MATHEMATICA MONTISNIGRI
Vol XLIII (2018)

ADAPTIVE CHEBYSHEV ITERATIVE METHOD

O.B. FEODORITOVA[1], N.D. NOVIKOVA[1] and V.T. ZHUKOV[1]

[1] Keldysh Institute of Applied Mathematics, Russian Academy of Sciences,
Moscow, Russia
Email: feodor@kiam.ru, nn@kiam.ru, zhukov@kiam.ru

Summary. We propose an adaptive Chebyshev iterative method for solving large sparse system of linear equations with a symmetric positive-definite matrix. For Chebyshev method the self-adaptive algorithm is capable of evaluating an unknown lower bound of the matrix. The adaptive procedure is based on extremal properties of Chebyshev's polynomials and allows us to achieve automatically the prescribed convergence rate of the Chebyshev method. The numerical examples show that the self-adaptive algorithm is efficient and well suited for work in parallel environment.

1 INTRODUCTION

We present an adaptive technique for the Chebyshev iteration method [1], [2] which uses to solve a large sparse system of linear equations with a symmetric positive-definite matrix. The disadvantage of this method is the need for preliminary knowledge of the lower and upper bounds of the spectrum of a matrix. Our adaptation procedure provides the estimation of an unknown lower bound of the spectrum. An estimation of upper bound is obtained by the Gershgorin theorem [3]. For the Chebyshev method we develop the self-adaptive algorithm which is capable of evaluating an unknown lower bound whenever the algorithm has not achieved optimal convergence rate with a given lower bound estimation. For the multigrid such algorithm adjusts smoothers and the coarsest level solver for achieving the prescribed rate of multigrid convergence [4]. The effective implementation of these multigrid elements with the help of the adaptive Chebyshev algorithm provides a high scalability of multigrid for massively parallel computers. In this paper, we focus on the study of the self-adaptive method for the solution of large sparse system of linear equations; the numerical examples are presented to show that this algorithm is efficient. This method is well suited for work in parallel environment, because of it does not require computing communication-intensive inner products (unlike the conjugate gradient method).

2 STATEMENT OF THE PROBLEM

Consider a large sparse system of linear equations $Au = f$, where $u, f \in \mathrm{R}^n$ are vectors and $A \in \mathrm{R}^{n\times n}$ is a symmetric matrix with eigenvalues $\lambda \in [\lambda_{min}\,; \lambda_{max}]$, $\lambda_{min} > 0$. Such systems usually arise from discretization of elliptic partial differential equations. The Chebyshev iterative method [1] can be expressed as

$$u_{k+1} = u_k + \tau_k \left(f - Au_k \right), \quad k = 0,\dots, p-1 \tag{1}$$

where u_0 is an initial guess, p is a number of iterations, τ_k, $k = 1,\dots,p$ is the optimal set of the iteration parameters. The error propagation operator $F_p\left(A\right)$ of the method (1) is defined

2010 Mathematics Subject Classification: 35J25, 47B39, 65Q10.
Key words and Phrases: Linear Systems, Chebyshev Iteration, Adaptation, Parallel Computing

by the optimal Chebyshev polynomial $F_p(\lambda)$ for the interval $[\lambda_{min}; \lambda_{max}]$. For any p the polynomial $F_p(\lambda)$ can be defined by translating the Chebyshev polynomial T_p from the interval $[\lambda_{min}; \lambda_{max}]$ to the interval $[-1; 1]$ and scaling so that its value at 0 is 1. The operator $F_p(A)$ transfers an initial guess u_0 and error $z_0 = u - u_0$ into new approximation u_p and new error $z_p = u - u_p$ correspondently due to formulas

$$u_p = F_p(A)\ u_0 + \left[I - F_p(A)\right] A^{-1} f, \qquad z_p = F_p(A)\ z_0\ ,$$

here I is identity operator. Therefore the optimal iteration process is defined by the Chebyshev polynomial $F_p(\lambda)$, which deviates least from zero on the interval $[\lambda_{min}; \lambda_{max}]$ under the condition $F_p(0) = 1$. This polynomial F_p is expressed by the standard polynomial T_p with the linear mapping

$$\lambda(x) = 0.5 \cdot ((\lambda_{min} + \lambda_{max}) - (\lambda_{max} - \lambda_{min}) \cdot x),$$

which maps the interval $[-1;1]$ into the interval $[\lambda_{min}; \lambda_{max}]$ and point $x = 1$ to point $t = \lambda_{min}$, point $\lambda_0 = 0$ maps to point $x_0 = x(\lambda_0) > 1$, where $x = x(\lambda)$ is the inverse to $\lambda(x)$.

