Calcolo (2011) 48: 47–59
DOI 10.1007/s10092-010-0027-4

Comparison of parameter choices in regularization algorithms in case of different information about noise level

Uno Hämarik · Reimo Palm · Toomas Raus

Received: 12 November 2009 / Accepted: 5 May 2010 / Published online: 25 September 2010

Abstract We consider linear ill-posed problems in Hilbert space with noisy data. The noise level may be given exactly or approximately or there may be no information about the noise level. We regularize the problem using the Landweber method, the Tikhonov method or the extrapolated version of the Tikhonov method. For all three cases of noise information we propose rules for choice of the regularization parameter. Extensive numerical experiments show the advantage of the proposed rules over known rules, including the discrepancy principle, the quasioptimality criterion, the Hanke-Raus rule, the Brezinski-Rodriguez-Seatzu rule and others. Numerical comparison also shows at which information about the noise level our rules for approximately given noise level should be preferred to other rules.

Keywords Ill-posed problem · Noise level · Regularization · Tikhonov method · Extrapolated Tikhonov method · Landweber method · Regularization parameter choice

Mathematics Subject Classification (1991) Primary 65J20 · Secondary 47A52

1 Introduction

Let $A\colon X \to Y$ be a linear bounded operator between real Hilbert spaces. We are interested in finding the minimum norm solution x_* of the equation

$$Ax = y_*, \quad y_* \in \mathcal{R}(A). \tag{1}$$

This paper is part of special issue devoted to the 2nd Dolomites Workshop on Constructive Approximation and Applications, 2009.

U. Hämarik (✉) · R. Palm · T. Raus
Faculty of Mathematics and Informatics, University of Tartu, J. Liivi 2, 50409 Tartu, Estonia
e-mail: uno.hamarik@ut.ee

We do not assume the range $\mathcal{R}(A)$ to be closed or the kernel $\mathcal{N}(A)$ to be trivial, so in general this problem is ill-posed. For solving ill-posed problems regularization methods are used [8, 28, 30–32]. As usual in studying ill-posed problems, we assume that instead of exact data y_*, noisy data y are given. We consider the knowledge about the noise level in one of the following forms.

1. Noise level is known fully: given is δ with $\|y - y_*\| \le \delta$.
2. Noise level is not known.
3. Noise level is known approximately: δ is given but it is not known whether the inequality $\|y - y_*\| \le \delta$ holds or not. For example, it may be known that with high probability $\delta/\|y - y_*\| \in [1/10, 10]$.

Typically only the cases 1, 2 are considered. Investigation of the case 3 is motivated by the fact that classical parameter choice rules that use the noise level (e.g. the discrepancy principle) need exact noise level: they fail in case of underestimated noise level and give large error in case of very moderate overestimation of the noise level. On the other hand, for heuristic rules that do not use the noise level, the convergence of approximate solutions as $\|y - y_*\| \to 0$ can not be guaranteed (see [1]). To our opinion, the knowledge about the noise level is often between the extreme cases 1 (full information) and 2 (no information) and thus the case 3 needs the attention. In [14, 15] we proposed the rule DM for the case 3, guaranteeing the convergence of approximate solutions to the exact solution, as $\delta \to 0$, provided that $\lim_{\delta\to 0} \frac{\|y-y_*\|}{\delta} \le C$, where C is an unknown constant. This rule contains two parameters about which in [14, 15] only general recommendations were given. In this paper we consider regularization by the Landweber iteration method, the Tikhonov method and the extrapolated variant of the Tikhonov method (see also [10, 12, 25], where conjugate gradient type methods are included as well) and optimize the parameters of the DM rule over the test problems of [20]. We find out numerically, what is the extent of over- or underestimation of the noise level at which the results in the rule DM are better than in heuristic rules (the quasioptimality criterion [28, 29] (see also [2–4, 9, 21, 22, 24]), the Hanke-Raus rule [18, 21], the Brezinski-Rodriguez-Seatzu rule [7] and our modifications [12] of these rules) and in the discrepancy principle (and in other, proposed rules).

2 Regularization methods

For the solution of the problem $Ax = y_*$ we consider basic regularization methods. In *Landweber method* (cf. [8, 32]) the approximations are computed iteratively as

$$x_n = x_{n-1} - \mu A^*(Ax_{n-1} - y) \quad (n = 1, 2, \ldots)$$

with $\mu \in (0, 2/\|A^*A\|)$ and with some $x_0 = \overline{x} \in X$, typically $\overline{x} = 0$. Many papers note that the Landweber method is not practical, since it needs too many iterations. If possible, we strongly recommend to implement this method by the operator form of iterations (also recommended in [30, 32]), which allows to compute $x_n = (I - A^*AC_k)\overline{x} + C_kA^*y$ for the indices $n = 2^k$ $(k = 1, 2, \ldots)$, using the operators

$$C_0 = \mu I, \qquad C_k = C_{k-1}(2I - A^*AC_{k-1}) \quad (k = 1, 2, \ldots).$$

The *m-iterated Tikhonov method* ($m = 1, 2, \ldots$; cf. [8, 32]) is defined as follows. Take $x_{0;\alpha} = \overline{x}$ and compute $x_{1;\alpha}, \ldots, x_{m;\alpha}$ iteratively from

$$\alpha x_{n;\alpha} + A^* A x_{n;\alpha} = \alpha x_{n-1;\alpha} + A^* y \quad (n = 1, \ldots, m); \tag{2}$$

the approximate solution of (1) is then $x_{m;\alpha}$. In case $m = 1$ (ordinary Tikhonov method) we use the notation $x_\alpha := x_{1;\alpha}$.

