



Monotonicity of error of regularized solution and its use for parameter choice

Uno Hämarik, Urve Kangro, Reimo Palm, Toomas Raus & Ulrich Tautenhahn

To cite this article: Uno Hämarik, Urve Kangro, Reimo Palm, Toomas Raus & Ulrich Tautenhahn (2014) Monotonicity of error of regularized solution and its use for parameter choice, Inverse Problems in Science and Engineering, 22:1, 10-30, DOI: [10.1080/17415977.2013.827185](https://doi.org/10.1080/17415977.2013.827185)

To link to this article: <https://doi.org/10.1080/17415977.2013.827185>



Published online: 20 Aug 2013.



Submit your article to this journal [↗](#)



Article views: 94



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)

Monotonicity of error of regularized solution and its use for parameter choice

Uno Hämarik^{a*}, Urve Kangro^a, Reimo Palm^a, Toomas Raus^a and Ulrich Tautenhahn^b

^aFaculty of Mathematics and Computer Science, University of Tartu, Liivi 2, 50409 Tartu, Estonia;

^bDepartment of Mathematic, University of Applied Sciences Zittau/Görlitz, P. O. Box 1454, 02754 Zittau, Germany

(Received 10 July 2013; accepted 13 July 2013)

We consider an ill-posed equation in a Hilbert space with a noisy operator and a noisy right-hand side. The noise level information is given in a general form, as a norm of a certain operator applied to the noise. We derive the monotone error rule (ME-rule) for the choice of the regularization parameter in many methods, giving parameter such that the error is monotonically increasing for larger parameters in the Tikhonov method and for smaller stopping indices in iteration methods. Regularization methods considered include Y -scale regularization in (iterated) Tikhonov method and in iteration methods (Landweber method, CG type methods, semi-iterative methods). We also consider modifications of the ME-rule and show in numerical experiments (test problems from Hansen's Regularization Toolbox, including the sideways heat equation) their advantages over the discrepancy principle.

Keywords: ill-posed problems; regularization methods; iteration methods; parameter choice; monotone error

AMS Subject Classifications: 47A52; 65J20; 65J10

1. Introduction

In this paper, we consider linear ill-posed problems

$$A_0x = y_0, \quad y_0 \in \mathcal{R}(A), \quad (1)$$

where A_0 is a bounded linear operator with non-closed range $\mathcal{R}(A)$ and X, Y are infinite-dimensional real Hilbert spaces with inner products (\cdot, \cdot) and norms $\|\cdot\|$. We are interested in the minimum-norm solution x^\dagger of problem (1) and assume that instead of *exact* data y_0 and A_0 , there are given *noisy* data $y \in Y$ and $A \in \mathcal{L}(X, Y)$ with

$$\|y_0 - y\| \leq \delta, \quad \|A_0 - A\| \leq h \quad (2)$$

and known noise levels δ, h . Later, we also consider the case of noise level information in more general form (4).

*Corresponding author. Email: uno.hamarik@ut.ee

Ill-posed problems (1) arise in a wide variety of problems in applied sciences. For their stable numerical solution, regularization methods are necessary, see [1,2]. Regularization methods include Tikhonov regularization $x_r = (A^*A + r^{-1}I)^{-1}A^*y$ with regularization parameter $r \in \mathbb{R}_+$ and iterative and projection methods, where the stopping index $r = n \in \mathbb{N}$ is the regularization parameter. Here, $A^* \in \mathcal{L}(Y, X)$ is the adjoint operator to $A \in \mathcal{L}(X, Y)$. Traditional regularization methods possess the property that in the case of exact data, the error $\|x_r^0 - x^\dagger\|$ of regularized solution x_r^0 , as a function of r , is monotonically decreasing for $r \rightarrow \infty$. This property is no longer true for the error $\|x_r - x^\dagger\|$ in the case of noisy data. The monotone decrease of the error $\|x_r - x^\dagger\|$ for growing r -values can only be guaranteed for small r . Typically, $\|x_r - x^\dagger\|$ diverges for $r \rightarrow \infty$. Therefore, a rule for the proper choice of the regularization parameter r is necessary.

In the monotone error rule (ME-rule) for choosing a proper regularization parameter, the idea consists in *searching for the largest computable regularization parameter $r = r_{ME}$ for which we can guarantee the ME-property: the error $\|x_r - x^\dagger\|$ is monotonically decreasing for $r \in (0, r_{ME}]$* . For the continuous regularization methods, this means that

$$\frac{d}{dr} \|x_r - x^\dagger\|^2 \leq 0 \quad \forall r \in (0, r_{ME}],$$

for the iteration methods and other methods with $r = n \in \mathbb{N}$ this means that

$$\|x_n - x^\dagger\| \leq \|x_{n-1} - x^\dagger\| \quad \forall n = 1, 2, \dots, n_{ME}. \quad (3)$$

From derivation of the ME-rule for concrete regularization methods, one can see for which perturbations of the operator and the right-hand side, the ME-rule gives the optimal regularization parameter.

Consider now the case of the exact operator. In theoretical works on ill-posed problems, often the worst-case error $\sup\{\|x_r - x^\dagger\| : y \in Y, Ax^\dagger = y_0, \|y - y_0\| \leq \delta\}$ is considered. Here, the exact data y_0 are fixed, supremum is over noisy data $y \in Y$. In applications, the exact data y_0 are unknown and noisy data y are known. Then, we are interested in finding the parameter r which minimizes the analogue of the worst-case error $\sup\{\|x_r - x\| : \tilde{y} \in Y, Ax = \tilde{y}, \|\tilde{y} - y\| \leq \delta\}$, where given data y are fixed, supremum is over “candidates of exact data” $\tilde{y} \in Y$. This analogue of the worst-case error is minimized by the parameter r_{ME} . This can be seen from derivation of the ME-rule for concrete regularization methods: for $r \in (0, r_{ME}]$, the error $\|x_r - x\|$ is decreasing for all “candidates of exact data” $\tilde{y} \in Y$ with $Ax = \tilde{y}$, but for $r > r_{ME}$ there exists $\tilde{y} \in Y$ with $Ax = \tilde{y}$ where $\|x_r - x\| > \|x_{r_{ME}} - x\|$.

All regularization methods considered in this paper have the property $x_r \in \mathcal{R}(A^*)$ because we need this property for the formulation of the ME rule.

The ME-rule was proposed and studied for continuous methods in [3–8], for iterative methods in [4,7,8], analogous rules were proposed in [9,10]. In this paper, we extend these results in the following directions: we consider the case of noisy operator ($h > 0$), the noise level information may be given in general form (4), we derive the ME-rule for a more general class of regularization methods (including the regularization in Y -scale,[11] also for iterative and semi-iterative methods). We also give convergence results and error estimates for the ME-rule.

In practical applications, often different regularization parameters are selected and the corresponding regularized solutions are studied online. With our ME-rule, we provide some help for the selection of suitable values of r , since we can guarantee that for $r < r_{ME}$ always $\|x_{r_{ME}} - x^\dagger\| < \|x_r - x^\dagger\|$ holds. This information may also be used for improving other

a posteriori rules R for the choice of the regularization parameter, since the error for the parameter $\max(r_R, r_{ME})$ is less than or equal to the error in rule R . Unfortunately, this observation does not help to improve the discrepancy principle, which gives the parameter $r_D \geq r_{ME}$, and for a smooth solution r_D is often too large in regularization methods with the finite qualification. In (iterated) Tikhonov method, the ME-rule is in contrast to the discrepancy principle a quasi-optimal parameter choice.

Note that typically the error of the regularized solution decreases monotonically also, somewhat further up to some $r_{opt} \geq r_{ME}$. Our numerical experiments suggest to use $r_{MEc} = cr_{ME}$ (or its integer part) with a certain $c > 1$.

The plan of this paper is as follows. In Sections 2 and 3, well-known regularization methods and parameter choice rules are introduced, using noise-level information (4). In Sections 4 and 5, the ME-rule is derived for the continuous regularization methods and for the iterative methods, respectively. In Section 6 convergence conditions and quasioptimality results of the ME-rule are given. The paper is finished by extensive numerical experiments.

2. Well-known regularization methods in Y -scale

We assume that the noise level information is given in the form

$$\|D(y_0 - y)\| \leq \delta, \quad \|D(A_0 - A)\| \leq h. \quad (4)$$

where D is a linear injective, possibly unbounded operator in Y with domain $\mathcal{D}(D)$. We assume that $y_0, y \in \mathcal{D}(D)$. The standard case (2) is special case with $D = I$, where I is the identity operator. In [12], the operator $D = L^{-t}$ with $t \in \mathbb{R}$ was used. Here, the operator L is a densely defined, unbounded, self-adjoint and strictly positive operator in the space Y , inducing Hilbert scale Y_s with norm $\|y\|_s := \|L^s y\|$. In the standard case (2), $t = 0$. The case $t < 0$ corresponds to large noise, the case $t > 0$ to small noise. The noise-level information (4) with different operators D (only the case $h = 0$ was considered) was used in works.[13–16]

In many papers, X_s -scale is considered with the generating operator L in the space X , the equation $Ax = y$ is transformed to the equation $L^{-2s} A^* Ax = L^{-2s} A^* y$ and the regularized solutions are constructed in the form $x_r = g_r(L^{-2s} A^* A)L^{-2s} A^* y$. Here, r is the regularization parameter and the function $g_r(\lambda)$ satisfies the conditions

$$\sup_{0 \leq \lambda \leq \Lambda} \left| \sqrt{\lambda} g_r(\lambda) \right| \leq \gamma_* \sqrt{r}, \quad r \geq 0, \quad (5)$$

$$\sup_{0 \leq \lambda \leq \Lambda} \lambda^p |1 - \lambda g_r(\lambda)| \leq \gamma_p r^{-p}, \quad r \geq 0, \quad 0 \leq p \leq p_0. \quad (6)$$

Here, p_0 , γ_* and γ_p are positive constants, Λ is at least the norm of the operator in the argument of the function g_r , $\gamma_0 = 1$ and the greatest value of p_0 , for which the inequality (6) holds is called the qualification of the method. We can formulate the ME rule for choice of parameter r only for regularized solutions $x_r \in \mathcal{R}(A^*)$. Therefore, we prefer to use the Y -scale instead of X scale. The Y -scale regularization was proposed by Egger [11,12]. Here, the equation $Ax = y$ is transformed into the equation

$$Bx = \bar{y}, \quad B = L^{-s} A, \quad \bar{y} = L^{-s} y,$$

assuming $B : X \rightarrow Y$ is bounded and $\bar{y} \in Y$. If $s \geq 0$ this is always satisfied; if $s < 0$, this means that $\mathcal{R}(A) \in Y_{-s}$ and $y \in Y_{-s}$. The regularized solutions are constructed in the form

$$x_r = g_r(B^*B)B^*\bar{y} = B^*g_r(BB^*)\bar{y} = A^*w_r, \quad w_r = L^{-s}g_r(BB^*)\bar{y}. \quad (7)$$

Special cases of regularization methods of this form are the following well-known methods (cf. [1,2,17]).

