Comparing parameter choice methods for regularization of ill-posed problems

Frank Bauer[a], Mark A. Lukas[*,b]

[a] *Fuzzy Logic Laboratorium Linz-Hagenberg, University of Linz, Softwarepark 21, 4232 Hagenberg, Austria*
[b] *Mathematics and Statistics, Murdoch University, South Street, Murdoch WA 6150, Australia*

Abstract

In the literature on regularization, many different parameter choice methods have been proposed in both deterministic and stochastic settings. However, based on the available information, it is not always easy to know how well a particular method will perform in a given situation and how it compares to other methods. This paper reviews most of the existing parameter choice methods, and evaluates and compares them in a large simulation study for spectral cut-off and Tikhonov regularization. The test cases cover a wide range of linear inverse problems with both white and colored stochastic noise. The results show some marked differences between the methods, in particular, in their stability with respect to the noise and its type. We conclude with a table of properties of the methods and a summary of the simulation results, from which we identify the best methods.

Key words: ill-posed problem, inverse problem, regularization parameter, spectral cut-off, Tikhonov regularization, truncated singular value decomposition
2000 MSC: 65J20, 65J22, 62H12

1. Introduction

Consider the linear inverse problem

$$Ax = y, \tag{1}$$

where A is a linear compact operator mapping between two separable Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$. In practical situations, only a noisy version y^δ of y is available as data. Because of the compactness of A, solving (1) for x is unstable, and one needs to regularize the problem to obtain a reasonable approximate solution [41].

The two most popular regularization methods are spectral cut-off regularization (also called truncated singular value decomposition) and Tikhonov regularization (also called

*Corresponding author.
Email addresses: frank.bauer.de@gmail.com (Frank Bauer), M.Lukas@murdoch.edu.au (Mark A. Lukas)

ridge regression or Wiener filtering in certain contexts). For both regularization methods, the choice of the regularization parameter is crucial to yield a good solution.

Over the last four decades, many quite different methods for choosing this parameter have been proposed. Most of these methods have been developed with some analytical justification, but this is usually confined to a particular framework, e.g. deterministic error, and under certain conditions. Therefore, it may not be known how the method performs in all practical situations and under general conditions.

Considering the number of parameter choice methods, there are relatively few comparative studies of them [1, 3, 9, 29, 45, 60, 63, 67, 89, 101, 113, 125, 131, 141], [71, chap. 7]. Although the existing studies are useful, most of them are fairly limited in terms of the detail of the assessments, the number of methods compared and/or the breadth of the problems considered. Furthermore, they do not consider all of the more recent methods that have been developed.

This paper aims to provide a comprehensive and up-to-date comparative study of parameter choice methods for spectral cut-off and Tikhonov regularization. We will review most of the existing methods, from both deterministic and stochastic settings, and make practical and objective comparisons through a large simulation study. To achieve this, we use a set of synthetic test problems that models a wide range of practical problems. These problems make a more objective test platform than the subjective choices usually made in previous studies. They also allow quick computation, so a large number of replicates are used to assess each parameter choice method. In particular, we investigate the variability (displayed graphically) of each method both for white noise and for colored noise of unknown color.

There are some connections between our study and the recent work of Palm [113] (see also [60]). This work discusses a number of parameter choice methods, with refinements, in a deterministic framework for several regularization methods, including spectral cut-off, Tikhonov regularization, Landweber iteration and conjugate gradient type methods. For numerical comparisons, it uses the different set of (subjective) test problems from Hansen [70], but, as discussed in Section 5, some similar conclusions are made about the performance of the methods that are in common.

However, our paper has a somewhat different focus. It is intended to give practitioners important guidelines on both the selection and implementation of an appropriate parameter choice method for their problem. It combines an exhaustive overview of the current state of the art with a solid foundation of experiments comparing the performance and stability of the methods. The experiments and detailed graphical reports are designed to have a predictive capacity, so that a practitioner can see how each method will perform in a situation close to their own.

The paper is organized as follows. In Section 2, we give a brief overview of the different frameworks for regularization of linear inverse problems, with the notation that will be used. Section 3 describes our evaluation process and the design of our simulation experiments, which includes the generation of A and x in (1) and the noise in y^δ. In Section 4, we give a short description of each parameter choice method and its properties, including extensions to other regularization methods, and then present the simulation results with comments. Finally, Section 5 provides summaries of the properties and simulation results for all the methods and makes conclusions about their performance.

For ease of reference, we list below the subsections for the 20 parameter choice methods that are considered (including five methods in Sections 4.5 and 4.6) with, in brackets,

some other names used in the literature. In most cases, the names we use are the descriptive names originally given to the methods. However, in a few cases, there was no original name, and, to achieve consistency in the naming, we have chosen an appropriate name reflecting the basis of the method (rather than using the authors' name(s)). Section 4.18 lists some other methods that, for certain reasons, are not assessed.

4.1. Discrepancy principle
4.2. Transformed discrepancy principle
4.3. Modified discrepancy principle (Raus–Gfrerer rule, minimum bound method)
4.4. Monotone error rule
4.5. Balancing principle, balancing principle (white), fast balancing principle
4.6. Hardened balancing principle, hardened balancing principle (white)
4.7. Quasi-optimality criterion
4.8. L-curve method
4.9. Modified discrepancy partner rule (Hanke–Raus rule)
4.10. Extrapolated error method
4.11. Normalized cumulative periodogram method
4.12. Residual method
4.13. Generalized maximum likelihood
4.14. Generalized cross-validation
4.15. Robust generalized cross-validation
4.16. Strong robust GCV
4.17. Modified generalized cross-validation

2. Preliminaries

In practice, an inverse problem is often discretized (either as the model or for computation) and/or only a finite set of discrete data is available. Using equation (1), we will consider the following three cases in one common framework:

Case C1. Infinite dimensional situation, where A is a compact linear operator mapping between two separable Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$.

Case C2. Finite dimensional situation, where A is a matrix with large condition number mapping between $\mathcal{X} = \mathbb{R}^p$ and $\mathcal{Y} = \mathbb{R}^m$, where $p \leq m$. Assume that $\mathrm{rank} A = p$. This case is often called a discrete ill-posed problem.

Case C3. Discrete data situation, where the underlying problem $A_\infty x = y$ is still infinite dimensional, but we only have measurements $y_i = y(t_i)$ at the m points t_i, $i = 1, \ldots, m$. In this case, we define A to be the finite rank operator between $\mathcal{X}$ and $\mathcal{Y} = \mathbb{R}^m$ with $(Ax)_i = A_\infty x(t_i)$, $i = 1, \ldots, m$, so $Ax = y \in \mathbb{R}^m$. Assume that $\mathrm{rank} A = m$. This case is known as a semi-discrete model.

In all cases, the element $y \in \mathcal{Y}$ is perturbed by noise, giving the data element y^δ.

2.1. Regularization methods

In this paper, we will concentrate on the two main regularization methods that are used for solving linear inverse problems – spectral cut-off and Tikhonov regularization. Detailed accounts of these methods can be found in [41, 54, 77].

The operator or matrix A admits a singular value decomposition $\{\sigma_k, u_k, v_k\}_{k\in\mathbb{N}}$, where $\{u_k\}_{k\in\mathbb{N}}$ and $\{v_k\}_{k\in\mathbb{N}}$ are orthonormal in $\mathcal{X}$ and $\mathcal{Y}$, respectively, $Au_k = \sigma_k v_k$, $A^* v_k = \sigma_k u_k$ and $\sigma_k > 0$ are in decreasing order. Then

$$Ax = \sum_{k=1}^{R} \sigma_k \langle x, u_k \rangle v_k,$$

where R is ∞, p and m, respectively, in cases C1, C2 and C3 above. Spectral cut-off regularization is defined by

$$x_n^\delta = \sum_{k=1}^{l(n)} \sigma_k^{-1} \langle y^\delta, v_k \rangle u_k, \tag{2}$$

where l is an ascending integer valued function. The traditional choice is $l(n) = n$, but a general l allows us to restrict the regularized solutions to an appropriate subset, thereby reducing the computation time significantly without affecting the results.

Tikhonov regularization has a continuous regularization parameter α, but in practice one often searches over a discrete set. Here we use a geometric sequence of parameter values $\alpha_n = \alpha_0 q^n$, where $0 < q < 1$ and $n = 1, 2, \ldots$. Tikhonov regularization is defined by the variational formulation

$$x_n^\delta = \operatorname*{argmin}_{x\in\mathcal{X}} \|Ax - y^\delta\|^2 + \alpha_n \|x\|^2$$

or, equivalently, by

$$x_n^\delta = (A^*A + \alpha_n I)^{-1} A^* y^\delta, \tag{3}$$

where A^* is the adjoint of A. This method can be extended easily to have penalty function $\|Lx\|^2$ for some linear operator $L : \mathcal{X} \to \mathcal{Z}$, in which case

$$x_n^\delta = (A^*A + \alpha_n L^*L)^{-1} A^* y^\delta.$$

Here and throughout, the norm meant by $\|\cdot\|$ refers to the Hilbert space in use and will be clear from the context.

Note that, for both spectral cut-off and Tikhonov regularization, a larger value of the index n corresponds to less smoothing. For both methods, let x_n^0 be the regularized solution in the case of noise-free data and let A_n^{-1} be the linear regularization operator that maps y^δ to x_n^δ, i.e. it holds that

$$x_n^0 = A_n^{-1} y \qquad \text{and} \qquad x_n^\delta = A_n^{-1} y^\delta.$$

2.2. Noise

There are three main types of additive noise models used in the study of inverse problems. For all three, we will denote

$$y^\delta = y + \delta\xi,$$

where ξ is an appropriately normalized noise element and $\delta > 0$ is the noise level.

The most common noise model in the classical inverse problems literature is deterministic noise [41], where $\xi \in \mathcal{Y}$ with $\|\xi\| \leq 1$, so $\|y^\delta - y\| \leq \delta$. This noise model is quite suitable to represent discretization errors, but it is rather poor for describing random measurement errors arising in practice.

A practical stochastic noise model for a discrete data vector $y^\delta \in \mathbb{R}^m$ (for cases C2 and C3) [142] is $y^\delta = y + \delta\xi$, where the components ξ_i are i.i.d. random variables with mean $\mathbb{E}\xi_i = 0$ and variance $\mathbb{E}\xi_i^2 = 1$. Then δ is the standard deviation of each error component $\delta\xi_i$ and $\mathbb{E}\|y^\delta - y\|^2 = \delta^2\mathbb{E}\|\xi\|^2 = m\delta^2$. This model can be extended to one involving correlated errors, where $\varepsilon := \delta\xi$ has covariance matrix $C = [\mathbb{E}(\varepsilon_i\varepsilon_j)]$.

A stochastic noise model can also be defined in an infinite dimensional setting (case C1) by using the singular value decomposition of A. Suppose that the Fourier coefficients $\langle y, v_k\rangle$ are known only as the sequence data

$$y_k^\delta = \langle y, v_k\rangle + \delta\xi_k = \sigma_k \langle x, u_k\rangle + \delta\xi_k, \quad k = 1, 2, \ldots,$$

where $\xi_k := \langle \xi, v_k\rangle$ are independent normal $\mathcal{N}(0, 1)$ random variables and ξ is a zero-mean weak Gaussian random element. This is called a continuous Gaussian white noise model [94]. In this case, it is easy to see that

$$\mathbb{E}\langle \xi, f\rangle \langle \xi, g\rangle = \sum_k f_k g_k = \langle f, g\rangle$$

for any pair $f, g \in \mathcal{Y}$, where $f_k = \langle f, v_k\rangle$ and $g_k = \langle g, v_k\rangle$. Note there is no bound for the error in $\mathcal{Y}$ here, since $\mathbb{E}\sum(y_k^\delta - \langle y, v_k\rangle)^2 = \sum \delta^2$ is infinite. Colored noise can be defined by introducing a covariance matrix K for the random variables ξ_k so that $\mathbb{E}(\xi_k\xi_l) = K_{kl}$, in which case we have $\mathbb{E}\langle \xi, f\rangle \langle \xi, g\rangle = \sum_{kl} f_k K_{kl} g_l$ for any pair $f, g \in \mathcal{Y}$. A simple choice is to assume K to be diagonal. Then, if the entries K_{kk} are increasing, it is called blue noise, and, if they are decreasing, it is called red noise.

For the finite dimensional case C2, if $y^\delta = y + \delta\xi$ with $\xi \sim \mathcal{N}(0, I)$, then, clearly, using the orthonormal singular vectors $v_k \in \mathbb{R}^m$ of A, the model can be written equivalently as a Gaussian white noise model with finite sequence data.

2.3. Assumptions on x

In most of the literature on regularization, it is assumed that x is a fixed (non-random) element of $\mathcal{X}$. It is known [41, 30, 97] that the error $\|x - x_n^\delta\|$ or the expected squared error $\mathbb{E}\|x - x_n^\delta\|^2$ (with respect to the noise distribution) in the regularized solution (2) or (3) depends on the abstract smoothness of the unknown solution x. The smoothness assumption made on x is called a source condition, and is usually of the form $x \in \mathcal{R}((A^*A)^s)$ for some $s > 0$.

In the Bayesian approach to inverse problems [44, 76, 85, 129], it is assumed that x is a random element of $\mathcal{X}$ with some prior distribution. These models can be formulated in any of the cases C1, C2 or C3 above. It is known [46, 77, 91, 142] that, with a Gaussian prior and independent Gaussian error distribution, the posterior mean given the data is the solution of a certain Tikhonov regularization problem. In a simple case, if $A : \mathbb{R}^p \to \mathbb{R}^m$ and we have independent Gaussian random variables $x \sim \mathcal{N}(0, \eta^2 I)$ and $\delta\xi \sim \mathcal{N}(0, \delta^2 I)$, then the posterior mean of x given the data vector y^δ is

$$\bar{x} = \left(A^*A + (\delta^2/\eta^2)I\right)^{-1} A^* y^\delta, \tag{4}$$

which is of the same form as (3). This approach extends to the representation $x = \sum c_i \phi_i$ with respect to an arbitrary orthonormal basis $\{\phi_i\}$ by assuming that $c \sim \mathcal{N}(0, \Gamma)$. Similarly, for an infinite dimensional space $\mathcal{X}$, one can use the prior distribution that the Fourier coefficients $\langle x, u_k \rangle$ are independent random variables with distribution $\mathcal{N}(0, \gamma(k))$ [9]. Note that, from (4), if δ and η were known, then δ^2/η^2 would be an appropriate choice of the regularization parameter; however, they are usually not known in practice.

2.4. Parameter choice method

A parameter choice method is a rule that assigns a value for the regularization parameter. In our situation, with a discrete set of parameters for Tikhonov as well as spectral cut-off regularization, the method selects a value for the index, which will be denoted by n_*.

Parameter choice methods can be classified according to the input they require. There are three basic types [15, 41]:

- a-priori method, where n_* is a function of δ and information about the smoothness of x;
- a-posteriori method, where $n_* = n_*(\delta, y^\delta)$; and
- data-driven method, where $n_* = n_*(y^\delta)$.

Because a-priori methods need information about x that is generally not known, they are not really practical and so will not be discussed here. A-posteriori methods are much more practical; if δ is not known exactly, then one can use an estimate in its place, though this may affect the performance. Data-driven methods have the significant advantage that they only require the data y^δ as input. In the literature on deterministic noise, these methods are sometimes called "heuristic methods", but this has a negative connotation that is not generally deserved. We will consider several methods of each of the second and third types. For these methods, if y^δ contains stochastic noise, then n_* is a random variable.

All the methods rely on computing an associated function $F(n)$. It is possible to distinguish two types of methods based on how $F(n)$ is used to define the parameter:

- n_* = point at which F falls below some threshold;
- n_* = point at which F is minimized.

Most of the methods of the first type have their origins and analysis in a deterministic setting, and they use a (sensitive) tuning parameter to give their best results. By contrast, most of the methods of the second type come either from a stochastic framework, usually with appropriate analysis, or are based only on heuristic ideas. Mostly, these methods do not use a (sensitive) tuning parameter.

2.5. Optimal regularization parameter

For the problem $Ax = y$ with data y^δ, we define the optimal regularization parameter (index) by

$$n_{opt} = \operatorname*{argmin}_n \|x - x_n^\delta\|.$$

If y^δ contains stochastic noise, then n_{opt} is a random variable. In our numerical experiments of each parameter choice method, we will assess the accuracy of the choice n_* by computing the inefficiency defined by

$$\|x - x^\delta_{n_*}\| / \|x - x^\delta_{n_{opt}}\|. \tag{5}$$

The closer this is to 1, the better is the parameter choice. Using stochastic noise with a large number of replicates of the problem, we can estimate the distribution of the inefficiencies and hence determine the performance of the method.

It is clear that, since x is unknown, a practical parameter choice method must use some other known or easily computed quantities. Many methods use the norm of the residual defined as $\|y^\delta - Ax^\delta_n\|$. If the data are finite and the norm is the Euclidean norm, this is the square root of the usual residual sum of squares, and so it is easily computed.

Clearly, the error $\|x - x^\delta_n\|$ can be bounded as

$$\|x - x^\delta_n\| \leq \|x - x^0_n\| + \|x^0_n - x^\delta_n\|. \tag{6}$$

Although this bound is not directly computable, several parameter choice methods use it in various indirect ways.

The first term on the right-hand side of (6) is the regularization error, i.e. the error that is purely due to the regularization operator acting on (noise-free) y. Since a regularization operator is a perturbation of the generalized inverse of A in which the perturbation decreases with less smoothing, we can expect this error to decrease as n increases. For many regularization methods, including spectral cut-off and Tikhonov regularization, one can bound the regularization error as

$$\|x - x^0_n\| \leq \varphi(n), \tag{7}$$

where φ is a decreasing function which depends on the source condition and the regularization method. The rate of decrease may improve with higher smoothness s, where $x \in \mathcal{R}((A^*A)^s)$, but possibly only up to some maximum value s_0. Here s_0 is called the qualification of the method, and, if $s_0 < \infty$, the method exhibits saturation. Tikhonov regularization has qualification $s_0 = 1$, while spectral cut-off has infinite qualification [41].

The second term on the right-hand side of (6) is called the propagated noise error. For many regularization methods, it can be bounded as

$$\|x^0_n - x^\delta_n\| \leq \delta \varrho(n), \tag{8}$$

where ϱ is a known increasing function of n, indicating that, with less smoothing, there is more influence of the data noise. For spectral cut-off regularization, (8) holds with $\varrho(n) = \sigma^{-1}_{l(n)}$ and, for Tikhonov regularization, (8) holds with $\varrho(n) = \alpha_n^{-1/2}$ [41].

When the noise is stochastic, it is usual to consider the expected squared error $\mathbb{E}\|x - x^\delta_n\|^2$, known as the risk. For noise with zero mean, instead of (6), the risk can be decomposed exactly into a sum of squared bias and variance terms as

$$\mathbb{E}\|x - x^\delta_n\|^2 = \|x - x^0_n\|^2 + \mathbb{E}\|x^0_n - x^\delta_n\|^2. \tag{9}$$

The squared bias can be bounded as in (7) and, under suitable assumptions, the variance can be expressed as

$$\mathbb{E}\|x_n^0 - x_n^\delta\|^2 = \delta^2 \varrho^2(n) \tag{10}$$

for some increasing function $\varrho(n)$. For white noise, the spectral cut-off solution (2) has variance

$$\delta^2 \varrho^2(n) = \delta^2 \mathbb{E}\|A_n^{-1}\xi\|^2 = \delta^2 \sum_{k=1}^{l(n)} \sigma_k^{-2} \tag{11}$$

and the Tikhonov regularized solution (3) has variance

$$\delta^2 \varrho^2(n) = \delta^2 \mathbb{E}\|A_n^{-1}\xi\|^2 = \delta^2 \sum [\sigma_k/(\sigma_k^2 + \alpha_n)]^2. \tag{12}$$

A much more detailed discussion of the above errors (including, e.g., minimax results) in various situations can be found in [41, 105, 9, 30, 97].

The analysis of a parameter choice method depends on the underlying case (C1, C2 or C3) and noise model. For case C1 and deterministic noise, it is usual to consider the behavior of the method as $\delta \to 0$. A regularization method with parameter choice n_* is said to be convergent [41] if $\sup \|x - x_{n_*}^\delta\| \to 0$ as $\delta \to 0$, where the supremum is taken over all y^δ satisfying $\|y - y^\delta\| \le \delta$, so this is a "worst-case" condition. A result of Bakushinskii [7] states that, for an ill-posed problem, a parameter choice rule that does not explicitly use the noise level (e.g. data-driven rules) cannot yield a convergent regularization method. This result (sometimes called the Bakushinskii veto) appears to have had more influence in the literature than is warranted; while the result is important for deterministic noise, it is not really appropriate for stochastic noise [15]. In this situation, as we shall see, there are data-driven rules yielding regularization methods that converge with respect to the risk and perform very well in practice.

In many papers, it is proved that a particular parameter choice method is order optimal for a certain smoothness class, i.e., if $\|x - x_{n_{opt}}^\delta\| = O(\delta^{p_{opt}})$, then also $\|x - x_{n_*}^\delta\| = O(\delta^{p_{opt}})$. Although this is important, it is not an ideal result since the method is possibly sub-optimal by an arbitrarily large factor depending on x [13].

For some methods, there are stronger results involving oracle inequalities [19, 24, 26, 25], which, for continuous white noise, have the form

$$\mathbb{E}\|x - x_{n_*}^\delta\|^2 \le C(\delta) \inf_n \mathbb{E}\|x - x_n^\delta\|^2 + \beta(\delta)$$

for all x in some class. These provide, for any noise level, a bound on the risk relative to the smallest possible value of the risk, and, ideally, $C(\delta)$ is close to 1 for all δ and $\beta(\delta)$ is much smaller than the first term. If $C(\delta) \to 1$ and $\beta(\delta) = o(\inf_n \mathbb{E}\|x - x_n^\delta\|^2)$ as $\delta \to 0$, then the method is said to be asymptotically optimal.

Similar results exist for some methods with a discrete noisy data vector $y^\delta \in \mathbb{R}^m$ in case C3, where the asymptotic analysis is as $m \to \infty$ with fixed variance δ^2. There are connections between results for the continuous white noise model and for discrete sampled data. In particular, for function estimation ($A = I$), it is known [22] that, under certain conditions, asymptotic results for the continuous white noise model as $\delta \to 0$ can be translated into asymptotic results for discrete data as $m \to \infty$.

3. Evaluation process

This section presents the design of our review and evaluation process to be used in the next section. For each parameter choice method, we will start by describing the origin and rationale of the method. Then we will state the mathematical requirements of the method and the algorithm that we use. This will be followed by a brief discussion of known theoretical and practical issues about the method, including whether the method works for other regularization methods. Finally we will present the results of the numerical experiments for the method.

