International Conference "Inverse and Ill-Posed Problems of Mathematical Physics",
dedicated to Professor M.M. Lavrent'ev on occasion of his 75-th birthday, August 20-25, 2007, Novosibirsk, Russia

Comparison of stopping rules for the conjugate gradient type methods in ill-posed problems

U. Hämarik*, R. Palm**

* Institute of Applied Mathematics,
University of Tartu,
Liivi 2, 50409 Tartu, Estonia
E-mail: uno.hamarik@ut.ee

** Institute of Computer Science,
University of Tartu,
Liivi 2, 50409 Tartu, Estonia
E-mail: reimo.palm@ut.ee

This work was supported by the Estonian Science Foundation (grant nr 5785).

We consider ill-posed problem $Au = f$, where A is a linear bounded operator between two Hilbert spaces. For solving this equation we use conjugate gradient method applied to the normal equation $A^*Au = A^*f$ or to the equation $AA^*w = f$ with $u = A^*w$ (see [4,7,11]). The corresponding methods are called CGLS and CGME, respectively. 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 $n = 0, 1, 2, \dots$

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

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

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

In iterative method CGLS the kth iterate u_k minimizes the discrepancy $\|f - Au\|$ among all u from the Krylov subspace $\mathrm{span}\{A^*f, A^*AA^*f, \dots, (A^*A)^{k-1}A^*f\}$ (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]). Here u_* is the solution of equation $Au = f$, nearest to the initial approximation u_0.

In practice often instead of the exact data f only an approximation f_δ is given (containing, for example, measurement errors). In ill-posed problems with noisy data for preventing unbounded magnification of the data error, the iterations should be stopped after a certain number n of steps. If the exact noise level δ with $\|f_\delta - f\| \le \delta$ is given, the proper choice of $n = n(\delta)$ guarantees the convergence $u_{n(\delta)} \to u_*$ as $\delta \to 0$, where u_n is an approximation found after n iterations. For many iteration methods this convergence is guaranteed by choice of n by the discrepancy principle: $n = n_D$ is the first index for which $\|r_n\| \le C\delta$, where $r_n = f - Au_n$ and $C > 1$ is a constant (see [7,11] for CGLS and [13] for other methods). For the CGME the convergence is guaranteed by the following rule of Hanke [7]: stop at the first index $n = n_{DH}$ for which $d_{DH} := \big[\sum_{i=0}^{n} \|r_i\|^{-2}\big]^{-1/2} \le C\delta$ with fixed $C > 1$. The monotone error rule (ME rule, see [4]) chooses n_{ME} in both methods as the first index n satisfying

$$d_{ME}(n) := \frac{(r_n + r_{n+1}, v_n)}{2\|v_n\|} \le \delta. \tag{3}$$

The ME rule can be formulated also for other iterative methods (see [6]), in all methods it has the property

$$\|u_n - u_*\| \le \|u_{n-1} - u_*\| \qquad \text{for } n = 1,\ 2,\ \dots,\ n_{ME}. \tag{4}$$

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 (4) the ME rule is preferable.

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 $C = 1.2$.

In many applications the noise level δ is given approximately: it holds $\|f_\delta - f\|/\delta \le C$ for $\delta \to 0$, where C is an unknown constant. In this case convergence $u_{n(\delta)} \to u_*$ ($\delta \to 0$) for iterative methods of Landweber and Lardy is guaranteed by stopping iterations by the following rule [5].

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 $[1, N]$.

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 functions $\varphi(n)$ and $t(n)$ by functions $\sqrt{\gamma_{n+1}}\,\|A^* r_n\|$ and $\gamma_{n+1}^s\|r_n\|$, respectively. 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 $b = 0.1$ and $s = 1/30$.

