Evaluation of Parameter Choice Methods for Regularization of Ill-Posed Problems in Geomathematics

Frank Bauer, Martin Gutting, and Mark A. Lukas

Contents

1 Introduction and Inverse Problems Considered . . . 1714
1.1 Introduction . . . 1714
1.2 Inverse Problems Considered . . . 1716
2 Preliminaries . . . 1717
2.1 Regularization Methods . . . 1718
2.2 Deterministic and Stochastic Noise . . . 1719
2.3 Assumptions on the Solution x (Source Condition) . . . 1720
2.4 Parameter Choice Method . . . 1720
2.5 Optimal Regularization Parameter . . . 1721
3 Evaluation Process and Experiments . . . 1723
3.1 Construction of Our Numerical Experiments . . . 1723
3.2 Error Comparison in Case of Optimal Solutions . . . 1729
4 Description and Evaluation of Methods . . . 1730
4.1 Discrepancy Principle . . . 1730
4.2 Transformed Discrepancy Principle . . . 1732
4.3 Modified Discrepancy Principle (MD Rule) . . . 1734
4.4 Monotone Error (ME) Rule . . . 1736
4.5 Varying Discrepancy Principle . . . 1739
4.6 Balancing Principle . . . 1740
4.7 Hardened Balancing Principle . . . 1743

F. Bauer
DZ Bank AG, Kapitalmärkte Handel, Quantitative Modelle, F/KHSQ, Frankfurt, Germany
e-mail: frank.bauer@dzbank.de

M. Gutting (✉)
Geomathematics Group, University of Siegen, Siegen, Germany
e-mail: martingutting@gmail.com

M.A. Lukas (✉)
Mathematics and Statistics, School of Engineering and Information Technology, Murdoch University, Murdoch, WA, Australia
e-mail: M.Lukas@murdoch.edu.au

© Springer-Verlag Berlin Heidelberg 2015

W. Freeden et al. (eds.), *Handbook of Geomathematics*,
DOI 10.1007/978-3-642-54551-1_99

4.8 Quasi-optimality Criterion . . . 1746
4.9 L-Curve Method . . . 1747
4.10 Extrapolated Error Method . . . 1750
4.11 Modified Discrepancy Partner (MDP) Rule . . . 1750
4.12 Normalized Cumulative Periodogram Method . . . 1752
4.13 Residual Method . . . 1753
4.14 Generalized Maximum Likelihood . . . 1755
4.15 Generalized Cross-Validation . . . 1757
4.16 Robust Generalized Cross-Validation . . . 1759
4.17 Strong Robust GCV . . . 1761
4.18 Modified Generalized Cross-Validation . . . 1763
4.19 Other Methods . . . 1764
5 Conclusion . . . 1765
5.1 Requirements and Properties . . . 1765
5.2 Average Performance . . . 1766
References . . . 1768

Abstract

Many different parameter choice methods have been proposed for regularization in both deterministic and stochastic settings. The performance of a particular method in a specific setting and its comparison to other methods is sometimes hard to predict. This chapter reviews many of the existing parameter choice methods and evaluates and compares them in a large simulation study for spectral cutoff and Tikhonov regularization.

The numerical tests deal with downward continuation, i.e., an exponentially ill-posed problem, which is found in many geoscientific applications, in particular those involving satellite measurements. A wide range of scenarios for these linear inverse problems are covered with regard to 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.

1 Introduction and Inverse Problems Considered

1.1 Introduction

Inverse problems are an important part of geomathematics, especially now in the age of satellite measurement technologies. Most applications want to use measurements made at satellite altitude (e.g., by satellite-to-satellite tracking (SST), satellite gravity gradiometry (SGG), magnetic field measurements) to determine a geoscientific quantity such as the gravitational potential or the magnetic field on the Earth's surface. This process involves some variant of downward continuation, which is known to be an exponentially ill-posed problem requiring regularization.

For further details, we refer to Freeden (1999), Freeden and Michel (2004), Freeden and Gerhards (2013), Lu and Pereverzev (2014), and the references therein.

These ill-posed problems can be expressed mathematically as a 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 (see, e.g., Engl et al. 1996).

The two most popular regularization methods are spectral cutoff regularization (also called truncated singular value decomposition) and Tikhonov regularization (also called ridge regression or Wiener filtering in certain contexts). Spectral cutoff uses projection onto a subspace spanned by singular vectors corresponding to the larger singular values of A, while Tikhonov regularization is a penalized variational approach which shrinks the contribution of the smaller singular values of A. For both regularization methods, the choice of the regularization parameter is crucial. Many different methods for choosing this parameter have been proposed, often with analytical justification in mind. However, this usually requires a particular setting, and it may not be known how the method performs in all practical situations and under general conditions.

The paper of Bauer and Lukas (2011) provides a comprehensive review of most of the existing parameter choice methods and compares them through a large simulation study. Other more limited comparative studies can be found in Wahba (1985), Kohn et al. (1991), Thompson et al. (1991), Hanke and Hansen (1993), Hansen (1994, 1998, Chap. 7), Lukas (1998b), Farquharson and Oldenburg (2004), Abascal et al. (2008), Åkesson and Daun (2008), Rust and O'Leary (2008), Correia et al. (2009), Hämarik and Raus (2009), Hämarik et al. (2009, 2011, 2012), Palm (2010), and Reichel and Rodriguez (2013).

Bauer and Lukas (2011) consider synthetic inverse problems with singular values having power-type decay. This paper is based on the survey of Bauer and Lukas (2011), but specifically treats exponentially ill-posed problems closely related to real applications in geomathematics. For these inverse problems, a thorough and up-to-date comparative study of parameter choice methods for spectral cutoff and Tikhonov regularization is provided, taking into account deterministic and stochastic settings. In particular, we investigate the variability of each method both for white noise and for colored noise. For ease of reference, we list below the subsections for the parameter choice methods that are considered. Further methods are briefly discussed in Sect. 4.19 where we also explain why they are left out in this paper.

4.1 Discrepancy principle
4.2 Transformed discrepancy principle
4.3 Modified Discrepancy Principle (MD rule)

4.4 Monotone error (ME) rule
4.5 Varying discrepancy principle
4.6 Balancing principle, balancing principle (white), fast balancing principle
4.7 Hardened balancing principle, hardened balancing principle (white)
4.8 Quasi-optimality criterion
4.9 L-curve method
4.10 Extrapolated error method
4.11 Modified discrepancy partner (MDP) rule
4.12 Normalized cumulative periodogram method
4.13 Residual method
4.14 Generalized maximum likelihood
4.15 Generalized cross-validation
4.16 Robust generalized cross-validation
4.17 Strong robust GCV
4.18 Modified generalized cross-validation

We introduce the downward continuation problem in Sect. 1.2. Section 2 summarizes regularization of linear inverse problems. In Sect. 3, the design of our numerical experiments and the main algorithms used for their generation are presented in brief; we refer to Bauer and Lukas (2011) for a more detailed explanation of those parts that are similar. All the parameter choice methods are briefly described in Sect. 4 and their corresponding results are displayed and commented on there. A summary and conclusions can be found in Sect. 5.

1.2 Inverse Problems Considered

Due to its compactness, the operator A of (1) admits a singular value decomposition (SVD) $\{\sigma_k, u_k, v_k\}_{k\in\mathbb{N}}$, 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. The decay rate of the singular values σ_k determines the degree of ill-posedness of the inverse problem. For the inverse problem of downward continuation, we have $\mathcal{X} = L^2(\Omega_R)$, i.e., the space of square-integrable functions on the sphere Ω_R of radius R (radius of a spherical Earth), and $\mathcal{Y} = L^2(\Omega_{R+h})$, i.e., the space of square-integrable functions on the sphere of radius $R + h$, where h is the satellite altitude (assuming a simplified spherical orbit). Then x stands for, e.g., the gravitational potential on the Earth's surface and y denotes the same potential at the satellite's orbit. The orthonormal systems are given in terms of spherical harmonics $Y_{m,j}$ of degree $m \in \mathbb{N}_0$ and order $j = -m, \ldots, m$ which are adjusted to the corresponding radii. Thus, $u_k = u_{k(m,j)} = \frac{1}{R} Y_{m,j}$, $v_k = v_{k(m,j)} = \frac{1}{R+h} Y_{m,j}$ and the singular values of the operator $A : \mathcal{X} \to \mathcal{Y}$ of (1) are given by

$$\sigma_k = \sigma_{k(m,j)} = \left(\frac{R}{R+h}\right)^m, \qquad m \in \mathbb{N}_0, j = -m, \ldots, m. \tag{2}$$

Note that this operator is the so-called upward continuation, its (generalized) inverse is the downward continuation. Obviously, the singular values in (2) are exponentially decreasing, i.e., downward continuation is an exponentially ill-posed problem. Since there are $2m + 1$ spherical harmonics of degree m and the singular values do not depend on the order j, the singular values occur with this multiplicity. Note that we use the notation $k(m, j)$ in (2) to indicate the renumbering, which is given by $k(m, j) = m^2 + m + 1 + j$. It is also well known that there are $(M + 1)^2$ spherical harmonics of degree M or less (see, e.g., Freeden and Gutting (2013) for a detailed introduction of spherical harmonics and their properties). It should also be noted that instead of L^2-spaces the theory of spherical Sobolev spaces can be applied.

The SST and SGG problems that are mentioned in the introduction can be investigated in simplified form such that y is the first or second radial derivative of the gravitational potential instead of the potential itself. This provides the following singular values for $m \in \mathbb{N}$, $j = -m, \ldots, m$:

$$\sigma^{\mathrm{SST}}_{k(m,j)} = \frac{m+1}{R+h}\left(\frac{R}{R+h}\right)^m, \qquad \sigma^{\mathrm{SGG}}_{k(m,j)} = \frac{(m+1)(m+2)}{(R+h)^2}\left(\frac{R}{R+h}\right)^m.$$

In both cases, the singular values are still exponentially decreasing as k or m tend to infinity. Further details can be found in, e.g., Freeden (1999), Freeden and Michel (2004), Freeden and Gerhards (2013), Freeden and Schreiner (2015), Lu and Pereverzev (2014), and the references therein. Another ill-posed problem closely related to downward continuation is the inverse gravimetric problem (see Freeden and Michel (2004), Michel (2014), and the references therein). While the problems discussed above are exponentially ill-posed in a formal sense, in practice, where there is a bound on the degree of the spherical harmonics, the problems may exhibit behavior of more moderately ill-posed problems. In particular, it is argued in Lu and Pereverzev (2014) that the SGG problem for the GOCE mission, where the degree is bounded by 300, can be considered as mildly (polynomially) ill-posed, with $\sigma^{\mathrm{SGG}}_{k(m,j)} \approx (m + 1/2)^{-5.5}$ for $100 \le m \le 300$. This observation also applies to the finite dimensional problems to be considered in our numerical experiments. For example, with $R = 6{,}371$ km and $h = 300$ km, the eigenvalues $\sigma_{k(m,j)}$ in (2) satisfy $0.5R^4(m + 1/2)^{-8.5} \le \sigma_{k(m,j)} \le 3R^4(m + 1/2)^{-8.5}$ for $100 \le m \le 300$.

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. Starting with (1), we distinguish the following three cases which are considered 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 $\mathcal{Y} = \mathbb{R}^q$, $p \leq q$. Assume that $\operatorname{rank}(A) = p$. This case is often called a discrete ill-posed problem.

Case C3. Discrete data situation, where the underlying problem is still infinite dimensional, but we only have measurements $y_i = y(t_i)$ at the q points $\{t_i\}_{i=1,\ldots,q}$. Then A is a finite rank operator between $\mathcal{X}$ and $\mathcal{Y} = \mathbb{R}^q$ with the correspondence $(Ax)_i = (A_\infty x)(t_i)$, $i = 1, \ldots, q$. Assume that $\operatorname{rank}(A) = q$. 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 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 cutoff and Tikhonov regularization. Detailed accounts of these methods can be found, e.g., in Groetsch (1984), Hofmann (1986), and Engl et al. (1996).

By the use of the singular value decomposition, we obtain

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

where R is ∞, p, and q, respectively, in cases C1, C2, and C3 above. Spectral cutoff regularization is defined by

$$x_n^\delta = \sum_{k=1}^{l(n)} \sigma_k^{-1} \langle y^\delta, v_k \rangle u_k, \tag{3}$$

where $l(\cdot)$ is an ascending integer-valued function. The traditional choice is $l(n) = n$, but a general $l(\cdot)$ 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 $\alpha_n = \alpha_0 q_\alpha^n$, where $0 < q_\alpha < 1$ and $n \in \mathbb{N}$. 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

$$x_n^\delta = (A^* A + \alpha_n I)^{-1} A^* y^\delta, \tag{4}$$

where A^* is the adjoint of A. 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 cutoff 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 Deterministic and Stochastic Noise

Several kinds of additive noise models are used in the study of inverse problems (cf. Eggermont et al. 2014 and the references therein). To describe the major ones, we will denote $y^\delta = y + \delta\xi$, where ξ is a normalized noise element and $\delta > 0$ is the noise level.

The most common noise model in the classical inverse problems literature is deterministic noise (cf. Engl et al. 1996), where $\xi \in \mathcal{Y}$ with $\|\xi\| \le 1$, so $\|y^\delta - y\| \le \delta$. 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 noise model for a discrete data vector $y^\delta \in \mathbb{R}^q$ (for cases C2 and C3, see, e.g., Wahba 1990) 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 $\mathbb{E}\|y^\delta - y\|^2 = \delta^2\mathbb{E}\|\xi\|^2 = q\delta^2$. This model can be extended to one involving correlated errors, where $\varepsilon_i = \delta\xi_i$ 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

$$y_k^\delta = \langle y, v_k\rangle + \delta\xi_k = \sigma_k \langle x, u_k\rangle + \delta\xi_k, \quad k \in \mathbb{N}, \tag{5}$$

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 (cf. Lepskij 1990). 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 $f, g \in \mathcal{Y}$ with $f_k = \langle f, v_k\rangle$ and $g_k = \langle g, v_k\rangle$. 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 $\xi \sim N(0, I)$, then, clearly, using the orthonormal singular vectors $v_k \in \mathbb{R}^q$ of A, the model can be written equivalently as a Gaussian white noise model with finite sequence data.

2.3 Assumptions on the Solution *x* (Source Condition)

In most of the literature on regularization, it is assumed that x is a fixed (nonrandom) element of $\mathcal{X}$. It is known (see, e.g., Cox 1988; Lukas 1988; Engl et al. 1996) 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 (3) or (4) 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 $s > 0$ (called Hölder type). However, this form is not suitable for a severely ill-posed problem, where the singular values of the operator have exponential decay, while the unknown solution has relatively low smoothness, in particular for downward continuation. For such problems, it is natural to use the logarithmic source condition $x \in \{u : u = \ln^{-p}(A^*A)^{-1}v, \|v\| \leq K\}$ (cf. Mair 1994; Hohage 2000; Pereverzev and Schock 2000). General source conditions are also used in Mathé and Pereverzev (2003) and Nair et al. (2003).

In the Bayesian approach to inverse problems (cf. Tarantola 1987; Evans and Stark 2002; Kaipio and Somersalo 2005; Hofinger and Pikkarainen 2007), 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 (cf. Larkin 1972; Hofmann 1986; Wahba 1990; Fitzpatrick 1991) 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 and the appropriate choice for the regularization parameter is determined by the variance of x and the noise level δ. However, in practice both are usually unknown.

2.4 Parameter Choice Method

A parameter choice method is a rule that assigns a value for the regularization parameter. In the case of a discrete set of parameters, the method selects a value for the index, which will be denoted by n_*. Parameter choice methods can be classified according to the input used to make the choice. There are three basic types (see, e.g., Engl et al. 1996; Bauer and Kindermann 2009):

- A priori method, i.e., n_* is a function of δ and information about the smoothness of x. Since information about x is required, but in practice not known, such methods are not discussed here.
- A posteriori method, i.e., $n_* = n_*(\delta, y^\delta)$. The noise level δ is required and has to be estimated if it is not known.
- Data-driven method, i.e., $n_* = n_*(y^\delta)$. The method only requires 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. Each method computes an associated function $F(n)$, and n_* is defined as either the point at which F falls below a threshold (type 1) or the minimizer of F (type 2). Most methods of type 1 originate from a deterministic setting and use a (sensitive) tuning parameter. The methods of type 2 mostly come from a stochastic framework or from heuristic ideas and work without (sensitive) tuning parameters.

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

$$\|x - x_{n_*}^\delta\| / \|x - x_{n_{opt}}^\delta\|. \tag{6}$$

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/estimated quantities. Many methods use the norm of the residual defined as $\|y^\delta - Ax_n^\delta\|$. 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.

Splitting the error $\|x - x_n^\delta\|$ such that

$$\|x - x_n^\delta\| \le \|x - x_n^0\| + \|x_n^0 - x_n^\delta\|, \tag{7}$$

the first term (regularization error) is bounded by a decreasing function $\varphi(n)$ which depends on the source condition and the regularization method. A higher smoothness s for $x \in \mathcal{R}((A^*A)^s)$ improves the rate of decrease. If this is only possible up to a maximal value s_0, this value is the so-called qualification and the method exhibits saturation. Tikhonov regularization has qualification $s_0 = 1$, while spectral cutoff has infinite qualification (see, e.g., Engl et al. 1996). Under a logarithmic source condition, slower (logarithmic) rates of decrease are obtained for standard regularizations, in particular for Tikhonov and spectral cutoff regularization (Mair 1994; Hohage 2000; Pereverzev and Schock 2000). It follows that Tikhonov regularization does not exhibit saturation here.

The second term (propagated noise error) on the right-hand side of (7) can often be bounded for regularization methods as

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

where ϱ is a known increasing function of n, indicating that, with less smoothing, there is more influence of the data noise. For spectral cutoff regularization, (8) holds with $\varrho(n) = \sigma_{l(n)}^{-1}$ and, for Tikhonov regularization, (8) holds with $\varrho(n) = \alpha_n^{-1/2}$ (cf. Engl et al. 1996).

In the case of stochastic noise, the risk, i.e., the expected squared error $\mathbb{E}\|x - x_n^\delta\|^2$, is considered. For noise with zero mean, instead of (7), the risk can be decomposed exactly into a sum of squared bias $\|x - x_n^0\|^2$ and variance terms $\mathbb{E}\|x_n^0 - x_n^\delta\|^2$, i.e.,

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

The squared bias can be bounded as before and, under suitable assumptions, the variance can be expressed as $\delta^2\varrho^2(n)$ for some increasing function $\varrho(n)$. For white noise, the spectral cutoff solution (3) has variance

$$\delta^2\varrho^2(n) = \delta^2\mathbb{E}\|A_n^{-1}\xi\|^2 = \delta^2\sum_{k=1}^{l(n)}\sigma_k^{-2} \tag{10}$$

and the Tikhonov regularized solution (4) has variance

$$\delta^2\varrho^2(n) = \delta^2\mathbb{E}\|A_n^{-1}\xi\|^2 = \delta^2\sum[\sigma_k/(\sigma_k^2 + \alpha_n)]^2. \tag{11}$$

A much more detailed discussion of the above errors (including minimax results) in various situations can be found in Cox (1988), Lukas (1988), Engl et al. (1996), Mair and Ruymgaart (1996), Mathé and Pereverzev (2003), Cavalier et al. (2004), Bauer (2007), and Hofmann and Mathé (2007).

