Mathematical Modelling and Analysis. Ed. A.Buikis
Volume 12 Number 1, 2007, pages 61–70
 ISSN 1392-6292 print, ISSN 1648-3510 online

ON RULES FOR STOPPING THE CONJUGATE GRADIENT TYPE METHODS IN ILL-POSED PROBLEMS[1]

U. HÄMARIK and R. PALM

University of Tartu

Liivi 2, 50409 Tartu, Estonia

E-mail: uno.hamarik@ut.ee; reimo.palm@ut.ee

Received October 3, 2006; revised November 19, 2006; published online February 10, 2007

Abstract. We consider stopping rules in conjugate gradient type iteration methods for solving linear ill-posed problems with noisy data. The noise level may be known exactly or approximately or be unknown. We propose several new stopping rules, mostly for the case of unknown noise level. Numerical comparison with known rules (discrepancy principle, montone error rule, L-curve rule, Hanke-Raus rule) shows that the new rules are competitive.

Key words: Ill-posed problems, conjugate gradient type iteration methods, noise level, stopping rule, discrepancy principle, monotone error rule, L-curve rule

1. Introduction

We consider an operator equation

$$Au = f_*\,, \quad f_* \in \mathcal{R}(A), \tag{1.1}$$

where A is a linear bounded operator between Hilbert spaces H and F. In general, the problem (1.1) is ill-posed (see [4, 21]): the range $\mathcal{R}(A)$ may be non-closed, the kernel $\mathcal{N}(A)$ may be non-trivial. In practice often instead of the exact data f_* only an approximation f is given (containing, for example, measurement errors). If an ill-posed problem is solved by some iterative method, typically on first iteration steps $n = 1, 2, \ldots$ the iterated approximation u_n approaches to the minimal-norm solution u_* of (1.1), the error $\|u_n - u_*\|$ has minimal value for some $n_{\rm opt}$ and increases for $n > n_{\rm opt}$. Therefore, the iterations should be stopped after a certain number n of steps. If the exact noise level δ with $\|f_* - f\| \le \delta$ is given, the proper choice of $n = n(\delta)$

[1] This work was supported by the Estonian Science Foundation (Grant No. 5785).

guarantees the convergence $u_{n(\delta)} \to u_*$ as $\delta \to 0$. For many iterative methods this convergence is guaranteed by choice of n by the discrepancy principle or by its modifications (see [1, 4, 5, 6, 8, 9, 10, 13, 16, 17, 18, 19, 21] or by the monotone error rule [6, 8, 9]. If there is no information about the noise level δ, then no rule can guarantee the convergence $u_{n(\delta)} \to u_*$, $\delta \to 0$ (see [2]). Nevertheless, iterations may be stopped e.g. by heuristic rules from [3, 4, 6, 10, 14]. In some applications the noise level δ is given approximately: it holds

$$\frac{\|f_* - f\|}{\delta} \le C \quad \text{for } \delta \to 0,$$

where C is an unknown constant. In this case convergence $u_{n(\delta)} \to u_*$ ($\delta \to 0$) for explicit and implicit iteration schemes is guaranteed by stopping iterations by rule from [7]. In [6] we formulated the monotone error rule and the analogue of rule [7] for conjugate gradient type methods.

In this paper we formulate many additional stopping rules, which do not use the noise level. Numerical comparison of these rules and known stopping rules (L-curve rule etc) shows good performance of new rules.

2. Conjugate Gradient Type Methods

The problem (1.1) can be solved by various different iterative methods. Simple iterative methods are

$$u_n = u_{n-1} + \beta_{n-1} A^*(f - Au_{n-1}), \quad n = 1, 2, \dots, \quad \beta_{n-1} > 0; \tag{2.1}$$

$$\beta_{n-1} u_n + A^* A u_n = \beta_{n-1} u_{n-1} + A^* f, \quad n = 1, 2, \dots, \quad \beta_{n-1} > 0. \tag{2.2}$$

For the initial approximation u_0 often $u_0 = 0$ is used.

