DISSERTATIONES MATHEMATICAE UNIVERSITATIS TARTUENSIS
59

DISSERTATIONES MATHEMATICAE UNIVERSITATIS TARTUENSIS
59

REIMO PALM

Numerical Comparison of Regularization Algorithms for Solving Ill-Posed Problems

TARTU UNIVERSITY
PRESS

Institute of Computer Science, Faculty of Mathematics and Computer Science, University of Tartu, Estonia.

Dissertation is accepted for the commencement of the degree of Doctor of Philosophy (PhD) on January 7, 2010 by the Council of the Institute of Computer Science, University of Tartu.

Supervisor:

Senior Lecturer, Cand. Phys. Math. Sc.	Uno Hämarik University of Tartu Tartu, Estonia

Opponents:

Prof. Dr.	Martin Hanke-Bourgeois Johannes Gutenberg Universität Mainz Institut für Mathematik Mainz, Germany
Prof., Cand. Phys. Math. Sc.	Jaan Janno Tallinn University of Technology Tallinn, Estonia

The public defense will take place on February 26, 2010 at 15:15 in Liivi 2-405.

The publication of this dissertation was financed by the Institute of Computer Science, University of Tartu.

ISSN 1024-4212
ISBN 978-9949-19-293-9 (trükis)
ISBN 978-9949-19-294-6 (PDF)

Tartu Ülikooli Kirjastus
http://www.tyk.ee/
Tellimus nr 11

CONTENTS

I. INTRODUCTION

Many problems in nature, science and engineering (for example, in signal and image processing, tomography, heat conduction, geophysics, inverse scattering) are ill-posed in the sense that small errors in data can cause a large error in the solution. Modelling of these problems often leads to integral equations of the first kind; also systems of linear equations, where the condition number of the matrix is very large, can be considered as ill-posed problems. In applications the data are typically noisy, containing unavoidable measurement errors. When dealing with well-posed problems, where the solution depends continuously on the data, the solving algorithms do not need to take into account the noise level of data. In ill-posed problems, however, it is crucial to take into account all available information about the solution and the noise. For solving ill-posed problems special methods, regularization methods [19, 47, 53, 89] have been developed. These methods contain a free parameter, which controls the amount of regularization. When applying regularization, one has to make two decisions: 1) select the regularization method and 2) choose a proper regularization parameter.

In this work we consider basic regularization methods: the methods of Tikhonov and Lavrentiev, their iterated variants, Landweber method, truncated singular value decomposition (TSVD) and conjugate gradient type methods (CG, CGLS, CGME) [19, 37]. The regularization parameter is the stopping index in iteration methods, the number of terms in TSVD, and a positive real number in remaining methods.

For theoretical study it is typically assumed that the exact noise level δ is known with $\|y - y_*\| \le \delta$, where y are given (measured) data and y_* are exact data. Then one can show that if the regularization parameter is chosen properly, in dependence of the noise level δ, then the regularized solution converges to the exact solution as $\delta \to 0$. This convergence is guaranteed for well-known parameter choice rules: discrepancy principle [65, 87–89], modified discrepancy principle [22, 73, 74], monotone error rule [84] and balancing principle [3,4,7–9,15–18,50,54–56,58–63,70], which has received much attention recently. Often these rules are implemented by computing regularized approximations for a sequence of parameters until a certain condition is fulfilled (in balancing principle a large sequence of regularized solutions is computed). In case of smooth solution more accurate than the Tikhonov approximation is its $m \ge 2$ times iterated variant but to compute a new approximation, the iterated Tikhonov method requires

solving m equations. Instead of iterated approximation we suggest to use extrapolated approximation, which is a linear combination of m Tikhonov approximations with different parameters and which requires solving only one equation to compute a new approximation. The accuracy of both approximations is the same. The extrapolated approximation is preferred to the iterated one also in case of Lavrentiev method, by the same motivation.

All regularization methods have difficulties, if no information about the noise level of the data is given. Then it is known [1] that no rule for choice of the regularization parameter can guarantee the convergence of regularized solutions to the exact one as the noise level of the data tends to zero. Nevertheless, heuristic rules, i.e. rules not using noise level δ, are very popular: quasioptimality criterion [2, 5, 6, 10, 23, 49, 51, 52, 85, 86] and its analog [69], generalized cross-validation (GCV) [24, 91], Hanke-Raus rule [42] and L-curve rule [44, 45]. The reason of popularity of these rules is that in applied ill-posed problems the noise level is usually not known exactly but if in classical parameter choice rules (discrepancy principle etc.) the supposed noise level is somewhat smaller than the actual noise level, then the error of approximate solution can be arbitrarily large. On the other hand, classical parameter choice rules are also quite sensitive to overestimation of the noise level. In this work we propose some strategies for constructing rules for choice of the regularization parameter, give several specific new rules and present the results of extensive numerical experiments on test problems of Hansen [43, 45], which are widely used in literature for testing regularization algorithms. Additional test problems are taken from papers [13, 14].

If no information about the noise level is known, then many heuristic rules, such as the quasioptimality criterion and its analog by Neubauer, Hanke-Raus rule, and Brezinski-Rodrigues-Seatzu rule choose the regularization parameter as the global minimizer of a certain function. These rules often give good results but sometimes fail. The main problem with these heuristic rules is that the global minimizers of the corresponding functions are sometimes very small, leading to very large error. Instead of global minimizer we propose to use a larger local minimizer, if there is a large maximum between them. The algorithm is to make computations with decreasing values of the parameter and stop the calculations, if the value of the function to be minimized is some fixed number of times larger than its minimal value. An alternative approach is to take the regularization parameter to be the largest local minimizer of a certain function. We also propose other modifications to known rules.

Consider parameter choice in (iterated) Tikhonov method in case of exactly given noise level. It is known that the monotone error rule gives a regularization parameter α_{ME} that is always larger than or equal to the optimal. However, to decrease this parameter, one cannot decrease the coefficient in front of the noise level, since then neither convergence nor order optimal

error estimate will be guaranteed. We use an alternative way: we modify the parameter α_{ME} to find its smaller estimate, optimizing coefficients on test problems. In our tests the estimated regularization parameter α_{MEe} gave much better results than the original parameter α_{ME}. The same ideas are also applicable to Landweber iteration method, where the monotone error rule and the discrepancy principle choose n_{ME} and n_{D} such that n_{ME} is smaller than or equal to the optimal stopping index and $n_{\mathrm{D}} - 1 \le n_{\mathrm{ME}} \le n_{\mathrm{D}}$. Numerical experiments recommend not to stop at n_{ME} or n_{D} but at $2n_{\mathrm{ME}}$ or $2n_{\mathrm{D}}$ or somewhat later.

In the Tikhonov method, in case of possible overestimation or moderate underestimation of the noise level we propose to use the rule R2e (post-estimation of the parameter from rule R2 [80]), which is less sensitive to false estimation of noise level than the discrepancy principle and the monotone error rule. In case of possible overestimation of the noise level, if underestimation is excluded, we propose to use $\alpha_{\mathrm{Me}} = \min(\alpha_{\mathrm{MEe}}, \alpha_{\mathrm{R2e}})$.

In some ill-posed problems the noise level of the data is known approximately. Recently, a rule DM for choice of the regularization parameter was proposed [33], which guarantees convergence of the approximate solution to the exact solution, as the noise level tends to zero, provided that the ratio of actual to presumed noise levels remains bounded. This rule contains two free parameters, whose values we found by optimizing on test problems.

In numerical experiments our rules gave better results than old rules. For example, in most cases the best of our heuristic rules and the rule DM in case of 100 times underestimated or overestimated noise level gave better results than the discrepancy principle in case of 2 times overestimated noise level. In Lavrentiev method the rule MEaql and heuristic rule QmC gave, on average, only 1% or 4% larger errors, respectively, than the optimal parameter (see Table 33). In method CGME, the heuristic rule DHP chose the optimal stopping index in most problems in case of smooth solution (see Table 45).

The main attention in this thesis is given to numerical analysis of regularization parameter choice rules with respect to their accuracy. While often the parameter choice rules are formulated together with proof that the error of the corresponding regularized solution tends to zero, as the noise level δ tends to zero, in real problems, however, the process $\delta \to 0$ cannot be accomplished; the range of δ's is limited by physical factors. In addition, theoretical error estimates often contain constants, whose values are unknown or hard to find but which are crucial in practical application of the algorithm. Numerical tests are a way to estimate the actual error at particular range of δ's. They also take into account changes in problem resulting from discretization, which inevitably occurs when reformulating the usually infinite-dimensional problem for solving on computer. We believe that a combination of mathematical analysis and execution on sample data sets is the best way to understand the performance of an algorithm.

Contribution of the thesis can be formulated as follows.

- Elaboration of extrapolated versions of Tikhonov and Lavrentiev methods, guaranteeing high accuracy in case of smooth solutions by moderate amount of computations.
- New parameter choice rules that work in wide range of practice-oriented problems and with various regularization methods. These rules include both the rules that use information about noise level as well as the rules that don't. Also they include rules that allow the noise level to be known only approximately.
- Strategies and principles to construct new rules, verified by performance tests of particular rules that are realizations of these strategies. Extensive numerical experiments show the advantage of resulting rules over known rules.
- Framework for numerical comparison of rules.

The set of test problems was not large, for larger set all conclusions may not hold but if problems from papers [13, 14] were added, the results remained similar.

Main results of this work are reported in talks at conferences Approximation Methods and Orthogonal Expansions (Kääriku 2003, Kääriku 2008), Mathematical Modelling and Analysis (Trakai 2005, Jūrmala 2006, Kääriku 2008), Inverse and Ill-Posed Problems of Mathematical Physics (Novosibirsk 2007), Inverse Problems: Modeling and Simulation (Fethiye 2008), Methods of Algebra and Analysis (Tartu 2008), Applied Inverse Problems (Vienna 2009).

In the following we make some notational conventions. Generic regularization parameter will be λ. In continuous methods we use traditional parameter α, so in these methods $\lambda = \alpha$. In simple iteration methods the iteration step n will be the regularization parameter, in this case $\lambda = 1/n$. When dealing with iteration methods, we write everywhere in expressions, where the regularization parameter λ is in subscript, simply n instead of $1/n$, so for example, the approximate solution found by an iteration method is $x_\lambda = x_n$. Wherever we have defined some approximation $x_.$, we automatically define $r_. = Ax_. - y$. Wherever we have defined the function $d_.(\lambda)$ (like $d_{\mathrm{D}}(\lambda) = \|r_\lambda\|$ in classical discrepancy principle), we also define the function $\varphi_.(\lambda) = d_.(\lambda)/\sqrt{\lambda}$ in case of non-self-adjoint problem and the function $\varphi_.(\lambda) = d_.(\lambda)/\lambda$ in case of self-adjoint problem—or vice versa, having the function $\varphi_.(\lambda)$, we assume that also $d_.(\lambda) = \sqrt{\lambda}\,\varphi_.(\lambda)$ or $d_.(\lambda) = \lambda\varphi_.(\lambda)$ for non-self-adjoint and self-adjoint problems, respectively, are defined.

II. REGULARIZATION METHODS

2.1. The problem

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

$$Ax = y_*, \qquad y_* \in \mathcal{R}(A) \tag{2.1}$$

in case where instead of exact data y_* only its approximation y is available. 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.

Since $y_* \in \mathcal{R}(A)$, the minimum norm solution x_* exists. In theoretical study, however, some additional information about x_* is needed. Often it is assumed that the solution x_* satisfies the *source condition*

$$\overline{x} - x_* = \mathcal{R}\big((A^*A)^p\big), \tag{2.2}$$

where $\overline{x}$ is some initial approximation of the solution of (2.1) and $p > 0$ is a constant (may be fractional). Regularization under generalized source conditions is studied, for example, in [66].

In some cases we consider problems with $y_* \notin \mathcal{R}(A)$ but $Qy_* \in \mathcal{R}(A)$, where Q is the orthoprojector $Y \to \overline{\mathcal{R}(A)}$. In this case we are interested in the minimum norm solution of equation

$$A^*Ax = A^*y_*.$$

2.2. Operator form of regularization methods

Ill-posedness of the problem means that small error in right hand side can lead to large error in computed solution. To solve the problem in stable way, we turn to regularization.

Many regularization methods can be uniformly treated as special cases of the general regularization method of the form [19, 89]

$$x_\lambda = (I - A^*Ag_\lambda(A^*A))\overline{x} + g_\lambda(A^*A)A^*y. \tag{2.3}$$

Here x_λ is the *regularized solution* and λ is a positive real number, called *regularization parameter*. A particular regularization method is determined by the family $\{g_\lambda : \lambda \in (0,\infty)\}$ of piecewise continuos *generating functions* defined on $[0, \|A^*A\|]$ and satisfying the following conditions:

$$\sup_{0\le t\le\|A^*A\|} |g_\lambda(t)| \le \gamma\lambda^{-1} \qquad (\lambda > 0), \tag{2.4}$$

$$\sup_{0\le t\le\|A^*A\|} t^p|1-tg_\lambda(t)| \le \gamma_p\lambda^p \qquad (\lambda>0,\ 0\le p\le p_0), \tag{2.5}$$

$$\sup_{0\le t\le\|A^*A\|} |t^{1/2}g_\lambda(t)| \le \gamma_*\lambda^{-1/2} \qquad (\lambda>0), \tag{2.6}$$

where γ, γ_p, γ_* and p_0 are positive constants. The largest p_0 for which the inequality (2.5) holds, is called the *qualification* of the method; it can be finite or infinite.

Denoting $\overline{r} = A\overline{x} - y$ and using the commutativity property (see [89]) $A^*Ag_\lambda(A^*A) = g_\lambda(A^*A)A^*A$, the equality (2.3) can also be written as

$$x_\lambda = \overline{x} - g_\lambda(A^*A)A^*\overline{r}. \tag{2.7}$$

From this we obtain an expression for the discrepancy $r_\lambda = Ax_\lambda - y$:

$$r_\lambda = (I - AA^*g_\lambda(AA^*))\overline{r} = K_\lambda(AA^*)\overline{r}, \tag{2.8}$$

where the *discrepancy functions* K_λ are defined as $K_\lambda(t) = 1 - tg_\lambda(t)$.

All abovementioned formulas and conditions can be adjusted to the case of self-adjoint problem, where $X = Y$ and $A = A^* \ge 0$, by simply omitting A^*. Thus instead of (2.3), (2.7), and (2.8), in this case we can compute the approximates and their discrepancies according to formulas

$$x_\lambda = (I - Ag_\lambda(A))\overline{x} + g_\lambda(A)y = \overline{x} - g_\lambda(A)\overline{r}, \tag{2.9}$$

$$r_\lambda = (I - Ag_\lambda(A))\overline{r} = K_\lambda(A)\overline{r}, \tag{2.10}$$

where the functions g_λ and K_λ are defined in the same way as before.

In *continuous regularization methods* the parameter λ can take every value in the interval $(0,\infty)$. In *iterative regularization methods* the parameter λ has only discrete values that can be associated with iteration steps.

2.3. Particular regularization methods

The most prominent regularization method is the Tikhonov method together with its iterated variant. The *m-iterated Tikhonov method* ($m = 1, 2, \ldots$) is defined as follows. Take $x_{0;\alpha} = \overline{x}$ and compute $x_{1;\alpha}, \ldots, x_{m;\alpha}$ iteratively from

$$\alpha x_{n;\alpha} + A^*Ax_{n;\alpha} = \alpha x_{n-1;\alpha} + A^*y \qquad (n = 1, \ldots, m); \tag{2.11}$$

the approximate solution of (2.1) is then $x_{m;\alpha}$.

The ordinary Tikhonov method ($m = 1$) takes $\overline{x} = 0$. In this case we also write x_α instead of $x_{1;\alpha}$ (and r_α instead of $r_{1;\alpha}$).

The m-iterated Tikhonov method is a special case of (2.3) with $\lambda = \alpha$ and $g_\lambda(t) = g_{m;\alpha}(t) = t^{-1}(1 - (1+\alpha^{-1}t)^{-m})$. Inequalities (2.4), (2.5), (2.6) are satisfied with $\gamma = m$, $\gamma_p = (p/m)^p(1-p/m)^{m-p}$, $\gamma_* = m^{1/2}$ and $p_0 = m$. In addition, $K_\lambda(t) = K_{m;\alpha}(t) = (1+\alpha^{-1}t)^{-m}$.

If $\|y - y_*\| \le \delta$, then under assumption (2.2) a proper choice of α (depending on δ) guarantees $\|x_{m;\alpha} - x_*\| = \mathcal{O}(\delta^{p/(p+1)})$ for all $p \le 2m$ (see

Theorem 3 of Section 3.4). We say that the maximal guaranteed accuracy of m-iterated Tikhonov method is $\mathcal{O}(\delta^{2m/(2m+1)})$.

We also introduce the *iteration operator*

$$B_\alpha = K_{m;\alpha}(AA^*)^{1/(2m)} = (1 + \alpha^{-1}AA^*)^{-1/2}, \tag{2.12}$$

which acts as "iteration by half step", as can be seen from the relations

$$\|B_\alpha r_{m;\alpha}\| = (r_{m;\alpha}, r_{m+1;\alpha})^{1/2} \qquad \text{and} \qquad B_\alpha^2 r_{m;\alpha} = r_{m+1;\alpha}.$$

The analogs of this operator can also be defined for other methods of finite qualification p_0 by formula $B_\lambda = K_\lambda(AA^*)^{1/(2p_0)}$. For methods of infinite qualification we let $B_\lambda = I$.

The self-adjoint analog of the m-iterated Tikhonov method in case $A^* = A \geq 0$ is the *m-iterated Lavrentiev method* ($m = 1, 2, \ldots$), which starts with $x_{0;\alpha} = \overline{x}$ and computes $x_{1;\alpha}, \ldots, x_{m;\alpha}$ from

$$(\alpha x_{n;\alpha} + Ax_{n;\alpha}) = \alpha x_{n-1;\alpha} + y \qquad (n = 1, \ldots, m);$$

the approximate solution of (2.1) is $x_{m;\alpha}$. This method is a special case of (2.9) with the same λ and $g_\lambda(t)$ as in m-iterated Tikhonov method. Therefore also the inequalities (2.4), (2.5), (2.6) are satisfied with the same constants. In particular, $p_0 = m$.

The maximal guaranteed accuracy of m-iterated Lavrentiev method is $\mathcal{O}(\delta^{m/(m+1)})$ (Theorem 4 of Section 3.4). Order optimal error bounds in Hilbert scales for a priori parameter choice and for discrepancy principle are given in [48].

The iteration operator B_α for m-iterated Lavrentiev method can be defined as

$$B_\alpha = K_{m;\alpha}(AA^*)^{1/m} = (1 + \alpha^{-1}A)^{-1}, \tag{2.13}$$

it satisfies $B_\alpha r_{m;\alpha} = r_{m+1;\alpha}$.

We now formulate some properties that are useful later; they can be proved by direct computation.

Proposition. *The functions $g_{m;\alpha}$ and $K_{m;\alpha}$ of m-iterated Tikhonov method (and of m-iterated Lavrentiev method) satisfy relations*

$$g_{m+1;\alpha}(t) - g_{m;\alpha}(t) = \alpha^{-1}K_{m+1;\alpha}(t), \tag{2.14}$$

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}g_{m;\alpha}(t) = -m\alpha^{-2}K_{m+1;\alpha}(t), \tag{2.15}$$

$$K_{m+1;\alpha}(t) = (1 + \alpha^{-1}t)^{-1}K_{m;\alpha}(t). \tag{2.16}$$

For m-iterated Tikhonov method a direct consequence of these properties is the equality

$$\alpha(x_{m;\alpha} - x_{m+1;\alpha}) = A^* r_{m+1;\alpha}, \tag{2.17}$$

since by (2.14)

$$g_{m+1;\alpha}(A^*A)A^*\overline{r} - g_{m;\alpha}(A^*A)A^*\overline{r} = \alpha^{-1}K_{m+1;\alpha}(A^*A)A^*\overline{r},$$

whose left hand side is by (2.7) equal to $x_{m;\alpha} - x_{m+1;\alpha}$ and right hand side by (2.8) to $\alpha^{-1}A^* r_{m+1;\alpha}$.

Let us now consider iterative methods. One of the simplest of these is the *Landweber method* or *explicit iteration scheme*. Let $\mu \in (0, 2/\|A^*A\|)$. Take $x_0 = \overline{x}$ and compute $x_1, x_2, \ldots$ iteratively from

$$x_n = x_{n-1} - \mu A^*(Ax_{n-1} - y) \qquad (n = 1, 2, \ldots). \tag{2.18}$$

This method is of the form (2.3) with $\lambda = 1/n$ and $g_\lambda(t) = g_n(t) = t^{-1}(1 - (1 - \mu t)^n)$. Inequalities (2.4), (2.5), (2.6) are satisfied with $\gamma = \mu$, $\gamma_p = (p/(\mu e))^p$, $\gamma_* = \mu^{1/2}$, and $p_0 = \infty$.

Many papers note that the Landweber method is not practical, since it needs too many iterations. Based on our computational experience, we strongly recommend to implement this method by operator form of iterations (as recommended in [87,89]), which allows to compute x_n for indices $n = m^k$ ($m \geq 2$; $k = 1, 2, \ldots$). Define the operators

$$C_0 = \mu I, \qquad C_k = C_{k-1} \sum_{j=0}^{m-1} (I - A^*AC_{k-1})^j \qquad (k = 1, 2, \ldots). \tag{2.19}$$

Then $x_n = (I - A^*AC_k)\overline{x} + C_kA^*y$. Typically this algorithm is applied with $m = 2$ (scheme of Schultz-Hotelling). In our numerical tests this algorithm was particularly effective, since the operator A was often a diagonal matrix.

Another iterative method is the *implicit iteration scheme*. Let $\alpha > 0$ be a constant. Take $x_0 = \overline{x}$ and compute $x_1, x_2, \ldots$ iteratively from

$$\alpha x_n + A^*Ax_n = \alpha x_{n-1} + A^*y \qquad (n = 1, 2, \ldots). \tag{2.20}$$

Here $\lambda = 1/n$ and $g_\lambda(t) = g_n(t) = t^{-1}(1 - (1 + \alpha^{-1}t)^{-n})$. Inequalities (2.4), (2.5), (2.6) are satisfied with $\gamma = \alpha^{-1}$, $\gamma_p = (\alpha p)^p$, $\gamma_* = \alpha^{-1/2} \sup_{0<t<\infty} t^{-1/2}(1 - e^{-t})$, and $p_0 = \infty$.

Also the implicit iteration scheme can be implemented via operator iterations (2.19) by taking $C_0 = (\alpha I + A^*A)^{-1}$; other formulas do not change.

Both Landweber method and the implicit iteration scheme can be viewed as discrete variants of a continuous regularization method, the *method of asymptotical regularization* or *method of Cauchy problem*, which finds the approximate solution of (2.1) as the solution of the problem

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}x_\alpha + A^*Ax_\alpha = A^*y, \qquad x_0 = \overline{x}. \tag{2.21}$$

Here $\lambda = \alpha$ and $g_\lambda(t) = g_\alpha(t) = t^{-1}(1 - e^{\alpha t})$. Inequalities (2.4), (2.5), (2.6) are satisfied with $\gamma = 1$, $\gamma_p = (p/e)^p$, $\gamma_* = \sup_{0<t<\infty} t^{-1/2}(1 - e^{-t}) \approx 0.6382$, and $p_0 = \infty$.

Of the form (2.3) is also the *spectral cut-off method*

$$x_\alpha = \int_\alpha^{\|A^*A\|} \frac{1}{\eta}\mathrm{d}E_\eta A^*y, \tag{2.22}$$

where $\{E_\eta\}$ is the spectral family of A^*A. In this method $\lambda = \alpha$ and $g_\lambda(t) = g_\alpha(t) = t^{-1}$, if $t \geq \alpha$, and 0 otherwise. Inequalities (2.4), (2.5), (2.6) are satisfied with $\gamma = \gamma_p = \gamma_* = 1$ and $p_0 = \infty$. For compact operators

this method is called *truncated singular value decomposition* (TSVD). The compact operator A has a singular value decomposition (σ_i, u_i, v_i), where σ_i are positive real numbers in decreasing order, $\{u_i\}$, $\{v_i\}$ are complete orthonormal systems in X, Y, respectively, and all these components are related to each other by relations $Au_i = \sigma_i v_i$, $A^* v_i = \sigma_i u_i$ $(i = 1, 2, \ldots)$. In TSVD we identify the regularization parameter with the number of terms used, finding the approximate solution of (2.1) as

$$x_n = \sum_{i=1}^{n} \frac{1}{\sigma_i} (y, v_i) u_i.$$

Here $g_\lambda(t) = g_n(t) = t^{-1}$, if $\sigma_n \geq t$, and 0 otherwise.

Wherever we refer to iteration methods in the following, we also include TSVD among them.

In self-adjoint problems self-adjoint variants of methods (2.18), (2.20), (2.21), (2.22) can be used, omitting A^* in formulas of these methods and also in operator iterations (2.19).

2.4. Conjugate gradient type methods

To find the approximate solution of $Ax = y$, we can use conjugate gradient type methods CGLS and CGME, which minimize the discrepancy or the error, respectively, in Krylov subspace.

The *CGLS method* is the method of conjugate gradients applied to symmetrized equation $A^*Ax = A^*y$. Let x_0 be an initial approximation (for example $x_0 = 0$) and $r_0 = Ax_0 - y$. To start the algorithm, additionally fix the initial values $u_0 = 0$ and $p_{-1} = \infty$. For $n = 1, 2, \ldots$ compute iteratively

$$p_{n-1} = A^* r_{n-1}, \quad \beta_{n-1} = \frac{\|p_{n-1}\|^2}{\|p_{n-2}\|^2}, \quad u_n = p_{n-1} + \beta_{n-1} u_{n-1},$$

$$v_n = Au_n, \quad \gamma_n = \frac{\|p_{n-1}\|^2}{\|v_n\|^2}, \quad x_n = x_{n-1} - \gamma_n u_n, \quad r_n = r_{n-1} - \gamma_n v_n.$$

The *CGME method* is the method of conjugate gradients applied to dually symmetrized problem $AA^*z = y$ with $x = A^*z$. Again let x_0 be an initial approximation and $r_0 = Ax_0 - y$. Fix the initial values $u_0 = 0$ and $r_{-1} = \infty$. For $n = 1, 2, \ldots$ compute iteratively

$$p_{n-1} = A^* r_{n-1}, \quad \beta_{n-1} = \frac{\|r_{n-1}\|^2}{\|r_{n-2}\|^2}, \quad u_n = p_{n-1} + \beta_{n-1} u_{n-1},$$

$$v_n = Au_n, \quad \gamma_n = \frac{\|r_{n-1}\|^2}{\|u_n\|^2}, \quad x_n = x_{n-1} - \gamma_n u_n, \quad r_n = r_{n-1} - \gamma_n v_n.$$

If we wish to compute the approximates x_n in these methods according to the relation $x_n = x_{n-1} - A^* z_n$, then we can replace $u_n = p_{n-1} + \beta_{n-1} u_{n-1}$ in above algorithms by two formulas $w_n = r_{n-1} + \beta_{n-1} w_{n-1}$, $u_n = A^* w_n$, whereby $w_0 = 0$.

The *conjugate gradient method* (CG) itself is defined for a self-adjoint non-negative operator A. Let x_0 be an initial approximation and $r_0 = Ax_0 - y$. Let $u_0 = 0$ and $r_{-1} = \infty$. For $n = 1, 2, \ldots$ compute

$$\beta_{n-1} = \frac{\|r_{n-1}\|^2}{\|r_{n-2}\|^2}, \quad u_n = r_{n-1} + \beta_{n-1} u_{n-1},$$

$$v_n = Au_n, \quad \gamma_n = \frac{\|r_{n-1}\|^2}{(u_n, v_n)}, \quad x_n = x_{n-1} - \gamma_n u_n, \quad r_n = r_{n-1} - \gamma_n v_n.$$

2.5. Extrapolation of Tikhonov and Lavrentiev methods

Extrapolation is a technique to increase the accuracy of regularization methods by finding the approximate solution not as one Tikhonov or Lavrentiev approximation but as a linear combination of several approximations.

Up to now, there are few papers that address the idea of extrapolation as a means to increase the accuracy of approximation methods in ill-posed problems. In [57] (see also [83]) the extrapolated Tikhonov method and a version of the extrapolated Lavrentiev method were proposed for systems of linear algebraic equations. The extrapolated Tikhonov method for operator equations with exact data was studied in [25]. In [11, 12] some other techniques for extrapolation of Tikhonov method for ill-conditioned linear systems were proposed. In case of noisy data the extrapolated Tikhonov method was studied in [28, 29].

Let $x_{\alpha_1}, \ldots, x_{\alpha_m}$ be Tikhonov approximations (Lavrentiev approximations) with pairwise different parameters $\alpha_1, \ldots, \alpha_m$. The *$m$-extrapolated Tikhonov (Lavrentiev) approximation* is given by

$$x_{\alpha_1,\ldots,\alpha_m} = \sum_{i=1}^{m} d_i x_{\alpha_i}, \qquad d_i = \prod_{\substack{j=1 \\ j\neq i}}^{m} \left(1 - \frac{\alpha_i}{\alpha_j}\right)^{-1}. \tag{2.23}$$

The coefficients d_i are chosen in such way that the leading terms in 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}.

For example, if $\alpha_1 = \alpha/2$, $\alpha_2 = \alpha$, $\alpha_3 = 2\alpha$, then

$$x_{\alpha_1,\alpha_2} = 2x_{\alpha/2} - x_\alpha, \qquad x_{\alpha_1,\alpha_2,\alpha_3} = \frac{8}{3}x_{\alpha/2} - 2x_\alpha + \frac{1}{3}x_{2\alpha}.$$

If the sequence of parameters (α_n) is defined as $\alpha_n = q^n$ ($q < 1$; $n = 0, 1, \ldots$), then

$$\begin{aligned}
x_{\alpha_n,\alpha_{n+1}} &= (1-q)^{-1}(-qx_{\alpha_n} + x_{\alpha_{n+1}}),\\
x_{\alpha_n,\alpha_{n+1},\alpha_{n+2}} &= (1-q)^{-2}(q^3(1+q)^{-1}x_{\alpha_n} - qx_{\alpha_{n+1}} + (1+q)^{-1}x_{\alpha_{n+2}}),\\
x_{\alpha_n,\alpha_{n+1},\alpha_{n+2},\alpha_{n+3}} &= (1-q)^{-3}(1+q)^{-1}(-q^6(1+q+q^2)^{-1}x_{\alpha_n}\\
&\qquad + q^3 x_{\alpha_{n+1}} - qx_{\alpha_{n+2}} + (1+q+q^2)^{-1}x_{\alpha_{n+3}}).
\end{aligned}$$

The following table shows rounded values of coefficients d_i for $m = 1, \ldots, 5$ at $q = 0.1$ (upper part) and $q = 0.9$ (lower part):

m	d_1	d_2	d_3	d_4	d_5
1	1				
2	-0.11	1.1			
3	$1.1 \cdot 10^{-3}$	-0.12	1.1		
4	$-1.1 \cdot 10^{-6}$	$1.2 \cdot 10^{-3}$	-0.12	1.1	
5	$1.1 \cdot 10^{-10}$	$-1.2 \cdot 10^{-6}$	$1.3 \cdot 10^{-3}$	-0.12	1.1
1	1				
2	-9	10			
3	38.4	-90.0	52.6		
4	-103	384	-474	194	
5	197	-1032	2019	-1748	565

In case of source-like solutions (2.2) the maximal guaranteed accuracy of m-extrapolated Tikhonov approximation is $\mathcal{O}(\delta^{2m/(2m+1)})$, versus the maximal guaranteed accuracy $\mathcal{O}(\delta^{2/3})$ of single Tikhonov approximation (cf. Theorem 12). For Lavrentiev method the accuracies are $\mathcal{O}(\delta^{m/(m+1)})$ and $\mathcal{O}(\delta^{1/2})$, respectively.

In [28] it is shown that the m-extrapolated Tikhonov approximation coincides with the approximation found by the non-stationary m times iterated Tikhonov regularization [40], which starts with $x = 0$ and computes the iterates $x_{\alpha_1}, x_{\alpha_1,\alpha_2}, \ldots, x_{\alpha_1,\alpha_2\ldots,\alpha_m}$ from

$$\alpha_n x_{\alpha_1,\ldots,\alpha_n} + A^*Ax_{\alpha_1,\ldots,\alpha_n} = \alpha_n x_{\alpha_1,\ldots,\alpha_{n-1}} + A^*y \quad (n = 1, \ldots, m), \quad (2.24)$$

taking $x_{\alpha_1,\ldots,\alpha_m}$ as the approximate solution of (2.1). Numerical experience shows that if m is large and α_n's are not very different, then the non-stationary iterated Tikhonov method is computationally more stable than the direct formula (2.23) and this is the way we have calculated extrapolated approximations in numerical tests of Chapter IV. The same remarks hold for Lavrentiev method.

Extrapolated Tikhonov approximation can be viewed as being of type (2.3) in two ways.

1. Let m be fixed and let the sequence of parameters $\alpha_1, \ldots, \alpha_m$ be computed on base of single α, for example from relations $\alpha_n = \alpha_{n-1}q$ ($n = 2, \ldots, m$; q fixed) and either $\alpha_1 = \alpha$ or $\alpha_m = \alpha$ or $\alpha_{\lfloor m/2 \rfloor} = \alpha$. Then $\lambda = \alpha$ and $g_\lambda(t) = g_\alpha(t) = t^{-1}(1 - \prod_{i=1}^{m}(1 + \alpha_i^{-1}t)^{-1})$. In this setting the extrapolated Tikhonov method is an analog of m-iterated Tikhonov method, where at different steps n different α_n are used. Under assumption (2.2) a proper choice of α gives the approximate solution $x_\alpha = x_{\alpha_1,\ldots,\alpha_m}$ with maximal accuracy $\mathcal{O}(\delta^{2m/(2m+1)})$.
2. Let $\alpha_1, \alpha_2, \ldots$ be a fixed infinite sequence. Compute approximations $x_{\alpha_1}, x_{\alpha_1,\alpha_2}, \ldots, x_{\alpha_1,\alpha_2,\ldots,\alpha_n}, \ldots$ Here the regularization parameter is

$\lambda = 1/n$ and $g_\lambda(t) = g_n(t) = t^{-1}(1 - \prod_{i=1}^{n}(1+\alpha_i^{-1}t)^{-1})$. In this case the extrapolated Tikhonov method is an analog of non-stationary iterated Tikhonov method [40] (implicit iteration scheme with parameter α_n at step n). Under assumption (2.2) a proper choice of n gives the approximate solution $x_n = x_{\alpha_1,\dots,\alpha_n}$ with accuracy $\mathcal{O}(\delta^{p/(p+1)})$ for all $p > 0$.

Extrapolation can also be used together with iterated version of Tikhonov (Lavrentiev) method. Let $\alpha_1, \dots, \alpha_m$ be pairwise different parameters and let $x_{j;\alpha_i}$ be j-iterated Tikhonov (Lavrentiev) approximation with parameter α_i ($i = 1, \dots, m$; $j = 1, \dots, M_i$). For different α_i different number of iterations $M_1, \dots, M_m$ may be used. The extrapolated approximation is given by

$$x_{\underbrace{\alpha_1,\dots,\alpha_1}_{M_1},\underbrace{\alpha_2,\dots,\alpha_2}_{M_2},\dots,\underbrace{\alpha_m,\dots,\alpha_m}_{M_m}} = \sum_{i=1}^{m}\sum_{j=1}^{M_i} d_{i,j}\, x_{j;\alpha_i},$$

where the coefficients $d_{i,j}$ can be uniquely determined from relation (see [28] and Theorem 14)

$$\sum_{i=1}^{m}\sum_{j=1}^{M_i} d_{i,j}\left(1+\frac{t}{\alpha_i}\right)^{-j} = \prod_{i=1}^{m}\left(1+\frac{t}{\alpha_i}\right)^{-M_i} \qquad (\forall t \in \mathbb{R}).$$

As with extrapolation of single Tikhonov (Lavrentiev) approximations, also here the extrapolated approximation can be found by non-stationary iterated Tikhonov (Lavrentiev) method, where M_1 iteration steps are performed with parameter α_1, then M_2 steps with parameter $\alpha_2, \dots, M_m$ steps with parameter α_m.

Extrapolation can increase the accuracy of approximate solution for example in case, where several approximations $x_{m;\alpha_n}$ in m-iterated Tikhonov method are computed for a sequence (α_n) of parameters until some condition is fulfilled, and traditionally a single approximation with maximal accuracy $\mathcal{O}(\delta^{2m/(2m+1)})$ is declared to be the final approximate solution (as, for example, in balancing principle, see Section 3.1.6). The accuracy of extrapolated approximation with m terms and with proper parameters is the same as the accuracy of m-iterated Tikhonov approximation. At the same time, extrapolation requires less computational work since at transition from $x_{m;\alpha_n}$ to $x_{m;\alpha_{n+1}}$ in m-iterated Tikhonov method m equations have to be solved, whereas at transition from $x_{\alpha_1,\dots,\alpha_m}$ to $x_{\alpha_2,\dots,\alpha_{m+1}}$ we need to solve only one additional equation.

III. RULES FOR CHOICE OF THE REGULARIZATION PARAMETER

When applying regularization methods, we need to properly choose the regularization parameter λ. If λ is too big, then the computed approximation x_λ is too close to the initial guess $\overline{x}$ (see (2.7) and (2.4)) but if λ is too small, then the numerical implementation tends to be unstable due to ill-posedness of the problem. There are several strategies (so-called rules) for choosing this parameter, using different kinds of information. The work reported in this thesis is mainly devoted to *a posteriori* rules. These rules can be classified according to how much information there is available about the noise level.

1. Noise level is known fully: $\|y - y_*\| \le \delta$, where δ is known.
2. Noise level is not known.
3. Noise level is known approximately: there holds $\lim_{\delta\to 0} \frac{\|y - y_*\|}{\delta} \le C$, where C is an unknown constant.

If the noise level is fully known (case 1), then for source-like solutions (2.2) the best error estimate (with the fastest convergence rate) that can be achieved with any method is [87, 89]

$$\|x_\lambda - x_*\| \le \text{const}\, \delta^{\frac{p}{p+1}}. \tag{3.1}$$

This error estimate is called *order optimal*. In a given method a proper choice of regularization parameter λ may guarantee the error estimate (3.1) for all $p \in [0, M]$ with some finite M or for all $p \in [0, \infty)$. Theoretical performance of a parameter choice rule may be characterized by maximal p for which the order optimal error estimate (3.1) holds.

In the following, λ_* denotes the optimal regularization parameter, i.e.

$$\lambda_* = \operatorname*{argmin}_{\lambda \ge 0} \{\|x_\lambda - x_*\|\}$$

(in iterative methods the minimum is taken over $\lambda \in \mathbb{N}$).

3.1. Parameter choice rules for known noise level

If the noise level δ with $\|y - y_*\| \le \delta$ is known, then many rules choose the regularization parameter λ as the solution of some equation $d(\lambda) = C\delta$, $C \ge 1$. We now consider several forms of $d(\lambda)$ for which these rules guarantee convergence $\|x_\lambda - x_*\| \to 0$ as $\delta \to 0$ and often also order-optimal error estimate, thereby referring to theorems of Section 3.4.

3.1.1. Discrepancy principle.

The most widely used parameter choice rule is the discrepancy principle [65, 87–89], which can be formulated as follows.

Define $d_{\mathrm{D}}(\lambda) = \|r_\lambda\|$ (this means that $d_{\mathrm{D}}(\alpha) = \|r_\alpha\|$ for continuous methods and $d_{\mathrm{D}}(n) = \|r_n\|$ for iterative methods).

- In a continuous regularization method choose the regularization parameter $\alpha = \alpha_{\mathrm{D}}$ as the solution of the equation $d_{\mathrm{D}}(\alpha) = C\delta$, where $C > 1$ is a fixed constant.
- In an iterative regularization method choose the regularization parameter $n = n_{\mathrm{D}}$ as the first n for which $d_{\mathrm{D}}(n) \le C\delta$, where $C > 1$ is a fixed constant.

The parameter found according to this rule can in principle be any positive number, no matter how large or small. In computation, however, the range of its values is finite, depending on chosen data type and possibly other factors. Thus, when making calculations, we supplement this rule with the following conditions.

- In case of continuous regularization method let $\alpha \in [\alpha_{\min}, \alpha_{\max}]$. If $d_{\mathrm{D}}(\alpha_{\max}) \le C\delta$, then take $\alpha_{\mathrm{D}} = \alpha_{\max}$. If $d_{\mathrm{D}}(\alpha) > C\delta$ on the whole segment $[\alpha_{\min}, \alpha_{\max}]$, then take $\alpha_{\mathrm{D}} = \alpha_{\min}$.
- In case of iterative regularization method let $n \in [n_{\min}, n_{\max}]$. If $d_{\mathrm{D}}(n_{\min}) \le C\delta$, then take $n_{\mathrm{D}} = n_{\max}$. If $d_{\mathrm{D}}(n) > C\delta$ on the segment $[n_{\min}, n_{\max}]$, then $n_{\mathrm{D}} = n_{\max}$.

Analogously we restrict other parameter choice rules.

In m-iterated Tikhonov method the discrepancy principle guarantees the order optimal error estimate (3.1) for all $p \le 2m - 1$ in case of source-like solutions (2.2) (see Theorem 3 of Section 3.4).