It should be pointed out that Bakushinskii (1984) 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 regularization method such that the worst-case error converges to 0 as $\delta \to 0$. This Bakushinskii veto is important for deterministic noise, but it is not really appropriate for stochastic noise (cf. Bauer and Kindermann 2009; Becker 2011). 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 (see also Bauer and Lukas 2011).

For some methods, there are stronger results involving oracle inequalities (see Cavalier et al. 2002; Candès 2006; Bauer and Reiß 2008; Cavalier 2008), which provide, for any noise level, a bound on the risk $\mathbb{E}\|x - x_n^\delta\|^2$ relative to the smallest possible value of the risk and allow the classification of methods as asymptotically optimal.

Similar results exist for some methods with a discrete noisy data vector $y^\delta \in \mathbb{R}^q$ in case C3, where the asymptotic analysis is as $q \to \infty$ 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 ($A = I$) and other problems, it is known 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 $q \to \infty$ (cf. Brown and Low 1996; Pensky and Sapatinas 2010).

3 Evaluation Process and Experiments

This section presents the design of the evaluation process and numerical experiments that will be used for each parameter choice method described in Sect. 4.

3.1 Construction of Our Numerical Experiments

Each parameter choice method will be assessed using the same large set of test problems corresponding to downward continuation. 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® whose notation we also use to describe the algorithms in this section.

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

- $M_{\max}$: maximal spherical harmonic degree, i.e., the number of singular values of A is $q = (M_{\max} + 1)^2$;
- h: varying heights in (2) (with $R = 6{,}371\,\text{km}$) determining the speed of the exponential decay behavior;
- ν : (polynomial) decay behavior (smoothness) of x;
- $\log(N2S)$: $\log_{10}$ of the noise-to-signal ratio $N2S = (\mathbb{E}\|y^\delta - y\|^2)^{1/2}/\|y\|$;
- ω : noise behavior (corresponds to the color).

Generation of the Operators

The operator A will be taken to be a random diagonal $q \times q$ matrix, with diagonal elements (i.e., eigenvalues or singular values) decaying like

$$a_{kk} \approx \sigma_k = \sigma_{k(m,j)} = \left(\frac{R}{R+h}\right)^m, \quad m \in \mathbb{N}_0, j = -m, \ldots, m,$$

where $k(m, j) = m^2 + m + 1 + j$. A larger value of h corresponds to a more ill-posed problem, all of which are exponentially ill-posed. The diagonal vector is generated by the following procedure:

```
for deg = 0 : M_max + 10
   OrderVector(deg^2 + 1 : (deg + 1)^2) = deg;
end
NumberEigenvalues = (M_max + 1)^2;
HelpVar = (6,371/(6,371 + h)).^OrderVector;
Perturb = exp(0.5 * randn(length(HelpVar), 1) − 0.5^2/2);
HelpOp = sort(HelpVar .* Perturb, 'descend');
A = HelpOp(1 : NumberEigenvalues);
```

For each parameter choice method, more than 160,000 trials are done using about 40,000 up to 120,000 eigenvalues each time, so speed and memory usage are very important issues. Note that for the maximal degree $M_{\max}$ we have to deal with $(M_{\max} + 1)^2$ eigenfunctions and $M_{\max}$ can be up to 350 (see Tables 1 and 2) in our tests. For this reason, we use the simplest possible form of an inverse problem, i.e., diagonal matrices. 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. In addition, for Gaussian white noise, there is no

Table 1 Test cases for Tikhonov regularization

h	$M_{\max}$	ν	$\log(N2S)$
300	350	2	−4.0
300	350	2	−6.0
300	350	3	−6.0
300	350	3	−8.0
300	350	4	−8.0
300	350	4	−10.0
300	350	5	−8.0
300	350	5	−10.0
500	250	2	−4.0
500	250	2	−6.0
500	250	3	−6.0
500	250	3	−8.0
500	250	4	−8.0
500	250	5	−8.0
500	250	5	−10.0
700	200	2	−6.0
700	200	3	−6.0
700	200	3	−8.0
700	200	4	−8.0
700	200	4	−10.0

Table 2 Test cases for ExpCutOff regularization

h	$M_{\max}$	ν	$\log(N2S)$
300	300	2	−4.0
300	300	2	−6.0
300	300	3	−6.0
300	300	3	−8.0
300	300	4	−8.0
300	300	4	−10.0
300	200	5	−8.0
300	200	5	−10.0
500	250	2	−4.0
500	250	2	−6.0
500	250	3	−6.0
500	250	3	−8.0
500	250	4	−8.0
500	200	5	−8.0
500	200	5	−10.0
700	200	2	−6.0
700	200	3	−6.0
700	200	3	−8.0
700	200	4	−8.0
700	200	4	−10.0

difference at all in using diagonal matrices to investigate stochastic behavior. This follows from the invariance of white noise under the orthogonal transformation involved in the SVD.

We also 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"; see, e.g., Colton and Kress 1998). Therefore, we used a slight random perturbation of the sequence σ_k. The procedure balances retaining the overall exponential decay behavior of the operator including the multiplicity of its singular values while providing some randomness in this component.

Solution Generation

Each time a solution x is generated, we use the following procedure (the variables $OrderVector$ and $NumberEigenvalues$ are the same as in Sect. 3.1:

$$
\begin{aligned}
&HelpVar = (OrderVector(1 : NumberEigenvalues) + 1).\hat{}(-\nu);\\
&Sign = 2 * \mathrm{ceil}(2. * \mathrm{rand}(NumberEigenvalues, 1)) - 3;\\
&Perturb = 1 + 0.1 * \mathrm{randn}(NumberEigenvalues, 1)\\
&x = Sign. * Perturb. * HelpVar;
\end{aligned}
$$

This can be interpreted as generating random Fourier coefficients x_k with decay behavior $|x_k| = |x_{k(m,j)}| \approx m^{-\nu}$ and random sign of equal probability. A larger value of ν gives a smoother solution x.

For such polynomial decay rates, the Sobolev lemma allows an interpretation in terms of differentiability, i.e., smoothness. In our setting with $2m+1$ eigenfunctions

(spherical harmonics) of polynomial degree m, this means that we require a decay of the order of $m^{-1-\varepsilon}$, for some $\varepsilon > 0$ (i.e., $\nu > 1$), to guarantee x is square integrable. Moreover, it is known (see, e.g., Freeden et al. 1998; Freeden 1999; Freeden and Michel 2004) that for $\nu > r + 2$ the function defined by the Fourier coefficients x_k, $|x_k| = |x_{k(m,j)}| \approx m^{-\nu}$, corresponds to an $x \in C^{(r)}$, $r \in \mathbb{N}_0$. Therefore, we only consider $\nu \geq 2$ in our tests.

An alternative way to generate solutions of different smoothness is through a source condition. However, using the standard source condition to define $x = (A^*A)^s x_0$ for some $x_0 \in \mathcal{X}$ is not suitable for the downward continuation problem as it would imply $x \in C^{(\infty)}$ which is of less interest in practice. A more appropriate logarithmic source condition could be used to generate the solutions, but it is easier to use the direct approach above. Furthermore, our approach also allows encapsulation in the software design, i.e., the solution should not "see" the operator and vice versa. Note that, because the theory of saturation is based on source conditions, it does not apply directly for our assumed form of x. Nevertheless, we see similar results and will call these saturation effects.

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 characterized by the parameter ν, and also the effect of the noise-to-signal ratio $N2S$. To avoid too much variability in these features of x for different replicates of x, we chose not to use a colored Gaussian variable for x as in Kaipio and Somersalo (2005) and Bauer (2007). 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.

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 $\delta = N2S * \|y\|/\sqrt{q}$. Then, each time a noise vector is generated, we employ the same procedure as in Bauer and Lukas (2011): 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. By adding the noise in a different space from the one used to generate $y = Ax$, we avoided potential inverse crimes.

For $\omega = 0$, the noise is simply Gaussian white noise. For $\omega \neq 0$, the noise is defined by a moving average process in which both the weights and order depend on ω. The noise has higher correlation for larger ω, and it has color red for $\omega > 0$ and color blue for $\omega < 0$ (the algorithm can be found in Bauer and Lukas 2011). This type of correlation is observed in many practical applications, for example, in geodesy (cf. Ditmar et al. 2003, 2007).

For our experiments, we considered two scenarios for the noise:

- White noise, i.e., $\omega = 0$
- Colored noise of unknown random color, i.e., ω is chosen as a pseudorandom variate, uniformly distributed in the interval $[-0.5, 0.5]$.

Regularization Operators

For the Tikhonov regularization parameter sequence $\alpha_n = \alpha_0 q_\alpha^n$, $n \in \mathbb{N}$, in (4), we used 100 values from 1 to 10^{-14} with logarithmic equal spacing, i.e., $\alpha_0 = 1.3849$ and $q_\alpha = 0.7221$. The set of test cases 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 1 and 10^{-14}, i.e., it is not near an endpoint.

For the sequence $l(n)$ of cutoff points in (3), i.e., for spectral cutoff regularization, we used the same procedure as in Bauer and Lukas (2011). The cutoff points are chosen such that there is a minimum spacing of 3 and, more importantly, the cutoff point $n+1$ is the first to possess an eigenvalue $\sigma_{k(n+1)} < \sigma_{k(n)}/1.07$, where $\sigma_{k(n)}$ is the eigenvalue corresponding to the cutoff point n. By this, we can avoid in most cases that there are two cutoff points corresponding to the same spherical harmonic degree m. This method is referred to as exponential spectral cutoff (ExpCutOff), and it also achieves the minimax rate for optimal parameter choice (cf. Bauer and Pereverzev 2005).

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 not near an endpoint.

Maximal Regularization Parameter

For most parameter choice methods, the use of a discrete set of regularization parameters with a fine enough resolution, as chosen above, does not alter the behavior of the method. 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 if the method minimizes some function.

For a few parameter choice methods, e.g., the balancing principle (Sect. 4.6), 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. For both spectral cutoff and Tikhonov regularization in a stochastic setting, it is reasonable to define the maximal index as

$$N = \max\{n | \varrho(n) < 0.5\varrho(\infty)\}, \tag{12}$$

where $\mathbb{E}\|x_n^0 - x_n^\delta\|^2 = \delta^2\varrho^2(n)$ and $\delta^2\varrho^2(\infty)$ is the supremum of the variance. We expect that $n_{opt} < N$ because if $n > N$, then it follows from (9) that $\mathbb{E}\|x - x_n^\delta\|^2 > 0.5\mathbb{E}\|x - x_\infty^\delta\|^2$, but 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 (10) and (11) for white noise, or a good estimate of it. For any noise color, if several independent data sets are available, then a good estimate is

$$\delta^2\varrho^2(n) \approx 2^{-1}\text{Mean}\{\|x_{n,i}^\delta - x_{n,j}^\delta\|^2, i \neq j\}. \tag{13}$$

In the experiments, we use four data sets to obtain N for the methods that require a maximal parameter. Often, two sets of data are sufficient (see Bauer 2007 for further details).

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, but for some methods, it has the beneficial effect of reducing the number of severely under-smoothed solutions (see also Bauer and Lukas 2011).

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 $l(N_1) = q$ for spectral cutoff and by $\alpha_{N_1} \approx \sigma_q^2$ 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 (see Bauer and Lukas 2011).

Runs and Organization of the Results

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 random colored noise, we generated 4 operators A as described in Sect. 3.1.
- For each operator A, we generated 8 solutions x as described in Sect. 3.1.
- For each pair of (A, x), we generated 8 times 8 different noisy data vectors y^δ as described in Sect. 3.1. In the colored noise scenario, each group of 8 data vectors has a different, randomly chosen color, and within the group, the color is the same.

This means that, for each test case and noise scenario, there are 2,048 inverse problems that need to be solved. The hierarchical structure was chosen in order to considerably reduce the computational cost. In total, this chapter is based on the solution of more than 3.4 million inverse problems each with an operator of at least 40,000 singular values, in most cases more than twice that number.

For each parameter choice method, we will display the simulation results, i.e., the inefficiencies as defined by (6), 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 $(h, \nu, \log(N2S))$, with $M_{\max}$ determined from Tables 1 and 2), a box plot (marking the lower and upper quartiles) shows the distribution of computed inefficiencies for the method, with blue whiskers showing the range. The whiskers have maximum length of 4 times the interquartile range, and outliers beyond this are marked with a red + symbol.
- For each test case, the red middle band in the box shows the median of the computed inefficiencies, and an open green dot shows the sample mean.

3.2 Error Comparison in Case of Optimal Solutions

To conclude this section, Fig. 1 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 (see Tables 1 and 2). The box plots are constructed in the same way as described in Sect. 3.1.

The optimal errors differ by several orders of magnitude across the test cases, mostly because of the different noise levels. In addition, there can be significant variability within one test case, especially for the colored noise scenario. 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 can see some saturation effects for Tikhonov regularization. The errors for the parameter sets $(300, 4, -8)$ and $(300, 5, -8)$ or $(500, 4, -8)$ and $(500, 5, -8)$ (to a lesser degree also $(300, 4, -10)$ and $(300, 5, -10)$) indicate there is no improvement as the smoothness increases beyond a certain value. Therefore, for our finite dimensional exponentially ill-posed problems, Tikhonov regularization behaves in a similar way as for a polynomially ill-posed problem, in line with the discussion at the end of Sect. 1.2.

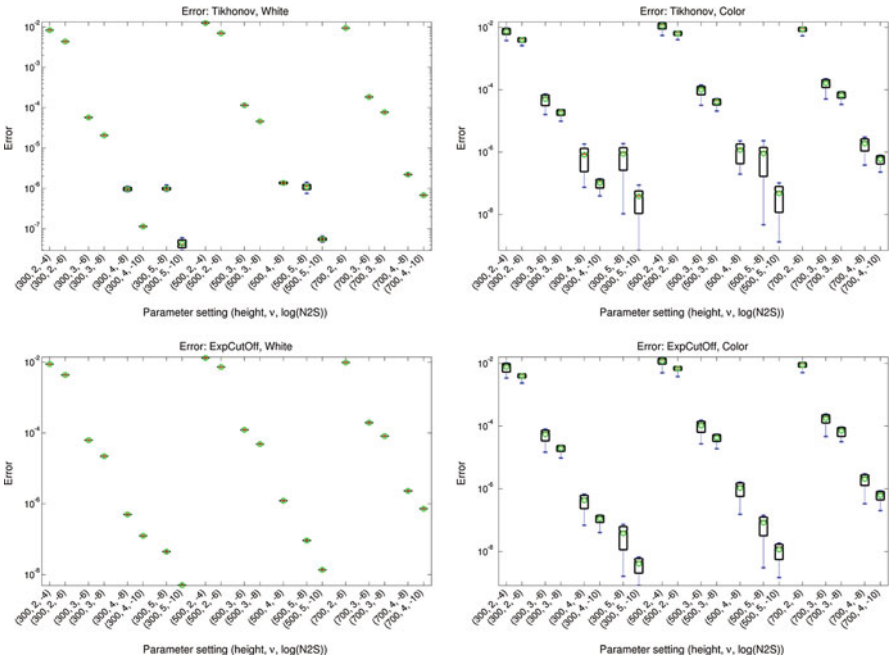

Fig. 1 Comparison of optimal errors

4 Description and Evaluation of Methods

In this section, we will review most of the available parameter choice methods and evaluate them according to the process outlined in Sect. 3. For each method, we will start by describing the origin and rationale of the method. Then we will state the input 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.

The methods are ordered as follows:

- The first group requires knowledge of the noise level δ. We provide the correct δ in all our tests even though this is usually not available in practice. An estimation of the noise level might induce further errors.
- The second group requires at least two independent sets of data as input.
- The third group requires no knowledge about the noise.

Many of the methods use a tuning parameter or some other parameter that must be chosen. For each of these methods, we have used the tuning parameters of Bauer and Lukas (2011) to allow an easier comparison of the results. However, the setting might not be optimal for the problems here, and the optimal settings for different problems might vary significantly. Section 4.19 contains a list of methods that were not included in this study.

4.1 Discrepancy Principle

The discrepancy principle, which was originally proposed by Phillips (1962) and then developed and analyzed by Morozov (1966, 1984), is one of the oldest and most widely used parameter choice procedures (cf. Engl et al. 1996 and references therein). The rationale 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 (see, e.g., Davies and Anderssen 1986; Lukas 1995; Vogel 2002) and a continuous stochastic setting (see Blanchard and Mathé 2012).

Method The method needs the following input:

- Norms of residuals $\{Ax_n^\delta - y^\delta\}_{n \le N}$ until a certain bound is satisfied
- Noise level δ
- Tuning parameter τ

In a deterministic setting with $\|y^\delta - y\| \le \delta$, the parameter choice n_* is the first n such that $\|Ax_n^\delta - y^\delta\| \le \tau\delta$, where $\tau \ge 1$ is a tuning parameter. In our stochastic

setting, with the error in each element of $y^\delta \in \mathbb{R}^q$ having standard deviation δ, the choice n_* is the first n such that

$$\|Ax_n^\delta - y^\delta\| \le \tau\delta\sqrt{q} \tag{14}$$

where we use $\tau = 1.5$.

Known Issues There has been a lot of work done on the convergence properties of this method (see, e.g., Groetsch 1984; Morozov 1984; Engl et al. 1996). The convergence rate in the deterministic setting for Tikhonov regularization is at the optimal order $O(\delta^{2s/(2s+1)})$ if $s \in (0, 1/2]$, but at the suboptimal order $O(\delta^{1/2})$ if $s > 1/2$. For spectral cutoff, the discrepancy principle gives the optimal order for any value of s. Under a logarithmic source condition, the discrepancy principle attains the optimal order for both Tikhonov and spectral cutoff regularization (see Hohage 2000, Pereverzev and Schock 2000; Nair et al. 2003; Mathé 2006). In the stochastic setting with a data vector $y^\delta \in \mathbb{R}^q$ containing uncorrelated errors of variance δ^2, for Tikhonov regularization, it is known (cf. Davies and Anderssen 1986; Lukas 1995; Vogel 2002) that as the sample size $q \to \infty$, 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, but otherwise it is order suboptimal (under-smoothing). Lukas (1998b) shows that for $\tau = 1$, the actual estimate is asymptotically unstable in a relative sense. For spectral cutoff, the "expected" estimate is order optimal for both the prediction risk and the $\mathcal{X}$-norm risk (cf. Vogel 2002). For a preconditioned Gaussian white noise model and general linear regularization, Blanchard and Mathé (2012) show that the discrepancy principle yields suboptimal rates as $\delta \to 0$.