(M1) The Tikhonov method $x_\alpha = (\alpha I + B^*B)^{-1}B^*\bar{y}$. Here, $r = \alpha^{-1}$, $g_r(\lambda) = (\lambda + r^{-1})^{-1}$, $p_0 = 1$, $\gamma_* = 1/2$, $\gamma_p = p^p(1-p)^{1-p}$.

(M2) Iterative variants of the Tikhonov method. Let $m \in \mathbb{N}$, $m \geq 1$, $x_{0,\alpha} \in X$ – initial approximation and

$$x_{n,\alpha} = (\alpha I + B^*B)^{-1}(\alpha x_{n-1,\alpha} + B^*\bar{y}) \quad (n = 1, 2, \dots, m).$$

$$\text{Here } r = \alpha^{-1}, \quad g_r(\lambda) = \frac{1}{\lambda} \left(1 - \left(\frac{1}{1+r\lambda} \right)^m \right), \quad p_0 = m, \quad \gamma_* = \sqrt{m}, \quad \gamma_p = \left(\frac{p}{m} \right)^p \left(1 - \frac{p}{m} \right)^{m-p}.$$

(M3) Explicit iteration scheme (Landweber's method). Let $x_0 = 0$,

$$x_n = x_{n-1} - \mu(B^*Bx_{n-1} - \bar{y}), \quad \mu \in (0, 1/\|B^*B\|), \quad n = 1, 2, \dots$$

$$\text{Here, } r = n, \quad g_r(\lambda) = \frac{1}{\lambda} \left(1 - (1 - \mu\lambda)^r \right), \quad p_0 = \infty, \quad \gamma_* = \sqrt{\mu}, \quad \gamma_p = \left(\frac{p}{\mu e} \right)^p.$$

(M4) Implicit iteration scheme. Let $\alpha > 0$ be a constant, $x_0 = 0$ and

$$x_n = (\alpha I + B^*B)^{-1}(\alpha x_{n-1} + B^*\bar{y}), \quad n = 1, 2, \dots$$

$$\text{Here, } r = n, \quad g_r(\lambda) = \frac{1}{\lambda} \left(1 - \left(\frac{\alpha}{\alpha + \lambda} \right)^r \right), \quad p_0 = \infty, \quad \gamma_* = \frac{b_0}{\sqrt{\alpha}}, \quad \text{where } b_0 = \sup_{0 < \lambda < \infty} \lambda^{-1/2} (1 - e^{-\lambda}) \approx 0.6382 \text{ and } \gamma_p = (\alpha p)^p.$$

(M5) The method of asymptotical regularization: approximation x_r solves the Cauchy problem

$$x'(r) + B^*Bx(r) = B^*\bar{y}, \quad x(0) = 0.$$

$$\text{Here, } g_r(\lambda) = \frac{1}{\lambda} (1 - e^{-r\lambda}), \quad p_0 = \infty, \quad \gamma_* = b_0, \quad \gamma_p = (p/e)^p.$$

In the iterated Tikhonov method, we may use different parameter α_n at every iteration step n ; then, we get the same approximation as in the extrapolated Tikhonov method [18,19]

$$x_{\alpha_1, \dots, \alpha_m} = \sum_{i=1}^m c_i x_{\alpha_i}, \quad c_i = \prod_{\substack{j=1 \\ j \neq i}}^m \frac{\alpha_j}{\alpha_j - \alpha_i}, \quad (8)$$

where x_{α_i} , $i = 1, \dots, m$ are the approximations of Tikhonov method with the parameters α_i .

Note that regularized solutions in X_s -scales and Y_s -scales variants of Tikhonov method have the form

$$x_\alpha = (\alpha I + L^{-2s} A^* A)^{-1} L^{-2s} A^* y, \quad x_\alpha = (\alpha I + A^* L^{-2s} A)^{-1} A^* L^{-2s} y, \quad (9)$$

and they minimize the corresponding functionals

$$\|Ax - y\|^2 + \alpha \|L^s x\|^2, \quad \|L^{-s}(Ax - y)\|^2 + \alpha \|x\|^2.$$

If the operator L acts in both spaces X and Y and the operators A^* and L commute, then these regularized approximations coincide. Much more widespread of these two approximations is the first one (cf.[20]). For the operators L and L^{-1} , one may use, for example, the differentiation and integration operators. The fractional powers L^s and L^{-s} can be implemented efficiently, e.g. via FFT or multi-level techniques.

Positive values of s are good for delaying saturation in Tikhonov and iterated Tikhonov methods,[12] negative values of s are good for preconditioning in the iteration methods that drastically decreases the number of iteration steps.[11]

3. Well-known rules for choice of the regularization parameter

For the choice of the regularization parameter in any regularization method under noise information (4), some estimate or approximation of value $\|D(y - Ax^\dagger)\|$ is needed. This value may be estimated as follows:

$$\|D(y - Ax^\dagger)\| = \|D(y - y_0 + (A_0 - A)x^\dagger)\| \leq \delta + h\|x^\dagger\| =: \Delta_1. \quad (10)$$

Typically, the exact value of $\|x^\dagger\|$ is unknown. Substitution of $\|x^\dagger\|$ by some upper bound $M \geq \|x^\dagger\|$ gives a rough estimate $\|D(y - Ax^\dagger)\| \leq \delta + hM =: \Delta_2$. If an upper bound M is not available, approximation of $\|D(y - Ax^\dagger)\|$ by $\Delta_3 := \delta + h\|x_\alpha\|$ (in iteration methods $\Delta_3 := \delta + h\|x_n\|$) may be used.

In the following, we formulate some *a posteriori* rules for choosing the regularization parameter which use $\Delta \in \{\Delta_1, \Delta_2, \Delta_3\}$ and are well known in the case $D = I, s = 0$. We assume $\mathcal{R}(AA^*) \subset \mathcal{D}(D)$.

- (1) *Discrepancy principle*. [1,2,17,21] In the continuous regularization methods, this principle (*D principle*) chooses the parameter $\alpha = \alpha_D$ as the solution of the equation

$$d_D(\alpha) := \|D(y - Ax_\alpha)\| = C\Delta \quad \text{with } C \geq 1. \quad (11)$$

In iteration methods, the discrepancy principle finds n_D as the first index for which $\|D(y - Ax_n)\| \leq C\Delta$.

- (2) *Modified discrepancy principle*. [22] In this rule for the methods M1, M2, the parameter $\alpha = \alpha_{MD}$ is chosen as the solution of the equation

$$d_{MD}(\alpha) := \left\| DL^s (I - g_\alpha(BB^*)BB^*)^{1/(2p_0)} L^{-s} (y - Ax_\alpha) \right\| = C\Delta \quad (12)$$

with $C \geq 1$, assuming $Y_{-s} \subset \mathcal{D}(D)$ (if $D = L^{-t}$ then a sufficient condition is $s \geq t$). Here, p_0 is the qualification of the regularization method, see (6). One may

use also the alternative MD' rule

$$d_{\text{MD}' }(\alpha) = (D(y - Ax_\alpha), DL^s(I - g_\alpha(BB^*)BB^*)^{1/p_0}L^{-s}(y - Ax_\alpha))^{1/2} = C\Delta. \quad (13)$$

In case $D = L^{-s}$, the functions $d_{\text{MD}}(\alpha)$ and $d_{\text{MD}' }(\alpha)$ coincide. For regularization methods with $p_0 = \infty$, both these forms of the modified discrepancy principle coincide with the discrepancy principle.

4. ME-rule for the continuous regularization methods

4.1. Derivation of the ME-rule

Let us consider the ME-rule for the continuous regularization methods. The prominent example of these methods is Tikhonov method where the regularization parameter is traditionally denoted by α . Therefore, in this section we use notation $\alpha = 1/r$ instead of r . We assume that the corresponding function $g_\alpha(\lambda)$ is differentiable with respect to α .

