8-30-2018

On the convergence of a heuristic parameter choice rule for Tikhonov regularization

Mark Gockenbach
Michigan Technological University

Elaheh Gorgin
Michigan Technological University

Follow this and additional works at: https://digitalcommons.mtu.edu/math-fp

Part of the Mathematics Commons

SIAM J. Sci. Comput.
Vol. 40, No. 4, pp. A2694–A2719

ON THE CONVERGENCE OF A HEURISTIC PARAMETER CHOICE RULE FOR TIKHONOV REGULARIZATION*

MARK S. GOCKENBACH† AND ELAHEH GORGIN‡

Abstract. Multiplicative regularization solves a linear inverse problem by minimizing the product of the norm of the data misfit and the norm of the solution. This technique is related to Tikhonov regularization with the parameter chosen to make the data misfit and regularization terms (of the Tikhonov objective function) equal. This suggests a heuristic parameter choice method, equivalent to the rule previously proposed by Reginska. Reginska's rule is well defined provided the data is sufficiently close to exact data and does not lie in the range of the operator. If a sufficiently large portion of the data error lies outside the range of the operator, then the solution defined by Reginska's rule converges weakly to the exact solution as the data error converges to zero. The regularization parameter converges to zero like the square of the norm of the data noise, leading to under-regularization for small noise levels. Nevertheless, the method performs well on a suite of test problems, as shown by comparison with the L-curve, generalized cross-validation, quasi-optimality, and Hanke–Raus parameter choice methods. A modification of the approach yields a heuristic parameter choice rule that is provably convergent (in the norm topology) under the restrictions on the data error described above, as long as the exact solution has a small amount of additional smoothness. On the test problems considered here, the modified rule outperforms all of the above heuristic methods, although it is only slightly better than the quasi-optimality rule.

Key words. inverse problems, Tikhonov regularization, convergence analysis

AMS subject classifications. 65J22, 65R32

DOI. 10.1137/17M1138698

1. Introduction.

The most popular regularization methods for inverse problems are based on optimization, with the objective function consisting of two parts. The first part measures how well the proposed solution fits the given data, and the second part (the regularization term) penalizes undesirable properties of the proposed solution (such as a large norm or nonsmoothness). Typically these two terms are added together with a weighting parameter multiplying the regularization term. A large parameter implies more regularization (that is, a smaller or smoother solution). The classic method of Tikhonov regularization, which we now describe, is a prime example of this approach.

We will discuss linear inverse problems of the form $Tx = y$, where $T : X \to Y$ is a bounded linear operator and X and Y are Hilbert spaces. We assume that there exist exact data $y^* \in Y$ and an exact solution $x^* \in X$ such that $Tx^* = y^*$, that $y \in Y$ is a measurement of y^*, and it is desired to estimate x^* by solving (in some sense) $Tx = y$. We will assume throughout that $y^* \neq 0$ (and hence also $x^* \neq 0$). If the null space $\mathcal{N}(T)$ is nontrivial, then we assume that $x^* \in \mathcal{N}(T)^\perp$. Since $Tx = y$ may not have a solution, it is natural to consider the least-squares problem

$$\min_{x\in X} \|Tx - y\|_Y^2. \tag{1}$$

*Submitted to the journal's Methods and Algorithms for Scientific Computing section July 14, 2017; accepted for publication (in revised form) May 25, 2018; published electronically August 30, 2018.
http://www.siam.org/journals/sisc/40-4/M113869.html
†Department of Mathematical Sciences, Michigan Technological University, 1400 Townsend Drive, Houghton, MI 49931-1295 (msgocken@mtu.edu).
‡Department of Mathematics, Minot State University, 500 University Avenue West, Minot, ND 58707 (elaheh.gorgin@minotstateu.edu).

However, (1) has a solution if and only if y belongs to the subspace $\mathcal{R}(T) \oplus \mathcal{R}(T)^\perp$, where $\mathcal{R}(T)$ denotes the range of T. When $\mathcal{R}(T)$ fails to be closed, $\mathcal{R}(T) \oplus \mathcal{R}(T)^\perp$ is a proper dense subspace of Y, and hence a measurement y of y^* is likely to lie outside of this set. Moreover, the *pseudoinverse* $T^\dagger$ of T, which maps $y \in D(T^\dagger) = \mathcal{R}(T)\oplus\mathcal{R}(T)^\perp$ to the unique solution of (1) of smallest norm (that is, to the minimum-norm least-squares solution of $Tx = y$), is unbounded when $\mathcal{R}(T)$ fails to be closed.

Tikhonov regularization addresses the shortcomings of the least-squares approach by replacing (1) with

$$\min_{x\in X} \|Tx - y\|_Y^2 + \lambda\|x\|_X^2, \tag{2}$$

which has a unique solution $x_{\lambda,y} = (T^*T + \lambda I)^{-1}T^*y$ for each $y \in Y$, provided the *regularization parameter* λ is positive. Tikhonov regularization can be viewed as replacing the unbounded operator $T^\dagger$ with the bounded operator $(T^*T + \lambda I)^{-1}T^*$. It can be shown that Tikhonov regularization is effective in the sense that $x_{\lambda,y} \to x^*$ as $y \to y^*$, provided λ is chosen appropriately. Since we will frequently refer to the operator $T^*T + \lambda I$, we introduce the notation

$$N_\lambda = T^*T + \lambda I. \tag{3}$$

A significant drawback for methods such as Tikhonov regularization is the need to choose the regularization parameter. Various approaches have been proposed to address this problem; these techniques can be classified depending on what information they use about the noisy data vector y. If λ is the regularization parameter and $\delta = \|y - y^*\|_Y$, where y^* is the exact data vector, then a parameter choice rule takes one of the following forms: $\lambda = \lambda(\delta)$, $\lambda = \lambda(\delta, y)$, and $\lambda = \lambda(y)$. Engl, Hanke, and Neubauer [6] call such rules a priori, a posteriori, and error-free, respectively. For the third type of parameter choice method, the term *heuristic* is also used.

It should be noted that one would not expect to know the exact value of $\|y-y^*\|_Y$. Thus, in practice, δ is taken as an estimate of this error, and it is usually assumed that $\|y - y^*\|_Y \le \delta$ holds.

One of the most fundamental facts about heuristic parameter choice methods is that they cannot be convergent. We say that a parameter choice method $\lambda = \lambda(\delta, y)$ is *convergent* if for each sequence $(\delta_n, y_n) \in [0,\infty) \times Y$ such that

$$\delta_n \to 0^+ \text{ and } \|y_n - y^*\|_Y \le \delta_n,$$

we have

$$x_{\lambda(\delta_n,y_n),y_n} \to x^* = T^\dagger y^*.$$

Bakushinskii [3] showed that a parameter choice rule of the form $\lambda = \lambda(y)$ cannot be convergent in this sense (because, as one can show, it would have to choose $\lambda = 0$ for $y \in \mathcal{R}(T)$, implying that the regularized solution is $T^\dagger y$ for $y \in \mathcal{R}(T)$; but $T^\dagger$ is unbounded on $\mathcal{R}(T)$). Nevertheless, it is possible to prove convergence for a heuristic parameter choice rule provided some assumptions are made about how the noisy data y converges to y^*. We will mention several existing results of this type below, and the convergence results in this paper are of this type.

A popular a posteriori method is the Morozov discrepancy principle [22], which chooses the regularized solution that produces an error in the data of the same size as the given estimate. Heuristic methods include the L-curve criterion (Hansen [13]; see also [16]), the generalized cross-validation (GCV) rule (Wahba [32]), the quasi-optimality criterion (Tikhonov and Arsenin [29, pp. 93–94]), and the Hanke–Raus rule (Hanke and Raus [12]).

The L-curve is a graph of the regularization term versus the residual in the data. This graph frequently has a characteristic L shape, and the corner corresponds to the point where further reduction in the residual comes only at the expense of a drastic increase in the regularization term (that is, in undesirable properties of the regularized solution). The L-curve criterion chooses the regularization parameter that corresponds to this corner (which is usually defined as the point of maximum curvature of the L-curve in a log-log plot).

GCV, which applies when Y is a finite-dimensional space, is based on minimizing an estimator of the predictive mean-square error $\|Tx_{\lambda,y} - y^*\|_Y^2$, the residual with respect to the exact data y^*. This estimator is

$$V(\lambda) = \left(\frac{\|Tx_{\lambda,y} - y\|_Y}{\operatorname{trace}\left(\frac{\lambda}{m}\left(TT^* + \lambda I\right)^{-1}\right)} \right)^2,$$

where m is the dimension of Y (see [6, section 4.5]).

The quasi-optimality criterion chooses the regularization parameter to minimize

$$\psi_\lambda = \lambda \left\| \frac{\partial}{\partial \lambda}(x_{\lambda,y}) \right\|_X .$$

This parameter choice method was apparently developed without a strong intuitive or theoretical basis; in fact, Morozov [23, pp. 239–240] wrote, "Unfortunately, it has not been possible to justify this technique for choosing the parameter although it is widely used for unstable problems." However, Kitagawa [21] and Hansen [13] have shown that the quasi-optimality approach seeks to approximately minimize the total error in $x_{\lambda,y}$. Kindermann and Neubauer [20] proved that the method is convergent under certain assumptions on the spectral properties of the noise in the data. In the case that T is a compact operator, these assumptions imply that the Fourier components of the noisy data do not decay to zero too quickly (that is, that the noisy data has significant high-frequency content). Neubauer generalized these results to a family of abstract regularization methods in [24]. Hämarik, Palm, and Raus [10] provide an analysis of a family of minimization-based strategies that include the quasi-optimality approach. Their results show that the quasi-optimality method converges at an optimal worst-case rate provided the minimum value of ψ_λ satisfies $\psi_\lambda \sim \delta/\sqrt{\lambda}$. However, it is not clear how to guarantee this condition. (The result of [10] is similar to that of Hanke–Raus, described in the next paragraph and in more detail following Theorem 8 below.)

The Hanke–Raus [12] rule chooses $\lambda > 0$ to minimize

$$\phi_\lambda = \lambda \left[\langle y, (TT^* + \lambda I)^{-3} y\rangle_Y\right]^{1/2} . \tag{4}$$

This heuristic method is derived from an optimal order a posteriori parameter choice rule, which in turn is a modification of the Morozov discrepancy principle. (For the original a posteriori rule, see [25], [7], or [5]. The heuristic rule (4) is derived in section 2.1 of [12].) Under the hypotheses of our Theorem 7 below, Theorem 3.3 of [12] shows that $x_{\lambda,y}$, with λ chosen by minimizing (4), converges to x^* in norm as $y \to y^*$. We discuss this in more detail following the proof of Theorem 8, but the key hypothesis is that $y \notin \overline{\mathcal{R}(T)}$ and, moreover, that y does not follow a path that is tangent to $\overline{\mathcal{R}(T)}$ as it approaches y^*.