The conjugate gradient type methods converge much faster than methods (2.1), (2.2): the error $\|u_n - u_*\|$ in these methods for $n \le n_{\text{opt}}$ is close to $\|u_{n^2} - u_*\|$ in methods (2.1), (2.2). In this paper we consider two iterative methods based on conjugate gradient method for various possibilities to symmetrize the problem (1.1). Application of the conjugate gradient method to the normal equation $A^*Au = A^*f$ or to the equation $AA^*w = f$ with $u = A^*w$ gives the methods called CGLS or CGME respectively. In iterative method CGLS the kth iterate u_k minimizes the discrepancy $\|f - Au\|$ among all u from the Krylov subspace

$$\operatorname{span}\{A^*f,\ A^*AA^*f,\ \dots,\ (A^*A)^{k-1}A^*f\}$$

as in the projection method of least squares. The kth iterate u_k in method CGME minimizes the error $\|u_* - u\|$ with u in the same Krylov subspace (as in the projection method of minimal error (see [12, 20])). Both algorithms start with fixing the starting values $u_0 = 0$, $r_0 = f$, $v_{-1} = 0$. Method CGLS also takes $p_{-1} = \infty$ and computes for every $n = 0, 1, 2, \dots$

$$\begin{aligned} &p_n = A^*r_n\,, \quad \sigma_n = \|p_n\|^2/\|p_{n-1}\|^2\,, \quad v_n = r_n + \sigma_n v_{n-1}\,, \\ &q_n = A^*v_n\,, \quad s_n = Aq_n\,, \quad \beta_n = \|p_n\|^2/\|s_n\|^2\,, \\ &u_{n+1} = u_n + \beta_n q_n\,, \quad r_{n+1} = r_n - \beta_n s_n\,. \end{aligned} \tag{2.3}$$

In CGME method one takes $r_{-1} = \infty$ and computes for every $n = 0, 1, 2, \ldots$

$$\begin{aligned} &\sigma_n = \|r_n\|^2/\|r_{n-1}\|^2\,, \quad v_n = r_n + \sigma_n v_{n-1}\,, \quad q_n = A^*v_n\,, \\ &\beta_n = \|r_n\|^2/\|q_n\|^2\,, \quad u_{n+1} = u_n + \beta_n q_n\,, \quad r_{n+1} = r_n - \beta_n Aq_n\,. \end{aligned} \tag{2.4}$$

In CGLS method the discrepancy $\|r_n\|$ decreases monotonically for all n but in CGME method numerical results show that $\|r_n\|$ decreases monotonically only at some first iterates.

3. Stopping Rules Using Noise Level

First we consider the case when the exact noise level δ is known with $\|f - f_*\| \le \delta$. Then the most prominent stopping rule is the discrepancy principle: we stop at the first index $n = n_D$ for which $\|r_n\| \le C\delta$, where $r_n = f - Au_n$ and $C > 1$ is a constant. For many methods, choice of $n = n_D$ by the discrepancy principle guarantees convergence $\|u_{n_D} - u_*\| \to 0$ $(\delta \to 0)$ and in case $u_* \in \mathcal{R}((A^*A)^{p/2})$ the order optimal error estimate

$$\|u_{n_D} - u_*\| \le c\delta^{p/(p+1)}, \quad \forall p < \infty.$$

These assertions were proved in [21] for method (2.1) if $\beta_n = \beta \in (0, 2/\|A^*A\|)$ (then (2.1) is called the explicit scheme of iteration method or Landweber method) and for method (2.2) if $\beta_n = \beta > 0$ (then (2.2) is called implicit scheme of iteration method), in [13] for method (2.2) with certain various β_n, in [8, 9] for methods (2.1), (2.2) with certain other β_n, in [1, 4, 5, 10, 16, 17, 18, 19] for method (2.3). As stated by Hanke in works [10, 11], in method (2.4) the discrepancy principle fails but above-mentioned assertions about convergence and error estimates hold for the following rule: stop at the first index $n = n_{DH}$ for which $d_{DH} := \big[\sum_{i=0}^{n} \|r_i\|^{-2}\big]^{-1/2} < C\delta$ with fixed $C > 1$. In numerical experiments we used $C = 1$ for the discrepancy principle and $C = 1.2$ in stopping rule with d_{DH}.

