ISSN 0494-7304 0132-053x

TARTU ÜLIKOOLI
TOIMETISED

УЧЕНЫЕ ЗАПИСКИ ТАРТУСКОГО УНИВЕРСИТЕТА

ACTA ET COMMENTATIONES UNIVERSITATIS TARTUENSIS

863

ПРОЕКЦИОННЫЕ МЕТОДЫ В ЗАДАЧАХ МАТЕМАТИЧЕСКОЙ ФИЗИКИ

PROJECTION METHODS FOR PROBLEMS OF MATHEMATICAL PHYSICS

Matemaatika- ja mehhaanika-alaseid töid

Труды по математике и механике

TARTU 1989

TARTU ÜLIKOOLI TOIMETISED
УЧЕНЫЕ ЗАПИСКИ ТАРТУСКОГО УНИВЕРСИТЕТА
ACTA ET COMMENTATIONES UNIVERSITATIS TARTUENSIS
Alustatud 1893.a. VIHIK 863 ВЫПУСК Основаны в 1893.г.

ПРОЕКЦИОННЫЕ МЕТОДЫ В ЗАДАЧАХ МАТЕМАТИЧЕСКОЙ ФИЗИКИ

PROJECTION METHODS FOR PROBLEMS OF MATHEMATICAL PHYSICS

Matemaatika- ja mehhaanika-alaseid töid
Труды по математике и механике

TARTU 1989

Ученые записки Тартуского университета.
Выпуск 863.
ПРОЕКЦИОННЫЕ МЕТОДЫ В ЗАДАЧАХ МАТЕМАТИЧЕСКОЙ ФИЗИКИ.
Труды по математике и механике.
На русском и английском языках.
Резюме на английском и русском языках.
Корректоры М.Тамм и П.Тороп.
Тартуский университет.
ЭССР, 202400, г.Тарту, ул.Юликооли, 18.
Ответственный редактор П.Уба.
Подписано к печати 29.08.89.
МВ 01599
Формат 60x90/16.
Бумага писчая.
Машинопись. Ротапринт.
Учетно-издательских листов 5,09. Печатных листов 6,25.
Тираж 350.
Заказ № 516.
Цена 1 руб.
Типография ТГУ, ЭССР, 202400, г.Тарту, ул. Тийги, 78.
2 - 2

Acta et commentationes universitatis Tartuensis, 863, 1989, 3-18.

ON THE REGULARIZATION OF THE RITZ-GALERKIN METHOD FOR SOLVING ILL-POSED PROBLEMS

Robert Plato (Berlin) and Gennadi Vainikko (Tartu)

In this paper we consider a class of regularization methods for a discretized version of an operator equation (which includes the case that the problem is ill-posed) with approximately given right-hand side. We propose an a priori- as well as an a posteriori-parameter choice method which is similar to the discrepancy principle of Ivanov-Morozov. From the results on fractional powers of self-adjoint operators we obtain convergence rates which are (in many cases) the same for both parameter choices.

1 Introduction

Let X be a Hilbert space and $A \in L_b(X)$, i.e. $A : X \to X$ is a bounded linear operator. We suppose $A = A^* \geq 0$ and consider the equation

$$Ax = y, \quad y \in R(A). \tag{1.1}$$

We assume that only an approximation $y_\epsilon \in Y$ to y is available with $\|y - y_\epsilon\| \leq \epsilon$, where $\epsilon > 0$ is a known error bound. To get an approximation to a solution of (1.1) we have to discretize the problem. For $h > 0$, let P_h be an orthogonal projection in X. In the following we will assume that

$$\|A(I - P_h)\| \to 0 \quad (h \to 0). \tag{1.2}$$

If $R(P_h)$ is finite-dimensional which is the most interesting case, it is sufficient and necessary for condition (1.2) to hold that A is a compact operator and that $P_h \to I$ $(h \to 0)$ pointwise on $N(A)^\perp$. If R(A) is non-closed, we have to use a regularization method. For example, we may choose h in dependence of ϵ and take the solution $x_h \in R(P_h)$ of the equation $P_h A P_h x_h = P_h y_\epsilon$ (we assume for the moment that it exists and is unique) as approximation to the solution x_* of (1.1) (see [3], [7], [9], [17]). Another, more favorable way is the use of regularization methods which are generated by Borel measurable functions

$$g_r : [0, a] \to \mathbb{R},$$

$r \geq 0, \|A\| \leq a$. We assume that the functions g_r satisfy the following conditions:

$$\sup_{0\leq t\leq a} t^p|1 - tg_r(t)| \leq \gamma_p r^{-p}, \quad r > 0, \quad 0 \leq p \leq p_0, \tag{1.3}$$

$$\sup_{0\leq t\leq a} |g_r(t)| \leq \gamma r, \quad r \geq 0, \tag{1.4}$$

where $p_0 > 0, \gamma_p$ and γ are constants. Let $x_0 \in X$ be an initial approximation. Then, with an appropriate parameter choice r (see Section 3), an approximation to the solution x_* of (1.1) which is nearest to x_0, i.e. $Ax_* = y, x_* - x_0 \in N(A)^\perp$, is given by $R_{h,r}y_\epsilon$ with

$$R_{h,r} = (I - g_r(A_h)A_h)P_h x_0 + g_r(A_h)P_h, \tag{1.5}$$

where $A_h = P_h A P_h : X \to X$. In Section 3 we propose an a priori parameter selection as well as an a posteriori parameter selection which is comparable to the discrepancy method of Ivanov-Morozov ([4], [8]): for fixed h choose r such that $\|(I - A_h R_{h,r})P_h y_\epsilon\| = d\epsilon$ with some constant $d > 1$.

In Section 2 we present some examples and in Section 4 some auxiliary results on fractional powers of selfadjoint positive operators which will be needed in the main Section 3, are given. Finally, in Section 5 some results on the regularization of projection methods for solving ill-posed problems involving non-selfadjoint operators are presented.

2 Examples and illustrations

2.1 Examples

1. The method of Lavrentiev: We have to determine the solution $x_{h,r}$ of the equation

$$(A_h + r^{-1}I)x = P_h y_\epsilon. \tag{2.1}$$

The method is of the form (1.5) with $x_0 = 0$ and $g_r(t) = (t + r^{-1})^{-1}$, $r > 0$. Conditions (1.3) and (1.4) hold with $p_0 = 1$.

2. The generalized method of Lavrentiev: Let $q \geq 0$. We have to determine the solution $x_{h,r}$ of the equation

$$(A_h^{q+1} + r^{-q-1}I)x = A_h^q P_h y_\epsilon. \tag{2.2}$$

The method is of the form (1.5) with $x_0 = 0$ and $g_r(t) = t^q/(t^{q+1} + r^{-q-1})$, $r > 0$. Conditions (1.3) and (1.4) hold with $p_0 = q + 1$.

3. The method of successive approximation (explicit scheme): Let $0 < \mu < \frac{2}{a}$. The algorithm which is due to Landweber is given by $x_{h,0} = P_h x_0$,

$$x_{h,r} = (I - \mu A_h)x_{h,r-1} + \mu P_h y_\epsilon, \quad r = 1, 2, ... \tag{2.3}$$

We have $R_{h,r}y_\epsilon = x_{h,r}$, where $R_{h,r}$ is of the class (1.5) with $g_r(t) = \frac{1}{t}[1 - (1 - \mu t)^r]$, $t \neq 0$. Conditions (1.3) and (1.4) hold for any $p_0 > 0$.

4. Implicit scheme: Let $0 < \mu$ be constant. The algorithm which is due to Fakeev-Lardy is given by $x_{h,0} = P_h x_0$,

$$(A_h + \mu I)x_{h,r} = \mu x_{h,r-1} + P_h y_\epsilon, \quad r = 1, 2, ... \tag{2.4}$$

We have $R_{h,r}y_\epsilon = x_{h,r}$, where $R_{h,r}$ is of the class (1.5) with $g_r(t) = \frac{1}{t}[1 - (\frac{\mu}{\mu+t})^r]$, $t \neq 0$. Conditions (1.3) and (1.4) hold for any $p_0 > 0$.

Other examples of regularization methods can be found in [15], [16] or [18]. Now let $\Phi_1, \Phi_2, ..., \Phi_n \in X$ be a basis of $R(P_h)$. We define

$$\begin{aligned} G_\Phi &= (<\Phi_i, \Phi_j>)_{i,j=1,...,n}, \\ B &= (<\Phi_i, A\Phi_j>)_{i,j=1,...,n}, \\ z &= (<\Phi_i, y_\epsilon>)_{i=1,...,n}. \end{aligned} \tag{2.5}$$

Let us first consider the method of Lavrentiev. The solution $x_{h,r}$ of the corresponding equation (2.1) can be expressed in the form

$$x_{h,r} = \sum_{j=1}^{n} c_j \Phi_j.$$

Some calculation shows that (2.1) is equivalent to the following system of equations for determining $c = (c_j)_{j=1,...,n}$:

$$(B + r^{-1}G_\Phi)c = z. \tag{2.6}$$

Concerning the method of successive approximation (2.3) and the implicit scheme (2.4) $x_{h,r}$ may be expressed again in the form

$$x_{h,r} = \sum_{j=1}^{n} c_j^r \Phi_j.$$

Some calculation shows that (2.3) resp. (2.4) is equivalent to the following sequence of systems of equations (2.7) resp. (2.8) for determining $c^r = (c_j^r)_{j=1,...,n}$, $r = 1, 2, ...$:

$$c^r = c^{r-1} - \mu(G_\Phi^{-1}Bc^{r-1} - G_\Phi^{-1}z), \tag{2.7}$$

$$(G_\Phi^{-1}B + \mu I_n)c^r = \mu c^{r-1} + G_\Phi^{-1}z, \tag{2.8}$$

where I_n is the identity matrix of order n. In both cases we have

$$c^0 = G_\Phi^{-1}((<\Phi_j, x_0>)_{j=1,\dots,n}).$$

2.2 Illustrations

Let us illustrate the method of Lavrentiev. We consider an integral equation

$$Ax(t) := \int_a^b k(t,s)x(s)\,ds = y(t) \quad (a \le t \le b).$$

The underlying space and the projections are chosen as follows.

$$X = L_2([a,b]), \quad <f,g>_X = \int_a^b f(s)g(s)\,ds.$$

We suppose that k(t,s) = k(s,t) for t, s $\in$ [a,b] and that

$$\int_a^b \int_a^b k(t,s)f(s)f(t)\,ds\,dt \ge 0 \quad \forall f \in X.$$

Let $\Phi_1, \Phi_2, \dots, \Phi_n \in X$ be piecewise constant functions such that $\Phi_j(s) = 1$ if $s \in [s_{j-1}, s_j]$ and Φ_j vanishes outside $[s_{j-1}, s_j]$. Here $s_j = a + jh$, $j = 0, \dots, n$ $(h = (b-a)/n)$. For this example and a sufficient smooth kernel k we have

$$\|A(I-P_h)\| = \|(I-P_h)A\| \le c_1 h, \quad c_1 = \frac{1}{\pi}\left(\int_a^b \int_a^b \left|\frac{\partial k(t,s)}{\partial t}\right|^2 dt\,ds\right)^{\frac{1}{2}}.$$

Here the matrices and vectors in (2.5) are $G_\Phi = hI_n, B = (b_{ij}), z = (z_i)_{i=1,\dots,n}$, where

$$\begin{aligned} b_{ij} &= \int_{s_{i-1}}^{s_i} \int_{s_{j-1}}^{s_j} k(t,s)\,ds\,dt \quad (i,j = 1,\dots,n), \\ z_i &= \int_{s_{i-1}}^{s_i} y_\epsilon(t)\,dt \quad (i = 1,\dots,n). \end{aligned}$$

With these functions the method of Lavrentiev (2.6) takes the form

$$\left(\frac{1}{h}B + r^{-1}I_n\right)c = \frac{1}{h}z. \tag{2.9}$$

Note that one has to compute the vectors and matrices exactly to stay within our theory which will be developed in the next section. If we use approximations $b_{ij} \approx h^2 k_{ij}$ with $k_{ij} = k(s_i - \frac{h}{2}, s_j - \frac{h}{2})$ and $z_i \approx hy_\epsilon(s_i - \frac{h}{2})$, then the system of equations (2.9) is approximated by

$$h\sum_{j=1}^n k_{ij}c_j + r^{-1}c_i = y_\epsilon(s_i - \frac{h}{2}), \quad i = 1,\dots,n.$$

We obtain the same system of equations using the rectangular rule for numerical quadrature and a collocation method for solving the Tikhonov equation $(A + r^{-1}I)x = y_\epsilon$, i.e.

$$\int_a^b k(t,s)x(s)\,ds + r^{-1}x(t) = y_\epsilon(t) \quad (a \le t \le b).$$

3 Parameter choice

In this section some appropriate parameter selection rules are introduced. We use some results on fractional powers of operaters which are stated in Section 4. We start with the case that the parameter r is selected a priori.

3.1 A priori parameter choice

Theorem 3.1 *Let $A \in L_b(X), A = A^* \geq 0, \|A\| \leq a$. Let $y \in R(A), \|y - y_\epsilon\| \leq \epsilon, x_0 \in X$ and x_* the solution of $Ax = y$, which is nearest to x_0. Let $P_h \in L_b(X)$ be an orthogonal projection and $A_h = P_h A P_h$. Suppose that $\|A(I - P_h)\| \leq \xi_h$. Assume that conditions (1.3) - (1.5) hold.*
1. Suppose that $P_h \to I$ pointwise, $\xi_h \to 0 \quad (h \to 0)$. If

$$r_{(h,\epsilon)}\epsilon \to 0, r_{(h,\epsilon)}\xi_h \leq C \text{ and } r_{(h,\epsilon)} \to \infty \quad (h \to 0, \epsilon \to 0),$$

then $R_{h,r_{(h,\epsilon)}} y_\epsilon \to x_* \ (h \to 0, \epsilon \to 0)$.
2. If $0 < p \leq p_0, x_* - x_0 = A^p z, \|z\| \leq \rho, x_* = A^p v, \|v\| \leq \rho$, and

$$C_1((\frac{\epsilon}{\rho})^{\frac{1}{p+1}} + \xi_h) \leq r^{-1} \leq C_2((\frac{\epsilon}{\rho})^{\frac{1}{p+1}} + \xi_h^{min\{\frac{1}{p},1\}})$$

with some positive constants C_1, C_2, then

$$\|x_* - R_{h,r} y_\epsilon\| \leq e_p((\rho\epsilon^p)^{\frac{1}{p+1}} + \rho\xi_h^{min\{p,1\}}).$$

e_p is independent of ϵ, h and ρ. $p \to e_p$ is bounded in (0,q] for each $q > 0$.

Proof. Let

$$S_{h,r} = I - g_r(A_h)A_h. \tag{3.1}$$

Then we have (see (1.5) for the definition of $R_{h,r}$)

$$P_h x_* - R_{h,r} y_\epsilon = S_{h,r} P_h (x_* - x_0) + g_r(A_h) P_h (A P_h x_* - y_\epsilon). \tag{3.2}$$

From (1.4) we know $\|g_r(A_h)\| \leq \gamma r$, from (3.2) we obtain

$$\|P_h x_* - R_{h,r} y_\epsilon\| \leq \|S_{h,r} P_h (x_* - x_0)\| + \gamma r (\xi_h \|(I - P_h) x_*\| + \epsilon). \tag{3.3}$$

The following inequality which follows from (1.3) will also be useful:

$$\|S_{h,r} A_h^p\| \leq \gamma_p r^{-p}, \quad 0 < p \leq p_0, r > 0. \tag{3.4}$$

1. To prove the first part of the theorem, it suffices to show the convergence of the right-hand side of (3.3) with the above given parameter choice. (3.4) shows that $\|S_{h,r} P_h A\| \to 0 \ (r \to \infty, h \to 0)$

and $\|S_{h,r}P_h\| \leq \gamma_0$. Using the Banach-Steinhaus theorem we get $\|S_{h,r}P_h(x_* - x_0)\| \to 0$ $(r \to \infty, h \to 0)$, since $x_* - x_0 \in N(A)^\perp = \overline{R(A)}$, and convergence of the first term of the right-hand side of (3.3) with the parameter choice which is stated in assertion 1 of the theorem is proved. The convergence of the second term of the right-hand side of (3.3) is an easy consequence of the choice of $r_{(h,\epsilon)}$.

2. Suppose that $x_* - x_0 = A^p z, \|z\| \leq \rho, x_* = A^p v, \|v\| \leq \rho$. By (3.3), (3.4), Corollary 4.2 and Lemma 4.3 (see Section 4) we have

$$
\begin{aligned}
\|P_h x_* - R_{h,r} y_\epsilon\| &\leq \|S_{h,r} P_h A^p\| \rho + \gamma r (c_p \rho \xi_h^{min\{p,1\}+1} + \epsilon) \\
&\leq (\|S_{h,r} A_h^p\| + \gamma_0 \|A^p - A_h^p\|) \rho + \gamma r (c_p \rho \xi_h^{min\{p,1\}+1} + \epsilon) \\
&\leq (\gamma_p r^{-p} + \gamma_0 b_p \xi_h^{min\{p,1\}}) \rho + \gamma r (c_p \rho \xi_h^{min\{p,1\}+1} + \epsilon). \qquad (3.5)
\end{aligned}
$$

Now the second part of the theorem is an easy consequence of the choice of r. □

Remark An appropriate discretization in dependence of the known error level ϵ of the data has to be chosen. If there is no information on the "smoothness" of x_* and $x_0 - x_*$, a choice of h such that $\xi_h \approx \epsilon$ is reasonable. Otherwise, a choice of h, such that $\xi_h^{min\{p,\frac{1}{p}\}+1} \approx \epsilon/\rho$ is appropriate, if $p \leq p_0$.

3.2 A posteriori parameter choice

We propose the following discrepancy principles. Let the conditions of Theorem 3.1 hold.

Rule 1. Let $1 < d_1 \leq d_2$.

1. If $\|A_h x_0 - P_h y_\epsilon\| \leq d_1 \epsilon$, then choose $r = 0$, i.e. take $P_h x_0$ as approximation.
2. If $\|A_h x_0 - P_h y_\epsilon\| > d_1 \epsilon$, then:
   a) Choose $0 < r \leq \xi_h^{-1} =: r_{max}$, such that

$$
\begin{aligned}
d_1 \epsilon &\leq \|(I - A_h R_{h,r}) P_h y_\epsilon\|, \qquad (3.6) \\
d_2 \epsilon &\geq \|(I - A_h R_{h,r}) P_h y_\epsilon\|. \qquad (3.7)
\end{aligned}
$$

   b) If there is no $r \leq r_{max}$, such that (3.7) holds, choose $r = r_{max}$.

Suppose that $r \to |1 - t g_r(t)|$ is decreasing for any $t \geq 0$. Then $r \to \|(I - A_h R_{h,r}) P_h y_\epsilon\|$ is a decreasing function. This can be shown by spectral methods. From (1.4) we know that $\|(I - A_h R_{h,0}) P_h y_\epsilon\| = \|A_h x_0 - P_h y_\epsilon\|$. Further, if $r \to |1 - t g_r(t)|$ is continuous, then $r \to$

$\|(I - A_h R_{h,r}) P_h y_\epsilon\|$ is a continuous function. Then our parameter selection rule is practicable.
The next parameter selection rule may be applied to iteration procedures. The parameter r may be restricted to the set of nonnegative integers.

Rule 2. Let $1 < d, 0 < \theta < 1$.
1. If $\|A_h x_0 - P_h y_\epsilon\| \leq d\epsilon$, then choose $r = 0$, i.e. take $P_h x_0$ as approximation.
2. If $\|A_h x_0 - P_h y_\epsilon\| > d\epsilon$, then:
a) Choose $0 < r \leq \xi_h^{-1} =: r_{max}$, such that there is a $s \in [\theta r, r]$ with

$$d\epsilon \leq \|(I - A_h R_{h,s}) P_h y_\epsilon\|, \tag{3.8}$$
$$d\epsilon \geq \|(I - A_h R_{h,r}) P_h y_\epsilon\|. \tag{3.9}$$

b) If there is no $r \leq r_{max}$, such that (3.9) holds, choose $r = r_{max}$ or $r = [r_{max}] + 1$. [x] denote the largest integer not greater than x.

To prove convergence rates for these parameter selection rules, we need the following lemma which assertion is based on an idea of T. Raus ([13], [14]).

Lemma 3.2 *Let $g_r : [0, a] \to \mathbb{R}$ be functions, such that (1.3) and (1.4) hold. Then there is a $\kappa \geq 0$ such that for $0 \leq r \leq r_1$ and $0 \leq t \leq a$ we have*

$$(1 - t g_r(t))^2 \leq \kappa((1 - t g_{r_1}(t))^2 + (r_1 t (1 - t g_r(t)))^2).$$

Proof. By (1.4) we know that $|1 - t g_{r_1}(t)| \geq 1 - t|g_{r_1}(t)| \geq 1 - \gamma t r_1$. Hence $|1 - t g_r(t)| \leq \gamma_0 \leq 2\gamma_0 |1 - t g_{r_1}(t)|$, if $r_1 t \leq (2\gamma)^{-1}$. Otherwise, if $r_1 t > (2\gamma)^{-1}$, then $|1 - t g_r(t)| \leq 2\gamma r_1 t |1 - t g_r(t)|$. □

Theorem 3.3 *Let $A \in L_b(X)$, $A = A^* \geq 0$, $\|A\| \leq a$, $y \in R(A)$, $\|y - y_\epsilon\| \leq \epsilon, x_0 \in X$ and x_* the solution of $Ax = y$ which is nearest to x_0. Let $P_h \in L_b(X)$ be an orthogonal projection and $A_h = P_h A P_h$. Suppose that $\|A(I - P_h)\| \leq \xi_h$. Suppose that conditions (1.3) – (1.5) hold for functions g_r with $p_0 > 1$ in (1.3) and that $r \to |1 - t g_r(t)|$ is decreasing for any $t \geq 0$. Let the parameter $r = r(h, \epsilon)$ be chosen according to Rule 1 or Rule 2.*
1. If $P_h \to I$ pointwise, $\xi_h \to 0$ $(h \to 0)$, then

$$R_{h,r} y_\epsilon \to x_* \ (h \to 0, \epsilon \to 0).$$

2. If $0 < p \leq p_0 - 1$, $x_* - x_0 = A^p z, \|z\| \leq \rho$ and $x_* = A^p v, \|v\| \leq \rho$, then

$$\|x_* - R_{h,r} y_\epsilon\| \leq e_p((\rho\epsilon^p)^{\frac{1}{p+1}} + \rho\xi_h^{min\{p,1\}}).$$

e_p is independent of ϵ, h and ρ.

Proof. First of all we introduce three important inequalities. Applying A_h to (3.2) leads to

$$A_h S_{h,r}(x_* - x_0) = (I - A_h R_{h,r}) P_h y_\epsilon + S_{h,r} P_h (A P_h x_* - y_\epsilon) \tag{3.10}$$

(for the definition of $S_{h,r}$ see (3.1)). Estimating (3.10) yields

$$\|A_h S_{h,r}(x_* - x_0)\| \le \|(I - A_h R_{h,r}) P_h y_\epsilon\| + \xi_h \|(I - P_h) x_*\| + \epsilon, \tag{3.11}$$

$$\|A_h S_{h,r}(x_* - x_0)\| \ge \|(I - A_h R_{h,r}) P_h y_\epsilon\| - \xi_h \|(I - P_h) x_*\| - \epsilon, \tag{3.12}$$

since $\|S_{h,r}\| \le \sup_{0 \le t \le a} |1 - tg_r(t)| \le 1$ ($r \to |1 - tg_r(t)|$ is decreasing and $1 - tg_0(t) = 1$, hence $|1 - tg_r(t)| \le 1$ $(0 \le t \le a, 0 < r)$). Further we will use again the inequality (see (3.3))

$$\|P_h x_* - R_{h,r} y_\epsilon\| \le \|S_{h,r} P_h (x_* - x_0)\| + \gamma r (\xi_h \|(I - P_h) x_*\| + \epsilon). \tag{3.13}$$

1. To prove the first part of the theorem it suffices to show the convergence of the right-hand side of (3.13) with the above given parameter choice.
(i) To show the convergence of the second term of (3.13) let us first assume that the parameter is chosen according to Rule 1. If $r = r(h, \epsilon) \ne 0$, then (3.6) holds. From inequalities (3.12) and $r\xi_h \le 1$ we obtain

$$\begin{aligned} (d_1 - 1)\epsilon &\le \|A_h S_{h,r}(x_* - x_0)\| + \xi_h \|(I - P_h) x_*\| \\ \Rightarrow r(d_1 - 1)\epsilon &\le r\|A_h S_{h,r}(x_* - x_0)\| + \|(I - P_h) x_*\|. \end{aligned} \tag{3.14}$$

Again using the Banach-Steinhaus theorem we get

$$r\|A_h S_{h,r}(x_* - x_0)\| \to 0 \ (h \to 0, \epsilon \to 0).$$

Here the condition $p_0 > 1$ is important. Since $\|(I - P_h)x_*\| \to 0$ $(h \to 0)$, we have $r\epsilon \to 0$ $(h \to 0, \epsilon \to 0)$ and $r\xi_h \|(I - P_h)x_*\| \to 0$ $(h \to 0, \epsilon \to 0)$, the convergence of the second term in (3.13) is proved, if the parameter is chosen according to Rule 1. If the parameter is chosen according to Rule 2, we similarly obtain $s\epsilon \to 0$ $(h \to 0, \epsilon \to 0)$ and $s\xi_h \|(I - P_h)x_*\| \to 0$ $(h \to 0, \epsilon \to 0)$. The estimate $r \le \theta^{-1} s$ then shows $r\epsilon \to 0$ $(h \to 0, \epsilon \to 0)$ and $r\xi_h \|(I - P_h)x_*\| \to 0$ $(h \to 0, \epsilon \to 0)$, the convergence of the second term in (3.13) is proved, if the parameter is chosen according to Rule 2.
(ii) If, for $r = r(h, \epsilon)$ chosen by Rule 1, we have $r \ge r_1 := (\epsilon^{\frac{1}{2}} + \xi_h)^{-1}$, then $r \to \infty$ $(h \to 0, \epsilon \to 0)$ and the convergence of the first term of

the right-hand side in (3.13) follows as in the proof of Theorem 3.1. Otherwise, $r < r_1$ and Lemma 3.2 yields

$$\begin{aligned}
\|S_{h,r}P_h(x_* - x_0)\|^2 &= \|(I - g_r(A_h)A_h)P_h(x_* - x_0)\|^2 \\
&\le \kappa(\|(I - g_{r_1}(A_h)A_h)P_h(x_* - x_0)\|^2 \\
&\quad + r_1^2\|A_h(I - g_r(A_h)A_h)(x_* - x_0)\|^2) \\
&= \kappa(\|S_{h,r_1}P_h(x_* - x_0)\|^2 + r_1^2\|A_hS_{h,r}(x_* - x_0)\|^2). \qquad (3.15)
\end{aligned}$$

Here $\|S_{h,r_1}P_h(x_* - x_0)\|^2 \to 0$ $(h \to 0, \epsilon \to 0)$ (see the proof of Theorem 3.1) and $r_1^2\|A_hS_{h,r}(x_* - x_0)\|^2 \le (\epsilon^{\frac{1}{2}} + \xi_h)^{-2}((d_2+1)\epsilon + \xi_h\|(I - P_h)x_*\|)^2 \to 0$ $(h \to 0, \epsilon \to 0)$ (note that in case $r < r_1$ we have $r < \xi_h^{-1}$, hence (3.7) holds). Hence the convergence of the first term of (3.13) is proved, if the parameter is chosen according to Rule 1. If the parameter is chosen by Rule 2, in (ii) we just have to substitute d_1 by d to prove convergence.

2. (i) We estimate the second term of the right-hand side of (3.13). As in the first part of the proof of Theorem 3.3 let us first assume that the parameter r is chosen according to Rule 1. If $r \le (\frac{\epsilon}{\rho})^{-\frac{1}{p+1}}$, then an estimate for the second term of (3.13) which is good enough follows as in the proof of Theorem 3.1. Otherwise $r > (\frac{\epsilon}{\rho})^{-\frac{1}{p+1}}$. Using (3.14) we get

$$\begin{aligned}
r(d_1 - 1)\epsilon &\le \{r[\,\|(I - g_r(A_h)A_h)A_h^{p+1}\| + \|A_hS_{h,r}(A_h^p - A^p)\|] \\
&\quad + c_p\xi_h^{min\{p,1\}}\}\rho \\
&\le \{r[\gamma_{p+1}r^{-p-1} + b_p\gamma_1r^{-1}\xi_h^{min\{p,1\}}] + c_p\xi_h^{min\{p,1\}})\}\rho \\
&\le \{\gamma_{p+1}(\frac{\epsilon}{\rho})^{\frac{p}{p+1}} + (b_p\gamma_1 + c_p)\xi_h^{min\{p,1\}}\}\rho. \qquad (3.16)
\end{aligned}$$

This estimate is good enough for our purpose. With an argument like that in the proof of Theorem 3.1 we get the same estimate, if the parameter is chosen according to Rule 2.

(ii) Now we will estimate the first term of the right-hand side of (3.13). If $r \ge ((\epsilon/\rho)^{\frac{1}{p+1}} + \xi_h)^{-1} = r_1$, then a good estimate follows as in the proof of Theorem 3.1. Otherwise from (3.11), (3.15), Corollary 4.2 and Lemma 4.3 we have

$$\begin{aligned}
\|S_{h,r}P_h(x_* - x_0)\|^2 &\le \kappa(\|S_{h,r_1}P_h(x_* - x_0)\|^2 + r_1^2\|A_hS_{h,r}(x_* - x_0)\|^2) \\
&\le \kappa((\|S_{h,r_1}A_h^pz\| + b_p\xi_h^{min\{p,1\}}\rho)^2 + r_1^2((d_2+1)\epsilon + c_p\xi_h^{min\{p,1\}+1}\rho)^2) \\
&\le \kappa((\gamma_p((\frac{\epsilon}{\rho})^{\frac{1}{p+1}} + \xi_h)^p + b_p\xi_h^{min\{p,1\}})^2\rho^2 \\
&\quad + ((d_2+1)(\rho\epsilon^p)^{\frac{1}{p+1}} + \rho c_p\xi_h^{min\{p,1\}})^2). \quad \square
\end{aligned}$$

Remark 1. If $p \leq 1$, then we have $r = O((\frac{\epsilon}{\rho})^{-\frac{1}{p+1}})$ which follows from inequalities (3.16) and $\xi_h^p \leq r^{-p}$.

2. The remark after the proof of Theorem 3.1 also applies to Theorem 3.3 (with $p_0 - 1$ instead of p_0).

3. In case of critical level of discrepancy ($d_1 = d_2 = 1$ in Rule 1, $d = 1$ in Rule 2) generally we obtain divergence, see [16].

4. The discrepancy level $d_1\epsilon + f_1\xi_h \leq \|(I - A_h R_{h,r}) P_h y_\epsilon\| \leq d_2\epsilon + f_2\xi_h$ with $1 < d_1 \leq d_2, 0 < f_1 \leq f_2$ leads to error estimate $\|x_* - R_{h,r}y_\epsilon\| = O((\epsilon + \xi_h)^{\frac{p}{p+1}})$ $(0 < p \leq p_0 - 1)$, that is worse than assertion 2 in Theorem 3.3.

4 Some results on fractional powers of linear operators

The proof of the next lemma which can be found in [18] is mainly based on the following formula (see [6]): Suppose $0 < \alpha < 1$. Let $B \in L_b(X)$ such that $B = B^* \geq 0$. Then

$$B^\alpha = \frac{\sin \alpha\pi}{\pi} \int_0^\infty t^{\alpha-1}(tI + B)^{-1} B \, dt.$$

If B is a compact operator this is not difficult to prove, if one uses the singular value decomposition of B and the formula

$$\lambda^\alpha = \frac{\sin \alpha\pi}{\pi} \int_0^\infty t^{\alpha-1}(t + \lambda)^{-1}\lambda \, dt, \quad 0 < \alpha < 1, \lambda \geq 0..$$

Lemma 4.1 *Suppose that $B \in L_b(X)$ with $B = B^* \geq 0$. If $p > 0$ and $a > 0$, then for any $B_h \in L_b(X)$ with $B_h = B_h^* \geq 0, \|B_h\| \leq a$ we have*

$$\|B^p - B_h^p\| \leq a_p \|B - B_h\|^{\min\{p,1\}}.$$

Here $a_p = \frac{4}{\pi}$, if $p \leq 1$, and $p \to a_p$ is bounded in $(0, p_0]$ for any $p_0 > 0$.

Corollary 4.2 *Suppose that $A \in L_b(X), A = A^* \geq 0$. If $p > 0$, then for any orthogonal projection $P \in L_b(X)$ we have*

$$\|A^p - (PAP)^p\| \leq b_p \|(I - P)A\|^{\min\{p,1\}},$$

$b_p = \frac{8}{\pi}$, if $p \leq 1$, and $p \to b_p$ is bounded in $(0, p_0]$ for any $p_0 > 0$.

Proof. This follows immediately from Lemma 4.1 with $B = A, B_h = PAP$, since $\|B - B_h\| \leq 2\|(I - P)A\|$. □

The proof of the following lemma can be found in [11].

Lemma 4.3 *Let $p > 0, A \in L_b(X)$ with $A = A^* \geq 0$ and $P \in L_b(X)$ be an orthogonal projection. Let $c_p = 1$, if $p \leq 1$, and $c_p = \|A\|^{p-1}$, if $p > 1$. Then the following inequality holds:*

$$\|(I-P)A^p\| \leq c_p\|A(I-P)\|^{min\{p,1\}}.$$

5 On the regularization of projection methods for ill-posed problems with nonselfadjoint operator

The proofs of the following theorems are presented in [12]. Let X and Y be real Hilbert spaces and $A \in L_b(X,Y)$, i.e. $A : X \to Y$ is a bounded linear operator. We consider the equation

$$Ax = y, \; y \in R(A). \tag{5.1}$$

We assume that only an approximation $y_\epsilon \in$ Y to y is available with $\|y - y_\epsilon\| \leq \epsilon$, where $\epsilon > 0$ is a known error bound. To get an approximation to a solution of (5.1) we use regularization methods which are generated by Borel measurable functions

$$g_r : [0,a] \to I\!R,$$

$r \geq 0, \|A\|^2 \leq a$. We assume that the functions g_r satisfy conditions (1.3) and

$$\sup_{0\leq t\leq a} \sqrt{t}|g_r(t)| \leq \gamma_* r^{\frac{1}{2}}, \quad r \geq 0, \tag{5.2}$$

where $p_0 > 0, \gamma_p$ and γ_* are constants. Let $x_0 \in X$ be an initial approximation. Then, with an appropriate parameter choice r an approximation to the solution x_* of (5.1), which is nearest to x_0, i.e. $Ax_* = y, x_* - x_0 \in N(A)^\perp$, is given by $R_{h,r}y_\epsilon$ with

$$R_{h,r} = (I - g_r(A_h^*A_h)A_h^*A_h)P_h x_0 + g_r(A_h^*A_h)A_h^*. \tag{5.3}$$

Here $A_h = Q_h A P_h : X \to Y$, where P_h and $Q_h, h > 0$, are orthogonal projections in X resp. Y.

5.1 Examples

1. The method of Tikhonov: We have to determine the solution $x_{h,r}$ of the equation

$$(A_h^*A_h + r^{-1}I)x = A_h^*y_\epsilon. \tag{5.4}$$

The method is of the form (5.3) with $x_0 = 0$ and $g_r(t) = (t + r^{-1})^{-1}$, $r > 0$. Conditions (1.3) and (5.2) hold with $p_0 = 1$.

2. The generalized method of Tikhonov: Let $q \geq -\frac{1}{2}$. We have to determine the solution $x_{h,r}$ of the equation

$$((A_h^* A_h)^{q+1} + r^{-q-1} I)x = (A_h^* A_h)^q A_h^* y_\epsilon, \tag{5.5}$$

if $q \geq 0$. Otherwise, the equation takes the form

$$(A_h^* A_h + r^{-q-1}(A_h^* A_h)^{-q})x = A_h^* y_\epsilon.$$

The method is of the form (5.3) with $x_0 = 0$ and $g_r(t) = t^q/(t^{q+1} + r^{-q-1})$, $r > 0$. Conditions (1.3) and (5.2) hold with $p_0 = q + 1$.

3. The method of successive approximation (explicit scheme): Let $0 < \mu < \frac{2}{a}$. The algorithm which is given by $x_{h,0} = P_h x_0$,

$$x_{h,r} = (I - \mu A_h^* A_h)x_{h,r-1} + \mu A_h^* y_\epsilon, \quad r = 1, 2, ... \tag{5.6}$$

We have $R_{h,r} y_\epsilon = x_{h,r}$, where $R_{h,r}$ is of the class (5.3) with $g_r(t) = \frac{1}{t}[1 - (1 - \mu t)^r]$, $t \neq 0$. Conditions (1.3) and (5.2) hold for any $p_0 > 0$.

4. Implicit scheme: Let $0 < \mu$ be constant. The algorithm is given by $x_{h,0} = P_h x_0$,

$$(A_h^* A_h + \mu I)x_{h,r} = \mu x_{h,r-1} + A_h^* y_\epsilon, \quad r = 1, 2, ... \tag{5.7}$$

We have $R_{h,r} y_\epsilon = x_{h,r}$, where $R_{h,r}$ is of the class (5.3) with $g_r(t) = \frac{1}{t}[1 - (\frac{\mu}{\mu+t})^r]$, $t \neq 0$. Conditions (1.3) and (5.2) hold for any $p_0 > 0$.

Other examples of regularization methods can be found in [7], [15], [16] or [18]. Now let $\Phi_1, \Phi_2, ..., \Phi_n \in X$ be a basis of $R(P_h)$ and $\Psi_1, \Psi_2, ..., \Psi_m \in Y$ be a basis of $R(Q_h)$. We define

$$\begin{aligned} G_\Phi &= (<\Phi_i, \Phi_j>)_{i,j=1,...,n}, \\ G_\Psi &= (<\Psi_i, \Psi_j>)_{i,j=1,...,m}, \\ B &= (<\Psi_i, A\Phi_j>)_{i=1,...,m,\ j=1,...,n}, \\ z &= (<\Psi_i, y_\epsilon>)_{i=1,...,m}. \end{aligned} \tag{5.8}$$

Let us first consider the method of Tikhonov. The solution $x_{h,r}$ of the corresponding equation (5.4) can be expressed in the form

$$x_{h,r} = \sum_{j=1}^{n} c_j \Phi_j.$$

Then (5.4) is equivalent to the following system of equations for determining $c = (c_j)_{j=1,...,n}$:

$$(B^T G_\Psi^{-1} B + r^{-1} G_\Phi)c = B^T G_\Psi^{-1} z. \tag{5.9}$$

Concerning the method of successive approximation (5.6) and the implicit scheme (5.7) $x_{h,r}$ may be expressed again in the form

$$x_{h,r} = \sum_{j=1}^{n} c_j^r \Phi_j.$$

Some calculation shows that (5.6) resp. (5.7) is equivalent to the following sequence of systems of equations (5.10) resp. (5.11) for determining $c^r = (c_j^r)_{j=1,\dots,n}$, $r = 1, 2, \dots$:

$$c^r = c^{r-1} - \mu(G_\Phi^{-1}B^T G_\Psi^{-1} B c^{r-1} - G_\Phi^{-1} B^T G_\Psi^{-1} z), \tag{5.10}$$

$$(B^T G_\Psi^{-1} B + \mu G_\Phi) c^r = \mu G_\Phi c^{r-1} + B^T G_\Psi^{-1} z. \tag{5.11}$$

In both cases we have $c^0 = G_\Phi^{-1}((< \Phi_j, x_0 >)_{j=1,\dots,n})$.

5.2 Parameter choice

We have $|A|^p = (A^*A)^{\frac{p}{2}}$.