3.1. Numerical experiments

Each parameter choice method will be assessed using the same large set of test problems. The results will be shown for all situations, independent of any prior knowledge about situations where the method does not work. For each parameter choice method, the experiments used the same random seed, so every method had exactly the same set of operators, solutions and noisy data to deal with. The experiments were implemented in MATLAB®.

The test problems are finite dimensional problems (case C2), where $A : \mathcal{X} \to \mathcal{Y}$ and $\mathcal{X} = \mathcal{Y} = \mathbb{R}^m$ with Euclidean norms. The problems are characterized by the following parameters:

- m : number of eigenvalues of A;
- μ : decay behavior of the eigenvalues of A;
- ν : decay behavior (smoothness) of x;
- $\log(N2S)$: $\log_{10}$ of the noise-to-signal ratio $N2S = (\mathbb{E}\|y^\delta - y\|^2)^{1/2}/\|y\|$; and
- ω : noise behavior (corresponds to the color).

3.1.1. Operator generation

The operator A will be taken to be a random diagonal $m \times m$ matrix, with diagonal elements (i.e. eigenvalues or singular values) decaying like $a_{kk} \approx k^{-\mu}$. A larger value of μ corresponds to a more ill-posed problem. The diagonal vector is generated by the following procedure (written in MATLAB® notation).

$HelpVar = (1 : (m + 100)).\hat{} - \mu;$
$Perturb = \exp(0.5 * \text{randn}(m + 100, 1) - 0.5^2/2);$
$HelpOp = \text{sort}(HelpVar .* Perturb, \text{'descend'});$
$A = HelpOp(1 : m);$

This choice of operator raises some natural questions which warrant discussion.

Why did we use diagonal matrices? Why did we not use the standard problem XYZ? For each parameter choice method, more than 300,000 trials are done, so speed is a very important issue. For this reason, we use the simplest possible form of an inverse problem. Because of the singular value decomposition, these diagonal problems are no less or more ill-posed than other discrete inverse problems. Furthermore, this approach enables us to see the effects of ill-posedness with almost no side-effects originating from numerical errors due to machine precision and other machine dependent errors.

Why did we not use $a_{kk} = k^{-\mu}$, or why did we not perturb the operator by the stochastic process XYZ? We wanted to ensure that we do not use, even accidently, specific features

of the operator that would help the inversion, but cannot be found in practice (called "inverse crimes" [28]). Therefore, we used a slight random perturbation of the sequence $k^{-\mu}$. The procedure above achieves a good balance between retaining the overall $k^{-\mu}$ behavior of the operator while providing some randomness in this component.

3.1.2. Solution generation

Each time a solution x is generated, we use the following procedure.

$HelpVar = (1:m).^{\wedge} - \nu;$
$Sign = 2 * \text{ceil}(2. * \text{rand}(m,1)) - 3;$
$Perturb = 1 + 0.1 * \text{randn}(m,1)$
$x = Sign .* Perturb .* HelpVar;$

This can be interpreted as generating random Fourier coefficients x_k with decay behavior $|x_k| \approx k^{-\nu}$ and random sign of equal probability. A larger value of ν gives a smoother solution x. Note that if $\nu > 1/2$, then (with probability 1) $\|x\|$ is bounded independent of the dimension m.

Why did we not define the solutions as $x = (A^*A)^s x_0$ for a fixed x_0 and different values of s, i.e. in the context of a source condition? Once the operator has been chosen as above, the decay behavior of x and the source condition form are in a one-to-one correspondence, so either could be used to define x. We chose the first approach as we used encapsulation in the software design, i.e., the solution should not "see" the operator and vice versa.

Why did we not use some colored Gaussian variable for x as in [9, 85]? Aren't the components too far away from 0? In our evaluation of the methods, we want to identify the effect of the smoothness of the solution, determined by the decay behavior of x and specified by the parameter ν, and also the effect of the noise-to-signal ratio $N2S$. To achieve this, there cannot be too much variability in these features of x for different replicates of x. For a Gaussian random variable, both the norm of x and the decay behavior of x vary over a large scale, which also affects the noise-to-signal ratio. The procedure above is a reasonable compromise in limiting the variability in the norm and decay behavior, while allowing x to be as flexible as possible.

3.1.3. Noise generation

We use a finite stochastic noise model. First, for each replicate $y = Ax$, the noise level δ is defined from the input noise-to-signal ratio $N2S$ as $\delta = N2S * \|y\|/\sqrt{m}$. Then, each time a noise vector is generated, we employ the following procedure.

$Points = \text{ceil}(1.5 * m);$
$NoiseTime = \text{randn}(Points, 1);$
$StdDev = \text{zeros}(21, 1);$
$StdDev(11) = 1;$
for $counter = 1 : \text{ceil}(\text{abs}(10 * \omega))$
  $NoiseTimeTemp = NoiseTime;$
  $NoiseTime(1 : Points - 1) = NoiseTime(1 : Points - 1, :)$
    $+\omega * NoiseTimeTemp(2 : Points);$
  $NoiseTime(Points) = NoiseTime(Points) + \omega * NoiseTimeTemp(1);$
  $NoiseTime(2 : Points) = NoiseTime(2 : Points)$
    $+\omega * NoiseTimeTemp(1 : Points - 1);$

$NoiseTime(1) = NoiseTime(1) + \omega * NoiseTimeTemp(Points)$;
$StdDevTemp = StdDev$;
$StdDev(2 : end) = StdDev(2 : end) + \omega * StdDevTemp(1 : end - 1)$;
$StdDev(1 : end - 1) = StdDev(1 : end - 1) + \omega * StdDevTemp(2 : end)$;
end
$NoiseTime = NoiseTime/\,\mathrm{norm}(StdDev)$;
$yTime = \mathrm{dct}([y; \mathrm{zeros}(Points - m, 1)])$;
$yPertTime = yTime + \delta * NoiseTime$;
$yHelp = \mathrm{idct}(yPertTime)$;
$y^{\delta} = yHelp(1 : m)$;

This means that we transform y via the discrete cosine transform to the time domain, add (possibly serially correlated) noise with standard deviation δ, and transform back. The degree of correlation and color is determined by ω. For $\omega = 0$, the for-loop section is not executed and the noise is simply Gaussian white noise. For $\omega \neq 0$, the noise $(\delta * NoiseTime)$ is defined by a moving average process in which both the weights and order ceil(abs$(10 * \omega)$) depend on ω. The noise has higher correlation for larger ω, and it has color red for $\omega > 0$ and color blue for $\omega < 0$. Figure 1 displays realizations of the noise for $\omega = 0$ (white noise), $\omega = 0.5$ (red noise) and $\omega = -0.5$ (blue noise).

For our experiments, we considered two scenarios for the noise. The first was simply white noise, i.e. $\omega = 0$ above. The second scenario was colored noise of unknown random color. For this we used the above procedure with ω chosen as a pseudo-random variate, uniformly distributed in the interval $[-0.5, 0.5]$.

Why did we generate and add the noise in the time domain? The for-loop section of the above procedure uses a simple and efficient technique to get positive or negative serial correlation, corresponding to the desired noise color. This type of correlation is observed in many practical applications, for example, in geodesy [35, 36]. Also, by adding the noise in a different space from the one used to generate $y = Ax$, we avoided potential inverse crimes.

3.1.4. Tikhonov regularization

For the regularization parameter sequence $\alpha_n = \alpha_0 q^n$, $n = 1, 2, \ldots$, in (3), we used 50 values from 10^1 to 10^{-12} with logarithmic equal spacing. This means that $\alpha_0 = 18.42$ and $q = 0.5429$.

The set of test cases we used for Tikhonov regularization is given in Table 1. The parameter values for these cases were chosen to achieve a balance between speed and the representation of a reasonably wide set of problems. In addition, the parameter values are constrained so that the optimal regularization parameter lies clearly between 10^1 and 10^{-12}, i.e. it is not near an endpoint.

3.1.5. Spectral cut-off regularization

For the sequence $l(n)$ of cut-off points in (2), we used the following procedure, which is similar to that described in [9].

$OldSubSample = 1 : 3 : m$;
$HelpVar = A(1)$
for $Element = 1 : \mathrm{length}(OldSubSample)$
$OldElement = OldSubSample(Element)$;

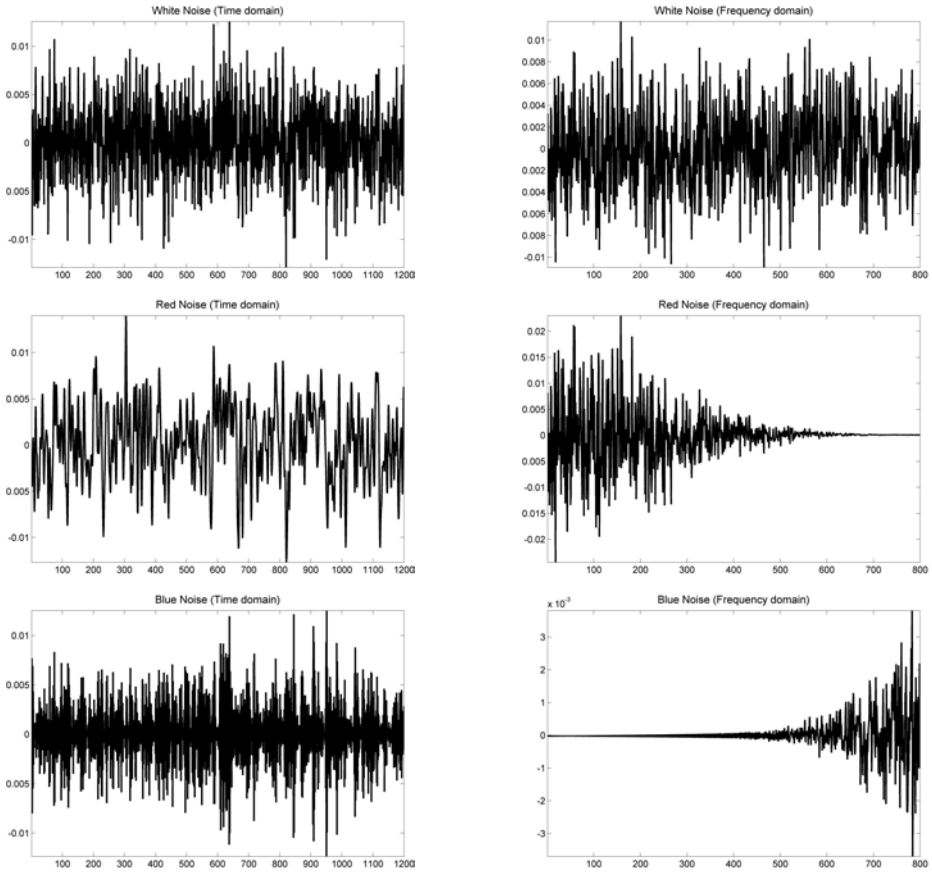

Figure 1: Different realizations of the noise, in the time domain (left) and frequency domain (right), for $\omega = 0$ (top), $\omega = 0.5$ (middle) and $\omega = -0.5$ (bottom)

if $A(OldElement) < HelpVar/1.07$
$SubSample = [SubSample\ \ OldElement]$;
$HelpVar = A(OldElement)/1.07$;
end
end

This means there is a minimum spacing of 3 between the cut-off points, and, furthermore, the corresponding eigenvalues form an exponentially decreasing sequence. Therefore, we will refer to this method as exponential spectral cut-off (ExpCutOff). One can easily show that, with this version of spectral cut-off, we also achieve the minimax rate for optimal parameter choice [18].

The set of test cases we used for ExpCutOff is given in Table 2. Again, the parameter values for these cases are constrained so that the optimal parameter n_{opt} has $l(n_{opt})$ clearly between 1 and m, i.e. not near an endpoint. Note that different test cases were used for Tikhonov and ExpCutOff regularization.

μ	m	ν	$\log(N2S)$
1	800	1	-1.0
1	800	1	-3.0
1	500	3	-1.0
1	500	3	-3.0
1	500	3	-5.0
1	300	5	-1.0
1	300	5	-3.0
3	500	1	-1.0
3	500	1	-3.0
3	300	1	-5.0
3	150	3	-3.0
3	150	3	-5.0
3	150	3	-7.0
3	150	5	-3.0
3	150	5	-5.0
5	300	1	-3.0
5	150	1	-5.0
5	150	3	-5.0
5	150	5	-7.0

Table 1: Test cases for Tikhonov regularization

μ	m	ν	$\log(N2S)$
1	800	1	-1.0
1	800	1	-3.0
1	500	3	-3.0
1	500	3	-5.0
1	500	3	-7.0
1	300	5	-5.0
1	300	5	-7.0
3	500	1	-3.0
3	300	1	-5.0
3	300	1	-7.0
3	150	3	-5.0
3	150	3	-7.0
3	150	3	-11.0
3	150	5	-11.0
5	300	1	-7.0
5	150	1	-11.0
5	120	3	-11.0
5	120	5	-11.0

Table 2: Test cases for ExpCutOff regularization

3.1.6. Maximal regularization parameter

For most parameter choice methods, the choice is defined either by incrementing $n = 1, 2, \ldots$ until a certain condition is satisfied or by minimizing some function for all n. For most of these methods, the use of a discrete set of regularization parameters does not alter the behavior of the method. One simply needs to have a fine enough resolution, as chosen above. Clearly, for the efficient implementation of these methods, it is useful to have a bound on the value of n_{opt} (i.e. a maximal regularization parameter), especially in order to restrict the search for a minimum.

The quasi-optimality criterion is defined by a minimization, but the behavior of this method is affected more seriously when using a discrete set of regularization parameters for Tikhonov regularization. For the method to work in this situation, one needs to constrain the value of n to be smaller than a suitable maximal index N that makes α_n greater than the smallest eigenvalue σ_m^2 of A^*A. This is to ensure that, numerically, the finite problem behaves like an ill-posed problem for all $n \leq N$. This is discussed further in Section 4.7.

For a few parameter choice methods, e.g. the balancing principle, a maximal index N is an essential input in the algorithm itself. The actual value of N is not crucial so long as $n_{opt} < N$ and, for the sake of computational efficiency, N is not too large.

To satisfy the above requirements, for both spectral cut-off and Tikhonov regularization in a stochastic setting, it is reasonable to define the maximal index as

$$N = \max\{n | \varrho(n) < 0.5\varrho(\infty)\}, \tag{13}$$

where $\delta^2\varrho^2(n)$ is the variance in (10) and $\delta^2\varrho^2(\infty)$ is its supremum. We can expect that $n_{opt} < N$ because, if $n \geq N$, then it follows from (9) (since $\|x - x_n^0\|^2$ is decreasing) that $\mathbb{E}\|x - x_n^\delta\|^2 \geq 0.25\mathbb{E}\|x - x_\infty^\delta\|^2$, and we would expect $\mathbb{E}\|x - x_{n_{opt}}^\delta\|^2$ to be much smaller than the right-hand side.

To obtain N in practice, one either has to have an analytic expression for $\delta^2\varrho^2(n)$, as in (11) and (12) for white noise, or a good estimate of it. It is known [10] that, for any noise color, if several independent data sets are available, then a good estimate of $\delta^2\varrho^2(n)$ is $2^{-1}\text{Mean}\{\|x_{n,i}^\delta - x_{n,j}^\delta\|^2, i \neq j\}$. In the experiments, we use four data sets to obtain N for the methods that require a maximal parameter.

Furthermore, so that all the parameter choice methods can be compared on an equal basis, we use the same maximal index N for all the methods. For many methods, the usage of N has almost no effect on the results. However, for some methods, the restriction that $n \leq N$ has the beneficial effect of reducing the number of severely under-smoothed solutions in some test cases. If the restriction has a noticeable effect, this will be noted in the description of the results.

If, in practice, only a single data set is available, then it may not be possible to estimate $\delta^2\varrho^2(n)$ if the noise is correlated with unknown covariance. Then one can define a maximal index N_1 by $l(N_1) = m$ for spectral cut-off and by $\alpha_{N_1} \approx \sigma_m^2$ for Tikhonov regularization. For the methods that perform much worse without the use of the maximal index N, the results for N and N_1 may be quite different. However, for the methods that perform essentially the same with or without the use of N, the results for N and N_1 will be very similar.

3.1.7. Runs

For each parameter choice method, we performed exactly the same experiments, constructed as follows. For each of the cases in Tables 1 and 2, and for both white noise and colored noise, we:

- generated 8 operators A as above;
- for each operator A, we generated 8 solutions x as above; and
- for each pair (A, x) (i.e. for each $y = Ax$), we generated 64 different noisy data vectors y^δ as above. In the colored noise scenario, the 64 vectors are made up of 8 groups of 8 vectors, where each group has a different color and within the group the color is the same.

This means that, for each test case and noise scenario, there are 4096 inverse problems that need to be solved. The hierarchical structure was chosen in order to considerably reduce the computational cost. In total, this article is based on the solution of more than 6 million inverse problems.

3.1.8. Plots

For each parameter choice method, we will display the simulation results in one figure, with four panels corresponding to Tikhonov and ExpCutOff regularization under both the white noise and colored noise scenarios. Each panel has the following features.

- For each test case (denoted by $(\mu, \nu, \log(N2S))$, with m determined from Tables 1 and 2), a box plot (marking the lower and upper quartiles) shows the distribution of computed inefficiencies (5) for the method, with whiskers showing the range.

The whiskers have maximum length of 4 times the interquartile range, and outliers beyond this are marked with a + symbol.

- For each test case, the middle band in the box shows the median of the computed inefficiencies and an open dot shows the sample mean.

3.2. Error comparison for optimal solutions

To conclude this section, fig. 2 shows the distribution of the optimal errors $\|x - x^{\delta}_{n_{opt}}\|$ for Tikhonov and ExpCutOff regularization under both the white and colored noise scenarios. The test cases and replicates are the same as those used for the simulations in the next section. The box plots are constructed in the same way as described above.

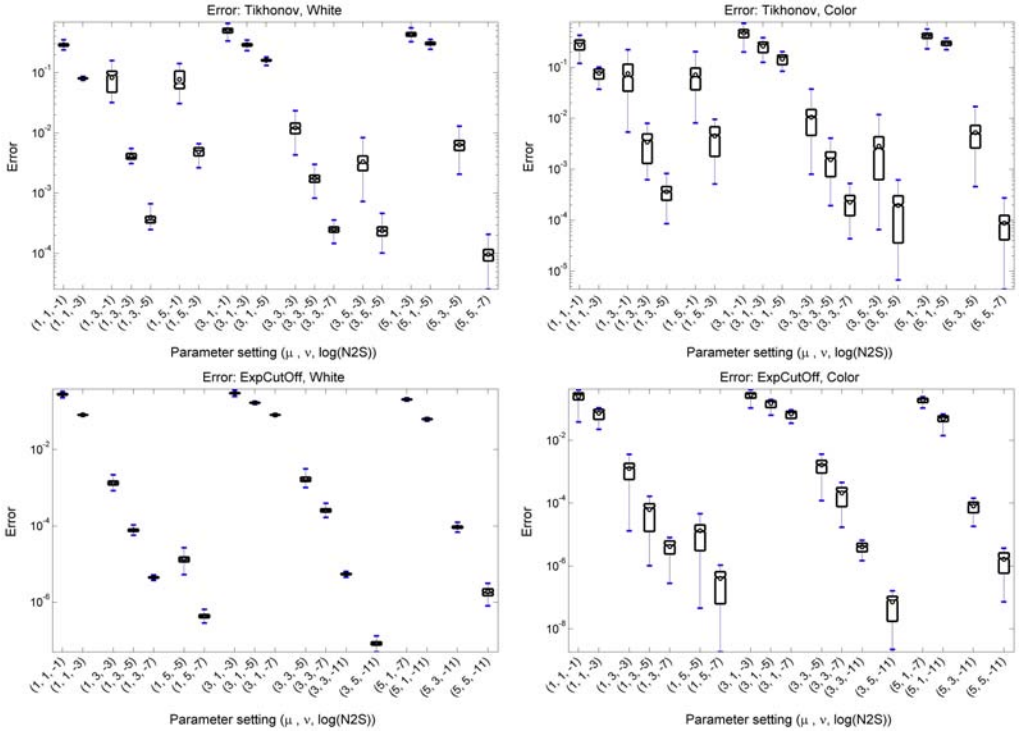

Figure 2: Comparison of optimal errors

As one can see, the optimal errors differ by several orders of magnitude across the test cases, some of which is simply due to the different noise levels. In addition, there can be significant variability within one test case, especially for the colored noise scenario. This is not surprising because, depending on whether the colored noise is at the blue end or red end, both regularization methods will find it easier or harder, respectively, to extract the solution x, compared to the white noise situation.

Furthermore, we clearly see some effects of the saturation property of Tikhonov regularization [41, 105]; in particular, the errors for the parameter sets $(1, 1, -3)$, $(1, 3, -3)$ and $(1, 5, -3)$ indicate there is no improvement as the smoothness increases beyond a certain value. Consistent with the theory, the results for spectral cut-off show no saturation effects.

4. Description and evaluation of methods

In this section, we will describe and evaluate the major parameter choice methods according to the process outlined in the previous section. The methods are presented in a specific order. Those in the first group require knowledge of the noise level, those in the second group require two or more independent sets of data as input, and those in the last group require no knowledge about the noise. Within each group, the methods are presented in either historical order or an order based on connections between the methods.

Many of the methods use a tuning parameter or some other parameter that must be chosen. For each of these methods, we have used any available information about the setting of the parameter and then chosen it to roughly optimize the method's performance in the simulations, on average over the test cases. However, the search for an optimal setting was not exhaustive, and the optimal settings for different problems might vary significantly.

At the end of the section is a list of methods that, for certain reasons, were not included in this study.

4.1. Discrepancy principle

The discrepancy principle, which was originally proposed by Phillips [114] and then developed and analyzed by Morozov [107, 108], is one of the oldest and most widely used parameter choice procedures ([41] and references therein). The rationale for the method is simply that for a good regularized solution, the norm of the residual should match the noise level δ of the data. Although the method was originally developed in a deterministic setting, it has also been studied in a discrete, stochastic setting [34, 99, 139].

Method. In a deterministic setting with $\|y^\delta - y\| \leq \delta$, the parameter choice n_* is the first n such that $\|Ax_n^\delta - y^\delta\| \leq \tau\delta$, where $\tau \geq 1$ is a tuning parameter. In a stochastic setting, with the error in each element of $y^\delta \in \mathbb{R}^m$ having standard deviation δ, the choice n_* is the first n such that

$$\|Ax_n^\delta - y^\delta\| \leq \tau\delta\sqrt{m}. \tag{14}$$

We use $\tau = 1.5$.