Kindermann [18] proved that the Hanke–Raus and quasi-optimality rules, along with certain other minimization-based rules, can be guaranteed to converge provided that the noise in the data is of a certain type. Specifically, suppose $\lambda = \lambda(y)$ is chosen by minimizing a functional $\psi(\lambda, y)$ and that the regularized solution is denoted by $R_\lambda y$. It is assumed that $y \to y^*$ in such a way that

$$\inf_{\lambda>0} \psi(\lambda, y) < \liminf_{\lambda\to 0} \psi(\lambda, y) \tag{5}$$

(that is, that only data y that satisfy (5) are admitted) and also that ψ satisfies

$$\lim_{(\lambda,y)\to(0,y^*)} \psi(\lambda, y) = 0 \ \Rightarrow\ \|R_\lambda y - R_\lambda y^*\|_X \to 0.$$

Under these assumptions, Kindermann proved that the parameter choice rule is convergent. In [19], Kindermann extended his analysis to finite-dimensional problems, deriving results that are independent of the discretization. In addition to the Hanke–Raus and quasi-optimality rules, he extended the analysis to the GCV rule, although obtaining only partial results.

Each of the methods described above has its own difficulties. The obvious problem with Morozov discrepancy principle is that a good estimate of the error in the data may not be available. While the L-curve criterion works well for many problems, it has been shown to perform poorly on certain problems (see [31], [11]), and there is little theory supporting this approach. For a discussion of the strengths and weaknesses of the L-curve criterion, see section 4.5 of the book by Engl, Hanke, and Neubauer [6], which also discusses the other heuristic methods considered here. The GCV functional is often nearly constant near the minimizer and can have multiple nearby local, non-global minimizers. The quasi-optimality method requires minimization of a function that frequently has multiple local minima (see [15, p. 183]); moreover, according to the authors' tests, sometimes two local minima give similar values of the objective function. Recent work by Raus and Hämarik [26] does show that one of the local minimizers of the quasi-optimality function is always *pseudo-optimal*, that is, its error is at most a constant times the sum of the regularization error and the perturbation error. They also propose algorithms for computing a good local minimizer.

Because of the difficulties associated with choosing an appropriate regularization parameter, van den Berg, van Broekhoven, and Abubakar [30] (see also [1]) proposed incorporating the regularization penalty into the objective function by multiplying the data misfit term by it, rather than by adding it. In our context, this implies seeking a nonzero local minimizer of

$$J(x; y) = \|x\|_X^2 \|Tx - y\|_Y^2. \tag{6}$$

This approach, which is described in detail below, does not require any regularization parameter and therefore avoids the problems discussed above. We will refer to this approach as *multiplicative regularization*. The work in [30] and [1] was in the context of total variation regularization, and the authors provided examples showing that the approach can work well in practice.

As we will show below, multiplicative regularization is closely related to a heuristic parameter choice rule for Tikhonov regularization, namely the rule that chooses $\lambda > 0$ to satisfy the fixed point equation

$$\lambda = \frac{\|Tx_{\lambda,y} - y\|_Y^2}{\|x_{\lambda,y}\|_X^2}. \tag{7}$$

Assuming such a value of λ exists, it depends on y only, not on δ; in other words, (7) defines a heuristic parameter choice rule that we will denote by $\lambda = \Lambda(y)$.

Reginska [27] analyzed a similar approach to regularization. She defined an objective function

$$\phi(\lambda) = \|x_{\lambda,y}\|_X^2 \|Tx_{\lambda,y} - y\|_Y^2$$

and proposed choosing λ by minimizing this function. (More generally, she also considered minimization of $\phi_\beta(\lambda) = \|x_{\lambda,y}\|_X^{2\beta} \|Tx_{\lambda,y} - y\|_Y^2$ for $\beta > 0$.) It is easy to see that any stationary point of ϕ satisfies (7), and therefore Reginska's formulation leads to the same parameter choice rule as does multiplicative regularization.

Multiplicative regularization has been studied in the Ph.D. dissertation of Orozco Rodríguez [28]. He compared the performance of multiplicative regularization with the L-curve criterion for an image deblurring problem and, emphasizing the original formulation as an optimization problem, derived conditions for the existence and identification of a nontrivial minimizer of $J(\cdot; y)$.

The purpose of this paper is to provide an analysis of (7) and, in particular, show that it is convergent under a simple restriction on the noise. Specifically, if $\overline{\mathcal{R}(T)}$ is a proper subspace of Y, $y \notin \overline{\mathcal{R}(T)}$, and y converges to y^* in such a way that the angle that y makes with $\overline{\mathcal{R}(T)}$ is sufficiently large, then $\Lambda(y) \sim \|y - y^*\|_Y^2$ as $y \to y^*$ (here $\Lambda(y)$ denotes the solution of (7)) and

$$x_{\Lambda(y),y} \to x^* \text{ weakly as } y \to y^*$$

(see Theorem 7). We can obtain this result because the assumptions on data mean that an estimate of δ such that $\delta \geq \|y - y^*\|_Y$ is implicitly available (although not used directly).

As is well known, to guarantee that $x_{\lambda,y} \to x^*$ strongly, λ must be asymptotically larger than $\|y - y^*\|_Y^2$; specifically, we need

$$\lambda \to 0^+ \text{ and } \frac{\|y - y^*\|_Y^2}{\lambda} \to 0 \text{ as } y \to 0.$$

We can improve on the weak convergence offered by (7) by modifying the fixed point equation to

$$\lambda = \frac{\|Tx_{\lambda,y} - y\|_Y^{2\mu}}{\|x_{\lambda,y}\|_X^{2\mu}}, \tag{8}$$

where $\mu \in (1/2, 1)$ is a constant. We will show that if the true solution x^* has some extra smoothness (for instance, if $x^* \in \mathcal{R}(T^*)$ or more generally $x^* \in \mathcal{R}((T^*T)^\nu)$ for $\nu > 0$), then (8) defines a unique parameter $\lambda = \Lambda_\mu(y)$ satisfying

$$\frac{\|y - y^*\|_Y^{2\mu}}{\Lambda_\mu(y)} \leq c$$

for some $c > 0$ and hence

$$x_{\Lambda_\mu(y),y} \to x^* \text{ strongly as } y \to y^*.$$

Moreover, we can derive the worst-case rate of convergence of $x_{\Lambda_\mu(y),y}$ to x^*, which turns out to be optimal for $\nu \in (0, 1/2)$ and suboptimal for $\nu \geq 1/2$. Once again, we must assume that $y \to y^*$ in such a way that $y \notin \overline{\mathcal{R}(T)}$ and y does not follow a path that is tangent to $\overline{\mathcal{R}(T)}$ as it approaches y^*.

The convergence results just described suggest that the parameter choice rules defined by (7) and (8) may not be effective when $\mathcal{R}(T)$ is dense in Y (since in that case y must lie in $\overline{\mathcal{R}(T)}$). More generally, the same difficulty is encountered if $Tx = y$ is discretized as $Ax = y$ ($A \in \mathbb{R}^{m\times n}$), where y lies in col(A). For problems presenting this issue, we propose to ensure that the discretization satisfies $m > n$, where m is the number of data samples collected. Implemented in this fashion, (7) and (8) appear to be effective even for such problems.

As noted above, (7) is essentially equivalent to Reginska's approach. However, it is emphasized that the rule (8) is not equivalent to minimizing Reginska's modified objective function. This rule appears not to have been considered previously.

In addition to proving the theoretical results described above, we perform numerical experiments to show that these heuristic rules are effective, comparing their performance to the L-curve, GCV, quasi-optimality, and Hanke–Raus criteria on a collection of 20 test problems. We will see that (7) does work well in many cases, although not quite as well as the L-curve and quasi-optimality approaches (it seems clearly superior to the GCV and Hanke–Raus methods, at least on our test problems). However, for μ close to but smaller than 1, (8) defines a parameter choice rule that seems to outperform the L-curve, GCV, and Hanke–Raus rules, and which is approximately as effective as the quasi-optimality rule.

2. Analysis of the parameter choice methods. As noted above, we assume that $T : X \to Y$ is a bounded linear operator from one Hilbert space to another, $y \in Y$ is given, and x is to be determined as an approximate solution of $Tx = y$. Multiplicative regularization determines x by solving

$$\min_{x\in X} J(x;y), \tag{9}$$

where $J(x;y) = \|x\|_X^2\|Tx-y\|_Y^2$. Similar to Tikhonov regularization, multiplicative regularization tries to identify a value of x that makes the residual $Tx-y$ small, while simultaneously not allowing x to be large. We notice, however, that (9) always has the global solution $x = 0$, which is not a meaningful solution to the inverse problem, so we interpret (9) as asking for a nonzero local minimizer of J.

To analyze (9), we notice that

$$J(x;y) = \|x\|_X^2\|Tx-y\|_Y^2 = \langle x,x\rangle_X\left(\langle x,T^*Tx\rangle_X - 2\langle T^*y,x\rangle_X + \langle y,y\rangle_Y\right).$$

We then have

$$\nabla J(x;y) = 2\left(\|Tx-y\|_Y^2 x + \|x\|_X^2\left(T^*Tx - T^*y\right)\right)$$

and, assuming $x \neq 0$,

$$\begin{aligned}\nabla J(x;y) = 0 &\Leftrightarrow \|Tx-y\|_Y^2 x + \|x\|_X^2\left(T^*Tx-T^*y\right) = 0\\ &\Leftrightarrow T^*Tx - T^*y + \frac{\|Tx-y\|_Y^2}{\|x\|_X^2}x = 0.\end{aligned}$$

This optimality condition reduces to the pair of simultaneous equations

$$T^*Tx + \lambda x = T^*y, \qquad \lambda = \frac{\|Tx-y\|_Y^2}{\|x\|_X^2}.$$

The first equation means that $x = x_{\lambda,y}$ for a certain value of λ; the second equation constrains that value of λ. It is a value of λ for which

$$\lambda = \frac{\|Tx_{\lambda,y} - y\|_Y^2}{\|x_{\lambda,y}\|_X^2} \;\Leftrightarrow\; \|Tx_{\lambda,y} - y\|_Y^2 = \lambda \|x_{\lambda,y}\|_X^2. \tag{10}$$

We see that multiplicative regularization is related to Tikhonov regularization when the regularization parameter λ is chosen by the rule (10). The reader will notice that this rule chooses λ by requiring the two terms in the regularized objective function to have equal value. Henceforth, we restrict our attention to this parameter choice method (rather than studying (9) directly). Since, as we pointed out above, (10) already arose in Reginska's work, we will refer to it as Reginska's parameter choice rule.

Given any $y \in Y$, we will always write $\overline{y}$ for the orthogonal projection of y onto $\overline{\mathcal{R}(T)}$ and $\hat{y} = y - \overline{y}$ for the orthogonal projection of y onto $\mathcal{R}(T)^{\perp}$. We will need several standard results about Tikhonov regularization and the operator N_λ defined in (3) (see, for example, Chapter 3 of [8]):

$$\|N_\lambda^{-1}T^*\| \le \frac{1}{2\sqrt{\lambda}} \text{ for all } \lambda > 0, \tag{11}$$

$$\|TN_\lambda^{-1}T^*\| \le 1 \text{ for all } \lambda > 0, \tag{12}$$

$$\sqrt{\lambda}N_\lambda^{-1}T^* \to 0 \text{ pointwise as } \lambda \to 0^+, \tag{13}$$

$$Tx_{\lambda,y} \to \overline{y} \text{ as } \lambda \to 0^+ \text{ for all } y \in Y, \tag{14}$$

$$x_{\lambda,y} - T^\dagger y = -\lambda N_\lambda^{-1}T^\dagger y \text{ for all } y \in D(T^\dagger), \tag{15}$$

$$\|x_{\lambda,y}\|_X \le \|T^\dagger y\|_X \text{ for all } y \in D(T^\dagger) \text{ and all } \lambda > 0. \tag{16}$$

Here is a preliminary result about Reginska's parameter choice rule.