In ordinary Lavrentiev method ($m = 1$) the discrepancy principle is not applicable and leads to divergence of approximate solutions but it can be used in iterated Lavrentiev method with $m \ge 2$. Then in case of source-like solutions it guarantees order optimal error estimate (3.1) for all $p \le m - 1$ (see Theorem 4). Note that in [67] the discrepancy principle for modifications of Lavrentiev method in Hilbert scales is investigated and in [64] the rule $d_{\mathrm{D}}(\alpha) = \delta^s$ with $s \in (0, 1)$ is proposed.

For method CGLS the discrepancy principle guarantees convergence and in case of source-like solution (2.2) also order optimal error estimates for all p [37, 68, 71].

Also in method CGME the discrepancy principle is not applicable and leads to divergence of approximate solutions. For this method Hanke proposed [37, 38] a version of the discrepancy principle $d_{\mathrm{DH}}(n) = C\delta$, where $d_{\mathrm{DH}}(n) = \big(\sum_{i=0}^{n} \|r_i\|^{-2}\big)^{-1/2}$. This rule guarantees the convergence and for source-like solutions (2.2) also order optimal error estimate (3.1).

Discrepancy principle can also be used in extrapolated Tikhonov method and in extrapolated Lavrentiev method or even in extrapolation of iterated variants of these methods. In [28] the following results are proved.

If in extrapolated Tikhonov approximation or in extrapolated Lavrentiev approximation the number m of terms is fixed and $\alpha_n = q_n\alpha$ with fixed q_n ($n = 1, \ldots, m$), then the function $d_{\mathrm{D}}(\alpha) = \|r_{\alpha_1,\ldots,\alpha_m}\|$ is monotonically increasing. If α_{D} is the solution of the equation $d_{\mathrm{D}}(\alpha) = C\delta$, then for $x_\alpha = x_{\alpha_1,\ldots,\alpha_m}$ with $\alpha = \alpha_{\mathrm{D}}$ there holds $\|x_\alpha - x_*\| \to 0$ as $\delta \to 0$ and in case (2.2) the error estimate (3.1) is valid with $p \le 2m - 1$ for extrapolated Tikhonov approximation and with $p \le m - 1$ for extrapolated Lavrentiev approximation (see Theorem 12).

If in extrapolated Tikhonov approximation or in extrapolated Lavrentiev approximation the sequence $\alpha_1 \ge \alpha_2 \ge \ldots$ is given, then the function $d_{\mathrm{D}}(n) = \|r_{\alpha_1,\ldots,\alpha_n}\|$ is monotonically decreasing. If $\sum_{n=1}^{\infty} \alpha_n^{-1} = \infty$, $\alpha_n^{-1} \le \sum_{i=1}^{n-1} \alpha_i^{-1}$, then there exists n for which $d_{\mathrm{D}}(n) \le C\delta$. Let n_{D} be the first n for which the last inequality holds. Then for $x_n = x_{\alpha_1,\ldots,\alpha_n}$ with $n = n_{\mathrm{D}}$ there holds $\|x_n - x_*\| \to 0$ as $\delta \to 0$ and in case (2.2) the error estimate (3.1) is valid with any $p > 0$ (see Theorem 13).

Let $\alpha_n = q_n\alpha$ ($n = 1, \ldots, m$) with m and q_n fixed. Let

$$x_\alpha = x_{\underbrace{\alpha_1,\ldots,\alpha_1}_{M_1},\underbrace{\alpha_2,\ldots,\alpha_2}_{M_2},\ldots,\underbrace{\alpha_m,\ldots,\alpha_m}_{M_m}}$$

be an approximation found by extrapolation of iterated Tikhonov method or iterated Lavrentiev method, where M_n is the number of iteration steps performed with parameter α_n. Define $d_{\mathrm{D}}(\alpha) = \|r_\alpha\|$. If α_{D} is the solution of the equation $d_{\mathrm{D}}(\alpha) = C\delta$, then $\|x_\alpha - x_*\| \to 0$ as $\delta \to 0$ and in case (2.2) the error estimate (3.1) is valid with $p \le 2(M_1 + \ldots + M_m) - 1$ for Tikhonov method and with $p \le M_1 + \ldots + M_m - 1$ for Lavrentiev method (see Theorem 14).

In all previous equalities and inequalities it is assumed that $C > 1$ but for practical purposes $C = 1$ also works well.

3.1.2. Modified discrepancy principle.

In m-iterated Tikhonov method and in m-iterated Lavrentiev method the following modification of the discrepancy principle (rule MD, also called Raus-Gfrerer rule) [22, 73, 74] can be used. Define $d_{\mathrm{MD}}(\alpha) = \|B_\alpha r_{m;\alpha}\|$, where B_α is the iteration operator (2.12) or (2.13) in non-self-adjoint or self-adjoint case, respectively. Choose the regularization parameter $\alpha = \alpha_{\mathrm{MD}}$ as the solution of the equation $d_{\mathrm{MD}}(\alpha) = C\delta$ with constant $C > 1$.

Thus, for m-iterated Tikhonov method $d_{\mathrm{MD}}(\alpha) = (r_{m;\alpha}, r_{m+1;\alpha})^{1/2}$ and for m-iterated Lavrentiev method $d_{\mathrm{MD}}(\alpha) = \|r_{m+1;\alpha}\|$. Unlike ordinary discrepancy principle, the modified discrepancy principle is order optimal for

full range: in case of source-like solutions (2.2) the error estimate (3.1) holds for all $p \le 2m$ in m-iterated Tikhonov method and for all $p \le m$ in m-iterated Lavrentiev method (including $m = 1$).

In m-iterated Tikhonov method the monotone error rule (considered in the following subsection) is better than the modified discrepancy principle: in [84] it is proved that $\alpha_{\mathrm{MD}} \ge \alpha_{\mathrm{ME}}$ (and in case $y \notin \mathcal{N}(A^*)$ even $\alpha_{\mathrm{MD}} > \alpha_{\mathrm{ME}}$), which implies the inequality $\|x_{\alpha_{\mathrm{MD}}} - x_*\| \ge \|x_{\alpha_{\mathrm{ME}}} - x_*\|$.

Using the rule MD in ordinary Tikhonov method ($m = 1$), we have to compute the discrepancy $r_{2;\alpha}$. Replacing this discrepancy of two times iterated approximation by discrepancy $r_{\alpha,q\alpha}$ of extrapolated approximation, we come to the function $d_{\mathrm{eMD}}(\alpha) = (r_\alpha, r_{\alpha,q\alpha})$. (Here and in the following, "e" in front of the rule name means using extrapolated approximations.) This can be generalized to extrapolated Tikhonov approximation, where the number m of terms is fixed. In [28] it is proved that if $\alpha_n = q_n\alpha$ with q_n fixed ($n = 1, \ldots, m+1$), then the function $d_{\mathrm{eMD}}(\alpha) = (r_{\alpha_1,\ldots,\alpha_m}, r_{\alpha_1,\ldots,\alpha_{m+1}})^{1/2}$ is monotonically decreasing. If $x_\alpha = x_{\alpha_1,\ldots,\alpha_m}$, where $\alpha = \alpha_{\mathrm{eMD}}$ is the solution of $d_{\mathrm{eMD}}(\alpha) = C\delta$, then $\|x_\alpha - x_*\| \to 0$ as $\delta \to 0$ and in case of source-like solutions (2.2) the error estimate (3.1) holds with $p \le 2m$ (cf. Theorem 15).

In extrapolated Tikhonov method we often choose the regularization parameter α from some sequence (α_n), for example, from the geometric sequence $\alpha_n = q^n$ ($n = 0, 1, \ldots$; $q < 1$). In this case we stop at first $\alpha = \alpha_n$ for which $d_{\mathrm{eMD}}(\alpha_n) \le C\delta$.

Analogous results also hold for m-extrapolated Lavrentiev method (for ordinary Lavrentiev method, if $m = 1$). Here $d_{\mathrm{eMD}}(\alpha) = \|r_{\alpha_1,\ldots,\alpha_{m+1}}\|$, where $\alpha_1, \ldots, \alpha_{m+1}$ are defined in the same way as above. If $\alpha = \alpha_{\mathrm{eMD}}$ is the solution of $d_{\mathrm{eMD}}(\alpha) = C\delta$, then for $x_\alpha = x_{\alpha_1,\ldots,\alpha_m}$ there holds $\|x_\alpha - x_*\| \to 0$ as $\delta \to 0$ and in case (2.2) the error estimate (3.1) is valid with $p \le m$ (see Theorem 16, also [72] for case $m = 1$).

3.1.3. Monotone error rule.

The monotone error rule [84] is based on the idea to search for the smallest computable value λ_{ME} of the regularization parameter λ, for which it can be guaranteed that the error $\|x_\lambda - x_*\|$ is a monotonically increasing function of λ for $\lambda \in (\lambda_{\mathrm{ME}}, \infty)$.

In case of continuous regularization methods of the form (2.3) let

$$d_{\mathrm{ME}}(\alpha) = \frac{(r_\alpha, -\frac{\mathrm{d}}{\mathrm{d}\alpha} g_\alpha(AA^*)\overline{r})}{\|\frac{\mathrm{d}}{\mathrm{d}\alpha} g_\alpha(AA^*)\overline{r}\|}.$$

Choose the regularization parameter $\alpha = \alpha_{\mathrm{ME}}$ as the solution of the equation $d_{\mathrm{ME}}(\alpha) = C\delta$. If the function $d_{\mathrm{ME}}(\alpha)$ is non-monotone, then choose $\alpha = \alpha_{\mathrm{ME}}$ as the largest solution of this equation. In case of exactly known noise level the best value of C is $C = 1$.

The name of the monotone error rule is justified by the property (Theorem 20)

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\|x_\alpha - x_*\| \geq 0 \quad \text{for all} \quad \alpha \in [\alpha_{\mathrm{ME}}, \infty). \tag{3.2}$$

Therefore, if $\alpha \in [\alpha_{\mathrm{ME}}, \infty)$, then $\|x_\alpha - x_*\| \geq \|x_{\alpha_{\mathrm{ME}}} - x_*\|$. This means that the optimal parameter α_* always satisfies $\alpha_* \leq \alpha_{\mathrm{ME}}$.

In m-iterated Tikhonov method we have due to (2.15) and (2.8)

$$d_{\mathrm{ME}}(\alpha) = \frac{(r_{m;\alpha}, r_{m+1;\alpha})}{\|r_{m+1;\alpha}\|} = \frac{\|B_\alpha r_{m;\alpha}\|^2}{\|B_\alpha^2 r_{m;\alpha}\|},$$

which is a monotone function of α. Monotone error rule in m-iterated Tikhonov method is order optimal for full range $p \leq 2m$ (see Theorem 3).

It is interesting to note that in case of asymptotical regularization we have $\frac{\mathrm{d}}{\mathrm{d}\alpha} g_\alpha(AA^*)\overline{r} = -\frac{1}{\alpha^2} r_\alpha$ and therefore $d_{\mathrm{ME}}(\alpha) = d_{\mathrm{D}}(\alpha)$.

In case of iterative methods of the form

$$x_{n+1} = x_n - A^* z_n, \qquad z_n \in Y \qquad (n = 0,\ 1,\ \ldots) \tag{3.3}$$

let

$$d_{\mathrm{ME}}(n) = \frac{(r_n + r_{n+1}, z_n)}{2\|z_n\|}. \tag{3.4}$$

Choose the regularization parameter $n = n_{\mathrm{ME}}$ as the first n for which $d_{\mathrm{ME}}(n) \leq \delta$.

This choice guarantees (Theorem 19) the monotonicity property

$$\|x_n - x_*\| \leq \|x_{n-1} - x_*\| \qquad \text{for } n = 1,\ \ldots,\ n_{\mathrm{ME}},$$

so $\|x_{n_{\mathrm{ME}}} - x_*\| \leq \|x_n - x_*\|$ for all $n = 0, 1, \ldots, n_{\mathrm{ME}}$. Also always $n_* \geq n_{\mathrm{ME}}$.

Of the form (3.3) are for example Landweber method with $z_n = \mu(Ax_n - y)$, the implicit iteration scheme with $z_n = (\alpha I + AA^*)^{-1}(Ax_n - y)$, and the methods CGLS and CGME with $z_n = y_n w_n$. This variant of the monotone error rule was proposed and studied for simple iteration methods in [26,36] and for methods CGLS and CGME in [27]. In Landweber method, in implicit iteration scheme and in some other methods of the form (3.3) $n_{\mathrm{D}} - 1 \leq n_{\mathrm{ME}} \leq n_{\mathrm{D}}$ (see [26,36]).

Sometimes a sequence of approximations of the form $x_n = A^* z_n$ ($n = 0, 1, \ldots$) arises during the computations. Since $x_{n+1} = x_n - A^*(z_n - z_{n+1})$, this sequence can be viewed as a special case of previously considered iterative regularization methods. Therefore the monotone error rule with

$$d_{\mathrm{ME}}(n) = \frac{(r_n + r_{n+1}, z_n - z_{n+1})}{2\|z_n - z_{n+1}\|} \tag{3.5}$$

guarantees monotone decrease of error for $n = 0,\ 1,\ \ldots,\ n_{\mathrm{ME}}$. In many regularization methods the elements z_n here can be computed separately and x_n can be found as $x_n = A^* z_n$.

This variant of monotone error rule can be applied for example to the sequence of Tikhonov approximations x_{α_n} with decreasing parameters $\alpha_1 >$

$\alpha_2 > \dots$ (this sequence arises, for example, when choosing the regularization parameter according to the balancing principle) or to the extrapolated Tikhonov approximations with fixed number of terms and varying $\alpha = \alpha_n$ or to the extrapolated Tikhonov approximations with varying number n of terms or to some other arbitrary sequence of extrapolated approximations.

In ordinary Tikhonov method ($m = 1$) we can replace the discrepancy of iterated approximation by the discrepancy of extrapolated approximation, as we did in modified discrepancy principle, and choose the parameter $\alpha = \alpha_{\mathrm{eME}}$ from equation $d_{\mathrm{eME}}(\alpha) = \delta$, where $d_{\mathrm{eME}}(\alpha) = \dfrac{(r_\alpha, r_{\alpha,q\alpha})}{\|r_{\alpha,q\alpha}\|}$. In [28] it is proved that $\|x_{\alpha_{\mathrm{eME}}} - x_*\| \to 0$ as $\delta \to 0$ and under assumption (2.2) the error estimate (3.1) holds for $p \le 2$ (cf Theorem 15). Generally, if m is the number of terms in extrapolated Tikhonov approximation and $\alpha_n = q_n\alpha$ with fixed q_n ($n = 1, \dots, m+1$), then for $x_\alpha = x_{\alpha_1,\dots,\alpha_m}$ we can choose the parameter $\alpha = \alpha_{\mathrm{eME}}$ from the equation $d_{\mathrm{eME}}(\alpha) = C\delta$, where

$$d_{\mathrm{eME}}(\alpha) = \frac{(r_{\alpha_1,\dots,\alpha_m}, r_{\alpha_1,\dots,\alpha_m,\alpha_{m+1}})}{\|r_{\alpha_1,\dots,\alpha_m,\alpha_{m+1}}\|}, \tag{3.6}$$

or, if it is assumed that $\alpha = q^n$ for some $n = 0, 1, \dots$, as the first $\alpha = q^n$ satisfying $d_{\mathrm{eME}}(q^n) \le C\delta$.

In extrapolated Tikhonov method with given sequence of parameters $\alpha_1 \ge \alpha_2 \ge \dots$ let $x_n = x_{\alpha_1,\dots,\alpha_n}$ and $d_{\mathrm{ME}}(n) = (r_n + r_{n+1}, r_{n+1})/(2\|r_{n+1}\|)$. In [28] it is proved that $d_{\mathrm{ME}}(n)$ is monotonically decreasing. If in addition $\sum_{n=1}^{\infty} \alpha_n^{-1} = \infty$, $\alpha_n^{-1} \le \sum_{i=1}^{n-1} \alpha_i^{-1}$, then there exists n for which $d_{\mathrm{ME}}(n) \le C\delta$ and if n_{ME} is the first n for which $d_{\mathrm{ME}}(n) \le C\delta$, then for x_n with $n = n_{\mathrm{ME}}$ there holds $\|x_n - x_*\| \to 0$ as $\delta \to 0$. In case of source-like solutions (2.2) also the error estimate (3.1) is valid for all $p > 0$. If $d_{\mathrm{D}}(n)$ is the discrepancy of the extrapolated approximation x_n, then $d_{\mathrm{D}}(n+1) \le d_{\mathrm{ME}}(n) \le d_{\mathrm{D}}(n)$, so the stopping indices n_{D} and n_{ME}, found by the discrepancy principle and the monotone error rule, respectively, satisfy $n_{\mathrm{D}} - 1 \le n_{\mathrm{ME}} \le n_{\mathrm{D}}$ (see Theorem 13).

Analogs of monotone error rule for Lavrentiev method. In our paper [35] an attempt was made to formulate the monotone error rule for Lavrentiev method. Unfortunately this has not succeeded but several analogs of the monotone error rule were proposed, which gave good results in numerical tests. Assume now $X = Y$, $A = A^* \ge 0$.

Straightforward is the analog of ME rule with function

$$d_{\mathrm{MEa}}(\alpha) = \frac{\|B_\alpha^{m+1} y\|^2}{\|B_\alpha^{m+2} y\|} = \frac{\|r_{m+1;\alpha}\|^2}{\|r_{m+2;\alpha}\|}.$$

As proved in [35], the function $d_{\mathrm{MEa}}(\alpha)$ is monotonically increasing with $d_{\mathrm{MEa}}(0) = \|Qy\|$ and $\lim\limits_{\alpha\to\infty} d_{\mathrm{MEa}}(\alpha) = \|y\|$, where Q is the orthoprojector of Y onto $\overline{\mathcal{R}(A)}$. Therefore if $\|Qy\| \le C\delta \le \|y\|$, then the equation

$d_{\mathrm{MEa}}(\alpha) = C\delta$ has a unique solution α_{MEa}. Under additional assumptions $\|r_{m+2,\alpha}\| \geq \|B_\alpha^{m+2}(y - y_*)\|$ for all $\alpha \geq \alpha_{\mathrm{MEa}}$ and $C \geq 2$ the convergence $\|x_{\alpha_{\mathrm{MEa}}} - x_{\alpha_*}\| \to 0$ as $\delta \to 0$ has been proven and for source-like solutions (2.2) the error estimate (3.1) has been shown to hold with $p \leq m$ (see Theorem 11). Numerical experiments suggest that with high probability the order optimality of rule MEa also holds without the additional assumption $\|r_{m+2,\alpha}\| \geq \|B_\alpha^{m+2}(y - y_*)\|$.

Unlike the monotone error rule in Tikhonov method, the rule MEa does not guarantee monotonicity of the error for $\alpha \geq \alpha_{\mathrm{MEa}}$.

Instead of MEa, we can use an analog of that rule, which can be derived as follows. Since by (2.10), (2.15), and (2.9)

$$d_{\mathrm{MEa}}(\alpha) = \frac{\|K_{m+1;\alpha}\overline{r}\|^2}{\|K_{m+2;\alpha}\overline{r}\|} = \frac{\alpha^4}{m^2} \cdot \frac{(m+1)}{\alpha^2} \cdot \frac{\left\|\frac{\mathrm{d}}{\mathrm{d}\alpha} g_{m;\alpha}(A)\overline{r}\right\|^2}{\left\|\frac{\mathrm{d}}{\mathrm{d}\alpha} g_{m+1;\alpha}(A)\overline{r}\right\|} = \alpha^2 \frac{m+1}{m^2} \frac{\left\|\frac{\mathrm{d}}{\mathrm{d}\alpha} x_{m;\alpha}\right\|^2}{\left\|\frac{\mathrm{d}}{\mathrm{d}\alpha} x_{m+1;\alpha}\right\|}$$

and the derivatives can be approximated as

$$\left\|\frac{\mathrm{d}}{\mathrm{d}\alpha} x_{m;\alpha}\right\| \approx \frac{\|x_{m;q\alpha} - x_{m;\alpha}\|}{\alpha(1-q)}, \qquad \left\|\frac{\mathrm{d}}{\mathrm{d}\alpha} x_{m+1;\alpha}\right\| \approx \frac{\|x_{m+1;q\alpha} - x_{m+1;\alpha}\|}{\alpha(1-q)}$$

with $q < 1$, the function $d_{\mathrm{MEa}}(\alpha)$ is approximated by function

$$d_{\mathrm{MEaq}} = \frac{\alpha}{1-q} \cdot \frac{m+1}{m^2} \cdot \frac{\|x_{m;q\alpha} - x_{m;\alpha}\|^2}{\|x_{m+1;q\alpha} - x_{m+1;\alpha}\|}. \tag{3.7}$$

We choose the regularization parameter $\alpha = \alpha_{\mathrm{MEaq}}$ as the solution of equation $d_{\mathrm{MEaq}}(\alpha) = C\delta$ with $C > 1$, or if the parameter is searched from the sequence of parameters $\alpha_n = \alpha_{n-1}q$ with $q < 1$, then as the largest α_n for which $d_{\mathrm{MEaq}}(\alpha_n) \leq C\delta$.

Let us now consider the question, what form the genuine monotone error rule in (iterated) Lavrentiev method should have. For simplicity, take now $\overline{x} = 0$. Let $y \in \mathcal{R}(A)$ and let x be the solution of $Ax = y$. Then $\overline{r} = -y = -Ax$. Using the equalities

$$\frac{\mathrm{d}}{\mathrm{d}\alpha} x_{m;\alpha} = -\frac{\mathrm{d}}{\mathrm{d}\alpha} g_{m;\alpha}(A)\overline{r} = m\alpha^{-2} K_{m+1;\alpha}(A)\overline{r} = -m\alpha^{-2} B_\alpha^{m+1} Ax$$

and assuming $\|y - y_*\| \leq \delta$, we get for m-iterated Lavrentiev method, along the lines of analogous derivation for m-iterated Tikhonov method (see Theorem 20):

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\alpha}\|x_{m;\alpha} - x_*\|^2 &= 2(x_{m;\alpha} - x_*, \frac{\mathrm{d}}{\mathrm{d}\alpha} x_{m;\alpha}) = 2m\alpha^{-2}(x_{m;\alpha} - x_*, -B_\alpha^{m+1}Ax) \\
&= 2m\alpha^{-2}(Ax_{m;\alpha} - Ax_*, -B_\alpha^{m+1}x) = 2m\alpha^{-2}(r_{m;\alpha} + y - y_*, -B_\alpha^{m+1}x) \\
&\geq 2m\alpha^{-2}[(r_{m;\alpha}, -B_\alpha^{m+1}x) - \|y - y_*\|\|B_\alpha^{m+1}x\|] \\
&\geq 2m\alpha^{-2}[(B_\alpha^m y, B_\alpha^{m+1}x) - \delta\|B_\alpha^{m+1}x\|].
\end{aligned}$$

Let

$$d_{\mathrm{ME}}(\alpha) = \frac{(B_\alpha^m y, B_\alpha^{m+1} x)}{\|B_\alpha^{m+1} x\|}. \tag{3.8}$$

If α_{ME} is the solution of $d_{\mathrm{ME}}(\alpha) = \delta$, then $\|x_{m;\alpha} - x_*\|$ is a monotonically increasing function of α for $\alpha \geq \alpha_{\mathrm{ME}}$.

Unfortunately $d_{\mathrm{ME}}(\alpha)$ cannot be computed in practice, since x is unknown in case $y \in \mathcal{R}(A)$ and even non-existent in case $y \notin \mathcal{R}(A)$. Substituting x in (3.8) by various approximate solutions of $Ax = y$, we get approximations of the monotone error rule function (3.8). The more accurate approximations we have for x, the larger is the probability that the resulting rule R yields a parameter α_{R} for which the error is an increasing function (i.e the monotonicity property (3.2) holds) for all $\alpha \geq \alpha_{\mathrm{R}}$.

Replacing x by ordinary Lavrentiev approximation $x_\alpha = \alpha^{-1} B_\alpha y$, we get the function $d_{\mathrm{MEa}}(\alpha)$. However, when finding the approximate solution of $Ax = y$, in contrast to the problem (2.1), we consider this problem as a problem with exact right hand side. Therefore, here a smaller regularization parameter $\nu\alpha$ with $\nu \leq 1$ may be preferred. Replacing x in (3.8) by Lavrentiev approximation $(\nu\alpha)^{-1} B_{\nu\alpha} y$, we get the function

$$d_{\mathrm{MEn}}(\alpha) = \frac{(B_\alpha^m y, B_\alpha^{m+1} B_{\nu\alpha} y)}{\|B_\alpha^{m+1} B_{\nu\alpha} y\|} = \frac{(r_{m;\alpha}, r_{\alpha,\ldots,\alpha,\nu\alpha})}{\|r_{\alpha,\ldots,\alpha,\nu\alpha}\|}, \quad \nu \leq 1,$$

where $r_{\alpha,\ldots,\alpha,\nu\alpha}$ has $m+1$ indices α. For $\nu = 1$ this function coincides with $d_{\mathrm{MEa}}(\alpha)$. If $y \in \mathcal{R}(A)$, then $(\nu\alpha)^{-1} B_{\nu\alpha} y \to x$ as $\nu\alpha \to 0$, so the function $d_{\mathrm{MEn}}(\alpha)$ approximates the function d_{ME} in process $\nu \to 0$. However, very small values of ν cause numerical instabilities; numerical experiments suggest to limit the values of ν to $\nu \geq 10^{-3}$. In tests the best values for non-smooth solutions were found to be around $\nu \approx 0.17$.

Alternatively, x in (3.8) can be replaced by iterated Lavrentiev approximation $x_{k;\alpha}$. Since $g_{k;\alpha}(t) = t^{-1}(1 - (1 + \alpha^{-1}t)^{-k}) = \alpha^{-1} \sum_{j=1}^{k} (1 + \alpha^{-1}t)^{-j}$, we have $x_{k;\alpha} = \alpha^{-1} \sum_{j=1}^{k} B_\alpha^j y$ and the function $d_{\mathrm{ME}}(\alpha)$ gets the form

$$d_{\mathrm{MEk}}(\alpha) = \frac{(r_{m;\alpha}, \sum_{j=1}^{k} r_{m+j+1;\alpha})}{\|\sum_{j=1}^{k} r_{m+j+1;\alpha}\|}.$$

For $k = 1$ this coincides with $d_{\mathrm{MEa}}(\alpha)$. If $y \in \mathcal{R}(A)$, then $x_{k;\alpha} \to x$ as $k \to \infty$, therefore the function $d_{\mathrm{MEk}}(\alpha)$ can be considered as the approximation of the function $d_{\mathrm{ME}}(\alpha)$. Instead of iterated Lavrentiev approximation, x in (3.8) can also be replaced by extrapolated Lavrentiev approximation $x_{\alpha_1,\ldots,\alpha_k}$ to form the function

$$d_{\mathrm{MEke}}(\alpha) = \frac{(r_{m;\alpha}, \sum_{j=1}^{k} r_{\alpha_1,\ldots,\alpha_{m+j+1}})}{\|\sum_{j=1}^{k} r_{\alpha_1,\ldots,\alpha_{m+j+1}}\|},$$

where $\alpha_1, \alpha_2, \ldots$ is some sequence of parameters, for example, $\alpha_n = \alpha_{n-1} q$

($q < 1$) and in addition $\alpha_1 = \alpha$ or the geometric mean of $\alpha_1, \dots, \alpha_{m+j+1}$ is (approximately) α.

In similar way we can modify the function $d_{\mathrm{MEaq}}(\alpha)$. Namely, in $d_{\mathrm{MEa}}(\alpha)$ all iterations are performed with the same parameter α but in $d_{\mathrm{MEn}}(\alpha)$ one iteration uses smaller parameter $\nu\alpha$. In function $d_{\mathrm{MEaq}}(\alpha)$, when choosing the parameter α from the sequence $\alpha_n = \alpha_n q$ with $q < 1$, this means that instead of α we can use the parameter $q^l\alpha$ with $l \in \mathbb{N}$. Replacing α in one side of the scalar product in the numerator $\|x_{m;q\alpha} - x_{m;\alpha}\|^2 = (x_{m;q\alpha} - x_{m;\alpha}, x_{m;q\alpha} - x_{m;\alpha})$ and in denominator by $q^l\alpha$, we get the function

$$d_{\mathrm{MEaql}}(\alpha) = \frac{\alpha}{1-q} \cdot \frac{m+1}{m^2} \cdot \frac{(x_{m;q\alpha} - x_{m;\alpha}, x_{m;q^{l+1}\alpha} - x_{m;q^l\alpha})}{\|x_{m+1;q^{l+1}\alpha} - x_{m+1;q^l\alpha}\|}.$$

In numerical experiments with $q = 1/1.2$ the optimal l was $l = 5$ for non-smooth solutions and $l = 4$ for source-like solutions (2.2) with $p = 1$.

Note that in [20, 21] a rule $d_{\mathrm{GN}}(\alpha) = C\delta$ with

$$d_{\mathrm{GN}}(\alpha) = \frac{(r_\alpha, r_{2;\alpha})}{\|r_{2;\alpha}\|}$$

was proposed.

3.1.4. Rule R1.

Previous rules have the disadvantage that they cannot be used, if the equation $Ax = y$ has only quasisolution (i.e $f \notin \mathcal{R}(A)$ but $Qf \in \mathcal{R}(A)$, where $Q\colon Y \to Y$ is the orthoprojector onto $\overline{\mathcal{R}(A)}$). The following rule was formulated in [77].

In case of continuous regularization methods (2.3) let

$$d_{\mathrm{R1},k}(\lambda) = \lambda^{-k}\big\||A^*|^{2k} B_\lambda^{2k+1} r_\lambda\big\| \qquad (k = 0,\ \frac{1}{2},\ 1,\ \frac{3}{2},\ \dots),$$

where $|A^*| = (AA^*)^{1/2}$. Choose the regularization parameter $\lambda = \lambda_{\mathrm{R1},k}$ as the smallest λ for which $d_{\mathrm{R1},k}(\lambda) = C\delta$. In general the function $d_{\mathrm{R1},k}(\lambda)$ is not monotone and the equation $d_{\mathrm{R1},k}(\lambda) = C\delta$ may have many solutions. The papers [75, 76] study an analog of this rule, where the largest solution of this equation was taken for the regularization parameter (instead of the smallest).

For m-iterated Tikhonov method we get

$$d_{\mathrm{R1},k}(\alpha) = \begin{cases} \alpha^{-k}\big((AA^*)^k r_{m+k;\alpha}, (AA^*)^k r_{m+k+1;\alpha}\big)^{1/2} & (k = 0,\ 1,\ \dots), \\ \alpha^{-k}\|A^*(AA^*)^{k-\frac{1}{2}} r_{m+k+\frac{1}{2};\alpha}\| & (k = \frac{1}{2},\ \frac{3}{2},\ \dots). \end{cases}$$

By induction, using the relations

$$A^* r_{m+i;\alpha} = \alpha(x_{m+i-1;\alpha} - x_{m+i;\alpha}),$$
$$A(x_{m+i-1;\alpha} - x_{m+i;\alpha}) = r_{m+i-1;\alpha} - r_{m+i;\alpha},$$

we can represent $d_{\mathrm{R1},k}(\alpha)$ in the form

$$d_{\mathrm{R1},k}(\alpha) = \begin{cases} \Big(\sum\limits_{i=0}^{k}(-1)^i\binom{k}{i}r_{m+i},\ \sum\limits_{i=0}^{k}(-1)^i\binom{k}{i}r_{m+i+1}\Big)^{1/2} & (k = 0,\ 1,\ \ldots), \\ \alpha^{1/2}\Big\|\sum\limits_{i=0}^{k+1/2}(-1)^i\binom{k+\frac{1}{2}}{i}x_{m+i}\Big\| & (k = \frac{1}{2},\ \frac{3}{2},\ \ldots). \end{cases}$$

From here we get the following practically usable formulas for smaller k:

$$\begin{aligned} d_{\mathrm{R1},0}(\alpha) &= d_{\mathrm{MD}}(\alpha) = (r_{m;\alpha}, r_{m+1;\alpha})^{1/2}, \\ d_{\mathrm{R1},1/2}(\alpha) &= \alpha^{-1/2}\|A^*r_{m+1;\alpha}\| = \alpha^{1/2}\|x_{m;\alpha} - x_{m+1;\alpha}\|, \\ d_{\mathrm{R1},1}(\alpha) &= (r_{m+1;\alpha} - r_{m+2;\alpha}, r_{m;\alpha} - r_{m+1;\alpha})^{1/2}, \\ d_{\mathrm{R1},3/2}(\alpha) &= \alpha^{1/2}\|x_{m;\alpha} - 2x_{m+1;\alpha} + x_{m+2;\alpha}\|. \end{aligned}$$

Note that another formulas for realization of rule R1 and other quasioptimal rules are given in [81].

For self-adjoint problems in case of continuous methods (2.9) let

$$d_{\mathrm{R1},k}(\lambda) = \lambda^{-k}\|A^kB_\lambda^{k+1}r_\lambda\| \qquad (k = 0,\ \frac{1}{2},\ 1,\ \frac{3}{2},\ \ldots).$$

Choose the regularization parameter $\lambda = \lambda_{\mathrm{R1},k}$ as the smallest λ for which $d_{\mathrm{R1},k} = C\delta$.

For m-iterated Tikhonov method and m-iterated Lavrentiev method the constant C is to be chosen to satisfy the inequality $C > \tilde{y}_k^{h(k)}$, where $\tilde{y}_k^s = (y_{k/\tau(s)})^{\tau(s)}$, $\tau(s) = 1 + (s+1)/m$, $h(k) = k - 1/2$ in m-iterated Tikhonov method and $h(k) = k$ in m-iterated Lavrentiev method.

Rule R1 in m-iterated Tikhonov method and in m-iterated Lavrentiev method is order optimal for full range $p \le 2m$ and $p \le m$, respectively (see Theorems 3, 4). The analog of rule R1, where the largest solution of the equation $d_{\mathrm{R1},k}(\alpha) = C\delta$ is taken to be the regularization parameter, guarantees the convergence $\|x_\alpha - x_*\| \to 0$ also in case, where the noise level is known approximately by the condition $\|y - y_*\|/\delta \le c$ as $\delta \to 0$, where c is an unknown constant (see [75, 76], cf. Theorem 10).

We use this rule in Section 3.3 as the first step in two-step rule DM. Rule R1 itself is not included in tables of Chapter IV, since rules ME, R2, Me gave better results.

3.1.5. Rule R2.

In [80] the following rule was proposed for Tikhonov method and for its iterated variant. Let

$$d_{\mathrm{R2}}(\alpha) = \frac{\|A^*r_{m+1;\alpha}\|^2\kappa(\alpha)^s}{\alpha^{1/2}(A^*r_{m+1;\alpha}, A^*r_{m+2;\alpha})},$$

where $\kappa(\alpha) = (1 + \alpha\|A\|^{-2})$, $s \in [0, 1]$. Choose the regularization parameter $\alpha = \alpha_{\mathrm{R2}}$ as the solution of the equation $d_{\mathrm{R2}}(\alpha) = C\delta$. Note that the

influence of the factor $\kappa(\alpha)$ is small for small α but for large α sometimes it is useful. Our numerical tests are always performed with $s = 1$.

As in Rule R1, this equation may have many solutions and both the smallest and the largest are of interest. We propose to take the largest solution; then typically $\alpha_{\mathrm{R2}} \geq \alpha_*$ and a somewhat smaller estimated parameter $\alpha_{\mathrm{R2e}} = b\alpha_{\mathrm{R2}}$ with $b < 1$ is often better.

Note that in order to avoid numerical instability when finding the smallest solution of $d_{\mathrm{R2}}(\alpha) = C\delta$, [80] recommends to find the smallest solution under condition $\alpha \geq (\gamma_*\delta/M)^2$, where M is an upper bound of $\|x_* - \overline{x}\|$.

The bounds for the constant C are not quite obvious. Originally in [80] the rule R2 was proposed with constant $C > \tilde{\gamma}^2_{1/4,m}$, where $\tilde{\gamma}_{k,s} = \dfrac{s^{s/2}(2k)^k}{(s+2k)^{k+s/2}}$. This gives $C > 0.3849$ for $m = 1$ and $C > 0.2862$ for $m = 2$. Based on our numerical tests, we recommend somewhat smaller constants: $C = 0.3$ for $m = 1$, $C = 0.2$ for $m = 2$, and $C = 0.13$ for $m = 3$.

In m-iterated Tikhonov method in case of (2.2) the rule R2 with the smallest solution of equation $d_{\mathrm{R2}}(\alpha) = C\delta$ guarantees the error estimate (3.1) for all $p \leq 2m-1$ and under certain mild additional assumption about the error of y, also for all $p \leq 2m$ (see Theorems 5, 6).

Due to (2.17) the expression of $d_{\mathrm{R2}}(\alpha)$ can be written as

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

In numerical tests this function is preferred because of better computational stability.

In case of extrapolated Tikhonov method let $\alpha_n = q^n\alpha$ ($q < 1$, $n = 0, 1, \ldots$) and

$$d_{\mathrm{eR2}}(\alpha) = \frac{\sqrt{\alpha}\|x_{\alpha_1,\ldots,\alpha_m} - x_{\alpha_1,\ldots,\alpha_{m+1}}\|^2\kappa(\alpha)^s}{(x_{\alpha_1,\ldots,\alpha_m} - x_{\alpha_1,\ldots,\alpha_{m+1}}, x_{\alpha_1,\ldots,\alpha_{m+1}} - x_{\alpha_1,\ldots,\alpha_{m+2}})^{1/2}}.$$

Choose the parameter $\alpha = \alpha_{\mathrm{eR2}}$ as the solution of the equation $d_{\mathrm{eR2}}(\alpha) = C\delta$ or, if it is assumed that $\alpha = q^n$ ($q < 1$; $n = 0, 1, \ldots$), as the first $\alpha = q^n$ for which $d_{\mathrm{eR2}}(q^n) \leq C\delta$.

3.1.6. Balancing principle.

The balancing principle [3, 4, 7–9, 15–18, 50, 54–56, 58–63, 70], originating from work by Lepskii about parameter estimation for stochastic regression, has received much attention in recent years. Sometimes the balancing principle is also called Lepskii type rule. As shown in [34], the balancing principle is closely related to rule R1.

Let $q < 1$ be a constant. Define the sequence of parameters $\lambda_0, \lambda_1, \ldots, \lambda_N$, where $\lambda_0 = \delta^2$, $\lambda_i = \lambda_0/q^i$ for $i = 1, 2, \ldots, N$, and N is the first index i for which $\lambda_i \geq 1$. There are several variants of the balancing principle for methods (2.3), in all of these the regularization parameter $\lambda = \lambda_{\mathrm{BP}}$ is

chosen as the first λ_i for which a certain condition is fulfilled. In [59, 60] this condition is

$$\|x_{\lambda_i} - x_{\lambda_{i+1}}\| > \frac{c\delta}{\sqrt{\lambda_i}} \tag{3.9}$$

and in [70]

$$\exists j \in \{1, \dots, i\}\colon \ \|x_{\lambda_j} - x_{\lambda_{i+1}}\| > \frac{c\delta}{\sqrt{\lambda_i}}. \tag{3.10}$$

In Tikhonov method $c = 2$ was proposed. Instead of a fixed constant c the paper [34] recommends to use c of the forms $c = c'(1-q)/\sqrt{q}$ and $c = c'(1-q^{i+1-j})$ in these formulas, respectively, and also gives some reasonable bounds of c' for different methods (see Theorem 9). In [34] it is also suggested to take $q \in [0.5, 0.9]$, since the values from this interval balance the size of constants in error estimates found in [34] with the amount of computational work needed to reach λ_i for which the condition of the balancing principle is satisfied. For Tikhonov method the balancing principle with the condition (3.9) or (3.10) is order optimal in case of source-like solutions (2.2) for $p \le 2$, if $c' \ge 3\sqrt{3}/16 \approx 0.3248$ or $c' \ge 1$, respectively (Theorems 7, 8).

Instead of (3.10), numerical experiments suggest to use the condition

$$\exists j \in \{i-1, i\}\colon \ \|x_{\lambda_j} - x_{\lambda_{i+1}}\| > \frac{c\delta}{\sqrt{\lambda_i}}, \tag{3.11}$$

since the calculations with (3.10) showed that very often (3.10) was satisfied with $j = i-1$ or $j = i$. This condition also avoids double loop and is therefore significantly faster to check. In Tikhonov method we recommend $c = 0.3(1-q^{i+1-j})q^{(j-i-1)/2}$. Also note that the value $\lambda_0 = \delta^2$ is problematic, since often the optimal λ is less than δ^2.

For methods (2.9) the balancing principle is the same as described above, only the analogs of inequalities (3.9), (3.10), (3.11), have λ_i in denominators, instead of $\sqrt{\lambda_i}$.

In [34] an analog of the balancing principle (3.9) was proposed for methods (2.3), where the largest $\lambda_n = q^n$ ($q < 1$; $n = 0, 1, \dots$) for which

$$\|x_{\lambda_i} - x_{\lambda_{i+1}}\| \le \frac{c\delta}{\sqrt{\lambda_i}}$$

was taken to be the regularization parameter. This rule guarantees convergence $\|x_\lambda - x_*\| \to 0$ for all c as $\delta \to 0$ under assumption that $\|y - y_*\|/\delta$ remains bounded. If $c > \gamma\tilde{\gamma}_{1/2}(1-q)/\sqrt{q}$ with $\tilde{\gamma}_{1/2} = \gamma_{1/(2+2/p_0)}^{1+1/p_0}$, then in case $\|y - y_*\| \le \delta$ for source-like solutions (2.2) this rule also gives order optimal error estimate for $p \le 2p_0$.

