda \leq \alpha, \\ \frac{1}{\lambda} & \lambda \geq \alpha. \end{cases}$$

Thus, using (2.7) and (2.6)

$$\begin{aligned} \|x_\alpha^\delta - x_\alpha\|^2 &= \sum_i \frac{\lambda_i}{(\lambda_i + \alpha)^2} |\langle e, u_i \rangle|^2 \\ &\leq \sum_{\lambda_i \leq \alpha} \frac{\lambda_i}{\alpha^2} |\langle e, u_i \rangle|^2 + \sum_{\lambda_i \geq \alpha} \frac{1}{\lambda_i} |\langle e, u_i \rangle|^2 \leq (1 + C_2) \sum_{\lambda_i \leq \alpha} \frac{\lambda_i}{\alpha^2} |\langle e, u_i \rangle|^2. \end{aligned}$$

Conversely, the ψ -expression can be estimated as

$$\begin{aligned} \|\rho_\alpha(AA^*)^{\frac{1}{2}}(y_\delta - y)\|^2 &= \sum_i \frac{\lambda_i \alpha}{(\lambda_i + \alpha)^3} |\langle e, u_i \rangle|^2 \\ &= \sum_{\lambda_i \leq \alpha} \frac{\lambda_i \alpha}{(\lambda_i + \alpha)^3} |\langle e, u_i \rangle|^2 + \sum_{\lambda_i \geq \alpha} \frac{\lambda_i \alpha}{\alpha^3} |\langle e, u_i \rangle|^2 \geq \sum_{\lambda_i \geq \alpha} \frac{\lambda_i}{\alpha^2} |\langle e, u_i \rangle|^2, \end{aligned}$$

which yields the statement. \square

REMARK 2.7. Note that the stability part of the simple-L curve method behaves similar to the QO-method, for which also the condition MC₂ has been postulated to obtain the analogous estimate. This is different to the HD- and HR-methods, where the condition MC₁ is sufficient [20].

The next step involves the approximation error:

PROPOSITION 2.8. *Suppose that $x^\dagger \neq 0$ satisfies a source condition (1.4) with $\mu \leq 1$. Then for $\alpha \in (0, \alpha_{\max})$ and $\rho_\alpha(\lambda)$ corresponding to the ψ_{SL} -functional, there is a constant C such that*

$$\|x_\alpha - x^\dagger\| \leq \frac{C}{\|A^*Ax^\dagger\|^{2\mu}} \|\rho_\alpha(AA^*)^{\frac{1}{2}}y\|^{2\mu}.$$

Proof. As $(\alpha + \lambda_i) \leq \alpha_{\max} + \|A\|^2 =: C_3$, we have that

$$\begin{aligned} \|\rho_\alpha(AA^*)^{\frac{1}{2}}y\|^2 &= \sum_i \frac{\alpha \lambda_i}{(\lambda_i + \alpha)^3} \lambda_i |\langle x^\dagger, v_i \rangle|^2 \\ &\geq \frac{\alpha}{C_3^3} \sum_i \lambda_i^2 |\langle x^\dagger, v_i \rangle|^2 = \frac{\alpha \|A^*Ax^\dagger\|^2}{C_3^3}. \end{aligned}$$

Conversely, from the classical convergence rate estimate (1.5), we obtain with a generic constant C that

$$\begin{aligned} \|x_\alpha - x^\dagger\| &\leq C\alpha^\mu \leq C \left(\frac{C_3^3}{\|A^*Ax^\dagger\|^2} \|\rho_\alpha(AA^*)^{\frac{1}{2}}y\|^2 \right)^\mu \\ &\leq \frac{C}{\|A^*Ax^\dagger\|^{2\mu}} \|\rho_\alpha(AA^*)^{\frac{1}{2}}y\|^{2\mu}. \end{aligned}$$

Moreover, we note that x^\dagger is a minimum-norm solution and thus in $N(A)^\perp$. Hence, if $x^\dagger \neq 0$, then $A^*Ax^\dagger \neq 0$. \square

If we impose a certain regularity assumption on x^\dagger , then it can be shown that the approximation part of ψ_{SL} , $\|\rho_\alpha(AA^*)^{\frac{1}{2}}y\|$, is an upper bound for the approximation error. The regularity assumption [20, 23] is similar to the Muckenhoupt-type condition but with the spectral parts interchanged:

$$(2.8) \quad \sum_{\lambda_i \leq \alpha} |\langle x^\dagger, v_i \rangle|^2 \leq D \sum_{\lambda_i \geq \alpha} \frac{\alpha}{\lambda_i} |\langle x^\dagger, v_i \rangle|^2.$$

For a comparison with other situations, we also state a different regularity condition that is also used in [20]:

$$\sum_{\lambda_i \leq \alpha} |\langle x^\dagger, v_i \rangle|^2 \leq D \sum_{\lambda_i \geq \alpha} \left(\frac{\alpha}{\lambda_i}\right)^2 |\langle x^\dagger, v_i \rangle|^2.$$

Obviously, the first of these conditions, (2.8), is weaker, and the second one implies the first. For the simple L-curve method, the weaker one suffices:

PROPOSITION 2.9. *Let x^\dagger satisfy the regularity condition (2.8). Then for $\alpha \in (0, \alpha_{\max})$ and $\rho_\alpha(\lambda)$ corresponding to the ψ_{SL} -functional, there is a constant C such that*

$$\|x_\alpha - x^\dagger\| \leq C \|\rho_\alpha(AA^*)^{\frac{1}{2}}y\|.$$

Proof. Using the splitting of the sums and (2.7), we have

$$\begin{aligned} \|x_\alpha - x^\dagger\|^2 &= \sum_i \frac{\alpha^2}{(\lambda_i + \alpha)^2} |\langle x^\dagger, v_i \rangle|^2 \\ &= \sum_{\lambda_i \leq \alpha} \frac{\alpha^2}{(\lambda_i + \alpha)^2} |\langle x^\dagger, v_i \rangle|^2 + \sum_{\lambda_i \geq \alpha} \frac{\alpha^2}{(\lambda_i + \alpha)^2} |\langle x^\dagger, v_i \rangle|^2 \\ &\leq \sum_{\lambda_i \leq \alpha} |\langle x^\dagger, v_i \rangle|^2 + \sum_{\lambda_i \geq \alpha} \frac{\alpha^2}{\lambda_i^2} |\langle x^\dagger, v_i \rangle|^2 \\ &\leq \sum_{\lambda_i \leq \alpha} |\langle x^\dagger, v_i \rangle|^2 + \sum_{\lambda_i \geq \alpha} \frac{\alpha}{\lambda_i} |\langle x^\dagger, v_i \rangle|^2. \end{aligned}$$

