Київський національний університет імені Тараса Шевченка

Журнал обчислювальної та прикладної

2015 **МАТЕМАТИКИ** 2 (119)

Заснований у 1965 році

Адреса редакції: 03022 Київ, пр. Глушкова, 4 д
Київський національний університет імені Тараса Шевченка,
факультет кібернетики, кафедра обчислювальної математики,
тел.: (044) 259-04-36, E-mail: opmjournal@gmail.com
http://www.opmj.univ.kiev.ua

Адреса редакції серії: 79000 Львів, вул. Університетська, 1
Львівський національний університет імені Івана Франка,
Кафедра обчислювальної математики,
тел.: (032) 239-43-91, E-mail: kom@franko.lviv.ua
http://jnam.lnu.edu.ua

Затверджено Вченою радою факультету кібернетики
від 28 вересня 2015 р. (протокол № 2)

Свідоцтво про державну реєстрацію КВ 4246 від 26.05.2000

Підписано до друку 28 вересня 2015 р.

Журнал обчислювальної та прикладної математики | **2015** **№ 2 (119)** | Journal of Computational & Applied Mathematics

UDC 517.9

DYNAMICS OF TERNARY STATISTICAL EXPERIMENTS WITH EQUILIBRIUM STATE

M. L. BERTOTTI, S. O. DOVGYI, D. KOROLIOUK

Резюме. Вивчаються сценарії динаміки моделі тринарних статистичних експериментів. Важливою особливістю моделі є умова балансу і наявність стаціонарного стану рівноваги (ρ). Це дозволяє використовувати різницеві рівняння для приростів ймовірностей для вивчення динаміки моделі. Дається класифікація сценаріїв еволюції моделі, які значно відрізняються один від одного в залежності від області значень основних параметрів моделі V_0 і ρ_0.

Abstract. We study the scenarios of dynamics ternary statistical experiments model. /par An important feature of the model is a balance condition and the presence of steady state (equilibrium). This allows to use difference equations for increments probabilities to study the dynamics of the model. /par We give a classification of scenarios of the model's evolution which are significantly different one from another depending on the domain of the values of the model basic parameters V_0 and ρ_0 (see. Proposition 1).

1. Building a model

We consider *statistical experiments* (SE) with persistent linear regression [1] with additional alternatives.

The basic idea of the model construction is to choose a main factor that determines the essential state of SE, supplemented by additional alternatives in the way that, the aggregation of the principal factor and its complementary alternatives completely describe the dynamics of CE on time.

The *basic characteristic* of the main factor and of the additional alternatives are their probabilities (frequencies): P_0 of the main factor and $P_1,\ P_2$ of the additional alternatives, for which the balance condition takes place:

$$P_0 + P_1 + P_2 = 1. \tag{1}$$

The dynamics of SE characteristics is determined by a *linear regression function* which specifies the values of SE characteristics in the next stage of observation, for given value probability at the present stage.

Consider a sequence of SE characteristics values which depends on the stage of observation, or, equivalently, on a discrete time parameter $k \geq 0$:

$$P(k) := (P_0(k), P_1(k), P_2(k)) \ , \quad k \geq 0,$$

and their increments at k-th time instant:

$$\Delta P(k+1) := P(k+1) - P(k) \ , \quad k \geq 0.$$

Key words. Binary statistical experiment, persistent regression, stabilization, stochas-tic approximation, exponential statistical experiment, exponential autoregression process.

Linear regression function of increments is given by a matrix which is generated by directing action parameters:

$$\Delta P(k+1) = -\widehat{\mathbb{V}}P(k)\ , \quad k \geq 0, \tag{2}$$

where

$$\begin{aligned} &\widehat{\mathbb{V}} := [\widehat{V}_{mn}\ ; \quad 0 \leq m, n \leq 2], \\ &\widehat{V}_{mm} = 2V_m\ , \quad \widehat{V}_{mn} = -V_n\ , \quad 0 \leq n \leq 2\ , \quad n \neq m. \end{aligned} \tag{3}$$

The directing action parameters V_0, V_1, V_2 satisfy the following inequality:

$$|V_m| \leq 1\ , \quad 0 \leq m \leq 2. \tag{4}$$

An important feature of SE is the presence of steady state ρ (equilibrium), which is determined by zero of the regression function of increments :

$$\widehat{\mathbb{V}}\rho = 0, \tag{5}$$

or in scalar form:

$$\widehat{V}_m\rho := \widehat{V}_{m0}\,\rho_0 + \widehat{V}_{m1}\,\rho_1 + \widehat{V}_{m2}\,\rho_2 = 0\ , \quad 0 \leq m \leq 2. \tag{6}$$

Of course, the following balance condition takes place:

$$\rho_0 + \rho_1 + \rho_2 = 1. \tag{7}$$

Next, we consider the fluctuations probabilities relative to equilibrium value

$$\widehat{P}_m(k) := P_m(k) - \rho_m\ , \quad 0 \leq m \leq 2. \tag{8}$$

The basic assumption. SE dynamics is determined by a difference equation for the main factor probabilities $\widehat{P}_0(k)$, and by the probabilities of additional alternatives $\widehat{P}_1(k)$ and $\widehat{P}_2(k)$

$$\Delta\widehat{P}(k+1) = -\widehat{\mathbb{V}}\widehat{P}(k)\ , \quad k \geq 0, \tag{9}$$

or in scalar form:

$$\Delta\widehat{P}_m(k+1) = \widehat{V}_{m0}\,\widehat{P}_0(k) + \widehat{V}_{m1}\,\widehat{P}_1(k) + \widehat{V}_{m2}\,\widehat{P}_2(k)\ , \quad 0 \leq m \leq 2\ , \quad k \geq 0. \tag{10}$$

Also the initial values have to be fixed:

$$\widehat{P}(0) = (\widehat{P}_0(0), \widehat{P}_1(0), \widehat{P}_2(0)).$$

Remark 1. *Considering equations (5) - (6) and the balance condition (7), we have explicit formulas for equilibrium:*

$$\begin{aligned} &\rho_m = V_m^{-1}/\overline{V}\ , \quad 0 \leq m \leq 2, \\ &\overline{V} := V_0^{-1} + V_1^{-1} + V_2^{-1}, \end{aligned}$$

or in other form:

$$\begin{aligned} &\rho_0 = V_1V_2/V\ , \quad \rho_1 = V_0V_2/V\ , \ \rho_2 = V_0V_1/V, \\ &V := V_1V_2 + V_0V_2 + V_0V_1. \end{aligned} \tag{11}$$

The validity of the formulas (11) and (12) can be easily confirmed by their substitution in equations (6) - (7). This is obvious the additional condition: $V \neq 0$.

Remark 2. *The dynamics determination, by linear regression function (9) - (10), in regression model of statistic experiments, does not envolves the balance condition (1), and the equilibrium (7) with additional restrictions:*

$$0 \le P_m(k) \le 1 \ , \quad 0 \le m \le 2 \ , \quad k \ge 0 \ ; \quad 0 \le \rho_m \le 1 \ , \quad 0 \le m \le 2$$

for solutions of difference equations (12), or, equivalently (20), and equations (5) - (6) for equilibriums.

2. The model interpretation

The model of SE is constructed in several stages. First, the main factor should be chosen, characterized by probability (or frequency, concentration etc.). So there exist supplementary alternatives, whose probabilities are complement to the main factor probability. In particular, having only one alternative, the binary models of SE are considered in the works [1,2,3] (see also [4,5]). The presence of two or more alternatives brings more difficulties in the analysis of SE.

With a full set of characteristics CE, the probabilities of the main factor and of additional alternatives satisfy the balance condition (1) or, equivalently, the balance condition (9), the dynamics of the probability of the main factor P_0, as well as of supplementary factors P_1, P_2 is given by the following difference equations for the probabilities of fluctuations for all $k \ge 0$:

$$\begin{aligned} \Delta\widehat{P}_0(k+1) &= V_1\widehat{P}_1(k) + V_2\widehat{P}_2(k) - 2V_0\widehat{P}_0(k) \ , \\ \Delta\widehat{P}_1(k+1) &= V_0\widehat{P}_0(k) + V_2\widehat{P}_2(k) - 2V_1\widehat{P}_1(k) \ , \\ \Delta\widehat{P}_2(k+1) &= V_0\widehat{P}_0(k) + V_1\widehat{P}_1(k) - 2V_2\widehat{P}_2(k) \ . \end{aligned} \tag{12}$$

The increment of probabilities fluctuations of the main and supplementary factors

$$\Delta\widehat{P}_m(k+1) := \widehat{P}_m(k+1) - \widehat{P}_m(k) \ , \quad 0 \le m \le 2 \ , \quad k \ge 0,$$

is determined by the values of guide action parameters ï V_0, V_1, V_2.

Remark 3. *The fluctuations of probabilities in (7) - (8) satisfy the balance condition:*

$$\widehat{P}_0(k) + \widehat{P}_1(k) + \widehat{P}_2(k) = 0 \ , \quad k \ge 0, \tag{13}$$

and by formula (8) one has:

$$\Delta\widehat{P}_m(k) = \Delta P_m(k) \ , \quad 0 \le m \le 2 \ , \quad k \ge 0. \tag{14}$$

The equation (12) characterizes two basic principles of alternatives interaction: *stimulation* (positive terms) and *containment* (negative term).

3. The model analysis

The existence of equilibrium point for the fluctuations increments regression function (5) provides the possibility to analyze the dynamics of CE (by $k \to \infty$) in view of the possible guide parameters values which satisfy the constraint (4).

The dynamics of the main factor probability is described by *several scenarios*.

Proposition 1. *The main factor probability $P_0(k)$, $k \geq 0$, determined by the solution of the difference equation (12), as well as by the basic assumption (9), with equilibrium (11), changes with increasing $k \to \infty$ by the following scenarios:*

***Attractive equilibrium**: $V_0 > 0,\ 0 < \rho_0 < 1$:*

$$\lim_{k\to\infty} P_0(k) = \rho_0; \tag{15}$$

***Repulsive equilibrium**: $V_0 < 0,\ 0 < \rho_0 < 1$:*

$$\lim_{k\to\infty} P_0(k) = \begin{cases} 1 & npu \quad P_0(0) > \rho_0; \\ 0 & npu \quad P_0(0) < \rho_0. \end{cases} \tag{16}$$

***Dominant equilibrium**: $\rho_0 \notin (0,1),\ V_0 < 0$:*

$$\lim_{k\to\infty} P_0(k) = 1; \tag{17}$$

***Degenerate equilibrium**: $\rho_0 \notin (0,1),\ V_0 > 0$:*

$$\lim_{k\to\infty} P_0(k) = 0; \tag{18}$$

Remark 4. *Of course, the main factor dynamics scenarios can be formulated by domain of values of the guide parameters V_0, V_1, V_2.*

Remark 5. *Similar scenarios for additional alternative dynamics take place by considering the values of parameters V_1, ρ_1 або V_2, ρ_2.*

4. Annexes

	$V_0 > 0$	$V_0 < 0$
$0 < \rho_0 < 1$	*Attractive equilibrium:* $P_0(k) \to \rho_0 \ , \ k \to \infty$	*Repulsive equilibrium:* $P_0(k) \to \begin{cases} 1, & P_0(0) > \rho_0, \\ 0, & P_0(0) < \rho_0, \end{cases} \ k \to \infty$
$\rho_0 > 1$	*Repulsive degradation:* $P_0(k) \to 0 \ , \ k \to \infty$	*Attractive domination:* $P_0(k) \to 1 \ , \ k \to \infty$
$\rho_0 < 0$	*Attractive degradation:* $P_0(k) \to 0 \ , \ k \to \infty$	*Repulsive domination:* $P_0(k) \to 1 \ , \ k \to \infty$

Fig. 1. Table of scenarios

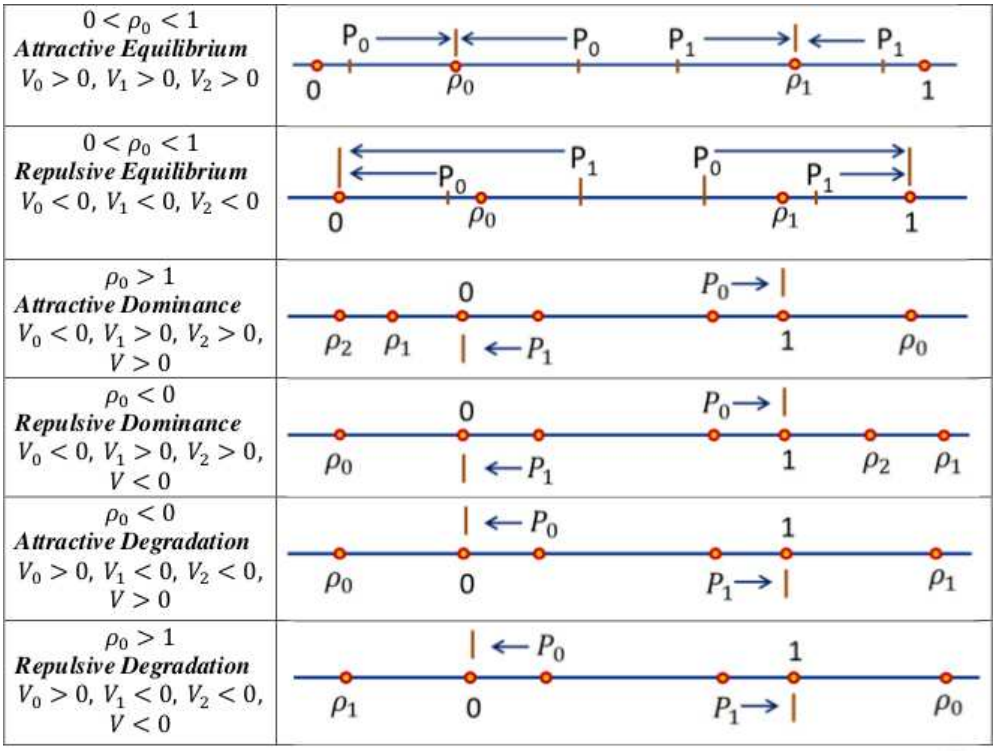


FIG. 2. Illustration of P_i limit behaviour

BIBLIOGRAPHY

1. Koroliouk D. Bi-component binary statistical experiments with persistent linear regression / D. Koroliouk // TVIMS. – 2014. – № 3. – P. 31-42.
2. Koroliouk D. Equilibrium processes in biomedical data analysis: Wright-Fisher model / D. Koroliouk, V. S. Koroliuk, N. Rosato // Cyb. and Syst. Anal. – 2014 – № 6. – P. 80-86.
3. Bertotti M. L. Stochastic behavioral model, / M. L. Bertotti, D. Koroliouk, V. S. Koroliuk // Preprint UniBZ. – 2014. (in Pubblication).
4. Arato M. Linear stochastic systems with constant coefficients. A statistical approach / M. Arato. – New-York: Springer, 1982. – 304 p.
5. Skorokhod A. V. Random Perturbation Methods with Applications in Science and Engineering / A. V. Skorokhod, F. C. Hoppensteadt, H. Salehi. – New-York: Springer, 2002. – 501 p.

M. L. BERTOTTI,
FREE UNIVERSITY OF BOZEN-BOLZANO, FACULTY OF SCIENCE AND TECHNOLOGY, PIAZZA UNIVERSITA' 5, 39100 BOZEN-BOLZANO, ITALY;
S. O. DOVGYI, D. KOROLIOUK,
INSTITUTE OF TELECOMMUNICATIONS AND GLOBAL INFORMATION SPACE UKR. ACAD. SCI., 13, CHOKOLOVSKIY BOULEVARD, KIEV, 03110, UKRAINE

Received 20.05.2015

Журнал обчислювальної та прикладної математики | **2015** **№ 2 (119)** | Journal of Computational & Applied Mathematics

UDC 517.988

INVARIANCE AND UNIQUENESS OF SOLUTIONS TO POLYNOMIAL INTERPOLATION PROBLEMS IN EUCLIDEAN SPACE

O. F. KASHPUR, V. V. KHLOBYSTOV

Резюме. В роботі розглянуто розв'язання задачі інтерполяції функції багатьох змінних в умовах недовизначеності. Одержано умови інваріантної розв'язуваності та єдиності розв'язку поставленої задачі.

Abstract. In this paper we consider solving of the interpolation problem as applied to many-variable function in the case of under-determinacy. The condition for invariant resolution and uniqueness of this problem is obtained.

1. Introduction

The fundamentals of general theory of operator's interpolation in abstract Hilbert spaces have been established in [1-3]. Then the authors also derived the conditions of invariant solvability for interpolation problems in the event when the solution is available at some or other operator's values in the nodes. The issue of convergence of interpolation processes and estimated accuracy of interpolation for the case of differential operators in Hilbert spaces are considered in [4].

Let X, Y be Hilbert spaces, μ - a Gaussian measure on X such that its first moment is equal to zero, B - the correlation operator of this measure (B belonging to trace-class ones), and $KerB = \emptyset$ [5, 6]. Assume also that Π_n be the set of operator polynomials $P_n : X \to Y$ of n-th power in the form

$$\Pi_n = \{P_n(x) : P_n(x) = L_0 + L_1 x + \cdots + L_n x^n\},$$

where $L_0 \in Y, L_k x^k = L_k(\underbrace{x, x, \ldots, x}_{k})$, and $L_k(x_1, x_2, \ldots, x_k)$ is the k-linear continuous symmetric operator form. Now introduce the scalar product on the set Π_n [2]:

$$(P_n^{(1)}, P_n^{(2)}) = \sum_{k=0}^{n} \int_X \cdots \int_X \Big((L_k^{(1)}(v_1, v_2, \ldots, v_k), L_k^{(2)}(v_1, v_2, \ldots, v_k) \Big)_Y \mu(dv_1)\mu(dv_2)\ldots\mu(dv_k),$$

where $(\cdot,\cdot)_Y$ is the scalar product in the Y-space, while $L_k^{(1)}$ and $L_k^{(2)}$ are k-linear continuous symmetric operator forms corresponding to the polynomials $P_n^{(1)}, P_n^{(2)} \in \Pi_n$ and $\|P_n\| = (P_n, P_n)^{1/2}$.

Key words. Hilbert space, Euclidean space, operator, interpolation polynom, invariance of solution.

2. Formulation and treatment of the interpolation problem in Hilbert space

For the operator $F : X \to Y$ set by its values $F(x_i)$ in the nodes $x_i, i = \overline{1,m}$ we have to find the unique operator polynomial $P_n \in \Pi_n$ that satisfies the interpolation conditions

$$P_n(x_i) = F(x_i), i = \overline{1,m}. \tag{1}$$

Introduce the following notation: $\Gamma = \| \sum_{p=0}^{n} (x_i, x_j)^p \|_{i,j=1}^{m}, 0^0 = 1, (\cdot,\cdot)$ is the scalar product in the X-space, Γ^+ is the Moore-Penrose pseudo-inverse matrix with respect to Γ, and E is identity matrix.

In [1-3], in the event of fulfillment of the necessary and sufficient conditions for solvability of operator interpolation task, such as

$$(E - \Gamma\Gamma^+)\overrightarrow{F} = \overrightarrow{0}, \overrightarrow{F} = \{F(x_i)\}_{i=1}^{m}. \tag{2}$$

the following unique interpolation polynomial of n-th power with minimal norm is constructed:

$$P_n(x) =< \overrightarrow{F}, \Gamma^+ \sum_{p=0}^{n} \{(x, x_i)^p\}_{i=1}^{m} >, \tag{3}$$

where $< \overrightarrow{\alpha}, \overrightarrow{\beta} >= \sum_{i=1}^{m} \alpha_i\beta_i$, $\alpha_i \in Y, \beta_i \in R_1$, i.e. $P_n(x)$ is a solution to the extremum task

$$\|P_n\| = \min \|Q_n\| = (<< \Gamma^+\overrightarrow{P_n}, \overrightarrow{P_n} >>)^{1/2}, Q_n \in \Pi_n^I, \overrightarrow{P_n} = (P_n(x_i))_{i=1}^{m}$$

and Π_n^I is the set of interpolation polynomial of n-th power.

We call an interpolation task invariantly solvable if it has a solution at arbitrary $\overrightarrow{F}$. Then, obviously, the matrix Γ in (2.2) has to be nonsingular. According to [7], an interpolation problem is invariantly solvable in Hilbert space if the interpolation nodes $x_i, i = \overline{1,m}$ are different and the condition

$$m \leqslant n + 1. \tag{4}$$

is met.

In practice, we often deal with approximation of many-variable functions. When such function is represented by a set of its values, one of approximation methods consists in polynomial interpolation. But there another problem arises: the conditions for existence and uniqueness of the interpolant are to be established.

In the tasks of object's identification based on its responses to input signals, of particular interest is the case when the information available is not sufficient: for example, the number of conditions is less than dimension of the space of polynomials used for seeking the solution in Euclidean space. This problem will be called underdetermined.

This work focuses on treatment of the interpolation problem as applied to many-variable functions in the case of under-determinacy, and on analysis of conditions for invariant resolution and uniqueness of the final result.

3. Solution of the interpolation problem in Euclidean spaces

To begin with, apply the above results of treatment of the interpolation problem to the case of Euclidean space E_2. Consider the interpolation of the function $f : E_2 \to R_1$ set by its values in nodes $\gamma_i = (x_i, y_i), i = \overline{1,m}$. Let us represent the solution in the form of interpolant with minimum norm:

$$P_n(x,y) =< \overrightarrow{f}, \Gamma^+ \sum_{p=0}^{n} \{(x_i x + y_i y)^p\}_{i=1}^m >, \tag{5}$$

where $\overrightarrow{f} = \{f(\gamma_i)\}_{i=1}^m, \Gamma = \| \sum_{p=0}^n (x_i x_j + y_i y_j)^p \|_{i,j=1}^m$. If inequality (2.4) holds and all nodes γ_i are different then $\Gamma^+ = \Gamma^{-1}$ (see [7]). In this work for the Euclidean space we obtain a stronger result for invertibility of the matrix Γ as compared to (2.4).

First we construct the solution to this problem based on the general interpolation theory of multivariable functions [8]. The required interpolation polynomial $P_n(x,y)$ will be written as

$$\begin{aligned} P_n(x,y) = a_{00} + a_{10}x + a_{01}y + a_{20}x^2 + a_{11}xy + a_{02}y^2 + \cdots + \\ + a_{n0}x^n + a_{n-1,1}x^{n-1}y + \cdots + a_{0n}y^n, \end{aligned} \tag{6}$$

and $a_{ik} \in R_1, i,k = \overline{0,n}$ are unknown coefficients. Denote by $p = (n+1)(n+2)/2$ the dimension of space of n-th power polynomials defined in E_2. To get the unique solution to the interpolation problem, we have to find the nodes $\gamma_i \in E_2, i = \overline{1,p}$ such that the determinant of the system of linear algebraic equations for a_{ik}

$$P_n(\gamma_i) = f(\gamma_i), i = \overline{1,p} \tag{7}$$

is always nonzero.

As shown in [9], it happens if for interpolation nodes we take the following system of points:

$$\begin{aligned} &(x_0,y_0),(x_1,y_0),\ldots,(x_{n-1},y_0),(x_n,y_0),\\ &(x_0,y_1),(x_1,y_1),\ldots,(x_{n-1},y_1)\\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ &(x_0,y_{n-1}),(x_1,y_{n-1}),\\ &(x_0,y_n),\\ &x_i \neq x_j, y_i \neq y_j \text{ as } i \neq j. \end{aligned} \tag{8}$$

Such selection of nodes gives us single-valued a_{ik}, and the interpolation polynomial (3.2) is feasible and unique.

Now apply the system of nodes (3.4) to set up the interpolant (3.1). Since the solution to the problem in this case is unique, interpolation polynomials in (3.2) and in minimum norm (3.1) are coincident. Consider next the entries of the matrix Γ

$$\sum_{p=0}^{n}(\gamma_i,\gamma_j)^p=\sum_{p=0}^{n}(\overline{x}_i\overline{x}_j+\overline{y}_i\overline{y}_j)^p=$$

$$=1+\overline{x}_i\overline{x}_j+\overline{y}_i\overline{y}_j+(\overline{x}_i\overline{x}_j)^2+2\overline{x}_i\overline{x}_j\overline{y}_i\overline{y}_j+(\overline{y}_i\overline{y}_j)^2+$$

$$+\cdots+(\overline{x}_i\overline{x}_j)^n+n(\overline{x}_i\overline{x}_j)^{n-1}\overline{y}_i\overline{y}_j+\cdots+n\overline{x}_i\overline{x}_j(\overline{y}_i\overline{y}_j)^{n-1}+(\overline{y}_i\overline{y}_j)^n,$$

where $(\overline{x}_i,\overline{y}_j)$ are the points of set (3.4). Introduce a set of vectors s_i defined as follows:

$$s_i=(1,\overline{x}_i,\overline{y}_i,\overline{x}_i^2,\sqrt{2}\overline{x}_i\overline{y}_i,\overline{y}_i^2),\ldots,\overline{x}_i^n,\sqrt{n}\overline{x}_i^{n-1}\overline{y}_i,\ldots,\sqrt{n}\overline{x}_i\overline{y}_i^{n-1},\overline{y}_i^n),\quad i=\overline{1,p} \tag{9}$$

and, in conformity to [9], are linearly independent. Then the matrix Γ takes the form of Gram's matrix

$$\Gamma=\begin{pmatrix}(s_1,s_1) & \ldots & (s_1,s_p)\\ \ldots & \ldots & \ldots\\ (s_p,s_1) & \ldots & (s_p,s_p)\end{pmatrix} \tag{10}$$

which is nonsingular. Since any subsystem of vectors (3.5) is also linearly independent and the matrix Γ is invertible, our interpolation task will be invariantly solvable and have a single solution in the form of an interpolating polynomial with minimum norm (3.1), where $\Gamma^+=\Gamma^{-1}$. Based on the above, the following theorem may be suggested.

Theorem 1. *Let the function $f:E_2\to R_1$ be set by its values $f(\gamma_i),i=\overline{1,m}$. If the interpolation nodes $\gamma_i,i=\overline{1,m}$ are so selected that the subsystem of vectors from (3.5) is linearly independent (representing, for example,a subset of points (3.4)), then an interpolation problem with two-dimensional function is invariantly solvable and has a single solution with minimum norm under the condition $m\le p$, where p is the dimension of space of polynomials in n-th power defined in E_2.*

Thus, with Theorem 3.1 taking into account, for the function $f:E_2\to R_1$ we obtained better results compared to inequality (2.4) (see [7]).

Example.Consider the derivation of an interpolational polynomial with minimum norm (3.1) of the second power $P_2(x,y)$. The interpolation nodes are selected from the set of points (3.4), so that

$$\gamma_1=(0,0),\gamma_2=(1,0),\gamma_3=(-1,0),$$
$$\gamma_4=(0,1),\gamma_5=(1,1),$$
$$\gamma_6=(0,-1)$$

Based on formula (3.5), the vectors s_i will be written as

$$s_1=(1,0,0,0,0,0),s_2=(1,1,0,1,0,0),s_3=(1,-1,0,1,0,0),$$
$$s_4=(1,0,1,0,0,1),s_5=(1,1,1,1,\sqrt{2},1),s_6=(1,0,-1,0,0,1) \tag{11}$$

Since the vectors $s_i,i=\overline{1,m}$ are linearly independent, the matrix Γ defined by formula (3.6) is invertible. So we come to the conclusion that in order to

construct the interpolant (3.1) we may select any subsystem of vectors (3.7), meaning that the interpolation problem is invariantly solvable and has a unique solution in the event when $m \leq 6$ (m is the number of nodes from set (3.4)). Compared to inequality (2.4), where $m \leq 3$, we obtain a better result.

As noted above, in practice we may encounter problems where the number of interpolation nodes and the function values in these nodes are less than p. In this case the interpolation task treated in classical manner [8] has nonunique solution.

If for solving this problem (at $m \leq p$) we use an interpolant with minimum norm from [1-3] and take the subsystem of vectors s_i from (3.5) for construction of the matrix Γ, then the solution will be invariant and unique. For our example we take $m = 4$ and the subsystem s_1, s_2, s_3, s_4 from (3.7). In this case the matrix Γ is invertible, the interpolation polynomial $P_2(\gamma)$ will be written as

$$P_2(\gamma) = P_2(x,y) =< \overrightarrow{f}, \Gamma^{-1} \sum_{p=0}^{2} \{(x_i x + y_i y)^p\}_{i=1}^{4} >= \sum_{i=1}^{4} l_i(\gamma) f(\gamma_i)$$

that satisfies the conditions $P_2(\gamma_i) = f(\gamma_i)$, where $l_i(\gamma) = l_i(x,y)$ are Lagrange fundamental polynomials of the second power, $l_i(\gamma_j) = \delta_{ij}$, δ_{ij} is the Kronecker symbol, $i, j = \overline{1,4}$, $l_1(x,y) = 1 - x^2 - 1/2y - 1/2y^2$, $l_2(x,y) = 1/2x + 1/2x^2$, $l_3(x,y) = -1/2x + 1/2x^2$, $l_4(x,y) = 1/2y + 1/2y^2$.

Now let us perform comparative analysis of the structure with two interpolants: that corresponding to the classical approach [8], and that suggested here for $m = p$. We choose the system of nodes from the set of points (3.4). In constructing the polynomial (3.2), the problem transforms into search for solutions of linear algebraic equations (3.3) with inverse matrix of general form. In the first case for the solution we use the Gauss method requiring $Q(m) = \frac{2}{3}m^3 + O(m^2)$ arithmetical operations. In the other case for constructing the polynomial (3.1) we have to define the vector

$$\Gamma^{-1} \sum_{p=0}^{n} \{(\overline{x}_i x + \overline{y}_i y)^p\}_{i=1}^{m} = z$$

which is equivalent to solving the system

$$\Gamma z = \sum_{p=0}^{n} \{(\overline{x}_i x + \overline{y}_i y)^p\}_{i=1}^{m} = l(x,y) \tag{12}$$

where $l(x,y)$ is the two-variable polynomial of n-th power. The solution to system (3.8) with its symmetric nonsingular matrix Γ will be sought by the square-root method demanding $Q(m) = \frac{1}{3}m^3 + O(m^2)$ arithmetic operations - with the constant at m^3 twice less than by the Gauss method.

Thus, when comparing the two methods for constructing the interpolation polynomial for the function $f : E_2 \to R_1$ we may conclude that when $m = p$ (m is the number of nodes, and p - dimension of the space of second-power polynomials in E_2) and the interpolation nodes selected correspond to system (3.4), then interpolants (3.1) and (3.2) are coincident, but the polynomial with

minimum norm is preferable due to less number of arithmetic operations, so that its formula is easier for applications.

If $m < p$ then for construction (3.2) under conditions (3.3) with nodes (3.4) the classic approach [8] does not ensure uniqueness of solution. On the other hand, polynomial interpolation (3.1) is invariant and unique. In fact, we have obtained a consistent formula making it possible to construct the interpolant of rather simple configuration.

The above results can be extended to the function of many variables $f : E_k \to R_1$, where E_k is k-dimensional Euclidean space. Let the solution of interpolation problem be sought in the space Π_{kn} where Π_{kn} is the space of k-variable polynomials of n-th power. Then, as noted in [8], we always can (find a system of nodes $(x_{i_1}, x_{i_2}, \ldots, x_{i_k}) \in E_k$ such that the task of interpolation of multivariable function will have a single solution while the system of vectors s_i can be written as

$$s_i = \left\{ \left(\frac{j!}{j_1! j_2! \cdots j_k!} \right)^{1/2} x_{i_1}^{j_1} x_{i_2}^{j_2} \cdots x_{i_k}^{j_k}, j_1 + j_2 + \cdots + j_k = j, 0! = 1 \right\}_{j=0}^{n}, \quad i = \overline{1, p} \tag{13}$$

where $p = (n+k)!/n!k!$. Then we may speak of generalization of Theorem 3.1.

Theorem 2. *Let the function $f : E_k \to R_1$ be given its values $f(\gamma_i), i = \overline{1, m}$. If the interpolation nodes γ_i choose so that the relevant subsystem from vectors (3.9) are linearly independent then in the space Π_{kn} interpolation problem of k-variables function with the condition $P_n(\gamma_i) = f(\gamma_i), i = \overline{1, m}$, $P_n \in \Pi_{kn}$ is invariantly solvable and its has a unique solution with minimum norm under the condition $m \leq p$, where p - the dimension of the space Π_{kn}.*

Bibliography

1. Makarov V. L. Foundations of polynomial operator interpolation theory / V. L. Makarov, V. V. Khlobystov. – Kyiv: Institute of Mathematics, National Academy of Sciences of Ukraine, 1998. – Vol. 24. (in Russian).
2. Makarov V. L. Interpolation of operators / V. L. Makarov, V. V. Khlobystov, L. A. Yanovich. – Kyiv: Naukova Dumka, 2000. (in Russian).
3. Makarov V. L. Methods of operator interpolation / V. L. Makarov, V. V. Khlobystov, L. A. Yanovich. – Kyiv: Institute of Mathematics, National Academy of Sciences of Ukraine, 2010. – Vol. 83.
4. Khlobystov V. V. On interpolation approximation of differentiable operators in Hilbert space / V. V. Khlobystov, T. N. Popovicheva // Ukrainian Mathematical Journal. – 2006. – Vol. 58, № 4. – P. 554-563. (in Russian).
5. Gikhman I. I. Theory of stochastic processes / I. I. Gikhman., A. V. Skorokhod. – Moscow: Nauka, 1971. (in Russian).
6. Yegorov A. D. Approximate methods for computation of continual integrals / A. D. Yegorov, P. I. Sobolevsky, L. A. Yanovich. – Minsk: Science and tecnics, 2010. – Vol. 24. (in Russian).
7. Chapko R. On the interpolation of a function on a bounded domain by its traces on parametric hypersurfaces / R. Chapko, C. Babenko, V. Khlobystov, V. Makarov. – // International Journal of Computer Mathematics. – 2014. – Vol. 91 (8). – P. 1673-1682.

8. Babenko K. I. Foundations of numerical analysis / K. I. Babenko // RC "Regular and chaotic dynamics". – 2002. (in Russian).
9. Berezin I. S. Methods of computations / I. S. Berezin, N. P. Zhidkov. – Moscow: Fizmatgiz, 1962. – Vol. 1. (in Russian).

O. F. Kashpur,
Taras Shevchenko National University of Kyiv,
4D, Glushkova Str., Kyiv, 03187, Ukraine;

V. V. Khlobystov
Institute of Mathematics, National Academy of Sciences,
3, Tereschenkivs'ka Str., Kyiv, 01601, Ukraine

Received 20.05.2015

Журнал обчислювальної та прикладної математики | 2015 № 2 (119) | Journal of Computational & Applied Mathematics

UDC 519.8

CONTINUOUS PROBLEMS OF OPTIMAL MULTIPLEX-PARTITIONING OF SETS WITHOUT CONSTRAINTS AND SOLVING METHODS

L. S. KORIASHKINA, A. P. CHEREVATENKO

Резюме. Розглядається неперервна лінійна задача оптимального мультиплексного розбиття множин у двох варіантах: з фіксованими центрами і з їх розміщенням. Описано методи розв'язання таких задач розбиття. Для задачі з фіксованими центрами оптимальний розв'язок знайдено аналітично у вигляді характеристичних вектор-функцій підмножин вищих порядків, що складають оптимальне мультиплексне розбиття заданої множини. Досліджено деякі властивості оптимальних мультиплексних розбиттів. Розв'язання задачі оптимального мультиплексного розбиття множини з розміщенням центрів зводиться до розв'язування скінченновимірної задачі мінімізації негладкої функції. Наведено результати розв'язання тестових задач. Продемонстрована можливість побудови діаграм Вороного вищих порядків у результаті формулювання та розв'язання неперервних задач мультиплексного розбиття множин з певними критеріями якості розбиття.

Abstract. We consider the continuous linear problem of optimal multiplex-partitioning of sets in two versions: with given coordinates of service centers or with their placing in a given region. The methods of solving such partitioning problems are described. For the problem with fixed centers the optimal solution was found analytically in the form of characteristic vector-functions of subsets of higher-order, which compose the optimal multiplex-partitioning of a given set. Some properties of optimal multiplex-partitions are investigated. The solution of the problem of optimal multiplex-partitioning of set with placing centers is reduced to the finite-dimensional problem of non-smooth function minimization. The results of the test problems are presented. We demonstrate the possibility of construction of higher order Voronoi diagrams via formulating and solving continuous problems of multiplex-partitioning of sets with some criterion of partitioning quality.

1. Introduction

The problems of optimal organization of service or manufacturing networks, including "Optimal set partitioning (OSP) problem", "Facility location problem", "Continuous Location-Allocation Problem", are actively studied over the past 50 years [1–15]. The main problem that can be solved using OSP models and methods is the arrangement of given region into several subregions served by only one service center. The criterion for choosing the optimal partition

Key words. sets partitioning of the k-th order, optimal multiplex-partitioning of set, Voronoi diagrams of higher orders, continuous problems of optimal sets partitioning, non-differentiable optimization.

may be the minimization of costs of service provision or obtainment. The majority of models of partitioning problems are discrete. In [3] it is shown that the discrete models and problems of partitioning - placement on a plane are NP-complete problems. The discrete problems of optimal sets partitioning and their solving methods are studied, in particular, in [3, 11].

The problems, in which the partitioned set is continuous, in scientific literature are called continuous problems of sets partitioning. Such problems are explored in [1, 2, 4–10]. Different formulations of continuous problems of optimal sets partitioning are presented in [14, 15]. There is described an unified approach, which underlies the methods and algorithms for solving such problems.

We present the mathematical formulations of problems of optimal partitioning of a given region into subregions, each of which covers customers that have the same k nearest service centers among N existing (or possible) centers. It is assumed that customers from each subregion can be served by any of the closest k centers.

The first mathematical models of continuous problems of optimal multiplex-partitioning of sets were presented in [16]. There was also substantiated the choice of name for a new class of partitioning problems. It was indicated that the order of partition can be pointed in the name of new partitioning problems. Similarly with computational geometry during the construction of a set of points that have the same set of k nearest centers among N existing (possible) ones the Voronoi cell of k-th order is obtained. The set of all such possible cells associated with N generator points (centers) is called Voronoi diagram of k-th order [17].

The name of a new class of problems takes into consideration the fact that the partitioning of customers (consumers) is carried out so that each subset is served by two, three or more service centers. There is an english term "duplex" (triplex») that in Russian (Ukrainian) translation means "which is designed for two (three) families", "multiplex" is a complex, compound. Thus, the name "problems of optimal multiplex -partitioning of sets" is total for all new OSP problems. Among them the problems of optimal duplex-partitioning of sets (continuous problems of optimal partitioning of sets of the second order), the problems of optimal triplex-partitioning of sets (continuous problems of optimal partitioning of sets of the third order) may be separated. For more detailed specification of multiplex-partitioning problems the words "continuous linear" can be added in its name considering accepted terminology of the theory of OSP problems [14, 15], where the first word means that the partitioned set is continuous, the second one indicates the property of functional and restrictions of the problem.

The difference between "multiplex-partitioning" and "multiple partitioning" is also denoted in [16]. In the first case the partitioning is associated with N homogeneous points called centers and the set is divided into subsets of points, which have the same set of k nearest neighbors among N centers. In the second case a regular partitioning of a given set is carried out for several times. It happens, for example, while solving multistep (multistage) OSP

problems, where service centers have different categories and customers must be partitioned for each category separately [18]. We also deal with multiple partitioning while solving multiproduct OSP problems, when each service center can provide multiple services (produce several items) and the partitioning of clients is performed for each service (product) separately [14, 15].

The methods of solvig continuous linear problems of optimal multiplex-partitioning of sets are based on the following general idea (similar to presented in [14, 15]): initial problems of optimal partitioning of sets are mathematically formulated as infinite-dimensional optimization problems and reduced to auxiliary finite-dimensional nonsmooth maximization problems or to nonsmooth maximin problems using the Lagrange functional and after that modern efficient methods of nondifferentiable optimization [19] are used to get their numerical solution. The feature of this approach for linear OSP problems is that the solution of initial infinite-dimensional optimization problems can be obtained analytically in explicit form and, at the same time, the obtained analytical expression can include parameters presented as optimal solutions to the above-mentioned auxiliary finite-dimensional optimization problems with nonsmooth objective functions.

The purpose of the article is to describe solvig methods of optimization problems of partitioning of a given region into subregions that cover customers with the same k nearest service centers among N existing (or possible) centers.