Known issues. There has been a lot of work done on the convergence properties of this method [41, 54, 108]. In the deterministic setting, for Tikhonov regularization with parameter $\alpha(\delta)$ chosen by the discrepancy principle, it is well known that $x_{\alpha(\delta)}^\delta$ converges to $x \in \mathcal{R}((A^*A)^s)$ as $\delta \to 0$ at the optimal order $O(\delta^{2s/(2s+1)})$ if $s \in (0, 1/2]$, but at the sub-optimal order $O(\delta^{1/2})$ if $s > 1/2$; i.e. it reaches saturation at $s = 1/2$. For spectral cut-off, the discrepancy principle gives the optimal order for any value of s.

In the stochastic setting with a data vector $y^\delta \in \mathbb{R}^m$ containing uncorrelated errors of variance δ^2, for Tikhonov regularization, it is known [34, 99, 139] that as the sample size $m \to \infty$, the "expected" discrepancy principle estimate has the optimal rate for the prediction risk $\mathbb{E}\|Ax_n^\delta - y\|^2$ (though, if $\tau = 1$, the constant makes it over-smoothing). It is also order optimal for the $\mathcal{X}$-norm risk $\mathbb{E}\|x_n^\delta - x\|^2$ if x is not too smooth relative to the operator A, i.e. $\nu \leq \mu + 1/2$ here, but otherwise it is order sub-optimal (under-smoothing). It is shown in [101] that for $\tau = 1$, the actual estimate is asymptotically

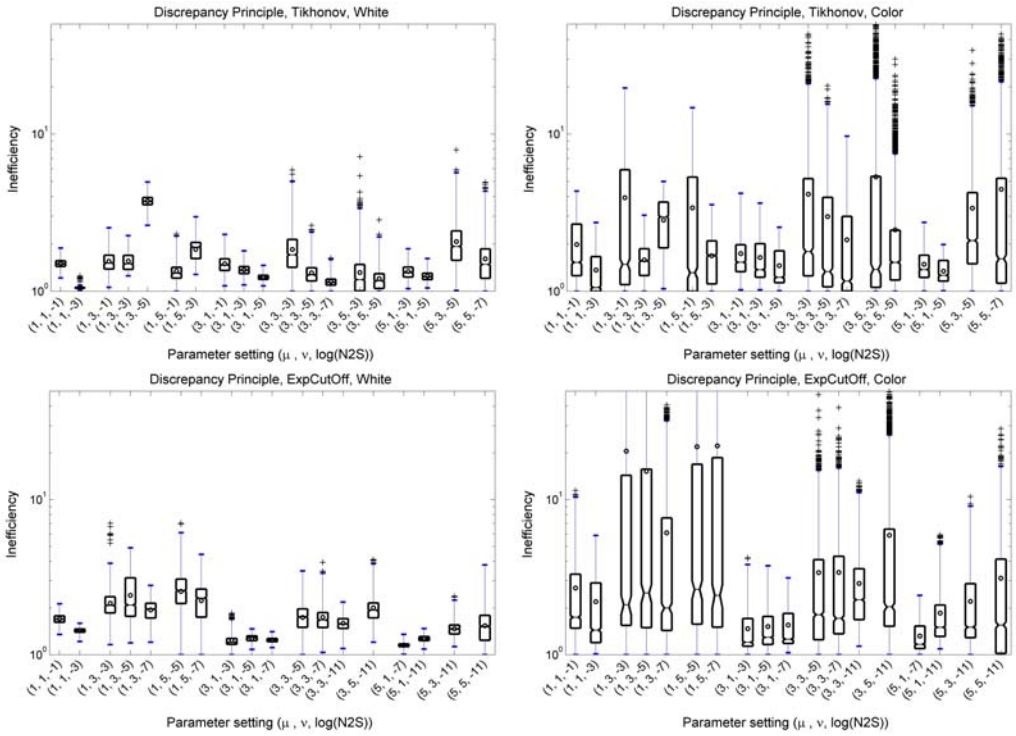

Figure 3: Inefficiencies of the discrepancy principle

unstable in a relative sense. For spectral cut-off, the "expected" estimate is order optimal for both the prediction risk and the $\mathcal{X}$-norm risk [139].

The discrepancy principle is one of the fastest methods available, since one only needs to compute the residuals until the bound (14) is satisfied. However, it has the serious drawback that it needs an accurate estimate of the noise level; even small misestimations can lead to very poor solutions [71, Chapter 7].

The discrepancy principle has also been applied to and analyzed for various iterative regularization methods for linear and nonlinear problems in the deterministic setting, including Landweber iteration, the conjugate gradient algorithm and the iteratively regularized Gauss–Newton method (see [41, 67, 42, 8, 81, 86] and the references therein). As in the linear case, the discrepancy principle applied to the iteratively regularized Gauss–Newton method yields sub-optimal convergence rates.

Numerics. From fig. 3, the discrepancy principle with $\tau = 1.5$ performs reasonably well for Tikhonov and spectral cut-off regularization with white noise, though there is some bias for spectral cut-off. With $\tau = 1$ there is less bias but much more variability, especially for Tikhonov regularization. The results for colored noise in fig. 3 are good to mediocre, with a lot of variability for many test cases.

4.2. Transformed discrepancy principle

Motivated by the instability of the discrepancy principle to an incorrect noise level, Raus [118, 119, 62] developed a parameter choice method in a deterministic setting where

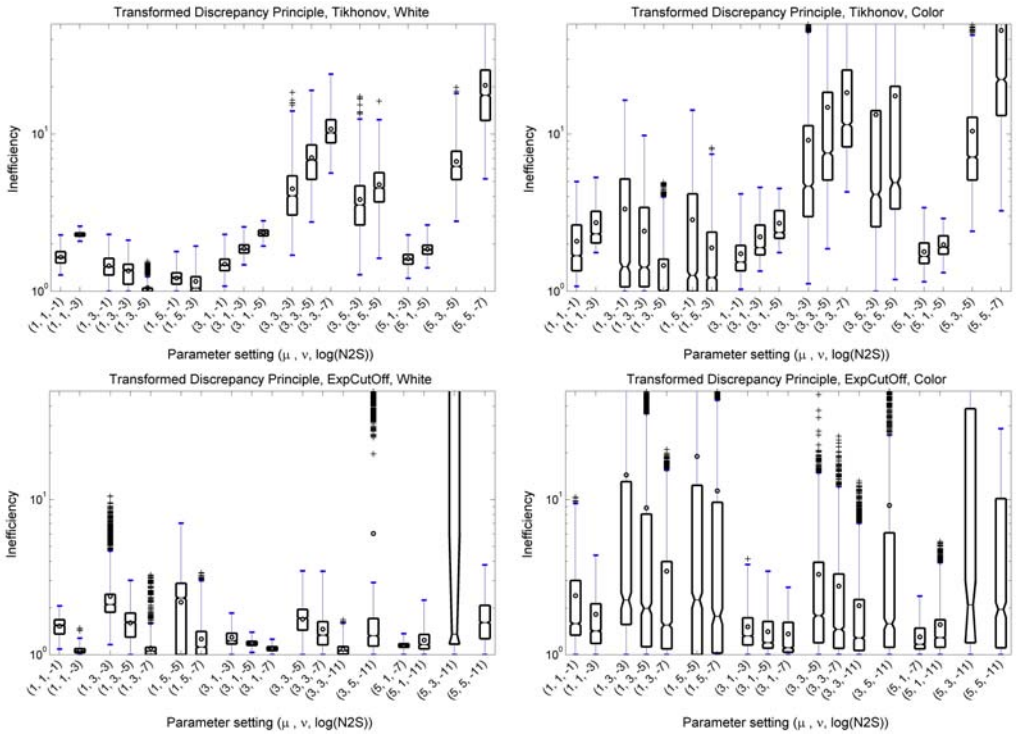

Figure 4: Inefficiencies of the transformed discrepancy principle

the noise level in the data y^δ is known only approximately as $\hat{\delta}$.

Method. For the stochastic case with $y^\delta \in \mathbb{R}^m$, it is assumed that a rough estimate $\hat{\delta}$ of the error standard deviation δ is known. For Tikhonov regularization, one computes n_* as the least integer n for which

$$\|A_n^{-1}(Ax_n^\delta - y^\delta)\| \leq b\hat{\delta}\sqrt{m}/\sqrt{\alpha_n}\,, \tag{15}$$

where $\alpha_n = \alpha_0 q^n$ and b is some constant satisfying $b > \gamma = ((1/4)^{1/4}(3/4)^{3/4})^2 = 0.3248$. We choose $b = 1.5\gamma$. For spectral cut-off, one computes n_* as the least integer n for which

$$\|A^*(Ax_n^\delta - y^\delta)\| \leq b\hat{\delta}\sqrt{m}\;\sigma_{l(n)}\,,$$

where b is some constant satisfying $b > \gamma = 1/2$. We choose $b = 1.5\gamma$.

Known issues. Note that the right-hand side of (15) is an approximate scaled bound of the propagated noise error $\|x_n^0 - x_n^\delta\| \leq \delta/\sqrt{\alpha_n}$. On the left-hand side of (15) is the norm of the residual transformed to the domain space under the approximate inverse A_n^{-1} of A. For this reason, we refer to this parameter choice method as the transformed discrepancy principle.

It was shown in [119] that, for deterministic noise, the method leads to optimal convergence rates when the noise level is known exactly, and it also converges under the assumption that $\|y - y^\delta\| = O(\hat{\delta})$ as $\hat{\delta} \to 0$. Consequently, it is more stable than the discrepancy principle. No knowledge of the solution smoothness is required. The method

was also defined and shown to be convergent for problems where the operator is only known approximately as A^η, where $\|A^\eta - A\| \leq \eta$. Like the discrepancy principle, the method can be applied easily to iterative regularization methods.

Numerics. We tried both $\hat{\delta} = \delta$ and $\hat{\delta} = \delta 10^{0.5-u}$, where u is a uniform pseudo-random variate on $[0, 1]$ (so $\log_{10}(\hat{\delta})$ is only accurate to one figure). In the case of Tikhonov regularization with white noise, the results for $\hat{\delta} = \delta 10^{0.5-u}$ were almost the same as those for $\hat{\delta} = \delta$. Otherwise, the performance of the method using $\hat{\delta} = \delta$ was much better. The results for this exact choice are displayed in fig. 4. Although the method performs well for some test cases, there are substantial differences across the test cases, giving a mediocre performance overall.

4.3. Modified discrepancy principle

The modified discrepancy principle (MD rule) was developed by Raus [116, 117] and by Engl and Gfrerer [40, 47] for Tikhonov regularization and other regularization methods in a continuous, deterministic setting (see also [39] and sections 4.4 and 5.1 in [41]). It was motivated by the desire to find a practical a-posteriori rule that yields optimal convergence rates. The basic idea of the rule is to minimize a bound on the squared error of the regularized solution derived from (6). In [100] the rule was adapted to the discrete, stochastic setting for Tikhonov regularization. The MD rule is also known as the Raus–Gfrerer rule and the minimum bound method.

Method. The MD rule was developed for regularization methods defined using the spectrum of A^*A by $x_\alpha^\delta = g_\alpha(A^*A)A^*y^\delta$, where $\lim_{\alpha\to 0} g_\alpha(\lambda) = 1/\lambda$. This includes Tikhonov regularization, for which $g_\alpha(\lambda) = 1/(\lambda + \alpha)$. For such methods, one can derive from (6) a bound on the squared error of the form

$$\|x - x_\alpha^\delta\|^2 \leq 2(\varphi^2(\alpha, y) + \delta^2 \varrho^2(\alpha)). \tag{16}$$

The minimizer of the bound is defined by the equation

$$f(\alpha, y) := -(\varphi^2)'(\alpha, y)/(\varrho^2)'(\alpha) = \delta^2, \tag{17}$$

and, by using y^δ in place of y, the parameter choice is defined by $f(\alpha, y^\delta) = \delta^2$ or $f(\alpha, y^\delta) = \tau^2\delta^2$ for a tuning parameter τ. To use the rule, we need to be able to compute $(\varphi^2)'(\alpha, y)$ and $(\varrho^2)'(\alpha)$, and this can be done effectively for Tikhonov and other regularization methods (see [40] and section 5.1 in [41]). The MD rule can also be applied to regularization methods with a discrete parameter, including spectral cut-off, with the derivatives above replaced by differences.

For Tikhonov regularization, the function $\varrho^2(\alpha)$ in the bound (16) is $\varrho^2(\alpha) = \alpha^{-1}$, and the parameter choice is defined by

$$\alpha \left| \left\langle Ax_\alpha^\delta - y^\delta, (A^*)^{-1} \frac{dx_\alpha^\delta}{d\alpha} \right\rangle \right|^{1/2} = \alpha^{3/2} \|(AA^* + \alpha I)^{-3/2} y^\delta\| = \tau\delta. \tag{18}$$

Using the discrete set $\{\alpha_n = \alpha_0 q^n\}$, we can approximate the derivative $dx_\alpha^\delta/d\alpha$ on the left-hand side of (18) with $(x_n^\delta - x_{n+1}^\delta)(-\alpha_n \log q)^{-1}$. For spectral cut-off regularization, the term $\delta^2\varrho^2(\alpha)$ in (16) is $\delta^2\sigma_{l(n)}^{-2}$, and the method can be adapted by using differences. Thus, the parameter choice n_* is defined as the first n such that

$$\beta_n \left| \left\langle Ax_n^\delta - y^\delta, (A^*)^{-1} \left(x_n^\delta - x_{n+1}^\delta\right) \right\rangle \right|^{1/2} \leq \tau\delta, \tag{19}$$

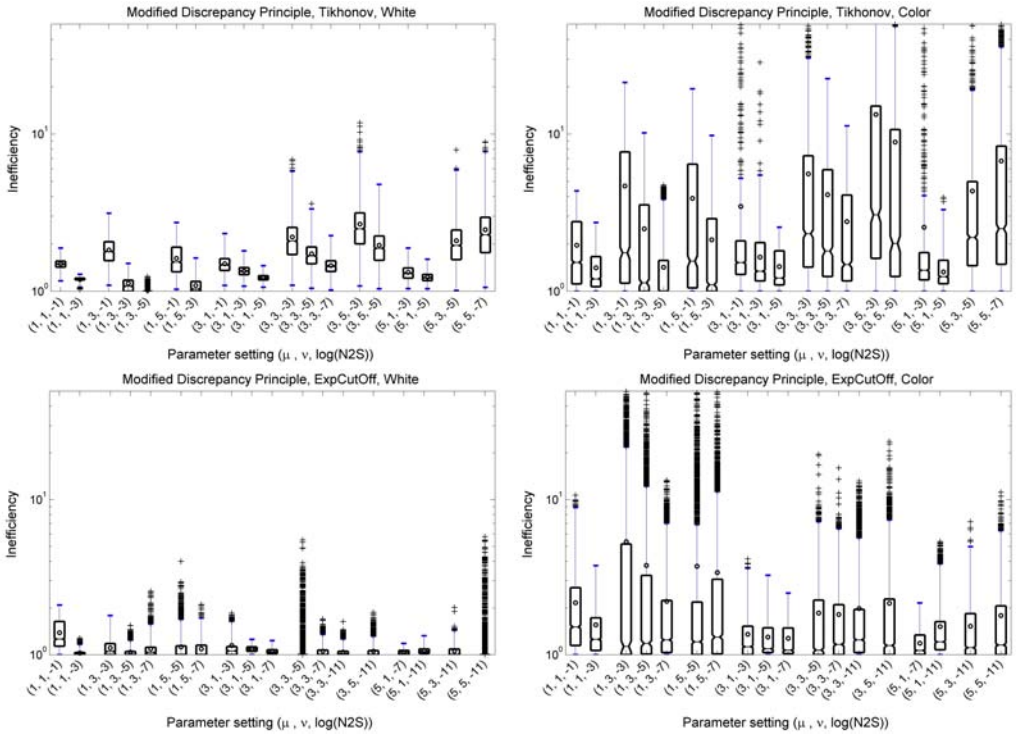

Figure 5: Inefficiencies of the modified discrepancy principle

where

$$\beta_n = \begin{cases} \alpha_n^{1/2}(-\log q)^{-1/2} & \text{for Tikhonov with } \alpha_n = \alpha_0 q^n, \\ (\sigma_{l(n+1)}^{-2} - \sigma_{l(n)}^{-2})^{-1/2} & \text{for spectral cut-off.} \end{cases}$$

For the tuning parameter, we use $\tau = 1.5$ for Tikhonov regularization and $\tau = 0.5$ for spectral cut-off regularization.

Known issues. The MD rule was a significant advance on the discrepancy principle because it achieves the optimal rate of convergence as $\delta \to 0$ for deterministic noise [40, 47], and it does so without any knowledge of the smoothness of the solution x. For Tikhonov regularization, this means that $\|x - x_{n_*}^{\delta}\| = O(\delta^{2/3})$, the best possible rate, if $x \in \mathcal{R}(A^*A)$. The rule also yields optimal rates for finite dimensional implementations [55] and when the operator is only known approximately [110].

The discrete, stochastic version of the MD rule, with a particular tuning constant, is asymptotically (as $m \to \infty$) equivalent to an unbiased risk method [100, 26], and, in expectation, it yields the optimal convergence rate. The unbiased risk method chooses the parameter by minimizing an unbiased (for white noise) estimate of the risk, i.e. the expected squared error. However, in [101] it is shown asymptotically and by simulations that both of these methods are unstable and have high variability; see also [27] for the unbiased risk method. By changing the tuning parameter, it is possible to improve the stability of the MD rule.

The MD rule can also be applied to iterative regularization methods, e.g. Landweber

iteration, for linear problems, and it achieves optimal convergence rates in a deterministic setting [40]. For nonlinear problems, the rule has been extended for Tikhonov regularization and shown to yield optimal rates [83, 127]. In addition, it has been proposed as the stopping rule for the iteratively regularized Gauss–Newton method, where again it yields optimal rates [82].

Numerics. As seen in fig. 5, the MD rule performs well in the white noise situation, especially for spectral cut-off regularization. However, it does not perform so well for colored noise, where in many test cases there is a lot of variability.

4.4. Monotone error rule

The monotone error (ME) rule was proposed in [4, 61, 130] for various regularization methods in a deterministic setting, and it was extensively discussed along with other similar parameter choice rules by Hämarik and Tautenhahn [64, 65]. The rule is based on the observation that, if n is too small (i.e. too much smoothing), then the error $\|x - x_n^\delta\|$ (like the regularization error $\|x - x_n^0\|$) decreases monotonically as n increases.

Method. For continuous regularization methods, the method is formulated by differentiating with respect to the regularization parameter α. The parameter choice α_* is the largest α such that

$$\frac{\left|\left\langle Ax_\alpha^\delta - y^\delta, \frac{d}{d\alpha} A^{*-1} A_\alpha^{-1} y^\delta \right\rangle\right|}{\|\frac{d}{d\alpha} A^{*-1} A_\alpha^{-1} y^\delta\|} \leq \tau\delta \quad \text{with} \quad \tau \geq 1.$$

In order to use it in our framework, we have generated a simple discretized version by replacing the differentials with adjacent differences. Then, in the stochastic setting with $y^\delta \in \mathbb{R}^m$ containing errors of standard deviation δ, the parameter choice n_* is the first n such that

$$\frac{\left|\left\langle Ax_n^\delta - y^\delta, (A^*)^{-1}\left(x_n^\delta - x_{n+1}^\delta\right)\right\rangle\right|}{\|(A^*)^{-1}\left(x_n^\delta - x_{n+1}^\delta\right)\|} \leq \tau\delta\sqrt{m}. \tag{20}$$

We take $\tau = 1.5$ for Tikhonov regularization and $\tau = 0.75$ for spectral cut-off regularization.

Known issues. For Tikhonov regularization (and iterated Tikhonov regularization) in a deterministic setting, the ME rule has some favourable properties [130]. If $\alpha > \alpha_*$, then the error $\|x - x_\alpha^\delta\|$ decreases monotonically as α is decreased, and so $\|x - x_{\alpha_*}^\delta\| < \|x - x_\alpha^\delta\|$, which provides a useful bound for parameter selection. Unlike the discrepancy principle, the ME rule is order optimal for the maximal range of the smoothness index (up to the qualification). In addition, for any noise level δ, it leads to smaller errors than the modified discrepancy principle (for the same tuning parameter). With precisely known δ, the optimal tuning parameter is $\tau = 1$. In the case of spectral cut-off regularization, no optimality results are known.

An alternative discretized version [63] to (20) defines n_* as the first n such that

$$\frac{\left|\left\langle (Ax_n^\delta + Ax_{n+1}^\delta)/2 - y^\delta, (A^*)^{-1}\left(x_n^\delta - x_{n+1}^\delta\right)\right\rangle\right|}{\|(A^*)^{-1}\left(x_n^\delta - x_{n+1}^\delta\right)\|} \leq \tau\delta\sqrt{m}.$$

This version has the advantage that, like the continuous version, the error in x_n^δ decreases monotonically as n is increased for $n < n_*$ [63].

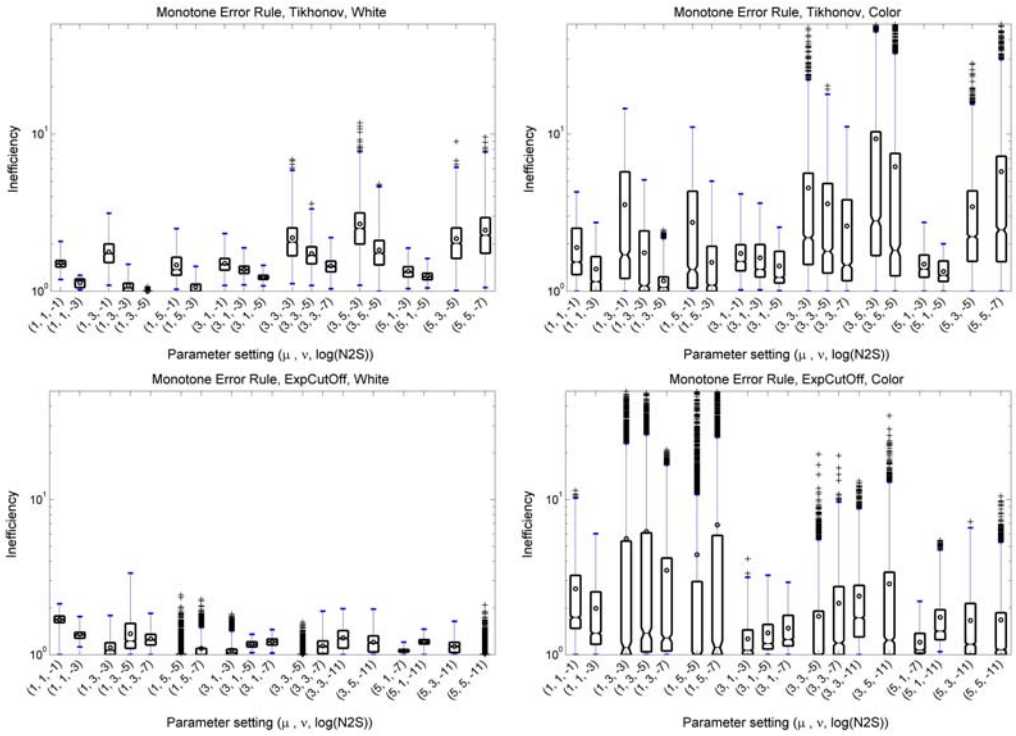

Figure 6: Inefficiencies of the monotone error rule

The ME rule can also be applied to iterative regularization methods, in particular Landweber iteration, for which it is order optimal [64, 65].