While for the approximation part of ψ_{SL} , using (2.7) again, we obtain

$$\begin{aligned} \|\rho_\alpha(AA^*)^{\frac{1}{2}}y\|^2 &= \sum_i \frac{\alpha \lambda_i}{(\lambda_i + \alpha)^3} \lambda_i |\langle x^\dagger, v_i \rangle|^2 \\ &\geq \sum_{\lambda_i \leq \alpha} \frac{\alpha \lambda_i^2}{(\lambda_i + \alpha)^3} |\langle x^\dagger, v_i \rangle|^2 + \sum_{\lambda_i \geq \alpha} \frac{\alpha \lambda_i^2}{(\lambda_i + \alpha)^3} |\langle x^\dagger, v_i \rangle|^2 \\ &\geq \frac{1}{2} \sum_{\lambda_i \leq \alpha} \frac{\lambda_i^2}{\alpha^2} |\langle x^\dagger, v_i \rangle|^2 + \frac{1}{2} \sum_{\lambda_i \geq \alpha} \frac{\alpha}{\lambda_i} |\langle x^\dagger, v_i \rangle|^2 \\ &\geq \frac{1}{2} \sum_{\lambda_i \geq \alpha} \frac{\alpha}{\lambda_i} |\langle x^\dagger, v_i \rangle|^2. \end{aligned}$$

Thus, the regularity condition (2.8) ensures the bound. \square

Together with Theorem 2.1 and the previous estimates, we arrive at the main theorem:

THEOREM 2.10. *Let the error satisfy a Muckenhoupt-type condition MC_2 , let x^\dagger satisfy a source condition (1.4) with $\mu \leq 1$, and let $\|x^\dagger\| \neq 0$.*

Then choosing the regularization parameter α^ as the minimizer of ψ_{SL} as in (1.10) yields the following error bounds*

$$\|x_{\alpha^*}^\delta - x^\dagger\| \leq C\delta^{\frac{2\tilde{\mu}}{2\tilde{\mu}+1}2\tilde{\mu}}, \quad \tilde{\mu} = \min\{\mu, \frac{1}{2}\}.$$

If, moreover, x^\dagger additionally satisfies a regularity condition (2.8), then the optimal-order (for $\mu \leq \frac{1}{2}$) estimate

$$\|x_{\alpha^*}^\delta - x^\dagger\| \leq C\delta^{\frac{2\tilde{\mu}}{2\tilde{\mu}+1}}, \quad \tilde{\mu} = \min\{\mu, \frac{1}{2}\},$$

holds.

REMARK 2.11. The convergence theorem for the simple-L method should be compared to the corresponding results for the HD, HR, and QO-rules in [20]: Essentially, the functional ψ_{SL} requires the same conditions as the QO-rule, but it only achieves the optimal order (in the best case when a regularity condition holds) up to $\mu \leq \frac{1}{2}$, while the QO-rule does this (under the same regularity condition) for all μ up to the saturation index $\mu = 1$. In this sense, the QO-rule is an improvement of the simple-L method. This is similar to the relations between HD and HR: the heuristic discrepancy method, ψ_{HD} , can also be only optimal up to $\mu \leq \frac{1}{2}$, while the Hanke-Raus method improves this up to $\mu = 1$. Thus, ψ_{SL} is related to ψ_{QO} in a similar way to how ψ_{HD} is related to ψ_{HR} .

2.3. Convergence for ψ_{SLR} . The previous analysis can be extended to the simple-L ratio method. We now consider a functional of the form

$$(2.9) \quad \psi(\alpha, y_\delta) = \phi(\alpha)\psi_{SL}(\alpha, y_\delta),$$

where ϕ is a nonnegative function. The simple-L ratio corresponds to $\phi(\alpha) = \frac{1}{\|x_\alpha^\delta\|}$. We have the following proposition (here, Id denotes the identity function $x \rightarrow x$):

PROPOSITION 2.12. *Let the error satisfy a Muckenhoupt-type condition MC_2 , and let (2.3) hold for ρ_α corresponding to ψ_{SL} . Suppose that α^* is selected by (2.9). Then the following error estimates hold: For $\bar{\alpha} \in (0, \alpha_{\max})$ arbitrary*

$$\|x_{\alpha^*}^\delta - x^\dagger\| \leq \frac{\phi(\bar{\alpha})}{\phi(\alpha^*)} (B(\bar{\alpha}) + V(\bar{\alpha})) + 2 \max\{\Phi(B(\bar{\alpha})), C_0\text{Id}(B(\bar{\alpha}))\} \quad \text{if } \alpha^* \leq \bar{\alpha},$$

$$\|x_{\alpha^*}^\delta - x^\dagger\| \leq C_0V(\bar{\alpha}) + \Phi \left[V(\bar{\alpha}) + \frac{\phi(\bar{\alpha})}{\phi(\alpha^*)} (V(\bar{\alpha}) + B(\bar{\alpha})) \right] \quad \text{if } \alpha^* \geq \bar{\alpha}.$$

Here, V and B are defined as in (2.4) and (2.5).