The articles [20,21] describe a unified approach to the construction of Voronoi diagrams that is based on the formulation of continuous problems of optimal partitioning of sets from an n-dimensional Euclidean space into subsets. The development of the theory of continious problems of optimal multiplex-partitioning of sets gives an opportunity to construct the Voronoi diagrams of higher orders and their different generalizations. We will demonstrate it below.

2. The mathematical formulations of continuous linear problems of optimal multiplex-partitioning of sets without constraints

Let Ω be a bounded Lebesgue measurable closed set in the space $E_n; \tau_i = (\tau_i^{(1)}, ..., \tau_i^{(n)})$ from Ω, for all $i = \overline{1, N}$, are some points, called "centers" (they can be fixed or subjected to determination).

We introduce the following notations: $\mathrm{N} = \{1, 2, ..., N\}$ is a set of all centers indeces; M (N, k) is a set of all k-elements subsets of the set N, $|\mathrm{M}\ (\mathrm{N},\ k)| = C_N^k = L$; $\sigma_l = \{j_1^l, j_2^l, ..., j_k^l\}$, $l = \overline{1, L}$ are elements of the set M(N, k). We associate each element σ_l from the set M(N, k) with some subset Ω_{σ_l} of points from Ω, $l = \overline{1, L}$. In its turn, subset Ω_{σ_l} is associated with a set of centers $\{\tau_{j_1^l}, \tau_{j_2^l}, ..., \tau_{j_k^l}\}$.

The collection of Lebesgue measurable subsets $\Omega_{\sigma_1}, \Omega_{\sigma_2}, ..., \Omega_{\sigma_L}$ from $\Omega \subset E_n$ (among which can be empty) will be called as a partition of the k-th order of the set Ω into disjoint subsets $\Omega_{\sigma_1}, \Omega_{\sigma_2}, ..., \Omega_{\sigma_L}$, if

$$\bigcup_{i=1}^{L} \Omega_{\sigma_i} = \Omega,$$

$$mes(\Omega_{\sigma_i} \cap \Omega_{\sigma_j}) = 0, \quad \sigma_i, \sigma_j \in \mathrm{M(N,}k), \quad i \neq j, \quad i, j = \overline{1, L},$$

where $mes(\cdot)$ means Lebesgue measure.

The subsets $\Omega_{\sigma_1}, \Omega_{\sigma_2}, ..., \Omega_{\sigma_L}$ of the set Ω we call as subsets of the k-th order of this set. Suppose $\Sigma_{\Omega}^{N,k}$ is a class of all possible **partitions of the k-th order** of the set Ω into disjoint subsets $\Omega_{\sigma_1}, \Omega_{\sigma_2}, ..., \Omega_{\sigma_L}$:

$$\Sigma_{\Omega}^{N,k} = \Big\{\overline{\omega} = \{\Omega_{\sigma_1}, ..., \Omega_{\sigma_L}\} : \bigcup_{i=1}^{L} \Omega_{\sigma_i} = \Omega;$$

$$mes(\Omega_{\sigma_i} \cap \Omega_{\sigma_j}) = 0,\ \sigma_i, \sigma_j \in \mathrm{M(N,}k),\ i \neq j,\ i, j = \overline{1, L}\Big\}.$$

Problem A1-k: Find

$$F\Big(\{\Omega_{\sigma_1}, ..., \Omega_{\sigma_L}\}\Big) \to \min_{\{\Omega_{\sigma_1}, ..., \Omega_{\sigma_L}\} \in \Sigma_{\Omega}^{N,k}},$$

$$F\Big(\{\Omega_{\sigma_1}, ..., \Omega_{\sigma_L}\}\Big) = \sum_{l=1}^{L} \int_{\Omega_{\sigma_l}} \sum_{i \in \sigma_l} (c(x, \tau_i)/w_i + a_i)\rho(x)dx$$

where $x = (x^{(1)}, ..., x^{(n)}) \in \Omega$; $\tau^N = (\tau_1, ..., \tau_i, ..., \tau_N) \in \Omega^N$, coordinates $\tau_i^{(1)}, ..., \tau_i^{(n)}$ of a center $\tau_i, i = \overline{1, N}$, are fixed; functions $c(x, \tau_i)$ are bounded defined on $\Omega \times \Omega$ measurable at x for any fixed $\tau_i = (\tau_i^{(1)}, ..., \tau_i^{(n)})$ from Ω for all $i = \overline{1, N}$; $\rho(x)$ is bounded measurable function integral on the set Ω ; $w_i > 0,\ a_i \geq 0,\ i = \overline{1, N}$, are given numbers.

The partition of the k-th order $\overline{\omega}^* = \{\Omega_{\sigma_1}^*, ..., \Omega_{\sigma_L}^*\}$ of $\Omega \subset E_n$ that affords minimum to the functional F, is called **optimal solution of the problem A1-k.**

If the centers $\tau_i, i = \overline{1, N}$ in the problem **A1-k** are not fixed in advance and there are some centers to be placed in a given set $\Omega \subset E_n$ along with finding its partition of the k-th order $\overline{\omega}^* = \{\Omega_{\sigma_1}^*, \Omega_{\sigma_2}^*, ..., \Omega_{\sigma_L}^*\}$, then we will have a new problem of optimal multiplex-partitioning of sets.

Problem A2-k: Find

$$\min_{\{\Omega_{\sigma_1}, ..., \Omega_{\sigma_L}\} \in \Sigma_{\Omega}^{N,k},\ \{\tau_1, ..., \tau_N\} \in \Omega^N} F\Big(\{\Omega_{\sigma_1}, ..., \Omega_{\sigma_L}\}, \{\tau_1, ..., \tau_N\}\Big),$$

where

$$F(\overline{\omega}, \tau^N) = F\Big(\{\Omega_{\sigma_1}, ..., \Omega_{\sigma_L}\}, \{\tau_1, ..., \tau_N\}\Big) = \\ = \sum_{l=1}^{L} \int_{\Omega_{\sigma_l}} \sum_{i \in \sigma_l} (c(x, \tau_i)/w_i + a_i)\rho(x)dx, \tag{1}$$

all functions and parameters are the same as in the problem **A1-k**; coordinates $\tau_i^{(1)}, ..., \tau_i^{(n)}$ of the centers $\tau_i, i = \overline{1, N}$, are unknown in advance.

An allowable pair $(\overline{\omega}^*, \tau_*^N) = \Big(\{\Omega_{\sigma_1}^*, \Omega_{\sigma_2}^*, ..., \Omega_{\sigma_L}^*\}, \{\tau_1^*, \tau_2^*, ..., \tau_N^*\}\Big)$ that affords minimum to the functional 1 is called **optimal solution of the problem A2-k.**

3. The solving method of the problems of optimal multiplex-partitioning of sets with fixed centers

By analogy with solving method of continuous linear OSP problems [14] first we write the initial problem **A1**-k as a problem of infinite-dimensional mathematical programming with Boolean variables.

Let $\overline{\omega} = \{\Omega_{\sigma_1}, ..., \Omega_{\sigma_l}, ..., \Omega_{\sigma_L}\}$ is some partition of the k-th order of the set Ω. For each point $x \in \Omega_{\sigma_l}, l = \overline{1, L}$, we introduce LN-dimensional vector $\lambda^l(x) = (\lambda_1^l(x), ..., \lambda_N^l(x))$, which coordinates are determined as follows:

$$\lambda_i^l(x) = \begin{cases} 1, & x \in \Omega_{\sigma_l} \ \& \ i \in \sigma_l, \\ 0, & \textit{in the other cases} \end{cases} \qquad i = \overline{1, N}, l = \overline{1, L}, \tag{2}$$

where $\sigma_l \in$ M(N, k), $\sigma_l = \{j_1^l, j_2^l, ..., j_N^l\}$ is the set of centers $\tau_{j_1^l}, \tau_{j_2^l}, ..., \tau_{j_k^l}$ indeces associated with a subset Ω_{σ_l}. Using these functions we introduce characteristic functions of the subsets Ω_{σ_l}, $l = \overline{1, L}$, forming the partition of the k-th order of the set Ω:

$$\chi_l(x) = \begin{cases} 1, & x \in \Omega_{\sigma_l}, \\ 0, & x \in \Omega \setminus \Omega_{\sigma_l}, \end{cases} \quad \Leftrightarrow \quad \chi_l(x) = \prod_{i=l, \ i \in \sigma_l}^{N} \lambda_i^l(x), l = \overline{1, L},$$

Therefore, the vector-function $\lambda^l(x) = (\lambda_1^l(x), ..., \lambda_N^l(x))$ defined on the set Ω with coordinates matched to 2 will be called as characteristic vector-function of the subset Ω_{σ_l} included into the partition of the k-th order of Ω (by analogy with the way as characteristic vector for a subset of a finite set in discrete mathematics is given).

Let us rewrite the problem **A1**-k in terms of characteristic functions of subsets that form the partition of the k-th order of the set Ω.

Problem B1-k. Find $\min\limits_{\lambda(\cdot) \in \Gamma_0^k} I(\lambda(\cdot))$,

$$I(\lambda(\cdot)) = \int\limits_{\Omega} \sum_{l=1}^{L} \Big(\sum_{i=1}^{N} (c(x, \tau_i)/w_i + a_i) \lambda_i^l(x) \Big) \rho(x) dx$$

$$\Gamma_0^k = \Big\{ \lambda(x) = (\lambda^1(x), ..., \lambda^l(x), ..., \lambda^L(x)) : \\ \lambda^l(x) = (\lambda_1^l(x), ..., \lambda_N^l(x)); \ \lambda_i^l(x) = 0 \vee 1, \\ i = \overline{1, N}, \ \sum_{i=1}^{N} \lambda_i^l(x) = k, \ l = \overline{1, L} \text{ a. e. for } x \in \Omega \Big\};$$

$\tau^N = (\tau_1, ..., \tau_N) \in \underbrace{\Omega \times ... \times \Omega}_{N} = \Omega^N$ is given vector.

Along with the problem **B1**-k we will consider the corresponding problem with variable values $\lambda_i^l(\cdot), \ i = \overline{1, N}, \ l = \overline{1, L}$, from segment $[0; 1]$.

Problem C1-k. Find $\min\limits_{\lambda(\cdot)\in\Gamma_1^k} I(\lambda(\cdot))$,

$$I(\lambda(\cdot)) = \int\limits_{\Omega} \sum_{l=1}^{L} \Big(\sum_{i=1}^{N} (c(x,\tau_i)/w_i + a_i)\lambda_i^l(x) \Big) \rho(x)dx$$

where

$$\Gamma_1^k = \Big\{ \lambda(x) = (\lambda^1(x), ..., \lambda^l(x), ..., \lambda^L(x)) : \lambda^l(x) = (\lambda_1^l(x), ..., \lambda_N^l(x)); 0 \le \lambda_i^l(x) \le 1, \; i = \overline{1,N}, \; \sum_{i=1}^{N} \lambda_i^l(x) = k, \; l = \overline{1,L} \text{ a. e. for } x \in \Omega \Big\};$$

$\tau^N = (\tau_1, ..., \tau_N) \in \Omega^N$ is given vector.

Obviously, $\Gamma_0^k \subset \Gamma_1^k$. It is easy to show that the set Γ_1^k is bounded closed convex set from the Hilbert space $L_2^{LN}(\Omega)$ with the norm

$$\|\lambda(\cdot)\| = \Big(\int\limits_{\Omega} \sum_{l=1}^{L} \sum_{n=1}^{N} [\lambda_i^l(x)]^2 \Big)^{1/2}.$$

The space $L_2^{LN}(\Omega)$ is reflexive. Fig. 1 depicts elements of the set Γ_1^2 corresponding to one of the points $x \in \Omega$.

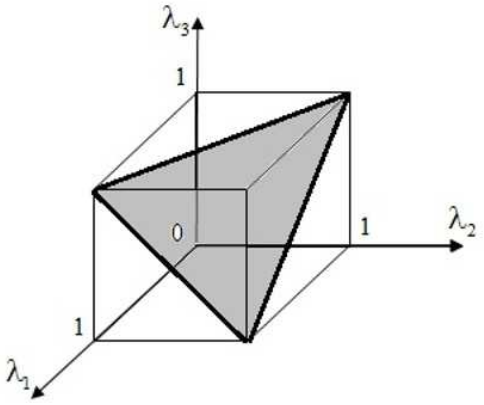


FIG. 1. The element of the set Γ_1^2 corresponding to each point $x \in \Omega$

The functional $I(\lambda(\cdot))$ is linear continuous about $\lambda(\cdot)$ on Γ_1^k at any fixed $\tau^N \in \Omega^N$.

The following statements are true.

Statement 3.1. At any fixed $\tau^N \in \Omega^N$ bounded closed convex set Γ_1^k from the Hilbert space $L_2^{LN}(\Omega)$ is slightly compact and contains at least one extreme point.

Statement 3.2. There is at least one extreme point among the set Γ_1^k of points, in which linear on $\lambda(\cdot)$ functional $I(\lambda(\cdot))$ reaches its minimum about $\lambda(\cdot)$ on the set Γ_1^k at any fixed $\tau^N \in \Omega^N$.

Statement 3.3. The extreme points of the set Γ_1^k are characteristic functions of the subsets of the k-th order $\Omega_{\sigma_1}, ..., \Omega_{\sigma_L}$ that form a partition of the k-th order of the set Ω at any fixed $\tau^N \in \Omega^N$.

The convex (linear) continuous functional $I(\lambda(\cdot))$ reaches its lower bound on a closed bounded convex set Γ_1^k from the Hilbert space $L_2^{LN}(\Omega)$ by the generalized Weierstrass theorem. Consequently, the problem **C1**-k has a solution.

Thus, there is at least one extreme point of Γ_1^k among the set of optimal solutions of the problem **C1**-k, and extreme points of Γ_1^k are characteristic functions of subsets of the k-th order $\Omega_{\sigma_1}, ..., \Omega_{\sigma_L}$ forming a partition of k-th order of the set Ω. A set of optimal solutions of the problem **C1**-k contains optimal solution of the problem **B1**-k. That is the solution of the last one reduces to the solution of the problem **C1**-k.

For the problem **C1**-k we form the Lagrange functional that includes restrictions $\sum\limits_{i=1}^{N} \lambda_i^l(x) = k$, $l = \overline{1, L}$:

$$W(\lambda(\cdot), \psi_0(\cdot)) =$$

$$= \int\limits_{\Omega} \sum_{l=1}^{L} \Big[\sum_{i=1}^{N} [c(x, \tau_i)/w_i + a_i] \rho(x) \lambda_i^l(x) + \psi_0^l(x) \Big(\sum_{i=1}^{N} \lambda_i^l(x) - k \Big) \Big] dx =$$

$$= \int\limits_{\Omega} \sum_{l=1}^{L} \Big\{ \sum_{i=1}^{N} [(c(x, \tau_i)/w_i + a_i) \rho(x) + \psi_0^l(x)] \lambda_i^l(x) - k \psi_0^l(x) \Big\} dx.$$

The functional $W(\lambda(\cdot), \psi_0(\cdot))$ is determined on the Cartesian product $\Lambda \times \Phi$, where

$$\Lambda = \{\lambda(\cdot) \in L_2^{LN}(\Omega) : \ 0 \le \lambda_i^l(x) \le 1, \ \forall x \in \Omega, \ i = \overline{1, N}, \ l = \overline{1, L}\};$$

$$\Phi = \{(\psi_0^1(\cdot), \psi_0^2(\cdot), ..., \psi_0^L(\cdot)) : \ \psi_0^l(\cdot) \in L_2(\Omega), \ l = \overline{1, L}\}.$$

The pair $(\hat{\lambda}(\cdot), \hat{\psi}_0(\cdot))$ is called **Lagrange functional** $W(\lambda(\cdot), \psi_0(\cdot))$ **saddle point** on set $\Lambda \times \Phi$, if $\forall \lambda(\cdot) \in \Lambda, \forall \psi_0(\cdot) \in \Phi$ the following inequality holds

$$W(\hat{\lambda}(\cdot), \psi_0(\cdot)) \le W(\hat{\lambda}(\cdot), \hat{\psi}_0(\cdot)) \le W(\lambda(\cdot), \hat{\psi}_0(\cdot)).$$

For each $x \in \Omega$ we introduce a function about $(LN + L)$ variables:

$$Q(\lambda(x), \psi_0(x)) = \sum_{l=1}^{L} \Big\{ \sum_{i=1}^{N} [(c(x, \tau_i)/w_i + a_i) \rho(x) + \psi_0^l(x)] \lambda_i^l(x) - k \psi_0^l(x) \Big\},$$

determined on the Cartesian product of sections $\Lambda_x \times \Phi_x$ of the sets Λ and Φ at $x \in \Omega$.

It is easy to prove, that in order to the admissible pair $\Big(\hat{\lambda}(\cdot), \hat{\psi}_0(\cdot)\Big) \in \Lambda \times \Phi$ would be a saddle point of Lagrange functional $W(\lambda(\cdot), \psi_0(\cdot))$, it is necessary

and sufficient for the following equality to be hold: a. e. for $x \in \Omega$

$$Q(\hat{\lambda}(x), \hat{\psi}_0(x)) = \max_{\psi_0(x)\in\Phi_x} \min_{\lambda(x)\in\Lambda_x} Q(\lambda(x), \psi_0(x)).$$

That means, that for each fixed $x \in \Omega$ the pair $(\hat{\lambda}(x), \hat{\psi}_0(x))$ must form the saddle point of function $Q(\lambda(x), \psi_0(x))$ on the set $\Lambda_x \times \Phi_x$.

Let x is arbitrary fixed point of set Ω. Because of the separability of function $Q(\lambda(x), \psi_0(x))$ about its parameters the following equality holds:

$$\max_{\psi_0(x)\in\Phi_x} \min_{\lambda(x)\in\Lambda_x} Q(\lambda(x), \psi_0(x)) =$$

$$= \max_{\psi_0(x)\in\Phi_x} \min_{\lambda(x)\in\Lambda_x} \sum_{l=1}^{L} \Big\{ \sum_{i=1}^{N} [(c(x,\tau_i)/w_i + a_i)\rho(x) + \psi_0^l(x)]\lambda_i^l(x) - k\psi_0^l(x) \Big\} =$$

$$= \sum_{l=1}^{L} \max_{\psi_0(x)\in\Phi_x} \Big\{ \sum_{i=1}^{N} \min_{0\le\lambda_i^l(x)\le 1} [(c(x,\tau_i)/w_i + a_i)\rho(x) + \psi_0^l(x)]\lambda_i^l(x) - k\psi_0^l(x) \Big\}.$$

The point $(\hat{\lambda}(x), \hat{\psi}_0(x))$ will be a saddle point for function $Q(\lambda(x), \psi_0(x))$ on the set $\Lambda_x \times \Phi_x$ then and only then, when the following conditions are performed:

1) $Q(\hat{\lambda}(x), \hat{\psi}_0(x)) = \min_{\lambda(x)\in\Lambda_x} Q(\lambda(x), \hat{\psi}_0(x))$;

2) $\frac{\partial Q(\hat{\lambda}(x), \hat{\psi}_0(x))}{\partial \psi_0^l} = 0 \quad \Leftrightarrow \quad \sum_{i=1}^{N} \hat{\lambda}_i^l(x) - k = 0 \quad \forall l = \overline{1, L}.$

The function $Q(\lambda(x), \psi_0(x))$ gets a minimum value at arbitrary fixed vector $\psi_0(x)$ in all admissible vectors $\lambda(x) \in \Lambda_x$,

$$\Lambda_x = \{\lambda = (\lambda_1^1, ..., \lambda_N^1, ..., \lambda_1^L, ..., \lambda_N^L) : \ 0 \ \le \ \lambda_i^l \ \le \ 1, i = \overline{1, N}, \ l = \overline{1, L}\},$$

in the point $\hat{\lambda}(x)$, which components are calculated by the formula: for each $i = \overline{1, N}, \ l = \overline{1, L}$

$$\hat{\lambda}_i^l(x) = \begin{cases} 1, & \text{if } (c(x,\tau_i)/w_i + a_i)\rho(x) + \psi_0^l(x) < 0, \\ 0, & \text{if } (c(x,\tau_i)/w_i + a_i)\rho(x) + \psi_0^l(x) > 0, \\ \alpha \in [0;1], & \text{if } (c(x,\tau_i)/w_i + a_i)\rho(x) + \psi_0^l(x) = 0. \end{cases} \tag{3}$$

Taking into account the fact that among all solutions of the problem **C1**-k we are interested only in those that are extreme points of the feasible set of problem's solutions, then because of an arbitrary choice of value $\alpha \in [0;1]$ for equalities we can assume that a particular case of 3 is the following formula: for each $i = \overline{1, N}, \ l = \overline{1, L}$

$$\hat{\lambda}_i^l(x) = \begin{cases} 1, & \text{if } (c(x,\tau_i)/w_i + a_i)\rho(x) + \psi_0^l(x) < 0, \\ 0, & \text{if } (c(x,\tau_i)/w_i + a_i)\rho(x) + \psi_0^l(x) > 0, \\ 0 \vee 1, & \text{if } (c(x,\tau_i)/w_i + a_i)\rho(x) + \psi_0^l(x) = 0. \end{cases} \tag{4}$$

At $\psi_0(x) = \hat{\psi}_0(x)$ coincidently with constraints 3, including 4, the following equalities are performed:

$$\sum_{i=1}^{N} \hat{\lambda}_i^l(x) - k = 0 \quad \forall l = \overline{1, L}. \tag{5}$$

It follows from 5 that for each fixed $l = \overline{1, L}$ among components of the vector $\hat{\lambda}^l(x) = (\hat{\lambda}_1^l(x), ..., \hat{\lambda}_N^l(x))$ in 4 exactly k components must be equal to 1. Let $\sigma_l = \{j_1^l, j_2^l, ..., j_k^l\}$ is the index set, for which the following inequalities hold:

$$\begin{gathered}(c(x, \tau_{j_m^l})/w_{j_m^l} + a_{j_m^l})\rho(x) + \hat{\psi}_0^l(x) \le 0,\ j_m^l \in \sigma_l, m = \overline{1, k}; \\ (c(x, \tau_i)/w_i + a_i)\rho(x) + \hat{\psi}_0^l(x) \ge 0,\ i \in \mathrm{N} \setminus \sigma_l. \end{gathered} \tag{6}$$

Then 4 can be written in the next form:

$$\hat{\lambda}_i^l(x) = \begin{cases} 1, & \text{if } i \in \sigma_l, \\ 0, & \text{if } i \in \mathrm{N} \setminus \sigma_l. \end{cases} \tag{7}$$

And thus, because of the arbitrary choice of point $x \in \Omega$ and number $l = \overline{1, L}$ the formula 7 determines the value of the characteristic vector-function of the subset Ω_{σ_l} of the k-th order in the point $x \in \Omega$ associated with a set of centers $\{\tau_{j_1^l}, \tau_{j_2^l}, ..., \tau_{j_k^l}\}$. That means that the formula 6 and 7 indicate the conditions of appurtenance of points x to the subset of the k-th order $\Omega_{\sigma_l},\ l = \overline{1, L}$.

We consider the system of inequalities :

$$\begin{cases} (c(x, \tau_{j_1^l})/w_{j_1^l} + a_{j_1^l})\rho(x) + \hat{\psi}_0^l(x) \le 0, \\ (c(x, \tau_{j_2^l})/w_{j_2^l} + a_{j_2^l})\rho(x) + \hat{\psi}_0^l(x) \le 0, \\ ... \\ (c(x, \tau_{j_k^l})/w_{j_k^l} + a_{j_k^l})\rho(x) + \hat{\psi}_0^l(x) \le 0, \\ -(c(x, \tau_i)/w_i + a_i)\rho(x) - \hat{\psi}_0^l(x) \le 0,\ i \in \mathrm{N} \setminus \sigma_l. \end{cases} \tag{8}$$

The system 8 is solvable, since the problem **C1**-k (as well as **B1**-k) has a solution. Summing in 8 each of the first k inequalities with each i-th inequality from the group $\mathrm{N} \setminus \sigma_l$ we obtain the following expressions: for each $l = \overline{1, L}$ and $j \in \sigma_l$

$$(c(x, \tau_j)/w_j + a_j)\rho(x) \le (c(x, \tau_i)/w_i + a_i)\rho(x), \quad \forall i \in \mathrm{N} \setminus \sigma_l.$$

Under the assumption that $\rho(x) \ge 0$ almost everywhere for $x \in \Omega$ we can write the formula for calculation of the characteristic functions of subsets of the k-th order $\Omega_{\sigma_l}^*,\ l = \overline{1, L}$ that form an optimal multiplex-partitioning of Ω as follows:
for each $l = \overline{1, L}$ the point x belongs to $\Omega_{\sigma_l}^*$, if the following inequalities hold

$$c(x, \tau_j)/w_j + a_j \le c(x, \tau_i)/w_i + a_i, \quad \forall j \in \sigma_l \text{ and } \forall i \in \mathrm{N} \setminus \sigma_l.$$

Thus, the following theorem is true.

Theorem 1. *In order a possible partition of the k-th order*

$$\overline{\omega}^* = \{\Omega^*_{\sigma_1}, ..., \Omega^*_{\sigma_L}\} \in \Sigma^{N,k}_{\Omega}$$

of the set Ω *is optimal for problem* ***A1****-k, it is necessary to fulfill an inequalities a.e. for* $x \in \Omega^*_{\sigma_l}$

$$c(x, \tau_j)/w_j + a_j \le c(x, \tau_i)/w_i + a_i, \quad \forall j \in \sigma_l \text{ and } \forall i \in N \setminus \sigma_l, \ l = \overline{1, L}. \quad (9)$$

Corollary 1. *Let in the problem* ***A1****-k function* $\rho(x) \ge 0$ *a.e. for* $x \in \Omega$, $\overline{\omega}^* = \{\Omega^*_{\sigma_1}, ..., \Omega^*_{\sigma_L}\} \in \Sigma^{N,k}_{\Omega}$ *is optimal partition, points* $x \in \Omega$ *belong to the boundary between the non-empty subsets of the k-th order* $\Omega^*_{\sigma_m}$ *and* $\Omega^*_{\sigma_l}$, $(m \ne l; m, l = \overline{1, L})$. *Then there is a subset of indices* $\zeta = \{j_1, ..., j_r\}$, $1 \le r < k$ *such that* $(\zeta \subset \sigma_l) \& (\zeta \subset \sigma_m)$ *and for each* $j \in \sigma_l \setminus \zeta$ *and* $i \in \sigma_m \setminus \zeta$ *the equality sign in 9 is achieved, i.e.:*

$$c(x, \tau_j)/w_j + a_j = c(x, \tau_i)/w_i + a_i, \quad \forall j \in \sigma_l \setminus \zeta \text{ and } \forall i \in \sigma_m \setminus \zeta. \quad (10)$$

Proof. Let $x \in \Omega$ is arbitrary fixed point, which belongs to the boundary between the non-empty subsets of the k-th order $\Omega^*_{\sigma_m}$ and $\Omega^*_{\sigma_l}$, $(m \ne l; m, l = \overline{1, L})$. Because of $x \in \Omega^*_{\sigma_m}$ the following inequalities system has a solution:

$$c(x, \tau_j)/w_j + a_j \le c(x, \tau_i)/w_i + a_i, \quad \forall j \in \sigma_m \text{ and } \forall i \in \mathrm{N} \setminus \sigma_m,$$

and by the fact that $x \in \Omega^*_{\sigma_m}$ the following inequalities are true:

$$c(x, \tau_j)/w_j + a_j \le c(x, \tau_i)/w_i + a_i, \quad \forall j \in \sigma_l \text{ and } \forall i \in \mathrm{N} \setminus \sigma_l.$$

It follows that

$$c(x, \tau_j)/w_j + a_j \le c(x, \tau_i)/w_i + a_i, \quad \forall j \in \sigma_m \cap \sigma_l, \ \forall i \in \mathrm{N} \setminus \sigma_m;$$

$$c(x, \tau_j)/w_j + a_j \le c(x, \tau_i)/w_i + a_i, \quad \forall j \in \sigma_m \cap \sigma_l, \ \forall i \in \mathrm{N} \setminus \sigma_l.$$

Let $\zeta = \sigma_m \cap \sigma_l$, $\zeta = \{j_1, ..., j_r\}$, $1 \le r < k$. Then $\forall p \in \sigma_l \setminus \zeta \quad c(x, \tau_p)/w_p + a_p \le c(x, \tau_i)/w_i + a_i \quad \forall i \in \mathrm{N} \setminus \sigma_l$.

On the other hand, since $p \in \mathrm{N} \setminus \sigma_m$, then $\forall i \in \sigma_m \quad c(x, \tau_p)/w_p + a_p \ge c(x, \tau_i)/w_i + a_i$, among them all indexes $i \in \sigma_m \setminus \zeta$. And thus, $\forall i \in \sigma_m \setminus \zeta$ and $\forall p \in \sigma_l \setminus \zeta$ the following double inequality is true:

$$c(x, \tau_i)/w_i + a_i \le c(x, \tau_p)/w_p + a_p \le c(x, \tau_i)/w_i + a_i.$$

It is possible only when $\forall i \in \sigma_m \setminus \zeta$ and $\forall p \in \sigma_l \setminus \zeta$ and the equality 10 holds, i.e.:

$$c(x, \tau_p)/w_p + a_p = c(x, \tau_i)/w_i + a_i.$$

The corollary 1 is proved.

Corollary 2. *Let in the problem* ***A1****-k function* $\rho(x) \ge 0$ *a.e. for* $x \in \Omega$, $\overline{\omega}^* = \{\Omega^*_{\sigma_1}, ..., \Omega^*_{\sigma_L}\} \in \Sigma^{N,k}_{\Omega}$ *is optimal partition, points* $x \in \Omega$ *are corner points of the partition, i.e.* x *belongs to the boundary between several non-empty subsets of the k-th order* $\Omega^*_{\sigma_m}$, $m \in \{l_1, l_2, ..., l_s\}$; $1 \le l_q \le L$, $q = \overline{1, s}$; $s > 2$. *Then there is a subset of indices* $\zeta = \{j_1, ..., j_r\}$, $1 \le r < k$ *such*

that $\zeta \subset \bigcap\limits_{q=\overline{1,s}} \sigma_{l_q}$ *and for each* $i \in \sigma_{l_p} \setminus \zeta, p = 1, 2, ..., s$ *there are indices* j_p *from the set* $\sigma_{l_p} \setminus \zeta,\ \forall q \neq p,\ q = 1, 2, ..., s$, *at which equal sign in 9 is achieved, i.e.:*

$$c(x, \tau_{j_p})/w_{j_p} + a_{j_p} = c(x, \tau_i)/w_i + a_i, \quad \forall q \neq p,\ q = 1, 2, ..., s. \tag{11}$$

The proof of the Corollary 2 is analogous to the proof of the Corollary 1.

Remark 1. *The necessary condition 9 is a sufficient condition of optimality for the problem* ***A1****-k because of* $I(\lambda(\cdot))$ *linearity.*

The Figures 2a, 2b are illustrations of the validity of the Corollary 1 and Theorem 1 in the case of optimal duplex and triplex partitioning of a square area with seven centers. Hereinafter, in order not to overload the figures the subsets of the k-th order are denoted as a set of indices $\sigma_l = \{j_1^l, j_2^l, ..., j_k^l\}$ of appropriate centers.

The implementation of the Corollary 2 for optimal triplex-partitioning of a square area with the same centers can be traced on the Fig. 2c. Let us describe this Figure in details. Let the point $x \in \Omega$ is a corner point of the partition, which lies on the border between the following subsets: $\Omega^*_{\{123\}}$, $\Omega^*_{\{237\}}$, $\Omega^*_{\{357\}}$, $\Omega^*_{\{135\}}$. The intersection of all indices sets corresponding to mentioned subsets of the third order is the set $\zeta = \{3\}$. The center τ_3 is really the closest one to a fixed point x among all seven predetermined centers. The remaining centers, which indices make up a set $\bigcup\limits_{q=\overline{1,s}} \sigma_{l_q} \setminus \zeta = \{1, 2, 5, 7\}$, are in the same distance from the point x (the shortest one without taking into account the distance between the center τ_3 and x).

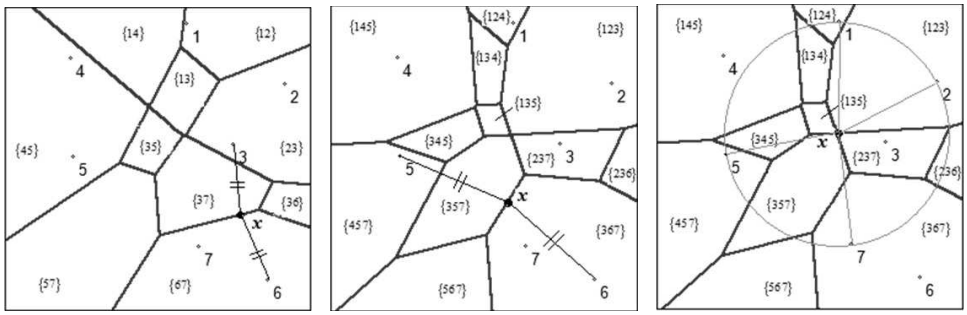


FIG. 2. Illustration of the equalities 10 (a, b) and 11 (c) for points of the boundary between subsets

Thus, from the Theorem 1 we see that the optimal solution of **B1**-k is reached on the vector-function $\lambda^*(x) = (\lambda^1_*(x), ..., \lambda^l_*(x), ..., \lambda^L_*(x))$, each component $\lambda^l(x)$ of which is calculated by the formula: a.e. for $x \in \Omega$

$$\lambda^l_{*i}(x) = \begin{cases} 1, & \text{if } c(x, \tau_i)/w_i + a_i \leq c(x, \tau_j)/w_j + a_j, \\ & \quad \text{simultaneously with } \forall i \in \sigma_l,\ j \in \mathrm{N} \setminus \sigma_l, \\ 0, & \text{in other cases} \qquad l = \overline{1, L},\ i = \overline{1, N}. \end{cases} \tag{12}$$

The functional of the problem **B1**-k at $\lambda(\cdot) = \lambda^*(\cdot)$ is noted as follows:

$$I(\lambda^*(\cdot)) = \int\limits_{\Omega} \min_{l=\overline{1,L}} \Big(\sum_{i \in \sigma_l} (c(x, \tau_i)/w_i + a_i) \Big) \rho(x) dx. \tag{13}$$

Remark 2. *Assume that for a certain point $x \in \Omega$ there are several sets of indices, for example, $\sigma_l = \{j_1^l, j_2^l, ..., j_k^l\}$ and $\sigma_q = \{j_1^q, j_2^q, ..., j_k^q\}$, under which system of inequalities 9 holds. It is possible only when these several indices sets have nonempty intersection and on the set of indices $\sigma_l \triangle \sigma_q$ ($\triangle$ is symmetric difference) an equal sign in the inequality 9 is achieved. Then the solution of the problem **C1**-k will consist of not one extreme point, but at least two ones. It is easy to see that the value of the functional 10 is the same for both extreme points. Since for the visual interpretation of the solution 12 and for the method implementation the selection of the certain extreme point (hence, the set Ω partition) is very important, then the ambiguity can be eliminated using conventional techniques: from several sets of indices $\sigma_l \triangle \sigma_q$, where $c(x, \tau_j)/w_j + a_j = c(x, \tau_i)/w_i + a_i$ is achieved, the smallest index is chosen.*

Usually, while formulating the OSP problems as a function $c(x, \tau_i)$ a particular case of Minkowski power distance $c(x, \tau_i) = \|x - \tau_i\|_p = \sqrt[p]{\sum\limits_{j=1}^{n} (x^j - \tau_i^j)^p}$ is selected: at $p = 2$ – Euclidean, at $p = 1$ – Manhattan (taxicab geometry), at $p = \infty$ – "domination" metrics (Chebyshev metrics).

The partitions of the 1-t, 2-d, 3-d order of the square area $\Omega \subset E_2$ with centers $\tau_i,\ i = 1, 2, ..., 8$ in case, when the function $c(x, \tau_i)$ in the functional 1 is Minkovsky distance at $p = 8;\ w_i = 1,\ a_i = 0,\ i = \overline{1,8}$, are presented on the Fig. 3. For each subset Ω_{σ_l} included in the multiplex partition of Ω (at $k = 2, 3$) it is defined a pair or a trio of indeces of corresponding centers. It is easy to notice that in the duplex partition only 14 (Fig. 3b) from $L = C_8^2 = 28$ subsets $\Omega_{\sigma_1}, \Omega_{\sigma_2}, ..., \Omega_{\sigma_L}$, which compose optimal Ω partition of the 2-d order, are nonempty. In the triplex partition (Fig. 3c) many subsets of the 3-rd order also were empty. The number of empty subsets included in the multiplex partition of the set depends not only on centers' location $\tau_i,\ i = \overline{1,N}$, its number, but also on the selection of metrics [16].

Remark 3. *If in the problem **A1**-k function $c(x, \tau_i)$ is Euclidean metric, $a_i = 0,\ w_i = 1,\ i = \overline{1,N};\ \rho(x) = 1\ \forall x \in \Omega$, then the optimal solution determined by vector-function $\lambda(\cdot) = \lambda^*(\cdot)$ as 12 turns out Voronoi diagram of the k-th order known in the computational geometry [17], i.e. such partition of set Ω into subsets $\Omega_1, ..., \Omega_L$ that:*

$$\bigcup_{i=1}^{L} \Omega_i = \Omega;\ mes(\Omega_i \cap \Omega_j) = 0,\ \forall i \neq j,\ i, j = \overline{1,L},$$

$$\Omega_m = \Big\{ x \in \Omega :\ \forall j \in T_m \quad c(x, \tau_j) < c(x, \tau_i),\ i \in N \setminus T_m \Big\},$$

where $T_m = \{i_1^m, i_2^m, ..., i_k^m\},\ m = \overline{1,L}$, are all possible k-element subsets of the set N of indeces.

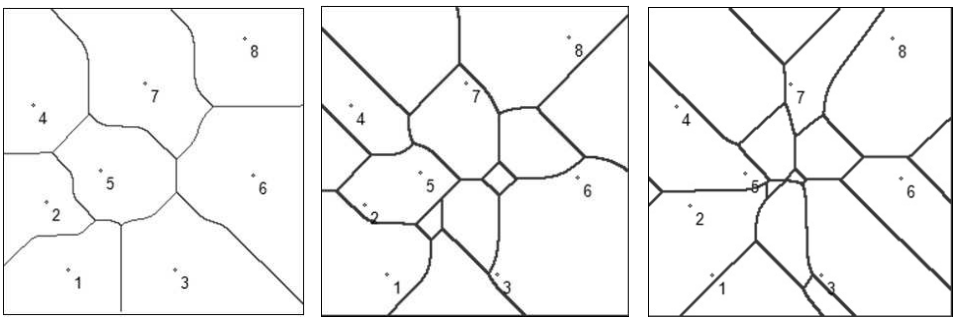


FIG. 3. The optimal partitions of k-th order of a square with 8 centers: $a-k=1;\ b-k=2;\ c-k=3$

Let us examine this fact to the problem **A1-2** (duplex OSP problem without constraints and with fixed centers) under initial data: $c(x,\tau_i)$ is Euclidean metric, $a_i=0,\ w_i=1, i=\overline{1,N};\ \rho(x)=1\ \forall x\in\Omega$. Under these conditions the formula 12 can be written as follows:

$$\lambda^l_{*i}(x)=\begin{cases}1, & \text{if} c(x,\tau_i)\le c(x,\tau_j), \text{simultaneously with} \forall i\in\sigma_l, j\in \mathrm{N}\setminus\sigma_l,\\ 0, & \text{in other cases.}\end{cases}$$

The optimal partition for this problem is shown on the Fig. 4. Suppose $x\in\Omega$ is arbitrary fixed point. Let us consider, for example, indexes sets $\sigma_q=\{7,8\}$, $\sigma_r=\{6,7\}$, $\sigma_m=\{6,8\}$. Then $\lambda^q(x)=\{0,0,0,0,0,0,1,1\}$, $\lambda^r_i(x)=0$, $\lambda^m_i=0$, $\forall i=\overline{1,N}$, because only for indexes $i\in\sigma_q$ condition is performed:

$$c(x,\tau_i)\le c(x,\tau_j),\ \forall j\in\{1,2,...,8\}\setminus\sigma_q.$$

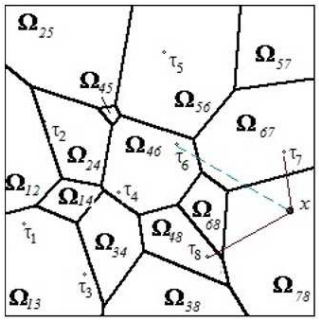


FIG. 4. The Voronoi diagram of second order

Suppose now, $\sigma_s=\{3,7\}$. On the Fig. 4 we can see, that there is no point $x\in\Omega$, for which at given centers $\tau_i, i=1,2,...,8$, ratios would be carried out:

$$c(x,\tau_3)\le c(x,\tau_j),\ \forall j\in\{1,2,...,8\}\setminus\sigma_s.$$

$$c(x,\tau_7)\le c(x,\tau_j),\ \forall j\in\{1,2,...,8\}\setminus\sigma_s.$$

Therefore, the subset of the 2-nd order Ω_{37} included in the duplex partition of the set Ω is empty.