LEMMA 1. *If $\|\hat{y}\|_Y > \frac{1}{2}\|y\|_Y$, then* (10) *has no solution.*

Proof. Assume that $y \in Y \setminus \overline{\mathcal{R}(T)}$ is given and $\lambda > 0$ satisfies (10). Since

$$\|Tx_{\lambda,y} - y\|_Y^2 = \|Tx_{\lambda,y} - \overline{y}\|_Y^2 + \|\hat{y}\|_Y^2,$$

(10) implies that

$$\|\hat{y}\|_Y^2 \le \lambda \|x_{\lambda,y}\|_X^2.$$

Applying (11), we see that

$$\lambda \|x_{\lambda,y}\|_X^2 = \lambda \|N_\lambda^{-1}T^*y\|_X^2 \le \lambda \left(\frac{1}{2\sqrt{\lambda}}\|y\|_Y\right)^2 = \frac{1}{4}\|y\|_Y^2.$$

It follows that if λ satisfies (10), then

$$\|\hat{y}\|_Y^2 \le \frac{1}{4}\|y\|_Y^2,$$

and the result follows. ∎

It follows that the most we can hope for is that (10) has a solution for all y sufficiently close to a given $y^* \in \mathcal{R}(T)$, which we prove below in Theorem 5. Writing $\Lambda(y)$ for this value of λ, we will also show that, under certain restrictions on the noisy data y, $\Lambda(y) \sim \|y - y^*\|_Y^2$. Since this value of λ is too small to guarantee strong convergence of $x_{\lambda,y}$ to x^*, we propose to modify (10) to

$$\lambda = \frac{\|Tx_{\lambda,y} - y\|_Y^{2\mu}}{\|x_{\lambda,y}\|_X^{2\mu}}, \tag{17}$$

where $\mu \in (1/2, 1)$ is a constant. We can perform much of the analysis of (10) and (17) together by allowing $\mu = 1$ in (17). It turns out to be easier to analyze (17) in the form

$$\lambda^{1/\mu} = \frac{\|Tx_{\lambda,y} - y\|_Y^2}{\|x_{\lambda,y}\|_X^2}. \tag{18}$$

We will refer to the rule that chooses λ to satisfy (17) or (18) as the modified Reginska, or MR, rule.

We begin by showing that (18) has a solution for all y sufficiently close to y^*. Equation (18) is equivalent to $f_\mu(\lambda, y) = \|\hat{y}\|_Y^2$, where $f_\mu : [0,\infty)\times Y \to \mathbb{R}$ is defined by

$$f_\mu(\lambda, y) = \begin{cases} \lambda^{1/\mu}\|x_{\lambda,y}\|_X^2 - \|Tx_{\lambda,y} - \overline{y}\|_Y^2 & \text{if } \lambda > 0, \\ 0 & \text{if } \lambda = 0. \end{cases}$$

LEMMA 2. *If $\mu \in (0, 1]$, then f_μ is continuous.*

Proof. It is straightforward to show that f_μ is continuous for $\lambda > 0$. Therefore, given $y_0 \in Y$, we must show that $f_\mu(\lambda, y) \to 0 = f_\mu(0, y_0)$ as $(\lambda, y) \to (0, y_0)$. We have

$$Tx_{\lambda,y} - \overline{y} = Tx_{\lambda,y_0} - \overline{y}_0 + Tx_{\lambda,y-y_0} - \overline{y} + \overline{y}_0 \to 0 \text{ as } (\lambda, y) \to (0, y_0)$$

(applying (12) and (14)). Also,

$$\begin{aligned}
\sqrt{\lambda}x_{\lambda,y} &= \sqrt{\lambda}x_{\lambda,y_0} + \sqrt{\lambda}x_{\lambda,y-y_0} \\
&\Rightarrow \sqrt{\lambda}\|x_{\lambda,y}\|_X \le \sqrt{\lambda}\|x_{\lambda,y_0}\|_X + \sqrt{\lambda}\|x_{\lambda,y-y_0}\|_X \\
&\Rightarrow \sqrt{\lambda}\|x_{\lambda,y}\|_X \le \sqrt{\lambda}\|N_\lambda^{-1}T^*y_0\|_X + \sqrt{\lambda}\|N_\lambda^{-1}T^*(y - y_0)\|_X \\
&\Rightarrow \sqrt{\lambda}\|x_{\lambda,y}\|_X \le \sqrt{\lambda}\|N_\lambda^{-1}T^*y_0\|_X + \frac{1}{2}\|y - y_0\|_Y \\
&\Rightarrow \lambda\|x_{\lambda,y}\|_X^2 \to 0 \text{ as } (\lambda, y) \to (0, y_0)
\end{aligned}$$

(by (11) and (13)). Since, for $\mu \in (0, 1)$, $\lambda^{1/\mu} = o(\lambda)$, it follows immediately that

$$\lambda^{1/\mu}\|x_{\lambda,y}\|_X^2 \to 0 \text{ as } (\lambda, y) \to (0, y_0)$$

and hence that

$$\lim_{(\lambda,y)\to(0,y_0)} f_\mu(\lambda, y) = 0 = f_\mu(0, y_0).$$

Thus f_μ is continuous. ∎

To prove that (18) has a solution for $1/2 < \mu < 1$, we have to assume that the true solution $x^* = T^\dagger y^*$ has some extra smoothness. We will use the following result.

LEMMA 3. *Suppose $x^* \in \mathcal{R}((T^*T)^\nu)$ for some $\nu > 0$. Then*

$$\nu \in \left(0, \frac{1}{2}\right) \;\Rightarrow\; \|Tx_{\lambda,y^*} - y^*\|_Y^2 = o\left(\lambda^{1+2\nu}\right)$$

and

$$\nu \ge \frac{1}{2} \;\Rightarrow\; \|Tx_{\lambda,y^*} - y^*\|_Y^2 \sim \lambda^2.$$

For a proof of this result, see Theorem 4.3, together with Remarks 4.15 and 4.19, of [6]. (Theorem 4.3 gives the corresponding "big-oh" estimate in the case that $\nu \in (0, 1/2)$, but Remark 4.19 suggests how to improve the estimate to "little-oh," as stated above.)

LEMMA 4.

1. *There exists* $\overline{\lambda} > 0$ *such that* $f_1(\overline{\lambda}, y^*) = 0$ *and* $f_1(\lambda, y^*) > 0$ *for all* λ *in the interval* $(0, \overline{\lambda})$.
2. *Suppose* $x^* \in \mathcal{R}((T^*T)^\nu)$ *and*

$$0 < \nu < \frac{1}{2} \text{ and } \frac{1}{1+2\nu} \le \mu < 1$$

or

$$\nu \ge \frac{1}{2} \text{ and } \frac{1}{2} < \mu < 1.$$

Then there exists $\overline{\lambda} > 0$ *such that* $f_\mu(\overline{\lambda}, y^*) = 0$ *and* $f_\mu(\lambda, y^*) > 0$ *for all* $\lambda \in (0, \overline{\lambda})$.

Proof.

1. Since $\|N_\lambda^{-1}\| \le \lambda^{-1}$, it follows that

$$x_{\lambda,y^*} \to 0 \text{ and } \lambda\|x_{\lambda,y^*}\|_X^2 \to 0 \text{ as } \lambda \to \infty$$

and hence that

$$\|Tx_{\lambda,y^*} - y^*\|_Y^2 \to \|y^*\|_Y^2 \text{ as } \lambda \to \infty.$$

Therefore,

$$f_1(\lambda, y^*) = \lambda\|x_{\lambda,y^*}\|_X^2 - \|Tx_{\lambda,y^*} - y^*\|_Y^2 \to -\|y^*\|_Y^2 < 0 \text{ as } \lambda \to \infty.$$

On the other hand,

$$\begin{aligned} f_1(\lambda, y^*) &= \lambda\left(\|x_{\lambda,y^*}\|_X^2 - \lambda^{-1}\|Tx_{\lambda,y^*} - y^*\|_Y^2\right) \\ &= \lambda\left(\|x_{\lambda,y^*}\|_X^2 - \lambda^{-1}\|Tx_{\lambda,y^*} - Tx^*\|_Y^2\right) \\ &= \lambda\left(\|x_{\lambda,y^*}\|_X^2 - \lambda^{-1}\|T\left(\lambda N_\lambda^{-1}x^*\right)\|_Y^2\right) \\ &= \lambda\big(\|x_{\lambda,y^*}\|_X^2 - \|\sqrt{\lambda}TN_\lambda^{-1}x^*\|_Y^2\big) \end{aligned}$$

(where we have used (15)). Since

$$\|x_{\lambda,y^*}\|_X \to \|x^*\|_X > 0 \text{ and } \sqrt{\lambda}TN_\lambda^{-1} \to 0 \text{ pointwise as } \lambda \to 0^+,$$

it follows that $f_1(\lambda, y^*) > 0$ for all $\lambda > 0$ sufficiently small. We can define

$$\overline{\lambda} = \sup\{\hat{\lambda} > 0 \,:\, f_1(\lambda, y^*) > 0 \text{ for all } \lambda \in (0, \hat{\lambda})\},$$

and the proof is complete.

2. As before, $f_\mu(\lambda, y^*) < 0$ for all $\lambda > 0$ sufficiently large. Suppose first that $\nu \ge 1/2$ and $\mu > 1/2$. By the preceding lemma, there exists $C > 0$ such that

$$\|Tx_{\lambda,y^*} - y^*\|_Y^2 \le C\lambda^2.$$

It follows that

$$f_\mu(\lambda, y^*) = \lambda^{1/\mu}\left(\|x_{\lambda,y^*}\|_X^2 - \lambda^{-1/\mu}\|Tx_{\lambda,y^*} - y^*\|_Y^2\right)$$
$$\geq \lambda^{1/\mu}\left(\|x_{\lambda,y^*}\|_X^2 - C\lambda^{2-1/\mu}\right).$$

Since $\|x_{\lambda,y^*}\|_X^2 \to \|x^*\|_X^2 > 0$ and $\lambda^{2-1/\mu} \to 0$ as $\lambda \to 0^+$, it follows that $f_\mu(\lambda, y^*) > 0$ for all $\lambda > 0$ sufficiently small.

Now suppose that $0 < \nu < 1/2$ and $\mu \geq 1/(1+2\nu)$. Then

$$f_\mu(\lambda, y^*) = \lambda^{1/\mu}\left(\|x_{\lambda,y^*}\|_X^2 - \frac{\|Tx_{\lambda,y^*} - y^*\|_Y^2}{\lambda^{1/\mu}}\right)$$
$$\geq \lambda^{1/\mu}\left(\|x_{\lambda,y^*}\|_X^2 - \frac{\|Tx_{\lambda,y^*} - y^*\|_Y^2}{\lambda^{1+2\nu}}\right).$$

Since $\|x_{\lambda,y^*}\|_X^2 \to \|x^*\|_X^2 > 0$ and $\lambda^{-(1+2\nu)}\|Tx_{\lambda,y^*} - y^*\|_Y^2 \to 0$ as $\lambda \to 0^+$ (by the previous lemma), it follows that $f_\mu(\lambda, y^*) > 0$ for all $\lambda > 0$ sufficiently small.