The polynomial $T_p(x)$ is the first kind Chebyshev polynomial that deviates least from zero on the interval [–1; 1] (see [1]):

$$T_p(x) = \cos(p \arccos x), \quad |x| \le 1; \qquad T_p(x) = \cosh(p\ \mathrm{Acosh} x), \quad |x| > 1.$$

If $|x| > 1$ then the notation $T_p(x) = \cosh(p\ \mathrm{Acosh} x)$ is correct taking into account the equality $T_p(-x) = (-1)^p T_p(x)$. Denote

$$q_p = \frac{1}{T_p(x_0)} = \frac{1}{ch(p\ arch\ x_0)}\ ,$$

then $F_p(\lambda) = q_p T_p(x)$ and for value q_p there is the representation:

$$q_p = \frac{2\rho_1^p}{1 + \rho_1^{2p}}, \quad \rho_1 = \frac{1 + \sqrt{\eta}}{1 - \sqrt{\eta}}, \quad \eta = \frac{\lambda_{min}}{\lambda_{max}}\ . \tag{2}$$

A number of iterations p is determined by the condition of achieving the prescribed accuracy ε according the criterion $\| r_p \| < \varepsilon \| r_0 \|$, where r_0 and $r_p = f - A \cdot y_p$ are the initial and final residuals. An estimate for p has the form (see [1])]:

$$p = p(\varepsilon, \lambda_{min} / \lambda_{max}) = \left\lceil ln\ (\varepsilon^{-1} + \sqrt{\varepsilon^{-2} - 1}) / ln \rho_1 \right\rceil, \tag{3}$$

here $\lceil r \rceil$ denotes the smallest integer greater or equal to r and the parameter ρ_1 is defined in the formulas (2).

The optimal parameters τ_k are defined by the set of zeros of T_p:

$$\beta_k \in K_p = \left\{ cos \frac{2i-1}{2p}\pi,\ i=1,\ldots,p \right\},\quad \tau_k^{-1} = \frac{\lambda_{max}+\lambda_{min}}{2} + \frac{\lambda_{max}-\lambda_{min}}{2}\beta_k,\quad k=1,\ldots,\ p,$$

where the set K_p is ordered to ensure computational stability [1]. It is clear that $\left|T_p(x)\right| \le 1$ when $|x| \le 1$, and the function $\left|T_p(x)\right|$ is strictly monotonic outside this interval, on hyperbolic branches. The last property is important to justify the adaptation procedure.

The following recurrence relations are true

$$T_0(x) = 1;\ T_1(x) = x;\ T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x),\quad k = 2,3,\ldots .$$

These relations can be used to construct polynomials of a given degree, but we prefer to construct the required polynomial relying on the zeros of the Chebyshev polynomial.

For the successful application of the Chebyshev iteration method we need to know the bounds of the spectral interval $[\lambda_{min};\lambda_{max}]$. Usually a good estimate of the low bound is missing. We suggest an iterative refinement procedure to find automatically appropriate lower bound. There are other approaches, for example, see [5, 6].

3 ADAPTATION

Further below λ_{min}, λ_{max} mean exact (usually unknown) lower and upper bounds of the spectrum and λ^*_{min}, λ^*_{max}, are their approximate evaluations. It is known that under the conditions

$$0 < \lambda^*_{min} \le \lambda_{min} \le \lambda_{max} \le \lambda^*_{max}$$

the Chebyshev method (1) with the inexact bounds λ^*_{min}, λ^*_{max} converges, but the convergence rate can be decreased. According to [1], for the convergence of the iterations with inexact bounds the necessary and sufficient conditions are

$$\lambda_{max} < \lambda^*_{max} + \lambda^*_{min},\quad 0 < \lambda^*_{min}. \tag{4}$$

For constructing the adaptive procedure we assume that

$$0 < \lambda_{min} \le \lambda^*_{min} \le \lambda_{max} \le \lambda^*_{max}. \tag{5}$$

The upper estimate λ^*_{max} is obtained by the Gershgorin theorem [3]. The conditions (5) guarantee convergence of the iterations, since then the conditions (4) are fulfilled.