Numerics. Figure 6 shows that the ME rule has mostly good to acceptable performance for all cases, though there is substantial variation for some cases in the colored noise scenario. The performance for Tikhonov regularization is slightly better than that of the modified discrepancy principle in fig. 5, which is consistent with the theory.

4.5. Balancing principle

The balancing principle, due to Lepskij [94], was originally derived for statistical estimation from direct observations in a white noise model. Since then it has been developed further for regularization of linear inverse problems [50, 135, 105, 106, 18] and nonlinear inverse problems [12, 13] in deterministic and stochastic settings. The notation we will use is taken from [9]. The principle aims to balance the known propagated noise error bound $\delta\varrho(n)$ in (8) (a monotonic function of the parameter) with the unknown regularization error (7) (which depends on x), by an adaptive procedure that employs a collection of differences of regularized solutions.

Method. One needs all regularized solutions $\{x_n^\delta\}_{n\le N}$ up to a certain maximal index N, which is also a required input. Furthermore, an upper bound $\delta\varrho(n)$ for the propagated noise error $\|x_n^0 - x_n^\delta\|$ (with correct noise level δ) is required; bounds are known for spectral cut-off and Tikhonov regularization (see after (8)). The noise can be either deterministic

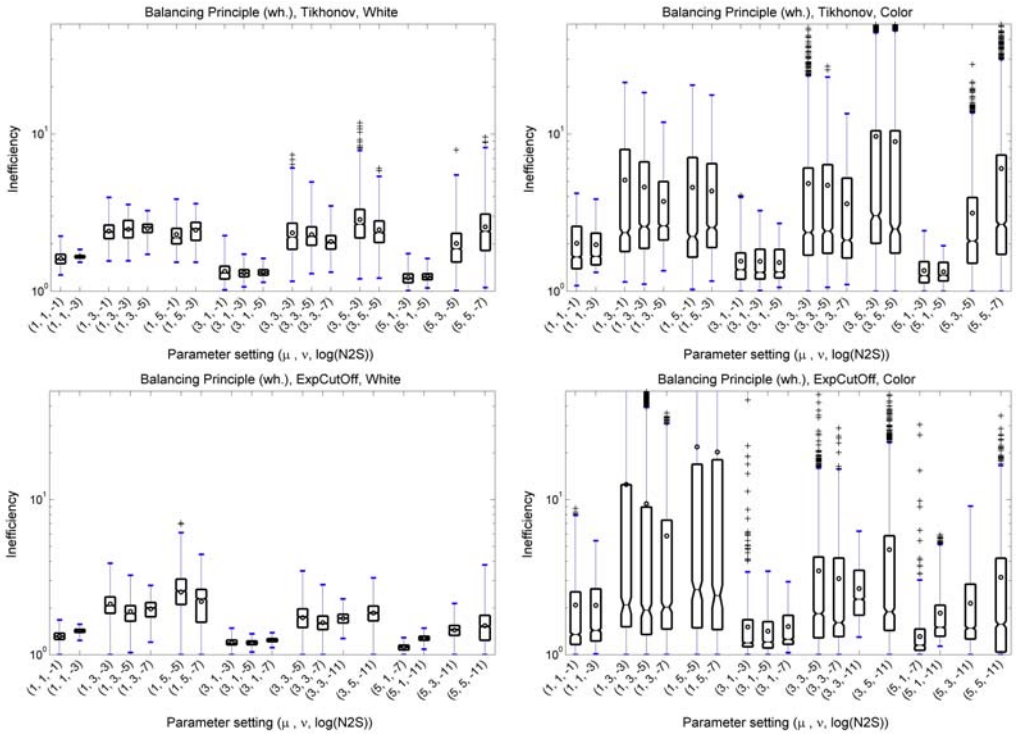

Figure 7: Inefficiencies of the balancing principle (white) using known δ

or stochastic white or colored. In the stochastic setting, a bound or estimate $\delta^2\varrho^2(n)$ of the variance $\mathbb{E}\|x_n^0 - x_n^\delta\|^2$ is required. If the noise covariance is known, then for $\delta^2\varrho^2(n)$ one can use a known expression of the variance. Alternatively, if one has two or more independent sets of data y_i^δ, then $\mathbb{E}\|x_n^0 - x_n^\delta\|^2$ can be estimated by $2^{-1}\text{Mean}\{\|x_{n,i}^\delta - x_{n,j}^\delta\|^2, i \neq j\}$ [9, 10]. In many situations, two sets of data are sufficient.

Define the balancing functional by

$$b(n) = \max_{n<k\leq N} \left\{4^{-1}\|x_n - x_k\|/(\delta\varrho(k))\right\}. \tag{21}$$

The smoothed balancing functional (which is monotonically decreasing) is defined as

$$B(n) = \max_{n\leq k\leq N} \{b(k)\}. \tag{22}$$

Then the parameter choice n_* is the first n such that

$$B(n) \leq \kappa.$$

In our implementation we take $\kappa = 1$. We will consider two input scenarios. Firstly, to address the situation of one data set with δ as the only known property of the noise, we use the known expressions (11) and (12) of $\delta^2\varrho^2(n)$ for the white noise model; we refer to this method as the balancing principle (white). Secondly, for $\delta^2\varrho^2(n)$ we use the

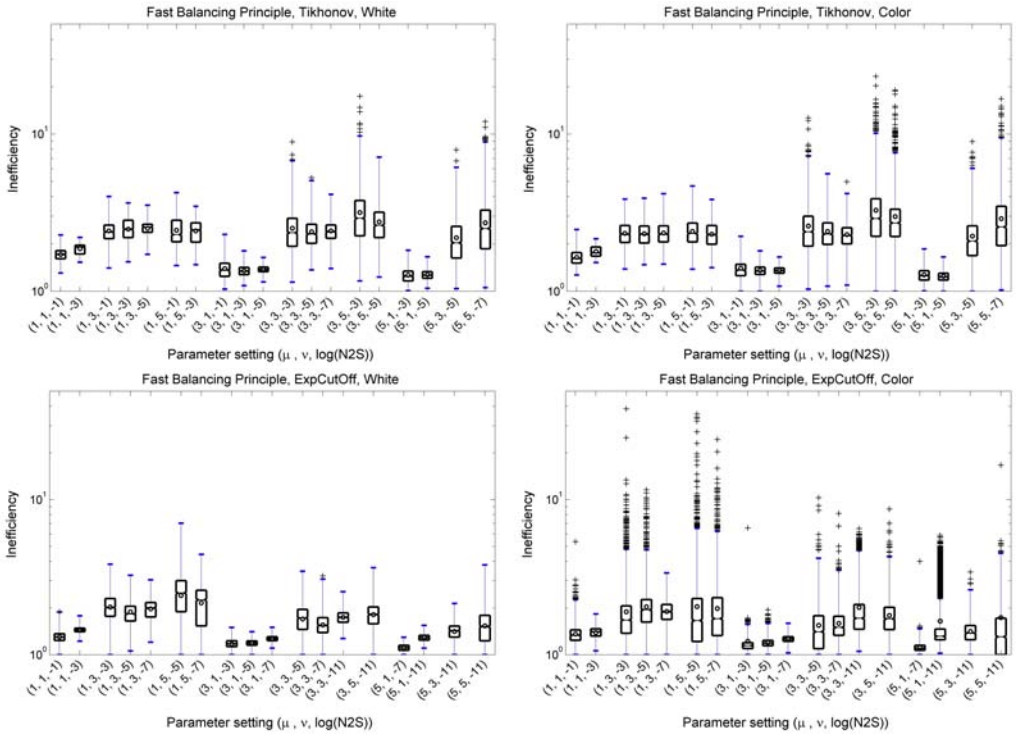

Figure 8: Inefficiencies of the fast balancing principle using four independent data sets

estimate $2^{-1}\text{Mean}\{\|x^\delta_{n,i} - x^\delta_{n,j}\|^2, i \neq j\}$ with four independent sets of data y^δ_i. In each situation, we use the maximal index N in (13).

Known issues. The balancing principle is one of the few parameter choice methods for which oracle inequalities for the error are known [120, 13], i.e. there are stronger results than rates of convergence alone. In particular, the balancing principle does not have a saturation problem, and, for stochastic noise, it is optimal up to at most a $\log(1/\delta)$ factor (though such a logarithmic behavior has not been observed in practice).

According to the theory, the constant κ should be 1 in the deterministic setting [105]. In the stochastic setting with continuous Gaussian white noise, κ should be $O((\log \delta^{-1})^{1/2})$ for mildly ill-posed problems and $O((\log\log \delta^{-1})^{1/2})$ for severely ill-posed problems [18]. The constant κ acts as a stability parameter. The method is very stable when κ is chosen sufficiently large, but it can be quite unstable if κ is chosen too small. A choice of $\kappa \in [0.5, 1.5]$ appears to give good results, independent of the noise situation and the inverse problem, a view that is also supported by the results in [10].

Instead of the balancing functional in (21), one can use the modified version

$$b(n) = \max_{n<k\leq \ell(n)} \left\{4^{-1}\|x_n - x_k\|/(\delta\varrho(k))\right\}, \tag{23}$$

where $\ell(n)$ is an increasing sequence for which $\varrho(\ell(n))/\varrho(n) \geq \beta > 1$ for some constant β. This definition requires fewer evaluation steps. It also gives a convergent method [17, 105] and performs well in practice. Our experience is that, provided β is big enough so that

$\ell(n) - n \geq 4$, the results are not distinguishable from those of the original version. In particular, in the experiments reported here, we used $\ell(n)$ defined by $\varrho(\ell(n))/\varrho(n) = 2$.

A related method, called the fast balancing principle [11], defines n_* as the first n such that

$$b(n) \leq \kappa, \tag{24}$$

where $b(n)$ is defined in (23). In a completely Bayesian setting for spectral cut-off [10] and in a restricted setting for Tikhonov regularization [11], it is known that this is an optimal method, which even theoretically does not have a $\log(1/\delta)$ factor in the error. Additionally, this method has the advantage of requiring a much lower number of computations than the original balancing principle. In fact, the computational cost is of the same order as that of the discrepancy principle. For the fast balancing principle with $\ell(n) = n+1$ in (23), optimality results are known [63] for Tikhonov and other regularization methods with deterministic noise, given either the actual or an approximate noise level.

For a white noise model, instead of using several independent data sets or the analytic expressions (11) and (12) of $\delta^2\varrho^2(n)$, one can use the Monte-Carlo estimate

$$\varrho^2(n) \approx \text{Mean}\|A_n^{-1}\hat{\xi}_i\|^2, \tag{25}$$

where $\hat{\xi}_i$ are synthetic vectors of independent standard normal pseudo-random variates, together with a known or estimated value of the noise variance δ^2. The same estimate applies to any regularization method with linear regularization operator A_n^{-1}.

Numerics. In all our experiments, the results for the balancing principle and the fast balancing principle with the same input are visually impossible to distinguish. Figure 7 displays the results for the balancing principle (white), which uses the known expressions (11) and (12) of $\delta^2\varrho^2(n)$ for the white noise model, so there is no adaptation to correlated errors (colored noise). In fig. 8 we display the results for the fast balancing principle using an estimate of $\delta^2\varrho^2(n)$ obtained from four independent data sets, which automatically adapts to the colored noise.

The results in figures 7 and 8 in the white noise situation are almost identical for both Tikhonov and spectral cut-off regularization; the method is stable and performs quite well, though in several cases (where $\nu \geq 3$) the estimates are somewhat biased. In fig. 8, the results for colored noise are quite similar to those for white noise, since here the method adapts to the noise color. By contrast, in fig. 7 there is considerable variability in the colored noise scenario. Interestingly, the results for spectral cut-off in fig. 7 are similar to those of the discrepancy principle in fig. 3.

4.6. Hardened balancing principle

The hardened balancing principle is a modified version of the balancing principle in the stochastic setting, and was first proposed in [9]. It uses the same input and computational steps as the original balancing principle, and so it can be computed parallel to this one. It has the advantage that it does not require a tuning parameter.

Method. One needs all regularized solutions $\{x_n^\delta\}_{n\leq N}$ up to a certain maximal index N, which is also a required input. Furthermore, an expression or approximation of the scaled variance $\varrho^2(n) = \delta^{-2}\mathbb{E}\|x_n^0 - x_n^\delta\|^2$, or any scalar multiple of this, is required. As for the balancing principle, an estimate of $\delta^2\varrho^2(n)$ can be obtained from two or more independent sets of data.

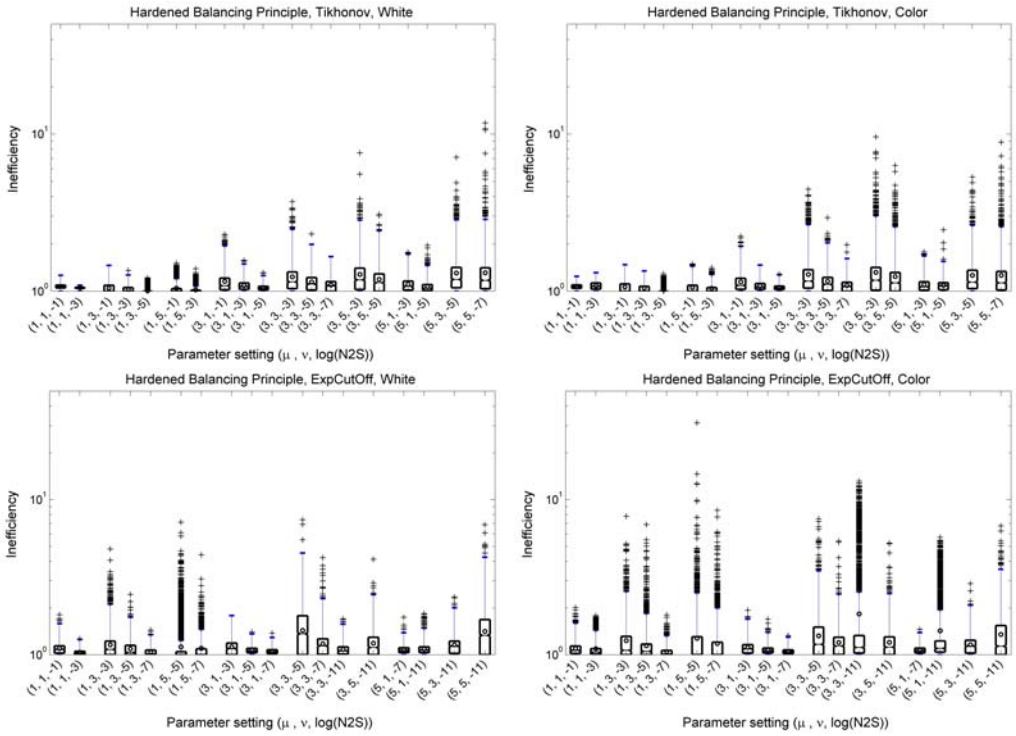

Figure 9: Inefficiencies of the hardened balancing principle using four independent data sets

Define the balancing functional $b(n)$ and smoothed balancing functional $B(n)$ as in (21) and (22), respectively. The parameter choice is

$$n_* = \operatorname*{argmin}_{n \le N} \left\{ B(n)\sqrt{\varrho(n)} \right\}.$$

Obviously, the same choice is obtained if any scalar multiple of $\varrho(n)$ is used in its place. We use the same two inputs for (scaled) $\varrho(n)$ as for the balancing principle. First we estimate $\delta^2\varrho^2(n)$ using $2^{-1}\text{Mean}\{\|x^\delta_{n,i} - x^\delta_{n,j}\|^2, i \neq j\}$ with four independent sets of data y^δ_i. Secondly, to address the situation of one data set with no knowledge about the noise, we use the known expressions (11) and (12) of $\delta^2\varrho^2(n)$ for the white noise model; we refer to this method as the hardened balancing principle (white). In each situation, we use the maximal index N in (13).

For better computational efficiency, as for the balancing principle, one can use the modified version of the balancing functional defined in (23) instead of (21). In our experiments, we used the modified version with $\ell(n)$ defined by $\varrho(\ell(n))/\varrho(n) = 2$.

Known issues. For spectral cut-off regularization, a version of the hardened balancing principle has been analyzed [19] in a Bayesian framework, with stochastic x and noise, and sequence data. It is shown that the parameter choice satisfies an oracle inequality and is optimal (up to a constant independent of the noise level) with respect to the risk for some error moment. As a consequence, the method is not prone to saturation.

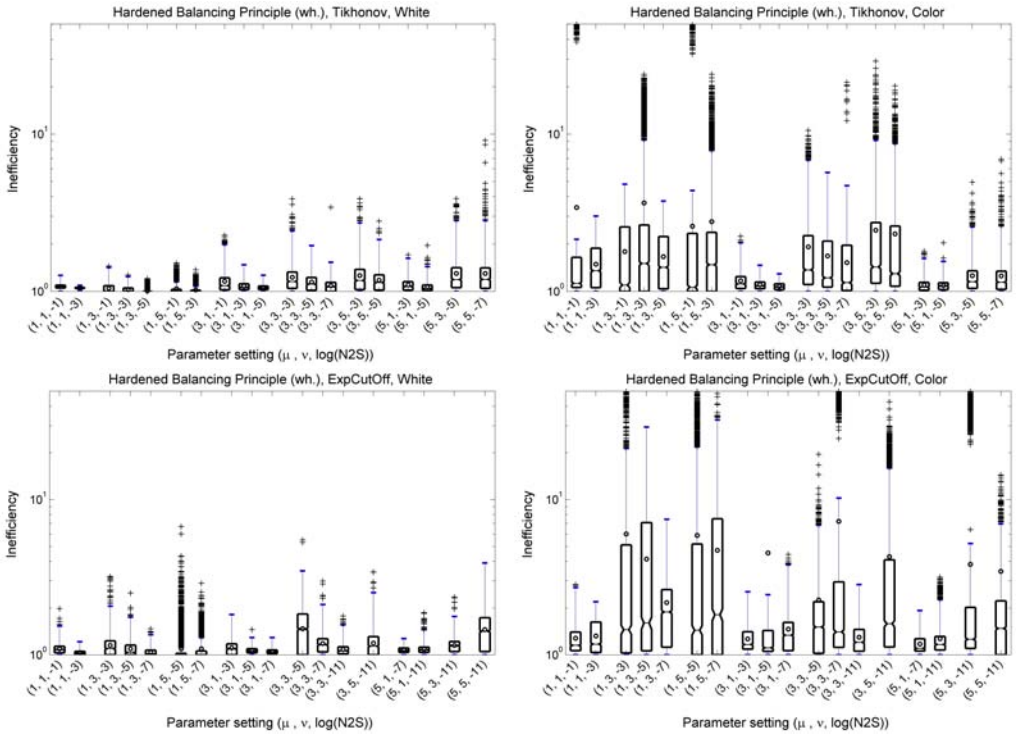

Figure 10: Inefficiencies of the hardened balancing principle (white)

Because the basis of the proofs is the same as for the quasi-optimality criterion, it is very likely that similar results hold for Tikhonov regularization, even with deterministic source conditions [14].

Numerical experiments in [9, 19] indicate that the method is very stable. Some care needs to be taken when the Tikhonov regularization parameter becomes smaller than the smallest eigenvalue of A^*A. However, this situation can be easily (and automatically) detected in practice by using a suitable maximal index, in particular N in (13).

All remarks in the subsection on the balancing principle concerning the choice of $\varrho(n)$ also hold true for the hardened balancing principle. In particular, for the white noise model, one can use the Monte-Carlo estimate (25) of $\varrho^2(n)$.

Numerics. Figure 9 shows that the hardened balancing principle using four independent data sets for variance estimation is extremely stable for both white and colored noise, and it has excellent inefficiency values in all the test cases for both Tikhonov and spectral cut-off regularization.

Figure 10 displays the results for the hardened balancing principle (white), which uses the variance expressions (11) and (12) for the white noise model. As expected, it performs very well in the white noise situation, with results that are almost identically to those in fig. 9. In the colored noise situation, there is much more variability but the performance is still quite good, especially in the cases where x is less smooth.

4.7. Quasi-optimality criterion

The quasi-optimality criterion [77, 133] is one of the oldest and simplest available parameter choice methods. It was originally introduced by Tikhonov and Glasko [134] and became better known from the continuous version proposed by Tikhonov and Arsenin [133]. A good overview of the method and its history can be found in [15].

Method. The parameter choice n_* is defined simply as

$$n_* = \operatorname*{argmin}_{n \leq N} \|x_n^\delta - x_{n+1}^\delta\|. \tag{26}$$

The continuous version for Tikhonov regularization defines the parameter choice by

$$\alpha_* = \operatorname{argmin} \left\| \alpha \frac{dx_\alpha^\delta}{d\alpha} \right\|.$$

Using a difference quotient in place of the derivative for the discrete parameters $\alpha_n = \alpha_0 q^n$, it is clear that these versions are consistent.

Known issues. For a discrete set of regularization parameters, the use of a suitable maximal index is essential. This is because the method is based on a discrete evaluation of a differential, and hence is very sensitive to a situation where the regularization operators A_n^{-1} are formally different, but are practically the same due to the finiteness of the considered problem. This happens especially when the regularization parameter α_n for Tikhonov regularization falls below the smallest eigenvalue of the operator A^*A. This issue does not apply to infinite dimensional inverse problems (Case C1) and, therefore, is rarely considered in the inverse problems literature. In order for the quasi-optimality criterion to be successful for spectral cut-off regularization, the cut-off points $l(n)$ need to be chosen carefully and far enough apart [19].

The noise can be stochastic (white or colored) or deterministic, but more care is advised in the deterministic setting because of the conditions on the error [14].