Let us consider the regularized approximation

$$x_\alpha = A^*w_\alpha, \quad z_\alpha := \frac{d}{d\alpha}w_\alpha \in Y, \quad (14)$$

assuming that $z_\alpha \in \mathcal{D}((D^{-1})^*)$. An example of a regularized approximation of this form is the approximation $x_\alpha = A^*w_\alpha$ with $w_\alpha = L^{-s}g_r(BB^*)\bar{y}$ from (7). The reformulation of the general idea of ME-rule in terms of $\alpha = 1/r$ means: *search in the approximation x_α for the smallest computable regularization parameter $\alpha = \alpha_{\text{ME}}$ for which we can guarantee that*

$$\frac{d}{d\alpha}\|x_\alpha - x^\dagger\|^2 \geq 0 \quad \text{for all } \alpha \in [\alpha_{\text{ME}}, \infty). \quad (15)$$

In order to guarantee this property, we estimate the derivative of the squared error $\|x_\alpha - x^\dagger\|^2$ with respect to α under condition (10) as follows:

$$\begin{aligned} \frac{1}{2} \frac{d}{d\alpha} \|x_\alpha - x^\dagger\|^2 &= (x_\alpha - x^\dagger, A^*z_\alpha) = (D(Ax_\alpha - y + y - Ax^\dagger), (D^{-1})^*z_\alpha) \\ &\geq \|(D^{-1})^*z_\alpha\| \left(\frac{(Ax_\alpha - y, z_\alpha)}{\|(D^{-1})^*z_\alpha\|} - (\delta + h\|x^\dagger\|) \right). \end{aligned} \quad (16)$$

This estimate leads us to following ME-rule for the continuous regularization method (14).

ME rule. Choose $\alpha = \alpha_{\text{ME}}$ as the largest solution of the equation

$$d_{\text{ME}}(\alpha) := \frac{(Ax_\alpha - y, z_\alpha)}{\|(D^{-1})^*z_\alpha\|} = \Delta \quad (17)$$

with $\Delta \in \{\Delta_1, \Delta_2\}$. If $d_{\text{ME}}(\alpha) > \Delta$ for all $\alpha > 0$ take $\alpha_{\text{ME}} = 0$. If $d_{\text{ME}}(\alpha) < \Delta$ for all $\alpha > 0$ take $x_\alpha = 0$ (it corresponds to $\alpha = \alpha_{\text{ME}} = \infty$). For regularized approximation (7), this equation has the form

$$d_{\text{ME}}(\alpha) := \frac{(Ax_\alpha - y, L^{-s} \frac{d}{d\alpha} g_\alpha(BB^*)L^{-s}y)}{\|(D^{-1})^*L^{-s} \frac{d}{d\alpha} g_\alpha(BB^*)L^{-s}y\|} = \Delta. \quad (18)$$

Here, the assumption $z_\alpha \in \mathcal{D}((D^{-1})^*)$ is satisfied if $Y_s \subset \mathcal{D}((D^{-1})^*)$.

Note that for $\Delta = \Delta_3$, the ME-property is not guaranteed but quasi-optimality retains (see Theorem 6.3). However, $\Delta = \Delta_3$ does not need the norm of the exact solution or a bound of it and in practice may give better results than $\Delta \in \{\Delta_1, \Delta_2\}$.

From definition of $d_{\text{ME}}(\alpha)$ follows the inequality $d_{\text{ME}}(\alpha) \leq \|D(y - Ax_\alpha)\|$, therefore $\alpha_D \leq \alpha_{\text{ME}}$ if $C = 1$ in the discrepancy principle. The derivation of the ME-rule in (16) uses only one inequality, which turns to the equality in the case

$$y - Ax^\dagger = -\frac{\delta + h\|x^\dagger\|}{\|(D^{-1})^*z_\alpha\|}(D^*D)^{-1}z_\alpha.$$

Therefore, in this case the ME-rule gives the optimal parameter, if the equation (18) has a unique solution.

4.2. ME-rule and modifications for the (iterated) Tikhonov regularization

Let $x_0 = 0$. Let $\rho_{m,\alpha}$ denote the discrepancy of the (iterated) Tikhonov approximation $x_\alpha = x_{m,\alpha}$, i.e. $\rho_{m,\alpha} := y - Ax_{m,\alpha}$.

Using the identities

$$1 - \lambda g_\alpha(\lambda) = \left(\frac{\alpha}{\lambda + \alpha}\right)^m \quad \text{and} \quad \frac{d}{d\alpha} g_\alpha(\lambda) = -\frac{m}{\alpha^2} \left(\frac{\alpha}{\lambda + \alpha}\right)^{m+1}$$

we obtain

$$L^{-s} \rho_{m,\alpha} = \bar{y} - Bx_{m,\alpha} = [I - BB^*g_\alpha(BB^*)]\bar{y} = [\alpha(BB^* + \alpha I)^{-1}]^m \bar{y}$$

and

$$\frac{d}{d\alpha} g_\alpha(BB^*)\bar{y} = -\frac{m}{\alpha^2} L^{-s} \rho_{m+1,\alpha}.$$

From these representations and from (18), we conclude that the functions $d_{\text{MD}'}(\alpha)$ in (13) and $d_{\text{ME}}(\alpha)$ for the MD'- and ME-rules have the form

$$d_{\text{MD}'}(\alpha) = (D\rho_{m,\alpha}, D\rho_{m+1,\alpha})^{1/2} \quad \text{and} \quad d_{\text{ME}}(\alpha) = \frac{(\rho_{m,\alpha}, L^{-2s} \rho_{m+1,\alpha})}{\|(D^{-1})^* L^{-2s} \rho_{m+1,\alpha}\|}. \quad (19)$$

In the case $D = I$, $h = s = 0$, convergence and order optimal error estimates for (iterated) Tikhonov method are proved in the case of MD'-rule in [5,22,23] and in the case of ME-rule in [3,5,6,8,24].

One may also use an analog of the ME-rule with the function

$$d_{\text{MEa}}(\alpha) = \frac{(D\rho_{m,\alpha}, D\rho_{m+1,\alpha})}{\|D\rho_{m+1,\alpha}\|}, \quad (20)$$

which coincides with ME-rule in the case $D = L^{-s}$.

Consider now some modifications of the ME-rule. We know from (15) that $\alpha_{\text{ME}} \geq \alpha_{\text{opt}} := \operatorname{argmin}\{\|x_{m,\alpha} - x^\dagger\|, \alpha \geq 0\}$. It means that typically somewhat a smaller parameter α_{ME}/c with proper $c > 1$ is better parameter than α_{ME} . For case $D = I$, $h = s = 0$ we optimized the value of c (and other constants below) in extensive numerical experiments and recommend the estimated parameter $\alpha_{\text{MEe}} = \alpha_{\text{ME}}/2.3$. Note that in the case of rough

estimate of the noise level, better rules than the ME-rule and the discrepancy principle are rules from the recently derived family of rules,[25] see also [26–29].

In the extrapolated Tikhonov method (8) $x_\alpha := x_{\alpha_1, \dots, \alpha_m}$, one may use the logarithmically uniform mesh of parameters $\alpha_n = \alpha q^{n-(m+1)/2}$, $n = 1, \dots, m$; $q < 1$ and to choose α by the same rules as in the iterated Tikhonov method.

Note that in [30], an analog of ME-rule in case $D = I$, $h = s = 0$ was proposed for the ($m \geq 1$ times iterated) Lavrentiev method, finding the parameter $\alpha = \alpha_{\text{MEa}}$ as solution of the equation $\|\rho_{m+1, \alpha}\|^2 / \|\rho_{m+2, \alpha}\| \leq C_m \delta$, where $C_1 = 1.55$, $C_2 = 1.6$. Numerical experiments in [30,31] showed that with this parameter choice, the error of regularized solution was in average only 5% larger (in modifications of this rule 3% larger) than with optimal parameter. Note also that several other analogs of ME-rule for the ($m \geq 1$ times iterated) Lavrentiev method were proposed and analysed in [32].

4.3. ME-rule for asymptotical regularization

In this regularization method, the regularized solution is given by x_α as the solution of the initial value problem

$$\frac{d}{dt}x(t) + B^*Bx(t) = B^*\bar{y} \quad \text{for } 0 < t \leq r, \quad x(0) = 0$$

with $r = 1/\alpha$. In this method, we have $g_\alpha(\lambda) = (1 - e^{-\lambda/\alpha})/\lambda$. Using the identities

$$\frac{d}{d\alpha}g_\alpha(\lambda) = -\frac{1}{\alpha^2}e^{-\lambda/\alpha} = -\frac{1}{\alpha^2}(1 - \lambda g_\alpha(\lambda))$$

we obtain for the method of asymptotical regularization

$$d_{\text{ME}}(\alpha) = \frac{(Ax_\alpha - y, L^{-2s}(Ax_\alpha - y))}{\|(D^{-1})^*L^{-2s}(Ax_\alpha - y)\|}.$$

In the case $s = 0$ and $D = I$, the ME-rule coincides with the discrepancy principle for which convergence and order optimal error estimates are well known (cf. [1,2,8,17]).

5. ME-rule for iterative regularization methods

5.1. Derivation of the ME-rule

For solving ill-posed problems (1), we now consider iteration methods of the general form

$$x_n = x_{n-1} + A^*z_{n-1}, \quad n = 1, 2, \dots \quad (21)$$

with $z_n \in Y$. The elements z_n characterize the special iteration method. We assume that $z_n \in \mathcal{D}((D^{-1})^*)$. Simple iteration methods are explicit iteration scheme (Landweber method) M3 with $z_n = \mu L^{-2s} \rho_n$, $\mu \in (0, 1/\|B^*B\|)$ and implicit iteration scheme M4 with $z_n = L^{-s}(\alpha I + BB^*)^{-1}L^{-s} \rho_n = \alpha^{-1}L^{-2s} \rho_{n+1}$, $\alpha > 0$, where $\rho_n = y - Ax_n$.