We also use extrapolated Tikhonov approximations, which are linear combinations of x_α with different α. Let $x_{\alpha_1}, \ldots, x_{\alpha_m}$ be Tikhonov approximations with pairwise different parameters $\alpha_1, \ldots, \alpha_m$. The *m-extrapolated Tikhonov approximation* is given by

$$x_{\alpha_1,\ldots,\alpha_m} = \sum_{i=1}^{m} d_i x_{\alpha_i}, \quad d_i = \prod_{\substack{j=1 \\ j \neq i}}^{m} \left(1 - \frac{\alpha_i}{\alpha_j}\right)^{-1} \tag{3}$$

(see [11, 13]; other algorithms for extrapolation can be found in [5, 6]). The coefficients d_i are chosen in such a way that the leading terms in the error expansion are eliminated. It is easy to see that $\sum_{i=1}^{m} d_i = 1$ but the coefficients have alternating signs, so $x_{\alpha_1,\ldots,\alpha_m}$ is not a convex combination of x_{α_i}. As shown in [11], the approximations (3) coincide with the nonstationary iterated Tikhonov approximations [17], using α_n instead of α on the step n in (2). For large m and similar α's this way is computationally more stable than the formula (3). We use logarithmically uniform mesh of parameters $\alpha_n = q^{n-1}$ ($q < 1$; $n = 1, 2, \ldots$).

The use of the extrapolated Tikhonov method is motivated by Theorems 3, 4 of Sect. 4, which indicate high accuracy. Compared with the iterated Tikhonov method the amount of computations is less: in a posteriori parameter choice at transition $x_{m;\alpha_i}$ to $x_{m;\alpha_{i+1}}$ we need to solve m equations in case of m-iterated Tikhonov method but only one equation in case of extrapolated approximation with m terms.

3 Rules for choice of the regularization parameter

An important problem, when applying regularization methods, is the proper choice of the regularization parameter. We consider parameter choice depending on the information about the noise level.

3.1 Rules for known noise level

In this subsection we assume that δ with $\|y - y_*\| \le \delta$ is given.

Discrepancy principle (rule D) [23, 30–32]; see Theorem 1. Let $C \ge 1$ be a fixed constant. In (iterated) Tikhonov method the regularization parameter $\alpha = \alpha_{\mathrm{D}}$ is the solution of the equation $\|r_{m;\alpha}\| = C\delta$, $r_{m;\alpha} := Ax_{m;\alpha} - y$. In iterative methods the regularization parameter $n = n_{\mathrm{D}}$ is chosen as the first n for which $\|r_n\| \le C\delta$, $r_n := Ax_n - y$.

Monotone error rule (ME-rule) [27]; see Theorem 1. In m-iterated Tikhonov method the ME-rule chooses $\alpha = \alpha_{\mathrm{ME}}$ from the equation

$$d_{\mathrm{ME}}(\alpha) = (r_{m;\alpha}, r_{m+1;\alpha}) / \|r_{m+1;\alpha}\| = \delta, \quad r_{m;\alpha} = Ax_{m;\alpha} - y,$$

where $(\cdot,\cdot)$ denotes the scalar product. The monotone error rule guarantees $\frac{\mathrm{d}}{\mathrm{d}\alpha}\|x_{m;\alpha}-x_*\|\geq 0$ for all $\alpha\in[\alpha_{\mathrm{ME}},\infty)$.

In case of iterative methods of the form

$$x_{n+1}=x_n-A^*z_n,\quad z_n\in Y\ (n=0,1,\ldots)$$

the ME-rule [10, 16] chooses the regularization parameter (the stopping index) $n=n_{\mathrm{ME}}$ as the first n for which

$$d_{\mathrm{ME}}(n)=(r_n+r_{n+1},z_n)/(2\|z_n\|)\leq\delta.$$

This guarantees the monotonicity property $\|x_n-x_*\|\leq\|x_{n-1}-x_*\|$ for all $n=1,2,\ldots,n_{\mathrm{ME}}$.

Rule R2. In [26] (see Theorem 2) it was proposed to choose $\alpha=\alpha_{R2}$ in (m-iterated) Tikhonov method as the smallest solution of the equation

$$d_{\mathrm{R2}}(\alpha)=\frac{\sqrt{\alpha}\|x_{m;\alpha}-x_{m+1;\alpha}\|^2\kappa(\alpha)}{(x_{m;\alpha}-x_{m+1;\alpha},x_{m+1;\alpha}-x_{m+2;\alpha})^{1/2}}=C_m\delta,\tag{4}$$

where $\kappa(\alpha)=(1+\alpha\|A\|^{-2})$, $C_1=0.3$, $C_2=0.2$. In this paper we choose α_{R2} as the largest solution of (4), which is better in case of inexactly given noise level.

Estimated parameters. Sometimes it is known either theoretically or practically that a parameter choice rule typically chooses a too large or too small parameter. Then it is reasonable to post-estimate the computed parameter, by decreasing or increasing it. We found the values of various constants in post-estimation by optimization on the test problems [20] with normalized operator (solving the equation $\|A\|^{-1}Ax=\|A\|^{-1}y_*$). We emphasize that these constants occurred to be good also for other test problems [7].

In (iterated) Tikhonov method we always have $\alpha_{\mathrm{ME}}\geq\alpha_{\mathrm{opt}}=\mathrm{argmin}\{\|x_\alpha-x_*\|,\ \alpha>0\}$. Our numerical experiments suggested to use the estimated parameter $\alpha_{\mathrm{MEe}}=\min(c_1\alpha_{\mathrm{ME}},c_2\alpha_{\mathrm{ME}}^{c_3})$ or $\alpha_{\mathrm{MEe'}}=c_4\alpha_{\mathrm{ME}}$ instead of α_{ME}, where $c_1=0.53$, $c_2=0.6$, $c_3=1.06$, $c_4=0.44$ (here the letter "e" refers to estimation). Our numerical tests showed that in most cases (especially for smaller α's) also $\alpha_{\mathrm{R2}}\geq\alpha_{\mathrm{opt}}$. The estimated parameter $\alpha_{\mathrm{R2e}}=0.5\alpha_{\mathrm{R2}}$ is usually better than α_{R2}.