3.1.7. Estimated parameters.

Sometimes it is known either theoretically or practically that a parameter choice rule has a bias, choosing typically too large or too small parameter. Therefore it is reasonable to post-estimate the computed parameter,

shifting it towards smaller or larger values. The amount of shift generally depends on the size of the original parameter, so different formulas can be applied. We describe here only the strategies we used to construct the final rules with post-estimated parameters. The rules themselves are formulated in Chapter IV together with results they give on test problems. Names of rules with post-estimated parameters have the letter "e" ("estimation") added to the end of original rules' names.

In Tikhonov method always $\alpha_{\mathrm{ME}} \geq \alpha_{\mathrm{opt}}$ (and often α_{ME} and α_{opt} are not very close). So it is reasonable to decrease the parameters found by ME-rule a little. Assuming that $\alpha_{\mathrm{opt}} < 1$, there exist $k \geq 1$ for which $\alpha_{\mathrm{opt}} = \alpha_{\mathrm{ME}}^k$ and $l \leq 1$ for which $\alpha_{\mathrm{opt}} = l\alpha_{\mathrm{ME}}$. In Tikhonov method our numerical experiments suggested to use the estimated parameter $\alpha_{\mathrm{MEe}} = \min(c_1\alpha_{\mathrm{ME}}, c_2\alpha_{\mathrm{ME}}^{c_3})$ instead of α_{ME}, where $c_1 = 0.53$, $c_2 = 0.6$, $c_3 = 1.06$. The same formula with similar constants c_1, c_2, c_3 gives quite good results also for extrapolated Tikhonov method with 2 or 3 terms.

Our numerical tests showed that in 90% of cases (typically for smaller α's) also $\alpha_{\mathrm{R2}} \geq \alpha_{\mathrm{opt}}$. Here we found that the estimated parameter $\alpha_{\mathrm{R2e}} = 0.5\alpha_{\mathrm{R2}}$ is usually better than α_{R2}.

Comparing the parameters α_{MEe} and α_{R2e}, we found that α_{MEe} is better of the two, if $\|y - y_*\|$ is equal to or slightly less than δ, but α_{R2e} is better, if $\|y - y_*\|$ is considerably less than δ. In both cases $\alpha_{\mathrm{Me}} = \min(\alpha_{\mathrm{MEe}}, \alpha_{\mathrm{R2e}})$ often chooses the best of them.

In Landweber method n_{D} and n_{ME} are close: $n_{\mathrm{D}} - 1 \leq n_{\mathrm{ME}} \leq n_{\mathrm{D}}$, so n_{ME} and mostly 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 $n_{\mathrm{ME}} = \mathrm{round}(cn_{\mathrm{ME}})$ instead of n_{D} and n_{ME}, respectively. The constant $c = 2.3$ is found by numerical experiments.

Also in CGLS $n_{\mathrm{ME}} \leq n_*$, so a larger index is preferred. We take $n_{\mathrm{MEe}} = \mathrm{round}(0.99n_{\mathrm{ME}}^{1.13})$. Also for discrepancy principle $n_{\mathrm{De}} = \mathrm{round}(1.02n_{\mathrm{D}}^{1.03})$ usually gives better results than n_{D}.

3.2. Parameter choice rules for unknown noise level

If the noise level is unknown, then no rule for choosing the regularization parameter α can guarantee the convergence $\|x_\alpha - x_*\| \to 0$ as $\delta \to 0$. Namely, it was shown by Bakushinskii [1] that for ill-posed problems with $\mathcal{R}(A) \neq \overline{\mathcal{R}(A)}$ the worst case error

$$\sup\{\|x_\lambda - x_*\| : y \in Y,\ \|y - y_*\| \leq \delta\}$$

can converge to 0 as δ goes to 0, only if the regularization parameter is chosen depending on the noise level. Nevertheless, some heuristic rules are rather popular, because they often work well in practice and because in applied ill-posed problems the exact noise level is often unknown. Typically these rules minimize (or maximize) certain functions. Well-known heuristic

rules are for example quasioptimality criterion [85, 86] (see also [2, 5, 6, 10, 23, 49, 51, 52]), generalized cross-validation (GCV) [91], L-curve rule [44, 45, 82] (for limitations of L-curve rule see [39, 90]), and Hanke-Raus rule [42].

3.2.1. Known rules.

An overview of heuristic rules can be found in [19, 41]. For our purposes we review some known rules that are similar to the rules that we will derive using strategies of Subsection 3.2.2.

A classical rule in case of unknown noise level is the *quasioptimality criterion*. In m-iterated Tikhonov method or in m-iterated Lavrentiev method let $\varphi_{\mathrm{Q}}(\alpha) = \|x_{m;\alpha} - x_{m+1;\alpha}\|$. Choose the parameter $\alpha = \alpha_{\mathrm{Q}}$ as the global minimizer of the function $\varphi_{\mathrm{Q}}(\alpha)$.

Note that $\varphi_{\mathrm{Q}}(\alpha) = d_{\mathrm{R}1,1/2}(\alpha)/\sqrt{\alpha}$.

Another version of the quasioptimality criterion minimizes the function $\varphi_{\mathrm{Qq}}(\alpha) = \|x_{m;\alpha} - x_{m;q\alpha}\|$, where q is a constant. The rule with this function is very natural to apply, if we choose the regularization parameter on the mesh (α_n), where $\alpha_n = q\alpha_{n-1}$; then the regularization parameter is α_n such that $d_{\mathrm{Qq}}(\alpha_n)$ is minimal.

The quasioptimality criterion can be carried over from m-iterated Tikhonov method to extrapolated Tikhonov method with m terms. Let $\alpha_n = q^n$ and $\varphi_{\mathrm{eQ}}(\alpha_i) = \|x_{\mathrm{seq}(m,q,\alpha_i)} - x_{\mathrm{seq}(m+1,q,\alpha_i)}\|$, where $\mathrm{seq}(k, q, \alpha)$ is a finite sequence of parameters defined by

$$\mathrm{seq}(k, q, \alpha) = (\alpha q^{-\left\lfloor \frac{k}{2} \right\rfloor}, \alpha q^{-\left\lfloor \frac{k}{2} \right\rfloor + 1}, \ldots, \alpha q^{\left\lceil \frac{k}{2} \right\rceil - 1}).$$

For Tikhonov method Neubauer [69] proposed a modification of the quasioptimality criterion: minimize the function $\varphi_{\mathrm{QN}}(\alpha) = \|x_{m;\alpha} - x_{2m;\alpha}\|$ in the interval $[m\sigma_{\min}, 1]$, where $\sigma_{\min}$ is the smallest eigenvalue of discretized version of the operator A^*A (we assume that $\|A\| = 1$).

For class of methods (2.3) this function has the form $\varphi_{\mathrm{QN}}(\lambda) = \|x_\lambda - \tilde{x}_\lambda\|$, where $\tilde{x}_\lambda = g_\lambda(A^*A)A^*Ax_\lambda$.

The *Hanke-Raus rule* [42] in m-iterated Tikhonov method finds the regularization parameter $\alpha = \alpha_{\mathrm{HR}}$ as the global minimizer of the function $\varphi_{\mathrm{HR}}(\alpha) = d_{\mathrm{MD}}(\alpha)/\sqrt{\alpha}$. In Landweber method the Hanke-Raus rule minimizes the function $\varphi_{\mathrm{HR}}(n) = n^{1/2}\|r_n\|$ for $n \geq 1$. The ideas of constructing the Hanke-Raus rule may also be used in TSVD, minimizing $\varphi_{\mathrm{HR}}(n) = \|r_n\|/\sigma_{n+1}$. In CGLS and in CGME the Hanke-Raus rule finds the global minimum of the function $\varphi_{\mathrm{HR}}(n) = \sqrt{\varrho_{n+1}}\,\|r_n\|$, where (ϱ_n) is a sequence constructed as follows: starting with $\kappa_{-1} = 0$, $\varrho_0 = 0$, compute $\kappa_n = 1 + \beta_n\kappa_{n-1}$, $\varrho_{n+1} = \varrho_n + \gamma_n\kappa_n$ $(n = 0, 1, \ldots)$.

In [13] it was proposed to minimize the function $\varphi_{\mathrm{BRS}}(n) = \|r_n\|^2/\|A^*r_n\|$ in iteration methods and $\varphi_{\mathrm{BRS}}(\alpha) = \|r_\alpha\|^2/(\alpha\|x_\alpha\|)$ in Tikhonov method.

In Regińska's version [82] of the well-known *L-curve rule* the regularization parameter is chosen as the minimizer of the function $\varphi_{\mathrm{L}}(\lambda) =$

$\|r_\lambda\|\|x_\lambda\|$. In our experiments the results with this rule were worse than those with rules presented in tables. We used the function $\varphi_L(\lambda)$ in TSVD for constructing rules HRL' and HRLm'.

3.2.2. Strategies for constructing new rules.

The main problem with many heuristic rules is that the global minimums of functions that these rules minimize occur at very small λ, leading to very large error. The reason of this is that the discretized versions (as used in computing) of several of these functions tend to zero as $\lambda \to 0$, due to Banach-Steinhaus theorem. For example, if the discretized problem has a unique solution, then the finite-dimensional analogs of $\varphi_{HR}(\lambda) = d_{MD}(\lambda)/\sqrt{\lambda}$ converge to 0 as $\lambda \to 0$ [42], this also holds for other functions $\varphi(\lambda) = d(\lambda)/\sqrt{\lambda}$, where $d(\lambda)$ is a function from previously considered rules that use noise level. So the global minimizer of these functions is $\lambda = 0$. Therefore, in practice it is important to bound the minimization interval from below.

In m-iterated Tikhonov method the papers [49] and [69] propose to find the minimizer of the quasioptimality function and the minimizer of a certain analog of the quasioptimality function on the interval $[\lambda_{min}, 1]$, where $\lambda_{min} = m\sigma_{min}$ and σ_{min} is the smallest eigenvalue of discretized analog of A^*A. It was noticed that the function $\varphi_{QN}(\alpha)$ is monotonically increasing for $\lambda < \lambda_{min}$ and has a large maximum near λ_{min}. However, our numerical experiments showed that this lower bound is sometimes too small (see column QN of Table 16 in Chapter IV). On the other hand, numerical experiments suggest that for problems with small condition number the reasonable bound should be much smaller. For example, we obtained better results with the lower bound $\alpha_{min} = (m\lambda_{min})^3$ in problems with $\mathrm{cond}(A) \le 10^6$.

In TSVD, when solving discrete $N \times N$ problems, we used $\lambda_{min} = \sigma_M$, with some $M \le N$, in several rules.

The problem of finding an appropriate *a priori* lower bound λ_{min} for λ's is hard. In general, instead of an *a priori* lower bound some *a posteriori* lower bound determined during computations may be more promising.

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

1. Climbing approach. Stop the computation at the point, where the value of a function $\varphi(\lambda)$ has become C times larger than its current minimum value. Take α at which the function has this minimum value as the regularization parameter. Suitable values of C for functions that we used are around 4 in Tikhonov method, 20–50 in Landweber method, 10–20 in CGLS, and 100 in CGME.

2. First local minimum. Stop the computation at the first local minimum of $\varphi(\lambda)$. In some functions $\varphi(\alpha)$, for example, in functions used in rule R2 or in quasioptimality criterion, the first local minimizer is too large, in this case often the first local minimum of the function $\varphi(\alpha)\alpha^c$ with $c \approx 1/3$ suits.

The climbing approach may also be used in TSVD method but here we preferred an alternative approach, stopping the computations at some fixed $n = M$ with $M \le N$.

In [31] we present theoretical and numerical arguments in favor of the approach that several heuristic rules can be viewed and new rules be formulated as partners of order optimal rules. Namely, many previously considered rules choose the regularization parameter in methods (2.3) from condition $d(\lambda) \approx \delta$. Under assumption $\overline{x} - x_* = (A^*A)^{p/2}w$, $\|w\| = \omega$ (and maybe, for example in case of rule R2 [80], under certain additional assumptions) the inequalities

$$\|x_\lambda - x_*\| \le \text{const}\,(\lambda^{p/2}\omega + \lambda^{-1/2}\delta), \quad \frac{d(\lambda)}{\sqrt{\lambda}} \le \text{const}\,(\lambda^{p/2}\omega + \lambda^{-1/2}\delta) \tag{3.12}$$

hold true for $0 \le p \le 2p_0$ (except in discrepancy principle, for which $0 \le p \le 2p_0 - 1$). The same estimates for $\|x_\lambda - x_*\|$ and $d(\lambda)/\sqrt{\lambda}$ motivate the following heuristic parameter choice rule. Let $\varphi(\lambda) = d(\lambda)/\sqrt{\lambda}$. Choose the regularization parameter λ as the minimizer of the function $\varphi(\lambda)$. Thus, if in case of known δ a rule chooses the regularization parameter λ as the solution of the equation $d(\lambda) = \delta$, then in case of unknown δ the parameter λ that minimizes the function $d(\lambda)/\sqrt{\lambda}$ may be a reasonable choice. This approach was used already in [42], where on base of the modified discrepancy principle (order optimal rule), the Hanke-Raus rule (heuristic rule) was derived.

The estimates (3.12) are minimized by $\lambda = (\delta/\omega)^{2/(p+1)}$. It has been proved in [31] (see Theorem 17) that under assumption (2.2) the approximate solution x_λ in considered methods satisfies the error estimate

$$\|x_\lambda - x_*\| \le \text{const}\left(1 + \frac{\|y - y_*\|}{\varphi(\lambda_0)\sqrt{\lambda}}\right)\omega^{1/(p+1)}\Delta_\lambda^{p/(p+1)}, \tag{3.13}$$

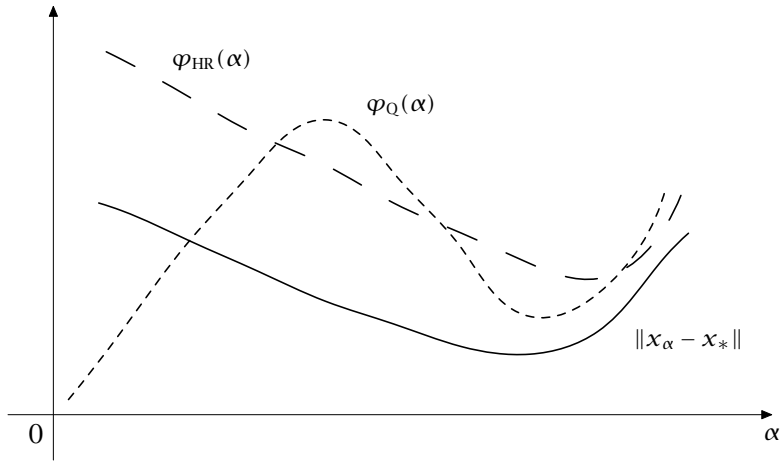
where $\Delta_\lambda = \max\{d(\lambda), C\|y - y_*\|\}$, $\varphi(\lambda) = d(\lambda)/\sqrt{\lambda}$ with $d(\lambda)$ from rules D, MD, ME, R1, R2, and $\lambda_0 = (\|y - y_*\|/\omega)^{2/(p+1)}$. To minimize this error estimate, $d(\lambda)/\lambda^{\frac{p+1}{2p}}$ should be minimized in region $d(\lambda) \le \text{const}\,\|y - y_*\|$; for large p the function to be minimized is close to $\varphi(\lambda)$.

Also, if λ is the global minimizer of $\varphi(\lambda)$, then a part of the error estimate (3.13) can be further estimated as

$$\frac{\|y - y_*\|}{\varphi(\lambda_0)\sqrt{\lambda}} = \frac{\varphi(\lambda)}{\varphi(\lambda_0)}\,\frac{\|y - y_*\|}{d(\lambda)} \le \frac{\|y - y_*\|}{d(\lambda)}.$$

Therefore, for the global minimizer of $\varphi(\lambda)$ the error estimate (3.13) is of optimal order, if $d(\lambda)$ is of the same order as $\|y - y_*\|$. In practice often

Figure 1. Typical behavior of functions $\varphi_{\mathrm{Q}}(\alpha)$ and $\varphi_{\mathrm{HR}}(\alpha)$ in Tikhonov method.



the function $d(\lambda)$ of an order optimal rule $d(\lambda) \approx \delta$ stays near the level $\delta \approx \|y - y_*\|$ after reaching this level.

On the other hand, if in error estimate (3.13) the inequality $d(\lambda) \geq \|y - y_*\|$ holds, then this estimate is of order $d(\lambda)^{p/(p+1)}$, which usually does not cause problems, since $d(\lambda)$ is a function that decreases to the level δ. But if $d(\lambda) \ll \|y - y_*\|$, then the error may be very large.

These arguments can also be applied to self-adjoint case, where we use the function $\varphi(\lambda) = d(\lambda)/\lambda$.

For CGLS in [42] on base of estimates [68]

$$\|x_n - x_*\| \leq \mathrm{const}\,(\varrho_n^{-p/2}\omega + \varrho_n^{1/2}\delta), \quad \|r_n\| \leq \mathrm{const}\,(\varrho_n^{-p/2-1/2}\omega + \delta)$$

the stopping index $\operatorname{argmin}\{\max\{1, \varrho_n\}^{1/2}\|r_n\|\}$ was proposed.

If two functions have different behaviors, then we can combine them to make use of desirable properties of both functions to choose the regularization parameter. For example, in (iterated) Tikhonov method the values of functions $\varphi_{\mathrm{Q}}(\alpha)$ and $\varphi_{\mathrm{R2}}(\alpha)$ are sometimes very small, if α is small, so the global minimum lies at too small α; on the other hand, the values of functions $\varphi_{\mathrm{HR}}(\alpha)$ and $\varphi_{\mathrm{BRS}}(\alpha)$ are much larger at small α, though the global minimizers of these function tend to be larger than the optimal α. Typical behavior of the functions $\varphi_{\mathrm{Q}}(\alpha)$ and $\varphi_{\mathrm{HR}}(\alpha)$ in Tikhonov method is illustrated by Figure 1. In rule QHR we choose the regularization parameter as the local minimum of $\varphi_{\mathrm{Q}}(\alpha)\kappa(\alpha)$ for which $\varphi_{\mathrm{HR}}(\alpha)$ is minimal. In rule HR2 we combine the functions $\varphi_{\mathrm{HR}}(\alpha)$ and $\varphi_{\mathrm{R2}}(\alpha)$ in such way that in the region, where $\varphi_{\mathrm{R2}}(\alpha)$ and $\varphi_{\mathrm{HR}}(\alpha)$ have similar values, the function $\varphi_{\mathrm{R2}}(\alpha)$ dominates, and in region, where $\varphi_{\mathrm{R2}}(\alpha)$ is significantly smaller than $\varphi_{\mathrm{HR}}(\alpha)$, the function $\varphi_{\mathrm{HR}}(\alpha)$ dominates. More precisely, the regularization parameter $\alpha = \alpha_{\mathrm{HR2}}$ is found as the global argmin of

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

In numerical tests the rule with this function gives good results but the best results with this approach were obtained with the function $\varphi_{\mathrm{BR2},\tau}(\alpha)$, which is similar to $\varphi_{\mathrm{HR2},\tau}(\alpha)$, except that it uses the function $\varphi_{\mathrm{BRS}}(\alpha)$ instead of $\varphi_{\mathrm{HR}}(\alpha)$.

These functions can also be used to choose the regularization parameter as a certain local minimizer of some other functions as in rules QHR2 and QHR2 (see Chapter IV).

Choosing the proper local minimizer in Tikhonov method can be assisted by the error estimate $\|x_\lambda - x_*\| \le (1 + W(\lambda)) \inf_{\lambda>0} \|\hat{x}_\lambda - x_*\|$, where

$$W(\lambda) = \sup_{\lambda_1 \le \mu \le \lambda_2} \frac{\mu \|x_\lambda - x_\mu\|}{\gamma \|A^* B_{\xi\mu}^2 r_\mu\|}$$

and $\hat{x}_\lambda$ is found by (2.3) with $\hat{y}$ instead of y, where $\hat{y}$ is an arbitrary element satisfying $\|Q_\eta(\hat{y} - y_*)\| = \|Q_\eta(y - y_*)\|$ and $(Q_\eta A(\overline{x} - x_*), \hat{y} - y_*) \ge 0$ $(0 \le \eta \le \|A^*A\|)$, Q_η is the spectral family of the operator AA^*, and $[\lambda_1, \lambda_2]$ is the interval containing the global minimizer of the function $f(\lambda) = \|\hat{x}_\lambda - x_*\|$ (see Theorem 18, also for values of ξ).

For Lavrentiev method an analog of $W(\lambda)$ is

$$W(\lambda) = \sup_{\lambda_1 \le \mu \le \lambda_2} \frac{\mu \|x_\lambda - x_\mu\|}{\gamma \|B_\mu r_\mu\|},$$

for CGLS the function

$$W(n) = \sup_{m \in M} \frac{\|x_m - x_n\|}{\varrho_{n+1} \|r_n\|}$$

may be used, where M is either $\{1, 2, \ldots, n_{\max}\}$ or a subset of this set (for computational reasons), for example M containing points 1, 2, ..., 10 and differences of other points form an arithmetic progression. Using this function could improve results in CGLS (Rule HRmWC in Chapter IV).

In TSVD method the function $\varphi_{\mathrm{Q}}(n) = \|x_n - x_{n+1}\|$ to be minimized in quasioptimality criterion, is oscillating in some problems and then only maximums of this function are near to the error of the approximate solution. We modified this function to $\varphi_{\mathrm{Qa}}(n) = \big(\sum_{i=n-2}^{n+2} \|x_n - x_i\|^2\big)^{1/2}$. The last function does not oscillate but sometimes it is smaller than both $\varphi_{\mathrm{Q}}(n)$ and the error. Therefore, instead of $\varphi_{\mathrm{Qa}}(n)$ we minimized the maximum of $\varphi_{\mathrm{Q}}(n)$ and $\varphi_{\mathrm{Qa}}(n)$.

In Landweber method, in TSVD, and in CGLS the Hanke-Raus rule minimizes the functions $\sqrt{n}\,\|r_n\|$, $\|r_n\|/\sigma_{n+1}$ and $\varrho_{n+1}\|r_n\|$, respectively. In these methods the Hanke-Raus rule tends to choose too small stopping index. With this reason, we modified the Hanke-Raus rule, subtracting the discrepancy at a later iteration step from the discrepancy at step n, so in Landweber method we minimized $\sqrt{n}(\|r_n\| - \|r_{2n+100}\|)$, in TSVD $(\|r_n\| - \|r_{1.5n+8}\|)/\sigma_{n+1}$, and in CGLS $\varrho_{n+1}(\|r_n\| - \|r_{2n+10}\|)$. In TSVD the

value of $\|r_{1.5n+8}\|$ for odd n was found as $\sqrt{\|r_{1.5n+7.5}\|\|r_{1.5n+8.5}\|}$. Analogous modifications were applied to BRS rule, which also tends to choose too small stopping index. These ideas can be useful in other functions as well, for example we modified the function $\varphi_{\mathrm{QN}}(n) = \|x_n - x_{2n}\|$ in Landweber method to $\varphi_{\mathrm{QNm}}(n) = \|x_n - x_{2n+100}\|$.

Besides minimization of some function one may use the observation that several monotone functions attain a certain level ("plateau") around the optimal stopping index λ_* and after that do not change much. This approach gave good results in conjugate gradient type methods. Examples of such functions are

$$(\sum_{i=0}^{n} \|x_i\|^{-2})^{-1/2}, \qquad d_{\mathrm{DH}}(n) = (\sum_{i=0}^{n} \|r_i\|^{-2})^{-1/2}, \qquad (\sum_{i=0}^{n-1} \|r_i - r_{i+1}\|^{-2})^{-1/2},$$

which are monotonically decreasing even if $\|x_n\|$ (in CGLS) and $\|r_n\|$ (in CGME) may not be. In rule DHP, used in CGME, we choose the stopping index n_{DHP} as the first n for which the function $d_{\mathrm{DH}}(n)$ decreases in next 10 steps by no more than 1.5 times.

3.3. Parameter choice rules for approximate noise level

Up to now we have considered cases, where the exact upper bound of the noise level is given or no information about the noise level is known at all. In some situations the noise level may be known approximately: instead of $\|y - y_*\| \le \delta$ it holds $\|y - y_*\|/\delta \le \mathrm{const}$ as $\delta \to 0$. Here the constant on the right hand side may be greater than 1, so the actual noise level may be underestimated. In this case usual rules for choosing the regularization parameter lead to divergence of approximate solutions.

The first rules (analogs of R1, where the largest λ is taken as the regularization parameter, instead of the smallest) for choice of the regularization parameter λ, guaranteeing $x_\lambda \to x_*$ as $\delta \to 0$ under assumption that $\|y - y_*\|/\delta$ remains bounded, were proposed in [75, 76]. The same convergence is valid also for the analog of the balancing principle, proposed in [34] (see the end of the Section 3.1.6).

The rules of [75, 76] guarantee convergence but better error estimates can be obtained, adding the second step. The corresponding rule was proposed in [33].

We now formulate a rule for many methods that minimizes a certain function in interval $[\underline{\lambda}, \overline{\lambda}]$, where the lower endpoint $\underline{\lambda}$ is found from another parameter choice rule that uses the supposed noise level δ. These rules use two-step strategy: 1) using approximate noise level δ, find $\underline{\lambda}$; 2) minimize certain function in $[\underline{\lambda}, \overline{\lambda}]$. We denote this rule by DM, where D refers to δ and M refers to minimization.

Rule DM for Tikhonov method finds the parameter $\alpha = \alpha_{\mathrm{DM}}$ according to the following two-step strategy. 1) Find $\underline{\alpha}$ as the maximal solution of

$d_{\mathrm{R1},1/2}(\alpha) = c_1\delta$. 2) Find $\alpha = \alpha_{\mathrm{DM}}$ as the minimizer of $\varphi_{\mathrm{R2}}(\alpha)\alpha^{c_2}$ on $[\underline{\alpha}, 1]$.

This rule can be generalized to extrapolated Tikhonov approximation with fixed number m of terms and with parameters $\alpha_n = q^n\alpha$ (q fixed; $n = 1, \ldots, m$). To this end replace $d_{\mathrm{R1},1/2}(\alpha)$ by $d_{\mathrm{eR1},1/2}(\alpha)$ in 1) and $\varphi_{\mathrm{R2}}(\alpha)$ by $\varphi_{\mathrm{eR2}}(\alpha)$ in 2).

For ordinary ($m = 1$) and extrapolated ($m > 1$) Tikhonov approximation x_α the choice of α from the rule DM guarantees convergence $\|x_\alpha - x_*\| \to 0$, as $\delta \to 0$, if $\lim \|y - y_*\|/\delta \le C$. In case of source-like solutions (2.2) the error estimate (3.1) holds with $p \le 2m$, if c_1 is large enough (theoretical bound 0.24).

Practical values of constants c_1, c_2 can be determined by computation and they are much smaller: we found that $c_1 = 0.001 \ldots 0.02$ and $c_2 = 0.03 \ldots 0.14$ give the smallest error.

Rule DM for m-iterated Tikhonov method ($m > 1$). Find $\underline{\alpha}$ as the maximal solution of $d_{\mathrm{R1},1/2}(\alpha) = c_1\delta$ with $c_1 = (m+1)^{m+1}/m^m$. 2) Fix $c_2 \in (0,1)$, $q \in (0,1)$ and find α_{DM} as the minimizer of $\varphi_{\mathrm{Qq}}(\alpha)\kappa(\alpha)^{1/(2m)}/\alpha^{c_2/2}$ in $[\underline{\alpha}, 0.4m + 0.6]$. If the equation in 1) has no solutions, then let $\underline{\alpha}$ be the largest local minimum of $d_{\mathrm{R1},1/2}(\alpha)$.

Rule DM for m-iterated Lavrentiev method ($m \ge 1$). 1) Find $\underline{\alpha}$ as the minimal solution of $(r_{\alpha,1}, Ar_{\alpha,2})/\sqrt{\alpha} = c_1\delta$. 2) Fix $c_2 \in (0,1)$, $q \in (0,1)$ and find $\alpha = \alpha_{\mathrm{DM}}$ as the minimizer of $\varphi_{\mathrm{Qq}}(\alpha)\kappa(\alpha)^{0.005}\alpha^{c_2}$ on $[\underline{\alpha}, m]$. We used $c_1 = 2.5 \cdot 10^{-6}$, $c_2 = 0.25$.

Rule DM for Landweber method. 1) Find N as the first $n \ge 1$, for which $\sqrt{n}\|A^*r_n\| \le c_1\delta$ with $c_1 = 1/\sqrt{2\mu e}$. 2) Fix $c_2 \in [0, 1/2]$ and choose $n = n_{\mathrm{DM}}$ as the minimizer of $n^{c_2}\|r_n\|$ on $[1, N]$. In self-adjoint problems the rule DM is as follows. 1) Find N as the first $n \ge 1$ for which $\sqrt{n}(Ar_n, r_n) \le c_1\delta$ with $c_1 = 1/\sqrt{2\mu e}$. 2) Fix $c_2 \in [0, 1]$ and choose $n = n_{\mathrm{DM}}$ as the minimizer of $n^{c_2}\|r_n\|$ on $[1, N]$.

Rule DM for CGLS. 1) Find N as the first n for which $\sqrt{\varrho_{n+1}}\,\|A^*r_n\| \le c_1\delta$. 2) Fix $c_2 \in (0, 1/2)$ and find $n = n_{\mathrm{DM}}$ as the minimizer of $\varrho_{n+1}^{c_2}(\|r_n\| - \|r_{2n+1}\|)$ on $[1, N]$ (we suggest $c_1 = 0.25$, $c_2 = 0.4$).

3.4. Theoretical results

Here we review some theoretical results concerning the choice of the regularization parameter λ in methods of the forms (2.3) and (2.9). Theorems 1–10 are known results, Theorems 11–18 are proved in our papers; for longer proofs we refer to corresponding papers.

From (2.4), (2.6) follows the error estimate $\|x_\lambda - x_*\| \le \Psi(\lambda)$ with

$$\Psi(\lambda) = \begin{cases} \|K_\lambda(A^*A)(\overline{x} - x_*)\| + y_*\delta/\lambda^{1/2} & \text{for method (2.3),} \\ \|K_\lambda(A)(\overline{x} - x_*)\| + y\delta/\lambda & \text{for method (2.9).} \end{cases}$$

In this estimate the first term dominates for large λ and the second term

for small λ. The quality of a parameter choice rule can be characterized by the following quasioptimality property.

Definition 1. *[78] A rule R for* a posteriori *choice of the regularization parameter $\lambda = \lambda_R$ is called* quasioptimal *(or* weakly quasioptimal*) if there exists a constant C (which does not depend on A, x_*, y) such that for each $y \in Y$, $\|y - y_*\| \le \delta$ there holds the error estimate*

$$\|x_{\lambda_R} - x_*\| \le C \inf_{\lambda \ge 0} \Psi(\lambda) + \mathcal{O}(\delta). \tag{3.14}$$

The error estimate (3.14) is satisfied, if the following stronger error estimate holds:

$$\|x_{\lambda_R} - x_*\| \le C' \inf_{\lambda \ge 0} \Psi(\lambda). \tag{3.15}$$

Theorem 1. *[78] Let $y \in \mathcal{R}(A)$, $\|y - y_*\| \le \delta$. Let the generating function $g_\lambda(t)$ in approximation* (2.3) *satisfy the conditions* (2.5) *with $p_0 = \infty$, $\gamma_0 = 1$,* (2.6)*, and* $\sup_{0 \le \lambda \le \|A^*A\|} t g_\lambda(t) \le 1$ *with* $\lim_{\lambda \to 0} t g_\lambda(t) = 1$ *($0 \le t \le \|A^*A\|$). Assume further that the function $\lambda \to g_\lambda(t)$ is continuous, nonnegative and monotonically increasing, the function $\lambda \to g_\lambda(t)/\lambda$ is monotonically decreasing and* $\frac{\mathrm{d}}{\mathrm{d}\lambda} g_\lambda(t) \le \gamma \hat{y}(1 - t g_\lambda(t))$, $\hat{y} = const$ *($0 \le t \le \|A^*A\|$, $\lambda \ge 0$). Then the rule D in method* (2.3) *is quasioptimal. If the generating function $g_\lambda(t)$ in approximation* (2.9) *satisfies the self-adjoint analogs of above conditions (with* (2.4) *instead of* (2.6) *and A instead of A^*A), then the rule D in method* (2.9) *is quasioptimal.*

Theorem 2. *[78] Let $y \in \mathcal{R}(A)$, $\|y - y_*\| \le \delta$. Let the generating function $g_\lambda(t)$ in approximation* (2.3) *satisfy the conditions* (2.5) *with $\gamma_0 = 1$,* (2.6)*. Assume further that the function $\lambda \to g_\lambda(t)$ is continuous, nonnegative and monotonically increasing, and satisfies the conditions* $\lim_{\lambda \to 0} t g_\lambda(t) = 1$, $\frac{\mathrm{d}}{\mathrm{d}\lambda} g_\lambda(t) \le \gamma \hat{y} \beta_\lambda(t)(1 - t g_\lambda(t))$, $\hat{y} = const$, *where* $\beta_\lambda(t) = (1 - t g_\lambda(t))^{1/p_0}$ *for methods with $p_0 < \infty$ and $\beta_\lambda(t) = 1$ for methods with $p_0 = \infty$, and* $\sup_{0 \le t \le \|A^*A\|} (t\beta_{\lambda_1}(t))^p (1 - t g_{\lambda_1}(t))(1 - t g_{\lambda_2}(t))^{-1} \le \tilde{\gamma}_p^{p-1} (\lambda_1^{-1} - \lambda_2^{-1})^{-p}$, $0 < \lambda_1 < \lambda_2$, $p \ge 0$, *where* $\tilde{\gamma}_p^s = (\gamma_{p/\tau(s)})^{\tau(s)}$, $\tau(s) = 1 + (s+1)/p_0$ *for methods with $p_0 < \infty$, $\tau(s) = 1$ for methods with $p_0 = \infty$ ($0 \le t \le \|A^*A\|$, $\lambda \ge 0$). Then the rules MD, ME, R1, B1, B2 in method* (2.3) *are quasioptimal. If the generating function $g_\lambda(t)$ in approximation* (2.9) *satisfies the self-adjoint analogs of above conditions (with* (2.4) *instead of* (2.6) *and A instead of A^*A), then the rules MD, ME, R1, B1, B2 in method* (2.9) *are quasioptimal.*

The following four theorems give the convergence results and error estimates for *a priori* parameter choice and for rules D, MD, ME, R1, R2.

Theorem 3. *[74, 76, 84, 88, 89] Let $y \in \mathcal{R}(A)$, $\|y - y_*\| \le \delta$. If in approximation* (2.3) *the regularization parameter $\lambda = \lambda(\delta)$ is chosen by* a priori *rule $\lambda \sim \delta^{\frac{2}{p+1}}$, then $\|x_{\lambda(\delta)} - x_*\| \to 0$ as $\delta \to 0$ and for source-like solutions* (2.2) *the error estimate* (3.1) *holds with $p \le 2p_0$. If in approximation* (2.3) *the*

regularization parameter $\lambda = \lambda(\delta)$ is chosen by one of rules D, MD, ME, R1, then $\|x_{\lambda(\delta)} - x_*\| \to 0$ as $\delta \to 0$ and for source-like solutions (2.2) *the error estimate* (3.1) *holds true with $p \le 2p_0 - 1$ in case of rule D or with $p \le 2p_0$ in case of rules MD, ME, R1.*

Theorem 4. *[73, 75, 87–89] Let $y \in \mathcal{R}(A)$, $\|y - y_*\| \le \delta$. If in approximation* (2.9) *the regularization parameter $\lambda = \lambda(\delta)$ is chosen by* a priori *rule $\lambda \sim \delta^{\frac{1}{p+1}}$, then $\|x_{\lambda(\delta)} - x_*\| \to 0$ as $\delta \to 0$ and for source-like solutions* (2.2) *the error estimate* (3.1) *holds with $p \le p_0$. If in approximation* (2.9) *the regularization parameter $\lambda = \lambda(\delta)$ is chosen by one of rules D (with assumption $p_0 > 1$), MD, R1, then $\|x_{\lambda(\delta)} - x_*\| \to 0$ as $\delta \to 0$ and for source-like solutions* (2.2) *the error estimate* (3.1) *holds true with $p \le p_0 - 1$ in case of rule D or with $p \le p_0$ in case of rules MD, R1.*

Theorem 5. *[80] Let $Qy \in \mathcal{R}(A)$, $\|y - y_*\| \le \delta$. If the regularization parameter $\alpha = \alpha(\delta)$ in m-iterated Tikhonov method is chosen as the smallest solution of the equation $d_{R2}(\alpha) = C\delta$ with $C > 1$, then $\|x_{\alpha(\delta)} - x_*\| \to 0$ as $\delta \to 0$ and for source-like solutions* (2.2) *the error estimate* (3.1) *holds true with $p \le 2m - 1$.*

Theorem 6. *[80] Let $Qy \in \mathcal{R}(A)$, $\|y - y_*\| \le \delta$. If the regularization parameter $\alpha = \alpha(\delta)$ in m-iterated Tikhonov method is the unique solution of equation $d_{R2}(\alpha) = C\delta$ with $C > 2$ and the inequality $\big\||A^*|B_\alpha^{3+2m}\overline{r}\big\| \ge \big\||A^*|B_\alpha^{3+2m}(y - y_*)\big\|$ is valid for all $\alpha \ge \alpha(\delta)$, then the error estimate* (3.15) *holds true with $C' = 2\sqrt{m}\,(C + \tilde{\gamma}_{1/2,2m+1}) + \max\Big(\dfrac{m}{C/2-1}, \dfrac{C/2-1}{m}\Big)^{1/2}$.*

Next three theorems concern the balancing principle (Section 3.1.6). The first two of them show that the two forms (3.9), (3.10) of the balancing principle are quasioptimal, the third establishes a monotonous dependence of error on the constant c in these two variants of balancing principle.

Theorem 7. *[34, 79] Let $Qy \in \mathcal{R}(A)$, $\|y - y_*\| \le \delta$. If the regularization parameter $\alpha = \alpha(\delta)$ is chosen according to the balancing principle* (3.9) *with $c > (1 - q)\gamma\tilde{\gamma}_{1/2}/(\gamma_*\sqrt{q})$, then for methods* (2.11), (2.18), (2.20), (2.21) *the error estimate* (3.15) *holds true.*

Theorem 8. *[34, 79] Let $y \in \mathcal{R}(A)$, $\|y - y_*\| \le \delta$. If the regularization parameter $\alpha = \alpha(\delta)$ is chosen according to the balancing principle* (3.10) *with $c > \gamma\gamma_1^{1/2}/\gamma_*$, then for methods* (2.11), (2.18), (2.20), (2.21) *the quasioptimal error estimate* (3.15) *holds true.*

Theorem 9. *[34] Let $y \in \mathcal{R}(A)$, $\|y - y_*\| \le \delta$. If the regularization parameter $\alpha = \alpha(c)$ in m-iterated Tikhonov method is chosen from the condition* (3.9) *with $c > (q^{-m} - 1)$ or from the condition* (3.10) *with $c > (q^{-m} - 1)(1 - q^{i+1-j})(1 - q)^{-1}$ then the error $\|x_{\alpha(c)} - x_*\|$ is a monotonically increasing function of parameter c; particularly $\|x_{\alpha(c)} - x_*\| > \|x_{\alpha_{ME}} - x_*\|$.*

For approximately given noise level we can say the following.

Theorem 10. *[32, 33] Let $y \in \mathcal{R}(A)$. Consider choice of the regularization parameter $\lambda = \lambda(\delta)$ according to the following rule.*

Rule P. *Let $s \in (0,1)$. Let $\sigma = 2$ for approximation* (2.3) *and $\sigma = 1$ for approximation* (2.9). *If $d_{R1,1/2}(1) \le C\delta$, then choose $\lambda(\delta) = 1$. Otherwise choose $\lambda(\delta)$ as the global minimizer of the function $f(\lambda) = \lambda^{-s/\sigma}\|B_\lambda r_\lambda\|$ on the interval $[\lambda_{R1,1/2}, 1]$.*

If $\|y - y_*\|/\delta \le$ const as $\delta \to 0$, then in methods (2.11), (2.18), (2.20), (2.21) *and in self-adjoint variants of these methods the rule P guarantees convergence $\|x_{\lambda(\delta)} - x_*\| \to 0$ as $\delta \to 0$ and for approximations* (2.3), (2.9) *the following error estimates hold true. 1) If $\|y - y_*\| \le \max\{\delta, \delta_0\}$, where $\delta_0 = \frac{1}{2}\|B_{\lambda(\delta)} r_{\lambda(\delta)}\|$, then* (3.15) *holds with $C' = 1/(1-\sigma s)$. 2) If $\max(\delta, \delta_0) < \|y - y_*\| \le \frac{1}{2}\|B_1 r_1\|$, then* (3.15) *holds with $C' =$ const$(\|y - y_*\|/\delta_0)^{1/(\sigma s)}$.*

The proofs of the following Theorems 11–19 can be found in our papers.