Theorem 5.1 *Let $A \in L_b(X,Y)$, $\|A\|^2 \le a$. Let $y \in R(A)$, $\|y - y_\epsilon\| \le \epsilon$, $x_0 \in X$ and x_* the solution of $Ax=y$, which is nearest to x_0. Let $P_h \in L_b(X)$ and $Q_h \in L_b(Y)$ be orthogonal projections and $A_h = Q_h A P_h$. Suppose that $\|A(I - P_h)\| \le \xi_h$ and $\|(I - Q_h)A\| \le \eta_h$. Let conditions (1.3), (5.2) and (5.3) hold.*
1. Suppose that $P_h \to I$ pointwise, $\xi_h \to 0, \eta_h \to 0$ $(h \to 0)$. If

$$r_{(h,\epsilon)}^{1/2}\epsilon \to 0,\ r_{(h,\epsilon)}^{1/2}\xi_h \le C \text{ and } r_{(h,\epsilon)} \to \infty \ (h \to 0, \epsilon \to 0),$$

then $R_{h,r_{(h,\epsilon)}} y_\epsilon \to x_*$ $(h \to 0, \epsilon \to 0)$.
2. If $0 < p \le 2p_0$, $x_* - x_0 = |A|^p z$, $\|z\| \le \rho$, $x_* = |A|^p v$, $\|v\| \le \rho$, and

$$C_1((\frac{\epsilon}{\rho})^{\frac{1}{p+1}} + \xi_h) \le r^{-1/2} \le C_2((\frac{\epsilon}{\rho})^{\frac{1}{p+1}} + \xi_h^{\min\{\frac{1}{p},1\}})$$

with some positive constants C_1, C_2. then

$$\|x_* - R_{h,r} y_\epsilon\| \le e_p((\rho\epsilon^p)^{\frac{1}{p+1}} + \rho(\xi_h^{\min\{p,1\}} + \eta_h^{\min\{p,2\}})).$$

e_p is independent of ϵ, h and ρ. $p \to e_p$ is bounded in $(0, p_0]$ for any $p_0 > 0$.

Theorem 5.2 *Let $A \in L_b(X,Y)$, $\|A\|^2 \le a$, $y \in R(A)$, $\|y - y_\epsilon\| \le \epsilon$, $x_0 \in X$ and x_* the solution of $Ax=y$. which is nearest to x_0. Let $P_h \in L_b(X)$ and $Q_h \in L_b(Y)$ be orthogonal projections and $A_h = Q_h A P_h$. Suppose that $\|A(I - P_h)\| \le \xi_h$ and $\|(I - Q_h)A\| \le \eta_h$. Suppose that conditions (1.3), (5.2) and (5.3) hold for functions g_r with $p_0 > \frac{1}{2}$ in*

(1.3) and that $r \to |1 - tg_r(t)|$ is decreasing for any $t \geq 0$. Let the parameter $r = r(h,\epsilon)$ be chosen according to Rule 1 or Rule 2 (see Section 3), where $P_h y_\epsilon$ is substituted by $Q_h y_\epsilon$ and r_{max} is set to ξ_h^{-2} resp. $[\xi_h^{-2}] + 1$.
1. If $P_h \to I$ pointwise, $\xi_h \to 0, \eta_h \to 0$ $(h \to 0)$, then

$$R_{h,r} y_\epsilon \to x_* \ (h \to 0, \epsilon \to 0).$$

2. If $0 < p \leq 2p_0 - 1$, $x_* - x_0 = |A|^p z, \|z\| \leq \rho$ and $x_* = |A|^p v, \|v\| \leq \rho$, then

$$\|x_* - R_{h,r} y_\epsilon\| \leq e_p((\rho\epsilon^p)^{\frac{1}{p+1}} + \rho(\xi_h^{min\{p,1\}} + \eta_h^{min\{p,2\}})).$$

e_p is independent of ϵ, h and ρ.

Remark Another a posteriori parameter selection which leads to better results in few cases is recently under research (see e.g. [1], [2], [5], [10], [13], [14]).

Acknowledgement. This work was written during the visit of the second author at the University of Kaiserslautern (December 1988 - March 1989), which was partially supported by DFG (Deutsche Forschungsgemeinschaft).

References

1. H.W. Engl and H. Gfrerer, A posteriori parameter choice for generalized regularization methods for solving linear ill-posed problems, Applied Numerical Mathematics 4 (1988), 395-417.

2. H. Gfrerer, An a-posteriori parameter choice for ordinary and iterated Tikhonov regularization of ill-posed problems leading to optimal convergence rates, Math. Comput. 49 (1987), 507-522, S5-S12.

3. U.A. Hämarik, Residue principle for choice of dimension solving ill-posed problems by projection methods, Uch. Zap. Tartu Gos. Univ. 672 (1984), 27-34.

4. V.K. Ivanov. Approximate solution of operator equations of the first kind. U.S.S.R. Comput. Maths. Math. Phys. 6 No. 6 (1966), 197-205.

5. J.T. King and A. Neubauer, A variant of finite-dimensional Tikhonov regularization with a-posteriori parameter choice, Computing 40 (1988), 91-109.

6. M. Krasnoselskii et al, Integral operators in spaces of summable functions (Noordjoff Int. Publ., Leydem, 1976).

7. A.K. Louis, Inverse und schlecht gestellte Probleme (Teubner, Stuttgart, 1989).

8. V.A. Morozov, On the solution of functional equations by the method of regularization, Soviet Math. Doklady 7 (1966), 414-417.

9. F. Natterer, Regularisierung schlecht gestellter Probleme durch Projektionsverfahren, Numer. Math. 28 No. 3 (1977), 511-522.

10. A. Neubauer, An a posteriori parameter choice for Tikhonov Regularization in the presence of modeling error, Applied Numerical Mathematics 4 (1988), 507-519.

11. R. Plato, Discretization and regularization of ill-posed problems, submitted.

12. R. Plato and G.M. Vainikko, On the regularization of projection methods for solving ill-posed problems, submitted.

13. T. Raus, On the discrepancy principle for the solution of ill-posed problems, Uch. Zap. Tartu Gos. Univ. 672 (1984), 16-26.

14. T. Raus, On the residue principle for the solution of ill-posed problems with non-selfadjoint operators, Uch. Zap. Tartu Gos. Univ. 715 (1984), 12-20.

15. G.M. Vainikko, The discrepancy principle for a class of regularization methods, U.S.S.R. Comput. Maths. Math. Phys. 22 No. 3 (1982), 1-19.

16. G.M. Vainikko, The critical level of discrepancy in regularization methods, U.S.S.R. Comput. Maths. Math. Phys. 23 No. 6 (1983), 1-9.

17. G.M. Vainikko and U.A. Hämarik, Projection methods and self-regularization in ill-posed problems. Izv. VUZ. Mat. 29 No. 10 (1985), 1-17.

18. G.M. Vainikko and A.Yu. Veretennikov. Iteration procedures in ill- posed problems (Nauka, Moscow, 1986).

О РЕГУЛЯРИЗАЦИИ МЕТОДА РИТЦА-ГАЛЕРКИНА ДЛЯ РЕШЕНИЯ НЕКОРРЕКТНО ПОСТАВЛЕННЫХ ЗАДАЧ

Р.Плато и Г.Вайникко

Резюме

В гильбертовом пространстве X рассматривается уравнение $Ax=y$, где оператор $A=A^*\geq 0$ известен точно, а вместо $y\in R(A)$ задан $y_\varepsilon\in X$, $\|y_\varepsilon - y\|\leq\varepsilon$. Пусть P_h $(h>0)$ – ортопроекторы в X, $\|A(I-P_h)\|\leq\xi_h\to 0$ $(h\to 0)$. Приближенное решение строится по формуле

$$x_{r,h,\varepsilon}=(I-g_r(A_h)A_h)P_hx_0+g_r(A_h)P_hy_\varepsilon,$$

где $x_0\in X$ – начальное приближение, $A_h=P_hAP_h$, а измеримые по Борелю функции $g_r\colon[0,a]\to\mathbb{R}$ $(a\geq\|A\|)$ удовлетворяют условиям (1.3) и (1.4). Ряд известных методов, например, метод Лаврентьева, его видоизменения, некоторые итерационные методы, примененные к дискретизированному при помощи метода Ритца–Галеркина уравнению $A_hx_h=P_hy_\varepsilon$, включаются в эту схему; указаны соответствующие системы уравнений для вычислений.

В статье указывается априорный выбор параметра регуляризации r, обеспечивающий на классах истокообразно представимых решений $x_*=A^pz$, $x_*-x_0=A^pv$ $(0<p<p_0)$ сходимость порядка

$$\|x_{r,h,\varepsilon}-x_*\|=O(\varepsilon^{p/(p+1)}+\xi_h^{\min\{p,1\}})$$

(см. теорему 3.1). Такой же порядок сходимости при $0<p\leq p_0-1$ достигается при выборе r по невязке на уровне $\|A_hx_{r,h,\varepsilon}-P_hy_\varepsilon\|=d\varepsilon$, $d>1$, с априорным ограничением $r\leq\xi_h^{-1}$ (см. теорему 3.3) Отметим, что уровень невязки $d\varepsilon+f\xi_h$ $(d>1,\ f>0)$ приводит к более грубой оценке.

Вкратце рассмотрен также случай проекционных методов для несамосопряженных задач.

A COLLOCATION METHOD WITH CUBIC SPLINES TO THE SOLUTION OF A MULTIDIMENSIONAL WEAKLY SINGULAR INTEGRAL EQUATION

Peep Uba

A collocation method with cubic splines of class C^2 for numerical solution of a two-dimensional weakly singular integral equation is constructed. To obtain the fourth degree of accuracy, a special non-uniform grid is used where (analogous for one-dimensional case in [2,6]) the degree of non-uniformity depends on the properties of the kernel of the integral operator. It is easy to generalize the present method for multidimensional equation on a parallelepiped.

1. **The smoothness of the solution** of the multidimensional weakly singular integral equation

$$u(z) = \int_G K(z,\zeta)u(\zeta)d\zeta + f(z), \quad z \in G, \tag{1}$$

where $G \subset \mathbb{R}^n$ is an open bounded domain, is investigated in [1,7] (see [3] too). In this part we refer to a special case of these results, where $n = 2$ and G is a two-dimensional rectangle $G = \{z \equiv (x,y) \in \mathbb{R}^2\colon 0 < x < b_1,\ 0 < y < b_2\}$.

We assume that the derivatives $D_z^\alpha D_\zeta^\beta K(z,\zeta)$, $|\alpha| + |\beta| \le 4$ are continuous on the set $(G \times G)\setminus\{z = \zeta\}$ and there exists a real number ν $(-\infty < \nu < 2)$ such that the estimations

$$\left.\begin{aligned} &\left|\left(\frac{\partial}{\partial x}\right)^{\alpha_1}\left(\frac{\partial}{\partial y}\right)^{\alpha_2}\left(\frac{\partial}{\partial x}+\frac{\partial}{\partial \xi}\right)^{\beta_1}\left(\frac{\partial}{\partial y}+\frac{\partial}{\partial \psi}\right)^{\beta_2} K(z\equiv(x,y),\zeta\equiv(\xi,\psi))\right| \le \\ &\le c\begin{cases}1+|\ln|z-\zeta||, & \nu+|\alpha|=0,\\ 1+|z-\zeta|^{-\nu-|\alpha|}, & \nu+|\alpha|\ne 0,\end{cases} \quad |\alpha|+|\beta|\le 4\end{aligned}\right\} \tag{2}$$

are valid (we keep in view the validity of estimations for each $\alpha = (\alpha_1,\alpha_2)$ and $\beta = (\beta_1,\beta_2)$ with $\alpha_j \ge 0$, $\beta_j \ge 0$, $|\alpha|+|\beta| \equiv \alpha_1+\alpha_2+\beta_1+\beta_2 \le 4$). Moreover, we assume that the kernel $K(z,\zeta)$ has the following smoothness:

$$\left.\begin{aligned} &\forall\, \varepsilon>0 \quad \forall\, \eta>0 \quad \exists\, \delta(\varepsilon,\eta)>0: \\ &z_1, z_2, \zeta \in G,\ |z_1-z_2| < \delta,\ |z_j-\zeta| \ge \eta \quad (j=1,2) \Rightarrow \\ &\Rightarrow \left|D_z^\alpha D_\zeta^\beta K(z_1,\zeta) - D_z^\alpha D_\zeta^\beta K(z_2,\zeta)\right| < \varepsilon \quad (|\alpha|+|\beta|\le 4).\end{aligned}\right\} \tag{3}$$

Here $D_z^\alpha D_\zeta^\beta K(z_j,\zeta) = \left[\left(\frac{\partial}{\partial x}\right)^{\alpha_1}\left(\frac{\partial}{\partial y}\right)^{\alpha_2}\left(\frac{\partial}{\partial \xi}\right)^{\beta_1}\left(\frac{\partial}{\partial \psi}\right)^{\beta_2}K(z,\zeta)\right]_{z=z_j\equiv(x_j,y_j)}$.

By $\rho_x(z) = min(x, b_1 - x)$ and $\rho_y(z) = min(y, b_2 - y)$ we denote the distances from the point $z \equiv (x,y)$ to the nearest of the opposite boundary lines of G, which are orthogonal to the axis x and the axis y respectively. Hence $\rho(z) = min(\rho_x(z), \rho_y(z))$ is the distance from the point z to the boundary line of domain G.

Let $C_{\square}^{4,\nu}(G)$ be a special weight class of functions $u \in C^4(G) \cap C(\bar{G})$, the derivatives of which satisfy the inequalities

$$|D^{\alpha} u(z)| \le c \begin{cases} 1, & |\alpha| < 2-\nu, \\ 1 + |ln\,\rho(z)|, & |\alpha| = 2-\nu, \\ \rho(z)^{2-\nu-|\alpha|}, & |\alpha| > 2-\nu. \end{cases} \tag{4}$$

in G, where the derivatives $\partial^k u(x,y)/\partial x^k$ and $\partial^k u(x,y)/\partial y^k$ can be extended to continuous functions on sets $0 < x < b_1$, $0 \le y \le b_2$ and $0 \le x \le b_1$, $0 < y < b_2$ respectively, and the estimates

$$\left|\frac{\partial^k u(x,y)}{\partial x^k}\right| \le c \begin{cases} 1 + |ln\,\rho_x(z)|, & k = 2-\nu, \\ \rho_x(z)^{2-\nu-k}, & k > 2-\nu, \end{cases}$$
$$\left|\frac{\partial^k u(x,y)}{\partial y^k}\right| \le c \begin{cases} 1 + |ln\,\rho_y(z)|, & k = 2-\nu, \\ \rho_y(z)^{2-\nu-k}, & k > 2-\nu, \end{cases} \tag{4'}$$

are valid. The constant c in inequalities certainly depends on u and has different values in different inequalities.

Lemma 1 (see [1], Theorem 2): Let $f \in C_{\square}^{4,\nu}(G)$ and the conditions (2), (3) be satisfied. If the equation (1) is solvable in $L_1(G)$, then all integrable solutions belong $C_{\square}^{4,\nu}(G)$.

2. Degree of the accuracy of interpolation. In domain $G = [0,b_1]\times[0,b_2]$ we define the grid $\Delta^r \equiv \Delta_x \times \Delta_y$ by

$$\left.\begin{aligned} \Delta_x &= \{ x_j : x_j = (b_1/2)(j/N_x)^r ;\ x_{j+N_x} = b_1 - x_{N_x - j},\ j = 0,1,\dots,N_x \}, \\ \Delta_y &= \{ y_j : y_j = (b_2/2)(j/N_y)^r ;\ y_{j+N_y} = b_2 - y_{N_y - j},\ j = 0,1,\dots,N_y \}, \end{aligned}\right\} \tag{5}$$

where the parameter $r \in R$, $r \ge 1$, characterizes the degree of nonuniformity of the grid and will be specified below. The formulas (5) give us the partition of the domain G into cells

$$G_{jk} = \{ (x,y) : x_j \le x \le x_{j+1},\ y_k \le y \le y_{k+1} \} \subset G$$
$$j = 0,1,\dots,2N_x - 1; \quad k = 0,1,\dots,2N_y - 1.$$

For short expressions we introduce the notations $N = (N_x, N_y)$ and $h = 1/min(N_x, N_y)$.

For a function $f(z)$, $z \in \bar{G}$, we construct a twice continuously differentiable function $S(f; x, y)$ on $\bar{G}$, which is a cubic polynomial of both variables x and y on each cell G_{jk} and which interpolates the $f(z)$ in the points of the grid Δ^r (we say that $S(f;x,y)$ is the two-dimensional interpolating cubic spline of defect 1; for detail see, for example [4]). It is well known that for the uniqueness of the interpolating cubic spline in addition to interpolating conditions one needs certain boundary conditions. As our aim is to interpolate the functions, the derivatives of which can have the singularities at the boundary of domain G (the solution of equation (1)), we choose these boundary conditions in the form

$$\left.\begin{aligned}&\left(\frac{\partial}{\partial x}\right)^3 S(f;x_k+0,y_j) = \left(\frac{\partial}{\partial x}\right)^3 S(f;x_k-0,y_j), \quad k=1,2N_x-1;\ j=0,1,\dots,2N_y,\\ &\left(\frac{\partial}{\partial y}\right)^3 S(f;x_k,y_j+0) = \left(\frac{\partial}{\partial y}\right)^3 S(f;x_k,y_j+0), \quad k=0,1,\dots,2N_x;\ j=1,2N_y-1,\end{aligned}\right\} \quad (6)$$

(about these and other boundary conditions see, for example [4], §3.3).

Lemma 2: Let $f \in C_{\alpha}^{4,\nu}(G)$. If $r = 4/(2-\nu)$, then, for the interpolating cubic spline $S(f;x,y)$, the estimation

$$\max_{(x,y)\in\bar{G}} |f(x,y) - S(f;x,y)| \le ch^4 \quad (7)$$

is valid, where c is independent of N.

<u>Proof</u>: We can present the difference $S(f;x,y) - f(x,y)$ in form (see [4],p.136) :

$$S(f;x,y) - f(x,y) = S[S[f(x,y);y] - f(x,y);x] + S[f(x,y);x] - f(x,y), \quad (8)$$

where $S[f(x,y);y]$ denotes the one-dimensional cubic spline of defect 1 on the grid Δ_y^r, which depends on the variable y and interpolates the function $f(x,y)$ in points y_j. The variable x in $S[f(x,y);y]$ is a parameter. The right side of equality (8) consists only of interpolations with one-dimensional cubic splines, the estimations of which we can determine with procedures presented in [2,6]. Namely, observe step by step the proof of Theorem 4.2 in [2] (or Theorem 2 in [6]) with $r = 4/(2-\nu)$ and consider (4'). We get

$$\max_{0\le y\le b_2} |S[f(x,y);y] - f(x,y)| \le c\,N_y^{-4} \le c_1 h^4 \quad (9)$$

independently from x and

$$\max_{0\le x\le b_1} |S[f(x,y);x] - f(x,y)| \le \tilde{c}\,N_x^{-4} \le c_2 h^4$$

independently from y. We also note that the projector in $C(\bar{G})$

$$P_x : f(x,y) \to S[f(x,y);x]$$

is uniformly bounded with respect to y. Now with regard to (9) we can estimate the first summand in (8)

$$\max_{0\le y\le b_2}\ \max_{0\le x\le b_1} |S[S[f(x,y);y] - f(x,y);x]| =$$

$$\|P_x(S[f(x,y);y] - f(x,y))\|_C \le \|P_x\|\,\|S[f(x,y);y] - f(x,y)\| \le$$

$$\le c'\max_{0\le y\le b_2}\ \max_{0\le x\le b_1} |S[f(x,y);y] - f(x,y)| \le c_3 h^4.$$

The lemma is proved.

Remark 1. Let $f \in C(G)$. Similarily to the results for one-dimensional case in [2,6] the estimation

$$\max_{(x,y)\in \overline{G}} |f(x,y) - S(f;x,y)| \le c\,\omega(f)$$

with $\omega(f) = \max\limits_{k,j} \max\limits_{z_1,z_2 \in G_{k,j}} |f(z_1) - f(z_2)|$ can be proved.

Remark 2. Let P_N denote the interpolation projector in $C(\overline{G})$, assigning to any continuous function $f \in C(\overline{G})$ its interpolant $S(f;x,y)$ satisfying boundary conditions (6). Due to the principle of uniform boundedness the sequence of operators $\{P_N\}$ is uniformly bounded.

3.The collocation method for integral equation. The approximate solution $u_N(z) \equiv u_N(x,y)$ of equation (1) we seek in form of a cubic spline on grid Δ^r. It is required that $u_N(z)$ should satisfy equation (1) in the interpolation points (x_k,y_j) and boundary conditions (compare with (6)) :

$$\left[u_N(z) - \int_G K(z,\zeta)u_N(\zeta)d\zeta - f(z)\right]_{z=(x_k,y_j)} = 0 \tag{10}$$

$$k = 0,1,\dots,2N_x\,;\; j = 0,1,\dots,2N_y$$

$$\left.\begin{aligned} \left(\frac{\partial}{\partial x}\right)^3 u_N(x_k+0,y_j) &= \left(\frac{\partial}{\partial x}\right)^3 u_N(x_k-0,y_j),\; j = 0,1,\dots,2N_y,\; k = 1,2N_x-1 \\ \left(\frac{\partial}{\partial y}\right)^3 u_N(x_j,y_k+0) &= \left(\frac{\partial}{\partial y}\right)^3 u_N(x_j,y_k-0),\; j = 0,1,\dots,2N_x,\; k = 1,2N_y-1 \end{aligned}\right\} \tag{11}$$

Theorem 1: Let $f \in C_{\square}^{4,\nu}(G)$ and the conditions (2) and (3) be satisfied. Suppose that the homogeneous equation corresponding to (1) has only the trivial solution in $L_1(G)$. Then there exist integers N_x^0 and N_y^0 such that with $N_x \ge N_x^0$ and $N_y \ge N_y^0$ the collocation equation (10) with conditions (11) has the unique solution $u_N(z)$ and

$$\max_{z\in\overline{G}} |u_N(z) - u(z)| \longrightarrow 0$$

as $N \to \infty$ (it is $N_x \to \infty$ and $N_y \to \infty$), where $u(z)$ is the solution of (1). If (in case $-2 \le \nu < 2$) . $= 4/(2-\nu)$, then

$$\max_{z\in\overline{G}} |u_N(z) - u(z)| \le ch^4. \tag{12}$$

<u>Proof:</u> Let K denote the integral operator of equation (1). Then (1) can be considered as the equation $u = Ku + f$ in the Banach space $C(G)$. The spline collocation conditions {(10),(11)} are equivalent to the solution of equation $u_N = P_N K u_N + P_N f$, where projector P_N is described in Remark 2. In virtue of Remarks occurs the strong convergense $P_N \to I$ as $N \to \infty$. The assumptions (2) and (3) ensure that the integral operator K is compact as an operator from $C(\overline{G})$ to $C(\overline{G})$. By means of standard arguments (see [5], Lemma 5.5) we conclude

$$\|K - P_N K\|_{C(\overline{G}) \to C(\overline{G})} \longrightarrow 0$$

as $N \to \infty$. Now the invertibility of $I - K$ ensures for sufficiently large N the invertibility of $I - P_N K : C(\bar{G}) \to C(\bar{G})$, whereby

$$\|u_N - u\| \le c\, \|P_N u - u\|.$$

From here and Lemma 2 we immediately get the estimation (12).

The theorem is proved.

4. Approximation with B-splines. The conditions $\{(10),(11)\}$ represent a system of linear equations whose exact form is determined by the choise of a basis in the subspace of cubic splines. Since the space of two-dimensional splines is a tensor product space of two one-dimensional spline spaces (see, for example [4]), we can seek u_N in form

$$u_N(x,y) = \sum_{k=-1}^{2N_x+1} \sum_{l=-1}^{2N_y+1} b_{kl} B_k(x) \bar{B}_l(y)$$

where b_{kl} are unknown and $B_k(x)$ is the one-dimensional cubic B-spline with support $[x_{k-2}, x_{k+2}]$ ($\bar{B}_l(y)$ respectively with support $[y_{l-2}, y_{l+2}]$). About the construction of these splines see [4] or [2,6]. We note that, in addition to the points of the grid Δ^r for the construction of B-splines, $6(2N_x+1)+6(2N_y+1)$ points from outside of domain $\bar{G}$ are necessary.

Now conditions $\{(10),(11)\}$ take the form of the following system of equations:

$$\sum_{k=i-1}^{i+1} \sum_{l=j-1}^{j+1} b_{kl} B_k(x_i) \bar{B}_l(y_j) =$$

$$= f(x_i, y_j) + \sum_{k=-1}^{2N_x+1} \sum_{l=-1}^{2N_y+1} b_{kl} \iint\limits_{G \cap G_{kl}} K(x_i, y_j, s, t) B_k(s) \bar{B}_l(t)\, ds\, dt$$

$$i = 0,1,\dots,2N_x; \quad j = 0,1,\dots,2N_y,$$

$$\sum_{l=j-1}^{j+1} \sum_{i=1}^{3} b_{il} B_i'''(x_1+0) \bar{B}_l(y_j) = \sum_{i=-1}^{2} \sum_{l=j-1}^{j+1} b_{il} B_i'''(x_1-0) \bar{B}_l(y_j), \qquad j = 0,1,\dots,2N_y,$$

$$\sum_{i=-2}^{1} \sum_{l=j-1}^{j+1} b_{2N_x+i,l} B_{2N_x+i}'''(x_{2N_x-1}+0) \bar{B}_l(y_l) =$$

$$= \sum_{i=-3}^{0} \sum_{l=j-1}^{j+1} b_{2N_x+i,l} B_{2N_x+i}'''(x_{2N_x-1}-0) \bar{B}_l(y_j), \qquad j = 0,1,\dots,2N_y,$$

$$\sum_{k=i-1}^{i+1} \sum_{j=0}^{3} b_{kj} B_k(x_i) \bar{B}_l'''(y_1+0) = \sum_{k=i-1}^{i+1} \sum_{j=-1}^{2} b_{kj} B_k(x_i) \bar{B}_l'''(y_1-0), \quad i = 0,1,\dots,2N_x,$$

$$\sum_{k=i-1}^{i+1} \sum_{j=-2}^{1} b_{k,2N_y+j} B_k(x_i) \bar{B}_{2N_y+j}'''(y_{2N_y-1}+0) =$$

$$= \sum_{k=i-1}^{i+1} \sum_{j=-3}^{0} b_{k,2N_y+j} B_{2N_y+j}'''(y_{2N_y-1}-0), \qquad i = 0,1,\dots,2N_x.$$

For the final determination of the matrix of the system of linear equations we need to select a sufficiently exact method for computation of integrals

$$\int_{x_{k-2}}^{x_{k+2}} \int_{y_{l-2}}^{y_{l+2}} K(x_i, y_j, s, t) B_k(s) \bar{B}_l(t) \, ds \, dt .$$

Theorem 1 justifies the convergence of the present method.

References

1. В а й н и к к о Г. Гладкость решения многомерного слабо сингулярного интегрального уравнения. В сб.: Методы решения интегральных, дифференциальных и операторных уравнений. Тарту, ТГУ, 1987, 3 -5.
2. В а й н и к к о Г., П е д а с А., У б а П. Методы решения слабо-сингулярных интегральных уравнений. – Тарту, ТГУ, 1984.
3. В а й н и к к о Г. Кусочно-полиномиальная аппроксимация решения многомерного слабо сингулярного интегрального уравнения// Уч. зап. Тарт. ун-та, 1988, вып. 833, 19-26.
4. З а в ь я л о в Ю.С., К в а с о в Б.И., М и р о ш н и ч е н к о В.Л. Методы сплайн-функций. М.: Наука, 1980.
5. К р а с н о с е л ь с к и й М.А., В а й н и к к о Г.М., З а б р е й к о П.П., Р у т и ц к и й Я.Б., С т е ц е н к о В.Я. Приближенное решение операторных уравнений. -М.: Наука, 1969.
6. У б а П. О сходимости интерполяционных кубических сплайнов на неравномерных сетках// Изв. АН ЭССР. Физ.-Матем., 1982, т.31, № 4, с. 399 - 409.
7. P i t k ä r a n t a J. Estimates for derivatives of solutions to weakly singular Fredholm integral equations// SIAM J. Math. Anal., 1980, 11, No 6, 952 -968.

МЕТОД КОЛЛОКАЦИИ С КУБИЧЕСКИМИ СПЛАЙНАМИ ДЛЯ РЕШЕНИЯ МНОГОМЕРНЫХ СЛАБО-СИНГУЛЯРНЫХ ИНТЕГРАЛЬНЫХ УРАВНЕНИЙ

П. Уба

Резюме

В работе рассматривается приближенное решение двумерного слабо-сингулярного интегрального уравнения (1), ядро $K(z,\zeta)$ которого удовлетворяет условиям (2) и (3). Предполагается, что свободный член уравнения $f(z)$, $z \in G$ ($G \in \mathbb{R}^2$, G - открытый прямоугольник) принадлежит специальному весовому пространству $C^{4,\nu}_{\square}(G)$ функций, которые удовлетворяют (4) и (4'). В [1] показано, что при сделанных предположениях все интегрируемые решения уравнения (1) принадлежат $C^{4,\nu}_{\square}(G)$, т.е. их производные могут иметь возле границы G особенности определенного порядка.

Приближенное решение уравнения (1) ищется в виде двумерного кубического сплайна минимального дефекта по обоим переменным. Требуется, чтобы в дополнение к коллокационным условиям (10) приближенное решение $u_N(z)$ удовлетворяло и краевым условиям (11), гарантирующим однозначность сплайна. Четвертый порядок сходимости метода достигается (теорема 1) специальным сгущением сетки возле границы области аналогично одномерному случаю, предложенному в [2,6].

Предложенный метод легко обобщаем для решения слабо-сингулярных интегральных уравнений на n-мерном ($n > 2$) параллелепипеде.

—

Acta et commentationes universitatis Tartuensis, 863, 1989, 26–30.

РЕШЕНИЕ ПАРАБОЛИЧЕСКОГО УРАВНЕНИЯ МЕТОДОМ СПЛАЙН-КОЛЛОКАЦИИ

Пирет Луйк, Энн Тамме, Галина Ханстейн

В работе построен алгоритм для решения параболического уравнения методом сплайн-коллокации, причем решение задачи аппроксимируется параболическими сплайнами. Исследуется устойчивость и сходимость метода. Аналогичные результаты на основе кубических сплайнов содержатся в статье [1].

Рассмотрим в прямоугольнике $D = \{(x, t):\ a \le x \le b,\ 0 \le t \le T\}$ решение задачи

$$\frac{\partial u}{\partial t} = Lu + f(x, t),$$
$$u(x, 0) = u^0(x), \quad u(a, t) = \varphi(t), \quad u(b, t) = \psi(t) \tag{1}$$

где

$$Lu = k(x)\,\frac{\partial^2 u}{\partial x^2} + r(x)\,\frac{\partial u}{\partial x} - q(x)\,u.$$

Пусть удовлетворены условия

$$k(x) > 0, \quad q(x) \ge 0 \quad \text{при} \quad x \in [a, b]. \tag{2}$$

Считаем, что k, r, q, u^0 непрерывны на отрезке $[a, b]$, φ, ψ непрерывны на $[0, T]$ и f непрерывна на D. Предположим, что задача (1) имеет на D решение $u(x, t)$.

Построим для решения задачи (1) численный алгоритм следующим образом. Введем узлы

$$\bar{x}_i = a + ih, \ i = -2, -1, \ldots, N+2; \quad h = (b-a)/N > 0;$$

$$t_n = n\tau, \quad n = 0, 1, \ldots, N_0, \quad \tau = [T/N_0] > 0.$$

Приближенное решение задачи ищем в виде квадратичного сплайна по x

$$\tilde{S}(x, t) = \sum_{j=0}^{N+1} \alpha_j(t)\, B_j(t),$$

где B_j – квадратичные B-сплайны:

$$B_j(x) = \frac{1}{2h^2}\begin{cases} (x-\bar{x}_{j-2})^2 & \text{при } x \in [\bar{x}_{j-2}, \bar{x}_{j-1}], \\ h^2 + 2(x-\bar{x}_{j-1})(\bar{x}_j - x) & \text{при } x \in [\bar{x}_{j-1}, \bar{x}_j], \\ (x-\bar{x}_{j+1})^2 & \text{при } x \in [\bar{x}_j, \bar{x}_{j+1}], \\ 0 & \text{при } x \notin (\bar{x}_{j-2}, \bar{x}_{j+1}). \end{cases}$$

Для определения коэффициентов α_j используем условия коллокации

$$\left[\frac{\partial\tilde S}{\partial t} - L\tilde S - f\right]_{x=x_i} = 0, \quad i = 1, 2,\dots, N,$$

и начальные и краевые условия

$$\tilde S(x_i, 0) = u^0(x_i), \quad \tilde S(a, t) = \varphi(t), \quad \tilde S(b, t) = \psi(t),$$

где $x_i = a + (i - 1/2)h = (\bar x_i + \bar x_{i-1})/2$. Таким образом получаем для определения α_j систему обыкновенных дифференциальных уравнений, которую решаем разностным методом

$$\begin{aligned} &\frac{S_i^n - S_i^{n-1}}{\tau} = \sigma(LS)_i^n + (1-\sigma)(LS)_i^{n-1} + \tilde f_i^n, \quad i = 1, 2,\dots, N, \\ &S(a, t_n) = \varphi^n, \quad S(b, t_n) = \psi^n, \quad n = 1, 2,\dots, N_0, \end{aligned} \tag{3}$$

где

$$S_i^n = S(x_i, t_n) = \sum_{j=0}^{N+1} \alpha_j^n B_j(x_i), \quad \tilde f_i^n = \sigma f_i^n + (1-\sigma) f_i^{n-1},$$

$$f_i^n = f(x_i, t_n), \quad \varphi^n = \varphi(t_n), \quad \psi^n = \psi(t_n), \quad 0 < \sigma \le 1.$$

Из соотношений (3) получается система линейных уравнений

$$\begin{aligned} &d_1(\sigma)\alpha_1^n + b_1(\sigma)\alpha_2^n = \tilde F_1^n, \\ &a_i(\sigma)\alpha_{i-1}^n + c_i(\sigma)\alpha_i^n + b_i(\sigma)\alpha_{i+1}^n = F_i^n, \quad i = 2,\dots, N-1, \\ &a_N(\sigma)\alpha_{N-1}^n + d_N(\sigma)\alpha_N^n = \tilde F_N^n, \end{aligned} \tag{4}$$

где

$$a_i(\sigma) = \frac{1}{8}(1 + \tau\sigma q_i) - \tau\sigma\left(\frac{k_i}{h^2} - \frac{r_i}{2h}\right),$$

$$c_i(\sigma) = \frac{3}{4}(1 + \tau\sigma q_i) + 2\tau\sigma\frac{k_i}{h^2},$$

$$b_i(\sigma) = \frac{1}{8}(1 + \tau\sigma q_i) - \tau\sigma\left(\frac{k_i}{h^2} + \frac{r_i}{2h}\right),$$

$$k_i = k(x_i), \quad r_i = r(x_i), \quad q_i = q(x_i),$$

$$d_1(\sigma) = c_1(\sigma) - a_1(\sigma), \quad d_N(\sigma) = c_N(\sigma) - b_N(\sigma),$$

$$F_i^n = a_i(\sigma-1)\alpha_{i-1}^{n-1} + c_i(\sigma-1)\alpha_i^{n-1} + b_i(\sigma-1)\alpha_{i+1}^{n-1} + \tau\tilde f_i, \quad i = 2,\dots, N-1,$$

$$\tilde F_1^n = d_1(\sigma-1)\alpha_1^{n-1} + b_1(\sigma-1)\alpha_2^{n-1} + \tau\tilde f_1^n + 2a_1(\sigma-1)\varphi^{n-1} - 2a_1(\sigma)\varphi^n,$$

$$\tilde F_N^n = a_N(\sigma-1)\alpha_{N-1}^{n-1} + d_1(\sigma-1)\alpha_N^{n-1} + \tau\tilde f_N^n + 2b_N(\sigma-1)\psi^{n-1} - 2b_N(\sigma)\psi^n.$$

Если выполнены условия (2) и

$$h\,|r(x)| \le 2k(x) \quad \text{при} \quad x \in [a, b],$$

то в системе (4) преобладает главная диагональ и, следовательно, эта система имеет единственное решение.

7*

Систему (4) можно записать в виде

$$A(\sigma)\alpha^n = A(\sigma-1)\alpha^{n-1} + \tau g^n, \quad n = 1, 2,\dots, N_0, \tag{5}$$

где $A(\sigma)$ – матрица этой системы,

$$\alpha^n = (\alpha_1^n, \alpha_2^n,\dots, \alpha_N^n)^T, \quad g^n = (g_1^n, \tilde f_2^n,\dots, \tilde f_{N-1}^n, g_N^n)^T,$$

$$g_1^n = \frac{2}{\tau}[a_1(\sigma-1)\varphi^{n-1} - a_1(\sigma)\varphi^n] + \tilde f_1^n,$$

$$g_N^n = \frac{2}{\tau}[b_N(\sigma-1)\psi^{n-1} - b_N(\sigma)\sigma^n] + \tilde f_N^n.$$

Вектор α^0 определяется как вектор коэффициентов интерполяционного квадратичного сплайна $S(x, 0)$, определенного условиями

$$S(x_i, 0) = u^0(x_i), \quad i = 1,\dots, N-1, \quad S(a, 0) = \varphi(0), \quad S(b, 0) = \psi(0).$$

Соотношение (5) является двухслойной разностной схемой (см., напр., [3]). Говорят, что эта схема равномерно устойчива, если для ее решения имеет место оценка

$$\|\alpha^n\| \le M_1\|\alpha^0\| + M_2\tau\sum_{j=1}^{n}\|g^j\|, \quad n = 1, 2,\dots, N_0,$$

где M_1 и M_2 – постоянные и

$$\|\alpha^n\| = \max_{1\le i\le N}|\alpha_i^n|.$$

Докажем устойчивость и сходимость рассматриваемого метода при $\sigma = 1$.

Теорема. Пусть $\sigma = 1$, удовлетворены условия (2) и

$$\frac{1}{h^2}(k_i \quad \frac{h}{2}|r_i|) \ge \frac{1}{8}, \quad i = 1, 2,\dots,N.$$

Тогда схема (5) равномерно устойчива. Если, кроме того, решение задачи (1) $u(x, t)$ имеет в D непрерывные производные по x до четвертого порядка и по t до второго порядка, то имеет место оценка

$$|S(x, t_n) - u(x, t_n)| \le c(h^2 + \tau), \quad x \in [a, b], \quad n = 1, 2,\dots, N_0, \tag{6}$$

где c - постоянная и

$$S(x, t_n) = \sum_{j=1}^{N+1}\alpha_j^n B_j(x), \quad \alpha_0^n = -\alpha_1^n + 2\varphi^n, \quad \alpha_{N+1}^n = -\alpha_N^n + 2\psi^n.$$

<u>Доказательство.</u> При $\sigma = 1$ система (5) принимает вид

$$A(1)\alpha^n = A(0)\alpha^{n-1} + \tau g^n.$$

Из принципа максимума следует существование обратной матрицы $[A(1)]^{-1}$ и оценка $\|[A(1)]^{-1}\| \le 1$. Учитывая еще равенство $\|A(0)\| = 1$, получаем

$$\|\alpha^n\| < \|\alpha^{n-1}\| + \tau\|g^n\|,$$

откуда следует оценка

$$\|\alpha^n\| \le \|\alpha^0\| + \tau\sum_{j=1}^{n}\|g^j\|, \quad n = 1, 2,\dots, N_0. \tag{8}$$

Таким образом, метод (5) при $\sigma = 1$ равномерно устойчив.