Thus, by solving the problems **A1**-k of optimal multiplex - partitioning of sets under different parameter values of the objective functional, we can get a higher order Voronoi diagrams and their generalizations: additively weighted ($\exists$ at least one $i : a_i \neq 0, i = \overline{1,N}$), multiplicatively weighted ($\exists$ at least one $j : w_j \neq 1$, obviously $\forall j : w_j \neq 0, j = \overline{1,N}$), additively and multiplicatively weighted (simultaniously $\exists$ at least one i and at least one $j : a_i \neq 0,\ w_j \neq 1,\ i,j = \overline{1,N}$).

For the problem **B2**-k, which is equivalent to **A2**-k but written in terms of characteristic functions of subsets that constitute the partition of the k-th order of a given set Ω, the following theorem holds.

Theorem 2. *The optimal solution of the problem **B2**-k has the following form: for $i = \overline{1,N}, l = \overline{1,L}$ and almost all $x \in \Omega$*

$$\lambda^l_{*i}(x) = \begin{cases} 1, & \text{if } c(x,\tau_{*i})/w_{*i} + a_i \leq c(x,\tau_{*j})/w_{*j} + a_j,\ i \in \sigma_l,\ j \in N \setminus \sigma_l, \\ 0, & \text{in other cases,} \end{cases}$$

in the capacity of $\tau_{*1}, ..., \tau_{*N}$ the optimal solution of the problem

$$G(\tau) \to \min_{\tau^N \in \Omega^N}, \tag{14}$$

is chosen, where

$$G(\tau) = \int_\Omega \min_{\sigma_l \in M(N,,k)} \sum_{i \in \sigma_l} [c(x,\tau_i)/w_i + a_i]\rho(x)dx. \tag{15}$$

Hence, with a help of the Theorem 2 solving the continuous problem of optimal multiplex-partitioning of sets is reduced to a finite-dimensional minimization problem 14 solving with non-differentiable function 15 by any known method of non-smooth optimization [19].

In article we present only the results of solving some problems of multiplex-partitioning of sets with centers placing. Fig. 4, 5, respectively, demonstrate the results of solving the optimal duplex and triplex partitioning of square area with centers placing under parameters: $c(x,\tau_i)$ is Euclidean metric, $a_i = 0,\ w_i = 1, i = \overline{1,N};\ \rho(x) = 1\ \forall x \in \Omega$. To solve finite-dimensional problem 14 with non-differentiable function 15, the algorithm of pseudo-gradients was used with space dilatation in the direction of the difference between two successive gradients; this algorithm is close to Shor's r-algorithm [19].

Due to the fact that the Shor's r-algorithm provides a search of non-differentiable function local minimum, and the problem 14 is multiextremal, then under different initial approximations various local solutions of the problem **A2**-k can be obtained. For example, in the case of solving this problem for $N = 15,\ k = 2, 3$ except of the optimal solutions depicted in the Table 1, can be obtained the solutions presented in Fig. 5.

The identification of the properties of optimal solutions of the problem **A2**-k under certain initial data is the direction of further research.

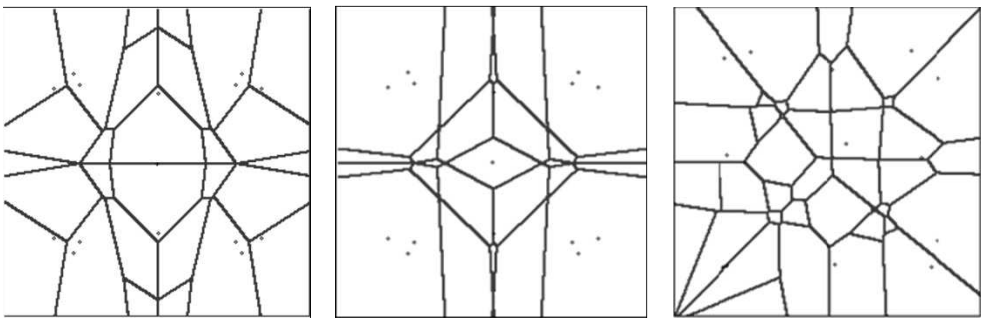


FIG. 5. The local solutions of the problem A2-k at $N = 15$: $a - k = 2$; $b, c - k = 3$

4. Conclusion

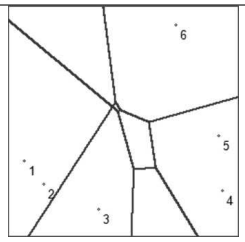
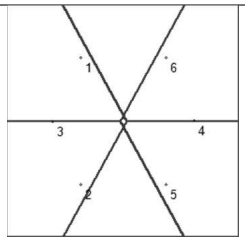
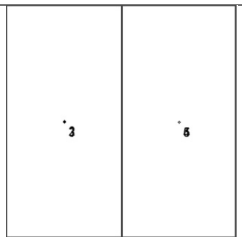
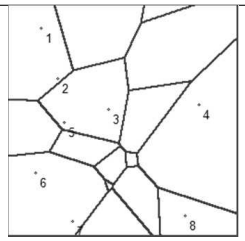
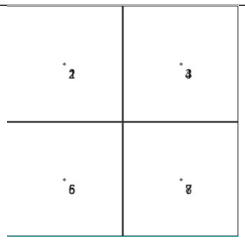
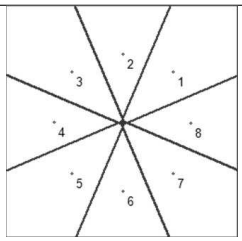
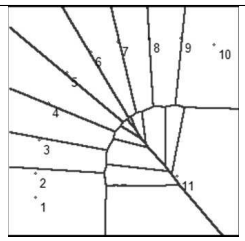
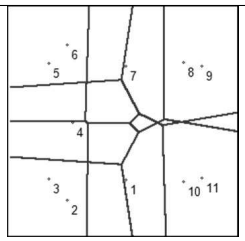
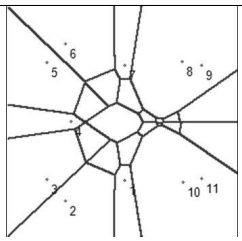
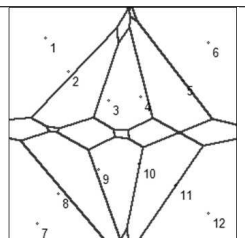
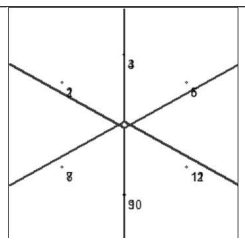
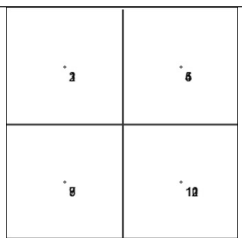
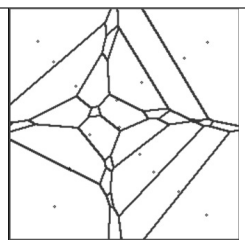
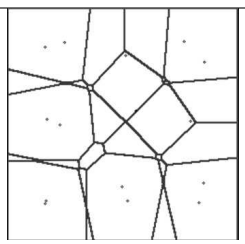
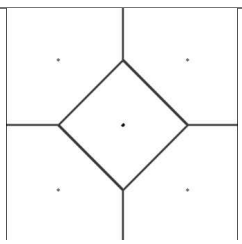
Thus, the solutions of continuous linear problems of optimal multiplex-partitioning of sets without restrictions with fixed centers and with their placing are obtained. In the latter case, the optimal solution of the multiplex partitioning problem contains unknown parameters that are obtained in the process of solving the finite-dimensional nonsmooth function minimization. The results of computational experiments are presented. The considered mathematical models can be attributed to the so-called minisum problems of partitioning-placement in terms of a quality criterion of multiplex-partitioning by analogy with the objectives of location-allocation problems of the graph theory [11,22].

We can consider a different form of the functional of multiplex-partitioning problem, for example:

$$F_1\Big(\{\Omega_{\sigma_1}, ..., \Omega_{\sigma_L}\}\Big) = \sum_{l=1}^{L} \int\limits_{\Omega_{\sigma_l}} \max_{i \in \sigma_l} (c(x, \tau_i)/w_i + a_i)\rho(x)dx.$$

In this case, the multiplex-partitioning problem is not linear and refers to the so-called minimax problems of partitioning-placement [11,22]. The development and substantiation of methods of solving these problems is one of the directions for further research in the theory of multiplex-partitioning of sets. We only note that even with this criterion the problems of optimal multiplex-partitioning of sets include as a particular case the continuous OSP problems studied in details in [14]. It is interesting to compare the solutions of the problems of optimal multiplex-partitioning of sets with different quality critera. It can be assumed that while solving the problem with placement of centers $\tau^N = (\tau_1, ..., \tau_i, ..., \tau_N) \in \Omega^N$ the functional $F_1\Big(\{\Omega_{\sigma_1}, ..., \Omega_{\sigma_L}\}\Big)$ will provide such their optimal location,that will be the solution of the optimal multiple covering of set $\Omega \subset E_n$ by circles with these centers [23].

TABL. 1. The optimal solutions of problems **A1**-k and **A2**-k

Centers number N	The partition of the k-th order of the set Ω		
	With fixed coordinates of the center	With the optimal location of centers in the set Ω	
	$k = 2$	$k = 2$	$k = 3$
6			
8			
11			
12			
15			

Bibliography

1. Francis R. L. Sufficient conditions for some optimum – property facility design / R. L. Francis // Oper. Res. – 1967. – Vol. 15., Issue 3. – P. 448-466.
2. Balas E. Set partitioning: a survey / E. Balas, M. W. Padberg // Comb. Optimiz. Lect. Summer Sch. Comb. Optimmiz. Urbino. – 1977. – Chichester e. a. – 1979. – P. 151-210.
3. Megiddo N. On the complexity of some common geometric location-problems / N. Megiddo, K. J. Supowit // SIAM Journal on Computing. – 1984. – P. 182-196.
4. Current J. Multiobjective analysis of facility location decisions / J. Current, H. Min, D. Schilling // European Journal of Operational Research. – 1990. – Vol. 49. – P. 295-307.
5. Drezner T. Replacing continuous demand with discrete demand in a competitive location model / T. Drezner, Z. Drezner // Naval Research Logistics. – 1997. – Vol. 44. – P. 81-95.
6. Bast H. The Area Partitioning Problem / H. Bast, S. Hert // 12th Canadian Conference on Computational Geometry. – 2000. – Fredericton, New Brunswick. – P. 163-171.
7. Dasci A. A continuous model for production-distribution system design. / A. Dasci, V. Verter // European Journal of Operational Research. – 2001. – Vol. 129. – P. 287-298.
8. Brimberg J. Decomposition strategies for large-scale continuous location-allocation problems / J. Brimberg, P. Hansen, N. Mladenovic // IMA Journal of Management Mathematics. – 2006. – Vol. 17. – P. 307-316.
9. Brimberg J. A continuous location-allocation problem with zone-dependent fixed cost / J. Brimberg, S. Salhi // Annals of Operations Research. – 2005. – Vol. 136. – P. 99-115.
10. Murat A. Continuous Analysis Framework for the Solution of Location-Allocation Problems with Dense Demand / A. Murat, V. Verter, G. A. Laporte. – Les Cahiers du GERAD, 2008. – 42 pp.
11. Farahani R. Z. Facility location. Concepts, models, algorithms and case studies / R. Z. Farahani, M. Hekmatfar. – Berlin, Heidelberg: Springer-Verlag, 2009. – 530 p.
12. Bischoff M. The Multi-Facility Location-Allocation Problem with Polyhedral Barriers / M. Bischoff, T. Fleischmann, K. Klamroth // http://www2.math.uni-wuppertal.de/ klamroth/publications/biflkl06.pdf
13. Durocher S. Geometric Facility Location under Continuous Motion. Bounded-Velocity Approximations to the Mobile Euclidean k-Centre and k-Median Problems / S. Durocher. – The University of British Columbia, 2006. – 260 p.
14. Kiseleva E. M. Continuous Problems of Optimal Set Partitioning: Theory, Algorithms, and Applications / E. M. Kiseleva, N. Z. Shor. – Kyiv: Naukova Dumka, 2005. – 564 p. (in Russian).
15. Kiseleva E. M. Models and Methods for Solving Continuous Problems of Optimal Set Partitioning: Linear, Nonlinear, and Dynamic Problems / E. M. Kiseleva, L. S. Koryashkina. – Naukova Dumka: Kyiv, 2013. – 606 p. (in Russian).
16. Koriashkina L. S. Extension of one class of infinite-dimensional optimization problems [in Ukrainian] / L. S. Koriashkina // Bulletin of Cherkassy University. Series "Applied Mathematics. Informatics". – 2015. – Issue 18 (351). – P. 28-38.
17. Preparata F. P. Computational Geometry: An Introduction (Texts and Monographs in Computer Science) / F. P. Preparata, M. I. Shamos. – New-York: Springer-Verlag, 1985. – 390 p.
18. Us S .A. Mathematical models of Multi-stage problems of locating production / S. A. Us, O. D. Stanina // Collected Scientific Papers on Problems of Applied Mathematics and Mathematical Modeling, Oles Honchar Dnipropetrovsk National University, Dnipropetrovsk. – 2014. – P. 258-268. (in Russian).
19. Shor N. Z. Algorithms of Nondifferentiable Optimization: Development and Application / N. Z. Shor, N. G. Zhurbenko, A. P. Likhovid, P. I. Stetsyuk // Cybernetics and Systems Analysis. – 01/2003. – Vol. 39 (4). – P. 537-548.
20. Kiseleva E. M. Theory of Continuous Optimal Set Partitioning Problems as a Universal Mathematical Formalism for Constructing Voronoi Diagrams and Their Generalizations.

I. Theoretical Foundations / E. M. Kiseleva, L. S. Koriashkina // Cybernetics and Systems Analysis. – May 2015. – Vol. 51, Issue 3. - P. 325-335.
21. Kiseleva E. M. The Theory of Continuous Optimal Set Partitioning Problems as a Universal Mathematical Formalism for Constructing the Voronoi Diagram and its Generalizations. II. Algorithms for constructing Voronoi Diagrams based on the theory of optimal partitioning of sets [in Russian] / E. M. Kiseleva, L. S. Koriashkina // Cybernetics and Systems Analysis. – May 2015. – Vol. 51, Issue 4. – P. 3-12.
22. Christofides N. Graph Theory – An Algorithmic Approach / N. Christofides. – New-York: Academic Press, 1975. – 415 p.
23. Kiseleva O. M. Constructive Algorithms for Solution Problems of Multiple Continuous Covering / O. M. Kiseleva, L. S. Koriashkina, O. O. Mykhalova // System Technologies. – 2014. – Issue 4 (93). – P. 3-13.

L. S. Koriashkina,
National Mining University,
19, Karl Marx Ave., Dnipropetrovsk, 49050, Ukraine;
A. P. Cherevatenko,
Oles Honchar Dnipropetrovsk National UniverSity,
35, Karl Marx Ave., Dnipropetrovsk, 49050, Ukraine

Received 10.06.2015

Журнал обчислювальної та прикладної математики | 2015 № 2 (119) | Journal of Computational & Applied Mathematics

UDC 519.6 + 517.983.54

REGULARIZATION OF ILL-POSED PROBLEMS IN HILBERT SPACE BY MEANS OF THE IMPLICIT ITERATION PROCESS

O. V. MATYSIK

Резюме. В роботі доведена збіжність методу з апостеріорним вибором числа ітерацій у вихідній нормі гільбертового простору в разі самоспряженого оператора, в припущенні, що похибки вносяться у праву частину рівняння. Отримано оцінку похибки методу і оцінку для апостеріорного моменту зупинки. Отримані результати можуть бути використані в теоретичних дослідженнях при розв'язуванні лінійних операторних рівнянь, а також при вирішенні прикладних некоректних задач, які зустрічаються в динаміці і кінетиці, математичній економіці, геофізиці, спектроскопії, системах повної автоматичної обробки та інтерпретації експериментів, діагностиці плазми, сейсмології, медицині.

Abstract. The article substantiates the convergence of the method with a posteriori choice of the number of iterations in the original norm of Hilbert space in case of a self-adjoint operator on the assumption of existing errors in the equation right-hand member. There has been secured error estimate of the method and the estimate of a posteriori stopping moment. The results obtained can be used in theoretic research while solving linear operator equations as well as in solving applied incorrect problems which occur in dynamics and kinetics, mathematical economics, geophysics, spectroscopy, systems of full automatic procession and interpretation of experiments, plasma diagnostics, seismology, medicine.

1. Introduction

The article calls attention to the implicit iteration method of solving ill-posed problems, described by iteration equations of type I in Hilbert space. The method represents a family of iterative schemes depending on parameter k.

The comparison of the suggested implicit method with the well-known explicit iteration method $x_{n+1,\delta} = x_{n,\delta} + \alpha\,(y_\delta - Ax_{n,\delta})$, $x_{0,\delta} = 0$ [1–8] demonstrates that the degrees of their optimum estimates coincide. The advantage of explicit methods lies in the fact that explicit methods do not require any operator inversion. They require only the calculation of the operator value on progressive approximation. In this sense the explicit method of [1–8] is preferred to the suggested implicit method. However, the recommended implicit method has a very important advantage. In the explicit method of [1–8] step α is constrained from above by the in equation $0 < \alpha \leq \dfrac{5}{4\,\|A\|}$, which may

Key words. Regularization, iteration method, incorrect problem, Hilbert space, self–conjugated and non self–conjugated approximately operator.

actually necessitate a great number of iterations. In the implicit method under consideration there are no restraints from above on the iteration parameter $b > 0$. It follows from this that the optimum estimate of the implicit method under consideration can be obtained as early as at the first iteration steps.

2. Problem statement

One deals with solving the equation

$$Ax = y \tag{1}$$

with the unbounded linear self-adjoint operator A operating in Hilbert space, on the assumption that zero belongs to the spectrum of this operator, though, generally speaking, it is not its characteristic value. According to the suggested hypotheses the problem of solving the equation (1) is incorrect. If the solution of the equation (1) really exists, then a new implicit iteration method is proposed for its finding:

$$\left(A^{2k} + B\right) x_{n+1} = Bx_n + A^{2k-1}y, x_0 = 0, k \in N, \tag{2}$$

where E is a unit operator, while B is a bounded auxiliary self-adjoint operator which is chosen for enhancing conditionality. Let's take operator $B = bE$, $b > 0$ as B. Usually the right-hand member of the equation is known with a certain accuracy δ, i.e. we know y_δ, for which $\|y - y_\delta\| \leq \delta$. That is why instead of (2) it is necessary to consider the approximation

$$\left(A^{2k} + B\right) x_{n+1,\delta} = Bx_{n,\delta} + A^{2k-1}y_\delta, x_{0,\delta} = 0, k \in N. \tag{3}$$

In what follows, the convergence of the method is understood as the statement that approximations (3) fit arbitrarily close the exact solution of the operator equation in case of the suitable choice of n and sufficiently small δ. In other words, method (3) is convergent if

$$\lim_{\delta \to 0} \left(\inf_n \|x - x_{n,\delta}\| \right) = 0.$$

If $b > 0$, the convergence for method (3) is proved in case of an accurate and approximate right-hand member of the equation, and on the assumption that the accurate solution of the equation is sourcewise representable, that is $x = A^{2s}z$, $s > 0$, there has been obtained a priori error estimate

$$\|x - x_{n,\delta}\| \leq \|x - x_n\| + \|x_n - x_{n,\delta}\| \leq \left(\frac{bs}{2kn}\right)^{\frac{s}{k}} \|z\| + 2k \left(\frac{n}{b}\right)^{\frac{1}{2k}} \delta,$$

$n \geq 1$ [9]. This error estimate has been optimized:

$$\|x - x_{n,\delta}\|_{opt} \leq (1 + 2s) \left(\frac{s}{k}\right)^{\frac{s(1-2k)}{k(1+2s)}} 2^{-\frac{s}{k(2s+1)}} \|z\|^{\frac{1}{2s+1}} \delta^{\frac{2s}{2s+1}}$$

and a priori stopping moment has been found

$$n_{opt} = 2^{-\frac{2s}{2s+1}} \left(\frac{s}{k}\right)^{\frac{2(s+k)}{2s+1}} b\|z\|^{\frac{2k}{2s+1}} \delta^{-\frac{2k}{2s+1}}.$$

It is evident that the optimum estimate does not depend on iteration parameter b, but n_{opt} does depend on b. Consequently, for reducing the calculating

procedure one should take b satisfying the condition $b > 0$ and proceed from the assumption that $n_{opt} = 1$. For that purpose it is enough to choose

$$b_{opt} = 2^{\frac{2s}{2s+1}} \left(\frac{s}{k}\right)^{-\frac{2(s+k)}{2s+1}} \|z\|^{-\frac{2k}{2s+1}} \delta^{\frac{2k}{2s+1}}.$$

The article [10] proves that provided $b > 0$, the iteration method (3) converges in the energy norm of Hilbert space $\|x\|_A = \sqrt{(Ax, x)}$, when one chooses the number of iterations n from the condition $\sqrt[4k]{n}\delta \to 0$ at $n \to \infty$, $\delta \to 0$. Without knowing the sourcewise representability of the exact solution, it is in the energy norm that there has been found a priori stopping moment $n_{opt} = b2^{-\frac{3+2k}{2}} k^{-\frac{1+2k}{2}} \|x\|^{2k} \delta^{-2k}$ and the conditions when the convergence in the energy norm results in the convergence in the original norm of Hilbert space H. In case of non-unique solution of the equation (1) the article [10] also proves that process (2) comes to the normal solution, i.e. the solution with a minimum norm.

3. Rule of stopping due to infinitesimal residual

When there is no information about the sourcewise representability of the exact solution, method (3) becomes ineffective, as it is impossible to get the error estimate and find the a priori stopping epoch in the original norm of Hilbert space. Nevertheless, one can make method (3) quite effective if one uses the following rule due to infinitesimal residual [3 – 4]. Here and in what follows, we shall consider that A is a bounded linear self-adjoint operator.

Let us set the stopping moment level $\varepsilon > 0$, $\varepsilon = b_1\delta$, $b_1 > 1$ and the moment m of stopping the iteration process (3) by condition

$$\|Ax_{n,\delta} - y_\delta\| > \varepsilon, (n < m), \|Ax_{m,\delta} - y_\delta\| \le \varepsilon. \tag{4}$$

Let us suppose that at initial approximation $x_{0,\delta}$ the residual is large enough, that is, larger than stopping level, i.e. $\|Ax_{0,\delta} - y_\delta\| > \varepsilon$. In what follows method (3) with stopping rule (4) is convergent provided

$$\lim_{\delta \to 0} \left(\inf_m \|x - x_{m,\delta}\|\right) = 0$$

Let us show the possible application of rule (4) to method (3). Consider the collection of functions $g_n(\lambda) = \frac{1}{\lambda}\left[1 - \frac{b^n}{(\lambda^{2k} + b)^n}\right] \ge 0$. By using the results of [9] it is easy to show that at $b > 0$ for $g_n(\lambda)$ the following conditions hold

$$\sup_{-M \le \lambda \le M} |g_n(\lambda)| \le 2k \left(\frac{n}{b}\right)^{1/(2k)}, n > 0, M = \|A\|, \tag{5}$$

$$\sup_{-M \le \lambda \le M} |1 - \lambda g_n(\lambda)| \le 1, n > 0, \tag{6}$$

$$1 - \lambda g_n(\lambda) \to 0, n \to \infty, \forall \lambda \in [-M, M], \tag{7}$$

$$\sup_{-M \le \lambda \le M} \left|\lambda^{2s}(1 - \lambda g_n(\lambda))\right| \le \left(\frac{bs}{2kn}\right)^{s/k}, kn > s, 0 \le s < \infty. \tag{8}$$

One finds valid

Lemma 1. *Let A be a bounded operator, $A = A^*$. Then for any $\omega \in H$ $(E - Ag_n(A))\,\omega \to 0, n \to \infty$.*

Proof. By using the integral expression of operator $A = \int\limits_{-M}^{M} \lambda dE_\lambda$, where $M = \|A\|$ and E_λ is the spectral function of operator A, we get

$$(E - Ag_n(A))\,\omega = \int\limits_{-M}^{M} (1 - \lambda g_n(\lambda))\, dE_\lambda \omega =$$

$$= \int\limits_{0}^{M} (1 - \lambda g_n(\lambda))\, dE_\lambda \omega + \int\limits_{-M}^{0} (1 - \lambda g_n(\lambda))\, dE_\lambda \omega = I_1 + I_2.$$

Let us break up the first of the integrals obtained into two integrals

$$I_1 = \int\limits_{0}^{\varepsilon_0} (1 - \lambda g_n(\lambda))\, dE_\lambda \omega + \int\limits_{\varepsilon_0}^{M} (1 - \lambda g_n(\lambda))\, dE_\lambda \omega.$$

Since $1 - \lambda g_n(\lambda) = \dfrac{b^n}{(\lambda^{2k} + b)^n} \le q^n(\varepsilon_0) < 1$ for all $\lambda \in [\varepsilon_0, M]$, we get

$$\left\| \int\limits_{\varepsilon_0}^{M} (1 - \lambda g_n(\lambda))\, dE_\lambda \omega \right\| \le q^n(\varepsilon_0) \left\| \int\limits_{\varepsilon_0}^{M} dE_\lambda \omega \right\| \le q^n(\varepsilon_0) \|\omega\| \to 0, n \to \infty.$$

On the basis of condition (6) we have

$$\left\| \int\limits_{0}^{\varepsilon_0} (1 - \lambda g_n(\lambda))\, dE_\lambda \omega \right\| \le \left\| \int\limits_{0}^{\varepsilon_0} dE_\lambda \omega \right\| \le \|E_{\varepsilon_0} \omega\| \to 0, \quad \varepsilon_0 \to 0,$$

because of the properties of spectral function [11]. Similarly to that, $I_2 \to 0$, $n \to \infty$. Consequently, $(E - Ag_n(A))\omega \to 0$, $n \to \infty$. Lemma 2.1 is proved.

There occurs

Lemma 2. *Let A be a bounded operator, $A = A^*$. Then for any $\vartheta \in \overline{R(A)}$ there exists correlation $n^{s/k} \left\| A^{2s}(E - Ag_n(A))\vartheta \right\| \to 0$ at $n \to \infty$, $0 \le s < \infty$.*

Proof. Since (8) is true, then

$$n^{s/k} \left\| A^{2s}(E - Ag_n(A)) \right\| \le n^{s/k} \sup_{-M \le \lambda \le M} \left| \lambda^{2s}(1 - \lambda g_n(\lambda)) \right| \le$$

$$\le n^{s/k} \gamma_s n^{-s/k} = \gamma_s,$$

where $\gamma_s = \left(\dfrac{bs}{2k} \right)^{s/k}$. Let us use Banach-Steingaus theorem [11, p. 151], according to which convergence $B_n u \to Bu$ at $n \to \infty$ for all $u \in H$ is realized

only when this convergence occurs in some compact subset in H and $\|B_n\|$, $n = 1, 2, \ldots$, are limited by the constant independent from n.

Let us take subset $R(A)$ as a compact one in $\overline{R(A)} = H$. We suppose that $s_1 = s + \dfrac{1}{2}$. Then for every $\vartheta = A\omega \in R(A)$ we have

$$n^{s/k} \left\| A^{2s}(E - Ag_n(A))\vartheta \right\| = n^{s/k} \left\| A^{2s+1}(E - Ag_n(A))\omega \right\| =$$

$$= n^{s/k} \left\| A^{2s_1}(E - Ag_n(A))\omega \right\| \le \gamma_{s_1} n^{\frac{-(s_1-s)}{k}} \|\omega\| = \gamma_{s_1} \|\omega\| n^{-1/(2k)} \to 0,$$

$n \to \infty$, as $s_1 < \infty$. Lemma 2.2 is proved.

There is validity in

Lemma 3. *Let A be a bounded operator, $A = A^*$. Provided for some sequence $n_p < \overline{n} = const$ and $\vartheta_0 \in \overline{R(A)}$ at $p \to \infty$ we get $\omega_p = A\left(E - Ag_{n_p}(A)\right)\vartheta_0 \to 0$, then $\vartheta_p = \left(E - Ag_{n_p}(A)\right)\vartheta_0 \to 0$.*

Proof. Due to (6) sequence ϑ_p is bounded $\|\vartheta_p\| \le 1, p \in N$. That is why out of this sequence in Hilbert space we can extract a weakly convergent subsequence $\vartheta_p - \to \vartheta, (p \in N' \subseteq N)$, then $A\vartheta_p - \to A\vartheta, (p \in N')$.

But by the data $\omega_p = A\vartheta_p \to 0, p \to \infty$, consequently, $A\vartheta = 0$. Since zero is not the characteristic value of operator A, then $\vartheta = 0$. Hence,

$$\|\vartheta_p\|^2 = \left(\vartheta_p, \left(E - Ag_{n_p}(A)\right)\vartheta_0\right) = (\vartheta_p, \vartheta_0) - \left(\vartheta_p, Ag_{n_p}(A)\vartheta_0\right) =$$

$$= (\vartheta_p, \vartheta_0) - \left(A\vartheta_p, g_{n_p}(A)\vartheta_0\right) =$$

$$= (\vartheta_p, \vartheta_0) - \left(\omega_p, g_{n_p}(A)\vartheta_0\right) \to (\vartheta, \vartheta_0) = 0, \quad \left(p \in N'\right),$$

since $\vartheta = 0, \omega_p \to 0, p \to \infty$ and by the data (5)

$$\left\| g_{n_p}(A) \right\| \le 2k \left(\frac{n_p}{b}\right)^{1/(2k)} \le 2k \left(\frac{\overline{n}}{b}\right)^{1/(2k)}$$

Thus, every weakly convergent subsequence of the bounded sequence ϑ_p mentioned above tends to zero according to the norm. Consequently, the whole sequence $\vartheta_p \to 0, p \to \infty$. Lemma 2.3 is proved.

If A is a bounded non self-adjoint operator, lemma 2.3 which is analogous to lemma 4 proves its validity.

Lemma 4. *Let A be a bounded non self-adjoint operator. If for some sequence $n_p < \overline{n} = const$ and $\vartheta_0 \in \overline{R(A)}$ at $p \to \infty$ we have*

$$\omega_p = A^*A\left(E - A^*Ag_{n_p}\left(A^*A\right)\right)\vartheta_0 \to 0,$$

then $\vartheta_p = \left(E - A^*Ag_{n_p}\left(A^*A\right)\right)\vartheta_0 \to 0$.

For proving lemma 2.4 it is necessary to go over to operator $A = A^*A$ and use lemma 2.3.

Let us use the proved lemmas for proving the following theorem.

Theorem 1. *Let A be a bounded operator, $A = A^*$, and let the stopping moment $m = m(\delta)$ in method (3) be chosen according to rule (4). Then $x_{m(\delta),\delta} \to x$ at $\delta \to 0$.*

Proof. In [9] we find that $x_{n,\delta} = A^{-1}\left[E - (CB)^n\right] y_\delta$, where

$$C = \left(A^{2k} + B\right)^{-1}.$$

That is why

$$\begin{aligned} x_{n,\delta} - x &= A^{-1}\left[E - (CB)^n\right] y_\delta - x = \\ &= A^{-1}\left[E - (CB)^n\right](y_\delta - y) + A^{-1}\left[E - (CB)^n\right] y - A^{-1}y = \\ &= A^{-1}\left[E - (CB)^n\right](y_\delta - y) - (CB)^n x = \\ &= g_n(A)(y_\delta - y) - (E - Ag_n(A))x, \end{aligned} \tag{9}$$

consequently,

$$Ax_{n,\delta} - y = Ax_{n,\delta} - Ax = -A(E - Ag_n(A))x + Ag_n(A)(y_\delta - y).$$

Let us consider

$$\begin{aligned} Ax_{n,\delta} - y_\delta &= -A(E - Ag_n(A))x + (y - y_\delta) + Ag_n(y_\delta - y) = \\ &= -A(E - Ag_n(A))x - (E - Ag_n(A))(y_\delta - y). \end{aligned} \tag{10}$$

On the strength of lemmas 2.1 and 2.2 we have

$$\|(E - Ag_n(A))x\| \to 0, n \to \infty, \tag{11}$$

$$\sigma_n = n^{1/(2k)} \|A(E - Ag_n(A))x\| \to 0, n \to \infty. \tag{12}$$

What is more, it follows from (5) and (6) that

$$\|g_n(A)(y_\delta - y)\| \le 2k \left(\frac{n}{b}\right)^{1/(2k)} \delta, \tag{13}$$

$$\|E - Ag_n(A)\| \le 1. \tag{14}$$

Let us use stopping rule (4). Then

$$\|Ax_{m,\delta} - y_\delta\| \le b_1\delta, \quad b_1 > 1$$

and from (10) and (14) we get

$$\begin{aligned} \|A(E - Ag_m(A))x\| &\le \|Ax_{m,\delta} - y_\delta\| + \|(E - Ag_m(A))(y_\delta - y)\| \le \\ &\le (b_1 + 1)\delta. \end{aligned} \tag{15}$$

For any $n < m$ $\|Ax_{n,\delta} - y_\delta\| > \varepsilon$, that is why

$$\|A(E - Ag_n(A))x\| \ge \|Ax_{n,\delta} - y_\delta\| - \|(E - Ag_n(A))(y - y_\delta)\| \ge (b_1 - 1)\,\delta.$$

Thus, for $\forall n < m$

$$\|A(E - Ag_n(A))x\| \ge (b_1 - 1)\,\delta. \tag{16}$$

From (12) and (16) at $n = m - 1$ we have

$$\frac{\sigma_{m-1}}{(m-1)^{1/(2k)}} = \|A(E - Ag_{m-1}(A))x\| \ge (b_1 - 1)\,\delta$$

or $(m-1)^{1/(2k)}\delta \le \dfrac{\sigma_{m-1}}{b-1} \to 0, \delta \to 0$ (because from (12) $\sigma_m \to 0, m \to \infty$). If in this case $m \to \infty$ at $\delta \to 0$, then using (9), we get

$$\|x_{m,\delta} - x\| \le \|(E - Ag_m(A))\,x\| + \|g_m(A)(y_\delta - y)\| \le$$

$$\leq \|(E - Ag_m(A))\, x\| + 2k \left(\frac{m}{b}\right)^{1/(2k)} \delta \to 0$$

at $m \to \infty, \delta \to 0$, since from (11)

$$\|(E - Ag_m(A))\, x\| \to 0, \quad m \to \infty.$$

Provided for some $\delta \to 0$ the sequence $m(\delta_n)$ turns out to be bounded, $x_{m(\delta_n),\delta_n} \to x, \delta_n \to 0$ is relevant in this case as well. Actually, from (15) we have

$$\left\|A\left(E - Ag_{m(\delta_n)}(A)\right) x\right\| \leq (b_1 + 1)\, \delta_n \to 0, \delta_n \to 0.$$

Hence, according to lemma 2.3 we get that

$$\left(E - Ag_{m(\delta_n)}(A)\right) x \to 0, \delta_n \to 0.$$

As a result

$$\left\|x_{m(\delta_n),\delta_n} - x\right\| \leq \left\|\left(E - Ag_{m(\delta_n)}(A)\right) x\right\| + 2k \left(\frac{m(\delta_n)}{b}\right)^{1/(2k)} \delta_n \to 0, \delta_n \to 0.$$

This proves theorem 2.5.

4. Error estimate

We have

Theorem 2. *Suppose the conditions of theorem 2.5 are fulfilled, operator A is positive and $x = A^{2s}z, s > 0$. Then the following estimates hold*

$$m \leq 1 + \frac{(2s+1)b}{4k} \left[\frac{\|z\|}{(b_1 - 1)\delta}\right]^{\frac{2k}{2s+1}},$$

$$\|x_{m,\delta} - x\| \leq [(b_1 + 1)\, \delta]^{\frac{2s}{2s+1}} \|z\|^{\frac{1}{2s+1}} + \\ + \frac{2k}{b^{1/(2k)}} \left\{1 + \frac{(2s+1)b}{4k} \left[\frac{\|z\|}{(b_1 - 1)\delta}\right]^{\frac{2k}{2s+1}}\right\}^{\frac{1}{2k}} \delta. \tag{17}$$

Proof. Since $x = A^{2s}z$, then

$$\|A(E - Ag_{m-1}(A))x\| = \left\|A^{2s+1}(E - Ag_{m-1}(A))z\right\| =$$

$$= \left\|\int_0^M \frac{\lambda^{2s+1} b^{m-1}}{(\lambda^{2k} + b)^{m-1}} dE_\lambda z\right\| \leq \left[\frac{(2s+1)b}{4k(m-1)}\right]^{\frac{2s+1}{2k}} \|z\|.$$

By using (16), we get

$$(b_1 - 1)\delta \leq \left[\frac{(2s+1)b}{4k(m-1)}\right]^{\frac{2s+1}{2k}} \|z\|.$$

Hence we have

$$m \leq 1 + \frac{(2s+1)b}{4k} \left[\frac{\|z\|}{(b_1 - 1)\delta}\right]^{\frac{2k}{2s+1}}.$$

With the help of moment inequality let us estimate

$$\|(E - Ag_m(A))x\| = \left\|A^{2s}(E - Ag_m(A))z\right\| \leq$$

$$\leq \left\|A^{2s+1}(E - Ag_m(A))z\right\|^{\frac{2s}{2s+1}} \|(E - Ag_m(A))z\|^{\frac{1}{2s+1}} \leq$$

$$\leq \|A(E - Ag_m(A))x\|^{\frac{2s}{2s+1}} \|z\|^{\frac{1}{2s+1}} \leq [(b_1+1)\delta]^{\frac{2s}{2s+1}} \|z\|^{\frac{1}{2s+1}}.$$

Then

$$\|x_{m,\delta} - x\| \leq \|(E - Ag_m(A))x\| + \|g_m(A)(y_\delta - y)\| \leq$$

$$\leq [(b_1+1)\delta]^{\frac{2s}{2s+1}} \|z\|^{\frac{1}{2s+1}} + 2k\left(\frac{m}{b}\right)^{\frac{1}{2k}} \delta \leq$$

$$\leq [(b_1+1)\delta]^{\frac{2s}{2s+1}} \|z\|^{\frac{1}{2s+1}} + \frac{2k}{b^{1/(2k)}} \left\{1 + \frac{(2s+1)b}{4k}\left[\frac{\|z\|}{(b_1-1)\delta}\right]^{\frac{2k}{2s+1}}\right\}^{\frac{1}{2k}} \delta.$$

This proves theorem 3.1.

Note 1. *The estimate procedure* (17) *is* $O\left(\delta^{\frac{2s}{2s+1}}\right)$ *and, as it follows from* [3], *it is optimal in the class of problems with sourcewise representable solutions.*

Note 2. *The knowledge of order* $2s > 0$ *of sourcewise representability of exact solution, which is used in theorem 2, is not required in practice as it does not hold for the rule of stopping due to infinitesimal residual. Theorem 2 states that the number of iterations m, supporting the optimum error order. But even if the sourcewise representability of the exact solution is missing, stopping due to residual provides the convergence of the method, as it is shown in theorem 1.*

Conclusion. The paper studies some properties of the suggested implicit iteration method of solving ill-posed problems: it proves the convergence of the method with the a posteriori choice of the iteration number in the original norm of Hilbert space. It also presents the obtained error estimate of the method and the estimate of a posteriori stopping moment.