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 (see Hansen 1998, Chap. 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 (see Hanke and Hansen 1993; Engl et al. 1996; Engl and Scherzer 2000; Bakushinsky and Smirnova 2005; Kaltenbacher et al. 2008; Jin and Tautenhahn 2009; and the references therein). It has also been analyzed for conjugate gradient iteration in a continuous stochastic setting by Blanchard and Mathé (2012).

Results From Fig. 2, for white noise, the discrepancy principle performs reasonably well for spectral cutoff regularization, as well as for Tikhonov regularization with less smooth solutions. However, significant saturation effects can be seen for $\nu \ge 4$. The results for colored noise in Fig. 2 are mediocre, with quite a bit of variation. However, the performance for Tikhonov regularization is not much worse than it is for white noise.

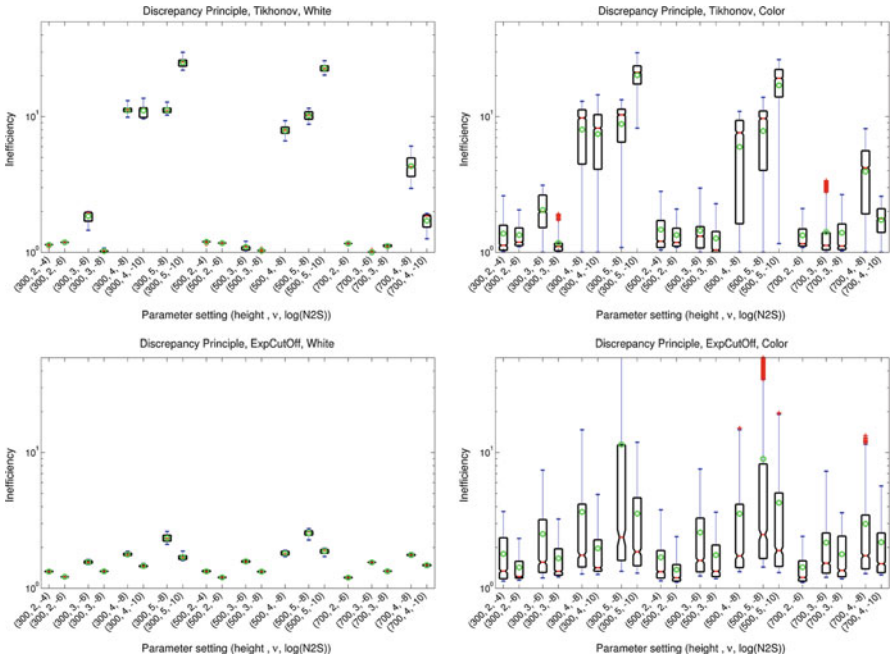

Fig. 2 Inefficiencies of the discrepancy principle

4.2 Transformed Discrepancy Principle

Motivated by the instability of the discrepancy principle to an incorrect noise level, Raus (1990, 1992) and Hämarik and Raus (2006) developed a parameter choice method in a deterministic setting where the noise level in the data y^δ is known only approximately as $\hat{\delta}$.

Method The transformed discrepancy principle needs the following input:

- Norms of transformed residuals $\{C(Ax_n^\delta - y^\delta)\}_{n \le N}$ where the operator C depends on A, the regularization method, and its qualification
- Rough estimate $\hat{\delta}$ of the noise level δ
- Tuning parameter b

For the stochastic case with $y^\delta \in \mathbb{R}^q$, 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

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

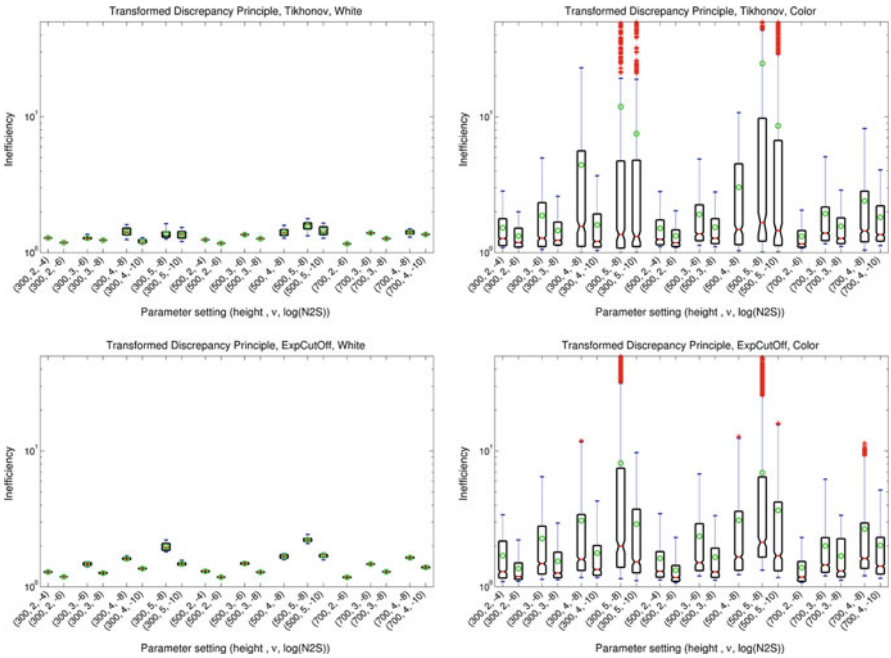

Fig. 3 Inefficiencies of the transformed discrepancy principle

where $\alpha_n = \alpha_0 q_\alpha^n$ and b is some constant satisfying $b > \gamma = ((1/4)^{1/4}(3/4)^{3/4})^2 \approx 0.3248$. For spectral cutoff, one computes n_* as the least integer n for which

$$\|A^*(Ax_n^\delta - y^\delta)\| \leq b\hat{\delta}\sqrt{q}\ \sigma_{l(n)}\,, \tag{16}$$

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

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 Raus (1992) 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 in this case. No knowledge of the solution smoothness is required. A stronger result involving a deterministic oracle inequality was shown in Raus and Hämarik (2007) for the usual noise assumption. The method was also defined and shown to be convergent for problems where the

operator is only known approximately as A^η, where $\|A^\eta - A\| \leq \eta$ (Raus 1992). Like the discrepancy principle, the method can be applied easily to iterative regularization methods.

Results Figure 3 shows that the transformed discrepancy principle has good to acceptable performance for all cases with white noise. Consistent with the theory, there is no saturation effect for Tikhonov regularization, so in many cases the results are significantly better than those for the discrepancy principle. However, there is substantial variation for some cases in the colored noise scenario. The results for spectral cutoff are slightly better than those of the discrepancy principle. In all situations, the results are very similar to those of the ME rule (Sect. 4.4).

4.3 Modified Discrepancy Principle (MD Rule)

The modified discrepancy principle (MD rule) was developed by Gfrerer (1987) and Engl and Gfrerer (1988) and independently by Raus (1984, 1985) for Tikhonov regularization and general regularization methods in a continuous, deterministic setting (see also Engl 1993 and Sects. 4.4 and 5.1 in Engl et al. 1996). The basic idea of the method is to minimize a bound on the squared error of the regularized solution derived from (7) to yield a practical a posteriori method with optimal convergence rates. It is also known as the Raus–Gfrerer rule or the minimum bound method. In Lukas (1998a), the method was adapted to the discrete, stochastic setting for Tikhonov regularization.

The method 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 (7) a bound on the squared error of the form

$$\|x - x_\alpha^\delta\|^2 \leq 2(\varphi^2(\alpha, y) + \delta^2\varrho^2(\alpha)). \tag{17}$$

The minimizer of (17) is defined by

$$f(\alpha, y) = -(\varphi^2)'(\alpha, y)/(\varrho^2)'(\alpha) = \delta^2. \tag{18}$$

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.

Method To use the MD rule, we need to be able to compute $(\varphi^2)'(\alpha, y)$ and $(\varrho^2)'(\alpha)$ which can be done effectively for Tikhonov and other regularization methods (see Engl and Gfrerer 1988 and Engl et al. 1996, Sect. 5.1). The method can also be applied to regularization methods with a discrete parameter, including spectral cutoff, with the derivatives above replaced by differences. The required input consists of the following:

- All regularized solutions $\{x_n^\delta\}_{n\le N}$ until a certain bound is satisfied
- Noise level δ
- Tuning parameter τ

For Tikhonov regularization, the function $\varrho^2(\alpha)$ in the bound (17) is $\varrho^2(\alpha) = \alpha^{-1}$, and the parameter choice is defined by

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

Using the discrete set $\{\alpha_n = \alpha_0 q_\alpha^n\}$, we can approximate the derivative $dx_\alpha^\delta/d\alpha$ on the left-hand side with $(x_n^\delta - x_{n+1}^\delta)(-\alpha_n \log q_\alpha)^{-1}$. For spectral cutoff regularization, the term $\delta^2\varrho^2(\alpha)$ in (17) 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

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

where

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

For the tuning parameter, we use $\tau = 1.5$ for Tikhonov regularization and $\tau = 0.5$ for spectral cutoff regularization. Note that, for spectral cutoff, this rule is different from the MD rule defined in Raus and Hämarik (2007) since the latter is just the usual discrepancy principle for this regularization method.

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, and it does so without any knowledge of the smoothness of the solution x (cf. Gfrerer 1987; Engl and Gfrerer 1988). The method also yields optimal rates for finite dimensional implementations (cf. Groetsch and Neubauer 1989) and when the operator is only known approximately (cf. Neubauer 1988). A deterministic oracle inequality is known for the MD rule with Tikhonov regularization (Raus and Hämarik 2007; Hämarik et al. 2012).

The discrete, stochastic version of the method, with a particular tuning constant, is asymptotically (as $q \to \infty$) equivalent to an unbiased risk method (see Lukas 1998a; Cavalier et al. 2002), 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 Lukas (1998b), it is shown asymptotically and by simulations that both of these methods are unstable and have high variability (see also Cavalier and Golubev (2006) for the unbiased risk method). By changing the tuning parameter, it is possible to improve the stability of the method.

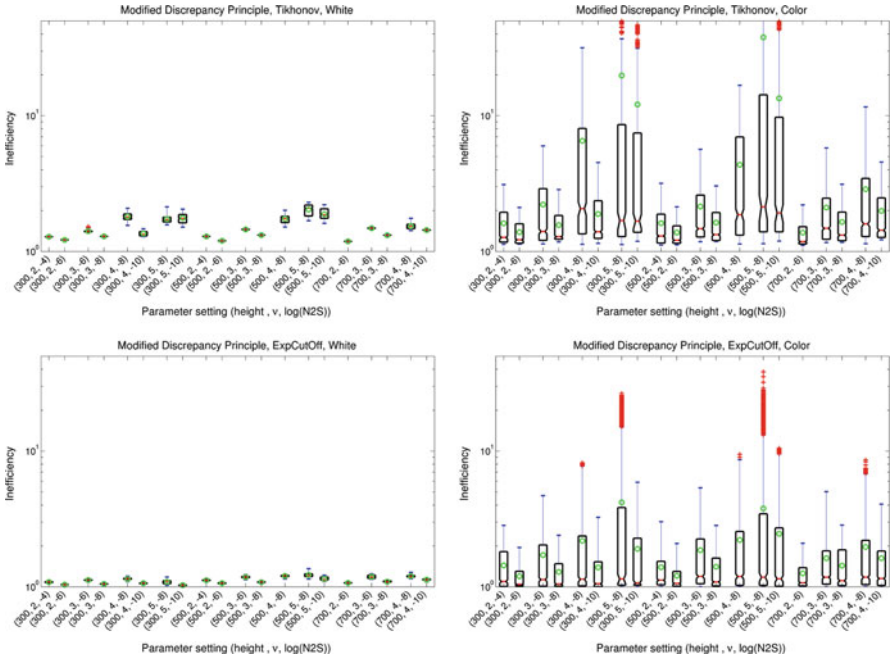

Fig. 4 Inefficiencies of the MD rule

The modified discrepancy principle can also be applied to iterative regularization methods for linear problems, and it achieves optimal convergence rates in a deterministic setting (cf. Engl and Gfrerer 1988). For nonlinear problems, the method has been extended for Tikhonov regularization and shown to yield optimal rates (see Scherzer et al. 1993; Jin and Hou 1999). In addition, it has been proposed in Jin (2000) as the stopping rule for the iteratively regularized Gauss–Newton method yielding again optimal rates.

Results As seen in Fig. 4, the MD rule performs very well in the white noise situation, especially for spectral cutoff regularization. However, it does not perform quite so well for colored noise; in particular, test cases with smoother solutions show a lot of variability. Consistent with the theory, the median results are much better than those of the discrepancy principle in the cases with smoother solutions.

4.4 Monotone Error (ME) Rule

The monotone error (ME) rule was proposed in Alifanov and Rumyantsev (1979), Hämarik and Raus (1999), and Tautenhahn and Hämarik (1999) for various regularization methods in a deterministic setting, and it was extensively discussed

along with other similar parameter choice rules by Hämarik and Tautenhahn (2001, 2003). 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 As input one needs the following:

- All regularized solutions $\{x_n^\delta\}_{n \le N}$ until a certain bound is satisfied
- Noise level δ
- Tuning parameter τ

In contrast to similar methods, no tuning parameter is required. However, in order to gain some security in the case that δ is only known roughly, a tuning parameter $\tau \ge 1$ is advisable. For continuous regularization methods, the algorithm is formulated by differentiating with respect to the regularization parameter α. The parameter choice α_* is the largest α such that

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

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 $y^\delta \in \mathbb{R}^q$ containing errors of standard deviation δ, the parameter choice n_* is the first n such that

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

We take $\tau = 1.5$ for Tikhonov regularization and $\tau = 0.75$ for spectral cutoff regularization.

Known Issues For Tikhonov regularization (and iterated Tikhonov regularization) in a deterministic setting, the ME rule has some favorable properties (cf. Tautenhahn and Hämarik 1999). 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 (Sect. 4.1), 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 rule (Sect. 4.3) for the same tuning parameter. With precisely known δ, the optimal tuning parameter is $\tau = 1$. In the case of spectral cutoff regularization, no optimality results are known.

In Hämarik and Raus (2009) an alternative discretized version to (20) defines n_* as the first n such that

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

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

The ME rule can also be applied to iterative regularization methods, in particular Landweber iteration, for which it is order optimal (see Hämarik and Tautenhahn 2001, 2003).

Results Figure 5 shows that the ME rule has mostly good to acceptable performance for all cases with white noise. 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 rule in Fig. 4, which is consistent with the theory. For both Tikhonov and spectral cutoff regularization, the performance is almost identical to that of the transformed discrepancy principle in Sect. 4.2.

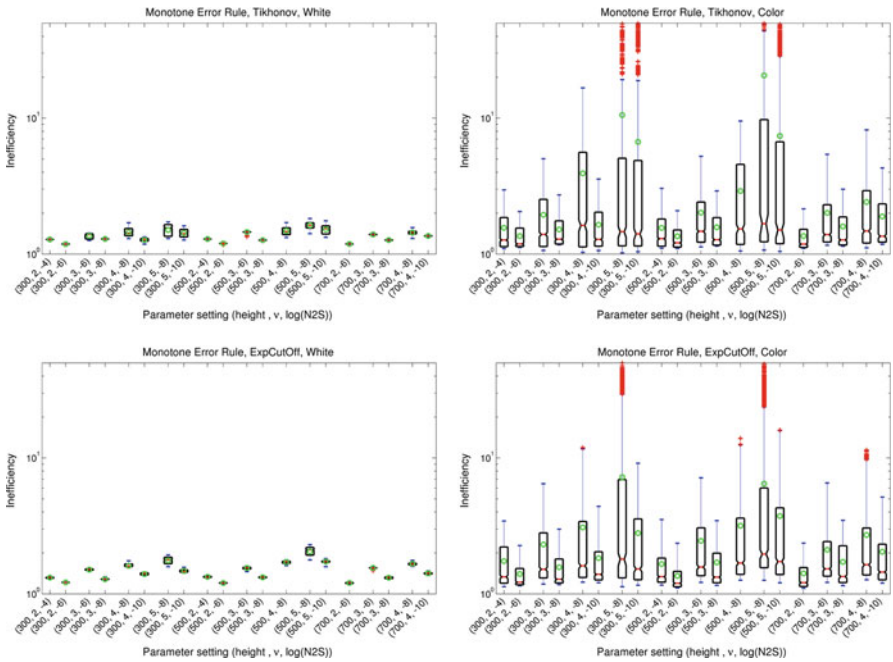

Fig. 5 Inefficiencies of the ME rule

4.5 Varying Discrepancy Principle

This method, due to Lu and Mathé (2014) and based on Blanchard and Mathé (2012), was developed by modifying the usual discrepancy principle to take into account stochastic noise and to achieve optimal convergence rates in a continuous stochastic setting. The modification involves a certain weighted form of the discrepancy with a varying weight.

Method The varying discrepancy principle needs the following input:

- Norms of weighted residuals $\{(\lambda_n I + A^*A)^{-1/2}A^*(Ax_n^\delta - y^\delta)\}_{n\le N}$
- The effective dimension, i.e., the trace of $(\lambda_n I + A^*A)^{-1}A^*A$
- The noise level δ, where the noise is assumed to be Gaussian white
- Tuning parameter τ

The parameter choice is the first n such that

$$\|(\lambda_n I + A^*A)^{-1/2}A^*(Ax_n^\delta - y^\delta)\| \le \tau\delta\sqrt{\operatorname{tr}((\lambda_n I + A^*A)^{-1}A^*A)},$$

where $\lambda_n = \alpha_n$ for Tikhonov regularization and $\lambda_n = \sigma_{l(n)}^2$ for spectral cutoff. For the tuning parameter $\tau > 1$, we use $\tau = 4$ for Tikhonov regularization and $\tau = 1.1$ for spectral cutoff.

Known Issues The rule was developed for general linear regularization methods as well as conjugate gradient iteration. In Blanchard and Mathé (2012) the discrepancy is weighted using a fixed weight parameter λ, and it is shown that this weighted discrepancy principle is order optimal as $\delta \to 0$ for an appropriate choice of λ depending on the solution smoothness. Based on this optimality result, Lu and Mathé (2014) proposed varying λ yielding the a posteriori method above. It is shown that the method is order optimal provided that the solution has a certain self-similarity property.

In Blanchard and Mathé (2012) and Lu and Mathé (2014), the regular methods are combined with an "emergency stop" condition, which defines a maximal regularization parameter. We have not implemented this condition here, but we have used the maximal regularization parameter that is common to all the methods evaluated.