При выводе оценки (6) используем интерполяционные квадратичные сплайны

$$\bar{S}(x, t_n) = \sum_{j=0}^{N+1} \bar{\alpha}_j^n B_j(x), \quad n = 0, 1, \dots, N_0,$$

определенные условиями

$$\bar{S}(x_i, t_n) = u(x_i, t_n), \quad i = 1, 2, \dots, N,$$

$$\bar{S}(a, t_n) = u(a, t_n) - \frac{h^4}{128} \frac{\partial^4(a, t_n)}{\partial x^4},$$

$$\bar{S}(b, t_n) = u(b, t_n) - \frac{h^4}{128} \frac{\partial^4(b, t_n)}{\partial x^4}.$$

Для такого сплайна имеют место оценки (см. [2])

$$\max_{a \le x \le b} |\bar{S}(x, t_n) - u(x, t_n)| = O(h^3), \tag{9}$$

$$\max_{1 \le i \le N} \left| \frac{\partial^s \bar{S}(x_i, t_n)}{\partial x^s} - \frac{\partial^s u(x_i, t_n)}{\partial x^s} \right| = O(h^2), \; s = 1, 2, \; n = 0, 1, \dots, N_0. \tag{10}$$

На правой стороне неравенства

$$|S(x, t_n) - u(x, t_n)| \le |S(x, t_n) - \bar{S}(x, t_n)| + |\bar{S}(x, t_n) - u(x, t_n)| \tag{11}$$

второй член оцениваем при помощи (9), а первый член следующим образом. Подставив $\bar{S}(x, t_n)$ в (3) получаем для $\bar{\alpha}^n = (\bar{\alpha}_1^n, \bar{\alpha}_2^n, \dots, \bar{\alpha}_N^n)^T$ уравнение

$$A(1)\bar{\alpha}^n = A(0)\bar{\alpha}^{n-1} + \tau g^n + \tau \varepsilon^n, \quad n = 1, 2, \dots, N_0 \tag{12}$$

При помощи (10) получается оценка $\|\varepsilon^n\| = O(h^2 + \tau)$. Вычисляя из (12) равенство (5), получаем

$$A(1)(\bar{\alpha}^n - \alpha^n) = A(0)(\bar{\alpha}^0 - \alpha^0) + \tau \varepsilon^n.$$

Из неравенства (8) следует

$$\|\bar{\alpha}^n - \alpha^n\| \le \|\bar{\alpha}^0 - \alpha^0\| + \tau \sum_{k=1}^{n} \|\varepsilon^n\| = O(h^2 + \tau), \quad n = 1, 2, \dots, N_0.$$

Пользуясь свойствами B-сплайнов

$$B_j(x) \ge 0, \qquad \sum_{j=0}^{N+1} B_j(x) = 1, \quad x \in [a, b],$$

оценим

$$\max_{a \le x \le b} |S(x, t_n) - \bar{S}(x, t_n)| = \max_{a \le x \le b} \left| \sum_{j=0}^{N+1} (\alpha_j^n - \bar{\alpha}_j^n) B_j(x) \right| \le$$

$$\le \max_{0 \le j \le N+1} |\alpha_j^n - \bar{\alpha}_j^n| = \max\left(\|\alpha^n - \bar{\alpha}^n\|, |\alpha_0^n - \bar{\alpha}_0^n|, |\alpha_{N+1}^n - \bar{\alpha}_{N+1}^n|\right) = O(h^2 + \tau).$$

Из этой оценки и из (11) и (9) следует оценка (6).

Теорема доказана.

Литература

1. Жанлав Т., Мирошниченко В. Метод сплайн-коллокации для параболических уравнений с непрерывными и разрывными коэффициентами // Вычислительные системы. 1981, вып. 87, с. 77-98.
2. Оя П., Рейтсекас А. О методах коллокации и подобластей квадратичными и кубическими сплайнами для краевых задач // Изв. АН Эст. ССР. Физ. Мат. 1987, т. 36, № 2, с. 118-128.
3. Самарский А. А. Введение в численные методы. Узд. 2-ое, перераб. и доп. – М., 1987.

SPLINE-COLLOCATION METHOD FOR SOLVING PARABOLIC EQATIONS

P. Luik, E. Tamme, G. Hanstein

Summary

We construct an algorithm for solving parabolic equation (1) using the spline-collocation method with quadratic splines. The stability and the convergence of the method is proved.

Acta et commentationes universitatis Tartuensis, 863, 1989, 31–39.

КУСОЧНО-ПОСТОЯННАЯ АППРОКСИМАЦИЯ РЕШЕНИЯ СЛАБО-ОСОБОГО ИНТЕГРАЛЬНОГО УРАВНЕНИЯ

Геннадий Вайникко, Арвет Педас

Рассматривается кусочно-постоянная аппроксимация решения линейного одномерного интегрального уравнения второго рода и проблемы собственных значений для указанных уравнений. Предполагается, что ядро $K(t,s)$ интегрального уравнения может при $s = t$ иметь логарифмическую или степенную интегрируемую особенность, а при $s = d$ ($d = const$) оно может иметь разрыв первого рода. Выводятся оценки погрешности приближенного решения.

I. Постановка задачи. Рассмотрим уравнение

$$u(t) = \int_0^b a(t,s)\varkappa(t-s)u(s)ds + f(t), \quad 0 \le t \le b. \tag{I}$$

Введём следующие условия:

(i) $f \in C^2[0,b]$;

(ii) $a \in C^2([0,b] \times ([0,b] \setminus \{d\}))$, $0 < d < b$,

причём $\partial^k a(t,s)/\partial s^k$ $(0 \le k < \beta)$ могут иметь при $s = d$ разрыв первого рода;

(iii) $\varkappa \in C^1([-b,b] \setminus \{0\})$,

причём при $-b \le t < 0$ и $0 < t \le b$ справедлива оценка[I)]

$$|\varkappa'(t)| \le c\,|t|^{-\beta}, \quad 0 < \beta < 2. \tag{2}$$

Заметим, что из (2) вытекают неравенства

$$|\varkappa(t)| \le c\left(|t|^{-\beta+1} + 1\right) \quad (\beta \ne 1), \tag{3}$$

$$|\varkappa(t)| \le c\left(|\ln|t|| + 1\right) \quad (\beta = 1). \tag{4}$$

I) Буквой c обозначаем положительные постоянные, которые в разных неравенствах могут принимать разные значения.

Зададим сетку точек $t_j = t_{j,n}$ отрезка $[0, b]$ такую, что

$$0 = t_0 < t_1 < \ldots < t_n = b, \tag{5}$$

$$d \in \{t_1, \ldots, t_{n-1}\}, \tag{6}$$

$$h_n \equiv \max_{1 \le j \le n} (t_j - t_{j-1}) \longrightarrow 0 \quad \text{при } n \to \infty. \tag{7}$$

Приближённое решение уравнения (I) будем искать в виде кусочно-постоянной функции

$$u_n(t) = \sum_{j=1}^{n} u_j \varphi_j(t), \tag{8}$$

где $\varphi_1(t) = 1$ при $t \in [t_0, t_1]$ и $\varphi_1(t) = 0$ при $t \notin [t_0, t_1]$; $\varphi_j(t) = 1$ при $t \in (t_{j-1}, t_j]$ и $\varphi_j(t) = 0$ при $t \notin (t_{j-1}, t_j]$, $j = 2, \ldots, n$. Неизвестные коэффициенты $u_1, \ldots, u_n$ определим по методу коллокации из системы линейных алгебраических уравнений

$$u_i = \sum_{j=1}^{n} \Big(\int_{t_{j-1}}^{t_j} a(s_i, s) \varkappa(s_i - s) ds \Big) u_j + f(s_i), \; i = 1, \ldots, n, \tag{9}$$

где

$$s_i = (t_{i-1} + t_i)/2, \; i = 1, \ldots, n. \tag{10}$$

<u>Теорема I</u>. Пусть выполнены условия (i)–(iii) и (5)–(7). Пусть уравнение (I) имеет единственное решение $u(t)$.

Тогда система уравнений (9) имеет при достаточно больших n единственное решение $(u_1, \ldots, u_n)$. Справедлива оценка

$$\max_{1 \le i \le n} |u_i - u(s_i)| \le c \cdot \varepsilon_n^2, \tag{11}$$

где

$$\varepsilon_n = \begin{cases} h_n & \text{при } 0 < \beta < 1, \\ h_n(|\ln h_n| + 1) & \text{при } \beta = 1, \\ h_n^{2-\beta} & \text{при } 1 < \beta < 2. \end{cases} \tag{12}$$

<u>Замечание I</u>. В условиях (i)–(iii) решение $u(t)$ уравнения (I) негладко. Точнее ([2,5]; ср. также [3], стр. 6–9), $u \in C[0, b] \cap C^2((0, b) \setminus \{d\})$, причём при $\beta \ne 1$

$(0<\beta<2)$ справедливы оценки[1)]

$$|u^{(k)}(t)|\leq c\left[t^{-\beta+2-k}+|t-d|^{-\beta+2-k}+(b-t)^{-\beta+2-k}\right],\ t\in(0,b)\setminus\{d\} \quad (13)$$

В случае $\beta=1$ справедливы оценки

$$|u'(t)|\leq c\left[\,|\ln t|+|\ln|t-d||+|\ln(b-t)|+1\right],\ t\in(0,b)\setminus\{d\}, \quad (14)$$

$$|u''(t)|\leq c\left[\frac{1}{t}+\frac{1}{|t-d|}+\frac{1}{b-t}\right],\ t\in(0,b)\setminus\{d\}. \quad (15)$$

Поэтому добиться высокого порядка точности приближённых методов для решения уравнения (I) в условиях (i)-(iii) довольно сложно. Равномерная по $t\in[0,b]$ оценка порядка $O(h_n)$ метода (8),(9) при специальном выборе сетки имеется в [6,3,5]. Для модифицированного метода квадратурных формул в [4] получена оценка порядка (II) при более жестких условиях гладкости на $a(t,s)$ и $\varkappa(t)$.

Замечание 2. Ослабим условие (i): пусть $f\in C[0,b]\cap C^2((0,b)\setminus\{d\})$ удовлетворяет оценкам, указанным в замечании I. Тогда (см. [2]) утверждения замечания I, а вместе с ними и теоремы I остаются в силе.

2. Доказательство теоремы I. Уравнение (I) рассмотрим как операторное уравнение $u=Tu+f$ в банаховом пространстве $C[0,b]$, а систему уравнений (9) как операторное уравнение

$$\underline{u}_n = T_n\underline{u}_n + p_n f \quad (16)$$

в банаховом пространстве m_n векторов вида $\underline{u}_n=(u_1,\dots,u_n)$ с нормой $\|\underline{u}_n\| = \max\limits_{1\leq i\leq n}|u_i|$. Здесь T и T_n - линейные вполне непрерывные операторы соответственно в пространствах $C[0,b]$ и m_n, задаваемые формулами

$$(Tu)(t)=\int_0^b a(t,s)\varkappa(t-s)u(s)ds,$$

$$(T_n\underline{u}_n)_i=\sum_{j=1}^{n}\left[\int_{t_{j-1}}^{t_j}a(s_i,s)\varkappa(s_i-s)ds\right]u_j,$$

I) Оценки (I3)-(I5) соответствуют случаю, когда $a(t,s)$ имеет при $s=d$ разрыв. Если $a(t,s)$ непрерывна, и в точке $s=d$ разрыв имеется у $\partial a(t,s)/\partial s$, то оценки (I3)-(I5) допускают уточнение (см. [5]). В (13) $k=1,2$.

а p_n - линейные непрерывные операторы из $C[0,b]$ в m_n задаваемые формулами

$$p_n u = (u(s_1), \ldots, u(s_n)) \quad (u \in C[0,b]).$$

Из ($\ddot{u}$)-($\dddot{u}$) следует, что последовательность операторов $\{T_n\}$ компактно сходится ([I], стр.32) к оператору T относительно связывающих отображений p_n :

$$T_n \longrightarrow T \qquad \text{компактно.} \tag{I7}$$

По теореме сходимости для операторных уравнений ([I], стр. 49) получаем, что при достаточно больших n уравнение (I6) имеет единственное решение $\underline{u}_n$ и справедлива оценка

$$\| \underline{u}_n - p_n u \|_{m_n} \leq c \cdot \| p_n T u - T_n p_n u \|_{m_n} . \tag{I8}$$

Для завершения доказательства теоремы I достаточно показать что

$$\| p_n T u - T_n p_n u \|_{m_n} \leq c \cdot \varepsilon_n^2 . \tag{I9}$$

Имеем

$$\| p_n T u - T_n p_n u \|_{m_n} = \max_{1 \leq i \leq n} \left| \sum_{j=1}^{n} \int_{t_{j-1}}^{t_j} K(s_i, s)[u(s) - u(s_j)] ds \right| \leq$$

$$\leq \max_{1 \leq i \leq n} \left(I_i^{(1)} + I_i^{(2)} + I_i^{(3)} + I_i^{(4)} \right),$$

где

$$K(t,s) = a(t,s) \varkappa(t-s),$$

$$I_i^{(1)} = \left| \sum_{j \in J(i)} \int_{t_{j-1}}^{t_j} K(s_i, s)[u(s) - u(s_j)] ds \right|,$$

$$I_i^{(2)} = \left| \sum_{j \in J} \int_{t_{j-1}}^{t_j} K(s_i, s)[u(s) - u(s_j)] ds \right|,$$

$$I_i^{(3)} = \left| \sum_{j \notin J(i), j \notin J} \int_{t_{j-1}}^{t_j} [K(s_i, s) - K(s_i, s_j)][u(s) - u(s_j)] ds \right|,$$

$$I_i^{(4)} = \left| \sum_{j \notin J(i), j \notin J} \int_{t_{j-1}}^{t_j} K(s_i, s_j)[u(s) - u(s_j)] ds \right|,$$

$$\mathcal{J}(i)=\{j: |t_j-s_i|\le h_n, |t_{j-1}-s_i|\le h_n\},$$

$$\mathcal{J}=\{j: t_{j-1}\le h_n, b-t_j\le h_n, |t_j-d|\le h_n, |t_{j-1}-d|\le h_n\}.$$

Покажем, что

$$I_i^{(k)}\le c\cdot\varepsilon_n^2,\quad i=1,\dots,n;\ k=1,\dots,4.$$

Обозначим

$$\gamma_i(h_n)=[0,b]\cap[s_i-h_n, s_i+h_n],$$

$$\eta_i(h_n)=\{s\in[0,b]: h_n\le s\le b-h_n, |s-d|\ge h_n, |s-s_i|\ge h_n\}.$$

Из (ii)–(iii) вытекает неравенство

$$\int_{\gamma_i(h_n)}|K(s_i,s)|\,ds\le c\cdot\varepsilon_n\quad(i=1,\dots,n).\qquad(20)$$

Из (13) и (14) в силу соотношения

$$u(s)-u(s_j)=\int_{s_j}^{s}u'(t)\,dt$$

вытекает неравенство

$$\max_{s\in[t_{j-1},t_j]}|u(s)-u(s_j)|\le c\cdot\varepsilon_n\quad(j=1,\dots,n).\qquad(21)$$

Поэтому

$$I_i^{(1)}\le c\cdot\varepsilon_n\cdot\sum_{j\in\mathcal{J}(i)}\int_{t_{j-1}}^{t_j}|K(s_i,s)|\,ds\le$$

$$\le c\cdot\varepsilon_n\cdot\int_{\gamma_i(2h_n)}|K(s_i,s)|\,ds\le c\cdot\varepsilon_n^2\quad(i=1,\dots,n).$$

Из (20) и (21) получаем также $I_i^2\le c\cdot\varepsilon_n^2,\ i=1,\dots,n$.
В силу (21) имеем

$$I_i^{(3)}\le c\cdot\varepsilon_n\cdot\sum_{j\notin\mathcal{J}(i),\,j\notin\mathcal{J}}\int_{t_{j-1}}^{t_j}|K(s_i,s)-K(s_i,s_j)|\,ds\le$$

9*

$$\leq c\cdot\varepsilon_n\cdot h_n\cdot\sum_{j\notin J(i),\,j\notin J}\int_{t_{j-1}}^{t_j}\sup_{0<\theta<1}\left|\frac{\partial K(s_i,\theta s+(1-\theta)s_j)}{\partial s}\right|ds.$$

Заметим, что для $j\notin J(i),\ s\in[t_{j-1},t_j],\ \theta\in[0;1]$ имеем

$$0<c_1\leq|s_i-[\theta s+(1-\theta)s_j]|\big/|s_i-s|\leq c_2,$$

поэтому оценка с учетом (2) примет вид

$$I_i^{(3)}\leq c\cdot\varepsilon_n\cdot h_n\cdot\int\limits_{\eta_i(h_n)}|s_i-s|^{-\beta}ds\leq c\cdot\varepsilon_n^2\quad(i=1,\dots,n).$$

Оценим $I_i^{(4)}$. Для $j\notin J(i),\ j\notin J$ имеем (см. (I0))

$$\int_{t_{j-1}}^{t_j}K(s_i,s_j)u'(s_j)(s-s_j)ds=0.$$

Поэтому

$$I_i^{(4)}\leq\sum_{j\notin J(i),\,j\notin J}\int_{t_{j-1}}^{t_j}|K(s_i,s_j)|\sup_{0<\theta<1}|u''(\theta s+(1-\theta)s_j)||s-s_j|^2ds.$$

В силу (I3),(I5),(3) и (4) отсюда получаем оценку

$$I_i^{(4)}\leq c\cdot\varepsilon_n^2\ ,\ i=1,\dots,n,$$

что завершает доказательство (I9) и теоремы I.

3. <u>Проблема собственных значений</u>. Рассмотрим уравнение

$$\lambda u(t)=\int_0^b a(t,s)\varkappa(t-s)u(s)ds. \tag{22}$$

Аппроксимирующую конечномерную задачу для (22) построим в виде

$$\lambda u_i=\sum_{j=1}^{n}\Big[\int_{t_{j-1}}^{t_j}a(s_i,s)\varkappa(s_i-s)ds\Big]u_j\ ,\ i=1,\dots,n, \tag{23}$$

где t_j и s_i определены соответственно в (5)-(7) и (I0).

Уравнение (22) рассмотрим как операторное уравнение $\lambda u=Tu$ в банаховом пространстве $C[0,b]$ и систему уравнений (23) как операторное уравнение $\lambda\underline{u}_n=T_n\underline{u}_n$ в банаховом пространстве m_n.

<u>Теорема 2</u>. Пусть выполнены условия ($\ddot{u}$)-($\dddot{u}$) и (5)-(7).

Тогда для каждого ненулевого собственного значения λ_0 уравнения (22) найдётся последовательность $\{\lambda_n\}$ собственных значений систем уравнений (23) такая, что $\lambda_n \to \lambda_0$ при $n \to \infty$. Обратно, каждая ненулевая предельная точка любой последовательности $\{\lambda_n\}$ собственных значений систем уравнений (23) является собственным значением уравнения (22).

<u>Доказательство</u>. Теорема 2 следует на основе (I7) непосредственно из общей теоремы о сходимости собственных значений ([I], стр. 68–74).

Следуя [3], введём обозначения для собственного подпространства

$$V = V(\lambda_0; T) = N(\lambda_0 I - T) = \{u \in C[0,b] : (\lambda_0 I - T)u = 0\}$$

и корневого подпространства

$$W = W(\lambda_0; T) = N((\lambda_0 I - T)^{\ell}).$$

Здесь I – тождественное отображение, а ℓ – ранг собственного значения λ_0, т.е. наименьшее натуральное число, для которого

$$N((\lambda_0 I - T)^{\ell}) = N((\lambda_0 I - T)^{\ell+1}).$$

Пусть $\delta > 0$ – такое число, что в круге $|\lambda - \lambda_0| \leq \delta$ нет других собственных значений уравнения (22), кроме λ_0. Из теоремы 2 следует, что при достаточно больших n в этот круг попадает хотя бы одно собственное значение задачи (23). Пусть $\lambda_n^{(i)}$ ($i = 1, \ldots, k_n$) – попарно различные собственные значения задачи (23), попавшие в указанный круг,

$$\nu_n^{(i)} = \dim W(\lambda_n^{(i)}; T_n)$$

– их корневые кратности. Обозначим

$$\hat{\lambda}_n = \Big[\sum_i \nu_n^{(i)} \lambda_n^{(i)}\Big] \cdot \Big[\sum_i \nu_n^{(i)}\Big]^{-1}$$

– это среднее арифметическое чисел $\lambda_n^{(i)}$ с учетом их корневых кратностей. Линейную оболочку корневых подпространств $W(\lambda_n^{(i)}; T_n)$, $i = 1, \ldots, k_n$, обозначим через $W_n = W(\lambda_0; T_n; \delta)$. Собственное подпространство матрицы T_n для $\lambda = \lambda_n^{(i)}$ обозначим через

$$V_n = V(\lambda_n^{(i)}; T_n) = N(\lambda_n^{(i)} I - T_n) = \{\underline{u}_n \in m_n : (\lambda_n^{(i)} I - T_n)\underline{u}_n = 0\}.$$

Теорема 3. Пусть выполнены условия ($\ddot{u}$)-($\dddot{u}$) и (5)-(7). Пусть $\lambda_n \to \lambda_0 \neq 0$,где λ_0 и λ_n - собственные значения уравнения (22) и систем уравнений (23) соответственно, причём λ_0 имеет ранг ℓ . Пусть, наконец, $\delta > 0$ - такое число, что в круге $|\lambda - \lambda_0| \leq \delta$ нет других собственных значений уравнения (22), кроме λ_0 .

Тогда справедливы следующие оценки

$$|\lambda_n - \lambda_0| \leq const \cdot \varepsilon_n^{2/\ell}, \quad |\hat{\lambda}_n - \lambda_0| \leq const \cdot \varepsilon_n^2,$$

$$\sup_{\underline{u}_n \in V_n, \|\underline{u}_n\|_{m_n} = 1} \ \inf_{u \in V} \ \max_{1 \leq i \leq n} |u_i - u(s_i)| \leq const \cdot \varepsilon_n^{2/\ell},$$

$$\sup_{\underline{u}_n \in W_n, \|\underline{u}_n\|_{m_n} = 1} \ \inf_{u \in W} \ \max_{1 \leq i \leq n} |u_i - u(s_i)| \leq const \cdot \varepsilon_n^2,$$

$$\sup_{u \in W, \|u\|_{C[0,b]} = 1} \ \inf_{\underline{u}_n \in W_n} \ \max_{1 \leq i \leq n} |u_i - u(s_i)| \leq const \cdot \varepsilon_n^2,$$

где ε_n - определенная в (I2) величина.

Доказательство. Теорема 3 непосредственно следует из (I7)-(I9) и общей теоремы о сходимости для проблемы собственных значений ([I], стр. 68-74).

Замечание 3. Если $a(t,s)$ или $\partial a(t,s)/\partial s$ имеет несколько точек разрыва $s = d_k$, $k = 1, \ldots, m$ (ср. ($\ddot{u}$)), то, включая их в сетку (ср. (5)-(7)), теоремы I-3 сохраняют силу.

Литература

I. В а й н и к к о Г.М. Анализ дискретизационных методов. Тарту, I976.

2. В а й н и к к о Г. Гладкость решения многомерного слабо-сингулярного интегрального уравнения. В сб. Методы решения интегральных, дифференциальных и операторных уравнений. Тарту, I987, 3-5.

3. В а й н и к к о Г., П е д а с А., У б а П. Методы решения слабо-сингулярных интегральных уравнений., Тарту, I984.

4. П е д а с А. О решении слабо-сингулярных уравнений методом механических квадратур с формулой трапеций. Уч.зап.Тартуск.ун-та, I986, 762, 89-97.

5. У б а П. Кусочно полиномальная аппроксимация для слабо-сингулярного интегрального уравнения с разрывном коэффициентом. В сб. Методы решения интегральных и операторных уравнений., Тарту, 1987, 6-8.

6. V a i n i k k o G., U b a P. A piecewise polynomial approximation to the solution of an integral equation with weakly singular kernel. J. Austral. Math. Soc., Series B, 1981, v. 22, 431-438.

A PIECEWISE CONSTANT APPROXIMATION TO THE SOLUTION OF A WEAKLY SINGULAR INTEGRAL EQUATION

G. Vainikko, A. Pedas

Summary

Let integral equation (1) satisfy conditions (i)-(iii). Introduce a grid on $[0, b]$ satisfying conditions (5)-(7) and consider the collocation method (8)-(10) with piecewise constant approximation to the solution of equation (1). Assuming that equation (1) has a unique solution $u(t)$, error estimate (11)-(12) is proved (Theorem 1). Similar results hold for eigenvalue problems (22) and (23), see Theorem 3.

Acta et commentationes universitatis Tartuensis, 863, 1989, 40-56.

ON THE CONVERGENCE OF EIGENVALUES BY APPROXIMATION OF THE PROBLEM

Otto Karma

Approximation of the eigenvalue problem $A(\lambda)u=0$ is studied with $A(\cdot)$ a holomorphic Fredholm operator-function. The discrete approximation scheme for spaces and the regular approximation scheme for operator-functions are used. The convergence rate of the approximate eigenvalues to the exact ones is estimated by the powers of the approximation error on the sphere of the generalized eigenspace. Two estimations are given - estimation (2.1) for every eigenvalue individually and estimation (2.2) for the weighed arithmetic mean of all eigenvalues converging to the one and the same exact eigenvalue. The article is shortened variant of [10]. See review in [1], and [2,3,8,9,13,15] on this topic, too.

1. General Assumptions and Basic Notions

1.1 (Exact problem). Let U, V be complex Banach spaces and $A(\cdot)$ be an operator-function from region (open connected set) $\Lambda\in\mathbb{C}$ to $\mathcal{B}(U,V)$ (Banach space of bounded linear operators with norm $\|A\|=\sup\{\|Au\|: u\in U, \|u\|=1\}$). Let us examine the eigenvalue problem

$$A(\lambda)u=0, \quad u\neq 0. \tag{1.1}$$

(Special cases are $A(\lambda)=A-\lambda I$ and $A(\lambda)=A-\lambda B$.)

The points $\lambda\in\Lambda$, for which the equation (1.1) has a solution, are the <u>eigenvalues</u> of $A(\cdot)$. These solutions themselves are the <u>eigenelements</u>.

Recall that $A\in\mathcal{B}(U,V)$ is a Fredholm operator if the dimension of the nullspace $\mathcal{N}(A)$ is finite, the range $\mathcal{R}(A)$ is closed and the *codim* $\mathcal{R}(A)$ is finite. Then the index of A is given by $ind A = dim\mathcal{N}(A) - codim\mathcal{R}(A)$.

The operator-function $A(\cdot):\Lambda\to\mathcal{B}(U,V)$ is <u>holomorphic</u> on Λ if it is continuous and complex-differentiable at every point of Λ. Such a function has at each point $\lambda_0\in\Lambda$ a unique Taylor expansion

$$A(\lambda)=\sum_{j=0}^{j=\infty}(\lambda-\lambda_0)^j A_j, \quad A_j=\frac{1}{j!}A^{(j)}(\lambda_0)=\frac{1}{j!}\frac{d^j}{d\lambda^j}A(\lambda_0)\in\mathcal{B}(U,V).$$

which is norm-convergent in some neighborhood of λ_0. If $\Lambda'\subset\Lambda$ is a bounded region with a smooth boundary $\Gamma\subset\Lambda$, then at each $\lambda_0\in\Lambda'$ we have $A_j=(2\pi)^{-1}\int_\Gamma(\lambda-\lambda_0)^{-j-1}A(\lambda)d\lambda$ (see e.g. [5]).

We assume that

a1) $A(\cdot)$ is a holomorphic function of λ on Λ,

a2) for every λ in Λ the operator $A(\lambda)$ is a Fredholm operator,

a3) the <u>resolvent set</u> $\rho(A)=\{\lambda\in\Lambda: \exists A(\lambda)^{-1}\in\mathcal{B}(U,V)\}$ is not empty.

From these assumptions a1)-a3) it follows that for every $\lambda\in\Lambda$ the index of $A(\lambda)$ is 0, the spectrum $\sigma(A)=\Lambda\setminus\rho(A)$ of $A(\cdot)$ has no cluster points in Λ, and $\sigma(A)$ consists only of eigenvalues of $A(\cdot)$. Moreover, the operator-function $A^{-1}(\cdot)$, defined on $\rho(A)$ by the equality $A^{-1}(\lambda)=A(\lambda)^{-1}$, is holomorphic on $\rho(A)$ and has poles of finite order $\varkappa(\lambda_0,A)$ at every point $\lambda_0\in\sigma(A)$ (see e.g. [4]).

1.2 (Approximation of the spaces). Let $\{X_n\}_{n\in\mathbb{N}}$ and $\{Y_n\}_{n\in\mathbb{N}}$ be two sequences of complex Banach spaces which, together with the sequences of the connecting operators $\{p_n:U\to X_n\}_{n\in\mathbb{N}}$ and $\{q_n:V\to Y_n\}_{n\in\mathbb{N}}$ form the discrete approximations for U and V, correspondingly. That is [14]:

$$\|p_nu\|_{X_n}\to\|u\|_U\quad(n\in\mathbb{N})\quad\forall u\in U,$$

$$\|q_nv\|_{Y_n}\to\|v\|_V\quad(n\in\mathbb{N})\quad\forall v\in V,$$

$$\|p_n(\alpha u+\alpha'u')-(\alpha p_nu+\alpha'p_nu')\|_{X_n}\to 0\quad(n\in\mathbb{N})\quad\forall u,u'\in U,\ \alpha,\alpha'\in\mathbb{C},$$

$$\|q_n(\alpha v+\alpha'v')-(\alpha q_nv+\alpha'q_nv')\|_{Y_n}\to 0\quad(n\in\mathbb{N})\quad\forall v,v'\in V,\ \alpha,\alpha'\in\mathbb{C}.$$

(Special cases are as follows: 1) $X_n\subset U$, p_n-linear projectors onto X_n, strongly converging to identity, 2) $U=C[0,1]$, X_n the set of $(n+1)$-dimensional vectors with maximum norm, $p_nu=(\,u(0),\,u(1/n),\,\dots,\,u(1)\,)$.) Typically in practical applications $p_n\in\mathcal{B}(U,X_n)$, $q_n\in\mathcal{B}(V,Y_n)$.

Denote infinite subsequences of $\mathbb{N}$ by N', N'', ... By the definition the sequence $\{x_n\}_{n\in N}$ of elements x_n with $x_n\in X_n$:

- is converging to the element $u\in U$ if $\|p_nu-x_n\|\to 0$ $(n\in N')$; we will write it $x_n\to u$ $(n\in N')$,

- is compact if for every subsequence $\{x_n\}_{n\in N''}$ with $N''\subset N'$ there exist $N'''\subset N''$ and $u\in U$ so that $x_n\to u$ $(n\in N''')$.

Note the following properties of this convergence :

$$x_n\to u\ (n\in N')\ \Rightarrow\ \|x_n\|\to\|u\|\ (n\in N'),$$

$$x_n\to 0\ (n\in N')\ \Rightarrow\ \|x_n\|\to 0\ (n\in N'),$$

$$x_n\to u,\ x_n\to u'\ (n\in N')\ \Rightarrow\ u=u',$$

$$x_n\to u,\ x_n'\to u',\ \alpha_n\to\alpha,\ \alpha_n'\to\alpha'\ (n\in N')\ \Rightarrow\ \alpha_nx_n+\alpha_n'x_n'\to\alpha u+\alpha'u'\ (n\in N').$$

The convergence defined by the operators p_n $(n\in\mathbb{N})$ coincide with the convergence defined by any operators p_n': $U\to X_n$ $(n\in\mathbb{N})$ such that $\|p_nu-p_n'u\|\to 0$ $(n\in\mathbb{N})$. The operators p_n' $(n\in\mathbb{N})$ can always be chosen to be linear on every fixed finite-dimensional subspace of U. Moreover, it is shown, that if U is separable and there exists a sequence of finite-dimensional operators converging strongly to identity in U, then p_n' $(n\in\mathbb{N})$ can

II

be chosen linear and continuous on the whole U.

1.3 (Approximation of the Eigenvalue Problem). Let the problem (1.1) be approximated by a sequence of problems

$$B_n(\lambda)x_n=0\ ,\ x_n\neq 0 \tag{1.2}$$

with $B_n(\cdot)$, $n\in\mathbb{N}$ the operaton-functions from the region $\Lambda\subseteq\mathbb{C}$ to $\mathcal{B}(X_n,Y_n)$ (the region Λ being the same region for all $B_n(\cdot)$ and $A(\cdot)$).

We assume that

b1) all $B_n(\cdot)$ are holomorphic functions of λ on Λ,

b2) for every $n\in\mathbb{N}$ and $\lambda\in\Lambda$ the operator $B_n(\lambda)$ is Fredholm operator with index 0,

b3) the sequence $\{B_n(\cdot)\}_{n\in\mathbb{N}}$ converges regularly to $A(\cdot)$ on Λ, i.e.:

b3.0) on every compact $\Lambda_0\subset\Lambda$ the norms $\|B_n(\lambda)\|$ are uniformly bounded: $\|B_n(\lambda)\|\le c(\Lambda_0)<\infty\ \forall n\in\mathbb{N},\lambda\in\Lambda_0$,

b3.1) for every $\lambda\in\Lambda$ the sequence of operators $\{B_n(\lambda)\}_{n\in\mathbb{N}}$ converges to $A(\lambda)$: $x_n\to u\ (n\in\mathbb{N})\Rightarrow B_n(\lambda)x_n\to A(\lambda)u\ (n\in\mathbb{N})$,

b3.2) for every $\lambda\in\Lambda$ the sequence of operators $\{B_n(\lambda)\}_{n\in\mathbb{N}}$ is regular:

$$\|x_n\|\le c,\ \{B_n(\lambda)x_n\}_{n\in\mathbb{N}} \text{ is compact } \Rightarrow \{x_n\}_{n\in\mathbb{N}} \text{ is compact.}$$

(Special case is: $X_n=Y_n=U=V$, $p_n=q_n\in\mathcal{B}(U,X_n)$ projectors, $A(\lambda)=I+K(\lambda)$ with $K(\cdot)$ holomorphic on Λ and $K(\lambda)$ compact for all $\lambda\in\Lambda$, $B_n(\lambda)=q_nA(\lambda)|_{X_n}$.)

It is useful to note that:

(1) from b1) and b3.0) it follows that on every compact $\Lambda_0\subset\Lambda$ the norms of the derivatives $\|B_n^{(j)}(\lambda)\|$ ($\lambda\in\Lambda_0$, $n\in\mathbb{N}$) are uniformly bounded for every fixed $j=1,2,\ldots$ and that the set of operator-functions $\{B_n(\cdot)\}_{n\in\mathbb{N}}$ is equicontinuous on Λ_0 i.e.

$$\exists c(\Lambda_0)<\infty\ :\ \|B_n(\lambda)-B_n(\lambda')\|\le c(\Lambda_0)|\lambda-\lambda'| \qquad \forall n\in\mathbb{N},\ \lambda,\ \lambda'\in\Lambda_0,$$

(2) if b1) holds, then the regular convergence on Λ is equivalent to the following assumption b3'):

b3') for every convergent sequence $\lambda_n\to\lambda\in\Lambda$ ($n\in\mathbb{N}$) the sequence of operators $\{B_n(\lambda_n)\}_{n\in\mathbb{N}}$ is regular and converges to the operator $A(\lambda)$.

2. Convergence of the Eigenvalues and the Rate of Convergence

2.1 (Convergence of Eigenvalues).

Lemma 2.1 [8,2.13.10]. Let a1)-b3) hold. Then for every compact $\Lambda_0\subset\rho(A)$ there exist an index $n(\Lambda_0)$ and a constant $c(\Lambda_0)$, so that

$$n\ge n(\Lambda_0)\Rightarrow \Lambda_0\subset\rho(B_n)\ ,\ \|B_n(\lambda)^{-1}\|\le c(\Lambda_0) \qquad \forall\lambda\in\Lambda_0\ ,$$

i.e. from the regular convergence on Λ follows the stabile convergence on every compact $\Lambda_0\subset\rho(A)$.

<u>Proof.</u> Let there exist for some $N'\subseteq\mathbb{N}$, contradictory to the assertion, sequences $(x_n)_{n\in N'}$ and $(\lambda_n)_{n\in N'}$ such that $\|x_n\|=1$, $\lambda_n\in\Lambda_0$, $B_n(\lambda_n)x_n\to 0$

$(n\in N')$. Using compactness of Λ_0, we have $\lambda_n \to \lambda_0 \in \Lambda_0$ $(n\in N''\subset N')$ and b3) yields compactness of $\{x_n\}_{n\in N''}$. Now if $x_n \to u$ $(n\in N'''\subset N'')$, then $\|x_n\|\to\|u\|$ $(n\in N''') \Rightarrow \|u\|=1$. However at the same time

$$B_n(\lambda_n)x_n \to A(\lambda_0)u \ (n\in N''') \Rightarrow A(\lambda_0)u=0 \Rightarrow u=0$$

contradictory with $\|u\|=1$.■

Theorem 2.1 [8,2,13,10]. Let a1)-b3) hold. Then

1) $\lambda_0\in\sigma(A) \Rightarrow \exists n(\lambda_0), \{\lambda_n\}_{n\in\mathbb{N}} : \lambda_n\in\sigma(B_n), \lambda_n\to\lambda_0 \ (n\in\mathbb{N},\ n\geq n(\lambda_0))$,

2) $\lambda_n\in\sigma(B_n)$, $\lambda_n\to\lambda_0\in\Lambda$ $(n\in N') \Rightarrow \lambda_0\in\sigma(A)$.

Proof. 1) For the boundary Γ of every quite small neighborhood G of λ_0 we have $\Gamma\subset\rho(A)$ and from the lemma 2.1 it follows that $\Gamma\subset\rho(B_n)$, $\|B_n(\lambda)^{-1}\|\leq c(\Gamma)$ $(\lambda\in\Gamma,\ n\geq n(\Gamma))$. Let for $N'\subset\mathbb{N}$ be $G\cap\sigma(B_n)=\emptyset$, i.e. $G\subset\rho(B_n)$. Then, from the principle of maximum of modulus [5], it follows that $\|B_n(\lambda)^{-1}\|\leq c(\Gamma)$ for $\lambda\in G$, $n\in N'$, $n\geq n(\Gamma)$. If now $A(\lambda_0)u^0=0$, $\|u^0\|=1$, then $B_n(\lambda_0)p_nu^0\to 0$, $\|p_nu^0\|\to 1$ $(n\in\mathbb{N})$ and at the same time

$$\|p_nu^0\|=\|B_n(\lambda_0)^{-1}B_n(\lambda_0)p_nu^0\|\to 0 \ (n\in N',\ n\geq n(\Gamma)).$$

2) From $B_n(\lambda_n)x_n=0$, $\|x_n\|=1$, $\lambda_n\to\lambda_0$ $(n\in N')$ and b3') we get that for some $N''\subset N'$ and $u\in U$ there holds $x_n\to u$ $(n\in N'')$. Then $0=B_n(\lambda_n)x_n\to A(\lambda_0)u$, $1=\|x_n\|\to\|u\|$ $(n\in N'')$, and we have $\lambda_0\in\sigma(A)$.■

2.2 (Stability of the Full Algebraic Multiplicity of the Eigenvalues). Let a1)-a3) hold. Denote $A_j=\frac{1}{j!}A^{(j)}(\lambda_0)$, $j=1,2,\ldots$ The vector $(u^0,u^1,\ldots,u^k)$ with $u^0\neq 0$ is a chain of generalized eigenelements or a Jordan chain of the length $k+1$ for $A(\cdot)$ at λ_0, if

$$A_0u^0=0,\ A_0u^1+A_1u^0=0,\ \ldots,\ A_0u^k+A_1u^{k-1}+\ldots+A_ku^0=0.$$

Generalized eigenspace $\mathcal{J}(A,\lambda_0)$ of $A(\cdot)$ at λ_0 is the closed linear hull of all generalized eigenelements of $A(\cdot)$ at λ_0. The order $\nu(u^0)$ of the eigenelement u^0 is the maximal length of the Jordan chains beginning with u^0. From a1)-a3) it follows that $\dim\mathcal{J}(A,\lambda_0)<\infty$ and

$$\max\{\nu(u^0)\mid u^0\in\mathcal{N}(A(\lambda_0)),\ u^0\neq 0\}=\varkappa(A,\lambda_0)<\infty,$$

where $\varkappa(A,\lambda_0)$ is the order of the λ_0 as a pole of $A^{-1}(\cdot)$.