Bibliography

1. Lavrentiev M. M. On some ill-posed problems of mathematical physics / M. M. Lavrentiev. – Novosibirsk : SO AS USSR, 1962. – 92 p. (in Russian).
2. Bakushinski A. B. A general method of regularizing algorithm development for the linear ill-posed equation in Hilbert space / A. B. Bakushinski // Computational mathematics and mathematical physics journal. – 1967. – Vol. 7, № 3. – P. 672-677. (in Russian).
3. Vainikko G. M. Iterative procedures in ill-posed problems / G. M. Vainikko, A. Yu. Veretennikov. – Moscow: Nauka, 1986. – 178 p. (in Russian).
4. Yemelin I. V. To the theory of ill-posed problems / I. V. Yemelin, M. A. Krasnoselsky // Reports of the USSSR Academy of Sciences. – 1979. – Vol. 244, № 4. – P. 805-808. (in Russian).
5. Yemelin I. V. Rule of stopping in the iteration procedures of solving ill-posed problems / I. V. Yemelin, M. A. Krasnoselsky // Automatics and telemechanics. – 1978. – № 12. – P. 59-63. (in Russian).
6. Denisov A. M. Introduction to the inverse problem theory / A. M. Denisov. – M.: Moscow State University, 1994. – 207 p. (in Russian).
7. Lavrentiev M. M. Operator theory and ill-posed problems / M. M. Lavrentiev, L. Ya. Saveliev. – Novosibirsk: Institute of mathematics publishing house, 1999. – 702 p. (in Russian).
8. Samarsky A. A. Numerical methods of solving inverse problems of mathematical physics / A. A. Samarsky, P. N. Vabishchevitch. – M.: Editorial URSS, 2004. – 480 p. (in Russian).

9. Matysik O. V. The implicit iteration method for solving linear equations with an unbounded operator / O. V. Matysik // Brest University Vestnik. Series 4. Physics. Mathematics. – 2013. – № 1. – P. 77-83. (in Russian).
10. Matysik O. V. On the approximate solution of linear equations with an unbounded operator in Hilbert space / O. V. Matysik // Brest University Vestnik. Series 4. Physics. Mathematics. – 2013. – № 2. – P. 87-92. (in Russian).
11. Lyusternik L. A. Elements of functional analysis / L. A. Lyusternik, V. I. Sobolev. – Moscow: Nauka, 1965. – 520 p. (in Russian).

O. V. Matysik,
Applied Mathematics and Programming Technologies Department,
Brest State University named after A. S. Pushkin,
21 Kosmonavtov boulevard, Brest, 224016, Belarus

Received 08.07.2015

Журнал обчислювальної та прикладної математики | **2015** **№ 2 (119)** | Journal of Computational & Applied Mathematics

UDC 519.6

GENERALIZATION OF THE KHOVANSKII'S METHOD FOR SOLVING MATRIX POLYNOMIAL EQUATIONS

A. M. NEDASHKOVSKA

Резюме. Розглянуто алгоритм розв'язування поліноміальних матричних рівнянь. Запропоновані рекурентні формули обчислення наближених розв'язків для рівнянь степеня n. Досліджено збіжність методу для рівнянь другого степеня. Наведено результати чисельних експериментів, що підтверджують справедливість теоретичних викладок.

Abstract. The article deals with the algorithm for solving the polynomial matrix equations. Recurrent formulas for calculating approximate solutions of equations of degree n are proposed. The convergence of the method for equations of the second degree has been researched and the results of the numerical experiments that confirm the validity of the calculations are provided.

1. Introduction

The method reduces itself to the consistent application of a certain matrix operator to the given vector and occupies a special place among various generalizations of continued fractions. In a simpler form, this method has been considered by Euler. He used it to calculate the approximate expression $x^{\frac{p}{q}}$. Here x is a known number, p and q are integers.

Euler's method has also been considered by Laurie, Kraft and Muller. But the possibility of practical use of the method hasn't been considered in these works. Later Khovanskii applied this method to the approximate value of the roots of some degrees and to find approximate solutions of polynomial equations over the field of real numbers.

In particular, the scheme of finding the roots of the equation

$$x^2 = u \tag{1}$$

has been considered in [1].

It has been shown that the solution of equation (1) can be found as the fraction $\frac{P_n}{Q_n}$, where valid values P_n and Q_n are interconnected by relations

$$\begin{pmatrix} P_n \\ Q_n \end{pmatrix} = \begin{pmatrix} a & u \\ 1 & a \end{pmatrix} \begin{pmatrix} P_{n-1} \\ Q_{n-1} \end{pmatrix} \quad (n = 1, 2, \ldots). \tag{2}$$

Here a is a free parameter.

The equation (2) implies that

$$\frac{P_n}{Q_n} = \frac{aP_{n-1} + uQ_{n-1}}{P_{n-1} + aQ_{n-1}},$$

Key words. Polynomial matrix equations; generalization of the Khovanskii's method; the convergence of the method.

that is

$$\frac{P_n}{Q_n} = \frac{a\frac{P_{n-1}}{Q_{n-1}} + u}{\frac{P_{n-1}}{Q_{n-1}} + a}. \tag{3}$$

Let limit $\lim\limits_{n\to\infty} \frac{P_{n-1}}{Q_{n-1}}$ exists and is finite. We denote it as x. Than from (3) we receive

$$x = \frac{ax + u}{x + a}$$

or

$$x^2 = u, x = \pm\sqrt{u}.$$

So, if the limit $\lim\limits_{n\to\infty} \frac{P_{n-1}}{Q_{n-1}}$ exists, then it may be equal to $\sqrt{u}$ or to $-\sqrt{u}$. Accordingly, if $u < 0$, then the process (2) diverges.

A similar scheme has been proposed in [1] for solving the quadratic equation $x^2+px+q=0$:

$$\begin{pmatrix} P_n \\ Q_n \end{pmatrix} = \begin{pmatrix} a & -q \\ 1 & a+p \end{pmatrix} \begin{pmatrix} P_{n-1} \\ Q_{n-1} \end{pmatrix} \quad (n = 1, 2, \ldots). \tag{4}$$

From (4) it follows that

$$\frac{P_n}{Q_n} = \frac{a\frac{P_{n-1}}{Q_{n-1}} - q}{\frac{P_{n-1}}{Q_{n-1}} + a + p}. \tag{5}$$

Let the limit $\lim\limits_{n\to\infty} \frac{P_{n-1}}{Q_{n-1}}$ exists and is finite. We denote it as x. Then from (5) we receive

$$x = \frac{ax - q}{x + a + p}$$

or

$$x^2 + px + q = 0, x_{1,2} = \frac{-p \pm \sqrt{p^2 - 4q}}{2}.$$

In [1] the conditions for the convergence of the iterative formulas (3) and (5) have been analysed.

2. The computational scheme of the method

Let us try to generalize the scheme proposed in [1] and apply it to solving the matrix equation

$$A_n X^n + A_{n-1} X^{n-1} + A_{n-2} X^{n-2} + \ldots + A_2 X^2 + A_1 X + A_0 = 0. \tag{6}$$

Here matrices $A_0, A_1, A_2, \ldots, A_{n-2}, A_{n-1}, A_n \in \Re^{m\times m}$ are given coefficients of equation (6) and $X \in \Re^{m\times m}$ is an unknown solution.

Suppose, that X is a non singular matrix and let us denote $Y_0 = X^{-1}$. .... After the right multiplication of the equation (6) with X^{-1} we receive

$$A_n X^{n-1} + A_{n-1} X^{n-2} + A_{n-2} X^{n-3} + \ldots + A_2 X + A_1 + A_0 Y_0 = 0. \tag{7}$$

Let $Y_1 = Y_0 X^{-1} = \left(X^{-1}\right)^2$ and let us right multiply the equation (7) with X^{-1} :

$$A_n X^{n-2} + A_{n-1} X^{n-3} + A_{n-2} X^{n-4} + \ldots + A_2 + A_1 Y_0 + A_0 Y_1 = 0.$$

Accordingly, after $(n-2)$ the right multiplication of the equation (6) with X^{-1} we get

$$A_nX^2 + A_{n-1}X + A_{n-2} + A_{n-3}Y_0 + \ldots + A_2Y_{n-5} + A_1Y_{n-4} + A_0Y_{n-3} = 0, \quad (8)$$

where

$$Y_0 = \left(X^{-1}\right)^1, Y_1 = \left(X^{-1}\right)^2, \ldots, Y_{n-3} = \left(X^{-1}\right)^{n-2}.$$

We introduce the parameter, a non singular matrix $L \in \Re^{m\times m}$ and left multiply equation (8) with L :

$$LA_nX^2 + LA_{n-1}X + LA_{n-2} + \ldots + LA_2Y_{n-5} + LA_1Y_{n-4} + LA_0Y_{n-3} = 0. \quad (9)$$

Obviously the equation (9) is equivalent to

$$\begin{aligned} LA_nX^2 + (LA_{n-1} + K - K)\,X + LA_{n-2} + \ldots + \\ + LA_2Y_{n-5} + LA_1Y_{n-4} + LA_0Y_{n-3} = 0. \end{aligned}$$

Here $K \in \Re^{m\times m}$ is a non singular matrix.

And it is evident that

$$LA_nX^2 + KX = (K - LA_{n-1})\,X - LA_{n-2} - \ldots - LA_1Y_{n-4} - LA_0Y_{n-3}$$

or

$$(LA_nX + K)\,X = (K - LA_{n-1})\,X - LA_{n-2} - \ldots - LA_1Y_{n-4} - LA_0Y_{n-3}.$$

Then, assuming $\det(K - LA_{n-1}) \neq 0$ we get

$$\begin{aligned} X = (LA_nX + K)^{-1}\big((K - LA_{n-1})\,X - LA_{n-2} - \ldots - \\ - LA_1Y_{n-4} - LA_0Y_{n-3}\big). \end{aligned} \quad (10)$$

Now let us consider the obvious equality

$$LA_nXX^{-1} + KY_0 = LA_n + KY_0$$

or

$$(LA_nX + K)\,Y_0 = LA_n + KY_0. \quad (11)$$

Then from (11) we get

$$Y_0 = (LA_nX + K)^{-1}\,(LA_n + KY_0)\,. \quad (12)$$

Applying similar transformations, we receive formulas for $Y_1, Y_2, \ldots, Y_{n-3}$ calculation:

$$\begin{aligned} Y_1 &= (LA_nX + K)^{-1}\,(LA_nY_0 + KY_1)\,; \\ Y_2 &= (LA_nX + K)^{-1}\,(LA_nY_1 + KY_2)\,; \\ &\vdots \\ Y_{n-3} &= (LA_nX + K)^{-1}\,(LA_nY_{n-2} + KY_{n-3})\,. \end{aligned} \quad (13)$$

Then, on the basis of the formulas (10),(12) and (13) we get an approximate calculation algorithm for solving the polynomial matrix equation (6):

1. Set the accuracy $\varepsilon > 0$;
2. Set the initial approximation, a non singular matrix $X_0 \in \Re^{m\times m}$;
3. Set the counter $n = 1$;

4. Calculate

$$Y_0^{(0)} = \left(X^{(0)^{-1}}\right)^1, Y_1^{(0)} = \left(X^{(0)^{-1}}\right)^2,$$

$$Y_2^{(0)} = \left(X^{(0)^{-1}}\right)^3, \ldots, Y_{n-3}^{(0)} = \left(X^{(0)^{-1}}\right)^{n-2};$$

5. Calculate

$$\begin{aligned}
Y_0^{(n)} &= \left(LA_nX^{(n-1)} + K\right)^{-1}\left(LA_n + KY_0^{(n-1)}\right),\\
Y_1^{(n)} &= \left(LA_nX^{(n-1)} + K\right)^{-1}\left(LA_nY_0^{(n)} + KY_1^{(n-1)}\right),\\
Y_2^{(n)} &= \left(LA_nX^{(n-1)} + K\right)^{-1}\left(LA_nY_1^{(n)} + KY_2^{(n-1)}\right),\\
&\vdots\\
Y_{n-3}^{(n)} &= \left(LA_nX^{(n-1)} + K\right)^{-1}\left(LA_nY_{n-2}^{(n)} + KY_{n-3}^{(n-1)}\right),\\
X^{(n)} &= \left(LA_nX^{(n-1)} + K\right)^{-1}\times\\
&\quad\times\left((K - LA_{n-1})X^{(n-1)} - LA_{n-2} - \ldots - LA_0Y_{n-3}^{(n)}\right);
\end{aligned} \tag{14}$$

6. Verify the condition $\left\|X^{(n)} - X^{(n-1)}\right\| < \varepsilon$. If this condition is not satisfied, , set the counter $n = n + 1$ and go to step 5, or else return $X^{(n)}$.

3. The convergence of the method for equations of the second power

Let us consider the equation

$$A_2X^2 + A_1X + A_0 = 0. \tag{15}$$

Like the equation (6) we left multiply (15) with a non singular diagonal matrix $L = l \cdot E, L \in \Re^{m\times m}$:

$$LA_2X^2 + (LA_1 + K - K)\,X + LA_0 = 0$$

or

$$(LA_2X + LA_1 + K)\,X = KX - LA_0. \tag{16}$$

Here $K = k \cdot E, K \in \Re^{m\times m}$ is non singular diagonal matrix.

Assuming that $\det(LA_2X + LA_1 + K) \neq 0$ from (16) we get

$$X = (LA_2X + LA_1 + K)^{-1}(KX - LA_0)$$

or as a recurrent formula

$$X^{(n)} = \left(LA_2X^{(n-1)} + LA_1 + K\right)^{-1}\left(KX^{(n-1)} - LA_0\right)\ (n = 1, 2, \ldots). \tag{17}$$

Let A and B be real, square $m \times m$ matrix with $\det B \neq 0$. Further multiplication operation $B^{-1}A$ will be written in the form of a matrix fraction $\frac{A}{B}$.

Inasmuch

$$X = \frac{KX - LA_0}{LA_2X + LA_1 + K} = \frac{kX - lA_0}{lA_2X + lA_1 + kE} = \frac{kX - lA_0}{lX + lA_2^{-1}A_1 + kA_2^{-1}} =$$

$$\frac{X-\frac{l}{k}A_0}{X+A_2^{-1}A_1+\frac{k}{l}A_2^{-1}}=\frac{X+A_2^{-1}A_1+\frac{k}{l}A_2^{-1}-\left(A_2^{-1}A_1+\frac{k}{l}A_2^{-1}+\frac{l}{k}A_0\right)}{X+A_2^{-1}A_1+\frac{k}{l}A_2^{-1}},$$

then

$$X=E-\frac{A_2^{-1}A_1+\frac{k}{l}A_2^{-1}+\frac{l}{k}A_0}{X+A_2^{-1}A_1+\frac{k}{l}A_2^{-1}}. \tag{18}$$

Let $P=A_2^{-1}A_1+\frac{k}{l}A_2^{-1}$ and $Q=\frac{l}{k}A_0$. Then (18) can be written as

$$X=E-\frac{P+Q}{X+P},$$

or as an infinite matrix continued fraction

$$X=E-\frac{P+Q}{P+E-\frac{P+Q}{P+E-\ldots}}. \tag{19}$$

The matrix continued fraction (18) also can be presented in a compact Prynhsheym's form

$$X=E-\frac{P+Q|}{|P+E}-\frac{P+Q|}{|P+E}-\frac{P+Q|}{|P+E}-\ldots \tag{20}$$

Let us consider the continued fraction with real elements. It is evident that

$$\begin{gathered}\frac{a_1|}{|b_1}+\frac{a_2|}{|b_2}+\frac{a_3|}{|b_3}+\ldots+\frac{a_n|}{|b_n}+\ldots=\\ =\frac{\frac{a_1}{b_1}\big|}{|1}+\frac{\frac{a_2}{b_1}\big|}{|b_2}+\frac{a_3|}{|b_3}+\ldots+\frac{a_n|}{|b_n}+\ldots\\ =\frac{\frac{a_1}{b_1}\big|}{|1}+\frac{\frac{a_2}{b_1b_2}\big|}{|1}+\frac{\frac{a_3}{b_2b_3}\big|}{|1}+\ldots+\frac{\frac{a_n}{b_{n-1}b_n}\big|}{|1}+\ldots\end{gathered} \tag{21}$$

Suppose that the matrix $(P+E)$ is non singular and in (20) we perform transformations similar to (21):

$$\begin{gathered}X=E-\frac{P+Q|}{|P+E}-\frac{P+Q|}{|P+E}-\frac{P+Q|}{|P+E}-\ldots-\frac{P+Q|}{|P+E}-\ldots=\\ =E-\frac{(P+E)^{-1}(P+Q)\big|}{|E}-\frac{(P+E)^{-1}(P+Q)\big|}{|P+E}-\\ -\frac{P+Q|}{|P+E}-\ldots-\frac{P+Q|}{|P+E}-\ldots=\\ =E-\frac{(P+E)^{-1}(P+Q)\big|}{|E}-\frac{(P+E)^{-2}(P+Q)\big|}{|E}-\\ -\frac{(P+E)^{-2}(P+Q)\big|}{|E}-\ldots-\frac{(P+E)^{-2}(P+Q)\big|}{|E}-\ldots\end{gathered} \tag{22}$$

In [2] Vorpitskyi's sufficient convergence sign has been generalized. It can be used to analyse the convergence of matrix continued fractions of the form (22):

Theorem 1. *Matrix branch continued fraction*

$$\sum_{k_1=1}^{n}\frac{A_{k_1}|}{|E}+\sum_{k_2=1}^{n}\frac{A_{k_1k_2}|}{|E}+\ldots+\sum_{k_1=1}^{n}\frac{A_{k_1k_2\ldots k_l}|}{|E}+\ldots$$

is absolutely convergent if the condition

$$\|A_{k_1k_2\ldots k_i}\|\le\frac{1}{4n}\,(i=1,2,3,\ldots;k_i=1,2,\ldots,n)$$

is true.

Let us apply Theorem 1 to the continued fraction (22). It is obvious that the branched continued fraction (22) will be convergent, if the condition

$$\left\|(P+E)^{-2}(P+Q)\right\| \leq \frac{1}{4} \tag{23}$$

is satisfied.

Substituting the values of P and Q in the formula (23) we get sufficient condition for the convergence of the matrix continued fraction (22):

$$\left\|\left(A_2^{-1}A_1+\frac{k}{l}A_2^{-1}+E\right)^{-2}\left(4A_2^{-1}A_1+\frac{4k}{l}A_2^{-1}+\frac{4l}{k}A_0\right)\right\| \leq 1.$$

4. Computational experiments

To test the effectiveness of the practical application of recurrent formula (14), a series of numerical experiments has been conducted in the FreeMat environment.

Example 1. Let us consider the polynomial matrix equation

$$A_2X^2+A_1X+A_0=0, \tag{24}$$

with

$$A_2=\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix},\ A_1=\begin{pmatrix}1&2&3\\2&3&4\\3&4&5\end{pmatrix},\ A_0=\begin{pmatrix}-13&-13&-14\\-16&-18&-18\\-20&-21&-23\end{pmatrix}.$$

Put $l=1, k=1$ and initial value

$$X_0=\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}$$

then using the recurrent formula (17) we obtain the following results

Tabl. 1. Example 1

ε	Number of iterations, n	Approximate solution, X_n	Norm of residual
0.1	15	$\begin{pmatrix}-8.9203&-9.9203&-9.9203\\-0.5083&0.4917&-0.5083\\8.9038&8.9038&9.9038\end{pmatrix}$	0.0848
0.01	19	$\begin{pmatrix}-8.9079&-9.9079&-9.9079\\-0.5065&0.4935&-0.5065\\8.8948&8.8948&9.8948\end{pmatrix}$	0.0056
0.001	22	$\begin{pmatrix}-8.9069&-9.9069&-9.9069\\-0.5064&0.4936&-0.5064\\8.8941&8.8941&9.8941\end{pmatrix}$	0.0007

These results show convergence of the iterative process (17) to the solution of equation (24),

$$X = \begin{pmatrix} -8.9070 & -9.9070 & -9.9070 \\ -0.5064 & 0.4936 & -0.5064 \\ 8.8942 & 8.8942 & 9.8942 \end{pmatrix}$$

with a decrease of ε.

Example 2. Now let us consider the polynomial matrix equation

$$A_2X^2 + A_1X + A_0 = 0, \tag{25}$$

with coefficients

$$A_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \; A_1 = \begin{pmatrix} -1 & 0 & 2 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 4 & 1 \\ 0 & 0 & 0 & -5 \end{pmatrix},$$

$$A_0 = \begin{pmatrix} -8 & -8 & -10 & -9 \\ -9 & -11 & -9 & -11 \\ -11 & -11 & -16 & -12 \\ -1 & -1 & -1 & 3 \end{pmatrix}.$$

Let $l = 1, k = 1$ and initial value

$$X_0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and we use the recurrent formula (17). We get the results from Table 2.

TABL. 2. Example 2

ε	Number of iterations, n	Approximate solution, X_n	Norm of residual
0.1	12	$\begin{pmatrix} -8.3232 & -9.3232 & -9.3232 & -9.3232 \\ 5.7750 & 6.7750 & 5.7750 & 5.7750 \\ 2.4210 & 2.4210 & 3.4210 & 2.4210 \\ -0.2323 & -0.2323 & -0.2323 & 0.7677 \end{pmatrix}$	0.0610
0.01	15	$\begin{pmatrix} -8.3335 & -9.3335 & -9.3335 & -9.3335 \\ 5.7775 & 6.7775 & 5.7775 & 5.7775 \\ 2.4216 & 2.4216 & 3.4216 & 2.4216 \\ -0.2288 & -0.2288 & -0.2288 & 0.7712 \end{pmatrix}$	0.0070
0.001	18	$\begin{pmatrix} -8.3323 & -9.3323 & -9.3323 & -9.3323 \\ 5.7773 & 6.7773 & 5.7773 & 5.7773 \\ 2.4216 & 2.4216 & 3.4216 & 2.4216 \\ -0.2293 & -0.2293 & -0.2293 & 0.7707 \end{pmatrix}$	0.0008

These results show convergence of the iterative process (17) to the solution of equation (25),

$$X = \begin{pmatrix} -8.3325 & -9.3325 & -9.3325 & -9.3325 \\ 5.7773 & 6.7773 & 5.7773 & 5.7773 \\ 2.4216 & 2.4216 & 3.4216 & 2.4216 \\ -0.2292 & -0.2292 & -0.2292 & 0.7708 \end{pmatrix}$$

with a decrease of ε.

5. Conclusions

The article deals with the modification of the method that was proposed by A.N. Khovanskii [1] for solving polynomial equations defined over the set of real numbers. Obtained computational scheme allows us to construct approximate solutions of the equation (6), that are considered over the ring of non commutative matrices. Sufficient conditions for the convergence of the iterative process for the equation of the second degree and software implementation of the method were presented. A number of numerical experiments confirm the applicability of the proposed scheme were conducted.

Bibliography

1. Khovanskii A. N. The Application of continued fractions and their generalizations to problems in approximation theory / A. N. Khovanskii, P. Noordhoff. – 1963 – 212 p.
2. Bodnar D. I. Branched Continued Fractions / D. I. Bodnar. – K: Naukova Dumka, 1986. – 176 p. (in Russian).

A. M. Nedashkovska,
Ivan Franko National University of Lviv,
1, Universytets'ka Str., Lviv, 79000, Ukraine

Received 20.05.2015

Журнал обчислювальної та прикладної математики | **2015** **№ 2 (119)** | Journal of Computational & Applied Mathematics

UDC 519.6

SPECIAL ESTIMATORS FOR CORRECTING SOME SOLUTIONS OF INTEGRAL EQUATIONS

B. OSTUDIN, YA. GARASYM, A. BESHLEY

Резюме. У роботі проведено аналіз чисельного розв'язування двовимірного інтегрального рівняння теорії потенціалу на незамкнених поверхнях. На прикладі аналізу конкретної модельної задачі показано, як, враховуючи специфіку початкових даних, вирішити проблему спеціального зображення самого інтегрального рівняння. Таке зображення дозволяє при побудові відповідної наближеної схеми суттєво спростити використання апріорної інформації про характер поведінки шуканого розв'язку. Останнє відіграє важливу роль у процесі реалізіації різних процедур уточнення отримуваних наближених розв'язків на основі спеціально побудованих оцінювачів. У роботі представлені результати чисельних експериментів.

Abstract. The numerical solution of two-dimensional integral equation on unclosed surfaces is analyzed in present paper. Such equations with weak singularities in the kernels are considered in potential theory. General problem of integral equation solving, and besides that special representation of considered equation, are exemplified by the model task, taking into account the specificity of initial date. In the process of appropriate numerical scheme constructing such a representation gives the possibility to essentially simplify the use of a priori information on desired solution. The last is important for objectifying various correction procedures of obtained results on the basis of special estimators. The results of numerical experiments are presented.

1. Introduction

In previous paper [2] with a similar research object various aspects of numerical schemes construction for solving integral equations of the first kind were considered. In this connection we had to deal with two-dimensional equations in the form as

$$(A\sigma)(M) \equiv \iint\limits_S \sigma(P)|M-P|^{-1}dS_P = U(M),\ M \in S, \tag{1}$$

where, in general case, S is an open Lipschitz surface; M and P are the points of Euclidean space $\mathbb{R}^3$. In present article, by solving one typical model problem, we analyze the proposed schemes adaptive possibilities for maximal taking account of desired solutions specificity in order to receive the results with preassigned accuracy. The equations of type (1) have been used in mathematical modelling of some boundary value problems in electron optics [3]. Ordinary generalization of (1) is an assumption that S is formed by the aggregate of m surfaces, so

Key words. Two-dimensional integral equation, weak singularity, rational representation, numerical scheme construction, correction of obtained results, special estimators.

that $S := \bigcup_{i=1}^{m} S_i$. In this case, we interpret $\sigma(P)$ as a desired total charge distribution density on S, that is $\sigma(P) := \{\sigma_i(P),\ P \in S_i\}_{i=1}^{m}$.

It is possible to research the solvability of integral equation (1) in various functional spaces. However, it should be taken into account the specificity of investigated physical phenomenon. In this connection, the modelling of electrostatic field in the substantially spatial setting foresees the account of desired charge distribution density $\sigma(P)$ behavior near the contour of unclosed surface S. As to right hand side of (1), we consider that $U(M),\ M \in S$, is the given boundary value of potential on an electrode which is actually simulated by a surface S ($U(M) \equiv const$). At last, the solvability of (1) can be expressed by the following inequalities [4, 6]:

$$m_1\|\sigma\|_{H_{00}^{-1/2}(S)} \le \|A\sigma\|_{H^{1/2}(S)} \le m_2\|\sigma\|_{H_{00}^{-1/2}(S)}\ (0 < m_1 \le m_2),$$

where $H^{1/2}(S)$ is a trace space, $H_{00}^{-1/2}(S)$ is dual space with respect to $H_{00}^{1/2}(S)$. Note that S is an open surface treated as a component of some close surface Σ. In addition, $H_{00}^{1/2}(S)$ is different from $H^{1/2}(S)$, and in the case of smooth S, relevant norm may be defined as

$$\|\sigma\|^2_{H_{00}^{1/2}(S)} = \|\sigma\|^2_{H^{1/2}(S)} + \|\rho^{-1/2}\sigma\|^2_{L_2(S)},$$

where $\rho(M)$ is the distance from $M \in S$ to the smooth edge ∂S.

2. The numerical scheme for model problem testing

Let us consider the calculation problem of plane-parallel condenser electrostatic field. From mathematical model point of view this condenser can be represented as a surface S, which is an aggregate of two parallel identical plates S_1 and S_2 situated symmetrically with respect to a coordinate plane XY, so that $S := S_1 \bigcup S_2$. The distance between them equals $2h$. Suppose that U_1 and U_2 are the given potential values on S_1 and S_2, respectively. The electrostatic treatment of problem (1) means that U_1 and U_2 are arbitrary constant. As we mentioned in [2], this problem is not trivial, and the results of calculation are especially sensitive with respect to variation of output data.

With a view to analyze integral equation (1) let us use such S_l representation

$$S_l := \Big\{(x,y,z)^\top \in \mathbb{R}^3 \Big| (x,y) \in [-1,1]^2;\ z = (-1)^{l-1}h;\ l = \overline{1,2};\ h > 0\Big\}. \quad (2)$$

According to (2), we can represent S in the form of congruent components combination:

$$S = \bigcup_{l=1}^{2}\Big(\bigcup_{k=1}^{4} S_{lk}\Big).$$

Taking into account subdivision of S_1 and S_2, integral equation (1), in its turn, can be formally represented as

$$\sum_{l=1}^{2}\sum_{k=1}^{4}\iint\limits_{S_{lk}} \sigma_{lk}(P)|P-M|^{-1}dS_p = U(M) = \begin{cases} U_1, & M \in S_1, \\ U_2, & M \in S_2, \end{cases} \quad (3)$$

where $\sigma_{lk}(P) := \sigma_{lk}(x,y)$ is the restriction of $\sigma(P)$ onto S_{lk};

$$M := (x_0, y_0, z_0 = \pm h)^\top;\ (x,y), \quad (x_0,y_0) \in [-1,1]^2.$$

Then, applying in (3) some changes of variables, we realize the conversion from integration over S to integration over its congruent constituent S_{11}. As a result, we get the system of eight linear integral equations with respect to unknown density $\sigma_j(x,y)(j=\overline{1,8})$, according to the chosen group of surface S symmetry:

$$\sum_{j=1}^{8} \iint\limits_{\triangle_1} \sigma_j(x,y) G_{|i-j|+1}(x,y;x_0,y_0;h)dxdy = U(M_i), \quad (i=\overline{1,8}). \tag{4}$$

Here, $\triangle_1 := [0,1]^2$; $M_i := \Big((-1)^{r-1}x_0, (-1)^{s-1}y_0, (-1)^{p-1}h\Big)^\top \in S_{pq}$; in this case $i := 4(p-1)+2(r-1)+s$, and $q := 2(r-1)+s$ with $p,r,s = 1,2$; M_i are the points of collocation; $(x_0,y_0) \in \triangle_1$. The point of integration is

$$P := \Big((-1)^{n-1}x, (-1)^{m-1}y, (-1)^{l-1}h\Big)^\top \in S_{lk};$$

in this case, $j := 4(l-1)+2(n-1)+m$, and $k := 2(n-1)+m$ with $l,n,m = 1,2$; and finally

$$G_{|i-j|+1}(x,y;x_0,y_0;h) := |P - M_i|^{-1}.$$

It is easy to see that the system of integral equations (4) may be written in the form of matrix operator equation

$$A\overline{\sigma} = \overline{U}, \tag{5}$$

where

$$\overline{\sigma} := (\sigma_1(x,y), \sigma_2(x,y), \ldots, \sigma_8(x,y))^\top,$$
$$\overline{U} := (U(M_1), U(M_2), \ldots, U(M_8))^\top;$$

and $A := (A_{ij})_{i,j=1}^8$, in this case, A_{ij} is an integral operator that acts by the rule

$$A_{ij}\sigma_j(M_i) \equiv \iint\limits_{\triangle_1} \sigma_j(x,y) G_{|i-j|+1}(x,y;x_0,y_0;h)dxdy.$$

Since an initial integral equation has an Abelian eighth order group of symmetry [7], then, we can split (5) into eight independent integral equations $A'\overline{\sigma}' = \overline{U}'$, where $A' := F \cdot A \cdot F^{-1}$, $\overline{\sigma}' := F\overline{\sigma}$, $\overline{U}' := F\overline{U}$. Here, $F := (F_{ij})_{i,j=1}^8$ is known matrix of Fourier transform [2,7]; $A' := (A'_i)_{i=1}^8$, in this case,

$$A'_i\sigma'_i(M_i) \equiv \iint\limits_{\triangle_1} \sigma'_i(x,y) R_i(x,y;x_0,y_0;h)dxdy,$$

$$R_i(x,y;x_0,y_0;h) := \sum_{j=1}^{8} F_{ij} G_{|i-j|+1}(x,y;x_0,y_0;h),$$

$$\sigma'_i(x,y) := \sum_{j=1}^{8} F_{ij}\sigma_j(x,y), \qquad U'(M_i) := \sum_{j=1}^{8} F_{ij}U(x,y).$$

Solving every of independent integral equations, as the final result, it is possible to reproduce $\sigma_j(x,y)$.

Then, without loss of generality let us consider one special case of integral equation (5) presentation. Namely, taking into account the antisymmetry of boundary values of potentials on condenser plates ($U := U_1 = -U_2$), and in accordance with this similar properties of (5) solutions, it is possible to represent (5) in the form as

$$(A\sigma)(x_0,y_0) \equiv \iint\limits_{\triangle_1} \sigma(x,y)\hat{R}(x,y;x_0,y_0;h)dxdy = U(x_0,y_0), \qquad (x_0,y_0) \in (0,1)^2, \tag{6}$$

where

$$\hat{R}(x,y;x_0,y_0;h) := \sum_{l=1}^{2}(-1)^{l-1}\sum_{p=1}^{2}\sum_{k=1}^{2}\Big\{4h^2(l-1)+ +\big[(-1)^{p-1}x+(-1)^{k-1}x_0\big]^2+\big[y+(-1)^k y_0\big]^2\Big\}^{-\frac{1}{2}}.$$

It is easy see that integral equation (6) is an equation with weak singularity in the kernel. In addition, (6) has mentioned singularity only in one item of the sum $\hat{R}(x,y;x_0,y_0;h)$, where $k=l=1,\ p=2$. Moreover, in the process of numerical scheme constructing it is necessary to take into consideration special behavior of desired solution only on S_{11}.

It is known [5] that desired solution $\sigma(x,y)$ has singularities in the neighborhood of S_{11} corner point and at the points which border on a straight edge of S_{11}. In the first case, the charge singularity is proportional to $\rho^{-0,7034}$, and, in the second case, the charge singularity is proportional to $\rho^{-0,5}$, where ρ is the distance from the vertex and straight edge of S_{11}, respectively. These singularities can be expressed by the following weight function

$$\frac{(1-x)^\gamma+(1-y)^\gamma}{\big[(1-x)(1-y)\big]^{1/2}} \qquad (\gamma = 0,2966).$$

This function is applied for mentioned singularities isolation in the notation of charge distribution density $\sigma(x,y)$. But such accounting of desired solution characteristics is rather complicated from practical point of view. So, we apply the different method, based on progressive analysis and correction of obtained results.

Using the collocation method under the condition of piecewise-constant approximation of desired density $\sigma(x,y)$, two-dimensional integral equation (6) was reduced to the following system of linear algebraic equations

$$\sum_{j=1}^{N_x}\sum_{i=1}^{N_y}\sigma_{ij}\int\limits_{x_i-\frac{H_x}{2}}^{x_i+\frac{H_x}{2}}\int\limits_{y_j-\frac{H_y}{2}}^{y_j+\frac{H_y}{2}}\hat{R}(x,y;x_0,y_0,h)dxdy = U(x_0,y_0),$$

where $H_x := N_x^{-1}$, $H_y := N_y^{-1}$ $(N_x,\ N_y \in \mathbb{N})$;

$$x_0 \in \left\{\frac{H_x}{2}(2i-1)\right\}_{i=1}^{N_x}, \quad y_0 \in \left\{\frac{H_y}{2}(2j-1)\right\}_{j=1}^{N_y};$$

σ_{ij} are approximate values of desired density $\sigma(x,y)$ at the points of collocation (x_0, y_0). In this case, we used uniform subdivision of S_{11} onto elements, that is $H_x = H_y$, and $N_x = N_y$.

3. A posteriori error estimation of (6) numerical solution under the condition of $\triangle_1$ irregular partition onto elements

In numerical solving of integral equation (6) the problem of obtained results error estimation is actual from practical point of view. Taking into account a priori information of desired density special behavior, the method based on experience proved to be the most acceptable. Let us note that stable results obtaining is also important problem independently of S_{11} uniform or nonuniform partition onto elements.

Let $\sigma_\varepsilon(P)$ be a numerical solution of integral equation (6) that belongs to the chosen approximation space. It generates approximate potential value at arbitrary point Q between charged condenser electrodes simulated by appropriate surfaces:

$$U_\varepsilon(Q) = (A\sigma_\varepsilon)(Q).$$

In addition, general error function e_U of integral equation (6) approximate solution may be represented as [1]

$$e_U = A\sigma - A\sigma_\varepsilon = A(\sigma - \sigma_\varepsilon) = Ae_\sigma,$$

where e_σ is a solution of such integral equation

$$(Ae_\sigma)(M) = U - (A\sigma_\varepsilon)(M), \quad M \in S_{11}. \tag{7}$$

Integral equation (6) solution has irregular behavior near the contour of unclosed surface S (essentially in the neighborhood of its corner points) [5]. Therefore, the reproduction of error function e_U, specified the level of boundary values satisfaction, is established onto elements D^e. These elements appear in the process of surface S sequential nonuniform partition (in the present case, its congruent component S_{11}). On D^e the function e_U may reach maximum values. Moreover, on D^e the function e_σ is approximately equal to its value at checking point T (see fig.1, fig.2):

$$e_\sigma(T) = \frac{U - (A\sigma_\varepsilon)(T)}{\int\limits_{D^e} |T-P|^{-1} dS_P}.$$

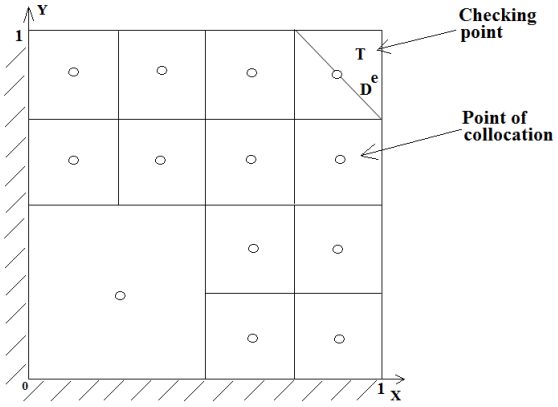


Fig. 1. Nonuniform partition

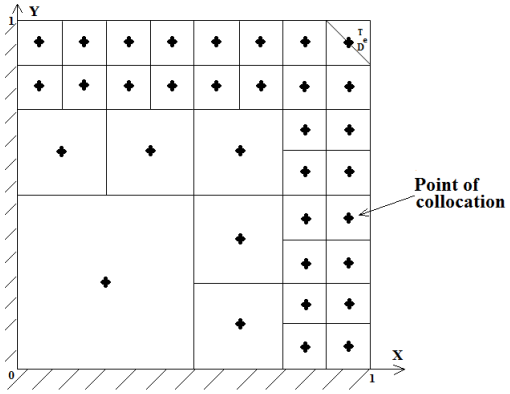


Fig. 2. Nonuniform partition in progress

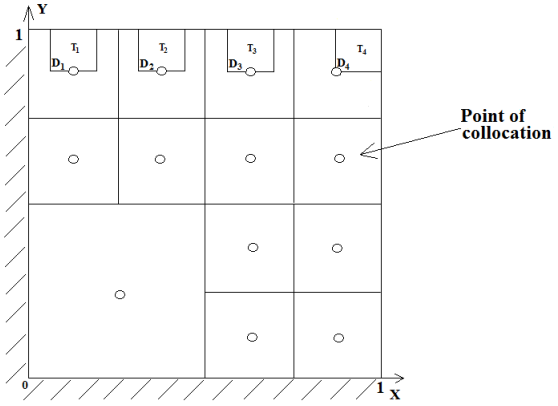


Fig. 3. Checking elements D_i

Selecting the furthest strategies of obtained results correction, it is possible to use various methodologies. Let us consider the method, different from proposed in the paper [2], which is sufficiently effective for two-dimensional integral equations numerical solution. The main idea of this strategy consists of the following. In the process of domain $\triangle_1$ nonuniform subdivision let us consider not

only one special element D^e but some set of elements where the desired function errors are inadmissible. Taking into account the symmetry of obtained results it is advisable to select such elements not far from the part of plate contour (for example, on the last horizontal layer). Let $D_1, D_2, \ldots, D_N$ $(N \in \mathbb{N})$ be above mentioned elements (see fig. 3). Then, if we use piece-wise approximation of e_σ and equation (7), it is possible to find solution error on every element D_i, $(i = \overline{1, N})$:

$$e_\sigma(T_i) = \frac{U - (A\sigma_\varepsilon)(T_i)}{\int_{D^e} |T_i - P|^{-1} dS_P}.$$

Let us denote by e_k the solution error e_σ on the element D_k $(k = \overline{1, N})$, that is $e_k = e_\sigma(T_k)$. Then, it needs to calculate the value ξ:

$$\xi = \sqrt{\frac{\sum\limits_{k=1}^{N} \|e_k\|^2}{N}}.$$

At that time for the completeness of domain $\triangle_1$ subdivision process the following condition must be fulfilled

$$\frac{\|e_k\|}{\xi} \cdot 100\% < TOL \quad \forall e_k,\ k = \overline{1, N}. \tag{8}$$

If the condition (8) is fulfilled only for certain elements D_k and appropriate errors e_k, then it is needed later on to eliminate such elements out of previous defined checking. Let us note that the disposition of elements D_k does not strictly allocate, so its sampling must be realized in various ways. In this connection, it is always necessary to control the obtained results of calculation.