In the monotone error rule (ME-rule), we search for a *largest computable iteration number* n_{ME} for which the monotonicity property (3)

$$\|x_n - x^\dagger\| \leq \|x_{n-1} - x^\dagger\| \quad \text{for all } n = 1, 2, \dots, n_{\text{ME}} \quad (22)$$

can be guaranteed. Exploiting (21) and (10), we obtain

$$\begin{aligned}
\|x_{n+1} - x^\dagger\|^2 - \|x_n - x^\dagger\|^2 &= \|x_n + A^*z_n - x^\dagger\|^2 - \|x_n - x^\dagger\|^2 \\
&= 2(x_n - x^\dagger, A^*z_n) + \|A^*z_n\|^2 \\
&= 2(D(Ax_n - y_0), (D^{-1})^*z_n) + \|A^*z_n\|^2 \\
&= 2(D(y - Ax^\dagger - \rho_n), (D^{-1})^*z_n) + \|A^*z_n\|^2 \\
&\leq 2\|(D^{-1})^*z_n\|(\delta + h\|x^\dagger\|) - 2(\rho_n, z_n) + \|A^*z_n\|^2 \\
&=: \Phi(z_n).
\end{aligned} \tag{23}$$

This estimate leads us to the following ME-rule for iteration methods (21).

ME rule. Choose n_{ME} as the first index n satisfying

$$d_{\text{ME}}(n) := \frac{2(\rho_n, z_n) - \|A^*z_n\|^2}{2\|(D^{-1})^*z_n\|} \leq \Delta \tag{24}$$

with $\Delta \in \{\Delta_1, \Delta_2\}$.

The derivation of the ME-rule in (22) uses only one inequality, which turns to the equality in the case

$$y - Ax^\dagger = \frac{\delta + h\|x^\dagger\|}{\|(D^{-1})^*z_n\|} (D^*D)^{-1}z_n.$$

Therefore, in this case the ME-rule gives the optimal parameter.

Note that for $\Delta = \Delta_3$, the ME-property is not guaranteed but for certain methods it is known that quasi-optimality still holds (see Theorem 6.3). However, typically the monotonicity of error also holds up to somewhat larger iteration numbers, hence $\Delta = \Delta_3$ may also give better results than $\Delta \in \{\Delta_1, \Delta_2\}$.

From definition of $d_{\text{ME}}(n)$ follows the inequality $d_{\text{ME}}(n) \leq \|D\rho_n\| = d_D(n)$.

Due to equality $\rho_{n+1} = \rho_n - AA^*z_n$, the function $d_{\text{ME}}(n)$ may be presented also in forms

$$d_{\text{ME}}(n) := \frac{(\rho_n + \rho_{n+1}, z_n)}{2\|(D^{-1})^*z_n\|} = \frac{2(\rho_{n+1}, z_n) + \|A^*z_n\|^2}{2\|(D^{-1})^*z_n\|}. \tag{25}$$

Inequalities (23), (24) may be used:

- (a) for stopping iterations (21), if for the first time $d_{\text{ME}}(n) \leq \Delta$;
- (b) for choosing z_n , minimizing the function $\Phi(z_n)$ in error estimate (23) (for example choosing steplength μ in (26));
- (c) for using several iteration methods switching from some faster method (for example CGLS) to a slower method (for example Landweber method): if inequality (23) does not guarantee decrease of error in the fast method, the accuracy of this approximation may be further increased by a slower method.

5.2. ME-rule in gradient-type methods

Let us consider gradient-type methods of the form (21) with $z_n = \mu_n L^{-2s} \rho_n$, where $\rho_n = y - Ax_n$ is the discrepancy and $\mu_n > 0$ is some properly chosen stepsize:

$$x_n = x_{n-1} + \mu_{n-1} A^* L^{-2s} (y - Ax_{n-1}), \quad n = 1, 2, \dots \tag{26}$$

Special cases of method (26) are the Landweber method with $\mu_n = \mu \in (0, 1/\|B^*B\|)$, the minimal error method with $\mu_n = \frac{\|L^{-s}\rho_n\|^2}{\|A^*L^{-2s}\rho_n\|^2}$ and the steepest descent method with $\mu_n = \frac{\|A^*L^{-2s}\rho_n\|^2}{\|L^{-s}AA^*L^{-2s}\rho_n\|^2}$.

Here, the ME-function (25) has the form

$$d_{\text{ME}}(n) = \frac{(\rho_n + \rho_{n+1}, L^{-2s}\rho_n)}{2\|(D^{-1})^*L^{-2s}\rho_n\|}.$$

The estimate (23) which led us to the ME-rule can also be exploited for finding stepsizes $\mu_n > 0$ in iteration methods (26) which guarantee that x_{n+1} is a better approximation for x^\dagger than x_n . Here, the stepsize μ_n may not only depend on y but also on the noise level Δ . Exploiting $z_n = \mu_n L^{-2s}\rho_n$, the estimate (23) obtains form

$$\|x_{n+1} - x^\dagger\|^2 - \|x_n - x^\dagger\|^2 \leq 2\mu_n(\|(D^*)^{-1}L^{-2s}\rho_n\|\Delta - \|L^{-s}\rho_n\|^2) + \mu_n^2\|A^*L^{-2s}\rho_n\|^2. \quad (27)$$

Therefore, x_{n+1} is a better approximation for x^\dagger than x_n , if

$$\Delta < \frac{\|L^{-s}\rho_n\|^2}{\|(D^*)^{-1}L^{-2s}\rho_n\|}, \quad 0 < \mu_n < \frac{2(\|L^{-s}\rho_n\|^2 - \|(D^*)^{-1}L^{-2s}\rho_n\|\Delta)}{\|A^*L^{-2s}\rho_n\|^2}.$$

Minimizing the right-hand side of (27) with respect to μ_n yields

$$\mu_n = \frac{\|L^{-s}\rho_n\|^2 - \|(D^*)^{-1}L^{-2s}\rho_n\|\Delta}{\|A^*L^{-2s}\rho_n\|^2}.$$

Substituting into (27) shows that for this stepsize, the improvement of the squared error can be estimated by

$$\|x_{n+1} - x^\dagger\|^2 - \|x_n - x^\dagger\|^2 \leq -\frac{(\|L^{-s}\rho_n\|^2 - \|(D^*)^{-1}L^{-2s}\rho_n\|\Delta)^2}{\|A^*L^{-2s}\rho_n\|^2}.$$

In case $D = I$, $h = s = 0$ last two relations were given in [8,10], but convergence results and order optimal error bounds for stopping many gradient methods [2,9,10,17,33] with the ME-rule were stated in [4,7–10,34].

5.3. ME-rule in conjugate gradient methods

Some gradient-type methods that do not fit into the class of methods (26) are *conjugate gradient methods* (see [9,10,33,36]).

Methods CGLS and CGME are the conjugate gradient method applied to $B^*Bx = B^*\bar{y}$ and $BB^*v = \bar{y}$ with $x = B^*v$, respectively. In both methods, we fix the initial approximation x_0 , the initial value $u_0 = 0$ and $\beta_0 = 0$, and compute $\tau_0 = \bar{y} - Bx_0$. Further, we compute in CGLS $p_0 = B^*\tau_0$, and for $n = 1, 2, \dots$ iteratively

$$\begin{aligned} u_n &= \beta_{n-1}u_{n-1} - \tau_{n-1}, & \gamma_n &= \frac{\|p_{n-1}\|^2}{\|BB^*u_n\|^2}, & x_n &= x_{n-1} - \gamma_n B^*u_n, \\ \tau_n &= \tau_{n-1} + \gamma_n BB^*u_n, & p_n &= B^*\tau_n, & \beta_n &= \frac{\|p_n\|^2}{\|p_{n-1}\|^2}. \end{aligned}$$

In the method CGME, one computes for $n = 1, 2, \dots$ iteratively

$$\begin{aligned} u_n &= \beta_{n-1}u_{n-1} - \tau_{n-1}, & \gamma_n &= \frac{\|\tau_{n-1}\|^2}{\|B^*u_n\|^2}, \\ x_n &= x_{n-1} - \gamma_n B^*u_n, & \tau_n &= \tau_{n-1} + \gamma_n B B^*u_n, & \beta_n &= \frac{\|\tau_n\|^2}{\|\tau_{n-1}\|^2} \end{aligned}$$

Both CGLS and CGME method have the form (21) with $z_{n-1} = -\gamma_n L^{-s}u_n$, hence ME-rule (24) can be used. In case $D = I$, $h = s = 0$, the ME-rule for CGLS and CGME methods was proposed and numerically tested in [38].

5.4. ME-rule for sequence of approximations

In contrast to iterative methods where approximation on step n is computed using approximation on previous step $n - 1$, one may consider an arbitrary sequence of approximations

$$x_n = A^*w_n, \quad n = 1, 2, \dots, \quad w_n \in Y \quad (28)$$

and ask which from these approximations to take as approximate solution.

THEOREM 5.1 *If in sequence (28) the index $n = n_{ME}$ is chosen as the first index n satisfying*

$$d_{ME}(n) = \frac{(\rho_n + \rho_{n+1}, w_{n+1} - w_n)}{2\|((D^{-1})^*(w_{n+1} - w_n))\|} \leq \Delta \quad (29)$$

with $\Delta \in \{\Delta_1, \Delta_2\}$, then ME-property (3) holds.

Proof This follows from (25) with $z_n = w_{n+1} - w_n$. □

To use the functional $d_{ME}(n)$, elements w_n are needed. They can be found by first computing w_n and later $x_n = A^*w_n$.

Note that the last theorem also gives an alternative way to derive the ME-rule for continuous regularization methods, using the sequence $\alpha_n = q^n$, $q < 1$ and taking the limit $q \rightarrow 1$.