Let $\dim\mathcal{N}(A(\lambda_0))=m$. The system of eigenelements $(u_1^0,\ldots,u_m^0)$ is a canonical system of eigenelements for $A(\cdot)$ at $\lambda_0\in\sigma(A)$ if $\nu(u_1^0)=\varkappa(A,\lambda_0)$ and u_j^0 is an eigenelement of maximal possible order in some direct complement in $\mathcal{N}(A(\lambda_0))$ of the linear hull of the eigenelements $u_1^0,\ldots,u_{j-1}^0$. The vector $\underline{\nu}(\lambda_0,A)=(\nu_1,\ldots,\nu_m)$ with $\nu_i=\nu(u_i^0)$, $i=1,\ldots,m$ is determined uniquely and we call it the vector of multiplicities of the eigenvalue λ_0. The number $\nu(\lambda_0,A)=\nu_1+\nu_2+\ldots+\nu_m$ is called the algebraic multiplicity of the eigenvalue λ_0 of the operator-function $A(\cdot)$.

Theorem 2.2 [8,15,10]. Let a1)-b3) hold and Λ_0 be a compact in Λ with the boundary $\Gamma\subset\rho(A)$. Let V_0 be a subspace of V such that

$$V_0\supseteq A(\lambda)\mathcal{J}(A,\lambda_0) \qquad \forall\lambda\in\Lambda,\ \lambda_0\in\sigma(A)\cap\Lambda_0$$

and let there exist operators $q'_n: V\to Y_n$, $n\in\mathbb{N}$ which are linear and continuous on V_0 and so that $\|q_n v-q'_n v\|\to 0$ $(n\in\mathbb{N})$ $\forall v\in V$. Then there exists an index $n(\Lambda_0)$ such that for all $n\geq n(\Lambda_0)$ we have $\Gamma\subset\rho(B_n)$ and

$$\sum_{\lambda\in\sigma(A)\cap\Lambda_0}\nu(\lambda,A)=\sum_{\lambda\in\sigma(B_n)\cap\Lambda_0}\nu(\lambda,B_n) \qquad (\text{here } \sum_{\lambda\in\emptyset}\ldots=0).$$

Proof. From Lemma 2.1 it follows that it is sufficient to prove Theorem 2.2 for the case $\sigma(A)\cap\Lambda_0=\{\lambda_0\}$ with Λ_0 in the small neighbourhood V_1 of λ_0. Moreover, we can assume that the convergence of sequences $\{y_n\}_{n\in\mathbb{N}}$ is given using the operators q'_n, $n\in\mathbb{N}$. In these assumptions Theorem 2.2 will be proven in paragraph 4.

2.3 (Asymptotic Estimation of the Convergence Rate of the Approximate Eigenvalues).

Theorem 2.3 ([10], see also [8,2]). Let a1)-b3) hold, $\lambda_0\in\sigma(A)$, and let $\Lambda_0\subset\Lambda$ be a compact so that $\Lambda_0\cap\sigma(A)=\{\lambda_0\}$. Let the $p'_n\in U\to X_n$, $q'_n\in V\to Y_n$, $n\in\mathbb{N}$ be the arbitrary operators such that

1) $\|p_n u-p'_n u\|\to 0$, $\|q_n v-q'_n v\|\to 0$ $(n\in\mathbb{N})$ $\quad\forall u\in U$, $v\in V$,

2) p'_n, $n\in\mathbb{N}$ are linear on the finite-dimensional subspace $\mathcal{J}(A,\lambda_0)\subset U$,

3) q'_n, $n\in\mathbb{N}$ are linear and continuous on some subspace $V_0\subseteq V$ with $\{v\in V\mid v\in A(\lambda)\mathcal{J}(A,\lambda_0),\ \lambda\in\Lambda\}\subset V_0$.

Then there exists an index $n(\Lambda_0)$ such that for every $n\geq n(\Lambda_0)$ we have

$$\Lambda_0\cap\sigma(B_n)\neq\emptyset, \qquad \sum_{\lambda\in\sigma(B_n)\cap\Lambda_0}\nu(\lambda,B_n)=\nu(\lambda_0,A)$$

and the following estimations hold :

$$1)\ |\lambda_n-\lambda_0|\leq c\varepsilon_n^{1/\varkappa(A,\lambda_0)} \qquad \forall\lambda_n\in\sigma(B_n)\cap\Lambda_0, \tag{2.1}$$

$$2)\ |\overline{\lambda}_n-\lambda_0|\leq c\varepsilon_n \ \text{ with } \overline{\lambda}_n=\sum_{\lambda\in\sigma(B_n)\cap\Lambda_0}(\nu(\lambda,B_n)/\nu(\lambda_0,A))\cdot\lambda, \tag{2.2}$$

where ε_n are defined in (3.17) together with (3.2) and, for estimation (2.2) in addition together with demand (4.12) (see Lemma 4.5, e.g.). The quantity ε_n can be chosen so that

$$\varepsilon_n\leq\overline{\varepsilon}_n=\sup\{\|(q'_nA(\lambda)-B_n(\lambda)p'_n)u\| \mid \lambda\in\Lambda_0,\ u\in\mathcal{J}(A,\lambda_0),\ \|u\|=1\}. \tag{2.3}$$

At that

$$\overline{\varepsilon}_n\to 0\ (n\in\mathbb{N}). \tag{2.4}$$

Proof. By Lemma 2.1 it is again sufficient to look Λ_0 in the small neighborhood of λ_0. Moreover, we can assume that the convergence of sequences $\{x_n\}_{n\in\mathbb{N}}$, $\{y_n\}_{n\in\mathbb{N}}$ is given by the operators p'_n and q'_n, $n\in\mathbb{N}$, respectively. In this notation the estimations (2.1) and (2.2) will be proven in paragraph 4. The estimation (2.3) and the convergence (2.4) are established in Lemma 3.6. ■

Note that from the linearity and continuity of p'_n on $\mathcal{J}(A,\lambda_0)$ and q'_n on V_0, and the convergence $\|p'_n u\| \to \|u\|$, $\|q'_n v\| \to \|v\|$ $(n\in\mathbb{N})$ $\forall u\in U, v\in V$, by the principle of uniform boundedness [7] we get uniform boundedness of operators p'_n and q'_n on $\mathcal{J}(A,\lambda_0)$ and V_0, respectively.

3. Reducing the Problem to the Case of Matrix-Functions

3.1 (Canonical system of Jordan polynomials). A holomorphic in some neighborhood of λ_0 function $u(\cdot): \Lambda \to U$ we call a Jordan function of order $r+1$ for $A(\cdot)$ at λ_0 if

1) $u(\lambda_0) \neq 0$,

2) $\frac{1}{j!}\frac{d^j}{d\lambda^j}[A(\lambda)u(\lambda)]_{\lambda=\lambda_0} = A_0 u^{(j)}(\lambda_0)/j! + A_1 u^{(j-1)}(\lambda_0)/(j-1)! + \ldots + A_j u(\lambda_0) = 0,$
$j = 0,1,\ldots,r$,

3) $\frac{d^{r+1}}{d\lambda^{r+1}}[A(\lambda)u(\lambda)]_{\lambda=\lambda_0} \neq 0.$

Denote the order of the Jordan function $u(\cdot)$ by $\nu(u(\cdot))$. Note that if $u(\cdot)$ is a Jordan function of order $r+1$, then the vector $(u(\lambda_0), u'(\lambda_0)/1!, \ldots, u^{(k)}(\lambda_0)/k!)$ with $k \le r$ is a Jordan chain of length $k+1$. And, if the vector $(u^0, u^1, \ldots, u^k)$ is a Jordan chain, then the polynomial $u^0 + (\lambda-\lambda_0)u^1 + \ldots + (\lambda-\lambda_0)^r u^r$ with $r \le k$ is a Jordan function of order $r+1$. So the order $\nu(u^0)$ of the eigenelement $u^0 \in \mathcal{N}(A(\lambda_0))$ is equal to the maximal order of Jordan functions $u(\cdot)$ for $A(\cdot)$ at λ_0 with $u(\lambda_0) = u^0$.

Let $\lambda_0 \in \sigma(A)$, $\dim \mathcal{N}(A(\lambda_0)) = m$ and $(u_1^0, \ldots, u_m^0)$ be a canonical system of eigenelements for $A(\cdot)$ at λ_0. A canonical system of Jordan polynomials for $A(\cdot)$ at λ_0 is a system $(u_1(\cdot), u_2(\cdot), \ldots, u_m(\cdot))$ of polynomials for which $u_i(\lambda_0) = u_i^0$, $\nu(u_i(\cdot)) = \nu(u_i^0)$ and the degree of $u_i(\cdot)$ is equal to $\nu(u_i^0) - 1$, $i = 1, \ldots, m$. Note that for such polynomials $u_i(\lambda) \in \mathcal{J}(A,\lambda_0)$ $\forall \lambda \in \Lambda$.

3.2 (Some auxiliary notions). We shall reduce the examination of the eigenvalue problem for operator-functions to that of matrix-functions. We use for that the standard construction for Fredholm operators. The difference is that for our purpose we need to use instead of fixed elements of $U^* = \mathcal{B}(U,\mathbb{C})$ and V the suitable functions of λ with values in X_n^* and V.

Let

1) $(u_1^0, \ldots, u_m^0)$ be a canonical system of eigenelements for $A(\cdot)$ at λ_0 and $(u_1(\cdot), u_2(\cdot), \ldots, u_m(\cdot))$ a canonical system of Jordan polynomials such that $u_i(\lambda_0) = u_i^0$,

2) $(g_n^1, \ldots, g_n^m)$, $n\in\mathbb{N}$ be systems of functionals $g_n^i \in X_n^* = \mathcal{B}(X_n, \mathbb{C})$ such that

$$\sup_{i,n} \|g_n^i\| \le c, \quad \langle p_n u_k^0, g_n^i \rangle = g_n^i(p_n u_k^0) \to \delta_{ik} \qquad (n\in\mathbb{N}), \quad i,k = 1,\ldots,m, \ (3.1)$$

and $(g_n^1(\cdot),\ldots,g_n^m(\cdot))$, $n\in\mathbb{N}$ be a system of holomorphic on Λ functions $g_n^i(\cdot):\Lambda\to X_n^*$ with $g_n^i(\lambda_0)=g_n^i$, $n\in\mathbb{N}$, $i=1,\ldots,m$ such that on every compact $\Lambda_0\subset\Lambda$ there holds

$$sup\{\|g_n^i(\lambda)\| \mid n\in\mathbb{N},\ \lambda\in\Lambda_0,\ i=1,\ldots,m\}\le c(\Lambda_0)<\infty.$$

3) $(v_1,\ldots,v_m)$ be a basis in some direct complement of the subspace $A(\lambda_0)U$ in V and $(v_1(\cdot),\ldots,v_m(\cdot))$ be a system of holomorphic on Λ functions with values in V such that $v_i(\lambda_0)=v_i$, $i=1,\ldots,m$,

4) the operators p_n, $n\in\mathbb{N}$ be linear on finite-dimensional subspace $\mathcal{H}(A,\lambda_0)$ and operators q_n, $n\in\mathbb{N}$ linear and continuous on some subspace $V_1\subseteq V$ such that $\{v=v_i(\lambda)\mid\lambda\in\Lambda\}\subseteq V_1$,

5) $P\in\mathcal{B}(U,U)$ be a projector onto $\mathcal{H}(A,\lambda_0)\subset U$.

To prove Theorems 2.2 and 2.3 we shall take

$$v_i(\lambda)=(\lambda-\lambda_0)^{-\nu(u_i(\,))}A(\lambda)u_i(\lambda)\ ,\quad i=1,\ldots,m. \tag{3.2}$$

Justification for this is the following lemma :

Lemma 3.2 [8,2,10]. Let $v_i(\cdot)$ be defined by (3.2). Then the elements $v_i=v_i(\lambda_0)$, $i=1,\ldots,m$ are linearly independent . If V^m is the linear hull of the elements $v_1,\ldots,v_m$, then $V^m\cap A(\lambda_0)U=\{0\}$ and V is the direct sum of $A(\lambda_0)U$ and V^m.

Proof. Recall that $codim\,A(\lambda_0)U=m$. Let there exist, contradictory to the assertions of Lemma 3.2, a linear combination $a_1v_1+\ldots+a_mv_m=v_0\in A(\lambda_0)U$ with $|a_1|+\ldots+|a_m|\ne0$. Let u_0 be the origin of v_0, i.e. $A(\lambda_0)u_0=v_0$. Denote $\nu_i=\nu(u_i(\cdot))$, $i=1,\ldots,m$ and $\nu=max\{\nu_i\mid i=1,\ldots,m,\ a_i\ne0\}$. Look at the polynomial

$$u(\lambda)=\sum_{i=1}^{i=m}\ a_i(\lambda-\lambda_0)^{\nu-\nu_i}u_i(\lambda)-(\lambda-\lambda_0)^{\nu}u_0.$$

For $u(\lambda_0)$ we have

$$u(\lambda_0)=\sum_{i:\nu_i=\nu}a_iu_i(\lambda_0)\ne0,\quad A(\lambda_0)u(\lambda_0)=0,\quad \nu(u(\lambda_0))=\nu.$$

(Recall that the system $(u_1^0,\ldots,u_m^0)$ is a canonical basis in $\mathcal{N}(A(\lambda_0))$.)

But at the same time direct calculations give (make use of the equalities $\frac{d^k}{d\lambda^k}[A(\lambda)u_i(\lambda)]_{\lambda=\lambda_0}=0,\ k<\nu_i,\ \ \frac{d^{\nu_i}}{d\lambda^{\nu_i}}[A(\lambda)u_i(\lambda)]_{\lambda=\lambda_0}=\nu_i!v_i$) :

$$\frac{d^j}{d\lambda^j}[A(\lambda)u(\lambda)]_{\lambda=\lambda_0}=0,\ j=1,\ldots,\nu-1,$$

$$\frac{d^\nu}{d\lambda^\nu}[A(\lambda)u(\lambda)]_{\lambda=\lambda_0}=\nu!(a_1v_1+\ldots+a_mv_m-v_0)=0\ ,$$

i.e. $\nu(u(\cdot))>\nu$ and, consequently, $\nu(u(\lambda_0))>\nu$. ∎

3.3 (Auxiliary operator-functions). Let us define operator-functions $K_n(\cdot)$, $R_n(\cdot):\Lambda\to\mathcal{B}(U,V)$ and $L_n(\cdot)$, $S_n(\cdot):\Lambda\to\mathcal{B}(X_n,Y_n)$, $n\in\mathbb{N}$ as follows :

$$K_n(\lambda)u=\sum_{i=1}^{i=m}\langle p_nPu,\ g_n^i(\lambda)\rangle v_i(\lambda),\quad R_n(\lambda)=A(\lambda)+K_n(\lambda)\ , \tag{3.3}$$

$$L_n(\lambda)x_n=\sum_{i=1}^{i=m}\langle x_n,\ g_n^i(\lambda)\rangle q_nv_i(\lambda),\quad S_n(\lambda)=B_n(\lambda)+L_n(\lambda). \tag{3.4}$$

Note that all these operator-functions are holomorphic on Λ and uniformly bounded on every compact $\Lambda_0 \subset \Lambda$ (and, hence, equicontinuous on every Λ_0). Obviously for every $\lambda \in \Lambda$ the operators $K_n(\lambda)$ and $L_n(\lambda)$ are finite-dimensional and, thus, the operators $R_n(\lambda)$ and $S_n(\lambda)$ are Fredholm operators with index 0. Moreover, the following lemma holds :

Lemma 3.3. In some neighborhood $\Lambda_1 \subset \Lambda$ of λ_0 for all quite large indices, say for $n \geq n_0$, all the operators $R_n(\lambda), S_n(\lambda)$, $n \in N$ are continuously invertible and the inverses are uniformly bounded:

$$sup\{\|R_n(\lambda)^{-1}\|, \|S_n(\lambda)^{-1}\| \,|\, \lambda \in \Lambda_1,\ n \geq n_0\}.$$

Proof. a) Contradictory to assertions of lemma 3.3 let there exist sequences $\{\lambda_n\}_{n \in N'}$ and $\{u^n\}_{n \in N'}$ such that $\lambda \to \lambda_0$, $\|u^n\|=1$, $R_n(\lambda_n)u^n \to 0$ $(n \in N')$. Then we also have $R_n(\lambda_0)u^n \to 0$ $(n \in N')$ and $A(\lambda_0)u^n \to 0$, $K_n(\lambda_0)u^n \to 0$ (the last because $V = A(\lambda_0)U \dotplus K_n(\lambda_0)U$).

Recall now that $A(\lambda_0)$ is a Fredholm operator. Therefore, from $A(\lambda_0)u^n \to 0, \|u^n\|=1$ $(n \in N')$ it follows that $\{u^n\}_{n \in N'}$ is compact in U. Let $u^n \to u^0$ $(n \in N'' \subset N')$; then $A(\lambda_0)u^0=0$, $\|u^0\|=1$ and $K_n(\lambda_0)u^0 \to 0$ $(n \in N'')$.

Represent u^0 as a linear combination of $u^0_1, \dots, u^0_m$: $u=\sum_{i=1}^{i=m}\alpha_i u^0_i$. Then $K_n(\lambda_0)u^0=\sum_{i=1}^{i=m}\alpha_i K_n(\lambda_0)u^0_i \to \sum_{i=1}^{i=m}\alpha_i v_i(\lambda_0)$. Therefore, we have $\sum_{i=1}^{i=m}\alpha_i v_i=0$, from which it follows that $\alpha_i=0$, $i=1,\dots,m$ and $u^0=0$, contradictory to $\|u^0\|=1$.

b) The proof for $S_n(\lambda)$ is similiar. Contradictory to the assertions of Lemma 3.3 let there exist sequences $\{\lambda_n\}_{n \in N'}$ and $\{x_n\}_{n \in N'}$, such that $\lambda_n \to \lambda_0$, $\|x_n\|=1$, $S_n(\lambda_n)x_n \to 0$ $(n \in N')$. Then we also have $S_n(\lambda_0)x_n \to 0$ $(n \in N')$. Note that $\{L_n(\lambda_0)x_n\}_{n \in N'}$ is a compact sequence because $q_n v_i(\lambda_0) \to v_i$ and $|\langle x_n, g^i_n(\lambda_0)\rangle| \leq c$ $(n \in N')$. Therefore the $\{B_n(\lambda_0)x_n\}_{n \in N'}$ is compact, too, and by b3.2) we get the compactness of the $\{x_n\}_{n \in N'}$. Let $x_n \to u^0$ $(n \in N'' \subset N')$. Then $\|u^0\|=1$, $B_n(\lambda_0)p_n u^0 \to A(\lambda_0)u^0$, $S_n(\lambda_0)p_n u^0 \to 0$ and $L_n(\lambda_0)p_n u^0 \to -A(\lambda_0)u^0$ $(n \in N'')$.

However, $L_n(\lambda_0)p_n u^0$ can converge only to the element of linear hull of elements $v_1(\lambda_0), \dots, v_m(\lambda_0)$. Therefore, $A(\lambda_0)u^0$ must be 0. Now we have $u^0=\sum_{i=1}^{i=m}\alpha_i u^0_i$, $L_n(\lambda_0)p_n u^0 \to \sum_{i=1}^{i=m}\alpha_i v_i(\lambda_0)=0$. From that it follows that $u^0=0$ is contradictory to $\|u^0\|=1$. ■

3.4 (Introduction of matrix-functions). Let us look at the equations

$$A(\lambda)u=0, \qquad B_n(\lambda)x_n=0 \tag{3.5}$$

at a neighborhood $\Lambda_1 \subset \Lambda$ of $\lambda_0 \in \sigma(A)$ and for $n \geq n_0$ such that the operators $R_n(\lambda), S_n(\lambda)$ are continuously invertible and the inverses are uniformly bounded on Λ_1 (see Lemma 3.3). Note that the operator functions $R_n^{-1}(\cdot)$ and $S_n^{-1}(\cdot)$ are holomorphic on Λ_1.

The equations (3.5) can be written, correspondingly, as follows :

$$R_n(\lambda)(I-R_n(\lambda)^{-1}K_n(\lambda))u=0,\quad S_n(\lambda)(I-S_n(\lambda)^{-1}L_n(\lambda))x_n=0,$$

i.e.

$$A(\lambda)u\equiv R_n(\lambda)[I-\sum_{i=1}^{i=m}\langle p_nP\cdot,\ g_n^i(\lambda)\rangle R_n^{-1}(\lambda)v_i(\lambda)]u=0, \tag{3.6}$$

$$B_n(\lambda)x_n\equiv S_n(\lambda)[I-\sum_{i=1}^{i=m}\langle\ \cdot\ ,\ g_n^i(\lambda)\rangle S_n^{-1}(\lambda)q_nv_i(\lambda)]x_n=0. \tag{3.7}$$

Define on Λ_1 the operator-functions $C_n'(\cdot):\Lambda_1\to\mathcal{B}(U,\mathbb{C}^m)$, $D_n'(\cdot):\Lambda_1\to\mathcal{B}(\mathbb{C}^m,U)$ and $C_n''(\cdot):\Lambda_1\to\mathcal{B}(X_n,\mathbb{C}^m)$, $D_n''(\cdot):\Lambda_1\to\mathcal{B}(\mathbb{C}^m,X_n)$ by the following relations:

$$C_n'(\lambda)u=(\ \langle p_nPu,\ g_n^1(\lambda)\rangle,\dots,\langle p_nPu,g_n^m(\lambda)\rangle\ ),$$

$$D_n'(\lambda)(\alpha_1,\dots,\alpha_m)=\sum_{i=1}^{i=m}\alpha_iR_n^{-1}(\lambda)v_i(\lambda),$$

$$C_n''(\lambda)x_n=(\ \langle x_n,\ g_n^1(\lambda)\rangle,\dots,\langle x_n,\ g_n^m(\lambda)\rangle\ ),$$

$$D_n''(\lambda)(\alpha_1,\dots,\alpha_m)=\sum_{i=1}^{i=m}\alpha_iS_n^{-1}(\lambda)q_nv_i(\lambda).$$

All these operator-functions are holomorphic on Λ_1, uniformly bounded , and equicontinuous on every compact $\Lambda_0\subset\Lambda_1$.

In this notation equations (3.6) and (3.7) can be written as follows :

$$A(\lambda)u\equiv R_n(\lambda)[I-D_n'(\lambda)C_n'(\lambda)]u=0, \tag{3.8}$$

$$B_n(\lambda)x_n\equiv S_n(\lambda)[I-D_n''(\lambda)C_n''(\lambda)]x_n=0. \tag{3.9}$$

Define now the operator-functions $M_n(\cdot),T_n(\cdot):\Lambda_1\to\mathcal{B}(\mathbb{C}^m,\mathbb{C}^m)$ as follows:

$$M_n(\cdot)=I-C_n'(\cdot)D_n'(\cdot),\quad T_n(\cdot)=I-C_n''(\cdot)D_n''(\cdot)\ . \tag{3.10}$$

Note that

$$(M_n(\lambda)(\alpha_1,\dots,\alpha_m))_r=\alpha_r-\langle p_nP\sum_{i=1}^{i=m}\alpha_iR_n^{-1}(\lambda)v_i(\lambda),\ g_n^r(\lambda)\rangle=$$

$$=\alpha_r-\sum_{i=1}^{i=m}\alpha_i\langle p_nPR_n^{-1}(\lambda)v_i(\lambda),\ g_n^r(\lambda)\rangle,$$

$$(T_n(\lambda)(\alpha_1,\dots,\alpha_m))_r=\alpha_r-\sum_{i=1}^{i=m}\alpha_i\langle S_n^{-1}(\lambda)q_nv_i(\lambda),\ g_n^r(\lambda)\rangle,$$

and $M_n(\lambda),T_n(\lambda)$ can be represented as matrices with elements

$$m_n^{rs}(\lambda)=\delta_{rs}-\ \langle p_nPR_n^{-1}(\lambda)v_s(\lambda),g_n^r(\lambda)\rangle, \tag{3.11}$$

$$t_n^{rs}(\lambda)=\delta_{rs}-\ \langle S_n^{-1}(\lambda)q_nv_s(\lambda),\ g_n^r(\lambda)\rangle. \tag{3.12}$$

These functions are holomorphic on Λ_1 , uniformly bounded, and equicontinuous on every compact $\Lambda_0\subset\Lambda_1$. (It is not important for us which norm is fixed on $\mathbb{C}^m$.)

3.5 (Reduction of the problem to the case of matrix-functions). We shall use the following lemma :

Lemma 3.5 [11]. Let U,V,W be complex Banach spaces and $A(\cdot)$, $R(\cdot)$, $C(\cdot)$, $D(\cdot)$, $M(\cdot)$ holomorphic on some region $\Lambda_1\subseteq\mathbb{C}$ operator-functions with values in $\mathcal{B}(U,V)$, $\mathcal{B}(U,V)$,$\mathcal{B}(U,W)$,$\mathcal{B}(W,U)$,$\mathcal{B}(W,W)$, correspondlingy. Let $\Lambda_1\subset\rho(R)$ and the following relations hold :

$$A(\cdot)=R(\cdot)(I-D(\cdot)C(\cdot)),\quad M(\cdot)=I-C(\cdot)D(\cdot). \tag{3.13}$$

Then

1) $u_0\neq 0,\ A(\lambda_0)u_0=0 \ \Rightarrow\ C(\lambda_0)u_0\neq 0,\ M(\lambda_0)[C(\lambda_0)u_0]=0,\ u_0=D(\lambda_0)C(\lambda_0)u_0,$

2) $w_0\neq 0,\ M(\lambda_0)w_0=0 \ \Rightarrow\ D(\lambda_0)w_0\neq 0,\ A(\lambda_0)[D(\lambda_0)w_0]=0, w_0=C(\lambda_0)D(\lambda_0)w_0,$

3) $u(\cdot)$ is a Jordan function of order r for $A(\cdot)$ at $\lambda_0\in\Lambda_1$ $\Rightarrow$

$\Rightarrow$ $C(\cdot)u(\cdot)$ is a Jordan function of order $r'\geq r$ for $M(\cdot)$ at $\lambda_0\in\Lambda_1$,

4) $w(\cdot)$ is a Jordan function of order r' for $M(\cdot)$ at $\lambda_0\in\Lambda_1$ $\Rightarrow$

$\Rightarrow$ $D(\cdot)w(\cdot)$ is a Jordan function of order $r\geq r'$ for $A(\cdot)$ at $\lambda_0\in\Lambda_1$,

5) $u(\cdot)$ is a Jordan function of maximal order $r=\nu(u(\lambda_0))$ for $A(\cdot)$ at $\lambda_0\in\Lambda$

$\Rightarrow$ $C(\cdot)u(\cdot)$ is a Jordan function of maximal order r' for $M(\cdot)$ at $\lambda_0\in\Lambda_1$ and $r'=r$,

6) $w(\cdot)$ is a Jordan function of maximal order $r'=\nu(w(\lambda_0))$ for $M(\cdot)$ at $\lambda_0\in\Lambda_1$

$\Rightarrow$ $D(\cdot)w(\cdot)$ is a Jordan function of maximal order r for $A(\cdot)$ at $\lambda_0\in\Lambda$ and $r=r'$,

7) $\sigma(A)=\sigma(M)$ and at each point $\lambda_0\in\sigma(A)$ the vectors of multiplicities for $A(\cdot)$ and $M(\cdot)$ coincide: $\underline{\nu}(\lambda_0,A)=\underline{\nu}(\lambda_0,M)$.

Proof. 1) If $u_0\neq 0$ and $A(\lambda_0)u_0=0$, then from (3.13) we get

$$[I-D(\lambda_0)C(\lambda_0)]\,u_0=0,\quad u_0=D(\lambda_0)C(\lambda_0)u_0\neq 0$$

and thus $C(\lambda_0)u_0\neq 0$.

2) If $w_0\neq 0$ and $M(\lambda_0)w_0=0$, then from (3.13) we get $w_0=C(\lambda_0)D(\lambda_0)w_0$ and thus $D(\lambda_0)w_0\neq 0$.

3,4) Recall at first that $u(\cdot)$ is a Jordan function of order r for $A(\cdot)$ at λ_0 if $u(\lambda_0)\neq 0$ and $A(\cdot)u(\cdot)$ has at λ_0 the zero of multiplicity r. Now the assertions 3) and 4) follow from the identities (we shall write here u instead of $u(\lambda)$, A instead of $A(\lambda)$ etc.):

$$M(Cu)=[I-CD]Cu=C[I-DC]u=CR^{-1}Au=(CR^{-1})Au, \tag{3.14}$$

$$A(Dw)=R[I-DC]Dw=RD[I-CD]w=(RD)Mw \tag{3.15}$$

(recall that all these operator-functions are holomorphic).

5,6) These assertions are direct conclusions of the 3) and 4) together.

7) This assertion is a direct conclusion of 5) and 6). ■

Corollary 3.5. For the operator-functions $A(\cdot), B_n(\cdot), M_n(\cdot)$ and $T_n(\cdot)$ in paragraph 3.4 we have

$$\sigma(A)=\sigma(M_n)\ ,\quad \sigma(B_n)=\sigma(T_n)$$

and for every $\lambda_0\in\sigma(A)$, $\lambda_n\in\sigma(B_n)$ there hold :

$$\underline{\nu}(\lambda_0,A)=\underline{\nu}(\lambda_0,M_n),\ \underline{\nu}(\lambda_n,B_n)=\underline{\nu}(\lambda_n,T_n),$$

$$\nu(\lambda_0,A)=\nu(\lambda_0,M_n),\ \nu(\lambda_n,B_n)=\nu(\lambda_n,T_n),$$

$$\varkappa(\lambda_0,A)=\varkappa(\lambda_0,M_n),\ \varkappa(\lambda_n,B_n)=\varkappa(\lambda_n,T_n).$$

13*

3.6 (Closeness of matrices $M_n(\cdot)$ and $T_n(\cdot)$). For us it will be important to estimate $\|M_n(\lambda)-T_n(\lambda)\|$ and $|detM_n(\lambda)-detT_n(\lambda)|$ on a compact $\Lambda_0\in\Lambda_1$. Denote

$$e_n^{rs}(\lambda)=m_n^{rs}(\lambda)-t_n^{rs}(\lambda),\ \varepsilon_n=\varepsilon_n(\Lambda_0)=max\{\,|e_n^{rs}(\lambda)|\ \big|\ \lambda\in\Lambda_0,\ r,s=1,\dots,m\}.$$

Then it is clear that

$$\|M_n(\lambda)-T_n(\lambda)\|\le c\varepsilon_n\,,\quad |detM_n(\lambda)-detT_n(\lambda)|\le c\varepsilon_n\quad \forall\lambda\in\Lambda_0, n\ge n_0 \tag{3.16}$$

(recall that $m_n(\lambda)$ and $t_n(\lambda)$ are uniformly bounded on Λ_0).

For $e_n^{rs}(\lambda)$ we have

$$e_n^{rs}(\lambda)=m_n^{rs}(\lambda)\ -t_n^{rs}(\lambda)=\langle p_nPR_n^{-1}(\lambda)v_s(\lambda)-S_n^{-1}(\lambda)q_n(\lambda)v_s(\lambda),\ g_n^r(\lambda)\rangle=$$
$$=\langle S_n^{-1}(\lambda)[S_n(\lambda)p_nP-q_nR_n(\lambda)]R_n^{-1}(\lambda)v_s(\lambda),\ g_n^r(\lambda)\rangle=$$
$$=\langle S_n^{-1}(\lambda)[B_n(\lambda)p_nP-q_nA(\lambda)]R_n^{-1}(\lambda)v_s(\lambda),\ g_n^r(\lambda)\rangle$$

and thus

$$\varepsilon_n=\max_{\lambda\in\Lambda_0,r,s=1,\dots m}|\langle S_n^{-1}(\lambda)[B_n(\lambda)p_nP-q_nA(\lambda)]R_n^{-1}(\lambda)v_s(\lambda),\ g_n^r(\lambda)\rangle|. \tag{3.17}$$

Let us examine the function $R_n^{-1}(\cdot)v_s(\cdot)$ if $v_s(\cdot)$ is determined by (3.2). We prove the following lemma:

Lemma 3.6 [8,2,10]. Let $v_l(\cdot)$ be defined by (3.2). Then

$$R_n^{-1}(\lambda)v_l(\lambda)\in\partial(A,\lambda_0)\quad\forall\lambda\in\Lambda_0,\ n\ge n_0$$

and

$$\varepsilon_n\le\bar\varepsilon_n=max\,\{\|(q_n'A(\lambda)-B_n(\lambda)p_n')u\|\ \big|\ \lambda\in\Lambda_0,\ u\in\partial(A,\lambda_0),\ \|u\|=1\,\}. \tag{3.18}$$

At that $\bar\varepsilon_n\to0$ ($n\in\mathbb{N}$).

<u>Proof</u>. For $\lambda\ne\lambda_0$ we can write

$$R_n^{-1}(\lambda)=A^{-1}(\lambda)[A(\lambda)-K_n(\lambda)+K_n(\lambda)]R_n^{-1}(\lambda)=A^{-1}(\lambda)[I+K_n(\lambda)R_n^{-1}(\lambda)]. \tag{3.19}$$

Note further that for $\lambda\ne\lambda_0$ we also have

$$A^{-1}(\lambda)v_s(\lambda)=(\lambda-\lambda_0)^{-\nu_s}u_s(\lambda)\in\partial(A,\lambda_0),$$

$$A^{-1}(\lambda)K_n(\lambda)u=\sum_{l=1}^{l=m}(\lambda-\lambda_0)^{-\nu_l}\langle p_nPu,\ g_n^l(\lambda)\rangle u_l(\lambda)\in\partial(A,\lambda_0)\qquad\forall u\in U.$$

Hence, for $\lambda\ne\lambda_0$ we have $R_n^{-1}(\lambda)v_s(\lambda)\in\partial(A,\lambda_0)$ and from the continuity of the $R_n^{-1}(\cdot)v_n(\cdot)$ we conclude that $R_n^{-1}(\lambda)v_s(\lambda)\in\partial(A,\lambda_0)$ $\forall\lambda\in\Lambda_0$.

Now from the uniform boundedness of $g_n^r(\cdot)$ and $S_n^{-1}(\cdot)$, $n\ge n_0$ on Λ_0 we get (3.18).

The convergence $\bar\varepsilon_n\to0$ ($n\in\mathbb{N}$) we prove by reductio ad absurdum proof. Let there exist the subsequences $(\lambda_n)_{n\in N'}\subset\Lambda_0$ and $(u_n)_{n\in N'}\subset\partial(A,\lambda_0)$ so that, contradictory to assertion, we have

$$\|\,[q_n'A(\lambda_n)-B_n(\lambda_n)p_n']u_n\|\ge\alpha>0,\quad \|u_n\|=1\quad(n\in N').$$

Recall that $A(\cdot)$ and $B_n(\cdot)$, $n\in\mathbb{N}$ are uniformly bounded and equicontinuous on Λ_0. So from the compactness of Λ_0 and $\{u\in\partial(A,\lambda_0)\,|\,\|u\|=1\}$ it follows that for some $N''\subseteq N'$, $\lambda^0\in\Lambda_0$ and $u^0\in\mathcal{J}(A,\lambda_0)$ we have

$$\| q_n' A(\lambda^0)u^0 - B_n(\lambda^0)p_n u^0\| \ge \alpha > 0 \ (n\in N'').$$

But this is a contradiction with assumption b3.1) by which

$$B_n(\lambda^0)p_n u^0 \to A(\lambda^0)u^0 \ (n\in \mathbf{N}).\blacksquare$$

4. The Proof of the Theorems 2.2 and 2.3

4.1 (**Common assumptions and discussions**). For both theorems 2.2 and 2.3 we shall:

1) define functions $v_i(\cdot)$ in paragraph 3.2 by relations (3.2) and take $V_1 = V_0$,

2) define the convergence of sequences $(x_n)_{n\in N}$ and $(y_n)_{n\in N}$ using the operators p_n' and q_n', $n\in\mathbf{N}$ which are linear and continuous on the subspaces $\mathcal{J}(A,\lambda_0)$ and V_0, correspondingly,

3) look at compact Λ_0 in a small neighborhood $\Lambda_1\subset\Lambda$ of $\lambda_0\in\sigma(A)$ and indices n so great that (a) the assertions of lemma 3.3 hold and (b) the elements $v_1(\lambda^0),\dots,v_m(\lambda^0)$ are linearly independent for all λ^0 in Λ_0 (this can be assumed because the elements $v_1(\lambda_0),\dots,v_m(\lambda_0)$ are linearly independent and the functions $v_1(\cdot),\dots,v_m(\cdot)$ are holomorphic on Λ_1).

We shall use the following theorem:

Theorem 4.1 (see [6,12]). Let $M(\cdot)$ be holomorphic in some neighborhood of $\lambda_0\in\mathbf{C}$ matrix-function and $\lambda_0\in\sigma(M)$. Then the algebraic multiplicity $\nu(\lambda_0,M)$ of λ_0 is equal to multiplicity of λ_0 as a zero of the holomorphic function $\det M(\cdot)$. ■

Denote $\nu(\lambda_0,A)=\sum_{\lambda\in\sigma(A)\cap\Lambda_0}\nu(\lambda,A)$. Then by Corollary 3.5 and Theorem 4.1 we have the following corollary:

Corollary 4.1. The following assertions hold:

1) $\nu(\Lambda_0,A)=\nu(\lambda_0,A)=\nu(\lambda_0,M_n)=$ (4.1)
=the multiplicity of λ_0 as a zero of the function $\det M(\cdot)$

2) $\nu(\Lambda_0,B_n)=\nu(\Lambda_0,T_n)=$the sum of the multiplicities of all zeroes of the function $\det T_n(\cdot)$ in Λ_0. ■ (4.2)

For the investigation of the zeros of the functions $\det T_n(\cdot)$ the following lemma will be used:

Lemma 4.1 [8,10]. Let Γ be the boundary of the compact Λ_0. Then there exists an index $n_1=n_1(\Gamma)$ so that for all $n\ge n_1$:

1) on Γ there exist inverses $M_n(\lambda)^{-1}$ and they are uniformly bounded for $n\ge n_1$, $\lambda\in\Gamma$,

2) $\sup\{|\det M_n(\lambda)^{-1}| \mid \lambda\in\Gamma, n\ge n_1\}=\sup\{|\det M_n(\lambda)|^{-1} \mid \lambda\in\Gamma, n\ge n_1\}\le c<\infty$, (4.3)

3) $\inf\{\,|\det M_n(\lambda)| \mid \lambda\in\Gamma,\ n\ge n_1\,\}\ge\alpha>0.$ (4.4)

<u>Proof.</u> We shall prove the existence and uniform boundedness of $M_n(\lambda)^{-1}$, $\lambda\in\Gamma, n\ge n_1$. From this it follows that all the elements of matrices

$M_n(\lambda)^{-1}$ are uniformly bounded and, therefore, (4.3) must hold, too. From (4.3) we immediately get (4.4).

Let there exist, contradictory to assertion 1) of Lemma 4.1, the subsequences $\{\lambda_n\}_{n\in N'}$ and $\{w_n\}_{n\in N'}$ with $\lambda_n\in\Gamma$, $w_n\in\mathbb{C}^m$, $\|w_n\|=1$ so that $M_n(\lambda_n)w_n\to 0$ $(n\in N')$. Using the compactness of Γ and $\{w_n\}_{n\in N'}$ we find $N''\subset N'$, $\lambda^0\in\Gamma$ and $w^0\in\mathbb{C}^m$ with $\|w^0\|=1$ so that $\lambda_n\to\lambda^0$, $w_n\to w^0$ $(n\in N'')$.