By comparing the parameters α_{MEe} and α_{R2e} numerically, we found that typically α_{MEe} is the better of the two, when $\delta/\|y-y_*\|\in[1,1.05]$ but α_{R2e} is better, when $\delta/\|y-y_*\|>1.05$. In both cases $\alpha_{\mathrm{Me}}=\min(\alpha_{\mathrm{MEe}},\alpha_{\mathrm{R2e}})$ or $\alpha_{\mathrm{Me'}}=\min(\alpha_{\mathrm{MEe'}},\alpha_{\mathrm{R2e}})$ typically chooses the best of these parameters.

In Landweber method n_{D} and n_{ME} are close: $n_{\mathrm{D}}-1\leq n_{\mathrm{ME}}\leq n_{\mathrm{D}}$ (see [16]), so n_{ME} and typically also n_{D} are smaller than the optimal stopping index n_*. Therefore it makes sense to use the estimated indices $n_{\mathrm{De}}=\mathrm{round}(cn_{\mathrm{D}})$ and $n_{\mathrm{MEe}}=\mathrm{round}(cn_{\mathrm{ME}})$ instead of n_{D} and n_{ME}, respectively, with $c>1$. Computations suggested the value $c=2.3$.

For shifted parameters $c\alpha_{\mathrm{ME}}$, $c\alpha_{\mathrm{R2}}$, cn_{ME}, cn_{D} with fixed $c>0$ the same convergence and convergence rate results hold as for parameters α_{ME}, α_{R2}, n_{ME}, n_{D} (see Theorems 1, 2). Theoretically the parameter $\alpha_{\mathrm{MEe'}}$ is more justified than α_{MEe} but α_{MEe} worked slightly better in tests.

3.2 Parameter choice rules for unknown noise level

If the noise level is unknown, then, as shown by Bakushinskii [1], no rule for choosing the regularization parameter can guarantee the convergence of the regularized solution to the exact one as $\delta \to 0$. Nevertheless, some heuristic rules that don't use δ are rather popular, because they often work well in practice and because in applied ill-posed problems the exact noise level is often unknown.

A classical heuristic rule is the *quasioptimality criterion*. In m-iterated Tikhonov method it chooses $\alpha = \alpha_Q$ as the global minimizer of the function $\varphi_Q(\alpha) = \|x_{m;\alpha} - x_{m+1;\alpha}\|$. Neubauer [24] proposed to minimize the function $\varphi_{QN}(\alpha) = \|x_{m;\alpha} - x_{2m;\alpha}\|$ for $\alpha \in [m\sigma_{\min}, 1]$, where $\sigma_{\min}$ is the smallest eigenvalue of discretized version of the operator A^*A (we assume $\|A\| = 1$), in m-iterated Tikhonov method, and the function $\{\|x_n - x_{2n}\|, n \geq 1\}$ in the Landweber method.

The *Hanke-Raus rule* [18] in m-iterated Tikhonov method finds the regularization parameter $\alpha = \alpha_{HR}$ as the global minimizer of the function $\varphi_{HR}(\alpha) = (r_{m;\alpha}, r_{m+1;\alpha})^{1/2}/\sqrt{\alpha}$. In Landweber method the Hanke-Raus rule minimizes $\varphi_{HR}(n) = n^{1/2}\|r_n\|$ for $n \geq 1$.

The paper [7] by Brezinski, Rodriguez, and Seatzu proposed to minimize the function $\varphi_{BRS}(n) = \|r_n\|^2/\|A^*r_n\|$ in iteration methods and the function $\varphi_{BRS}(\alpha) = \|r_\alpha\|^2/(\alpha\|x_\alpha\|)$ in Tikhonov method. In GCV-rule [33] and L-curve rule [19, 20] functions of other types are minimized of maximized. Our numerical experiments of Section 6 do not contain the results for L-curve rule and GCV-rule, since the rules presented in tables gave essentially better results.

We combined the functions $\varphi_{HR}(\alpha)$ and $\varphi_{R2}(\alpha)$ and found $\alpha = \alpha_{HR2}$ and $\alpha = \alpha_{BR2}$ as the global argmins of

$$\varphi_{HR2,\tau}(\alpha) = \varphi_{R2}(\alpha)^{(\varphi_{R2}(\alpha)/\varphi_{HR}(\alpha))^\tau}\varphi_{HR}(\alpha)^{1-(\varphi_{R2}(\alpha)/\varphi_{HR}(\alpha))^\tau} \tag{5}$$

and $\varphi_{BR2,\tau}(\alpha)$, respectively, where $\varphi_{R2}(\alpha) = d_{R2}(\alpha)/\sqrt{\alpha}$ and $\varphi_{BR2,\tau}(\alpha)$ is derived from (5) by substituting $\varphi_{HR}(\alpha)$ with $\varphi_{BRS}(\alpha)$. Here $\tau \in (0, 1)$.

To consider different regularization methods in parallel, in the following we use the common parameter $\lambda = 1/n$ for iteration methods and $\lambda = \alpha$ for (iterated) Tikhonov method.

The abovementioned heuristic rules take the global minimizer of a certain function $\varphi(\lambda)$ to be the regularization parameter. Heuristic rules often give good results but sometimes fail, especially if the global minimizer of function $\varphi(\lambda)$ is a very small λ with very large error. Our proposal is to use a local minimizer at larger λ instead of the global minimizer, if there is a large maximum between them.

Based on numerical evidence, we propose the following strategies [12] to stop the computations. We make computations for decreasing sequence of λ's, starting from a certain initial value (usually $\lambda = 1$).

1. Climbing approach. Stop the computation at the point, where the value of the function $\varphi(\lambda)$ has become C or more times larger than its currently found minimal value. Take λ at which the function has minimal value as the regularization parameter. Suitable values of C for functions that we used are around 3 in Tikhonov method, and 20–50 in Landweber method.

2. First local minimum. Stop the computation at the first local minimum of a certain function $\psi(\lambda)$. We often used $\psi(\lambda)=\varphi(\lambda)\lambda^c$ with $c\approx 1/3$.

Note also that in iteration methods the original heuristic rules HR, QN, BRS often stopped at too small n. In our modifications HRmC, QNmC, BRSC the minimum was shifted towards larger n.