4. The analysis of numerical experiments

Example 1. Illustration of calculation stability and analysis of results reliability Using piece-wise approximation of $\sigma(P)$ (charge distribution density) for $N_x = N_y = 40$ (the number of collocation points is 1600) we obtained the following results:

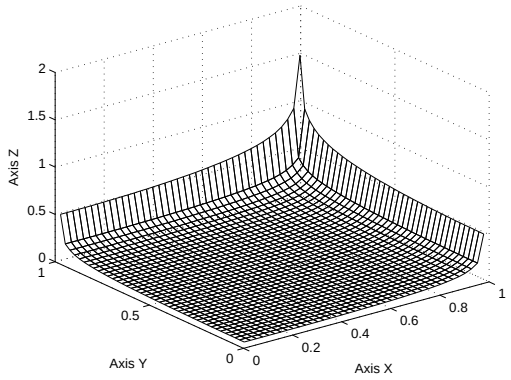


Fig. 4. Charge distribution density. $N_x = N_y = 40$

Let us note, that uniform subdivision of S_{11} can be selected so that there exists a point of collocation which will be present at the next division area. For example, the following divisions of $N_x = N_y = 6$, $N_x = N_y = 18$ contain the collocation point with coordinates (0.75 0.75). Justification of approximation schemes stability and hence the approximate solution results of integral equation solving are shown in the Tabl. 1. Approximative values of density $\sigma(x, y)$ at the checking points are not much different from the values which were obtained in the previous step of division.

Tabl. 1. Charge distribution density. Illustration of calculation stability

Point of collocation (x,y)	$N_x = N_y$		
	6	18	54
(0.250,0.250)	0.0531427	0.0519844	0.0515285
(0.250,0.917)	0.2005168	0.1459617	0.1482432
(0.917,0.250)	0.2005168	0.1459617	0.1482432
(0.917,0.917)	0.3955893	0.2588432	0.2699086

Absolute error e_U of reproduced boundary values for $N_x = N_y = 40$ is represented in the following figure:

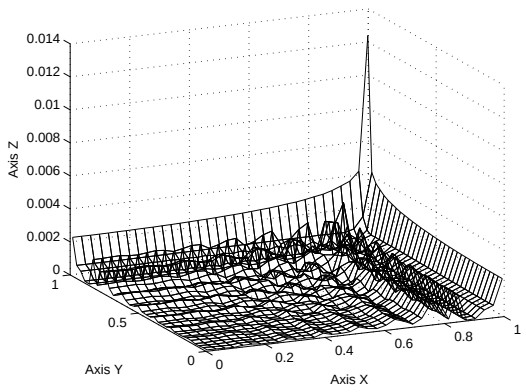


Fig. 5. Absolute error of boundary values. $N_x = N_y = 40$

Example 2. Illustration of nonuniform partition approach. The comparison between approaches Nonuniform partition is applied for better approximation of charge distribution density function and decreasing error function, especially near the contour of unclosed surface. Two parameters are important for this approach: the first one is initial partition of the surface, and the second is the number or steps of nonuniform partition; these parameters affect to the results of calculation. Absolute error of reprodused boundary values is shown in the Fig. 6, in the case when initial partition is $N_x = N_y = 2$. The number of iterations (steps) for nonuniform partition is 9.

The results in this figure reflect the impact of initial partition parameter to the error function: error was reduced near the contour of surface but was not decreased onto others elements. So, next figure displays the results of calculation with different initial partition.

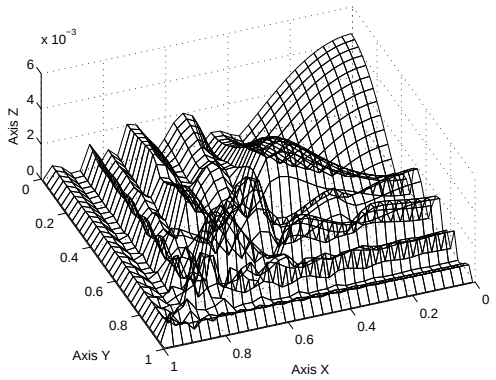


Fig. 6. Absolute error of reproduced boundary values. Nonuniform partition

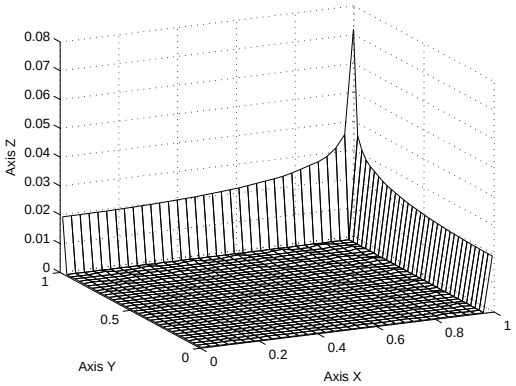


Fig. 7. Absolute error of boundary values. First partition ($N_x = N_y = 8$)

The Fig. 7 presents an absolute error of boundary values for the first partition $N_x = N_y = 8$ and the number of iteration for nonuniform division is 6.

The following two tables represent comparing of surface partition approaches (uniform and nonuniform) and summarize obtained results. The tables contain values of error function at checking points near the contour and comparison of these tables concludes that nonuniform partition is more effective for solving integral equations of such type.

Tabl. 2. Uniform partition

y/x	0.85	0.95	0.995	0.9995
0.85	0.00303	0.00124	0.08051	0.09443
0.95	-	0.02228	0.10008	0.11801
0.995	-	-	0.15808	0.17186
0.9995	-	-	-	0.18295

The Tabl. 2 represents $N_x = N_y = 8$. The number of collocation points is 64.

In the Tabl. 3 initial partition $N_x = N_y$ is equal to 2. The number of steps for nonuniform partition is 4. The number of collocation points is 79.

Tabl. 3: Nonuniform partition

y/x	0.85	0.95	0.995	0.9995
0.85	0.00385	0.01239	0.04937	0.06633
0.95	-	0.02971	0.04802	0.07266
0.995	-	-	0.110306	0.13008
0.9995	-	-	-	0.14557

So, by the example of the concrete model problem solving it is shown how, taking into account the specificity of initial data, to solve the problem of integral equation special representation. In the process of appropriate numerical scheme constructing such a representation gives the possibility to essentially simplify the use of a priori information on desired solution. The last is important for objectifying various correction procedures of obtained results on the basis of special estimators. With the help of proposed estimators the effective solution of initial integral equation were received.

Bibliography

1. Eriksson K. Introduction to Adaptiv Method for Differential Equation / K. Errikson, D. Estep, P. Hansbo, C. Johnson // Acta Numerica. – 1995. – P. 1-54.
2. Garasym Ya. S. On practical attainment of preassigned accuracy of results in the process of numerical analysis of some two-dimensional integral equations / Ya. S. Garasym, B. A. Ostudin // Jour. of Comput. and Appl. Math. – 2013. – № 3 (113). – P. 18-25.
3. Il'in V. P. The Numerical Methods of Emissive Electron-Optical Systems Optimization / V. P. Il'in, V. A. Kateshov, Yu. V. Kulikov, M. A. Monastyrskii. – Novosibirsk: Nauka, 1987. – 192 p. (in Russian).
4. Lions J.-L. Nonhomogeneous boundary-value problems and applications. Vol. 1. / J.-L. Lions, E. Magenes. – Berlin: Springer-Verlag, 1972.
5. Morrison J. A. Charge singularity at the corner of a flat plate / J. A. Morrison, J. A. Lewis // SIAM J. Appl. Math.– 1976.– Vol. 31, № 2.– P. 233-250.
6. Sybil Yu. M. Three dimensional elliptic boundary value problems for an open Lipschtz surface / Yu. M. Sybil // Матем. студії.– 1997.– Vol. 8, № 2.– P. 79-96.
7. Zakharov E. V. Abelian groups of finite order in numerical analysis of potential theory linear boundary value problems / E. V. Zakharov, S. I. Safronov, R. P. Tarasov // Journal of Numerical Mathematics and Mathematical Physics. – 1992. – Vol. 2, № 1. – P. 40-58. (in Russian).

B. Ostudin, Ya. Garasym, A. Beshley,
Ivan Franko National University of Lviv,
1, Universytets'ka Str., Lviv, 79000, Ukraine

Received 08.07.2015

Журнал обчислювальної та прикладної математики | 2015 № 2 (119) | Journal of Computational & Applied Mathematics

UDC 519.6

ARITHMETICAL COMPLEXITY OF MODIFIED FULLY DISCRETE PROJECTION METHOD FOR THE PERIODIC INTEGRAL EQUATIONS

E. V. SEMENOVA

Резюме. Розглядається задача скорочення обсягу інформаційних затрат при розв'язанні періодичних інтегральних рівнянь з мінімальною похибкою. Для цього пропонується деяка модифікація повністю дискретного проекційного методу. Доведено, що ця модифікація зберігає найкращу точність чисельного методу в метриці соболєвських просторів з обсягом арифметичних дій $N \log N$ за порядком.

Abstract. The reduction of arithmetical operations for the solving of periodic integral equations with minimal error bound is considered. For this some modification of a fully discrete projection method was proposed. It was proved that such modification guarantees the best possible accuracy of the numerical method in the metric of Sobolev spaces with the order of arithmetical operations $N \log N$.

1. Introduction

Periodic integral equations are frequently found in various problems of natural sciences that can be described by a boundary value problems such as Laplace or Helmholz equations. To illustrate this, we rewrite Dirichlet problem for Laplace equation on the simply connected domain Ω. So it takes the form

$$\triangle G(X) = 0, \quad X \in \Omega, \tag{1}$$

$$G(X) = g(X), \quad X \in \Gamma = \partial\Omega, \tag{2}$$

where Γ is a smooth boundary of domain Ω and function g is continuous. As it is well-known (see [8]), the problem (1) has a unique solution under quite natural condition on Γ. Solving (1) by direct method, using the representation of the function $G(X)$, $X \in \Omega$ in the form of a simple-layer potential, we derive to a boundary integral equation

$$Su = g, \tag{3}$$

where S is a single layer operator with logarithmic kernel and $u = \dfrac{\partial G}{\partial n}$ is a normal derivation on the boundary. Note that by so-called Cauchy data $(G|_\Gamma, \dfrac{\partial G}{\partial n}\big|_\Gamma)$ we can easily find the function $G(X)$ for $X \in \Omega$. Thus for solving boundary value problem (1) it is necessary to solve periodic integral equation of the first kind (3). It is such kind of problem that will be the object of our investigation. Periodic integral equations are well-known and various aspects of their solving in the metric of Sobolev spaces were investigated, for example,

Key words. Periodic integral equations, fully discrete projection method, GMRES.

in [2], [4], [7]. The most widely-used approaches for numerical solving of periodic integral equations are fully discrete collocation and projection methods that applied together with selfregularization principle. In the paper we will consider modification of a fully discrete projection method that was firstly proposed for solving the integral Symm equation (see Example 1) in [4] and extended on the class of pseudodifferentional equation in [12]. Moreover we introduce some additional projection in the method to reduce amount of arithmetical operations.

2. Statement of the problems

In the space $L_2(0,1)$ we consider the following integral equation

$$\mathcal{A}u(t) = f(t), \quad t \in [0,1], \tag{4}$$

where f is 1- periodic function and operator $\mathcal{A}$ has the form

$$\mathcal{A} = \sum_{p=0}^{q} A_p, \quad A_p u(t) = \int_0^1 k_p(t-s) a_p(t,s) u(s) ds. \tag{5}$$

Let's denote by $C^\infty = C^\infty([0,1]^2)$ the space C^∞ of smooth 1-biperiodic functions of both variables. Suppose that $a_p \in C^\infty([0,1]^2)$, $p = 0, \dots, q$, and

$$a_0(t,t) \neq 0, \forall t \in [0,1]. \tag{6}$$

Moreover assume that $k_p(t)$ is 1 - periodic function with known Fourier coefficients $\hat{k}_p(n)$ by trigonometric basis for each $p = 0, \dots, q$. Additionally we suppose that for some $\alpha \in \mathbb{R}$ and $\beta > 0$ the following inequalities

$$c_{00}|n|^\alpha \leq |\hat{k}_0(n)| \leq c_0|n|^\alpha, \quad n \in \mathbb{Z}/0, \tag{7}$$

$$|\hat{k}_0(n) - \hat{k}_0(n-1)| \leq c\underline{n}^{\alpha-\beta}, \quad n \in \mathbb{Z}, \tag{8}$$

$$|\hat{k}_p(n)| \leq c\underline{n}^{\alpha-\beta}, \quad n \in \mathbb{Z}, p = 1, \dots, q, \tag{9}$$

hold true, where c, c_0, $c_{00} > 0$ and

$$\underline{n} = \begin{cases} |n|, & n \in \mathbb{Z}/0 \\ 1, & n = 0 \end{cases}.$$

Denote by H^{λ_1} and H^{λ_1,λ_2}, $-\infty < \lambda_1, \lambda_2 < \infty$, Hilbert spaces of 1-periodic functions and 1-biperiodic functions with the norm

$$\|u\|_{\lambda_1} := \left(\sum_{n \in \mathbb{Z}} |\underline{n}|^{2\lambda_1} |\hat{u}(n)|^2 \right)^{1/2} < \infty,$$

$$\|a\|_{\lambda_1,\lambda_2} := \left(\sum_{(k,l) \in \mathbb{Z}^2} |\underline{k}|^{2\lambda_1} |\underline{l}|^{2\lambda_2} |\hat{a}(k,l)|^2 \right)^{1/2} < \infty$$

respectively. Here

$$\hat{u}(n) = \int_0^1 e_{-n}(t) u(t) dt, \quad \hat{a}(k,l) = \int_0^1 \int_0^1 e_{-k}(t) e_{-l}(s) a(t,s) dt ds$$

are Fourier coefficients of functions $u(t)$ and $a(t,s)$ by trigonometric basis $\{e_k\}_{k=-\infty}^{+\infty}$, where $e_k(t) = e^{i2\pi kt}$, $t \in [0,1]$.

In general case in the space $H^0 = L_2(0,1)$ operator $\mathcal{A}$ is compact and problem is unstable. But for considered class of equations (4) with (6)-(9) it is possible to choose appropriate pair of spaces to regularized problem. As it was shown in [7, Theorem 6.3.1], operator $\mathcal{A}$ under our assumptions creates isomorphism between H^λ and $H^{\lambda-\alpha}$ for any $\lambda \in \mathbb{R}$. That is why if $f \in H^{\lambda-\alpha}$ the equation (4) has unique solution $u \in H^\lambda$. Let's consider more precisely the structure of (4). Following [7, Ch.6], we rewrite the equation (4) in a such way

$$\mathcal{A}u = Du + \sum_{p=1}^{q} A'_p u = f', \tag{10}$$

where $Du = \int_0^1 k_0(t-s)u(s)ds$, $A'_0 := A'_0 = \frac{1}{a_0(t,t)}\int_0^1 k(t-s)(a_0(t,s) - a_0(t,t))u(s)ds$, $A'_p := A'_p = \frac{A_p}{a_0(t,t)}$ for $p = 1..q$ and $f := f' = \frac{f}{a_0(t,t)}$. Note that $D \in \mathcal{L}(H^\lambda, H^{\lambda-\alpha})$ is performing the isomorphism between the spaces H^λ and $H^{\lambda-\alpha}$ and operators $A_p \in \mathcal{L}(H^\lambda, H^{\lambda-\alpha+\beta}), p = 0,..,q$ are compact on the pair of spaces H^λ and $H^{\lambda-\alpha}$. Further we will deal with equation (10) instead of (4).

Thereafter for all $\lambda \le \mu$ there are constants $c'_\lambda, c''_\lambda > 0$, such that for any $v \in H^\lambda$ the following inequality

$$c'_\lambda \|v\|_\lambda \le \|\mathcal{A}v\|_{\lambda-\alpha} \le c''_\lambda \|v\|_\lambda \tag{11}$$

holds true.

Further we assume that exact solution of equation (4) belongs to some Sobolev spaces, namely $u \in H^\mu$ for some $\mu > \alpha + 1/2$ and $\|u\|_\mu \le 1$. Then due to conditions (11) we have that $f \in H^{\mu-\alpha}$ and $\|f\|_{\mu-\alpha} \le c''_\mu$.

Note that classical elliptic pseudodifferential equations are included in the class of equations (4) with conditions (6)- (9) (see for detail [6]). Below we rewrite the examples of some equations that satisfy the conditions (6)- (9).

Example 1. The typical example of equation from the class under consideration is an integral Symm's equation

$$\mathcal{A}u(t) := \int_0^1 k_0(t-s)u(s)ds + \int_0^1 a_1(t,s)u(s)ds = f(t), \tag{12}$$

$$k_0(t-s) = \log|\sin\pi(t-s)|, \tag{13}$$

$$a_1(t,s) = \begin{cases} \log\frac{|\gamma(t)-\gamma(s)|}{|\sin\pi(t-s)|}, & t \ne s \\ \\ \log(|\gamma'(t)/\pi|), & t = s \end{cases}.$$

As it is known, the kernel $a_1(t,s)$ of operator A_1 presents the C^∞-smooth and 1-biperiodic function and Fourier coefficients k_0 have the view

$$\hat{k}_0(n) = \begin{cases} \frac{1}{2|n|}, & n \in \mathbb{Z}/0 \\ \log 2, & n = 0. \end{cases}$$

It is evident that conditions (6)-(9) are satisfied for $a_0(t,s) = k_1(t,s) \equiv 1$, $\alpha = -1$ and any $\beta > 0$.

Example 2. The integral equation

$$\int_0^1 |x(t) - x(s)|^2 \log|x(t) - x(s)| u(s) ds = f(t), \quad t \in [0,1],$$

arises for solving biharmonic Dirichlet problems in the bounded domain with smooth Jordan boundary (see for more detailed information, for example, [1], [7, Ch. 6]). Rewrite the equation in the form

$$\int_0^1 k_0(t-s) a_0(t,s) u(s) ds + \int_0^1 a_1(t,s) u(s) ds = f(t),$$

where

$$a_0(t,s) = \frac{|x(t) - x(s)|^2}{\sin^2 \pi(t-s)} \quad \text{for} \quad t \neq s, \quad a_0(t,t) = \frac{|x'(t)|^2}{\pi^2},$$

$$a_1(t,s) = |x(t) - x(s)|^2 \log \frac{|x(t) - x(s)|}{|\sin \pi(t-s)|} \quad \text{for} \quad t \neq s, \quad a_1(t,t) \equiv 0,$$

$$k_0(t) = \sin^2 \pi t \log |\sin \pi t|.$$

The Fourier coefficients k_0 are known and have the following view $\hat{k}_0(0) = -\frac{1}{2}\log 2 + \frac{1}{4}$,
$\hat{k}_0(\pm 1) = \frac{1}{4}\log 2 - \frac{3}{16}$,

$$\hat{k}_0(n) = \frac{1}{4|n|(n^2-1)}, \quad |n| \geq 2.$$

It is easy to see that conditions (7)-(9) satisfied for $\alpha = -3, \beta = 1$. Thus, the equation under consideration is also included in the investigated class of problems.

To make more precise the smoothness properties of functions $a_p, p = 0, \ldots, q$, we introduce in consideration the space of Gevre function of Roumieu type (see [3, p.112]):

$$G_{\eta_1,\eta_2} = \Big\{ a \in C^\infty : \|a\|^2_{\eta_1,\eta_2} := \\ := \sum_{k,l=-\infty}^{\infty} |\hat{a}(k,l)|^2 e^{2\eta_2(|k|^{1/\eta_1} + |l|^{1/\eta_1})} < \infty \Big\}, \quad \eta_1, \eta_2 > 0. \tag{14}$$

Note that with $\eta_1 = 1$ by (14) it follows that function $a(t,s)$ has analytic continuations in both variables into the strip $\{z : z = t + is, |s| < \frac{\eta_2}{2\pi}\}$ of complex plane. Further suppose that $a_p \in G_{\eta_1,\eta_2}, p = 0, \ldots, q$, for some $\eta_1 \geq 1$ and $\eta_2 > 0$. It should be noted that condition (14) doesn't restrict the class of equation under consideration but allows to take better into account the smoothness of kernels a_p. At first such assumption for a_p was proposed in the paper [4], which considered particular case of mentioned class of periodic integral equations, namely Symm integral equation.

In the paper we state the aim to reduce the amount of arithmetical operations of fully discrete projection method for solving (4) with conditions (7)-(9) and (14). For that we propose modification of the method that should not influence

on the best error accuracy of solution for a priori case of choosing regularization parameter.

3. Auxiliary statements

For further presentation of our results we will use the following notations. Let's introduce n-dimensional subspaces of trigonometric polynomials

$$\mathcal{T}_N = \{u_N : u_N(t) = \sum_{k \in Z_N} c_k e_k(t)\},$$

$$Z_N = \left\{k : -\frac{N}{2} < k \le \frac{N}{2}, k = 0, \pm 1, \pm 2, \dots\right\}. \tag{15}$$

Denote by P_N and P_Ω orthogonal projectors

$$P_N u(t) = \sum_{k \in Z_N} \hat{u}(k) e_k(t) \in \mathcal{T}_N,$$

$$P_{\Omega_N} a(t,s) = \sum_{l,k \in \Omega_N} \hat{a}(k,l) e_k(t) e_l(s) \in \mathcal{T}_N \times \mathcal{T}_N,$$

where Ω_N is some domain on coordinate plane restricted by square $(-N/2, N/2] \times (-N/2, N/2]$. Also denote by Q_N and $Q_{N,N}$ interpolation projectors, such that $Q_N u(t) \in \mathcal{T}_N$, $Q_{N,N} a(t,s) \in \mathcal{T}_N \times \mathcal{T}_N$ and on the uniform grid it holds true

$$(Q_N u)(jN^{-1}) = u(jN^{-1}), \quad j = 1, 2, \dots, N,$$

$$(Q_{N,N} a)(jN^{-1}, iN^{-1}) = a(jN^{-1}, iN^{-1}), \quad j, i = 1, 2, \dots, N.$$

It is well-known (see, for example, [7, Ch.8]), that

$$\|u - P_N u\|_\lambda \le \left(\frac{N}{2}\right)^{\lambda - \mu} \|u\|_\mu, \quad \lambda \le \mu, \quad u \in H^\mu, \tag{16}$$

$$\|u - Q_N u\|_\lambda \le c_{\lambda,\mu} N^{\lambda-\mu} \|u\|_\mu, \quad 0 \le \lambda \le \mu, \quad \mu > \frac{1}{2}, \quad u \in H^\mu, \tag{17}$$

where $c_{\lambda,\mu} = \left(\frac{1}{2}\right)^{\lambda-\mu} \gamma_\mu$, and $\gamma_\mu = \left(1 + 2\sum_{j=1}^{\infty} \frac{1}{j^{2\mu}}\right)^{\frac{1}{2}}$.

Moreover, for any $v_N \in \mathcal{T}_N$ according to inverse Bernshtein inequality it holds

$$\|v_N\|_\mu \le \left(\frac{N}{2}\right)^{\mu - \lambda} \|v_N\|_\lambda, \quad \lambda \le \mu. \tag{18}$$

4. Discretization of operator A_p, $p = 0, \dots, q$

Note that operator D has simple structure and doesn't need any additional discretization. Thus we need to discretize only operators A_p for each $p = 0, \dots, q,$. This will be done further.

Let's consider the following domain of coordinate plane

$$D_M^{\eta_1} = \{(k,l) : |k|^{1/\eta_1} + |l|^{1/\eta_1} < \left(\frac{M}{2}\right)^{\frac{1}{\eta_1}}, \ k, l = 0, \pm 1, \pm 2 \dots\} \tag{19}$$

Note that $D_M^{\eta_1} \subseteq D_M^1$ for all $\eta_1 \ge 1$.

Assume that the discrete information about kernels $a_p(t,s)$ and right hand side f is given in the knots of uniform grids $\left(\frac{j_1}{M}, \frac{j_2}{M}\right)$, where $j_1, j_2 = 1..M$.

Let's approximate the kernels a_p in the following way

$$a_{p,M} = P_{D_M^{\eta_1}} Q_{M,M} a_p, \tag{20}$$

where $P_{D_M^{\eta_1}}$ is ortoprojector on span of vectors $\{e_i, e_j\}$ such that $(i,j) \in D_M^{\eta_1}$.

Then the operators $A_{p,M}$ can be approximate by

$$A_{p,M}u(t) = \int_0^1 k_p(t-s)a_{p,M}(t,s)u(s)ds. \tag{21}$$

where function $a_{p,M}$ has the form (20). To find the approximative properties of operator (21) we state the following auxiliary lemmas.

Lemma 1. *Let $a \in G_{\eta_1,\eta_2}$ for $\eta_1 \geq 1$, then for $\forall \lambda_1, \lambda_2$ and*

$$M > 2\left(\frac{\max\{\lambda_1,\lambda_2\}\eta_1}{\eta_2}\right)^{\eta_1}$$

it holds true

$$\|a - P_{D_M^{\eta_1}} a\|_{\lambda_1,\lambda_2} \leq \left(\frac{M}{2}\right)^{\lambda_1+\lambda_2} e^{-2\eta_2\left(\frac{M}{2}\right)^{1/\eta_1}} \|a\|_{\eta_1,\eta_2}.$$

Proof. We rewrite the norm of element $a - P_{D_M^{\eta_1}} a$ in the following way

$$\begin{aligned}\|a - P_{D_M^{\eta_1}} a\|^2_{\lambda_1,\lambda_2} &\leq \|\sum_{|k|>0} \sum_{l:(k,l)\notin D_M^{\eta_1}} \hat{a}(k,l)e_k(t)e_l(s)\|^2_{\lambda_1,\lambda_2} = \\ &= \sum_{|k|>0} \sum_{l:(k,l)\notin D_M^{\eta_1}} |k|^{2\lambda_1}|l|^{2\lambda_2}|\hat{a}(k,l)|^2 = \\ &= \sum_{|k|>0} \sum_{l:(k,l)\notin D_M^{\eta_1}} |k|^{2\lambda_1}|l|^{2\lambda_2}|\hat{a}(k,l)|^2 e^-_{k,l}e^+_{k,l} =: S_1,\end{aligned}$$

where $e^{\pm}_{k,l} = e^{\pm 2\eta_2(|k|^{1\eta_1}+|l|^{1/\eta_1})}$. Further it is worth to estimate the norm of S_1 depending on values k and l.

At first we consider the case then $|k| < \frac{M}{2}$, $|l| < \frac{M}{2}$ and $(k,l) \notin D_M^{\eta_1}$. In the view of fact that $\max_{k,l\notin D_M^{\eta_1}} |k|^{2\lambda_1}|l|^{2\lambda_2}e^-_{k,l} = (\frac{M}{2})^{2(\lambda_1+\lambda_2)}e^{-4\eta_1\left(\frac{M}{2}\right)^{1/\eta_2}}$ we have

$$\begin{aligned}S_1 &= \sum_{|k|<\frac{M}{2}} \sum_{|l|<\frac{M}{2}:(k,l)\notin D_M^{\eta_1}} |k|^{2\lambda_1}|l|^{2\lambda_2}|\hat{a}(k,l)|^2 e^-_{k,l}e^+_{k,l} = \\ &= \left(\frac{M}{2}\right)^{2(\lambda_1+\lambda_2)} e^{-4\eta_2\left(\frac{M}{2}\right)^{1/\eta_1}} \|a\|^2_{\eta_1,\eta_2}.\end{aligned}$$

Let consider the element S_1 for the case $|k| < \frac{M}{2}$, $|l| \geq \frac{M}{2}$ and $(k,l) \notin D_M^{\eta_1}$, then

$$S_1 = \sum_{|k|<\frac{M}{2}} |k|^{2\lambda_1} \sum_{|l|\geq\frac{M}{2}:(k,l)\notin D_M^{\eta_1}} |l|^{2\lambda_2}|\hat{a}(k,l)|^2 e^-_{k,l}e^+_{k,l}.$$

Since the function $x^{2\nu}e^{-2\eta_2 x^{\frac{1}{\eta_1}}}$ has the maximum in the point $x_1 = \left(\frac{\nu\eta_1}{\eta_2}\right)^{\eta_1}$, then for all

$$|l| > \frac{M}{2} \geq \left(\frac{\lambda_2\eta_1}{\eta_2}\right)^{\eta_1}$$

it holds true

$$|l|^{2\lambda_2}e^{-2\eta_2|l|^{1/\eta_1}} < \left(\frac{M}{2}\right)^{2\lambda_2} e^{-2\eta_2\left(\frac{M}{2}\right)^{1/\eta_1}}.$$

With account of this we have

$$S_1 = \sum_{|k|<\frac{M}{2}} |k|^{2\lambda_1}e^{-2\eta_2|k|^{1/\eta_1}} \sum_{|l|\geq\frac{M}{2}:(k,l)\notin D_M^{\eta_1}} |l|^{2\lambda_2}|\hat{a}(k,l)|^2 e^{-}k,le_{k,l}^{+} \leq$$

$$\leq \left(\frac{M}{2}\right)^{2(\lambda_1+\lambda_2)} e^{-4\eta_2\left(\frac{M}{2}\right)^{1/\eta_1}} \|a\|^2_{\eta_1,\eta_2}.$$

For the third case when $|k| > \frac{M}{2}$, $l < \frac{M}{2}$, $(k,l) \notin D_M^{\eta_1}$ the estimation of S_1 can be found similar to the second one, namely we get

$$S_1 = \sum_{|k|>\frac{M}{2}} |k|^{2\lambda_1}e^{-2\eta_2|k|^{1/\eta_1}} \sum_{0<|l|<\frac{M}{2}:(k,l)\notin D_M^{\eta_1}} |l|^{2\lambda_2}e^{2\eta_2|l|^{1/\eta_1}}|\hat{a}(k,l)|^2 e_{k,l}^{+}$$

$$\leq \left(\frac{M}{2}\right)^{2(\lambda_1+\lambda_2)} e^{-4\eta_2\left(\frac{M}{2}\right)^{1/\eta_1}} \|a\|^2_{\eta_1,\eta_2}.$$

And in the last case when $|k| > \frac{M}{2}$, $|l| > \frac{M}{2}$, the element S_1 can be easily estimated as in the cases above, namely we have

$$S_1 = \sum_{|k|\geq\frac{M}{2}} \sum_{|l|\geq\frac{M}{2}} |k|^{2\lambda_1}|l|^{2\lambda_2}e_{k,l}^{-}|\hat{a}(k,l)|^2 e_{k,l}^{+}$$

$$\leq \left(\frac{M}{2}\right)^{2(\lambda_1+\lambda_2)} e^{-4\eta_2\left(\frac{M}{2}\right)^{1/\eta_1}} \|a\|^2_{\eta_1,\eta_2}$$

for $M \geq \left(\frac{\max\{\lambda_1,\lambda_2\}\eta_1}{\eta_2}\right)^{\eta_1}$.

Summarizing all cases considered above, we arrive to statement of lemma.

Lemma 2. *Let $a \in G_{\eta_1,\eta_2}$ for $\eta_1 \geq 1$, then for $\lambda_1, \lambda_2 > 1/2$ and*

$$M > 2\left(\frac{\max\{\lambda_1,\lambda_2\}\eta_1}{\eta_2}\right)^{\eta_1}$$

it holds true

$$\|a - P_{D_M^{\eta_1}}Q_{M,M}a\|_{\lambda_1,\lambda_2} \leq c_1\left(\frac{M}{2}\right)^{\lambda_1+\lambda_2} e^{-2\eta_2\left(\frac{M}{2}\right)^{1/\eta_1}} \|a\|_{\eta_1,\eta_2},$$

where $c_1 = z_1 + 1$, $z_1 := z_1(\lambda_1,\lambda_2) = \gamma_{\lambda_1} + \gamma_{\lambda_2} + \gamma_{\lambda_1}\gamma_{\lambda_2}$.

Proof. Due to simple transformation we have

$$\|a - P_{D_M^{\eta_1}} Q_{M,M} a\|_{\lambda_1,\lambda_2} \le \|P_{D_M^{\eta_1}}(a - Q_{M,M}a)\|_{\lambda_1,\lambda_2} + \|(I - P_{D_M^{\eta_1}})a\|_{\lambda_1,\lambda_2}.$$

For the further estimation we need the previous result, that was obtained in Lemma 2 [10]. Namely, for $\lambda_1, \lambda_2 > 1/2$ it holds true

$$\|a - Q_{M,M}a\|_{\lambda_1,\lambda_2} \le z_1 \left(\frac{M}{2}\right)^{\lambda_1+\lambda_2} e^{-2\eta_2\left(\frac{M}{2}\right)^{1/\eta_1}} \|a\|_{\eta_1,\eta_2}.$$

Using inequality above and lemma 1 we have

$$\|a - P_{D_M^{\eta_1}} Q_{M,M} a\|_{\lambda_1,\lambda_2} \le \|a - P^{\eta_1}_{D_M} a\|_{\lambda_1,\lambda_2} + \|P_{D_M^{\eta_1}}(Q_{M,M}a - a)\|_{\lambda_1,\lambda_2} \le$$

$$z_1 \left(\frac{M}{2}\right)^{\lambda_1+\lambda_2} e^{-2\eta_2\left(\frac{M}{2}\right)^{1/\eta_1}} \|a\|_{\eta_1,\eta_2} + \left(\frac{M}{2}\right)^{\lambda_1+\lambda_2} e^{-2\eta_2\left(\frac{M}{2}\right)^{1/\eta_1}} \|a\|_{\eta_1,\eta_2} \le$$

$$\le c_1 \left(\frac{M}{2}\right)^{\lambda_1+\lambda_2} e^{-2\eta_2\left(\frac{M}{2}\right)^{1/\eta_1}} \|a\|_{\eta_1,\eta_2},$$

what was to be proved.

For the further analysis we need following results

Proposition 2. *[7, Lemma 6.1.3] Let $k(t)$ be 1 - periodic function such that*

$$|\hat{k}(n)| \le c_0 \underline{n}^{\alpha} \quad n \in \mathbb{Z}. \tag{22}$$

Then for any $\lambda > \frac{1}{2}$ it fulfils

$$\left\| \int_0^1 k(t-s)v(t,s)ds \right\|_{\lambda-\alpha} \le c_0 2^{\lambda-\alpha+1} \gamma_{\lambda-\alpha} \|v\|_{\lambda,\lambda-\alpha},$$

where c_0 is some constant and $v(t,s)$ is 1-biperiodic function in Sobolev space $H^{\lambda,\lambda-\alpha}$.

Proposition 3. *[7, Lemma 6.1.1] For any $\lambda_1, \lambda_2 \ge \frac{1}{2}$, $u, a \in H^{\lambda_1,\lambda_2}$ it holds true*

$$\|au\|_{\lambda_1,\lambda_2} \le z_2 \|a\|_{\lambda_1,\lambda_2} \|u\|_{\lambda_1,\lambda_2},$$

where $z_2 := z_2(\lambda_1,\lambda_2) = 2^{\lambda_1+\lambda_2+2}\gamma_{\lambda_1}\gamma_{\lambda_2}$.

Further we need the following additional bounds. Namely using the propositions 2, 3 and integral representation of A_p it is easy to find that for any $\lambda_1 > 1/2$ and $\lambda_2 > 1/2$

$$\|A_p\|_{\lambda_1,\lambda_2} \le z_3 \|a_p\|_{\lambda_1,\lambda_2}, \tag{23}$$

where $z_3 := z_3(\lambda_1,\lambda_2) = 2^{\lambda_1+1}\gamma_{\lambda_1} z_2(\lambda_1,\lambda_2)$ is some increasing function. Now we are ready to prove the error of approximation for the operator $A_p \in \mathcal{L}(H^{\lambda}, H^{\lambda-\alpha})$ by $A_{p,M}$. The corresponding result is formulated in the lemma 3.

Lemma 3. *Let A_p has the form (5) for all $p = 0, \dots, q$ and the conditions (6)-(9) are fulfilled. Moreover we assume that $a_p \in G_{\eta_1,\eta_2}$, $p = 0..q$ for $\eta_1 \geq 1$ and $\eta_2 > 0$. Then for all $\lambda > \max\{\frac{1}{2}, \frac{1}{2} + \alpha\}$ and $M > 2\left(\frac{\eta_1}{\eta_2}\max\{\lambda, \lambda - \alpha\}\right)^{\eta_1}$ it holds true*

$$\|A_p - A_{p,M}\|_{\lambda,\lambda-\alpha} \leq c_2\|a\|_{\eta_1,\eta_2}\left(\frac{M}{2}\right)^{2\lambda-\alpha} e^{-2\eta_2\left(\frac{M}{2}\right)^{1/\eta_1}},$$

where $c_2 = c_1c_0 2^{\lambda-\alpha+1}\gamma_{\lambda-\alpha}z_2$.

Proof. Taking into account Lemma 1, the Propositions 2 and 3, we have

$$\begin{aligned}\|(A - A_{p,M})\|_{\lambda-\alpha} &= \|\int_0^1 k_p(t-s)(a_p - P_{D_m^{\eta_1}}Q_{M,M}a_p)(t,s)u(s)ds\|_{\lambda-\alpha} \leq \\ &\leq c_0 2^{\lambda-\alpha+1}\gamma_{\lambda-\alpha}\|(a_p - P_{D_m^{\eta_1}}Q_{M,M}a_p)(t,s)u(s)\|_{\lambda-\alpha} \leq \\ &\leq c_0 2^{\lambda-\alpha+1}\gamma_{\lambda-\alpha}z_2\|a_p - P_{D_m^{\eta_1}}Q_{M,M}a_p\|_{\lambda,\lambda-\alpha}\|u(s)\|_\lambda \leq \\ &\leq c_2\left(\frac{M}{2}\right)^{2\lambda-\alpha} e^{-2\eta_2\left(\frac{M}{2}\right)^{\frac{1}{\eta_1}}}\|a_p\|_{\eta_1,\eta_2}\|u(s)\|_\lambda,\end{aligned}$$

which was to be proved.

Corollary 3. *From Lemma 3 follows that*

$$\|\sum_{p=0}^{q} A_p - A_{p,M}\|_{\lambda,\lambda-\alpha} \leq c_2(q+1)\max_p\{\|a_p\|_{\eta_1,\eta_2}\}\left(\frac{M}{2}\right)^{2\lambda-\alpha} e^{-2\eta_2\left(\frac{M}{2}\right)^{\frac{1}{\eta_1}}}$$

Now we are ready to propose fully discrete method for solving equations under consideration.

5. Fully Discrete Projection Method

Taking into account representation (10), we approximate $\mathcal{A}$ as follows

$$\mathcal{A}_M = D + P_l\sum_{p=0}^{q} A_{p,M}P_l, \tag{24}$$

where $l = N^\tau$, for some $0 < \tau < 1$. Note that our approximate variant of $\mathcal{A}$ is distinguished from respective approximation from [12] by using additional projections P_l and $P_{D_m^{\eta_1}}$. Such projection helps to bound the amount of arithmetical operations. The right-hand side of equation (4) we approximate as following

$$f_N := Q_N f,$$

where $N > M$. The main idea of the fully discrete projection method (FDPM) for equation (4) consists in solving the equation

$$\mathcal{A}_M u_N := Du_N + P_l\sum_{p=0}^{q} A_{p,M}P_l u_N = Q_N f, \tag{25}$$

where $A_{p,M}$ has the view (21) and $u_N \in \mathcal{T}_N$ is considered as approximate solution of (4). Note that by virtue of (7) and (8), it holds true $A_{p,M} \in \mathcal{L}(H^\lambda, H^{\lambda-\alpha+\beta})$, $p = 0, \dots, q$.