From the equicontinuity of $M_n(\cdot)$ on Γ we have now, also, $M_n(\lambda^0)w^0\to 0$ $(n\in N'')$. And using an identity similiar to (3.15) we get:

$$A(\lambda^0)D_n'(\lambda^0)w^0=R_n(\lambda^0)D_n'(\lambda^0)M_n(\lambda^0)w^0\to 0 \ (n\in N''). \tag{4.5}$$

Since $\lambda^0\in\Gamma\subset\rho(A)$, we have also $D_n'(\lambda^0)w^0\to 0$ $(n\in N'')$. And from this we get $R_n(\lambda^0)D_n'(\lambda^0)w^0\to 0$ $(n\in N'')$. Therefore, if $w^0=(\alpha_1,\dots,\alpha_m)$, then

$$\Sigma_{i=1}^{i=m}\alpha_i R_n(\lambda^0)R_n^{-1}(\lambda^0)v_i(\lambda^0)=\Sigma_{i=1}^{i=m}\alpha_i v_i(\lambda^0)\to 0.$$

From the linear independence of the elements $v_i(\lambda^0)$, $i=1,\dots,m$, it follows now that $\alpha_i=0$, $i=1,\dots,m$. Therefore $w^0=0$, contradictory to $\|w^0\|=1$.■

4.2 (Proof of Theorem 2.2). Theorem 2.2 is a straightforward consequence of Corollary 4.1 and Theorem of Rouché for the functions $det M_n(\cdot)$ and $det T_n(\cdot)$ on Λ_0.The fulfillment of the assumptions of Theorem of Rouché follows from estimation (3,16), Lemma 3.6, and Lemma 4.1.

4.3 (Proof of estimation (2.1) in Theorem 2.3). To prove the estimation (2.1), we shall use the following lemma:

<u>Lemma</u> 4.3 [8,10]. Let W be a complex Banach space, and $M_n(\cdot)$, $T_n(\cdot)$ $(n\in\mathbb{N})$ be holomorphic in some region $\Lambda_1\subseteq\mathbb{C}$ operator-functions with values in $\mathscr{B}(W,W)$. Let $\Lambda_0=\{\lambda\in\mathbb{C} \mid |\lambda-\lambda_0|\le\delta\}\subset\Lambda_1$, Γ be the boundary of Λ_0 and the following assumptions hold:

1) $\sigma(M_n)\cap\Lambda_1=\{\lambda_0\}$, $\varkappa(\lambda_0,M_n)=\varkappa \quad \forall n\in\mathbb{N}$ (4.6)

with $\varkappa(\lambda_0,M_n)$ the order of λ_0 as a pole of $M_n(\ )$,

2) $\sup_{n\in\mathbb{N},\lambda\in\Gamma}\|M_n(\lambda)\|\le c<\infty$, $\quad \sup_{n\in\mathbb{N},\lambda\in\Gamma}\|M_n(\lambda)^{-1}\|\le c<\infty$, (4.7)

3) $\eta_n=\eta_n(\Gamma)=\sup_{\lambda\in\Gamma}\|M_n(\lambda)-T_n(\lambda)\|\to 0$ $(n\in\mathbb{N})$. (4.8)

Then for almost all $n\in\mathbb{N}$ the following estimation holds:

$$\max_{\lambda\in\sigma(T_n)\cap\Lambda_0}|\lambda-\lambda_0|\le c(\eta_n)^{1/\varkappa}. \tag{4.9}$$

<u>Proof</u>. From (4.6) it follows that the function $(\cdot-\lambda_0)^{\varkappa}M_n^{-1}(\cdot)$ is holomorphic on Λ_1. Thus, by the principle of maximum of modulus and (4.7) we have

$$\sup_{\lambda\in\Lambda_0}\|(\lambda-\lambda_0)^{\varkappa}M_n(\lambda)^{-1}\|=\sup_{\lambda\in\Gamma}\|(\lambda-\lambda_0)^{\varkappa}M_n(\lambda)^{-1}\|\le c_1<\infty. \tag{4.10}$$

Therefore, $\inf_{\lambda\in\Lambda_0}\|(\lambda-\lambda_0)^{-\varkappa}M_n(\lambda)\|\geq 1/c_1$ and, if for some $\lambda\neq\lambda_0$ in Λ_0 we have

$$\|(\lambda-\lambda_0)^{-\varkappa}M_n(\lambda)-(\lambda-\lambda_0)^{-\varkappa}T_n(\lambda)\| < 1/c_1\,, \tag{4.11}$$

then the operator $(\lambda-\lambda_0)^{-\varkappa}T_n(\lambda)$ will be invertible (as an operator which is sufficiently close to the invertible operator). Hence, the operator $T_n(\lambda)$ will be invertible for such $\lambda\neq\lambda_0$.

By (4.8) we have $\eta_n\to 0$ ($n\in\mathbb{N}$). Thus, for large n we shall have $\delta^{-\varkappa}\eta_n< 1/c_1$ and the operators $T_n(\lambda)$ will be invertible in the annulus

$$(c_1\eta_n)^{1/\varkappa} < |\lambda-\lambda_0| \leq \delta$$

(for these λ we have (4.11) fulfilled).

Therefore, $\lambda\in\sigma(T_n)\cap\Lambda_0 \Rightarrow |\lambda-\lambda_0|\leq c\eta_n^{1/\varkappa}$. ■

The estimation (2.1) in Theorem 2.3 is a straightforward consequence of Corollary 3.5 and Lemma 4.3. The fulfillment of the assumptions of Lemma 4.3 follows from Corollary 3.5, Lemma 4.1, estimation (3.16) and Lemma 3.6.

4.4 (**Proof of the estimation (2.2) in Theorem 2.3**). To prove the estimation (2.2), we shall use the following lemma:

Lemma 4.4 [9,10]. Let $\{f_n(\cdot)\}_{n\in N}$ and $\{h_n\}_{n\in N}$ be holomorphic in some region $\Lambda_1\subset\mathbb{C}$ functions. Let $\Lambda_0=\{\lambda\,\big|\,|\lambda-\lambda_0|\leq\delta\}\subset\Lambda_1$, Γ be the boundary of Λ_0, and the following assertions hold:

1) every function $f_n(\cdot)$ has in Λ_0 only one zero $\lambda=\lambda_0$ and the multiplicity of λ_0 is ν for every $f_n(\cdot)$, $n\in\mathbb{N}$,

2) $$\varliminf_{n\in\mathbb{N}} |f_n^{(\nu)}(\lambda_0)|=\beta>0, \tag{4.12}$$

3) $$\sup_{n\in\mathbb{N},\lambda\in\Gamma}|f_n(\lambda)|\leq c,\quad \sup_{n\in\mathbb{N},\lambda\in\Gamma}|1/f_n(\lambda)|\leq c, \tag{4.13}$$

4) $$\eta_n=\eta_n(\Gamma)=\sup_{\lambda\in\Gamma}|f_n(\lambda)-h_n(\lambda)|\to 0\quad (n\in\mathbb{N}). \tag{4.14}$$

Denote by $\sigma(h_n)$ the set of zeros of the $h_n(\cdot)$, by $\nu(\lambda,h_n)$ the multiplicity of the zero λ of $h_n(\cdot)$, and by $\overline{\lambda}$ the sum $\sum_{\lambda\in\Lambda_0\cap\sigma(h_n)}[\nu(\lambda,h_n)/\nu]\cdot\lambda$.

Then $\Lambda_0\cap\sigma(h_n)\neq\varnothing$ and $|\overline{\lambda}-\lambda_0|\leq c\eta_n$ for almost all $n\in\mathbb{N}$. ■

The estimation 2.2 in Theorem 2.3 is a straightforward consequence of Corollary 3.5 and Lemma 4.4 for $f_n(\cdot)=\det M_n(\cdot)$, $h_n(\cdot)=\det T_n(\cdot)$. Assumptions 1,3) and 4) follow from Corollary 3.5, Lemma 4.1 estimation (3.16) and Lemma 3.6. To apply lemma 4.4, we must yet guarantee assumption 2) there. For this we use the freedom in the choice of functions $g_n^1(\cdot)$ and choose them so that (4.12) is fulfilled. In Lemma 4.5 in the next paragraph we shall demonstrate one possibility of such a choice.

4.5 (**One auxiliary result**). Denote by $f(\lambda)\,{}^{(l)}_{\lambda=\lambda_0}$ the value of the l-th derivative of the function $f(\cdot)$ at the $\lambda=\lambda_0$.

Lemma 4.5 [10]. Let in paragraph 3.2 the functions $v_s(\cdot)$, $s=1,\dots,m$ be determined by (3.2) and the functions $g_n^r(\cdot)$, $r=1,\dots,m$, $n\in\mathbb{N}$ as follows:

$$g_n^r(\lambda)=g_n^r+(\lambda-\lambda_0)g_n^{r1}+..+(\lambda-\lambda_0)^{\varkappa}g_n^{r\varkappa}, \tag{4.15}$$

with g_n^r from paragraph 3.2, $\varkappa=max\,\{\nu_i=\nu(u_i^0)\,|\,i=1,\dots,m\}$ and

$$g_n^{rk}=\sum_{l=1}^{l=m}\alpha_{nl}^{rk}g_n^l,\quad k=1,\dots,\varkappa \tag{4.16}$$

where α_{nl}^{rk} are determined by following equalities:

$$\alpha_{nl}^{rk}=-\sum_{j=1}^{j=k}\langle p_nPu_l^j,g_n^{r,k-j}\rangle,\qquad g_n^{r0}=g_n^r,\quad l=1,\dots,m. \tag{4.17}$$

Then for $\nu=\nu_1+\dots+\nu_m$ we have

$$det M_n(\lambda)|_{\lambda=\lambda_0}^{(\nu)}\to 1\quad(n\in\mathbb{N}). \tag{4.18}$$

Proof. 1) Let us denote $\varphi_n^{sr}(\lambda)=\langle p_nPu_s(\lambda),g_n^r(\lambda)\rangle$, $s,r=1,\dots,m$. By (3.1) we have $\varphi_n^{sr}(\lambda_0)\to\delta_{sr}$ $(n\in\mathbb{N})$. Let us prove that

$$\varphi_n^{sr}(\lambda)|_{\lambda=\lambda_0}^{(l)}\to 0\quad(n\in\mathbb{N}),\qquad s,r=1,\dots,m,\quad l=1,\dots \tag{4.19}$$

For this recall that $u_s(\lambda)=u_s^0+(\lambda-\lambda_0)u_s^1+\dots+(\lambda-\lambda_0)^{\varkappa}u_s^{\varkappa}$ with $u_s^j=0$ for $j\geq\nu_s$. Define , also, $g_n^{rk}=0$ for $k>\varkappa$. Then the direct calculations give:

$$\varphi_n^{sr}(\lambda)|_{\lambda=\lambda_0}^{(l)}=\sum_{j=0}^{j=l}\binom{l}{j}\langle p_nPu_s(\lambda)|_{\lambda=\lambda_0}^{(j)},\,g_n^r(\lambda)|_{\lambda=\lambda_0}^{(l-j)}\rangle=$$

$$=\sum_{j=0}^{j=l}l!\,\langle p_nPu_s^j,\,g_n^{r,l-j}\rangle=l!\,\langle p_nPu_s^0,\,g_n^{r,l-j}\rangle+l!\sum_{j=1}^{j=l}\langle p_nPu_s^j,\,g_n^{r,l-j}\rangle=$$

$$=l!\,\langle p_nPu_s^0,\,\sum_{i=1}^{i=m}\alpha_{ni}^{rl}g_n^i\rangle-l!\,\alpha_{ns}^{rl}=l!\sum_{i=1}^{i=m}\alpha_{ni}^{rl}\varphi_n^{si}(\lambda_0)-l!\,\alpha_{ns}^{rl}.$$

Therefore, we have

$$\varphi_n^{sr}(\lambda)|_{\lambda=\lambda_0}^{(l)}\to\;l!\,\alpha_{ns}^{rl}-l!\,\alpha_{ns}^{rl}=0\quad(n\in\mathbb{N}). \tag{4.20}$$

2) Let us prove now that for $r,s=1,\dots,m$ there hold:

$$m_n^{rs}(\lambda)|_{\lambda=\lambda_0}^{(l)}\to 0\ (n\in\mathbb{N}),\ l=1,\dots,\nu_s-1,\qquad m_n^{rs}(\lambda)|_{\lambda=\lambda_0}^{(\nu_s)}\to\delta_{rs}\ (n\in\mathbb{N}). \tag{4.21}$$

Examine the functions $\psi_n^{sr}(\cdot)$: $\psi_n^{sr}(\lambda)=(\lambda-\lambda_0)^{\nu_s}[\delta_{rs}-m_n^{rs}(\lambda)]$. Clearly,

$$\psi_n^{sr}(\lambda)|_{\lambda=\lambda_0}^{(l)}=0,\ l=1,\dots,\nu_s-1,\qquad \psi_n^{sr}(\lambda)|_{\lambda=\lambda_0}^{(\nu_s)}=m_n^{rs}(\lambda)-\delta_{rs}. \tag{4.22}$$

Using the definition of $m_n^{rs}(\lambda)$ in (3.11) and $v_s(\lambda)$ in (3.2) gives us:

$$\psi_n^{sr}(\lambda)=(\lambda-\lambda_0)^{\nu_s}\langle p_nPR_n^{-1}(\lambda)v_s(\lambda),\,g_n^r(\lambda)\rangle=\langle p_nPR_n^{-1}(\lambda)A(\lambda),\,g_n^r(\lambda)\rangle.$$

Now from $A(\lambda)=R_n(\lambda)-K_n(\lambda)$ and the definition of $K_n(\lambda)$ in (3.3) it follows:

$$\psi_n^{sr}(\lambda)=\langle p_nPu_s(\lambda),g_n^r(\lambda)\rangle-\langle p_nPR_n^{-1}(\lambda)\sum_{i=1}^{i=m}\langle p_nPu_s(\lambda),g_n^i(\lambda)\rangle v_i(\lambda),g_n^r(\lambda)\rangle=$$

$$=\varphi_n^{sr}(\lambda)-\sum_{i=1}^{i=m}\langle p_nPu_s(\lambda),\,g_n^i(\lambda)\rangle\,\langle p_nPR_n^{-1}(\lambda)v_i(\lambda),\,g_n^r(\lambda)\rangle=$$

$$=\varphi_n^{sr}(\lambda)-\sum_{i=1}^{i=m}\varphi_n^{si}(\lambda)(\lambda-\lambda_0)^{-\nu_i}\psi_n^{ir}(\lambda)=\varphi_n^{sr}(\lambda)-\sum_{i=1}^{i=m}\varphi_n^{si}(\lambda)[\delta_{ir}-m_n^{ri}(\lambda)].$$

So we have

$$\sum_{l=1}^{l=m}\varphi_n^{sl}(\lambda)m_n^{rl}(\lambda)=\psi_n^{sr}(\lambda). \tag{4.23}$$

From (4.23), (4.22) and (4.19) we get

$$\sum_{l=1}^{l=m}\varphi_n^{sl}(\lambda_0)m_n^{rl}(\lambda_0)=\psi_n^{sr}(\lambda_0)=0$$

and, therefore, $m_n^{rs}(\lambda_0)\to 0$ $(n\in\mathbf{N})$ because $\varphi_n^{sl}(\lambda_0)\to\delta_{sl}$ $(n\in\mathbf{N})$.

For $l=1,\dots,\nu_s$ calculating the l-th derivative gives us:

$$\sum_{l=1}^{l=m}\sum_{j=0}^{j=l}\binom{l}{j}\varphi_n^{sl}(\lambda)\Big|_{\lambda=\lambda_0}^{(j)}m_n^{rl}(\lambda)\Big|_{\lambda=\lambda_0}^{(l-j)}=\psi_n^{sr}(\lambda)\Big|_{\lambda=\lambda_0}^{(l)}=0, \quad l=1,\dots,\nu_s-1,$$

$$\sum_{l=1}^{l=m}\sum_{j=0}^{j=l}\binom{l}{j}\varphi_n^{sl}(\lambda)\Big|_{\lambda=\lambda_0}^{(j)}m_n^{rl}(\lambda)\Big|_{\lambda=\lambda_0}^{(l-j)}=\psi_n^{sr}(\lambda)\Big|_{\lambda=\lambda_0}^{(l)}=\delta_{sr}-m_n^{rs}(\lambda_0), \quad l=\nu_s.$$

From these equalities we get (4.21) using (4.19) together with $\varphi_n^{sl}(\lambda_0)\to\delta_{sl}$, $m_n^{rs}(\lambda_0)\to 0$ $(n\in\mathbf{N})$.

3) Let us prove now that (4.18) holds. For this we note that the function $\det M_n(\cdot)$ is the sum of products of m elements where in every product all elements are from different rows. Using the equalities (4.21), it is not difficult to see that the ν-th derivative $(\nu=\nu_1+\dots+\nu_m)$ of the product $m_n^{11}(\lambda)\cdot\ldots\cdot m_n^{ll}(\lambda)\cdot\ldots\cdot m_n^{mm}(\lambda)$ converges to 1 and the ν-th derivatives of all the other products converge to 0. From this we get (4.18). ∎

References

1. Вайникко Г. М. Регулярная сходимость операторов и приближенное решение уравнений. В сб.: Итоги науки и техники. Математический анализ. Т. 16, 1978, 5-53.
2. Вайникко Г. М., Карма О. О. О быстроте сходимости приближенных методов в проблеме собственных значений с нелинейным вхождением параметра//Ж. вычисл. матем. и матем. физ., 1974, т. 14, № 6, 1393-1408.
3. Вайникко Г., Карма О. Оценка скорости сходимости собственных значений в приближенных методах//Уч. зап. Тартуск. ун-та, 1988, вып. 833, 75-83.
4. Гохберг И. Ц. О линейных операторах, аналитический зависящих от параметра//Докл. АН СССР, 1951, т. 78, № 4, 629-632.
5. Данфорд Н., Шварц Дж. Линейные операторы. Общая теория. -М.: ИИЛ, 1962.
6. Ени В. М. О кратности собственного вектора и кратности характеристического числа матричного пучка//Известия АН МССР, 1965, № 7, 94-97.
7. Иосида К. Функциональный анализ. -М.: Мир, 1967.

8. К а р м а О. О. Асимптотические оценки погрешности приближенных характеристических значений голоморфных фредгольмовых оператор-функций//Ж. вычисл. матем. и матем.физ., 1971, т. 11, № 3, 559-568.
9. К а р м а О. Некоторые асимптотические оценки близости корней голоморфных функций//Уч. зап. Тартуск. ун-та, 1979, вып. 492, 37-46.
10. К а р м а О. О. Аппроксимация в проблеме собственных значений с голоморфной зависимостью оператора от параметра I// Рукопись деп. в ВИНИТИ 11.01.82 г. № 130-82 Деп.
11. К р е й н С. Г., Т р о ф и м о в В. П. О нетеровых операторах, голоморфно зависящих от параметров//Тр. матем. фак. Воронежск. ун-та, 1970, 63-85.
12. М а р к у с А. С., С и г а л Е. И. О кратности характеристического числа аналитической оператор-функции//Матем. исследования, 1970, т. 5, вып. 3, 129-147.
13. G r i g o r i e f f R. D. Diskrete Approximation von Eigenwertproblemen. I: Qualitative Konvergenz, II: Konvergenzordnung//Numer. Math., 1975, 24, 355-374, 415-433.
14. S t u m m e l F. Diskrete Konvergenz linearer Operatoren I//Math. Annal., 1970, 190, 45-92.
15. W o l f R. Zur Stabilität der algebraischen Vielfachheit von Eigenwerten von holomorphen Fredholm-Operatorfunktionen//Appl. Analysis, 1979, 9, 165-177.

О СХОДИМОСТИ СОБСТВЕННЫХ ЗНАЧЕНИЙ ПРИ АППРОКСИМАЦИИ ЗАДАЧИ

О. Карма

Резюме

В статье рассматривается проблема собственных значений для голоморфных фредгольмовых оператор-функций. Используется схема дискретной аппроксимации банаховых пространств и регулярная аппроксимация оператор-функций. Приводятся доказательства асимптотических оценок скорости сходимости к точному собственному значению как каждого собственного значения приближенной задачи в отдельности, так и для их взвешенного арифметического. Статья является укороченным вариантом статьи [10].

Acta et commentationes universitatis Tartuensis, 863, 1989, 57-70.

РЕГУЛЯРИЗОВАННЫЙ ИТЕРАЦИОННЫЙ МЕТОД С ОТНОШЕНИЕМ РЕЛЕЯ

Таймо Саан

В данной работе исследуется скорость сходимости итерационного метода с отношением Релея. Этот метод обладает свойством монотонного убывания норм невязок и для эрмитовых и для неэрмитовых матриц. В случае нормальной матрицы метод имеет пятый порядок сходимости. При матрице простой структуры для данного метода сходимость линейная. В случае симметричной матрицы этот метод исследован в [4].

В работах [2], [5], [6] и [7] исследована быстрота сходимости итерационных методов RQI (Rayleigh Quotient Iteration) и OTSI (Ostrowski's Two-Sided Iteration) для отыскания собственных значений и собственных векторов матрицы T. В случае эрмитовой матрицы T эти методы совпадают и обладают свойством монотонного убывания норм невязок (см. [2], [7]). В случае ненормальной матрицы T указанные методы свойством монотонного убывания норм невязок не обладают.

Пусть дана комплексная $(n\times n)$-матрица T. Через $\sigma(T)=\{\lambda_1,\lambda_2,\dots,\lambda_r\}$ и $\{X_1,X_2,\dots,X_r\}$, $r\le n$, обозначим множество попарно различных собственных значений матрицы T и соответствующие им корневые подпространства.

<u>Определение</u> 1. Отношением Релея для матрицы T называется функция $\rho(v)=(Tv,v)/(v,v)$, где $v\in\mathbb{C}^n\setminus\{0\}$.

Для отыскания собственного значения λ матрицы T и соответствующего ему собственного вектора x используем следующий итерационный метод. Сначала выберем единичный вектор v_0. Затем для $k=0,1,2,\dots$ повторим следующие шаги:

(I) вычислим $\rho_k=\rho(v_k)=(Tv_k,v_k)$;

(II) решим уравнение $[\alpha_k+(T^*-\bar{\rho}_k)(T-\rho_k)]w_k=v_k$, где $\alpha_k>0$ и $\lim\limits_{k\to\infty}\alpha_k=0$;

(III) формируем $v_{k+1}=w_{k+1}/\|w_{k+1}\|$.

Шаги (II) и (III) можно объединить в один,

(II') $v_{k+1}=\tau_k^2[\alpha_k+(T^*-\overline{\rho}_k)(T-\rho_k)]^{-1}v_k$, где

$\tau_k^2=\|[\alpha_k+(T^*-\overline{\rho}_k)(T-\rho_k)]^{-1}v_k\|^{-1}$.

Рассмотрим сначала локальную сходимость итерационного метода без регуляризирующего члена $(\alpha_k=0)$. Дополнительно к обозначениям $\sigma(T)=\{\lambda_1,\lambda_2,\dots,\lambda_r\}$ и $\{X_1,X_2,\dots,X_r\}$ вводим множество корневых векторов $\{Y_1,Y_2,\dots,Y_r\}$ матрицы T^*, соответствующих значениям $\overline{\lambda}_1,\overline{\lambda}_2,\dots,\overline{\lambda}_r$.

Из высшей алгебры известны следующие факты.

Факт 1. Если x_i собственный вектор матрицы T, соответствующий λ_i, а y_j собственный вектор матрицы T^*, соответствующий $\overline{\lambda}_j$, $\overline{\lambda}_j\neq\lambda_i$, то $(x_i,y_j)=0$ (см. [5, стр.21]).

Обозначим алгебраические кратности собственных значений λ_i через n_i, тогда $n_1+n_2+\dots n_r=n$.

Факт 2. Если корневое подпространство X_i состоит только из собственных векторов, т.е. $X_i=\{x|Tx=\lambda_i x\}$, то Y_i состоит только из собственных векторов матрицы T^* и можно выбирать базисные векторы $x_{i1},x_{i2},\dots,x_{in_i}$ и $y_{i1},y_{i2},\dots,y_{in_i}$ такие, что

$$(x_{ij},y_{ik})=0 \quad \text{при } j\neq k,$$

$$(x_{ij},y_{ik})\neq 0 \quad \text{при } j=k$$

(см. [1, стр.244]).

Допустим, что собственное значение λ_1 такое, что соответствующее ему корневое подпространство X_1 состоит только из собственных векторов. Обозначим через P проектор Рисса в $\mathbb{C}^n$, проектирующий на X_1. Тогда проектор P^* проектирует на собственное подпространство матрицы T^*, соответсвующее собственному значению $\overline{\lambda}_1$,т.е. на Y_1.

Любой вектор $x_1\in X_1$, $\|x_1\|=1$ можно представить в виде

$$x_1=P^*x_1+(I-P^*)x_1=\|P^*x_1\|\frac{P^*x_1}{\|P^*x_1\|}+(I-P^*)x_1.$$

Обозначим $a=\|P^*x_1\|$, $y_1=P^*x_1/\|P^*x_1\|$ и $\hat{y}=(I-P^*)x_1$. Отметим, что y_1 будет нормированный собственный вектор матрицы T^*. Поступим с вектором y_1 так же как и с вектором x_1.

$$y_1=Py_1+(I-P)y_1=\|Py_1\|\frac{Py_1}{\|Py_1\|}+(I-P)y_1.$$

Так как P проектор Рисса, то $Py_1/\|Py_1\|=x_1$. Будем пользоваться и обозначениями $b=\|Py_1\|$, $\hat{x}=(I-P)y_1$.

Теперь можно собственные векторы x_1 и y_1 представить в виде

$$x_1=ay_1+\hat{y}\ ,\quad y_1=bx_1+\hat{x}, \tag{1}$$

где $\hat{y}\in Y_1{}^{T}=Y_2+Y_3+\ldots+Y_r$ и $\hat{x}\in X_1{}^{T}=X_2+X_3+\ldots+X_r$.

Убедимся, что $a=b=1/(x_1,y_1)\in\mathbb{R}$. Для этого преобразуем

$$(x_1,y_1)=(x_1,\frac{P^*x_1}{\|P^*x_1\|})=\frac{1}{\|P^*x_1\|}(x_1,P^*x_1)=\frac{1}{\|P^*x_1\|}(Px_1,x_1)=\frac{1}{\|P^*x_1\|}=\frac{1}{a},$$

$$(x_1,y_1)=(\frac{Py_1}{\|Py_1\|},y_1)=\frac{1}{\|Py_1\|}(Py_1,y_1)=\frac{1}{\|Py_1\|}(y_1,P^*y_1)=\frac{1}{\|Py_1\|}=\frac{1}{b}.$$

Так как $\|x_1\|=\|y_1\|=1$, то $|(x_1,y_1)|\le1$ и $\|P^*x_1\|=1/(x_1,y_1)\ \ge1$.

Хорошей вычислимой мерой точности пары $[\rho_k,v_k]$, $\|v_k\|=1$, как собственной пары для матрицы T является вектор невязки $r_k=(T-\rho_k)v_k$.

Теорема 1. Пусть последовательность $\{v_k\}$ такая, что $\mathrm{dist}(v_k,X_1)\to0$ при $k\to\infty$. Если подпространство X_1 состоит только из собственных векторов и в итерационном процессе (I)-(III) $\alpha_k=0$, $k=0,1,2\ldots$, то имеет место сходимость $v_k\to x_1$, $x_1\in X_1$ и

1) в случае $\|P^*x_1\|>1$ выполняются

$$\lim_{k\to\infty}\frac{|\lambda_1-\rho_{k+1}|}{|\lambda_1-\rho_k|}=\lim_{k\to\infty}\frac{\|r_{k+1}\|}{\|r_k\|}=q,\quad \text{где } q=1-1/\|P^*x_1\|^2;$$

2) в случае $\|P^*x_1\|=1$ выполняются

$$\lim_{k\to\infty}\frac{|\lambda_1-\rho_{k+1}|}{|\lambda_1-\rho_k|^5}\le\frac{1}{\mu^2}\quad\text{и}\quad\lim_{k\to\infty}\frac{\|r_{k+1}\|}{\|r_k\|^5}\le\frac{1}{\mu^3}|\rho(z)-\lambda_1|,$$

где μ – минимальное отличающееся от нуля собственное значение матрицы $(T^*-\bar{\lambda}_1)(T-\lambda_1)$ и z некоторый соответствующий ему собственный вектор. Равенства в оценках выполняются, когда начальный вектор v_0 не ортогонален множеству $\{w|\ \ (T^*-\bar{\lambda}_1)(T-\lambda_1)w=\mu w\}$ и $\rho(z)\ne\lambda_1$.

<u>Доказательство</u>. Рассмотрим случай $\|P^*x_1\|>1$. Представим приближение v_k в виде

$$v_k=\beta_k(x_1^k+\varepsilon_k z_k), \tag{2}$$

где $x_1^k=Pv_k/\|Pv_k\|\in X_1$, $z_k\in X_1^T$, $\|z_k\|=1$ и коэффициент β_k гарантирует, что вектор v_k нормированный. Преобразуем

$$\rho_k-\lambda_1=(Tv_k,v_k)-\lambda_1(v_k,v_k)=((T-\lambda_1)v_k,v_k)=$$

$$=((T-\lambda_1)\beta_k(x_1^k+\varepsilon_k z_k),\beta_k(x_1^k+\varepsilon_k z_k))=$$

$$=|\beta_k|^2((T-\lambda_1)\varepsilon_k z_k,x_1^k+\varepsilon_k z_k)=$$

$$=|\beta_k|^2\varepsilon_k((T-\lambda_1)z_k,x_1^k)+O(|\varepsilon_k|^2). \tag{3}$$

Обозначим $y_1^k=P^*x_1^k/\|P^*x_1^k\|$ и представим векторы x_1^k и y_1^k в виде

$$x_1^k=a_k y_1^k+\hat{y}_k\ ,\quad y_1^k=b_k x_1^k+\hat{x}_k, \tag{4}$$

где $\hat{y}_k\in Y_1^T$, $\hat{x}_k\in X_1^T$ и $a_k=b_k=1/(x_1^k,y_1^k)$.

Если для какого-то индекса k приближение $\rho_k=\lambda_1$, то из итерационного шага (II) следует, что v_{k+1} будет точный собственный вектор. Продолжая итерационный процесс, получим, что $[\rho_m,v_m]$ будет точная собственная пара для всех $m\geq k+1$.

Дальнейшие рассуждения проводятся в случае $\rho_k\neq\lambda_1$. Так как $\alpha_k=0$, $k=0,1,2,\ldots$, то приближение v_{k+1} в итерационном процессе можно представить в виде

$$v_{k+1}=\tau_k^2[(T^*-\bar{\rho}_k)(T-\rho_k)]^{-1}v_k.$$

Учитывая представления (2) и (4), преобразуем

$$v_{k+1}=\tau_k^2(T-\rho_k)^{-1}(T^*-\bar{\rho}_k)^{-1}\beta_k(x_1^k+\varepsilon_k z_k)=$$

$$=\tau_k^2\beta_k(T-\rho_k)^{-1}(T^*-\bar{\rho}_k)^{-1}(a_k y_1^k+\hat{y}_k+\varepsilon_k P^*z_k+\varepsilon_k(I-P^*)z_k)=$$

$$=\tau_k^2\beta_k(T-\rho_k)^{-1}\left(\frac{a_k y_1^k+\varepsilon_k P^*z_k}{\bar{\lambda}_1-\bar{\rho}_k}+(T^*-\bar{\rho}_k)^{-1}(\hat{y}_k+\varepsilon_k(I-P^*)z_k)\right)=$$

$$=\frac{\tau_k^2\beta_k}{\bar{\lambda}_1-\bar{\rho}_k}(T-\rho_k)^{-1}(a_k y_1^k+\varepsilon_k P^*z_k+(\bar{\lambda}_1-\bar{\rho}_k)(T^*-\bar{\rho}_k)^{-1}(\hat{y}_k+\varepsilon_k(I-P^*)z_k))=$$

$$=\frac{\tau_k^2\beta_k a_k}{\bar{\lambda}_1-\bar{\rho}_k}(T-\rho_k)^{-1}(y_1^k+t_k),$$

где $t_k=\dfrac{1}{a_k}(\varepsilon_k P^*z_k+(\bar{\lambda}_1-\bar{\rho}_k)(T^*-\bar{\rho}_k)^{-1}(\hat{y}_k+\varepsilon_k(I-P^*)z_k))$.

Так как $\hat{y}_k+\varepsilon_k(I-P^*)z_k\ \in Y_1^T$, то $\|(T^*-\bar{\rho}_k)^{-1}(\hat{y}_k+\varepsilon_k(I-P^*)z_k)\|\leq\mathrm{const}$ и, учитывая еще равенство (3), получим, что $\|t_k\|=O(|\varepsilon_k|)$. Теперь

$$v_{k+1}=\frac{\tau_k^2\beta_k a_k}{\bar\lambda_1-\bar\rho_k}(T-\rho_k)^{-1}(y_1^k+t_k)=$$

$$=\frac{\tau_k^2\beta_k a_k}{\bar\lambda_1-\bar\rho_k}(T-\rho_k)^{-1}(b_k x_1^k+\hat{x}_k+Pt_k+(I-P)t_k)=$$

$$=\frac{\tau_k^2\beta_k a_k}{\bar\lambda_1-\bar\rho_k}\left(\frac{b_k x_1^k+Pt_k}{\lambda_1-\rho_k}+(T-\rho_k)^{-1}(\hat{x}_k+(I-P)t_k)\right)=$$

$$=\frac{\tau_k^2\beta_k a_k b_k}{|\lambda_1-\rho_k|^2}\left(x_1^k+\frac{1}{b_k}Pt_k+\frac{(\lambda_1-\rho_k)}{b_k}(T-\rho_k)^{-1}(\hat{x}_k+(I-P)t_k)\right).$$

Так как $x_1^k+\frac{1}{b_k}Pt_k\in X_1$ и $(T-\rho_k)^{-1}(\hat{x}_k+(I-P)t_k)\in X_1{}^{\tau}$, то учитывая представление (2), получим

$$x_1^{k+1}=\frac{x_1^k+\frac{1}{b_k}Pt_k}{\|x_1^k+\frac{1}{b_k}Pt_k\|}\quad \text{и}\quad z_{k+1}=\frac{(T-\rho_k)^{-1}(\hat{x}_k+(I-P)t_k)}{\|(T-\rho_k)^{-1}(\hat{x}_k+(I-P)t_k)\|}. \tag{5}$$

Теперь вектор v_{k+1} приобретает вид

$$v_{k+1}=\frac{\tau_k^2\beta_k a_k b_k\|x_1^k+\frac{1}{b_k}Pt_k\|}{|\lambda_1-\rho_k|^2}\left(x_1^{k+1}+\frac{(\lambda_1-\rho_k)\|(T-\rho_k)^{-1}(\hat{x}_k+(I-P)t_k)\|}{b_k\|x_1^k+\frac{1}{b_k}Pt_k\|}z_{k+1}\right). \tag{6}$$

Из предположения теоремы $\mathrm{dist}(v_k,X_1)\to 0$ при $k\to\infty$ следует, что $\beta_k\to 1$ и

$$1=\|v_k\|^2=|\beta_k|^2\|x_1^k+\varepsilon_k z_k\|^2=|\beta_k|^2(1+O(|\varepsilon_k|)).$$

Так как $\|t_k\|=O(|\varepsilon_k|)$, $\|P\|<\infty$ и $1/b_k=(x_1^k,y_1^k)$, то

$$\|x_1^k+\frac{1}{b_k}Pt_k\|=1+O(|\varepsilon_k|).$$

Опираясь на представления (2), (3) и (6), вычислим

$$\frac{\varepsilon_{k+1}}{\varepsilon_k}=\frac{(\lambda_1-\rho_k)\|(T-\rho_k)^{-1}(\hat{x}_k+(I-P)t_k)\|}{\varepsilon_k b_k\|x_1^k+\frac{1}{b_k}Pt_k\|}=$$

$$=-\frac{|\beta_k|^2\varepsilon_k[((T-\lambda_1)z_k,x_1^k)+O(|\varepsilon_k|)]\,\|(T-\rho_k)^{-1}(\hat{x}_k+(I-P)t_k)\|}{\varepsilon_k b_k\|x_1^k+\frac{1}{b_k}Pt_k\|}=$$

$$=-\frac{((T-\lambda_1)z_k,x_1^k)\|(T-\rho_k)^{-1}(\hat{x}_k+(I-P)t_k)\|}{b_k}(1+O(|\varepsilon_k|))=$$

$$=-\frac{((T-\lambda_1)z_k,x_1^k)\|(T-\rho_k)^{-1}\hat{x}_k\|}{b_k}(1+O(|\varepsilon_k|)).$$

Учтем, что $z_k=\frac{(T-\rho_{k-1})^{-1}(\hat{x}_{k-1}+(I-P)t_{k-1})}{\|(T-\rho_{k-1})^{-1}(\hat{x}_{k-1}+(I-P)t_{k-1})\|}$ и преобразуем

$$\frac{\varepsilon_{k+1}}{\varepsilon_k}=-\frac{((T-\lambda_1)(T-\rho_{k-1})^{-1}(\hat{x}_{k-1}+(I-P)t_{k-1}),x_1^k)\|(T-\rho_k)^{-1}\hat{x}_k\|}{b_k\|(T-\rho_{k-1})^{-1}(\hat{x}_{k-1}+(I-P)t_{k-1})\|}(1+O(|\varepsilon_k|))=$$

$$=-\frac{((T-\lambda_1)(T-\rho_{k-1})^{-1}\hat{x}_{k-1},x_1^k)\|(T-\rho_k)^{-1}\hat{x}_k\|}{b_k\|(T-\rho_{k-1})^{-1}\hat{x}_{k-1}\|}(1+O(|\varepsilon_{k-1}|))=$$

$$=-\frac{(\hat{x}_{k-1},x_1^k)\|(T-\rho_k)^{-1}\hat{x}_k\|}{b_k\|(T-\rho_{k-1})^{-1}\hat{x}_{k-1}\|}(1+O(|\varepsilon_{k-1}|))=$$

$$=-\frac{(\hat{x}_{k-1},x_1^k)}{b_k}(1+O(|\varepsilon_{k-1}|))=-\frac{(y_1^{k-1}-b_{k-1}x_1^{k-1},x_1^k)}{b_k}(1+O(|\varepsilon_{k-1}|)).$$

Поскольку $\|Pt_k\|=O(|\varepsilon_k|)$, то из (5) следует, что $\|x_1^k-x_1^{k+1}\|=O(|\varepsilon_k|)$. Так как $y_1^k=P^*x_1^k/\|P^*x_1^k\|$, то $\|y_1^k-y_1^{k+1}\|=O(|\varepsilon_k|)$ и $|b_k-b_{k+1}|=|1/(x_1^k,y_1^k)-1/(x_1^{k+1},y_1^{k+1})|=O(|\varepsilon_k|)$. Получим

$$\frac{\varepsilon_{k+1}}{\varepsilon_k}=-\frac{(y_1^{k-1}-b_{k-1}x_1^{k-1},x_1^k)}{b_k}(1+O(|\varepsilon_{k-1}|))=$$

$$=-\frac{(y_1^k-b_kx_1^k,x_1^k)}{b_k}(1+O(|\varepsilon_{k-1}|))=-\left(\frac{(x_1^k,y_1^k)}{b_k}-1\right)(1+O(|\varepsilon_{k-1}|))=$$

$$=1-|(x_1^k,y_1^k)|^2+O(|\varepsilon_{k-1}|)=1-\frac{1}{\|P^*x_1^k\|^2}+O(|\varepsilon_{k-1}|). \tag{7}$$

Так как X_1 состоит только из собственных векторов, то $q_1\equiv\max\limits_{x\in X_1}(1-1/\|P^*x\|^2)<1$. Следовательно,

$$\left|\frac{\varepsilon_{k+1}}{\varepsilon_k}\right|\le q_1+O(|\varepsilon_{k-1}|)\le q_2<1.$$

Убедимся, что $v_k\to x_1$, $x_1\in X_1$ при $k\to\infty$. Для этого покажем, что $\sum\limits_{k=1}^{\infty}\|v_k-v_{k+1}\|<\infty$. В силу $\beta_k=1+O(|\varepsilon_k|)$, $x_1^{k+1}=\frac{x_1^k+\frac{1}{b_k}Pt_k}{\|x_1^k+\frac{1}{b_k}Pt_k\|}$, и соотношения $\left|\frac{\varepsilon_{k+1}}{\varepsilon_k}\right|\le q_2<1$, оценим

$$\sum_{k=1}^{\infty}\|v_k-v_{k+1}\|=\sum_{k=1}^{\infty}\|\beta_k(x_1^k+\varepsilon_kz_k)-\beta_{k+1}(x_1^{k+1}+\varepsilon_{k+1}z_{k+1})\|=$$

$$=\sum_{k=1}^{\infty}\|\beta_k(x_1^k+\varepsilon_kz_k)-\beta_{k+1}(x_1^k+\varepsilon_kz_k)+\beta_{k+1}(x_1^k+\varepsilon_kz_k)-\beta_{k+1}(x_1^{k+1}+\varepsilon_{k+1}z_{k+1})\|\le$$

$$\le \sum_{k=1}^{\infty} |\beta_k-\beta_{k+1}| \|x_1^k+\varepsilon_k z_k\| + \sum_{k=1}^{\infty} |\beta_{k+1}| \|x_1^k+\varepsilon_k z_k - x_1^{k+1}-\varepsilon_{k+1} z_{k+1}\| \le$$

$$\le \sum_{k=1}^{\infty} |O(|\varepsilon_k|)| \|x_1^k+\varepsilon_k z_k\| + \sum_{k=1}^{\infty} |\beta_{k+1}| |O(|\varepsilon_k|)| \le \mathrm{const} \sum_{k=1}^{\infty} |\varepsilon_k| \le$$

$$\le \mathrm{const}\, |\varepsilon_1| \sum_{k=0}^{\infty} q_2{}^k = \mathrm{const}\, |\varepsilon_1|/(1-q_2) < \infty.$$

В итоге получается, что $v_k \to x_1$ при $k \to \infty$. Вычислим (см. (1))

$$(\hat{x},x_1)=(y_1-bx_1,x_1)=(y_1,x_1)-1/(x_1,y_1) =$$

$$=[|(x_1,y_1)|^2-1]/(x_1,y_1) \ne 0.$$

Опираясь на результаты (3), (5) и (7), вычислим

$$\frac{\lambda_1-\rho_{k+1}}{\lambda_1-\rho_k} = \frac{|\beta_{k+1}|^2 \varepsilon_{k+1}((T-\lambda_1)z_{k+1},x_1^{k+1})+O(|\varepsilon_{k+1}|^2)}{|\beta_k|^2 \varepsilon_k((T-\lambda_1)z_k,x_1^k)+O(|\varepsilon_k|^2)} =$$

$$= \frac{\varepsilon_{k+1}((T-\lambda_1)z_{k+1},x_1^{k+1})+O(|\varepsilon_{k+1}|^2)}{\varepsilon_k((T-\lambda_1)z_k,x_1^k)+O(|\varepsilon_k|^2)} = \frac{\varepsilon_{k+1}(\hat{x},x_1)+O(|\varepsilon_k|^2)}{\varepsilon_k(\hat{x},x_1)+O(|\varepsilon_{k-1}|^2)} = q+O(|\varepsilon_{k-1}|), \tag{8}$$

где $q=1-1/\|P^*x_1\|^2$.