Proof. Let $\alpha^* \leq \bar{\alpha}$. Then from the previous estimates for ψ_{SL} , the minimization property of $\phi(\alpha)\psi_{SL}(\alpha, y_\delta)$, and by the monotonicity of B , we have

$$\begin{aligned} \|x_{\alpha^*}^\delta - x^\dagger\| &\leq \Phi(\psi_{SL}(\alpha^*, x^\dagger)) \leq \Phi(B(\alpha^*)) \leq \Phi(B(\bar{\alpha})), \\ \phi(\alpha^*)\|x_{\alpha^*}^\delta - x_{\alpha^*}\| &\leq C_0\phi(\alpha^*)\psi_{SL}(\alpha^*, e) \\ &\leq C_0\phi(\alpha^*)\psi_{SL}(\alpha^*, y_\delta) + C_0\phi(\alpha^*)\psi_{SL}(\alpha^*, x^\dagger) \\ &\leq C_0\phi(\bar{\alpha})\psi(\bar{\alpha}, y_\delta) + C_0\phi(\alpha^*)B(\bar{\alpha}) \\ &\leq C_0\phi(\bar{\alpha})B(\bar{\alpha}) + C_0\phi(\bar{\alpha})V(\bar{\alpha}) + C_0\phi(\alpha^*)B(\bar{\alpha}). \end{aligned}$$

For $\alpha^* \geq \bar{\alpha}$, with the same arguments and by the monotonicity of V , we obtain

$$\begin{aligned} \|x_{\alpha^*}^\delta - x_{\alpha^*}\| &\leq C_0 \psi_{SL}(\alpha^*, e) \leq C_0 V(\alpha^*) \leq C_1 V(\bar{\alpha}), \\ \Phi^{-1}(\|x_\alpha - x^\dagger\|) &\leq \phi(\alpha^*) \frac{\psi_{SL}(\alpha^*, y_\delta)}{\phi(\alpha^*)} + \psi_{SL}(\alpha^*, e) \leq \phi(\bar{\alpha}) \frac{\psi_{SL}(\bar{\alpha}, y_\delta)}{\phi(\alpha^*)} + V(\alpha^*) \\ &\leq \frac{\phi(\bar{\alpha})}{\phi(\alpha^*)} B(\bar{\alpha}) + \left(\frac{\phi(\bar{\alpha})}{\phi(\alpha^*)} + 1\right) V(\bar{\alpha}). \quad \square \end{aligned}$$

THEOREM 2.13. *Under the same conditions as Theorem 2.10 and if α^* is chosen by the simple-L ratio-method, then the error bounds of Theorem 2.10 hold if δ is sufficiently small.*

Proof. Since we assume a source condition (1.4), we have from Proposition 2.8 that $\Phi(x) = x^\xi$, where $\xi \leq 1$. The error estimates can be rewritten as

$$\begin{aligned} \phi(\alpha^*) \|x_{\alpha^*}^\delta - x^\dagger\| &\leq \phi(\bar{\alpha}) (B(\bar{\alpha}) + V(\bar{\alpha})) + \phi(\alpha^*) 2 \max\{\Phi, C_1 \text{Id}\}(B(\bar{\alpha})) \quad \text{if } \alpha^* \leq \bar{\alpha}, \\ \phi(\alpha^*) \|x_{\alpha^*}^\delta - x^\dagger\| &\leq C_1 \phi(\alpha^*) V(\bar{\alpha}) \\ &\quad + \Phi \left[\phi(\alpha^*)^{\frac{1}{\xi}} V(\bar{\alpha}) + \phi(\alpha^*)^{\frac{1}{\xi}-1} \phi(\bar{\alpha}) (V(\bar{\alpha}) + B(\bar{\alpha})) \right] \quad \text{if } \alpha^* \geq \bar{\alpha}. \end{aligned}$$

For the simple-L ratio method, we have $\phi(\alpha) = \frac{1}{\|x_\alpha^\delta\|}$. We take $\bar{\alpha}$ as the optimal order choice $\bar{\alpha} \sim \delta^{\frac{2}{2\bar{\mu}+1}}$, which implies $x_{\bar{\alpha}}^\delta \rightarrow x^\dagger$, and hence for δ sufficiently small, we obtain that $\phi(\bar{\alpha}) \sim \frac{1}{\|x^\dagger\|}$. From the standard theory it follows that $\alpha \rightarrow \|x_\alpha^\delta\|$ is monotonically decreasing, hence $\phi(\alpha)$ is monotonically increasing. Thus, with some constant C

$$\phi(\alpha^*) \leq \phi(\alpha_{\max}) \leq C.$$

In any case, the expressions $\phi(\alpha^*)^{\frac{1}{\xi}}$, $\phi(\bar{\alpha})$, and, since $\xi \leq 1$, also $\phi(\alpha^*)^{\frac{1}{\xi}-1}$ remain bounded. It follows that

$$\phi(\alpha^*) \|x_{\alpha^*}^\delta - x^\dagger\| \leq C' \max\{\Phi, \text{Id}\} \left(C \delta^{\frac{2\bar{\mu}}{2\bar{\mu}+1}} \right),$$

with different constants C, C' . Moreover, since

$$\phi(\alpha^*) = \frac{1}{\|x_{\alpha^*}^\delta\|} \geq \frac{1}{\|x_{\alpha^*}^\delta - x^\dagger\| + \|x^\dagger\|},$$

we have that

$$\frac{\|x_{\alpha^*}^\delta - x^\dagger\|}{\|x_{\alpha^*}^\delta - x^\dagger\| + \|x^\dagger\|} \leq C' \max\{\Phi, \text{Id}\} \left(C \delta^{\frac{2\bar{\mu}}{2\bar{\mu}+1}} \right).$$

Since $\frac{x}{x+\|x^\dagger\|} \sim x$ for x small, this yields estimates of the same order as before. \square

The reason for requiring a small δ is that the expression $\frac{\|x_{\alpha^*}^\delta - x^\dagger\|}{\|x_{\alpha^*}^\delta - x^\dagger\| + \|x^\dagger\|}$ is bounded by 1. Hence if the right-hand side (which is of the order of the optimal convergence) is large, then the estimate holds trivially true but the content is negligible.

2.4. Extension to other regularization methods. We note that the simplification of the curvature of the L-curve relies heavily on Tikhonov regularization, which is the only regularization method for which formula (1.7) holds true. For general regularization schemes, the expression for the curvature becomes rather complicated.