Convergence of Tikhonov regularization with the quasi-optimality criterion was shown in [93] in the discrete (i.e. not ill-posed, but only ill-conditioned) case with deterministic noise. The question of convergence in infinite-dimensional spaces was discussed in [49], where abstract conditions were given.

Although it has been used successfully in a number of practical situations [115], the first more concrete proofs for the quasi-optimality criterion were provided only quite recently - in [19] for spectral cut-off regularization in a Bayesian setting, and in [14] for Tikhonov regularization with either deterministic or stochastic noise. In both papers, oracle inequalities are derived under certain conditions, which show that the method is near-optimal and is not prone to saturation. One of the conditions is that the noise has weight in all frequency components (so, e.g., it is not band-limited), which is usually true in practice. A more theoretical analysis, including convergence properties, in some deterministic settings has been derived in [88, 111].

The method is not as stable as the hardened balancing principle, as there is no compensation for statistical outliers. On the other hand, no information about the problem needs to be known.

For iterative regularization methods, the situation regarding the quasi-optimality criterion is still quite unclear, and, for nonlinear ones, it is completely unclear. There are

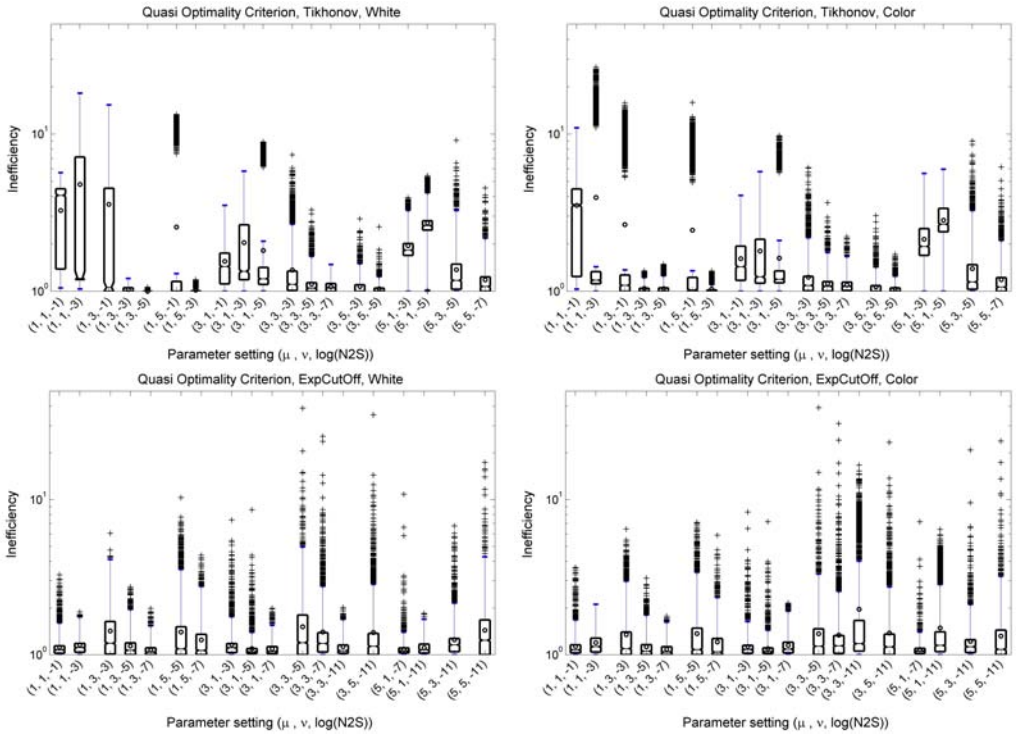

Figure 11: Inefficiencies of the quasi-optimality criterion

some convergence results [88, 111], but no oracle inequalities are known. For Landweber iteration (where A_n^{-1} represents n Landweber steps), the modified criterion defined by

$$n_* = \operatorname*{argmin}_{n \le N} \|x_n^\delta - A_n^{-1} A x_n^\delta\| \tag{27}$$

was analyzed in [111]. In order to see the similarity to the quasi-optimality criterion (26), consider the choice

$$n_* = \operatorname*{argmin}_{n \le N} \|x_{n+1}^\delta - A_n^{-1} A x_{n+1}^\delta\|.$$

This should numerically behave almost the same as (27) due to the very slow convergence of Landweber iteration, and it coincides exactly with the quasi-optimality criterion in the spectral cut-off case.

Numerics. As seen in fig. 11, the quasi-optimality criterion performs very well for spectral cut-off regularization with either white or colored noise. It is less stable and less consistent across the test cases for Tikhonov regularization. However, keeping in mind that it is computationally the simplest of the available methods and that it does not require the noise level, the general performance is remarkable.

4.8. L-curve method

The L-curve method, proposed by Hansen [69, 71, 75], is based on the long-known fact that a log-log parametric plot of $(\|Ax_n^\delta - y^\delta\|, \|x_n^\delta\|)$ often has a distinct L-shape [92].

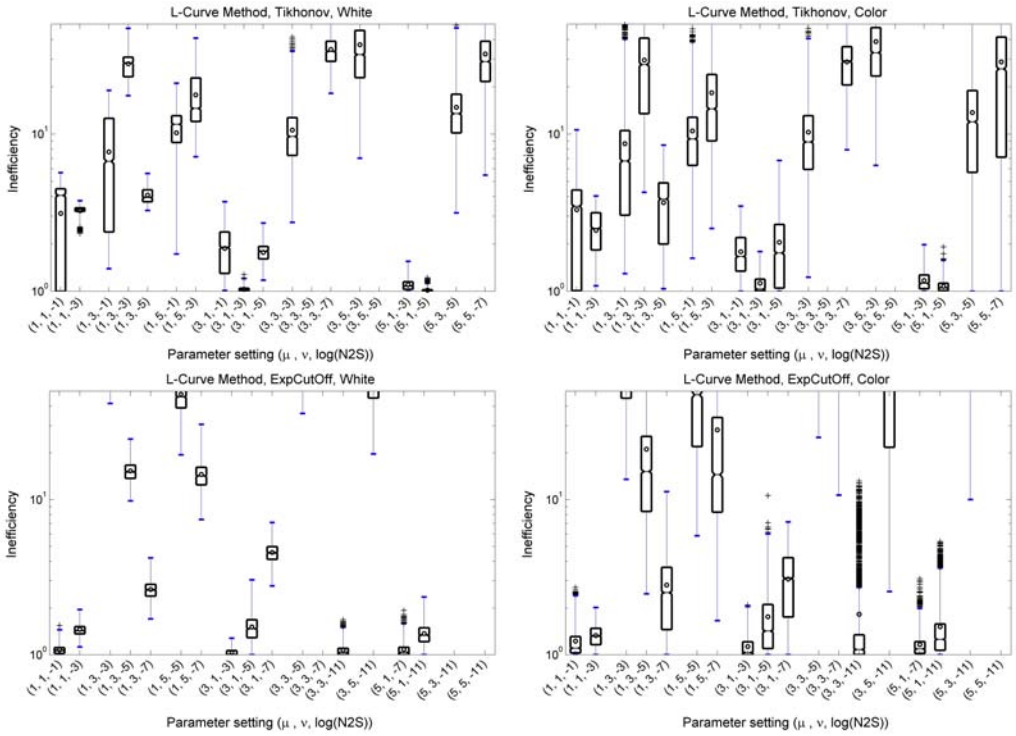

Figure 12: Inefficiencies of the L-curve method

Points on the vertical part correspond to large n (under-smoothed solutions) and those on the horizontal part correspond to small n (over-smoothed solutions), which suggests that the "corner point" of the L-curve should define a good value of the parameter n. Due to its simplicity and intuitive appeal, the method became popular in a number of application areas.

Method. There are several variants of the method; here we use the version proposed by Reginska [122, 41]. The parameter choice is

$$n_* = \operatorname*{argmin}_{n \le N} \left\{ \|Ax_n^\delta - y^\delta\| \|x_n^\delta\| \right\}.$$

This defines the "corner point" as the point on the L-curve where the "tangent" has slope -1 [122]. The method can be extended by minimizing $\|Ax_n^\delta - y^\delta\| \|x_n^\delta\|^\tau$ [122], where τ is a tuning constant, but we will not consider this here.

Known issues. One of the problems with the L-curve approach is that the "corner point" is not a well-defined notion, and several algorithms have evolved with different definitions. The first algorithm for Tikhonov regularization used the maximum curvature of the L-curve [75], which can be computed efficiently as in [23]. The algorithm above [122] is a simpler alternative, which can also be applied with discrete regularization parameters. Our experience (see also [84]) is that these algorithms have similar performance. Other algorithms are given in [73] and the references therein.

There is currently no rigorous justification of the L-curve method, but there is quite

a lot of numerical experience. For several (but not all) problems it has been observed to give a reasonably good and robust parameter choice, and it can cope with correlated errors [1, 3, 29, 67, 71, 73, 75].

However, it is known theoretically that the L-curve method (from [75]) has serious limitations [41, 146]. First, the L-curve corner may not even exist [122]. Also, the method is not convergent in the deterministic setting as the noise level $\delta \to 0$, with significant under-smoothing for smooth solutions [66]. Furthermore, the method is not convergent in the stochastic discrete data setting as the sample size goes to ∞, where it leads to over-smoothing [138, 139]. These effects have been observed in numerical experiments [72, 84].

The L-curve method can also be extended to iterative regularization methods, in particular the conjugate gradient method [67, 71], Krylov methods [87] and the iteratively regularized Gauss–Newton method [45].

Numerics. As seen in fig. 12, for both Tikhonov and spectral cut-off regularization, the L-curve method has a very unreliable and erratic performance over the set of problems considered. It performs well in the few cases where ν is small relative to μ, but very poorly in other cases where ν is larger (the solution is smooth) and the noise level is small, consistent with the theoretical result in [66]. The behavior for colored noise quite closely resembles the behavior for white noise.

If the L-curve method is used without a maximal index (i.e. N in (13)), then the results for the test cases with good performance in fig. 12 are unchanged, but the results for most of the other cases are significantly worse, with much larger medians and means for the inefficiencies. This shows the importance of using an appropriate maximal index for the L-curve method, a recommendation also made in [71, Section 7.7].

4.9. Modified discrepancy partner rule

In [68], Hanke and Raus developed a general approach to construct a data-driven rule (requiring no a-priori knowledge) from an order-optimal rule (that requires the noise level δ) in the deterministic setting (see also [60]). The approach uses a bound on the function defining the order-optimal rule to implicitly bound the error in the regularized solution. In particular, they applied the approach to the MD rule for Tikhonov regularization in Section 4.3 to get the modified discrepancy partner (MDP) rule.

Method. As for the MD rule, one has to compute the function f in equation (17). The MDP rule also uses the function ϱ in the bound (8), but this is already required in the derivation of f. For Tikhonov regularization with parameter $\alpha_n = \alpha_0 q^n$, we have $\varrho(n) = \alpha_n^{-1/2}$, and, for spectral cut-off regularization, we have $\varrho(n) = \sigma_{l(n)}^{-1}$. The parameter choice n_* is the minimizer of $\varrho(n)\eta(n)$, where $\eta(n) = (f(n))^{1/2}$ is the left-hand side of (19), that is

$$n_* = \operatorname*{argmin}_{n \le N} \{ \varrho(n)\beta_n \left| \left\langle A x_n^\delta - y^\delta, (A^*)^{-1} \left(x_n^\delta - x_{n+1}^\delta \right) \right\rangle \right|^{1/2} \}. \tag{28}$$

Known issues. For Tikhonov regularization, Hanke and Raus [68] derive a bound for the error which shows that, if $\eta(n_*)$ is of the same order as $\|y^\delta - y\|$, then the MDP rule achieves the optimal rate of convergence as $\delta \to 0$. They suggest that the value of $\eta(n_*)$ be monitored, and, if it is significantly less than the presumed noise level, then the parameter choice n_* should be rejected. In numerical experiments for Tikhonov

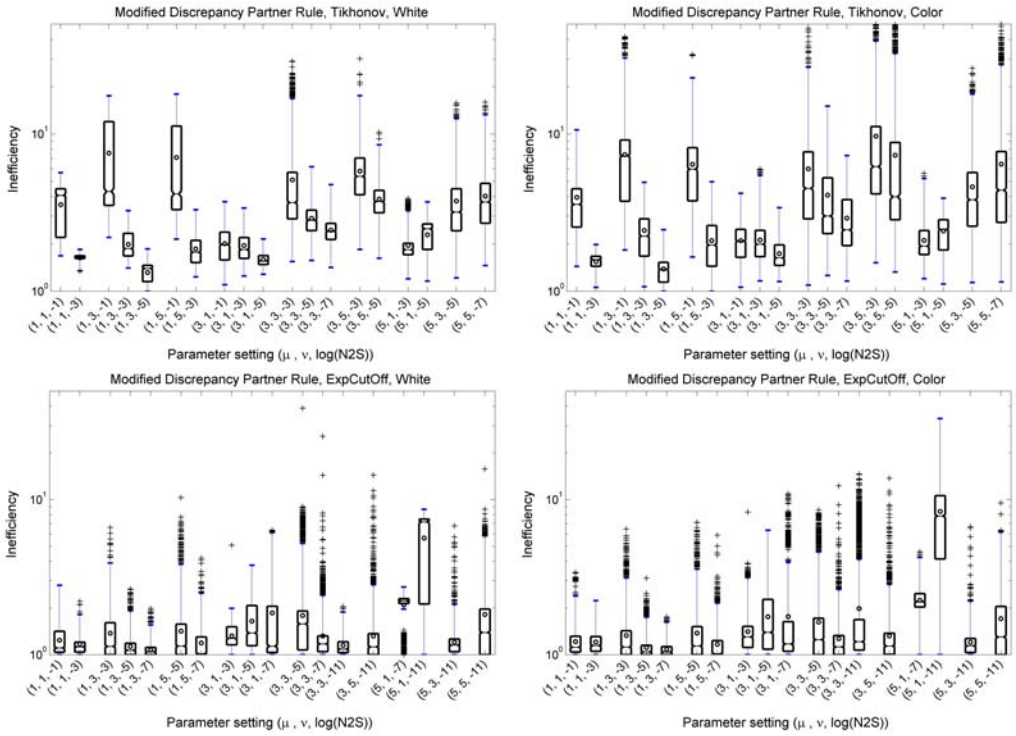

Figure 13: Inefficiencies of the modified discrepancy partner rule

regularization in [68], the MDP rule gives errors that, although greater, are less than twice that of the underlying MD rule. It is observed, however, that the MDP rule is not convergent (over-smoothing) in the discrete case as the sample size goes to ∞.

It is shown in [60, 68] that the proposed approach is quite general, and it can be applied with order-optimal rules for other regularization methods for linear problems, in particular Landweber iteration and the method of conjugate gradients, to give useful data-driven rules.

Numerics. As seen in fig. 13, the MDP rule has good performance for spectral cut-off in both the white and colored noise situations. For Tikhonov regularization, it has only mediocre performance, which is worse but not very much worse than for the MD rule in fig. 5. The behavior for colored noise is quite similar to that for white noise. If the MDP rule is applied without a maximal index, the results are similar to those in fig. 13, except for Tikhonov regularization with $\mu = 1$, where the results are much worse.

4.10. Extrapolated error method

This method, developed by Brezinski, Rodriguez and Seatzu [20, 21] for discrete ill-posed problems, chooses the regularization parameter by minimizing an estimate of the 2-norm error $\|x - x_n^\delta\|$ found by an extrapolation procedure.

Method. The parameter choice is

$$n_* = \operatorname{argmin}\{\|r_n\|^2/\|A^* r_n\|\}, \quad r_n = y^\delta - Ax_n^\delta. \tag{29}$$

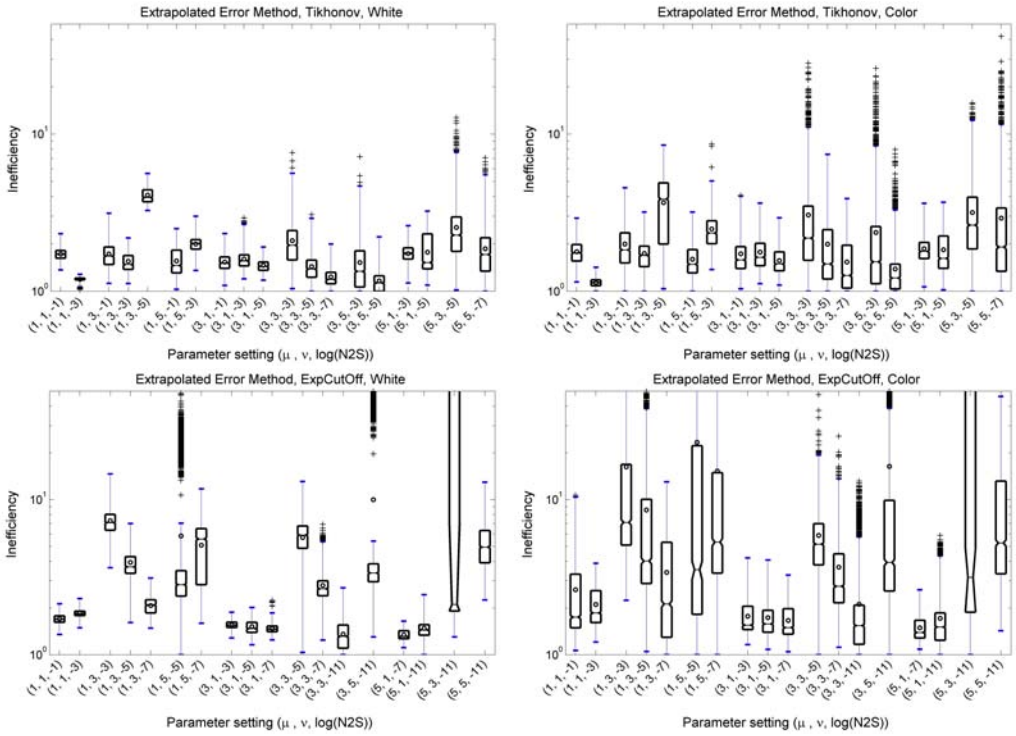

Figure 14: Inefficiencies of the extrapolated error method

Known issues. The expression $\|r_n\|^2/\|A^*r_n\|$ in (29) is one of a family of extrapolated estimates of the error $\|x-x_n^\delta\|$ [20, 21]. In practice, to calculate the expression accurately for small α_n in Tikhonov regularization, one should replace A^*r_n by the equivalent term $\alpha_n x_n^\delta$ [20]. There is currently no convergence analysis to justify the method, but numerical results in [20] indicate that it is quite robust.

Numerics. As seen in fig. 14, the extrapolated error method performs quite well for Tikhonov regularization, but poorly in many cases for spectral cut-off. If the method is used without a maximal index, the results for Tikhonov regularization are much worse. We also assessed a related method proposed in [21], which has $\|r_n\|\|A^*r_n\|/\|AA^*r_n\|$ in place of the expression in (29), but overall it gave significantly worse results than the method above.

4.11. Normalized cumulative periodogram method

For the finite dimensional situation with white noise, Rust [124] suggested using the periodogram of the residual vector as a diagnostic tool. Building on this, Hansen et. al. [74] developed a parameter choice method, called the normalized cumulative periodogram (NCP) method, and Rust and O'Leary [125] proposed a similar method. The basis of these methods is to make the residual as close as possible to white noise.

Method. The (unscaled) periodogram of the residual vector $r_n = y^\delta - Ax_n^\delta$ is defined as the vector p with elements $p_k = |\mathrm{dft}(r_n)|^2$, $k = 1, \ldots, m$, where dft denotes the

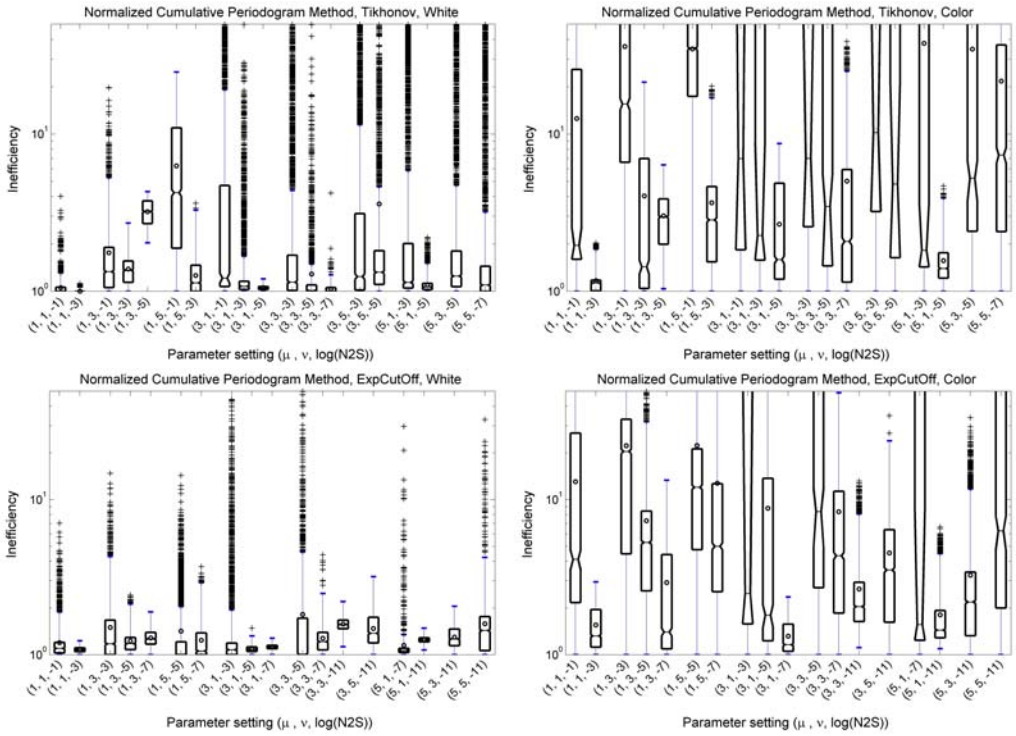

Figure 15: Inefficiencies of the NCP method

discrete Fourier transform. Define the normalized cumulative periodogram as the vector $c = c(r_n)$ with elements

$$c_i = \|(p_2, \ldots, p_{i+1})\|_1 / \|(p_2, \ldots, p_m)\|_1, \quad i = 1, \ldots, m-1,$$

where $\|\cdot\|_1$ is the l_1 norm, and let v be the vector with elements $v_i = i/(m-1)$. Then the NCP parameter choice is

$$n_* = \underset{n \le N}{\operatorname{argmin}} \|v - c(r_n)\|_1.$$

Known issues. There is currently no convergence analysis of the NCP method. As shown in [74], the underlying assumption that the residual resembles white noise holds only approximately for a range of values of the Tikhonov regularization parameter α that are not too small, so it is likely that there are limitations for the method. However, the NCP method and its variant in [125] have been shown to perform well on several test problems [74, 125]. It is clear from the basis of the NCP method, that the noise should be white. In principle, the method can also be applied to iterative regularization methods.