Учитывая (2) и (3), переформируем вектор невязки

$$r_k=(T-\rho_k)v_k=(T-\rho_k)\beta_k(x_1^k+\varepsilon_k z_k)=$$

$$=\beta_k(\lambda_1-\rho_k)x_1^k+\beta_k\varepsilon_k(T-\rho_k)z_k=$$

$$=-\beta_k|\beta_k|^2\varepsilon_k((T-\lambda_1)z_k,x_1^k)x_1^k+\beta_k\varepsilon_k(T-\rho_k)z_k+O(|\varepsilon_k|^2)=$$

$$=-\varepsilon_k((T-\lambda_1)z_k,x_1^k)x_1^k+\varepsilon_k(T-\rho_k)z_k+O(|\varepsilon_k|^2). \tag{9}$$

Вычислим

$$\|-(\hat{x},x_1)x_1+\hat{x}\|^2= \|-(y_1-bx_1,x_1)x_1+y_1-bx_1\|^2=$$

$$= \|-(y_1,x_1)x_1+bx_1+y_1-bx_1\|^2= \|-(y_1,x_1)x_1+y_1\|^2=$$

$$= |(y_1,x_1)|^2-(y_1,x_1)(x_1,y_1)-(x_1,y_1)(y_1,x_1)+1=$$

$$= 1-|(x_1,y_1)|^2= q.$$

Опираясь на результаты (5), (7) и (9), вычислим

$$\frac{\|r_{k+1}\|}{\|r_k\|} = \frac{\|-\varepsilon_{k+1}((T-\lambda_1)z_{k+1},x_1^{k+1})x_1^{k+1}+\varepsilon_{k+1}(T-\rho_{k+1})z_{k+1}+O(|\varepsilon_{k+1}|^2)\|}{\|-\varepsilon_k((T-\lambda_1)z_k,x_1^k)x_1^k+\varepsilon_k(T-\rho_k)z_k+O(|\varepsilon_k|^2)\|} =$$

$$=\frac{|\varepsilon_{k+1}|\|-(\hat{x},x_1)x_1+\hat{x}\|}{|\varepsilon_k|\|-(\hat{x},x_1)x_1+\hat{x}\|}+O(|\varepsilon_{k-1}|)= q+O(|\varepsilon_{k-1}|). \quad (10)$$

В пределе $k\to\infty$ результаты (8) и (10) завершают доказательство случая 1.

Приступим теперь к доказательству случая 2, т.е. $\|P^*x_1\|=1$. Убедимся, что $x_1=y_1=P^*x_1/\|P^*x_1\|$,

$$\|x_1-y_1\|^2= \|x_1-\frac{P^*x_1}{\|P^*x_1\|}\|^2= \|x_1-P^*x_1\|^2=$$

$$= 1-(x_1,P^*x_1)-(P^*x_1,x_1)+\|P^*x_1\|^2=0.$$

Рассмотрим случай, когда для каждого $x\in X_1$ выполнено $x=P^*x/\|P^*x\|$, т.е. $T^*x=\bar{\lambda}_1x$. Учитывая еще представление вектора v_k в виде (2), преобразуем

$$\rho_k-\lambda_1= (Tv_k,v_k)-\lambda_1(v_k,v_k)= ((T-\lambda_1)v_k,v_k)=$$

$$= ((T-\lambda_1)\beta_k(x_1^k+\varepsilon_kz_k),\beta_k(x_1^k+\varepsilon_kz_k))=$$

$$= |\beta_k|^2((T-\lambda_1)\varepsilon_kz_k,x_1^k+\varepsilon_kz_k)=$$

$$= |\beta_k|^2(\varepsilon_kz_k,(T^*-\bar{\lambda}_1)(x_1^k+\varepsilon_kz_k))=$$

$$= |\beta_k|^2(\varepsilon_kz_k,(T^*-\bar{\lambda}_1)\varepsilon_kz_k)=$$

$$= |\beta_k|^2|\varepsilon_k|^2((T-\lambda_1)z_k,z_k). \quad (11)$$

Вычислим

$$v_{k+1}= \tau_k^2(T-\rho_k)^{-1}(T^*-\bar{\rho}_k)^{-1}v_k=$$

$$= \tau_k^2(T-\rho_k)^{-1}(T^*-\bar{\rho}_k)^{-1}\beta_k(x_1^k+\varepsilon_kz_k)=$$

$$= \tau_k^2\beta_k(T-\rho_k)^{-1}[\frac{x_1^k}{\bar{\lambda}_1-\bar{\rho}_k}+\varepsilon_k(T^*-\bar{\rho}_k)^{-1}z_k]=$$

$$= \tau_k^2\beta_k[\frac{x_1^k}{|\lambda_1-\rho_k|^2}+\varepsilon_k(T-\rho_k)^{-1}(T^*-\bar{\rho}_k)^{-1}z_k]=$$

$$= \frac{\tau_k^2\beta_k}{|\lambda_1-\rho_k|^2}(x_1^k+|\lambda_1-\rho_k|^2\varepsilon_k\|(T-\rho_k)^{-1}(T^*-\bar{\rho}_k)^{-1}z_k\|z_{k+1}), \quad (12)$$

где $z_{k+1}=\dfrac{(T-\rho_k)^{-1}(T^*-\bar{\rho}_k)^{-1}z_k}{\|(T-\rho_k)^{-1}(T^*-\bar{\rho}_k)^{-1}z_k\|}$.

Сравнивая представления (2) и (12), видим, что в данном случае сохраняется направление вектора Pv_k даже в случае, когда собственное значение λ_1 кратное, т.е. $x_1=x_1^k$ для каждого $k=0,1,2,\dots$.

Обозначим через μ минимальное, отличающееся от нуля собственное значение матрицы $(T^*-\bar\lambda_1)(T-\lambda_1)$. Так как матрица $(T^*-\bar\lambda_1)(T-\lambda_1)$ положительно полуопределенная и $\mu\neq0$, то $\mu>0$. Проведем в подпространстве $X_1^\top$ следующее разбиение $X_1^\top=Z\oplus Z^\top$, где $Z=\{z\in X_1^\top|(T^*-\bar\lambda_1)(T-\lambda_1)z=\mu z\}$.

Предположим, что вектор z_k неортогонален подпространству Z. Убедимся, что при $k\to\infty$ имеет место сходимость $z_k\to z$, $z\in Z$.

Рассмотрим вначале случай, когда собственное значение μ однократное. Представим вектор z_k в виде

$$z_k=\gamma_k(\hat z_k+\delta_k t_k), \tag{13}$$

где $\hat z_k\in Z$, $\|\hat z_k\|=1$, $t_k\in Z^\top$, $\|t_k\|=1$ и γ_k гарантирует единичность вектора $\hat z_k$. Найдем вектор

$$[(T^*-\bar\rho_k)(T-\rho_k)]^{-1}z_k=\gamma_k(T-\rho_k)^{-1}(T^*-\bar\rho_k)^{-1}(\hat z_k+\delta_k t_k)=$$

$$=\gamma_k\{\frac{1}{\mu}(T-\rho_k)^{-1}(T^*-\bar\rho_k)^{-1}(T^*-\bar\lambda_1)(T-\lambda_1)\hat z_k+\delta_k(T-\rho_k)^{-1}(T^*-\bar\rho_k)^{-1}t_k\}=$$

$$=\frac{\gamma_k}{\mu}\{(T-\rho_k)^{-1}(T^*-\bar\rho_k)^{-1}(T^*-\bar\lambda_1)(T-\lambda_1)\hat z_k+\mu\delta_k\|[(T^*-\bar\rho_k)(T-\rho_k)]^{-1}t_k\|t_{k+1}\}, \tag{14}$$

где $t_{k+1}=\dfrac{[(T^*-\bar\rho_k)(T-\rho_k)]^{-1}t_k}{\|[(T^*-\bar\rho_k)(T-\rho_k)]^{-1}t_k\|}$.

Из предположения теоремы $\mathrm{dist}(v_k,X_1)\to0$ при $k\to\infty$ следует, что $\rho_k\to\lambda_1$. Так как $t_k\in Z^\top$, то получим $\varlimsup\limits_{k\to\infty}\|[(T^*-\bar\rho_k)(T-\rho_k)]^{-1}t_k\|\leq 1/\mu_1$, где μ_1 собственное значение матрицы $(T^*-\bar\lambda_1)(T-\lambda_1)$ больше чем μ.

Так как $\mu_1>\mu$ и $z_{k+1}=\dfrac{[(T^*-\bar\rho_k)(T-\rho_k)]^{-1}z_k}{\|[(T^*-\bar\rho_k)(T-\rho_k)]^{-1}z_k\|}$ (см. (12)), то в силу соотношений (13) и (14) для достаточно больших индексов k получим, что $\delta_{k+1}\leq\dfrac{\mu}{\mu_1}\delta_k(1+O(|\varepsilon_k|^2))$. Следовательно, $\delta_k\to0$ при $k\to\infty$ и $z_k\to z$, $z\in Z$, причем сходимость будет линейная.

Рассмотрим теперь случай, когда собственное значение μ матрицы $(T^*-\bar\lambda_1)(T-\lambda_1)$ имеет кратность $m>1$. Так как матрица $(T^*-\bar\lambda_1)(T-\lambda_1)$ эрмитова, то подпространство Z определяется только собственными векторами и из них можно выбирать базис ортонормированных векторов. Аналогично случаю $m=1$ можно получить, что компоненты вектора z_k не являющиеся собственными векторами матрицы $(T^*-\bar\lambda_1)(T-\lambda_1)$, соответствующие μ, будут стремиться к нулю при $k\to\infty$. Следовательно, все точки накопления последовательности $\{z_k\}$ имеют вид $z=\sum_{i=1}^{m} a_i\xi_i$, где $\{\xi_1,\xi_2,\dots,\xi_m\}$ является ортонормированным базисом в подпространстве Z. Так как вектор z нормированный, то коэффициенты a_i должны удовлетворять условию $\sum_{i=1}^{m}|a_i|^2=1$. Так как в пределе $k\to\infty$ направление векторов z_k не меняется, то имеет место и сходимость $z_k\to z$ при $k\to\infty$.

Из предположений теоремы вытекает, что $\beta_k\to 1$ при $k\to\infty$ и из (2), (11) и (12) получим

$$\lim_{k\to\infty}\frac{|\varepsilon_{k+1}|}{|\varepsilon_k|^5}=\lim_{k\to\infty}\frac{|\lambda_1-\rho_k|^2|\varepsilon_k|\,\|(T-\rho_k)^{-1}(T^*-\bar\rho_k)^{-1}z_k\|}{|\varepsilon_k|^5}=$$

$$=\lim_{k\to\infty}\frac{|\beta_k|^4|\varepsilon_k|^4|((T-\lambda_1)z_k,z_k)|^2\,\|(T-\rho_k)^{-1}(T^*-\bar\rho_k)^{-1}z_k\|}{|\varepsilon_k|^4}=$$

$$=\lim_{k\to\infty}|((T-\lambda_1)z_k,z_k)|^2\|[(T^*-\bar\lambda_1)(T-\lambda_1)]^{-1}z_k\|=|((T-\lambda_1)z,z)|^2\frac{1}{\mu}. \tag{15}$$

Если $((T-\lambda_1)z,z)\neq 0$, то при помощи (11) и (15) можем оценить

$$\lim_{k\to\infty}\frac{|\lambda_1-\rho_{k+1}|}{|\lambda_1-\rho_k|^5}=\lim_{k\to\infty}\frac{|\beta_{k+1}|^2|\varepsilon_{k+1}|^2|((T-\lambda_1)z_{k+1},z_{k+1})|}{|\beta_k|^{10}|\varepsilon_k|^{10}|((T-\lambda_1)z_k,z_k)|^5}=$$

$$=\lim_{k\to\infty}\frac{|\beta_{k+1}|^2|((T-\lambda_1)z_k,z_k)|^4\|[(T^*-\bar\lambda_1)(T-\lambda_1)]^{-1}z_k\|^2|((T-\lambda_1)z_{k+1},z_{k+1})|}{|\beta_k|^{10}|((T-\lambda_1)z_k,z_k)|^5}=\frac{1}{\mu^2}. \tag{16}$$

Если $((T-\lambda_1)z_k,z_k)=0$, то в силу соотношения (11) справедливо $\rho_k=\lambda_1$. Из итерационного шага (II) следует, что v_{k+1} будет точный собственный вектор матрицы T. Продолжая итерационный процесс, получим, что $[\rho_m,v_m]$ будет точная собственная пара для всех $m\geq k+1$.

Если $((T-\lambda_1)z_k,z_k)\to 0$, то аналогично пределу (16) получим

$$\lim_{k\to\infty}\frac{|\lambda_1-\rho_{k+1}|}{|\lambda_1-\rho_k|^5}\leq\frac{1}{\mu^2}. \tag{16'}$$

Представляя вектор v_k в виде (2), и, учитывая $z_k\perp X_1$, найдем норму вектора невязки

$$\|r_k\|^2=\|(T-\rho_k)v_k\|^2=\|(T-\rho_k)\beta_k(x_1+\varepsilon_k z_k)\|^2=$$

$$=|\beta_k|^2\|(T-\rho_k)x_1+\varepsilon_k(T-\rho_k)z_k\|^2=$$

$$=|\beta_k|^2|\lambda_1-\rho_k|^2+|\beta_k|^2|\varepsilon_k|^2\|(T-\rho_k)z_k\|^2. \tag{17}$$

Из соотношений (11), (15) и (17) следует

$$\lim_{k\to\infty}\frac{\|r_{k+1}\|}{\|r_k\|^5}=\lim_{k\to\infty}\left\{\frac{|\beta_{k+1}|^2|\lambda_1-\rho_{k+1}|^2+|\beta_{k+1}|^2|\varepsilon_{k+1}|^2\|(T-\rho_{k+1})z_{k+1}\|^2}{(|\beta_k|^2|\lambda_1-\rho_k|^2+|\beta_k|^2|\varepsilon_k|^2\|(T-\rho_k)z_k\|^2)^5}\right\}^{1/2}=$$

$$=\lim_{k\to\infty}\frac{\|(T-\rho_{k+1})z_{k+1}\|}{\|(T-\rho_k)z_k\|^5}\frac{|\varepsilon_{k+1}|}{|\varepsilon_k|^5}=\lim_{k\to\infty}\frac{|\varepsilon_{k+1}|\|(T-\lambda_1)z\|}{|\varepsilon_k|^5\|(T-\lambda_1)z\|^5}=$$

$$=|((T-\lambda_1)z,z)|^2\frac{1}{\mu}\left\{\frac{|((T^*-\bar{\lambda}_1)(T-\lambda_1)z,z)|}{|((T^*-\bar{\lambda}_1)(T-\lambda_1)z,z)|^5}\right\}^{1/2}=$$

$$=|((T-\lambda_1)z,z)|^2\frac{1}{\mu}\left\{\frac{|(\mu z,z)|}{|(\mu z,z)|^5}\right\}^{1/2}=|((T-\lambda_1)z,z)|^2\frac{1}{\mu}\left\{\frac{\mu}{\mu^5}\right\}^{1/2}=$$

$$=|((T-\lambda_1)z,z)|^2\frac{1}{\mu^3}=\frac{1}{\mu^3}|\rho(z)-\lambda_1|. \qquad (18)$$

Мы увидели, что из предположения неортогональности z_k подпространству Z следуют утверждения теоремы. Если $z_k \perp Z$, то видоизменения, которые нужно привести к доказательству, довольно очевидны. Так как z_k будет сходиться к собственному вектору матрицы $(T^*-\bar{\lambda}_1)(T-\lambda_1)$, соответствующее собственному значению большему μ, то оценки улучшаются. Нетрудно увидеть также, что если z_0 (тогда и v_0) неортогонален подпространству Z, то и z_k в итерационном процессе будет неортогональным подпространству Z.

Мы завершили доказательство для случая 2, если для каждого $x_1 \in X_1$ следует $T^*x_1=\bar{\lambda}_1 x_1$. Но если найдется вектор $x \in X_1$, для которого $T^*x \neq \bar{\lambda}_1 x$, то вычисляя вектор v_{k+1}, нетрудно увидеть, что в итерационном процессе будут расти компоненты собственных векторов $x \in X_1$, для которых $\|P^*x\|>1$. Следовательно, наш итерационный процесс не может сходиться к собственному вектору x_1, для которого $\|P^*x_1\|=1$. Но если такие компоненты отсутствуют, то будет верным предыдущее доказательство. В итоге мы завершили доказательство всей теоремы. □

Рассмотрим теперь локальную сходимость итерационного метода с регуляризирующим членом $(\alpha_k \neq 0)$.

Теорема 2. Пусть подпространство X_1 состоит только из собственных векторов. Если в итерационном процессе (I)-(III) $\mathrm{dist}(v_k, X_1)\to 0$ при $k\to\infty$, то выбор последовательности $\{\alpha_k\}$ такой, что $\lim\limits_{k\to\infty}\frac{\alpha_k}{|\lambda_1-\rho_k|^2}=0$, гарантирует сходимость $v_k \to x_1$, $x_1 \in X_1$.

1. Если $\|P^*x_1\|>1$, то

$$\lim_{k\to\infty}\frac{|\lambda_1-\rho_{k+1}|}{|\lambda_1-\rho_k|}=\lim_{k\to\infty}\frac{\|r_{k+1}\|}{\|r_k\|}=q,\quad \text{где } q=1-1/\|P^*x_1\|^2.$$

2. Если $\|P^*x_1\|=1$, то

$$\lim_{k\to\infty}\frac{|\lambda_1-\rho_{k+1}|}{|\lambda_1-\rho_k|^5}=\text{const},\qquad \lim_{k\to\infty}\frac{\|r_{k+1}\|}{\|r_k\|^5}=\text{const}.$$

<u>Доказательство</u>. Случай 1. Обозначим через $A_k=(T^*-\bar\rho_k)(T-\rho_k)$. Из доказательства теоремы 1 следует существование A_k^{-1} и $\|A_k^{-1}\|\le c/|\lambda_1-\rho_k|^2$, где c некоторая константа. Теперь из предположения $\lim\limits_{k\to\infty}\frac{\alpha_k}{|\lambda_1-\rho_k|^2}=0$ следует, что

$$\lim_{k\to\infty}\|\alpha_kA_k^{-1}\|\le\lim_{k\to\infty}\frac{\alpha_kc}{|\lambda_1-\rho_k|^2}=0.$$

В силу последнего результата при $k\to\infty$ справедливо

$$(\alpha_k+A_k)^{-1}v_k=[A_k(\alpha_kA_k^{-1}+I)]^{-1}v_k=\sum_{j=0}^{\infty}(-\alpha_kA_k^{-1})^jA_k^{-1}v_k\to A_k^{-1}v_k,\ \text{т.е.}$$

$$[\alpha_k+(T^*-\bar\rho_k)(T-\rho_k)]^{-1}v_k\to[(T^*-\bar\rho_k)(T-\rho_k)]^{-1}v_k.$$

Опираясь теперь на доказательство теоремы 1, получим, что утверждение случая 1 справедливо.

Случай 2. Преобразуем итерационный шаг (II'),

$$v_{k+1}=\tau_k^2[\alpha_k+(T^*-\bar\rho_k)(T-\rho_k)]^{-1}\beta_k(x_1+\varepsilon_kz_k)=$$

$$=\frac{\tau_k^2\beta_k}{[\alpha_k+|\lambda_1-\rho_k|^2]}\left\{x_1+\varepsilon_k[\alpha_k+|\lambda_1-\rho_k|^2][\alpha_k+(T^*-\bar\rho_k)(T-\rho_k)]^{-1}z_k\right\}=$$

$$=\frac{\tau_k^2\beta_k}{[\alpha_k+|\lambda_1-\rho_k|^2]}\left\{x_1+\varepsilon_k[\alpha_k+|\lambda_1-\rho_k|^2]\|[\alpha_k+(T^*-\bar\rho_k)(T-\rho_k)]^{-1}z_k\|z_{k+1}\right\},\tag{19}$$

где $z_{k+1}=\dfrac{[\alpha_k+(T^*-\bar\rho_k)(T-\rho_k)]^{-1}z_k}{\|[\alpha_k+(T^*-\bar\rho_k)(T-\rho_k)]^{-1}z_k\|}$.

Представление (19) аналог представления (12). Так как $\lim\limits_{k\to\infty}\frac{\alpha_k}{|\lambda_1-\rho_k|^2}=0$ и $\|[\alpha_k+(T^*-\bar\rho_k)(T-\rho_k)]^{-1}z_k\|\le M<\infty$, то рассуждения теоремы 1 можно перенести и на этот случай. В итоге получим, что сохраняется сходимость пятого порядка. □

В предположениях теоремы 2 фигурирует требование $\lim\limits_{k\to\infty}\frac{\alpha_k}{|\lambda_1-\rho_k|^2}=0$. Так как величина $|\lambda_1-\rho_k|^2$ нам не известна, то возникает вопрос, как выбирать последовательность $\{\alpha_k\}$ в реальном итерационном процессе.

Проанализируем вначале, что означает требование $\lim\limits_{k\to\infty}\frac{\alpha_k}{|\lambda_1-\rho_k|^2}=0$ для случая 1 теоремы 2. Учитывая результаты из доказательства теоремы 1, вычислим

$$\lim_{k\to\infty}\frac{|\lambda_1-\rho_k|^2}{\|r_k\|^2}=\lim_{k\to\infty}\frac{|\varepsilon_k((T-\lambda_1)z_k,x_1^k)+O(|\varepsilon_k|^2)|^2}{\|-\varepsilon_k((T-\lambda_1)z_k,x_1^k)x_1^k+\varepsilon_k(T-\rho_k)z_k+O(|\varepsilon_k|^2)\|^2}=$$

$$=\frac{|((T-\lambda_1)z,x_1)|^2}{\|-((T-\lambda_1)z,x_1)x_1+(T-\lambda_1)z\|^2}=\frac{|(\hat{x},x_1)|^2}{\|-(\hat{x},x_1)x_1+\hat{x}\|^2}=$$

$$=\frac{(1-|(x_1,y_1)|^2)^2}{(1-|(x_1,y_1)|^2)|(x_1,y_1)|^2}=\frac{q}{1-q}. \qquad (20)$$

Из равенства (20) следует, что предварительные условия $\lim\limits_{k\to\infty}\frac{\alpha_k}{|\lambda_1-\rho_k|^2}=0$ и $\lim\limits_{k\to\infty}\frac{\alpha_k}{\|r_k\|^2}=0$ равносильны. Так как величину $\|r_k\|^2$ можно вычислять, то в реальном итерационном процессе последовательность $\{\alpha_k\}$ можно выбирать так, что $\lim\limits_{k\to\infty}\frac{\alpha_k}{\|r_k\|^2}=0$.

Для случая 2 требование $\lim\limits_{k\to\infty}\frac{\alpha_k}{|\lambda_1-\rho_k|^2}=0$, как правило, равносильно требованию $\lim\limits_{k\to\infty}\frac{\alpha_k}{\|r_k\|^4}=0$ (из (11) и (17) следует, что $|\lambda_1-\rho_k|^2\asymp|\varepsilon_k|^4$ и $\|r_k\|^4\asymp|\varepsilon_k|^4$).

Литература

1. Воеводин В. В. Линейная алгебра. – М.: Наука, 1980.
2. Парлетт Б. Симметричная проблема собственных значений. – М.: Мир, 1983., 384 с.
3. Саан Т. Регуляризованный итерационный метод для отыскания собственных значений и собственных векторов симметричной матрицы. // Уч. зап. Тарт. ун-та., 1987., №762., с. 59-68.
4. Саан Т. Регуляризованный итерационный метод для отыскания собственных значений и собственных векторов неэрмитовой матрицы. // Уч. зап. Тарт. ун-та., 1988. , №833. , с.84-90.

5. Уилкинсон Дж. Х. Алгебраическая проблема собственных значений. – М.: Наука, 1970., 564 с.
6. Ostrowski A. M. On the Convergence of the Rayleigh Quotient Iteration for the Computation of Characteristic Roots and Vectors I-VI. // Arch. Rational Mech. Anal., 1958/59., 1-4., 1: pp. 233-241, 2: pp. 423-428, 3: pp. 325-340, 3: pp. 341-347, 3: pp. 472-481, 4: pp. 153-165.
7. Parlett B. N. The Rayleigh Quotient Iteration and Some Generalizations for Nonnormal Matrices. // Mathematics of Computation., 1974., 28., pp. 679-693.

THE REGULARIZED ITERATIVE METHOD WITH THE RAYLEIGH QUOTIENT

T.Saan

Summary

The Rayleigh Quotient Iteration (RQI) suits well for finding the eigenvalues and eigenvectors of normal matrices. Its excellent global property is existent due to the monotonic decrease in the norms of the residuals. This property fails for non-normal matrices. The method (I)-(III) has minimal residual property for any matrices (see [2]). In this work the asymptotic covergence rate of the method (I)-(III) is investigated. In the normal case the asymptotic convergence rate for this method is quintic and in the non-defective case linear.

РЕГУЛЯРИЗОВАННЫЙ ИТЕРАЦИОННЫЙ МЕТОД С ОБОБЩЕННЫМ ОТНОШЕНИЕМ РЕЛЕЯ

Таймо Саан

В данной работе исследован метод с обобщенным отношением Релея. В случае нормальной матрицы этот метод имеет пятый порядок сходимости. При матрице простой структуры для этого метода гарантирована квадратичная сходимость. Подробно исследуется сходимость метода, когда матрица не имеет простую структуру. Оказывается, если собственный вектор, к которому ведет итерационный процесс, принадлежит подпространству, имеющему корневые векторы выше чем первого порядка, то исследованный метод имеет линейную сходимость.

В работах [1], [2] и [3] исследован один регуляризованный итерационный метод с отношением Релея для отыскания собственного значения и соответствующего ему собственного вектора для матрицы. В случае нормальной матрицы этот метод имеет пятый порядок сходимости (см. [1], [3]). Но в случае ненормальной матрицы порядок сходимости, как правило, не лучше линейного.

1. Скорость сходимости для матрицы простой структуры.

Пусть дана комплексная $(n\times n)$-матрица T. Через $\sigma(T)=\{\lambda_1,\lambda_2,\dots,\lambda_r\}$, $\{X_1,X_2,\dots,X_r\}$ и $\{Y_1,Y_2,\dots,Y_r\}$, $r\le n$, обозначим множество попарно различных собственных значений матрицы T, соответствующие им корневые подпространства и корневые подпространства матрицы T^*, соответствующие значениям $\bar\lambda_1,\bar\lambda_2,\dots\bar\lambda_r$. Введем и обозначения $X_1{}^\top=X_2+X_3+\dots+X_r$ и $Y_1{}^\top=Y_2+Y_3+\dots+Y_r$.

<u>Определение</u>. Обобщенным отношением Релея для матрицы T называется функция $\rho(v,u)=(Tv,u)/(v,u)$, где $v,u\in\mathbb{C}^n$ и $(v,u)\ne0$.

Для отыскания собственного значения λ матрицы T и соответствующих ему правого и левого собственного вектора используем следующий итерационный метод. Выберем единичные векторы v_0 и u_0, $(v_0,u_0)\ne0$. Для $k=0,1,2,\dots$ повторим следующие шаги:

(I) вычислим $\rho_k=\rho(v_k,u_k)=\dfrac{(Tv_k,u_k)}{(v_k,u_k)}$;

(II) решим уравнения

$[\alpha_k+(T^*-\bar\rho_k)(T-\rho_k)]\hat v_{k+1}=v_k$,

$[\alpha_k+(T-\rho_k)(T^*-\bar\rho_k)]\hat u_{k+1}=u_k$,

где $\alpha_k>0$ и $\lim\limits_{k\to\infty}\alpha_k=0$;

(III) формируем $v_{k+1}=\hat v_{k+1}/\|\hat v_{k+1}\|$ и $u_{k+1}=\hat u_{k+1}/\|\hat u_{k+1}\|$.

Если $(v_{k+1},u_{k+1})=0$, то выбираем другие начальные векторы v_0 и u_0 и начинаем итерационный процесс сначала.

Шаги (II) и (III) можно объединить в один,

$$v_{k+1}=\tau_k^2[\alpha_k+(T^*-\bar\rho_k)(T-\rho_k)]^{-1}v_k,\ \text{где}\ \tau_k^2=\|[\alpha_k+(T^*-\bar\rho_k)(T-\rho_k)]^{-1}v_k\|^{-1},$$

$$u_{k+1}=\varkappa_k^2[\alpha_k+(T-\rho_k)(T^*-\bar\rho_k)]^{-1}u_k,\ \text{где}\ \varkappa_k^2=\|[\alpha_k+(T-\rho_k)(T^*-\bar\rho_k)]^{-1}u_k\|^{-1}.$$

Замечание. В теоремах данной работы имеется предположение: пусть последовательности $\{v_k\}$ и $\{u_k\}$, построенные по методу (I)-(III), такие, что $v_k\to x_1$ и $u_k\to y_1$ при $k\to\infty$, где $Tx_1=\lambda_1x_1$ и $T^*y_1=\bar\lambda_1y_1$. Из доказательств соответствующих теорем вытекает, что это предположение выполнено, например, если начальные приближения v_0 и u_0 довольно близкие к собственным векторам и ρ_0 близкое к собственному значению λ_1.

Введем векторы невязки $r_k=(T-\rho_k)v_k$ и $p_k=(T^*-\bar\rho_k)u_k$.

Теорема 1.1. Пусть последовательности $\{v_k\}$ и $\{u_k\}$, построенные по методу (I)-(III) такие, что $v_k\to x_1$ и $u_k\to y_1$ при $k\to\infty$, где $Tx_1=\lambda_1x_1$, $T^*y_1=\bar\lambda_1y_1$, $\|x_1\|=\|y_1\|=1$ и $(x_1,y_1)\neq0$. Если $\lim\limits_{k\to\infty}\dfrac{\alpha_k}{|\lambda_1-\rho_k|^2}=0$, то выполняются

$$\lim_{k\to\infty}\frac{|\lambda_1-\rho_{k+1}|}{|\lambda_1-\rho_k|^2}=\text{const},\quad \lim_{k\to\infty}\frac{\|r_{k+1}\|}{\|r_k\|\|p_k\|}=\text{const},\quad \lim_{k\to\infty}\frac{\|p_{k+1}\|}{\|r_k\|\|p_k\|}=\text{const}.$$

В случае $T^*x_1=\bar\lambda_1x_1$ и $Ty_1=\lambda_1y_1$ выполняются

$$\lim_{k\to\infty}\frac{|\lambda_1-\rho_{k+1}|}{|\lambda_1-\rho_k|^5}=\text{const},\quad \lim_{k\to\infty}\frac{\|r_{k+1}\|}{\|r_k\|^3\|p_k\|^2}=\text{const},\quad \lim_{k\to\infty}\frac{\|p_{k+1}\|}{\|r_k\|^2\|p_k\|^3}=\text{const}.$$

Доказательство. Обозначим через P проектор Рисса в $\mathbb{C}^n$, проектирующий на X_1. Тогда проектор P^* проектирует на Y_1. Представим приближения v_k и u_k в виде (см. [3, представление (2)])

$$v_k=\beta_k(x_1^k+\varepsilon_k z_k), \quad u_k=\gamma_k(y_1^k+\eta_k w_k), \tag{1.1}$$

где $x_1^k\in X_1$, $y_1^k\in Y_1$, $z_k\in X_1^{\top}$, $w_k\in Y_1^{\top}$ и коэффициенты β_k и γ_k гарантируют, что векторы v_k и u_k будут нормированы. Отметим, что в предположениях теоремы имеют место сходимости $\varepsilon_k\to 0$ и $\eta_k\to 0$ при $k\to\infty$. Преобразуем

$$\rho_k-\lambda_1=\frac{(T\beta_k(x_1^k+\varepsilon_k z_k),\gamma_k(y_1^k+\eta_k w_k))}{(\beta_k(x_1^k+\varepsilon_k z_k),\gamma_k(y_1^k+\eta_k w_k))}-\lambda_1=$$

$$=\frac{\lambda_1(x_1^k,y_1^k)+\lambda_1(x_1^k,\eta_k w_k)+\lambda_1(\varepsilon_k z_k,y_1^k)+\varepsilon_k\bar{\eta}_k(Tz_k,w_k)}{(x_1^k+\varepsilon_k z_k,y_1^k+\eta_k w_k)}-\lambda_1=$$

$$=\frac{\varepsilon_k\bar{\eta}_k((T-\lambda_1)z_k,w_k)}{(x_1^k,y_1^k)+\varepsilon_k\bar{\eta}_k(z_k,w_k)}. \tag{1.2}$$

Приступим к случаю, когда равенства $T^*x_1=\bar{\lambda}_1 x_1$, $Ty_1=\lambda_1 y_1$ не выполнены. Рассмотрим сначала вариант, когда $\alpha_k=0$ для $k=0,1,2,\ldots$. Найдем вектор v_{k+1}. Для этого используем обозначения и результаты работы [3]. Вектор v_{k+1} можно представить в виде (см. [3, формулы (2) и (6)])

$$v_{k+1}=\beta_{k+1}\left\{x_1^{k+1}+\frac{(\lambda_1-\rho_k)\|(T-\rho_k)^{-1}(\hat{x}_k+(I-P)t_k)\|}{b_k\|x_1^k+\frac{1}{b_k}Pt_k\|}z_{k+1}\right\},$$

где $b_k=1/(x_1^k,y_1^k)$, $\hat{x}_k=y_1^k-b_k x_1^k$ и $t_k=(\varepsilon_k P^* z_k+(\bar{\lambda}_1-\bar{\rho}_k)(T^*-\bar{\rho}_k)^{-1}(\hat{y}_k+\varepsilon_k(I-P^*)z_k))/b_k$, $\hat{y}_k=x_1^k-a_k y_1^k$ и $a_k=b_k$. Учитывая соотношение (1.2), получим

$$v_{k+1}=\beta_{k+1}\left\{x_1^{k+1}+\frac{\varepsilon_k\bar{\eta}_k((T-\lambda_1)z_k,w_k)}{(x_1^k,y_1^k)+\varepsilon_k\bar{\eta}_k(z_k,w_k)}\,\frac{\|(T-\rho_k)^{-1}(\hat{x}_k+(I-P)t_k)\|}{b_k\|x_1^k+\frac{1}{b_k}Pt_k\|}z_{k+1}\right\}. \tag{1.3}$$

В силу соотношений (1.1) и (1.3) получим

$$\overline{\lim_{k\to\infty}}\frac{|\varepsilon_{k+1}|}{|\varepsilon_k||\eta_k|}=\overline{\lim_{k\to\infty}}\frac{|\varepsilon_k\bar{\eta}_k((T-\lambda_1)z_k,w_k)|}{|(x_1^k,y_1^k)+\varepsilon_k\bar{\eta}_k(z_k,w_k)|}\,\frac{\|(T-\rho_k)^{-1}(\hat{x}_k+(I-P)t_k)\|}{b_k\|x_1^k+\frac{1}{b_k}Pt_k\||\varepsilon_k||\eta_k|}=$$

$$=\overline{\lim_{k\to\infty}}|((T-\lambda_1)z_k,w_k)|\|(T-\rho_k)^{-1}\hat{x}_k\|=\mathrm{const}. \tag{1.4}$$

Проводя аналогичные рассуждения и для приближения u_{k+1}, получим

$$\overline{\lim_{k\to\infty}}\frac{|\eta_{k+1}|}{|\varepsilon_k||\eta_k|}=\mathrm{const}. \tag{1.5}$$

Отметим, что как для метода с отношением Релея, так и для метода данной работы при $k\to\infty$ имеет место сходимость

$$z_k \to z \equiv \frac{(T-\lambda_1)^{-1}\hat{x}}{\|(T-\lambda_1)^{-1}\hat{x}\|},$$

где $\hat{x}=y_1-(1/(x_1,y_1))x_1$. Аналогично

$$w_k \to w \equiv \frac{(T^*-\bar{\lambda}_1)^{-1}\hat{y}}{\|(T^*-\bar{\lambda}_1)^{-1}\hat{y}\|},$$

где $\hat{y}=x_1-(1/(x_1,y_1))y_1$.