With the same definition of the L-curve, the curvature can be calculated as

$$\gamma = \frac{\rho\eta}{(\rho'^2\eta^2 + \rho^2\eta'^2)^{\frac{3}{2}}} (\eta''\eta\rho'\rho - \rho''\rho\eta'\eta - \eta'^2\rho'\rho + \rho'^2\eta'\eta).$$

For Tikhonov regularization, this can be simplified by the formula $\rho' = -\alpha\eta'$, but for other regularization methods, this is no longer possible. Similar as above, however, we introduce the variable $\zeta = \frac{\rho\eta'}{\rho'\eta}$, which for Tikhonov regularization agrees with the definition given in Lemma 1.1. Then we obtain that

$$(2.10) \quad \begin{aligned} \gamma &= \frac{1}{|\rho'\eta|^3 (1 + \zeta^2)^{\frac{3}{2}}} (\eta''\eta^2\rho'\rho^2 - \rho''\rho^2\eta'\eta^2 - \zeta^2(\rho'\eta)^3 + \zeta(\rho'\eta)^3) \\ &= \left[\frac{\eta\eta''}{\eta'^2} - \frac{\rho''\eta}{\eta'\rho'} \right] \frac{\zeta^2}{(1 + \zeta^2)^{\frac{3}{2}}} + \frac{-\zeta^2 + \zeta}{(1 + \zeta^2)^{\frac{3}{2}}}. \end{aligned}$$

(For Tikhonov regularization, the identity $\rho' = -\alpha\eta'$ yields that $\rho'' = -\eta' + \frac{\rho'}{\eta}\eta''$ and consequently formula (1.8).)

Thus, a fully analogous functional corresponding to ψ_{SLR} would be to minimize the reciprocal of the expression in brackets in (2.10). However, due to the appearance of several second-order derivative terms, such a method would probably not be qualified then to be named “simple”.

We try to simplify the expression for asymptotic regularization (cf. [8]), which is a continuous version of classical Landweber iteration. The method is defined via an initial value problem in Hilbert spaces,

$$x(t)' = A^*p(t), \quad x(0) = 0, \quad t \geq 0,$$

where $p(t) = y_\delta - Ax(t)$. The regularized solution is given by

$$x_\alpha^\delta = x\left(\frac{1}{\alpha}\right).$$

When the derivative $x'(t)$ is replaced by a forward difference, this yields exactly Landweber iteration.

For this method, we have the identities

$$p' = -Ax', \quad x' = A^*p \quad \Rightarrow \quad p_\alpha^{\delta'} = -AA^*p.$$

Thus,

$$\begin{aligned} \eta' &= 2 \langle x, x' \rangle = 2 \langle Ax, p \rangle, \\ \rho' &= 2 \langle p, p' \rangle = -2 \langle p, Ax' \rangle = -2 \langle x', x' \rangle = -2 \|A^*p\|^2, \\ \eta'' &= 2 \langle Ax', p \rangle + 2 \langle Ax, p' \rangle = -\rho' - 2 \langle Ax, Ax' \rangle, \\ \rho'' &= -4 \langle A^*p, A^*p' \rangle = -4 \langle A^*p, A^*p' \rangle = 4 \langle A^*p, A^*AA^*p \rangle. \end{aligned}$$

As the curvature is independent of the parameterization, we may use the variable t in place of α to calculate it. Hence, the expression in brackets in (2.10) can then be written as

$$\frac{\eta}{\eta'} \left(\frac{\eta''}{\eta'} - \frac{\rho''}{\rho'} \right) = \frac{\|x\|^2}{2 \langle Ax, p \rangle} \left(\frac{\|A^*p\|^2}{\langle Ax, p \rangle} - \frac{\langle Ax, p' \rangle}{\langle Ax, p \rangle} - \frac{2 \langle A^*p, A^*AA^*p \rangle}{\|A^*p\|^2} \right).$$

The last expression $\frac{2 \langle A^*p, A^*AA^*p \rangle}{\|A^*p\|^2}$ is bounded by $\|A^*A\|$. Thus, the only way that the L-curve can have a large curvature is when $\langle Ax, p \rangle = \eta'$ is small. This essentially leads again to the

simple L-curve method with the minor difference that the derivative is taken with respect to the t -variable, i.e., the parameter α^* would be selected via $\alpha^* := \frac{1}{t_*}$ with t_* being the argmin over t of

$$\psi(t) := \langle Ax(t), p(t) \rangle = \langle Ax(t), y_\delta - Ax(t) \rangle.$$

This expression is quite similar (but not completely identical) to the generalization of the simple L-curve suggested in (1.13) for general regularization methods.

By analogy, we may transfer these results to Landweber iteration, where derivatives are replaced by finite differences. The simple L-curve method would then be defined by minimizing

$$(2.11) \quad \psi(k) = \langle Ax_k, y_\delta - Ax_k \rangle \sim \langle x_k, x_{k+1} - x_k \rangle,$$

over the iteration indices k . Clearly, this can be considered a discrete variant of ψ_{SL} , where the derivative $\alpha \frac{\partial}{\partial \alpha}$ is replaced by a finite difference. Another possibility for defining a simple L-curve method is to use (1.12)–(1.13) for general regularization methods via their filter functions. In case of Landweber iteration this leads to a similar functional as in (2.11), namely

$$\psi(k) = \langle x_k, x_{2k} - x_k \rangle.$$

Of further special interest is to use these methods for nonlinear (e.g., convex) Tikhonov regularization, where x_α^δ is defined as the minimizer of

$$(2.12) \quad x \rightarrow \|Ax - y_\delta\|^2 + \alpha R(x),$$

with a general convex regularization functional R . For an analysis of several heuristic rules in this context, see [25]. Note that the L-curve method is then defined by analogy as a plot of $(\log(R(x_\alpha^\delta)), \log(\|Ax_\alpha^\delta - y_\delta\|))$. It has been applied with success in such a context, e.g., in [38]. One should be cautioned, however, that here it is not necessarily true that x_α^δ is differentiable with respect to α , and moreover, $R(x_\alpha^\delta)$ can be 0, hence the L-graph in its logarithmic form may not be defined in that case. If R is smooth, then the formula (1.7) still holds with $\eta(\alpha) = R(\alpha)$, and we may define a simple-L method as the minimization of

$$\psi_{SL}(\alpha) = -\alpha \frac{\partial}{\partial \alpha} R(x_\alpha^\delta).$$

However, for convex Tikhonov regularization it is preferable—due to a possible lack of differentiability—to replace the derivative $\alpha \frac{\partial}{\partial \alpha}$ by an alternative expressions. One way is to use a finite difference approximation on a logarithmic grid yielding

$$(2.13) \quad \psi_{SL}(\alpha) = R(x_{\alpha_{n+1}, \delta}) - R(x_{\alpha_n, \delta}), \quad \alpha = \alpha_0 q^n \quad q < 1.$$

Another way is to replace the derivative by expressions obtained via Bregman iteration. In this case, the functional would be

$$\psi_{SL}(\alpha) = R(x_\alpha^{\delta II}) - R(x_\alpha^\delta),$$

where $x_\alpha^{\delta II}$ is the second Bregman iterate; cf. [25]. Both methods can also be understood as a kind of quasi-optimality method, where the “strict metric” $d(x, y) = |R(x) - R(y)|$ (cf. [10]) is used for measuring convergence (a similar method has been tested in [22]). Note that we may similarly adapt the simple-L ratio functional as

$$(2.14) \quad \psi_{SLR} = \frac{R(x_\alpha^{\delta II}) - R(x_\alpha^\delta)}{R(x_\alpha^\delta)},$$

with the notation as before.