Numerics. Figure 15 shows that the NCP method has quite good performance in the white noise situation, though there is evidence of saturation for Tikhonov regularization, as well as many outliers. By contrast, the method has poor performance in the colored

noise scenario for both Tikhonov and spectral cut-off regularization. If the NCP method is used without a maximal index, many of the extreme outliers are even more extreme, resulting in larger means for the inefficiencies.

4.12. Residual method

The residual method was introduced in [16] for spectral cut-off regularization in an infinite dimensional Bayesian setting; see also [9]. The method is based on minimizing a certain weighted form of the norms of the residuals, where the weighting penalizes under-smoothing parameter values.

Method. Let $B = A(I - AA_n^{-1})$. The parameter choice is defined by

$$n_* = \operatorname*{argmin}_{n \le N} \left\{ \frac{\|Ax_n^\delta - y^\delta\|}{(\operatorname{trace} B^*B)^{1/4}} \right\}. \tag{30}$$

Known issues. Assuming an appropriate behavior for the random element x, the residual method for spectral cut-off is a convergent method as $\delta \to 0$ [16], and this holds with noise of unknown moderate color. The rate results show that the performance improves as the degree of ill-posedness of the problem increases, though it is not optimal.

It is clear from the definition (30) that this choice is similar in form to generalized cross-validation (32). For spectral cut-off regularization, the trace can be evaluated easily. For Tikhonov regularization, computation of the trace might be rather expensive if m is large, but more efficient stochastic trace estimators can be used [48, 78, 52].

Numerics. Figure 16 shows that the residual method has good performance in most cases for both Tikhonov and spectral cut-off regularization, though it gives poor results for some more well-posed cases ($\mu = 1$). Consistent with the theory, the method has better performance when the problem is more ill-posed, and in these cases it performs quite well for both white and colored noise. If the residual method is used without a maximal index, the results for $\mu = 1$ are much worse.

4.13. Generalized maximum likelihood

As discussed in Section 2, the Tikhonov regularized solution for discrete data with independent Gaussian errors can be interpreted as a Bayes estimate of x if x is endowed with the prior of a certain zero mean Gaussian stochastic process [43, 142]. Using this interpretation, Wahba [141] derived the generalized maximum likelihood (GML) estimate (see also [5, 33, 145]). In the case where $A : \mathbb{R}^m \to \mathbb{R}^m$ has full rank and the Euclidean norm is used for regularization, the GML estimate (which is then an ordinary maximum likelihood estimate) is based on $y \sim \mathcal{N}(0, b(AA^* + \lambda I))$ for a constant b.

Method. The GML parameter estimate is defined by

$$n_* = \operatorname*{argmin}_{n \le N} \left\{ \frac{\|Ax_n^\delta - y^\delta\|^2}{(\det{}^+ (I - AA_n^{-1}))^{1/m_1}} \right\}, \tag{31}$$

where $m_1 = \operatorname{rank}(I - AA_n^{-1})$ and $\det^+$ is the product of the nonzero eigenvalues.

Known issues. For spectral cut-off regularization, clearly $\det^+ (I - AA_n^{-1}) = 1$, so the GML function is just the residual sum of squares. Therefore, the GML method is not suitable since it will always severely under-smooth.

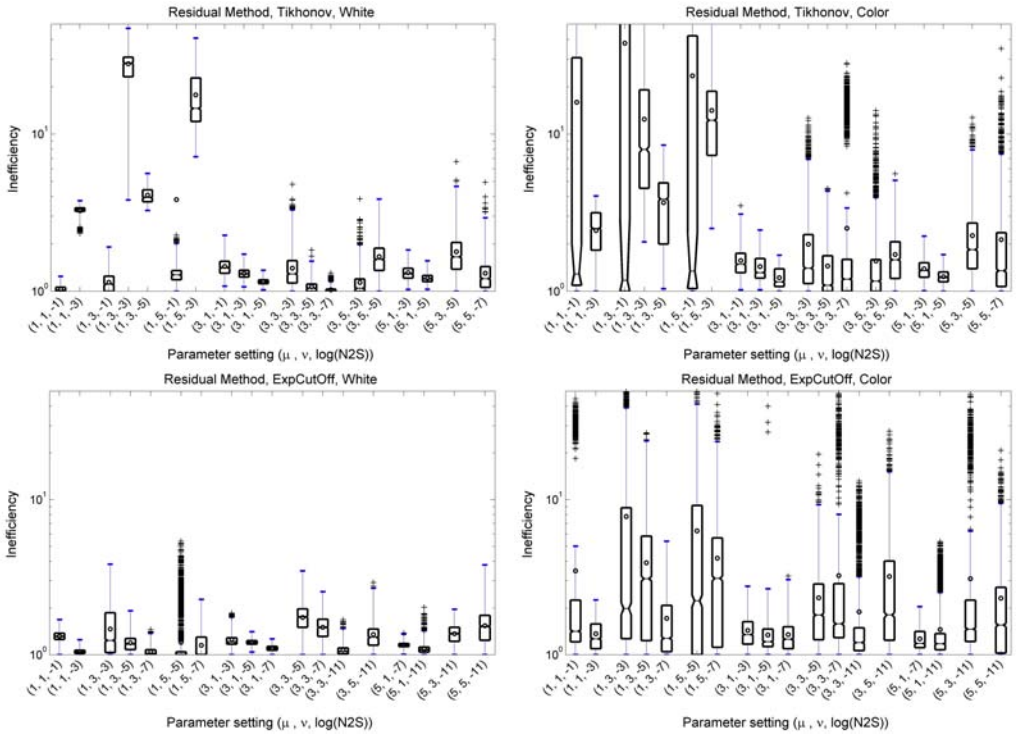

Figure 16: Inefficiencies of the residual method

For Tikhonov regularization with uncorrelated errors in the data, the "expected" GML estimate has been analyzed in [99, 141]. As the sample size $m \to \infty$, the estimate is asymptotically optimal with respect to the prediction risk for "rough" solutions, i.e. solutions x that behave like a realization of the prior stochastic process. However, the estimate is asymptotically sub-optimal and under-smoothing for solutions which are smoother than the minimum required by $x \in \mathcal{X}$, i.e. those satisfying $\nu > 1/2$ here. In numerical studies for spline smoothing [89], the GML method performed well and was more stable than GCV. A geometric explanation of this stability is given in [37], where it is also shown that GML can suffer serious bias.

It is known that GML tends to under-smooth when the errors are positively correlated [112]. The method can be extended to deal with correlated errors, where the correlation is known or parametrically specified, and it performs quite well in this situation for certain smoothing problems [112, 144].

For Tikhonov regularization, the $\det^+$ term in (31) can be computed using a SVD of A if m is not too large.

Numerics. As seen in fig. 17 (and noted above), the GML method does not give reasonable results for spectral cut-off regularization. For Tikhonov regularization with white noise, GML is stable and performs very well in most test cases. It does not perform as well in those cases where $\nu > \mu + 1/2$, for which the minimizer of the prediction risk is not so close to the minimizer of the $\mathcal{X}$-norm risk. For the colored noise scenario, GML

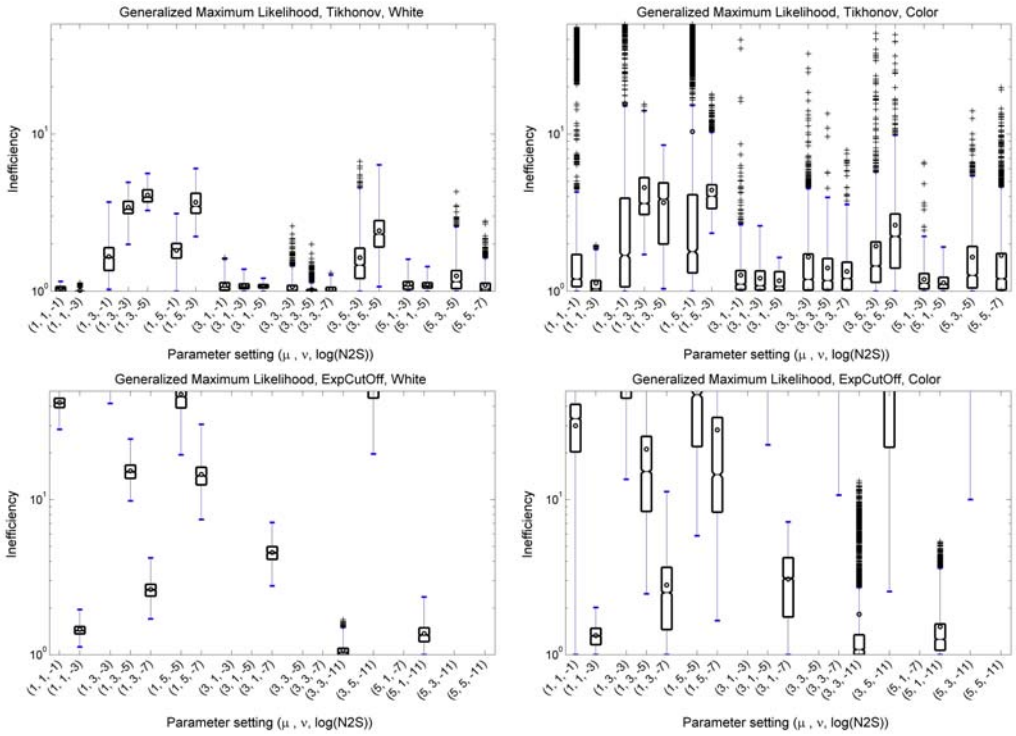

Figure 17: Inefficiencies of generalized maximum likelihood

performs quite well in most cases, but there are many outliers. If GML is used without a maximal index, many of the extreme outliers for colored noise when $\mu = 1$ are even more extreme, resulting in larger means for the inefficiencies.

4.14. Generalized cross-validation

Generalized cross-validation (GCV), due to Wahba [51, 140], is a popular method for practical problems with discrete data and stochastic noise. It originates from the older method of ordinary cross-validation, whose rationale is to consider all the "leave-one-out" regularized solutions and choose the parameter that minimizes the average of the squared prediction errors in using each solution to predict the missing data value. The calculations can be done without computing all the regularized solutions. By using a certain weighting of the prediction errors, Wahba [31, 51, 140, 142] derived the GCV method, which has the advantage of being invariant under orthogonal transformations of the data.

Method. The GCV parameter estimate is defined by

$$n_* = \underset{n \le N}{\operatorname{argmin}} \left\{ \frac{\|Ax_n^\delta - y^\delta\|^2}{(m^{-1} \operatorname{tr}(I - AA_n^{-1}))^2} \right\}, \tag{32}$$

where tr denotes the trace of the matrix.

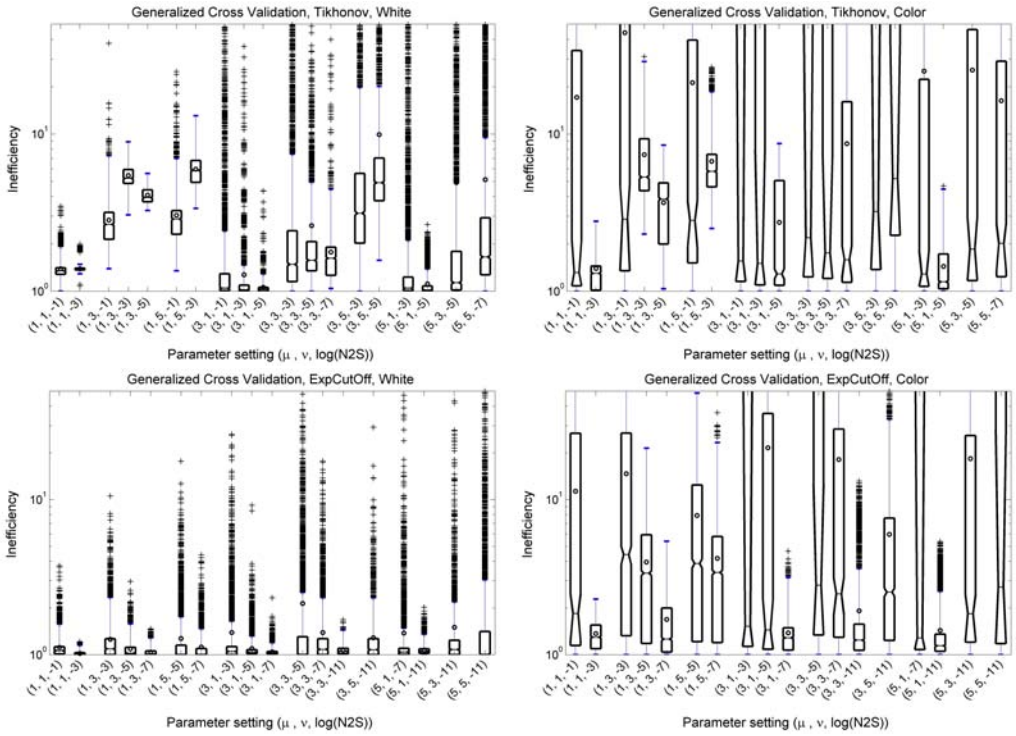

Figure 18: Inefficiencies of generalized cross-validation

Known issues. GCV is closely related to and behaves like the unbiased prediction risk method (also known as Mallows C_p or C_L) [37, 96, 142]. This uses a certain estimate of the prediction risk $\mathbb{E}\|Ax_n^\delta - y\|^2$ that is unbiased if the errors are uncorrelated. Therefore, the GCV estimate is close to being unbiased as a minimizer of the prediction risk.

For Tikhonov regularization, it is known [57, 95, 98] that, with uncorrelated errors, GCV is asymptotically optimal with respect to the prediction risk as the number of data points $m \to \infty$, i.e. the inefficiency goes to 1. In addition, if the unknown solution x is not too smooth relative to the operator, then GCV is order optimal for the $\mathcal{X}$-norm risk $\mathbb{E}\|x_n^\delta - x\|^2$ [98, 139, 143]. The condition required for this here is $\nu \leq \mu + 1/2$, and otherwise GCV is order sub-optimal. This saturation effect is not as serious as it may seem because (using an approximation of ν) one can choose the order of regularization (order of the Sobolev space $\mathcal{X}$) so that the condition is satisfied [98]. In fact, one can use GCV to choose the order of regularization [142].

For spectral cut-off regularization, it is known [96, 137],[139, Chapter 7] that the GCV method is asymptotically optimal for both the prediction risk and the $\mathcal{X}$-norm risk.

The GCV method has been used widely and has been observed to perform very well for reasonably large data sets with uncorrelated errors (white noise). However, it is known [37, 90, 89, 102, 104, 132, 142] that for smaller data sets or correlated errors of red noise type, the method is rather unstable, often resulting in under-smoothing. Graphically, the GCV function in (32) can be very flat near its minimum, it can have multiple local

minima and the global minimum can be at the extreme endpoint for under-smoothing.

The term $\operatorname{tr}(AA_n^{-1})$ in the GCV function is a measure of the degrees of freedom in the regularized solution. For spectral cut-off regularization, it is simply $l(n)$, the number of terms in the expansion of x_n^δ. For Tikhonov regularization, it is harder to compute; although SVD is a convenient approach [142, 71], it is not efficient for large m. There is a more efficient algorithm based on bidiagonalization [38] or tridiagonalization [58] (see also [71] and [142, Chapter 11]). One can also estimate the trace efficiently using stochastic (Monte-Carlo) algorithms [48, 78, 52]. Other efficient algorithms exist for special problems, in particular spline smoothing [79].

Several extensions of GCV have been proposed to deal with correlated data in certain smoothing problems, where the correlation is known or parametrically specified, and good results have been obtained for large sample sizes [112, 144]. The GCV method has also been extended in other directions, including to non-Gaussian data [57] and to wavelet thresholding [80]. It can also be applied to iterative regularization methods, in particular the conjugate gradient method [67, 71], Krylov methods [87], the ART method [126] and the iteratively regularized Gauss–Newton method [59, 142].

Some other parameter choice methods proposed in the literature have been shown to be closely related to GCV, in particular the Akaike information criterion (AIC) [2, 43]. In view of its similarity to GCV, we do not consider this here.

Numerics. As seen in fig. 18, GCV mostly performs well for both Tikhonov and spectral cut-off regularization with white noise. It does not perform so well for Tikhonov regularization in the cases where (μ, ν) equals $(1, 3)$, $(1, 5)$ and $(3, 5)$. These are the cases affected by saturation (since $\nu > \mu + 1/2$), for which the minimizer of the prediction risk is not so close to the minimizer of the $\mathcal{X}$-norm risk. There is no such problem for spectral cut-off, consistent with the theory.

By contrast with the white noise situation, GCV performs very poorly in the colored noise experiment. For both white and colored noise, there are many outliers due to the instability of the method. If GCV is used without a maximal index, many of the extreme outliers are even more extreme, resulting in larger means for the inefficiencies.

4.15. Robust generalized cross-validation

In order to overcome the instability of GCV, a robust GCV (RGCV) method has been developed [102, 123]. Like GCV, the RGCV method has a good rationale in terms of the influence of data values on the regularized solution [102].

Method. The RGCV parameter estimate is defined by

$$n_* = \operatorname*{argmin}_{n \le N} \left\{ \frac{\|Ax_n^\delta - y^\delta\|^2}{(m^{-1} \operatorname{tr}(I - AA_n^{-1}))^2} \left(\gamma + (1-\gamma) m^{-1} \operatorname{tr}((AA_n^{-1})^2) \right) \right\}, \qquad (33)$$

where $\gamma \in (0, 1)$ is a robustness parameter. We take $\gamma = 0.1$.

Known issues. Note that for $\gamma = 1$, the RGCV method is just GCV. As γ is decreased, the method becomes more robust and it is less likely to choose a large value of n (i.e. to under-smooth) [102]. The effect of the last term of the RGCV function in (33) is clear graphically. Compared to GCV, the RGCV function has significantly higher curvature at the minimum point and so the method is much more stable.

The last term in (33) can also be explained by the fact that, for uncorrelated errors with variance δ^2, we have $\delta^2 \operatorname{tr}((AA_n^{-1})^2) = \mathbb{E}\|Ax_n^\delta - EAx_n^\delta\|^2$, the variance of Ax_n^δ. In

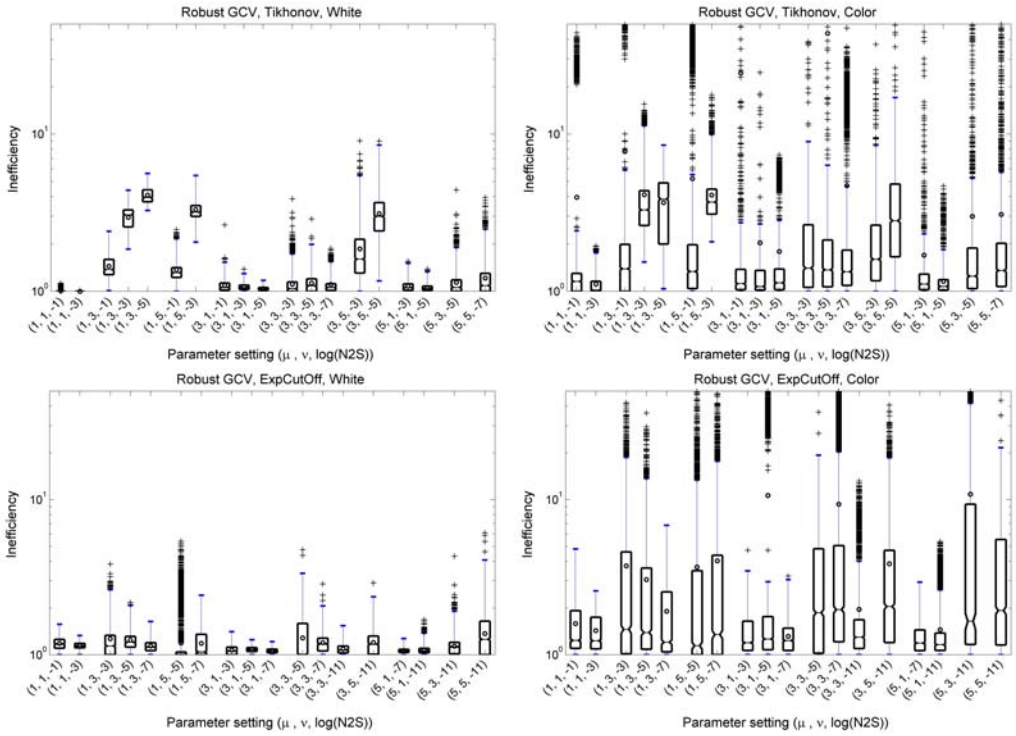

Figure 19: Inefficiencies of robust generalized cross-validation

fact, for Tikhonov regularization, it is known [102, 103] that as $m \to \infty$, the RGCV function is consistent with a weighted sum of the prediction risk $\mathbb{E}\|Ax_n^\delta - y\|^2$ and the variance $\mathbb{E}\|Ax_n^\delta - EAx_n^\delta\|^2$, with weights γ and $1-\gamma$, respectively. Therefore the RGCV method places extra weight on reducing the variability of the regularized solution. As $m \to \infty$, the RGCV estimate has the same optimal order as the GCV estimate, but it has a different constant resulting in a slightly larger value [102, 103]. Consequently, the method does not suffer as badly as GCV from the saturation effect.

For spectral cut-off regularization, it is clear that $\mathrm{tr}((AA_n^{-1})^2) = \mathrm{tr}(AA_n^{-1})$ is simply the number of terms in the expansion of x_n^δ. For Tikhonov regularization, if a SVD is used to compute $\mathrm{tr}(AA_n^{-1})$ in (33), then it requires very little extra work to compute $\mathrm{tr}((AA_n^{-1})^2)$. If a bidiagonalization is used to compute $\mathrm{tr}(AA_n^{-1})$ [38], then $\mathrm{tr}((AA_n^{-1})^2)$ can also be computed efficiently as in [53].

It can be expected that RGCV can be extended in the same way as GCV to iterative regularization methods.

Numerics. As seen in fig. 19, RGCV is stable and mostly performs very well in the white noise situation. For Tikhonov regularization, there is evidence of the saturation effect when $\nu > \mu + 1/2$, but it is smaller than for GCV. The results for colored noise are mostly good and a big improvement compared to GCV, but there are still many outliers. If RGCV is used without a maximal index, many of the extreme outliers for colored noise are even more extreme, resulting in larger means for the inefficiencies.