In either case, the proof follows as before. ∎

We can now prove the existence of solutions of (10) and (17). In the following theorem, $B_\epsilon(y^*)$ denotes the open ball of radius ϵ centered at y^*.

THEOREM 5.

1. *There exist $\epsilon > 0$ and $\lambda^* > 0$ such that for all $y \in B_\epsilon(y^*) \setminus \overline{\mathcal{R}(T)}$, there exists $\lambda \in (0, \lambda^*)$ such that $\lambda\|x_{\lambda,y}\|_Y^2 = \|Tx_{\lambda,y} - y\|_Y^2$.*
2. *Suppose $x^* \in \mathcal{R}((T^*T)^\nu)$ and*

$$0 < \nu < \frac{1}{2} \text{ and } \frac{1}{1+2\nu} \leq \mu < 1$$

or

$$\nu \geq \frac{1}{2} \text{ and } \frac{1}{2} < \mu < 1.$$

Then there exist $\epsilon > 0$ and $\lambda^* > 0$ such that for all $y \in B_\epsilon(y^*) \setminus \overline{\mathcal{R}(T)}$, there exists $\lambda \in (0, \lambda^*)$ such that $\lambda\|x_{\lambda,y}\|_Y^{2\mu} = \|Tx_{\lambda,y} - y\|_Y^{2\mu}$.

Proof.

1. Let $\overline{\lambda}$ be the value from the previous lemma; then we have

$$f_1(0, y^*) = f_1(\overline{\lambda}, y^*) = 0$$

and $f_1(\lambda, y^*) > 0$ for all $\lambda \in (0, \overline{\lambda})$. Since $f_1(\cdot, y^*)$ is continuous, it achieves its maximum on $[0, \overline{\lambda}]$; let $M = \max\{f_1(\lambda, y^*) : 0 \leq \lambda \leq \overline{\lambda}\}$ and define

$$\lambda^* = \sup\{\hat{\lambda} > 0 : f_1(\lambda, y^*) < M \text{ for all } \lambda \in (0, \hat{\lambda})\}.$$

Since f_1 is continuous, there exists $\epsilon_1 > 0$ such that for all $y \in B_{\epsilon_1}(y^*)$, $f_1(\lambda^*, y) \geq M/2$. Define

$$\epsilon = \min\left\{\epsilon_1, \sqrt{\frac{M}{2}}\right\}.$$

If $y \in B_\epsilon(y^*) \setminus \overline{\mathcal{R}(T)}$, then

$$f_1(\lambda^*, y) \geq \frac{M}{2} \geq \epsilon^2 > \|\hat{y}\|_Y^2 > 0.$$

Since $f_1(\lambda, y) \to 0$ as $\lambda \to 0^+$, it follows that

$$\Lambda = \sup\{\hat{\lambda} \in (0, \lambda^*) \,:\, f_1(\lambda, y) < \|\hat{y}\|_Y^2 \text{ for all } \lambda \in (0, \hat{\lambda})\}$$

is well defined and satisfies $\Lambda \in (0, \lambda^*)$, $f_1(\Lambda, y) = \|\hat{y}\|_Y^2$, and $f_1(\lambda, y) < \|\hat{y}\|_Y^2$ for all $\lambda \in (0, \Lambda)$. Since $\lambda\|x_{\lambda,y}\|_X^2 = \|Tx_{\lambda,y} - y\|_Y^2$ if and only if $f_1(\lambda, y) = \|\hat{y}\|_Y^2$, the proof is complete.

2. The proof is essentially the same as that of the first part. ∎

From now on, $\epsilon > 0$ and λ^* will have the values from the previous theorem and we always assume that $\mu \in (1/2, 1]$. We define $\Lambda_\mu : B_\epsilon(y^*) \setminus \overline{\mathcal{R}(T)} \to (0, \lambda^*)$ by the condition that $\lambda = \Lambda_\mu(y)$ is the smallest solution of (18) in $(0, \lambda^*)$. When $\mu = 1$, we will sometimes write $\Lambda(y)$ in place of $\Lambda_1(y)$.

Corollary 6. *Let* $\{y_n\} \subset B_\epsilon(y^*) \setminus \overline{\mathcal{R}(T)}$ *satisfy* $y_n \to y^*$ *as* $n \to \infty$. *Then* $\Lambda_\mu(y_n) \to 0$ *as* $n \to \infty$.

Proof. We will write $\lambda_n = \Lambda_\mu(y_n)$ for each $n \in \mathbb{Z}^+$. Since $\{\lambda_n\} \subset (0, \lambda^*)$, without loss of generality, there exists $\hat{\lambda} \in [0, \lambda^*]$ such that $\lambda_n \to \hat{\lambda}$. Since f_μ is continuous, it follows that

$$\lim_{n\to\infty} f_\mu(\lambda_n, y_n) = f_\mu(\hat{\lambda}, y^*).$$

But $f_\mu(\lambda_n, y_n) = \|\hat{y}_n\|_Y^2 \to 0$ as $n \to \infty$ and therefore $f_\mu(\hat{\lambda}, y^*) = 0$. Since the only value $\lambda \in [0, \lambda^*]$ such that $f_\mu(\lambda, y^*) = 0$ is $\lambda = 0$, it follows that $\hat{\lambda} = 0$ and $\Lambda_\mu(y_n) \to 0$ as $n \to \infty$. ∎

We now define

$$S = S_{y^*,\epsilon,s} = \{y \in B_\epsilon(y^*) \,:\, \|\hat{y}\|_Y \geq s\|y - y^*\|_Y\},$$

where $s \in (0, 1)$ is a constant. Notice that if $y \in S$, then the angle θ between y and $\overline{\mathcal{R}(T)}$ satisfies $\sin(\theta) \geq s$. We can now prove that (10) and (17) define convergent parameter choice rules if y approaches y^* from within S.

Theorem 7.

1. *Let* s *satisfy* $1/2 < s < 1$ *and consider* $S = S_{y^*,\epsilon,s}$. *If* $\{y_n\} \subset S$ *and* $y_n \to y^*$, *then*

$$\Lambda(y_n) \sim \|y_n - y^*\|_Y^2 \text{ as } n \to \infty.$$

Moreover,

$$x_{\Lambda(y_n),y_n} \to x^* \text{ weakly as } n \to \infty.$$

2. *Suppose* $x^* \in \mathcal{R}((T^*T)^\nu)$ *and*

$$0 < \nu < \frac{1}{2} \text{ and } \frac{1}{1+2\nu} \leq \mu < 1$$

or

$$\nu \geq \frac{1}{2} \text{ and } \frac{1}{2} < \mu < 1.$$

For any $s > 0$,

$$\{y_n\} \subset S_{y^*,\epsilon,s}, \quad y_n \to y^* \;\Rightarrow\; \Lambda_\mu(y_n) \sim \|y_n - y^*\|_Y^{2\mu} \text{ as } n \to \infty.$$

Moreover,

$$\{y_n\} \subset S_{y^*,\epsilon,s}, \quad y_n \to y^* \;\Rightarrow\; x_{\Lambda_\mu(y_n),y_n} \to x^* \text{ as } n \to \infty,$$

where now the convergence is in the norm topology.

Proof.

1. We must show that there exist constants $c_1 > 0$ and $c_2 > 0$ such that

$$c_1 \leq \frac{\Lambda(y_n)}{\|y_n - y^*\|_Y^2} \leq c_2 \text{ for all } n \in \mathbb{Z}^+.$$

To prove that such a constant c_1 exists, we argue by contradiction and assume (without loss of generality) that

$$\frac{\Lambda(y_n)}{\|y_n - y^*\|_Y^2} \to 0 \text{ as } n \to \infty.$$

For any $y \in Y$, we have $x_{\lambda,y} = x_{\lambda,y} - x_{\lambda,y^*} + x_{\lambda,y^*} = N_\lambda^{-1}(y - y^*) + x_{\lambda,y^*}$ and therefore, applying (11) and (16),

$$\begin{aligned}\|x_{\lambda,y}\|_X &\leq \|N_\lambda^{-1}T^*(y - y^*)\|_X + \|x_{\lambda,y^*}\|_X \\ &\leq \frac{1}{2\sqrt{\lambda}}\|y - y^*\|_Y + \|x^*\|_X,\end{aligned}$$

which implies

$$\lambda\|x_{\lambda,y}\|_X^2 \leq \frac{1}{4}\|y - y^*\|_Y^2 + \sqrt{\lambda}\|x^*\|_X\|y - y^*\|_Y + \lambda\|x^*\|_X^2.$$

We apply this inequality to $y = y_n$ with $\lambda = \lambda_n = \Lambda(y_n)$. Then, since

$$\Lambda(y_n) = o(\|y_n - y^*\|_Y^2),$$

we obtain

$$\sqrt{\lambda_n}\|x^*\|_X\|y_n - y^*\|_Y = o(\|y_n - y^*\|_Y^2) \text{ and } \lambda_n\|x^*\|_X^2 = o(\|y_n - y^*\|_Y^2).$$

It follows that for any $\delta > 0$,

$$\lambda\|x_{\lambda_n,y_n}\|_X^2 \leq \left(\frac{1}{4} + \delta\right)\|y_n - y^*\|_Y^2 \text{ for all } n \text{ sufficiently large.}$$

On the other hand, we have

$$\|Tx_{\lambda_n,y_n} - y_n\|_Y^2 = \|Tx_{\lambda_n,y_n} - \overline{y}_n\|_Y^2 + \|\hat{y}_n\|_Y^2 \geq \|\hat{y}_n\|_Y^2.$$

Therefore,

$$\lambda_n\|x_{\lambda_n,y_n}\|_X^2 = \|Tx_{\lambda_n,y_n} - y_n\|_Y^2 \geq \|\hat{y}_n\|_Y^2 \geq s^2\|y_n - y^*\|_Y^2$$

and we obtain

$$s^2\|y_n - y^*\|_Y^2 \leq \lambda_n\|x_{\lambda_n,y_n}\|_X^2 \leq \left(\frac{1}{4} + \delta\right)\|y_n - y^*\|_Y^2 \text{ for all } n \text{ sufficiently large.} \Rightarrow\ s^2 \leq \frac{1}{4} + \delta.$$

Moreover, this must hold for all $\delta > 0$. Since $s > 1/2$ by assumption, this is a contradiction. It follows that there must exist $c_1 > 0$ such that

$$\frac{\Lambda(y_n)}{\|y_n - y^*\|_Y^2} \geq c_1 \text{ for all } n \in \mathbb{Z}^+.$$

It now follows from the standard theory of Tikhonov regularization that $x_{\lambda_n,y_n} \to x^*$ weakly. To be specific,

$$y_n \to y^*, \ \lambda_n \to 0, \text{ and } \frac{\|y_n - y^*\|_Y^2}{\lambda_n} \le c_1^{-1} \ \Rightarrow \ x_{\lambda_n,y_n} \to x^* \text{ weakly}$$

(see, for example, Theorem 3.20 of [8]).
Now we show that there exists $c_2 > 0$ such that

$$\frac{\Lambda(y_n)}{\|y_n - y^*\|_Y^2} \le c_2 \text{ for all } n \in \mathbb{Z}^+.$$

Still writing $\lambda_n = \Lambda(y_n)$, $x_{\lambda_n,y_n} \to x^*$ weakly implies that

$$\begin{aligned}\|x^*\|_X &\le \liminf_{n\to\infty} \|x_{\lambda_n,y_n}\|_X \\ &\Rightarrow \liminf_{n\to\infty}(2\|x_{\lambda_n,y_n}\|_X^2 - \|x^*\|_X^2) \ge \|x^*\|_X^2.\end{aligned}$$