3. Numerical tests. We perform some numerical tests of the proposed methods with the noise-levels 0.01%, 0.1%, 1%, 5%, 10%, 20%, 50%. Here, the first pair is classified as “small”, the second pair as “medium”, and the last triple is classified as “large”. For each noise-level, we performed 10 experiments. We tested the method ψ_{SL} (simple-L), ψ_{SLR} (simple-L ratio), the QO-method, and the original L-curve method defined by maximizing the curvature.

A general observation was that whenever the L-curve showed a clear corner, then the selected parameter by both ψ_{SL} and ψ_{SLR} was very close to that corner, which confirms the idea of those methods being simplifications of the L-curve method. Note, however, that closeness on the L-curve does not necessarily mean that the selected parameter is close as well since the parameterization around the corner becomes “slow”.

We compare the four methods, namely, the two new simple-L rules, the QO-rule, and the original L-curve according to their total error for the respective selected α and calculate the ratio of the obtain error to the best possible error:

$$(3.1) \quad J(\alpha^*) := \frac{d(x_{\alpha^*}^\delta, x^\dagger)}{\inf_{\alpha} d(x_{\alpha}^\delta, x^\dagger)},$$

where one would typically compute J with $d(x, y) := \|x - y\|$ for the case of linear regularization.

3.1. Linear Tikhonov regularization. We begin with classical Tikhonov regularization, in which case we compute the regularized solution as (1.2).

3.1.1. Diagonal operator. At first we consider a diagonal operator A with singular values having polynomial decay: $\sigma_i = i^{-s}$, $i = 1, \dots, n$, for some value s and consider an exact solution also with polynomial decay $\langle x^\dagger, v_i \rangle = (-1)^i i^{-\tau}$, where τ was adapted to have a certain source condition (1.4) with index μ satisfied. The size of the diagonal matrix $A \in \mathbb{R}^{n \times n}$ was chosen as $n = 500$. Furthermore, we added random noise (colored Gaussian noise) $\langle e, u_i \rangle = \delta i^{-0.6} \tilde{e}_i$, where \tilde{e}_i are standard normally distributed values.

Table 3.1 displays the median of the values of J over 10 experiments with different random noise realizations and for varying smoothness indices μ . The table provides some information about the performance of the rules. Based on additional numbers not presented here, we can state some conclusions:

- The simple-L and simple-L ratio outperform the other rules for small smoothness index $\mu = 0.25$ and small data noise. Except for very large δ , the simple-L ratio is slightly better than the simple-L curve. For very large δ , the simple-L method works but is inferior to QO while the simple-L ratio method fails then.
- For high smoothness index, the QO-rule outperforms the other rules and it is the method of choice then.
- The original L-curve method often fails for small δ . For larger δ it works often only acceptably. Only in situations when δ is quite large ($> 20\%$) we found several instances when it outperforms all other rules.

A similar experiment was performed for a higher smoothing operator by setting $s = 4$ with similar conclusions. We note that theory has indicated that for $\mu = 0.5$, the simple-L curve is order optimal without any additional condition on x^\dagger while for the QO-rule this happens at $\mu = 1$. One would thus expect that the simple-L rule perform better for $\mu = 0.5$. However, this was not the case (only for $\mu \leq 0.25$) and the reason is unclear. (We did not do experiments with an x^\dagger that does not satisfy the regularity condition (2.8), though). Still, the result that the simple-L methods perform better for small μ is backed by the numerical results.

TABLE 3.1

Tikhonov regularization, diagonal operator: median of ratio (3.1) of errors rules over 10 runs.

	simple-L	simple-L rat.	QO	L-curve
<i>s = 2, μ = 0.25</i>				
<i>δ small</i>	1.02	1.02	1.03	9.49
<i>δ medium</i>	1.01	1.02	1.08	1.78
<i>δ large</i>	1.79	1.06	1.18	1.15
<i>δ = 50%</i>	1.97	3.64	1.42	1.46
<i>s = 2, μ = 0.5</i>				
<i>δ small</i>	1.48	1.48	1.01	50.68
<i>δ medium</i>	1.66	1.72	1.07	3.78
<i>δ large</i>	1.78	1.59	1.01	2.52
<i>δ = 50%</i>	3.09	5.07	1.48	1.90
<i>s = 2, μ = 1</i>				
<i>δ small</i>	3.88	3.88	1.07	77.12
<i>δ medium</i>	2.01	2.01	1.07	7.98
<i>δ large</i>	1.57	1.66	1.08	2.33
<i>δ = 50%</i>	2.97	4.07	1.27	1.32

TABLE 3.2

Tikhonov regularization, rotational blur operator: median of ratio (3.1) of errors rules over 10 runs.

	simple-L	simple-L rat.	QO	L-curve
<i>δ small</i>	1.29	1.29	74.04	9.43
<i>δ medium</i>	1.05	1.09	1.19	2.59
<i>δ large</i>	1.02	1.01	1.10	1.02
<i>δ = 50%</i>	1.00	65.21	1.05	1.00

3.1.2. Examples from IR tools. For the next scenario, we consider a rotational blurring operator from IR Tools [11], namely `PRblur` which outputs a sparse operator (which we chose as 1024×1024) and seek to reconstruct the satellite image solution provided in the package. The data is corrupted with white Gaussian noise which is chosen such that $\|e\|/\|x^\dagger\|$ yields the relative noise level. Note that the operator and solution are normalized such that $\|A\| = \|x^\dagger\| = 1$ and our parameter search is restricted to the interval $\alpha \in [10^{-10}, \|A\|^2]$. Similarly as for the previous experiment, in Table 3.2, we record the median of the values of J over 10 different experiments with varying random noise realizations.