Note that one can get λ^*_{min} by applying the power method to the operator $B = I - \left(1/\lambda^*_{max}\right)A$. After finding the maximal eigenvalue μ_{max} of this operator, we obtain $\lambda^*_{min} = \left(1-\mu_{max}\right)\lambda^*_{max}$. We use another procedure, which is more efficient; and it is naturally integrated in the iterative process. More important, our procedure can be applied for finding the low bound of a linear operator A not only on the whole space R^n but as well as on some subspaces of A (for instance, on a high frequency subspace defined by coarse-grid correction in the multigrid method, or an invariant subspace which includes the exact solution and the initial guess for the solution). Note, we are developing such an adaptive procedure because in multigrid framework it can be used to implement efficient polynomial smoothers and coarsest grid solver in parallel environment.

An estimate λ^*_{min} is updated during the outer iterative process, or adaptation cycle numbered $k = 1,\dots$. To start adaptation we need to set an initial guess λ^*_{min} greater than the exact value λ_{min}, $\lambda^*_{min} \ge \lambda_{min}$. As an initial guess we can take the Rayleigh-Ritz ratio $\lambda^*_{min} = \lambda_{RR} = (Av, v)/(v, v)$ with any nonzero function v. According to the known property of this relation, the required inequality $\lambda^*_{min} \ge \lambda_{min}$ is true. A simple practical choice is $\lambda^*_{min} = \lambda_{\max}/6$.

Let us solve the linear system with the specified accuracy ε. Our adaptation algorithm is as follows. On the step k of adaptation we denote the current approximate value as λ^*_{min}, and a new approximation to the exact bound λ_{min} as λ^*_{new}. To find this new value λ^*_{new} we set much lower additional tolerance $\varepsilon_1 = \ \varepsilon$. Some typical values are $\varepsilon_1 = 10^{-2}$, $\varepsilon \ = 10^{-10}$, for example. We implement one step of the Chebyshev algorithm with specified input λ^*_{min}, λ^*_{max}, ε_1 and then compute

$$\delta = \| r_p \| / \| r_0 \|, \quad r_p = F_p(A) r_0 \ .$$

Let the value δ exceed the specified tolerance ε_1. Then the maximal eigenvalue $\beta_{max}(F_p(A))$ of the operator $F_p(A)$ achieves at the point $\lambda = \lambda_{min}$: $\beta_{max}(F_p) = F_p(\lambda_{min})$. Therefore this maximal eigenvalue can be found approximately by the known power method applied to $F_p(A)$. As iterative approximations to $\beta_{max}(F_p)$ we take the resulting ratio $\delta = \|r_p\| / \|r_0\|$ of the residual norms. If on a step k of adaptation the obtained accuracy $\delta = \delta_k$ does not exceed the specified tolerance ε_1 (it means the estimate λ^*_{min} is close to λ_{min}), we can implement next step of adaptation with the previous data λ^*_{min}, λ^*_{max}, ε_1, or can change the accuracy tolerance ε_1, taking the new value $\varepsilon_1 = \varepsilon/(\delta_1 \delta_2 \dots \delta_k)$. The last choice of the additional tolerance ε_1 leads to achieving the desired total accuracy ε in the next step of adaptation.