The last theorem may be applied as an alternative or a complement to any other parameter choice rule for finding a regularized approximation from the sequence in the form $x_n = A^*w_n$, $n = 1, 2, \dots$ (for example in (iterated) Tikhonov method). Note that computation of the sequence of all approximate solutions for $n = 1, 2, \dots, N$ with some N is required, for example, in the balancing principle (see [11,35,37,39]). This theorem guarantees that the stopping index $n_{ME,B} = \max(n_{ME}, n_B)$ is at least as good as the index n_B from arbitrary balancing principle: $\|x_{n_{ME,B}} - x^\dagger\| \leq \|x_{n_B} - x^\dagger\|$.

5.5. ME-rule for semi-iterative regularization methods

Let p_n be residual polynomials orthogonal in interval $[0, \|B^*B\|]$ with respect to some measure (see [40]). Then $p_0 = 1$,

$$p_n(\lambda) = p_{n-1}(\lambda) + (1 - \mu_n)(p_{n-2}(\lambda) - p_{n-1}(\lambda)) - \omega_n \lambda p_{n-1}(\lambda), \quad n \geq 2$$

with some constants μ_n and ω_n , where $\mu_1 = 1$. Then, semi-iterative approximations are found as

$$\begin{aligned} x_1 &= x_0 + \omega_1 A^* L^{-2s} (y - Ax_0), \\ x_n &= x_{n-1} + (1 - \mu_n)(x_{n-2} - x_{n-1}) + \omega_n A^* L^{-2s} (y - Ax_{n-1}), \quad n \geq 2. \end{aligned}$$

In case $x_0 = A^* w_0$, we have $x_n = A^* w_n$ with $w_1 = w_0 + \omega_1 L^{-2s} (y - AA^* w_0)$,

$$w_n = w_{n-1} + (1 - \mu_n)(w_{n-2} - w_{n-1}) + \omega_n L^{-2s} (y - AA^* w_{n-1}), \quad n \geq 2. \quad (30)$$

Thus, the iterations of the semi-iterative methods may be stopped by the ME-rule (29). We give formula for w_n for the following examples of semi-iterative methods (see [40]).

(1) The Chebyshev method of Stiefel

$$w_n = \frac{2n}{n+1} w_{n-1} - \frac{n-1}{n+1} w_{n-2} + \frac{4n}{n+1} L^{-2s} (y - AA^* w_{n-1}), \quad n = 1, 2, \dots$$

(2) The Chebyshev method of Nemirovskii and Polyak

$$w_1 = \frac{4}{3} L^{-2s} (y - AA^* w_0),$$

$$\begin{aligned} w_n &= 2 \frac{2n-1}{2n+1} w_{n-1} - \frac{2n-3}{2n+1} w_{n-2} + 4 \frac{2n-1}{2n+1} L^{-2s} (y - AA^* w_{n-1}), \\ n &= 2, 3, \dots \end{aligned}$$

(3) The ν -method of Brakhage is method where w_n in (30) has for fixed $\nu > 0$ the coefficients

$$\begin{aligned} \mu_1 &= 1, \quad \mu_n = 1 + \frac{(n-1)(2n-3)(2n+2\nu-1)}{(n+2\nu-1)(2n+4\nu-1)(2n+2\nu-3)}, \quad n = 2, 3, \dots \\ \omega_1 &= \frac{4\nu+2}{4\nu+1}, \quad \omega_n = \frac{4(2n+2\nu-1)(n+\nu-1)}{(n+2\nu-1)(2n+4\nu-1)}, \quad n = 2, 3, \dots \end{aligned}$$

5.6. ME-rule in implicit iteration methods

Let us consider implicit iteration methods of the form

$$\begin{aligned} x_{n+1} &= x_n + A^* L^{-s} (BB^* + \alpha_n I)^{-1} L^{-s} (y - Ax_n) \\ &= (B^* B + \alpha_n I)^{-1} (\alpha_{n-1} x_n + B^* L^{-s} y), \quad n = 0, 1, \dots \end{aligned} \quad (31)$$

This method has form (21) with

$$z_n = L^{-s} (\alpha_n I + BB^*)^{-1} L^{-s} \rho_n = \alpha_n^{-1} L^{-2s} \rho_{n+1}. \quad (32)$$

Therefore, the function $d_{ME}(n)$ in (25) attains the form

$$d_{ME}(n) = \frac{(\rho_n + \rho_{n+1}, L^{-2s} \rho_{n+1})}{\|(D^{-1})^* L^{-2s} \rho_{n+1}\|}.$$

In case $D = I$, $h = s = 0$, ME-rule for method (31) was studied in [4,7,8,34].

Let us now consider the question of choosing the parameter α_n such that the guaranteed improvement of accuracy of approximation x_{n+1} over x_n would be maximal.

PROPOSITION 5.2 *In regularization method (31), the guaranteed improvement $-\Phi(z_n)$ of accuracy in estimate (23) is maximized for z_n in (32) with α_n as solution α of the equation*

$$\begin{aligned} & \alpha \|(D^{-1})^* L^{-s} (\alpha I + BB^*)^{-1} L^{-s} \rho_n\| \|(\alpha I + BB^*)^{-3/2} L^{-s} \rho_n\|^2 \\ & = ((D^{-1})^* L^{-s} (\alpha I + BB^*)^{-1} L^{-s} \rho_n, (D^{-1})^* L^{-s} (\alpha I + BB^*)^{-2} L^{-s} \rho_n) \Delta_1. \end{aligned} \quad (33)$$

Proof Substitution of z_n in (32) to function $\Phi(z_n)$ in (23) shows that minimization of function $\Phi(z_n)$ is equivalent to minimization of the function

$$\begin{aligned} \phi(\alpha) := & 2\|(D^{-1})^* L^{-s} (\alpha I + BB^*)^{-1} \tau_n\| \Delta_1 - 2(\tau_n, (\alpha I + BB^*)^{-1} \tau_n) \\ & + \|B^* (\alpha I + BB^*)^{-1} \tau_n\|^2, \end{aligned}$$

where $\tau_n := L^{-s} \rho_n$. Differentiation with use of formula $\frac{1}{2} \frac{d}{d\alpha} \|\cdot\| = \left(\frac{1}{2} \frac{d}{d\alpha} \|\cdot\|^2\right) / (2\|\cdot\|)$ gives

$$\begin{aligned} \frac{1}{2} \phi'(\alpha) = & - \frac{((D^{-1})^* L^{-s} (\alpha I + BB^*)^{-1} \tau_n, (D^{-1})^* L^{-s} (\alpha I + BB^*)^{-2} \tau_n) \Delta_1}{\|(D^{-1})^* L^{-s} (\alpha I + BB^*)^{-1} \tau_n\|} \\ & + \|(\alpha I + BB^*)^{-1} \tau_n\|^2 - \|B^* (\alpha I + BB^*)^{-3/2} \tau_n\|^2. \end{aligned}$$

The equality

$$\begin{aligned} \|(\alpha I + BB^*)^{-1} \tau_n\|^2 & = ((\alpha I + BB^*) \tau_n, (\alpha I + BB^*)^{-3} \tau_n) \\ & = \alpha \|(\alpha I + BB^*)^{-3/2} \tau_n\|^2 + \|B^* (\alpha I + BB^*)^{-3/2} \tau_n\|^2 \end{aligned}$$

shows that $\phi'(\alpha) = 0$ iff α solves the equation (33). \square

THEOREM 5.3 *In the case $D = L^{-s}$ in regularization method (31), the guaranteed improvement $-\Phi(z_n)$ of accuracy in estimate (23) is maximized, if α_n is chosen by the analogue of the discrepancy principle $\|L^{-s} \rho_{n+1}\| = \Delta_1$.*

Proof Indeed, in the case $D = L^{-s}$, the equation (33) for finding $\alpha = r_n^{-1}$ has the form $\alpha_n \|(\alpha_n I + BB^*)^{-1} L^{-s} \rho_n\| = \Delta_1$, which is equivalent to $\|L^{-s} \rho_{n+1}\| = \Delta_1$ due to the equality $\alpha_n (\alpha_n I + BB^*)^{-1} L^{-s} \rho_n = L^{-s} \rho_{n+1}$. \square

6. Convergence and quasi-optimality for the ME-rule

In the ME-rule, considered in previous sections, the regularization parameter r in continuous regularization methods is found as the solution (the largest solution in case of many solutions) of a certain equation $d(r, y, A) = \Delta$ with $\Delta \in \{\Delta_1, \Delta_2\}$. In regularization methods, where the regularization parameter is a natural number as in iteration methods, we choose r_{ME} as the first $r = 1, 2, \dots$ for which $d(r, y, A) \leq \Delta$. Let us denote by x_r^0 the regularized solutions for exact data y_0, A_0 . We give for the ME-rule, the following convergence result.

THEOREM 6.1 Let $\Delta \in \{\Delta_1, \Delta_2\}$. Let x_r^0, x_r satisfy the following properties:

- (1) $\|x_r^0 - x^\dagger\| \rightarrow 0$ as $r \rightarrow \infty$;
- (2) $\|x_{r_1}^0 - x_{r_1}\| \leq \|x_{r_2}^0 - x_{r_2}\|$ (for $r_1 \leq r_2$);
- (3) if $r(\Delta) \leq \text{const} < \infty$ as $\Delta \rightarrow 0$, then $\|x_{r(\Delta)} - x_{r(\Delta)}^0\| \rightarrow 0$ as $\Delta \rightarrow 0$;
- (4) if for some sequence $r_k \leq \text{const} < \infty$ ($k \rightarrow \infty$) it holds true $d(r_k, y_0, A_0) \rightarrow 0$ ($k \rightarrow \infty$), then $\|x_{r_k}^0 - x^\dagger\| \rightarrow 0$ ($k \rightarrow \infty$).