Note that the simple-L rules outperform the other ones in the majority of cases. However, the margin of improvement compared to the QO rule is not large. We observed that for small noise, the QO rule often overestimates the optimal parameter. All rules performed quite well, and the L-curve method in particular showed noticeable improvement as the noise level increased. The simple-L ratio method failed for 50% noise which may indicate that the smallness condition on δ in Theorem 2.13 is non-negligible.

REMARK 3.1. Whilst using the IRtools package, we encountered a occurring problem, especially for the tomography operator (to be discussed), whereby the error and functional plots did not display the typical shape one would expect. In particular, the stability error did not “blow up” as $\alpha \rightarrow 0$. Similar problems have been encountered for the L-curve method, which fails to show a clear corner point.

Remarkably, the illustrated convergence theory for heuristic rules via Muckenhoupt-type conditions can elucidate these problems quite well: The theoretical explanation for this appears to be that the tomography operator in IRtools is *not sufficiently ill-conditioned* (or “too well-

TABLE 3.3
Tikhonov regularization, tomography operator: median of ratio (3.1) of errors rules over 10 runs.

	simple-L	simple-L rat.	QO	L-curve
δ medium	1.36	1.36	4.32	1.68
δ large	1.38	1.39	1.20	2.71
$\delta = 50\%$	2.37	1.99	1.63	4.00

posed”), i.e., the singular values do not decay sufficiently fast. It can be shown that here the MC-conditions may not be satisfied for typical noise, which is a cause for possible failure of the heuristic rules. (Note that the MC-conditions require the noise to be more “irregular” than the range of A , and if A is less smoothing, then it is more likely that it belongs to the range of A). This reasoning also matches with the observation that if the ill-posedness is increased by employing a stronger regularization (see, e.g., the results with TV-regularization below), then the difficulties disappear. Interesting to note is that such problems do not occur for the operators in Hansen’s Regularization Tools problems [16], which appear to be sufficiently smoothing.

We consider now the tomography operator from the IR Tools package, i.e., `PRTOMO` with the true solution being a Shepp Logan phantom. The operator in question is of size 8100×1024 . In Table 3.3, one can find a record of the median values of J for 10 different realizations of each noise level, where we omit the “small” noise case in this scenario as the problem is not sufficiently ill-conditioned so that for small noise levels, $\alpha_* = \alpha_{\min}$ yields the best choice. We also subsequently switch the 1% noise level consideration for 2%. (We revert back to 1% noise level in the subsequent scenarios).

In this scenario, the L-curve method actually performed worse for larger noise and most often chose $\alpha_* = \alpha_{\min}$. All rules, however, were guilty of quite often underestimating the optimal parameter, which is a result of a lack of a sufficiently high ill-conditioning (or, in other words, of the MC-conditions not being satisfied) mentioned in the paragraph above.

3.2. Convex Tikhonov regularization. We now investigate the heuristic rules for convex Tikhonov regularization, i.e., we consider x_α^δ as the minimizer of the functional (2.12) with a nonquadratic penalty R . Note that the convergence theory of the present paper does not cover this case. For the HD, HR, and QO-rules, some convergence results of the theory in [20] have been extended to the convex case in [25].

Henceforth, the simple-L methods consist of minimizing the functionals (2.12) and (2.14). Note that we did consider (2.13) as an alternative “convexification” of the simple L-curve method, but the former method appeared to yield more fruitful results, and we therefore opted to stick with it.

3.2.1. ℓ^1 regularization. To begin with, we consider $R = \|\cdot\|_1$ and the rotational blur operator as before (of the same size as our earlier configuration, too), but this time we would like to reconstruct a sparse solution x^\dagger , and therefore we opt for the `sppattern` solution from the IRtools package which is a sparse image of geometric shapes. We choose Gaussian white noise as before, corresponding to the respective noise levels. Note that we compute a minimizer via FISTA [3]. In this case, we measure the error by the ℓ^1 norm, i.e., we compute J with $d(x, y) := \|x - y\|_1$.

In our experiments, we observed that the values of the aforementioned simple-L functionals were particularly small, therefore on occasion yielding negative values due to numerical errors. This problem was easily rectified however by taking the absolute value of (2.12) and (2.14), respectively, which is theoretically equivalent to the original functionals in any case. For the quasi-optimality functional, we opted to use the so-called right quasi-optimality

TABLE 3.4
 ℓ^1 regularization, rotational blur operator: median of ratio (3.1) of errors rules over 10 runs.

	simple-L	simple-L rat.	QO	L-curve
δ small	2.07	2.07	1.18	2.19
δ medium	4.82	4.82	1.04	4.68
δ large	1.00	1.00	1.00	6.73
$\delta = 50\%$	1.08	1.08	1.08	10.37

TABLE 3.5
 $\ell^{\frac{3}{2}}$ regularization, diagonal operator: median of ratio (3.1) of errors rules over 10 runs.

	simple-L	simple-L rat.	QO	L-curve
$s = 2, \mu = 0.25$				
δ small	6.59	6.59	1.02	471.03
δ medium	1.95	1.95	1.18	31.22
δ large	1.10	1.10	1.07	1.11
$\delta = 50\%$	1.13	1.21	1.11	1.22
$s = 2, \mu = 0.5$				
δ small	14.41	14.41	1.00	8.91
δ medium	2.05	2.05	1.01	115.00
δ large	1.09	1.09	1.03	1.15
$\delta = 50\%$	4.72	5.61	1.01	1.80
$s = 2, \mu = 1$				
δ small	20.51	20.51	1.46	4.29
δ medium	1.36	1.36	1.33	107.77
δ large	1.14	1.14	1.40	1.34
$\delta = 50\%$	7.06	9.28	1.08	1.53

rule [25]. For selecting the parameter according to the L-curve method of Hansen, maximizing the curvature via (1.7) is no longer an implementable strategy as R is now non-smooth. However, it was still possible to compute the corner point due to the discretization of the problem. In Table 3.4, one may find a recording of the results.