The new approximation λ^{new}_{min} is evidently defined as a unique root of the algebraic equation $F_p(\lambda) - \delta = 0$ on the hyperbolic monotonic branch of the Chebyshev polynomial $F_p(\lambda)$, i.e. on the interval $[0; \lambda^*_{min}]$. Finally, we get the following formulas to compute a new approximation λ^{new}_{min} to the λ_{min} using the input parameters λ^*_{min}, λ^*_{max}, ε_1 and the obtained accuracy δ:

$$\lambda^*_{new} = 0.5\lambda^*_{max} \left(1 + \eta - (1-\eta)x^*\right),$$

where

$$x^* = ch\left(\ln[\,y + \sqrt{y^2 - 1}\,]/p\right), \quad y = \delta\,\frac{1+\rho^{2p}}{2\rho^p}, \qquad \rho = \frac{1+\sqrt{\eta}}{1-\sqrt{\eta}}, \quad \eta = \frac{\lambda^*_{min}}{\lambda^*_{max}}\,.$$

If the desired accuracy ε is not achieved, then we implement next step of adaptation with the parameters λ^*_{min}, λ^*_{max}, ε_1. In favour of this version of the power method, (instead of applying it to the operator $B = I - \left(1/\lambda^*_{max}\right)A$), we mention the following important property of the optimal Chebyshev polynomials: according to [7], outside the optimal interval the Chebyshev polynomials have opposite extreme property, namely, they deviate most from zero outside the optimal interval.

Let us consider the 7-point discretization of the Laplace operator in the unit cube on a uniform cubic grid with mesh size h and zero boundary condition. The extremal eigenvalues of this operator A are known [1]:

$$\lambda_{min} = \frac{12}{h^2} sin^2 \frac{\pi h}{2} \approx 3\pi^2, \quad \lambda_{max} = \frac{12}{h^2} cos^2 \frac{\pi h}{2} \approx \frac{12}{h^2}\,.$$

The linear mapping $\lambda' = \lambda h^2$ maps the interval $\left[\lambda_{min};\, \lambda_{max}\right]$ into the interval $\left[h^2\lambda_{min};\; h^2\lambda_{max}\right] \approx [0;\, 12]$. To illustrate the process of finding the new approximate λ^{new}_{min} we take an initial guess $\lambda^*_{min} = 0.25/h^2 \approx 0.02\lambda_{max}$ and the accuracy $\varepsilon_1 = 0.25$. For these data $p = 7$ and the corresponding polynomial $F_p(\lambda)$ is shown on Fig. 1 on the interval $[0;\, 12]$.

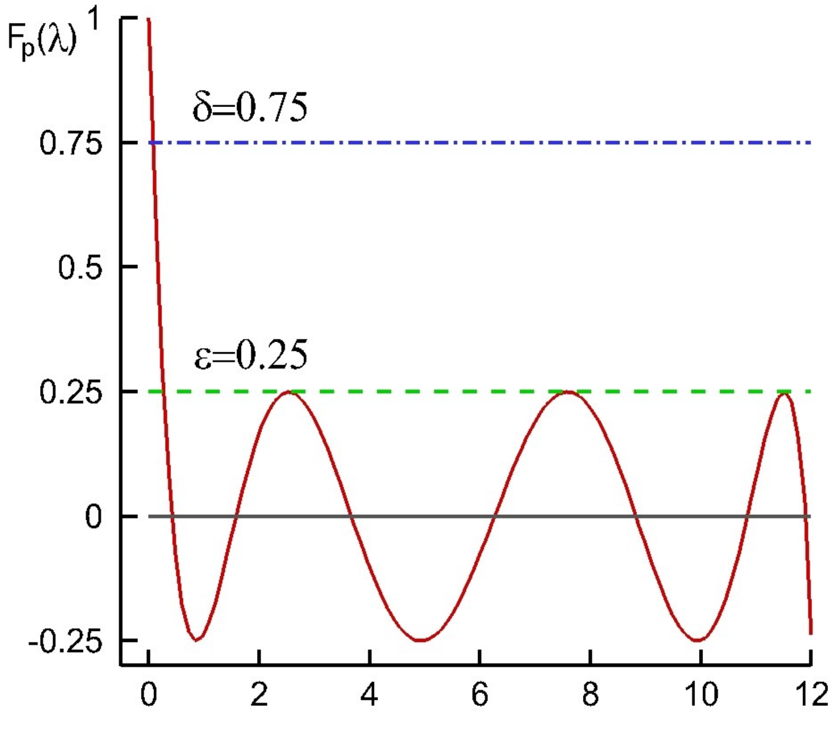

Fig. 1. Function $F_p(\lambda)$ on the interval $[0;\, 12]$

Fig. 2 shows the process of solution of the equation $F_p(\lambda)=\delta$ on the interval $[0;\ 0.75]$. The point d is the current value $d=\lambda^*_{min}$, the point c is the update value $c=\lambda^{new}_{min}$ obtaining as a root of the equation $F_p(\lambda)=\delta$.