3.3 Parameter choice rules for approximate noise level

In case of approximate noise level we follow a two-step strategy. The corresponding rule is denoted by DM, where D refers to the first step (that uses δ), and M refers to the minimization on the second step.

Rule DM for m-iterated Tikhonov method. 1) Find $\underline{\alpha}$ as the maximal solution of $\alpha^{1/2}\|x_{m;\alpha}-x_{m+1;\alpha}\|=c_1\delta$ with $c_1>(2m+2)^{-m/(2m+2)}\,((2m+1)/(2m+2))^{m+1/2}$. 2) Find $\alpha=\alpha_{\mathrm{DM}}$ as the minimizer of $\varphi_{\mathrm{R2}}(\alpha)\alpha^{c_2}$ on $[\underline{\alpha},1]$, where $\varphi_{\mathrm{R2}}(\alpha)=d_{\mathrm{R2}}(\alpha)/\sqrt{\alpha}$, $c_2\in(0,1/2)$. This is a slight modification of the rule DM^0 [14, 15] (see Theorem 5), where the function $\varphi_{\mathrm{HR}}(\alpha)\alpha^{c_2}$ is minimized on the second step.

In our computations the best parameters were produced by small constants $c_1\in[0.001,0.02]$ and $c_2\in[0.03,0.14]$.

Rule DM for Landweber method [14, 15]. 1) Find N as the first n for which $n^{1/2}\|A^*r_n\|\le c_1\delta$ with $c_1>1/\sqrt{2\mu e}$. 2) Find $n=n_{\mathrm{DM}}$ as the minimizer of $n^{c_2}\|r_n\|$ on $[1,N]$, where $c_2\in(0,1/2)$. In this paper we obtain better results by minimizing $n^{c_2}(\|x_n-x_{2n+100}\|)$ in the second step.

4 Convergence and convergence rate

If the noise level of the data is known exactly or approximately, then the following theoretical results are known for the parameter choice rules considered here.

Theorem 1 [27, 31, 32] *Let $y\in\mathcal{R}(A)$, $\|y-y_*\|\le\delta$. If in Landweber approximation or in m-iterated Tikhonov approximation the regularization parameter $\lambda=\lambda(\delta)$ is chosen by one of the rules D, ME, then $\|x_{\lambda(\delta)}-x_*\|\to 0$ as $\delta\to 0$ and for source-like solutions*

$$x_*-\overline{x}=\mathcal{R}\big((A^*A)^{p/2}\big), \tag{6}$$

the error estimate

$$\|x_\lambda-x_*\|\le \mathrm{const}\,\delta^{p/(p+1)} \tag{7}$$

holds true in Landweber method for all $p>0$ and in m-iterated Tikhonov method for $p\le 2m-1$ in case of rule D or for $p\le 2m$ in case of rule ME.

Theorem 2 [26] *Let $A^*y\in\mathcal{R}(A^*A)$, $\|y-y_*\|\le\delta$. If the regularization parameter $\alpha=\alpha(\delta)$ in m-iterated Tikhonov method is chosen by the rule R2, then $\|x_{\alpha(\delta)}-x_*\|\to 0$ as $\delta\to 0$ and for source-like solutions (6) the error estimate (7) holds true with $p\le 2m-1$.*

Theorem 3 [11] *Let* $y \in \mathcal{R}(A)$, $\|y - y_*\| \le \delta$. *Let* $x_m = x_{\alpha_1,\ldots,\alpha_m}$ *be an extrapolated Tikhonov approximation. If* $m = m_D$ *is the first number with* $d_D(m) = \|Ax_m - y\| \le C\delta$, *then* $\|x_m - x_*\| < \|x_{m-1} - x_*\|$ $(m = 1, 2, \ldots, m_D - 1)$. *For* $m = m_D$ $\|x_m - x_*\| \to 0$ *as* $\delta \to 0$ *and for source-like solutions* (6) *for* $\lambda = m = m_D$ *the error estimate* (7) *holds for all* $p > 0$.

Theorem 4 [11] *Let* $y \in \mathcal{R}(A)$, $\|y - y_*\| \le \delta$. *Let* $x_\alpha = x_{\alpha_1,\ldots,\alpha_m}$ *be an extrapolated Tikhonov approximation with* $\alpha_n = q_n\alpha$; m *and* q_n *fixed* $(n = 1, \ldots, m + 1)$. *Let* $C > 1$. *If* α *is chosen as the solution of* $(Ax_{\alpha_1,\ldots,\alpha_m} - y, Ax_{\alpha_1,\ldots,\alpha_{m+1}} - y)^{1/2} = C\delta$, *then* $\|x_\alpha - x_*\| \to 0$ *as* $\delta \to 0$ *and for source-like solutions* (6) *the error estimate* (7) *with* $\lambda = \alpha$ *holds for all* $p \le 2m$.

Theorem 5 [14, 15] *Let the parameter* $\alpha = \alpha(\delta)$ *in (iterated) Tikhonov method be chosen by the rule* DM^0. *If* $\|y - y_*\|/\delta \le const$ *as* $\delta \to 0$, *then the rule* DM^0 *guarantees the convergence* $\|x_{\alpha(\delta)} - x_*\| \to 0$ *as* $\delta \to 0$ *and the following error estimates hold true.* 1) *If* $\|y - y_*\| \le \max\{\delta, \delta_0\}$, *where* $\delta_0 = \frac{1}{2}(r_{m;\alpha(\delta)}, r_{m+1;\alpha(\delta)})^{1/2}$, *then*

$$\|x_\alpha - x_*\| \le C \inf_{\alpha \ge 0}\left(\|x_\alpha^0 - x_*\| + \sqrt{m}\,\alpha^{-1/2}\delta\right) \tag{8}$$

holds with $C = 1/(1 - 2c_2)$, *where* x_α^0 *is an (iterated) Tikhonov approximation with* y_* *instead of* y. 2) *If* $\max(\delta, \delta_0) < \|y - y_*\| \le \frac{1}{2}(r_{m;1}, r_{m+1;1})^{1/2}$, *then* (8) *holds with* $C = const(\|y - y_*\|/\delta_0)^{1/(2c_2)}$.