Следовательно, существуют не только верхние пределы (1.4) и (1.5), но и соответствующие пределы. Опираясь дополнительно на эти пределы и на соотношение (1.2), получим

$$\lim_{k\to\infty}\frac{|\lambda_1-\rho_{k+1}|}{|\lambda_1-\rho_k|^2}=\text{const}. \tag{1.6}$$

Вычислим нормы векторов невязок r_k и p_k,

$$\|r_k\|=\|(T-\rho_k)v_k\|=\|(T-\rho_k)\beta_k(x_1^k+\varepsilon_k z_k)\|=$$

$$=|\beta_k|\|(\lambda_1-\rho_k)x_1^k+\varepsilon_k(T-\rho_k)z_k\|=$$

$$=|\beta_k|\|O(|\varepsilon_k||\eta_k|)x_1^k+\varepsilon_k(T-\rho_k)z_k\|=$$

$$=|\beta_k||\varepsilon_k|\|(T-\rho_k)z_k\|+O(|\varepsilon_k||\eta_k|),$$

$$\|p_k\|=|\gamma_k||\eta_k|\|(T^*-\bar{\rho}_k)w_k\|+O(|\varepsilon_k||\eta_k|). \tag{1.7}$$

Из предела (1.4) и норм (1.7) следует

$$\lim_{k\to\infty}\frac{\|r_{k+1}\|}{\|r_k\|\|p_k\|}=\lim_{k\to\infty}\frac{|\beta_{k+1}||\varepsilon_{k+1}|\|(T-\rho_{k+1})z_{k+1}\|}{|\beta_k||\varepsilon_k|\|(T-\rho_k)z_k\||\gamma_k||\eta_k|\|(T^*-\bar{\rho}_k)w_k\|}=$$

$$=\lim_{k\to\infty}\frac{|\varepsilon_{k+1}|}{|\varepsilon_k||\eta_k|}\,\frac{\|(T-\lambda_1)z\|}{\|(T-\lambda_1)z\|\|(T^*-\bar{\lambda}_1)w\|}=\text{const}. \tag{1.8}$$

Повторим аналогичные рассуждения и для нахождения предела

$$\lim_{k\to\infty}\frac{\|p_{k+1}\|}{\|r_k\|\|p_k\|}=\text{const}. \tag{1.9}$$

Пределы (1.6), (1.8) и (1.9) завершают доказательство, когда $\alpha_k=0$ для $k\geq 0$. Легко убедиться, что скорость сходимости сохраняется, когда последовательность $\{\alpha_k\}$ такая, что $\lim\limits_{k\to\infty}\frac{\alpha_k}{|\lambda_1-\rho_k|^2}=0$ (см. [3]).

Для основного случая утверждения теоремы доказаны.

Рассмотрим случай, когда $T^*x_1=\bar{\lambda}_1 x_1$ и $Ty_1=\lambda_1 y_1$. Исходя из вектора $v_k=\beta_k(x_1^k+\varepsilon_k z_k)$ вычислим (см. [3, соотношение (19)])

$$v_{k+1}=\beta_{k+1}\{x_1^k+\varepsilon_k[\alpha_k+|\lambda_1-\rho_k|^2]\|[\alpha_k+(T^*-\bar{\rho}_k)(T-\rho_k)]^{-1}z_k\|z_{k+1}\}. \tag{1.10}$$

Выпишем аналогичное представление для вектора u_{k+1},

$$u_{k+1}=\gamma_{k+1}\{y_1^k+\eta_k[\alpha_k+|\lambda_1-\rho_k|^2]\|[\alpha_k+(T-\rho_k)(T^*-\bar{\rho}_k)]^{-1}w_k\|w_{k+1}\}. \tag{1.11}$$

Опираясь на соотношения (1.2),(1.10),(1.11) и на выбор последовательности (α_k), можем убедиться, что утверждения теоремы в этом случае справедливы. Рассуждения проводятся аналогично доказательству теоремы 2 из [3].

Если выполнено одно и только одно равенство из $T^*x_1=\bar{\lambda}_1 x_1$, $Ty_1=\lambda_1 y_1$, то $(x_1,y_1)=0$. А последнее противоречит предположениям теоремы. В итоге мы закончили доказательство теоремы. □

2. Скорость сходимости при наличии корневых векторов

Исследуем случай, когда корневое подпространство X_1 содержит корневую цепочку длиной больше единицы. Исследуем сперва вариант, когда в подпространстве X_1 только один собственный вектор (в точности до коэффициента) и ему соответствует корневая цепочка длиной m. Эти корневые векторы обозначим через $x_{11},x_{12},\ldots,x_{1m}$, т.е.

$$\begin{aligned}
&(T-\lambda_1)x_{11}=0 \qquad (\|x_{11}\|=1),\\
&(T-\lambda_1)x_{12}=x_{11},\\
&\ldots\ldots\ldots\ldots\ldots\ldots\\
&(T-\lambda_1)x_{1m}=x_{1,m-1}.
\end{aligned}$$

Отметим, что в корневом подпространстве Y_1, т.е. корневое подпространство матрицы T^*, соответсвующее собственному значению $\bar{\lambda}_1$, найдется базис из корневых векторов $y_{11},y_{12},\ldots,y_{1m}$ такой, что

1) $$\begin{aligned}
&(T^*-\bar{\lambda}_1)y_{11}=0 \qquad (\|y_{11}\|=1),\\
&(T^*-\bar{\lambda}_1)y_{12}=y_{11},\\
&\ldots\ldots\ldots\ldots\ldots\ldots\\
&(T^*-\bar{\lambda}_1)y_{1m}=y_{1,m-1};
\end{aligned}$$

2) $(x_{1i},y_{1j})=0$ при $i+j\neq m+1$

$(x_{1i},y_{1j})\neq 0$ при $i+j=m+1$.

Теорема 2.1. Пусть в итерационном процессе $\alpha_k=0$, $k=0,1,2,\dots$. Если $v_k\to x_{11}$ и $u_k\to y_{11}$ при $k\to\infty$, то

$$\lim_{k\to\infty}\frac{\rho_{k+1}-\lambda_1}{\rho_k-\lambda_1}=\lim_{k\to\infty}\frac{\|(T-\rho_{k+1})v_{k+1}\|}{\|(T-\rho_k)v_k\|}=\lim_{k\to\infty}\frac{\|(T^*-\bar\rho_{k+1})u_{k+1}\|}{\|(T^*-\bar\rho_k)u_k\|}=\frac{m-1}{m},$$

где m длина корневой цепочки.

<u>Доказательство</u>. Представим приближения v_k и u_k в виде

$$v_k=\sum_{j=1}^{m}c_{jk}x_{1j}+\varepsilon_k z_k,\qquad u_k=\sum_{j=1}^{m}\bar d_{jk}y_{1j}+\delta_k w_k,\tag{2.1}$$

где $z_k\in X_1^{\,\tau}$, $w_k\in Y_1^{\,\tau}$. Из предположений теоремы следует, что $c_{2k},\dots,c_{mk},\bar d_{2k},\dots,\bar d_{mk},\varepsilon_k,\delta_k\to0$ при $k\to\infty$. Используем множества $\sigma(T)=\{\lambda_1,\dots,\lambda_r\}$, $\{X_1,\dots,X_r\}$ и $\{Y_1,\dots,Y_r\}$, где X_i, $i=1,2,\dots,r$ корневое подпространство матрицы T, соответствующее собственному значению λ_i, а Y_i, $i=1,2,\dots,r$ корневое подпространство матрицы T^*, соответствующее собственному значению $\bar\lambda_i$. Обозначим алгебраические кратности собственных значений λ_i через m_i, тогда

$$\sum_{i=1}^{r}m_i=n\quad(m_1=m\ \text{-кратность собственного значения}\ \lambda_1).$$

Выделим в корневых подпространствах X_i и Y_i, $i=2,\dots,r$ аналогичные базисы как и в подпространствах X_1 и Y_1. Обозначим базисные векторы через x_{ip} и y_{ip}, $p\le m_i$. Представим векторы x_{1j} и y_{1j} следующим образом:

$$x_{1j}=\sum_{i=1}^{r}\sum_{p=1}^{m_i}a_{jip}y_{ip},\qquad y_{1j}=\sum_{i=1}^{r}\sum_{p=1}^{m_i}b_{jip}x_{ip}.\tag{2.2}$$

Найдем

$$t_{k+1}=(T^*-\bar\rho_k)^{-1}v_k=(T^*-\bar\rho_k)^{-1}\sum_{j=1}^{m}c_{jk}x_{1j}+(T^*-\bar\rho_k)^{-1}\varepsilon_k z_k=$$

$$=(T^*-\bar\rho_k)^{-1}\sum_{j=1}^{m}c_{jk}\sum_{i=1}^{r}\sum_{p=1}^{m_i}a_{jip}y_{ip}+\varepsilon_k(T^*-\bar\rho_k)^{-1}z_k.$$

Выделим из вектора t_{k+1} компонент, который принадлежит к подпространству Y_1,

$$t_{k+1}=(T^*-\bar\rho_k)^{-1}\sum_{j=1}^{m}c_{jk}\sum_{p=1}^{m}a_{j1p}y_{1p}+(T^*-\bar\rho_k)^{-1}\sum_{j=1}^{m}c_{jk}\sum_{i=2}^{r}\sum_{p=1}^{m_i}a_{jip}y_{ip}+\varepsilon_k(T^*-\bar\rho_k)^{-1}z_k=$$

$$=(T^*-\bar{\rho}_k)^{-1}\sum_{j=1}^{m} c_{jk}\sum_{p=1}^{m} a_{j1p}\, y_{1p} + (T^*-\bar{\rho}_k)^{-1}\hat{z}_k, \tag{2.3}$$

где $\hat{z}_k \in Y_1^{\top}$.

Найдем из первого члена формулы (2.3) компонент $j=1$,

$$\Sigma_1 \equiv (T^*-\bar{\rho}_k)^{-1} c_{1k} \sum_{p=1}^{m} a_{11p}\, y_{1p}. \tag{2.4}$$

Для этого обозначим

$$(T^*-\bar{\rho}_k)^{-1}\sum_{p=1}^{m} a_{11p}\, y_{1p} = s. \tag{2.5}$$

Отождествим $a_p \equiv a_{11p}$, $p \le m$. Так как вектор $s \in Y_1$, то можем его представить в виде

$$s=\sum_{j=1}^{m} s_j y_{1j}.$$

Умножим формулу (2.5) с обеих сторон на матрицу $(T^*-\bar{\rho}_k)$ и преобразуем

$$\sum_{p=1}^{m} a_p\, y_{1p} = (T^*-\bar{\rho}_k)s = (T^*-\bar{\rho}_k)\sum_{j=1}^{m} s_j y_{1j} =$$

$$=\sum_{j=1}^{m} s_j (T^*-\bar{\rho}_k) y_{1j} = s_1(\bar{\lambda}_1-\bar{\rho}_k)y_{11} + \sum_{j=2}^{m} s_j (T^*-\bar{\rho}_k) y_{1j} =$$

$$= s_1(\bar{\lambda}_1-\bar{\rho}_k)y_{11} + \sum_{j=2}^{m} (T^*-\bar{\lambda}_1+\bar{\lambda}_1-\bar{\rho}_k) y_{1j} =$$

$$= s_1(\bar{\lambda}_1-\bar{\rho}_k)y_{11} + \sum_{j=2}^{m} s_j y_{1,j-1} + \sum_{j=2}^{m} s_j(\bar{\lambda}_1-\bar{\rho}_k)y_{1j}$$

$$=\sum_{j=1}^{m} s_{j+1} y_{1j} + (\bar{\lambda}_1-\bar{\rho}_k)\sum_{j=1}^{m} s_j y_{1j}. \tag{2.6}$$

Умножим равенство (2.6) справа на x_{1q}, $q=1,2,\dots,m$,

$$\sum_{p=1}^{m} a_p (y_{1p}, x_{1q}) = \sum_{j=1}^{m-1} s_{j+1}(y_{1j}, x_{1q}) + (\bar{\lambda}_1-\bar{\rho}_k)\sum_{j=1}^{m} s_j (y_{1p}, x_{1q}), \quad q=1,2,\dots,m.$$

Учитывая, что $(y_{1p}, x_{1q})=0$ при $p+q\ne m+1$ и $(y_{1p}, x_{1q})\ne 0$ при $p+q=m+1$, последнее равенство приобретает более простой вид.

$$\begin{cases} a_p = s_{p+1} + (\bar\lambda_1 - \bar\rho_k) s_p, & p=1,2,\dots,m-1, \\ a_m = (\bar\lambda_1 - \bar\rho_k) s_m \end{cases} \tag{2.7}$$

Вычислим из системы (2.7) по-очереди коэффициенты $s_m,\dots,s_1$, получим

$$s_{m-q} = -\sum_{l=1}^{q+1} a_{m-q-l+1} (\bar\rho_k - \bar\lambda_1)^{-l}, \quad q=0,1,\dots,m-1, \text{ т.е.}$$

$$s_q = -\sum_{l=1}^{m-q+1} a_{q-l+1} (\bar\rho_k - \bar\lambda_1)^{-l}, \quad q=1,2,\dots,m. \tag{2.8}$$

Вычислим коэффициент a_m. Исходим из вектора (см. (2.2))

$$x_{11} = \sum_{l=1}^{r} \sum_{p=1}^{m_l} a_{1lp} y_{lp}.$$

Умножим это представление на вектор x_{11}, получим

$$(x_{11}, x_{11}) = \sum_{l=1}^{r} \sum_{p=1}^{m_l} a_{1lp} (y_{lp}, x_{11}) = \sum_{p=1}^{m} a_{11p} (y_{1p}, x_{11}) = a_{11m} (y_{1m}, x_{11}).$$

Так как вектор x_{11} единичный, то

$$a_m = a_{11m} = 1/(y_{1m}, x_{11}). \tag{2.9}$$

Из сходимостей $v_k \to x_{11}$ и $u_k \to y_{11}$ при $k \to \infty$ вытекает, что $\rho_k \to \lambda_1$ и, следовательно,

$$s_q = -a_m (\bar\rho_k - \bar\lambda_1)^{-m+q-1} + O((\bar\rho_k - \bar\lambda_1)^{-m+q}), \quad q=1,2,\dots,m. \tag{2.10}$$

Опираясь на соотношения (2.4),(2.5) и (2.10), найдем

$$\Sigma_1 = -c_{1k} a_m (\bar\rho_k - \bar\lambda_1)^{-m} y_{11} + O((\bar\rho_k - \bar\lambda_1)^{-m+1}) \tilde y_{1k},$$

где $\tilde y_{1k} \in Y_1$ и $\|\tilde y_{1k}\| = 1$. Повторим аналогичные рассуждения, как для нахождения Σ_1, так и для компонентов Σ_j, $j=2,3,\dots,m$ формулы (2.3). Получим

$$\Sigma_j = -c_{jk} a_{j1m} (\bar\rho_k - \bar\lambda_1)^{-m} y_{11} + O((\bar\rho_k - \bar\lambda_1)^{-m+1}) \tilde y_{jk}, \tag{2.11}$$

где $\tilde y_{jk} \in Y_1$ и $\|\tilde y_{jk}\| = 1$, $j=1,2,\dots,m$. Учитывая теперь соотношения (2.3) и (2.11), вектор t_{k+1} приобретает следующий вид

$$t_{k+1} = -(\bar\rho_k - \bar\lambda_1)^{-m} \sum_{j=1}^{m} c_{jk} a_{j1m} y_{11} + O((\bar\rho_k - \bar\lambda_1)^{-m+1}) \tilde y_k + (T^* - \bar\rho_k)^{-1} z_k,$$

где $\tilde{y}_k \in Y_1$, $\|\tilde{y}_k\|=1$.

Найдем сумму $\sum_{j=1}^{m} c_{jk} a_{j1m}$. Исходим из $x_{1j}=\sum_{l=1}^{r}\sum_{p=1}^{m_l} a_{jlp} y_{lp}$ (см.(2.2)) и умножим этот вектор на x_{11}, получим

$$(x_{1j},x_{11})=\sum_{l=1}^{r}\sum_{p=1}^{m_l} a_{jlp}(y_{lp},x_{11})=\sum_{p=1}^{m} a_{j1p}(y_{1p},x_{11})=a_{j1m}(y_{1m},x_{11}).$$

Из последнего равенства вытекает, что $a_{j1m}=(x_{1j},x_{11})/(y_{1m},x_{11})=$ $=a_m(x_{1j},x_{11})$. Следовательно, в силу (2.1) получим

$$\sum_{j=1}^{m} c_{jk} a_{j1m}=\sum_{j=1}^{m} c_{jk} a_m(x_{1j},x_{11})=a_m\Big(\sum_{j=1}^{m} c_{jk} x_{1j},x_{11}\Big)=a_m(Pv_k,x_{11}),$$

где P проектор Рисса, проектирующий на корневое подпространство X_1. Теперь вектор t_{k+1} приобретет вид

$$t_{k+1}=-(\bar{\rho}_k-\bar{\lambda}_1)^{-m} a_m(Pv_k,x_{11})[y_{11}+O(\bar{\rho}_k-\bar{\lambda}_1)\tilde{y}_k+O((\bar{\rho}_k-\bar{\lambda}_1)^m)(T^*-\bar{\rho}_k)^{-1}\hat{z}_k]. \quad (2.12)$$

Найдем вектор

$$t_{k+2}=(T-\rho_k)^{-1}y_{11}=(T-\rho_k)^{-1}\sum_{l=1}^{r}\sum_{p=1}^{m_l} b_{1lp}x_{lp}=$$

$$=(T-\rho_k)^{-1}\sum_{p=1}^{m} b_{11p}x_{1p}+(T-\rho_k)^{-1}\sum_{l=2}^{r}\sum_{p=1}^{m_l} b_{1lp}x_{lp}=$$

$$=(T-\rho_k)^{-1}\sum_{p=1}^{m} b_{11p}x_{1p}+(T-\rho_k)^{-1}\tilde{x}_k, \quad (2.13)$$

где $\tilde{x}_k \in X_1^{\top}$. Обозначим

$$(T-\rho_k)^{-1}\sum_{p=1}^{m} b_{11p}x_{1p}=g \quad (2.14)$$

и для краткости отождествим $b_p=b_{11p}$, $p=1,2,\dots,m$. Представим $g=\sum_{j=1}^{m} g_j x_{1j}$ и повторим аналогичные рассуждения, как и для преобразования формулы (2.5). Получим

$$g_j=-(\rho_k-\lambda_1)^{-m+j-1}b_m+O((\rho_k-\lambda_1)^{-m+j}), \quad j=1,2,\dots,m. \quad (2.15)$$

Отметим, что коэффициент b_m можно вычислить аналогично коэффициенту a_m (см. (2.9)),

$$b_m=b_{11m}=1/(x_{1m},y_{11}). \quad (2.16)$$

Следовательно,

$$t_{k+2}=\sum_{j=1}^{m}[-(\rho_k-\lambda_1)^{-m+j-1}b_m+O((\rho_k-\lambda_1)^{-m+j})]x_{1j}+(T-\rho_k)^{-1}\tilde{x}_k=$$

$$=-(\rho_k-\lambda_1)^{-m}b_m\Big\{\sum_{j=1}^{m}[(\rho_k-\lambda_1)^{j-1}+O((\rho_k-\lambda_1)^{j})]x_{1j}+O((\rho_k-\lambda_1)^{m})(T-\rho_k)^{-1}\tilde{x}_k\Big\}.$$

Опираясь теперь на соотношения (2.3), (2.12), (2.13), (2.14) и (2.15), выпишем приближение v_{k+1},

$$v_{k+1}=\tau_k^2[(T^*-\bar{\rho}_k)(T-\rho_k)]^{-1}v_k=\tau_k^2(T-\rho_k)^{-1}(T^*-\bar{\rho}_k)^{-1}v_k=\tau_k^2(T-\rho_k)^{-1}t_{k+1}=$$

$$=-\tau_k^2(\bar{\rho}_k-\bar{\lambda}_1)^{-m}a_m(Pv_k,x_{11})(T-\rho_k)^{-1}[y_{11}+O(\bar{\rho}_k-\bar{\lambda}_1)\bar{y}_k+O((\bar{\rho}_k-\bar{\lambda}_1)^{m})(T^*-\bar{\rho}_k)^{-1}\hat{z}_k]=$$

$$=-\tau_k^2(\bar{\rho}_k-\bar{\lambda}_1)^{-m}a_m(Pv_k,x_{11})\{t_{k+2}+(T-\rho_k)^{-1}[O(\bar{\rho}_k-\bar{\lambda}_1)\bar{y}_k+O((\bar{\rho}_k-\bar{\lambda}_1)^{m})(T^*-\bar{\rho}_k)^{-1}\hat{z}_k]\}=$$

$$=\frac{\tau_k^2a_mb_m(Pv_k,x_{11})}{|\lambda_1-\rho_k|^{2m}}\{\sum_{j=1}^{m}[(\rho_k-\lambda_1)^{j-1}+O((\rho_k-\lambda_1)^{j})]x_{1j}+O((\bar{\rho}_k-\bar{\lambda}_1)^{m})z_{k+1}\}=$$

$$=\frac{\tau_k^2a_mb_m(Pv_k,x_{11})}{|\lambda_1-\rho_k|^{2m}}\{\sum_{j=1}^{m}c_jx_{1j}+O((\bar{\rho}_k-\bar{\lambda}_1)^{m})z_{k+1}\},\tag{2.17}$$

где $c_j=(\rho_k-\lambda_1)^{j-1}+O((\rho_k-\lambda_1)^{j}),\quad j=1,2,\dots,m.$

Приступим к нахождению приближения u_{k+1}. Для этого проведем аналогичные рассуждения как и для нахождения вектора v_{k+1}. Получим (см. (2.17))

$$u_{k+1}=\varkappa_k^2[(T-\rho_k)(T^*-\bar{\rho}_k)]^{-1}u_k=$$

$$=\frac{\varkappa_k^2a_mb_m(y_{11},P^*u_k)}{|\lambda_1-\rho_k|^{2m}}\{\sum_{j=1}^{m}\bar{d}_jy_{1j}+O((\rho_k-\lambda_1)^{m})w_{k+1}\},\tag{2.18}$$

где $d_j=(\rho_k-\lambda_1)^{j-1}+O((\rho_k-\lambda_1)^{j}),\quad j=1,2,\dots,m.$

Для обобщенного отношения Релея $\rho(v,u)=(Tv,u)/(v,u)$ справедливо $\rho(\alpha v,\beta u)=\rho(v,u)$, если только $\alpha,\beta\neq0$. Так как $a_m,b_m\neq0$ (см. (2.9) и (2.16)), то

$$\rho_{k+1}-\lambda_1=\frac{(Tv_{k+1},u_{k+1})}{(v_{k+1},u_{k+1})}-\lambda_1=\frac{((T-\lambda_1)v_{k+1},u_{k+1})}{(v_{k+1},u_{k+1})}=$$

$$=\frac{((T-\lambda_1)\sum_{j=1}^{m}c_jx_{1j},\sum_{j=1}^{m}\bar{d}_jy_{1j})+O(|\lambda_1-\rho_k|^{2m})}{(\sum_{j=1}^{m}c_jx_{1j},\sum_{j=1}^{m}\bar{d}_jy_{1j})+O(|\lambda_1-\rho_k|^{2m})}=$$

$$=\frac{\left(\sum_{j=2}^{m} c_j x_{1,j-1},\sum_{j=1}^{m}\bar d_j y_{1j}\right)+O(|\lambda_1-\rho_k|^{2m})}{\sum_{j=1}^{m} c_j d_{m+1-j}(x_{1j},y_{1,m+1-j})+O(|\lambda_1-\rho_k|^{2m})}=$$

$$=\frac{\sum_{j=1}^{m-1} c_{j+1} d_{m+1-j}(x_{1j},y_{1,m+1-j})+O(|\lambda_1-\rho_k|^{2m})}{\sum_{j=1}^{m} c_j d_{m+1-j}(x_{1j},y_{1,m+1-j})+O(|\lambda_1-\rho_k|^{2m})}. \qquad (2.19)$$

Учитывая вид коэффициентов c_j, d_j, $j=1,2,\dots,m$ в представлениях (2.17) и (2.18), вычислим

$$c_j d_{m+1-j}=[(\rho_k-\lambda_1)^{j-1}+O((\rho_k-\lambda_1)^j)][(\rho_k-\lambda_1)^{m-j}+O((\rho_k-\lambda_1)^{m+1-j})]=$$

$$=(\rho_k-\lambda_1)^{m-1}+O((\rho_k-\lambda_1)^m), \qquad j=1,2,\dots,m.$$

Аналогично

$$c_{j+1} d_{m+1-j}=(\rho_k-\lambda_1)^m+O((\rho_k-\lambda_1)^{m+1}), \qquad j=1,2,\dots,m-1.$$

Опираясь на определение корневых векторов, получим $(x_{11},y_{1m})=(x_{1j},y_{1,m+1-j})$ для $j=1,2,\dots,m$. В силу формулы (9.19) найдем

$$\frac{\rho_{k+1}-\lambda_1}{\rho_k-\lambda_1}=\frac{\sum_{j=1}^{m-1}[(\rho_k-\lambda_1)^m+O((\rho_k-\lambda_1)^{m+1}]}{(\rho_k-\lambda_1)\sum_{j=1}^{m}[(\rho_k-\lambda_1)^{m-1}+O((\rho_k-\lambda_1)^m]}=$$

$$=\frac{(m-1)(\rho_k-\lambda_1)^m+O((\rho_k-\lambda_1)^{m+1})}{m(\rho_k-\lambda_1)^m+O((\rho_k-\lambda_1)^{m+1})}=\frac{m-1}{m}+O(\rho_k-\lambda_1). \qquad (2.20)$$

В пределе $k\to\infty$ отношение (2.20) завершает доказательство одной части теоремы.

Используя соотношения (2.17) и (2.20), преобразуем

$$(T-\rho_{k+1})v_{k+1}=(T-\lambda_1)v_{k+1}+(\lambda_1-\rho_{k+1})v_{k+1}=$$

$$=\frac{\tau_k^2 a_m b_m (Pv_k,x_{11})}{|\lambda_1-\rho_k|^{2m}}\{(T-\lambda_1)((\rho_k-\lambda_1)x_{12}+O((\rho_k-\lambda_1)^2))+(\lambda_1-\rho_{k+1})(x_{11}+O(\rho_k-\lambda_1))\}=$$

$$=\frac{\tau_k^2 a_m b_m (Pv_k,x_{11})}{|\lambda_1-\rho_k|^{2m}}\{(\rho_k-\lambda_1)x_{11}+(\lambda_1-\rho_{k+1})x_{11}+O((\rho_k-\lambda_1)^2)\}=$$

$$=\frac{\tau_k^2 a_m b_m (Pv_k, x_{11})}{|\lambda_1-\rho_k|^{2m}}\{(\rho_k-\lambda_1)x_{11}+\frac{m-1}{m}(\lambda_1-\rho_k)x_{11}+O((\rho_k-\lambda_1)^2)\}=$$

$$=\frac{\tau_k^2 a_m b_m (Pv_k, x_{11})}{|\lambda_1-\rho_k|^{2m}}\{\frac{1}{m}(\rho_k-\lambda_1)x_{11}+O((\rho_k-\lambda_1)^2)\}. \tag{2.21}$$

Так как вектор v_{k+1} единичный, то из представления (2.17) вытекает

$$\frac{\tau_k^2 a_m b_m (Pv_k, x_{11})}{|\lambda_1-\rho_k|^{2m}}=1+O(\rho_k-\lambda_1). \tag{2.22}$$

Найдем и вектор невязки $(T-\rho_k)v_k$. Опираясь теперь на результаты (2.20), (2.21) и (2.22), получим

$$\lim_{k\to\infty}\frac{\|(T-\rho_{k+1})v_{k+1}\|}{\|(T-\rho_k)v_k\|}=\lim_{k\to\infty}\frac{|\lambda_1-\rho_k|}{|\lambda_1-\rho_{k-1}|}=\frac{m-1}{m}.$$

Точно так же проведем и нахождение предела

$$\lim_{k\to\infty}\frac{\|(T^*-\bar\rho_{k+1})u_{k+1}\|}{\|(T^*-\bar\rho_k)u_k\|}=\frac{m-1}{m}.$$

В итоге мы получили доказательство всей теоремы. □

Рассмотрим теперь случай, когда корневое подпространство X_1 содержит q собственных векторов. Проведем в подпространстве X_1 разбиение $X_1=X_{11}+X_{12}+\ldots+X_{1q}$, где X_{1i}, $i\le q$ линейная оболочка некоторого собственного вектора и выходящих из него корневых векторов. Обозначим длины корневых цепочек через m_i, $i\le q$.

Если $m_i=1$ для каждого $i\le q$, то из пункта 1 известна по крайней мере квадратичная сходимость итерационного процесса.

Рассмотрим случай, когда найдется $m_i>1$, $i\le q$. Пусть на k-ом шаге приближенный вектор v_k имеет компоненты всех X_{1i}, $i\le q$. В таком случае в итерационном процессе будут расти компоненты таких собственных векторов, которые имеют самые длинные корневые цепочки. Если таких собственных векторов один (в точности до коэффициента), то итерационный процесс сходится к этому собственному вектору. Если их больше, то итерационный процесс сходится к некоторому собственному вектору, принадлежащему к линейной оболочке этих собственных векторов. Если компоненты корневых векторов, имеющие самые длинные корневые цепочки будут отсутствовать, то скорость сходимости будет лучше.

Теорема 2.2. Пусть последовательности $\{v_k\}$ и $\{u_k\}$, построенные по методу (I)-(III), такие, что $v_k \to x_1$ и $u_k \to y_1$ при $k\to\infty$, где $Tx_1=\lambda_1 x_1$, $T^* y_1=\bar\lambda_1 y_1$ и $\|x_1\|=\|y_1\|=1$. Выберем последовательность $\{\alpha_k\}$ такую, что $\lim\limits_{k\to\infty}\frac{\alpha_k}{|\lambda_1-\rho_k|^{2m}}=0$, где m длина корневой цепочки, соответствующая собственному вектору x_1.

1. Если $T^* x_1=\bar\lambda_1 x_1$ и $Ty_1=\lambda_1 y_1$, то

$$\lim_{k\to\infty}\frac{|\lambda_1-\rho_{k+1}|}{|\lambda_1-\rho_k|^5}=\text{const},\quad \lim_{k\to\infty}\frac{\|r_{k+1}\|}{\|r_k\|^3\|p_k\|^2}=\text{const},\quad \lim_{k\to\infty}\frac{\|p_{k+1}\|}{\|r_k\|^2\|p_k\|^3}=\text{const}.$$

2. Если длина корневой цепочки, соответствующая вектору x_1, равна единице, то

$$\lim_{k\to\infty}\frac{|\lambda_1-\rho_{k+1}|}{|\lambda_1-\rho_k|^2}=\text{const},\quad \lim_{k\to\infty}\frac{\|r_{k+1}\|}{\|r_k\|\|p_k\|}=\text{const},\quad \lim_{k\to\infty}\frac{\|p_{k+1}\|}{\|r_k\|\|p_k\|}=\text{const}.$$

3. Если длина корневой цепочки, соответствующая вектору x_1, равна m, то

$$\lim_{k\to\infty}\frac{\lambda_1-\rho_{k+1}}{\lambda_1-\rho_k}=\lim_{k\to\infty}\frac{\|r_{k+1}\|}{\|r_k\|}=\lim_{k\to\infty}\frac{\|p_{k+1}\|}{\|p_k\|}=\frac{m-1}{m}.$$

<u>Доказательство</u>. Справедливость случаев 1 и 2 следует из пункта 1.

Утверждения случая 3 верны для $\alpha_k=0$, $k=0,1,2,\ldots$. Убедимся, что при $\lim\limits_{k\to\infty}\frac{\alpha_k}{|\lambda_1-\rho_k|^{2m}}=0$ сохраняется линейная сходимость с коэффициентом $\frac{m-1}{m}$. Обозначим через $A_k=(T^*-\bar\rho_k)(T-\rho_k)$. Из доказательства теоремы 1 следует существование A_k^{-1} и $\|A_k^{-1}\|\le\frac{c}{|\lambda_1-\rho_k|^{2m}}$, где c некоторая константа. Из предположения $\lim\limits_{k\to\infty}\frac{\alpha_k}{|\lambda_1-\rho_k|^{2m}}=0$ следует, что $\lim\limits_{k\to\infty}\|\alpha_k A_k^{-1}\|\le\lim\limits_{k\to\infty}\frac{\alpha_k c}{|\lambda_1-\rho_k|^{2m}}=0$. В силу последнего предела при $k\to\infty$ справедливо

$$(\alpha_k+A_k)^{-1}v_k=[A_k(\alpha_k A_k^{-1}+I)]^{-1}v_k=\sum_{i=0}^{\infty}(-\alpha_k A_k^{-1})^i A_k^{-1}v_k\to A_k^{-1}v_k,\ \text{т.е.}$$

$$[\alpha_k+(T^*-\bar\rho_k)(T-\rho_k)]^{-1}v_k\to[(T^*-\bar\rho_k)(T-\rho_k)]^{-1}v_k.$$

Последний результат говорит о том, что такой выбор последовательности $\{\alpha_k\}$ сохраняет линейную сходимость с коэффициентом $\frac{m-1}{m}$. □

В работе [3] исследовано, как выбирать последовательность $\{\alpha_k\}$ при практической реализации на ЭВМ для случаев, когда подпространство X_1 имеет нормальную или простую структуру.

Теперь покажем, что для случая 3 теоремы 2 предположения $\lim\limits_{k\to\infty}\frac{\alpha_k}{|\lambda_1-\rho_k|^{2m}}=0$ и $\lim\limits_{k\to\infty}\frac{\alpha_k}{\|r_k\|^m\|p_k\|^m}=0$ равносильны. Из (2.21) и (2.22) вытекает, что

$$\|r_{k+1}\|=\frac{1}{m}|\lambda_1-\rho_k|+O(|\lambda_1-\rho_k|^2).$$

Аналогично

$$\|p_{k+1}\|=\frac{1}{m}|\lambda_1-\rho_k|+O(|\lambda_1-\rho_k|^2).$$

Вычислим

$$\lim_{k\to\infty}\frac{\alpha_k}{\|r_k\|^m\|p_k\|^m}=\lim_{k\to\infty}\frac{\alpha_k}{[\frac{1}{m}|\lambda_1-\rho_{k-1}|+O(|\lambda_1-\rho_{k-1}|^2)]^{2m}}=$$

$$=\lim_{k\to\infty}\frac{\alpha_k m^{2m}}{|\lambda_1-\rho_{k-1}|^{2m}}=m^{2m}\lim_{k\to\infty}\frac{|\lambda_1-\rho_k|^{2m}}{|\lambda_1-\rho_{k-1}|^{2m}}\lim_{k\to\infty}\frac{\alpha_k}{|\lambda_1-\rho_k|^{2m}}=$$

$$=m^{2m}\left(\frac{m-1}{m}\right)^{2m}\lim_{k\to\infty}\frac{\alpha_k}{|\lambda_1-\rho_k|^{2m}}=(m-1)^{2m}\lim_{k\to\infty}\frac{\alpha_k}{|\lambda_1-\rho_k|^{2m}}. \tag{2.23}$$

Из (2.23) следует, что $\lim\limits_{k\to\infty}\frac{\alpha_k}{\|r_k\|^m\|p_k\|^m}=0$ и $\lim\limits_{k\to\infty}\frac{\alpha_k}{|\lambda_1-\rho_k|^{2m}}=0$ равносильны. Следовательно, при реализации метода на ЭВМ в случае 3 можно последовательность $\{\alpha_k\}$ выбирать так, что выполняется $\lim\limits_{k\to\infty}\frac{\alpha_k}{\|r_k\|^m\|p_k\|^m}=0$.

Литература

1. С а а н Т. Регуляризованный итерационный метод для отыскания собственных значений и собственных векторов симметричной матрицы. // Уч. зап. Тарт. ун-та., 1987., №762., с. 59-68.
2. С а а н Т. Регуляризованный итерационный метод для отыскания собственных значений и собственных векторов неэрмитовой матрицы. // Уч. зап. Тарт. ун-та., 1988., №833., с. 84-90.
3. С а а н Т. Итерационный метод с отношением Релея. // Уч. зап. Тарт. ун-та., 1989., №863., с. 57-70.
4. O s t r o w s k i A. M. On the Convergence of the Rayleigh Quotient Iteration for the Computation of Characteristic Roots and Vectors I-VI. // Arch. Rational Mech. Anal., 1958/59., 1-4., 1: pp. 233-241, 2: pp. 423-428, 3: pp. 325-340, 3: pp. 341-347, 3: pp. 472-481, 4: pp. 153-165.

THE REGULARIZED ITERATIVE METHOD WITH THE GENERALIZED RAYLEIGH QUOTIENT

T.Saan

Summary

The regularized iterative method with the Rayleigh quotient for finding the eigenvalues and eigenvectors of matrices was investigated in papers [1], [2], [3]. In the case of normal matrices this method has quintic convergence rate. In the case of non-normal matrices the asymptotic convergence is, at best, linear. In this work the asymptotic convergence of the method with the generalized Rayleigh quotient is analyzed. In the normal case the asymptotic convergence rate for this method is quintic, in the non-defective case quadratic and in the defective case linear.

Acta et commentationes universitatis Tartuensis, 863, 1989, 86-94.

О ВЫБОРЕ ПАРАМЕТРА ДЛЯ КЛАССА МЕТОДОВ РЕГУЛЯРИЗАЦИИ В УСЛОВИЯХ СЛУЧАЙНЫХ ОШИБОК

Тоомас Раус

В статье рассматривается одна возможность выбора параметра для класса методов решений линейных некорректных задач в случае случайных ошибок исходных данных. Предполагается, что свободный член уравнения $Au=f_z$ имеет вид $f_z = f + z$, где элемент z — одна реализация некоторой случайной величины ξ и нам известно вероятностное распределение величины $\|\xi\|$. Статья является обобщением работы [1], в которой выбор параметра изучался для метода Лаврентьева.

1. Постановка задачи. Рассмотрим вероятностное пространство (Ω,F,P). Пусть $\xi = \xi(\omega)$, $\omega \in \Omega$, – случайная величина со значениями в сепарабельном гильбертовом пространстве F, наделенным борелевской σ-алгеброй β. Хорошо известно (см. [2]), что если ξ: $(\Omega,\Phi) \to (F,\beta)$ – случайная величина, то норма $\|\xi\|$: $(\Omega,\Phi) \to (R^1,\beta^1)$ – тоже случайная величина и мы можем рассматривать выражения типа $P(\|\xi\| \le \delta) \ge \alpha$.

Рассмотрим уравнение

$$Au = f, \tag{1}$$

где $A \in \mathscr{L}(H,F)$ - линейный непрерывный оператор, H и F - сепарабельные гильбертовые пространства. Допускается незамкнутость области значений $R(A) \subset F$ и нетривиальность ядра $N(A)$. Предполагается, что вместо f задан $f_z = f+z$, где элемент z - одна реализация случайной величины ξ (то есть $z = \xi(\omega_z)$, $\omega_z \subset \Omega$) и нам известны числа $1 = \alpha_1 \ge \alpha_2 \ge \ldots \ge \alpha_n$, $\delta_1 \ge \delta_2 \ge \ldots \ge \delta_n$ такие, что

$$P(\|\xi\| \le \delta_k) \ge \alpha_k, \quad k = 1,2,\ldots,n \tag{2}$$

Также предполагается, что $E\xi = 0$.

Для решения уравнения (1) рассмотрим класс методов, подробно изученный в [3]. Пусть $\{g_r\}_{r\in(0,\infty)}$ - семейство ограниченных, измеримых по Борелью функций $g_r : [0,a] \to R$ таких, что $\|A\| \le a$ и

$$\sup_{0\le\lambda\le a} |g_r(\lambda)| \le \gamma r \qquad (r > 0), \tag{3}$$

$$\sup_{0\le\lambda\le a} \lambda^p |1-\lambda g_r(\lambda)| \le \gamma_p r^{-p} \qquad (r>0,\ 0\le p\le p_0,\ p_0>0), \qquad (4)$$

где γ, $\gamma_p = \text{const}$. В случае $H = F$, $A = A^* \ge 0$, $\|A\| \le a$ приближенное решение уравнения (1) построим в виде

$$u_r^z = (I - Ag_r(A))u_0 + g_r(A)f_z,$$

где $u_0 \in H$ - начальное приближение и I - единичный оператор, а в случае несамосопряженного оператора в виде

$$u_r^z = (I - A^*Ag_r(A^*A))u_0 + g_r(A^*A)A^*f_z,$$

где A^* – сопряженный к оператору A оператор. Примерами данного класса методов являются такие хорошо известные методы как методы Лаврентьева и Тихонова, соответственно в виде

$$u_r = (I+rA)^{-1}u_0 + r(I+rA)^{-1}f_z,$$

$$u_r^z = (I+rA^*A)^{-1}u_0 + r(I+rA^*A)^{-1}f_z,$$

а также явный и неявный итерационный метод.