Then, $\|x_{r_{ME}} - x^\dagger\| \rightarrow 0$ as $\Delta \rightarrow 0$.

Proof For every r it holds

$$\|x_r - x^\dagger\| \leq \|x_r - x_r^0\| + \|x_r^0 - x^\dagger\|. \quad (34)$$

First we consider the main case where $r_{ME}(\Delta) \rightarrow \infty$ as $\Delta \rightarrow 0$. Then, $\|x_{r_{ME}}^0 - x^\dagger\| \rightarrow 0$ due to the assumption (1). It remains to show that $\|x_{r_{ME}}^0 - x_{r_{ME}}\| \rightarrow 0$ as $\Delta \rightarrow 0$. Let $r_0(\Delta)$ be a monotonically increasing function giving some *a priori* regularization parameter guaranteeing convergence $\|x_{r_0} - x^\dagger\| \rightarrow 0$ as $\Delta \rightarrow 0$. If $r_{ME} \leq r_0$, then due to assumption (2)

$$\|x_{r_{ME}}^0 - x_{r_{ME}}\| \leq \|x_{r_0}^0 - x_{r_0}\|.$$

If $r_{ME} \geq r_0$, then ME-property says that

$$\|x_{r_{ME}} - x^\dagger\| \leq \|x_{r_0} - x^\dagger\|.$$

Both estimates converge to 0 as $\Delta \rightarrow 0$.

Consider now the case if $r_{ME}(\Delta) \leq \text{const}$ ($\Delta \rightarrow 0$). Then, for $r = r_{ME}$ both terms in the right-hand side of (34) converge as $\Delta \rightarrow 0$ due to assumptions (3) and (4). \square

Assumptions (1)–(3) of Theorem 6.1 are general concerning regularization method. Only the last assumption (4) uses concrete form of the function $d(r, y, A)$, for regularization method (7) it can be proved similarly to Lemma 3.2 in [2, p. 66].

Consider now the case $D = L^{-s}$. The problem $Ax = y$ may be rewritten in form $Bx = \bar{y}$ with operator $B \in L(X, Y)$ and for x_r in (7)

$$x_r = B^* g_r(BB^*)\bar{y} = g_r(B^*B)B^*\bar{y} \quad (35)$$

the standard regularization theory applies. We get the following results.

THEOREM 6.2 Let (4) hold with $D = L^{-s}$. Let x_r be defined by one of methods M1–M5. Then, choice of r by one of rules D, MD, ME or MEe guarantees convergence $\|x_r - x^\dagger\| \rightarrow 0$ as $\Delta \rightarrow 0$ and under additional assumption $x^\dagger \in \mathcal{R}((B^*B)^{p/2})$ the order optimal error estimate $\|x_r - x^\dagger\| \leq O(\Delta^{\frac{p}{p+1}})$ holds with $p \leq 2p_0 - 1$ for rule D, with $p \leq 2p_0$ for rules MD, ME and MEe.

Proof Let $D = L^{-s}$. Then, results of this theorem for discrepancy principle follow from the results in [1,2,17], for MD rule in [22,41]. Let $C = 1$ in rules D, MD. In methods M1, M2 it holds $d_{MD}(\alpha) \leq d_{ME}(\alpha) \leq d_D(\alpha)$, therefore $\alpha_D \leq \alpha_{ME} \leq \alpha_{MD}$ ([5–8]), in methods

M3, M4 $n_D - 1 \leq n_{ME} \leq n_D$ ([4,7,8]), in method M5 $d_{ME}(\alpha) = d_D(\alpha)$. Therefore, results for ME rule follow from results for D rule and MD rule. The same results hold for parameter $r_{MEe} = \text{const } r_{ME}$. \square

Let us consider the quasi-optimality results for the regularization methods. To characterize the quality of the rule of an *a posteriori* regularization parameter choice, we use in the following the quasi-optimality property (see [41,42]). We say that a rule R for *a posteriori* choice of the regularization parameter $r = r(R)$ is quasi-optimal for a given regularization method $x_r = A^* g_r(AA^*)y$, if there exists a constant C (not depending on A_0, A, x^\dagger, y) such that for $\|y - y_0\| \leq \delta, \|A - A_0\| \leq h$ the error estimate

$$\|x_{r(R)} - x^\dagger\| \leq C \inf_{r \geq 0} \psi(r) + \mathcal{O}(\delta + h\|x^\dagger\|), \quad (36)$$

holds, where the function

$$\psi(r) = \|(I - A^* A g_r(A^* A))x^\dagger\| + \gamma_* \sqrt{r}(\delta + h\|x^\dagger\|) \quad (37)$$

is an upper bound of the error $\|x_r - x^\dagger\|$. The following result holds.

THEOREM 6.3 *Let $D = L^{-s}$. The modified discrepancy principle with $\Delta \in \{\Delta_1, \Delta_2, \Delta_3\}$ and $C > 1$ is quasi-optimal rule for solving the problem $Bx = \bar{y}$ by methods M1–M5: the inequality (36) holds where the operators A, A^* are replaced by operators B, B^* in (37). This quasi-optimality is also guaranteed for ME-rule with $\Delta \in \{\Delta_1, \Delta_2\}$, but also with $\Delta = \Delta_3$ if in (17) and (24) Δ is replaced by $C\Delta$ with $C > 1$.*

Proof Applying results of [41] (Theorem 3) (see also [34](Theorem 2, Remark 2)) to the approximation (35) instead of $x_r = A^* g_r(AA^*)y$, we get the assertions of theorem about MD rule, but also for ME rule with $\Delta \in \{\Delta_1, \Delta_2, \Delta_3\}$ if in (17) and (24) $C\Delta$ with $C > 1$ is used. In case $\Delta \in \{\Delta_1, \Delta_2\}$, the increase of $C \geq 1$ in right-hand side $C\Delta$ in (17) and (24) leads to decrease of parameter $r \leq r_{ME}$, therefore, to increase of the error $\|x_r - x^\dagger\|$. \square

7. Numerical examples

Our tests are performed on the well-known set of test problems by Hansen [43]: *baart, deriv2, foxgood, gravity, heat, ilaplace, phillips, shaw, spikes, wing*. In all tests, we used the discretization parameter $n = 100$.

For example, the test problem *heat* (see [43,44]) represents the heat conduction in one-dimensional semi-infinite rod $[0, \infty)$ with the thermal diffusivity 1. The initial temperature of the rod is 0 and the boundary condition at the point $\xi = 0$ is not known. Instead, one measures the temperature at $\xi = 1$ for the time $t \in [0, 1]$ and tries to determine the temperature at $\xi = 0$ from this measurement. (NB! In this paragraph and in the following paragraph, t and s have different meaning from the rest of the paper.)

This results in the Volterra integral equation of the first kind

$$\int_0^t \frac{(t-s)^{-3/2}}{2\sqrt{\pi}} \exp\left(-\frac{1}{4(t-s)}\right) x(s) ds = y(t), \quad t \in [0, 1],$$

where $y(t)$ is the measured temperature at $\xi = 1$ and $x(t)$ is the temperature at $\xi = 0$.

The integral equation is discretized by means of a simple collocation method on the points $t_i = i/n$ and the midpoint rule with $n = 100$ points. An exact discrete solution z with components

$$z_i = \begin{cases} 75t_i^2, & \text{if } t_i \leq 0.1 \\ 0.75 + (20t_i - 2)(3 - 20t_i), & \text{if } 0.1 \leq t_i < 0.15 \\ 0.75 \exp(-2(20t_i - 3)), & \text{if } 0.15 < t_i \leq 0.5 \\ 0, & \text{otherwise} \end{cases}$$

is constructed, and then the right-hand side b is computed as $b = Az$.

Since the performance of methods and rules generally depends on the smoothness p of the exact solution, we complemented the standard solutions x^\dagger of the (now discrete) test problems with smoothed solutions $(A^*A)^{p/2}x^\dagger$ ($p = 2$) in all test problems (computing the right-hand side as $A((A^*A)^{p/2}x^\dagger)$). After discretization, all problems were scaled (normalized) in such a way that the Euclidian norms of the operator and the right-hand side were 1. We used operator $D = L^{-t}$ with $t \in \{-0.5, 0, 0.5\}$. We computed $L^{-t}y_0$ and added a normally distributed noise such that $\|L^{-t}(y - y_0)\|$ had the values $0.3, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}$. Here, $\|\cdot\|$ means the Euclidean norm.

We generated 10 noise vectors and used these vectors in all of the problems. The problems were regularized by the Tikhonov method in Y_s -scales (the second formula in (9)), where the regularization parameters were chosen according to the rules that we wanted to compare. We used the Y_s -scale with the operator L acting as $Lu = u - u''$. In the discrete form, we used the approximating matrix (l_{ij}) with the elements $l_{ii} = 2n^2 + 1, i = 1, \dots, n, l_{i,i+1} = l_{i+1,i} = -n^2, i = 1, \dots, n-1, l_{ij} = 0$ otherwise.

Since in these model equations the exact solution is known, it is possible to find the regularization parameter $\alpha = \alpha_*$ which gives the smallest error: $\|x_{\alpha_*} - x^\dagger\| = \min_{\alpha > 0} \{\|x_\alpha - x^\dagger\|\}$. For every rule R, the error ratio $\|x_{\alpha_R} - x^\dagger\|/\|x_{\alpha_*} - x^\dagger\|$ describes the performance of the rule R on this particular problem. For better comparison of the cases with different s , in the denominator we always use $s = 0$. To compare the rules or to present their properties, the following tables show averages of these error ratios over various parameters of the data set (problems 1–10, noise levels δ , noise vectors).