It follows that there exists n_0 sufficiently large that

$$2\|x_{\lambda_n,y_n}\|_X^2 - \|x^*\|_X^2 \ge \frac{1}{2}\|x^*\|_X^2 \text{ for all } n \ge n_0.$$

Since x_{λ_n,y_n} is the minimizer of $\|Tx - y_n\|_Y^2 + \lambda_n\|x\|_X^2$,

$$\begin{aligned}\|Tx_{\lambda_n,y_n} - y_n\|_Y^2 + \lambda_n\|x_{\lambda_n,y_n}\|_X^2 &\le \|Tx^* - y_n\|_Y^2 + \lambda_n\|x^*\|_X^2 \\ &= \|y_n - y^*\|_Y^2 + \lambda_n\|x^*\|_X^2.\end{aligned}$$

Moreover, we have $\|Tx_{\lambda_n,y_n} - y_n\|_Y^2 = \lambda_n\|x_{\lambda_n,y_n}\|_X^2$. Therefore, for $n \ge n_0$,

$$\begin{aligned}2\lambda_n\|x_{\lambda_n,y_n}\|_X^2 &\le \|y_n - y^*\|_Y^2 + \lambda_n\|x^*\|_X^2 \\ &\Rightarrow (2\|x_{\lambda_n,y_n}\|_X^2 - \|x^*\|_X^2)\lambda_n \le \|y_n - y^*\|_Y^2 \\ &\Rightarrow \frac{1}{2}\|x^*\|_X^2\lambda_n \le \|y_n - y^*\|_Y^2 \text{ for all } n \ge n_0 \\ &\Rightarrow \lambda_n \le \frac{2}{\|x^*\|_X^2}\|y_n - y^*\|_Y^2 \text{ for all } n \ge n_0.\end{aligned}$$

It follows that there exists $c_2 > 0$ such that

$$\lambda_n \le c_2\|y_n - y^*\|_Y^2 \text{ for all } n \in \mathbb{Z}^+$$

and the proof is complete.

2. We must show that there exist constants $c_1 > 0$ and $c_2 > 0$ such that

$$c_1 \le \frac{\Lambda_\mu(y_n)}{\|y_n - y^*\|_Y^{2\mu}} \le c_2 \text{ for all } n \in \mathbb{Z}^+.$$

To prove that such a constant c_1 exists, we argue by contradiction and suppose that there exist $s > 0$ and $\{y_n\} \subset S = S_{y^*,\epsilon,s}$ such that

$$y_n \to y^* \text{ and } \lambda_n = o(\|y_n - y^*\|_Y^{2\mu}),$$

where $\lambda_n = \Lambda_\mu(y_n)$. We now argue much as in the proof of the first part of the theorem. We have

$$\begin{aligned}
&\lambda_n^{1/\mu}\|x_{\lambda_n,y_n}\|_X^2 \\
&\quad \leq \lambda_n^{1/\mu}\left(\|x_{\lambda_n,y_n-y^*}\|_X + \|x_{\lambda_n,y^*}\|_X\right)^2 \\
&\quad = \lambda_n^{1/\mu}\|x_{\lambda_n,y_n-y^*}\|_X^2 + 2\lambda_n^{1/\mu}\|x_{\lambda_n,y^*}\|_X\|x_{\lambda_n,y_n-y^*}\|_X + \lambda_n^{1/\mu}\|x_{\lambda_n,y^*}\|_X^2 \\
&\quad \leq \lambda_n^{1/\mu}\left(\frac{1}{2\sqrt{\lambda_n}}\|y_n-y^*\|_Y\right)^2 + 2\lambda_n^{1/\mu}\|x^*\|_X\left(\frac{1}{2\sqrt{\lambda_n}}\|y_n-y^*\|_Y\right) + \lambda_n^{1/\mu}\|x^*\|_X^2 \\
&\quad = \frac{\lambda_n^{1/\mu-1}}{4}\|y_n-y^*\|_Y^2 + \lambda^{1/\mu-1/2}\|x^*\|_X\|y_n-y^*\|_Y + \lambda_n^{1/\mu}\|x^*\|_X^2.
\end{aligned}$$

Since $\lambda_n = o(\|y_n - y^*\|_Y^{2\mu})$, it is easy to see that each of the three terms on the right is $o(\|y_n - y^*\|_Y^2)$. Therefore, there exists a sequence $\{\alpha_n\}$ of positive numbers such that

$$\alpha_n \to 0 \text{ and } \lambda_n^{1/\mu}\|x_{\lambda_n,y_n}\|_X^2 \leq \alpha_n\|y_n - y^*\|_Y^2 \text{ for all } n \in \mathbb{Z}^+.$$

On the other hand,

$$\begin{aligned}
\lambda_n^{1/\mu}\|x_{\lambda_n,y_n}\|_X^2 = \|Tx_{\lambda_n,y_n} - y_n\|^2 &= \|Tx_{\lambda_n,y_n} - \overline{y}_n\|^2 + \|\hat{y}_n\|_Y^2 \\
&\geq \|\hat{y}_n\|_Y^2 \\
&\geq s^2\|y_n - y^*\|_Y^2.
\end{aligned}$$

It follows that

$$s^2\|y_n - y^*\|_Y^2 \leq \lambda_n^{1/\mu}\|x_{\lambda_n,y_n}\|_X^2 \leq \alpha_n\|y_n - y^*\|_Y^2 \text{ for all } n \in \mathbb{Z}^+.$$

Since $\alpha_n \to 0$ and $s > 0$, this is impossible; thus we have obtained the desired contradiction. This shows that there exists $c_1 > 0$ such that

$$c_1 \leq \frac{\lambda_n}{\|y_n - y^*\|_Y^{2\mu}} \text{ for all } n \in \mathbb{Z}^+.$$

Because $\mu \in (0, 1)$, it follows immediately that

$$\frac{\|y_n - y^*\|_Y^2}{\lambda_n} = \frac{\|y_n - y^*\|_Y^{2\mu}}{\lambda_n}\|y_n-y^*\|_Y^{2-2\mu} \leq c_1^{-1}\|y_n-y^*\|_Y^{2-2\mu} \to 0 \text{ as } n \to \infty.$$

Therefore, by the standard theory of Tikhonov regularization (for example, Theorem 3.19 of [8]),

$$y_n \to y^*,\ \lambda_n \to 0, \text{ and } \frac{\|y_n - y^*\|_Y^2}{\lambda_n} \to 0 \ \Rightarrow\ \|x_{\lambda_n,y_n} - x^*\|_X \to 0.$$

Now we wish to show that there exists $c_2 > 0$ such that

$$\frac{\lambda_n}{\|y_n - y^*\|_Y^{2\mu}} \leq c_2 \text{ for all } n \in \mathbb{Z}^+.$$

We will argue by contradiction and assume, without loss of generality, that

$$\frac{\|y_n - y^*\|_Y^{2\mu}}{\lambda_n} \to 0 \text{ as } n \to \infty.$$

This is equivalent to

$$\frac{\|y_n - y^*\|_Y^2}{\lambda_n^{1/\mu}} \to 0 \text{ as } n \to \infty. \tag{19}$$

We have

$$\lambda_n^{1/\mu}\|x_{\lambda_n,y_n}\|_X^2 - \|Tx_{\lambda_n,y_n} - \overline{y}_n\|_Y^2 = \|\hat{y}_n\|_Y^2 \le \|y_n - y^*\|_Y^2,$$

which implies that

$$\frac{\|y_n - y^*\|_Y^2}{\lambda_n^{1/\mu}} \ge \|x_{\lambda_n,y_n}\|_X^2 - \frac{\|Tx_{\lambda_n,y_n} - \overline{y}_n\|_Y^2}{\lambda_n^{1/\mu}}.$$

Since $x_{\lambda_n,y_n} \to x^*$ as $n \to \infty$, we will obtain the desired contradiction if we can prove that

$$\frac{\|Tx_{\lambda_n,y_n} - \overline{y}_n\|_Y^2}{\lambda_n^{1/\mu}} \to 0 \text{ as } n \to \infty.$$

We have

$$\frac{\|Tx_{\lambda_n,y_n} - \overline{y}_n\|_Y^2}{\lambda_n^{1/\mu}} \le \frac{2\|Tx_{\lambda_n,y^*} - y^*\|_Y^2}{\lambda_n^{1/\mu}} + \frac{2\|Tx_{\lambda_n,y_n-y^*} - (\overline{y}_n - y^*)\|_Y^2}{\lambda_n^{1/\mu}}. \tag{20}$$

Next, we show that

$$\frac{\|Tx_{\lambda_n,y^*} - y^*\|_Y^2}{\lambda_n^{1/\mu}} \to 0 \text{ as } n \to \infty.$$

We must consider two cases. If $0 < \nu < 1/2$, then Lemma 3 yields

$$\|Tx_{\lambda_n,y^*} - y^*\|_Y^2 = o(\lambda_n^{1+2\nu}),$$

which, in turn, implies that

$$\frac{\|Tx_{\lambda_n,y^*} - y^*\|_Y^2}{\lambda_n^{1/\mu}} \le \frac{\|Tx_{\lambda_n,y^*} - y^*\|_Y^2}{\lambda_n^{1+2\nu}} \to 0 \text{ as } n \to \infty.$$

If $\nu \ge 1/2$ and $\mu > 1/2$, then Lemma 3 implies that

$$\|Tx_{\lambda_n,y^*} - y^*\|_Y^2 = O(\lambda_n^2).$$

Therefore,

$$\frac{\|Tx_{\lambda_n,y^*} - y^*\|_Y^2}{\lambda_n^{1/\mu}} = O(\lambda_n^{2-\mu^{-1}}) \to 0 \text{ as } n \to \infty,$$

and the result holds in this case also.

We also have

$$\|Tx_{\lambda_n,y_n-y^*} - (\overline{y}_n - y^*)\|_Y^2 = \|(TN_\lambda^{-1}T^* - I)(\overline{y}_n - y^*)\|_Y^2 \le \|\overline{y}_n - y^*\|_Y^2$$

since $\|TN_\lambda^{-1}T^* - I\| \le 1$. It then follows from (19) that

$$\frac{\|Tx_{\lambda_n,y_n-y^*} - (\overline{y}_n - y^*)\|_Y^2}{\lambda_n^{1/\mu}} \le \frac{\|\overline{y}_n - y^*\|_Y^2}{\lambda_n^{1/\mu}} \le \frac{\|y_n - y^*\|_Y^2}{\lambda_n^{1/\mu}} \to 0 \text{ as } n \to \infty.$$

Therefore, by (20), we have

$$\frac{\|Tx_{\lambda_n,y_n} - \overline{y}_n\|_Y^2}{\lambda_n^{1/\mu}} \to 0 \text{ as } n \to \infty,$$

which completes the proof by contradiction. ∎

We now derive the rate of convergence for the MR rule. We note that, for x^* belonging to $\mathcal{R}((T^*T)^\nu)$, $0 < \nu < 1$, the optimal worst-case rate of convergence of $x_{\lambda,y}$ to x^* is

$$\|x_{\lambda,y} - x^*\|_X = o\left(\|y - y^*\|_Y^{2\nu/(2\nu+1)}\right),$$

and it is $\|x_{\lambda,y} - x^*\|_X = O(\|y - y^*\|_Y^{2/3})$ if $\nu = 1$ (see, for instance, [6] or [8]).