5 Test problems

Our tests are performed on the well-known set of test problems by Hansen [20]: *baart*, *deriv2*, *foxgood*, *gravity*, *heat*, *ilaplace*, *phillips*, *shaw*, *spikes*, *wing* (problems 1–10). In all tests we used discretization parameter 100.

Since the performance of methods and rules generally depends on the smoothness p of exact solution in (6), we complemented the standard solutions x_* of (now discrete) test problems with smoothened solutions $|A|^p x_*$, $p = 2$ (computing the right-hand side as $A(|A|^p x_*)$). After discretization all problems were scaled (normalized) in such a way that the Euclidean norms of the operator and the right-hand side were 1. On base of exact data y_* we formed the noisy data y, where $\|y - y_*\|$ had values $0.5, 10^{-1}, \ldots, 10^{-6}$. In most cases the noise $y - y_*$ added to y_* had uniform distribution, where components of the noise were uncorrelated. Besides this we used correlated noise, where the components of noise vector had nonzero correlation. As in [4], the amount of correlation was determined by randomly choosing the parameter $\omega \in [-0.5, 0.5]$, where negative and positive ω correspond to noise that has dominantly higher or lower frequencies in the frequency domain. Results for correlated noise are given only in Table 2, since in iteration methods and in rules that do not use δ, the correlation influenced the results only a little. We also made computations with normally distributed noise, the results were similar.

We generated 10 noise vectors and used these vectors in all problems. The problems were regularized, using different methods, choosing the regularization parameters by the rules that we wanted to compare. In our experiments we also took into account the possibility of over- or underestimation of the noise level: we used $\delta = d\|y - y_*\|$ with $d \in [1/64, 64]$. Thus $d = 1$ corresponds to the exact noise level.

Since in model equations the exact solution is known, it is possible to find the regularization parameter $\lambda = \lambda_*$, which gives the smallest error: $\|x_{\lambda_*} - x_*\| = \min_{\lambda>0}\{\|x_\lambda - x_*\|\}$. For every rule R the error ratio $\|x_{\lambda_R} - x_*\|/\|x_{\lambda_*} - x_*\|$ describes the performance of the rule R on this particular problem. To compare the rules or to present their properties, the following tables show averages A and medians M of these error ratios over various parameters of the data set (problems 1–10, smoothness indices p, noise levels δ, runs).

6 Numerical results

Comparing the methods, we found that in Landweber method (we used $\mu = 1$) the minimal errors were slightly smaller in case $p = 0$ and 3–5 times smaller in case $p = 2$ than in Tikhonov method.

In tables below we compare different parameter choice rules. If the noise level is known, we use the rules T1–T4 below in Tikhonov method and rules T5, T6 in extrapolated Tikhonov method.

(T1) Discrepancy principle: α_D is the solution of $\|r_{m;\alpha}\| = \delta$.
(T2) Rule MEe: take $\alpha_{MEe} = \min(0.53\alpha_{ME}, 0.6\alpha_{ME}^{1.06})$.
(T3) Rule R2e: take $\alpha_{R2e} = 0.5\alpha_{R2}$.
(T4) Rule Me: using T2, T3, take $\alpha_{Me} = \min(\alpha_{MEe}, \alpha_{R2e})$.
(T5) Rule 2Me is the rule from T4, applied to the extrapolated Tikhonov method with $m = 2$.
(T6) Rule maxDe. Let $m = m_D$ be the first number with $d_D(m) = \|Ax_{\alpha_1,\ldots,\alpha_m} - y\| \le \delta$. Take $m_{maxDe} = \text{round}(1.1m_D)$.

Tables 1, 2, 3, 5 contain the averages and medians of error ratios for Tikhonov method, Tables 2, 4, 5 for extrapolated Tikhonov method. For better comparison of these tables also in Tables 4, 5 and in columns 2Me, maxDe of Table 2 the denominators of the error ratios are the errors of the best single Tikhonov approximations.

Table 1 confirms the disadvantages of the discrepancy principle: saturation for $p \ge 1$ (see large error ratios for $p = 2$) and sensitivity to inexact noise level (discrepancy principle totally fails in case of underestimated noise level; results for 1.05–2 times overestimated noise level are much larger than for $d = 1$). The rule R2e is much more accurate than the discrepancy principle in smooth case ($p \ge 1$) and/or in case

Table 1 Results in Tikhonov method for discrepancy principle (upper part) and rule R2e (lower part) for various d; $p = 0$ (left), $p = 2$ (right)

d	0.6	0.8	0.9	1	1.05	1.1	1.3	2	0.6	0.8	0.9	1	1.05	1.1	1.3	2
A	$\gg 1$			1.19	1.44	1.56	1.83	2.25	$\gg 1$			2.77	2.29	2.25	2.39	2.79
M				1.05	1.17	1.25	1.42	1.67				1.87	1.60	1.65	1.76	1.91
A	1.75	1.36	1.37	1.39	1.40	1.42	1.47	1.60	1.50	1.13	1.11	1.11	1.10	1.10	1.11	1.19
M	1.09	1.08	1.09	1.10	1.10	1.11	1.15	1.28	1.11	1.07	1.06	1.05	1.05	1.04	1.04	1.04

Table 2 Results in Tikhonov method in case of uncorrelated noise (upper part) and correlated noise (lower part); $p=0$ (left), $p=2$ (right)