Применение этих методов в случае случайных ошибок по сравнению с детерминированными ошибками рассмотрено до сих пор мало. В работах М.А.Красносельского и А.Ю.Веретенникова [3,4] изучалась сходимость по вероятности для итерационных методов со случайными ошибками, которые возникают в силу погрешностей округления в вычислении итераций. Отметим также работу В.А.Морозова [5], посвященную вопросу оптимального выбора параметра для метода Тихонова в случае известных дисперсий σ_k^2, $k = 1,2,\dots$, координат случайного вектора $f_z - f$. В литературе наиболее популярными объектами исследования являлись чисто статистические методы, но реализация таких методов требовала наличия априорной информации о решении, например, о принадлежности решения некоторому априорно заданному множеству корректности. В данной статье предполагается, что в нашем распоряжении имеется лишь вероятностная информация (2) о величине погрешностей исходных данных.

Оценим сначала дисперсию $E\|\xi\|^2$ случайной величины ξ. Имеет место

Лемма 1. Пусть выполнены условия (2). Тогда справедливо неравенство

$$E\|\xi\|^2 \le \sigma^2,$$

где

$$\sigma^2 = \sum_{k=1}^{n} (\alpha_k - \alpha_{k+1})\delta_k^2, \qquad \alpha_{n+1} = 0.$$

Доказательство. Обозначим $\Omega_k = \{\omega: \delta_{k+1} < \|\xi(\omega)\| \le \delta_k\}$, $1 \le k \le n$, $\delta_{n+1} = 0$. Тогда имеем

$$E\|\xi\|^2 = \int_\Omega \|\xi(\omega)\|^2 dP = \sum_{k=1}^n \int_{\Omega_k} \|\xi(\omega)\|^2 dP \le \sum_{k=1}^n \delta_k^2 P(\delta_k < \|\xi\| \le \delta_{k+1}). \tag{5}$$

Обозначим $x_k = P(\delta_{k+1} < \|\xi\| \le \delta_k)$. Имеем $\sum_{k=1}^n x_k = 1$. В силу неравенства $P(\|\xi\| \le \delta_2) \ge \alpha_2$ имеем $x_1 \le 1 - \alpha_2 = \alpha_1 - \alpha_2$, в силу неравенства $P(\|\xi\| \le \delta_3) \ge \alpha_3$ имеем $x_1 + x_2 \le \alpha_1 - \alpha_3 = (\alpha_1 - \alpha_2) + (\alpha_2 - \alpha_3)$; в общем случае

$$\sum_{k=1}^l x_k = \sum_{k=1}^l (\alpha_k - \alpha_{k+1}), \qquad l = 1,2,\dots,n.$$

Теперь, учитывая, что $\delta_1 \ge \delta_2 \ge \dots \ge \delta_n$, получим

$$\sum_{k=1}^n \delta_k^2 x_k \le \max_{\substack{x_k,\ k=1,2,\dots,n \\ \sum_{k=1}^l x_k \le \sum_{k=1}^l (\alpha_k - \alpha_{k+1}),\ l=1,2,\dots,n}} \sum_{k=1}^n \delta_k^2 x_k =$$

$$= \delta_1^2 (\alpha_1 - \alpha_2) + \max_{\substack{x_k,\ k=2,3,\dots,n \\ \sum_{k=2}^l x_k \le \sum_{k=2}^l (\alpha_k - \alpha_{k+1}),\ l=2,3,\dots,n}} \sum_{k=2}^n \delta_k^2 x_k =$$

$$= \dots = \sum_{k=1}^n \delta_k^2 (\alpha_k - \alpha_{k+1}),$$

что вместе с (5) доказывает лемму 1.

2. Выбор параметра. Введем оператор B_r, который зависит от квалификации метода p_0 (см. неравенство (4)), обозначая

$$B_r = \begin{cases} (I - A g_r(A))^{1/p_0}, & \text{если } p_0 < \infty, \\ I, & \text{если } p_0 = \infty \end{cases}$$

в случае $A = A^* \ge 0$ и

$$B_r = \begin{cases} (I - A^*A g_r(A^*A))^{1/2p_0}, & \text{если } p_0 < \infty, \\ I, & \text{если } p_0 = \infty \end{cases}$$

в случае несамосопряженного оператора.

Для методов, в случае которых приближенное u_r^z можно вычислить для всех вещественных $r \ge 0$ (например, методы Лаврентьева и Тихонова), обозначим через $\bar r_k$, $k = 1,2,\dots,n$, любое значение параметра, при котором выполняются неравенства

$$b_1 \delta_k \le \|B_r(Au_r^z - f_z)\| \le b_2 \delta_k, \tag{6}$$

где $b_2 \ge b_1 > 1$ — заданные постоянные, причем $b_2/b_1 < \min\limits_{1\le k\le n} \delta_k/\delta_{k+1}$. Если приближенное u_r^z можно вычислить только для целых чисел (например, итерационные методы), то под r_k понимаем наименьшее целое число, при котором выполняется неравенство

$$\|B_r(Au_r^z - f_z)\| \le b_2\delta_k, \tag{7}$$

где $b_2 > 1$. Таким образом, параметры $\bar r_k$ определяются по модифицированному принципу невязки (см. [6,7]). Если такого $\bar r_k$ не существует (это возможно, когда $\|f_z - f\| > \delta_k$), то положим $\bar r_k = \infty$. Заметим, что в силу условия $P(\|f_z - f\| \le \delta_1) = 1$ по крайней мере $\bar r_1$ с вероятностью 1 конечно. В силу монотонности функции $f(r) = \|B_r(Au_r^z - f_z)\|$ и неравенства $b_1\delta_k > b_2\delta_{k+1}$, $1 \le k \le n-1$, имеем $\bar r_k \le \bar r_{k+1}$. Обозначим

$$m = \begin{cases} k_0, \text{ если } \bar r_{k_0-1} < \infty \text{ и } \bar r_{k_0} = \infty, \\ n, \text{ если } \bar r_n < \infty \end{cases}$$

и под r_k понимаем величину $r_k = \max\{1, \bar r_{k_0}\}$, $k = 1,2,\dots,m$.

Определим функцию $F(r)$ на отрезке $[r_1, r_m]$. Если $r_j \le r < r_{j+1}$, $1 \le j \le m-1$ или $r = r_{j+1} = r_n < \infty$, то в случае $A = A^* \ge 0$

$$F(r) = \sum_{k=1}^{j}(\alpha_k - \alpha_{k+1})\left(\frac{\|u_r^z - u_{r_k}^z\|}{\gamma r_k \delta_k}\right)^2 \max\{1, (\delta_k/\sigma)^2\} + \alpha_{j+1}\|B_r(Au_r^z - f_z)\|/\sigma,$$

и в случае несамосопряженной задачи

$$F(r) = \sum_{k=1}^{j}(\alpha_k - \alpha_{k+1})\left(\frac{\|u_r^z - u_{r_k}^z\|}{\gamma_*\sqrt{r_k}\,\delta_k}\right)^2 \max\{1, (\delta_k/\sigma)^2\} + \alpha_{j+1}\|B_r(Au_r^z - f_z)\|/\sigma.$$

Сформулируем теперь правила выбора параметра. Для методов, для которых $\bar r_k$ определяется из условия (6), используем

Правило П1. Зададим числа C_0, b_1, b_2, где $C_0 \ge 0$, $b_2 \ge b_1 > 1$, $b_2/b_1 < \min\limits_{1\le k\le n-1} \delta_k/\delta_{k+1}$. В качестве параметра регуляризации $r(z) = r(f_z, \{\alpha_k\}_1^n, \{\delta_k\}_1^n)$ выберем любое число $r \in [r_1, r_m]$ такое, что

$$F(r) \le \operatorname{glob}\min_{r_1 \le s \le r_m} F(s) + C_0.$$

В противном случае используем

Правило П2. Зададим числа C_0, b_2, где $C_0 \ge 0$, $b_2 > 1$. В качестве параметра регуляризации $r(z) = r(f_z, \{\alpha_k\}_1, \{\delta_k\}_1)$ выберем любое целое число $r \in [r_1, r_m]$ такое, что

$$F(r) \le \operatorname{glob}\min_{r_1 \le s \le r_m,\ s \in N} F(s) + C_0.$$

Число C_0 задается в правилах исходя из вычислительных рассуждений, целесообразно выбрать C_0 меньше единицы. Нетрудно увидеть, что при заданных числах C_0, b_1, b_2, функция $F(r)$ вычислима, и что для разных z_1 и z_2 параметры $r(z_1)$ и $r(z_2)$ будут в общем случае разными.

Выведем теперь одну оценку для функции $F(r)$, которая облегчит вычисление параметра $r(z)$. Представим функцию $F(r)$ в виде $F(r) = D(r) + G(r)$, где при $r_j \le r \le r_{j+1}$

$$D(r) = \sum_{k=1}^{j} (\alpha_k - \alpha_{k+1}) \left(\frac{\|u_r^z - u_{r_k}^z\|}{\gamma r_k \delta_k} \right)^2 \max\{1, (\delta_k/\sigma)^2\}$$

$$G(r) = \alpha_{j+1} \|B_r(Au_r^z - f_z)\| / \sigma.$$

Так как $\|B_r(Au_r^z - f_z)\|$ – убывающая, а $\|u_r^z - u_{r_k}^z\|$ – возрастающая при $r \ge r_k$ функция, то $G(r)$ и $D(r)$ – соответственно убывающая и возрастающая функции. Таким образом, мы можем оценить функцию $F(r)$ на любом отрезке $[r', r''] \subset [r_1, r_m]$ следующим образом:

$$G(r') + D(r'') \le F(r) \le G(r'') + D(r'), \quad r \in [r', r''].$$

С учетом последнего неравенства представим один возможный алгоритм для вычисления параметра $r(z)$.

1. Зададим шаг R и вычислим функции $D(r)$ и $G(r)$ в точках $r^k = r_1 + (k - 1)R$, $k = 1,2,\ldots,k_0$, где k_0 – наименьший индекс, при котором $D(r^{k_0}) \ge G(r^1) = F(r^1)$.
2. Для каждого отрезка $[r^k, r^{k+1}]$, $k = 1,2,\ldots,k_0 - 1$, вычислим величину $T^k = G(r^k) + D(r^{k+1})$.
3. Находим индекс s_0, для которого $T_{s_0} = \min\limits_{1 \le k \le k_0} T_k$.
4. Если

$$G(r^{s_0}) + D(r^{s_0+1}) - G(r^{s_0+1}) - D(r^{s_0}) \le C_0, \tag{8}$$

то в качестве $r(z)$ выберем любое число $r \in [r^{s_0}, r^{s_0+1}]$.

5. Если (8) не выполнено, то выбросим те отрезки $[r^k, r^{k+1}]$, для которых

$$G(r^{k+1}) + D(r^k) - C_0 \ge G(r^{s_0}) + D(r^{s_0+1}).$$

6. Остальные отрезки (т.е. отрезки, для которых последнее неравенство не выполнено) делим пополам и продолжим вычисления, начиная с пункта 2.

3. Сходимость функции риска. Приведем сперва один результат о модифицированном принципе невязки [6,7], который понадобится нам в дальнейшем. Пусть $\|f_z - f\| \le \delta_k$ и параметр $\bar{r}_k = r(\delta_k)$ выбран по модифицированному принципу невязки (6) или (7). Тогда в случае $A = A^* \geq 0$

$$\|u_{\bar r_k} - u_*\| \to 0, \qquad \bar r_k\, \delta_k \to 0 \quad \text{при } \delta_k \to 0 \tag{9}$$

и в случае несамосопряженной задачи

$$\|u_{\bar r_k} - u_*\| \to 0, \qquad \sqrt{\bar r_k}\, \delta_k \to 0 \quad \text{при } \delta_k \to 0 . \tag{10}$$

Как обычно, в случае случайных ошибок погрешность приближенного решения будем характеризовать так называемой функцией риска $E\|u^{\xi}_{r(\xi)} - u_*\|^2$. Под f_ξ в дальнейшем будем понимать элемент $f + \xi(\omega)$, где $\xi(\omega)$ – произвольная реализация случайной величины ξ, а через u_r обозначим u^{ξ}_r.

Теорема. Пусть параметры $r = r(\xi)$ выбраны по правилам П1 или П2. Тогда, если $\sigma \to 0$, то

$$E\|u_{r(\xi)} - u_*\|^2 \to 0.$$

Доказательство будем проводить только для самосопряженной задачи. Рассуждения в случае несамосопряженного оператора совершенно аналогичные. Буквой C обозначим в доказательстве постоянные (в разных местах значения C, вообще говоря, разные).

1. Аналогично оценке дисперсии (см. лемма 1) получим

$$E\|u_{r(\xi)} - u_*\|^2 \leq \sum_{k=1}^{n} (\alpha_k - \alpha_{k+1}) \sup_{f_\xi,\ \|f_\xi - f\| \leq \delta_k} \|u_{r(\xi)} - u_*\|^2. \tag{11}$$

Очевидно, что $\sigma = \left(\sum_{k=1}^{n} (\alpha_k - \alpha_{k+1})\, \delta_k^2\right)^{1/2} \to 0$ только тогда, когда $(\alpha_k - \alpha_{k+1})\delta_k^2 \to 0$ при каждом $k = 1,2,\dots,n$. При этом считаем, что $\delta_1 \geq \delta_2 \geq \dots \geq \delta_n$ при всех σ. В силу неравенства (11) для доказательства теоремы достаточно рассмотреть произвольный элемент f_ξ такой, что $\|f_\xi - f\| \leq \delta_k$ и установить сходимость

$$(\alpha_k - \alpha_{k+1})\, \|u_{r(\xi)} - u_*\|^2 \to 0 \quad \text{при } \sigma \to 0 \tag{12}$$

для любого $k \in \{1,2,\dots,n\}$.

2. Итак, рассмотрим f_ξ такой, что $\|f_\xi - f\| \leq \delta_k$, $k \in \{1,2,\dots,n\}$. Предположим сначала, что $r(\xi) \geq r_k$. Имеем

$$\begin{aligned}(\alpha_k - \alpha_{k+1})\|u_{r(\xi)} - u_*\|^2 &\leq \\ \leq 2(\alpha_k - \alpha_{k+1})\|u_{r(\xi)} - u_{r_k}\|^2 &+ 2(\alpha_k - \alpha_{k+1})\|u_{r_k} - u_*\|^2.\end{aligned} \tag{13}$$

Покажем, что

$$(\alpha_k - \alpha_{k+1})\|u_{r_k} - u_*\|^2 \to 0, \quad \text{если } \sigma \to 0. \tag{14}$$

В силу (9) имеем $\|u_{\bar r_k} - u_*\| \leq C$ и $\|u_{\bar r_k} - u_*\| \to 0$, если $\delta_k \to 0$. Поскольку $r_k \leq \bar r_k + 1$, то $\|u_{r_k} - u_*\| \leq \|u_{\bar r_k} - u_*\| + \gamma\, \delta_k$, и, значит, $\|u_{r_k} - u_*\| \leq C$ и $\|u_{r_k} - u_*\| \to 0$, если $\delta_k \to 0$. Так как при $\sigma \to 0$ должно быть $\alpha_k - \alpha_{k+1} \to 0$

или $\delta_k \to 0$, то теперь сходимость (14) становится очевидной.

Докажем, что

$$(\alpha_k - \alpha_{k+1})\|u_{r(\xi)} - u_{r_k}\|^2 \to 0, \text{ если } \sigma \to 0. \tag{15}$$

Так как $F(r(\xi)) \le \mathop{\text{glob min}}_{r_1 \le s \le r_m} F(s) + C_0 \le F(r_k) + C_0$, то

$$\begin{aligned}\sum_{i=1}^{j}(\alpha_i - \alpha_{i+1})\left(\frac{\|u_{r(\xi)} - u_{r_i}\|}{\gamma r_i \delta_i}\right)^2 \max\{1, (\delta_i/\sigma)^2\} \le \\ \le \|B_{r_k}(Au_{r_k} - f_\xi)\| / \sigma + C_0,\end{aligned} \tag{16}$$

где в силу $r(\xi) \ge r_k$ имеем $j \ge k$. С учетом неравенств

$$\|B_{r_k}(Au_{r_k} - f_\xi)\| \le \|B_{\bar r_k}(Au_{\bar r_k} - f_\xi)\| \le b_2\delta_k$$

теперь из (16) вытекает, что

$$(\alpha_k - \alpha_{k+1})\|u_{r(\xi)} - u_{r_k}\|^2 \le (\gamma r_k \delta_k)^2 (b_2\delta_k/\sigma + C_0)\min\{1, (\sigma/\delta_k)^2\}. \tag{17}$$

Пусть теперь $\delta_k \to 0$ при $\sigma \to 0$. Тогда сходимость (15) следует из (17) в силу сходимости $r_k\delta_k \to 0$ (см. (9)) и неравенства

$$(b_2\delta_k/\sigma + C_0)\min\{1, (\sigma/\delta_k)^2\} \le C_0 + b_2\min\{\delta_k/\sigma, \sigma/\delta_k\} \le C_0 + b_2 \le C.$$

Если $\alpha_k - \alpha_{k+1} \to 0$ и $\delta_k \ge C$ в процессе $\sigma \to 0$, то справедливы соотношения $r_k\delta_k \le C$ и

$$\begin{aligned}(b_2\delta_k/\sigma + C_0)\min\{1, (\sigma/\delta_k)^2\} = \\ = (b_2\delta_k/\sigma + C_0)(\sigma/\delta_k)^2 \to 0, \text{ если } \sigma \to 0\end{aligned}$$

и сходимость (15) опять вытекает из (17). Таким образом, если $r(\xi) \ge r_k$, то в силу (13) - (15) получим сходимость (12).

3. Предположим теперь, что $r(\xi) < r_k$. Если $\alpha_k - \alpha_{k+1} \to 0$ при $\sigma \to 0$, то с учетом неравенств

$$\|u_{r(\xi)} - u_*\|^2 \le \|G_{r(\xi)}(u_0 - u_*)\| + \gamma r(\xi)\delta_k \le \|u_0 - u_*\| + \gamma r_k\delta_k \le C$$

утверждение (12) очевидно.

Если $\delta_k \to 0$ при $\sigma \to 0$, то покажем сначала, что

$$\|B_{r(\xi)}(Au_{r(\xi)} - f_\xi)\| \to 0. \tag{18}$$

Пусть s – наименьший индекс, при котором $\delta_s \to 0$ в процессе $\sigma \to 0$. Очевидно, что $s \le k$. Обозначим $r_\sigma = \max\{1, \bar r_\sigma\}$, где параметр $\bar r_\sigma$ выбран по модифицированному принципу невязки, т.е.

$$b_1\sigma \le \|B_{\bar r_\sigma}(Au_{\bar r_\sigma} - f_\xi)\| \le b_2\sigma$$

или

$$\|B_{\bar r_\sigma}(Au_{\bar r_\sigma} - f_\xi)\| \le b_2\sigma, \quad \|B_{\bar r_\sigma - 1}(Au_{\bar r_\sigma - 1} - f_\xi)\| > b_2\sigma.$$

Обозначим $r_0 = \min\{r_\sigma, r_s\}$ и $\delta_0 = \max\{\delta_s, \sigma\}$. Если $r(\xi)$ такое, что $r_0 \le r(\xi) \le r_k$, то

$$\|B_{r(\xi)}(Au_{r(\xi)} - f_\xi)\| \le \|B_{r_0}(Au_{r_0} - f_\xi)\| \le b_2\delta_0$$

и поскольку $\delta_0 \to 0$ при $\sigma \to 0$, то (18) справедлива.

Рассмотрим теперь случай $r(\xi) \le r_0$. Исходя из неравенства $F(r(\xi)) \le F(r_0) + C_0$, получим

$$\|B_{r(\xi)}(Au_{r(\xi)} - f_\xi)\| \le (\alpha_{k_0+1}\|B_{r_0}(Au_{r_0} - f_\xi)\| + C_0\sigma)/\alpha_{j+1} + \\ + \frac{\sigma}{\alpha_{j+1}}\left[\sum^{k_0}(\alpha_i - \alpha_{i+1})\left(\frac{\|u_{r_0} - u_{r_i}\|}{\gamma r_i \delta_i}\right)^2 \max\{1, (\delta_i/\sigma)^2\}, \right. \tag{19}$$

где в силу $r(\xi) \le r_0$ имеем $j \ge k_0$ и $\alpha_{j+1} \ge \alpha_{k+1}$, а по определению параметра r_0 имеем $k_0 \le s - 1$, $\delta_i > \sigma$, $\max\{1, (\delta_i/\sigma)^2\} = (\delta_i/\sigma)^2$. С учетом неравенств

$$\sum_{i=1}^{k_0}(\alpha_i - \alpha_{i+1})(\delta_i/\sigma)^2 \le 1, \quad \|B_{r_0}(Au_{r_0} - f_\xi)\| \le b_2\delta_0,$$

теперь из (19) получим неравенство

$$\|B_{r(\xi)}(Au_{r(\xi)} - f_\xi)\| \le b_2\delta_0 + \frac{\sigma}{\alpha_{k_0+1}}\left(\max_{1,\, 1\le i\le k_0}\left(\frac{\|u_{r_0} - u_{r_i}\|}{\gamma r_i \delta_i}\right)^2 + C_0\right). \tag{20}$$

Нетрудно увидеть, что

$$\|u_{r_0} - u_{r_i}\|/(r_i\delta_i) \le C, \tag{21}$$

если $i \le k_0$. Действительно, так как $i < s$, то по определению индекса s имеем $\delta_i \ge C$, $r_i \le C$ и, значит, $\|u_{r_0} - u_{r_i}\| \le \|u_{r_0}\| + \|u_{r_i}\| \le 2\|u_{r_k}\| \le C$. Покажем еще, что число α_{k_0+1} не сходится к нулю в процессе $\sigma \to 0$. Действительно, поскольку $\delta_i \ge C$, то $\alpha_i - \alpha_{i+1} \to 0$, если $i < s$ и, следовательно,

$$\alpha_{k_0+1} \ge \alpha_s = 1 - \sum_{i=1}^{s-1}(\alpha_i - \alpha_{i+1}) \to 1 \text{ при } \sigma \to 0. \tag{22}$$

Теперь из (20) - (22) непосредственно вытекает сходимость (18) при $\sigma \to 0$.

Наконец, докажем, что

$$\|u_{r(\xi)} - u_*\|^2 \to 0, \quad \text{если} \quad \|B_{r(\xi)}(Au_{r(\xi)} - f_\xi)\| \to 0. \tag{23}$$

Имеем

$$B_{r(\xi)}(Au_{r(\xi)} - f_\xi) = B_{r(\xi)}G_{r(\xi)}A(u_0 - u_*) + B_{r(\xi)}G_{r(\xi)}(f_\xi - f).$$

Поскольку $\|B_{r(\xi)}G_{r(\xi)}(f_\xi - f)\| \le \|f_\xi - f\| \le \delta_k \to 0$, если $\sigma \to 0$, то, следовательно, $\|B_{r(\xi)}G_{r(\xi)}A(u_0 - u_*)\| \to 0$. В книге [3] доказано, что если при $n \to \infty$ имеем $G_{r_n}A(u_0 - u_*) \to 0$, то $G_{r_n}(u_0 - u_*) \to 0$. Аналогично доказывается, что если при $n \to \infty$ имеем $B_{r_n}G_{r_n}A(u_0 - u_*) \to 0$, то $G_{r_n}(u_0 - u_*) \to 0$. Значит, $\|G_{r(\xi)}(u_0 - u_*)\| \to 0$ и с учетом неравенств

$$\|u_{r(\xi)} - u_*\| \le \|G_{r(\xi)}(u_0 - u_*)\| + \gamma r(\xi)\delta_k \le$$

$$\le \|G_{r(\xi)}(u_0 - u_*)\| + \gamma r_k \delta_k$$

получим теперь сходимость (23).

Таким образом, мы показали, что $(\alpha_k - \alpha_{k+1})\|u_{r(\xi)} - u_{r_k}\|^2 \to 0$ при $\sigma \to 0$ и в случае $r(\xi) < r_k$. Теорема доказана.

Литература

1. Раус Т. О выборе параметра регуляризации в случае случайных ошибок данных // Уч. зап. Тартуск. ун-та, 1988, вып. 833, с. 107–122.
2. Ширяев А.Н. Вероятность – М.: Наука, 1980. – 576 с.
3. Вайникко Г.М., Веретенников А.Ю. Итерационные процедуры в некорректных задачах – М.: Наука, 1986. – 184 с.
4. Веретенников А.Ю., Красносельский М.А. Регуляризующие правила останова в условиях случайных ошибок // Докл. АН СССР, 1983, т.268, № 3, с. 521-524.
5. Морозов В.А. Об оптимальной регуляризации операторных уравнений // Ж. вычисл. матем. и матем. физ., т.10, № 4, с. 1341 - 1349.
6. Раус Т. О принципе невязки при решении некорректных задач // Уч. зап. Тартуск. ун-та, 1984, вып. 672, с. 16 - 26.
7. Раус Т. О принципе невязки при решении некорректных задач с несамосопряженным оператором // Уч. зап. Тартуск. ун-та, 1985, вып. 715, с. 12 - 20.

ABOUT THE CHOICE OF PARAMETER FOR THE CLASS OF REGULARIZATION METHODS IN CASE OF RANDOM ERRORS

T. Raus

Summary

Let H,F be separable Hilbert spaces, $A \in L(H,F)$. To regularize ill- posed problem (1), when the errors of the right hand term are random (see (2)), a class of regularization methods is applied. We give a prescription for parameter choice (see rules П1, П2) and prove the theorem of regularization.

Acta et commentationes universitatis Tartuensis, 863, 1989, 95-99.

ОПТИМАЛЬНЫЙ ВЫБОР ПАРАМЕТРА В ИТЕРИРОВАННЫХ ВАРИАНТАХ МЕТОДА ТИХОНОВА НА КЛАССЕ ИСТОКООБРАЗНО ПРЕДСТАВИМЫХ РЕШЕНИЙ

Тоомас Кихо

В статье изучается вопрос об оптимальном по точности выбора параметра регуляризации α в итерированных вариантах метода Тихонова при решении операторного уравнения с линейным непрерывным оператором в гильбертовых пространствах. Найдены классы истокообразно представимых решений, на которых оптимален по точности m–шаговый итерированный вариант метода Тихонова с подходящим образом выбранным параметром.

Пусть H,F - гильбертовые пространства и $A\in\mathcal{L}(H,F)$ - линейный непрерывный оператор с незамкнутой области значений $\mathcal{R}(A)\subset F$, действующий из H в F. Рассмотрим операторное уравнение

$$Au = f, \tag{1}$$

где $f\in F$. Пусть вместо точного свободного члена f нам известен $f_\delta\in F$, такой, что $\|f-f_\delta\|\le\delta$, $\delta>0$. Приближенное решение уравнения (1) найдем m-шаговым итерированным вариантом метода Тихонова:

$$\begin{aligned} &u_\alpha = u_{m\alpha},\\ &u_{i\alpha} = (\alpha I+A^*A)^{-1}(\alpha u_{i-1,\alpha}+A^*f_\delta), \qquad i=1,\dots,m. \\ &u_{0\alpha} = u_0, \end{aligned} \tag{2}$$

Здесь $u_0\in H$ - начальное приближение, а $\alpha>0$ – параметр регуляризации. Заметим, что при $m=1$ и $u_0=0$ мы из (2) получим обыкновенный метод Тихонова

$$u_\alpha=(\alpha I+A^*A)^{-1}A^*f_\delta. \tag{3}$$

Введем класс истокообразно представимых элементов

$$M_{p\rho u_0}=\{u\in H : u-u_0=|A|^p v,\ \|v\|\le\rho\},\quad p>0,\ \rho>0.$$

Приведенный метод (2) (и (3)) укладывается в более общий класс методов. А именно, методы в виде

$$u_\alpha=\left(I-A^*Ag_\alpha(A^*A)\right)u_0+g_\alpha(A^*A)A^*f_\delta, \qquad (4)$$

где $g_\alpha(A^*A)$ - операторы, которые порождены функциями $g_\alpha(\lambda)=\frac{1}{\alpha}g(\frac{\lambda}{\alpha})$, $\lambda\geq 0$, здесь $g[0,\infty)\mapsto\mathbb{R}$ – непрерывно дифференцируемая функция, удовлетворяющая условиям

$$\sup_{\lambda\geq 0}|g(\lambda)|<\infty,$$

$$\sup_{\lambda\geq 0}\lambda^p|1-\lambda g(\lambda)|<\infty \qquad \text{при } p\in[0,p_0],\ p_0\geq 0.$$

Если $g(\lambda)=\frac{1}{\lambda}\left[1-\frac{1}{(1+\lambda)^m}\right]$, то придем к методу (2) с $p_0=m$ (см. [1] с.31).

Метод (4) называем оптимальным по точности, если

$$\sup_{u\in M_{\rho\rho u_0},f_\delta\in F:\|Au-f_\delta\|\leq\delta}\|u_\alpha-u\|\leq\rho^{\frac{1}{p+1}}\delta^{\frac{p}{p+1}}. \qquad (5)$$

Отметим, что для любого метода P (т.е. отображения P: $F\mapsto H$), вообще говоря, имеет место неравенство

$$\Delta(P,\delta)\equiv\sup_{u\in M_{\rho\rho u_0},f_\delta\in F:\|Au-f_\delta\|\leq\delta}\|u_\alpha-u\|\leq\rho^{\frac{1}{p+1}}\delta^{\frac{p}{p+1}}\equiv\omega(\delta)$$

по крайней мере для тех δ при которых $\left(\frac{\delta}{\rho}\right)^{\frac{2}{p+1}}\in\sigma(A^*A)$ (σ – спектр оператора). Таким образом разумна следующая терминология (см.[1] с.11-12; [2] с.8). Скажем, что метод P *оптимальный по точности*, если $\Delta(P,\delta)\leq\omega(\delta)$; *оптимального порядка*, если $\Delta(P,\delta)\leq c\ \omega(\delta)$, $c=\mathrm{const}>1$; *асимптотически оптимальный*, если $\varliminf_{\delta\to 0}\Delta(P,\delta)/\omega(\delta)=1$.

В [2] с.60 сформулирована теорема для оптимальности метода (4), откуда вытекает, что если при некотором $p\in(0,2m]$ функция

$$\Psi_{pm}(\lambda)\equiv\frac{(p+1)\lambda^p}{[\sqrt[m]{p+1}-1]^p(1+\lambda)^{2m}}+\frac{(p+1)[\sqrt[m]{p+1}-1][1-(1+\lambda)^{-m}]^2}{p\lambda}\leq 1 \qquad (6)$$

при всех $\lambda\geq 0$, то для приближения найденного методом (2) при

$$\alpha=\frac{1}{[\sqrt[m]{p+1}-1]^p}\left(\frac{\delta}{\rho}\right)^{\frac{2}{p+1}} \qquad (7)$$

справедлива оценка (5), т.е. метод (2) с выбором параметра (7) оптимален по точности на классе $M_{\rho\rho u_0}$, по крайней мере для тех $\delta>0$, для которых $\left(\frac{\delta}{\rho}\right)^{\frac{2}{p+1}}\in\sigma(A^*A)$; и если при некотором $\lambda_0\geq 0$

справедливо противоположное неравенство $\Psi_{pm}>1$, то метод (2) не будет оптимальным на классе $M_{p\rho u_0}$ ни при каком выборе параметра $\alpha=\alpha(\delta,M_{p\rho u_0})$. Точка $\mu_{pm}\equiv\sqrt[m]{p+1}-1$ стационарна для Ψ_{pm} и имеет место равенство

$$\Psi_{pm}(\mu_{pm})=1. \tag{8}$$

Известно также, что при $p>2m$ m-шаговый итерированный вариант метода Тихонова теряет даже оптимальный порядок (см.[2] с.61).

Исследуем неравенство (6) в более удобном виде

$$\Psi_{pm}(\lambda)=\frac{p+1}{[\sqrt[m]{p+1}-1]^p(1+\lambda)^{2m}}\left[\lambda^p+\frac{[\sqrt[m]{p+1}-1]^{p+1}}{p}\lambda B_m^2(\lambda)\right]\le 1, \tag{9}$$

где $B_m(\lambda)$ – полином степени $m-1$:

$$B_m(\lambda)=\binom{m}{1}+\binom{m}{2}\lambda+\ldots+\binom{m}{m}\lambda^{m-1}.$$

Известно, что обыкновенный метод Тихонова (3) оптимален по точности для всех $p\in(0,2]$ (см.[2] с.61). Действительно, для $m=1$ условие (9) принимает простой вид

$$\Psi_{p1}(\lambda)\equiv\frac{p+1}{p^p(1+\lambda)^2}\left[\lambda^p+p^p\lambda\right]\le 1, \qquad \lambda\ge 0,$$

справедливость которого легко проверяется при всех $p\in(0,2]$. Для $m>1$ аналитически выяснить при каких p справедливо неравенство (9) трудно. Например, при $m=2$ условие (9) принимает вид

$$\Psi_{p2}(\lambda)=\frac{p+1}{[\sqrt{p+1}-1]^p(1+\lambda)^4}\left[\lambda^p+\frac{[\sqrt{p+1}-1]^{p+1}}{p}\lambda(2+\lambda)\right]\le 1, \quad \lambda\ge 0,$$

при $m=3$

$$\Psi_{p3}(\lambda)=\frac{p+1}{[\sqrt[3]{p+1}-1]^p(1+\lambda)^6}\left[\lambda^p+\frac{[\sqrt[3]{p+1}-1]^{p+1}}{p}\lambda(3+3\lambda+\lambda^2)\right]\le 1,$$

$\lambda\ge 0$, итд.

Для малых $p>0$ функция Ψ_{pm} $(m=1,2,\ldots)$ удовлетворяет условию (9), так как имеет точно одну стационарную точку $\mu_{pm}\equiv\sqrt[m]{p+1}-1$ в которой (в силу (8)) $\Psi_{pm}(\mu_{pm})=1$. Действительно, в процессе $p\to 0$ имеет место сходимость $\mu_{pm}\to 0$ и

$$\lim_{p\to 0}\Psi_{pm}(\lambda)=\frac{1}{(1+\lambda)^{2m}}\left[1+\frac{1}{m\lambda}\left((1+\lambda)^m-1\right)^2\right]\equiv\Psi_{om}(\lambda),$$

производное которого при $\lambda>0$

$$\frac{d}{d\lambda}\Psi_{om}(\lambda)=\frac{-2m}{(1+\lambda)^{2m+1}}-\frac{1}{m\lambda}\left[\frac{2m((1+\lambda)^m-1)}{(1+\lambda)^{2m+1}}+\frac{((1+\lambda)^m-1)^2}{\lambda(1+\lambda)^{2m}}\right]<0,$$

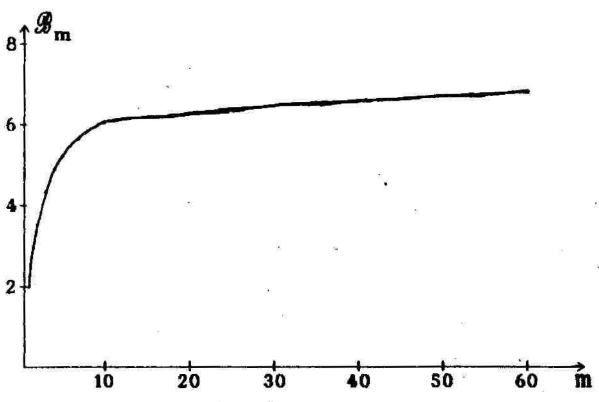
$$\Psi_{om}(0)=1,$$

иными словами, исследуемая функция монотонно убывает ($\lambda>0$) и имеет единственную стационарную точку $\mu_{pm}\to 0$. Значит, найдутся числа $\mathcal{B}_m \in (0,2m]$, такие, что , если $0<p<\mathcal{B}_m$, то выбором параметра (7) обеспечивается оптимальность по точности m-шагового метода Тихонова (2) на классе $M_{p\rho u_0}$. Более того, каждый итерационный шаг в (2) улучшает метод, расширяет пределы выбора p: $\mathcal{B}_{m+1} \ge \mathcal{B}_m \ge \mathcal{B}_1 = 2$, m=1,2,.... Рост чисел $\mathcal{B}_m$ при $m\to\infty$ ограничен – см. замечание 1 и приведенные численные значения в таблице 1.

Таблица 1

m	1	2	3	4	5	6	7	8	9	10	100	1000	∞
$\mathcal{B}_m$	2	3.58	4.38	4.88	5.23	5.48	5.67	5.83	5.95	6.05	7.00	7.11	7.12

Рис.1



Замечание 1. Утверждением найденным числам $\mathcal{B}_m$ является следующий результат (см.[2] с.62). Пусть $\|A\|^2 \le a$, $G:[0,a]\mapsto R$ непрерывная функция, $0 < G(\lambda) < \frac{2}{\lambda}$, $\lambda\in[0, a]$. Если $0<p<7.124$, то для итерационных приближений

$$u_n = u_{n-1} - G(A^*A)A^*(Au_{n-1} - f_\delta), \qquad n=1,2,..., \tag{10}$$

$\left(n=n(\delta,\rho)=\mathrm{int}\left[\frac{\ln(1+p)}{G(0)}\left(\frac{\rho}{\delta}\right)^{\frac{2}{p+1}}\right]\right)$ имеет место оценка

$$\overline{\lim_{\delta\to 0}} \sup_{u\in M_{p\rho u_0},\, f_\delta\in F:\|Au-f_\delta\|\le\delta} \|u_n - u\| / \left(\rho^{\frac{1}{p+1}} \delta^{\frac{p}{p+1}}\right) \le 1,$$

т.е. метод (10) асимптотически оптимален на классе $M_{p\rho u_0}$.

Замечание 2. Для уравнения (1) с самосопряженным неотрицательным оператором в гильбертовом пространстве H придем к итерированному варианту метода Лаврентьева, для которого оптимальность по точности изучена в [3].

Литература

1. В а й н и к к о Г. Методы решения линейных некорректно поставленных задач в гильбертовых пространствах. – Тарту: ТГУ, 1982. - 112 с.
2. В а й н и к к о Г. М., В е р е т е н н и к о в А.Ю. Итерационные процедуры в некорректных задачах. – М. : Наука, 1986. – 184 с.
3. К и х о Т. Оптимальный по точности выбор параметра в итерированных вариантах метода Лаврентьева на классе истокообразно представимых решений // Уч. зап. Тартуск. ун-та, 1988, вып. 833, с. 97–106.

OPTIMAL CHOICE OF THE PARAMETER IN THE ITERATED VERSIONS OF THE TIKHONOV METHOD ON THE SOURCE CLASSES OF SOLUTIONS

T. Kiho

Summary

Let A be a linear continuous operator from a Hilbert space H into another Hilbert space F. Let the range of the operator $\mathcal{R}(A)\subset F$ be non-closed. Consider the operator equation Au=f and the m-step iterated version of the Tikhonov method (2) for it. Introduce the source classes of solutions

$$M_{p\rho u_0}=\{u\in H : u-u_0=|A|^p v, \|v\|\le\rho\}, \quad p>0, \ \rho>0.$$

In this paper the numerical values of p and the corresponding choice of the regularization parameter α are indicated, for which method (2) is optimal on $M_{p\rho u_0}$.

СОДЕРЖАНИЕ CONTENTS