For $x^* \in \mathcal{R}((T^*T)^\nu)$, Theorem 7 shows that the MR rule is guaranteed to be convergent if μ satisfies

$$\frac{1}{1+2\nu} \le \mu < 1 \text{ if } 0 < \nu < \frac{1}{2}$$

or

$$\frac{1}{2} < \mu < 1 \text{ if } \nu \ge \frac{1}{2}.$$

The following theorem derives the corresponding rate of convergence.

Theorem 8.

1. *If $x^* \in \mathcal{R}((T^*T)^\nu)$ for some $\nu \in (0, 1/2)$ and $\mu = 1/(1+2\nu)$, then, for any $s > 0$, $\{y_n\} \subset S_{y^*,\epsilon,s}$ and $y_n \to y^*$ imply*

   $$\|x_{\Lambda_\mu(y_n),y_n} - x^*\|_X = O\left(\|y_n - y^*\|_Y^{2\nu/(2\nu+1)}\right) \text{ as } n \to \infty.$$

   Moreover, this value of μ gives the optimal worst-case rate of convergence for $x^* \in \mathcal{R}((T^*T)^\nu)$.
2. *If $x^* \in \mathcal{R}((T^*T)^\nu)$ for some $\nu \ge 1/2$, $\epsilon \in (0, 1/2)$, and $\mu = 1/2 + \epsilon$, then, for any $s > 0$, $\{y_n\} \subset S_{y^*,\epsilon,s}$ and $y_n \to y^*$*

   $$\|x_{\Lambda_\mu(y_n),y_n} - x^*\|_X = O\left(\|y_n - y^*\|_Y^{1/2-\epsilon}\right) \text{ as } n \to \infty.$$

Proof. Given μ satisfying the requirements of Theorem 7, we have

$$\lambda_n = \Lambda_\mu(y_n) \sim \|y_n - y^*\|_Y^{2\mu}.$$

Therefore,

$$\begin{aligned}\|x_{\lambda_n,y_n} - x^*\|_X &\le \|x_{\lambda_n,y_n} - x_{\lambda_n,y^*}\|_X + \|x_{\lambda_n,y^*} - x^*\|_X \\ &= O\left(\frac{\|y_n - y^*\|_Y}{\sqrt{\lambda_n}}\right) + O\left(\lambda_n^\nu\right) \\ &= O\left(\|y_n - y^*\|_Y^{1-\mu}\right) + O\left(\|y_n - y^*\|_Y^{2\nu\mu}\right) \\ &= O\left(\|y_n - y^*\|_Y^{\min\{1-\mu,2\nu\mu\}}\right).\end{aligned}$$

Suppose now that $0 < \nu < 1/2$. It is easy to show that the solution of

$$\begin{aligned}&\max\ \min\{1-\mu, 2\nu\mu\} \\ &\text{s.t. } \mu \ge \frac{1}{1+2\nu}\end{aligned}$$

is $\mu = 1/(1+2\nu)$, which yields $\min\{1-\mu, 2\nu\mu\} = 2\nu/(2\nu+1)$ and hence

$$\|x_{\lambda_n,y_n} - x^*\|_X = O\left(\|y_n - y^*\|_Y^{2\nu/(2\nu+1)}\right).$$

This proves the first result. On the other hand, if $\nu \geq 1/2$ and $\mu = 1/2 + \epsilon$, then $\min\{1-\mu, 2\nu\mu\} = 1/2 - \epsilon$, and the second result follows. ∎

Theorem 8 suggests that the rate of convergence achieved by the MR rule can be arbitrarily close to $O(\|y_n - y^*\|_Y^{1/2})$, but that this latter rate cannot be obtained. The MR rule yields the optimal worst-case rate of convergence only for $0 < \nu < 1/2$; for $\nu \geq 1/2$, the rate of convergence is suboptimal.

Hanke and Raus also use the hypothesis $\{y_n\} \subset S_{y^*,\epsilon,s}$ to analyze their heuristic parameter choice rule. In Theorem 3.1, Corollary 3.2, and Theorem 3.3 of [12], they showed that $\{y_n\} \subset S_{y^*,\epsilon,s}$ and $y_n \to y^*$ imply that $x_{\lambda,y} \to x^*$ (in norm) when λ is chosen by the Hanke–Raus rule. Moreover, they showed that for all $\nu \in (0,1]$ there exists a constant $c > 0$ (depending on x^*) such that

$$\|x_{\lambda,y} - x^*\|_X \leq \frac{c}{s}\delta_*^{2\nu/(2\nu+1)}. \tag{21}$$

Here $\delta_* = \max\{\eta_\lambda, \|y - y^*\|_Y\}$, where

$$\eta_\lambda = \left[\left\langle y, (TT^* + \lambda I)^{-3} y\right\rangle_Y\right]^{1/2}.$$

If it were possible to prove that $\eta_\lambda \sim \|y - y^*\|_Y$, this result would yield the optimal worst-case rate of convergence for all $\nu \in (0,1]$. However, this has not been proved. (We recall that there is another analysis of the Hanke–Raus rule, due to Kindermann [18] and discussed in section 1, that uses different assumptions about the allowable data. We also note that Hämarik, Palm, and Raus [10] proved an estimate of the form (21), valid for a family of minimization-based parameter choice rule that include the quasi-optimality approach. As in the Hanke–Raus estimate, δ_* includes a quantity that must be of order δ to obtain the optimal rate of convergence. However, the analysis in [10] does not rely on the assumption that $\{y_n\} \subset S_{y^*,\epsilon,s}$.)

In spite of this theoretical foundation, the Hanke–Raus rule did not perform as well in our experiments as most of the other parameter choice methods considered in this paper; in particular, it performed poorly compared to the Reginska and MR rules.

The condition that the noisy data satisfy $\|\hat{y}\|_Y \geq s\|y - y^*\|_Y$ for some $s > 1/2$ is sufficient to guarantee weak convergence, in rule (10), as $y \to y^*$. Although we cannot prove that this condition is necessary, it is not difficult to show that $\|x_{\Lambda(y_n),y_n}\|_X$ can grow without bound if $\|\hat{y}_n\|_Y$ is too small compared to $\|y_n - y^*\|_Y$. We will demonstrate this by example. We choose any sequence $\{\overline{y}_n\} \subset \mathcal{R}(T)$ such that

$$\overline{y}_n \to y^* \text{ and } \|T^\dagger \overline{y}_n\|_X \to \infty \text{ as } n \to \infty$$

and consider data of the form $\overline{y}_n + z$, where $z \in \mathcal{R}(T)^\perp$ remains to be chosen. (We are assuming that $\mathcal{R}(T)$ is not closed, so that $T^\dagger$ is unbounded and it is possible to choose $\{\overline{y}_n\}$ to satisfy these conditions.) For each fixed value of n, we have

$$\Lambda(\overline{y}_n + z) \to 0 \text{ as } z \to 0$$

and hence

$$x_{\Lambda(\overline{y}_n+z),\overline{y}_n} \to T^\dagger \overline{y}_n \text{ as } z \to 0.$$

Now choose z sufficiently small that

$$\|x_{\Lambda(\overline{y}_n+z),\overline{y}_n}\|_X \geq \frac{1}{2}\|T^\dagger \overline{y}_n\|_X$$

and define y_n by $y_n = \overline{y}_n + z$ for that value of z. Since $x_{\Lambda(y_n),y_n} = x_{\Lambda(y_n),\overline{y}_n}$ for each n, it follows that

$$\|x_{\Lambda(y_n),y_n}\|_X \to \infty \text{ as } n \to \infty,$$

as desired.

The fact that $\|y-y^*\|_Y^2/\Lambda(y)$ is only bounded (when y is restricted to lie in the set S) as $y \to y^*$ suggests that (10) will produce under-regularized solutions when the error $\|y-y^*\|_Y$ is small. We will perform numerical experiments to determine how it performs, in practice, on discretized inverse problems.

Before we proceed to the numerical experiments, we must deal with the fact that (10) and (17) do not define parameter values when $\|y-\overline{y}\|_Y$ is too large. In this case, we wish to choose a value of λ that comes as close as possible to satisfying the fixed point equation (10) or (17). Numerical experience suggests that we do this by minimizing

$$\log\left(\frac{\|Tx_{\lambda,y}-y\|_Y^{2\mu}}{\|x_{\lambda,y}\|_X^{2\mu}}\right) - \log(\lambda).$$

Specifically, we take the smallest local minimizer of this function as our definition of λ when (10) or (17) has no solution.

One other issue that must be addressed: If the noisy data vector y happens to lie in $\overline{\mathcal{R}(T)}$, then the parameter choice rules defined here do not apply—the smallest solution of either (10) or (17) is $\lambda = 0$. If we discretize $Tx = y$ to obtain a matrix-vector equation $Ax = y$, then this problem arises when $y \in \text{col}(A)$ (the column space of A). If $A \in \mathbb{R}^{m\times n}$, $\text{rank}(A) < m$, and the noise in y is random, then the probability that y lies in $\text{col}(A)$ is nearly zero and the difficulty does not arise. However, if $Tx = y$ is only mildly ill-posed, then the corresponding matrix A might have full column rank. In such a case (which we encounter in some of our numerical experiments), a discretization for which $m = n$ is problematic. Therefore, in our experiments, we ensure that A has more rows than columns; then $\text{col}(A)$ is a proper subspace of $\mathbb{R}^m$ and a vector y that contains random noise is unlikely to lie in $\text{col}(A)$. (To be clear, this was done only when necessary.) This is equivalent to choosing a discretization of the solution space X that is not too fine compared to the amount of data that can be collected.

If $\overline{\mathcal{R}(T)} = Y$, then Theorem 7 does not apply; hence the theorem is not applicable to every problem. Nevertheless, by discretizing $Tx = y$ in such a way as to obtain $Ax = y$, $A \in \mathbb{R}^{m\times n}$ with $m > n$, the method appears to work even in such cases.

3. Numerical experiments.

To test the performance of Reginska's rule and the modified Reginska rule, we applied it to 20 test problems, all of which are discretizations of first-kind integral equations. Sixteen are one-dimensional problems, 11 chosen from Hansen's suite of test problems [14] (we omitted the problem **parallax**, for which no exact solution is available, and the two-dimensional image-reconstruction problems **blur** and **tomo**), and 5 from various research papers found in the literature. The other 4 problems are two-dimensional integral equations defined by various kernels.

Each (discretized) problem is of the form $Ax = y$, where the exact data vector y^* and solution x^* are known and the dimensions m, n of the problem ($A \in \mathbb{R}^{m\times n}$, $y \in \mathbb{R}^m$) can be chosen. For each test problem, we chose values of m, n and then generated noisy vectors y for $n_v = 8$ different relative noise levels

$$\delta = \frac{\|y - y^*\|}{\|y^*\|},$$

namely, $\delta = 2 \cdot 10^{-1}, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}$ (using the Euclidean norm). The components of the noise vector $y - y^*$ were uniformly distributed pseudorandom numbers selected from an interval centered at zero and generated by MATLAB (using the default random number generator in version 9.0). For each noise level, we generate $n_e = 10$ instances of the test problem. We regard each noise level for each test problem as one experiment and ranked the parameter choice methods as described below.