Lemma 4. *Let the conditions of Lemma 3 be satisfied and* $f \in H^{\mu-\alpha}$. *Moreover operator* $\mathcal{A}_M$ *has the form (24). Then for all* $l \sim N^\tau$, $\tau \in [\frac{\mu-\lambda}{\mu-\lambda+\beta}, 1)$ *and* $\max\{\alpha+1/2, 1/2\} < \lambda < \mu$ *it holds true*

$$\|(\mathcal{A}-\mathcal{A}_{\mathcal{M}})\|_{\lambda,\lambda-\alpha} \le c_3 \left(\frac{N}{2}\right)^{\lambda-\mu} + c_4 e^{-2\eta_2\left(\frac{M}{2}\right)^{1/\eta_1}} \left(\frac{M}{2}\right)^{2\lambda-\alpha},$$

where

$$c_3 := 2(q+1)\max_p\{\|a_p\|_{\mu,\mu+\beta-\alpha}\} z_3(\mu, \mu+\beta-\alpha),$$

$$c_4 = c_2(q+1)\max_p\{\|a_p\|_{\eta_1,\eta_2}\}.$$

Proof. Due to simple transformation we have

$$\begin{aligned} \mathcal{A}-\mathcal{A}_M = (I-P_l)\sum_{p=0}^{q} A_p u + \\ + P_l\left(\sum_{p=0}^{q} A_{p,M} - \sum_{p=0}^{q} A_p\right) P_l + P_l \sum_{p=0}^{q} A_p (I-P_l). \end{aligned} \tag{26}$$

Consider each summand separately.

By virtue of the fact that $A_p \in \mathcal{L}(H^\lambda, H^{\lambda-\alpha+\beta})$ for $p = 0..q$ and taking into account (16) and (23) we find that

$$\|(I-P_l)\sum_{p=0}^{q} A_p u\|_{\lambda-\alpha} \le \left(\frac{l}{2}\right)^{\lambda-\mu-\beta} \|\sum_{p=0}^{q} A_p u\|_{\mu-\alpha+\beta} \le$$

$$\le \left(\frac{l}{2}\right)^{\lambda-\mu+\beta} (q+1)\max_p\{\|a_p\|_{\mu,\mu+\beta-\alpha}\} z_3(\mu, \mu+\beta-\alpha).$$

Because of $l = N^\tau$ and $N^{\tau(\lambda-\mu-\beta)} \le N^{\lambda-\mu}$ for $\tau \in \left[\frac{\mu-\lambda}{\mu-\lambda+\beta}; 1\right)$, one can derive the estimate

$$\|(I-P_l)\sum_{p=0}^{q} A_p u\|_{\lambda-\alpha} \le (q+1)\max_p\{\|a_p\|_{\mu,\mu+\beta-\alpha}\} z_3(\mu, \mu+\beta-\alpha)\left(\frac{N}{2}\right)^{\lambda-\mu}.$$

Similar estimate holds for third summand from (26), namely

$$\|P_l(\sum_{p=0}^{q} A_p (I-P_l)u)\|_{\lambda-\alpha} \le \|\sum_{p=0}^{q} A_p\|_{\lambda-\beta,\lambda-\alpha} \|(I-P_l)u\|_{\lambda-\alpha} \le$$

$$\le \left(\frac{l}{2}\right)^{\lambda-\mu-\beta} (q+1)\max_p\{\|a_p\|_{\lambda-\beta,\lambda-\alpha}\} z_3(\lambda-\beta, \lambda-\alpha) \le$$

$$\le \left(\frac{N}{2}\right)^{\lambda-\mu} (q+1)\max_p\{\|a_p\|_{\lambda-\beta,\lambda-\alpha}\} z_3(\lambda-\beta, \lambda-\alpha).$$

The second summand from (26) we estimate with help of Lemma 3:

$$\|P_l(\sum_{p=0}^{q} A_{p,M} - \sum_{p=0}^{q} A_p)P_l\|_{\lambda,\lambda-\alpha} \leq$$

$$\leq c_2(q+1)\max_p\{\|a_p\|_{\eta_1,\eta_2}\}e^{-2\eta_2\left(\frac{M}{2}\right)^{1/\eta_1}}\left(\frac{M}{2}\right)^{2\lambda-\alpha}.$$

Combing the corresponding bounds we get the statement of the lemma.

Lemma 5. *Let the conditions of Lemma 3 are fulfilled. Then for any* $\lambda \in (\max\{\alpha+1/2, 1/2\}, \mu)$ *and for sufficiently small* N *and* M *such that*

$$c_3\left(\frac{N}{2}\right)^{\lambda-\mu} + c_4 e^{-2\eta_2\left(\frac{M}{2}\right)^{1/\eta_1}}\left(\frac{M}{2}\right)^{2\lambda-\alpha} < \frac{c'_\lambda}{2}$$

it holds true

$$\|v\|_\lambda \leq d_\lambda \|\mathcal{A}_M v\|_{\lambda-\alpha},$$

where $d_\lambda = \frac{2}{c'_\lambda}$.

Proof. Using the inequality (11) and lemma 4 we have

$$\|v\|_\lambda \leq \frac{1}{c'_\lambda}\|\mathcal{A}v\|_{\lambda-\alpha} \leq \frac{1}{c'_\lambda}(\|\mathcal{A}_M v\|_{\lambda-\alpha} + \|(\mathcal{A}-\mathcal{A}_M)v\|_{\lambda-\alpha}) \leq$$

$$\leq \frac{1}{c'_\lambda}\frac{\|\mathcal{A}_M v\|_{\lambda-\alpha}}{1-\frac{1}{c'_\lambda}\left(c_3\left(\frac{N}{2}\right)^{\lambda-\mu} + c_4 e^{-2\eta_2\left(\frac{M}{2}\right)^{1/\eta_1}}\left(\frac{M}{2}\right)^{2\lambda-\alpha}\right)} \leq \frac{2}{c'_\lambda}\|\mathcal{A}_M v\|_{\lambda-\alpha},$$

which was to be proved.

The estimation of accuracy for FDPM on the class of problems (4)-(9) with nonperturbed input data is established in the following assertion (see for detail [10]).

Theorem 1. *Let the conditions (6)- (9) are fulfilled, and operator* $\mathcal{A}_M$ *has the form (24). Then for any* $\lambda \in (\max\{1/2+\alpha, 1/2\}, \mu)$, $\mu > \alpha + 1/2$ *and for all*

$$\begin{gathered} M, N : M > 2\left(\frac{\eta_1}{\eta_2}\max\{\lambda, \lambda-\alpha\}\right)^{\eta_1}, \\ c_3\left(\frac{N}{2}\right)^{\lambda-\mu} + c_4 e^{-2\eta_2\left(\frac{M}{2}\right)^{1/\eta_1}}\left(\frac{M}{2}\right)^{2\lambda-\alpha} < \frac{c'_\lambda}{2} \end{gathered} \tag{27}$$

it holds true

$$\|u-u_N\|_\lambda \leq c_5\left(\frac{N}{2}\right)^{\lambda-\mu} + c_6 e^{-2\eta_2\left(\frac{M}{2}\right)^{1/\eta_1}}\left(\frac{M}{2}\right)^{2\lambda-\alpha}, \tag{28}$$

where $c_5 = 1 + d_\lambda c''_\lambda + d_\lambda c_3 + d_\lambda c''_\mu \gamma_{\mu-\alpha}$, $c_6 = d_\lambda c_4$.

Proof. Using the inequality (16) and $\|u\|_\mu \leq 1$ we find

$$\|u-u_N\|_\lambda \leq \|u-P_N u\|_\lambda + \|P_N u - u_N\|_\lambda \leq \left(\frac{N}{2}\right)^{\lambda-\mu} + \|P_N u - u_N\|_\lambda \tag{29}$$

Using Lemma 5 it is easy to find the bounds for second summand in (29), namely

$$\|P_N u - u_N\|_\lambda \leq d_\lambda \|\mathcal{A}_M(P_N u - u_N)\|_{\lambda-\alpha} \leq$$

$$\leq d_\lambda(\|(\mathcal{A}-\mathcal{A}_M)P_Nu\|_{\lambda-\alpha}+\|\mathcal{A}_Mu_N-\mathcal{A}P_Nu\|_{\lambda-\alpha})\leq$$
$$\leq d_\lambda(\|(\mathcal{A}-\mathcal{A}_M)P_Nu\|_{\lambda-\alpha}+\|Q_Nf-f\|_{\lambda-\alpha}+\|\mathcal{A}(P_Nu-u)\|_{\lambda-\alpha}).$$

Taking into account the lemma 4, inequalities (11), (16), (17) and the fact that $\|u\|_\mu\leq 1$ we have

$$\|P_Nu-u_N\|_\lambda\leq d_\lambda\Big(c_3\left(\frac{N}{2}\right)^{\lambda-\mu}+c_4e^{-2\eta_2(\frac{M}{2})^{1/\eta_1}}\left(\frac{M}{2}\right)^{2\lambda-\alpha}+$$
$$+c''_\mu\gamma_{\mu-\alpha}\left(\frac{N}{2}\right)^{\lambda-\mu}+c''_\lambda\left(\frac{N}{2}\right)^{\lambda-\mu}\Big).$$

Substituting the bound above in (29) we obtain the desired estimation.

Corollary 4. *As follows from (16), the optimal error of recovering the elements from $u\in H^\mu$, $\lambda<\mu$ is the following*

$$\|u-u_n\|_\lambda\leq n^{\lambda-\mu}\|u\|_\mu,$$

where $u_n\in\mathcal{T}_n$ is some approximation. From Theorem 1 follows that for $M\asymp\log^{\eta_1}N$ we have $\|u-u_N\|_\lambda\asymp\left(\frac{N}{2}\right)^{\lambda-\mu}$, that establish optimality of the method.

6. Calculation of arithmetical operations

Let construct the matrix corresponding to the element $P_lA_{p,M}P_lu_N(t)$. Using the fact that $\int k_0(t-s)e_i(s)ds=\hat{k}_0(i)e_i(t)$ we have

$$\begin{aligned}
&P_lA_{p,M}P_lu_N(t)=P_l\int_0^1k_p(t-s)Q_{M,M}a_p(t,s)P_lu_N(s)ds=\\
&=P_l\int_0^1k_p(t-s)\sum_{m,k\in D_M^{\eta_1}}\widehat{Q_{M,M}a_p}(m,k)e_m(t)e_k(s)\sum_{i\in\mathbb{Z}_l}\hat{u}(i)e_i(s)ds=\\
&=P_l\sum_{m,k\in D_M^{\eta_1},i\in\mathbb{Z}_l}\widehat{Q_{M,M}a_p}(m,k)\hat{u}(i)e_m(t)\int_0^1k_p(t-s)e_{k+i}(s)ds=\\
&=P_l\sum_{m,k\in D_M^{\eta_1},i\in\mathbb{Z}_l}\widehat{Q_{M,M}a_p}(m,k)\hat{k}_p(k+i)\hat{u}(i)e_{m+k+i}(t)=
\end{aligned}\tag{30}$$

To obtain the matrix form of FDPM (25) one can make the following substitution

$$\left|\begin{array}{c}m+k+i\to m\\ k+i\to k\end{array}\right|$$

and as the result get

$$P_lA_{p,M}P_lu_N(t)=\sum_{m\in\mathbb{Z}_l}\left[\sum_{i\in\mathbb{Z}_l}\Lambda_{m,i}^{p,\eta_1}\hat{u}(i)\right]e_m(t),$$

where

$$\Lambda_{m,i}^{p,\eta_1}=\sum_{(m-k,k-i)\in D_M^{\eta_1},k\in\mathbb{Z}_{M+l}}\widehat{Q_{M,M}a_p}(m-k,k-i)\hat{k}_p(k).$$

Thus, the equation (25) can be rewritten as the system of linear equations

$$D\bar{u} + \sum_{p=0}^{q} \Lambda^{p,\eta_1} \bar{u} = \bar{f}, \tag{31}$$

where $\bar{u} = \{\hat{u}(i)\}_{i\in\mathbb{Z}_N}$ is Fourier coefficient of desired solution, $\bar{f} = \{\hat{f}(i)\}_{i\in\mathbb{Z}_N}$ is Fourier coefficient for right-hand side and $\Lambda^{p,\eta_1} = \{\Lambda^{p,\eta_1}_{m,i}\}_{m,i\in\mathbb{Z}_l}$.

Proposition 4. *Calculation of matrix Λ^{p,η_1} requires $N\log N$ arithmetical operations (a.o.) by the order.*

Proof. Since $D^{\eta_1}_M \subset D^1_M$ for $\eta_1 \geq 1$, then the biggest amount of arithmetical operations is needed for calculation of matrix $\Lambda^{p,1}$ and we consider this case below. Since $(m-k, k-i) \in D^1_M$, then by the definition of the set D^1_M we have that $m - i \in \mathbb{Z}_M$. Let $l = m - i$ and calculate the amount of a.o. for element $\Lambda^{p,1}_{m,i}$ near diagonal l. For that rewrite the element $\Lambda^{p,1}_{m,i}$ in the following way

$$\Lambda^{p,1}_{m,m-l} = y_m = \sum_{\mathbb{Z}_{M+l}} \widehat{Q_{M,M}a_p}(m-k, l-(m-k))\hat{k}_p(k) = \sum_{k\in\mathbb{Z}_{M+l}} \alpha(m-k)k_p(k).$$

Using FFT, we can construct the element $\Lambda^{p,1}_{m,m-l}$ for all $m \in \mathbb{Z}_{M+l}$ with $(M+l)\log(M+l)$ a.o. by the order. Because of $l \in \mathbb{Z}_M$, the total amount of a.o. for constructing elements of matrix $\Lambda^{p,1}$ is $M(M+l)\log(M+l)$. Taking into account the fact that $l\log l \sim N$ for $\tau \in [\frac{\mu-\lambda}{\mu-\lambda+\beta}, 1)$ we arrive to the required result.

Let's calculate the amount of arithmetical operations that is necessary to construct all the elements from equation (31).

- For the element $\widehat{Q_{M,M}a_p}(i,j)$ we apply the relation

$$\widehat{Q_{M,M}a_p}(i,j) = \frac{1}{M^2}\sum_{l_1=1}^{M}\sum_{l_2=1}^{M} a_p(l_1M^{-1}, l_2M^{-1})e_i(l_1M^{-1})e_j(l_2M^{-1})$$

  that can be calculated for all $i, j \in \mathbb{Z}_M$ with the help of FFT by $M^2\log M$ arithmetical operations.
- the elements of the vector $\bar{f}$ can be calculated by the relation

$$\hat{f}(i) = \frac{1}{N}\sum_{=1}^{N} f(lN^{-1})e_i(lN^{-1})$$

  with the help of FFT by $N\log N$ a.o.
- the elements of Λ^{p,η_1} for $l = N^\tau$ can be calculated by $(N\log N)$ a.o. (see proposition 4).

Summarizing all items above, we can conclude that the total amount of a.o. for constructing all elements from (25) is $N\log N$ by the order.

7. Perturbed input data

Following [7], suppose that instead of functions $a_p(t,s), p = 0,\ldots,q$ and $f(t)$ we are given only some their pertubations $a_{p,\varepsilon}(t,s), p = 0,\ldots,q$, and $f_\delta(t)$ is

such that in the points of uniform grids it fulfils

$$M^{-2}\left(\sum_{i,j=1}^{M}|a_{p,\varepsilon}(iM^{-1},jM^{-1})-a_p(iM^{-1},jM^{-1})|\right)^{\frac{1}{2}}\le\varepsilon,\quad p=0,\dots,q,$$

$$N^{-1}(\sum_{j=1}^{N}|f_\delta(jN^{-1})-f(jN^{-1})|^2)^{1/2}\le\delta\|f\|_{\mu-\alpha}.$$

It is easy to show (see, for example, [7, p.100]), that mentioned estimations are equivalent to

$$\|Q_{M,M}(a_p-a_{p,\varepsilon})\|_{0,0}\le\varepsilon,\quad p=0,\dots,q, \tag{32}$$

$$\|Q_N(f_\delta-f)\|_0\le\delta\|f\|_{\mu-\alpha} \tag{33}$$

respectively. Then taking into account perturbation of input data the FDPM for equation (10) becomes

$$\mathcal{A}_{M,\varepsilon}u_{N,\varepsilon,\delta}=Du_{N,\varepsilon}+P_l\sum_{p=0}^{q}A_{p,M,\varepsilon}P_lu_{N,\varepsilon,\delta}=Q_Nf_\delta, \tag{34}$$

where $A_{p,M,\varepsilon}v(s)=\int_0^1 k_p(t-s)P_{D_M^{\eta_1}}Q_{M,M}a_{p,\varepsilon}(t,s)v(s)ds$ and $u_{N,\varepsilon,\delta}\in\mathcal{T}_N$ is approximate solution.

We pose the problem to solve equations (4) and (10) with perturbed input data as (32) and (33) with minimal amount of discrete information (i.e. set of values for functions $f_\delta(t)$ and $a_{p,\varepsilon}(t,s)$ in the points of uniform grid). At the same time arithmetical expenses should be less in comparison with methods known earlier (see, for example, [7] and [12]).

To achieve the aim of our investigation at first we state some auxiliary estimations.

Lemma 6. *Let estimation (32) is satisfied then for any* $\lambda\ge\max\{1/2,\alpha+1/2\}$ *it holds true*

$$\|\mathcal{A}_M-\mathcal{A}_{M,\varepsilon}\|_{\lambda,\lambda-\alpha}\le c_7\left(\frac{M}{2}\right)^{2\lambda-\alpha}\varepsilon,$$

where $c_7=c_0 2^{\lambda-\alpha+1}\gamma_{\lambda-\alpha}z_2(\lambda,\lambda-\alpha)(q+1)$.

It is easy to find that

$$(\mathcal{A}_M-\mathcal{A}_{M,\varepsilon})u=P_l(\sum_{p=0}^{q}A_{p,M}-A_{p,M,\varepsilon})P_lu.$$

Using Proposition 2, 3, inequalities (18) and (32) we have

$$\|(\mathcal{A}_M-\mathcal{A}_{M,\varepsilon})v\|_{\lambda-\alpha}\le$$

$$\le\|\sum_{p=0}^{q}P_l\int_0^1 k(t-s)P_{D_{D_M^{\eta_1}}}Q_{M,M}(a_{p,\varepsilon}-a_p)(t,s)P_lv(s)ds\|_{\lambda-\alpha}\le$$

$$\le c_0 2^{\lambda-\alpha+1}\gamma_{\lambda-\alpha}z_2(\lambda,\lambda-\alpha)\sum_{p=0}^{q}\|Q_{M,M}(a_{p,\varepsilon}-a_p)\|_{\lambda,\lambda-\alpha}\|P_lv\|_\lambda\le$$

$$\leq (q+1)c_0 2^{\lambda-\alpha+1}\gamma_{\lambda-\alpha}z_2(\lambda,\lambda-\alpha)\left(\frac{M}{2}\right)^{2\lambda-\alpha}\varepsilon\|v\|_\lambda,$$

which is the required result.

Lemma 7. *Let estimation (32) is satisfied and $\mathcal{A}_{M,\varepsilon}$ has the form (34). Then for M such that*

$$d_\lambda c_7\left(\frac{M}{2}\right)^{2\lambda-\alpha}\varepsilon \leq \frac{1}{2} \tag{35}$$

operator $\mathcal{A}_{M,\varepsilon}$ is invertible between spaces H^λ and $H^{\lambda-\alpha}$ and the following holds true

$$\|u\|_\lambda \leq 2d_\lambda\|\mathcal{A}_{M,\varepsilon}u\|_{\lambda-\alpha}. \tag{36}$$

The lemma can be proved in a similar way as lemma 5 by using the statements of lemmas 5 and 6.

Lemma 8. *Let the conditions (6)-(9) and (32), (33) fulfil and $a \in G_{\eta_1,\eta_2}$, $\eta_1 \geq 1, \eta_2 > 0$. Then for all $\lambda \in (\max\{1/2, \alpha+1/2\}, \mu)$ it holds true*

$$\|u_N - u_{N,\delta,\varepsilon}\|_\lambda \leq c_8\left(\frac{N}{2}\right)^{\lambda-\alpha}\delta + c_9\left(\frac{M}{2}\right)^{2\lambda-\alpha}\varepsilon,$$

where $c_8 = 2d_\lambda c''_\mu$ and $c_9 = c_{10}c_7 2d_\lambda$ with $c_{10} \leq 2 + d_\lambda(c''_\lambda + \frac{c'_\lambda}{2} + c''_\mu\gamma_{\mu-\alpha})$.

Proof. Using Lemmas 7 and 6, inequality (18) and (33) we find

$$\begin{aligned}
&\|u_N - u_{N,\delta,\varepsilon}\|_\lambda \leq 2d_\lambda\|\mathcal{A}_{M,\varepsilon}(u_N - u_{N,\delta,\varepsilon})\|_{\lambda-\alpha} \leq \\
&\leq 2d_\lambda\|\mathcal{A}_M u_N - \mathcal{A}_{M,\varepsilon}u_N\|_{\lambda-\alpha} + +2d_\lambda\|Q_N f - Q_N f_\delta\|_\lambda \leq \\
&\leq 2d_\lambda\left(\left(\frac{N}{2}\right)^{\lambda-\alpha}\delta\|f\|_{\mu-\alpha} + c_7\left(\frac{M}{2}\right)^{2\lambda-\alpha}\varepsilon\|u_N\|_\lambda\right).
\end{aligned} \tag{37}$$

Using (28) and (27) we bound the norm of element u_N as follows:

$$\begin{aligned}
&\|u_N\|_\lambda \leq \|u\|_\lambda + \|u - u_N\|_\lambda \leq \\
&\leq \|u\|_\lambda + c_5\left(\frac{N}{2}\right)^{\lambda-\mu} + c_6 e^{-2\eta_2\left(\frac{M}{2}\right)^{1/\eta_1}}\left(\frac{M}{2}\right)^{2\lambda-\alpha} \leq c_{10}.
\end{aligned}$$

Substituting the estimation above in (37) and taking into account (11) we derive desired estimation.

8. Selection of the discretization levels

Generalizing the results of the previos section we rewrite general estimation of error for FDPM. By virtue of Theorem 1 and Lemma 7, the accuracy of method (34) is estimated as

$$\begin{aligned}
\|u - u_{N,\delta,\varepsilon}\|_\lambda &\leq \|u - u_N\|_\lambda + \|u_N - u_{N,\delta,\varepsilon}\|_\lambda \leq \\
&\leq c_5\left(\frac{N}{2}\right)^{\lambda-\mu} + c_6 e^{-\eta_2\left(\frac{M}{2}\right)^{1/\eta_1}}\left(\frac{M}{2}\right)^{2\lambda-\alpha} + \\
&+ c_8\left(\frac{N}{2}\right)^{\lambda-\alpha}\delta + c_9\left(\frac{M}{2}\right)^{2\lambda-\alpha}\varepsilon.
\end{aligned} \tag{38}$$

Further following the paper [12] we consider the problem to select such levels of discretization N and M that minimize the error bound (38). Here we consider only the case then smoothness of parameters μ is known precisely (a priori case).

1. **A priori selection of parameter.** The problem of a priori selection of discretization levels was described in detail in [12] for class of equations under consideration. Here we slightly modify FDPM. However, as we can see below, it doesn't influence on the best accuracy of the method.

Further we denote by $[q]$ the integer part of number q and formulate the theorem that establishes a priori rule for choosing discretization parameter.

Theorem 2. *Let the conditions (6)-(9) fulfil and input data are perturbed as (33) and (32). Then for any* $\lambda \in (\max\{1/2, \alpha + 1/2\}, \mu)$, $\mu > \alpha + 1/2$ *with choosing the discretization parameters according to rule*

$$\bar{M} = \left[2 \left(\frac{1}{2\eta_2} \log \frac{c_{13}}{\varepsilon} \right)^{\eta_1} \right], \tag{39}$$

$$\bar{N} = \left[2 \left(\frac{c_8 \delta}{c_5} \right)^{\frac{1}{\alpha - \mu}} \right] \tag{40}$$

the error bound of the method (34) has the form

$$\|u - u_{N,\delta,\varepsilon}\|_\lambda \le c_{11} \delta^{\frac{\mu - \lambda}{\mu - \alpha}} + c_{12} \varepsilon \log^{\eta_1 (2\lambda - \alpha)} \frac{c_{13}}{\varepsilon}, \tag{41}$$

where

$$c_{11} = (c_8)^{\frac{\lambda - \mu}{\alpha - \mu}} c_5^{\frac{\lambda - \alpha}{\mu - \alpha}}, \quad c_{12} = \frac{c_6}{c_{13}} \left(\frac{1}{2\eta_2} \right)^{\eta_1 (2\lambda - \alpha)}$$

and

$$c_{13} = \frac{c_1}{c_{10}} \max_p \{\|a\|_{\eta_1, \eta_2}\}.$$

Proof. Direct substitution (39) and (40) in (38) gives the statement of theorem.

Remark 4. *It is evident that condition (35) fulfils with choosing M according (39) for sufficiently small ε. Let's check that condition (27) also holds true. From (39) it follows that*

$$c_{13} e^{-2\eta_2 \left(\frac{M}{2}\right)^{1/\eta_1}} = \varepsilon.$$

Then taking into account the relation above and (40) we can conclude that condition (27) takes place sor sufficiently small ε.

2. **Fast solving of FDPM (34).** Following [6] for fast solving (34), we propose to use GMRES. Such approach for solving problem under consideration has been detailed in [6] and here we only rewrite main points. Denote by

$$S_N := D + P_l \sum_{p=0}^{q} A_{p,M,\varepsilon} P_l.$$

It is evident that S_N is invertable operator (see lemma 7) that acts in $\mathcal{T}_N$. Thus according to theory we can apply GMRES with operator S_N and right-hand

side f_N with respect to the space H^α. The procedure concludes in constructing sequence u_{n_N} that satisfies the condition for $n = 1, 2, \ldots$

$$\|S_N u_{n_N} - f_N\|_\alpha = \min_{u \in \mathcal{K}_n(S_N, f_N)} \|S_N u - f_N\|_\alpha,$$

where $\mathcal{K}_n(S_N, f_N)$ is well-known Krylov space. As the stopping rule we consider the discrepancy principle

$$\|S_N u_{n_N} - f_N\|_\alpha \le c\delta \|f_N\|_\alpha, \tag{42}$$

where u_{n_N} is n-iteration of GMRES that we consider as approximation for u_N.

Now we are ready to establish the accuracy of GMRES approximation for our class of problems.

Theorem 3. *Suppose that $N, M \to 0$. Let n be the first number for which the condition (42) fulfils. Then the accuracy of GMRES applied to equation (34) is the following*

$$\|u_{N,\delta,\varepsilon} - u_{N_n}\|_\lambda \le 2 d_\lambda \left(\frac{N}{2}\right)^{\lambda-\alpha} \delta \|f_N\|_\alpha. \tag{43}$$

Moreover we have that $n = O(\log(N))$.

Proof. Using Lemma 6 we have that

$$\|u_{N,\delta,\varepsilon} - u_{N_n}\|_\lambda \le d_\lambda \|\mathcal{A}_{M,\varepsilon}(u_{N,\delta,\varepsilon} - u_{N_n})\|_{\lambda-\alpha} \le d_\lambda \|f_N - \mathcal{A}_{M,\varepsilon} u_{N_n}\|_{\lambda-\alpha}.$$

Further applying the inequalities (18) and (42) one can obtain

$$\|u_{N,\delta,\varepsilon} - u_{N_n}\|_\lambda \le 2 d_\lambda \left(\frac{N}{2}\right)^{\lambda-\alpha} \delta \|f_N\|_\alpha,$$

what was to be proved.

Remark 5. *As we can see from Theorem 3 the accuracy of FDPM method in combination with GMRES is the following*

$$\|u - u_{N_n}\|_\lambda \le O(\delta^{\frac{\mu-\lambda}{\mu-\alpha}} + \varepsilon \log^{\eta_1(2\lambda-\alpha)} \frac{1}{\varepsilon}).$$

Such accuracy of FDPM in the case of $\varepsilon = 0$ is optimal by the order (see [11]).

Remark 6. *For the realization of GMRES we need at every iteration to compute a matrix-vector product $S_N f_N$. Due to the structure of S_N as (34) and relation (30), the calculation can be performed by $l \cdot M^2$ operations. Since $M = O(\log N)$ (see corollary 4), then due to $N = l \log l$ for $l = N^\tau, \tau \in [\frac{\mu-\lambda}{\mu-\lambda+\beta}, 1)$ we have that constructing of matrix-vector product $S_N f_N$ requires $N \log N$ a.o. Moreover, as it is known, for realization of GMRES $O(nl)$ floating-point operations must be computed at the n-th iteration, i.e on the n-th step we need $O(N \log N)$ a.o. Thus total amount of a.o. for solving (10) is limited by $O(N \log N)$ by the order.*

Remark 7. *Let us suppose that $\varepsilon \ge c\delta$ and calculate the amount of necessary discrete information for equation (4) to implement the proposed method (34) with the accuracy (41). It is evident that in that case M does not exceed the magnitude $O(\log(N))$. So, for the discretization of $A_{p,\varepsilon}$ less than $O(\log^2 N)$*

values of kernels $a_{p,\varepsilon}(t,s)$ in the points of the uniform grid should be used. Note, that in the monograph [7] for the realization of the fully discrete projection method (34) at $M = N$ the order of discrete information was estimated as $O(N \log N)$.

Bibliography

1. Cheng G. Boundary Element Methods / G. Cheng, J. Shou. – San Diego: Academic Press, 1992.
2. Harbrecht H. Self-regularization by projection for noisy pseudodifferential equations of negative order / H. Harbrecht, S. V. Pereverzev, R. Schneider // Numer. Math. – 2003. – Vol. 95. – P. 123-143.
3. Gorbachuk V. I. Boundary value problems for operator differential equations / V. I. Gorbachuk, M. L. Gorbachuk. – Dordrecht, Boston, London: Kluwer Academic Publishers, 1991.
4. Pereverzev S. V. On the characterization of self-regularization properties of a fully discrete projection method for Symm's integral equation / S. V. Pereverzev, S. Prossdorf // J. Integral Equations Appl. – 2000. – Vol. 12. – P. 113-130.
5. Pereverzev S. V. On the adaptive selection of the parameter in regularization of ill-posed problems / S. V. Pereverzev, E. Schock // SIAM J. Numer. Anal. – 2005. – Vol. 43. – P. 2060-2076.
6. Plato R. On the fast fully discretized solution of intergral and pseudo-differential equations on smooth curves / R. Plato, G. Vainikko // Calcolo. – 2001. – Vol. 38. – P. 13-36.
7. Saranen J. Periodic Integral and Pseudodifferential Equations with Numerical Approximation / J. Saranen, G. Vainikko. – Berlin: Springer, 2002.
8. Saranen J. Fast solvers of integral and pseudodifferential equations on closed curves / J. Saranen, G. Vainikko // Math. Comp. – 1998. – Vol. 67. – P. 1473-1491.
9. Solodky S. G. Error bounds of a fully discrete projection method for Symm's integral equation / S. G. Solodky, E. V. Lebedeva // Comp. Method Appl. Math. – 2007. – Vol. 7. – P. 255-263.
10. Semenova E. V. A The accuracy of Fully Discrete Projection Method on one class of PD equation / E. V. Semenova, E. A. Volonets // Dynamical system. – 2012. – Vol. 2. – P. 309-321.
11. Solodky S. G. On optimal order accuracy of solving Symm's integral equation / S. G. Solodky, E. V. Semenova // Zh. Vychisl. Mat. i Mat. Fiz. – 2012. – Vol. 52. – P. 1–11.
12. Solodky S. G. A class of periodic integral equation with numerical solving by a fully discrete projection method / S. G. Solodky, E. V. Semenova // UMV. – 2014. – Vol. 11. – P. 400-416.

E. V. Semenova,
Institute of Mathematics, National Academy of Sciences,
3, Tereschenkivs'ka Str., Kyiv, 01601, Ukraine;

Received 08.07.2015

Журнал обчислювальної та прикладної математики | 2015 № 2 (119) | Journal of Computational & Applied Mathematics

UDC 519.6

COMBINED NEWTON-KURCHATOV METHOD UNDER THE GENERALIZED LIPSCHITZ CONDITIONS FOR THE DERIVATIVES AND DIVIDED DIFFERENCES

S. M. SHAKHNO

Резюме. Доведено локальну збіжність комбінованого ітераційного процесу, побудованого на основі методу Ньютона і методу лінійної інтерполяції Курчатова, для розв'язування нелінійних операторних рівнянь в банаховому просторі за узагальнених умов Ліпшиця для похідних першого порядку і поділених різниць першого та другого порядку. Визначено радіус кулі збіжності і швидкість збіжності методу, знайдено область єдиності розв'язку нелінійного рівняння.

Abstract. The local convergence of combined iterative process, built on the basis of Newton's method and Kurchatov's method of linear interpolation, for solving of nonlinear operator equations in Banach space under the generalized Lipschitz conditions for the derivative of the first order and divided differences of the first and second order is proved. The radius of the convergence ball and convergence order of the method are determined, the ball of uniqueness of the solution of nonlinear equation is found.

1. Introduction

In this study we are concerned with the problem of approximating a locally unique solution $x^* \in D$ of equation

$$F(x) + G(x) = 0, \tag{1}$$

where F is a Fréchet-differentiable nonlinear operator on an open convex subset D of a Banach space X with values in a Banach space Y, and $G : D \to Y$ is a continuous nonlinear operator.

Let x, y be two points of D. A linear operator from X into Y, denoted $\delta G(x, y)$, which satisfies the condition

$$\delta G(x, y)(x - y) = G(x) - G(y) \tag{2}$$

is called a divided difference of G at points x and y.

Let x, y, z be three points of D. A operator $\delta G(x, y, z)$ will be called a divided difference of the second order of the operator G at the points x, y and z , if it satisfies the condition

$$\delta G(x, y, z)(y - z) = \delta G(x, y) - \delta G(x, z). \tag{3}$$

Key words. Banach space, Newton's method; Kurchatov's method; Combined iterative method; Divided difference; Local convergence; Convergence order; Generalized Lipschitz condition.

A well-known simple difference method for solving nonlinear equations $F(x) = 0$ is the Secant method

$$x_{n+1} = x_n - (\delta F(x_{n-1}, x_n))^{-1} F(x_n),\ n = 0,1,2,\ldots, \tag{4}$$

where $\delta F(x_{n-1}, x_n)$ is a divided difference of the first order and x_0, x_{-1} are given.

Secant method for solving nonlinear operator equations in a Banach space was explored by the authors [5,14,15,19,30] under the condition that the divided differences of a nonlinear operator F satisfy the Lipschitz (Hölder) condition with constant L of type

$$\|\delta F(x, y) - \delta F(u, v)\| \leq L(\|x - u\| + \|y - v\|).$$

In [11] it was proposed one-point iterative Secant-type method with memory. In [29] it was explored the Kurchatov method under the classical Lipschitz conditions for the divided differences of the first and second order and it was determined the quadratic convergence of it. The iterative formula of Kurchatov method has the form [4, 5, 18, 29]

$$x_{n+1} = x_n - (\delta F(2x_n - x_{n-1}, x_{n-1}))^{-1} F(x_n),\ n = 0,1,2,\ldots, \tag{5}$$

where $\delta F(u, v)$ is a divided difference of the first order, x_0, x_{-1} are given.

In paper [20] Potra investigated the three-point difference method with convergence order $1.839\ldots$ for classical Lipschitz conditions for divided differences of the first and second order [20]

$$\begin{aligned} &x_{n+1} = x_n - A_n^{-1} F(x_n), \\ &A_n = \delta F(x_n, x_{n-1}) + \delta F(x_{n-2}, x_n) - \delta F(x_{n-2}, x_{n-1}),\ n = 0,1,2,\ldots, \end{aligned} \tag{6}$$

x_0, x_{-1}, x_{-2} are given. This method first has been proposed for scalar nonlinear equations by Traub in [30].

Regarding the local convergence of Newton method, Traub and Wozniakowski in [31] and Wang in [33] gave the best estimate of the radii of convergence balls when the first derivatives are Lipschitz continuous around a solution.

Besides, there are a lot of the works on the weakness and/or extension of the hypothesis made on the nonlinear operators; see works of Argyros, Ezquerro, Hernandez, Rubio, Gutierrez, Wang, Li [5, 12, 13, 32–35] and references therein. In particular, Wang introduced in [33] the notions of generalized Lipschitz conditions or Lipschitz conditions with L average, where instead of constant L it is used some positive integrable function.

The center Lipschitz condition with L average in the inscribe sphere makes us unify the convergence criteria containing the Kantorovich theorem and the Smale α-theory, while the radius Lipschitz conditions with L average unify the estimates of the radii of convergence balls for operators with Lipschitz continuous first derivatives and analytic operators [10, 32, 33].

In our work [27] for the first time we have introduced a similar generalized Lipschitz condition for the operator of the first order divided difference and under this condition the convergence of Secant method was studied and was found that its convergence order is $(1 + \sqrt{5})/2$. In the paper [26] we have

introduced a generalized Lipschitz condition also for the divided differences of the second order and we have studied the local convergence of Kurchatov method (5).

Note that in many papers such as [1,2,7,16,21], the authors investigated the Secant and Secant-type methods under the generalized conditions for the first divided differences of the form

$$\|(\delta F(x,y) - \delta F(u,v)))\| \leq \omega(\|x-y\|, \|u-v\|) \quad \forall x,y,u,v \in D, \tag{7}$$

where $\omega : \mathbf{R}_+ \times \mathbf{R}_+ \longrightarrow \mathbf{R}_+$ is continuous nondecreasing function in their two arguments. Under these same conditions in the work of Argyros [4] it is proven semilocal convergence of Kurchatov method and in [22] of Ren and Argyros the semilocal convergence of combined Kurchatov method and Secant method. In both cases only the linear convergence of the methods is received.

In this paper, we study the local convergence of the combined Newton–Kurchatov method

$$x_{n+1} = x_n - (F'(x_n) + \delta G(2x_n - x_{n-1}, x_{n-1}))^{-1}(F(x_n) + G(x_n)), \quad n = 0,1,2,\ldots, \tag{8}$$

where $F'(u)$ is a Fréchet derivative, $\delta G(u,v)$ is a divided difference of the first order, x_0, x_{-1} are given, which is built on the basis of the mentioned Newton and Kurchatov methods under relatively weak, generalized Lipschitz conditions for the derivatives and divided differences of nonlinear operators. Setting $G(x) \equiv 0$, we receive the results for Newton method [33], and when $F(x) \equiv 0$ we get the known results for Kurchatov method [26].

We first proposed the method (8) in the paper [28]. Semilocal convergence of the method (8) under the classical Lipschitz conditions is studied in the mentioned article, but there was determined the convergence only with the order $(1+\sqrt{5})/2$.

In this article we prove the quadratic order of convergence of the method (8), which is higher than the convergence order $(1+\sqrt{5})/2$ for the Newton–Secant method [5,8,9,23]

$$x_{n+1} = x_n - (F'(x_n) + \delta G(x_{n-1}, x_n))^{-1}(F(x_n) + G(x_n)), \quad n = 0,1,2,\ldots, \tag{9}$$

Method (9) was proposed in [9] and proved its convergence with the order $(1+\sqrt{5})/2$ under the classical Lipschitz conditions for the first derivative $F'(x)$ and bounded norm of the second-order divided difference $\delta G(x,y,z)$. The same order of convergence was received in [5] with weaker conditions - classical Lipschitz conditions for the first derivative $F'(x)$ and the first-order divided difference $\delta G(x,y)$.

Note that in the work [23] we have considered combined Newton–Secant method (9) and we have proposed a methodology of studying the convergence of combined methods for solving nonlinear equations with nondifferentiable operator under the relatively weak, generalized Lipschitz conditions for the first derivatives and divided differences of nonlinear operators. Under the same

conditions in [24] it was studied the convergence of the combined two-step method for the solution of nonlinear equations.

The results of the numerical study of the method (8) and other combined methods on the test problems are provided in our works [25, 28].

2. Local convergence of Newton–Kurchatov method

Lets denote $B(x_0, r) = \{x : \|x - x_0\| < r\}$ an open ball of radius $r > 0$ with center at point $x_0 \in D$, $B(x_0, r) \subset D$.

Condition on the divided difference operator $\delta F(x, y)$

$$\|\delta F(x, y) - \delta F(u, v)\| \leq L(\|x - u\| + \|y - v\|) \quad \forall x, y, u, v \in D \tag{10}$$

is called Lipschitz condition in domain D with constant $L > 0$. If the condition is being fulfilled

$$\|\delta F(x, y) - F'(x_0)\| \leq L(\|x - x_0\| + \|y - x_0\|) \quad \forall x, y \in B(x_0, r), \tag{11}$$

then we call it the center Lipschitz condition in the ball $B(x_0, r)$ with constant L.