Let us consider now the monotone error rule (ME rule). For iteration methods of the form

$$u_n = u_{n-1} + A^*z_n, \qquad n = 1, 2, \ldots \tag{3.1}$$

the ME rule is the following: choose n_{ME} as the first index n satisfying

$$d_{ME}(n) := \frac{(r_n + r_{n+1}, z_n)}{2\|z_n\|} \le \delta. \tag{3.2}$$

Theorem 1. *The ME rule has the property*

$$\|u_n - u_*\| < \|u_{n-1} - u_*\| \qquad \text{for } n = 1, 2, \ldots, n_{ME}. \tag{3.3}$$

Proof. Using (3.1) we obtain

$$\begin{aligned}\|u_n - u_*\|^2 - \|u_{n-1} - u_*\|^2 &= (u_{n-1} + u_n - 2u_*, A^* z_{n-1}) \\ &= (2(f - f_*) - (r_{n-1} + r_n), z_{n-1}) \\ &\le 2\|z_{n-1}\|\Big\{\delta - \frac{(r_{n-1} + r_n, z_{n-1})}{2\|z_{n-1}\|}\Big\}.\end{aligned}$$

■

Note that iteration methods (2.1), (2.2), (2.3), and (2.4) have form (3.1) with $z_n = \beta_n r_n$, $z_n = (\beta_n I + AA^*)^{-1} r_n$, $z_n = \beta_n v_n$, and $z_n = \beta_n v_n$, respectively. In methods (2.1), (2.2) the convergence $\|u_{n_{ME}} - u_*\| \to 0$ ($\delta \to 0$) and in case $u_* \in \mathcal{R}((A^*A)^{p/2})$ the error estimate $\|u_{n_{ME}} - u_*\| \le c\delta^{p/(p+1)}$ for all $p < \infty$ was stated in [8, 9]. For the CGLS method in [1] a rule similar to ME rule was proposed and convergence $\|u_n - u_*\| \to 0$ ($\delta \to 0$) was stated but corresponding stopping index is smaller than n_{ME}, hence due to (3.3) the ME rule is preferable.

However, iterating by formula (3.1), on different iteration steps different forms of element z_n may be used for constructing better approximation u_n than u_{n-1}, if (3.2) is fulfilled. For instance, after finding $u_{n_{ME}}$ by the CGLS method or by the CGME method this approximation can be further improved by methods (2.1), (2.2), until in these methods (3.2) is satisfied first time.

We experimented numerically also with rule, which chooses in CGLS method n_{DD} as the first $n = 1, 2, \ldots$, for which $\|r_{n+1} - r_n\| \le C\delta^{1.5}$, using the value $C = 0.8$.

Consider now the case, when noise level is known approximately: δ is given, for which it holds

$$\frac{\|f - f_*\|}{\delta} \le C, \quad \delta \to 0$$

with unknown constant C. In [7] for explicit and implicit iterative schemes the following stopping rule R was formulated.

Rule R. Let $0 \le s \le 1/2$. Find N as the first n for which

$$\varphi(n) \equiv \sqrt{n}\|A^* r_n\| \le b\delta$$

with constant b large enough. Find the stopping index n_R as the location of the global minimum of the function $t(n) = n^s\|r_n\|$ on the interval $[1, N]$.

For explicit and implicit iteration schemes in [7] convergence $\|u_{n_R} - u_*\| \to 0$ ($\delta \to 0$) was proved and error estimates (which are quasioptimal in case $\|f - f_*\| \le \delta$) were given.