	MEe		Me		2Me		maxDe		MEe		Me		2Me		maxDe	
d	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
A	1.15	2.02	1.16	1.59	1.24	1.51	1.17	2.25	1.11	2.82	1.12	1.19	0.68	0.67	0.52	2.67
M	1.03	1.54	1.03	1.28	1.03	1.17	1.03	1.66	1.07	2.25	1.07	1.04	0.66	0.62	0.48	1.95
A	1.18	2.18	1.18	1.70	1.28	1.61	1.19	2.42	1.11	4.70	1.14	1.59	0.73	1.01	0.50	4.37
M	1.03	1.56	1.04	1.29	1.04	1.21	1.04	1.75	1.05	2.51	1.09	1.08	0.66	0.71	0.44	2.16

of overestimated noise level. In contrast to the discrepancy principle and other rules, the rule R2e also allows moderate underestimation of the noise level. In case of exact noise level ($d=1$) the discrepancy principle was better than R2e but as Table 2 shows, MEe is even better. The rule Me is good in both cases $d=1$, $d=2$. Note that the errors in rules R2 and ME are typically 10–15% larger than in post-estimated versions R2e and MEe respectively; in rules MEe' and Me' (see page 50) they are about 1–2% larger than in rules MEe and Me. Note also that in rule Me the dependence of the error on the precision of noise level information is much less significant in case of smooth solution ($p=2$). If $p=2$, then the rule 2Me in 2-extrapolated Tikhonov approximation gave the average that is typically better than the average of the best single Tikhonov approximation for both exact noise level and 2 times overestimated noise level. The nonstationary iterated Tikhonov method with the rule maxDe was better than 2Me in case of exact noise level but essentially worse in case of overestimated noise level.

The lower part of Table 2 is the analog of its upper part for correlated noise of data (the correlation parameter ω is chosen randomly from $[-0.5, 0.5]$ with uniform distribution). The medians in the lower part are similar to the medians in the upper part but the averages are 3–6% larger in case $p=0$ and 10–60% larger in case $p=2$ (an exception is the rule MEe).

Heuristic rules T7–T11 for Tikhonov method and its extrapolated version are described below.

- (T7) Rules HR and BRS: α_{HR} and α_{BRS} are the global minimizers of the functions $\varphi_{\mathrm{HR}}(\alpha)$ and $\varphi_{\mathrm{BRS}}(\alpha)$, respectively.
- (T8) Rule QN: $\alpha_{\mathrm{QN}} = \operatorname{argmin}\{\varphi_{\mathrm{QN}}(\alpha),\ \alpha \in [m\sigma_{\min}, 1]\}$.
- (T9) The rules R2C, QC, BRSC choose the parameter by the climbing approach in the functions $\varphi_{\mathrm{R2}}(\alpha)$, $\varphi_{\mathrm{Q}}(\alpha)$, and $\varphi_{\mathrm{BRS}}(\alpha)$ with $C=3.1$, $C=2.7$, and $C=3$, respectively.
- (T10) Rules Q1, DR21 and BRS1 choose the regularization parameter as the first (i.e. the largest) local minimum of the functions $\varphi_{\mathrm{Q}}(\alpha)\alpha^{1/3}$, $\varphi_{\mathrm{D}}(\alpha)^{1/4}\varphi_{\mathrm{R2}}(\alpha)^{3/4}\alpha^{1/3}$ ($\varphi_{\mathrm{D}}(\alpha)^{0.9}\varphi_{\mathrm{R2}}(\alpha)^{0.1}\alpha^{0.4}$ in extrapolated Tikhonov method), and $\varphi_{\mathrm{BRS}}(\alpha)\alpha^{0.56}$ ($\varphi_{\mathrm{BRS}}(\alpha)\alpha^{0.66}$ in extrapolated Tikhonov method). Here, $\varphi_{\mathrm{D}}(\alpha) = \|r_{m,\alpha}\|/\sqrt{\alpha}\|$.
- (T11) Rules HR2 and BR2 choose the parameters as the global minimizers of the functions $\varphi_{\mathrm{HR2},\tau}(\alpha)$ with $\tau = 0.06$ and $\varphi_{\mathrm{BR2},\tau}(\alpha)$ with $\tau=0.05$, respectively.

Table 3 Results in Tikhonov method for heuristic rules, $p = 0$

	HR	QN	QC	R2C	BRS	BRSC	Q1	DR21	BRS1	HR2	BR2
A	1e+4	1e+3	1.39	1.38	1e+4	2.10	1.49	1.44	1.67	1.49	1.46
M	2.21	1.13	1.08	1.09	1.84	1.59	1.12	1.10	1.30	1.12	1.11

Table 4 Results in 2-extrapolated Tikhonov approximation

	$p=0$					$p=2$				
	QC	R2C	BRSC	DR21	BRS1	QC	R2C	BRSC	DR21	BRS1
A	1.47	1.46	2.34	1.61	1.81	0.79	0.85	1.54	0.68	0.90
M	1.09	1.08	1.67	1.14	1.35	0.65	0.69	1.11	0.62	0.71

Table 5 Results in Tikhonov method (upper part) and in 2-extrapolated Tikhonov method (lower part) for Rule DM, $c_1 = 0.001$, $c_2 = 0.03$

	$p=0$							$p=2$						
d	1/64	1/16	1/4	1	4	16	64	1/64	1/16	1/4	1	4	16	64
A	1.42	1.37	1.34	1.32	1.29	1.29	1.48	1.49	1.49	1.49	1.49	1.41	1.28	1.31
M	1.07	1.07	1.07	1.06	1.06	1.07	1.11	1.20	1.20	1.20	1.20	1.19	1.18	1.16
A	3.41	1.96	1.82	1.41	1.41	1.36	1.54	2.96	0.88	0.88	0.88	0.83	0.74	0.71
M	1.07	1.07	1.07	1.07	1.07	1.08	1.11	0.71	0.70	0.70	0.70	0.68	0.66	0.64

Table 4 shows that heuristic rules in extrapolated Tikhonov approximations gave good results and these approximations were more accurate than best single Tikhonov approximation in case $p = 2$.

If we compare the results of the rule DM in Tikhonov method in Table 5 with the results of the best heuristic rules QC, R2C in Table 3, we may conclude that taking into account the information about approximate noise level improves results, if $d \in [1/16, 16]$.

The stopping index in Landweber method in case of known noise level was chosen by the rules L1, L2 below.