In Tables 1 and 2 we compare the following rules. Earlier we have introduced the discrepancy principle D (see page 5, formula (11)), the rule MD' (page 7, (19)), the monotone error rule ME (page 7, (19)) and the rule MEa as the analog of the ME-rule (page 7, (20)). In rule MEe (recall motivation on page 7), we chose the regularization parameter as $\alpha_{\text{MEe}} = \alpha_{\text{ME}}/c$ with $c = 2.3 \cdot 10^{c_1}, c_1 = (s-t)(1.5 + 1.3s - 0.5t - 0.8s^2 + 2st - t^2)$; this form of the constant was obtained by extensive numerical experiments on different values of s and t . In rule MEae, we chose the parameter $\alpha_{\text{MEae}} = \alpha_{\text{MEa}}/c$, where $c = 2.3$ for $t = 0$ and $c = 3$ for $t = -0.5$ or $t = 0.5$.

Note that in the case $s > t$, the ME rule (unlike rules D, MD', MEa) allows mild underestimation of the noise level. In the columns ME08 and ME07, we present the error ratios obtained by using the noise levels 0.8δ and 0.7δ instead of δ . Typically, ME08 and ME07 give better results than ME rule, but sometimes they can give large errors. If instead of the exact noise level approximate noise level is given, we recommend using the MEe-rule with $s > t$. In the case $s = t$, all considered rules failed if the noise level was underestimated.

In Table 1, we present the results separately for large δ 's (1 means $\delta \geq 10^{-2}$) and small δ 's (s means $\delta \leq 10^{-3}$). We present results for $s = t$ and $s = t + 0.5$. For a given t in most

Table 1. Error ratios for different rules.

t	s	p	δ	D	MD'	ME	ME08	ME07	MEe	MEa	MEae
-0.5	-0.5	0	l	1.20	1.38	1.28			1.15	1.28	1.17
-0.5	-0.5	0	s	1.44	2.41	1.48			1.23	1.48	1.19
-0.5	-0.5	2	l	4.10	4.90	4.35			3.73	4.35	3.90
-0.5	-0.5	2	s	16.2	18.1	14.9			10.4	14.9	10.2
-0.5	0	0	l	3e+2	2e+2	1.97	1.80	1.70	1.26	1e+2	1e+2
-0.5	0	0	s	2.84	2.72	2.06	1.86	15.5	1.31	1.89	1.89
-0.5	0	2	l	4e+4	2e+4	3.56	2.92	2.58	1.27	2e+4	2e+4
-0.5	0	2	s	5e+3	2e+3	2.29	2.03	2.47	1.54	1e+3	1e+3
0	0	0	l	1.15	1.37	1.28			1.13	1.28	1.13
0	0	0	s	1.22	1.51	1.38			1.19	1.38	1.19
0	0	2	l	1.24	1.53	1.29			1.09	1.29	1.09
0	0	2	s	3.66	1.91	1.57			1.11	1.57	1.11
0	0.5	0	l	1.61	1.48	2.13	1.93	1.98	1.32	1.34	1.26
0	0.5	0	s	1.34	1.55	2.00	1.82	3.30	1.36	1.46	1.37
0	0.5	2	l	1.54	1.68	3.20	2.61	2.31	1.19	1.52	1.24
0	0.5	2	s	3.31	3.78	6.97	5.96	5.89	2.95	3.55	3.33
0.5	0.5	0	l	0.92	1.02	0.98			0.90	0.98	0.89
0.5	0.5	0	s	1.18	1.35	1.29			1.15	1.29	1.12
0.5	0.5	2	l	0.61	0.77	0.66			0.60	0.66	0.66
0.5	0.5	2	s	1.18	1.66	1.47			1.13	1.47	1.10
0.5	1	0	l	0.99	1.02	1.56	1.46	1.40	1.14	1.04	0.90
0.5	1	0	s	1.24	1.37	1.77	1.64	2.08	1.22	1.32	1.20
0.5	1	2	l	0.54	0.76	2.37	1.98	1.76	1.31	0.66	0.58
0.5	1	2	s	1.56	1.90	3.22	2.85	3.00	1.70	1.76	1.56

cases, the best choice of s is $s = t$, in which case all of the considered rules gave similar results. But in the case $s = t = -0.5$, we could not obtain good results for smooth solutions, especially in case of small noise levels. Then, the error function had a very sharp decrease at its minimum, which was hard to locate. Therefore, in the case $t = -0.5$ we recommend taking $s = 0$. Then, the ME and MEe rules gave good results. However, other rules failed in this case in the problems *baart*, *foxgood*, and *wing*.

In most problems, using other values of s and t than $s = t$ and $s = t + 0.5$ caused slightly worse results in the rules ME and MEe and essentially worse results in rules D and MD'. However, in the problems *deriv2* and *phillips* in the case of a smooth solution ($p \geq 2$), increasing s increased the accuracy of rules D, MD', MEe for all t (in problem *deriv2* the accuracy of rule D increased for s up to 0.5).

Table 1 shows that the discrepancy principle works well in the case $s = t = 0$, $s = t = 0.5$ but is not as good for $p > 1$. Typically, the MEe-rule gives the best results. We made computations besides MD' rule (13) also with MD rule (12), the results were almost the same.

Table 2 presents the results for the problem *heat*. In column 'min', we present averages of minimal errors over 10 runs for concrete s and t . Unlike Table 1, the rules D, MD', MEa, MEae gave almost as good results as the rule MEe. These rules were not as good as the rule

Table 2. Results for the problem *heat*.

t	s	p	δ	min	D	MD'	ME	ME08	ME07	MEe	MEa	MEae
-0.5	-0.5	0	10^{-1}	1.08e-1	1.21	1.58	1.35			1.19	1.35	1.27
-0.5	-0.5	0	10^{-4}	9.35e-3	1.05	1.25	1.16			1.05	1.16	1.04
-0.5	-0.5	2	10^{-1}	5.56e-2	5.05	5.96	5.12			4.88	5.12	5.37
-0.5	-0.5	2	10^{-4}	9.85e-4	24.3	19.4	16.3			13.5	16.3	14.0
-0.5	0	0	10^{-1}	9.34e-2	1.14	1.49	1.58	1.34	1.22	1.12	1.33	1.33
-0.5	0	0	10^{-4}	9.48e-3	1.12	1.25	1.32	1.26	1.23	1.03	1.18	1.18
-0.5	0	2	10^{-1}	1.28e-2	1.04	3.25	2.79	2.07	1.74	1.19	3.10	3.10
-0.5	0	2	10^{-4}	7.45e-5	1.29	3.38	1.45	1.27	1.17	1.71	2.77	2.77
0	0	0	10^{-1}	5.97e-1	1.07	1.24	1.15			1.02	1.15	1.02
0	0	0	10^{-4}	3.34e-2	1.00	1.26	1.14			1.01	1.14	1.01
0	0	2	10^{-1}	8.67e-2	1.15	1.22	1.07			1.14	1.07	1.14
0	0	2	10^{-4}	1.15e-3	2.50	1.40	1.18			1.04	1.18	1.04
0	0.5	0	10^{-1}	5.71e-1	1.11	1.24	1.89	1.66	1.51	1.15	1.17	1.04
0	0.5	0	10^{-4}	3.25e-2	1.04	1.27	1.17	1.06	1.17	1.21	1.16	1.01
0	0.5	2	10^{-1}	4.93e-2	0.62	0.81	1.69	1.47	1.35	0.57	0.72	0.60
0	0.5	2	10^{-4}	9.89e-4	0.89	1.19	1.33	1.05	8.52	1.23	1.05	0.88
0.5	0.5	0	10^{-1}	1.18	0.90	0.93	0.91			0.87	0.91	0.87
0.5	0.5	0	10^{-4}	1.64e-1	0.95	1.18	1.08			0.90	1.08	0.88
0.5	0.5	2	10^{-1}	1.41e-1	0.34	0.39	0.36			0.35	0.36	0.37
0.5	0.5	2	10^{-4}	7.21e-3	0.61	0.79	0.71			0.61	0.71	0.63
0.5	1	0	10^{-1}	1.15	0.92	0.94	1.14	1.13	1.11	0.98	0.93	0.89
0.5	1	0	10^{-4}	1.55e-1	1.03	1.20	1.30	1.10	7.04	0.92	1.13	0.92
0.5	1	2	10^{-1}	1.15e-1	0.28	0.36	0.84	0.67	0.59	0.31	0.33	0.29
0.5	1	2	10^{-4}	6.79e-3	0.63	0.80	1.21	1.00	0.86	0.67	0.72	0.58

MEe only in the case of smooth solution ($p = 2$): rule D for $t = s = 0$ and rules MD', MEa, MEae for $t = -0.5, s = 0$.

Both in Tables 1 and 2 in case $t = 0.5, s = 1$, the best results were obtained by using the MEae rule.

Some remarks about numerical realization of algorithms. In our numerical computations, the discretization parameter n was 100. This low dimension allowed us to use the advantages of singular value decomposition of the matrix for fast computations. Note also that instead of the solution of the Equations (11), (19) and (20) we made computations on the α -sequence $\alpha_i = 0.9^{-i}, i = 0, 1, \dots$ and used for the corresponding regularization parameter the first α_i for which the left hand side of the equation was smaller than the right-hand side.

The computations are made with the exact operator, since the inexact operator is typically accompanied by a significant overestimation of the noise level. In this case, it is preferable to use the rules from the article [25] where the exact operator was considered. We plan to extend these results to the case of a noisy operator in a forthcoming paper.

It is worthwhile to think about the possibility to choose the operator L depending on the operator A .

Note that the ME-rule can also be used in projection methods. It was studied for the exact operator in [45]. Generalizations to noisy operators are to be published in a further paper.