In iterative method CGLS we find the stopping index n_R by an analogue of Rule R with $s \in [0, 1]$ and by replacing the function $\varphi(n)$ by function

$\sqrt{\gamma_{n+1}}\|A^*r_n\|$. Here γ_n is found iteratively as follows: starting with $\kappa_{-1} = 0$, $\gamma_0 = 0$, compute $\kappa_n = 1 + \sigma_n\kappa_{n-1}$, $\gamma_{n+1} = \gamma_n + \beta_n\kappa_n$ for every $n = 0, 1, 2, \ldots$ In numerical experiments we used $b = 0.4$ and $s = 0.2$.

4. Stopping Rules not Using Noise Level

Consider now the case when there is no information about the noise level δ. Then it is principally impossible to formulate a stopping rule with convergence property $\|u_n - u_*\| \to 0$ $(\delta \to 0)$ (see [2]). Nevertheless, various rules work well in many cases.

1. In literature much attention is paid to L-curve rule: the norm of the discrepancy $\|r_n\|$ and the norm of the iterated approximation $\|u_n\|$ are plot on log-log scale and the corner point of this L-shaped curve is found. We used the algorithm from [3], where all triangles are considered with fixed first vertice, corresponding to $n = 0$, and where two other vertices change in all possible ways. For n_L the middle vertice with minimal angle is taken.
2. In Hanke-Raus rule [14] the stopping index $n = n_{HR}$ is found as a location of the global minimum of the function $\sqrt{\gamma_{n+1}}\|r_n\|$. Note that in [14], for iteration methods in form $u_n = g_n(A^*A)A^*f$ with function $g_n(\lambda)$ approximating $1/\lambda$, the analogous rule was proposed: here the stopping index is a location of the global minimum of the function $\sqrt{g_{n+1}(0)}\|r_n\|$. In [14] for this stopping rule also error estimates are given.
3. Let us consider rules which minimize some function as Hanke-Raus rule does. For the CGME method we used rule RM as an analogue of rule R: at first N was found as global minimizer of the function $\sqrt{\gamma_{n-2}} \cdot d_{DH}(n-3)$ (we noticed that global minimizer of the function $\sqrt{\gamma_{n+1}} \cdot d_{DH}(n)$ was in most cases smaller than n_{opt}), and after that n_{RM} was found as the minimizer of $\|r_n\|$ on interval $[0, N]$.
   For the CGME method a good choice of n is also the global minimizer n_{DM} of the discrepancy function $\|r_n\|$. In both methods CGLS, CGME one may use n_{SD} as the global minimizer of the function $\|u_n\| \cdot \|r_n\|$ (S and D refer to "solution approximation" and "discrepancy"). In CGME method we used also n_{DHNM} as the global minimizer of the function $n^s d_{DH}(n)$ with $s = 0.9$ (in DHNM, DH refers to $d_{DH}(n)$, N to n, and M to "minimization").
   In numerical experiments we noticed that the maximums of $\|u_n - u_{n-1}\|$ were close to error $\|u_n - u_*\|$. It motivated us to choose n_{SDS} as global minimizer of some function approximating maximums of $\|u_n - u_{n-1}\|$. We minimized the function

$$\psi(n) = \left[\sum_{i=1}^{n} \|u_i - u_{i-1}\|^{16} i^k \Big/ \sum_{i=1}^{n} i^k\right]^{1/16}$$

   with $k = 16 + 16(k_1+1)(n_{\max} - i)^{k_1}/n_{\max}^{k_1}$. Here $n_{\max}$ is maximum number of iterations; for k_1 we took 4 in CGLS method and 3 in CGME method.

Large exponents 16 in $\psi(n)$ emphasize maximums of $\|u_i - u_{i-1}\|$. The function $\psi(n)^{16}$ is weighted average of $\|u_i - u_{i-1}\|^{16}$, $i = 1, 2, \ldots, n$, where the terms with i close to n have larger weights. Note that functions $\sqrt{\gamma_{n+1}}\, d_{DH}(n)$ and $\psi(n)$ predict well not only location of $n_{\rm opt}$ but also the behaviour of $\|u_n - u_*\|$ for all n.