(L1) Discrepancy principle: n_D is the first $n \geq 1$ for which $d_D(n) \leq \delta$.
(L2) Rule De: find n_D from L1 and take $n_{De} = \text{round}(2.3n_D)$.

Table 6 shows that in Landweber method the error typically continued to decrease monotonically long after the discrepancy reached the level δ: an about twice larger index gave smaller errors. As the right part of the table shows, continuing iterations after the discrepancy reaches the level δ is especially essential in case of smooth solution and/or if the noise level is overestimated.

The following heuristic rules L3–L8 stop Landweber iterations without using the noise level.

(L3) Hanke-Raus rule: $n_{HR} = \text{argmin}\{\sqrt{n}\,\|r_n\|, n \geq 1\}$.
(L4) Rule HRmC: n_{HRmC} chooses $n \geq 1$ in the function $\sqrt{n}(\|r_n\| - \|r_{2n+100}\|)$ by the climbing approach with $C = 50$.
(L5) Rule QN: $n_{QN} = \text{argmin}\{\|x_n - x_{2n}\|, n \geq 1\}$.

Table 6 Results in Landweber method for $p=0$ (left), $p=2$ (right)

	D			De			D			De		
d	1	1.1	2	1	1.1	2	1	1.1	2	1	1.1	2
A	1.36	1.85	2.50	1.22	1.58	2.06	1.73	5.31	13.1	1.74	2.64	4.67
M	1.20	1.45	1.98	1.05	1.21	1.62	1.33	2.48	5.42	1.14	1.14	2.10

Table 7 Results in Landweber method for $p=0$ (left), $p=2$ (right)

	HR	HRmC	QN	QNmC	BRS	BRSC	HR	HRmC	QN	QNmC	BRS	BRSC
A	2e+4	2.41	2e+4	1.77	2e+4	1.75	5e+4	6.86	6e+4	3.98	5e+4	3.43
M	2.17	1.11	1.55	1.12	2.26	1.17	4.47	1.31	1.96	1.20	6.68	1.20

Table 8 Results in Landweber method for Rule DM, $c_1=0.1$, $c_2=0.001$

	Case $p=0$								Case $p=2$							
d	1/27	1/9	1/3	1	2	3	9	27	1/27	1/9	1/3	1	2	3	9	27
A	5.35	4.88	1.63	1.63	1.74	1.84	2.39	3.06	41.2	40.4	2.45	2.49	3.11	3.41	10.3	17.4
M	1.14	1.14	1.12	1.12	1.24	1.32	1.68	2.20	1.33	1.26	1.18	1.12	1.16	1.27	2.08	4.00

(L6) Rule QNmC (modification of rule QN): n_{QNmC} chooses $n \geq 1$ in the function $\|x_n - x_{2n+100}\|$ by the climbing approach with $C=20$.
(L7) Rule BRS: n_{BRS} is the global minimizer of $\varphi_{\mathrm{BRS}}(n)$.
(L8) Rule BRSC: n_{BRSC} chooses $n \geq 1$ in function $\|r_n\| \cdot (\|r_n\| - \|r_{2n+100}\|)/\|A^* r_n\|$ by the climbing approach with $C=15$.

From Tables 3, 7 we see that the modifications of known heuristic rules are essentially better than the original rules. Note that the medians of error ratios of modified heuristic rules in Table 7 are better than the medians of the discrepancy principle at exact noise level in Table 6. Heuristic rules HR, QN, and BRS in Landweber method, and HR and BRS in Tikhonov method failed in weakly ill-posed problems *deriv2* and *phillips*. In Tikhonov method the quasioptimality criterion failed in severely ill-posed problem *heat* but gave reasonably good results in problems *deriv2* and *phillips*.

Comparing the results of Landweber method for the rule DM in Table 8 with the results of best heuristic rules QNmC, BRSC in Table 7, we see that the rule DM is better, if $d \in [1/3, 2]$.

The best heuristic rules (not using the noise level) performed only slightly worse than non-heuristic rules at the exact noise level but better than the same rules at 1.1 times (in Tikhonov method 1.05 times) overestimated noise level. The Rule DM that uses the approximate knowledge about the noise level often gave better results in case of d_1 times overestimated and underestimated noise level than the discrepancy principle in case of essentially smaller d_2 times overestimated noise level—this holds in Tikhonov method for $d_1=50$, $d_2=1.05$ (for $d=50$, average and median error ratios were 1.43 and 1.09 in DM rule) and in Landweber method for $d_1=3$, $d_2=1.1$.

We made numerical experiments of Tables 2, 3, 4 also for $p=1$, then the extrapolated Tikhonov approximation with a posteriori parameter choice typically gave better results than the best single Tikhonov approximation.

Similar computations were made on problems of [7] and the results were similar, except the problem *pascal*, where the error ratios were essentially smaller for $p = 0$ and essentially larger for $p = 2$.

Much more numerical examples can be found in [25], where also conjugate gradient type methods and the truncated singular value decomposition method are included.

The conclusion of this work are the following recommendations. If there is some reason to assume that the solution is smooth, then we strongly recommend to use the extrapolated Tikhonov method instead of Tikhonov method. The choice of the regularization parameter depends on the information about the noise level $\|y - y_*\|$. If there is some guess δ of noise level $\|y - y_*\|$, then the value of $d = \delta/\|y - y_*\|$ is essential. We recommend to use the rule Me, if we are sure that $d \in [1, 1.3]$ (or the smoothness index $p \geq 2$ and $d \in [1, 3]$), and the rule DM in case of more approximate information $d \in [1/16, 16]$. Otherwise we recommend to use the rules QC or R2C but if $\|r_{m;\alpha_{\mathrm{QC}}}\|$ or $\|r_{m;\alpha_{\mathrm{R2C}}}\|$ is evidently less than the noise level, we recommend to reduce the constant C in the climbing approach (for example, replacing C by $(C + 1)/2$). In Landweber method we recommend the rule De in case $d \in [1, 1.1]$ and the rule DM in case $d \in [1/3, 2]$. Otherwise the rules QNmC, BRSC can be recommended but if $\|r_{n_{\mathrm{QNmC}}}\|$ or $\|r_{n_{\mathrm{BRSC}}}\|$ is evidently less than the noise level, reducing the constant C is advisable.