Results As seen in Fig. 6, there are significant saturation effects for $\nu \ge 4$ in the Tikhonov case. Apart from that, the performance with white noise is good, almost as good as for the MD rule in Sect. 4.3. For colored noise, the varying discrepancy principle does not perform quite so well; in particular, test cases with smoother solutions show a lot of variability.

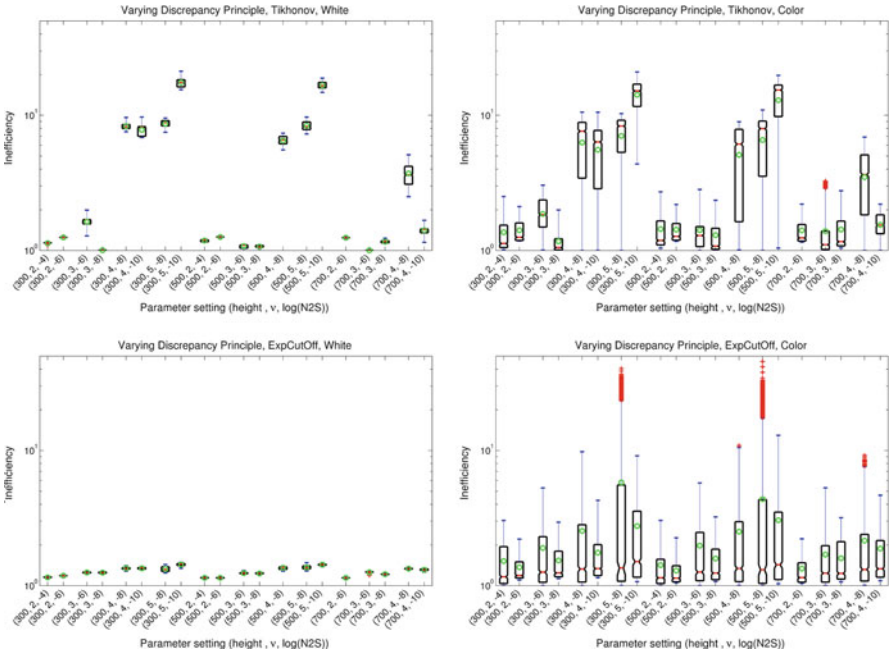

Fig. 6 Inefficiencies of the varying discrepancy principle

4.6 Balancing Principle

The balancing principle, due to Lepskij (1990), 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 (see, e.g., Goldenshluger and Pereverzev 2000; Tsybakov 2000; Mathé and Pereverzev 2003; Bauer and Pereverzev 2005; Mathé and Pereverzev 2006) and nonlinear inverse problems (cf. Bauer and Hohage 2005; Bauer et al. 2009) in deterministic and stochastic settings. The principle aims to balance the known propagated noise error bound $\delta\varrho(n)$ in (8) with the unknown regularization error (7) by an adaptive procedure that employs a collection of differences of regularized solutions.

Method As input one needs the following:

- Maximal index N.
- All regularized solutions $\{x_n^\delta\}_{n\le N}$ up to N.
- An upper bound $\delta\varrho(n)$ for the propagated noise error $\|x_n^0 - x_n^\delta\|$; bounds are known for spectral cutoff and Tikhonov regularization (see after (8)). In the stochastic setting, a bound or estimate $\delta^2\varrho^2(n)$ of the variance $\mathbb{E}\|x_n^0 - x_n^\delta\|^2$.

- Noise level δ (and the covariance in the stochastic setting if known). Then one can use known expressions for $\delta\varrho(n)$. 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 (13).
- Tuning constant κ.

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

1. There is one data set with δ as the only known property of the noise, in which case we use the known expressions (10) and (11) of $\delta^2\varrho^2(n)$ for the white noise model; we refer to this method as the balancing principle (white). These results are given in Fig. 7.
2. Four independent sets of data y_i^δ are available, in which case we use the estimate (13) for $\delta^2\varrho^2(n)$. Figure 8 shows the results for the fast balancing variant that is described in the next paragraph.

In each situation, we use the maximal index N as defined in (12).

Known Issues The balancing principle is one of the few parameter choice methods for which oracle inequalities for the error are known (cf. Raus and Hämarik 2007; Bauer et al. 2009), 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 (which has not been observed in practice).

According to the theory (cf. Mathé and Pereverzev 2003), the constant κ should be 1 in the deterministic setting. 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 (cf. Bauer and Pereverzev 2005).

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.

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

$$b(n) = \max_{n<k\leq \ell(n)} \left\{4^{-1}\|x_n - x_k\|/(\delta\varrho(k))\right\}, \tag{22}$$

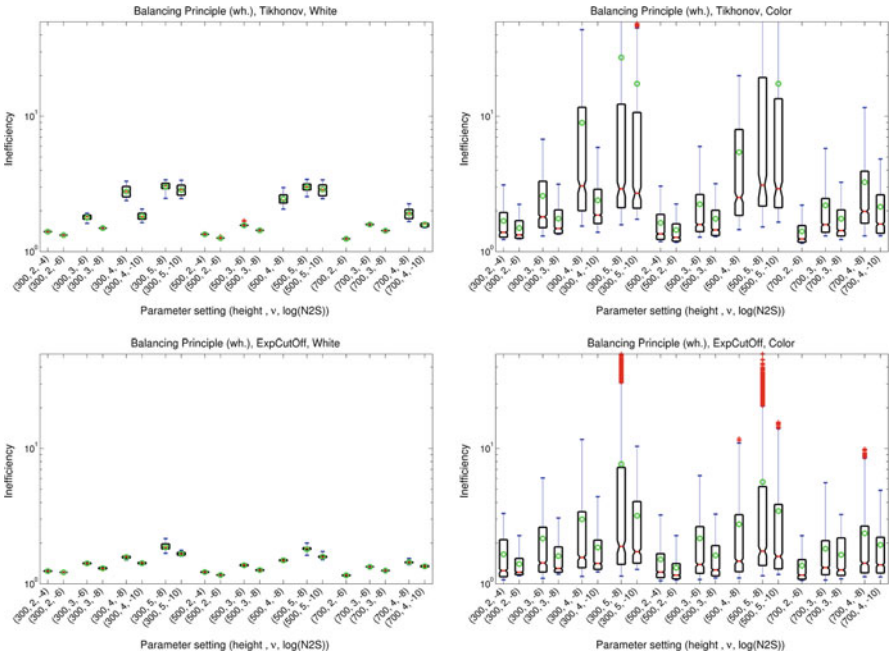

Fig. 7 Inefficiencies of the balancing principle (white) using known δ

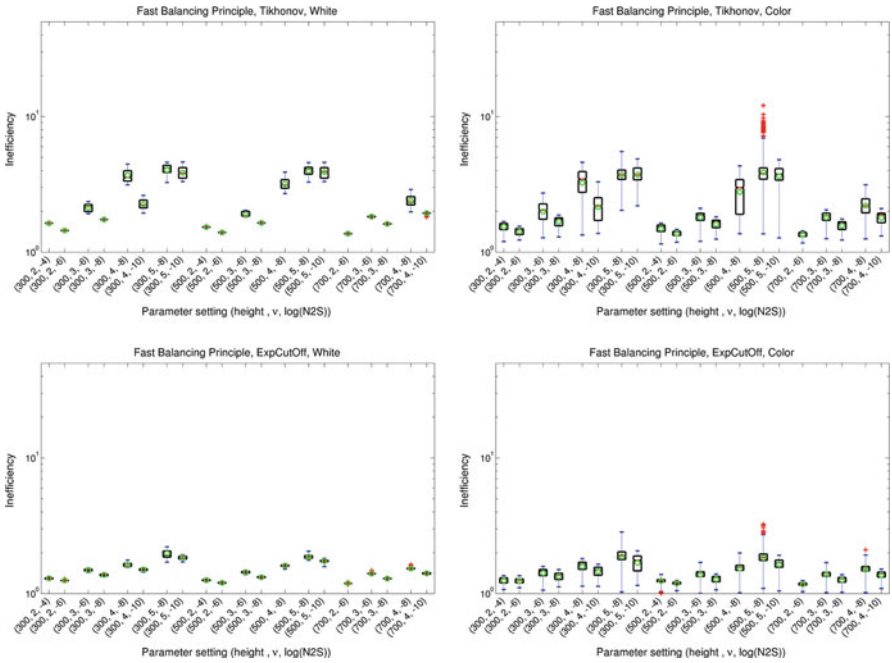

Fig. 8 Inefficiencies of the fast balancing principle using four independent data sets

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 while also giving a convergent method (see Mathé and Pereverzev 2003; Bauer and Munk 2007). 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 the experiments reported here, we used $\ell(n)$ defined by $\varrho(\ell(n))/\varrho(n) = 2$. The choice of n_* as the first n with $b(n) \leq \kappa$ ($b(n)$ as in (22)) is called the fast balancing principle. This method has the advantage of requiring a much lower number of computations than the original version. For the fast balancing principle with $\ell(n) = n+1$ in (22), optimality results are known (cf. Hämarik and Raus 2009) 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 (10) and (11) of $\delta^2\varrho^2(n)$, one can use the Monte Carlo estimate $\varrho^2(n) \approx \text{Mean}\|A_n^{-1}\hat{\xi}_i\|^2$, where $\hat{\xi}_i$ are synthetic vectors of independent standard normal pseudorandom 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}.

Results In our experiments, the results for the balancing principle and its fast version with the same input are very hard to distinguish, and therefore we included only the figure for the fast variant (Fig. 8). Both are listed in Table 4 in our conclusion (Sect. 5).

The results in Figs. 7 and 8 for the white noise situation are almost identical for both Tikhonov and spectral cutoff regularization (slightly better for balancing principle (white) in Fig. 7). The method is stable and performs quite well. In Fig. 8, the results for colored noise are quite similar (with only slightly more variation) to those for white noise, since here the method adapts automatically to the noise color. On the other hand, in Fig. 7, there are a considerable number of outliers for the colored noise scenario, in particular for smoother solutions x.

4.7 Hardened Balancing Principle

The hardened balancing principle is a modified version of the balancing principle of Sect. 4.6 in the stochastic setting and was first proposed in Bauer (2007). It has the advantage that it does not require a tuning parameter.

Method The input is the same as for the balancing principle in Sect. 4.6 without the tuning parameter and the noise level, so it can be computed parallel to this one. 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 (so δ can be unknown), is required. Define the balancing functional $b(n)$ as in (21) and smoothed balancing functional $B(n) = \max_{n\leq k\leq N}\{b(k)\}$. The parameter choice is

$$n_* = \operatorname*{argmin}_{n \le N} \left\{ B(n) \sqrt{\varrho(n)} \right\}. \tag{23}$$

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:

1. To address the situation of one data set with no knowledge about the noise, we use the known expressions (10) and (11) of $\delta^2\varrho^2(n)$ for the white noise model. We refer to this method as the hardened balancing principle (white), and its results are depicted in Fig. 9.
2. We estimate $\delta^2\varrho^2(n)$ using (13) with four independent sets of data y_i^δ. Figure 10 shows the results.

In each situation, we use the maximal index N in (12). For better computational efficiency, as for the balancing principle, one can use the modified version of the balancing functional defined in (22) 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 cutoff regularization, a version of the hardened balancing principle has been analyzed by Bauer and Reiß (2008) in a Bayesian framework, with stochastic x and noise, and sequence data. It is shown that the parameter choice

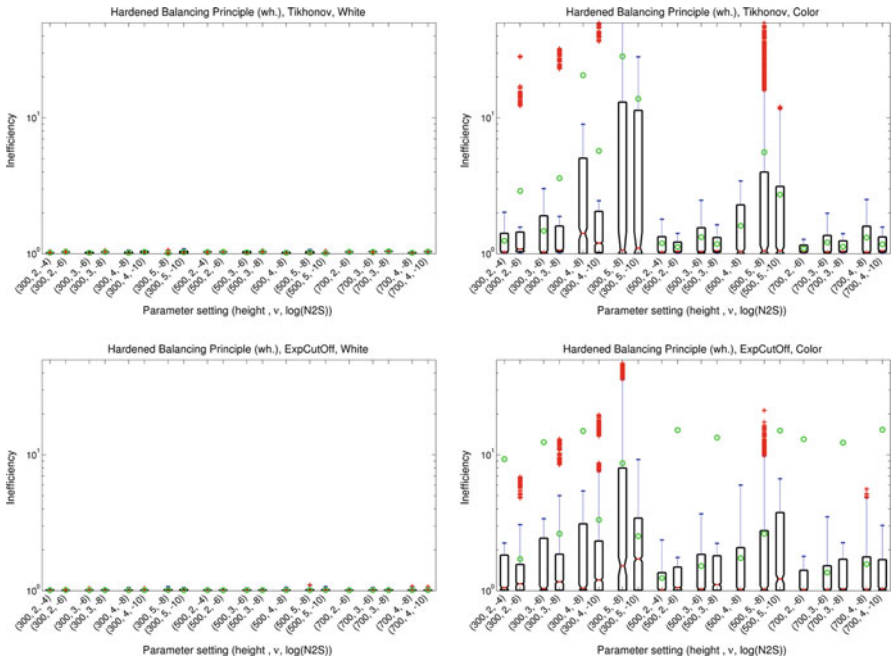

Fig. 9 Inefficiencies of the hardened balancing principle (white)

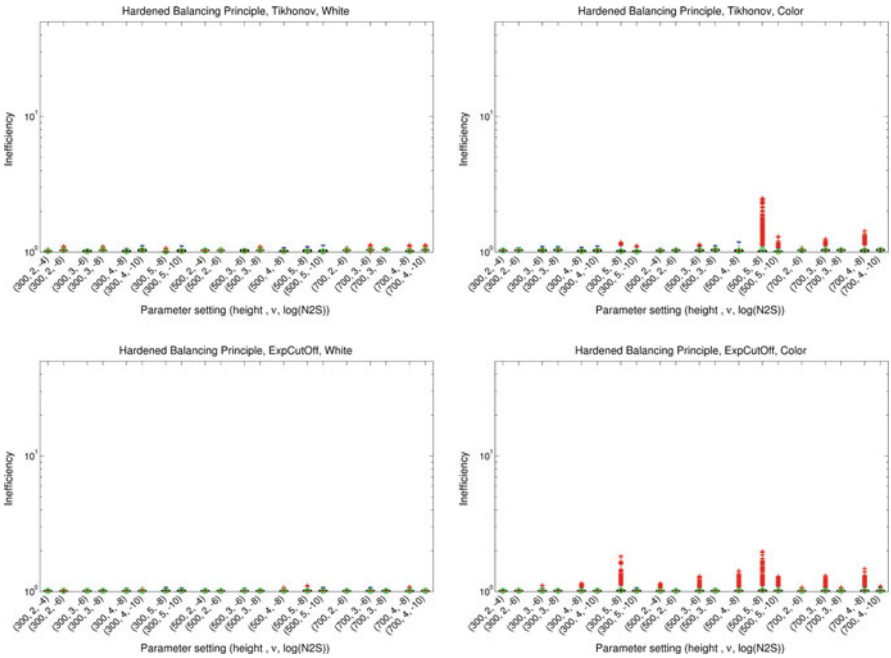

Fig. 10 Inefficiencies of the hardened balancing principle using four independent data sets

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. 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 (cf. Bauer and Kindermann 2008).

Numerical experiments in Bauer (2007) and Bauer and Reiß (2008) 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 choosing N as in (12).

All remarks in Sect. 4.6 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 of $\varrho^2(n)$ as for the balancing principle in Sect. 4.6.

Results Figure 9 displays the results for the hardened balancing principle (white), which uses the variance expressions (10) and (11) for the white noise model. As expected, it performs very well in the white noise situation, with results that are almost identical to those in Fig. 10. In the colored noise situation, there is much more variability – note that there are extreme outliers that lead to some large means in the spectral cutoff case. The performance for Tikhonov regularization is better in the cases with less smooth solutions.

Figure 10 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 efficiency values in all the test cases for both Tikhonov and spectral cutoff regularization.

4.8 Quasi-optimality Criterion

The quasi-optimality criterion (see, e.g., Tikhonov and Arsenin 1977; Hofmann 1986) is one of the oldest and simplest available parameter choice methods. It was originally introduced by Tikhonov and Glasko (1965) and became better known from the continuous version proposed by Tikhonov and Arsenin (1977). A good overview of the method and its history can be found in Bauer and Kindermann (2009).

Method As input for a certain minimization, one needs the following:

- Maximal index N, a suitable value is defined in (12)
- All regularized solutions $\{x_n^\delta\}_{n\le N}$ up to N

The noise level does not need to be known, and there is no tuning parameter. The parameter choice n_* is defined simply as

$$n_* = \operatorname*{argmin}_{n\le N} \|x_n^\delta - x_{n+1}^\delta\|. \tag{24}$$

The continuous version for Tikhonov regularization defines the parameter choice by $\alpha_* = \operatorname{argmin} \left\|\alpha \frac{d}{d\alpha} x_\alpha^\delta\right\|$. Using a difference quotient in place of the derivative for the discrete parameters $\alpha_n = \alpha_0 q_\alpha^n$, 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 cutoff regularization, the cutoff points $l(n)$ need to be chosen carefully and far enough apart (cf. Bauer and Reiß 2008).

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 (see Bauer and Kindermann 2008).

Convergence of Tikhonov regularization with the quasi-optimality criterion was shown in Leonov (1979) 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 Glasko and Kriksin (1984), where abstract conditions were given.

Although it has been used successfully in a number of practical situations (see, e.g., Pohl et al. 2001), the first more concrete proofs for the quasi-optimality criterion were provided only quite recently – in Bauer and Reiß (2008) for spectral cutoff regularization in a Bayesian setting and in Bauer and Kindermann (2008) 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 was derived in Kindermann and Neubauer (2008) and Neubauer (2008) and generalized in Kindermann (2011).

For iterative regularization methods, less is known about the behavior of the quasi-optimality criterion. There are some convergence results in Kindermann and Neubauer (2008), Neubauer (2008), and Kindermann (2011), for deterministic noise, in particular a slow convergence rate for Landweber iteration.

Results As seen in Fig. 11, the quasi-optimality criterion has excellent performance across all the test cases for spectral cutoff and Tikhonov regularization with either white or colored noise. Keeping in mind that it is computationally the simplest of the available methods and that it does not require the noise level, the performance is remarkable.