4. Besides of minimization of some function one may use the observation that several monotone functions attain certain level ("plateau") around $n_{\rm opt}$ and after that do not change much. We choose n_{DHP}, n_{SHP}, and n_{DDHP} as the first n, for which the functions

$$d_{DH}(n) = \Big[\sum_{i=0}^{n} \|r_i\|^{-2}\Big]^{-1/2}, \quad \Big[\sum_{i=0}^{n} \|u_i\|^{-2}\Big]^{-1/2}, \quad \Big[\sum_{i=0}^{n-1} \|r_{i+1} - r_i\|^{-2}\Big]^{-1/2},$$

respectively, decreased in next 10 steps no more than C times. We used for these functions C values 1.5, 1.3, and 2, respectively.

5. Numerical Experiments

We solved 12 test problems, 10 of which were from [15] and the other two were slight modifications of these.

For the supposable noise level the values $\delta = 10^{-i}$ with $i = 1, \ldots, 6$ were taken and instead of the exact data f_* randomly perturbed data were used with actual noise level $\|f - f_*\| = d\delta$ where the values of d were 1 and 100.

The problems were discretized by the collocation method with 256 piecewise constant basis functions on a uniform mesh and solved by the methods CGLS and CGME 16 times. In numerical experiments we found the optimal stopping index $n_{\rm opt}$ as an index n which minimizes the error $\|u_n - u_*\|$ on the interval $[1, n_{\max}]$. We used $n_{\max} = 200$.

In Tables 1 and 2 we give for the method CGLS root-mean-squares of ratios $\|u_n - u_*\|/\|u_{n_{\rm opt}} - u_*\|$ over all 16 runs and over all δ values, for cases $d = 1$ and $d = 100$, respectively, where the stopping index n was chosen by the rule given in the first row. In Tables 3 and 4, corresponding results are given for the CGME method. Tables 2 and 4 do not contain results for n_D, n_{DH}, and n_{ME}, since these rules do not suit for the case of inexact noise level (they did not stop within 200 iterations).

In case of exactly given noise level ($d = 1$) in method CGLS all 4 rules that use the noise level hold first 4 places. In contrast to this situation, in method CGME both rules ME and DH that use noise level were surprisingly outperformed by 2 rules not using noise level.

After summarizing the cases $d = 1$ and $d = 100$, the best three rules were R, HR and SDS for CGLS method, and DHP, RM and DM for CGME method. The rules RM and DM differ only in interval for minimization of $\|r_n\|$: the intervals are $[0, N]$ and $[0, \infty)$, respectively. Tables 3 and 4 show that in most cases the results for these rules coincide but additional work done in RM for finding N is justified in some problems.

Table 1. CGLS, $d = 1$, root-mean-squares of error ratios.

Problem	D	ME	R	DD	SDS	HR	SHP	SD	L
baart	1.21	1.77	1.77	1.44	1.72	2.47	1.58	1.40	1.39
baart2	1.23	1.77	1.77	1.49	1.68	2.54	1.69	1.44	2e4
deriv2	1.07	1.16	3.73	1.97	1.50	1.71	1.33	1.56	9e3
foxgood	1.70	2.61	3.00	3.21	8.08	6.87	27.37	67.03	19.00
gravity	1.46	2.73	1.98	2.48	1.87	3.25	14.97	15.58	7.55
heat	1.12	1.36	1.13	1.27	1.38	2.06	15.15	3.37	2.67
heat2	1.15	2.02	1.27	5.60	1.46	2.77	25.98	28.73	1e3
ilaplace	1.07	1.18	1.13	1.18	1.12	1.29	1.37	1.03	1.04
phillips	1.18	1.95	1.75	1.46	1.67	3.92	7.63	30.69	9.01
shaw	1.41	2.63	2.47	2.51	2.04	2.89	3.27	1.75	1.69
spikes	1.00	1.01	1.01	1.02	1.01	1.01	1.11	1.00	1.01
wing	1.07	1.13	1.20	1.27	1.40	1.48	1.10	1.11	1e6
average	1.22	1.78	1.85	2.07	2.08	2.69	8.55	12.89	9e4

Table 2. CGLS, $d = 100$, root-mean-squares of error ratios.