We compared the solutions produced by the Reginska and MR rules with those produced by the L-curve, GCV, quasi-optimality, and Hanke–Raus criteria. The L-curve and GCV solutions were computed using Hansen's Regularization Toolbox [14] (with one slight modification described below) and the others by Matlab code implemented in a similar manner. In particular, values of λ were sought in the interval $[16\epsilon\sigma_1, \sigma_1]$, where σ_1 is the largest singular value of A and ϵ is machine epsilon. For the quasi-optimality and Hanke–Raus criteria, the global minimizer of the corresponding objective function was chosen if it lay in the interior of the interval $[16\epsilon\sigma_1, \sigma_1]$. Otherwise, the value of λ corresponding to the smallest (interior) local minimum was chosen. For the L-curve and GCV methods, Hansen's algorithms were used, except that the search interval was always chosen to be $[16\epsilon\sigma_1, \sigma_1]$. (In some cases, his code would choose a more restricted interval, but our experiments showed that the larger interval gave better overall results, at least for these test problems.)

Our experiments suggest that the performance of the MR rule is relatively insensitive to the value of μ in the interval $[0.85, 0.95]$. For the numerical results presented here, we used $\mu = 0.93$.

Since each experiment consists of ten trials, we have to rank the methods somehow. We used three measures:

- the mean error ratio;
- the median error ratio;
- a Borda-type count.

The error ratio is defined to be the error in $x_{\lambda,y}$ divided by the error in the optimal Tikhonov solution $x_{\lambda^*,y}$:

$$\frac{\|x_{\lambda,y} - x^*\|_{L^2}}{\|x_{\lambda^*,y} - x^*\|_{L^2}}. \tag{22}$$

Note that we can compute $x_{\lambda^*,y}$ because x^* is known in our test problems. The Borda count was computed as follows: the methods were ranked on each trial from first to last (with ties allowed). A method that was, for example, first three times, second six times, and third one time would have a Borda count of $3 \cdot 1 + 6 \cdot 2 + 1 \cdot 3 = 18$. For each criterion (including the Borda count), we define "better" to mean at least 10% better; when the difference is less than 10%, we regard the performance of the two methods as essentially the same. One method was regarded as better than another if it was better on at least two of the three measures defined above.

The one-dimensional test problems are **baart**, **baker3**, **deriv2**, **foxgood**, **gravity**, **groetsch2.3**, **groetsch2.5**, **heat**, **ilaplace**, **indramm**, **phillips**, **shaw**, **spikes**,

ursell, **wazwaz2**, and **wing**, and the two-dimensional problems are **WangXiao**, **Gaussian2Dver1**, **Gaussian2Dver2**, and **LogisticKernel2D**. The appendix contains a description or a reference for each test problem.

Overall results. We performed 160 experiments (20 test problems with 8 noise levels for each problem). Table 1 shows how many times each parameter choice ranked first, second, and so forth. For these test problems, the MR rule performed the best, though not by a wide margin, followed by the quasi-optimality and L-curve criteria. Reginska's rule was somewhat less effective, while the Hanke–Raus and GCV rules were much less effective.

Table 2 shows the mean and median error ratio and the Borda count for each method over all 1600 trials. These data also suggest that the MR rule performs the best on these test problems, although the quasi-optimality rule has a slightly smaller median error ratio and also a slightly smaller Borda count. The L-curve and Reginska rules are next best, with the Hanke–Raus and GCV rules appearing least effective.

Comparison of the mean and median error ratios suggests that some of the methods fail badly for certain trials. Table 3 shows the number of times each method produced an error ratio greater than R for $R = 10, 100, 1000$. The results show that the MR is the most robust method, followed by the L-curve, Reginska, and Hanke–Raus rules. The quasi-optimality rule is noticeably less robust, and the GCV approach is the least robust of all.

Finally, Tables 4 and 5 show the results for all methods, organized by noise level. For larger errors, the L-curve and quasi-optimality criteria define the best

Table 1

The number of times each parameter choice rule achieved each rank in the 160 *experiments. Ties were allowed. (Thus each row sums to* 160*, but the columns need not sum to* 160*.)*

Method	1st	2nd	3rd	4th	5th	6th
MR	92	20	23	23	2	0
Quasi-optimality	87	16	13	32	5	7
L-curve	84	24	12	22	17	1
Reginska	79	22	18	17	11	13
Hanke–Raus	55	9	10	11	42	33
GCV	30	9	10	19	41	51

Table 2

The mean and median error ratios and total Borda count for all methods over all trials.

Method	Mean error ratio	Median error ratio	Borda count
MR	1.743	1.181	2439
Quasi-optimality	5.052	1.123	2334
L-curve	1.995	1.212	2652
Reginska	2.152	1.229	2875
Hanke–Raus	3.066	1.411	3377
GCV	489.3	1.447	3707

Table 3

The number of times each method produced an error ratio greater than R.

Method	$R = 10$	$R = 100$	$R = 1000$
GCV	128	27	15
Quasi-optimality	73	42	0
Hanke–Raus	58	0	0
Reginska	43	0	0
L-curve	30	0	0
MR	7	0	0

Table 4
The results for all trials, organized by noise level δ (larger values of δ).

δ	Method	Mean error ratio	Median error ratio	Borda count
$2 \cdot 10^{-1}$	L-curve	1.434	1.140	269
	Quasi-optimality	1.391	1.144	294
	Reginska	1.570	1.207	322
	MR	1.657	1.243	374
	GCV	2422	1.313	422
	Hanke–Raus	2.462	1.845	510
10^{-1}	Quasi-optimality	1.512	1.148	297
	L-curve	1.686	1.136	285
	Reginska	1.648	1.154	289
	MR	1.840	1.228	351
	GCV	946.9	1.415	453
	Hanke–Raus	2.841	1.934	535

Table 5
The results for all trials, organized by noise level δ (smaller values of δ).

δ	Method	Mean error ratio	Median error ratio	Borda count
10^{-2}	MR	1.532	1.074	266
	L-curve	1.527	1.166	288
	Quasi-optimality	2.459	1.142	289
	Reginska	1.611	1.152	303
	GCV	435.0	1.669	452
	Hanke–Raus	4.968	1.822	504
10^{-3}	Quasi-optimality	7.371	1.142	282
	MR	2.007	1.149	302
	L-curve	2.262	1.236	359
	Reginska	2.360	1.253	392
	Hanke–Raus	6.828	1.405	436
	GCV	45.99	1.589	502
10^{-4}	MR	1.888	1.236	299
	Quasi-optimality	7.214	1.123	309
	L-curve	2.336	1.236	380
	Hanke–Raus	2.080	1.363	373
	Reginska	2.845	1.250	414
	GCV	50.87	1.646	507
10^{-5}	MR	1.986	1.103	297
	Hanke–Raus	1.777	1.236	355
	Quasi-optimality	6.828	1.125	307
	L-curve	2.433	1.236	370
	Reginska	2.514	1.242	403
	GCV	5.252	1.410	472
10^{-6}	MR	1.561	1.100	281
	Quasi-optimality	6.923	1.095	290
	Hanke–Raus	1.886	1.236	345
	L-curve	2.406	1.236	343
	Reginska	2.522	1.236	361
	GCV	4.547	1.411	450
10^{-7}	MR	1.475	1.051	269
	Quasi-optimality	6.722	1.064	266
	Hanke–Raus	1.688	1.156	319
	L-curve	1.879	1.230	358
	Reginska	2.148	1.236	391
	GCV	3.850	1.405	449

rules, at least on these test problems. Since the MR rule was devised to ensure strong convergence as the noise level δ goes to zero, it is not surprising that it is the best method for smaller values of δ. Nevertheless, it performs reasonably well for larger values of δ (and better than the Hanke–Raus and GCV rules). As expected from the fact that $\Lambda(y) \sim \|y - y^*\|_Y^2$ as $y \to y^*$, Reginska's rule is less effective for small values of δ (although at no noise level is it the worst method).

4. Discussion. The method of multiplicative regularization suggests a method for choosing the regularization parameter in Tikhonov regularization. We have presented an analysis that shows that this method, which we call Reginska's rule because it is equivalent to the approach in [27], is well defined provided the data y is sufficiently close to exact data $y^* \in \mathcal{R}(T)$. The analysis shows that the regularization parameter converges to zero like $O(\delta^2)$, where δ is the noise level in the data, provided the component of y that is orthogonal to $\overline{\mathcal{R}(T)}$ is sufficiently large compared to $\|y - y^*\|_Y$. This shows that $x_{\lambda,y}$ is guaranteed (under the given conditions on the noise) to converge weakly to x^*, but it also suggests that the method will produce under-regularized estimates.

On a collection of 20 test problems, the Reginska's rule performed reasonably well when compared to four popular parameter choice rules. However, unsurprisingly, its relative performance deteriorated for smaller noise levels.

When it performs poorly, Reginska's rule tends to choose regularization parameters that are too small. For this reason, a modification is proposed: In place of

$$\lambda = \frac{\|Tx_{\lambda,y} - y\|_Y^2}{\|x_{\lambda,y}\|_X^2},$$

we can define λ by the fixed point equation

$$\lambda = \frac{\|Tx_{\lambda,y} - y\|_Y^{2\mu}}{\|x_{\lambda,y}\|_X^{2\mu}},$$

where $1/2 < \mu < 1$. A value of μ in the range $0.85 \le \mu \le 0.95$ seems to work well and results in a method that outperformed Reginska's rule, the Hanke–Raus rule, and the GCV approach, and slightly outperformed the quasi-optimality and L-curve criteria.

We wish to emphasize that the numerical experiments included in this study are not extensive enough to determine which heuristic parameter choice rule is most effective. With only 20 problems in the test set, the outcome can be changed by making small changes in the collection. For instance, the quasi-optimality method performed especially poorly on problem **groetsch2.5**, whereas the MR rule did poorly on the problems **gravity** and **phillips**. By including or omitting a few problems, we could produce a set of test problems that favors either of the two rules.

Nevertheless, the MR rule seems to be at least competitive with other popular heuristic parameter choice rules. Moreover, this method has a strong theoretical foundation: as long as y does not follow a path tangent to $\overline{\mathcal{R}(T)}$ in converging to y^*, $x_{\lambda,y}$ is guaranteed to converge (in norm) to x^*, assuming $x^* \in \mathcal{R}((T^*T)^\nu$ for $\nu > 0$ sufficiently large. The worst-case rate of convergence is optimal provided $0 < \nu < 1/2$ and $\mu = 1/(1 + 2\nu)$; for $\nu \ge 1/2$, the rate of convergence is suboptimal.

A more extensive numerical comparison of parameter choice rules, including the Reginska, quasi-optimality, and GCV rules, was given in [4]. (Note that the authors referred to Reginska's rule as the L-curve criterion, since it can be viewed as one approach to choosing the corner of the L-curve.) In that study, the quasi-optimality approach was the best, GCV was noticeably less effective, and Reginska's rule was much less effective than GCV. However, it is difficult to compare their results with ours because their test problems were so different from ours. The authors considered randomly generated ill-conditioned systems $Ax = b$ with the singular values of $A \in \mathbb{R}^{n\times n}$ of the form $\sigma_k \approx k^{-\mu}$. The chosen values of n and μ were such that all the test matrices A were of full numerical rank and most had a condition number of less than 10^8. By contrast, of the 20 test matrices considered in this paper, 15 were numerically singular and most of these had a large null space. As noted above, it is expected that Reginska's rule will perform poorly on a square system based on a full-rank matrix.