Next Theorem 11 shows that the rule MEa (Section 3.1.3) for Lavrentiev method is quasioptimal.

Theorem 11. *[35] Let $y \in \mathcal{R}(A)$, $\|y - y_*\| \le \delta$. If the regularization parameter $\alpha = \alpha(\delta)$ in m-iterated Lavrentiev method is chosen by Rule MEa with $C > 2$ and the inequality $\|B_\alpha^{m+2}\overline{r}\| \ge \|B_\alpha^{m+2}(y - y_*)\|$ is satisfied for all $\alpha \ge \alpha(\delta)$, then the quasioptimal error estimate* (3.15) *holds true with $C' = C + 1 + \max\Big(\frac{m}{C/2-1}, \frac{C/2-1}{m}\Big)^{1/2}$.*

The following three theorems concern parameter choice in extrapolated Tikhonov and extrapolated Lavrentiev approximations.

Theorem 12. *[28, 30] Let $y \in \mathcal{R}(A)$, $\|y - y_*\| \le \delta$. Let m be fixed and let $\alpha_n = q_n\alpha$ with q_n fixed ($n = 1, \dots, m, m+1$). Let $x_\alpha = x_{\alpha_1,\dots,\alpha_m}$ be an extrapolated Tikhonov approximation or an extrapolated Lavrentiev approximation. Then the functions $d_D(\alpha)$, $d_{eMD}(\alpha)$ are monotonically decreasing. Let $C > 1$. If α is chosen from the discrepancy principle $d_D(\alpha) = C\delta$, then $\|x_\alpha - x_*\| \to 0$ as $\delta \to 0$ and for source-like solutions* (2.2) *the error estimate* (3.1) *holds true with $p \le 2m - 1$ in extrapolated Tikhonov method and with $p \le m - 1$ in extrapolated Lavrentiev method. If α is chosen from the modified discrepancy principle $d_{eMD}(\alpha) = C\delta$, then $\|x_\alpha - x_*\| \to 0$ as $\delta \to 0$ and for source-like solutions* (2.2) *the error estimate* (3.1) *holds true with $p \le 2m$ in extrapolated Tikhonov method and with $p \le m$ in extrapolated Lavrentiev method.*

Theorem 13. *[28] Let $y \in \mathcal{R}(A)$, $\|y - y_*\| \le \delta$. Let the sequence $\alpha_1 \ge \alpha_2 \ge \dots$ be given and let $x_n = x_{\alpha_1,\dots,\alpha_n}$ be an extrapolated Tikhonov approximation. Then the functions $d_D(n)$, $d_{ME}(n)$ are monotonically decreasing and $d_D(n+1) < d_{ME}(n) < d_D(n)$ for all n. Let n_D, n_{ME} be the first numbers with $d_D(n) \le C\delta$, $d_{ME}(n) \le C\delta$ respectively, with $C > 1$. Then $n_D - 1 \le n_{ME} \le n_D$ and $\|x_n - x_*\| < \|x_{n-1} - x_*\|$ ($n = 1, 2, \dots, n_{ME}$). If a monotonically decreasing infinite sequence $\alpha_1, \alpha_2, \dots$ satisfies conditions $\sum_{i=1}^{\infty} \alpha_i^{-1} = \infty$, $\alpha_n^{-1} \le$ const $\sum_{i=1}^{n-1} \alpha_i^{-1}$, then the existence of finite n_D and n_{ME} is*

guaranteed and for $n \in \{n_D, n_{ME}\}$ $\|x_n - x_*\| \to 0$ *as* $\delta \to 0$*; for source-like solutions* (2.2) *also the error estimate* (3.1) *holds true for all* $p > 0$.

Theorem 14. *[28,30] Let* $y \in \mathcal{R}(A)$, $\|y - y_*\| \le \delta$. *Let* m, M_n *be fixed and let* $\alpha_n = q_n \alpha$ *with fixed* q_n $(n = 1, \ldots, m)$. *Let* $x_\alpha = x_{\underbrace{\alpha_1,\ldots,\alpha_1}_{M_1},\underbrace{\alpha_2,\ldots,\alpha_2}_{M_2},\ldots,\underbrace{\alpha_m,\ldots,\alpha_m}_{M_m}}$ *be an extrapolated Tikhonov approximation or an extrapolated Lavrentiev approximation. If the regularization parameter* $\alpha = \alpha(\delta)$ *is chosen by the discrepancy principle* $d_D(\alpha) = C\delta$ *with* $C > 1$, *then* $\|x_\alpha - x_*\| \to 0$ *as* $\delta \to 0$ *and for source-like solutions* (2.2) *the error estimate* (3.1) *holds true with* $p \le 2(M_1 + M_2 + \ldots + M_n) - 1$ *in extrapolated Tikhonov method and with* $p \le M_1 + M_2 + \ldots + M_n - 1$ *in extrapolated Lavrentiev method.*

Next two theorems deal with the approach, where in rules MD and ME the discrepancy of additionally iterated approximation is replaced by a proper linear combination (cf. rules eMD and eME in Sections 3.1.2, 3.1.3, respectively).

Theorem 15. *[28] Let* $y \in \mathcal{R}(A)$, $\|y - y_*\| \le \delta$. *Let* $x_{m;\alpha}$ *and* $x_{m;q\alpha}$ *be approximations found by* m*-iterated Tikhonov method (by Tikhonov method, if* $m = 1$*) and let*

$$v_{m;\alpha} = (1 - q^{-m})^{-1} x_{m;\alpha} + (1 - q^m)^{-1} x_{m;q\alpha}. \tag{3.16}$$

Denote $s_{m;\alpha} = Av_{m;\alpha} - y$. *If the regularization parameter* $\alpha = \alpha(\delta)$ *is chosen by rules* $(r_{m;\alpha}, s_{m;\alpha})^{1/2} = C\delta$ *with* $C > 1$ *or* $(r_{m;\alpha}, s_{m;\alpha})/\|s_{m;\alpha}\| = C\delta$ *with* $C > (q^{-m} - 1)/(m(1 - q))$, *then* $\|x_{m;\alpha} - x_*\| \to 0$ *as* $\delta \to 0$ *and for source-like solutions* (2.2) *the error estimate* (3.1) *holds true with* $p \le 2m$.

Theorem 16. *[28] Let* $y \in \mathcal{R}(A)$, $\|y - y_*\| \le \delta$. *Let* $x_{m;\alpha}$ *and* $x_{m;q\alpha}$ *be approximations found by* m*-iterated Lavrentiev method (by Lavrentiev method, if* $m = 1$*) and let* $v_{m;\alpha}$ *be their linear combination* (3.16). *Denote* $s_{m;\alpha} = Av_{m;\alpha} - y$. *If the regularization parameter* $\alpha = \alpha(\delta)$ *is chosen by the rule* $\|s_{m;\alpha}\| = C\delta$ *with* $C > 1$, *then* $\|x_{m;\alpha} - x_*\| \to 0$ *as* $\delta \to 0$ *and for source-like solutions* (2.2) *the error estimate* (3.1) *holds true with* $p \le m$.

For theoretical justification of the minimization strategy (Section 3.2.2) the following results can be used.

Theorem 17. *[31] Let* $y \in \mathcal{R}(A)$. *Let the function* $g_\lambda(t)$ *in method* (2.3) *be continuous, non-negative and monotonically increasing for each* $t \ge 0$ *and satisfy conditions* (2.5), (2.6) *and* $\frac{\mathrm{d}}{\mathrm{d}\lambda} g_\lambda(t) \le \gamma \hat{y} \beta_\lambda(t)(1 - t g_\lambda(t))$, $\hat{y} = const$, *where* $\beta_\lambda(t) = (1 - t g_\lambda(t))^{1/p_0}$ *for methods with* $p_0 < \infty$ *and* $\beta_\lambda(t) = 1$ *for methods with* $p_0 = \infty$ $(0 \le t \le \|A^*A\|, \lambda \ge 0)$. *Then for source-like solutions* $x_* = \overline{x} + (A^*A)^{p/2} w$, $\|w\| = \omega$ *the error estimate*

$$\|x_\lambda - x_*\| \le c\Big(1 + \frac{\|y - y_*\|}{\varphi(\lambda_0)\sqrt{\lambda}}\Big)\omega^{1/(p+1)} \Delta_\lambda^{p/(p+1)}, \quad \Delta_\lambda = \max\{d(\lambda), C\|y - y_*\|\}$$

holds for all $\lambda > 0$, *where* $\varphi(\lambda) = d(\lambda)/\sqrt{\lambda}$, $\lambda_0 = (\|y - y_*\|/\omega)^{2/(p+1)}$ *and* $0 < p \le 2p_0 - 1$ *in case* $d(\lambda) = d_D(\lambda)$ *or* $d(\lambda) = d_{ME}(\lambda)$ *or* $d(\lambda) = d_{R2}(\lambda)$,

and $0 < p \le 2p_0$ *in case* $d(\lambda) = d_{MD}(\lambda)$ *or* $d(\lambda) = d_{R1,k}(\lambda)$, *and* $C = 1$ *for rules D, MD, ME,* $C = \tilde{y}_{1/2}^0$ *for rule R2,* $C = \tilde{y}_k^{k-1/2}$ *for rule R1,k.*

Theorem 18. *[31] Let* $y \in \mathcal{R}(A)$. *Let* $\hat{x}_\lambda$ *be an approximate solution of* $Ax = \hat{y}$, *found by m-iterated Tikhonov method, where* $\hat{y}$ *is an element for which* $\|Q_\mu(\hat{y} - y_0)\| = \|Q_\mu(y - y_0)\|$, $(Q_\mu A(\overline{x} - x_*), \hat{y} - y_0) \ge 0$ $(0 \le \mu \le \|A^*A\|)$, Q_μ *is the spectral family of operator* AA^*. *If the global minimizer of the function* $f(\lambda) = \|\hat{x}_\lambda - x_*\|$ *lies in the interval* $[\lambda_1, \lambda_2]$, *then* $\|x_\lambda - x_*\| \le (1 + W(\lambda)) \inf\limits_{\lambda>0} \|\hat{x}_\lambda - x_*\|$, *where*

$$W(\lambda) = \sup_{\lambda_1 \le \mu \le \lambda_2} \frac{\mu \|x_\lambda - x_\mu\|}{\gamma \|A^* B_{\xi\mu}^2 r_\mu\|},$$

$\xi = 1$ *for* $m = 1$, $\xi = (1 + (m/(m+1))^{m+1})^{-1}$ *for* $m \ge 2$.

The proofs of the following two theorems are short and we present these proofs.

Theorem 19. *[36] Let* $y \in \mathcal{R}(A)$, $\|y - y_*\| \le \delta$. *If* $d_{ME}(n) \ge \delta$ *in iterative regularization methods of the form* (3.3), *then* $\|x_{n+1} - x_*\| \le \|x_n - x_*\|$.

Proof.

$$\begin{aligned}
&\|x_n - x_*\|^2 - \|x_{n+1} - x_*\|^2 \\
&\quad = (x_n + x_{n+1} - 2x_*, x_n - x_{n+1}) = (x_n + x_{n+1} - 2x_*, A^* z_n) \\
&\quad = (Ax_n + Ax_{n+1} - 2Ax_*, z_n) = (r_n + r_{n+1} + 2(y - y_*), z_n) \\
&\quad \ge (r_n + r_{n+1}, z_n) - 2\|y - y_*\| \|z_n\| \ge (r_n + r_{n+1}, z_n) - 2\delta \|z_n\|
\end{aligned}$$

Therefore, if $d_{\mathrm{ME}}(n) \ge \delta$, then $\|x_n - x_*\| \ge \|x_{n+1} - x_*\|$. □

Theorem 20. *[84] Let* $y \in \mathcal{R}(A)$, $\|y - y_*\| \le \delta$. *If* $d_{ME}(\alpha) \ge \delta$ *in continuous regularization methods of the form* (2.3), *then the function* $e(\alpha) = \|x_\alpha - x_*\|$ *is increasing at* α.

Proof.

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\alpha}\|x_\alpha - x_*\|^2 = 2(x_\alpha - x_*, \frac{\mathrm{d}}{\mathrm{d}\alpha} x_\alpha) \\
&\quad = 2(x_\alpha - x_*, -\frac{\mathrm{d}}{\mathrm{d}\alpha} g_\alpha(A^*A)A^*\overline{r}) = 2(x_\alpha - x_*, -A^* \frac{\mathrm{d}}{\mathrm{d}\alpha} g_\alpha(AA^*)\overline{r}) \\
&\quad = 2(Ax_\alpha - Ax_*, -\frac{\mathrm{d}}{\mathrm{d}\alpha} g_\alpha(AA^*)\overline{r}) = 2(r_\alpha + y - y_*, -\frac{\mathrm{d}}{\mathrm{d}\alpha} g_\alpha(AA^*)\overline{r}) \\
&\qquad \ge 2\big[(r_\alpha, -\frac{\mathrm{d}}{\mathrm{d}\alpha} g_\alpha(AA^*)\overline{r}) - \|y - y_*\| \|\frac{\mathrm{d}}{\mathrm{d}\alpha} g_\alpha(AA^*)\overline{r}\|\big] \\
&\qquad\qquad \ge 2\big[(r_\alpha, -\frac{\mathrm{d}}{\mathrm{d}\alpha} g_\alpha(AA^*)\overline{r}) - \delta \|\frac{\mathrm{d}}{\mathrm{d}\alpha} g_\alpha(AA^*)\overline{r}\|\big].
\end{aligned}$$

Hence, if $d_{\mathrm{ME}}(\alpha) \ge \delta$, then $\frac{\mathrm{d}}{\mathrm{d}\alpha}\|x_\alpha - x_*\| \ge 0$. □

IV. NUMERICAL RESULTS

4.1. Test problems, noise generation, and general remarks

Numerical computations, the results of which we present here, are made with parametrized test problems that allow to form the matrix, exact solution vector, and corresponding right hand side of prescribed dimension. Most tests are performed with the set of test problems by Hansen [43, 45, 46], which is becoming a *de facto* standard in studying numerics of ill-posed problems. Table 1 describes the problems used in experiments, showing the problems' names, descriptions and condition numbers at discretization parameter 100. Other parameters, if existent, had default values that can be considered being appropriate for ill-posed problems.

We have made an attempt to computationally estimate the smoothness parameter p in (2.2) for default solutions of these problems. For this we solved the problems 100 times at every noise level 10^{-1}, $10^{-1.01}$, ..., 10^{-6}, using discretization parameter 1000: these values should as much as possible minimize distortive effects caused by discretization and particular noise vectors. At each noise level δ we computed the average over all noise vectors of errors of approximate solutions found by Tikhonov method with monotone error rule and by Landweber method with discrepancy principle. Assuming that the error $e(\delta)$ is of the form $e(\delta) \approx c\delta^{p/(p+1)}$ with c constant, we calculated the best p by method of least squares (taking logarithms of both sides before). In Table 1 these results are in the column Est p. The column Maxerr shows maximal relative errors of $c\delta^{p/(p+1)}$ with respect to $e(\delta)$. It turned out in calculations that perturbations in $e(\delta)$, compared to the function $c\delta^{p/(p+1)}$ were quite large even for such small δ as 10^{-6}. This shows that theoretical convergence results about rules should be used with caution at least in region $\delta \in [10^{-6}, 10^{-1}]$.

In addition, we used some more or less artificial test problems from [13], whose matrices are described in Table 2. As in [13], we combined these matrices with 6 solution vectors of Table 3. In the following, the cases, where some other set of test problems (except Hansen's) was used, are specially noted.

The discretization parameter (number of rows/columns of resulting matrix) is typically 100. On one hand, this value is large enough to reveal the characteristic properties of problems, methods and rules; on the other hand it is small enough, so that the computer could generate sufficiently large amount of data to make various comparisons. We have also experi-

Table 1. Hansen's test problems [43,45,46] used in numerical tests, together with smoothness estimations of default solutions.

Nr	Problem	cond_{100}	selfadj	Est p	Maxerr	Description
1	baart	5e+17	no	0.17	20%	(Artificial) Fredholm integral equation of the first kind
2	deriv2	1e+4	yes	0.20	1%	Computation of the second derivative
3	foxgood	1e+19	yes	0.63	25%	A problem that does not satisfy the discrete Picard condition
4	gravity	3e+19	yes	0.51	4%	A gravity surveying problem
5	heat	2e+38	no	0.52	32%	Inverse heat equation
6	ilaplace	9e+32	no	0.04	6%	Inverse Laplace transform
7	phillips	2e+6	yes	0.69	21%	An example problem by Phillips
8	shaw	5e+18	yes	0.22	31%	An image reconstruction problem
9	spikes	3e+19	no	0.00	1%	Test problem whose solution is a pulse train of spikes
10	wing	1e+20	no	0.06	11%	Fredholm integral equation with discontinuous solution

Table 2. Test problems from [13].

Nr	Problem	cond_{100}	selfadj	Description
11	gauss	6e+18	yes	Test problem with Gauss matrix $a_{ij} = \sqrt{\frac{\pi}{2\sigma}} e^{-\frac{\sigma}{2(i-j)^2}}$ with $\sigma = 0.01$
12	hilbert	4e+19	yes	Test problem with Hilbert matrix $a_{ij} = \frac{1}{i+j-1}$
13	lotkin	2e+21	no	Test problem with Lotkin matrix (same as Hilbert matrix, except $a_{1j} = 1$)
14	moler	2e+4	yes	Test problem with Moler matrix $A = B^T B$, where $b_{ii} = 1$, $b_{i,i+1} = 1$, and $b_{ij} = 0$ otherwise
15	pascal	1e+60	yes	Test problem with Pascal matrix $a_{ij} = \binom{i+j-2}{i-1}$
16	prolate	1e+17	yes	Test problem with a symmetric, ill-conditioned Toeplitz matrix

Table 3. Solution vectors for test problems of [13]. Here N is discretization parameter of the problem (usually 100) and $i = 1, \ldots, N$.

Description	$\overline{x}_i$	Description	$\overline{x}_i$
constant	1	sinusoidal	$\sin \frac{2\pi(i-1)}{N}$
linear	$\frac{i}{N}$	linear+sinusoidal	$\frac{i}{N} + \frac{1}{4} \sin \frac{2\pi(i-1)}{N}$
quadratic	$\left(\frac{i - \left\lfloor \frac{N}{2} \right\rfloor}{\left\lceil \frac{N}{2} \right\rceil} \right)^2$	step function	$\begin{cases} 0, & \text{if } i \le \left\lfloor \frac{N}{2} \right\rfloor \\ 1, & \text{if } i > \left\lfloor \frac{N}{2} \right\rfloor \end{cases}$

mented with other values of the discretization parameter but for example the results with parameter 1000 were quite similar to the results with parameter 100.

Since the performance of methods and rules generally depends on the smoothness p of exact solution in (2.2), we complemented the standard solutions x_* of (now discrete) test problems with smoothened solutions $|A|^p x_*$ with $p = 0,\ 0.25,\ 0.5,\ 0.75,\ 1,\ 1.5,\ 2,\ 4,\ 8$. In the following tables p always means this additional smoothness and is not related to initial smoothness in Table 1. The right hand side of the equation was computed as $y_* = A|A|^p x_*$. After discretization all problems were scaled (normalized) in such way that the Euclidian norms of the operator and right-hand side were 1.

On base of exact data y_* we formed the noisy data y, where $\|y - y_*\| = \delta$ for $\delta = 0.5,\ 10^{-1},\ \ldots,\ 10^{-6}$. In most cases the noise $y - y_*$ added to y_* had uniform distribution (we preferred this to normal distribution since it is more consistent with the usual assumption $\|y - y_*\| \le \delta$ made in studying ill-posed problems). Besides this we used correlated noise, where the components of noise vector had nonzero correlation. The amount of correlation was determined by randomly choosing the parameter $\omega \in [-0.5, 0.5]$, where $\omega = 0$ corresponds to white noise, $\omega = -0.5$ corresponds to noise, which has dominantly high frequencies in frequency domain (blue noise), and $\omega = 0.5$ corresponds to noise with dominantly lower frequencies (red noise). If correlated noise was used, it is mentioned in heading of the corresponding table.

To enlarge the common base of comparisons, we generated 10 noise vectors and saved them beforehand, so that at different runs the same 10 noise vectors were used in all problems. Then the problems were regularized using different methods, choosing the regularization parameters by rules that we wanted to compare. In experiments we also took into account the possibility of over- or underestimation of the noise level: although the actual noise level was δ, we applied the rules as if the noise level was $d\delta$, where $d \in [0.01, 100]$. Thus, if $d > 1$, then the noise level was overestimated.

To speed up the computations, the discrete problem of the form $Ax = y$ was further transformed in the following way. Using a fast-working standard function of the programming language, the singular value decomposition of A was calculated as $A = U\tilde{A}V^T$, where $\tilde{A}$ is a diagonal matrix, U, V are orthogonal matrices and ^{T} means transposition. Then the problem $U\tilde{A}V^T x = y$ was replaced by the problem $\tilde{A}\tilde{x} = \tilde{y}$ with $\tilde{x} = V^T x$ and $\tilde{y} = U^T y$. This, however, brought a small decrease in computational stability but its influence was noticeable only at very small values of the regularization parameter (close to machine precision).

As Table 1 shows, half of Hansen's problems are self-adjoint. Methods, which are also applicable in non-self-adjoint problems, were used in all problems 1–10. Methods for self-adjoint problems were used only in

problems 2, 3, 4, 7, 8 of Hansen (exception is Table 33).

Taking into account normalization of problems, the interval of regularization parameters was taken to be $[10^{-30}, 1]$ for methods of Tikhonov and Lavrentiev, or $[\underline{\alpha}, 1]$ with $\underline{\alpha} > 10^{-30}$ if numerical instabilities occurred in the interval $[10^{-30}, \underline{\alpha})$ in these methods, for example, if a theoretically positive expression turned out to be numerically negative. If in Landweber method a parameter choice rule did not stop the iterations earlier, iterations were stopped at 2^{60}. In CGLS and CGME maximal number of iterations was 4000. This upper bound of iterations was, however, not necessary in most problems.

In the following sections we present several rules with particular numerical constants. These constants are always in some sense optimal: either they have been optimized on data set considered here or on some larger data set (additional problems (mainly artificial), additional δ's, larger number of noise vectors), or they have been selected with the aim to balance the behavior of rules between different smoothness indices (with smoothness $p = 0$ having the largest weight) or different noise level ill-estimation indices (with indices close to $d = 1$ having the largest weight). The constants were optimized mainly for uncorrelated noise.

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 rule R on this particular problem. To compare the rules or to present their properties, the following tables show the averages of these error ratios over various parameters of data set (problems, smoothness indices p, noise levels δ, runs). Besides averages of error ratios we also computed root-mean-squares of error ratios for our numerical experiments but the better-worse relationships of methods and rules remained mostly the same.

Numerical tests were made with GNU Octave, a freeware analog of Matlab, primarily intended for numerical computations.

4.2. Comparison of potential of methods

Tables 4 and 5 compare the potential of methods at noise level $\delta = 10^{-4}$, showing the averages of minimal relative errors $\|x_{\lambda_*} - x_*\| / \|x_*\|$ over 10 runs for $p = 0$ in case of uncorrelated noise and correlated noise, respectively. Tables 6 and 7 contain analogous results for $p = 2$. The best results for every problem are shown in bold. For other δ's the better-worse relationships of methods remained the same. Typically the best results were produced by methods of Landweber, TSVD, and CGLS. If $p = 0$, then the results of Tikhonov method were close to the best.

Note different typical behavior of error in methods CGLS and CGME (Fig-

Table 4. 10^3 times the averages of minimal relative errors in methods for $\delta = 10^{-4}$, $p = 0$. Self-adjoint variants of methods are marked with *.

Probl.	Tikh	*Lavr	Landw	*Landw	TSVD	CGLS	CGME	*CG
1	62.7	–	**61.9**	–	89.6	89.3	116	–
2	**107**	123	**107**	115	117	109	129	115
3	4.95	26.0	**4.51**	17.2	5.61	5.54	8.27	16.4
4	7.12	21.5	6.79	14.9	**6.71**	7.01	14.8	14.4
5	18.0	–	**16.9**	–	17.3	17.0	20.9	–
6	70.7	–	69.7	–	70.8	**69.5**	96.3	–
7	5.12	17.1	4.69	9.73	**4.38**	4.74	8.44	8.04
8	31.1	61.5	**30.9**	50.4	35.8	35.2	47.4	48.6
9	**788**	–	**788**	–	802	796	823	–
10	364	–	**363**	–	446	445	595	–

Table 5. 10^3 times the averages of minimal relative errors in methods for $\delta = 10^{-4}$, $p = 0$ (correlated noise).

Probl.	Tikh	*Lavr	Landw	*Landw	TSVD	CGLS	CGME	*CG
1	54.7	–	**53.9**	–	67.2	67.0	116	–
2	98.0	117	**96.3**	110	99.1	96.5	119	110
3	3.42	25.7	**2.91**	16.9	4.45	4.12	8.35	15.9
4	5.48	21.1	**5.33**	14.6	5.75	5.60	14.6	14.2
5	16.3	–	**15.7**	–	16.1	**15.7**	20.8	–
6	**55.9**	–	**55.9**	–	61.1	59.6	96.2	–
7	4.15	16.8	3.77	9.60	**3.37**	3.96	8.25	7.85
8	30.7	61.5	**30.2**	50.5	34.7	34.4	47.5	48.6
9	**781**	–	**781**	–	793	789	822	–
10	357	–	**356**	–	451	451	595	–

Figure 2. Errors of approximate solutions computed by TSVD on truncation steps 0, 1, ..., 60, by CGLS, CGME on iteration steps 0, 1, ..., 60 and by Landweber method on iteration steps 2^0, 2^1, ..., 2^{60} in problem baart at $\delta = 10^{-4}$, $p = 0$; —– TSVD, ····· CGLS, – – – CGME, - - - - - Landweber.

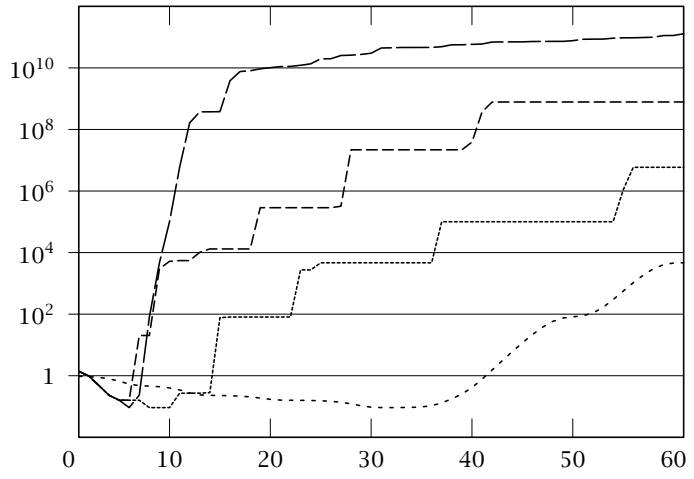


Table 6. 10^3 times the averages of minimal relative errors in methods for $\delta = 10^{-4}$, $p = 2$.

Probl.	Tikh	*Lavr	Landw	*Landw	TSVD	CGLS	CGME	*CG
1	0.35	–	**0.12**	–	0.20	0.20	0.22	–
2	0.80	13.4	0.30	1.73	0.25	**0.24**	0.73	1.61
3	0.30	14.0	**0.04**	1.51	0.08	0.08	0.19	1.00
4	0.55	13.9	0.17	1.22	**0.15**	0.18	0.60	1.03
5	1.19	–	0.59	–	**0.40**	0.55	1.13	–
6	0.51	–	0.19	–	**0.17**	0.19	0.59	–
7	0.71	13.6	0.12	1.13	**0.07**	0.09	0.22	1.26
8	0.51	14.1	**0.10**	1.57	**0.10**	**0.10**	0.12	1.38
9	0.48	–	**0.19**	–	0.25	0.24	0.91	–
10	0.24	–	**0.08**	–	0.12	0.12	0.29	–

Table 7. 10^3 times the averages of minimal relative errors in methods for $\delta = 10^{-4}$, $p = 2$ (correlated noise).

Probl.	Tikh	*Lavr	Landw	*Landw	TSVD	CGLS	CGME	*CG
1	0.35	–	**0.15**	–	0.16	0.18	0.22	–
2	0.84	12.9	0.36	1.72	**0.25**	0.37	0.73	1.60
3	0.32	13.9	**0.05**	1.50	0.07	0.10	0.21	1.00
4	0.59	13.7	**0.14**	1.20	0.15	0.16	0.59	1.01
5	1.13	–	0.52	–	**0.41**	0.49	1.15	–
6	0.56	–	0.23	–	**0.17**	0.21	0.61	–
7	0.72	13.2	0.14	1.11	**0.09**	0.10	0.19	1.22
8	0.54	13.9	0.09	1.53	**0.08**	0.10	0.17	1.37
9	0.60	–	0.31	–	**0.27**	0.35	0.93	–
10	0.28	–	0.12	–	**0.11**	0.15	0.24	–

ure 2): in both methods the error of approximate solution decreases quickly before the optimal stopping index but in CGME it begins to increase very quickly afterwards, whereas in CGLS it begins to increase with some delay.

In CGLS the number of iterations was much smaller than in Landweber method: in our tests the mean and median of optimal stopping indices were 14 and 5 in CGLS, versus $1.4 \cdot 10^{13}$ and 1038 in Landweber method. But since we implemented the Landweber method using operator iterations (2.19) with $m = 2$, the overall computing time was the shortest in Landweber method, followed by Lavrentiev and Tikhonov method; somewhat more time was needed in CGLS and CGME.

4.3. Results in Tikhonov method

Now we compare rules in each method separately. Adding the letters e or m to the name of some rule means post-estimation or modification of the corresponding rule, respectively. Adding the letters C or 1 to the name of some rule means climbing approach or first local minimum (see Section 3.2.2).

4.3.1. Comparison of refinements of known rules.

Table 8 presents the results for known and modified rules that use full information about noise level.

T1) Discrepancy principle: α_{D} is the solution of the equation $d_{\mathrm{D}}(\alpha) = \delta$.

T2) Monotone error rule: α_{ME} is the solution of the equation $d_{\mathrm{ME}}(\alpha) = \delta$.

T3) Rule MEe: using α_{ME} from T2, take $\alpha_{\mathrm{MEe}} = \min(0.53\alpha_{\mathrm{ME}}, 0.6\alpha_{\mathrm{ME}}^{1.06})$.

T4) Rule R2: α_{R2} is the largest solution of the equation $d_{\mathrm{R2}}(\alpha) = 0.3\delta$.

T5) Rule R2e: using α_{R2} from T4, take $\alpha_{\mathrm{R2e}} = 0.5\alpha_{\mathrm{R2}}$.

T6) Rule Me: using α_{MEe}, α_{R2e} from T3, T5, take $\alpha_{\mathrm{Me}} = \min(\alpha_{\mathrm{MEe}}, \alpha_{\mathrm{R2e}})$.

Here and in the following, the column R in tables, where R is a name of a rule, shows the average (over all free parameters: runs, δ's, problems 1–10 unless noted otherwise, p's) of error ratios $\|x_{\lambda_{\mathrm{R}}} - x_*\| / \|x_{\lambda_*} - x_*\|$ for regularization parameter λ_{R} in method under consideration. Recall that here the denominator is the error of regularized solution at optimal regularization parameter. The column heading R,2 means that the results in this column are computed with 2 times overestimated noise level $2\delta = 2\|y - y_*\|$, instead of $\delta = \|y - y_*\|$.

Table 8 confirms the disadvantages of discrepancy principle: saturation (large values of D for $p \geq 1$) and sensitivity to inexact noise level (results in column D,2 are much larger than in column D for $p \leq 1$). We also see that the estimated parameters from rules MEe and R2e give better results than the parameters from ME and R2, respectively. However, if the error level used in rules was 2 times larger than the actual, the results of rules ME and MEe were not so good.

Note also that the column Me almost coincides with column MEe and the column Me,2 with column R2e,2 (this also holds in single problems, see Tables 16 and 17). For $p \geq 1.5$ the rules MEe, R2e, and Me gave remarkably small averages. Although the results of MEe and Me were similar, the maximums of Me were smaller, especially at larger p (for example, 2.53 and 2.18, respectively, for $p = 2$).

Table 8. Means of error ratios for rules in case of exact noise level and in case of 2 times overestimated noise level at various smoothness indices p.

p	D	D,2	ME	ME,2	MEe	MEe,2	R2	R2,2	R2e	R2e,2	Me	Me,2
0	1.19	2.23	1.33	2.54	1.15	2.01	1.54	1.84	1.39	1.59	1.16	1.58
0.25	1.64	3.82	1.91	4.58	1.56	3.37	2.93	3.50	2.63	3.04	1.57	2.55
0.5	1.69	4.09	2.02	5.65	1.56	3.92	2.46	3.33	2.02	2.59	1.58	2.59
0.75	1.61	3.65	1.79	5.93	1.35	3.66	2.03	2.92	1.57	2.08	1.36	2.08
1	1.81	3.14	1.59	5.94	1.18	3.34	1.75	2.46	1.33	1.65	1.19	1.65
1.5	2.50	2.77	1.45	5.66	1.13	2.93	1.38	1.94	1.14	1.28	1.14	1.28
2	2.83	2.76	1.39	5.45	1.12	2.75	1.26	1.68	1.11	1.17	1.13	1.17
4	3.01	2.80	1.38	5.43	1.12	2.72	1.22	1.59	1.11	1.14	1.13	1.14
8	3.03	2.79	1.38	5.37	1.12	2.69	1.22	1.58	1.11	1.14	1.13	1.14
mean	2.14	3.12	1.58	5.17	1.26	3.04	1.75	2.32	1.49	1.74	1.26	1.69

Table 9. Means of error ratios for rules in case of exact noise level and in case of 2 times overestimated noise level at various smoothness indices p (positively correlated noise).

p	D	D,2	ME	ME,2	MEe	MEe,2	R2	R2,2	R2e	R2e,2	Me	Me,2
0	1.28	2.16	1.45	2.44	1.19	1.86	1.44	1.72	1.32	1.49	1.20	1.48
0.25	1.63	3.20	1.93	3.88	1.54	2.82	2.38	2.88	2.16	2.49	1.55	2.12
0.5	1.65	3.31	1.97	4.54	1.48	3.17	2.07	2.70	1.80	2.14	1.50	2.14
0.75	1.54	2.81	1.73	4.56	1.32	2.80	1.64	2.25	1.46	1.67	1.40	1.67
1	1.68	2.34	1.54	4.35	1.17	2.42	1.38	1.84	1.25	1.33	1.23	1.33
1.5	2.19	2.21	1.43	4.02	1.13	2.11	1.18	1.47	1.20	1.13	1.22	1.13
2	2.45	2.45	1.43	4.02	1.16	2.06	1.20	1.33	1.32	1.18	1.34	1.18
4	2.65	2.34	1.39	4.02	1.12	2.02	1.16	1.30	1.23	1.13	1.25	1.13
8	2.64	2.28	1.38	3.85	1.12	1.94	1.15	1.27	1.24	1.12	1.26	1.12
mean	1.97	2.57	1.58	3.96	1.25	2.36	1.51	1.86	1.44	1.52	1.33	1.48

Table 10. Means of error ratios for rules in case of exact noise level and in case of 2 times overestimated noise level at various smoothness indices p (correlated noise).

p	D	D,2	ME	ME,2	MEe	MEe,2	R2	R2,2	R2e	R2e,2	Me	Me,2
0	1.23	2.43	1.38	2.78	1.17	2.16	1.64	1.97	1.46	1.69	1.17	1.69
0.25	1.69	4.98	1.99	6.09	1.59	4.42	3.71	4.46	3.23	3.81	1.59	3.28
0.5	1.82	7.49	2.17	9.71	1.56	7.16	4.90	6.19	4.16	5.03	1.58	5.03
0.75	1.69	6.52	1.88	10.9	1.36	6.75	3.76	5.37	2.86	3.80	1.41	3.80
1	1.96	5.02	1.60	10.5	1.18	5.79	3.09	4.41	2.26	2.92	1.21	2.92
1.5	2.73	4.78	1.44	11.0	1.13	5.64	2.25	3.30	1.79	2.05	1.18	2.05
2	3.02	3.94	1.41	8.68	1.16	4.27	1.67	2.21	1.61	1.56	1.31	1.56
4	3.59	5.44	1.35	13.4	1.14	6.59	2.00	2.86	1.87	1.82	1.23	1.82
8	3.27	3.65	1.39	9.86	1.14	4.70	1.55	2.02	1.53	1.39	1.23	1.39
mean	2.33	4.92	1.62	9.22	1.27	5.27	2.73	3.64	2.31	2.67	1.32	2.61

Table 11. Means of error ratios for rules in case of exact noise level and in case of 2 times overestimated noise level at various smoothness indices p (negatively correlated noise).

p	D	D,2	ME	ME,2	MEe	MEe,2	R2	R2,2	R2e	R2e,2	Me	Me,2
0	1.18	3.39	1.32	3.87	1.17	3.02	2.25	2.72	1.95	2.30	1.18	2.29
0.25	1.74	6.48	2.02	8.03	1.61	5.83	4.80	5.76	4.19	4.89	1.67	4.23
0.5	1.70	9.11	1.96	12.9	1.46	8.89	5.41	7.44	4.24	5.66	1.50	5.66
0.75	1.85	9.76	1.97	16.7	1.41	10.3	5.58	8.14	4.01	5.73	1.41	5.73
1	2.22	9.17	1.75	19.1	1.23	10.5	5.19	7.72	3.39	4.92	1.23	4.92
1.5	3.15	6.93	1.44	16.5	1.15	8.50	3.16	4.87	1.95	2.84	1.15	2.84
2	3.58	5.84	1.42	13.8	1.14	6.85	2.16	3.33	1.45	1.93	1.17	1.93
4	4.25	7.19	1.37	19.3	1.14	9.39	2.41	3.71	1.51	2.11	1.16	2.11
8	4.27	7.58	1.40	19.1	1.14	9.41	2.54	3.98	1.58	2.27	1.15	2.27
mean	2.66	7.27	1.63	14.4	1.27	8.08	3.72	5.30	2.70	3.63	1.29	3.55

Table 10 is an analog of Table 8 for correlated noise of data (correlation parameter ω is chosen randomly from $[-0.5, 0.5]$ with uniform distribution). The columns D, ME, MEe, Me in Table 10 are approximately 10% larger than in Table 8 but other columns are 50%–80% larger. Tables 9, 11 are analogs of Table 10 for correlated noise of data with positive correlation ($\omega \in [0, 0.5]$) and negative correlation ($\omega \in [-0.5, 0]$), respectively. In case of correlated noise with exactly known noise level (especially in case of positively correlated noise) the rule MEe turned out to perform better than the rule Me. Since the mutual relationships of rules remained the same for various types of correlations, we present only one table for correlated noise ($\omega \in [-0.5, 0.5]$) in the following.

Tables 12, 13 show analogous results for different variants of balancing principle in Tikhonov method ($\lambda = \alpha$). Here we formed an increasing sequence α_0, α_1, …, α_N, where $\alpha_0 = \delta^2$, $\alpha_i = \alpha_{i-1}/q$ with $q = 0.9$, and α_N is the first α_i, which is greater than or equal to 1. Parameters α_{B1} and α_{B1^*} were chosen as the first α_i, for which (3.9) holds with $c = 2$ and $c = 3\sqrt{3}(1-q)/(16\sqrt{q})$, respectively; α_{B2} and α_{B2^*} were chosen as the first α_m, for which (3.10) holds with $c = 2$ and $c = (1-q^{i+1-j})/q$, respectively; α_{B3} was chosen as the first α_i, for which (3.11) holds with $c = 0.3(1-q^{i+1-j})q^{(j-i-1)/2}$.

Table 12. Means of error ratios for balancing principle.

p	B1	B1,2	B1*	B1*,2	B2	B2,2	B2*	B2*,2	B3	B3,2
0	6.27	7.48	1.90	2.35	3.82	4.58	2.71	3.34	1.85	2.29
0.25	12.6	15.7	3.19	4.13	7.46	9.05	4.96	6.48	3.08	4.01
0.5	20.4	27.0	3.67	5.07	10.4	13.8	6.12	8.24	3.51	4.87
0.75	25.9	36.3	3.52	5.21	12.2	16.8	6.48	9.25	3.37	4.96
1	29.9	41.9	3.38	5.08	13.5	18.9	6.56	9.76	3.21	4.83
1.5	33.0	48.1	3.11	4.84	14.2	20.8	6.33	9.66	2.96	4.56
2	33.0	48.8	2.99	4.67	14.1	20.9	6.09	9.44	2.82	4.40
4	33.6	49.6	3.01	4.70	14.4	21.2	6.07	9.33	2.80	4.36
8	33.8	49.4	3.02	4.68	14.5	21.1	6.01	9.31	2.79	4.32
mean	25.4	36.0	3.09	4.53	11.6	16.3	5.70	8.31	2.93	4.29

Table 13. Means of error ratios for balancing principle (correlated noise).

p	B1	B1,2	B1*	B1*,2	B2	B2,2	B2*	B2*,2	B3	B3,2
0	7.17	8.59	2.05	2.57	4.26	5.16	2.99	3.72	2.00	2.49
0.25	17.4	21.9	4.24	5.53	9.87	12.2	6.58	8.46	4.16	5.34
0.5	30.2	39.8	6.83	8.92	16.3	21.1	10.4	13.3	6.76	8.62
0.75	48.9	68.2	6.56	9.58	22.5	31.3	11.9	17.0	6.64	9.11
1	56.3	80.2	6.23	9.17	24.1	34.8	11.7	17.0	6.81	8.76
1.5	67.8	95.8	6.07	9.32	28.6	42.0	12.4	19.5	7.95	8.78
2	50.0	74.5	4.63	7.38	22.3	32.4	9.70	15.2	8.69	6.97
4	85.6	126	7.42	11.6	35.9	53.5	15.0	23.1	18.6	10.7
8	69.1	102	5.50	8.53	28.1	42.4	11.1	17.3	16.5	7.82
mean	48.1	68.5	5.50	8.07	21.3	30.5	10.2	14.9	8.68	7.63

As can be seen from Tables 12, 13, the balancing principle with original large constants $c = 2$ gives significantly larger error in rules B1, B2 than the rule B3 and rules B1*, B2* with smaller constants. However, the results for balancing principle were worse than results for rules MEe, R2e, Me in Table 8, especially when the noise is correlated (Table 13). We also experimented with constant 0.25 instead of 0.3 and obtained better results with rule (3.11) in case of uncorrelated noise but in case of correlated noise the results were significantly worse. As the last two columns of Table 13 show, in case of larger p even larger constant would be better.