Problem	R	HR	SDS	SHP	SD	DD	L
baart	1.61	2.74	1.81	1.54	1.62	2e2	8e5
baart2	1.68	2.71	1.77	1.66	1.56	99.18	1e6
deriv2	1.26	1.64	6.08	7.80	7.80	3e4	5e5
foxgood	3.22	10.59	8.42	27.25	13.35	4e2	3e5
gravity	1.63	3.53	1.73	4.15	15.40	48.42	7.23
heat	1.17	1.97	1.19	5.40	3.40	7e2	4e3
heat2	1.23	2.61	23.88	9.44	10.76	9e3	8e4
ilaplace	1.46	1.73	1.10	1.16	1.26	2e2	1.19
phillips	1.65	4.24	1.39	3.72	30.64	3e2	8.80
shaw	1.33	2.43	1.77	1.35	1.48	5.82	1.16
spikes	1.01	1.03	1.02	1.10	1.01	1.01	1.01
wing	1.11	1.36	2.20	2.28	2.28	2e4	1e8
average	1.53	3.05	4.36	5.57	7.55	5e3	1e7

For $d = 1$ the rule R gave the stopping index n_R near the end of the search interval $[1, N]$, for $d = 100$ the index n_R lies at the beginning of this interval.

The aim of the numerical experiments was comparison of parameter choice rules, not methods but some remarks about the relationship of errors in these methods can be made:

- The best error ratios in Tables 3, 4 are smaller than in Tables 1, 2 but typically error for n_{opt} is in method CGLS smaller than in method CGME. The ratios of errors for CGME and CGLS methods were in interval $[0.95, 5]$, averages of these ratios over test problems were in interval $[1.5, 2]$.

Table 3. CGME, $d = 1$, root-mean-squares of error ratios.

Problem	RM	DHP	DH	ME	DHNM	DM	SDS	HR	DDHP	L
baart	1.00	1.00	1.07	1.00	1.07	1.00	1.68	1.84	26.05	1.00
baart2	1.00	1.00	1.07	1.00	1.06	1.00	1.69	1.84	3.09	1.00
deriv2	1.01	1.03	1.05	1.07	1.91	4.67	1.21	1.36	1.20	9e3
foxgood	1.00	1.00	1.14	1.09	1.00	1.00	4.68	4.78	59.34	1.17
gravity	1.03	1.02	1.10	1.17	1.29	1.03	1.17	1.32	3.14	1.32
heat	1.04	1.29	1.13	1.14	1.09	1.04	1.36	1.44	2.26	1.13
heat2	1.03	1.00	1.13	1.24	2.31	1.03	1.22	1.30	1.35	1.57
ilaplace	1.01	1.02	1.04	1.03	1.03	1.01	1.09	1.12	1.12	1.02
phillips	1.14	1.11	1.14	1.16	1.17	1.14	1.23	1.45	3.90	3.33
shaw	1.03	1.01	1.10	1.04	1.01	1.03	1.17	1.32	2.91	1.01
spikes	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.01	1.11	1.00
wing	1.00	1.01	1.02	1.04	1.00	1.00	1.28	1.31	1.20	1.03
average	1.02	1.04	1.08	1.08	1.24	1.33	1.56	1.67	8.89	7e2

Table 4. CGME, $d = 100$, root-mean-squares of error ratios.