4.9 L-Curve Method

The L-curve method, proposed by Hansen (1992, 1998) and Hansen and O'Leary (1993), 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 (cf. Lawson and Hanson 1974). 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 As input to minimize a certain function, one needs the following:

- Norms of all residuals $\{Ax_n^\delta - y^\delta\}_{n \le N}$
- Norms of the regularized solutions $\{x_n^\delta\}_{n \le N}$

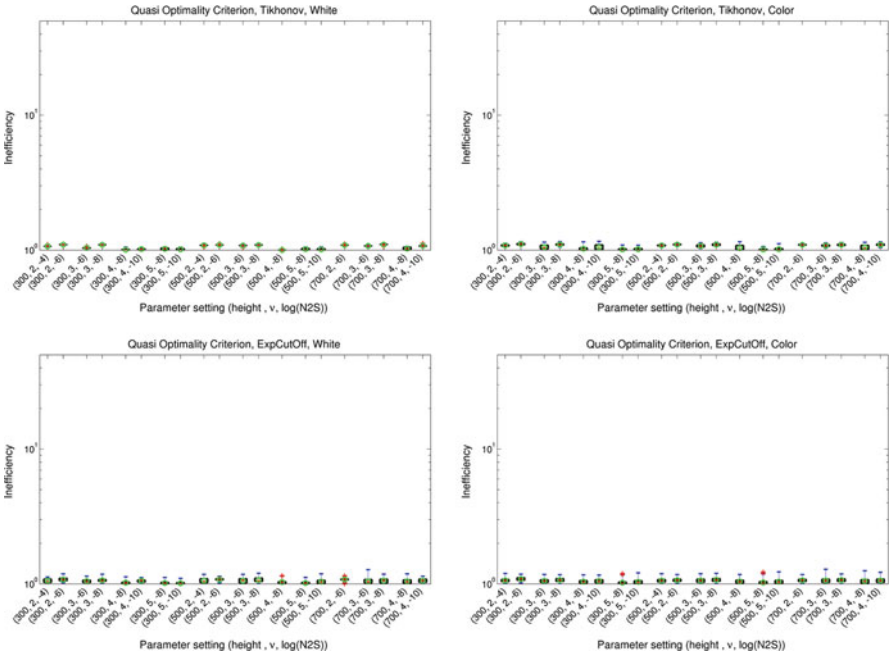

Fig. 11 Inefficiencies of the quasi-optimality criterion

The noise level does not need to be known. The parameter choice is defined by the product of the norms of the residual and regularized solution, i.e.,

$$n_* = \operatorname*{argmin}_{n \le N} \left\{ \|Ax_n^\delta - y^\delta\| \|x_n^\delta\| \right\} . \tag{25}$$

Here the "corner point" is defined by the slope of its "tangent" being -1 as in Reginska (1996) (see also Engl et al. 1996). The generalizations minimize $\|Ax_n^\delta - y^\delta\| \|x_n^\delta\|^\tau$ (see Reginska 1996; Lu and Mathé 2013), 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 (see Hansen and O'Leary 1993; Calvetti et al. 2002). The above algorithm by Reginska (1996) is a simpler alternative, which can also be applied with discrete regularization parameters. Our experience (see also Johnstone and Gulrajani 2000) is that these algorithms have similar performance. Other algorithms are given in Hansen et al. (2007) and the references therein.

Recently, Lu and Mathé (2013) derived the first rigorous optimality result for the L-curve criterion. They showed, in the deterministic setting, that if the L-curve

functional defined above has an interior global minimizer, then this choice satisfies a certain optimality property. However, there is no guarantee that there will be such a minimizer. In many (but not all) problems, the L-curve method (in its various forms) has been observed to give a reasonably good and robust parameter choice, and it can cope with correlated errors (see, e.g., Hanke and Hansen 1993; Hansen and O'Leary 1993; Hansen 1998; Hansen et al. 2007; Abascal et al. 2008; Åkesson and Daun 2008; Correia et al. 2009).

However, it is known theoretically that the L-curve method (from Hansen and O'Leary 1993) has serious limitations (cf. Engl et al. 1996; Yagola et al. 2002): the L-curve corner may not even exist and it is not convergent, neither in the deterministic setting (under-smoothing) nor in the stochastic setting (over-smoothing). These effects have been observed in numerical experiments (see, e.g., Johnstone and Gulrajani 2000; Hansen 2001). Moreover, finding the L-curve corner is hard to automate, in particular with discrete regularization parameters. For more details, we refer to Hanke (1996), Reginska (1996), and Vogel (1996, 2002).

The L-curve method can also be extended to various iterative regularization methods (see Hanke and Hansen 1993, Hansen 1998, Kilmer and O'Leary 2001, Farquharson and Oldenburg 2004, Kindermann 2011).

Results As seen in Fig. 12, the L-curve method has very poor performance across all the test cases for both Tikhonov and spectral cutoff regularization. Note that "missing" parts in the panels indicate that the inefficiency is very large!

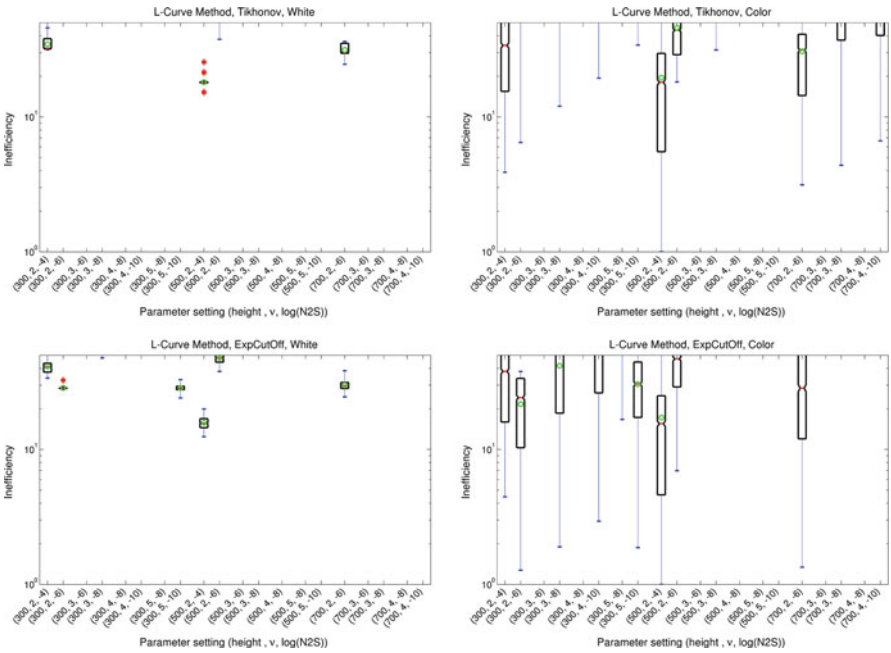

Fig. 12 Inefficiencies of the L-curve method

4.10 Extrapolated Error Method

This method, developed by Brezinski et al. (2008, 2009) 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 input required for a certain minimization is:

- Norms of all residuals $\{r_n = y^\delta - Ax_n^\delta\}_{n \le N}$
- Norms of transformed residuals $\{A^* r_n\}_{n \le N}$

The parameter choice is

$$n_* = \operatorname*{argmin}_{n \le N} \{\|r_n\|^2 / \|A^* r_n\|\}. \tag{26}$$

Known Issues The expression $\|r_n\|^2/\|A^* r_n\|$ in (26) is one of a family of extrapolated estimates of the error $\|x - x_n^\delta\|$ (see Brezinski et al. 2008, 2009). In practice, to calculate the expression accurately for small α_n in Tikhonov regularization, one should replace $A^* r_n$ by the equivalent term $\alpha_n x_n^\delta$. There is currently no convergence analysis to justify the method, but numerical results in Brezinski et al. (2008) indicate that it is quite robust.

Results As seen in Fig. 13, for white noise, the extrapolated error method performs reasonably well for spectral cutoff and for certain test cases of Tikhonov regularization but rather poorly for smoother solutions ($\nu \ge 4$). With colored noise, there are a significant number of poor outliers for smoother solutions. Note that the mean for the $(700, 4, -10)$ setting in spectral cutoff with white or colored noise is too large to fit into the panels. The results are similar to, but slightly worse than, the results of the discrepancy principle in Fig. 2. However, they show less variation for colored noise.

We also assessed a related method proposed in Brezinski et al. (2009), which has $\|r_n\| \|A^* r_n\| / \|AA^* r_n\|$ in place of the expression in (26), but overall it gave much worse results than the method above.

4.11 Modified Discrepancy Partner (MDP) Rule

Hanke and Raus (1996) 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 Hämarik et al. 2009, 2011; Lu and Mathé 2013). 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 Sect. 4.3 to get the modified discrepancy partner (MDP) rule.

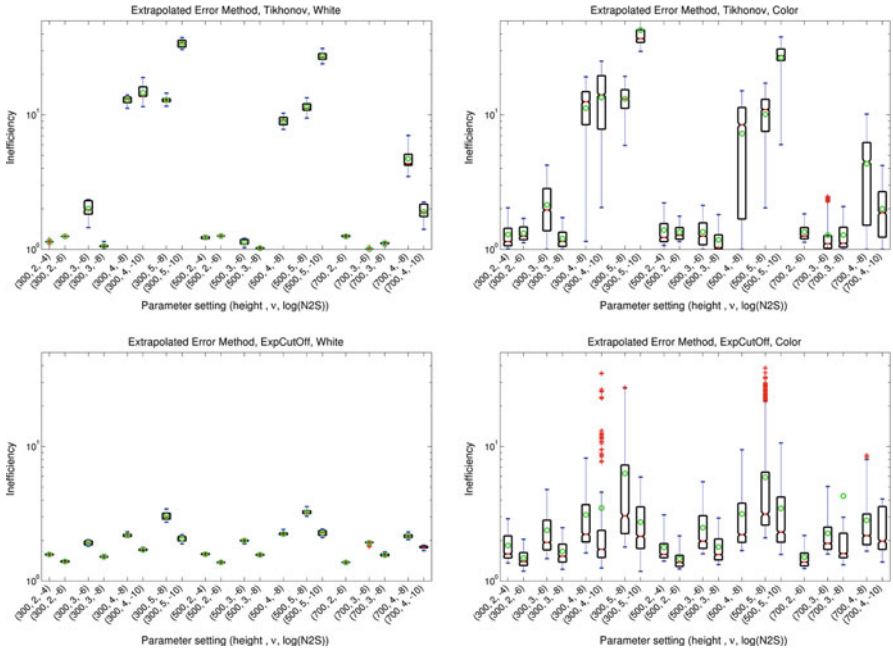

Fig. 13 Inefficiencies of the extrapolated error method

Method The MDP rule requires the following input for a certain minimization:

- All regularized solutions $\{x_n^\delta\}_{n \le N}$

Neither the noise level δ nor any tuning parameter is necessary. As for the MD rule, one has to compute the function f in Eq. (18). 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 $\alpha_n = \alpha_0 q_\alpha^n$, we have $\varrho(n) = \alpha_n^{-1/2}$, and, for spectral cutoff 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,

$$n_* = \operatorname*{argmin}_{n \le N} \{\varrho(n)\beta_n \left|\left\langle Ax_n^\delta - y^\delta, (A^*)^{-1}\left(x_n^\delta - x_{n+1}^\delta\right)\right\rangle\right|^{1/2}\}. \tag{27}$$

Known Issues For Tikhonov regularization, Hanke and Raus (1996) 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 regularization in Hanke and Raus (1996), the MDP rule gives errors that,

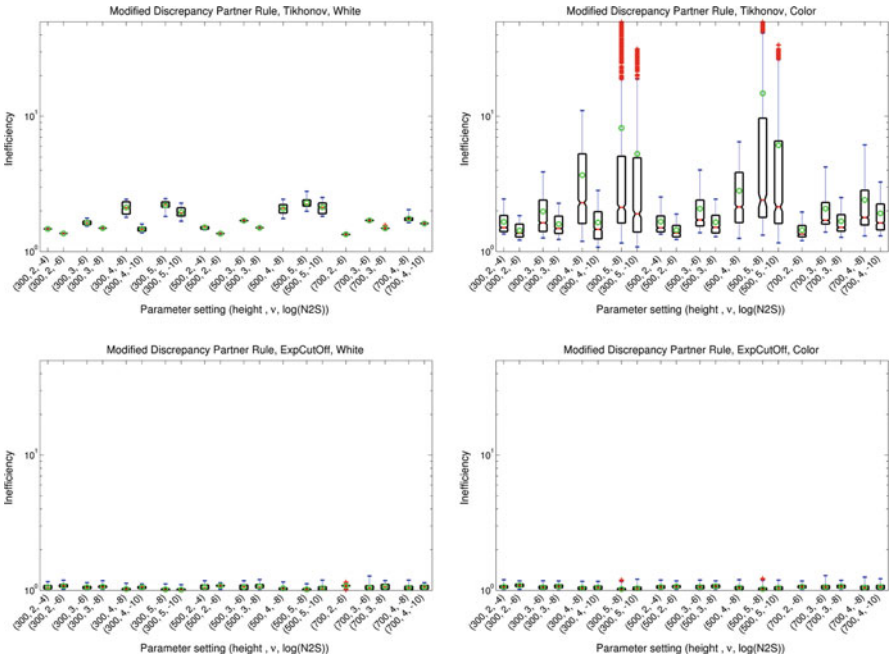

Fig. 14 Inefficiencies of the MDP rule

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 Hämarik et al. (2009) and Hanke and Raus (1996) that the proposed approach is quite general, and it can be applied with order-optimal rules for other regularization methods for linear problems to give useful data-driven rules. A more theoretical analysis, including convergence properties, in some deterministic settings was derived in Kindermann (2011), also for iterative regularization methods.

Results As seen in Fig. 14, the MDP rule has excellent performance across all the test cases for spectral cutoff in both the white and colored noise situations. For Tikhonov regularization and white noise, it has performance similar (slightly worse) to the MD rule in Fig. 4. The behavior for colored noise is better than the MD rule, but test cases with smoother solutions still show a lot of variability.

4.12 Normalized Cumulative Periodogram Method

For the finite dimensional situation with white noise, Rust (2000) suggested using the periodogram of the residual vector as a diagnostic tool. Building on this, Hansen et al. (2006) developed a parameter choice method, called the normalized

cumulative periodogram (NCP) method, and Rust and O'Leary (2008) proposed a similar method. The basis of these methods is to make the residual as close as possible to white noise.

Method As input to minimize a certain function, the following is required:

- Norms of all residual vectors $\{Ax_n^\delta - y^\delta\}_{n \le N}$ where $y^\delta \in \mathbb{R}^q$

The noise level does not need to be known. Note that the method is formulated for stochastic white noise. The (unscaled) periodogram of the residual vector $r_n = y^\delta - Ax_n^\delta$ is defined as the vector p with elements $p_k = |\mathrm{dft}(r_n)|^2$, $k = 1, \ldots, q$, where dft denotes the discrete Fourier transform. Define the normalized cumulative periodogram as the vector $c = c(r_n)$ with elements

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

where $\|\cdot\|_1$ is the l_1 norm, and let v be the vector with elements $v_i = i/(q-1)$. 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 Hansen et al. (2006), 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 Rust and O'Leary (2008) have been shown to perform well on several test problems. 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 methods.

Results Figure 15 shows that the NCP method has quite good performance in the white noise situation except for Tikhonov regularization with smoother solutions ($\nu \ge 4$). By contrast, the method has poor performance in the colored noise scenario for both Tikhonov and spectral cutoff regularization.

4.13 Residual Method

The residual method was introduced in Bauer and Mathé (2011) for spectral cutoff regularization in an infinite dimensional Bayesian setting (see also Bauer 2007). The method is based on minimizing a certain weighted form of the norms of the residuals, where the weighting penalizes under-smoothing parameter values.

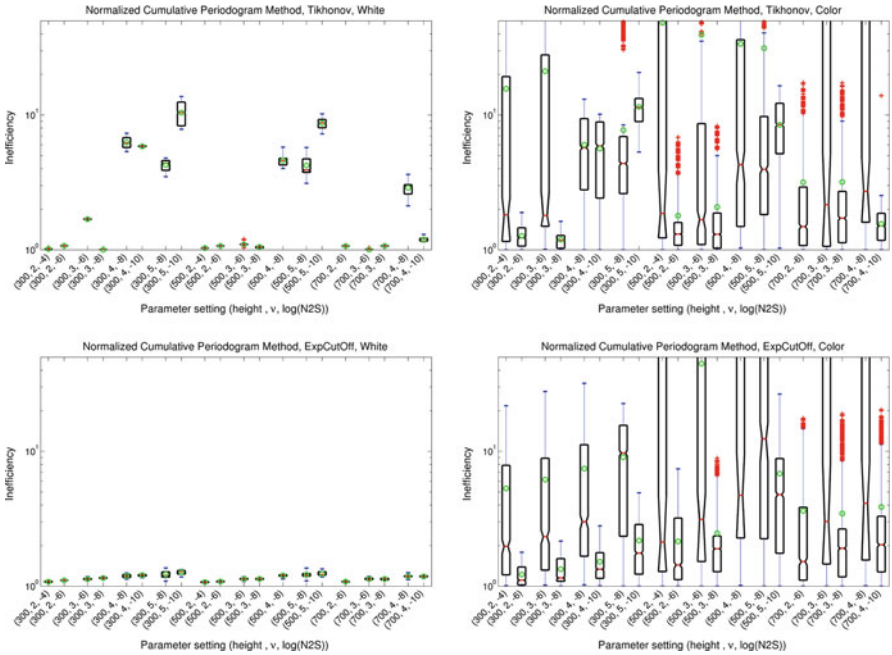

Fig. 15 Inefficiencies of the NCP method

Method As input to minimize a certain function, one needs the following:

- Norms of all residuals $\{Ax_n^\delta - y^\delta\}_{n \le N}$ of the regularized solutions
- A trace function of the operator and the regularization operators

The noise level does not need to be known. Define the operator $B = A(I - A_n^{-1}A)$. The parameter choice is defined by

$$n_* = \operatorname*{argmin}_{n \le N} \left\{ \frac{\|Ax_n^\delta - y^\delta\|}{(\operatorname{trace} B^* B)^{1/4}} \right\} . \tag{28}$$

Known Issues Assuming an appropriate behavior for the random element x, the residual method for spectral cutoff is a convergent method as $\delta \to 0$ (cf. Bauer and Mathé 2011), 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 (28) that this choice is similar in form to generalized cross-validation (30) of Sect. 4.15. For spectral cutoff regularization, the trace can be evaluated easily. For Tikhonov regularization, computation of the trace might be rather expensive if q is large, but more efficient stochastic trace estimators can be used (see, e.g., Girard 1989; Hutchinson 1989; Golub and von Matt 1997).