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 [3,4,7,8,10]. In Hanke-Raus rule [8] the stopping index $n = n_{HR}$ is found as a location of the global minimum of the function $\sqrt{\gamma_{n+1}}\,\|r_n\|$. In literature much attention is paid to L-curve rule [9]: the points $(\|r_n\|, \|u_n\|)$, $n = 0, \ldots, n_{\max}$, are plot on xy-plane and the corner point of this L-shaped curve is found. The maximum number of iterations $n_{\max}$ should be fixed earlier. We used an analogue of the algorithm from [3] with $n_{\max} = 200$, where all triangles are considered with fixed first vertice $(\|r_0\|, \|u_0\|)$ and with fixed last vertice $(\|r_{n_{\max}}\|, \|u_{n_{\max}}\|)$, and where the middle vertices are $(\|r_1\|, \|u_1\|), \ldots, (\|r_{n_{\max}-1}\|, \|u_{n_{\max}-1}\|)$. For n_L the index of the middle vertice with minimal angle is taken. Note that in [10] modification of L-curve rule was proposed using instead of n_L some neighbouring local minimizer n_{LM} of $\|u_n - u_{n-1}\|$. Due to oscillatory behaviour of the function $\|u_n - u_{n-1}\|$ difference between n_L and n_{LM} was small in our numerical experiments.

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} (the optimal n)), and after that n_{RM} was found as the minimizer of $\|r_n\|$ on interval $[1, N]$.

For the CGME method a good choice of n is also the global minimizer n_{DM} of the discrepancy function $\|r_n\|$. 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 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

$$\psi(n) = \left[\sum_{i=1}^{2n} \|u_i - u_{i-1}\|^{16} w(n,i) \Big/ \sum_{i=1}^{2n} w(n,i)\right]^{\frac{1}{16}}, \qquad w(n,i) = \left(\frac{i(2n-i)}{n^2}\right)^{\frac{200}{\sqrt{n}}}.$$

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

Besides of minimization of some function one may use the observation that several monotone functions attain certain level ("plateau") around n_{opt} and after that do not change much. Among these functions, the most useful were $d_{DH}(n)$ and $\big[\sum_{i=0}^{n} \|u_i\|^{-2}\big]^{-1/2}$. We choosed n_{DHP} and n_{SHP} as the first n for which these functions decreased in next 10 steps no more than 1.5 and 1.2 times, respectively.

We made numerical experiments with these rules [4], taking for supposable noise level the values $\delta = 10^{-i}$ with $i = 1, \ldots, 6$ and using instead of the exact data f randomly perturbed data with actual noise level

$\|f_\delta - f\| = d\delta$ where the values of d were 1 and 100. In case of exactly given noise level ($d = 1$) in method CGLS all 4 rules D (with $C = 1$), ME, R, DD that use the noise level hold first 4 places. In contrast to this situation, in method CGME both rules ME and DH (with $C = 1.2$) that use noise level were surprisingly outperformed by rules RM, DHP not using noise level.

Numerical results ordered the rules starting from the best ones as following: for CGLS in case $d = 1$ the order was D, ME, R, DD, SDS, HR, SHP, L, and in case $d = 100$ R, HR, SDS, SHP, DD, L. For the second method CGME the order was RM, DHP, DH, ME, DHNM, SDS, HR, L in case $d = 1$, and DHP, DM, RM, HR, DHNM, SDS, L in case $d = 100$. Note thad rules D, DH, and ME do not suit for the case of inexact noise level (in case $d = 100$ they did not stop within $n_{\max} = 200$ iterations).

The rules RM and DM differ only in interval for minimization of $\|r_n\|$: the intervals are $[1, N]$ and $[1, \infty)$, respectively. In most cases the results for these rules coincided but additional work done in RM for finding N was justified in some problems.

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 lied at the beginning of this interval.

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] 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.

[5] 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.

[6] 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.

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

[8] 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.

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

[10] S. Morigi, L. Reichel, F. Sgallari, and F. Zama. Iterative methods for ill-posed problems and semiconvergent sequences. *J. of Comput. and Applied Math.*, 193(1):157 – 167, 2006.

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

[12] 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).

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