In the future we plan to consider wider class of test problems and other noise distributions.

References

1. Bakushinskii, A.B.: Remarks on choosing a regularization parameter using the quasi-optimality and ratio criterion. Comp. Math. Math. Phys. **24**(4), 181–182 (1984)
2. Bauer, F., Kindermann, S.: The quasi-optimality criterion for classical inverse problems. Inverse Probl. **24**, 035002 (2008)
3. Bauer, F., Kindermann, S.: Recent results on the quasi-optimality principle. J. Inverse Ill-Posed Probl. **17**(1), 5–18 (2009)
4. Bauer, F., Lukas, M.A.: Comparing parameter choice methods for regularization of ill-posed problems. Preprint, http://bmath.de/Docs/mainLowRes.pdf (2010)
5. Brezinski, C., Redivo-Zaglia, M., Rodriguez, G., Seatzu, S.: Extrapolation techniques for ill-conditioned linear systems. Numer. Math. **81**(1), 1–29 (1998)
6. Brezinski, C., Redivo-Zaglia, M., Rodriguez, G., Seatzu, S.: Multi-parameter regularization techniques for ill-conditioned linear systems. Numer. Math. **94**(2), 203–228 (2003)
7. Brezinski, C., Rodriguez, G., Seatzu, S.: Error estimates for linear systems with applications to regularization. Numer. Algorithms **49**(14), 85–104 (2008)
8. Engl, H.W., Hanke, M., Neubauer, A.: Regularization of Inverse Problems. Mathematics and Its Applications, vol. 375. Dordrecht, Kluwer (1996)
9. Glasko, V., Kriksin, Y.: On the quasioptimality principle for linear ill-posed problems in Hilbert space. Zh. Vychisl. Mat. Mat. Fiz. **24**(11), 1603–1613 (1984) In Russian
10. Hämarik, U., Palm, R.: On rules for stopping the conjugate gradient type methods in ill-posed problems. Math. Model. Anal. **12**(1), 61–70 (2007)
11. Hämarik, U., Palm, R., Raus, T.: Use of extrapolation in regularization methods. J. Inverse Ill-Posed Probl. **15**(3), 277–294 (2007)
12. Hämarik, U., Palm, R., Raus, T.: On minimization strategies for choice of the regularization parameter in ill-posed problems. Numer. Funct. Anal. Optim. **30**(9–10), 924–950 (2009)
13. Hämarik, U., Palm, R., Raus, T.: Extrapolation of Tikhonov regularization method. Math. Model. Anal. **15**(1), 55–68 (2010)

14. Hämarik, U., Raus, T.: Choice of the regularization parameter in ill-posed problems with rough estimate of the noise level of data. WSEAS Trans. Math. **4**(2), 76–81 (2005)
15. Hämarik, U., Raus, T.: 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 (2006)
16. Hämarik, U., Tautenhahn, U.: On the monotone error rule for parameter choice in iterative and continuous regularization methods. BIT Numer. Math. **41**(5), 1029–1038 (2001)
17. Hanke, M., Groetsch, C.W.: Nonstationary iterated Tikhonov regularization. J. Optim. Theory Appl. **98**(1), 37–53 (1998)
18. Hanke, M., Raus, T.: A general heuristic for choosing the regularization parameter in ill-posed problems. SIAM J. Sci. Comput. **17**(4), 956–972 (1996)
19. Hansen, P.C.: Analysis of discrete ill-posed problems by means of the L-curve. SIAM Rev. **34**(4), 561–580 (1992)
20. Hansen, P.C.: Regularization tools: A Matlab package for analysis and solution of discrete ill-posed problems. Numer. Algorithms **6**(1), 1–35 (1994)
21. Jin, B., Lorenz, D.A.: Heuristic parameter-choice rules for convex variational regularization based on error estimates. SIAM J. Numer. Anal. **48**(3), 1208–1229 (2010)
22. Kindermann, S., Neubauer, A.: On the convergence of the quasioptimality criterion for (iterated) Tikhonov regularization. Inverse Probl. Imaging **2**(2), 291–299 (2008)
23. Morozov, V.A.: On the solution of functional equations by the method of regularization. Sov. Math. Dokl. **7**, 414–417 (1966)
24. Neubauer, A.: The convergence of a new heuristic parameter selection criterion for general regularization methods. Inverse Probl. **24**, 055005 (2008)
25. Palm, R.: Numerical comparison of regularization algorithms for solving ill-posed problems. PhD thesis, University of Tartu (2010). http://hdl.handle.net/10062/14623
26. Raus, T., Hämarik, U.: New rule for choice of the regularization parameter in (iterated) Tikhonov method. Math. Model. Anal. **14**(2), 187–198 (2009)
27. Tautenhahn, U., Hämarik, U.: The use of monotonicity for choosing the regularization parameter in ill-posed problems. Inverse Probl. **15**(6), 1487–1505 (1999)
28. Tikhonov, A.N., Arsenin, V.Y.: Solutions of Ill-Posed Problems. Wiley, New York/Toronto (1977)
29. Tikhonov, A.N., Glasko, V.B., Kriksin, Y.: On the question of quasioptimal choice of a regularized approximation. Sov. Math. Dokl. **20**, 1036–1040 (1979)
30. Vainikko, G.: Methods for Solution of Linear Ill-Posed Problems in Hilbert Spaces. University of Tartu (1982). In Russian
31. Vainikko, G.M.: The discrepancy principle for a class of regularization methods. USSR Comput. Math. Math. Phys. **22**(3), 1–19 (1982)
32. Vainikko, G.M., Veretennikov, A.Yu.: Iteration Procedures in Ill-Posed Problems. Nauka, Moscow (1986). In Russian
33. Wahba, G.: Practical approximate solutions to linear operator equations when the data are noisy. SIAM J. Numer. Anal. **14**(4), 651–657 (1977)