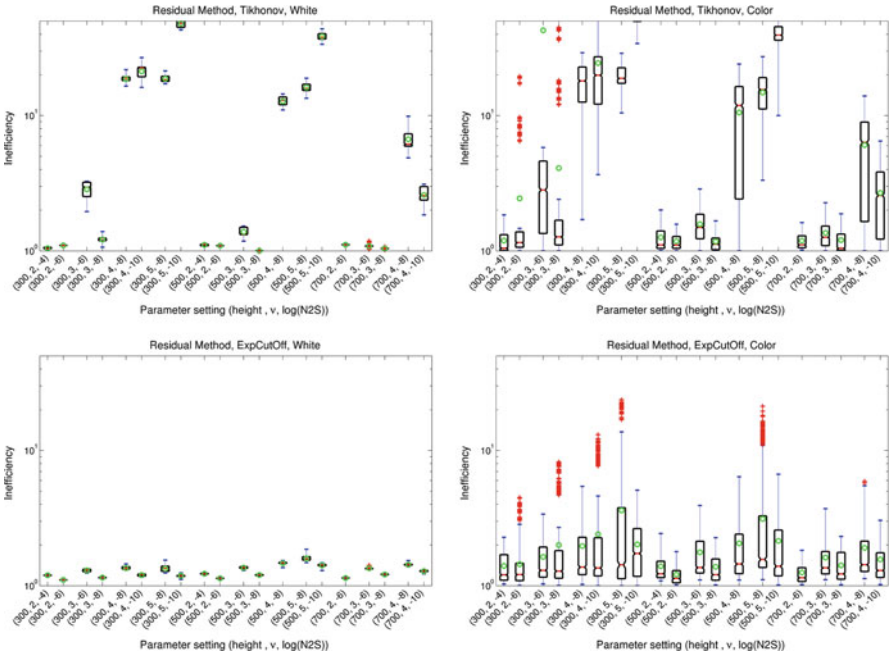

Fig. 16 Inefficiencies of the residual method

Results Figure 16 shows that the residual method has good performance for spectral cutoff regularization with white noise and mediocre performance with several outliers for colored noise. Tikhonov regularization shows some severe saturation effects. The method has slightly better performance when the problem is more ill-posed (larger values of the height h), and in these cases, it copes quite well with unknown colored noise. The results are similar to those of the discrepancy principle in Fig. 2 (slightly better for spectral cutoff, a bit worse for Tikhonov) and similar (slightly worse) to those of GML for Tikhonov (see Sect. 4.14 below).

4.14 Generalized Maximum Likelihood

As discussed in Sect. 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 (cf. Eubank 1988; Wahba 1990). Using this interpretation, Wahba (1985) derived the generalized maximum likelihood (GML) estimate (see also Anderssen and Bloomfield 1974; Davies 1982; Wecker and Ansley 1983). In the case where $A : \mathbb{R}^q \to \mathbb{R}^q$ 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 $y \sim \mathcal{N}(0, b(AA^T + \lambda I))$ for a constant b.

Method As input to minimize a certain function, the following is needed:

- Sums of squares of all the residuals $\{Ax_n^\delta - y^\delta\}_{n\le N}$ where $y^\delta \in \mathbb{R}^q$
- All eigenvalues of $I - AA_n^{-1}$, where AA_n^{-1} is the influence matrix mapping y^δ to Ax_n^δ

The method is formulated for stochastic white noise where knowledge of the noise level is not required. GML can be applied for all methods (in particular Tikhonov regularization) where these input data are known, but it is not appropriate for spectral cutoff regularization, as seen below. The GML parameter estimate is defined by

$$n_* = \underset{n\le N}{\operatorname{argmin}} \left\{ \frac{\|Ax_n^\delta - y^\delta\|^2}{(\det{}^+ (I - AA_n^{-1}))^{1/q_1}} \right\}, \tag{29}$$

where $q_1 = \operatorname{rank}(I - AA_n^{-1})$ and $\det^+$ is the product of the nonzero eigenvalues.

Known Issues For spectral cutoff, clearly, $\det^+ (I - AA_n^{-1}) = 1$, so the GML function is just the residual sum of squares. Therefore, the GML method will always severely under-smooth. For Tikhonov regularization of polynomially ill-posed problems with uncorrelated errors in the data, the "expected" GML estimate has been analyzed in Wahba (1985) and Lukas (1995). As the sample size $q \to \infty$, 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 suboptimal and under-smoothing for solutions which are "smoother" than the minimum required by $x \in \mathcal{X}$. In numerical studies for spline smoothing (cf. Kohn et al. 1991), the GML method performed well and was more stable than GCV (see Sect. 4.15). A geometric explanation of this stability is given in Efron (2001), 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 (see Opsomer et al. 2001). 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 (cf. Wang 1998; Opsomer et al. 2001). Note that for Tikhonov regularization, the $\det^+$ term in (29) can be computed using a SVD of A if q is not too large.

Results As seen in Fig. 17 (and noted above), the GML method does not give reasonable results for spectral cutoff regularization. Note that "missing" entries correspond to very large inefficiencies. For Tikhonov regularization with white noise, GML performs well when ν is small, but poorly when $\nu \ge 4$, consistent with the asymptotic results. The method also performs quite well when ν is small in the colored noise scenario, but otherwise it has poor performance.

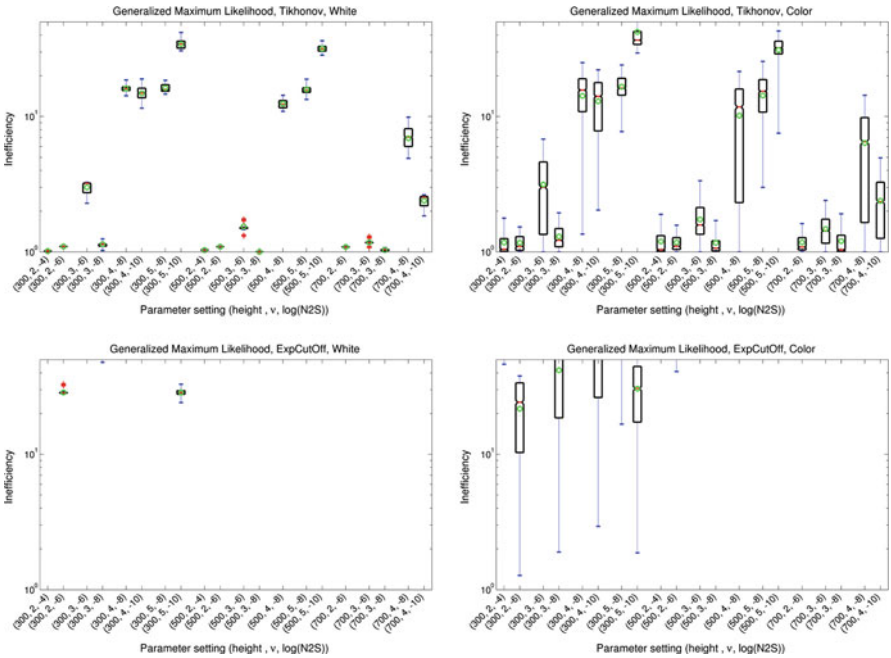

Fig. 17 Inefficiencies of the generalized maximum likelihood

4.15 Generalized Cross-Validation

Generalized cross-validation (GCV), due to Wahba (1977), 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, Craven and Wahba (1979), Golub et al. (1979), and Wahba (1977, 1990) derived the GCV method, which has the advantage of being invariant under orthogonal transformations of the data.

Method As input to minimize a certain function, one needs:

- Sums of squares of all the residuals $\{Ax_n^\delta - y^\delta\}_{n \le N}$ where $y^\delta \in \mathbb{R}^q$
- The trace of the influence matrix AA_n^{-1} mapping y^δ to Ax_n^δ

The noise level does not need to be known. As discussed below, GCV is best suited to stochastic white noise. The GCV parameter estimate is defined by

$$n_* = \operatorname*{argmin}_{n \le N} \left\{ \frac{\|Ax_n^\delta - y^\delta\|^2}{(q^{-1}\operatorname{tr}(I - AA_n^{-1}))^2} \right\}. \tag{30}$$

Known Issues GCV is closely related to and behaves like the unbiased prediction risk method (also known as Mallows C_p or C_L; see Li 1987; Wahba 1990; Efron 2001). 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.

It has been analyzed asymptotically for polynomially ill-posed problems. For Tikhonov regularization, it is known from, e.g., Li (1986), Lukas (1993), and Gu (2002), that, with uncorrelated errors, GCV is asymptotically optimal with respect to the prediction risk as the number of data points $q \to \infty$, 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 $\mathbb{E}\|x_n^\delta - x\|^2$ (cf. Wahba and Wang 1990; Lukas 1993; Vogel 2002). This saturation effect is not as serious as it may seem since one can use GCV to choose the order of regularization (Wahba 1990).

For spectral cutoff regularization, it is known, e.g., from Li (1987) and Vogel (1986, 2002, Chap. 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 (see, e.g., Thompson et al. 1989, Wahba 1990, Kohn et al. 1991, Efron 2001, Kou and Efron 2002, Lukas 2006, 2010) 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 (30) 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, i.e., $l(n)$ for spectral cutoff regularization. For Tikhonov regularization, it is harder to compute, e.g., by SVD (cf. Hansen 1998; Wahba 1990), bi-/tridiagonalization (cf. Eldén 1984; Gu et al. 1989; Wahba 1990, Chap. 11; Hansen 1998), or trace estimation using stochastic (Monte Carlo) algorithms (cf. Girard 1989; Golub and von Matt 1997; Hutchinson 1989). Other efficient algorithms exist for special problems, in particular spline smoothing (see Hutchinson and de Hoog 1985).

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, e.g., in Wang (1998) and Opsomer et al. (2001). The GCV method has also been extended in other directions, including to non-Gaussian data (cf. Gu 2002) and to wavelet thresholding in Jansen et al. (1997). It can also be applied to iterative regularization methods (see Wahba 1990, Hanke and Hansen 1993, Hansen 1998, Haber and Oldenburg 2000, Kilmer and O'Leary 2001, Santos and De Pierro 2003).

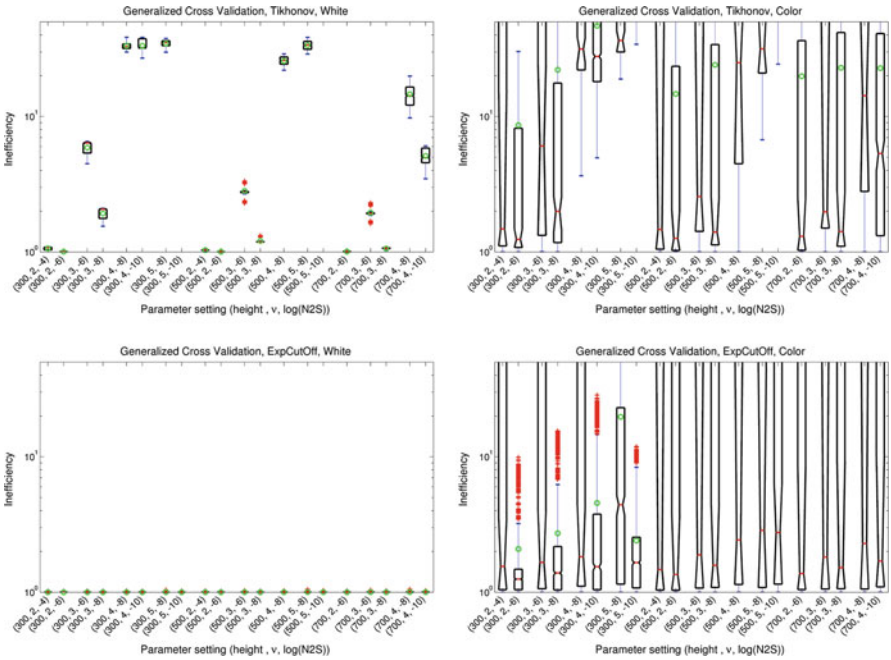

Fig. 18 Inefficiencies of generalized cross-validation

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) of Akaike (1973) and Eubank (1988). In view of its similarity to GCV, we do not consider this here.

Results As seen in Fig. 18, in the white noise situation, GCV has excellent performance across all the test cases for spectral cutoff, and it performs very well for Tikhonov regularization with solutions of lowest smoothness ($\nu = 2$) and reasonably well for $\nu = 3$. However, it performs poorly for Tikhonov regularization with significant saturation effects in the cases where $\nu \geq 4$. There is no such problem for spectral cutoff, consistent with the theory. By contrast with the white noise situation, GCV performs very poorly in all the colored noise experiments.

4.16 Robust Generalized Cross-Validation

In order to overcome the instability of GCV, a robust GCV (RGCV) method has been developed in Robinson and Moyeed (1989) and Lukas (2006). Like GCV, the RGCV method has a good rationale in terms of the influence of data values on the regularized solution.

Method As input, one needs the same as for GCV (Sect. 4.15) and:

- The trace of the square of the influence matrix $(AA_n^{-1})^2$
- A robustness parameter $\gamma \in (0, 1)$

The noise level does not need to be known. The RGCV parameter estimate is defined by minimizing a certain function:

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

where $\gamma \in (0, 1)$ is the robustness parameter (here: $\gamma = 0.1$). Note that with $\gamma = 1$ the RGCV method is just GCV.

Known Issues As γ is decreased, the method becomes more robust and it is less likely to choose a large value of n (see Lukas 2006). The effect of the last term of the RGCV function in (31) 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 (31) can also be explained by the fact that, for uncorrelated errors with variance δ^2, we have $\delta^2 \operatorname{tr}((AA_n^{-1})^2) = \mathbb{E}\|Ax_n^\delta - \mathbb{E}Ax_n^\delta\|^2$, the variance of Ax_n^δ. In fact, for Tikhonov regularization of polynomially ill-posed problems, it is known from Lukas (2006, 2008) that as $q \to \infty$, the RGCV function is consistent with a weighted sum of the prediction risk $\mathbb{E}\|Ax_n^\delta - y\|^2$ and the variance $\mathbb{E}\|Ax_n^\delta - \mathbb{E}Ax_n^\delta\|^2$, with weights γ and $1-\gamma$, respectively. Therefore, the RGCV method places extra weight on reducing the variability of the regularized solution. As $q \to \infty$, the RGCV estimate has the same optimal order as the GCV estimate, but it has a different constant resulting in a slightly larger value (cf. Lukas 2006, 2008). Consequently, the method does not suffer as badly as GCV from the saturation effect.

For spectral cutoff regularization, it is clear that $\operatorname{tr}((AA_n^{-1})^2) = \operatorname{tr}(AA_n^{-1})$ is simply the number of terms in the expansion of x_n^δ. For Tikhonov regularization, $\operatorname{tr}((AA_n^{-1})^2)$ can be computed by the use of SVD or bidiagonalization (cf. Grad and Zakrajšek 1972). It can be expected that RGCV can be extended in the same way as GCV to iterative regularization methods.

Results As seen in Fig. 19, in the white noise situation, RGCV performs very well across all the test cases for spectral cutoff and also for Tikhonov regularization with solutions of lowest smoothness. However, for $\nu \ge 4$, there is evidence of the same saturation effect described for GCV, but the effect is slightly smaller here. The results for colored noise are mediocre, but a big improvement compared to GCV in Fig. 18. Note that there are extreme outliers leading to some means that are too large to fit in the panels.

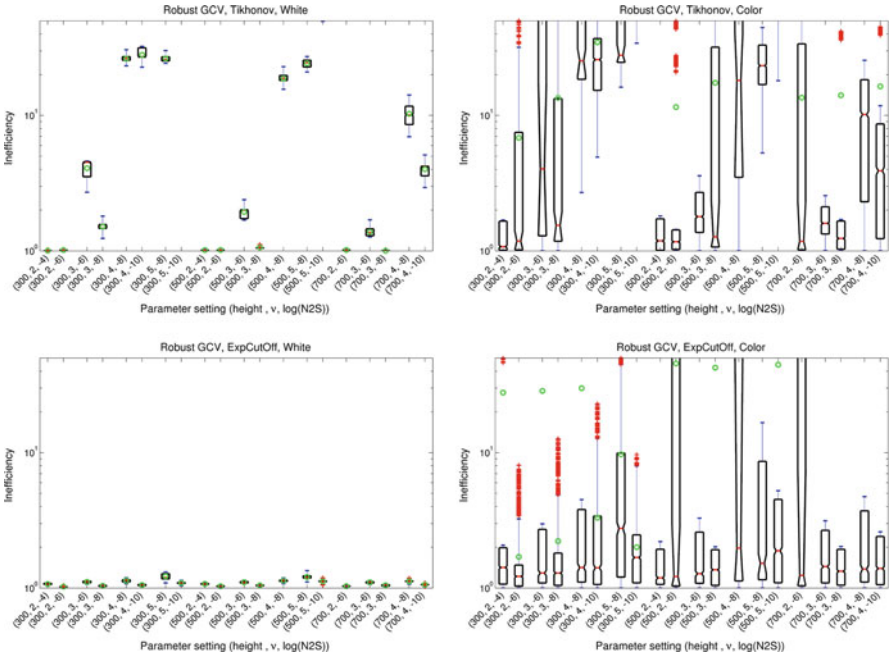

Fig. 19 Inefficiencies of robust generalized cross-validation

4.17 Strong Robust GCV

The RGCV method in Sect. 4.16 is one of a family of robust GCV methods developed in Lukas (2008) 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 (see Lukas 2008).

Method As input, one needs the same as for GCV (Sect. 4.15) and:

- The trace of $A_n^{-1} * A_n^{-1}$
- A robustness parameter $\gamma \in (0, 1)$

The noise level does not need to be known. The R_1GCV parameter estimate is defined by minimizing a certain function:

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

where $\gamma \in (0, 1)$ is the robustness parameter (here: $\gamma = 0.95$). Note that for $\gamma = 1$ the R_1GCV method is just GCV.

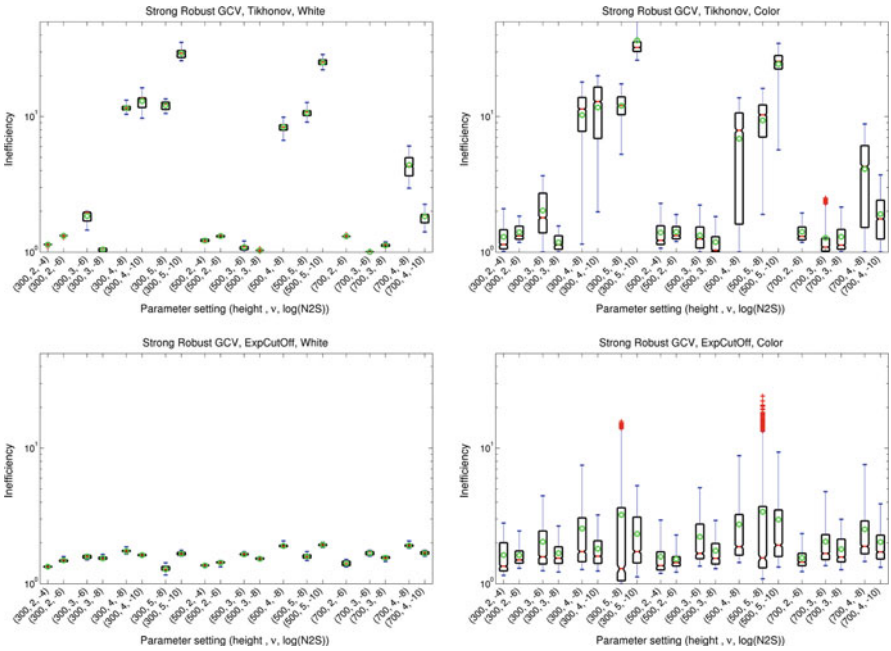

Fig. 20 Inefficiencies of strong robust generalized cross-validation

Known Issues 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 (cf. Lukas 2008). The last term in (32) can be explained by the fact that for uncorrelated errors with variance δ^2, we have the variance of x_n^δ given by $\mathbb{E}\|x_n^\delta - \mathbb{E}x_n^\delta\|^2 = \delta^2 \operatorname{tr}((A_n^{-1} * A_n^{-1}))$. In fact, for Tikhonov regularization of polynomially ill-posed problems, it is known from Lukas (2008) that as $q \to \infty$, the R_1GCV function in (32) 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 $\mathcal{X}$-norm risk (see Lukas 2008), 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 (cf. Lukas 2010).