As mentioned already, the simple-L functionals produced very small values and were somewhat oscillatory, i.e., they were prone to exhibiting multiple local minima. Our algorithm selected the smallest interior minimum, but in some plots, we observed that there were larger local minima which would have corresponded to a more accurate estimation of the optimal parameter. It should be noted that for medium noise, the L-curve was quite “hit & miss” and for larger noise, quite unsatisfactory.

3.3. $\ell^{\frac{3}{2}}$ regularization. Continuing with the theme of convex Tikhonov regularization and more specifically with ℓ^p regularization for $p \neq 2$, we now consider (2.12) with $R = \|\cdot\|_{\frac{3}{2}}$ (i.e., $p = \frac{3}{2}$). The forward operator $A : \ell^{\frac{3}{2}}(\mathbb{N}) \rightarrow \ell^2(\mathbb{N})$ is a diagonal operator with polynomially decaying singular values as before, i.e., $\sigma_i = i^{-s}$, and we also consider a solution with polynomial decay $\langle x^\dagger, v_i \rangle = (-1)^i i^{-\tau}$ and add random noise $\langle e, u_i \rangle = \delta i^{-0.6} \tilde{e}_i$. The size of the operator in question is 625×625 . Note that in this scenario, we are easily able to compute the Tikhonov solution and second Bregman iterate as we have a closed form solution of the associated proximal mapping operator; see [25].

Results are compiled in Table 3.5 and the following observations are noted:

- Barring the quasi-optimality rule, all methods were generally subpar in case of small noise for all tested smoothness indices. In general, the quasi-optimality rule would

TABLE 3.6
TV regularization, tomography operator: median of ratio (3.1) of errors rules over 10 runs.

	simple-L	simple-L rat.	QO	L-curve
δ small	6.86	6.86	504.24	5.65
δ medium	59.55	6.67	61.85	8.43
δ large	8.17	6.98	8.20	9.73
$\delta = 50\%$	2.35	2.47	2.35	9.99

appear to be the best performing overall at least, although trumped on a few occasions.

- The “sweet spot” for both simple-L methods appears to be medium to large noise. Overall, at least, they appear to perform marginally better for smaller smoothness indices μ . The original L-curve method performs quite well for larger noise, as has been observed in other experiments, but the margin for error is quite large for smaller noise levels.

3.4. TV regularization. We now suppose that x_α^δ is the minimizer of (2.12) with $R = |\cdot|_{TV}$ the total variation seminorm. Note that for numerical implementation, the above functional is often discretized as $R(x) = \sum \|\nabla x\|_1$, with ∇ denoting a (e.g., forward) difference operator. The functional is minimized using FISTA with the proximal mapping operator for the total variation seminorm being computed by a fast Newton-type method as in [25]. In this case, we compute the error with respect to α via the so-called strict metric

$$d_{\text{strict}}(x_\alpha^\delta, x^\dagger) := |R(x_\alpha^\delta) - R(x^\dagger)| + \|x_\alpha^\delta - x^\dagger\|_1,$$

which was suggested in, e.g., [22], and we subsequently record the values of J with $d = d_{\text{strict}}$, the results of which are provided in Table 3.6. The operator in question is the tomography one arising from `PRTOMO` with the same configuration as before. We add white Gaussian noise, corresponding to the respective noise levels.

We note the following observations: All the functionals were oscillatory, exhibiting local minima which were much more pronounced compared to the linear case. This oscillatory behavior is often a cause for a selection of a false parameter; cf. the subpar results in Table 3.6. Contrary to the linear regularization case, the QO-rule is not necessarily always robust for convex Tikhonov regularization, which is consistent with the numerical findings of [25]. An inspection of this table reveals that the simple-L ratio method appears to be the best performing overall which we also observed in other experiments involving TV regularization not recorded here.

3.5. Summary. To summarize the numerical results presented above, the simple-L methods are near optimal for linear Tikhonov regularization in case of low smoothness of the exact solution. Moreover, the simple-L rule in particular edges the simple-L ratio rule, but the margin of difference is small and only apparent for larger noise levels.

We also considered convex Tikhonov regularization for which the simple-L functionals had to be adapted from their original forms. In any case, they were successfully implemented and demonstrated satisfactory results. Interesting to note however, was that in this setting, the simple-L ratio method appeared to present itself as the slightly superior one of the two variants.

4. Conclusion. In conclusion, we reduced the standard L-curve method for parameter selection to a minimization problem of an error estimating surrogate functional from which two new parameter choice rules were born: the simple-L and simple-L ratio methods. The rules allow for an analysis yielding convergence rates for Tikhonov regularization under a

Muckenhoupt-type condition MC_2 , akin to that required for the quasi-optimality rule, but they saturate early like the heuristic discrepancy rule.

The subsequent numerical experiments furthermore verified that the simple-L methods are not only capable of substituting as parameter choice rules for the L-curve method, but also outperform it the majority of the time, performing often similarly and sometimes (especially for low smoothness) even superior to the quasi-optimality rule, whilst being much easier to implement than the original L-curve method.

Acknowledgement. We would like to thank the anonymous referees for their useful comments.