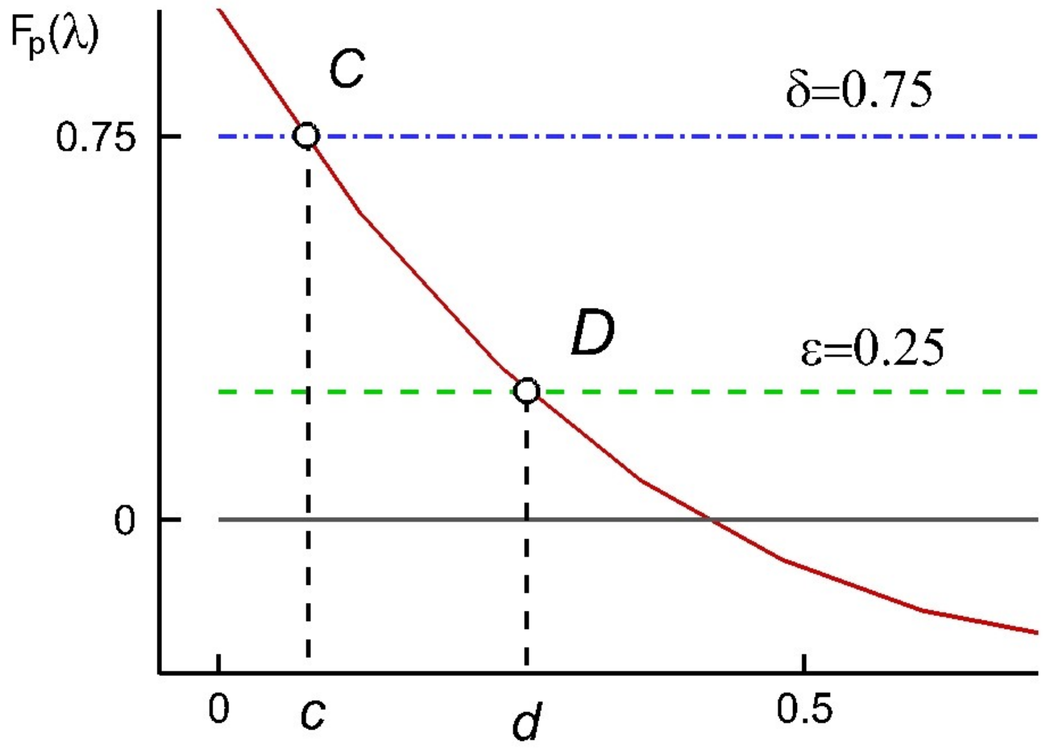

Fig. 2. Computation of the new approximate λ^{new}_{min}

4 NUMERICAL EXAMPLES

Test 1. Let us consider the 7-point discretization of the Poisson equation in the cube $[0;\ \pi]^3$ on the uniform cubic grid $N_h^3=128^3$ with mesh size $h=\pi/N_h$ and zero Dirichlet boundary condition. The extremal eigenvalues of the discrete Laplace operator are known:

$$\lambda_{min}=3, \qquad \lambda_{max}=\frac{12}{h^2}cos^2\frac{\pi h}{2}\approx\lambda^*_{max}=12\cdot N_h^2/\pi^2\approx 2\cdot 10^4.$$

Table 1 shows the characteristics of the convergence of approximations λ^*_{min} to the exact value $\lambda_1=3$ in the adaptive process with the specified parameters $\varepsilon=4\cdot10^{-8}$ and $\varepsilon_1=10^{-2}$, $\lambda^*_{min}=\lambda_{max}/6$. One can see that the convergence is fast enough, in spite of the rough initial guess λ^*_{min}. The total number of the iterations is obtained by summing the second column of Table 1, and it is equal to 818. The Chebyshev method with the tolerance $\varepsilon=4\cdot10^{-8}$ and the exact bounds $\lambda_{min}=3$, $\lambda_{max}\approx 2\cdot 10^4$ requires 722 iterations.