If a SVD is used to compute $\operatorname{tr}(AA_n^{-1}))$ in (31), then it requires very little extra work to compute $T_n = \operatorname{tr}((A_n^{-1} * A_n^{-1})$ in (32). 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 $T_\alpha = -(d/d\alpha)(\operatorname{tr}(AA_\alpha^{-1})))$ (cf. Lukas 2008). It can be expected that R_1GCV can be extended in the same way as GCV to iterative regularization methods.

Results Figure 20 shows that R_1GCV mostly performs reasonably well across all the test cases for spectral cutoff in both the white and colored noise situations. For Tikhonov regularization, it performs well with solutions of lower smoothness ($\nu \leq 3$), but there is still a significant saturation effect. The results for colored noise are much better than for GCV or RGCV, but the results for spectral cutoff and white noise are a little worse than RGCV.

4.18 Modified Generalized Cross-Validation

The modified GCV method involves a simple modification of the GCV function that is designed to stabilize the method (cf. Cummins et al. 2001; Vio et al. 2004).

Method The required inputs are the same as for GCV (Sect. 4.15) plus:

- A stabilization parameter $c > 1$

The noise level does not need to be known. The modified GCV estimate is defined by

$$n_* = \underset{n \leq N}{\operatorname{argmin}} \left\{ \frac{\|Ax_n^\delta - y^\delta\|^2}{(q^{-1}\operatorname{tr}(I - cAA_n^{-1}))^2} \right\}, \tag{33}$$

where $c > 1$ is the stabilization parameter (here: $c = 3$). When $c = 1$, the method reduces to GCV.

Known Issues The effect of the factor c can be explained in terms of the degrees of freedom for the regularized solution, defined as $df = \operatorname{tr}(AA_n^{-1})$, where AA_n^{-1} is the influence matrix (cf. Cummins et al. 2001). Clearly the factor introduces a pole at q/c in the objective function as a function of df, which constrains the value of n so that $df < q/c$ and modifies the function's shape to prevent under-smoothing.

The modified GCV method for Tikhonov regularization is closely related to the RGCV method of Sect. 4.16 in the sense that, under appropriate conditions, they are asymptotically equivalent as the sample size $q \to \infty$ (see Lukas 2008). Clearly, the modified GCV estimate can be computed in the same way as the GCV estimate of Sect. 4.15. It can be expected that the modified GCV method can be extended in the same way as GCV to iterative regularization methods.

Results As seen in Fig. 21, the modified GCV method has reasonably good performance across all the test cases for spectral cutoff in both the white and colored noise situations. For Tikhonov regularization, it performs well in the cases with less smooth solutions, but, like GCV, it suffers from the saturation effect. With $c = 3$ here, the results for the colored noise situation are similar (slightly better) to those of R_1GCV in Fig. 20.

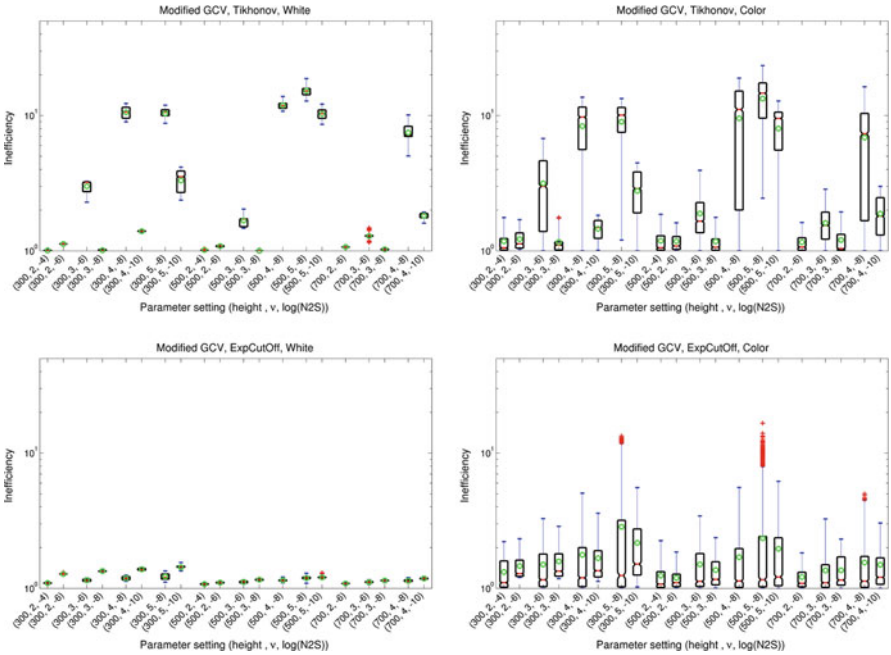

Fig. 21 Inefficiencies of modified generalized cross-validation

4.19 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) of Akaike (1973) behaves like GCV (cf. Eubank 1988).
  - The unbiased prediction risk method (also known as Mallows C_P or C_L) behaves like GCV (see Li 1987; Wahba 1990; Efron 2001).
  - The unbiased risk method of Lukas (1998a) and Cavalier and Golubev (2006) behaves like the modified discrepancy principle with a particular tuning constant.
  - Rule R1 of Hämarik and Raus (2009) behaves like the fast balancing principle with $\ell(n) = n + 1$ in (22).
- The method does not generalize in an obvious way to both Tikhonov regularization and spectral cutoff regularization:
  - Arcangeli's principle (see Arcangeli 1966; Groetsch and Schock 1984; Nair and Rajan 2002), which is an early method developed for Tikhonov regularization with deterministic noise

  - Rule R2 of Raus and Hämarik (2009) and Hämarik et al. (2011, 2012), which was developed for (iterated) Tikhonov regularization with deterministic noise
  - Risk hull method of Cavalier and Golubev (2006), which was developed for spectral cutoff 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 of Cavalier and Golubev (2006),
  - Modified balancing principle of Spokoiny and Vial (2009).

5 Conclusion

In this section, we will summarize the requirements and properties of the parameter choice methods described in Sect. 4 and then compare them with respect to their average performance in our numerical experiments. For standard uses, this (in combination with Bauer and Lukas 2011) 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 methods are based on principles or rationales that carry over to other regularizations, and so, if a regularization behaves like Tikhonov or spectral cutoff regularization, then similar results can be expected. Thus, our results for Tikhonov and spectral cutoff regularization are seen as model results for 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 the three groups according to the input information required as listed in the beginning of Sect. 4. We provide the following columns in Table 3:

"Spaces"	Function space(s) in which norms/inner products are computed.
"Setting"	Data setting, i.e., infinite data possible (∞) or finite data required (q).
"Tuning"	Use of tuning parameters (yes/no).
"Complexity"	Computational complexity, i.e., all regularized solutions needed (N) or bisection possible (log).
"Cal."	Requirements for expensive calculations such as traces (tr) or determinants (det) in each step.

Table 3 Requirements and general properties of the methods. (D(±) for the L-curve method means that there are both positive and negative results)

Method	Spaces	Tuning	Setting	Complexity	Cal.	Proofs	Gen.
Disc. princ.	$\mathcal{Y}$	y	∞	log	–	D+S	l+nl
Trans. disc.	$\mathcal{X}$	y	∞	log	–	D	–
MD rule	$\mathcal{Y}$	y	∞	log	–	D+S	l+nl
ME rule	$\mathcal{Y}$	y	∞	log	–	D	–
Vary. disc.	$\mathcal{X}$	y	∞	log	tr	S	l
Bal. pr. (wh.)	$\mathcal{X}$	y	∞	N	tr	D+S	l+nl
Balancing pr.	$\mathcal{X}$	y	∞	N	–	D+S	l+nl
Fast balancing	$\mathcal{X}$	y	∞	log	–	D+S	l
Hardened bal.	$\mathcal{X}$	n	∞	N	–	S	–
Hard. bal. (wh.)	$\mathcal{X}$	n	∞	N	tr	S	–
Quasi-opt.	$\mathcal{X}$	n	∞	log	–	D+S	l
L-curve	$\mathcal{X}, \mathcal{Y}$	n	∞	log	–	D(±)	–
MDP rule	$\mathcal{Y}$	n	∞	log	–	D	l
Extrap. err.	$\mathcal{X}, \mathcal{Y}$	n	q	log	–	–	–
NCP meth.	$\mathcal{Y}$	n	q	log	–	–	–
Residual meth.	$\mathcal{Y}$	n	∞	log	tr	S	–
GML	$\mathcal{Y}$	n	q	log	det	S	–
GCV	$\mathcal{Y}$	n	q	log	tr	S	l
Robust GCV	$\mathcal{Y}$	y	q	log	tr	S	–
Str. rob. GCV	$\mathcal{Y}$	y	q	log	tr	S	–
Modified GCV	$\mathcal{Y}$	y	q	log	tr	S	–

"Proofs" Availability of proofs in a deterministic (D) and/or stochastic (S) setting.

"Gen." Generalizations to other regularization methods (mostly Landweber) for linear (l) or nonlinear (nl) problems.

For further details (input, origin), see the corresponding sections. Note that the optimal use of a tuning parameter usually improves the results for a certain class of problems, but setting it incorrectly can give very poor results. Setting the parameter is especially difficult if there is no theory about its effect on the regularized solution or if one cannot necessarily check the validity of a certain setting. In practice, methods without (or with very robust) tuning parameters are usually preferable. All statements given are to the best knowledge of the authors; some might change in the course of further research. For further details, we refer to Bauer and Lukas (2011).

5.2 Average Performance

For each method evaluated in Sect. 4, there are four displayed figure panels (Tikhonov/ExpCutOff with white/colored noise) with a box plot for every test

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

	ExpCutOff, White		ExpCutOff, Color		Tikhonov, White		Tikhonov, Color	
Method	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Disc. princ.	1.59	1.51	3.14	1.52	5.91	1.53	4.83	1.54
Trans. disc.	1.47	1.43	2.66	1.44	**1.32**	**1.31**	4.16	**1.30**
MD rule	**1.12**	**1.11**	**1.88**	**1.13**	1.49	1.42	5.98	1.42
ME rule	1.48	1.44	2.63	1.48	1.36	1.38	**3.80**	1.37
Vary. disc.	1.27	1.26	2.20	1.26	4.60	1.35	3.92	1.42
Bal. pr. (wh.)	1.41	1.36	2.51	1.38	1.91	1.60	7.73	1.59
Balancing pr.	1.41	1.36	1.38	1.35	1.91	1.60	1.83	1.58
Fast balancing	1.48	1.43	1.43	1.42	2.37	1.90	2.24	1.88
Hardened bal.	**1.01**	**1.01**	**1.01**	**1.01**	**1.03**	**1.02**	**1.03**	**1.03**
Hard. bal. (wh.)	1.01	1.01	7.50	**1.04**	**1.02**	**1.02**	4.92	**1.04**
Quasi-opt.	1.05	1.04	**1.06**	1.05	1.06	1.07	**1.06**	1.08
L-curve	27,349	1,548	30,140	1,602	2,676	1,024	5,211	1,031
MDP rule	1.05	1.04	**1.06**	1.06	1.71	1.61	3.28	1.63
Extrap. err.	42.45	1.85	60.85	1.92	7.11	1.57	7.31	1.57
NCP meth.	1.16	1.14	25.08	2.08	3.03	1.13	19.65	1.84
Residual meth.	1.28	1.25	1.87	1.33	9.87	1.95	18.29	2.03
GML	77,361	3,257	73,677	3,140	8.25	2.02	8.31	1.97
GCV	**1.00**	**1.00**	1,487	1.66	17.42	3.87	316.4	3.95
Robust GCV	1.09	1.08	717.8	1.39	13.72	3.08	178.8	2.85
Str. rob. GCV	1.60	1.58	2.15	1.58	6.49	1.60	6.61	1.54
Modified GCV	1.19	1.15	1.63	1.16	4.34	1.56	3.87	1.60

case, showing the sample mean(I) (green $\circ$) and sample median(I) (red $-$) of the computed inefficiencies I. The average performance of the method across all the test cases is measured using both mean(mean(I)) and median(median(I)), and these results are displayed in Table 4, which again has three groups according to the input information required.

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 for white and colored noise:

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

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

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 recent numerical studies of Palm (2010), Bauer and Lukas (2011), Hämarik et al. (2011), and Reichel and Rodriguez (2013). The three of these other than Bauer and Lukas (2011) use the set of test problems of Hansen (1994) with uniformly distributed white noise and colored noise. Comparing the results here with Bauer and Lukas (2011), the multiplicity of the eigenvalues in our geomathematically motivated tests seems to have a stabilizing influence on many methods, as they clearly show less variance and fewer outliers here.

From Table 4, some methods performed much worse for 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 which have good performance for colored noise. This is especially true as the methods that performed well for colored noise also performed quite well for the white noise situation. It should be remarked that often the application of the regularization method requires much more computational effort than the evaluation of the parameter choice. Therefore, it can be advisable to use several parameter choice methods to help find the best choice of the parameter.

References

Abascal J-F, Arridge SR, Bayford RH, Holder DS (2008) Comparison of methods for optimal choice of the regularization parameter for linear electrical impedance tomography of brain function. Physiol Meas 29:1319–1334

Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: Second international symposium on information theory (Tsahkadsor, 1971). Akadémiai Kiadó, Budapest, pp 267–281

Åkesson EO, Daun KJ (2008) Parameter selection methods for axisymmetric flame tomography through Tikhonov regularization. Appl Opt 47:407–416

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

Anderssen RS, Bloomfield P (1974) Numerical differentiation procedures for non-exact data. Numer Math 22:157–182

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

Bakushinskii AB (1984) Remarks on choosing a regularization parameter using the quasi-optimality and ratio criterion. USSR Comput Math Math Phys 24:181–182

Bakushinsky A, Smirnova A (2005) On application of generalized discrepancy principle to iterative methods for nonlinear ill-posed problems. Numer Funct Anal Optim 26(1):35–48

Bauer F (2007) Some considerations concerning regularization and parameter choice algorithms. Inverse Probl 23(2):837–858

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

Bauer F, Kindermann S (2008) The quasi-optimality criterion for classical inverse problems. Inverse Probl 24:035002, 20

Bauer F, Kindermann S (2009) Recent results on the quasi-optimality principle. J Inverse Ill-Posed Probl 17(1):5–18

Bauer F, Lukas MA (2011) Comparing parameter choice methods for regularization of ill-posed problems. Math Comput Simul 81(9):1795–1841

Bauer F, Mathé P (2011) Parameter choice methods using minimization schemes. J Complex 27:68–85

Bauer F, Munk A (2007) Optimal regularization for ill-posed problems in metric spaces. J Inverse Ill-Posed Probl 15(2):137–148

Bauer F, Pereverzev S (2005) Regularization without preliminary knowledge of smoothness and error behavior. Eur J Appl Math 16(3):303–317

Bauer F, Reiß M (2008) Regularization independent of the noise level: an analysis of quasi-optimality. Inverse Probl 24:055009, 16

Bauer F, Hohage T, Munk A (2009) Iteratively regularized Gauss–Newton method for nonlinear inverse problems with random noise. SIAM J Numer Anal 47(3):1827–1846

Becker SMA (2011) Regularization of statistical inverse problems and the Bakushinskii veto. Inverse Probl 27:115010, 22

Blanchard G, Mathé P (2012) Discrepancy principle for statistical inverse problems with application to conjugate gradient iteration. Inverse Probl 28:115011, 23

Brezinski C, Rodriguez G, Seatzu S (2008) Error estimates for linear systems with applications to regularization. Numer Algorithms 49(1–4):85–104

Brezinski C, Rodriguez G, Seatzu S (2009) Error estimates for the regularization of least squares problems. Numer Algorithms 51(1):61–76

Brown LD, Low MG (1996) Asymptotic equivalence of nonparametric regression and white noise. Ann Stat 24(6):2384–2398

Calvetti D, Hansen PC, Reichel L (2002) L-curve curvature bounds via Lanczos bidiagonalization. Electron Trans Numer Anal 14:20–35

Candès EJ (2006) Modern statistical estimation via oracle inequalities. Acta Numer 15:257–325

Cavalier L (2008) Nonparametric statistical inverse problems. Inverse Probl 24(3):034004, 19

Cavalier L, Golubev Y (2006) Risk hull method and regularization by projections of ill-posed inverse problems. Ann Stat 34(4):1653–1677

Cavalier L, Golubev GK, Picard D, Tsybakov AB (2002) Oracle inequalities for inverse problems. Ann Stat 30(3):843–874

Cavalier L, Golubev Y, Lepski O, Tsybakov A (2004) Block thresholding and sharp adaptive estimation in severely ill-posed inverse problems. Theory Probab Appl 48(3):426–446

Colton D, Kress R (1998) Inverse acoustic and electromagnetic scattering theory, 2nd edn. Springer, Berlin

Correia T, Gibson A, Schweiger M, Hebden J (2009) Selection of regularization parameter for optical topography. J Biomed Opt 14(3):034044, 11

Cox DD (1988) Approximation of method of regularization estimators. Ann Stat 16(2):694–712

Craven P, Wahba G (1979) Smoothing noisy data with spline functions. Numer Math 31:377–403

Cummins DJ, Filloon TG, Nychka D (2001) Confidence intervals for nonparametric curve estimates: toward more uniform pointwise coverage. J Am Stat Assoc 96(453):233–246

Davies AR (1982) On the maximum likelihood regularization of Fredholm convolution equations of the first kind. In: Baker C, Miller G (eds) Treatment of integral equations by numerical methods. Academic, London, pp 95–105

Davies AR, Anderssen RS (1986) Improved estimates of statistical regularization parameters in Fourier differentiation and smoothing. Numer Math 48:671–697