However L in Lipschitz conditions can be not a constant, and can be a positive integrable function. In this case, if for $x^* \in D$ inverse operator $[F'(x^*)]^{-1}$ exists, then the conditions (10) and (11) for $x_0 = x^*$ can be replaced respectively for

$$\begin{aligned} &\|F'(x^*)^{-1}(\delta F(x, y) - \delta F(u, v)))\| \leq \\ &\leq \int_0^{\|x-y\|+\|u-v\|} L(z)dz \quad \forall x, y, u, v \in D \end{aligned} \tag{12}$$

and

$$\begin{aligned} &\|F'(x^*)^{-1}(\delta F(x, y) - F'(x^*))\| \leq \\ &\leq \int_0^{\|x-x^*\|+\|y-x^*\|} L(z)dz \quad \forall x, y \in B(x^*, r). \end{aligned} \tag{13}$$

Simultaneously Lipschitz condition (12) – (13) are called generalized Lipschitz conditions or Lipschitz conditions with the L average.

Similarly, we introduce the generalized Lipschitz condition for the divided difference of the second order

$$\begin{aligned} &\|F'(x^*)^{-1}(\delta F(u, x, y) - \delta F(v, x, y))\| \leq \\ &\leq \int_0^{\|u-v\|} N(z)dz \; \forall x, y, u, v \in B(x^*, r), \end{aligned} \tag{14}$$

where N is a positive integrable function.

Remark 8. *Note than the operator F is Fréchet differentiable on D when the Lipschitz conditions (10) or (12) are fulfilled $\forall x, y, u, v \in D$ (the divided differences $\delta F(x, y)$ are Lipschitz continuous on D) and $\delta F(x, x) = F'(x)$ $\forall x \in D$ [3].*

The radius of the convergence ball and the convergence order of the combined Newton–Kurchatov method (8) are determined in next theorem.

Theorem 1. *Let F and G be continuous nonlinear operators defined in open convex domain D of a Banach space X with values in the Banach space Y. Lets suppose, that: 1) $H(x) \equiv F(x) + G(x) = 0$ has a solution $x^* \in D$, for which there exists a Fréchet derivative $H'(x^*)$ and it is invertible; 2) F has the Fréchet derivative of the first order, and G has divided differences of the first and second order on $B(x^*, 3r) \subset D$, which are satisfying on $B(x^*, 3r)$ the generalized Lipschitz conditions*

$$\|H'(x^*)^{-1}(F'(x) - F'(x^\tau))\| \le \int_{\tau\rho(x)}^{\rho(x)} L_1(u)du,\ 0 \le \tau \le 1, \tag{15}$$

$$\|H'(x^*)^{-1}(\delta G(x,y) - \delta G(u,v))\| \le \int_0^{\|x-u\|+\|y-v\|} L_2(z)dz, \tag{16}$$

$$\|H'(x^*)^{-1}(\delta G(u,x,y) - \delta G(v,x,y))\| \le \int_0^{\|u-v\|} N(z)dz, \tag{17}$$

where $x^\tau = x^* + \tau(x - x^*)$, $\varrho(x) = \|x - x^*\|$, L_1, L_2 and N are positive nondecreasing integrable functions and $r > 0$ satisfies the equation

$$\frac{\frac{1}{r}\int_0^r L_1(u)udu + \int_0^r L_2(u)du + 2r\int_0^{2r} N(u)du}{1 - \Big(\int_0^r L_1(u)du + \int_0^{2r} L_2(u)du + 2r\int_0^{2r} N(u)du\Big)} = 1. \tag{18}$$

Then for all $x_0, x_{-1} \in B(x^*, r)$ the iterative method (8) is correctly defined and the generated by it sequence $\{x_n\}_{n\ge0}$, which belongs to $B(x^*, r)$, converges to x^* and satisfies the inequality

$$\begin{aligned}
&\|x_{n+1} - x^*\| \le \\
&\le \Big\{\frac{1}{\rho(x_n)}\int_0^{\rho(x_n)} L_1(u)udu + \int_0^{\rho(x_n)} L_2(u)du + \\
&+ \int_0^{\|x_n - x_{n-1}\|} N(u)du\|x_n - x_{n-1}\|\Big\} \times \\
&\times \Big\{1 - \Big(\int_0^{\rho(x_n)} L_1(u)du + \int_0^{2\rho(x_n)} L_2(u)du + \\
&+ \int_0^{\|x_n - x_{n-1}\|} N(u)du\|x_n - x_{n-1}\|\Big)\Big\}^{-1}\|x_n - x^*\|.
\end{aligned} \tag{19}$$

Proof. First we show that

$$f(t) = \frac{1}{t^2}\int_0^t L_1(u)udu, \quad g(t) = \frac{1}{t}\int_0^t L_2(u)du$$

in $h(t)=\frac{1}{t}\int_0^t N(u)du$ monotonically nondecreasing with respect to t. Indeed, under the monotony of L_1, L_2, N we have

$$\Big(\frac{1}{t_2^2}\int_0^{t_2}-\frac{1}{t_1^2}\int_0^{t_1}\Big)L_1(u)udu=\Big(\frac{1}{t_2^2}\int_{t_1}^{t_2}+\Big(\frac{1}{t_2^2}-\frac{1}{t_1^2}\Big)\int_0^{t_1}\Big)L_1(u)udu\geq$$

$$\geq L(t_1)\Big(\frac{1}{t_2^2}\int_{t_1}^{t_2}+\Big(\frac{1}{t_2^2}-\frac{1}{t_1^2}\Big)\int_0^{t_1}\Big)udu=L_1(t_1)\Big(\frac{1}{t_2^2}\int_0^{t_2}-\frac{1}{t_1^2}\int_0^{t_1}\Big)udu=0,$$

$$\Big(\frac{1}{t_2}\int_0^{t_2}-\frac{1}{t_1}\int_0^{t_1}\Big)L_2(u)du=\Big(\frac{1}{t_2}\int_{t_1}^{t_2}+\Big(\frac{1}{t_2}-\frac{1}{t_1}\Big)\int_0^{t_1}\Big)L_2(u)du\geq$$

$$\geq L_2(t_1)\Big(\frac{1}{t_2}\int_{t_1}^{t_2}+\Big(\frac{1}{t_2}-\frac{1}{t_1}\Big)\int_0^{t_1}\Big)du=L_2(t_1)\Big(\frac{t_2-t_1}{t_2}+t_1\Big(\frac{1}{t_2}-\frac{1}{t_1}\Big)\Big)=0$$

for $0<t_1<t_2$. So, $f(t)$, $g(t)$ are nondecreasing with respect to t. Similarly we get for $h(t)$.

We denote by A_n linear operator $A_n=F'(x_n)+\delta G(2x_n-x_{n-1},x_{n-1})$. Easy to see that if $x_n, x_{n-1}\in B(x^*,r)$, then $2x_n-x_{n-1}, x_{n-1}\in B(x^*,3r)$. Then A_n is invertible and the inequality holds

$$\begin{aligned}\|A_n^{-1}H^{'}(x^*)\|&=\|[I-(I-H^{'}(x^*)^{-1}A_n)]^{-1}\|\leq\\&\leq\Big(1-\Big(\int_0^{\rho(x_n)}L_1(u)du+\int_0^{2\rho(x_n)}L_2(u)du+\\&+\int_0^{\|x_n-x_{n-1}\|}N(u)du\|x_n-x_{n-1}\|\Big)\Big)^{-1}.\end{aligned}\tag{20}$$

Indeed from the formulas (15)–(17) we get

$$\begin{aligned}\|I-H^{'}(x^*)^{-1}A_n\|&=\|H^{'}(x^*)^{-1}(F'(x^*)-F'(x_n)+\delta G(x^*,x^*)-\\&-\delta G(x_n,x_n)+\delta G(x_n,x_n)-\delta G(2x_n-x_{n-1},x_{n-1})\|)\leq\\&\leq\int_0^{\rho(x_n)}L_1(u)du+\|H^{'}(x^*)^{-1}(\delta G(x^*,x^*)-\delta G(x_n,x_n)+\\&+\delta G(x_n,x_n)-\delta G(x_n,x_{n-1})+\delta G(x_n,x_{n-1})-\delta G(2x_n-x_{n-1},x_{n-1}))\|\leq\\&\leq\int_0^{\rho(x_n)}L_1(u)du+\int_0^{2\rho(x_n)}L_2(u)du+\\&+\|H^{'}(x^*)^{-1}(\delta G(x_n,x_{n-1},x_n)-\delta G(2x_n-x_{n-1},x_{n-1},x_n))(x_n-x_{n-1})\|\leq\\&\leq\int_0^{\rho(x_n)}L_1(u)du+\int_0^{2\rho(x_n)}L_2(u)du+\int_0^{\|x_n-x_{n-1}\|}N(u)du\|x_n-x_{n-1}\|.\end{aligned}$$

From the definition r (18) we get

$$\begin{aligned}&\int_0^r L_1(u)du+\int_0^{2r}L_2(u)du+2r\int_0^{2r}N(u)du=\\&=1-\frac{1}{r}\int_0^r L_1(u)du-\int_0^r L_2(u)du-2r\int_0^{2r}N(u)du<1.\end{aligned}\tag{21}$$

Using the Banach theorem on inverse operator [17], we get formula (20). Then we can write

$$\begin{aligned}
&\|x_{n+1}-x^*\| = \|x_n - x^* - A_n^{-1}(F(x_n)-F(x^*)+G(x_n)-G(x^*))\| = \\
&= \Big\|-A_n^{-1}\Big(\int_0^1 (F'(x_n^\tau)-F'(x_n))d\tau + \delta G(x_n,x^*) - \\
&- \delta G(2x_n - x_{n-1}, x_{n-1}))(x_n - x^*)\Big\| \le \\
&\le \|A_n^{-1}H'(x^*)\|\Big(\|H'(x^*)^{-1}\int_0^1\int_{\tau\rho(x_n)}^{\rho(x_n)} L_1(u)dud\tau + \\
&+ \|H'(x^*)^{-1}(+\delta G(x_n,x^*) - \delta G(2x_n - x_{n-1}, x_{n-1}))\|\Big)\|x_n - x^*\|.
\end{aligned} \tag{22}$$

According to the condition (15)–(17) of the theorem we get

$$\|H'(x^*)^{-1}\Big(\int_0^1\int_{\tau\rho(x_n)}^{\rho(x_n)} L_1(u)dud\tau + \delta G(x_n,x^*) - A_n\Big)\| =$$

$$= \frac{1}{\rho(x_n)}\int_0^{\rho(x_n)} L_1(u)udu + \|H'(x^*)^{-1}(\delta G(x_n,x^*) - \delta G(x_n,x_n) +$$

$$+\delta G(x_n,x_n) - \delta G(x_n,x_{n-1}) + \delta G(x_n,x_{n-1}) - \delta G(2x_n - x_{n-1},x_{n-1}))\| \le$$

$$\le \frac{1}{\rho(x_n)}\int_0^{\rho(x_n)} L_1(u)udu + \|H'(x^*)^{-1}(\delta G(x_n,x^*) - \delta G(x_n,x_n))\| +$$

$$+\|H'(x^*)^{-1}(\delta G(x_n,x_{n-1},x_n) - \delta G(2x_n - x_{n-1},x_{n-1},x_n))(x_n - x_{n-1})\| \le$$

$$\le \frac{1}{\rho(x_n)}\int_0^{\rho(x_n)} L_1(u)udu + \int_0^{\rho(x_n)} L_2(u)du +$$

$$+ \int_0^{\|x_n - x_{n-1}\|} N(u)du\|x_n - x_{n-1}\|.$$

From (20) and (22) shows that fulfills (19). Then from (19) and (18) we get

$$\|x_{n+1}-x^*\| < \|x_n - x^*\| < ... < \max\{\|x_0 - x^*\|, \|x_{-1} - x^*\|\} < r.$$

Therefore, the iterative process (5) is correctly defined and the sequence that it generates belongs to $B(x^*, r)$. From the last inequality and estimates (19) we get $\lim\limits_{n\to\infty}\|x_n - x^*\| = 0$. Since the sequence $\{x_n\}_{n\ge 0}$ converges to x^*, then

$$\|x_n - x_{n-1}\| \le \|x_n - x^*\| + \|x_{n-1} - x^*\| \le 2\|x_{n-1} - x^*\|$$

and $\lim\limits_{n\to\infty}\|x_n - x_{n-1}\| = 0$.

The theorem is proven.

Corollary 5. *The order of convergence of the iterative procedure (8) is quadratic.*

Proof. Lets denote $\rho_{\max} = \max\{\rho(x_0), \rho(x_{-1})\}$. Since $g(t)$ and $h(t)$ are monotonically nondecreasing, then with taking into account the expressions

$$\frac{1}{\rho(x_n)} \int_0^{\rho(x_n)} L_1(u)udu = \frac{\int_0^{\rho(x_n)} L_1(u)udu\rho(x_n))}{(\rho(x_n))^2} \leq$$

$$\leq \frac{\int_0^{\rho_{\max}} L_1(u)udu\rho(x_n)}{(\rho_{\max})^2} =: A_1\rho(x_n),$$

$$\int_0^{\rho(x_n)} L_2(u)du = \frac{\int_0^{\rho(x_n)} L_2(u)du\rho(x_n)}{\rho(x_n)} \leq \frac{\int_0^{\rho_{\max}} L_2(u)du\rho(x_n)}{\rho_{\max}} =: A_2\rho(x_n),$$

$$\int_0^{\|x_n - x_{n-1}\|} N(u)du = \frac{\int_0^{\|x_n - x_{n-1}\|} N(u)du\|x_n - x_{n-1}\|}{\|x_n - x_{n-1}\|} <$$

$$< \frac{\int_0^{\|x_0 - x_{-1}\|} N(u)du\|x_n - x_{n-1}\|}{\|x_0 - x_{-1}\|} =: A_3\|x_n - x_{n-1}\|$$

and

$$\Big(1 - \Big(\int_0^{\rho(x_n)} L_1(u)du + 2\int_0^{\rho(x_n)} L_2(u)du +$$

$$+ \int_0^{\|x_n - x_{n-1}\|} N(u)du\|x_n - x_{n-1}\|\Big)\Big)^{-1} <$$

$$< \Big(1 - \Big(\int_0^{\rho_{\max}} L_1(u)du + 2\int_0^{\rho_{\max}} L_2(u)du +$$

$$+ \int_0^{\|x_0 - x_{-1}\|} N(u)du\|x_0 - x_{-1}\|\Big)\Big)^{-1} =: A_4,$$

from the inequality (19) follows

$$\|x_{n+1} - x^*\| \leq A_4(A_1\rho(x_n) + A_2\rho(x_n) + A_3\|x_n - x_{n-1}\|^2)\|x_n - x^*\|.$$

or

$$\|x_{n+1} - x^*\| \leq C_3\|x_n - x^*\|^2 + C_4\|x_n - x_{n-1}\|^2\|x_n - x^*\|. \tag{23}$$

Here $A_k, k = 1, ..., 4, C_3, C_4$ are some positive constants.

Suppose now that the order of convergence of the iterative process (8) is lower 2, therefore there exist $C_5 \geq 0$ and $N > 0$, that for all $n \geq N$ the inequality holds

$$\|x_n - x^*\| \geq C_5\|x_{n-1} - x^*\|^2.$$

Since

$$\|x_n - x_{n-1}\|^2 \leq (\|x_n - x^*\| + \|x_{n-1} - x^*\|)^2 \leq 4\|x_{n-1} - x^*\|^2,$$

then from (23) we get

$$\begin{aligned}\|x_{n+1}-x^*\| &\le C_3\|x_n-x^*\|^2+4C_4\|x_{n-1}-x^*\|^2\|x_n-x^*\| \\ &\le (C_3+4C_4/C_5)\|x_n-x^*\|^2=C_6\|x_n-x^*\|^2.\end{aligned} \tag{24}$$

But inequality (24) means that the order of convergence not lower 2. Thus, the convergence rate of sequence $\{x_n\}_{n\ge 0}$ to x^* is quadratic.

Next theorem determines the ball of uniqueness of the solution x^* of (1) in $B(x^*,r)$.

Theorem 2. *Lets assume that: 1) $H(x)\equiv F(x)+G(x)=0$ has a solution $x^*\in D$, in which there exists a Fréchet derivative $H'(x^*)$ and it is invertible; 2) F has a continuous Freć het derivative in $B(x^*,r)$, F' satisfies the generalized Lipschitz condition*

$$\|H'(x^*)^{-1}(F'(x)-F'(x^*))\|\le\int_0^{\rho(x)}L_1(u)du\quad\forall x\in B(x^*,r),$$

the divided difference $\delta G(x,y)$ satisfies the generalized Lipschitz condition

$$\|H'(x^*)^{-1}(\delta G(x,x^*)-G'(x^*))\|\le\int_0^{\rho(x)}L_2(u)du\quad\forall x\in B(x^*,r),$$

where L_1 and L_2 are positive integrable functions. Let $r>0$ satisfy

$$\frac{1}{r}\int_0^r(r-u)L_1(u)du+\int_0^r L_2(u)du\le 1.$$

Then the equation $H(x)=0$ has a unique solution x^* in $B(x^*,r)$.

Proof analogous to [23,24].

3. Corollaries

In the study of iterative methods the traditional assumption is that the derivatives and/or the divided differences satisfy the classical Lipschitz conditions. Assuming that L_1, L_2 and N are constants, we get from theorem 2.1 and 3.1 important corollaries, which are of interest on its own.

Corollary 6. *Let's assume that: 1) $H(x)\equiv F(x)+G(x)=0$ has a solution $x^*\in D$, in which there exists Fréchet derivative $H'(x^*)$ and it is invertible; 2) F has a continuous Fréchet derivative and G has divided differences of the first and second order $\delta G(x,y)$ and $\delta G(x,y,z)$ in $B(x^*,3r)\subset D$, which satisfy the Lipschitz condition*

$$\|H'(x^*)^{-1}(F'(x)-F'(x^*+\tau(x-x^*)))\|\le(1-\tau)L_1\|x-x^*\|,$$

$$\|H'(x^*)^{-1}(\delta G(x,y)-\delta G(u,v))\|\le L_2(\|x-u\|+\|y-v\|),$$

$$\|H'(x^*)^{-1}(\delta G(u,x,y)-\delta G(v,x,y))\|\le N\|u-v\|,$$

where $x,y,u,v\in B(x^*,r)$, L_1, L_2, N are positive numbers and r is the positive root of the equation

$$\frac{L_1r/2+L_2r+4Nr^2}{1-L_1r-2L_2r-4Nr^2}=1.$$

Then Newton-Kurchatov method (5) converges for all $x_{-1}, x_0 \in B(x^*, r)$ *and there fulfills*

$$\|x_{n+1} - x^*\| \leq \frac{(L_1/2 + L_2)\|x_n - x^*\| + N\|x_n - x_{n-1}\|^2}{1 - \left(L_1 + 2L_2\|x_n - x^*\| + N\|x_n - x_{n-1}\|^2\right)}.$$

Moreover, r *is the best of all possible.*

Note that the received r coincides with the value of $r = \dfrac{2}{3L_1}$ for Newton method for solving equation $F(x) = 0$ [20, 31, 33] and with $r = 2/(3L_2 + \sqrt{9L_2^2 + 32N})$ for Kurchatov method for solving the equation $G(x) = 0$, as derived in [29].

Corollary 7. *Suppose that: 1)* $H(x) \equiv F(x) + G(x) = 0$ *has a solution* $x^* \in D$, *in which there exists the Fréchet derivative* $H'(x^*)$ *and it is invertible; 2)* F *has continuous derivative and* G *has divided difference* $\delta G(x, x^*)$ *in* $B(x^*, r) \subset D$, *which satisfy the Lipschitz conditions*

$$\|H'(x^*)^{-1}(F'(x) - F'(x^*))\| \leq L_1\|x - x^*\|,$$
$$\|H'(x^*)^{-1}(\delta G(x, x^*) - G'(x^*))\| \leq L_2\|x - x^*\|$$

for all $x \in B(x^*, r)$, *where* L_1 *and* L_2 *are positive numbers and* $r = \dfrac{2}{L_1 + 2L_2}$. *Then the equation* $H(x) = 0$ *has a unique solution* x^* *in the open ball* $B(x^*, r)$. *Moreover, the given* r *is the best of all possible and does not depend on* F *and* G.

Note that the resulting radius of the uniqueness ball of the solution coincides with $r = \dfrac{2}{L_1}$ for Newton method for solving the equation $F(x) = 0$ [33] and with $r = \dfrac{1}{L_2}$ for Kurchatov method for solving the equation $G(x) = 0$ [29].

4. Conclusions

In the papers [5, 15, 29] it was studied the local convergence of Secant and Kurchatov methods in the case of fulfilment of Lipschitz conditions for the divided differences, which hold some Lipschitz constants. In the work [33] it has been justified the convergence of Newton method for the generalized Lipschitz conditions for the Fréchet derivative of the first order. We explored the local convergence of Newton-Kurchatov method under the generalized Lipschitz conditions for Fréchet derivative of differentiable part of the operator and the divided differences of the nondifferentiable part, in which instead of Lipschitz constants some positive integrable functions are used. Our results contain the known ones as partial cases.

Bibliography

1. Amat S. On the local convergence of secant-type methods / S. Amat // Intern. J. Comput. Math. – 2004. – Vol. 81. – P. 1153-1161.
2. Amat S. On a higher order Secant method / S. Amat, S. Busquier // Appl. Math. Comput. – 2003. – Vol. 141. – P. 321-329.

3. Argyros I. K. On the secant method / I. K. Argyros // Publ. Math. Debrecen. – 1993. – Vol. 43. – P. 233-238.
4. Argyros I. K. A Kantorovich-type analysis for a fast iterative method for solving nonlinear equations / I. K. Argyros // J. Math. Anal. Appl. – 2007. – Vol. 332. – P. 97-108.
5. Argyros I. K. Convergence and applications of Newton-type iterations / I. K. Argyros. – New-York: Springer–Verlag Publ., 2008.
6. Argyros I. K. Improving the rate of convergence of Newton methods on Banach spaces with a convergence structure and applications / I. K. Argyros // Appl. Math. Lett. – 1997. – Vol. 10 – P. 21-28.
7. Argyros I. K. Chebyshev-Secant type methods for non-differentiable operator / I. K. Argyros, J. A. Ezquerro, J. M. Gutiérrez, M. A. Hernández, S. Hilout // Milan J. Math. – 2013. – Vol. 81. – P. 25-35.
8. Argyros I. K. The theorie and application of iteration methods / I. K. Argyros, F. Szidarovszky. – Bosa Raton, Florida: C.R.C. Press, 1993.
9. Catinas E. On some iterative methods for solving nonlinear equations / E. Catinas // Revue d'Analyse Numérique et de Théorie de l'Approximation. – 199. – Vol. 23. – P. 47-53.
10. Chong C. Convergence behavior of Gauss-Newton's method and extensions of the Smale point estimate theory / C. Chong, N. Hu, J. Wang // Journal of Complexity. – 2010. – Vol. 26. – P. 268-295.
11. Ezquerro J. A. Solving non-differentiable equations by a new one-point iterative method with memory / J. A. Ezquerro, M. Grau-Sánchez, M. Hernández // Journal of Complexity. – 2012. – Vol. 28. – P. 48-58.
12. Ezquerro J. A. Generalised differentiability conditions for Newton's method / J. A. Ezquerro, M. Hernández // IMA J. Numer. Anal. – 2002. – Vol. 22. – P. 187-205.
13. Gutiérrez J. M. Newton's method under weak Kantorovich conditions / J. M. Gutiérrez, M. A. Hernández // IMA J. Numer Anal. – 2000. – Vol. 20. – P. 521-532.
14. Hernandez M. A. The Secant method and divided differences Hölder continuous / M. A. Hernandez, M. J. Rubio // Applied Mathematics and Computation. – 2001. – Vol. 124. – P. 139-149.
15. Hernandez M. A. A uniparametric family of iterative processes for solving nondifferentiable equations / M. A. Hernandez, M. J. Rubio // J. Math. Anal. Appl. – 2002. – Vol. 275. – P. 821-834.
16. Hernandez M. A. The Secant method for nondifferentiable operators / M. A. Hernandez, M. J. Rubio // Appl. Math. Lett. – 2002. – Vol. 15. – P. 395-399.
17. Kantorovich L. V. Functional Analysis / L. V. Kantorovich, G. P. Akilov. – Oxford: Pergamon Press, 1982.
18. Kurchatov V. A. On a method of linear interpolation for the solution of functional equations / V. A. Kurchatov // Dokl. Akad. Nauk SSSR. – 1971. – Vol. 198 (3). – P. 524-526. (in Russian); translation in Soviet Math. Dokl. – 1971. – Vol. 12. – P. 835-838.
19. Ortega J. M. Iterative Solution of Nonlinear Equations in Several Variables / J. M. Ortega, W. C. Rheinboldt. – New-York: Academic Press, 1970.
20. Potra F. A. On an iterative algorithm of order 1.839... for solving nonlinear operator equations / F. A. Potra // Numer. Funct. Anal. Optim. – 1985. – Vol. 7 (1). – P. 75-106.
21. Ren H. New sufficient convergence conditions of the Secant method nondifferentiable operators / H. Ren // Appl. Math. Comput. – 2006. – Vol. 182. – P. 1255-1259.
22. Ren H. A new semilocal convergence theorem with nondifferentiable operators / H. Ren, I. K. Argyros // J. Appl. Math. Comput. – 2010. – Vol. 34. – P. 39-46.
23. Shakhno S. Convergence of combined Newton–Secant method and uniqueness of the solution of nonlinear equations / S. Shakhno // Visnyk Ternopil National Technical University. – 2013. – Vol. 69. – P. 242-252. (in Ukrainian).
24. Shakhno S. M. Convergence of the two-step combined method and uniqueness of the solution of nonlinear operator equations / S. M. Shakhno // J. Comput. Appl. Math. – 2014. – Vol. 261. – P. 378-386.

25. Shakhno S. M. Analysis of convergence of a combined method for the solution of nonlinear equations / S. M. Shakhno, I. V. Melnyk, H. P. Yarmola // Journal of Mathematical Sciences. – 2014. – Vol. 201. – P. 32-43.
26. Shakhno S. M. Method of linear interpolation of Kurchatov under generalized Lipschitz conditions for divided differences of first and second order / S. M. Shakhno // Visnyk Lviv. Univ. Ser. Mech. Math. – 2012. – Vol. 77. – P. 235-242. (in Ukrainian).
27. Shakhno S. M. On the Secant method under generalized Lipschitz conditions for the divided difference operator / S. M. Shakhno // PAMM – Proc. Appl. Math. Mech. – 2007. – Vol. 7. – P. 2060083-2060084.
28. Shakhno S. M. Twopoint method for solving nonlinear equation with nondifferentiable operator / S. M. Shakhno, H. P. Yarmola // Mat. Stud. – 2011.v– Vol. 36. – P. 213-220. (in Ukrainian).
29. Shakhno S. M. On a Kurchatov's method of linear interpolation for solving nonlinear equations / S. M. Shakhno // PAMM – Proc. Appl. Math. Mech. – 2004. – Vol. 4. – P. 650-651.
30. Traub J. F. Iterative Methods for the Solution of Equations / J. F. Traub. – Englewood Cliffs, NY: Prentice-Hall, Inc., 1964.
31. Traub J. F. Convergence and complexity of Newton iteration for operator equations / J. F. Traub, H. Woźniakowski // J. Assoc. Comput. Mach. – 1979. – Vol. 26. – P. 250-258.
32. Wang X. H. Convergence of Newton's method and inverse function theorem in Banach space / X. H. Wang // Math. Comp. – 1999. – Vol. 68. – P. 169-186.
33. Wang X. H. Convergence of Newton's method and uniqueness of the solution of equations in Banach space / X. H. Wang // IMA Journal of Numerical Analysis. – 2000. – Vol. 20. – P. 123-134.
34. Wang X. H. Local and global behavior for algorithms of solving equations / X. H. Wang, C. Li // Chinese Sci. Bull. – 2001. – Vol. 46. – P. 444-451.
35. Wang X. H. Convergence of Newton's method and uniqueness of the solution of equations in Banach space II / X. H. Wang, C. Li // Acta Math. Sin. (Engl. Ser.). – 2003. – Vol. 19. – P. 405-412.
36. Zabrejko P. P. The majorant method in the theory of Newton–Kantorovich approximations and the Pták error estimates / P. P. Zabrejko, D. F. Nguen // Numer. Funct. Anal. Optim. – 1987. – Vol. 9. – P. 671-684.

S. M. Shakhno,
Ivan Franko National University of Lviv,
1, Universytets'ka Str., Lviv, 79000, Ukraine

Received 08.07.2015

Журнал обчислювальної та прикладної математики | **2015** **№ 2 (119)** | Journal of Computational & Applied Mathematics

UDC 519.642

ABOUT MINIMAL INFORMATIONAL EFFORTS BY SOLVING EXPONENTIALLY ILL-POSED PROBLEMS

S. G. SOLODKY, E. V. SEMENOVA

Резюме. Розглядаються питання інформаційної складності для експоненціально некоректних задач. Дослідження виконані для інтегральних рівнянь Фредгольма першого роду з оператором скінченної гладкості. Запропоновані проекційні схеми дозволяють досягти оптимальний порядок точності для апостеріорного вибору параметра регуляризації за принципом рівноваги. Крім того такий підхід зберігає мінімальний обсяг інформаційних затрат.

Abstract. The issue of informational complexity for exponentially ill-posed problems is considered. The investigation is performed for Fredholm integral equations of the first kind with finite-smoothness operators. The proposed projection method allows to achieve optimal order accuracy for a posteriori selection of regularization parameter by balancing principle. Moreover such approach saves minimal volume of informational efforts.

1. Introduction

Nowadays for numerical method one of the most important issues is reduction of informational and computational efforts while saving approximation accuracy. These questions are studied in the framework of Informational Based Complexity Theory founded by J. Traub and H. Wozniakowski (see e.g. [18], [19]). The basic object of this theory is the information complexity, i.e. minimal amount of discrete information required to solve the problem with given accuracy. It was found that such amount depends on the smoothness properties of the problem. Particularly, for ill-posed problems presented by the first-kind operator equations $Ax = f$ the relation between smoothness of operator A and solution x is of primary importance. In the case of moderately ill-posed problems, when A and x are related by means of power function (i.e. A and x belong to the same smoothness scale), different efficient numeric approaches were proposed in [10], [12], [13], [14]. Owing to previous papers the exact order estimates of informational complexity for wide classes of moderately ill-posed problems (see, for example, [8]) were obtained. At the same time, much attention is paid to severely ill-posed problems where the solution has essentially worse smoothness in comparison with that of operator. Usually, in these cases A and x are related by means of logarithmic function but the corresponding equations are called exponentially ill-posed problems. For the

Key words. Severely ill-posed problems, minimal radius of Galerkin information, balancing principle.

first time severely ill-posed problems were considered by B.A. Mair [4]. Afterwards, these investigations were continued by T.Hohage [3], M.Y. Kokurin and A.B. Bakushinski [2], S.V. Pereverzev and E. Schock [17] and also in [15], [16].

It should be noted that for a long time the issue of improving effectiveness of numerical solving severely ill-posed problems (in sense of IBC theory) was not considered due to its complicatedness. The first step was done in [6], where the standard Galerkin discretization scheme was used to construct projective methods for solving different classes of problems including severely ill-posed ones. However, it was found that this approach does not provide minimal amount of computational efforts. Further investigations (see [16]) showed that amount of discrete information can be reduced in comparison with [6] for exponentially ill-posed integral equations with finite-smoothness kernels. It was done in [16] due to a modification of Galerkin scheme. In the case of a priori choice of regularization parameter it allowed not only to improve results of [6], but also provided minimal order of information efforts for mentioned Fredholm equations. The present paper is devoted to numerical solving exponentially ill-posed problems as in [16] for the case of a posteriori choice of regularization parameter. It will be shown that the absence of exact information about smoothness of solution does not influence informational complexity of problems under consideration.

2. Statement of the problem

Consider an integral equation of the first kind

$$Ax = f, \tag{1}$$

where $Ax(t) = \int_0^1 a(t,\tau)x(\tau)d\tau, \quad t \in [0,1]$, is acting continuously in $L_2 = L_2(0,1)$. Suppose that Range(A) is not closed in L_2 and $f \in$ Range(A).

Assume that instead of f we are given only $f_\delta \in L_2$ such that $\|f - f_\delta\| \le \delta$. Since, solution of problem (1) in general is not unique, we take solution of (1) with minimal norm in L_2 as element for approximation and denote it as $x^\dagger$.

Usually we call the equation (1) as severely ill-posed problem if its solution has essentially worse smoothness than that of elements from Range(A). As a rule in such case the solution $x^\dagger$ is said to satisfy the source conditions of logarithmic type and the corresponding equation (1) is called an exponentially ill-posed problem. To describe the smoothness property of solution we consider the set of smooth functions $M_p(A)$, which has the form

$$M_p(A) := \{u : u = \ln^{-p}(A^*A)^{-1}v, \ \|v\| \le \rho\}, \tag{2}$$

where $\rho, p > 0$ are some positive parameters and A^* is an adjoint operator to A. The exact information about smoothness, namely the value of p, is usually not available by practical experiment. So it should be assumed that the minimal-norm solution $x^\dagger$ belongs to the set

$$M(A) := \cup_{p\in(0,p_1]} M_p(A), \tag{3}$$

where $p_1 < \infty$ is an upper bound for possible values of p.

For constructing an effective numerical method for solving (1) we also need to describe smoothness properties of A. To this end let consider some orthonormal

basis $\{e_i(t)\}_1^\infty$ in L_2 and denote by P_m orthogonal projection onto linear span of elements $\{e_i(t)\}_1^m$ such that

$$P_m u(t) = \sum_{i=1}^{m} (u, e_i) e_i(t).$$

Further we introduce the class of operators

$$\mathcal{H}_\gamma^r = \left\{ A : \|A\| \le \gamma_0, \sum_{n+m=1}^{\infty} \hat{a}_{n,m}^2 (\underline{n} \cdot \underline{m})^{2r} \le \gamma_1^2 \right\},$$

where $r > 0$, $\hat{a}_{n,m}^2 = \int_0^1 \int_0^1 e_n(t) e_m(\tau) a(t,\tau) d\tau dt$, $\gamma_0 \le e^{-1}, \gamma = (\gamma_0, \gamma_1), \underline{n} = 1$ if $n = 0$ and $\underline{n} = n$ otherwise. As an example of operator from the class mentioned one can present integral operator A' that has the same structure as (1) with kernel $a'(t,\tau)$ that has mixed partial derivatives up to order r by each variables and for $i, j = 0, 1, \ldots, r$ it holds true that

$$\int_0^1 \int_0^1 \left[\frac{\partial^{i+j} a'(t;\tau)}{\partial t^i \partial \tau^j} \right]^2 dt d\tau < \infty.$$

It is known [7], that there is such set $\gamma = (\gamma_0; \gamma_1)$ that $A' \in \mathcal{H}_\gamma^r$. Further we assume that $A \in \mathcal{H}_\gamma^r$ for some values of γ with $\gamma_0 \le e^{-1}$.

Every projection scheme for discretization of equation (1) with perturbed right-hand side can be associated with a set of following functionals

$$(Ae_j, e_i), \quad (i,j) \in \Omega, \tag{4}$$

$$(f_\delta, e_k), \quad k \in \omega, \quad \omega = \{i : (i,j) \in \Omega\}, \tag{5}$$

where Ω is a bounded domain in the coordinate plane. The inner products (4) and (5) are called the Galerkin functionals about equation (1). We denote as $\mathrm{Card}(\Omega)$ the total amount of indexes for (4). Note that in the case of the Fredholm integral operator A the Galerkin functionals (4) and (5) become the Fourier coefficients by basis $\{e_i(t)\}_{i=1}^\infty$ for the kernel and right-hand side correspondingly. In the framework of this paper it is assumed that discrete information about equation (1) is given in the view of sets (4) and (5). Thus the projection methods for solving (1) are more suitable and will be investigated further. The first projection methods for ill-posed problems were proposed in [12] where rectangle $Q_{n,m} = [1,n] \times [1,m]$ was considered as domain Ω. Further this approach was improved by [11] due to the reduction of discretization domain $Q_{n,m}$ (it was replaced by so-called hyperbolic cross) with saving necessary accuracy of approximation. This idea will be used further for constructing an economical projection scheme (see section 3).

Further we call any mapping $\mathcal{P} = \mathcal{P}(\Omega) : L_2 \to L_2$ as projection method that by means of the set of Galerkin functional (4) gives an element $\mathcal{P}(A_\Omega) f_\delta \in L_2$. This elements can be interpreted as approximative solution of (1). In general such mapping can be nonlinear and discontinuous. Let define the error of projection method $\mathcal{P}(\Omega)$ for solving (1) with $A \in \mathcal{H}_\gamma^r$ and $x^\dagger \in M(A)$ in the

standard way

$$e_\delta\left(\mathcal{H}_\gamma^r, M(A), \mathcal{P}(\Omega)\right) = \sup_{A\in\mathcal{H}_\gamma^r}\ \sup_{x^\dagger\in M(A)}\ \sup_{f_\delta:\|f-f_\delta\|\le\delta} \|x^\dagger - \mathcal{P}(A_\Omega)f_\delta\|.$$

The minimal radius of Galerkin information we set as

$$R_{N,\delta}\left(\mathcal{H}_\gamma^r, M(A)\right) = \inf_{\substack{\Omega,\\ \mathrm{Card}(\Omega)\le N}}\ \inf_{\mathcal{P}(\Omega)} e_\delta\left(\mathcal{H}_\gamma^r, M(A), \mathcal{P}(\Omega)\right),$$

where N is maximal amount of discrete information (4).

The value $R_{N,\delta}\left(\mathcal{H}_\gamma^r, M(A)\right)$ is very important one and describes the minimal possible error (among the whole projection methods) on all classes of equations under consideration with using not more than N Galerkin functionals. At first the order bounds for minimal radius of Galerkin information for ill-posed problems with Holder-type smooth solutions were found by S.V. Pereverzyev and S.G. Solodky in [8]. Further for different classes of ill-posed problems the similar bounds were established in [16], [10] and others. Among mentioned papers we emphasize [16] where the minimal radius of Galerkin information was found for solving severely ill-posed problems (1) with operators $A \in \mathcal{H}_\gamma^r$ and smooth solutions from (2). In other words, in [16] only a priori case for choosing regularization parameter was considered. In the present paper we extend the set of possible solutions up to (3). Thus, we need to introduce a posteriori way for selecting regularization parameter and correct rule for discretization. Besides we set the goal to save both the order for minimal radius of Galerkin information and the accuracy estimation of the projection methods as it is in [16].

3. Method for solving

A modified projection scheme will be applied for economical discretization of operator A. The point of such scheme is to take as discretization domain Ω the hyperbolic cross of the form

$$\Gamma_{b,n} = \{1\}\times[1;2^{bn}] \cup_n^{k=1} (2^{k-1};2^k]\times[1;2^{bn-k}] \subset [1;2^n]\times[1;2^{bn}],$$

where $1 < b \le 2$, $n \in \mathbb{N}$. For simplicity of our computations we consider bn as the integer number. Then by approximative operator to A we understand the following finite-dimensional mapping

$$A_n = P_1 A P_{2^{bn}} + \sum_{k=1}^{n} (P_{2^k} - P_{2^{k-1}}) A P_{2^{bn-k}}. \tag{6}$$

Denote by N the total amount of integer pairs $(i,j) \in \Gamma_{b,n}$. It is known (see [16]) that $N := \mathrm{Card}(\Gamma_{b,n}) = c' 2^{bn} n$ for $1/2 \le c' \le 3/2$. The approximation properties of (6) for the operator class $\mathcal{H}_\gamma^r$ were investigated in [16] and we rewrite them below. So, for any $A \in \mathcal{H}_\gamma^r$ it holds true

$$\|A_n^* A_n - A^* A\| \le C_1 2^{-brn}, \tag{7}$$

$$\|(P_{2^n}A - A_n)\ln^{-p}(A^*A)^{-1}v\| \le \frac{C_2 2^{-brn}}{(brn\ln 2)^{p-1}}, \tag{8}$$

where

$$C_1 = \gamma_1 \max\{\gamma_1, \gamma_0\}\Big[3 + \frac{2^{2r+1}}{2^r - 1}\Big], \quad C_2 = \gamma_1\rho(\ln 2)^{-1}\frac{2^r}{r}\beta(p),$$

$$\beta(p) = \frac{1}{p-1}\left(\frac{(b-1)^{1-p}}{b^{1-p}} - 1\right),$$

for $p \neq 1$, and $\beta(1) = \ln\frac{b}{b-1}$.