Table 14 shows the results of the same rules as in Table 8 with the difference that the approximate solution is computed by 2-iterated Tikhonov method. Here, the proper constants in rules are somewhat different than those in ordinary Tikhonov method.

T1-2) Discrepancy principle: α_{D} is the solution of $d_{\mathrm{D}}(\alpha) = \delta$.

T2-2) Monotone error rule: α_{ME} is the solution of $d_{\mathrm{ME}}(\alpha) = \delta$.

T3-2) Rule MEe: using α_{ME} from T2-2, take $\alpha_{\mathrm{MEe}} = \min(0.8\alpha_{\mathrm{ME}}, 0.7\alpha_{\mathrm{ME}}^{1.04})$.

T4-2) Rule R2: α_{R2} is the largest solution of $d_{\mathrm{R2}}(\alpha) = 0.22\delta$.

T5-2) Rule R2e: using α_{R2} from T4-2, take $\alpha_{\mathrm{R2e}} = 0.9\alpha_{\mathrm{R2}}$.

T6-2) Rule Me: using $\alpha_{\mathrm{MEe}}, \alpha_{\mathrm{R2e}}$ from T3-2, T5-2, take $\alpha_{\mathrm{Me}} = \min(\alpha_{\mathrm{MEe}}, \alpha_{\mathrm{R2e}})$.

Table 14. Means of error ratios in 2-iterated Tikhonov method, compared to minimums of ordinary Tikhonov method.

p	D	D,2	ME	ME,2	MEe	MEe,2	R2	R2,2	R2e	R2e,2	Me	Me,2
0	1.26	2.34	1.29	2.39	1.16	2.05	1.56	1.81	1.53	1.77	1.16	1.77
0.25	1.74	4.15	1.81	4.28	1.55	3.48	2.99	3.45	2.95	3.37	1.55	2.89
0.5	1.72	4.82	1.84	5.11	1.54	4.09	2.39	3.39	2.31	3.28	1.54	3.27
0.75	1.39	4.67	1.53	4.99	1.25	3.68	1.94	2.80	1.85	2.66	1.26	2.66
1	1.08	4.24	1.23	4.73	0.98	3.20	1.64	2.26	1.56	2.11	0.98	2.11
1.5	0.76	3.34	0.91	4.02	0.74	2.51	1.06	1.66	0.99	1.52	0.74	1.52
2	0.59	2.54	0.70	3.26	0.59	1.90	0.79	1.16	0.74	1.04	0.59	1.04
4	0.46	1.64	0.48	2.33	0.45	1.24	0.48	0.66	0.46	0.59	0.45	0.59
8	0.46	1.58	0.47	2.27	0.44	1.21	0.47	0.64	0.45	0.57	0.44	0.57
mean	1.05	3.26	1.14	3.71	0.97	2.59	1.48	1.98	1.43	1.88	0.97	1.82

Table 15. Means of error ratios in 2-iterated Tikhonov method, compared to minimums of ordinary Tikhonov method (correlated noise).

p	D	D,2	ME	ME,2	MEe	MEe,2	R2	R2,2	R2e	R2e,2	Me	Me,2
0	1.31	2.57	1.34	2.63	1.19	2.21	1.66	1.94	1.63	1.90	1.18	1.89
0.25	1.81	5.40	1.89	5.59	1.63	4.51	3.76	4.37	3.69	4.27	1.63	3.74
0.5	1.87	8.21	1.99	8.63	1.61	7.14	4.82	6.17	4.69	5.99	1.63	5.98
0.75	1.46	8.58	1.59	9.19	1.30	6.88	3.91	5.32	3.75	5.07	1.32	5.07
1	1.09	7.25	1.24	8.08	0.98	5.38	3.11	4.16	2.95	3.90	0.99	3.90
1.5	0.77	6.49	0.90	7.85	0.73	4.84	2.00	3.07	1.85	2.80	0.73	2.80
2	0.61	4.03	0.69	5.09	0.64	2.95	1.27	1.75	1.19	1.59	0.73	1.59
4	0.44	3.89	0.44	5.69	0.43	3.01	0.97	1.54	0.88	1.35	0.48	1.35
8	0.46	2.34	0.46	3.86	0.45	1.94	0.66	0.91	0.62	0.82	0.50	0.82
mean	1.09	5.42	1.17	6.29	1.00	4.32	2.46	3.25	2.36	3.08	1.02	3.02

Again, the error ratios are computed as the error of 2-iterated Tikhonov method with chosen parameter divided by the error of ordinary Tikhonov method with best parameter. The results in Table 14 show that if $x_* \in \mathcal{R}(A^*)$, then in most cases the error of 2-iterated approximation by rules MEe and Me was smaller than the error of the best single Tikhonov approximations. For rules R2, R2e, and Me at large p this holds even when the noise level is 2 times overestimated. Table 15 is an analog of Table 14 for correlated noise. The differences of Tables 14 and 15 are the same as the differences of Tables 8 and 10.

4.3.2. Minimization strategy.

Tables 16–20 compare means of error ratios by problems for rules that choose the regularization parameter by the minimization strategy of Section 3.2. In Tables 16, 17 the results for rules D, MEe, R2e, Me by problems are provided for reference (here and in the following tables the columns using δ and columns not using δ are separated by |). To test the performance of rules on other problems, we have added the results on test problems [13] of Table 2. The first row labeled 'mean' presents arithmetic means of error ratios over problems 1–10, the second row 'mean' gives the same means over problems 11–16.

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 (rule of Neubauer [69]): α_{QN} is the minimizer of the function $\varphi_{\mathrm{QN}}(\alpha)$ on the interval $[m\sigma_{\min}, 1]$, where $\sigma_{\min}$ is the smallest eigenvalue of

Table 16. Means of error ratios for $p = 0$ by problems. Problems 1–10 are from [43, 45, 46], problems 11–16 are from [13] (see Tables 1 and 2).

Probl.	D	D,2	MEe	MEe,2	R2e	R2e,2	Me	Me,2	HR	QN	BRS	QHR
1	1.41	2.37	1.38	2.26	1.67	2.00	1.38	2.00	2.76	1.56	2.61	2.79
2	1.19	1.74	1.02	1.33	1.09	1.16	1.05	1.16	960	1.95	960	1.20
3	1.35	5.88	1.38	5.15	2.75	3.48	1.37	3.48	8.18	2.18	5.23	9.54
4	1.14	2.38	1.08	2.11	1.18	1.47	1.08	1.47	2.82	1.13	2.08	1.10
5	1.06	1.53	1.03	1.37	1.05	1.15	1.04	1.15	1.70	1e+4	1.35	1.80
6	1.25	1.89	1.16	1.70	1.27	1.45	1.16	1.45	2.05	1.20	1.87	1.27
7	1.03	1.98	1.04	1.82	1.05	1.19	1.04	1.19	1e+5	1.08	1e+5	1.09
8	1.30	2.25	1.25	2.06	1.45	1.65	1.25	1.65	2.58	1.44	2.25	2.43
9	1.02	1.05	1.02	1.05	1.03	1.04	1.02	1.04	1.07	1.04	1.06	1.06
10	1.19	1.38	1.18	1.36	1.39	1.40	1.18	1.34	1.56	1.43	1.55	1.55
mean	1.19	2.25	1.15	2.02	1.39	1.60	1.16	1.59	1e+4	1e+3	1e+4	2.38
11	1.19	2.12	1.13	1.91	1.21	1.43	1.13	1.43	2.44	1.18	1.87	1.43
12	1.39	2.16	1.26	1.94	1.54	1.73	1.26	1.73	3.04	1.80	2.45	2.64
13	1.59	2.44	1.29	2.28	1.60	1.85	1.29	1.84	4.79	3.21	2.90	3.35
14	1.06	1.77	1.08	1.52	1.26	1.25	1.19	1.25	3e+3	1.60	3e+3	17.2
15	1.02	1.04	1.02	1.03	1.03	1.04	1.02	1.03	1.06	1.05	1.06	1.06
16	1.50	2.07	1.57	2.08	1.36	1.52	1.36	1.52	2.62	1.35	1.56	2.21
mean	1.29	1.93	1.23	1.79	1.33	1.47	1.21	1.47	502	1.70	502	4.65

Table 17. Means of error ratios for $p = 2$ by problems.

Probl.	D	D,2	MEe	MEe,2	R2e	R2e,2	Me	Me,2	HR	QN	BRS	QHR
1	2.93	3.21	1.11	4.08	1.09	1.31	1.08	1.31	7.22	1.72	2.69	1.30
2	3.38	2.64	1.20	1.83	1.16	1.10	1.22	1.10	7e+3	1.04	7e+3	1.05
3	3.59	3.96	1.12	4.12	1.12	1.36	1.12	1.36	7.03	1.26	3.42	1.15
4	2.83	2.27	1.07	2.13	1.06	1.05	1.07	1.05	3.69	1.12	2.23	1.12
5	2.44	2.04	1.12	1.47	1.18	1.06	1.20	1.06	2.18	7e+4	2.40	1.05
6	2.28	2.15	1.06	2.28	1.05	1.07	1.06	1.07	3.91	1.12	2.00	1.13
7	3.03	2.28	1.09	1.74	1.12	1.05	1.13	1.05	2e+5	1.06	2e+5	1.06
8	2.48	2.40	1.05	2.55	1.04	1.08	1.05	1.08	4.42	1.27	2.17	1.15
9	2.39	2.92	1.07	3.10	1.05	1.25	1.06	1.25	5.30	1.18	2.55	1.18
10	2.36	4.05	1.19	4.92	1.18	1.55	1.16	1.55	8.54	1.75	3.25	1.29
mean	2.77	2.79	1.11	2.82	1.11	1.19	1.12	1.19	2e+4	7e+3	2e+4	1.15
11	2.53	2.06	1.06	2.06	1.05	1.03	1.06	1.03	3.53	1.10	2.03	1.10
12	2.26	2.62	1.10	3.07	1.07	1.22	1.08	1.22	5.32	1.25	2.26	1.16
13	2.72	3.04	1.14	3.21	1.15	1.37	1.14	1.37	5.44	1.27	2.70	2.33
14	3.84	3.29	1.23	2.54	1.12	1.18	1.23	1.18	2e+4	1.05	2e+4	1.05
15	3.56	34.4	2.08	27.3	2.50	6.24	1.86	6.24	43.0	5.43	27.6	4.09
16	1.61	1.58	1.35	1.97	1.03	1.02	1.03	1.02	2.90	1.14	1.63	1.09
mean	2.75	7.83	1.33	6.69	1.32	2.01	1.23	2.01	3e+3	1.87	3e+3	1.81

the matrix A^*A.

T9) Rule QHR: α_{QHR} is the local minimizer of the function $\varphi_{\mathrm{Q}}(\alpha)\kappa(\alpha)$ such that the function $\varphi_{\mathrm{HR}}(\alpha)$ is minimal.

T10) The rules QC, R2C, 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 $C = 4$.

T11) Rules QC' and R2C' choose the parameter by the climbing approach in the functions $\varphi_{\mathrm{Q}}(\alpha)$ and $\varphi_{\mathrm{R2}}(\alpha)$ with $C = 4$ but unlike the rules of T10, here the choice is made only from local minimizers of the corresponding functions on the interval $(\alpha_{\min}, 1)$; endpoints are excluded.

T12) Rules Q1, D1, R21, DR21 and BRS1 choose the parameter as the first (the largest) local minimum of the functions $\varphi_{\mathrm{Q}}(\alpha)\alpha^{0.36}$, $\varphi_{\mathrm{D}}(\alpha)\kappa(\alpha)\alpha^{0.3}$, $\varphi_{\mathrm{R2}}(\alpha)\alpha^{0.42}$, $\varphi_{\mathrm{D}}(\alpha)^{0.2}\varphi_{\mathrm{R2}}(\alpha)^{0.8}\alpha^{0.36}$, and $\varphi_{\mathrm{BRS}}(\alpha)\alpha^{0.56}$, respectively.

T13) Rules HR2 and BR2 choose the parameter as the global minimizer of the functions $\varphi_{\mathrm{HR2},\tau}(\alpha)$ and $\varphi_{\mathrm{BR2},\tau}(\alpha)$ with $\tau = 0.07$.

T14) Rules QHR2 and QBR2 choose the local minimizer of the function $\varphi_{\mathrm{Q}}(\alpha)\kappa(\alpha)$ for which the functions $\varphi_{\mathrm{HR2},\tau}(\alpha)$ and $\varphi_{\mathrm{BR2},\tau}(\alpha)$ with $\tau = 0.04$ are minimal.

The best heuristic rules not using δ were comparable to rules using δ: the results were worse in case of exact δ but even better in case of 2 times overestimated noise level.

The biggest challenge to all rules turned out to be the problem 15 in case $p = 2$; the problem 14 was hard to heuristic rules Q1, R21, QHR2, QBR2 in case $p = 0$.

Note that in case of smooth solution (Table 17) the rule QHR, containing neither noise level information nor any parameters, gave very good results.

Table 18. Means of errors by problems for rules of minimization strategy, $p = 0$.

Probl.	QC	R2C	BRSC	QC'	R2C'	Q1	D1	R21	DR21	BRS1	HR2	BR2	QHR2	QBR2
1	1.56	1.84	2.61	1.55	1.83	2.24	2.05	2.20	1.75	2.02	1.85	1.95	1.75	1.73
2	1.55	1.41	1.35	1.55	1.40	1.06	1.61	1.17	1.94	1.59	1.61	1.25	1.20	1.20
3	2.18	2.11	5.23	2.15	2.09	1.89	3.75	2.13	2.48	3.60	2.70	2.80	2.08	2.89
4	1.13	1.11	2.08	1.12	1.11	1.45	1.54	1.78	1.06	1.47	1.11	1.08	1.10	1.11
5	1.34	1.19	1.35	1.30	1.19	1.30	1.20	1.78	1.28	1.14	1.14	1.26	1.16	1.17
6	1.20	1.18	1.87	1.19	1.17	1.22	1.55	1.34	1.21	1.46	1.31	1.17	1.23	1.20
7	1.08	1.08	1.61	1.07	1.09	1.19	1.25	1.46	1.10	1.17	1.08	1.12	1.09	1.08
8	1.44	1.45	2.25	1.43	1.44	1.71	1.78	1.68	1.43	1.70	1.60	1.45	1.49	1.42
9	1.04	1.05	1.06	1.04	1.05	1.06	1.05	1.06	1.03	1.04	1.05	1.05	1.05	1.04
10	1.43	1.42	1.55	1.43	1.42	1.83	1.48	1.81	1.47	1.48	1.47	1.47	1.47	1.48
mean	1.39	1.38	2.10	1.38	1.38	1.49	1.72	1.64	1.47	1.67	1.49	1.46	1.36	1.43
11	1.18	1.16	1.87	1.17	1.16	1.36	1.45	1.50	1.14	1.41	1.20	1.13	1.18	1.16
12	2.01	1.92	2.45	1.94	1.91	2.27	1.81	2.26	1.51	1.77	2.19	2.10	2.07	1.97
13	3.25	3.24	2.90	3.25	3.24	4.15	3.42	4.17	3.14	1.92	3.39	2.71	2.05	2.05
14	1.85	1.76	2.07	1.84	1.76	19.8	1.87	19.6	1.86	1.74	2.06	1.58	16.7	16.7
15	1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.06	1.06
16	1.35	1.34	1.56	1.35	1.34	1.59	1.32	1.62	1.27	1.33	1.50	1.29	1.41	1.31
mean	1.78	1.75	1.99	1.77	1.75	5.04	1.82	5.04	1.66	1.54	1.90	1.65	4.08	4.04

Table 19. Means of errors by problems for rules of minimization strategy, $p = 0$ (correlated noise).

Probl.	R2C	QC	BRSC	R2C'	QC'	Q1	D1	R21	DR21	BRS1	HR2	BR2	QHR2	QBR2
1	1.52	1.80	2.51	1.52	1.80	2.23	2.00	2.20	1.74	1.96	1.82	1.91	1.65	1.71
2	1.22	1.18	1.32	1.21	1.17	1.15	1.46	1.28	1.77	1.44	1.78	1.30	1.16	1.16
3	2.31	2.05	5.85	2.28	2.06	1.93	4.11	2.19	2.62	3.97	2.78	2.88	2.35	2.77
4	1.16	1.15	2.29	1.15	1.15	1.54	1.71	2.26	1.16	1.63	1.20	1.15	1.14	1.14
5	1.39	1.32	1.51	1.35	1.32	1.37	2.43	1.98	3.87	1.60	1.26	1.60	1.23	1.31
6	1.22	1.18	1.94	1.21	1.18	1.25	1.61	1.31	1.28	1.54	1.40	1.26	1.26	1.24
7	1.08	1.06	1.79	1.06	1.07	1.26	2.27	2.05	4.18	1.85	1.90	1.09	1.07	1.06
8	1.35	1.32	2.29	1.33	1.32	1.57	1.80	1.57	1.40	1.70	1.52	1.48	1.49	1.41
9	1.04	1.05	1.06	1.04	1.05	1.06	1.05	1.06	1.03	1.04	1.05	1.05	1.05	1.04
10	1.44	1.44	1.56	1.44	1.44	1.84	1.49	1.82	1.48	1.49	1.48	1.48	1.47	1.48
mean	1.37	1.36	2.21	1.36	1.36	1.52	1.99	1.77	2.05	1.82	1.62	1.52	1.39	1.43
11	1.17	1.17	2.21	1.17	1.17	1.45	1.72	1.69	1.28	1.66	1.30	1.20	1.22	1.19
12	2.40	2.28	3.13	2.33	2.27	2.65	2.26	2.61	1.88	2.22	2.71	2.53	2.46	2.39
13	3.07	3.06	2.78	3.07	3.07	3.96	3.32	4.01	3.04	1.87	3.29	2.60	2.04	2.02
14	2.02	1.89	2.23	2.01	1.88	23.1	2.05	22.9	2.00	1.81	2.44	1.80	21.3	21.3
15	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05
16	1.47	1.44	1.87	1.46	1.44	1.72	1.53	1.76	1.38	1.54	1.65	1.36	1.51	1.41
mean	1.86	1.82	2.21	1.85	1.81	5.66	1.99	5.67	1.77	1.69	2.07	1.76	4.93	4.89

Table 20. Means of errors by problems for rules of minimization strategy, $p = 2$.

Probl.	R2C	QC	BRSC	R2C'	QC'	Q1	D1	R21	DR21	BRS1	HR2	BR2	QHR2	QBR2
1	1.72	1.90	2.69	1.76	1.95	1.75	2.20	1.90	1.77	1.95	1.23	1.34	1.30	1.34
2	1.04	1.17	4.46	1.08	1.23	1.23	11.3	1.52	16.1	6.64	1.11	1.10	1.05	1.05
3	1.26	1.38	3.42	1.29	1.44	1.64	2.94	1.83	2.12	2.58	1.13	1.13	1.17	1.15
4	1.12	1.25	2.23	1.15	1.30	1.56	2.46	2.03	2.47	2.26	1.14	1.15	1.12	1.12
5	1.04	1.17	2.40	1.07	1.22	1.32	3.13	1.67	3.25	2.92	1.12	1.11	1.05	1.05
6	1.12	1.25	2.00	1.16	1.30	1.70	1.98	1.97	1.95	1.84	1.14	1.15	1.13	1.13
7	1.06	1.17	2.54	1.09	1.23	1.26	3.23	1.49	3.20	3.03	1.12	1.11	1.06	1.06
8	1.27	1.40	2.17	1.31	1.45	1.84	2.09	2.04	1.99	1.90	1.17	1.24	1.21	1.21
9	1.18	1.33	2.55	1.22	1.38	1.60	2.26	1.92	1.93	1.97	1.20	1.20	1.18	1.18
10	1.75	1.86	3.25	1.78	1.91	1.87	2.38	2.00	1.72	2.00	1.26	1.46	1.29	1.48
mean	1.26	1.39	2.77	1.29	1.44	1.58	3.39	1.84	3.65	2.71	1.16	1.20	1.16	1.18
11	1.10	1.22	2.03	1.13	1.27	1.59	2.25	1.97	2.34	2.10	1.12	1.15	1.10	1.10
12	1.25	1.39	2.26	1.28	1.44	1.65	1.99	1.87	1.74	1.78	1.16	1.24	1.17	1.21
13	1.27	1.36	2.70	1.29	1.41	1.60	2.42	1.83	1.99	2.19	1.31	1.38	1.31	1.38
14	1.05	1.17	3.41	1.08	1.24	1.29	4.05	1.55	5.54	3.65	1.11	1.11	1.05	1.05
15	4.94	5.10	27.6	5.00	5.17	4.38	16.0	4.58	5.03	11.4	4.05	3.23	4.17	4.82
16	1.14	1.22	1.63	1.16	1.25	1.32	2.03	1.47	2.20	1.98	1.10	1.17	1.12	1.12
mean	1.79	1.91	6.61	1.82	1.96	1.97	4.79	2.21	3.14	3.85	1.64	1.55	1.65	1.78

4.3.3. Rule R2e.

Tables 21, 22 compare the error ratios of rules D and R2e for different values of noise level uncertainty coefficient d. In case of overestimated noise level ($d > 1$) the rule R2e is significantly better than the discrepancy principle. In contrast to the discrepancy principle and other rules, the rule R2e also allows moderate underestimation of the noise level. The same conclusions can be made about rule R2e in 2-extrapolated Tikhonov method (see Table 59).

Comparison of Tables 21 and 22 shows that for rule R2e the error ratios in case of correlated noise are larger than the error ratios in case of uncorrelated noise but the advantage over discrepancy principle still holds.

4.3.4. Rule DM.

Tables 23, 24 show the results for rule DM with parameters $c_1 = 0.002$, $c_2 = 0.03$ in Tikhonov method at various coefficients d of noise level ill-estimation. It can be seen that the means of error ratios do not change with d as much as in rule R2e, allowing wider range of over- or underestimation of the noise level. Table 25 compares some particular sets of parameters c_1, c_2 in rule DM for $p = 0$ and $p = 2$. If we have more information about the noise level, parameters of the first row may be used, since they give smaller error ratios for d close to 1. If we do not have much information, parameters for which the rule is less sensitive to ill-estimation of the noise level should be preferred.

Table 21. Means of error ratios for rules D (upper part) and R2e (lower part) in case, where the estimated noise level is d times larger than the actual noise level.

	$p \setminus d$	0.5	0.6	0.8	1	1.3	1.6	2	3	5	10
Rule D	0				1.19	1.82	2.05	2.23	2.56	3.06	3.69
	0.25				1.64	2.91	3.41	3.82	4.62	5.67	6.75
	0.5				1.69	3.01	3.62	4.09	5.00	6.44	8.42
	0.75	$\gg 1$			1.61	2.61	3.19	3.65	4.57	6.26	8.44
	1				1.81	2.28	2.76	3.14	4.01	5.69	7.80
	1.5				2.50	2.25	2.56	2.77	3.38	4.78	6.54
	2				2.83	2.37	2.61	2.76	3.27	4.52	6.08
	4				3.01	2.45	2.70	2.80	3.26	4.49	5.95
	8				3.03	2.46	2.69	2.79	3.24	4.45	5.89
	mean				2.14	2.46	2.84	3.12	3.77	5.04	6.62
Rule R2e	0	3.84	1.75	1.36	1.39	1.46	1.52	1.59	1.76	2.09	2.49
	0.25	4.77	2.38	2.58	2.63	2.80	2.91	3.04	3.34	3.88	4.67
	0.5	5.17	2.49	1.98	2.02	2.19	2.37	2.59	3.11	3.79	4.81
	0.75	4.50	2.29	1.54	1.57	1.70	1.87	2.08	2.55	3.28	4.46
	1	4.73	1.93	1.29	1.33	1.41	1.51	1.65	2.01	2.74	4.00
	1.5	4.64	1.70	1.16	1.14	1.16	1.21	1.28	1.53	2.05	2.90
	2	4.43	1.53	1.14	1.11	1.11	1.13	1.17	1.35	1.75	2.37
	4	4.07	1.54	1.14	1.11	1.10	1.11	1.14	1.29	1.64	2.13
	8	4.08	1.54	1.14	1.11	1.10	1.11	1.14	1.28	1.63	2.10
	mean	4.47	1.91	1.48	1.49	1.56	1.63	1.74	2.02	2.54	3.32

Table 22. Means of error ratios for rules D (upper part) and R2e (lower part) in case, where the estimated noise level is d times larger than the actual noise level (correlated noise).

	$p \setminus d$	0.5	0.6	0.8	1	1.3	1.6	2	3	5	10
Rule D	0				1.23	1.97	2.22	2.43	2.80	3.38	4.11
	0.25				1.69	3.83	4.49	4.98	5.95	7.22	8.69
	0.5				1.82	5.89	6.86	7.49	8.73	10.7	13.4
	0.75	$\gg 1$			1.69	4.59	5.65	6.52	8.25	11.4	15.4
	1				1.96	3.59	4.45	5.02	6.24	8.46	12.0
	1.5				2.73	3.24	4.08	4.78	6.50	10.0	14.3
	2				3.02	2.93	3.54	3.94	4.97	7.25	9.29
	4				3.59	3.92	4.88	5.44	6.96	10.2	14.5
	8				3.27	2.83	3.36	3.65	4.61	6.70	9.97
	mean				2.33	3.64	4.39	4.92	6.11	8.37	11.3
Rule R2e	0	5.46	1.99	1.49	1.47	1.55	1.61	1.69	1.89	2.27	2.75
	0.25	4.72	2.85	3.22	3.29	3.47	3.61	3.81	4.22	5.00	6.01
	0.5	5.23	4.30	4.07	4.17	4.43	4.69	5.03	5.77	6.88	8.21
	0.75	4.84	3.07	2.69	2.84	3.14	3.40	3.80	4.68	6.00	8.03
	1	3.39	2.39	2.02	2.20	2.39	2.61	2.92	3.57	4.84	6.79
	1.5	5.48	2.19	1.57	1.61	1.73	1.85	2.05	2.57	3.58	5.48
	2	4.89	1.82	1.44	1.42	1.43	1.47	1.56	1.78	2.43	3.59
	4	3.66	1.97	1.43	1.47	1.54	1.63	1.82	2.21	3.19	4.40
	8	3.88	1.68	1.29	1.27	1.28	1.31	1.39	1.58	2.11	2.87
	mean	4.62	2.47	2.14	2.19	2.33	2.47	2.67	3.14	4.03	5.35

Table 23. Means of error ratios in Tikhonov method with Rule DM, $c_1 = 0.002$, $c_2 = 0.03$ for $p = 0$ (left), $p = 2$ (right), where the used noise level is d times the actual noise level.

	Case $p = 0$, values of d							Case $p = 2$, values of d						
Probl.	0.01	0.1	0.5	1	2	10	100	0.01	0.1	0.5	1	2	10	100
1	1.46	1.46	1.46	1.46	1.49	1.69	2.51	1.93	1.93	1.93	1.84	1.74	1.33	3.33
2	1.56	1.56	1.34	1.08	1.08	1.07	1.25	1.22	1.22	1.22	1.22	1.22	1.22	1.43
3	2.02	2.02	2.02	2.02	2.02	1.84	5.88	1.55	1.55	1.55	1.55	1.55	1.26	3.23
4	1.12	1.12	1.12	1.12	1.12	1.11	1.62	1.33	1.33	1.33	1.33	1.33	1.25	1.75
5	1.66	1.16	1.16	1.10	1.10	1.10	1.17	1.21	1.21	1.21	1.21	1.21	1.21	1.15
6	1.16	1.16	1.16	1.16	1.16	1.16	1.44	1.33	1.33	1.33	1.33	1.33	1.29	1.82
7	1.11	1.11	1.11	1.11	1.11	1.11	1.36	1.21	1.21	1.21	1.21	1.21	1.21	1.37
8	1.39	1.39	1.39	1.39	1.39	1.46	2.06	1.49	1.49	1.49	1.49	1.49	1.26	2.02
9	1.03	1.03	1.03	1.03	1.03	1.03	1.05	1.42	1.42	1.42	1.42	1.42	1.26	2.38
10	1.42	1.42	1.42	1.42	1.42	1.47	1.54	2.20	2.20	2.20	1.85	1.62	1.28	3.85
mean	1.39	1.34	1.32	1.29	1.29	1.30	1.99	1.49	1.49	1.49	1.44	1.41	1.26	2.23
11	1.16	1.16	1.16	1.16	1.16	1.16	1.56	1.30	1.30	1.30	1.30	1.30	1.25	1.69
12	1.43	1.43	1.43	1.43	1.43	1.47	2.13	1.50	1.50	1.50	1.50	1.50	1.26	2.46
13	2.41	2.41	2.41	2.41	2.41	2.42	3.80	1.46	1.46	1.46	1.46	1.46	1.33	2.81
14	3.28	1.83	1.66	1.56	1.45	1.36	1.70	1.23	1.23	1.23	1.23	1.23	1.21	1.92
15	1.05	1.05	1.05	1.05	1.06	1.06	1.06	6.04	6.04	5.22	4.38	3.60	3.04	18.3
16	1.36	1.36	1.36	1.36	1.33	1.35	2.21	1.32	1.32	1.32	1.32	1.32	1.16	1.81
mean	1.78	1.54	1.51	1.50	1.47	1.47	2.08	2.14	2.14	2.01	1.87	1.74	1.54	4.83

Table 24. Means of error ratios in Tikhonov method with Rule DM, $c_1 = 0.002$, $c_2 = 0.03$ for $p = 0$ (left), $p = 2$ (right), where the used noise level is d times the actual noise level (correlated noise).

	values of d for $p = 0$							values of d for $p = 2$						
Probl.	0.01	0.1	0.5	1	2	10	100	0.01	0.1	0.5	1	2	10	100
1	1.44	1.44	1.44	1.48	1.50	1.75	2.62	1.82	1.81	1.79	1.72	1.72	1.93	5.69
2	2.09	1.32	1.12	1.08	1.08	1.10	1.57	2.10	1.51	1.33	1.22	1.22	1.26	2.25
3	2.03	2.03	2.03	1.98	1.98	2.33	7.29	1.78	1.78	1.78	1.79	1.80	1.79	5.12
4	1.16	1.16	1.16	1.16	1.16	1.18	1.97	1.30	1.30	1.30	1.29	1.29	1.42	3.11
5	1.83	1.32	1.18	1.17	1.20	1.20	1.44	1.27	1.27	1.27	1.27	1.26	1.28	1.92
6	1.17	1.17	1.17	1.17	1.17	1.22	1.59	1.29	1.29	1.29	1.28	1.29	1.44	3.40
7	1.10	1.10	1.10	1.10	1.10	1.11	1.66	1.24	1.24	1.24	1.23	1.23	1.32	2.49
8	1.29	1.29	1.29	1.29	1.30	1.45	2.13	1.52	1.48	1.48	1.47	1.47	1.59	4.17
9	1.03	1.03	1.03	1.03	1.03	1.03	1.06	1.30	1.30	1.28	1.29	1.29	1.45	3.73
10	1.44	1.44	1.44	1.44	1.44	1.49	1.60	3.46	3.46	3.48	3.34	3.39	3.22	8.07
mean	1.46	1.33	1.30	1.29	1.30	1.39	2.29	1.71	1.64	1.62	1.59	1.60	1.67	4.00
11	1.19	1.19	1.19	1.19	1.18	1.26	2.01	1.28	1.28	1.28	1.27	1.27	1.40	3.23
12	1.70	1.70	1.71	1.72	1.72	1.93	2.92	1.47	1.47	1.44	1.45	1.45	1.59	4.61
13	2.04	2.04	2.08	2.09	2.09	2.43	3.87	1.60	1.60	1.57	1.58	1.58	2.03	5.09
14	3.21	2.03	1.74	1.57	1.51	1.43	1.93	2.04	1.30	1.30	1.29	1.30	1.38	2.79
15	1.05	1.05	1.05	1.05	1.05	1.05	1.05	19.5	19.5	18.9	17.5	16.9	15.2	23.4
16	1.48	1.48	1.48	1.47	1.46	1.58	3.07	1.35	1.35	1.30	1.27	1.21	1.18	3.20
mean	1.78	1.58	1.54	1.52	1.50	1.61	2.48	4.54	4.42	4.30	4.06	3.95	3.80	7.05

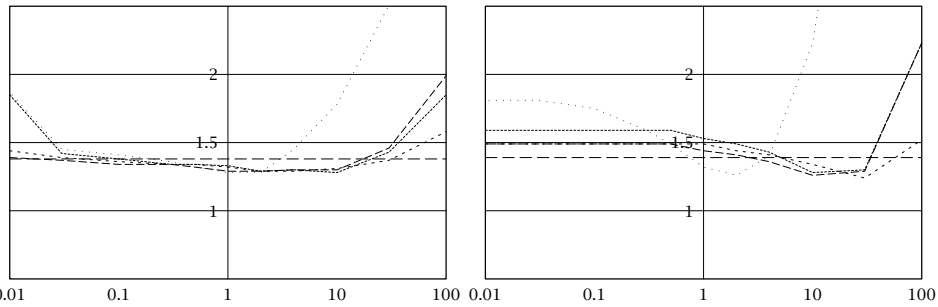
Table 25. Averages of error ratios over problems of Hansen for rule DM with different c_1, c_2; $p = 0$ (upper) and $p = 2$ (lower).

				values of d										
p	Nr	c_1	c_2	0.01	0.03	0.1	0.3	0.5	1	2	4	10	30	100
0	I	0.02	0.14	1.87	1.45	1.41	1.36	1.32	1.28	1.28	1.45	1.78	2.51	3.49
0	II	0.002	0.07	1.85	1.42	1.38	1.34	1.34	1.33	1.29	1.30	1.28	1.43	1.85
0	III	0.002	0.03	1.39	1.37	1.34	1.34	1.32	1.29	1.29	1.30	1.30	1.46	1.99
0	IV	0.001	0.03	1.44	1.39	1.36	1.34	1.34	1.32	1.29	1.29	1.31	1.37	1.58
2	I	0.02	0.14	1.81	1.81	1.75	1.59	1.49	1.32	1.26	1.40	2.23	4.61	9.38
2	II	0.002	0.07	1.59	1.59	1.59	1.59	1.59	1.53	1.49	1.43	1.28	1.30	2.23
2	III	0.002	0.03	1.49	1.49	1.49	1.49	1.49	1.44	1.41	1.36	1.26	1.29	2.23
2	IV	0.001	0.03	1.49	1.49	1.49	1.49	1.49	1.49	1.44	1.41	1.34	1.24	1.52

The rule DM differs from the rule R2C essentially only by the strategy of choosing the endpoint of minimization interval: rule DM uses noise level information and solves the equation $d_{\mathrm{R1},1/2}(\alpha) = C\delta$, rule R2C determines the endpoint by increase condition of $\varphi_{\mathrm{R2}}(\alpha)$, not using δ. Figure 3 compares the rule DM with rule R2C, illustrating dependence of error on d. In practice this comparison may help to decide, whether additional efforts to more precisely estimate the actual noise level will be worth of increased accuracy. In Tables 18 and 20 the rule R2C gives averages 1.38 for $p = 0$ and 1.39 for $p = 2$, hence DM is superior over R2C, if $d \in [0.1, 10]$ in case $p = 0$ or if $d \in [3, 30]$ in case $p = 2$.

For analogous results in extrapolated Tikhonov approximation, see Section 4.8.4.

Figure 3. Error ratios in Tikhonov method for rules R2C and DM: $p = 0$ (left) and $p = 2$ (right); ····· DM I, ····· DM II, – – – DM III, – · – · – DM IV, — — R2C.



In practice we may have an approximate knowledge about how much the supposed noise level $d\delta$ differs from the actual noise level δ. This knowledge may be used to choose the constants c_1 and c_2 in rule DM. In Table 26 each cell contains constants c_1 and c_2 that give the smallest maximum of averages over all Hansen's problems with $p = 0$, if we assume that $d \in [d_{\min}, d_{\max}]$. Table 27 shows the corresponding minimums and maximums of these averages. Computations were made on discrete set of values $c_1 = 10^{-4}, 10^{-3.8}, \ldots, 10^{-1}$ and $c_2 = 10^{-1}, 10^{-1.1}, \ldots, 10^{-3}$.

Table 26. Values of constants c_1 (lower left) and c_2 (upper right) in rule DM, giving the smallest maximum of averages over range $[d_{\min}, d_{\max}]$ of noise ill-estimation coefficient d over problems of Hansen with $p = 0$.

$d_{\min}$ \ $d_{\max}$	1	2	4	10	30	100
0.01	5.0e-2 2.5e-2	1.0e-2 1.0e-2	1.0e-2 6.3e-3	1.0e-2 2.5e-3	1.0e-2 1.0e-3	2.0e-3 4.0e-4
0.03	2.0e-2 2.5e-2	5.0e-2 1.0e-2	3.2e-2 6.3e-3	1.0e-2 2.5e-3	1.0e-2 1.0e-3	1.3e-2 4.0e-4
0.1	5.0e-2 2.5e-2	2.5e-2 1.0e-2	2.0e-2 6.3e-3	5.0e-2 2.5e-3	1.0e-2 1.0e-3	1.0e-2 2.5e-4
0.3	1.0e-2 4.0e-3	5.0e-2 1.0e-2	4.0e-2 6.3e-3	2.0e-2 2.5e-3	5.0e-2 1.0e-3	1.0e-2 2.5e-4
0.5	5.0e-2 6.3e-3	1.0e-2 2.5e-3	5.0e-2 6.3e-3	2.5e-2 2.5e-3	5.0e-2 1.0e-3	1.0e-2 2.5e-4
1	1.0e-1 2.5e-2	5.0e-2 2.5e-3	1.0e-2 1.0e-3	5.0e-2 2.5e-3	2.5e-2 1.0e-3	5.0e-2 2.5e-4

Table 27. Minimal and maximal averages in rule DM over all problems of Hansen, corresponding to c_1 and c_2 in Table 26, $p = 0$.

$d_{\min}$ \ $d_{\max}$	1	2	4	10	30	100
0.01	1.28..1.34	1.29..1.34	1.29..1.36	1.29..1.39	1.29..1.41	1.31..1.44
0.03	1.29..1.33	1.28..1.34	1.29..1.34	1.29..1.36	1.29..1.39	1.29..1.42
0.1	1.28..1.30	1.29..1.31	1.29..1.33	1.28..1.34	1.29..1.34	1.29..1.39
0.3	1.29..1.29	1.28..1.30	1.29..1.30	1.29..1.33	1.28..1.34	1.29..1.36
0.5	1.29..1.29	1.29..1.30	1.28..1.30	1.29..1.31	1.28..1.34	1.29..1.34
1	1.28..1.28	1.29..1.29	1.29..1.29	1.28..1.30	1.29..1.31	1.28..1.34

We also made computations with the variant of rule DM, which minimizes $\varphi_{\mathrm{Q}}(\alpha)\alpha^{c_2}$ instead of $\varphi_{\mathrm{R2}}(\alpha)\alpha^{c_2}$ on the second step but in case of very small or very large d the results were essentially worse.

4.4. Results in Lavrentiev method

For studying rules in Lavrentiev method we limit our test set to self-adjoint problems of Hansen, problems of [13] are included only in Table 33.

4.4.1. Analogs of monotone error rule.

Tables 28–30 show the results of analogs of the monotone error rule in Lavrentiev method, where the regularization parameter is chosen as the solution α_{R} of equation $d_{\mathrm{R}}(\alpha) = C\delta$, where R is GN, MD, MEa, MEn, MEk, MEke, MEaq, MEaql. Values of additional parameters in functions $d_{\mathrm{R}}(\alpha)$ are given in the first column of every table, the parameter q was always $1/1.2$. For comparison the tables also contain the results for iterated Lavrentiev method: 2-iterated with the modified discrepancy principle and 3-iterated Lavrentiev method with the discrepancy principle (both with $\overline{x} = 0$).

Table 28. Averages and maximums of error ratios in Lavrentiev method for various analogs of the monotone error rule, $p = 0$ (uncorrelated noise left, correlated noise right).