Ditmar P, Kusche J, Klees R (2003) Computation of spherical harmonic coefficients from gravity gradiometry data to be acquired by the GOCE satellite: regularization issues. J Geod 77(7–8):465–477

Ditmar P, Klees R, Liu X (2007) Frequency-dependent data weighting in global gravity field modeling from satellite data contaminated by non-stationary noise. J Geod 81(1):81–96

Efron B (2001) Selection criteria for scatterplot smoothers. Ann Stat 29(2):470–504

Eggermont PN, LaRiccia V, Nashed MZ (2014) Noise models for ill-posed problems. In: Freeden W, Nashed MZ, Sonar T (eds) Handbook of geomathematics, 2nd edn. Springer, Heidelberg

Eldén L (1984) A note on the computation of the generalized cross-validation function for ill-conditioned least squares problems. BIT 24:467–472

Engl HW (1993) Regularization methods for the stable solution of inverse problems. Surv Math Ind 3(2):71–143

Engl HW, Gfrerer H (1988) A posteriori parameter choice for general regularization methods for solving linear ill-posed problems. Appl Numer Math 4(5):395–417

Engl HW, Scherzer O (2000) Convergence rate results for iterative methods for solving nonlinear ill-posed problems. In: Colton D, Engl H, Louis A, McLaughlin J, Rundell W (eds) Survey on solution methods for inverse problems. Springer, New York, pp 7–34

Engl HW, Hanke H, Neubauer A (1996) Regularization of inverse problems. Kluwer, Dordrecht

Eubank RL (1988) Spline smoothing and nonparametric regression. Marcel Dekker, New York

Evans SN, Stark PB (2002) Inverse problems as statistics. Inverse Probl 18(4):R55–R97

Farquharson CG, Oldenburg DW (2004) A comparison of automatic techniques for estimating the regularization parameter in non-linear inverse problems. Geophys J Int 156:411–425

Fitzpatrick BG (1991) Bayesian analysis in inverse problems. Inverse Probl 7(5):675–702

Freeden W (1999) Multiscale modelling of spaceborne geodata. Teubner, Leipzig

Freeden W, Gerhards C (2013) Geomathematically oriented potential theory. Chapman & Hall/CRC, Boca Raton

Freeden W, Gutting M (2013) Special functions of mathematical (geo-)physics. Birkhäuser, Basel

Freeden W, Michel V (2004) Multiscale potential theory (with applications to geoscience). Birkhäuser, Boston

Freeden W, Schreiner M (2015) Satellite gravity gradiometry (SGG): from scalar to tensorial solution. In: Freeden W, Nashed MZ, Sonar T (Eds) Handbook of Geomathematics, 2nd Edn. Springer

Freeden W, Gervens T, Schreiner M (1998) Constructive approximation on the sphere. Oxford University Press, Oxford

Gfrerer H (1987) An a posteriori parameter choice for ordinary and iterated Tikhonov regularization of ill-posed problems leading to optimal convergence rates. Math Comput 49(180):507–522

Girard D (1989) A fast "Monte-Carlo cross-validation" procedure for large least squares problems with noisy data. Numer Math 56(1):1–23

Glasko V, Kriksin Y (1984) On the quasioptimality principle for linear ill-posed problems in Hilbert space. Vychisl Math Math Fiz 24:1603–1613

Goldenshluger A, Pereverzev S (2000) Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations. Probl Theory Relat Fields 118(2):169–186

Golub GH, von Matt U (1997) Generalized cross-validation for large scale problems. J Comput Graph Stat 6(1):1–34

Golub GH, Heath M, Wahba G (1979) Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics 21:215–223

Grad J, Zakrajšek E (1972) LR algorithm with Laguerre shift for symmetric tridiagonal matrices. Comput J 15(3):268–270

Groetsch CW (1984) The theory of Tikhonov regularization for Fredholm equations of the first kind. Pitman, Boston

Groetsch CW, Neubauer A (1989) Regularization of ill-posed problems: optimal parameter choice in finite dimensions. J Approx Theory 58(2):184–200

Groetsch CW, Schock E (1984) Asymptotic convergence rate of Arcangeli's method for ill-posed problems. Appl Anal 18:175–182

Gu C (2002) Smoothing spline ANOVA models. Springer, New York

Gu C, Bates DM, Chen Z, Wahba G (1989) 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(4):457–480

Haber E, Oldenburg DW (2000) A GCV based method for nonlinear ill-posed problems. Comput Geosci 4:41–63

Hämarik U, Raus T (1999) On the a posteriori parameter choice in regularization methods. Proc Estonian Acad Sci Phys Math 48(2):133–145

Hämarik U, Raus T (2006) On the choice of the regularization parameter in ill-posed problems with approximately given noise level of data. J Inverse Ill-Posed Probl 14(3):251–266

Hämarik U, Raus T (2009) About the balancing principle for choice of the regularization parameter. Numer Funct Anal Optim 30(9–10):951–970

Hämarik U, Tautenhahn U (2001) On the monotone error rule for parameter choice in iterative and continuous regularization methods. BIT 41(5):1029–1038

Hämarik U, Tautenhahn U (2003) On the monotone error rule for choosing the regularization parameter in ill-posed problems. In: Lavrent'ev MM et al (ed) Ill-posed and non-classical problems of mathematical physics and analysis. VSP, Utrecht, pp 27–55

Hämarik U, Palm R, Raus T (2009) On minimization strategies for choice of the regularization parameter in ill-posed problems. Numer Funct Anal Optim 30(9–10):924–950

Hämarik U, Palm R, Raus T (2011) Comparison of parameter choices in regularization algorithms in case of different information about noise level. Calcolo 48:47–59

Hämarik U, Palm R, Raus T (2012) A family of rules for parameter choices in Tikhonov regularization of ill-posed problems with inexact noise level. J Comput Appl Math 236: 2146–2157

Hanke M (1996) Limitations of the L-curve method in ill-posed problems. BIT 36(2):287–301

Hanke M, Hansen PC (1993) Regularization methods for large-scale problems. Surv Math Ind 3(4):253–315

Hanke M, Raus T (1996) A general heuristic for choosing the regularization parameter in ill-posed problems. SIAM J Sci Comput 17(4):956–972

Hansen PC (1992) Analysis of discrete ill-posed problems by means of the L-curve. SIAM Rev 34(4):561–580

Hansen PC (1994) Regularization tools: a Matlab package for analysis and solution of discrete ill-posed problems. Numer Algorithms 6:1–35

Hansen PC (1998) Rank-deficient and discrete ill-posed problems. SIAM, Philadelphia

Hansen PC (2001) The L-curve and its use in the numerical treatment of inverse problems. In: Johnstone PR (ed) Computational inverse problems in electrocardiography. WIT, Southampton, pp 119–142

Hansen PC, O'Leary DP (1993) The use of the L-curve in the regularization of discrete ill-posed problems. SIAM J Sci Comput 14(6):1487–1503

Hansen PC, Kilmer ME, Kjeldsen RH (2006) Exploiting residual information in the parameter choice for discrete ill-posed problems. BIT 46(1):41–59

Hansen PC, Jensen TK, Rodriguez G (2007) An adaptive pruning algorithm for the discrete L-curve criterion. J Comput Appl Math 198(2):483–492

Hofinger A, Pikkarainen HK (2007) Convergence rate for the Bayesian approach to linear inverse problems. Inverse Probl 23(6):2469–2484

Hofmann B (1986) Regularization of applied inverse and ill-posed problems. Teubner, Leipzig

Hofmann B, Mathé P (2007) Analysis of profile functions for general linear regularization methods. SIAM J Numer Anal 45(3):1122–1141

Hohage T (2000) Regularization of exponentially ill-posed problems. Numer Funct Anal Optim 21:439–464

Hutchinson M (1989) A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines. Commun Stat Simul Comput 18(3):1059–1076

Hutchinson MF, de Hoog FR (1985) Smoothing noisy data with spline functions. Numer Math 47:99–106

Jansen M, Malfait M, Bultheel A (1997) Generalized cross validation for wavelet thresholding. Signal Process 56(1):33–44

Jin Q-N (2000) On the iteratively regularized Gauss–Newton method for solving nonlinear ill-posed problems. Math Comput 69(232):1603–1623

Jin Q-N, Hou Z-Y (1999) On an a posteriori parameter choice strategy for Tikhonov regularization of nonlinear ill-posed problems. Numer Math 83(1):139–159

Jin Q, Tautenhahn U (2009) On the discrepancy principle for some Newton type methods for solving nonlinear inverse problems. Numer Math 111(4):509–558

Johnstone PR, Gulrajani RM (2000) Selecting the corner in the L-curve approach to Tikhonov regularization. IEEE Trans Biomed Eng 47(9):1293–1296

Kaipio J, Somersalo E (2005) Statistical and computational inverse problems. Springer, New York

Kaltenbacher B, Neubauer A, Scherzer O (2008) Iterative regularization methods for nonlinear ill-posed problems. Walter de Gruyter, Berlin

Kilmer ME, O'Leary DP (2001) Choosing regularization parameters in iterative methods for ill-posed problems. SIAM J Matrix Anal Appl 22(4):1204–1221

Kindermann S (2011) Convergence analysis of minimization-based noise level-free parameter choice rules for linear ill-posed problems. Electron Trans Numer Anal 38:233–257

Kindermann S, Neubauer A (2008) On the convergence of the quasioptimality criterion for (iterated) Tikhonov regularization. Inverse Probl Imaging 2(2):291–299

Kohn R, Ansley CF, Tharm D (1991) The performance of cross-validation and maximum likelihood estimators of spline smoothing parameters. J Am Stat Assoc 86:1042–1050

Kou SC, Efron B (2002) Smoothers and the C_p, generalized maximum likelihood, and extended exponential criteria: a geometric approach. J Am Stat Assoc 97(459):766–782

Larkin FM (1972) Gaussian measure in Hilbert space and applications in numerical analysis. Rocky Mt J Math 2:379–421

Lawson CL, Hanson RJ (1974) Solving least squares problems. Prentice-Hall, Englewood Cliffs

Leonov A (1979) Justification of the choice of regularization parameter according to quasi-optimality and quotient criteria. USSR Comput Math Math Phys 18(6):1–15

Lepskij O (1990) On a problem of adaptive estimation in Gaussian white noise. Theory Probab Appl 35(3):454–466

Li K-C (1986) Asymptotic optimality of C_L and generalized cross-validation in ridge regression with application to spline smoothing. Ann Stat 14:1101–1112

Li K-C (1987) Asymptotic optimality for C_p, C_L, cross-validation and generalized cross-validation: discret index set. Ann Stat 15:958–975

Lu S, Mathé P (2013) Heuristic parameter selection based on functional minimization: optimality and model function approach. Math Comput 82(283):1609–1630

Lu S, Mathé P (2014) Discrepancy based model selection in statistical inverse problems. J Complex 30(3):290-308

Lu S, Pereverzev S (2014) Multiparameter regularization in downward continuation of satellite data. In: Freeden W, Nashed MZ, Sonar T (eds) Handbook of geomathematics, 2nd edn. Springer, Heidelberg

Lukas MA (1988) Convergence rates for regularized solutions. Math Comput 51(183):107–131

Lukas MA (1993) Asymptotic optimality of generalized cross-validation for choosing the regularization parameter. Numer Math 66(1):41–66

Lukas MA (1995) On the discrepancy principle and generalised maximum likelihood for regularisation. Bull Aust Math Soc 52(3):399–424

Lukas MA (1998a) Asymptotic behaviour of the minimum bound method for choosing the regularization parameter. Inverse Probl 14(1):149–159

Lukas MA (1998b) Comparisons of parameter choice methods for regularization with discrete noisy data. Inverse Probl 14(1):161–184

Lukas MA (2006) Robust generalized cross-validation for choosing the regularization parameter. Inverse Probl 22(5):1883–1902

Lukas MA (2008) Strong robust generalized cross-validation for choosing the regularization parameter. Inverse Probl 24:034006, 16

Lukas MA (2010) Robust GCV choice of the regularization parameter for correlated data. J Integral Equ Appl 22(3):519–547

Mair BA (1994) Tikhonov regularization for finitely and infinitely smoothing operators. SIAM J Math Anal 25:135–147

Mair BA, Ruymgaart FH (1996) Statistical inverse estimation in Hilbert scales. SIAM J Appl Math 56(5):1424–1444

Mathé P (2006) What do we learn from the discrepancy principle? Z Anal Anwend 25(4):411–420

Mathé P, Pereverzev SV (2003) Geometry of linear ill-posed problems in variable Hilbert spaces. Inverse Probl 19(3):789–803

Mathé P, Pereverzev SV (2006) Regularization of some linear ill-posed problems with discretized random noisy data. Math Comput 75(256):1913–1929

Michel V (2014) Tomography: problems and multiscale solutions. In: Freeden W, Nashed MZ, Sonar T (eds) Handbook of geomathematics, 2nd edn. Springer, Heidelberg

Morozov VA (1966) On the solution of functional equations by the method of regularization. Soviet Math Dokl 7:414–417

Morozov VA (1984) Methods for solving incorrectly posed problems. Springer, New York

Nair MT, Rajan MP (2002) Generalized Arcangeli's discrepancy principles for a class of regularization methods for solving ill-posed problems. J Inverse Ill-Posed Probl 10(3):281–294

Nair MT, Schock E, Tautenhahn U (2003) Morozov's discrepancy principle under general source conditions. Z Anal Anwend 22:199–214

Neubauer A (1988) An a posteriori parameter choice for Tikhonov regularization in the presence of modeling error. Appl Numer Math 4(6):507–519

Neubauer A (2008) The convergence of a new heuristic parameter selection criterion for general regularization methods. Inverse Probl 24(5):055005, 10

Opsomer J, Wang Y, Yang Y (2010) Nonparametric regression with correlated errors. Stat Sci 16:134–153

Palm R (2010) Numerical comparison of regularization algorithms for solving ill-posed problems. PhD thesis, University of Tartu, Estonia

Pensky M, Sapatinas T (2010) On convergence rates equivalency and sampling strategies in functional deconvolution models. Ann Stat 38(3):1793–1844

Pereverzev S, Schock E (2000) Morozov's discrepancy principle for Tikhonov regularization of severely ill-posed problems in finite dimensional subspaces. Numer Funct Anal Optim 21: 901–916

Phillips D (1962) A technique for the numerical solution of certain integral equations of the first kind. J Assoc Comput Mach 9:84–97

Pohl S, Hofmann B, Neubert R, Otto T, Radehaus C (2001) A regularization approach for the determination of remission curves. Inverse Probl Eng 9(2):157–174

Raus T (1984) On the discrepancy principle for the solution of ill-posed problems. Uch Zap Tartu Gos Univ 672:16–26

Raus T (1985) The principle of the residual in the solution of ill-posed problems with nonselfadjoint operator. Tartu Riikl Ül Toimetised 715:12–20

Raus T (1990) 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, pp 73–87

Raus T (1992) About regularization parameter choice in case of approximately given error bounds of data. In: Vainikko G (ed) Methods for solution of integral equations and ill-posed problems. Tartu University, Tartu, pp 77–89

Raus T, Hämarik U (2007) On the quasioptimal regularization parameter choices for solving ill-posed problems. J Inverse Ill-Posed Probl 15(4):419–439

Raus T, Hämarik U (2009) New rule for choice of the regularization parameter in (iterated) Tikhonov method. Math Model Anal 14:187–198

Reginska T (1996) A regularization parameter in discrete ill-posed problems. SIAM J Sci Comput 17(3):740–749
Reichel L, Rodriguez G (2013) Old and new parameter choice rules for discrete ill-posed problems. Numer Algorithms 63:65–87
Robinson T, Moyeed R (1989) Making robust the cross-validatory choice of smoothing parameter in spline smoothing regression. Commun Stat Theory Methods 18(2):523–539
Rust BW (2000) Parameter selection for constrained solutions to ill-posed problems. Comput Sci Stat 32:333–347
Rust BW, O'Leary DP (2008) Residual periodograms for choosing regularization parameters for ill-posed problems. Inverse Probl 24(3):034005, 30
Santos R, De Pierro A (2003) A cheaper way to compute generalized cross-validation as a stopping rule for linear stationary iterative methods. J Comput Graph Stat 12(2):417–433
Scherzer O, Engl H, Kunisch K (1993) Optimal a posteriori parameter choice for Tikhonov regularization for solving nonlinear ill-posed problems. SIAM J Numer Anal 30(6):1796–1838
Spokoiny V, Vial C (2009) Parameter tuning in pointwise adaptation using a propagation approach. Ann Stat 37:2783–2807
Tarantola A (1987) Inverse problem theory: methods for data fitting and model parameter estimation. Elsevier, Amsterdam
Tautenhahn U, Hämarik U (1999) The use of monotonicity for choosing the regularization parameter in ill-posed problems. Inverse Probl 15(6):1487–1505
Thompson AM, Kay JW, Titterington DM (1989) A cautionary note about crossvalidatory choice. J Stat Comput Simul 33:199–216
Thompson AM, Brown JC, Kay JW, Titterington DM (1991) A study of methods for choosing the smoothing parameter in image restoration by regularization. IEEE Trans Pattern Anal Machine Intell 13:3326–3339
Tikhonov A, Arsenin V (1977) Solutions of ill-posed problems. Wiley, New York
Tikhonov A, Glasko V (1965) Use of the regularization method in non-linear problems. USSR Comput Math Math Phys 5(3):93–107
Tsybakov A (2000) On the best rate of adaptive estimation in some inverse problems. C R Acad Sci Paris Sér I Math 330(9):835–840
Vio R, Ma P, Zhong W, Nagy J, Tenorio L, Wamsteker W (2004) Estimation of regularization parameters in multiple-image deblurring. Astron Astrophys 423:1179–1186
Vogel CR (1986) 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(1):109–117
Vogel CR (1996) Non-convergence of the L-curve regularization parameter selection method. Inverse Probl 12(4):535–547
Vogel CR (2002) Computational methods for inverse problems. SIAM, Philadelphia
Wahba G (1977) Practical approximate solutions to linear operator equations when the data are noisy. SIAM J Numer Anal 14(4):651–667
Wahba G (1985) A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem. Ann Stat 13:1378–1402
Wahba G (1990) Spline models for observational data. SIAM, Philadelphia
Wahba G, Wang YH (1990) When is the optimal regularization parameter insensitive to the choice of the loss function? Commun Stat Theory Methods 19(5):1685–1700
Wang Y (1998) Smoothing spline models with correlated random errors. J Am Stat Assoc 93: 341–348
Wecker WE, Ansley CF (1983) The signal extraction approach to nonlinear regression and spline smoothing. J Am Stat Assoc 78(381):81–89
Yagola AG, Leonov AS, Titarenko VN (2002) Data errors and an error estimation for ill-posed problems. Inverse Probl Eng 10(2):117–129