Hämarik, Palm, and Raus [10] studied the performance of numerous parameter-choice rules, including the quasi-optimality and Hanke–Raus rules, on ten problems from Hansen's test set. With respect to the quasi-optimality rule, their results are generally consistent with ours, but they found the Hanke–Raus rule to perform much worse than we did.

Appendix: Test problems. The problems **baart**, **deriv2**, **foxgood**, **gravity**, **heat**, **ilaplace**, **phillips**, **shaw**, **spikes**, **ursell**, and **wing** are taken from Hansen's collection [14] of test problems. Each problem was discretized so that the approximate solution lay in $\mathbb{R}^{100}$ and, for all but two, so that the data also lay in $\mathbb{R}^{100}$. The problems **deriv2**, **heat**, and **phillips** were discretized to produce matrices $A \in \mathbb{R}^{200\times 100}$ because the matrices are full rank, or nearly so. The problem **ursell** is defined by the same operator as in [14], but a square-integrable solution (namely, $f(t) = t(1-t)$) was chosen.

baker3 [2]. Discretization of the integral equation

$$\int_0^1 e^{st} f(t)\,dt = \frac{e^{s+1}-1}{s+1}, \quad 0 < s < 1.$$

The exact solution is $f(t) = e^t$. The discretization is performed by the midpoint rule on a uniform mesh of 100 elements.

groetsch2.3 [9]. Discretization of the integral equation

$$\int_0^{100} \frac{se^{-s^2/(4t)}}{2\sqrt{\pi}t^{3/2}} f(t)\,dt = g(s), \quad 0 < s < 100.$$

The exact solution is

$$f(t) = 40 + 5\cos((100-t)/5) + 2.5\cos(2(100-t)/2.5) + 1.25\cos(4(100-t)/2).$$

The operator is discretized by the midpoint rule on a mesh with 200 elements and the exact data is generated by applying the discretized operator to the exact solution.

groetsch2.5 [9]. Discretization of the integral equation

$$\int_0^{\pi} k(s,t)f(t)\,dt = g(s), \quad 0 < s < \pi,$$

where the kernel k is defined by

$$k(s,t) = -\frac{2}{\pi}\sum_{n=1}^{\infty}\frac{\sin(ns)\sin(nt)}{n}.$$

For our computations, we approximate k by the first 100 terms of this series, and discretize the integral equation using the midpoint rule on a uniform mesh of 100 elements. The exact solution is $f(t) = t(\pi - t)$ and the exact data is generated by applying the discretized operator to the exact solution.

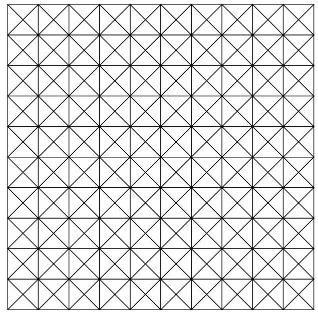

FIG. 1. *The mesh for problems* **WangXiao**, **Gaussian2Dver1**, **Gaussian2Dver2**, *and* **LogisticKernel2D**.

indramm [17]. Discretization of the integral equation

$$\int_0^1 e^{-st} f(t)\,dt = \frac{1-(s+1)e^{-s}}{s^2}, \quad 0 < s < 1.$$

The exact solution is $f(t) = t$. The equation is discretized by the midpoint rule on a uniform mesh with 100 elements.

wazwaz2 [34]. Discretization of the integral equation

$$\int_0^\pi \cos(s-t) f(t)\,dt = \frac{\pi}{2}\cos(s), \quad 0 < s < \pi.$$

The exact solution is $f(t) = \cos(t)$. The equation is discretized by the midpoint rule on a uniform mesh with 100 elements.

The 4 remaining test problems are all first-kind integral equations defined on the unit square $(0,1)\times(0,1)$; that is, each asks for an estimate of f in the equation

$$\int_0^1 \int_0^1 k(x,y,s,t) f(s,t)\,ds\,dt = g(x,y),\ (x,y) \in (0,1)\times(0,1),$$

from a measurement of the right-hand side g. The kernel k and the data g differ for each problem. Each was discretized by projecting the kernel onto the tensor-product finite element space defined by continuous piecewise linear function on the mesh shown in Figure 1. This mesh has 400 triangular elements and 221 nodes, which means that the discretized operator is represented by a 221×221 matrix.

WangXiao [33]. The kernel is $k(x,y,s,t) = e^{-80[(x-s-0.5)^2+(y-t-0.5)^2]}$ and the exact solution is

$$f(s,t) = \left(\frac{e^{-(s-0.3)^2/0.03} + e^{-(s-0.7)^2/0.03}}{0.9550408} - 0.052130913\right) e^{-(t-0.3)^2/0.03}.$$

Gaussian2Dver1. The kernel is $K(x,y,s,t) = k_0(x-s, y-t, 0.15)$, where

$$k_0(s,t,\sigma) = \frac{1}{2\pi\sigma^2} e^{-(s^2+t^2)/(2\sigma^2)}.$$

The exact solution is $f(s,t) = k_0(s-0.25, t-0.5, 0.1)$.

Gaussian2Dver2. The kernel is the same as in the previous problem, except with $\sigma = 0.3$. The solution is $f(s,t) = k_0(s-0.65, t-0.35, 0.15)$.

LogisticKernel2D. The kernel is

$$k(x, y, s, t) = \frac{2e^{-(x-s)}}{(1 + e^{-(x-s)} + e^{-(y-t)})^3},$$

and the exact solution is $f(s, t) = k_0(s - 0.75, t - 0.8, 0.1) + k_0(s - 0.2, t - 0.6, 0.15)$.

REFERENCES

[1] A. Abubakar and P. M. van den Berg, *Total variation as a multiplicative constraint for solving inverse problems*, IEEE Trans. Image Process., 10 (2001), pp. 1384–1392.

[2] C. T. H. Baker, L. Fox, D. F. Mayers, and K. Wright, *Numerical solution of Fredholm integral equations of the first kind*, Comput. J., 7 (1964), pp. 141–148.

[3] A. B. Bakushinskii, *Remarks on choosing a regularization parameter using the quasi-optimality and ratio criterion*, USSR Comput. Math. Math. Phys., 24 (1984), pp. 181–182.

[4] F. Bauer and M. A. Lukas, *Comparing parameter choice methods for regularization of ill-posed problems*, Math. Comput. Simulation, 81 (2011), pp. 1795–1841.

[5] H. W. Engl and H. Gfrerer, *A posteriori parameter choice for generalized regularization methods for solving linear ill-posed problems*, Appl. Numer. Math., 4 (1988), pp. 395–417.

[6] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Math. Appl. 375, Kluwer Academic Publishers Group, Dordrecht, 1996.

[7] H. Gfrerer, *An a posteriori parameter choice for ordinary and iterated Tikhonov regularization of ill-posed problems leading to optimal convergence rates*, Math. Comp., 49 (1987), pp. 507–522.

[8] M. S. Gockenbach, *Linear Inverse Problems and Tikhonov Regularization*, MAA Press, Washington, 2016.

[9] C. W. Groetsch, *Integral equations of the first kind, integral equations, and regularization: a crash course*, J. Phys.: Conf. Ser., 73 (2007), 012001, http://stacks.iop.org/1742-6596/73/i=1/a=012001.

[10] U. Hämarik, R. Palm, and T. Raus, *On minimization strategies for choice of the regularization parameter in ill-posed problems*, Numer. Funct. Anal. Optim., 30 (2009), pp. 924–950.

[11] M. Hanke, *Limitations of the L-curve method in ill-posed problems*, BIT, 36 (1996), pp. 287–301.

[12] M. Hanke and T. Raus, *A general heuristic for choosing the regulariztion parameter in ill-posed problems*, SIAM J. Sci. Comput., 17 (1996), pp. 956–972.

[13] P. C. Hansen, *Analysis of discrete ill-posed problems by means of the L-curve*, SIAM Rev., 34 (1992), pp. 561–580.

[14] P. C. Hansen, *Regularization tools: A Matlab package for analysis and solution of discrete ill-posed problems*, Numer. Algorithms, 6 (1994), pp. 1–35.

[15] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems*, SIAM, Philadelphia, 1998.

[16] 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.

[17] S. W. Indratno and A. G. Ramm, *An iterative method for solving Fredholm integral equations of the first kind*, Int. J. Comput. Sci. Math., 2 (2009), pp. 354–379.

[18] 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.

[19] S. Kindermann, *Discretization independent convergence rate for noise level-free parameter choice rules for the regularization of ill-conditioned problems*, Electron. Trans. Numer. Anal., 40 (2013), pp. 58–81.

[20] S. Kindermann and A. Neubauer, *On the convergence of the quasioptimality criterion for (iterated) Tikhonov regularization*, Inverse Probl. Imaging, 2 (2008), pp. 291–299.

[21] T. Kitagawa, *A deterministic approach to optimal regularization—the finite dimensional case*, Japan J. Appl. Math., 4 (1987), pp. 371–391.

[22] V. A. Morozov, *On the solution of functional equations by the method of regularization*, Soviet Math. Dokl., 7 (1966), pp. 414–417.

[23] V. A. Morozov, *Methods for Solving Incorrectly Posed Problems*, Springer-Verlag, New York, 1984.

[24] A. Neubauer, *The convergence of a new heuristic parameter selection criterion for general regularization methods*, Inverse Problems, 24 (2008), 055005, doi:10.1088/0266-5611/24/5/055005.

[25] T. Raus, *The principle of the residual in the solution of ill-posed problems with nonselfadjoint operator*, Uch. Zap. Tartu Gos. Univ., 715 (1985), pp. 12–20 (in Russian).

[26] T. Raus and U. Hämarik, *Heuristic parameter choice in Tikhonov method from minimizers of the quasi-optimality function*, in New Trends in Parameter Identification for Mathematical Models, B. Hofmann, A. Leitão, and J. P. Zubelli, eds., Birkhäuser, Cham, 2018, pp. 227–244.

[27] T. Reginska, *A regularization parameter in discrete ill-posed problems*, SIAM J. Sci. Comput., 17 (1996), pp. 740–749.

[28] J. A. O. Rodríguez, *Regularization methods for inverse problems*, Ph.D. thesis, University of Minnesota, Minneapolis, MN, 2011.

[29] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*, Wiley, New York, 1977.

[30] P. M. van den Berg, A. L. van Broekhoven, and A. Abubakar, *Extended contrast source inversion*, Inverse Problems, 15 (1999), pp. 1325–1344.

[31] C. R. Vogel, *Non-convergence of the L-curve regularization parameter selection method*, Inverse Problems, 12 (1996), pp. 535–547.

[32] G. Wahba, *Practical approximate solutions to linear operator equations when the data are noisy*, SIAM J. Numer. Anal., 14 (1977), pp. 651–667.

[33] Y.-F. Wang and T.-Y. Xiao, *Fast convergent algorithms for solving 2D integral equations of the first kind*, in Numerical Treatment of Multiphase Flows in Porous Media, Z. Chen, R. E. Ewing, and Z.-C. Shi, eds., Springer, New York, 2010, pp. 333–344.

[34] A.-M. Wazwaz, *The regularization method for Fredholm integral equations of the first kind*, Comput. Math. Appl., 61 (2011), pp. 2981–2986.