REFERENCES

- [1] A. BAKUSHINSKII, *Remarks on choosing a regularization parameter using the quasioptimality and ratio criterion*, USSR Comput. Maths. Math. Phys., 24 (1984), pp. 181–182.
- [2] F. BAUER AND M. A. LUKAS, *Comparing parameter choice methods for regularization of ill-posed problems*, Math. Comput. Simulation, 81 (2011), pp. 1795–1841.
- [3] A. BECK AND M. TEOULLE, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), pp. 183–202.
- [4] C. BREZINSKI, G. RODRIGUEZ, AND S. SEATZU, *Error estimates for linear systems with applications to regularization*, Numer. Algorithms, 49 (2008), pp. 85–104.
- [5] ———, *Error estimates for the regularization of least squares problems*, Numer. Algorithms, 51 (2009), pp. 61–76.
- [6] D. CALVETTI, P. C. HANSEN, AND L. REICHEL, *L-curve curvature bounds via Lanczos bidiagonalization*, vol. 14, 2002, pp. 20–35.
- [7] P. COLLI FRANZONE, L. GUERRI, S. TENTONI, C. VIGANOTTI, S. BARUFFI, S. SPAGGIARI, AND B. TACCARDI, *A mathematical procedure for solving the inverse potential problem of electrocardiography. Analysis of the time-space accuracy from in vitro experimental data*, Math. Biosci., 77 (1985), pp. 353–396.
- [8] H. W. ENGL, M. HANKE, AND A. NEUBAUER, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996.
- [9] G. FRASSO AND P. H. C. EILERS, *L- and V-curves for optimal smoothing*, Stat. Model., 15 (2015), pp. 91–111.
- [10] K. FRICK, D. A. LORENZ, AND E. RESMERITA, *Morozov’s principle for the augmented Lagrangian method applied to linear inverse problems*, Multiscale Model. Simul., 9 (2011), pp. 1528–1548.
- [11] S. GAZZOLA, P. C. HANSEN, AND J. G. NAGY, *IR Tools: a MATLAB package of iterative regularization methods and large-scale test problems*, Numer. Algorithms, 81 (2019), pp. 773–811.
- [12] U. HÄMARIK, R. PALM, AND T. RAUS, *Comparison of parameter choices in regularization algorithms in case of different information about noise level*, Calcolo, 48 (2011), pp. 47–59.
- [13] ———, *A family of rules for parameter choice in Tikhonov regularization of ill-posed problems with inexact noise level*, J. Comput. Appl. Math., 236 (2012), pp. 2146–2157.
- [14] M. HANKE, *Limitations of the L-curve method in ill-posed problems*, BIT, 36 (1996), pp. 287–301.
- [15] P. C. HANSEN, *Analysis of discrete ill-posed problems by means of the L-curve*, SIAM Rev., 34 (1992), pp. 561–580.
- [16] ———, *Regularization tools: a Matlab package for analysis and solution of discrete ill-posed problems*, Numer. Algorithms, 6 (1994), pp. 1–35.
- [17] ———, *Rank-Deficient and Discrete Ill-Posed Problems*, SIAM, Philadelphia, 1998.
- [18] ———, *The L-curve and its use in the numerical treatment of inverse problems*, in Computational Inverse Problems in Electrocardiography, P. Johnston, ed., vol. 5 of Advances in Computational Bioengineering Series, Southhampton, 2001, WIT Press, pp. 119–142.
- [19] P. C. HANSEN AND D. P. O’LEARY, *The use of the L-curve in the regularization of discrete ill-posed problems*, SIAM J. Sci. Comput., 14 (1993), pp. 1487–1503.
- [20] S. KINDERMANN, *Convergence analysis of minimization-based noise level-free parameter choice rules for linear ill-posed problems*, Electron. Trans. Numer. Anal., 38 (2011), pp. 233–257.
<http://etna.ricam.oeaw.ac.at/vol.38.2011/pp233-257.dir/pp233-257.pdf>
- [21] ———, *Discretization independent convergence rates for noise level-free parameter choice rules for the regularization of ill-conditioned problems*, Electron. Trans. Numer. Anal., 40 (2013), pp. 58–81.
<http://etna.ricam.oeaw.ac.at/vol.40.2013/pp58-81.dir/pp58-81.pdf>
- [22] S. KINDERMANN, L. D. MUTIMBU, AND E. RESMERITA, *A numerical study of heuristic parameter choice rules for total variation regularization*, J. Inverse Ill-Posed Probl., 22 (2014), pp. 63–94.
- [23] S. KINDERMANN AND A. NEUBAUER, *On the convergence of the quasioptimality criterion for (iterated) Tikhonov regularization*, Inverse Probl. Imaging, 2 (2008), pp. 291–299.

- [24] S. KINDERMANN, S. PEREVERZYEV, JR., AND A. PILIPENKO, *The quasi-optimality criterion in the linear functional strategy*, Inverse Problems, 34 (2018), pp. 075001, 24.
- [25] S. KINDERMANN AND K. RAIK, *Convergence of heuristic parameter choice rules for convex Tikhonov regularisation*, Preprint on arXiv, 2019. <https://arxiv.org/abs/1905.06828>, 2019
- [26] D. KRAWCZYK-STANĀDO AND M. RUDNICKI, *Regularization parameter selection in discrete ill-posed problems—the use of the U-curve*, Int. J. Appl. Math. Comput. Sci., 17 (2007), pp. 157–164.
- [27] C. L. LAWSON AND R. J. HANSON, *Solving Least Squares Problems*, Prentice-Hall, Englewood Cliffs, 1974.
- [28] S. LU AND P. MATHÉ, *Heuristic parameter selection based on functional minimization: optimality and model function approach*, Math. Comp., 82 (2013), pp. 1609–1630.
- [29] A. NEUBAUER, *The convergence of a new heuristic parameter selection criterion for general regularization methods*, Inverse Problems, 24 (2008), Art. 055005, 10 pages.
- [30] R. PALM, *Numerical Comparison of Regularization Algorithms for Solving Ill-Posed Problems*, PhD. Thesis, Institute of Computer Science, University of Tartu, Tartu, 2010.
- [31] T. RAUS, *An a posteriori choice of the regularization parameter in case of approximately given error bound of data*, Tartu Riikl. ÜI. Toimetised, 913 (1990), pp. 73–87.
- [32] T. RAUS AND U. HÄMARIK, *Q-curve and area rules for choosing heuristic parameter in Tikhonov regularization*, Preprint on arXiv, 2018. <https://arxiv.org/abs/1809.02061>
- [33] ———, *On the quasioptimal regularization parameter choices for solving ill-posed problems*, J. Inverse Ill-Posed Probl., 15 (2007), pp. 419–439.
- [34] ———, *On numerical realization of quasioptimal parameter choices in (iterated) Tikhonov and Lavrentiev regularization*, Math. Model. Anal., 14 (2009), pp. 99–108.
- [35] T. REGIŃSKA, *A regularization parameter in discrete ill-posed problems*, SIAM J. Sci. Comput., 17 (1996), pp. 740–749.
- [36] L. REICHEL AND G. RODRIGUEZ, *Old and new parameter choice rules for discrete ill-posed problems*, Numer. Algorithms, 63 (2013), pp. 65–87.
- [37] C. R. VOGEL, *Non-convergence of the L-curve regularization parameter selection method*, Inverse Problems, 12 (1996), pp. 535–547.
- [38] L. XU, L. LI, W. WANG, AND Y. GAO, *CT image reconstruction algorithms based on the Hanke Raus parameter choice rule*, Inverse Probl. Sci. Eng., 28 (2020), pp. 87–103.