Rule	Opt C	Avg	Max	Monind	Avg	Max	Monind
GN	1.059	1.19	6.31	0.145	1.34	28.6	0.135
MD	1.143	1.19	6.31	0.172	1.33	28.6	0.154
MEa	1.364	1.07	2.26	0.145	1.14	10.1	0.138
MEn, $\nu = 0.17$	1.096	1.05	1.97	0.096	1.11	8.59	0.094
MEk, $k = 2$	1.251	1.10	2.14	0.235	1.15	8.59	0.221
MEke, $k = 1$	1.421	1.05	2.26	0.046	1.12	10.1	0.042
MEke, $k = 2$	1.280	1.05	2.26	0.026	1.12	10.1	0.023
MEaq	1.209	1.08	2.26	0.159	1.14	10.1	0.150
MEaql, $l = 5$	1.004	1.01	1.25	0.036	1.05	5.21	0.042
2-Lavr w/ MD	1.187	0.98	4.30	0.000	1.08	19.6	0.000
3-Lavr w/ D	1.530	0.94	2.68	0.000	1.00	12.1	0.000

Table 29. Averages and maximums of error ratios in Lavrentiev method for various analogs of the monotone error rule, $p = 1$ (uncorrelated noise left, correlated noise right).

Rule	Opt C	Avg	Max	Monind	Avg	Max	Monind
GN	1.021	1.20	2.29	0.158	1.23	3.46	0.099
MD	1.263	1.02	1.08	0.105	1.08	1.70	0.209
MEa	1.583	1.03	1.09	0.135	1.07	1.17	0.209
MEn, $\nu = 0.4$	1.287	1.01	1.05	0.096	1.02	1.16	0.086
MEk, $k = 2$	1.365	1.07	1.16	0.222	1.11	1.26	0.278
MEke, $k = 1$	1.661	1.02	1.15	0.025	1.02	1.16	0.087
MEke, $k = 2$	1.416	1.01	1.08	0.005	1.01	1.27	0.041
MEaq	1.430	1.05	1.13	0.164	1.08	1.20	0.220
MEaql, $l = 4$	1.041	1.01	1.05	0.097	1.02	1.14	0.092
2-Lavr w/ MD	1.352	0.58	1.00	0.000	0.61	1.02	0.000
3-Lavr w/ D	1.856	0.63	1.23	0.000	0.65	0.98	0.000

Table 30. Averages and maximums of error ratios in Lavrentiev method for various analogs of the monotone error rule in case of rough δ, $p = 0$.

	$d = 2$		$d = 3$		$d = 10$		$d = 100$	
Rule	Avg	Max	Avg	Max	Avg	Max	Avg	Max
GN	1.46	8.18	1.66	10.4	2.49	18.1	4.97	33.6
MD	1.38	8.18	1.56	10.4	2.34	16.7	4.80	33.6
MEa	1.19	4.11	1.29	5.50	1.83	12.8	3.57	28.6
MEn, $\nu = 0.17$	1.27	3.54	1.41	4.76	2.00	12.8	3.74	30.2
MEk, $k = 2$	1.17	3.59	1.27	5.50	1.81	12.8	3.53	28.6
MEke, $k = 1$	1.22	4.11	1.36	6.31	1.93	14.0	3.75	30.2
MEke, $k = 2$	1.27	4.11	1.40	5.50	2.02	14.0	3.76	30.2
MEaq	1.18	4.11	1.27	5.50	1.81	12.8	3.54	28.6
MEaql, $l = 5$	1.20	1.79	1.32	2.61	1.86	8.18	3.46	27.0
2-Lavr w/ MD	1.14	6.19	1.30	8.46	1.99	15.1	4.07	31.7
3-Lavr w/ D	1.01	4.25	1.14	6.31	1.71	12.8	3.53	29.1

The optimal constants C in these rules were found by optimization on self-adjoint problems of Hansen at uncorrelated noise, over all δ's and 10 runs. Tables 28 and 29 correspond to cases $p = 0$, and $p = 1$, respectively (in the latter case the original solution x_* of each problem was replaced by Ax_*, and the right-hand side was computed as $A(Ax_*)$). In addition to averages and maximums of error ratios over self-adjoint problems of Hansen, averages of the monotonicity indicators $\max(\alpha_{\text{mon}} - \alpha_{\text{R}}, 0)/\alpha_{\text{mon}}$, where $\alpha_{\text{mon}} = \min\{\alpha' : \frac{\mathrm{d}}{\mathrm{d}\alpha}\|u_\alpha - u_*\| \geq 0, \alpha \geq \alpha'\}$, are given in additional columns. The smaller this indicator is, the larger is the number of cases, where the error is a monotonically increasing function of α for $\alpha \geq \alpha_{\text{R}}$. Four zeros in monotonicity indicators show that in corresponding approximations the chosen parameter α_{R} was always larger than α_{mon} in Lavrentiev method.

In Table 28 the rules GN and MD gave the largest averages and maximums of error ratios, in Table 29 the rule GN gave large averages and maximums, the modified discrepancy principle gave large maximum of error ratios. As expected, the methods with higher qualification (2 or 3 times iterated Lavrentiev methods and extrapolated Lavrentiev methods with 2 or 3 terms have the qualification 2 or 3, respectively) had small averages of error ratios in smooth case.

Rules MEa and MEaq were good in Lavrentiev method but their modifications were even slightly better. Note that in smooth case the modified discrepancy principle in 3-extrapolated Lavrentiev approximation gave 20% smaller averages and maximums of error ratios than the discrepancy principle (these results are not included in tables).

In Table 30 we provide results for non-smooth case ($p = 0$), if the supposed noise level is $d\delta$ and the rules are applied with the optimal constant C. Again, two first rules gave the largest numbers. Setting aside the first two rules, other rules gave smaller maximums of error ratios than 2-iterated Lavrentiev method in case of all d and 3-iterated Lavrentiev method in case $d = 1, 2$. The smallest values of averages and maximums of error ratios in Lavrentiev method were obtained by rules MEaql, MEn with $\nu = 0.17$ and MEke with $k = 2$; the rule MEaql had especially good performance in non-smooth case. However, the set of problems was not large enough for more far-reaching conclusions.

Note that for choosing the parameter α in m-iterated Lavrentiev method many rules require additional iterations for computation of iterated Lavrentiev approximations $x_{n;\alpha}$ with $n > m$. But then $x_{n;\alpha}$ itself can be considered to be an approximate solution of (2.1). Our rule MEa requires computation of $x_{m+2;\alpha}$ but if $x_{m+2;\alpha}$ has been computed, one can take the approximate solution to be $x_{m+1;\alpha}$ with choice of α from the modified discrepancy principle

$$\|Ax_{m+2;\alpha} - y\| = C\delta, \qquad C \geq 1. \tag{4.1}$$

or $x_{m+2;\alpha}$ with choice of α from the discrepancy principle $\|Ax_{m+2;\alpha} - y\| = C\delta$, giving the same α as (4.1). The corresponding means of error ratios are shown in two last rows of Tables 28–30. If the solution has source-like representation (2.2), then for $x_{m+1;\alpha}$ and for $x_{m+2;\alpha}$ with α from (4.1) the error estimate (3.1) holds with $p \le m+1$, for $x_{m;\alpha}$ with the proper α the error estimate (3.1) holds with $p \le m$ (but remind large maximums in Table 30).

Instead of iterated approximations, one can also use extrapolated approximations. Note that the rule MEk needs computing of $x_{m+k+1;\alpha}$ but rules MEaq, MEaql use $x_{m+1;\alpha}$ with different α and the rule MEke uses $x_{m;\alpha}$ with different α. Actually, if $x_{m;\alpha}$ with different α's are available and the solution is smooth, then using extrapolated iterated Lavrentiev approximation $x_{m;\alpha_1,\dots,\alpha_n}$ instead of $x_{m;\alpha}$ is strongly recommended due to much higher qualification (the qualification is nm, see [28]).

4.4.2. Rule DM.

Tables 31 and 32 show the results for the rule DM with parameters $c_1 = 2.5 \cdot 10^{-6}$, $c_2 = 0.25$ at various levels d of noise level ill-estimation. The results were very good both in case of correlated and uncorrelated noise.

Table 31. Means of error ratios in Lavrentiev method with Rule DM, $c_1 = 2.5 \cdot 10^{-6}$, $c_2 = 0.25$ for $p = 0$ (left), $p = 1$ (right), where the used noise level is d times the actual noise level.

	Case $p = 0$, values of d							Case $p = 1$, values of d						
Probl.	0.01	0.1	0.5	1	2	10	100	0.01	0.1	0.5	1	2	10	100
2	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.03	1.03	1.03	1.03	1.03	1.03	1.03
3	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.02	1.02	1.02	1.02	1.02	1.02	1.02
4	1.05	1.05	1.05	1.05	1.05	1.05	1.06	1.04	1.04	1.04	1.04	1.04	1.04	1.04
7	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.03	1.03	1.03	1.03	1.03	1.03	1.03
8	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.02	1.02	1.02	1.02	1.02	1.02	1.02
mean	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.03	1.03	1.03	1.03	1.03	1.03	1.03

Table 32. Means of error ratios in Lavrentiev method with Rule DM, $c_1 = 2.5 \cdot 10^{-6}$, $c_2 = 0.25$ for $p = 0$ (left), $p = 1$ (right), where the used noise level is d times the actual noise level (correlated noise).

	Case $p = 0$, values of d							Case $p = 1$, values of d						
Probl.	0.01	0.1	0.5	1	2	10	100	0.01	0.1	0.5	1	2	10	100
2	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.03	1.03	1.03	1.03	1.03	1.03	1.03
3	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.03	1.03	1.03	1.03	1.03	1.03	1.03
4	1.08	1.08	1.08	1.08	1.09	1.10	1.12	1.08	1.08	1.08	1.08	1.08	1.06	1.04
7	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.03	1.03	1.03	1.03	1.03	1.03	1.03
8	1.01	1.01	1.01	1.01	1.01	1.01	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02
mean	1.03	1.03	1.03	1.03	1.03	1.03	1.04	1.04	1.04	1.04	1.04	1.04	1.03	1.03

4.4.3. General comparison of rules.

Table 33 contains the means of error ratios in Lavrentiev method by problems. The following rules are included.

Lv1) Rule MD: α_{MD} is the solution of the equation $d_{\mathrm{MD}}(\alpha) = 1.143\delta$.

Lv2) Rules MEa, MEn, MEaql: α_{MEa} is the solution of $d_{\mathrm{MEa}}(\alpha) = 1.364\delta$; α_{MEn} is the solution of $d_{\mathrm{MEn}}(\alpha) = 1.096\delta$ with $\nu = 0.17$, and α_{MEaql} is the solution of $d_{\mathrm{MEaql}}(\alpha) = 1.004\delta$ with $l = 5$ and $q = 1/1.2$.

Lv3) Rules QC and QmC choose the parameters α_{QC} and α_{QmC} as the global minimizers of the functions $\varphi_{\mathrm{Qq}}(\alpha)(1 + \alpha\|A\|^{-1})$ and $\alpha^{-1}\|r_{m+1;\alpha}\| \cdot (r_{m+1;\alpha}, r_{m+2;\alpha})^{1/2}/(r_{m+2;\alpha}, r_{m+3;\alpha})^{1/2}(1 + \alpha\|A\|^{-1})$, respectively, with $q = 1/1.2$, using the climbing approach with $C = 1.5$ and $C = 2.5$.

In general, the error ratios given by Lavrentiev method are better than those found by Tikhonov method but note that Tables 4–7 show somewhat lower overall potential of this method. Also Lavrentiev method is applicable only to self-adjoint problems.

Table 33. Means of error ratios in Lavrentiev method by problems, $p = 0$.

Probl.	MD	MD,2	MEa	MEa,2	MEn	MEn,2	MEaql	MEaql,2	QC	QmC
2	1.76	2.17	1.21	1.59	1.18	1.56	1.04	1.29	1.07	1.05
3	1.07	1.16	1.05	1.09	1.03	1.20	1.00	1.18	1.01	1.00
4	1.04	1.19	1.04	1.08	1.01	1.20	1.00	1.19	1.01	1.00
7	1.03	1.21	1.04	1.08	1.01	1.21	1.01	1.20	1.00	1.00
8	1.02	1.15	1.01	1.09	1.01	1.16	1.00	1.15	1.03	1.02
mean	1.19	1.38	1.07	1.19	1.05	1.27	1.01	1.20	1.03	1.01
11	1.03	1.16	1.03	1.09	1.01	1.19	1.00	1.17	1.01	1.00
12	1.01	1.11	1.01	1.07	1.00	1.12	1.00	1.10	1.39	1.15
14	1.20	1.46	1.05	1.20	1.04	1.26	1.01	1.17	1.31	1.14
15	1.00	1.01	1.00	1.01	1.00	1.01	1.00	1.01	1.02	1.02
16	1.03	1.31	1.01	1.14	1.06	1.30	1.05	1.29	1.01	1.00
mean	1.05	1.21	1.02	1.10	1.02	1.17	1.01	1.15	1.15	1.07

4.5. Results in Landweber method

Tables 34–36 contain the averages of error ratios for the Landweber method with $\mu = 1$, using the following stopping rules. We implemented Landweber method by operator iterations with $m = 2$. The stopping index in rules was computed exactly, by gradually refining the mesh $n = 2^k$, taking the closest previously found approximation as a new initial approximation on each refinement step.

Ln1) Discrepancy principle: n_{D} is the first index $n \geq 1$ for which $d_{\mathrm{D}}(n) \leq \delta$.

Ln2) Rule De: find n_{D} from Ln1 and take $n_{\mathrm{De}} = \mathrm{round}(2.3n_{\mathrm{D}})$.

Ln3) Rule MEe: find n_{ME} as the first $n \geq 1$ for which $d_{\mathrm{ME}}(n) \leq \delta$ and take $n_{\mathrm{MEe}} = \mathrm{round}(2.3n_{\mathrm{ME}})$.

Ln4) Hanke-Raus rule: $n_{\mathrm{HR}} = \mathrm{argmin}\{\sqrt{n}\,\|r_n\|, n \geq 1\}$.

Table 34. Means of error ratios in Landweber method for $p = 0$.

Probl.	D	D,2	De	De,2	MEe	MEe,2	HR	HRmC	QN	QNmC	BRS	BRSC
1	1.46	2.47	1.40	2.36	1.40	2.36	2.71	2.50	1.80	1.75	2.60	2.03
2	1.30	1.91	1.05	1.43	1.05	1.43	976	1.95	976	1.59	976	1.62
3	1.92	6.36	1.76	5.00	1.75	5.01	7.75	2.98	5.01	3.63	6.29	3.61
4	1.42	2.84	1.16	2.13	1.17	2.30	2.70	1.83	1.27	1.61	2.90	1.40
5	1.22	1.89	1.05	1.53	1.05	1.54	1.67	7.39	1e+4	2.37	1.84	2.17
6	1.34	1.96	1.24	1.72	1.24	1.73	2.00	1.39	1.24	1.30	2.11	1.27
7	1.33	2.79	1.08	1.96	1.08	1.97	2e+5	1.44	2e+5	1.49	2e+5	1.39
8	1.40	2.36	1.29	2.03	1.29	2.06	2.58	2.11	1.60	1.56	2.51	1.55
9	1.02	1.05	1.02	1.05	1.02	1.05	1.07	1.07	1.05	1.04	1.07	1.05
10	1.19	1.37	1.18	1.35	1.18	1.35	1.55	1.41	1.48	1.41	1.49	1.41
mean	1.36	2.50	1.22	2.06	1.23	2.08	2e+4	2.41	2e+4	1.77	2e+4	1.75

Table 35. Means of error ratios in Landweber method for $p = 0$ (correlated noise).

Probl.	D	D,2	De	De,2	MEe	MEe,2	HR	HRmC	QN	QNmC	BRS	BRSC
1	1.46	2.39	1.37	2.27	1.38	2.28	2.63	3.39	1.71	3.06	2.53	1.90
2	1.64	2.47	1.19	1.67	1.20	1.67	911	2.44	911	2.69	911	2.16
3	2.10	7.88	1.90	5.98	1.89	5.99	10.0	3.13	5.91	3.44	7.84	3.67
4	1.47	3.23	1.20	2.45	1.21	2.60	3.01	1.88	1.32	1.33	3.34	1.36
5	1.34	2.20	1.16	1.79	1.16	1.79	1.86	7.97	1e+4	2.87	2.03	2.29
6	1.31	2.01	1.20	1.78	1.20	1.78	2.04	1.57	1.24	1.26	2.14	1.25
7	1.28	3.04	1.10	2.10	1.11	2.11	3e+5	2.48	3e+5	2.49	3e+5	1.82
8	1.36	2.41	1.24	2.07	1.24	2.09	2.64	2.32	1.67	1.47	2.51	1.48
9	1.02	1.05	1.02	1.05	1.02	1.05	1.07	1.09	1.05	1.04	1.07	1.05
10	1.20	1.38	1.18	1.36	1.18	1.36	1.56	1.41	1.49	1.41	1.49	1.41
mean	1.42	2.81	1.26	2.25	1.26	2.27	3e+4	2.77	3e+4	2.11	3e+4	1.84

Table 36. Means of error ratios in Landweber method for $p = 2$.

Probl.	D	D,2	De	De,2	MEe	MEe,2	HR	HRmC	QN	QNmC	BRS	BRSC
1	2.78	27.5	2.96	11.8	2.92	11.9	52.5	24.2	6.05	7.84	19.3	6.55
2	1.22	4.03	1.23	2.21	1.23	2.22	2e+4	1.29	2e+4	1.64	2e+4	1.39
3	1.71	28.3	2.91	8.25	2.91	8.28	38.6	8.04	6.67	8.25	18.2	7.94
4	1.50	6.22	1.26	2.29	1.22	2.33	3.38	2.99	1.49	3.02	4.88	1.48
5	1.19	3.56	1.18	1.83	1.15	1.83	2.34	1.26	1e+5	1.49	2.89	1.16
6	1.64	6.82	1.22	3.00	1.20	3.02	4.11	3.55	1.37	2.34	6.59	1.56
7	1.44	5.95	1.25	2.04	1.23	2.06	5e+5	1.59	5e+5	1.81	5e+5	1.32
8	1.60	7.55	1.27	3.01	1.26	3.20	3.89	8.20	1.93	2.47	5.50	2.22
9	1.78	10.4	1.38	3.94	1.36	4.14	6.76	5.59	1.57	1.76	7.92	1.65
10	2.39	31.0	2.74	8.34	2.74	8.36	33.9	11.9	9.62	9.15	19.4	8.99
mean	1.73	13.1	1.74	4.67	1.72	4.73	5e+4	6.86	6e+4	3.98	5e+4	3.43

Ln5) Rule HRmC (modification of the Hanke-Raus rule) chooses n_{HRmC} as the index $n \geq 1$, which minimizes the function $\sqrt{n}\,(\|r_n\| - \|r_{2n+100}\|)$, using the climbing approach with $C = 50$.

Ln6) Rule QN (rule of Neubauer): $n_{\mathrm{QN}} = \operatorname{argmin}\{\|x_n - x_{2n}\|, n \geq 1\}$.

Ln7) Rule QNmC (modification of Neubauer's rule) chooses n_{QNmC} as $n \geq 1$, which minimizes $\|x_n - x_{2n+100}\|$, using the climbing approach with $C = 20$.

L8) Rule BRS (rule of Brezinski-Rodriguez-Seatzu): n_{BRS} is the global minimizer of $\varphi_{\mathrm{BRS}}(n)$.

L9) Rule BRSC (modification of rule BRS) chooses n_{BRSC} as the index $n \geq 1$, which minimizes the function $\|r_n\|(\|r_n\| - \|r_{2n+100}\|)/\|A^* r_n\|$, using the climbing approach with $C = 15$.

Since the stopping index of the monotone error rule differs from the stopping index of the discrepancy principle by at most 1, the columns De and MEe in Table 34 almost coincide.

In case of smooth solution with $p = 2$ the average in column D,2 in Table 36 is 7.5 times larger than the average in column D but the ratio of averages in columns De,2 and De is only 2.7.

For the rule DM we found by varying parameters c_1 and c_2 that the best values were $c_1 = 0.35$ and $c_2 = 0.49$ but the results were worse than for rule QNmC that does not use δ.

4.6. Results in TSVD

Tables 37–39 contain the averages of error ratios in the method of truncated singular value decomposition, using the following truncation rules. If $m + 1/2 \in \mathbb{N}$, then the expression $\|r_m\|$ means $\sqrt{\|r_{m-0.5}\|\,\|r_{m+0.5}\|}$. Some rules contain an *a priori* upper bound of the truncation level: if the dimensions of the discrete problem are $N \times N$, then make computations only for $n \leq M(N,k)$, where $M(N,k) = \max(\max(\operatorname{round}((1.02-k)N - 2, 1), N-k))$.

SP1) Discrepancy principle: n_{D} is the first index n for which $d_{\mathrm{D}}(n) \leq \delta$.

SP2) Rule Q: $n_{\mathrm{Q}} = \operatorname{argmin}\{\|x_n - x_{n+1}\|\} = \operatorname{argmin}\{\sigma_{n+1}^{-1}|(y, v_{n+1})|\}$.

SP3) Rule Qm chooses n_{Qm} as the global minimizer of the function $\max\Big(\big(\sum_{n-2\leq i\leq n+2} \|x_n - x_i\|^2\big)^{1/2}, \|x_n - x_{n+1}\|\Big)$.

SP4) Rule HR': in a discrete $N \times N$ problem choose $n_{\mathrm{HR'}}$ as $n \leq M(N,10)$ for which $\varphi_{\mathrm{HR}}(n) = \|r_n\|/\sigma_{n+1}$ is minimal.

SP5) Rule HRm: $n_{\mathrm{HRm}} = \operatorname{argmin}\{(\|r_n\| - \|r_{1.5n+8}\|)/\sigma_{n+1}\}$.

SP6) Rule HRL': in a discrete $N \times N$ problem choose $n_{\mathrm{HRL'}}$ as $n \leq M(N,10)$ for which $\|r_n\|\sqrt{\|x_n\|/\sigma_{n+1}}$ is minimal (geometric mean of functions minimized in rules HR and L).

SP7) Rule HRLm': in a discrete $N \times N$ problem choose $n_{\mathrm{HRLm'}}$ as $n \leq M(N,5)$ for which $(\|r_n\| - \|r_{1.5n+10}\|)\sqrt{\|x_n\|/\sigma_{n+1}}$ is minimal.

Table 37. Means of error ratios in TSVD for $p = 0$.

Probl.	D	D,2	Q	Qm	HR'	HRm	HRL'	HRLm'	BRS'	BRSm
1	1.35	1.72	1.43	1.95	2.44	1.63	1.91	1.29	2.75	1.78
2	1.28	1.83	3.50	1.17	2.41	4.03	2.43	3.81	3.00	4.27
3	1.31	2.82	2.51	1.97	4.64	2.16	3.66	2.16	8.37	2.99
4	1.33	2.63	2.03	1.23	2.75	1.57	2.42	1.48	3.65	1.93
5	1.24	1.95	2.82	1.17	1.81	1.36	1.57	1.29	2.07	1.48
6	1.21	1.86	1.40	1.24	1.96	1.32	1.80	1.24	2.10	1.38
7	1.27	3.01	161	1.41	3.99	1.56	1.88	1.43	3.42	1.49
8	1.19	1.79	1.98	1.82	2.60	1.69	1.84	1.29	2.86	1.79
9	1.02	1.04	1.04	1.07	1.09	1.03	1.05	1.02	1.09	1.03
10	1.06	1.20	1.32	1.40	1.38	1.33	1.33	1.23	1.47	1.32
mean	1.23	1.98	17.9	1.44	2.51	1.77	1.99	1.62	3.08	1.95
11	1.31	2.32	9.78	1.29	2.60	1.65	2.53	1.45	3.51	1.88
12	1.22	1.74	3.16	1.97	2.96	1.75	1.82	1.43	2.98	1.69
13	1.34	2.02	2.07	1.81	3.07	1.69	2.18	1.39	3.87	1.99
14	1.35	2.11	37.6	1.42	1.97	1.62	2.03	1.65	2.23	1.78
15	1.25	1.01	1.02	1.03	1.03	1.02	1.02	1.02	1.03	1.02
16	1.55	2.26	55.3	7.15	2.81	3.03	1.82	2.33	3.22	11.0
mean	1.34	1.91	18.2	2.45	2.41	1.79	1.90	1.55	2.81	3.22

Table 38. Means of error ratios in TSVD for $p = 0$ (correlated noise).

Probl.	D	D,2	Q	Qm	HR'	HRm	HRL'	HRLm'	BRS'	BRSm
1	1.19	1.80	1.49	2.06	2.57	1.66	2.02	1.41	2.83	1.86
2	1.66	2.41	37.7	1.29	3.22	6.10	4.73	5.69	4.22	6.48
3	1.32	2.98	2.00	1.61	5.46	1.91	3.99	1.61	8.75	2.98
4	1.36	2.95	3.24	1.21	3.01	1.65	2.68	1.46	3.91	1.93
5	1.34	2.17	4.56	1.31	1.93	1.51	2.77	1.44	2.20	1.72
6	1.20	1.90	1.32	1.24	2.05	1.34	1.80	1.22	2.20	1.51
7	1.31	3.40	215	2.04	4.59	1.34	8.41	1.45	3.91	1.92
8	1.17	1.89	1.74	1.87	2.72	1.62	1.92	1.32	3.01	1.75
9	1.01	1.05	1.04	1.07	1.09	1.03	1.05	1.02	1.09	1.03
10	1.04	1.20	1.32	1.41	1.38	1.32	1.33	1.20	1.47	1.30
mean	1.26	2.18	27.0	1.51	2.80	1.95	3.07	1.78	3.36	2.25
11	1.27	2.58	12.2	1.25	2.90	1.62	2.77	1.47	3.81	1.88
12	1.22	1.85	1.87	2.04	3.13	1.73	1.91	1.42	3.14	1.74
13	1.31	2.38	2.22	1.79	3.61	1.76	2.55	1.44	4.55	2.02
14	1.29	2.32	96.3	4.27	2.32	1.71	8.03	1.82	2.54	1.89
15	1.10	1.01	1.02	1.04	1.03	1.02	1.02	1.01	1.03	1.02
16	1.71	2.64	51.9	11.2	3.05	3.28	2.20	2.89	3.68	14.3
mean	1.32	2.13	27.6	3.60	2.67	1.85	3.08	1.68	3.13	3.80

Table 39. Means of error ratios in TSVD for $p = 2$.

Probl.	D	D,2	Q	Qm	HR'	HRm	HRL'	HRLm'	BRS'	BRSm
1	1.37	5.48	3.39	2.05	28.8	5.38	5.48	2.60	50.8	10.8
2	1.54	4.38	21.2	2.23	3.49	1.56	3.48	1.66	4.38	2.71
3	2.27	5.02	2.41	10.3	36.8	5.02	5.02	4.34	181	5.02
4	1.55	8.26	1.62	1.73	7.33	1.81	1.94	1.76	8.26	2.12
5	1.31	4.89	23.2	1.58	3.18	1.51	2.27	1.38	5.71	1.96
6	1.51	4.39	2.45	1.64	5.83	2.13	2.95	1.88	12.5	2.89
7	1.25	2.49	1e+4	2.38	2.49	2.08	2.49	1.86	19.8	2.28
8	1.25	3.51	4.58	3.13	3.51	2.31	3.21	1.71	22.8	3.00
9	1.27	5.79	1.64	2.10	5.79	2.02	5.79	1.84	22.6	5.54
10	1.39	2.92	2.92	1.95	22.3	2.92	22.3	2.92	1e+3	19.7
mean	1.47	4.71	1e+3	2.91	12.0	2.67	5.50	2.20	163	5.60
11	1.30	4.57	871	1.51	4.77	2.15	4.14	1.69	13.5	3.08
12	1.36	4.43	2.56	2.22	8.17	2.38	4.43	2.02	25.3	4.25
13	1.44	4.11	2.27	3.92	6.44	1.95	4.29	1.64	22.2	3.66
14	1.59	4.15	1e+3	2.88	4.30	1.86	2.55	1.97	7.59	2.40
15	2.02	1.02	1.02	4.68	1.02	1.02	1.02	1.02	1.02	1.02
16	1.49	2.74	1e+3	2.52	2.70	2.41	1.67	1.82	8.09	6.96
mean	1.53	3.50	520	2.95	4.57	1.96	3.02	1.69	12.9	3.56

SP8) Rule BRS': in a discrete $N \times N$ problem choose $n_{\mathrm{BRS'}}$ as $n \le M(N, 20)$ for which $\|r_n\|^2 / \|A^* r_n\|$ is minimal.

SP9) Rule BRSm: $n_{\mathrm{BRSm}} = \operatorname{argmin}\{(\|r_n\| - \|r_{1.5n+15}\|)^2 / \|A^* x_n\|\}$.

The upper bound of the truncation level was needed because some functions started to decrease very quickly, attaining their minimums at $n = N$, where the error of approximate solution was large. The length of the interval, where this behavior occurred, depended weakly on the discretization level N.

4.7. Results in conjugate gradient type methods

The following tables present the averages of error ratios $\|x_{n_{\mathrm{R}}} - x_*\| / \|x_{n_*} - x_*\|$ for rules in methods CGLS and CGME. The denominator $\|x_{n_*} - x_*\|$ refers to the minimal error in the corresponding method (CGLS or CGME).

4.7.1. Rules for CGLS.

For CGLS the following rules were compared. Rules CL1–CL5 use noise level information, rules CL6–CL9 do not.

CL1) Discrepancy principle: n_{D} is the first index n for which $d_{\mathrm{D}}(n) \le \delta$.

CL2) Rule De. Using n_{D} from CL1, find $n_{\mathrm{De}} = \mathrm{round}(1.02 n_{\mathrm{D}}^{1.05})$.

CL3) Rule ME: n_{ME} is the first index n for which $d_{\mathrm{ME}}(n) \le \delta$.

CL4) Rule MEe: using n_{ME} from CL3, find $n_{\mathrm{MEe}} = \mathrm{round}(0.99 n_{\mathrm{ME}}^{1.13})$.

CL5) Rule Me: using n_{D} and n_{ME} from CL1 and CL3, respectively, find $n_{\mathrm{Me}} = \mathrm{round}(\max(n_{\mathrm{D}}^{1.043}, n_{\mathrm{ME}}^{1.07}))$.

Table 40. Means of error ratios in CGLS for $p = 0$.

Probl.	D	D,2	De	De,2	ME	ME,2	MEe	MEe,2	Me	Me,2	HR	QNmC	HRmC	HRmWC
1	1.40	1.86	1.42	1.86	1.66	2.27	1.66	2.27	1.40	1.86	2.52	2.28	1.87	1.52
2	1.20	1.74	1.06	1.43	1.38	2.09	1.08	1.47	1.07	1.46	912	3.05	1.62	1.62
3	1.29	2.82	1.43	2.75	2.15	4.88	1.96	4.81	1.29	2.75	5.29	27.8	2.34	2.34
4	1.22	2.24	1.20	2.13	1.71	3.61	1.27	3.03	1.19	2.09	2.22	1.59	1.45	1.45
5	1.13	1.81	1.05	1.62	1.31	2.04	1.07	1.65	1.06	1.63	1.60	1.33	1.45	1.12
6	1.23	1.79	1.21	1.74	1.54	2.39	1.34	2.20	1.21	1.76	2.12	1.46	1.42	1.42
7	1.17	2.39	1.16	2.36	1.74	3.49	1.48	3.15	1.17	2.29	2.34	27.4	1.41	1.41
8	1.19	1.87	1.19	1.72	1.49	2.62	1.38	2.34	1.17	1.79	2.12	1.95	1.74	1.49
9	1.02	1.04	1.03	1.04	1.03	1.06	1.03	1.06	1.02	1.04	1.07	1.06	1.04	1.02
10	1.06	1.20	1.06	1.20	1.16	1.25	1.13	1.24	1.06	1.20	1.48	1.72	1.38	1.38
mean	1.19	1.88	1.18	1.78	1.52	2.57	1.34	2.32	1.16	1.79	93.3	6.96	1.57	1.48

Table 41. Means of error ratios in CGLS for $p = 0$ (correlated noise).

Probl.	D	D,2	De	De,2	ME	ME,2	MEe	MEe,2	Me	Me,2	HR	QNmC	HRmC	HRmWC
1	1.23	1.89	1.23	1.89	1.71	2.35	1.71	2.35	1.23	1.89	2.58	2.44	1.97	1.60
2	1.49	2.25	1.24	1.71	1.75	2.73	1.24	1.66	1.23	1.75	875	3.31	1.49	3.65
3	1.35	3.12	1.40	2.95	2.20	5.19	1.78	5.01	1.33	2.95	5.97	42.6	2.07	2.07
4	1.26	2.60	1.19	2.46	1.94	4.00	1.46	3.33	1.23	2.37	2.50	1.64	1.43	1.43
5	1.27	2.07	1.22	1.84	1.44	2.33	1.18	1.87	1.21	1.85	1.79	1.41	1.94	1.30
6	1.18	1.82	1.14	1.76	1.57	2.39	1.35	2.18	1.18	1.79	2.12	1.36	1.47	1.36
7	1.16	2.63	1.15	2.56	2.03	3.80	1.58	3.40	1.16	2.47	2.72	32.0	1.38	1.39
8	1.18	1.96	1.18	1.80	1.55	2.75	1.34	2.46	1.18	1.86	2.18	1.99	1.61	1.30
9	1.01	1.05	1.01	1.05	1.03	1.06	1.02	1.06	1.01	1.05	1.07	1.06	1.04	1.02
10	1.05	1.21	1.05	1.21	1.15	1.25	1.13	1.24	1.05	1.21	1.47	1.74	1.35	1.35
mean	1.22	2.06	1.18	1.92	1.64	2.78	1.38	2.46	1.18	1.92	89.8	8.96	1.57	1.65

Table 42. Means of error ratios in CGLS for $p = 2$.

Probl.	D	D,2	De	De,2	ME	ME,2	MEe	MEe,2	Me	Me,2	HR	QNmC	HRmC	HRmWC
1	1.29	4.58	1.29	4.58	1.95	8.31	1.95	8.31	1.29	4.58	20.5	118	2.33	2.33
2	1.18	2.92	1.26	2.92	1.88	5.94	1.75	5.67	1.26	2.92	2e+4	18.4	2.54	2.54
3	1.46	4.56	1.46	4.56	1.64	13.9	1.64	13.9	1.46	4.56	52.9	573	6.83	6.83
4	1.16	4.76	1.36	3.47	1.96	6.32	1.69	4.63	1.32	3.47	4.48	3.22	1.55	1.55
5	1.16	3.09	1.12	2.63	1.54	4.69	1.23	3.74	1.13	2.54	2.69	2.14	1.24	1.24
6	1.42	4.00	1.55	2.88	1.99	9.46	1.59	6.76	1.24	3.01	4.48	4.58	1.85	1.85
7	1.22	4.00	1.20	4.00	1.97	5.58	1.38	5.18	1.23	4.00	2.24	5.01	1.55	1.55
8	1.13	2.70	1.13	2.70	1.70	5.51	3.32	6.31	1.13	2.70	10.3	59.1	3.68	3.68
9	1.35	6.68	1.45	4.95	2.36	10.2	1.64	8.47	1.31	4.95	5.94	7.21	1.50	1.50
10	1.42	2.88	1.42	2.88	2.18	9.87	2.18	9.87	1.42	2.88	55.3	2e+3	24.9	24.9
mean	1.28	4.02	1.32	3.56	1.92	7.98	1.84	7.29	1.28	3.56	2e+3	312	4.79	4.79

CL6) Hanke-Raus rule: $n_{\mathrm{HR}} = \operatorname{argmin}\{\varphi_{\mathrm{HR}}(n)\}$.

CL7) Rule QNmC (analog of Neubauer's rule): n_{QNmC} is the minimizer of the function $\varphi_{\mathrm{QNm}}(n) = \|x_n - \overline{x}_n\|$, using the climbing strategy with $C = 20$, where $\overline{x}_n$ is the solution of $A\overline{x}_n = Ax_n$ found after $n + 1$ steps in CGLS method.

CL8) Rule HRmC: n_{HRmC} is the minimizer of the function $\varphi_{\mathrm{HRm}}(n) = \sqrt{\varrho_{n+1}}\,(\|r_n\| - \|r_{2n+10}\|)$, using the climbing strategy with $C = 10$.

CL9) Rule HRmWC: n_{HRmWC} is the minimizer of the function, which coincides with $\varphi_{\mathrm{HRm}}(n)$ at indices $n \le 3$ and with W for $n \ge 500$, with smooth transition from φ_{HRm} to W on steps from 3 to 500. Computations are made on the interval $[0, N]$, where N is determined using the climbing strategy with $C = 20$.

From Tables 40–42 we see that the rule Me with the estimated parameter gives slightly better results than rules D and ME in case $p = 0$ both for exact and 2 times overestimated noise level and both for uncorrelated and correlated noise. Also note that our modifications to the rule HR were able to improve the results in most problems in case $p = 0$. Adding the function W was justified in many problems, if $p = 0$ (rules HRmC and HRmWC) but if p was 2, then the results remained unchanged.

4.7.2. Rules for CGME.

Tables 43–45 compare the following rules in CGME.

CM1) Rule DH: n_{DH} is the first index for which $d_{\mathrm{DH}}(n) \le 1.2\delta$.

CM2) Rule ME: n_{ME} is the first index n for which $d_{\mathrm{ME}}(n) \le \delta$.

CM3) Rule MEe: using n_{ME} from CM2, find $n_{\mathrm{MEe}} = \operatorname{round}(0.85 n_{\mathrm{ME}}^{1.06})$.

CM4) Hanke-Raus rule: $n_{\mathrm{HR}} = \operatorname{argmin}\{\varphi_{\mathrm{HR}}(n)\}$.

CM5) Rule HRC: n_{HRC} is the minimizer of the function $\varphi_{\mathrm{HR}}(n)$, using the climbing strategy with $C = 100$.

CM6) Rule DHP: n_{DHP} is the first index n for which the function $d_{\mathrm{DH}}(n)$ decreases in next 10 steps by no more than $C = 1.5$ times (plateau strategy).

CM7) Rule RMC: n_{RMC} is the minimizer of $d_{\mathrm{D}}(n)$ on $[0, N]$, where N minimizes $\varrho_{n-2}^{1/2}\, d_{\mathrm{DH}}(n-3)$ using the climbing strategy with $C = 3$.

The error ratios for CGME are better than for CGLS but as Tables 4–7 show, the method CGLS itself generally has smaller minimal errors. Noteworthy is remarkably good performance of the discrepancy principle (DH in Tables 43–45, also D in Tables 40–42) that is hard to beat by solely refining D and ME. In case $p = 0$ the heuristic delta-free rules RMC and DHP gave similar results than the discrepancy principle in most problems but surprisingly in case $p = 2$ they gave best possible results in 90% of runs (Table 45). The rule DHP was not so good in case of correlated noise (Table 44).

Note that some other stopping rules for CGLS and CGME methods are proposed and numerically tested in [27].

Table 43. Means of error ratios in CGME for $p = 0$.

Probl.	DH	DH,2	ME	ME,2	MEe	MEe,2	HR	HRC	DHP	RMC
1	1.02	1.17	1.10	1.23	1.10	1.23	1.74	1.74	1.01	1.01
2	1.14	1.55	1.36	1.92	1.23	1.70	686	1.27	4.03	1.04
3	1.07	1.24	1.00	2.55	1.00	2.55	4.19	4.19	1.00	1.00
4	1.05	1.51	1.07	1.68	1.08	1.68	1.61	1.61	1.00	1.04
5	1.05	1.24	1.14	1.39	1.11	1.36	1.38	1.38	1.21	1.09
6	1.11	1.16	1.08	1.40	1.07	1.40	1.48	1.48	1.00	1.10
7	1.05	1.41	1.14	1.52	1.14	1.50	1.64	1.66	1.00	1.16
8	1.02	1.20	1.09	1.52	1.09	1.52	1.51	1.50	1.00	1.13
9	1.00	1.02	1.01	1.02	1.01	1.02	1.04	1.04	1.00	1.02
10	1.03	1.03	1.00	1.07	1.00	1.07	1.27	1.32	1.00	1.00
mean	1.06	1.25	1.10	1.53	1.08	1.50	70.2	1.72	1.33	1.06

Table 44. Means of error ratios in CGME for $p = 0$ (correlated noise).

Probl.	DH	DH,2	ME	ME,2	MEe	MEe,2	HR	HRC	DHP	RMC
1	1.01	1.16	1.09	1.24	1.09	1.24	1.72	1.70	1.01	1.00
2	1.40	2.00	1.79	2.60	1.54	2.25	613	1.08	9.72	1.04
3	1.07	1.26	1.04	2.52	1.04	2.52	3.95	3.95	1.00	1.00
4	1.06	1.50	1.10	1.69	1.10	1.69	1.58	1.57	1.42	1.04
5	1.16	1.34	1.24	1.51	1.21	1.48	1.39	1.13	1.59	1.12
6	1.08	1.16	1.09	1.40	1.09	1.40	1.46	1.42	1.66	1.08
7	1.10	1.43	1.17	1.58	1.19	1.56	1.65	1.69	1.71	1.17
8	1.02	1.22	1.11	1.54	1.11	1.54	1.50	1.47	1.51	1.13
9	1.00	1.02	1.01	1.02	1.01	1.02	1.04	1.04	1.06	1.02
10	1.02	1.02	1.00	1.07	1.00	1.07	1.25	1.29	1.00	1.00
mean	1.09	1.31	1.16	1.62	1.14	1.58	62.8	1.64	2.17	1.06

Table 45. Means of error ratios in CGME for $p = 2$.