Step of adaptation	Number of iterationss	Achieved accuracy δ	λ^*_{min}
1	7	0.210	3307.007
2	10	0.452	1532.265
3	19	0.385	405.1740
4	33	0.398	129.7234
5	59	0.363	40.92577
6	100	0.321	14.03311
8	162	0.152	5.361031
9	212	0.016	3.126278
10	216	0.0099	3.000035

Table 1. Adaptation for Test 1

Test 2. Let us consider the 7-point discretization of the Poisson equation $\Delta u = 4$ in the domain $[-0.25;\ 1.25]\times[0;\ 1]\times[0;\ 1]$. We define the right hand side and the Dirichlet boundary condition by the exact solution $u(x,y,z) = x^2 + y^2$. The minimal eigenvalue of the discrete operator is $\lambda_{min} \approx 24$. On the sequence of the uniform grids with the number of nodes $N_h^3 = 16^3,\ 32^3,\ 64^3,\ 128^3$ for $\varepsilon = 10^{-12}$, $\varepsilon_1 = 10^{-2}$ and the initial guess $\lambda^*_{min} = \lambda_{max}/6$ we obtain almost the exact value λ_{min} in the adaptation process, see Fig. 3.

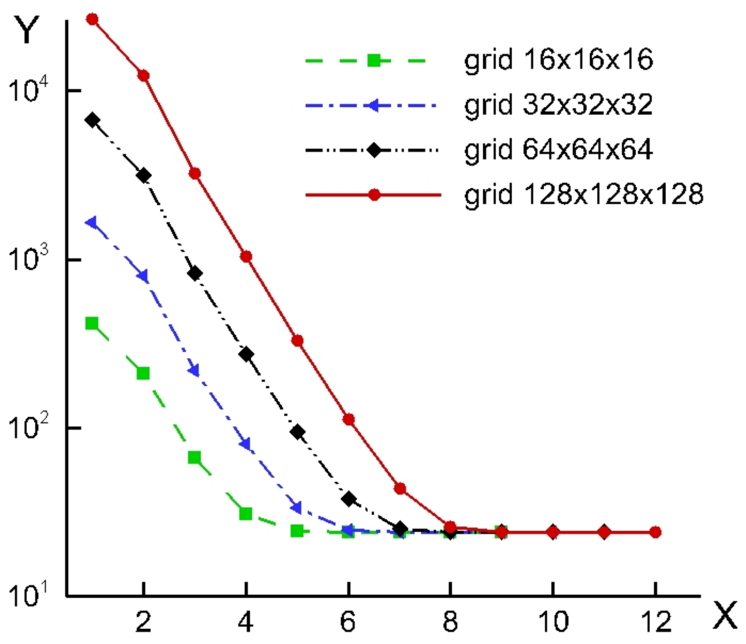

Fig. 3. Adaptation process for Test 2

The X-axis labels are the number of the adaptation steps; the Y-axis labels indicate the estimate of the lower bound λ^*_{min} on each adaptation step. Each grid with $N_h^3 = 16^3,\ 32^3$ nodes requires 6 adaptation steps; each grid with $N_h^3 = 64^3,\ 128^3$ nodes requires 8 steps to obtain well enough lower bound estimation.

Test 3. This test problem is derived by discretizing the elliptic partial differential equation with anisotropic discontinuous coefficients: $div(k\,grad\,u) = f$, see [8]. This equation is solved in the region $\Omega \equiv [-0.25;\ 1.25]\times[0;\ 1]\times[0;\ 1]$. This region is divided into four sub-regions $\Omega_i,\ i = 1,\ldots,4$ by the following way:

$$\Omega_1 = \{(x,y,z)\in\Omega:\ y \le 0.5,\ z \le 0.5\},\quad \Omega_2 = \{(x,y,z)\in\Omega : y > 0.5,\ z \le 0.5\},$$
$$\Omega_3 = \{(x,y,z)\in\Omega : y > 0.5,\ z > 0.5\},\quad \Omega_4 = \{(x,y,z)\in\Omega : y \le 0.5, z > 0.5\}.$$