4.16. Strong robust GCV

The robust GCV (RGCV) method in Section 4.15 is one of a family of robust GCV methods developed in [103] that also includes the strong robust GCV method, denoted R_1GCV. Like GCV and RGCV, the R_1GCV method has a good rationale in terms of the influence of data values on the regularized solution [103].

Method. The R_1GCV parameter estimate is defined by

$$n_* = \underset{n \le N}{\operatorname{argmin}} \left\{ \frac{\|Ax_n^\delta - y^\delta\|^2}{(m^{-1}\operatorname{tr}(I - AA_n^{-1}))^2} \left(\gamma + (1-\gamma) m^{-1} \operatorname{tr}((A_n^{-1\,*} A_n^{-1})\right) \right\}, \qquad (34)$$

where $\gamma \in (0,1)$ is a robustness parameter. We take $\gamma = 0.95$.

Known issues. Note that for $\gamma = 1$, the R_1GCV method is just GCV. As γ is decreased, the method becomes more robust and it is less likely than GCV and (generally) RGCV to choose a large value of n (i.e. to under-smooth) [103].

The last term in (34) can be explained by the fact that for uncorrelated errors with variance δ^2, we have $\delta^2 \operatorname{tr}((A_n^{-1\,*} A_n^{-1})) = \mathbb{E}\|x_n^\delta - \mathbb{E}x_n^\delta\|^2$, the variance of x_n^δ. In fact, for Tikhonov regularization, it is known [103] that as $m \to \infty$, the R_1GCV function in (34) is consistent with a weighted sum of the prediction risk $\mathbb{E}\|Ax_n^\delta - y\|^2$ and the variance $\mathbb{E}\|x_n^\delta - \mathbb{E}x_n^\delta\|^2$, with weights γ and $1-\gamma$, respectively. Since this variance is measured in a stronger norm than for the variance $\mathbb{E}\|Ax_n^\delta - \mathbb{E}Ax_n^\delta\|^2$, R_1GCV places more weight than even RGCV on reducing the variability of the regularized solution. If the solution x is sufficiently smooth, then the R_1GCV choice of regularization parameter is order optimal with respect to the $\mathcal{X}$-norm risk [103], and if the solution is less smooth, then it behaves somewhere between this rate and the optimal rate with respect to the prediction risk. The method also has good finite sample and asymptotic properties for problems with correlated errors [104].

If a SVD is used to compute $\operatorname{tr}(AA_n^{-1}))$ in (33), then it requires very little extra work to compute $T_n := \operatorname{tr}((A_n^{-1\,*} A_n^{-1})$ in (34). For Tikhonov regularization, T_n can also be computed as $T_n = -D(\operatorname{tr}(AA_n^{-1}))$, where D is the divided difference with respect to the regularization parameter (since with continuous Tikhonov regularization parameter α, we have $T_\alpha = -(d/d\alpha)(\operatorname{tr}(AA_\alpha^{-1})))$ [103].

It can be expected that R_1GCV can be extended in the same way as GCV to iterative regularization methods.

Numerics. Figure 20 shows that R_1GCV is stable and mostly performs well for Tikhonov regularization in both the white and colored noise situations, with fewer outliers than both GCV and RGCV. There is little sign of a saturation effect. For spectral cut-off, the performance is mediocre with quite a bit of bias.

4.17. Modified generalized cross-validation

The modified GCV method involves a simple modification of the GCV function that is designed to stabilize the method [32, 136].

Method. The modified GCV parameter estimate is defined by

$$n_* = \underset{n \le N}{\operatorname{argmin}} \left\{ \frac{\|Ax_n^\delta - y^\delta\|^2}{(m^{-1}\operatorname{tr}(I - cAA_n^{-1}))^2} \right\},$$

where $c > 1$ is a stabilization parameter. Here we use $c = 3$.

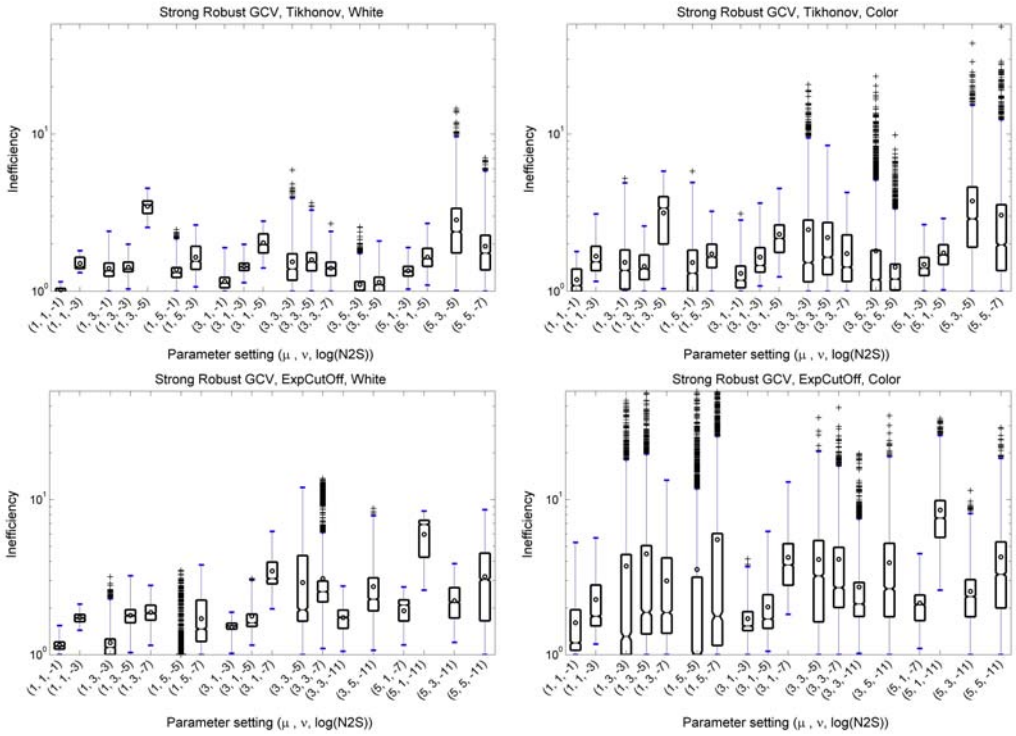

Figure 20: Inefficiencies of strong robust generalized cross-validation

Known issues. When $c = 1$ the method reduces to GCV. The effect of the factor c can be explained [32] in terms of the degrees of freedom for the regularized solution, defined as $df = \mathrm{tr}(AA_n^{-1})$, where AA_n^{-1} is the influence matrix. Clearly the factor introduces a pole at m/c in the objective function as a function of df, which constrains the value of n so that $df < m/c$ and modifies the function's shape to prevent under-smoothing.

The modified GCV method for Tikhonov regularization is closely related to the RGCV method in the sense that, under appropriate conditions, they are asymptotically equivalent as the sample size $m \to \infty$ [103], with the parameters c and γ related by

$$1/\gamma = 1 + (c-1)4\mu/(2\mu - 1), \quad \mu > 1/2.$$

For $\mu = 2$, the value $\gamma = 0.1$ used for RGCV in Section 4.15 corresponds to $c = 4.375$. From experimentation, it appears that $c = 3$ is a good value for our general situation.

Clearly, the modified GCV estimate can be computed in the same way as the GCV estimate; see Section 4.14. It can be expected that the modified GCV method can be extended in the same way as GCV to iterative regularization methods.

Numerics. As seen in fig. 21, the modified GCV method is stable and usually performs well in the white noise situation. For Tikhonov regularization, there is some evidence of the saturation effect for GCV when $\nu > \mu + 1/2$. The results for colored noise are quite good, but there are many outliers.

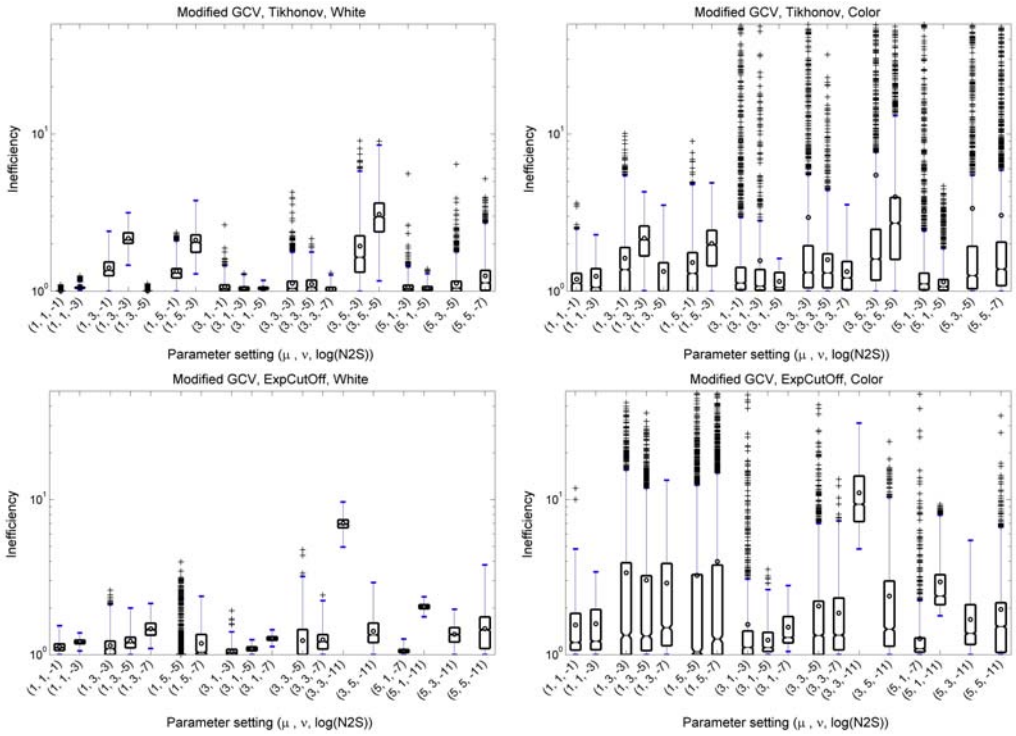

Figure 21: Inefficiencies of modified generalized cross-validation

4.18. Other methods

There are a few parameter choice methods that, for certain reasons, we did not consider in our study. These methods and the reasons are listed below. It should be noted that a method's omission does not mean that it performs poorly.

- The method is known to behave very much like another method in the study.
  - Akaike information criterion (AIC) [2]: behaves like GCV [43]
  - unbiased prediction risk method (also known as Mallows C_P or C_L): behaves like GCV [37, 96, 142]
  - unbiased risk method [27, 100]: behaves like the modified discrepancy principle with a particular tuning constant
  - rule R1 of Raus [63]: behaves like the fast balancing principle with $l(n) = n+1$ in (23)
- The method does not generalize in an obvious way to both spectral cut-off and Tikhonov regularization.
  - Arcangeli's principle [6, 56, 109], which is an early method developed for Tikhonov regularization with deterministic noise
  - rule R2 of Raus [121], which was developed for (iterated) Tikhonov regularization with deterministic noise

  - risk hull method [27], which was developed for spectral cut-off regularization with Gaussian white noise
- The method is difficult to automate.
  - some versions of the L-curve method
- The method requires a heavy load of precomputations, which makes it difficult to test in our large-scale simulation experiments. Such methods usually make use of a precise specification of the stochastic noise model, which in practice (at least at the necessary precision) is not known.
  - risk hull method [27]
  - modified balancing principle [128]

5. Summary and conclusions

In this section, we will summarize the requirements and properties of the parameter choice methods described in Section 4, and then compare them with respect to their average performance in our numerical experiments.

For standard uses, this should serve as a practical guide through the jungle of different methods. However, for special problems, the performance of the methods might be quite different. In addition, for each method, there are implementation issues which might make one or other method more practical in certain situations.

Most of the parameter choice methods considered here are based on some principle or rationale that is independent of the regularization method, and, in practice, they can be applied to various regularization methods. Therefore, it is reasonable to suppose that each method will have a similar performance when applied to regularization methods with similar behavior depending on the parameter. Hence, our simulation results for Tikhonov and spectral cut-off regularization may serve as model results for other regularization methods of low and high qualification, respectively.

5.1. Requirements and properties

The requirements and general properties of each method are summarized in Table 3. The methods are split into three groups according to the input information required; those in the first group require the noise level, those in the second group require at least two independent data sets and those in the third group do not require any a-priori knowledge. The abbreviations used in Table 3 are defined in the list below.

- (Or.) Mathematical origin of the method
  - (FA) functional analytic background
  - (St) stochastic background
  - (–) no clear origin can be identified
- (Sp.) The function space in which the method works
  - (X) the solution space $\mathcal{X}$, i.e. norms and/or inner products in $\mathcal{X}$ need to be evaluated
  - (Y) the data space $\mathcal{Y}$ (mostly using the residual), i.e. norms and/or inner products in $\mathcal{Y}$ need to be evaluated

Method	Or.	Sp.	Inp.	Tun.	Set.	Com.	Cal.	PD	PS	Gen.
Discrepancy Principle	FA	Y	δ	y	∞	log	–	y	–	l+nl
Transformed Disc. Pr.	FA	Y	δ	y	∞	log	–	y	–	–
Modified Discrepancy Pr.	FA	Y	δ	y	∞	log	–	y	y	l+nl
Monotone Error Rule	FA	Y	δ	y	∞	log	–	y	–	–
Balancing Pr. (white)	St	X	δ	y	∞	N	tr	y	y	l+nl
Balancing Principle	St	X	2	y	∞	N	–	–	y	l+nl
Fast Balancing Principle	St	X	2	y	∞	log	–	–	y	–
Hardened Balancing Pr.	–	X	2	n	∞	N	–	–	y	–
Hard. Balancing Pr. (wh.)	–	X	–	n	∞	N	tr	–	y	–
Quasi-Optimality Criterion	FA	X	–	n	∞	log	–	y	y	l
L-Curve Method	–	XY	–	n	∞	log	–	–	–	–
Modified Disc. Partner R.	FA	Y	–	n	∞	log	–	y	–	l
Extrapolated Error Meth.	–	Y	–	n	m	log	–	–	–	–
N. Cum. Periodogram M.	St	Y	–	n	m	log	–	–	–	–
Residual Method	–	Y	–	n	∞	log	tr	–	y	–
General. Max. Likelihood	St	Y	–	n	m	log	det	–	y	–
General. Cross-Validation	St	Y	–	n	m	log	tr	–	y	l
Robust GCV	St	Y	–	y	m	log	tr	–	y	–
Strong Robust GCV	St	Y	–	y	m	log	tr	–	y	–
Modified GCV	St	Y	–	y	m	log	tr	–	y	–

Table 3: Requirements and general properties of the methods

  - (XY) there is a need to evaluate norms or inner products both in $\mathcal{X}$ and $\mathcal{Y}$

- (Inp.) Input information required for the method
  - (δ) the noise level δ
  - (2) at least two (or better more) independent data sets
  - (–) no additional information at all
- (Tun.) The method involves a tuning parameter(s). While the optimal use of a tuning parameter usually improves the results for a certain class of problems, setting it incorrectly can give very poor results. Setting the parameter is especially difficult if there is is no theory about its effect on the regularized solution. In practice, where one cannot necessarily check the validity of a certain setting, methods without (or with very robust) tuning parameters are usually preferable.
  - (y) yes
  - (n) no
- (Set.) Data setting for which the method works
  - (m) finite data setting required
  - (∞) infinite data setting possible, as well as a finite one
- (Com.) Computational complexity: number of regularized solutions that need to be computed

  - (log) the method allows (in principle) for bisection or other methods to reduce the amount of computation to a logarithmic number $\log(N)$ of regularized solutions, where N is the maximal number
  - (N) all regularized solutions need to be computed
- (Cal.) Requirement for expensive calculations like trace computations
  - (tr)/(det) requires computing the trace or determinant in each step
  - (–) does not require any additional expensive calculations
- (PD) Proof is available in a deterministic setting (i.e. the noise and solution are deterministic)
  - (y) yes (Some restrictions may be required – see the known issues for the method in Section 4.)
  - (–) no proof known yet
- (PS) Proof is available in a stochastic setting (i.e. at least the noise is modeled as a random variable)
  - (y) yes
  - (–) no proof known yet
- (Gen.) Generalizations with theory are known for more than spectral cut-off or Tikhonov regularization of linear problems
  - (l) proofs for other regularization methods (mostly Landweber iteration) for linear problems are known
  - (nl) proofs for nonlinear problems (mostly with the iteratively regularized Gauss–Newton method or Landweber iteration) are known
  - (–) no proofs for further situations are known

5.2. Average performance

For each method evaluated in Section 4, there are four displayed figure panels (one for each regularization method and noise scenario) with a box plot for every test case, showing the sample mean(I) ($\circ$) and sample median(I) (-) of the computed inefficiencies I. For each of these four situations, we will measure the average performance of the method across all the test cases using both mean(mean(I)) and median(median(I)). The results are displayed in Table 4, which again has three groups according to the input information required.

Note that some methods performed significantly better for one of the regularization methods than the other. The best methods, i.e. those giving the smallest mean and those giving the smallest median, in the three groups and four situations are marked using boldface. In most cases, the same method is best for both the mean and median, with both values close to the ideal value of 1. Several other methods, especially in the colored noise situation, have a high mean and low median, which means that, while the method mostly performs well, it also generates a significant number of very poor outliers.

In each of the three groups, the following methods seem to be the best performers.

- Noise level δ is known accurately: modified discrepancy principle (ExpCutOff), discrepancy principle (Tikhonov) and monotone error rule (Tikhonov).

Method	ExpCO, White		ExpCO, Color		Tikh., White		Tikh., Color	
	Mean	Med.	Mean	Med.	Mean	Med.	Mean	Med.
Discrepancy Principle	1.71	1.62	6.64	1.73	**1.57**	**1.36**	**2.60**	1.49
Transformed Disc. Pr.	149	1.28	181	1.57	4.08	1.85	8.23	1.90
Modified Discrepancy Pr.	**1.09**	**1.01**	**2.22**	**1.16**	1.61	1.47	3.91	1.52
Monotone Error Rule	1.19	1.15	2.82	1.18	1.59	1.44	3.01	**1.47**
Balancing Pr. (white)	1.63	1.56	5.61	1.59	2.03	2.19	3.93	2.23
Balancing Principle	1.60	1.53	1.67	1.40	2.01	2.15	2.02	2.14
Fast Balancing Principle	1.61	1.53	1.62	1.41	2.14	2.31	2.13	2.29
Hardened Balancing Pr.	**1.14**	**1.07**	**1.22**	**1.07**	**1.12**	**1.07**	**1.13**	**1.07**
Hard. Balancing Pr. (wh.)	**1.14**	1.07	3.20	1.37	**1.12**	1.07	1.86	1.14
Quasi-Optimality Criterion	1.23	1.08	**1.28**	**1.08**	1.87	1.07	**1.78**	**1.09**
L-Curve Method	783	9.38	853	8.80	86.74	9.66	151	8.91
Modified Disc. Partner R.	1.65	1.14	1.84	1.14	3.30	2.50	4.04	2.47
Extrapolated Error Meth.	151	2.10	182	2.44	1.78	1.52	2.09	1.65
N. Cum. Periodogram M.	1.37	1.15	168	3.00	6.44	1.14	643	2.97
Residual Method	1.27	1.19	2.87	1.44	3.94	1.29	6.86	1.35
General. Max. Likelihood	11346	49.88	7968	49.59	1.67	1.08	3.43	1.20
General. Cross-Validation	1.37	**1.04**	717	1.84	5.48	1.57	797	1.86
Robust GCV	1.16	1.10	290	1.28	1.64	1.06	150	1.34
Strong Robust GCV	2.28	1.82	3.59	1.99	1.63	1.40	1.96	1.46
Modified GCV	1.59	1.20	2.73	1.32	1.37	**1.05**	2.58	1.26

Table 4: Mean and median inefficiencies of the methods over all the test cases

- Several independent data sets are available: hardened balancing principle.
- No extra information: hardened balancing principle (white), quasi-optimality criterion, modified discrepancy partner rule (ExpCutOff), robust GCV, strong robust GCV (Tikhonov) and modified GCV.

It should be noted that, in most situations, the best methods that do not require the noise level performed better than the methods that use the noise level (i.e. those in the first group). This indicates that one should not use the "known δ" methods for the sake of performance, but there may be another reason, e.g. computational efficiency, for doing so.

The conclusions here are consistent with those in the numerical study of Palm [113], which uses the set of test problems of Hansen [70] with uniformly distributed white noise and colored noise. As above, from the methods using the noise level in [113], both the discrepancy principle and the monotone error rule performed well for Tikhonov regularization. The discrepancy principle also performed well in [113] for spectral cut-off, but our MD and ME rules were not assessed. From the data-driven methods considered in [113], as above, adapted versions of the quasi-optimality criterion performed well for both spectral cut-off and Tikhonov regularization. (The hardened balancing principle and robust GCV were not implemented there.)

From Table 4, some methods performed much worse in the colored noise situation than for white noise. In practical applications, where one normally has only a vague idea

of the underlying noise structure, it is usually advisable to opt for those methods that have good performance for colored noise. This is especially true as the methods that performed well for colored noise also performed quite well in the white noise situation.

Acknowledgements

We would like to thank all colleagues who have provided us with hints for relevant literature or have helped us by explaining the particulars of their methods.

Furthermore, the first author would like to gratefully acknowledge the financial support of the Graduiertenkolleg 1023 'Identification in Mathematical Models', University of Göttingen, the Upper Austrian Technology and Research Promotion and the University of Linz.

References

[1] J.F.P.J. Abascal, S.R. Arridge, R.H. Bayford, D.S. Holder, Comparison of methods for optimal choice of the regularization parameter for linear electrical impedance tomography of brain function, Physiol. Meas. 29 (2008) 1319–1334.

[2] H. Akaike, Information theory and an extension of the maximum likelihood principle, in: Second International Symposium on Information Theory (Tsahkadsor, 1971), Akadémiai Kiadó, Budapest, 1973, pp. 267–281.

[3] E.O. Åkesson, K.J. Daun, Parameter selection methods for axisymmetric flame tomography through Tikhonov regularization, Appl. Opt. 47 (2008) 407–416.

[4] O. Alifanov, S. Rumyantsev, On the stability of iterative methods for the solution of linear ill-posed problems, Sov. Math., Dokl. 20 (1979) 1133–1136.

[5] R.S. Anderssen, P. Bloomfield, Numerical differentiation procedures for non-exact data, Numer. Math. 22 (1974) 157–182.

[6] R. Arcangeli, Pseudo-solution de l'équation $Ax = y$., C.R. Acad. Sci. Paris, Sér. A 263 8 (1966) 282–285.