Because the problem under consideration is ill-posed we need some regularization method to guarantee stability of approximations. In the framework of the paper we stabilize equation (1) following [1]. So, we construct an inverse operator to (1) by means of so-called generating function $g(\lambda)$. The function $g_\alpha(\lambda)$ is Borel measurable on the interval $[0, \gamma_0^2]$ and the following conditions are satisfied

$$\sup_{0<\lambda\le\gamma_0^2} \sqrt{\lambda}\ |g_\alpha(\lambda)| \le \tfrac{\chi_*}{\sqrt{\alpha}}, \tag{9}$$

$$\sup_{0<\lambda\le\gamma_0^2} |1 - \lambda g_\alpha(\lambda)|\ln^{-p}\lambda^{-1} \le \chi\ln^{-p}\tfrac{1}{\alpha}, \quad 0 < p < p_1, \tag{10}$$

where χ, χ_* are some positive constants independent of α. Then as the approximate solution we take

$$x^\delta_{\alpha,n} = g_\alpha(A_n^*A_n)A_n^*P_{2^n}f_\delta. \tag{11}$$

There are many well-known regularization methods satisfying (9). In particular, we can mention Tikhonov's method (with $g_\alpha(\lambda) = (\alpha + \lambda)^{-1}$), Landweber's method (with $g_\alpha(\lambda) = \lambda^{-1}[1 - (1 - \mu\lambda)^{1/\alpha}]$, $0 < \mu < 2$), and Showalter's method (with $g_\alpha(\lambda) = \lambda^{-1}(1 - \exp(-\lambda/\alpha))$).

In the paper [16] the error bound for (11) was found. For completeness we rewrite the stretch of proof.

Theorem 1 ([16]). *Let approximate solution has the form (11). Then on the class of equations (1) with $A \in \mathcal{H}^r_\gamma, x^\dagger \in M_p(A)$ for any $p > 0$ the following holds true*

$$\|x^\dagger - x^n_{\alpha,\delta}\| \le \tag{12}$$

$$\le \chi\rho\ln^{-p}\tfrac{1}{\alpha} + \tfrac{\chi_*}{\sqrt{\alpha}}\left[\delta + \|(P_{2^n}A - A_n)\ln^{-p}(A^*A)^{-1}v\|\right] + \tag{13}$$

$$+\chi\rho C_3\ln^{-p}\|A^*A - A_n^*A_n\|^{-1}, \tag{14}$$

where $C_3 = \begin{cases} 1, & 0 < p \le 1 \\ 1 + 4(5p)^p, & p > 1 \end{cases}$.

Proof. The error for (11) can be divided onto two terms

$$x^\dagger - x^n_{\alpha,\delta} := x^\dagger - g_\alpha(A_n^*A_n)A_n^*P_{2^n}f_\delta =$$

$$= (x^\dagger - g_\alpha(A_n^*A_n)A_n^*P_{2^n}f) + g_\alpha(A_n^*A_n)A_n^*P_{2^n}(f - f_\delta). \tag{15}$$

Owing to (9) we estimate the second term as following

$$\|g_\alpha(A_n^*A_n)A_n^*P_{2^n}(f-f_\delta)\| \le \frac{\chi_*\delta}{\sqrt{\alpha}}.$$

The first term we rewrite as

$$\begin{aligned} x^\dagger & -g_\alpha(A_n^*A_n)A_n^*P_{2^n}Ax^\dagger = \\ &= x^\dagger - g_\alpha(A_n^*A_n)A_n^*A_nx + g_\alpha(A_n^*A_n)A_n^*(A_n - P_{2^n}A)x^\dagger = \\ &= [\ln^{-p}(A_n^*A_n)^{-1}v - g_\alpha(A_n^*A_n)A_n^*A_n\ln^{-p}(A_n^*A_n)^{-1}v] + \\ &+(I - g_\alpha(A_n^*A_n)A_n^*A_n)(\ln^{-p}(A^*A)^{-1}v - \ln^{-p}(A_n^*A_n)^{-1}v) + \\ &+g_\alpha(A_n^*A_n)A_n^*(A_n - P_{2^n}A)x^\dagger. \end{aligned} \tag{16}$$

Then by (9) we immediately get

$$\begin{aligned} \|x^\dagger & -g_\alpha(A_n^*A_n)A_n^*P_{2^n}f\| \le \\ &\le \chi\rho\ln^{-p}\tfrac{1}{\alpha} + \tfrac{\chi_*}{\sqrt{\alpha}}\|(A_n - P_{2^n}A)x^\dagger\| + \\ &+\|(I - g_\alpha(A_n^*A_n)A_n^*A_n)(\ln^{-p}(A^*A)^{-1}v - \ln^{-p}(A_n^*A_n)^{-1}v)\| \le \\ &\le \chi\rho\ln^{-p}\tfrac{1}{\alpha} + \tfrac{\chi_*}{\sqrt{\alpha}}\|(P_{2^n}A - A_n)x^\dagger\| + \\ &+\chi\|\ln^{-p}(A^*A)^{-1}v - \ln^{-p}(A_n^*A_n)^{-1}v\|. \end{aligned}$$

Using the following relation (see [5, Theorem 4])

$$\left|\ln^{-p}\frac{1}{s} - \ln^{-p}\frac{1}{t}\right| \le C_3\ln^{-p}|s-t|^{-1},$$

where $|s-t| < e^{-1}$ for $s,t \in (0;e^{-1}]$, we have

$$\begin{aligned} \|x^\dagger & -x_{\alpha,\delta}^n\| \le \\ &\le \chi\rho\ln^{-p}\tfrac{1}{\alpha} + \tfrac{\chi_*}{\sqrt{\alpha}}\left[\delta + \|(P_{2^n}A - A_n)\ln^{-p}(A^*A)^{-1}v\|\right] + \\ &+\chi C_3\rho\ln^{-p}\|A^*A - A_n^*A_n\|^{-1}, \end{aligned}$$

that has to be proved. □

Remark 9. *Let consider the function $\beta(p)$ which is included in the bound (8). The analysis of behavior of $\beta(p)$ shows that it is continuous monotonically increasing function. Thus we have that for all $0 < p \le p_1$ the following inequality holds true*

$$\beta(p) \le \beta(p_1) = \begin{cases} \frac{1}{p_1-1}\left(\frac{(b-1)^{1-p_1}}{b^{1-p_1}} - 1\right), & p_1 \ne 1, \\ \ln\frac{b}{b-1}, & p_1 = 1. \end{cases}$$

To minimize the error bound (12) we take discretization parameter n according to the rule

$$(br\ln 2)n2^{-brn} = \delta. \tag{17}$$

The equality means that as discretization value n we take the number which is rounded up to solution of (17). Taking into account (17) and remark 9 the estimations (7) and (8) can be rewritten in the following way

$$\|A_n^*A_n - A^*A\| \le C_1\delta, \tag{18}$$

$$\|(P_{2^n}A - A_n)\ln^{-p}(A^*A)^{-1}v\| \le C_4\delta, \tag{19}$$

where $C_4 = \gamma_1\rho(\ln 2)^{-1}\frac{2^r}{r}\beta(p_1)$.

Due to (18) and (19) the error bound (12) can be represented as follows

$$\|x^\dagger - x^n_{\alpha,\delta}\| \le \chi\rho\ln^{-p}\frac{1}{\alpha} + \chi_*(1+C_4)\frac{\delta}{\sqrt{\alpha}} + \chi C_3\rho\ln^{-p}(C_1\delta)^{-1}. \tag{20}$$

Obviously, that for $\alpha_0 = \ln(\delta^{-1})(C_1\delta)^2$ we have

$$\ln^{-p}(C_1\delta)^{-1} = 2^p\ln^{-p}\left(\ln(\delta^{-1})(\ln(\delta^{-1})C_1^2\delta^2)^{-1}\right) = 2^p\ln^{-p}\left(\ln(\delta^{-1})(\alpha_0)^{-1}\right).$$

In this way for all $\alpha \ge \alpha_0$ it holds true that

$$\ln^{-p}\alpha^{-1} \ge \ln^{-p}\alpha_0^{-1} > \frac{1}{2^p}\ln^{-p}(C_1\delta)^{-1}.$$

Let denote by $\eta_1(\alpha) = C_5\ln^{-p}\frac{1}{\alpha}$ and $\eta_2(\alpha) = C_6\frac{\delta}{\sqrt{\alpha}}$, where $C_5 = \chi\rho + \chi C_3\rho 2^p$ and $C_6 = \chi_*(1+C_4)$. Thus error bound (20) can be rewritten as follows

$$\|x^\dagger - x^n_{\alpha,\delta}\| \le \eta_1(\alpha) + \eta_2(\alpha), \tag{21}$$

where the functions $\eta_1(\alpha)$ and $\eta_2(\alpha)$ for $\alpha \to \infty$ are monotone increasing and decreasing convex functions respectively.

4. A posteriori selection of regularization parameter

Fix some real number $q > 1$ and define by D_M the set of possible values for the parameter α:

$$D_M = \{\alpha_i = \alpha_0(q^2)^i, i = 1, 2, ..., M\},$$

where $\alpha_0 = \ln(\delta^{-1})(C_1\delta)^2$, $M = \left[\frac{\log\alpha_0^{-1}}{2\log q}\right]$. Then according to the balancing principle (see, for example, [9]) selection of index i_+ for parameter α is realized by the rule

$$i_+ = \max\{i : \alpha_i \in D_M^+\}, \tag{22}$$

where

$$D_M^+ = \{\alpha_i \in D_M : \|x^\delta_{\alpha_i,n} - x^\delta_{\alpha_j,n}\| \le 4\eta_2(\alpha_j), \quad j = 1, ..., i\}.$$

Further we introduce the auxiliary values

$$\alpha_* := \max\{\alpha_i \in D_M : \eta_1(\alpha_i) \le \eta_2(\alpha_i)\},$$

$$\hat{\alpha} = \{\alpha_i \in D_M : \quad \eta_1(\alpha_i) = \eta_2(\alpha_i)\}.$$

Theorem 2. *Let $A \in \mathcal{H}_r^\gamma$ and $x^\dagger \in M(A)$. Then for the projection method (11), (17), (22) the following error bound*

$$\|x^\dagger - x^\delta_{\alpha_+,n}\| \le 6q\eta_1(\hat{\alpha}) \tag{23}$$

takes place.

Proof. Let check that $\alpha_* \le \alpha_+$. Due to (21) it holds true that for all $\alpha \le \alpha_*$

$$\|x^\delta_{\alpha,n} - x^\delta_{\alpha_*,n}\| \le \|x^\dagger - x^\delta_{\alpha,n}\| + \|x^\dagger - x^\delta_{\alpha_*,n}\| \le \eta_1(\alpha) + \eta_2(\alpha) + \eta_1(\alpha_*) + \eta_2(\alpha_*) \le$$
$$\le 2\eta_2(\alpha) + 2\eta_2(\alpha_*) \le 4\eta_2(\alpha).$$

Consequently $\alpha_* \in D^+_M$ and $\alpha_* \le \alpha_+$.

Taking into account definitions of α_* and α_+, from (22) and (21) we have

$$\|x^\dagger - x^\delta_{\alpha_+,n}\| \le \|x^\dagger - x^\delta_{\alpha_*,n}\| + \|x^\delta_{\alpha_*,n} - x^\delta_{\alpha_+,n}\| \le 6\eta_2(\alpha_*).$$

It is evident that $\alpha_* \le \hat\alpha \le q^2\alpha_*$ then we find

$$\|x^\dagger - x^\delta_{\alpha_+,n}\| \le 6\eta_2(\alpha_*) = 6q\eta_2(\alpha_* q^2) \le 6q\eta_2(\hat\alpha) = 6q\eta_1(\hat\alpha),$$

which was to be proved. □

Theorem 3. *Let $A \in \mathcal{H}^\gamma_r$ and $x^\dagger \in M(A)$. Then error bound for the projection method (11), (17), (22) is the following*

$$\|x^\dagger - x^\delta_{\alpha_+,n}\| \le 6q\kappa_p \ln^{-p}\delta^{-1}, \tag{24}$$

where κ_p is some constant that does not depend on δ.

Proof. It is easy to find that

$$\hat\alpha \le \left(\frac{C_6}{C_5}\delta\right)^{\frac{2}{1+2p}},$$

then from (23) we have

$$\|x^\dagger - x^\delta_{\alpha_+,n}\| \le 6q\ln^{-p}\hat\alpha^{-1} \le 6q\ln^{-p}\left(\frac{c_6}{c_5}\delta\right)^{-\frac{2}{1+2p}} = 6q\kappa_p\ln^{-p}\delta^{-1}. \qquad \square$$

Remark 10. *It is well-known (see, for instance [17]) that for severely ill-posed problems any approximation method guaranteing accuracy $O(\ln^{-p}\delta^{-1})$ is optimal by the order on the whole set of solutions (3). Thus, theorem 3 shows that our method (11), (22), (17) saves optimal order of accuracy.*

5. Minimal radius of Galerkin information

Now we are ready to prove the upper bound for $R_{N,\delta}\left(\mathcal{H}^r_\gamma, M(A)\right)$.

Theorem 4. *Let $A \in \mathcal{H}^\gamma_r$ and $x^\dagger \in M(A)$. The parameters n and α for (11) are chosen according to (17) and (22) respectively. Then for sufficiently small δ the following inequality*

$$R_{N,\delta}\left(\mathcal{H}^r_\gamma, M(A)\right) \le c_p \ln^{-p} N^{2r}$$

holds true where $c_p = 6q\kappa_p\left(\frac{r(1-\mu)-\mu}{2r}\right)^{-p}$ and $\forall\mu : r(1-\mu)-\mu > 0$.

Proof. By virtue of (17) we have

$$br\ln 2(2^{bn}n)^{-r}n^{r+1} = \delta.$$

Using the relation $N = c'2^{bn}n$ we get

$$(c')^{-r}(br\ln 2)^{-1}N^r n^{-r-1} = \delta^{-1}. \tag{25}$$

By evident relation

$$n \le \frac{\ln N}{b \ln 2}$$

and (25) we have

$$\delta^{-1} = \frac{(c')^{-r} N^r}{(\ln 2) b r n^{r+1}} \ge \frac{(c')^{-r} N^r (b \ln 2)^{r+1}}{br \ln 2 (\ln N)^{r+1}} = \frac{(c')^{-r} N^r (b \ln 2)^r}{r (\ln N)^{r+1}}. \qquad (26)$$

Starting with some N it is holds true that $\ln N \le N^{\mu}$, then for any $\mu > 0$ we have

$$\frac{N^r (b \ln 2)^r}{(c')^r r (\ln N)^{r+1}} \ge \frac{N^r (c' b \ln 2)^r}{(c')^r r N^{\mu(r+1)}} = N^{r - r\mu - \mu} \frac{(c' b \ln 2)^r}{(c')^r r} =$$

$$= N^{r \frac{r(1-\mu)-\mu}{r}} \frac{(c' b \ln 2)^r}{(c')^r r}.$$

Taking into account the relation above from (26) we have

$$N^{r \frac{r(1-\mu)-\mu}{r}} \frac{(b \ln 2)^r}{(c')^r r} \le \delta^{-1}.$$

Without loss of generality we suppose that $\mu : r(1-\mu) - \mu > 0$. Then taking logarithm from inequality above one can find

$$\frac{r(1-\mu) - \mu}{2r} \ln N^{2r} \le \ln \delta^{-1}.$$

Hence, the error estimation (24) takes the form

$$\|x^{\dagger} - x^{\delta}_{\alpha_+, n}\| \le 6 q \kappa_p \ln^{-p} \delta^{-1} \le c_p \ln^{-p} N^{2r}.$$

Due to definition for $R_{N,\delta}\left(\mathcal{H}^r_{\gamma}, M(A)\right)$ we get

$$R_{N,\delta}\left(\mathcal{H}^r_{\gamma}, M(A)\right) \le c_p \ln^{-p} N^{2r},$$

which was to be proved. □

Theorem 5. *Let $A \in \mathcal{H}^{\gamma}_r$ and $x^{\dagger} \in M(A)$, then*

$$\frac{1}{2^{p+1}} \ln^{-p} N^{2r} \le R_{N,\delta}\left(\mathcal{H}^r_{\gamma}, M(A)\right) \le c_p \ln^{-p} N^{2r},$$

where $N \asymp \delta^{-\frac{1}{r}} \ln^{\frac{r+1}{r}} \delta^{-1}$. Indicated order $O(\ln^{-p} N^{2r})$ is achieved in the framework of projection method (11) , (17), (22).

Proof. It is known (see, for instance [16]) that for all $p > 0$ it fulfills $R_{N,\delta}\left(\mathcal{H}^r_{\gamma}, M_p(A)\right) \ge \tilde{c}_p \ln^{-p} N^{2r}$, where $\tilde{c}_p = 2^{-p-1}$. By virtue of definition for the sets $M_p(A)$ and $M(A)$ the following inequality holds true

$$R_{N,\delta}\left(\mathcal{H}^r_{\gamma}, M_p(A)\right) \le R_{N,\delta}\left(\mathcal{H}^r_{\gamma}, M(A)\right).$$

Due to Theorem 4 we immediately get statement of the theorem. □

Remark 11. *From Theorem 5 it follows that our approach gives optimal error bound with amount of discrete information in the form of Galerkin functionals (4).*

Remark 12. *Let consider the set* $M'(A) = \cup_{p\in[1,p_1]} M_p(A) \subset M(A)$. *If we assume that* $x^\dagger \in M'(A)$, *then the relation (17) should be replaced by the following*

$$(\ln 2) br 2^{-brn} = \delta, \tag{27}$$

with saving bounds (18) and (19). As we can see below, such selection of discretization parameter allows to reduce amount of discrete information by logarithmic multiplier.

Theorem 6. *Let* $A \in \mathcal{H}_r^\gamma$ *and* $x^\dagger \in M'(A)$. *The parameters* n *and* α *for (11) are chosen according to (27) and (22) respectively. Then for sufficiently small* δ *it holds true*

$$\frac{1}{2^{p+1}} \ln^{-p} N^{2r} \le R_{N,\delta}\left(\mathcal{H}_\gamma^r, M'(A)\right) \le c_p \ln^{-p} N^{2r},$$

where $N \asymp \delta^{-\frac{1}{r}} \ln \delta^{-1}$.

Proof. The proving of the theorem completely repeats as ones for Theorems 4 and 5. □

Remark 13. *Comparing Theorems 5 and 6 we can conclude that due to restriction of the set of possible solutions we obtain reduction of amount of discrete information by logarithmic multiplier (compare the values* N *in both theorems).*

Bibliography

1. Bakushinski A. B. A general method of constructing regularizing algorithms for a linear ill-posed equation in Hilbert Space / A. B. Bakushinski // USSR Comput. Math. Math. Phys. – 1967. – Vol. 7. – P. 672-677.
2. Bakushinski A. B. Iterative methods for solution of ill-posed operator equations with smooth operators / A. B. Bakushinski, M. Yu. Kokurin. – Moscow: Editorial URSS, 2002.
3. Hohage T. Regularization of exponentially ill-posed problems / T. Hohage // Numer. Funct. Anal. Optim. – 2000. – Vol. 21. – P. 439-464.
4. Mair B. A. Tikhonov regularization for finitely and infinitely smoothing operators / B. A. Mair // SIAM J. Math. Anal. – 1994. – Vol. 25. – P. 135-147.
5. Mathe P. Moduli of continuity for operator valued functions /P. Mathe, S.V. Pereverzev // Numer. Funct. Anal. Optim. – 2002. – Vol. 23. – P. 623-631.
6. Mathe P. Discretization strategy for ill-posed problems in variable Hilbert scales / P. Mathe, S. V. Pereverzev // Inverse Problems. – 2003. – Vol. 19. – P. 1263-1277.
7. Pereverzev S. V. Optimization of methods for approximate solution of operator equations / S. V. Pereverzev. – N-Y: Nova Science Publishers, 1996.
8. Pereverzev S. V. The minimal radius of Galerkin information for the Fredholm problems of first kind / S. V. Pereverzev, S. G. Solodky // J. Complexity. – 1996. – Vol. 12. – P. 401-415.
9. Pereverzev S. V. On the adaptive selection of the parameter in regularization of ill-posed problems / S. V. Pereverzev, E. Schock // SIAM J. Numer. Anal. – 2005. – Vol. 43. – P. 2060-2076.
10. Solodky S. G. The optimal approximations for solving linear ill-posed problems / S. G. Solodky // J. Complexity. – 2001. – Vol. 17. – P. 98-116.
11. Pereverzev S. V. Optimization of Projection Methods for Solving Ill-Posed Problems / S. V. Pereverzev // Computing. – 1999. – Vol. 55. – P. 113-124.
12. Plato R. On the Regularization of Projection Methods for solving Ill-posed Problems / R. Plato, G. Vainikko // Numer. Math. – 1990. – Vol. 57. – P. 63-79.

13. Solodky S. G. A generalized projection scheme for solving ill-posed problems / S. G. Solodky // J. Inverse Ill-Posed Probl. – 1999. – Vol. 7. – P. 185-200.
14. Solodky S. G. Bounds of information expenses in constructing projection methods for solving ill-posed problems / S. G. Solodky, E. V. Lebedeva // Comp. Method Appl. Math. – 2006. – Vol. 6. – P. 87-93.
15. Solodky S. G. On the a posteriori choice of regularization parametr in the solution of severelly ill-posed problems / S. G. Solodky, A. V. Grushevaya // Journal of Mathematical Science. – 2012. – Vol. 181. – P. 98-105.
16. Solodky S. G. The minimal radius of Galerkin information for severely ill-posed problems / S. G. Solodky, G. L. Myleiko // Journal of Inverse and Ill-Posed Problems. – 2014. – Vol. 5. – P. 735-757.
17. Tautenhahn U. Optimality for ill-posed problems under general source condition / U. Tautenhahn // Numer. Funct. Anal. and Optimiz. – 1998. – Vol. 19. – P. 377-398.
18. Traub J. F. Information-Based Complexity / J. F. Traub, G. Wasilkowski, H. Wozniakowski. – Boston: Academic Press, 1988.
19. Wozniakowski H. Information based complexity / H. Wozniakowski // Ann. Rev. Comput. Sci. – 1986. – Vol. 1. – P. 319-380.

S. G. Solodky, E. V. Semenova,
Institute of Mathematics, National Academy of Sciences,
3, Tereschenkivs'ka Str., Kyiv, 01601, Ukraine;

Received 10.06.2015

Журнал обчислювальної та прикладної математики | **2015** **№ 2 (119)** | Journal of Computational & Applied Mathematics

UDC 519.6

DIFFERENCE METHODS FOR SOLVING INVERSE EIGENVALUE PROBLEM

H. P. YARMOLA

Резюме. В роботі розглянуто обернену задачу на власні значення. Для чисельного розв'язування задачі застосовано метод хорд і метод лінійної інтерполяції (метод Курчатова). На відміну від методу Ньютона, ці ітераційні процеси використовують лише значення оператора з двох попередніх ітерацій та не потребують аналітично заданих похідних. Запропоновані методи застосовано для розв'язування обернених задач на власні значення різного типу. Розглянуті ітераційні процеси порівнюються з методом Ньютона за кількістю операцій, потрібних для обчислення першої поділеної різниці та похідної детермінанта.

Abstract. In this paper an inverse eigenvalue problem is considered. Secant method and method of the linear interpolation (Kurchatov's method) are applied for the numerical solution of this problem. Unlike Newton's method, these methods use only values of the operator at two previous iterations and do not require analytical derivatives. Proposed methods are used for solving different types of inverse eigenvalue problems. Considered iterative processes are compared with the Newton's method by the number of operations required to compute the first divided difference and derivative of determinant.

1. Introduction

An inverse eigenvalue problem (IEP) is to determine a matrix from a given spectral data. These problems arise in many applications, including control design, system identification, structure analysis and so on. There are special cases of inverse eigenvalue problems. Let's consider the following problems.

General IEP. Let $A_i = \{a^i_{jk}\}$ be complex $n \times n$ matrices for $i = \overline{0,n}$ and $\lambda = (\lambda_1, \ldots, \lambda_n)^T \in \mathbb{C}^n$. Find the vector $p = (p_1, p_2, \ldots, p_n) \in \mathbb{C}^n$, such that matrix

$$A(p) = A_0 + \sum_{i=1}^{n} p_i A_i$$

has eigenvalues $\lambda_1, \ldots, \lambda_n$. This problem involves classic partial cases of additive and multiplicative inverse eigenvalue problems.

Additive IEP. Let A be a given complex $n \times n$ matrix and $\lambda = (\lambda_1, \ldots, \lambda_n)^T \in \mathbb{C}^n$. Find the diagonal matrix $D = diag(p_1, p_2, \ldots, p_n)$, $p_i \in \mathbb{C}$, $i = \overline{1,n}$, such that matrix $A + D$ has eigenvalues $\lambda_1, \ldots, \lambda_n$.

Multiplicative IEP. Let A be a given complex $n \times n$ matrix and $\lambda = (\lambda_1, \ldots, \lambda_n)^T \in \mathbb{C}^n$. Find the diagonal matrix $D = diag(p_1, p_2, \ldots, p_n)$, $p_i \in \mathbb{C}$, $i = \overline{1,n}$, such that matrix AD has eigenvalues $\lambda_1, \ldots, \lambda_n$.

Key words. Inverse eigenvalue problem, Secant method, Kurchatov's method.

There are a large of literature on conditions for the solvability of inverse eigenvalue problems, different approaches and numerical methods for its solving [1–3,6,8]. We use approach, which calculates the zeros of the nonlinear function

$$F(p) = \begin{bmatrix} \det(A(p) - \lambda_1 I) \\ \vdots \\ \det(A(p) - \lambda_n I) \end{bmatrix}, \tag{1}$$

where $\lambda_1, \ldots, \lambda_n$ are given eigenvalues and $\lambda_i \neq \lambda_j$ for $i \neq j$.

Vector $p = \{p_1, p_2, \ldots, p_n\}^T \in \mathbb{C}^n$ is a solution of the inverse eigenvalue problem if and only if

$$F(p) = 0. \tag{2}$$

In papers [1,8] the Newton's method is used for solving systems of nonlinear equations (2) with $F(p)$ as (1). It is known that the application of Newton's method requires the calculation of the first derivative of determinant at every iteration. To calculate this derivative some authors use Trace-Theorem of Davidenko or LU decomposition of matrix [1,7,8].

In this work we apply difference methods for solving inverse eigenvalue problem, including Secant method and method of linear interpolation (Kurchatov's method), assuming the existence of a solution. These methods do not require analytical derivatives and can be applied to a wider range of problems.

2. Algorithms of difference methods

A well-known simple difference method for solving nonlinear equations is the Secant method

$$p^{(k+1)} = p^{(k)} - F(p^{(k-1)}; p^{(k)})^{-1} F(p^{(k)}) \tag{3}$$

with convergence order $\dfrac{1+\sqrt{5}}{2}$. An other method is the quadratically convergent Kurchatov's method

$$p^{(k+1)} = p^{(k)} - F(2p^{(k)} - p^{(k-1)}; p^{(k-1)})^{-1} F(p^{(k)}). \tag{4}$$

In formulas (3) and (4) $F(x;y)$ is a divided difference of the first order of F at the points x and y. Convergence analysis of difference methods (3) and (4) for solving nonlinear operator equations was conducted by the authors [4,5,9,10].

Let F be a nonlinear operator defined on a subset D of a linear space X with values in a linear space Y and let x, y be two points of D. A linear operator from X into Y, denoted as $F(x;y)$, which satisfies the condition

$$F(x;y)(x-y) = F(x) - F(y)$$

is called a divided difference of the first order of F at the points x and y. In the case of systems of nonlinear equations the divided difference $F(x;y)$ is $n \times n$ matrix. Its elements are calculated by the following formula:

$$F(x;y)_{i,j} = \frac{F_i(x_1, \ldots, x_j, y_{j+1}, \ldots, y_n) - F_i(x_1, \ldots, x_{j-1}, y_j, \ldots, y_n)}{x_j - y_j},$$
$$i, j = \overline{1,n}.$$

From (1) and the last formula we see that to calculate the elements of vector F and matrix of divided differences we need to calculate determinants of matrices. To calculate the determinant we apply the LU decomposition of matrix, as in [7,8]. Let $D(\lambda)$ be a matrix whose elements are functions of λ. Then for a fixed value $\lambda = \lambda_m$ we can calculate $D = LU$ or $PD = LU$, where P is a permutation matrix, $\det P = (-1)^q$, q is a number of permutations and

$$\det D = \det L \det U = \prod_{i=1}^{n} u_{ii} \tag{5}$$

or

$$\det D = \det P \det L \det U = (-1)^q \prod_{i=1}^{n} u_{ii}. \tag{6}$$

Algorithm of the Secant method for solving IEP.

1. Choose initial approximations $p^{(-1)}$ and $p^{(0)}$.
2. For $k = 0$ until convergence, do:
   (a) Compute LU decomposition of matrices $D_i = A(p^{(k)}) - \lambda_i I$, $i = \overline{1,n}$, $D_i^{'} = A(p') - \lambda_i I$, $D_i^{''} = A(p'') - \lambda_i I$, $(i,j = \overline{1,n})$, where
   $$p' = (p_1^{(k-1)}, \ldots, p_j^{(k-1)}, p_{j+1}^{(k)}, \ldots, p_n^{(k)}),$$
   $$p'' = (p_1^{(k-1)}, \ldots, p_{j-1}^{(k-1)}, p_j^{(k)}, \ldots, p_n^{(k)}).$$
   (b) Compute $F_i(p^{(k)}) = \det(D_i)$, $i = \overline{1,n}$ by formula (5) or (6) and form vector $F(p^{(k)})$.
   (c) Compute $F_i(p') = \det(D_i^{'})$, $F_i(p'') = \det(D_i^{''})$, $i,j = \overline{1,n}$ by formula (5) or (6) and form matrix $F(p^{(k-1)}; p^{(k)})$, where
   $$F(p^{(k-1)}; p^{(k)})_{i,j} = \frac{F_i(p') - F_i(p'')}{p_j^{(k-1)} - p_j^{(k)}}, (i,j = \overline{1,n}).$$
   (d) Compute $p^{(k+1)}$ by the formula (3).

Algorithm of the Kurchatov's method for solving IEP.

1. Choose initial approximations $p^{(-1)}$ and $p^{(0)}$.
2. For $k = 0$ until convergence, do:
   (a) Compute LU decomposition of matrices $D_i = A(p^{(k)}) - \lambda_i I$, $i = \overline{1,n}$, $D_i^{'} = A(p') - \lambda_i I$, $D_i^{''} = A(p'') - \lambda_i I$, $(i,j = \overline{1,n})$, where
   $$p' = (2p_1^{(k)} - p_1^{(k-1)}, \ldots, 2p_j^{(k)} - p_j^{(k-1)}, p_{j+1}^{(k-1)}, \ldots, p_n^{(k-1)}),$$
   $$p'' = (2p_1^{(k)} - p_1^{(k-1)}, \ldots, 2p_{j-1}^{(k)} - p_{j-1}^{(k-1)}, p_j^{(k-1)}, \ldots, p_n^{(k-1)}).$$
   (b) Compute $F_i(p^{(k)}) = \det(D_i)$, $i = \overline{1,n}$ by formula (5) or (6) and form vector $F(p^{(k)})$.
   (c) Compute $F_i(p') = \det(D_i^{'})$, $F_i(p'') = \det(D_i^{''})$, $i,j = \overline{1,n}$ by formula (5) or (6) and form matrix $F(2p^{(k)} - p^{(k-1)}; p^{(k-1)})$, where
   $$F(2p^{(k)} - p^{(k-1)}; p^{(k-1)})_{i,j} = \frac{F_i(p') - F_i(p'')}{2(p_j^{(k)} - p_j^{(k-1)})}, (i,j = \overline{1,n}).$$

(d) Compute $p^{(k+1)}$ by the formula (4).

Note, that matrices D_i, D_i', D_i'' can coincide with each other. In this case LU decomposition and determinant can be calculated only once and thus the amount of computation is reduced.

Next we consider the computational complexity of proposed algorithms. Let compute the amount of operations (multiplications and division) required to compute divided differences. It is known that to get LU decomposition of matrix and compute its determinant by formula (5) it is need $\frac{n^3+2n-3}{3}$ operations [7, 8]. In the same articles it is shown that to compute the first derivative of determinant it is required n^3+n^2-n operations.

To compute divided difference of determinant using LU decomposition it is required $\frac{2n^3+4n-3}{3}$ operations for Secant method and $\frac{2n^3+7n-3}{3}$ operations for Kurchatov's method.

From these assessments we conclude that the difference methods are more effective than Newton's method by the amount of calculations in one iteration. However, the number of iterations for difference methods usually is greater than for Newton's method, in particular for the Secant method.

3. Numerical experiments

In this section we present results of Secant and Kurchatov's methods and compare with results of Newton's method. We consider inverse eigenvalue problems with distinct eigenvalues. All vectors will be written as row-vectors. To apply the methods (3) and (4) we need to set the additional approximation $p^{(-1)}$. To get good starting values it was chosen in the following way: $p^{(-1)} = p^{(0)} + 10^{-4}$. The iterations of considered iterative processes were stopped when $\|p^{(k+1)} - p^{(k)}\|_\infty < \varepsilon$ or $\|F(p^{(k+1)})\|_\infty < \varepsilon$, $\varepsilon = 10^{-9}$.

Example 3.1 Consider the general inverse eigenvalue problem [1]. Let $n = 5$,

$$A_0 = \begin{pmatrix} 2 & -0.08 & 0 & 0 & 0 \\ -0.03 & 2 & -0.08 & 0 & 0 \\ 0 & -0.03 & 2 & -0.08 & 0 \\ 0 & 0 & -0.03 & 2 & -0.08 \\ 0 & 0 & 0 & -0.03 & 2 \end{pmatrix},$$

$$R = \sum_{i=1}^{n} r_i e_i^T = \begin{pmatrix} 1 & 0 & 0.01 & -0.02 & 0.03 \\ -0.03 & 1 & 0 & 0.01 & -0.02 \\ 0.02 & -0.03 & 1 & 0 & 0.01 \\ -0.01 & 0.02 & -0.03 & 1 & 0 \\ 0 & -0.01 & 0.02 & -0.03 & 1 \end{pmatrix}$$

and $A_i = r_i e_i^T$, $i = 1, \ldots, 5$, where e_i – i-th unit vector. The given eigenvalues are $\lambda = (\delta,\ 1-\delta,\ 2+\delta,\ 3-\delta,\ 4)$.

Let $\delta = 0$ and $p^{(0)} = (-2,\ -1,\ 0,\ 1,\ 2)$. Then Newton's method converge to a solution

$$p^* = (1.99279,\ 1.00257,\ 0.00237,\ -0.99786,\ -1.99987).$$

Using the same starting point $p^{(0)}$, we found a different solution

$$p^* = (-2.00240,\ -0.99800,\ 0.00236,\ 1.00271,\ 1.99533)$$

by Secant and Kurchatov's methods.

Let $\delta = 0.441$. Then Newton's method, methods (3) and (4) converge to a solution

$$p^* = (-1.56910,\ -1.43181,\ 0.49205,\ 0.51127,\ 1.99758)$$

with the starting point $p^{(0)} = (-2,\ -1,\ 0,\ 1,\ 2)$. The received results are displayed in the Table 1.

TABL. 1. The numerical results for example 3.1

	Iterations, k	$\|p^{(k)} - p^{(k-1)}\|_\infty$	$\|F(p^{(k)})\|_\infty$
Newton's method	10	5.73238×10^{-10}	8.07568×10^{-15}
Kurchatov's method	10	9.88872×10^{-11}	4.12121×10^{-15}
Secant method	14	2.31415×10^{-11}	8.07565×10^{-15}

Example 3.2 Consider an additive inverse eigenvalue problem with distinct eigenvalues [3]. Here n = 8,

$$A_0 = \begin{pmatrix} 0 & 4 & -1 & 1 & 1 & 5 & -1 & 1 \\ 4 & 0 & -1 & 2 & 1 & 4 & -1 & 2 \\ -1 & -1 & 0 & 3 & 1 & 3 & -1 & 3 \\ 1 & 2 & 3 & 0 & 1 & 2 & -1 & 4 \\ 1 & 1 & 1 & 1 & 0 & 1 & -1 & 5 \\ 5 & 4 & 3 & 2 & 1 & 0 & -1 & 6 \\ -1 & -1 & -1 & -1 & -1 & -1 & 0 & 7 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 0 \end{pmatrix},\ A_i = e_i e_i^T,\ i = \overline{1,8}.$$

The eigenvalues of the problem 3.2 are $\lambda^* = (10,\ 20,\ 30,\ 40,\ 50,\ 60,\ 70,\ 80)$.

TABL. 2. The numerical results for example 3.2

	Secant method	Kurchatov's method
k	$\|p^{(k)} - p^*\|_\infty$	$\|p^{(k)} - p^*\|_\infty$
0	8.68150	8.68150
1	2.31065	2.31079
2	1.10171	0.59989
3	0.23738	0.05708
4	0.02958	0.00171
5	0.00085	5.26419×10^{-6}
6	4.20674×10^{-6}	5.07569×10^{-10}
7	6.34913×10^{-10}	

Proposed methods converge to a solution

$$p^* = (11.907888,\ 19.705522,\ 30.545498,\ 40.062657,$$
$$51.587140,\ 64.702131,\ 70.170676,\ 71.318499)$$

with the starting point $p^{(0)} = (10, 20, 30, 40, 50, 60, 70, 80)$. The result was obtained in 8 (Secant method) and 7 (Kurchatov's method) iterations. The nature of the convergence of the considered numerical methods is shown in Table 2.

Applying difference methods (3) and (4) to this problem with the starting point $p^{(0)} = (10, 80, 70, 50, 60, 30, 20, 40)$ we find the following solution in 7 iterations:

$$p^* = (11.461354, 78.880829, 68.353400, 49.878330,$$
$$59.168918, 30.410470, 24.834324, 37.012374).$$

So, difference methods can be applied for solving inverse eigenvalue problems. Also these methods are simple in program implementation.

Bibliography

1. Biegler-Konig F. W. A Newton iteration process for inverse eigenvalue problems / F. W. Biegler-Konig // Numer. Math. – 1981. – Vol. 37, № 3. – P. 349-354.
2. Biegler-Konig F. W. Sufficient conditions for the solubility of inverse eigenvalue problems / F. W. Biegler-Konig // Linear Algebra Appl. – 1981. – Vol. 40. – P. 89-100.
3. Friedland S. The formulation and analysis of numerical methods for inverse eigenvalue problems / S. Friedland, J. Nocedal, M. L. Overton // J. Numer. Anal. – 1987. – Vol. 24, № 3. – P. 634-667.
4. Hernandez M. A. The Secant method and divided differences Hölder continuous / M. A. Hernandez, M. J. Rubio // Applied Mathematics and Computation. – 2001. – Vol. 124. – P. 139-149.
5. Kurchatov V. A. On a method of linear interpolation for the solution of functional equations / V. A. Kurchatov // Dokl. Akad. Nauk SSSR. – 1971. – Vol. 198, № 3. – P. 524-526. (in Russian).
6. Li L. Sufficient conditions for the solvability of algebraic inverse eigenvalue problems / L. Li // Linear Algebra Appl. – 1995. – Vol. 221. – P. 117-129.
7. Podlevskyi B. Calculating the exact derivatives of matrix determinant / B. Podlevskyi // Visnyk Lviv Univ. Ser. Appl. Math. Comp. Sci. – 2013. – № 20. – P. 42-48. (in Ukrainian).
8. Podlevskyi B. M. Newton's method for solving the inverse spectral problems / B. M. Podlevskyi, O. S. Yaroshko // J. Mathematical Sciences. – 2013. – Vol. 174, № 3. – P. 156-165.
9. Shakhno S. M. Application of nonlinear majorants for investigation of the Secant method for solving nonlinear equations / S. M. Shakhno // Matematychni Studii. – 2004. – Vol. 22, № 1. – P. 79-86. (in Ukrainian).
10. Shakhno S. M. On the difference method with quadratic convergence for solving nonlinear operator equations / S. M. Shakhno // Matematychni Studii. – 2006. – Vol. 26, № 1. – P. 105-110. (in Ukrainian)

H. P. Yarmola,
Ivan Franko National University of Lviv,
1, Universytets'ka Str., Lviv, 79000, Ukraine

Received 10.06.2015

Taras Shevchenko National University of Kyiv

Journal of Computational and Applied

2015 **M A T H E M A T I C S** 2 (119)

Founded in 1965

C O N T E N T S