The diffusion tensor is diagonal, $k(x,y,z) = diag\{k_x^i, k_y^i, k_z^i\}$. It has a discontinuity on the internal interfaces and is continuous each subregion $i = 1,...,4$:

$$k_x^1 = k_x^2 = k_x^3 = k_x^4 = 1,$$
$$k_y^1 = 10,\quad k_y^2 = 0.1,\quad k_y^3 = 0.01,\quad k_y^4 = 100,$$
$$k_z^1 = 0.01,\ k_z^2 = 100,\ k_z^3 = 10,\quad k_z^4 = 0.1.$$

The exact solution in each subdomain $i = 1,...,4$ has the form $u(x,y,z) = \alpha_i \sin(2\pi x)\sin(2\pi y)\sin(2\pi z)$; the continuity of the solution and the flux provides a choice $\alpha_1 = 0.1,\ \alpha_2 = 10,\ \alpha_3 = 100,\ \alpha_4 = 0.01$.

We set the homogeneous Neumann condition on two sides $x = -0.25,\ \ x = 1.25$ of the region and the Dirichlet boundary conditions on other sides. The right-hand side of the equation and boundary data are determined from the exact solution. Note that a rough estimate of the lower bound of the spectrum of the operator has the form $\lambda^*_{min} \approx 8\left(\min k_y(x,y,z)/l_y^2 + \min k_z(x,y,z)/l_z^2\right) = 0.16$.

Fig. 4 shows the adaptation process for the parameters $\varepsilon = 10^{-12},\ \ \varepsilon_1 = 10^{-2}$ on the sequence of the grids with $N_h^3 = 16^3,\ 32^3,\ 64^3,\ 128^3$ nodes. The initial guess is $\lambda^*_{min} = \lambda^*_{\max}/6$ on each grid. The X-axis labels are the number of the adaptation steps; the Y-axis labels indicate the estimate of the estimate λ^*_{min} on each adaptation step. For reasons of clarity the Y-axes is limited by the value 10^4. The grids $N_h^3 = 16^3,\ 32^3$ require 9 adaptation steps; the grids $N_h^3 = 64^3,\ 128^3$ require 10 adaptation steps. One can see the estimate $\lambda^*_{\min}$ begins to stabilize after 6 adaptation steps on the values from $\lambda^*_{\min} \approx 207$ to $\lambda^*_{\min} \approx 290$, depending on the grid. Such a result is satisfactory, since the variable coefficients of the test problem leads to differences in the discrete approximations. In this adaptation process at the same time we solve the linear system with the tolerance $\varepsilon = 10^{-12}$. The standard Chebyshev method with the rough estimate $\lambda^*_{min} = 0.16$ requires computational cost in many $(20-25)$ times more.