Probl.	DH	DH,2	ME	ME,2	MEe	MEe,2	HR	HRC	DHP	RMC
1	1.09	1.43	1.00	1.43	1.00	1.43	31.3	31.3	1.00	1.00
2	1.17	1.56	1.11	2.09	1.11	2.09	5e+3	74.5	1.02	1.00
3	1.00	1.43	1.00	1.43	1.00	1.43	170	170	1.00	1.00
4	1.41	1.34	1.00	1.37	1.00	1.37	1.37	1.37	1.00	1.00
5	1.05	1.78	1.31	2.37	1.31	2.37	1.59	1.59	1.01	1.01
6	1.01	1.75	1.22	2.99	1.22	2.99	2.85	2.85	1.00	1.01
7	1.00	1.77	1.43	3.20	1.43	3.20	1.75	1.75	1.00	1.00
8	1.03	1.71	1.00	1.71	1.00	1.71	12.6	12.6	1.00	1.00
9	1.10	1.43	1.25	2.05	1.25	2.05	2.27	2.27	1.00	1.00
10	1.00	1.40	1.00	1.43	1.00	1.43	1e+3	1e+3	1.00	1.00
mean	1.09	1.56	1.13	2.01	1.13	2.01	699	178	1.00	1.00

4.7.3. Rule DM.

If the noise level is known approximately, then for CGLS, we propose the two-step rule DM with constants $c_1 = 0.25$, $c_2 = 0.4$.

Tables 46, 47 illustrate behavior of this rule at various degrees of noise level over- or underestimation. Comparing with Table 40, we see that at $d = 1$, $p = 0$ this rule gives an average that is worse than the average of best rules that use noise level but better than the best average of a δ-free rule. The right hand sides of Tables 46, 47 show that if $p = 2$, then the rule DM looses its forgiveness of coarse estimation of noise level more quickly than in case $p = 0$.

Comparing Tables 46 and 47 with each other, we see that in case of correlated noise the usable domain of DM is essentially narrower. If $p = 2$, then middle columns (d close to 1) are better in case of uncorrelated noise but at 100 times underestimated noise level ($d = 0.01$) the results are better in case of correlated noise.

Table 46. Rule DM in CGLS for $p = 0$ (left), $p = 2$ (right), where the supposed noise level is d times the actual noise level.

	Case $p = 0$, values of d							Case $p = 2$, values of d						
Probl.	0.01	0.1	0.5	1	2	10	100	0.01	0.1	0.5	1	2	10	100
1	5.94	3.46	2.95	1.43	1.51	2.25	3.20	15.6	15.2	8.82	1.28	1.88	8.31	41.2
2	29.5	4.70	1.58	1.30	1.14	1.60	4.50	395	1.59	1.59	1.56	1.61	5.94	18.8
3	3.27	3.27	2.07	2.18	2.70	4.39	8.14	1.77	1.77	1.48	1.44	2.15	13.9	72.2
4	1.49	1.49	1.36	1.33	1.60	3.66	8.15	2.78	2.78	2.33	1.21	1.19	6.32	28.5
5	14.3	14.3	2.26	1.29	1.12	1.83	3.98	35.1	35.1	1.39	1.26	1.24	3.79	15.9
6	1.63	1.63	1.23	1.33	1.45	2.20	3.28	2.44	2.44	1.76	1.55	1.97	6.17	25.1
7	1.29	1.29	1.29	1.37	1.42	3.60	8.99	1.46	1.46	1.44	1.46	1.67	4.60	25.4
8	1.71	1.70	1.31	1.36	1.62	2.41	4.11	5.56	5.56	2.56	1.27	1.48	5.51	27.5
9	1.51	1.35	1.02	1.02	1.03	1.06	1.12	5.35	2.53	1.51	1.40	1.74	7.68	34.4
10	26.4	1.55	1.21	1.22	1.23	1.36	1.46	3e+3	109	1.57	1.57	2.87	9.87	45.4
mean	8.70	3.47	1.63	1.38	1.48	2.44	4.69	316	17.8	2.45	1.40	1.78	7.21	33.4

Table 47. Rule DM in CGLS for $p = 0$ (left), $p = 2$ (right), where the supposed noise level is d times the actual noise level (correlated noise).

	values of d for $p = 0$							values of d for $p = 2$						
Probl.	0.01	0.1	0.5	1	2	10	100	0.01	0.1	0.5	1	2	10	100
1	5.29	5.24	1.61	1.54	1.62	2.32	3.27	16.1	13.0	6.40	1.75	5.42	9.42	65.9
2	19.0	3.16	1.59	1.16	1.19	1.85	6.20	154	2.29	1.81	1.78	2.11	6.72	28.2
3	4.98	4.41	1.88	2.33	2.79	4.73	8.97	10.0	10.2	3.88	2.67	2.47	11.6	50.1
4	1.94	1.92	1.59	1.70	1.90	4.08	8.86	1.76	1.60	2.24	2.09	1.98	10.2	52.8
5	7.82	7.43	2.16	1.57	1.29	2.07	4.50	13.2	1.42	1.51	1.74	1.97	6.43	27.4
6	1.64	1.47	1.34	1.36	1.43	2.21	3.25	1.74	1.76	2.05	2.49	2.94	12.4	44.1
7	1.48	1.26	1.35	1.52	1.67	4.07	9.63	1.36	1.38	2.06	2.67	3.49	8.79	49.9
8	1.74	1.72	1.54	1.46	1.64	2.53	4.30	6.30	5.55	4.85	3.65	3.62	12.9	63.4
9	310	1.03	1.03	1.02	1.03	1.06	1.11	1.31	1.41	1.47	2.16	2.78	8.05	50.0
10	2.32	1.87	1.41	1.23	1.23	1.33	1.46	413	134	120	9.43	2.74	9.38	40.9
mean	35.7	2.95	1.55	1.49	1.58	2.63	5.16	61.9	17.3	14.6	3.04	2.95	9.59	47.3

4.8. Results in extrapolated Tikhonov method

Since extrapolation forms the approximate solution on base of several single Tikhonov approximations, in numerical experiments of this section we searched the regularization parameter from the sequence $\Omega = (\alpha_n)$, where $\alpha_n = q^n$ ($q = 0.9$; $n = 0, 1, \ldots$). The task of finding the solution of an equation $d(\alpha) = C\delta$ was thus replaced by the task of finding the first α in this sequence for which the inequality $d(\alpha) \le C\delta$ holds. The extrapolated approximation with m terms, corresponding to regularization parameter α, was constructed using Tikhonov approximations $x_{q^i\alpha}$ with $i = -\left\lfloor \frac{m}{2} \right\rfloor$, $-\left\lfloor \frac{m}{2} \right\rfloor + 1, \ldots, \left\lceil \frac{m}{2} \right\rceil - 1$. Then α lies exactly in the middle of the sequence $q^i\alpha$ of parameters, if m is odd.

To maintain a common base of comparison, all error ratios were computed by dividing the error of computed approximation by the minimal error of single Tikhonov approximation. In this sense the results are comparable with the results of Section 4.3.

4.8.1. Rules using noise level.

Tables 48–55 show the results for the following rules using noise level information.

Te1) Rule De. Let α_{D} be the first α_n with $d_{\mathrm{D}}(\alpha_n) \le \delta$ in ordinary Tikhonov method and $\alpha_{m\mathrm{D}}$ be the first α_n with $d_{\mathrm{D}}(\alpha_n) \le \delta$ in m-extrapolated Tikhonov method. Take $\alpha_{\mathrm{De}} = \alpha_{m\mathrm{De}}$ to be the nearest $\alpha \in \Omega$ to $\alpha_{m\mathrm{D}}^{c_{m,1}}\alpha_{\mathrm{D}}^{c_{m,2}}$, where $(c_{2,1}, c_{2,2}) = (1.22, -0.12)$, $(c_{3,1}, c_{3,2}) = (1.16, -0.04)$. We also recommend $(c_{4,1}, c_{4,2}) = (1.11, -0.01)$, $(c_{5,1}, c_{5,2}) = (1.1, 0)$. The exponent less than 1 of α_{D} was good only for $p \ge 1$.

Te2) Rule MEe. Let α_{ME} be the first α_n with $d_{\mathrm{eME}}(\alpha_n) \le \delta$ in m-extrapolated Tikhonov method (see (3.6)). Take $\alpha_{\mathrm{MEe}} = \alpha_{m\mathrm{ME}}$ to be the nearest $\alpha \in \Omega$ to $\min(0.5\alpha_{\mathrm{ME}}, 0.6\alpha_{\mathrm{ME}}^{1.08})$.

Te3) Rule MEs. Let α_{ME} be the first α_n for which $d_{\mathrm{ME}}(n) \le \delta$ on sequence (x_n) (see (3.5)), where x_n is an m-extrapolated approximation corresponding to parameter α_n. Take $\alpha_{\mathrm{MEs}} = \alpha_{m\mathrm{MEs}}$ to be the nearest $\alpha \in \Omega$ to $\min(0.5\alpha_{\mathrm{ME}}, 0.6\alpha_{\mathrm{ME}}^{1.08})$.

Te4) Rule R2e. Let α_{R2} be the first α with $d_{\mathrm{R2}}(\alpha) \le C\delta$ in m-iterated Tikhonov method, where $C = 0.2$ in case $m = 2$ and $C = 0.13$ in case $m = 3$. Take $\alpha_{\mathrm{R2e}} = \alpha_{m\mathrm{R2e}}$ to be the nearest $\alpha \in \Omega$ to $0.5\alpha_{\mathrm{R2}}$.

Te5) Rule Me. Using α_{MEe}, α_{R2e} from Te2, Te4, respectively, take $\alpha_{\mathrm{Me}} = \min(\alpha_{\mathrm{MEe}}, \alpha_{\mathrm{R2e}})$.

Te6) Rule maxD. Take α_{maxD} to be the first α_n for which $d_{\mathrm{D}}(n) \le \delta$ on sequence (x_n), where x_n is an n-extrapolated approximation formed using single Tikhonov approximations $x_{\alpha_1}, \ldots, x_{\alpha_n}$.

Te7) Rule maxDe. Using α_{maxD} from Te6, take α_{maxDe} to be the nearest $\alpha \in \Omega$ to $\alpha_{\mathrm{maxD}}^{1.1}$.

Table 48. Means of error ratios for 2-extrapolated Tikhonov approximation.

p	De	De,2	MEe	MEe,2	MEs	MEs,2	R2e	R2e,2	Me	Me,2
0	1.16	2.15	1.15	1.87	1.15	1.88	1.43	1.52	1.24	1.51
0.25	1.59	3.65	1.58	3.12	1.58	3.15	2.78	2.92	1.79	2.40
0.5	1.55	4.44	1.55	3.64	1.55	3.67	2.11	2.42	1.74	2.42
0.75	1.25	4.25	1.27	3.12	1.26	3.16	1.56	1.83	1.41	1.83
1	1.00	3.92	1.02	2.54	1.01	2.59	1.30	1.41	1.15	1.41
1.5	0.74	3.31	0.79	1.91	0.78	1.95	0.85	0.88	0.83	0.88
2	0.59	2.70	0.65	1.37	0.64	1.40	0.69	0.67	0.69	0.67
4	0.45	1.87	0.50	0.85	0.50	0.86	0.54	0.48	0.55	0.48
8	0.45	1.80	0.49	0.83	0.49	0.83	0.53	0.47	0.54	0.47
mean	0.98	3.12	1.00	2.14	0.99	2.16	1.31	1.40	1.10	1.34

Table 49. Means of error ratios for 2-extrapolated Tikhonov approximation (correlated noise).

p	De	De,2	MEe	MEe,2	MEs	MEs,2	R2e	R2e,2	Me	Me,2
0	1.18	2.31	1.17	2.00	1.17	2.02	1.54	1.61	1.28	1.61
0.25	1.61	4.76	1.61	4.04	1.61	4.07	3.32	3.63	1.77	3.06
0.5	1.54	7.78	1.53	6.46	1.52	6.50	4.21	4.64	1.75	4.64
0.75	1.31	7.89	1.32	5.91	1.32	5.96	2.73	3.65	1.49	3.65
1	0.99	6.39	1.01	4.31	1.00	4.39	2.14	2.61	1.13	2.61
1.5	0.74	6.49	0.77	3.64	0.77	3.72	1.35	1.57	0.87	1.57
2	0.59	4.25	0.72	2.17	0.72	2.22	1.00	1.04	0.90	1.04
4	0.42	4.48	0.49	2.07	0.49	2.09	0.69	0.71	0.63	0.71
8	0.44	2.74	0.51	1.35	0.50	1.35	0.61	0.58	0.62	0.58
mean	0.98	5.23	1.02	3.55	1.01	3.59	1.95	2.23	1.16	2.16

Table 50. Means of error ratios for 3-extrapolated Tikhonov approximation.

p	De	De,2	MEe	MEe,2	MEs	MEs,2	R2e	R2e,2	Me	Me,2
0	1.17	2.05	1.16	1.83	1.16	1.82	1.43	1.59	1.16	1.58
0.25	1.64	3.49	1.60	3.07	1.61	3.04	2.80	3.04	1.61	2.50
0.5	1.57	4.22	1.58	3.57	1.58	3.54	2.27	2.83	1.58	2.82
0.75	1.26	3.88	1.28	3.01	1.28	2.96	1.59	2.34	1.28	2.31
1	0.99	3.46	1.01	2.40	1.02	2.35	1.36	1.51	1.02	1.51
1.5	0.72	2.87	0.77	1.73	0.78	1.69	0.84	1.02	0.78	1.02
2	0.54	2.25	0.61	1.11	0.62	1.07	0.68	0.67	0.63	0.67
4	0.34	1.40	0.44	0.48	0.44	0.48	0.46	0.40	0.47	0.40
8	0.33	1.28	0.41	0.42	0.42	0.42	0.44	0.39	0.45	0.39
mean	0.95	2.77	0.98	1.96	0.99	1.93	1.32	1.53	1.00	1.47

Table 51. Means of error ratios for 3-extrapolated Tikhonov approximation (correlated noise).

p	De	De,2	MEe	MEe,2	MEs	MEs,2	R2e	R2e,2	Me	Me,2
0	1.18	2.20	1.17	1.96	1.17	1.95	1.53	1.70	1.18	1.69
0.25	1.63	4.48	1.63	3.96	1.63	3.92	3.46	3.77	1.64	3.20
0.5	1.53	7.31	1.56	6.27	1.56	6.22	4.31	5.21	1.58	5.19
0.75	1.32	7.29	1.35	5.78	1.36	5.69	3.17	4.65	1.39	4.58
1	0.99	5.53	1.01	4.02	1.02	3.97	2.41	2.81	1.04	2.81
1.5	0.71	5.74	0.74	3.24	0.75	3.11	1.37	1.75	0.78	1.75
2	0.59	3.54	0.75	1.78	0.75	1.75	0.96	1.03	0.81	1.03
4	0.34	3.48	0.45	1.13	0.46	1.11	0.53	0.57	0.50	0.57
8	0.35	2.07	0.46	0.66	0.46	0.66	0.49	0.46	0.50	0.46
mean	0.96	4.63	1.01	3.20	1.02	3.15	2.03	2.44	1.05	2.36

Table 52. Means of error ratios for extrapolated approximation, $n = \max$ (uncorrelated noise left, correlated noise right).

p	maxD	maxD,2	maxDe	maxDe,2	maxD	maxD,2	maxDe	maxDe,2
0	1.29	2.49	1.17	2.25	1.34	2.73	1.19	2.42
0.25	1.80	4.45	1.63	3.82	1.88	5.97	1.66	5.14
0.5	1.81	5.25	1.60	4.65	1.96	9.09	1.60	8.31
0.75	1.48	5.22	1.28	4.47	1.56	9.52	1.36	8.27
1	1.15	4.84	0.99	4.00	1.16	8.52	1.00	6.80
1.5	0.81	4.03	0.71	3.34	0.81	7.29	0.70	5.97
2	0.58	3.18	0.52	2.67	0.66	4.85	0.60	3.97
4	0.31	1.93	0.30	1.78	0.31	4.70	0.31	4.34
8	0.26	1.55	0.26	1.47	0.28	2.21	0.29	2.02
mean	1.06	3.66	0.94	3.16	1.11	6.10	0.97	5.25

Rules Te1–Te5 are intended for choosing the parameter in m-extrapolated approximation (m fixed), rules Te6, Te7 can be used in extrapolated approximation with maximum number of single Tikhonov approximations with parameters $\alpha_1, \alpha_2, \ldots, \alpha_{\mathrm{maxD}}$ (resp. α_{maxDe}).

The rules MEe and MEs always gave similar results. At exact δ the rules MEe and MEs are somewhat better than ME but in case of 2 times overestimated noise level the rule Me is better.

As Tables 48–52 show, if $x_* \in \mathcal{R}(A^*)$, then in most cases the error of extrapolated approximation with *a posteriori* parameter choice was smaller than the error of the best single Tikhonov approximation. Table 52 shows the advantage of the approximation with maximum number of terms for large p.

In Tables 53, 54 the error ratios of 3-extrapolated approximation with rule R2e and max-extrapolated approximation with rule maxDe are given for every problem (also including the problems from [13]). In most problems the error ratios decreased with increasing p, especially if $p \geq 1$.

Table 55 shows the averages of error ratios for problem 'heat'.

We do not present the results for the discrepancy principle and monotone error rule since the refined rules De and MEe were somewhat better.

Table 53. Means of error ratios in 3-extrapolated Tikhonov approximation with rule MEe by problems.

p	baart	deriv2	foxgood	gravity	heat	ilaplace	phillips	shaw	spikes	wing	gauss	hilbert	lotkin	moler	pascal	prolate
0	1.42	1.02	1.54	1.04	1.00	1.13	0.98	1.26	1.02	1.18	1.10	1.23	1.22	1.07	1.02	1.49
0.25	2.49	1.02	1.61	1.01	0.98	1.13	0.95	1.42	1.39	4.01	1.07	1.33	1.26	1.08	2.07	1.66
0.5	4.09	1.03	1.44	0.94	0.96	1.07	0.88	1.30	1.56	2.49	1.02	1.28	1.21	1.06	4.34	1.41
0.75	1.74	1.03	1.37	0.89	0.93	1.01	0.82	1.18	1.52	2.27	0.94	1.21	1.11	1.02	7.37	1.19
1	1.53	0.99	0.96	0.80	0.89	0.94	0.71	0.92	1.16	1.19	0.84	1.11	1.02	0.96	13.4	0.99
1.5	0.96	0.84	0.66	0.63	0.82	0.74	0.54	0.75	0.93	0.80	0.65	0.91	0.79	0.78	3.85	0.71
2	0.65	0.65	0.49	0.53	0.73	0.58	0.47	0.59	0.69	0.73	0.53	0.73	0.64	0.59	2.90	0.58
4	0.42	0.37	0.40	0.42	0.46	0.44	0.41	0.41	0.48	0.55	0.45	0.47	0.43	0.34	2.52	0.52
8	0.40	0.36	0.40	0.41	0.41	0.42	0.40	0.40	0.42	0.53	0.44	0.43	0.38	0.33	2.52	0.52
mean	1.52	0.81	0.99	0.74	0.80	0.83	0.68	0.91	1.02	1.53	0.78	0.97	0.90	0.80	4.44	1.01

Table 54. Means of error ratios for extrapolated Tikhonov approximation with rule maxDe.

p	baart	deriv2	foxgood	gravity	heat	ilaplace	phillips	shaw	spikes	wing	gauss	hilbert	lotkin	moler	pascal	prolate
0	1.43	1.03	1.47	1.12	1.01	1.18	1.01	1.30	1.02	1.19	1.18	1.29	1.24	1.01	1.02	1.60
0.25	2.54	1.02	1.55	1.10	0.97	1.21	0.96	1.55	1.43	3.97	1.18	1.45	1.33	0.98	1.99	1.92
0.5	4.23	0.99	1.38	1.00	0.90	1.18	0.87	1.43	1.64	2.36	1.12	1.42	1.27	0.94	4.00	1.67
0.75	1.82	0.94	1.26	0.92	0.84	1.11	0.81	1.27	1.67	2.19	1.04	1.38	1.13	0.87	6.53	1.46
1	1.58	0.87	0.84	0.81	0.78	1.02	0.68	0.93	1.21	1.19	0.93	1.25	1.03	0.80	11.9	1.26
1.5	0.92	0.68	0.58	0.59	0.69	0.74	0.50	0.73	0.98	0.69	0.66	1.01	0.74	0.62	3.06	0.92
2	0.56	0.49	0.37	0.45	0.57	0.54	0.42	0.55	0.63	0.63	0.52	0.76	0.55	0.43	1.74	0.71
4	0.27	0.25	0.24	0.30	0.30	0.33	0.33	0.30	0.35	0.37	0.35	0.36	0.28	0.20	1.49	0.55
8	0.24	0.22	0.23	0.27	0.24	0.28	0.29	0.26	0.28	0.30	0.31	0.28	0.21	0.18	1.49	0.53
mean	1.51	0.72	0.88	0.73	0.70	0.84	0.65	0.93	1.02	1.43	0.81	1.02	0.86	0.67	3.69	1.18

Table 55. Means (over all p) of error ratios and errors for problem heat.

δ	D	D,2	Me	Me,2	2MEe	2MEe,2	3De	3De,2	$\|x_{\lambda_*} - x_*\|$
0.5	1.06	2.49	1.07	1.92	0.99	1.12	0.87	1.19	3.91e-1
10^{-1}	1.12	2.17	1.06	1.88	0.93	1.52	0.83	2.33	2.10e-1
10^{-2}	1.26	1.55	1.07	1.57	0.85	1.28	0.77	1.72	7.14e-2
10^{-3}	1.48	1.28	1.08	1.28	0.80	1.01	0.71	1.26	2.15e-2
10^{-4}	1.81	1.26	1.10	1.13	0.78	0.78	0.68	0.92	7.19e-3
10^{-5}	2.51	1.47	1.13	1.06	0.77	0.70	0.67	0.74	3.18e-3
10^{-3}	3.45	2.19	1.15	1.04	0.75	0.67	0.66	0.68	1.51e-3
mean	1.81	1.77	1.09	1.41	0.84	1.01	0.74	1.26	

4.8.2. Rules not using noise level.

Tables 56–58 contain the results for rules using minimization strategy in extrapolated Tikhonov method with 2 or 3 terms (for analogous results in single Tikhonov approximation see Tables 18–20). The lower part shows the means of error ratios in problems of Table 2 with all solution vectors of Table 3. The rules were selected as the best rules for single Tikhonov approximation.

Te8) Rules QC, R2C, BRSC choose the parameter by the climbing approach in the functions $\varphi_{\mathrm{eQ}}(\alpha)$, $\varphi_{\mathrm{eR2}}(\alpha)$, and $\varphi_{\mathrm{BRS}}(\alpha)$, respectively, where $C = 7$ in 2-extrapolated Tikhonov approximation and $C = 7.5$ in 3-extrapolated Tikhonov approximation.

Te9) Rules DR21 and BRS1 choose the parameter as the largest local minimum of the functions $\varphi_{\mathrm{D}}(\alpha)^{0.9}\varphi_{\mathrm{eR2}}(\alpha)^{0.1}\alpha^{0.4}$ and $\varphi_{\mathrm{BRS}}(\alpha)\alpha^{c}$, where $c = 0.58$ or $c = 0.61$ in 2- or 3-extrapolated approximation, respectively.

Rules Te8 require a somewhat larger constant C in conjunction with the climbing approach than their analogs T10 of Section 4.3.2. Note also the difference in weighted geometric averages in rules DR21 in Te9 and T12.

In most cases the results remained good. The rule DR21 was good in case $p = 0$ and essentially worse in case $p = 2$. Note also that the problem 14 was hard to most rules in case $p = 0$ and the problem 15 in case $p = 2$. The error in problem 15 has a very sharp minimum, if $p = 2$, while the functions used in our rules are quite smooth.

Table 56. Means of error ratios for 2-extrapolated and 3-extrapolated Tikhonov approximation, $p = 0$.

	2-extrapolated Tikhonov					3-extrapolated Tikhonov				
Probl.	QC	R2C	BRSC	DR21	BRS1	QC	R2C	BRSC	DR21	BRS1
1	1.56	1.61	2.63	2.18	2.69	1.91	1.86	2.63	2.62	2.72
2	1.62	1.58	1.47	1.59	1.36	1.62	1.52	1.47	1.59	1.52
3	2.43	2.37	7.08	2.99	4.14	2.79	2.85	7.18	3.19	14.5
4	1.11	1.09	2.23	1.15	1.67	1.10	1.16	2.20	1.16	1.69
5	1.71	1.69	1.42	1.83	1.17	1.70	1.67	1.43	1.75	1.18
6	1.19	1.18	1.85	1.28	1.50	1.19	1.19	1.82	1.28	1.49
7	1.09	1.08	1.78	1.07	1.33	1.09	1.06	1.79	1.08	1.35
8	1.52	1.51	2.30	1.48	1.94	1.52	1.45	2.27	1.58	2.25
9	1.04	1.04	1.06	1.05	1.07	1.04	1.04	1.06	1.05	1.07
10	1.42	1.42	1.55	1.51	1.86	1.42	1.44	1.54	1.85	1.87
mean	1.47	1.46	2.34	1.61	1.87	1.54	1.52	2.34	1.72	2.96
11	1.18	1.17	2.01	1.20	1.58	1.19	1.22	2.03	1.25	1.59
12	1.99	1.95	2.83	2.05	2.30	1.98	2.33	2.84	2.06	2.30
13	3.26	3.25	3.32	3.14	3.52	3.27	3.06	3.33	3.20	3.54
14	1.99	1.93	2.22	1.91	20.3	2.07	2.22	5.67	5.37	20.5
15	1.06	1.06	1.06	1.06	1.06	1.06	1.05	1.06	1.06	1.06
16	1.66	1.65	2.27	1.74	1.84	1.70	1.92	2.34	2.76	4.18
mean	1.86	1.84	2.28	1.85	5.10	1.88	1.97	2.88	2.62	5.52

Table 57. Means of error ratios for 2-extrapolated and 3-extrapolated Tikhonov approximation, $p = 0$ (correlated noise).

	2-extrapolated Tikhonov					3-extrapolated Tikhonov				
Probl.	QC	R2C	BRSC	DR21	BRS1	QC	R2C	BRSC	DR21	BRS1
1	1.55	1.56	2.54	2.12	2.64	1.90	1.86	2.53	2.59	2.67
2	1.47	1.44	1.33	1.88	1.27	1.65	1.52	1.32	1.88	1.31
3	2.62	2.61	8.38	3.24	4.60	2.89	2.85	8.55	3.43	15.5
4	1.16	1.15	2.49	1.31	1.87	1.16	1.16	2.45	1.33	1.89
5	1.72	1.67	1.59	13.4	1.55	1.70	1.67	1.60	13.4	2.09
6	1.21	1.20	1.94	1.34	1.57	1.20	1.19	1.91	1.34	1.56
7	1.04	1.05	1.95	4.21	1.46	1.04	1.06	2.34	4.27	1.48
8	1.41	1.41	2.32	1.46	1.91	1.45	1.45	2.29	1.54	2.21
9	1.04	1.04	1.06	1.05	1.06	1.04	1.04	1.06	1.05	1.06
10	1.44	1.44	1.55	1.52	1.87	1.45	1.44	1.55	1.86	1.88
mean	1.47	1.46	2.52	3.15	1.98	1.55	1.52	2.56	3.27	3.17
11	1.20	1.20	2.40	1.42	1.86	1.21	1.22	2.39	1.47	1.87
12	2.37	2.33	3.57	2.54	2.86	2.35	2.33	3.59	2.56	2.87
13	3.07	3.06	3.33	3.04	3.44	3.10	3.06	3.35	3.12	3.48
14	2.28	2.13	2.35	2.15	23.7	2.38	2.22	6.95	6.66	23.9
15	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05
16	1.86	1.88	2.77	2.00	2.23	1.92	1.92	3.02	3.59	5.21
mean	1.97	1.94	2.58	2.03	5.86	2.00	1.97	3.39	3.08	6.40

Table 58. Means of error ratios for 2-extrapolated and 3-extrapolated Tikhonov approximation, $p = 2$.

	2-extrapolated Tikhonov					3-extrapolated Tikhonov				
Probl.	QC	R2C	BRSC	DR21	BRS1	QC	R2C	BRSC	DR21	BRS1
1	1.10	1.17	2.42	0.68	1.51	1.07	1.17	2.43	0.89	1.69
2	0.61	0.67	1.06	0.62	0.74	0.59	0.65	0.99	0.57	0.73
3	0.82	0.86	1.88	0.53	1.11	0.64	0.81	1.59	0.62	1.11
4	0.65	0.69	0.90	0.60	0.61	0.63	0.67	0.86	0.56	0.63
5	0.72	0.79	0.97	0.74	0.73	0.70	0.78	0.92	0.71	0.71
6	0.70	0.76	1.12	0.65	0.75	0.67	0.75	1.15	0.62	0.81
7	0.54	0.58	0.74	0.54	0.54	0.51	0.55	0.67	0.48	0.51
8	0.81	0.87	1.29	0.66	0.83	0.81	0.90	1.30	0.64	0.96
9	0.82	0.87	2.02	0.75	1.21	0.81	0.87	1.99	0.74	1.30
10	1.19	1.28	3.04	1.02	1.87	0.93	1.02	2.57	1.22	1.83
mean	0.79	0.85	1.54	0.68	0.99	0.74	0.82	1.45	0.71	1.03
11	0.63	0.69	0.85	0.60	0.61	0.61	0.67	0.86	0.57	0.64
12	0.91	0.97	2.06	0.79	1.34	0.94	1.00	2.20	0.80	1.43
13	0.89	0.89	1.89	0.83	1.30	0.87	0.87	1.83	0.83	1.36
14	0.57	0.62	1.42	0.59	0.90	0.55	0.59	1.30	0.59	0.87
15	3.62	3.64	14.5	3.63	6.72	3.62	3.63	6.88	3.62	5.11
16	0.66	0.68	0.62	0.64	0.60	0.63	0.65	0.63	0.59	0.59
mean	1.21	1.25	3.55	1.18	1.91	1.20	1.23	2.28	1.17	1.67

4.8.3. Rule R2e.

Table 59 shows the means of error ratios for rule R2e in 2-extrapolated Tikhonov method at various p and d. Comparison of Table 59 with 21, where results for ordinary Tikhonov method are given, shows that in 2-extrapolated Tikhonov method, the rule R2e is much more sensitive to underestimation of the noise level.

Table 59. Means of error ratios in 2-extrapolated Tikhonov method with the rule R2e, using the noise level that is d times the actual noise level.

$p \setminus d$	0.5	0.6	0.8	1	1.3	1.6	2	3	5	10
0	205	25.9	1.90	1.43	1.42	1.46	1.52	1.63	1.90	2.26
0.25	306	34.0	3.16	2.78	2.74	2.83	2.92	3.11	3.56	4.28
0.5	404	46.9	2.63	2.11	2.01	2.14	2.42	2.87	3.50	4.11
0.75	321	38.8	2.37	1.56	1.55	1.64	1.83	2.19	2.83	3.65
1	390	33.0	2.09	1.30	1.29	1.35	1.41	1.61	2.13	3.20
1.5	488	27.1	1.22	0.85	0.83	0.85	0.88	1.09	1.45	2.02
2	388	23.2	0.83	0.69	0.67	0.67	0.67	0.73	0.92	1.38
4	397	6.24	0.57	0.54	0.52	0.50	0.48	0.48	0.53	0.64
8	395	5.52	0.56	0.53	0.51	0.49	0.47	0.47	0.52	0.61
mean	366	26.7	1.70	1.31	1.28	1.32	1.40	1.57	1.93	2.46

4.8.4. Rule DM.

Tables 60, 61 present the results for the rule DM in 2-extrapolated Tikhonov method at various levels of noise level ill-estimations d. These tables are analogs of Tables 23, 24 of ordinary Tikhonov method. If $p = 2$, then the rule DM is able to choose the regularization parameter in such way that the 2-extrapolated Tikhonov approximation is better, on average, than the best single Tikhonov approximation even in wider range of d than $[0.1, 10]$ for uncorrelated noise (Table 60). For correlated noise the results are not so good but for d close to 1 the error still depends only weakly on d.

Table 62 compares some sets of parameters c_1, c_2 in rule DM.

Figure 4 compares the rule DM with the rule R2C, illustrating the dependence of the error on d.

In Tables 56 and 58 the rule R2C gives averages 1.46 for $p = 0$ and 0.85 for $p = 2$, in 2-extrapolated Tikhonov method, hence DM is superior over R2C, if $d \in [0.3, 20]$ in case $p = 0$ or $d \in [1, 50]$ in case $p = 2$.

Table 60. Means of error ratios in 2-extrapolated Tikhonov method with the rule DM, $c_1 = 0.002$, $c_2 = 0.03$ for $p = 0$ (left), $p = 2$ (right), where the used noise level is d times the actual noise level.

	Case $p = 0$, values of d							Case $p = 2$, values of d						
Probl.	0.01	0.1	0.5	1	2	10	100	0.01	0.1	0.5	1	2	10	100
1	1.53	1.53	1.53	1.54	1.56	1.71	2.55	1.24	1.24	1.24	1.17	1.08	0.87	2.13
2	7.93	5.04	1.57	1.57	1.56	1.09	1.20	15.9	0.68	0.68	0.68	0.68	0.67	0.94
3	2.55	2.55	2.55	2.55	2.55	2.27	6.26	0.88	0.88	0.88	0.88	0.89	0.60	1.77
4	1.11	1.11	1.11	1.11	1.11	1.08	1.62	0.70	0.70	0.70	0.70	0.70	0.65	0.87
5	6.42	2.37	1.15	1.15	1.10	1.10	1.17	0.79	0.79	0.79	0.79	0.79	0.79	0.82
6	1.19	1.19	1.19	1.19	1.19	1.19	1.44	0.77	0.77	0.77	0.77	0.77	0.74	1.05
7	1.08	1.08	1.08	1.08	1.08	1.08	1.34	0.58	0.58	0.58	0.58	0.58	0.58	0.72
8	1.45	1.45	1.45	1.45	1.46	1.49	2.12	0.92	0.92	0.92	0.92	0.86	0.76	1.20
9	1.04	1.04	1.04	1.04	1.04	1.04	1.06	0.89	0.89	0.89	0.89	0.89	0.79	1.58
10	1.42	1.42	1.42	1.42	1.43	1.48	1.54	1.34	1.34	1.34	1.13	1.05	0.77	2.59
mean	2.57	1.88	1.41	1.41	1.41	1.35	2.03	2.40	0.88	0.88	0.85	0.83	0.72	1.37
11	1.17	1.17	1.17	1.17	1.17	1.16	1.58	0.70	0.70	0.70	0.70	0.70	0.68	0.86
12	1.56	1.56	1.56	1.56	1.56	1.64	2.32	1.03	1.03	1.03	1.03	1.03	0.87	1.72
13	1.79	1.79	1.79	1.79	1.79	1.77	2.63	0.95	0.95	0.95	0.95	0.95	0.80	1.71
14	18.3	4.73	2.92	2.08	1.84	1.56	1.61	7.61	0.62	0.62	0.62	0.62	0.60	1.23
15	1.05	1.05	1.05	1.06	1.06	1.06	1.06	3.71	3.69	3.54	3.42	3.25	2.63	11.1
16	1.46	1.46	1.46	1.46	1.44	1.57	2.33	0.71	0.71	0.71	0.71	0.71	0.64	0.78
mean	4.22	1.96	1.66	1.52	1.48	1.46	1.92	2.45	1.28	1.26	1.24	1.21	1.04	2.90

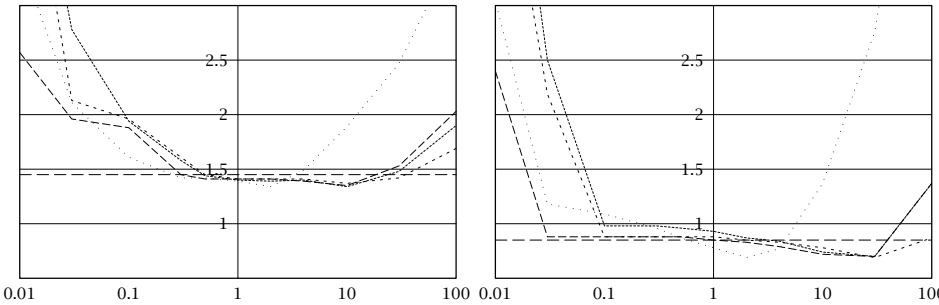
Table 61. Means of error ratios in 2-extrapolated Tikhonov method with the rule DM, $c_1 = 0.002$, $c_2 = 0.03$ for $p = 0$ (left), $p = 2$ (right), where the used noise level is d times the actual noise level (correlated noise).

	values of d for $p = 0$							values of d for $p = 2$						
Probl.	0.01	0.1	0.5	1	2	10	100	0.01	0.1	0.5	1	2	10	100
1	2.72	1.50	1.52	1.52	1.54	1.69	2.48	1.40	1.35	1.33	1.21	1.12	1.18	3.70
2	6.56	2.34	1.88	1.69	1.17	1.09	1.45	12.4	1.00	0.81	0.68	0.68	0.73	1.55
3	2.59	2.59	2.59	2.53	2.54	3.01	8.13	1.12	1.12	1.11	1.10	1.11	0.92	2.67
4	1.16	1.16	1.16	1.16	1.16	1.19	1.97	0.67	0.67	0.67	0.67	0.66	0.71	1.64
5	4.34	2.10	1.49	1.17	1.20	1.20	1.45	0.80	0.80	0.80	0.80	0.80	0.81	1.41
6	1.18	1.18	1.18	1.18	1.18	1.23	1.60	0.74	0.74	0.73	0.73	0.74	0.81	2.01
7	1.45	1.05	1.05	1.05	1.05	1.07	1.62	0.59	0.59	0.59	0.59	0.59	0.63	1.39
8	1.32	1.32	1.32	1.32	1.33	1.48	2.18	0.83	0.83	0.81	0.82	0.82	0.91	2.57
9	1.03	1.03	1.03	1.03	1.03	1.04	1.06	0.80	0.80	0.80	0.81	0.81	0.90	2.56
10	1.44	1.44	1.44	1.45	1.45	1.49	1.55	2.87	2.86	2.84	2.72	2.50	2.25	4.50
mean	2.38	1.57	1.47	1.41	1.37	1.45	2.35	2.22	1.07	1.05	1.01	0.98	0.99	2.40
11	1.21	1.21	1.21	1.21	1.21	1.29	2.04	0.66	0.66	0.66	0.65	0.65	0.69	1.63
12	1.87	1.87	1.87	1.88	1.88	2.14	3.26	1.09	1.09	1.07	1.07	1.07	1.09	3.25
13	1.60	1.60	1.61	1.62	1.63	1.77	2.75	1.08	1.08	1.08	1.07	1.01	1.11	3.13
14	20.4	5.02	2.90	2.11	1.96	1.55	1.84	9.15	0.87	0.70	0.70	0.70	0.77	1.82
15	1.05	1.05	1.05	1.05	1.05	1.05	1.05	17.1	17.4	17.2	17.0	16.8	15.2	12.6
16	1.62	1.62	1.64	1.63	1.62	1.93	3.17	0.75	0.75	0.75	0.74	0.74	0.68	1.32
mean	4.63	2.06	1.71	1.58	1.56	1.62	2.35	4.97	3.64	3.58	3.54	3.50	3.26	3.96

Table 62. Averages of error ratios over problems of Hansen for the rule DM for different c_1, c_2; $p = 0$ (upper) and $p = 2$ (lower).

				d										
p	Nr	c_1	c_2	0.01	0.03	0.1	0.3	0.5	1	2	4	10	30	100
0	I	0.02	0.14	3.45	2.11	1.61	1.42	1.42	1.39	1.34	1.49	1.89	2.47	3.53
0	II	0.002	0.07	5.28	2.78	1.94	1.58	1.44	1.40	1.39	1.40	1.34	1.48	1.90
0	III	0.002	0.03	2.57	1.96	1.88	1.45	1.41	1.41	1.41	1.39	1.35	1.53	2.03
0	IV	0.001	0.03	4.99	2.13	1.96	1.61	1.45	1.41	1.41	1.41	1.37	1.42	1.69
2	I	0.02	0.14	3.04	1.18	1.09	0.96	0.87	0.78	0.69	0.77	1.37	2.76	5.86
2	II	0.002	0.07	5.99	2.50	0.98	0.98	0.96	0.93	0.87	0.84	0.74	0.70	1.37
2	III	0.002	0.03	2.40	0.88	0.88	0.88	0.88	0.85	0.83	0.79	0.72	0.70	1.37
2	IV	0.001	0.03	4.95	2.19	0.88	0.88	0.88	0.88	0.85	0.83	0.78	0.69	0.87

Figure 4. Error ratios in 2-extrapolated Tikhonov method for rules R2C and DM: $p = 0$ (left) and $p = 2$ (right); ····· DM I, ····· DM II, – – – DM III, - - - - - DM IV, — — R2C.