8. Conclusion

We have considered a linear ill-posed problem $A_0x = y_0$ with the operator $A_0 \in \mathcal{L}(X, Y)$. A noisy right-hand side y and an operator A with noise levels $\|D(y_0 - y)\| \leq \delta$, $\|D(A_0 - A)\| \leq h$, where D is certain operator, are given. We derived the monotone error rule (ME rule) for parameter choice in many regularization methods in Y_{-s} -scale, generated by powers $s \in \mathbb{R}$ of certain operator L . These regularization methods include the (iterated) Tikhonov method and many iterative methods (Landweber method, steepest descent method, conjugate gradient-type methods, semi-iterative methods, implicit iteration method). It was shown for which data the ME rule gives the optimal parameter. For a class of methods, the convergence and quasi-optimality results are given. In extensive numerical experiments, the ME-rule and its modification MEe gave good results, whereby in the case $D = L^{-t}$ with $t < s$ these two rules allowed also moderate underestimation of the noise level (other rules fail in case of underestimated noise level).

Acknowledgements

This work was started while the first author held an appointment as Visiting Professor at the University of Applied Sciences Zittau/Görlitz from June until August 2010 and was supported by DFG (Deutsche Forschungsgemeinschaft) under a cooperative bilateral research project grant. In addition, the work was supported by the Estonian Science Foundation, Grant 9120.

References

- [1] Engl HW, Hanke M, Neubauer A. Regularization of inverse problems. Vol. 375, Mathematics and its applications. Dordrecht: Kluwer; 1996.
- [2] Vainikko GM, Veretennikov AY. Iteration procedures in ill-posed problems. Moscow: Nauka; 1986 (in Russian).
- [3] Tautenhahn U. On a new parameter choice rule for ill-posed inverse problems. In: Lipitakis EA, editor. HERCMA '98: Proc. 4th Hellenic-European Conference on Computer Mathematics and its Applications. Athens: Lea Publishers; 1999. p. 118–125.
- [4] Hämarik U. Monotonicity of error and choice of the stopping index in iterative regularization methods. In: Pedas A, editor. Differential and integral equations: theory and numerical analysis. Tartu: Estonian Mathematical Society; 1999. p. 15–30.
- [5] Hämarik U, Raus T. On the a posteriori parameter choice in regularization methods. Proc. Estonian Acad. Sci. Phys. Math. 1999;48:133–145.
- [6] Tautenhahn U, Hämarik U. The use of monotonicity for choosing the regularization parameter in ill-posed problems. Inverse Probl. 1999;15:1487–1505.
- [7] Hämarik U, Tautenhahn U. On the monotone error rule for parameter choice in iterative and continuous regularization methods. BIT Numerical Math. 2001;41:1029–1038.
- [8] Hämarik U, Tautenhahn U. On the monotone error rule for choosing the regularization parameter in ill-posed problems. In: Lavrentiev MM, Kabanikhin SI, editors. Ill-posed and non-classical problems of mathematical physics and analysis. Utrecht: VSP; 2003. p. 27–55.
- [9] Alifanov OM, Rummyantsev SV. On the stability of iterative methods for the solution of linear ill-posed problems. Soviet Math. Dokl. 1979;20:1133–1136.
- [10] Alifanov OM, Artyukhin EA, Rummyantsev SV. Extreme methods for solving ill-posed problems with applications to inverse heat transfer problems. New York: Begell House; 1995.
- [11] Egger H. Y -scale regularization. SIAM J. Numer. Anal. 2008;46:419–436.
- [12] Egger H. Regularization of inverse problems with large noise. J. Phys.: Conf. Ser. 2008;124:9pp.

- [13] Morozov VA. Regularization under large noise. *Comput. Math. Math. Phys.* 1996;36:1175–1181.
- [14] Eggermont PPB, LaRiccia VN, Nashed MZ. On weakly bounded noise in ill-posed problems. *Inverse Prob.* 2009;25:14 p.
- [15] Mathé P, Tautenhahn U. Enhancing linear regularization to treat large noise. *J. Inverse Ill-Posed Prob.* 2011;19:859–879.
- [16] Mathé P, Tautenhahn U. Regularization under general noise assumptions. *Inverse Prob.* 2011;27:035016 (15 p).
- [17] Vainikko GM. The discrepancy principle for a class of regularization methods. *USSR Comp. Math. Math. Phys.* 1982;22:1–19.
- [18] Hämarik U, Palm R, Raus T. Use of extrapolation in regularization methods. *J. Inverse Ill-Posed Prob.* 2007;15:277–294.
- [19] Hämarik U, Palm R, Raus T. Extrapolation of Tikhonov regularization method. *Math. Model. Anal.* 2010;15:55–68.
- [20] Tautenhahn U. Error estimates for regularization methods in Hilbert scales. *SIAM J. Numer. Anal.* 1996;33:2120–2130.
- [21] Morozov VA. On the solution of functional equations by the method of regularization. *Soviet Math. Dokl.* 1966;7:414–417.
- [22] Raus T. On the discrepancy principle for solution of ill-posed problems with non-selfadjoint operators. *Acta et comment. Univ. Tartuensis.* 1985;715:12–20 (in Russian).
- [23] Gfrerer H. An a posteriori parameter choice for ordinary and iterated Tikhonov regularization of ill-posed problems leading to optimal convergence rates. *Math. Comp.* 1987;49:507–522.
- [24] Tautenhahn U. On order optimal regularization under general source conditions. *Proc. Est. Acad. Sci. Phys. Math.* 2004;53:116–123.
- [25] Hämarik U, Palm R, Raus T. A family of rules for parameter choice in Tikhonov regularization of ill-posed problems with inexact noise level. *J. Comput. Appl. Math.* 2012;236:2146–2157.
- [26] Hämarik U, Palm R, Raus T. Comparison of parameter choices in regularization algorithms in case of different information about noise level. *Calcolo.* 2011;48:47–59.
- [27] Hämarik U, Kangro U, Palm R, Raus T. On parameter choice in the regularization of ill-posed problems with rough estimate of the noise level of the data. In: *Numerical analysis and applied mathematics ICNAAM 2012.* AIP Conference Proceedings, 1479. New York (NY): American Institute of Physics; 2012. p. 2332–2335.
- [28] Hämarik U, Raus T. On the choice of the regularization parameter in ill-posed problems with approximately given noise level of data. *J. Inverse Ill-Posed Prob.* 2006;14:251–266.
- [29] Hämarik U, Palm R, Raus T. On minimization strategies for choice of the regularization parameter in ill-posed problems. *Numer. Funct. Anal. Opt.* 2009;30:924–950.
- [30] Hämarik U, Raus T, Palm R. On the analog of the monotone error rule for parameter choice in the (iterated) Lavrentiev regularization. *Comput. Meth. Appl. Math.* 2008;8:237–252.
- [31] Palm R. Numerical comparison of regularization algorithms for solving ill-posed problems. University of Tartu; 2010. Available from: <http://hdl.handle.net/10062/14623>.
- [32] Hämarik U, Palm R, Raus T. A family of rules for the choice of the regularization parameter in the Lavrentiev method in the case of rough estimate of the noise level of the data. *J. Inverse Ill-Posed Prob.* 2012;20:831–854.
- [33] Gilyazov SF, Goldman NL. Regularization of ill-posed problems by iteration methods. Vol. 499, *Mathematics and its applications.* Dordrecht: Kluwer; 2000.
- [34] Hämarik U, Raus T. On the choice of the stopping index in iteration methods for solving problems with noisy data. In: Lipitakis EA, editor. *HERCMA 2001: Proceedings of the 5th Hellenic-European conference on computer mathematics and its applications,* Athens, Greece, September 20–22, 2001. Athens: Lea Publishers; 2002. p. 524–529.
- [35] Mathé P, Pereverzev SV. Geometry of linear ill-posed problems in variable Hilbert scales. *Inverse Prob.* 2003;19:789–803.

- [36] Hanke M. Conjugate gradient type methods for ill-posed problems. Harlow: Longman Scientific & Technical; 1995.
- [37] Pereverzev SV, Schock E. On the adaptive selection of the parameter in the regularization of ill-posed problems. *SIAM J. Numer. Anal.* 2005;43:2060–2076.
- [38] Hämarik U, Palm R. On rules for stopping the conjugate gradient type methods in ill-posed problems. *Math. Model. Anal.* 2007;12:61–70.
- [39] Hämarik U, Raus T. About the balancing principle for choice of the regularization parameter. *Numer. Funct. Anal. Opt.* 2009;30:951–970.
- [40] Hanke M. Accelerated Landweber iterations for the solution of ill-posed equations. *Numer. Math.* 1991;60:341–373.
- [41] Raus T, Hämarik U. On the quasi-optimal rules for the choice of the regularization parameter in case of a noisy operator. *Adv. Comput. Math.* 2012;36:221–233.
- [42] Hämarik U, Palm R, Raus T. On the quasioptimal regularization parameter choices for solving ill-posed problems. *J. Inverse Ill-Posed Prob.* 2007;15:419–439.
- [43] Hansen PC. Regularization tools: a Matlab package for analysis and solution of discrete ill-posed problems. *Numerical Algorithms.* 1994;6:1–35.
- [44] Eldén L. The numerical solution of a non-characteristic Cauchy problem for a parabolic equation. In: *Numerical Treatment of Inverse Problems in Differential and Integral Equations, Proceedings of an International Workshop, Heidelberg, 1982.* Boston: Birkhäuser; 1983. p. 246–268.
- [45] Hämarik U, Avi E, Ganina A. On the solution of ill-posed problems by projection methods with a posteriori choice of the discretization level. *Math. Model. Anal.* 2002;7:241–252.