Problem	DHP	DM	RM	HR	DHNM	SDS	DDHP	L
baart	1.00	1.00	1.00	1.65	3.77	1.55	1.49	1.09
baart2	1.00	1.00	1.00	1.65	3.88	1.56	1.51	1.09
deriv2	1.00	1.00	1.00	1.20	2e3	1e3	1e4	1e6
foxgood	1.00	1.00	1.00	4.81	1.21	4.87	3.26	1.67
gravity	1.00	1.01	1.01	1.33	1.20	1.28	1.77	1.01
heat	1.02	1.03	1.03	1.23	28.90	1.16	32.87	2e4
heat2	1.00	1.02	1.02	1.16	1e3	7e2	2e3	1e5
ilaplace	1.00	1.01	1.01	1.08	1.18	1.18	1.03	1.02
phillips	1.00	1.11	1.11	1.57	1.15	1.31	1.82	1.27
shaw	1.01	1.02	1.02	1.34	1.00	1.37	2.90	1.17
spikes	1.00	1.00	1.00	1.02	1.00	1.02	1.11	1.00
wing	1.01	1.00	1.00	1.26	4e3	2e4	9e4	3e10
average	1.00	1.02	1.02	1.61	6e2	2e3	9e3	2e9

- The choice of constants in used rules may depend on the discretization parameter and on the number of iteration steps; the values used in this paper are representative for our set of test problems.
- The rule SDS may have reserves for improvement if we change the form of the function ψ or adjust its parameters. Also in L-curve rule there may be room for improvement. Often the L-curve rule failed in case of large δ'.

References

[1] O. M. Alifanov, E. A. Artyukhin and S. V. Rumyantsev. *Extreme Methods for Solving Ill-Posed Problems with Applications to Inverse Heat Transfer Problems*. Begell House, New York, 1995.

[2] A. B. Bakushinskii. Remarks on the choice of regularization parameter from quasioptimality and relation tests. *Zh. Vychisl. Mat. i Mat. Fiz.*, **24**(8), 1258–1259, 1984. (in Russian)

[3] J. L. Castellanos, S. Gómez and V. Guerra. The triangle method for finding the corner of the L-curve. *Applied Numerical Mathematics*, **43**, 359–373, 2002.

[4] H. W. Engl, M. Hanke and A. Neubauer. *Regularization of Inverse Problems*. Kluwer, Dordrecht, 1996.

[5] S. F. Gilyazov and N. L. Goldman. *Regularization of Ill-Posed Problems by Iteration Methods*. Kluwer, Dordrecht, 2000.

[6] U. Hämarik and R. Palm. Comparison of stopping rules in conjugate gradient type methods for solving ill-posed problems. In: *Proc. of the 10th International Conference MMA2005&CMAM2, Trakai*. Technika, 285–291, 2005.

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

[8] 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.

[9] U. Hämarik and U. Tautenhahn. On the monotone error rule for choosing the regularization parameter in ill-posed problems. In: M. M. Lavrentiev and S. I. Kabanikhin(Eds.), *Proc. of the International Conference, Samarkand, Uzbekistan 2000, Ill-Posed and Non-Classical Problems of Mathematical Physics and Analysis*, VSP, Utrecht-Boston, 27–55, 2003.

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

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

[12] M. Hanke. On Lanczos based methods for the regularization of discrete ill-posed problems. *BIT Numerical Mathematics*, **41**(5), 1008–1018, 2001.

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

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

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

[16] 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.

[17] R. Plato. The method of conjugate residuals for solving the Galerkin equations associated with symmetric positive semidefinite ill-posed problems. *SIAM J. Numer. Anal.*, **35**(4), 1621–1645, 1998.

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

[19] R. Plato and G. Vainikko. On the fast and fully discretized solution of integral and pseudo-differential equations on smooth curves. *Calcolo*, **38**(1), 25–48, 2001.

[20] G. Vainikko and U. Hämarik. Projection methods and self-regularization in ill-posed problems. *Izv. Vyssh. Uchebn. Zaved. Mat.*, **10**(3), 3–17, 1985. (in Russian)

[21] G. Vainikko and A. Veretennikov. *Iteration Procedures in Ill-Posed Problems.* Nauka, Moscow, 1986. (in Russian)