V. CONCLUSIONS AND DISCUSSION

5.1. Recommendation of methods

In most test problems for both $p = 0$ and $p = 2$ the Landweber method and TSVD had the highest potential to give the best accuracy of regularized solution, in other problems their results lied close to the best (Tables 4, 5, 6, 7). To harness the potential of the Landweber method, however, one has to use additional techniques such as operator iterations (2.19), since in its original form (2.18) the Landweber method converges quite slowly. This sets certain restrictions on the matrix of the linear system. If using operator iterations (2.19) is not computationally feasible, and the size of the problem is small or moderate, we recommend TSVD. If the size is large, then we recommend either CGLS or Tikhonov method. The CGLS method works faster in conjunction with rules that find the regularization parameter as the first index for which a certain condition is satisfied, as opposed to rules which minimize expressions. In Tikhonov method both solving of an equation and minimizing a function in an interval can be implemented efficiently, using, for example, method of chords or Newton's method but the Tikhonov method has quite low qualification $p_0 = 1$. In case of smooth solution the advantage of Landweber method, TSVD and CGLS over other methods is more pronounced (Tables 6, 7).

Remaining methods performed not so good in our tests: in Tables 4–7 the method CGME was always beaten by the method CGLS. The Lavrentiev method gave more or less acceptable results in limited number of self-adjoint problems with low smoothness index p.

If a sequence of Tikhonov approximations arises during computations and we have a ground to assume that the smoothness index p of the solution is large ($p > 1$), then we strongly recommend to use extrapolation. As Tables 48, 49 show, in this case even an extrapolated approximation with 2 terms, where the regularization parameter is chosen from an *a posteriori* rule using noise level, has an accuracy that is better than the accuracy of the best single Tikhonov approximation. For rules R2e and Me this holds even when the noise level is overestimated 2 times, if the noise is uncorrelated (or if $p > 2$ in case of correlated noise). Also the best heuristic rules of Table 58 gave averages below 1, if $p = 2$. The results for larger p were quite uniform over problems, see Table 53, except for problems, in which the error had a very sharp minimum.

Similar conclusions can be made in case of iterated Tikhonov method, see Tables 14, 15, but compared to extrapolated Tikhonov approximation, parameter choice in iterated Tikhonov method may require more computational work. For example, if parameters are chosen from the sequence $\alpha_n = q^n$, then transition from α_n to α_{n+1} requires solving m equations in m-iterated Tikhonov approximations but only one equation in extrapolated Tikhonov approximations with m terms (since Tikhonov approximations with parameters α_n, α_{n-1}, ..., α_{n-m+1} are computed earlier).

5.2. Recommendation of rules

5.2.1. Case of exactly known noise level.

If the bound of the noise level is known ($\|y - y_*\| \le \delta$ with δ known), then in Tikhonov method and its iterated and extrapolated variants we recommend the rule Me. Tables 8, 9, 10, 11, 14, 15, 48, 49, 50, 51 show that although in case of exactly known noise level the rules MEe and MEs gave somewhat better results than the rule Me, in case of 2 times overestimated noise level the rule Me was clearly better. This did not depend on whether the noise was correlated or uncorrelated.

In Lavrentiev method we recommend the rule MEaql with $l = 5$ for $p = 0$ and with $l = 4$ for $p = 1$, since in Tables 28, 29, it gave the smallest averages *and* the smallest maximums. In addition, Table 30 shows that the rule MEaql was also reasonably insensitive to moderate overestimation of the noise level. The rule MEaql requires computing Lavrentiev approximations at different values of the regularization parameter, so we may form extrapolated approximations that give smaller averages of error but as Table 28 shows, maximums of error can be quite large at $p = 0$. Also, it can be seen from Table 30 that the extrapolated approximation can lose its advantage over single Lavrentiev approximation with rule MEaql the more we overestimate the actual noise level. Extrapolated approximations may be useful for larger p, see Table 29, where both averages and maximums of 2-extrapolated Lavrentiev approximation with the rule MD are small among rules compared.

In Landweber method the estimated parameters from rules De and MEe show clear advantage over discrepancy principle both for uncorrelated and correlated noise (Tables 34, 35, 36). At overestimated noise level the rule De seems to be slightly better than MEe.

In TSVD method the discrepancy principle works well, post-estimation of the truncation level did not improve it essentially.

In conjugate gradient type methods CGLS and CGME the discrepancy principle D and its variant DH work very well (Tables 40, 41, 42, 43, 44, 45). In CGLS the accuracy of the discrepancy principle can be somewhat improved by using post-estimation as in rules De and Me (Tables 40, 41).

In extrapolated Tikhonov approximation the best results at exact noise level were produced by the rule De (Tables 48, 49, 50, 51) but it is very sensible to overestimation of the noise level. In this respect the two similar rules MEe and MEs may be preferred. Even better stability is provided by the rule Me but its results at exact noise level were not so good, compared with others, especially for large p. For small p the rule Me in 3-extrapolated Tikhonov method gave very good results: its averages of error ratios were among the smallest both in case of uncorrelated noise and correlated noise and both at exact noise level and 2 times overestimated noise level.

The difference in performance of rules under uncorrelated noise or correlated noise reveals itself most prominently in case of 2 times overestimated noise level. Comparing the performance of rules under different noise models, we see from Tables 8, 9, 10, 11, 14, 15, 48, 49, 50, 51 about Tikhonov method and its extrapolated variant that while in case of exactly known noise level the averages of error ratios for correlated noise were somewhat larger that the same averages for uncorrelated noise, in case of 2 times overestimated noise level the averages of error ratios for correlated noise were much larger. For iteration methods this dependence on precision of noise level estimation is not so emphasized (Tables 34, 35, 40, 41, 43, 44).

5.2.2. Case of unknown noise level.

If the noise level is unknown, then, in the Tikhonov method, we recommend to use the climbing strategy. Comparing the columns QC and BRSC of Table 18 with columns QN and BRS of Table 16, respectively, we see that the results in corresponding columns in Tables 18 and 16 are almost the same in most problems but in some problems rules QN and BRS fail. This holds for both $p = 0$ and $p = 2$ (Tables 20 and 17). Stopping at the first local minimum can avoid large errors in many problems and is therefore also an option but as the columns Q1 and BRS1 of Tables 18, 19, 20 show, the results were not so good compared with columns QC and BRSC of these tables.

For Tikhonov method in case $p = 0$ we can recommend the rules R2C and R2C', which gave the smallest overall averages in our tests both for correlated and uncorrelated noise (Tables 18, 19). Also rules QC and QC' can be recommended. The rules R2C and QC gave the best results also in extrapolated variants of Tikhonov method (Tables 56, 57, 58).

The rules QHR2 and QBR2 gave very good results in problems 1–10 in Tables 18, 19, 20 but they needed considerably more computation time than rules that did not minimize a function unconditionally in the whole computation interval (in our case $[10^{-30}, 1]$).

Of special interest is the rule QHR, which, not using any parameters, gave a very small average at $p = 2$ in Table 17.

For Landweber method the rules HRmC and QNmC can be preferred to their unmodified counterparts HR and QN, since in most problems they had better accuracy and, in particular, avoided large errors (Tables 34, 35, 36). The rule QNmC is notable for its quite good overall average that is smaller than the average of rules D, De, and MEe (rules that use δ) in case of 2 times overestimated noise level.

For TSVD we recommend the rule HRLm'.

For CGLS the rule HRmC can be recommended (Tables 40, 41, 42). In some problems the rule HRmWC gave better averages but computing with this rule was very time-consuming and in many cases the results were the same as with rule HRmC.

In CGME the rule RMC gave the best overall averages for both uncorrelated and correlated noise (Tables 43, 44, 45). The rule DHP gave very good (even optimal) results in most test problems.

For extrapolated Tikhonov method the above-mentioned rules QC and R2C are recommended. The rule DR21 for 2-extrapolated Tikhonov method was close to the best in many test problems (Tables 56, 57) but in some problems the errors were intolerably large. For $p = 2$ this rule gave the best results among rules compared (Table 58), so it can be used, if we know that the solution is sufficiently smooth.

If the noise level is unknown, then the results by most of our rules in all methods depend only a little on whether noise is correlated or not (Tables 18, 19, 34, 35, 40, 41, 43, 44, 56, 57).

5.2.3. Case of approximately known noise level.

For Tikhonov method the rule DM gave results that were very good at exactly known noise level and still acceptable, when the noise level was up to 100 times over- or underestimated (Tables 23, 24).

If the probable bounds of over- or underestimation are known, then Table 26 gives the best values of constants c_1 and c_2 in our tests with Table 27 showing the corresponding minimums and maximums of averages of error ratios over the supposed interval of noise ill-estimation coefficient d (for uncorrelated noise).

For Landweber method we could not find the rule that works better than the best heuristic rule QNmC (Table 34) in our numerical tests. There also remains the task of formulating and testing the DM rule for TSVD.

For CGLS the rule DM gave the results that were intermediate between the results of best rules that use full noise level information and best rules that do not use this information at all (Tables 46, 47). This rule works best, if $p = 0$ and the noise is uncorrelated. If $p > 0$, then this rule requires more precise estimation of noise level, especially in case of correlated noise.

In extrapolated Tikhonov method the rule DM can be applied with the same constants c_1, c_2 as in ordinary Tikhonov method (Tables 60, 61), in

case of smooth solution the rule DM can give the extrapolated approximation that is more accurate than single Tikhonov approximation with optimal parameter. In extrapolated Tikhonov method the rule DM may be more sensitive to values of parameters c_1, c_2, if the noise level is severely over- or underestimated (Table 62).

5.3. About the set of test problems

We are aware of the limitations that are associated with our set of test problems. However, some points can be made to assume that the results remain valid also for other problems.

- Many rules that we propose are based on theoretical foundation (see Section 3.4). Theoretical explanation of many heuristic rules that we propose is given in Section 3.2.2: for many rules the theoretical ground is Theorem 17.
- The set of test problems of Hansen is very diverse, containing problems with quite different characteristics. Still, in many rules we can choose the constants in such way that these rules work uniformly well in all test problems, in whole range of noise levels δ and smoothness indices p, and both for uncorrelated and correlated noise.
- On a later stage of this work in many methods experiments were made on additional very different artificial problems (11–16). The heuristic rules worked well in problems 11–16 with the constants optimized for problems 1–10 or with slightly changed constants.

The representativeness of the set of test problems is essential and if there were a wider representative set of test problems available, then it would be essential to use it. Very important is the representativeness of the set of test problems for heuristic parameter choice rules that do not use noise level. It would also be interesting to test the non-heuristic rules on additional problem sets.

In addition, it is essential to consider other possibilities of noise generation. Namely from papers [49, 69] it is known that some heuristic rules (quasioptimality criterion and its modification by Neubauer) guarantee the convergence of regularized solutions (despite Bakushinskii's result [1]), if the noise distribution satisfies certain conditions.

The exact formulations of rules are not considered final, especially in case of unknown noise level. The results presented in previous tables indicate that there may be more resources for improvement in case of iteration methods and TSVD.

APPENDIX

A.1. A systematics of parameter choice rules in Tikhonov method

Many expressions of functions in stopping rules for m-iterated Tikhonov method contain scalar products of type

$$\big((AA^*)^{i_1} r_{m+j_1;\alpha}, (AA^*)^{i_2} r_{m+j_2;\alpha}\big),$$

where i_1, j_1, i_2, j_2 are some nonnegative integers. Since this scalar product is equal to

$$\begin{aligned}\big((AA^*)^{i_1} K_{m+j_1;\alpha}(AA^*)\overline{r}, (AA^*)^{i_2} K_{m+j_2;\alpha}(AA^*)\overline{r}\big)\\ = \big((AA^*)^{i_1+i_2} K_{m;\alpha}(AA^*)\overline{r}, K_{m+j_1+j_2;\alpha}(AA^*)\overline{r}\big),\end{aligned}$$

only two of four indices are independent. Therefore we can define quantities

$$\Psi_{i,j} = \alpha^{-i/2}\big((AA^*)^i r_{m;\alpha}, r_{m+j;\alpha}\big)^{1/2} \qquad (i,\ j = 0,\ 1,\ \dots). \tag{1.1}$$

Proposition. *For all positive integers i, j holds*

$$\Psi_{i,j}^2 = \Psi_{i-1,j-1}^2 - \Psi_{i-1,j}^2. \tag{1.2}$$

Proof. Using (2.17), we have

$$\begin{aligned}\Psi_{i,j}^2 &= \alpha^{-i}\big((AA^*)^{i-1} r_{m;\alpha}, (AA^*) r_{m+j;\alpha}\big)\\ &= \alpha^{-i}\big((AA^*)^{i-1} r_{m;\alpha}, \alpha A(x_{m+j-1;\alpha} - x_{m+j;\alpha})\big)\\ &= \alpha^{-(i-1)}\big((AA^*)^{i-1} r_{m;\alpha}, r_{m+j-1;\alpha} - r_{m+j;\alpha}\big) = \Psi_{i-1,j-1}^2 - \Psi_{i-1,j}^2. \quad \square\end{aligned}$$

Corollary. *If $i \le j$, then*

$$\Psi_{i,j}^2 = \sum_{k=0}^{i} (-1)^{i-k} \binom{i}{k} \Psi_{0,j-k}^2.$$

Proof. Induction on i, using (1.2). $\square$

Using the last corollary, it is possible to calculate $\Psi_{i,j}$ in such way that no application of AA^* to $r_{m;\alpha}$ is needed. Computational experience shows that this is numerically more stable than applying the direct formula (1.1). Hereby one needs to compute additional discrepancies $r_{m+1;\alpha}, \dots, r_{m+j-1;\alpha}$ but they have to be computed also in (1.1) to get $r_{m+j;\alpha}$.

Note that $d_{\mathrm{D}}(\alpha) = \Psi_{0,0}$, $d_{\mathrm{ME}}(\alpha) = \Psi_{0,1}^2/\Psi_{0,2}$, $d_{\mathrm{R1},k}(\alpha) = \Psi_{2k,2k+1}$, and $d_{\mathrm{R2}}(\alpha) = \kappa(\alpha)\Psi_{1,2}^2/\Psi_{1,3}$.

A.2. Index of rules for choice of the regularization parameter

In the following table, iTikh, eTikh, iLavr, eLavr mean iterated and extrapolated variants of Tikhonov and Lavrentiev method.

BIBLIOGRAPHY

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

[2] F. Bauer. Some considerations concerning regularization and parameter choice algorithms. *Inverse Problems*, 23:837–858, 2007.

[3] F. Bauer and T. Hohage. A Lepskij-type stopping rule for regularized Newton methods. *Inverse Problems*, 21(6):1975–1991, 2005.

[4] F. Bauer, T. Hohage, and A. Munk. Iteratively regularized Gauss-Newton method for nonlinear inverse problems with random noise. *SIAM J. Numer. Anal.*, 47:1827–1846, 2009.

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

[6] F. Bauer and S. Kindermann. Recent results on the quasioptimality principle. *J. Inv. and Ill-Posed Problems*, 17(1):5–18, 2009.

[7] F. Bauer, P. Mathé, and S. Pereverzev. Local solutions to inverse problems in geodesy: The impact of the noise covariance structure upon the accuracy of estimation. *J. of Geodesy*, 81(1):39–51, 2007.

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

[9] F. Bauer and S. Pereverzev. Regularization without preliminary knowledge of smoothness and error behaviour. *European J. Appl. Math.*, 16(3):303–317, 2005.

[10] F. Bauer and M. Reiss. Regularisation independent of the noise level: an analysis of quasi-optimality. *Inverse Problems*, 24:(055009), 2008.

[11] C. Brezinski, M. Redivo-Zaglia, G. Rodriguez, and S. Seatzu. Extrapolation techniques for ill-conditioned linear systems. *Numer. Math.*, 81(1):1–29, 1998.

[12] C. Brezinski, M. Redivo-Zaglia, G. Rodriguez, and S. Seatzu. Multiparameter regularization techniques for ill-conditioned linear systems. *Numer. Math.*, 94(2):203–228, 2003.

[13] C. Brezinski, G. Rodriguez, and S. Seatzu. Error estimates for linear systems with applications to regularization. *Numer. Algor.*, 49:85–104, 2008.

[14] C. Brezinski, G. Rodriguez, and S. Seatzu. Error estimates for the regularization of least squares problems. *Numer. Algor.*, 51:61–76, 2009.

[15] H. Cao, M. Klibanov, and S. V. Pereverzev. A Carleman estimate and the balancing principle in the quasi-reversibility method for solving the Cauchy problem for the Laplace equation. *Inverse Problems*, 25(3):035005 (21 pp), 2009.

[16] H. Cao and S. Pereverzev. The balancing principle for the regularization of elliptic Cauchy problems. *Inverse Problems*, 23(5):1943–1961, 2007.

[17] Z. Chen, Y. Xu, and H. Yang. Fast collocation methods for solving ill-posed integral equations of the first kind. *Inverse Problems*, 24(6):065007 (21 pp), 2008.

[18] P. P. B. Eggermont, V. N. LaRiccia, and M. Z. Nashed. On weakly bounded noise in ill-posed problems. *Inverse Problems*, 25(11):(115018), 2009.

[19] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*, volume 375 of *Mathematics and Its Applications*. Kluwer Academic Publishers, 1996.

[20] S. George and M. T. Nair. An optimal order yielding discrepancy for simplified regularization of ill-posed problems in Hilbert scales. *Int. J. Math. and Math. Sci.*, 2003(39):2487–2499, 2003.

[21] S. George and M. T. Nair. An optimal order yielding discrepancy principle for simplified regularization in Hilbert scales: finite dimensional realizations. *Int. J. Math. and Math. Sci.*, 2004(37):1973–1996, 2004.

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

[23] V. Glasko and Y. Kriksin. On the quasioptimality principle for linear ill-posed problems in Hilbert space. *Zh Vychisl. Mat. Mat. Fiz.*, 24:1603–1613, 1984. In Russian.

[24] G. H. Golub and U. von Matt. Generalized cross-validation for large-scale problems. *J. Comput. Graph. Statist.*, 6(1):1–34, 1997.

[25] C. W. Groetsch and J. T. King. Extrapolation and the method of regularization for generalized inverses. *J. Approx. Theory*, 25(3):233–247, 1979.

[26] U. Hämarik. Monotonicity of error and choice of the stopping index in iterative regularization methods. In A. Pedas, editor, *Differential and Integral Equations: Theory and Numerical Analysis*, pages 15–30, Tartu, 1999. Estonian Mathematical Society.

[27] U. Hämarik and R. Palm. On rules for stopping the conjugate gradient type methods in ill-posed problems. *Mathematical Modelling and Analysis*, 12(1):61–70, 2007.

[28] U. Hämarik, R. Palm, and T. Raus. Use of extrapolation in regularization methods. *J. Inverse and Ill-Posed Problems*, 15(3):277–294, 2007.

[29] U. Hämarik, R. Palm, and T. Raus. Extrapolation of Tikhonov and Lavrentiev regularization methods. *Journal of Physics: Conference Series*, 135(012048):8 pp, 2008.

[30] U. Hämarik, R. Palm, and T. Raus. Extrapolation of Tikhonov and Lavrentiev regularization methods. In *ICIPE2008, International Conference on Inverse Problems in Engineering: Theory and Practice. Dourdan (Paris), France, June 15–19, 2008*, pages 51.1–51.8, 2008.

[31] U. Hämarik, R. Palm, and T. Raus. On minimization strategies for choice of the regularization parameter in ill-posed problems. *Numerical Functional Analysis and Optimization*, 30(9&10):924–950, 2009.

[32] U. Hämarik and T. Raus. Choice of the regularization parameter in ill-posed problems with rough estimate of the noise level of data. *WSEAS Transactions on Mathematics*, 4(2):76–81, 2005.

[33] U. Hämarik and T. Raus. On the choice of the regularization parameter in ill-posed problems with approximately given noise level of data. *J. Inv. and Ill-Posed Problems*, 14(3):251–266, 2006.

[34] U. Hämarik and T. Raus. About the balancing principle for choice of the regularization parameter. *Numerical Functional Analysis and Optimization*, 30(9&10):951–970, 2009.

[35] U. Hämarik, T. Raus, and R. Palm. On the analog of the monotone error rule for parameter choice in the (iterated) Lavrentiev regularization. *Computational Methods in Applied Mathematics*, 8(3):237–252, 2008.

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

[37] M. Hanke. *Conjugate Gradient Type Methods for Ill-Posed Problems*. Longman House, Harlow, 1995.

[38] M. Hanke. The minimal error conjugate gradient method is a regularization method. *Proc. Amer. Math. Soc.*, 123(11):3487–3497, 1995.

[39] M. Hanke. Limitations of the l-curve method in ill-posed problems. *BIT*, 36(2):287–301, 1996.

[40] M. Hanke and C. W. Groetsch. Nonstationary iterated Tikhonov regularization. *J. Optim. Theory Appl.*, 98(1):37–53, 1998.

[41] M. Hanke and P. C. Hansen. Regularization methods for large-scale problems. *Surveys on Mathematics for Industry*, 3:253–315, 1993.

[42] M. Hanke and T. Raus. A general heuristic for choosing the regularization parameter in ill-posed problems. *SIAM Journal on Scientific Computing*, 17(4):956–972, 1996.

[43] P. C. Hansen. Regularization tools version 4.1 (for Matlab version 7.3). `http://www2.imm.dtu.dk/~pch/Regutools/`, 2010-01-13.

[44] P. C. Hansen. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Rev.*, 34(4):561–580, 1992.

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

[46] P. C. Hansen. Test matrices for regularization methods. *SIAM J. on Scientific Computing*, 16:506–512, 1995.

[47] P. C. Hansen. *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*. SIAM, Philadelphia, 1998.

[48] J. Janno and U. Tautenhahn. On Lavrentiev regularization for ill-posed problems in Hilbert scales. *Numer. Funct. Anal. Optim.*, 24(5-6):531–555, 2003.

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

[50] R. D. Lazarov, S. Lu, and S. V. Pereverzev. On the balancing principle for some problems of numerical analysis. *Numerische Mathematik*, 106:659–689, 2007.

[51] A. S. Leonov. Justification of the choice of regularization parameter according to quasi-optimality and quotient criteria. *USSR Comput. Math. Math. Phys.*, 18:1–15, 1979.

[52] A. S. Leonov. On the accuracy of Tikhonov regularizing algorithms and quasioptimal selection of a regularization parameter. *Sov. Math. Dokl.*, 44:711–716, 1991.

[53] A. K. Louis. *Inverse und schlecht gestellte Probleme*. Teubner Studienbücher, 1989.

[54] S. Lu and S. V. Pereverzev. Numerical differentiation from a viewpoint of regularization theory. *Math. Comp.*, 75(256):1853–1870, 2006.

[55] S. Lu and S. V. Pereverzev. Sparse recovery by the standard Tikhonov method. *Numer. Math*, 112:403–424, 2009.

[56] S. Lu, S. V. Pereverzev, and R. Ramlau. An analysis of Tikhonov regularization for nonlinear ill-posed problems under a general smoothness assumption. *Inverse Problems*, 23(1):217–230, 2007.

[57] G. I. Marchuk and V. V. Shaidurov. *Difference methods and their extrapolations*. Springer-Verlag, New York, 1983.

[58] P. Mathé. The Lepskii principle revisited. *Inverse Problems*, 22(3):L11–L15, 2006.

[59] P. Mathé and S. V. Pereverzev. Discretization strategy for linear ill-posed problems in variable Hilbert scales. *Inverse Problems*, 19(6):1263–1277, 2003.

[60] P. Mathé and S. V. Pereverzev. Geometry of linear ill-posed problems in variable Hilbert scales. *Inverse Problems*, 19(3):789–803, 2003.

[61] P. Mathé and S. V. Pereverzev. The discretized discrepancy principle under general source conditions. *Journal of Complexity*, 22(3):371–381, 2006.

[62] P. Mathé and S. V. Pereverzev. Regularization of some linear ill-posed problems with discretized random noisy data. *Math. Comp.*, 75(256):1913–1929, 2006.

[63] P. Mathé and U. Tautenhahn. Interpolation in variable Hilbert scales with application to inverse problems. *Inverse Problems*, 22(6):2271–2297, 2006.

[64] S. Morigi, L. Reichel, and F. Sgallari. An iterative Lavrentiev regularization method. *BIT Numerical Mathematics*, 46:589–606, 2006.

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

[66] M. T. Nair, S. Pereverzev, and U. Tautenhahn. Regularization in Hilbert scales under general smoothing conditions. *Inverse Problems*, 21(6):1851–1869, 2005.

[67] M. T. Nair and U. Tautenhahn. Convergence rates for Lavrentiev-type regularization in Hilbert scales. *Computational Methods in Applied Mathematics*, 8(3):279–293, 2008.

[68] A. S. Nemirovski. The regularizing properties of the conjugate gradient method in ill-posed problems. *USSR Comp. Math. Math. Phys.*, 26(2):7–16, 1986.

[69] A. Neubauer. The convergence of a new heuristic parameter selection criterion for general regularization methods. *Inverse Problems*, 24:(055005), 2008.

[70] S. V. Pereverzev and E. Schock. On the adaptive selection of the parameter in the regularization of ill-posed problems. *SIAM J. Numerical Analysis*, 43(5):2060–2076, 2005.

[71] R. Plato. The conjugate gradient method for linear ill-posed problems with operator perturbations. *Numer. Algorithms*, 20(1):1–22, 1999.

[72] R. Plato and U. Hämarik. On pseudo-optimal parameter choices and stopping rules for regularization methods in Banach spaces. *Numer. Funct. Anal. Optim.*, 17:181–195, 1996.

[73] T. Raus. The principle of the residual in the solution of ill-posed problems. *Acta et comment. Univ. Tartuensis*, 672:16–26, 1984. In Russian.

[74] T. Raus. The principle of the residual in the solution of ill-posed problems with non-selfadjoint operator. *Acta et comment. Univ. Tartuensis*, 715:12–20, 1985. In Russian.

[75] T. Raus. An a posteriori choice of the regularization parameter in case of approximately given error bound of data. *Acta et comment. Univ. Tartuensis*, 913:73–87, 1990.

[76] T. Raus. About regularization parameter choice in case of approximately given error bounds of data. *Acta et comment. Univ. Tartuensis*, 937:77–89, 1992.

[77] T. Raus. On the a posteriori parameter choice for the determination of a quasisolution. In A. Pedas, editor, *Differential and Integral Equations: Theory and Numerical Analysis*, pages 81–96. Estonian Mathematical Society, Tartu, 1999.

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

[79] T. Raus and U. Hämarik. About the balancing principle for choice of the regularization parameter. *Journal of Physics: Conference Series*, 135:(012087), 2008.

[80] T. Raus and U. Hämarik. New rule for choice of the regularization parameter in (iterated) Tikhonov method. *Mathematical Modelling and Analysis*, 14(2):187–198, 2009.

[81] T. Raus and U. Hämarik. On numerical realization of quasioptimal parameter choices in (iterated) Tikhonov and Lavrentiev regularization. *Mathematical Modelling and Analysis*, 14(1):99–108, 2009.

[82] T. Regińska. A regularization parameter in discrete ill-posed problems. *SIAM J. Sci. Comput.*, 17(3):740–749, 1996.

[83] V. V. Saidurov. Continuation with respect to the parameter in the method of regularization. In *Numerical methods of linear algebra*, pages 77–85. Nauka, Novosibirsk, 1973. In Russian.

[84] U. Tautenhahn and U. Hämarik. The use of monotonicity for choosing the regularization parameter in ill-posed problems. *Inverse Problems*, 15(6):1487–1505, 1999.

[85] A. N. Tikhonov and V. Y. Arsenin. *Solution of Ill-Posed Problems*. John Wiley&Sons, New York-Toronto, 1977.

[86] A. N. Tikhonov, V. B. Glasko, and Y. Kriksin. On the question of quasioptimal choice of a regularized approximation. *Sov. Math. Dokl.*, 20:1036–40, 1979.

[87] G. Vainikko. *Methods for Solution of Linear Ill-Posed Problems in Hilbert Spaces*. University of Tartu, 1982. In Russian.

[88] G. M. Vainikko. The principle of the residual for a class of regularization methods. *USSR Comput. Math. Math. Phys.*, 22(3):1–19, 1982.

[89] G. M. Vainikko and A. Yu. Veretennikov. *Iteration Procedures in Ill-Posed Problems*. Nauka, Moscow, 1986. In Russian.

[90] C. R. Vogel. Non-convergence of the l-curve regularization parameter selection method. *Inverse problems*, 12(4):535–547, 1996.

[91] G. Wahba. Practical approximate solutions to linear operator equations when the data are noisy. *SIAM J. Numer. Anal.*, 14(4):651–657, 1977.

SISUKOKKUVÕTE

Mittekorrektsete ülesannete regulariseerimisalgoritmide numbriline võrdlus

Töö on pühendatud mittekorrektsete ülesannete regulariseerimismeetodite parameetrivalikureeglite numbrilisele analüüsile täpsuse seisukohalt.

Töö sisu peatükkide kaupa on järgmine.

Esimeses peatükis (sissejuhatuses) kirjeldatakse probleeme ja nendele töös pakutud lahendusi.

Teises peatükis antakse ülevaade mittekorrektsete ülesannete lahendusmeetoditest, mida töös vaadeldakse. Need meetodid on Tihhonovi meetod koos itereeritud kujuga, Lavrentjevi meetod, Landweberi meetod, spektraallõike meetod, kaasgradientide tüüpi meetodid CGLS ja CGME ning Tihhonovi ja Lavrentjevi meetodi ekstrapoleeritud variandid.

Kolmandas peatükis kirjeldatakse mitmesuguseid regulariseerimisparameetri valikureegleid nende meetodite jaoks, sealhulgas tuuakse välja reeglite need teoreetilised omadused, mis on olulised arvutuste tegemise juures. Ühtlasi pakutakse välja mitmed strateegiad uute reeglite konstrueerimiseks nii teadaoleva kui ka mitteteadaoleva veataseme korral.

Neljandas peatükis on esitatud arvutuseksperimentide tulemused eelmises peatükis formuleeritud reeglite ning eelmises peatükis esitatud strateegiate realiseeringutena saadud mitmesuguste uute reeglite kohta ülalpool nimetatud meetodites nii teadaoleva, mitteteadaoleva kui ka ligikaudselt teadaoleva veataseme korral. Tulemused näitatavad, et uued reeglid annavad paremaid tulemusi kui senised.

Viies peatükk sisaldab kokkuvõtet saadud tulemustest ja nendel põhinevaid soovitusi, millist regulariseerimismeetodi ja parameetrivalikureegli kombinatsiooni on otstarbekas kasutada.

ACKNOWLEDGEMENT

Of all people who have helped me, directly or indirectly, I would like to especially thank Uno Hämarik and Gennadi Vainikko.

This work was partially supported by the Estonian Science Foundation, grant no. 7489.

CURRICULUM VITAE

Reimo Palm

Citizenship:	Estonia
Born:	January 21, 1973, Viljandi
Address:	Institute of Computer Science, Liivi 2, 50904 Tartu
Contacts:	(+372) 375482, reimo.palm@ut.ee

Education

1991–1996	Faculty of Mathematics, University of Tartu, BSc and MSc in mathematics
1996–2001	Faculty of Mathematics, University of Tartu, PhD studies in computer science

Professional employment

2000–2002	Institute of Computer Science, University of Tartu, researcher
since 2003	Institute of Computer Science, University of Tartu, assistant

Scientific work

Main fields of interest:

- extrapolation of regularization methods for solving ill-posed problems;
- stopping rules in conjugate gradient type methods
- strategies for specifying the value of regularization parameters.

ELULOOKIRJELDUS

Reimo Palm

Kodakondsus:	Eesti
Sünniaeg ja -koht:	21. jaanuar 1973, Viljandi
Aadress:	Arvutiteaduse instituut, Liivi 2, 50904 Tartu
Kontaktandmed:	(+372) 375482, reimo.palm@ut.ee

Haridus

1991–1996	Tartu ülikooli matemaatikateaduskond, BSc ja MSc matemaatikas
1996–2001	Tartu ülikooli matemaatikateaduskond, doktoriõpe informaatikas

Erialane teenistuskäik

2000–2002	Tartu ülikooli arvutiteaduse instituudi teadur
alates 2003	Tartu ülikoooli arvutiteaduse instituudi assistent

Teadustegevus

Peamised uurimisvaldkonnad:

- mittekorrektsete ülesannete regulariseerimismeetodite ekstrapoleerimine;
- peatumisreegli valik kaasgradientide tüüpi iteratsioonimeetodites;
- parameetrivaliku täpsustamise strateegiad.

LIST OF ORIGINAL PUBLICATIONS

1. U. Hämarik and R. Palm. Comparison of stopping rules in conjugate gradient type methods for solving ill-posed problems. In *MMA2005 Proceedings: 10th International Conference "Mathematical Modelling and Analysis" & 2nd International Conference "Computational Methods in Applied Mathematics", Trakai, Lithuania, June 1–5, 2005*, pages 285–291, Trakai, Lithuania, 2005. Technika.
2. U. Hämarik and R. Palm. On rules for stopping the conjugate gradient type methods in ill-posed problems. *Mathematical Modelling and Analysis*, 12(1):61–70, 2007.
3. U. Hämarik, R. Palm, and T. Raus. Use of extrapolation in regularization methods. *J. Inverse and Ill-Posed Problems*, 15(3):277–294, 2007.
4. U. Hämarik and R. Palm. Comparison of stopping rules for the conjugate type iteration methods in ill-posed problems. In *Inverse and Ill-posed Problems of Mathematical Physics, Novosibirsk, August 20–25*, pages 1–3, 2007.
5. U. Hämarik, R. Palm, and T. Raus. Extrapolation of Tikhonov and Lavrentiev regularization methods. In *ICIPE2008, International Conference on Inverse Problems in Engineering: Theory and Practice. Dourdan (Paris), France, June 15–19, 2008*, number 51, pages 1–8, 2008.
6. U. Hämarik, T. Raus, and R. Palm. On the analog of the monotone error rule for parameter choice in the (iterated) Lavrentiev regularization. *Computational Methods in Applied Mathematics*, 8(3):237–252, 2008.
7. U. Hämarik, R. Palm, and T. Raus. Extrapolation of Tikhonov and Lavrentiev regularization methods. *Journal of Physics: Conference Series*, 135(012048):8 pp, 2008.
8. U. Hämarik, R. Palm, and T. Raus. On minimization strategies for choice of the regularization parameter in ill-posed problems. *Numerical Functional Analysis and Optimization*, 30(9&10):924–950, 2009.
9. U. Hämarik, R. Palm, and T. Raus. Extrapolation of Tikhonov regularization method. *Mathematical Modelling and Analysis*, 15(1):(14 pp), 2010. Accepted.
10. U. Hämarik, R. Palm, and T. Raus. On parameter choice in regularization algorithms in case of different noise information *Calcolo*, 15 pp, 2010. Submitted.

The author's contribution in all papers [1–10] is their numerics part (including experiment design, solving stability issues, programming, and collection of results), participation in formulating and testing of new parameter choice strategies (in collaboration with U. Hämarik), and preparation of manuscripts.

DISSERTATIONES MATHEMATICAE UNIVERSITATIS TARTUENSIS

1. Mati Heinloo. The design of nonhomogeneous spherical vessels, cylindrical tubes and circular discs. Tartu 1991, 23 p.
2. Boris Komrakov. Primitive actions and the Sophus Lie problem. Tartu 1991, 14 p.
3. Jaak Heinloo. Phenomenological (continuum) theory of turbulence. Tartu 1992, 47 p.
4. Ants Tauts. Infinite formulae in intuitionistic logic of higher order. Tartu 1992, 15 p.
5. Tarmo Soomere. Kinetic theory of Rossby waves. Tartu 1992, 32 p.
6. Jüri Majak. Optimization of plastic axisymmetric plates and shells in the case of Von Mises yield condition. Tartu 1992, 32 p.
7. Ants Aasma. Matrix transformations of summability and absolute summability fields of matrix methods. Tartu 1993, 17 p.
8. Helle Hein. Optimization of plastic axisymmetric plates and shells with piece-wise constant thickness. Tartu 1993, 29 p.
9. Toomas Kiho. Study of optimality of iterated Lavrentiev method and its generalizations. Tartu 1994, 24 p.
10. Arne Kokk. Joint spectral theory and extension of non-trivial multiplicative linear functionals. Tartu 1995, 166 p.
11. Тоомас Лепикулт. Автоматизированный расчет задач динамики жесткопластических конструкций. Тарту 1995, 88 с.
12. Сандер Ханнус. Параметрическая оптимизация пластических цилиндрических оболочек с учетом геометрической и физической нелинейности. Тарту 1995, 75 с.
13. Sergei Tupailo. Hibert's epsilon-symbol in predicative subsystems of analysis. Tartu 1996, 132 p.
14. Enno Saks. Analysis and optimization of elastic-plastic shafts in torsion. Tartu 1996, 94 p.
15. Valdis Laan. Pullbacks and flatness properties of acts. Tartu 1999, 88 p.
16. Märt Põldvere. Subspaces of Banach spaces having Phelps' uniqueness property. Tartu 1999, 72 p.

17. Jelena Ausekle. Compactness of operators in Lorentz and Orlicz sequence spaces. Tartu 1999, 70 p.
18. Krista Fischer. Structural mean models for analyzing the effect of compliance in clinical trials. Tartu 1999, 124 p.
19. Helger Lipmaa. Secure and efficient time-stamping systems. Tartu 1999, 56 p.
20. Jüri Lember. Consistency of empirical k-centres. Tartu 1999, 146 p.
21. Ella Puman. Optimization of plastic conical shells. Tartu 2000, 102 p.
22. Kaili Müürisep. Eesti keele arvutigrammatika: süntaks. Tartu 2000, 107 lk.
23. Varmo Vene. Categorical programming with inductive and conductive types. Tartu 2000, 116 p.
24. Olga Sokratova. Ω-rings, their flat and projective acts with some applications. Tartu 2000, 121 p.
25. Maria Zeltser. Investigation of double sequence spaces by soft and hard analytical methods. Tartu 2001, 153 p.
26. Ernst Tungel. Optimization of plastic spherical shells. Tartu 2001, 90 p.
27. Tiina Puolakainen. Eesti keele arvutigrammatika: morfoloogiline ühestamine. Tartu 2001, 137 p.
28. Rainis Haller. $M(r,s)$-inequalities. Tartu 2002, 77 p.
29. Jan Villemson. Size-efficient interval time stamps. Tartu 2002, 81 p.
30. Eno Tõnisson. Solving of expression manipulation exercises in computer algebra systems. Tartu 2002, 91 p.
31. Mart Abel. Structure of Gelfand-Mazur algebras. Tartu 2003, 92 p.
32. Vladimir Kuchmei. Affine completeness of some Ockham algebras. Tartu 2003, 99 p.
33. Olga Dunajeva. Asymptotic matrix methods in statistical inference problems. Tartu 2003, 77 p.
34. Mare Tarang. Stability of the spline collocation method for Volterra integro-differential equations. Tartu 2004, 88 p.
35. Tatjana Nahtman. Permutation invariance and reparameterizations in linear models. Tartu 2004, 89 p.
36. Märt Möls. Linear mixed models with equivalent predictors. Tartu 2004, 69 p.
37. Kristiina Hakk. Approximation methods for weakly singular integral equations with discontinuous coefficients. Tartu 2004, 136 p.
38. Meelis Käärik. Fitting Sets to Probability Distributions. Tartu 2005, 89 p.
39. Inga Parts. Piecewise polynomial collocation methods for solving weakly singular integro-differential equations. Tartu 2005, 140 p.

40. Natalia Saealle. Convergence and summability with speed of functional series. Tartu 2005, 90 p.
41. Tanel Kaart. The reliability of linear mixed models in genetic studies. Tartu 2006, 123 p.
42. Kadre Torn. Shear and bending response of inelastic structures to dynamic loads. Tartu 2006, 141 p.
43. Kristel Mikkor. Uniform factorization for compact subsets of Banach spaces of operators. Tartu 2006, 72 p.
44. Darja Saveljeva. Quadratic and cubic spline collocation for Volterra integral equations. Tartu 2006, 116 p.
45. Kristo Heero. Path Planning and Learning Strategies for Mobile Robots in Dynamic Partially Unknown Environments. Tartu 2006, 122 p.
46. Annely Mürk. Optimization of inelastic plates with cracks. Tartu 2006, 136 p.
47. Annemai Raidjõe. Sequence spaces defined by modulus functions and superposition operators. Tartu 2006, 97 p.
48. Olga Panova. Real Gelfand-Mazur algebras. Tartu 2006, 81 p.
49. Härmel Nestra. Iteratively defined transfinite trace semantics and program slicing with respect to them. Tartu 2006, 118 p.
50. Margus Pihlak. Approximation of multivariate distribution functions. Tartu 2007, 81 p.
51. Ene Käärik. Handling dropouts in repeated measurements using copulas. Tartu 2007, 98 p.
52. Artur Sepp. Affine models in mathematical finance: an analytical approach. Tartu 2007, 148 p.
53. Marina Issakova. Solving of linear equations, linear inequalities and systems of linear equations in interactive learning environment. Tartu 2007, 169 p.
54. Kaja Sõstra. Restriction estimator for domains. Tartu 2007, 103 p.
55. Kaarel Kaljurand. Attempto Controlled English as a Semantic Web Language. Tartu 2008, 161 p.
56. Mart Anton. Mechanical modeling of IPMC actuators at large deformations. Tartu 2008, 123 p.
57. Evely Leetma. Solution of smoothing problems with obstacles. Tartu 2009, 81 p.
58. Ants Kaasik. Estimating ruin probabilities in the Cramér-Lundberg model with heavy-tailed claims. Tartu 2009, 136 p.