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

[8] A. Bakushinsky, A. Smirnova, On application of generalized discrepancy principle to iterative methods for nonlinear ill-posed problems, Numer. Funct. Anal. Optim. 26 (2005) 35–48.

[9] F. Bauer, Some considerations concerning regularization and parameter choice algorithms, Inverse Problems 23 (2007) 837–858.

[10] F. Bauer, Applying Lepskij-balancing in practice, http://arxiv.org/abs/1008.0657 (2010).

[11] F. Bauer, Parameter choice by fast balancing, http://arxiv.org/abs/1008.0620 (2010).

[12] F. Bauer, T. Hohage, A Lepskij-type stopping rule for regularized Newton methods, Inverse Problems 21 (2005) 1975–1991.

[13] F. Bauer, T. Hohage, A. Munk, Iteratively regularized Gauss–Newton method for nonlinear inverse problems with random noise, SIAM J. Numer. Anal. 47 (2009) 1827–1846.

[14] F. Bauer, S. Kindermann, The quasi-optimality criterion for classical inverse problems, Inverse Problems 24 (2008) 035002.

[15] F. Bauer, S. Kindermann, Recent results on the quasi-optimality principle, J. Inverse Ill-Posed Prob. 17 (2009) 5–18.

[16] F. Bauer, P. Mathé, Parameter choice methods using minimization schemes, J. Complexity (2010) doi:10.1016/j.jco.2010.10.001.

[17] F. Bauer, A. Munk, Optimal regularization for ill-posed problems in metric spaces, J. Inverse Ill-Posed Probl. 15 (2007) 137–148.

[18] F. Bauer, S. Pereverzev, Regularization without preliminary knowledge of smoothness and error behavior, European J. Appl. Math. 16 (2005) 303 – 317.

[19] F. Bauer, M. Reiß, Regularization independent of the noise level: an analysis of quasi-optimality, Inverse Problems 24 (2008) 055009.

[20] C. Brezinski, G. Rodriguez, S. Seatzu, Error estimates for linear systems with applications to regularization, Numer. Algorithms 49 (2008) 85–104.

[21] C. Brezinski, G. Rodriguez, S. Seatzu, Error estimates for the regularization of least squares problems, Numer. Algorithms 51 (2009) 61–76.
[22] L.D. Brown, M.G. Low, Asymptotic equivalence of nonparametric regression and white noise., Ann. Stat. 24 (1996) 2384–2398.
[23] D. Calvetti, P.C. Hansen, L. Reichel, L-curve curvature bounds via Lanczos bidiagonalization, Elect. Trans. Numer. Anal. 14 (2002) 20–35.
[24] E.J. Candès, Modern statistical estimation via oracle inequalities, Acta Numerica 15 (2006) 257–325.
[25] L. Cavalier, Nonparametric statistical inverse problems, Inverse Problems 24 (2008) 034004, 19.
[26] L. Cavalier, G.K. Golubev, D. Picard, A.B. Tsybakov, Oracle inequalities for inverse problems, Ann. Statist. 30 (2002) 843–874.
[27] L. Cavalier, Y. Golubev, Risk hull method and regularization by projections of ill-posed inverse problems, Ann. Statist. 34 (2006) 1653–1677.
[28] D. Colton, R. Kress, Inverse Acoustic and Electromagnetic Scattering Theory. 2nd ed., Springer, Berlin, 1998.
[29] T. Correia, A. Gibson, M. Schweiger, J. Hebden, Selection of regularization parameter for optical topography, J. Biomed. Opt. 14 (2009) 034044.
[30] D.D. Cox, Approximation of method of regularization estimators, Ann. Statist. 16 (1988) 694–712.
[31] P. Craven, G. Wahba, Smoothing noisy data with spline functions, Numer. Math. 31 (1979) 377–403.
[32] D.J. Cummins, T.G. Filloon, D. Nychka, Confidence intervals for nonparametric curve estimates: Toward more uniform pointwise coverage, J. Amer. Statist. Assoc. 96 (2001) 233–246.
[33] A.R. Davies, On the maximum likelihood regularization of Fredholm convolution equations of the first kind, in: C. Baker, G. Miller (Eds.), Treatment of Integral Equations by Numerical Methods, Academic Press, London, 1982, pp. 95–105.
[34] A.R. Davies, R.S. Anderssen, Improved estimates of statistical regularization parameters in Fourier differentiation and smoothing, Numer. Math. 48 (1986) 671–697.
[35] P. Ditmar, R. Klees, X. Liu, Frequency-dependent data weighting in global gravity field modeling from satellite data contaminated by non-stationary noise, J. Geod. 81 (2007) 81–96.
[36] P. Ditmar, J. Kusche, R. Klees, Computation of spherical harmonic coefficients from gravity gradiometry data to be acquired by the GOCE satellite: regularization issues, J. Geod. 77 (2003) 465–477.
[37] B. Efron, Selection criteria for scatterplot smoothers, Ann. Statist. 29 (2001) 470–504.
[38] L. Eldén, A note on the computation of the generalized cross-validation function for ill-conditioned least squares problems, BIT 24 (1984) 467–472.
[39] H.W. Engl, Regularization methods for the stable solution of inverse problems, Surv. Math. Ind. 3 (1993) 71–143.
[40] H.W. Engl, H. Gfrerer, A posteriori parameter choice for general regularization methods for solving linear ill-posed problems, Appl. Numer. Math. 4 (1988) 395–417.
[41] H.W. Engl, H. Hanke, A. Neubauer, Regularization of Inverse Problems, Kluwer, Dordrecht, 1996.
[42] H.W. Engl, O. Scherzer, Convergence rate results for iterative methods for solving nonlinear ill-posed problems, in: D. Colton, H. Engl, A. Louis, J. McLaughlin, W. Rundell (Eds.), Survey on Solution Methods for Inverse Problems, Springer, New York, 2000, pp. 7–34.
[43] R.L. Eubank, Spline Smoothing and Nonparametric Regression, Marcel Dekker, New York, 1988.
[44] S.N. Evans, P.B. Stark, Inverse problems as statistics, Inverse Problems 18 (2002) R55–R97.
[45] C.G. Farquharson, D.W. Oldenburg, A comparison of automatic techniques for estimating the regularization parameter in non-linear inverse problems, Geophys. J. Int. 156 (2004) 411–425.
[46] B.G. Fitzpatrick, Bayesian analysis in inverse problems, Inverse Problems 7 (1991) 675–702.
[47] H. Gfrerer, An a posteriori parameter choice for ordinary and iterated Tikhonov regularization of ill-posed problems leading to optimal convergence rates, Math. Comput. 49 (1987) 507–522.
[48] D. Girard, A fast "Monte-Carlo cross-validation" procedure for large least squares problems with noisy data, Numer. Math. 56 (1989) 1–23.
[49] V. Glasko, Y. Kriksin, On the quasioptimality principle for linear ill-posed problems in Hilbert space, Vychisl. Mat. Mat. Fiz. 24 (1984) 1603–1613.
[50] A. Goldenshluger, S. Pereverzev, Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations, Prob. Th. Rel. Fields 118 (2000) 169–186.
[51] G.H. Golub, M. Heath, G. Wahba, Generalized cross-validation as a method for choosing a good ridge parameter 21 (1979) 215–223.
[52] G.H. Golub, U. von Matt, Generalized cross-validation for large scale problems, J. Comput. Graph.

Statist. 6 (1997) 1–34.
[53] J. Grad, E. Zakrajšek, LR algorithm with Laguerre shift for symmetric tridiagonal matrices, Computer J. 15 (1972) 268–270.
[54] C.W. Groetsch, The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind, Pitman, Boston, 1984.
[55] C.W. Groetsch, A. Neubauer, Regularization of ill-posed problems: Optimal parameter choice in finite dimensions, J. Approx. Th. 58 (1989) 184–200.
[56] C.W. Groetsch, E. Schock, Asymptotic convergence rate of Arcangeli's method for ill-posed problems., Applicable Anal. 18 (1984) 175–182.
[57] C. Gu, Smoothing Spline ANOVA Models, Springer, New York, 2002.
[58] C. Gu, D.M. Bates, Z. Chen, G. Wahba, The computation of generalized cross-validation functions through Householder tridiagonalization with applications to the fitting of interaction spline models, SIAM J. Matrix Anal. Appl. 10 (1989) 457–480.
[59] E. Haber, D.W. Oldenburg, A GCV based method for nonlinear ill-posed problems, Comput. Geosci. 4 (2000) 41–63.
[60] U. Hämarik, R. Palm, T. Raus, On minimization strategies for choice of the regularization parameter in ill-posed problems, Numer. Funct. Anal. Optim. 30 (2009) 924–950.
[61] U. Hämarik, T. Raus, On the a posteriori parameter choice in regularization methods, Proc. Estonian Acad. Sci. Phys. Math. 48 (1999) 133–145.
[62] U. Hämarik, T. Raus, On the choice of the regularization parameter in ill-posed problems with approximately given noise level of data, J. Inverse Ill-Posed Probl. 14 (2006) 251–266.
[63] U. Hämarik, T. Raus, About the balancing principle for choice of the regularization parametertion, Numer. Funct. Anal. Optim. 30 (2009) 951–970.
[64] U. Hämarik, U. Tautenhahn, On the monotone error rule for parameter choice in iterative and continuous regularization methods, BIT 41 (2001) 1029–1038.
[65] U. Hämarik, U. Tautenhahn, On the monotone error rule for choosing the regularization parameter in ill-posed problems, in: M. M. Lavrent'ev et al. (Ed.), Ill-posed and Non-classical Problems of Mathematical Physics and Analysis, VSP, Utrecht, 2003, pp. 27–55.
[66] M. Hanke, Limitations of the L-curve method in ill-posed problems, BIT 36 (1996) 287–301.
[67] M. Hanke, P.C. Hansen, Regularization methods for large-scale problems, Surv. Math. Ind. 3 (1993) 253–315.
[68] M. Hanke, T. Raus, A general heuristic for choosing the regularization parameter in ill-posed problems, SIAM J. Sci. Comput. 17 (1996) 956–972.
[69] P.C. Hansen, Analysis of discrete ill-posed problems by means of the L-curve, SIAM Rev. 34 (1992) 561–580.
[70] P.C. Hansen, Regularization tools: A matlab package for analysis and solution of discrete ill-posed problems, Numer. Algorithms 6 (1994) 1–35.
[71] P.C. Hansen, Rank-Deficient and Discrete Ill-Posed Problems, SIAM, Philadelphia, 1998.
[72] P.C. Hansen, The L-curve and its use in the numerical treatment of inverse problems, in: P. R. Johnstone (Ed.), Computational Inverse Problems in Electrocardiography, WIT Press, Southampton, 2001, pp. 119–142.
[73] P.C. Hansen, T.K. Jensen, G. Rodriguez, An adaptive pruning algorithm for the discrete L-curve criterion, J. Comput. Appl. Math. 198 (2007) 483–492.
[74] P.C. Hansen, M.E. Kilmer, R.H. Kjeldsen, Exploiting residual information in the parameter choice for discrete ill-posed problems, BIT 46 (2006) 41–59.
[75] P.C. Hansen, D.P. O'Leary, The use of the L-curve in the regularization of discrete ill-posed problems, SIAM J. Sci. Comput. 14 (1993) 1487–1503.
[76] A. Hofinger, H.K. Pikkarainen, Convergence rate for the Bayesian approach to linear inverse problems, Inverse Problems 23 (2007) 2469–2484.
[77] B. Hofmann, Regularization of Applied Inverse and Ill-Posed Problems, Teubner, Leipzig, 1986.
[78] M. Hutchinson, A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines, Commun. Stat. Simulation Comput. 18 (1989) 1059–1076.
[79] M.F. Hutchinson, F.R. de Hoog, Smoothing noisy data with spline functions, Numer. Math. 47 (1985) 99–106.
[80] M. Jansen, M. Malfait, A. Bultheel, Generalized cross validation for wavelet thresholding, Signal Process. 56 (1997) 33–44.
[81] Q. Jin, U. Tautenhahn, On the discrepancy principle for some Newton type methods for solving nonlinear inverse problems, Numer. Math. 111 (2009) 509–558.
[82] Q.n. Jin, On the iteratively regularized Gauss–Newton method for solving nonlinear ill-posed

problems, Math. Comput. 69 (2000) 1603–1623.
[83] Q.n. Jin, Z.y. Hou, On an a posteriori parameter choice strategy for Tikhonov regularization of nonlinear ill-posed problems, Numer. Math. 83 (1999) 139–159.
[84] P.R. Johnstone, R.M. Gulrajani, Selecting the corner in the L-curve approach to Tikhonov regularization, IEEE Trans. Biomed. Eng. 47 (2000) 1293–1296.
[85] J. Kaipio, E. Somersalo, Statistical and Computational Inverse Problems, Springer, New York, 2005.
[86] B. Kaltenbacher, A. Neubauer, O. Scherzer, Iterative Regularization Methods for Nonlinear Ill-Posed Problems, Walter de Gruyter, Berlin, 2008.
[87] M.E. Kilmer, D.P. O'Leary, Choosing regularization parameters in iterative methods for ill-posed problems, SIAM J. Matrix Anal. Applic. 22 (2001) 1204–1221.
[88] S. Kindermann, A. Neubauer, On the convergence of the quasioptimality criterion for (iterated) Tikhonov regularization, Inverse Probl. Imaging 2 (2008) 291–299.
[89] R. Kohn, C.F. Ansley, D. Tharm, The performance of cross-validation and maximum likelihood estimators of spline smoothing parameters, J. Amer. Statist. Assoc. 86 (1991) 1042–1050.
[90] S.C. Kou, B. Efron, Smoothers and the C_p, generalized maximum likelihood, and extended exponential criteria: a geometric approach, J. Amer. Statist. Assoc. 97 (2002) 766–782.
[91] F.M. Larkin, Gaussian measure in Hilbert space and applications in numerical analysis, Rocky Mt. J. Math. 2 (1972) 379–421.
[92] C.L. Lawson, R.J. Hanson, Solving Least Squares Problems, Prentice-Hall, Englewood Cliffs, N. J., 1974.
[93] A. Leonov, Justification of the choice of regularization parameter according to quasi-optimality and quotient criteria, U.S.S.R. Comput. Math. Math. Phys. 18 (1979) 1–15.
[94] O. Lepskij, On a problem of adaptive estimation in Gaussian white noise., Theory Probab. Appl. 35 (1990) 454–466.
[95] K.C. Li, Asymptotic optimality of C_L and generalized cross-validation in ridge regression with application to spline smoothing, Ann. Statist. 14 (1986) 1101–1112.
[96] K.C. Li, Asymptotic optimality for C_p, C_L, cross-validation and generalized cross-validation: Discrete index set, Ann. Statist. 15 (1987) 958–975.
[97] M.A. Lukas, Convergence rates for regularized solutions, Math. Comput. 51 (1988) 107–131.
[98] M.A. Lukas, Asymptotic optimality of generalized cross-validation for choosing the regularization parameter, Numer. Math. 66 (1993) 41–66.
[99] M.A. Lukas, On the discrepancy principle and generalised maximum likelihood for regularisation, Bull. Aust. Math. Soc. 52 (1995) 399–424.
[100] M.A. Lukas, Asymptotic behaviour of the minimum bound method for choosing the regularization parameter, Inverse Problems 14 (1998) 149–159.
[101] M.A. Lukas, Comparisons of parameter choice methods for regularization with discrete noisy data, Inverse Problems 14 (1998) 161–184.
[102] M.A. Lukas, Robust generalized cross-validation for choosing the regularization parameter, Inverse Problems 22 (2006) 1883–1902.
[103] M.A. Lukas, Strong robust generalized cross-validation for choosing the regularization parameter, Inverse Problems 24 (2008) 034006, 16.
[104] M.A. Lukas, Robust GCV choice of the regularization parameter for correlated data, J. Integral Equations Appl. 22 (2010) 519–547.
[105] P. Mathé, S.V. Pereverzev, Geometry of linear ill-posed problems in variable Hilbert spaces, Inverse Problems 19 (2003) 789–803.
[106] P. Mathé, S.V. Pereverzev, Regularization of some linear ill-posed problems with discretized random noisy data, Math. Comput. 75 (2006) 1913–1929.
[107] V.A. Morozov, On the solution of functional equations by the method of regularization, Soviet Math. Dokl. 7 (1966) 414–417.
[108] V.A. Morozov, Methods for Solving Incorrectly Posed Problems, Springer-Verlag, New York, 1984.
[109] M.T. Nair, M.P. Rajan, Generalized Arcangeli's discrepancy principles for a class of regularization methods for solving ill-posed problems, J. Inverse Ill-Posed Probl. 10 (2002) 281–294.
[110] A. Neubauer, An a posteriori parameter choice for Tikhonov regularization in the presence of modeling error, Appl. Numer. Math. 4 (1988) 507–519.
[111] A. Neubauer, The convergence of a new heuristic parameter selection criterion for general regularization methods, Inverse Problems 24 (2008) 055005, 10.
[112] J. Opsomer, Y. Wang, Y. Yang, Nonparametric regression with correlated errors, Statist. Sci. 16 (2001) 134–153.

[113] R. Palm, Numerical comparison of regularization algorithms for solving ill-posed problems, Ph.D. thesis, University of Tartu, Estonia, 2010.
[114] D. Phillips, A technique for the numerical solution of certain integral equations of the first kind, J. Assoc. Comput. Mach. 9 (1962) 84–97.
[115] S. Pohl, B. Hofmann, R. Neubert, T. Otto, C. Radehaus, A regularization approach for the determination of remission curves, Inverse Probl. Eng. 9 (2001) 157–174.
[116] T. Raus, On the discrepancy principle for the solution of ill-posed problems, Uch. Zap. Tartu. Gos. Univ. 672 (1984) 16–26.
[117] T. Raus, The principle of the residual in the solution of ill-posed problems with nonselfadjoint operator, Tartu Riikl. Ül. Toimetised (1985) 12–20.
[118] T. Raus, An a posteriori choice of the regularization parameter in case of approximately given error bound of data, in: Pedas, A. (Ed.), Collocation and Projection Methods for Integral Equations and Boundary Value Problems, Tartu University, Tartu, 1990, pp. 73–87.
[119] T. Raus, About regularization parameter choice in case of approximately given error bounds of data, in: G. Vainikko (Ed.), Methods for solution of integral equations and ill-posed problems, Tartu University, Tartu, 1992.
[120] T. Raus, U. Hämarik, On the quasioptimal regularization parameter choices for solving ill-posed problems, J. Inverse Ill-Posed Probl. 15 (2007) 419–439.
[121] T. Raus, U. Hämarik, New rule for choice of the regularization parameter in (iterated) Tikhonov method, Math. Model. Anal. 14 (2009) 187–198.
[122] T. Reginska, A regularization parameter in discrete ill-posed problems, SIAM J. Sci. Comput. 17 (1996) 740–749.
[123] T. Robinson, R. Moyeed, Making robust the cross-validatory choice of smoothing parameter in spline smoothing regression, Commun. Stat. Theory Methods 18 (1989) 523–539.
[124] B.W. Rust, Parameter selection for constrained solutions to ill-posed problems, Comput. Sci. Stat. 32 (2000) 333–347.
[125] B.W. Rust, D.P. O'Leary, Residual periodograms for choosing regularization parameters for ill-posed problems, Inverse Problems 24 (2008) 034005, 30.
[126] R. Santos, A. De Pierro, A cheaper way to compute generalized cross-validation as a stopping rule for linear stationary iterative methods, J. Comput. Graph. Stat. 12 (2003) 417–433.
[127] O. Scherzer, H. Engl, K. Kunisch, Optimal a posteriori parameter choice for Tikhonov regularization for solving nonlinear ill-posed problems, SIAM J. Numer. Anal. 30 (1993) 1796–1838.
[128] V. Spokoiny, C. Vial, Parameter tuning in pointwise adaptation using a propagation approach, Ann. Statist. 37 (2009) 2783–2807.
[129] A. Tarantola, Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation, Elsevier, Amsterdam, 1987.
[130] U. Tautenhahn, U. Hämarik, The use of monotonicity for choosing the regularization parameter in ill-posed problems, Inverse Problems 15 (1999) 1487–1505.
[131] A.M. Thompson, J.C. Brown, J.W. Kay, D.M. Titterington, A study of methods for choosing the smoothing parameter in image restoration by regularization, IEEE Trans. Pattern Anal. Machine Intell. 13 (1991) 3326–3339.
[132] A.M. Thompson, J.W. Kay, D.M. Titterington, A cautionary note about crossvalidatory choice, J. Statist. Comput. Simul. 33 (1989) 199–216.
[133] A. Tikhonov, V. Arsenin, Solutions of Ill-Posed Problems, Wiley, New York, 1977.
[134] A. Tikhonov, V. Glasko, Use of the regularization method in non-linear problems, U.S.S.R. Comput. Math. Math. Phys. 5 (1965) 93–107.
[135] A. Tsybakov, On the best rate of adaptive estimation in some inverse problems, C. R. Acad. Sci., Paris, Sér. I, Math. 330 (2000) 835–840.
[136] R. Vio, P. Ma, W. Zhong, J. Nagy, L. Tenorio, W. Wamsteker, Estimation of regularization parameters in multiple-image deblurring, Astron. Astrophys. 423 (2004) 1179–1186.
[137] C.R. Vogel, Optimal choice of a truncation level for the truncated SVD solution of linear first kind integral equations when data are noisy, SIAM J. Numer. Anal. 23 (1986) 109–117.
[138] C.R. Vogel, Non-convergence of the L-curve regularization parameter selection method, Inverse Problems 12 (1996) 535–547.
[139] C.R. Vogel, Computational Methods for Inverse Problems, SIAM, Philadelphia, 2002.
[140] G. Wahba, Practical approximate solutions to linear operator equations when the data are noisy, SIAM J. Numer. Anal. 14 (1977) 651–667.
[141] G. Wahba, A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem, Ann. Statist. 13 (1985) 1378–1402.

[142] G. Wahba, Spline Models for Observational Data, SIAM, Philadelphia, 1990.
[143] G. Wahba, Y.H. Wang, When is the optimal regularization parameter insensitive to the choice of the loss function?, Comm. Statist. Theory Methods 19 (1990) 1685–1700.
[144] Y. Wang, Smoothing spline models with correlated random errors, J. Amer. Statist. Assoc. 93 (1998) 341–348.
[145] W.E. Wecker, C.F. Ansley, The signal extraction approach to nonlinear regression and spline smoothing, J. Amer. Statist. Assoc. 78 (1983) 81–89.
[146] A.G. Yagola, A.S. Leonov, V.N. Titarenko, Data errors and an error estimation for ill-posed problems, Inverse Probl. Eng. 10 (2002) 117–129.