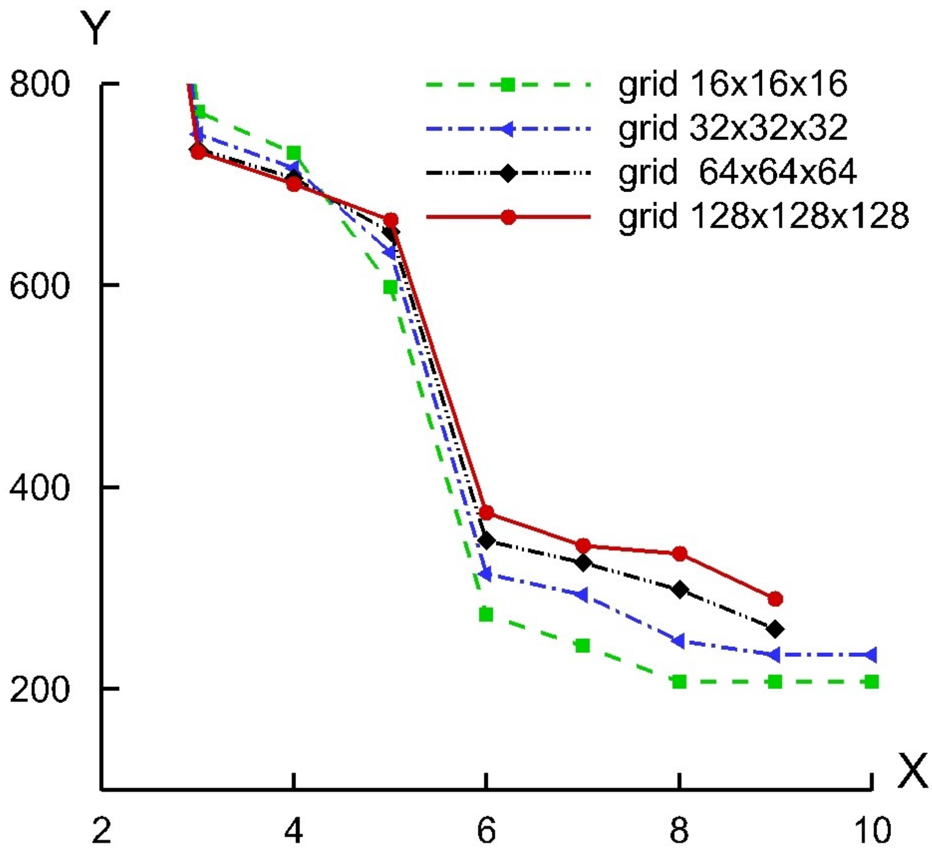

Fig. 4. Adaptation process for Test 3

One can see that for high accuracy, $\varepsilon \approx 10^{-7} - 10^{-12}$, the computational costs of the proposed adaptive Chebyshev algorithm exceed the ideal case of the Chebyshev method with the exact bounds on 15% – 40%. Extra cost depends on the number of grid nodes, and then the larger grid, the smaller extra cost.

5 CONCLUSION

We present the adaptive procedure for the Chebyshev iterative method. In some cases the adaptive algorithm may be preferable to traditional approaches. It is worth mentioning the problem of solving a very large linear system in parallel environment, application in multigrid methods on smoothing stages and for solving coarsest grid equations. The adaptive Chebyshev algorithm is easy to implement and integrate within existing computer codes, it performs identically on serial and parallel computers, and it requires only well-executed matrix-vector multiplications. Numerical experiments are performed for verification of the adaptive technique and they confirmed efficiency of the proposed adaptive approach.

Acknowledgments: This work was supported by the Russian Science Foundation, project № 17-71-30014

REFERENCES

[1] A.N. Samarskii and E.S. Nikolayev, *Solution Methods for Grid Equations, Vol. II: Iterative Methods*, Birkhauser, Basel, (1989).
[2] L.A. Hageman and D.M. Young, *Applied Iterative Methods*, Academic Press, New York-London, (1981).
[3] F.R. Gantmacher, *Theory of Matrices*, Chelsea, New York, (1959).
[4] O.B. Feodoritova, N.D. Novikova, V.T. Zhukov. "Multigrid method for diffusion equations based on adaptive smoothing", *Mathematica Montisnigri*, **36,** 14–26, (2016).
[5] M.V. Popov, Yu.A. Poveschenko, I.V. Popov, V.A. Gasilov, A.V. Koldoba, "Algorithm of competing processes for the Richardson iteration method with the Chebyshev parameters", XIV International Seminar Mathematical Model & Modeling in Laser-Plasma Processes &Advanced Science Technologies. Program and Abstracts. Moscow, Russia, 71, (2016).
[6] D. Calvetti, G.H. Golub, L. Reichel, "Adaptive Chebyshev iterative methods for non-symmetric linear systems based on modified moments", *Numerische Mathematik*, **67**, 1, (1994).
[7] P.K. Suetin, *Klassicheskie ortogonalnye mnogochleny*, Nauka, Moscow, (1979).
[8] R. Eymard, G. Henry, R. Herbin, F. Hubert, R. Klofkorn, G. Manzini, "3D Benchmark on Discretization Schemes for Anisotropic Diffusion Problems on General Grids", HAL Id: hal-00580549 https://hal.archives-ouvertes.fr/hal-0058054, (2011).

The results were presented at the 17-th International seminar "Mathematical models and modeling in laser-plasma processes and advanced science technologies" (May 26 – June 2, 2018, Budva, Montenegro).

Received September 1, 2018